Chemoinformatische Datenintegration zur Analyse und Vorhersage von Interaktionen von kleinen organischen Molekülen mit deren Biomolekülen

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von

Björn-Oliver Gohlke
aus Berlin

2016

Diese Arbeit wurde im Zeitraum von Juli 2012 bis Juni 2016 unter der Leitung von PD Dr. Robert Preißner an der Charité - Universitätsmedizin Berlin angefertigt.

1. Gutachter: PD Dr. Robert Preißner (Charité - Universitätsmedizin Berlin)

2. Gutachter: Prof. Dr. Udo Heinemann (Freie Universität Berlin)

Disputation am 11.01.2017

Inhaltsverzeichnis

Zusammenfassung

In den vergangenen Jahren ist die Zahl an frei verfügbaren Bioaktivitäts-Datenbanken stetig gestiegen. Bekannte Beispiele sind ChEMBL, PubChem BioAssay und BindingDB. Jedoch sind Datenqualität und -integrität in diesen Datenbanken nicht immer gewährleistet.
Da die Qualität jedoch ein entscheidendes Kriterium für die Implementierung neuer *in-silico*-Vorhersagealgorithmen ist, ist diese Dissertation diesem Thema gewidmet und zeigt auf, wie unter Verwendung von eigens erstellten spezifischen, integrativen Datenquellen die Vorhersagerate verschiedener *in-silico* Methoden verbessert werden kann. Hierfür wurden die Informationen aus den zuvor genannten Datenbanken integriert, überprüft und normalisiert. Sie wurden anschließend als Datengrundlage für verschiedene chemoinformatische Methoden verwendet, um das Verständnis der Interaktion kleiner organischer Moleküle mit deren Biomolekülen zu verbessern und grundlegende Bindungsmechanismen aufzuklären.
Um in diesem Kontext einen Mehrwert gegenüber den bereits bekannten Datenbanken zu generieren, wurde eine spezialisierte Text-Mining-Pipeline entwickelt, die es ermöglicht, Interaktionen zwischen kleinen organischen Molekülen und deren Biomolekülen zu identifizieren. Durch ein in die Pipeline eingebettetes Struktur-Synonym-Mapping, sowie einen manuellen Validierungsschritt wurde größtmöglicher Wert auf Datenintegrität gelegt. Dadurch konnte sichergestellt werden, dass der Datensatz frei von Redundanzen ist. Die anschließende Normalisierung der Interaktions-Daten erfolgte durch verschiedene Iterationsschritte, um die vorhandenen Strukturinformationen zu vereinheitlichen.
Basierend auf den so generierten Daten wurde im folgenden Schritt der Wirkmechanismus der kleinen organischen Strukturen identifiziert und zugewiesen. Dies ist unerlässlich, um zwischen Agonist und Antagonist unterscheiden zu können. In beiden Fällen wird das gleiche Biomolekül gebunden, jedoch in verschiedenen Bindungstaschen und mit unterschiedlichen biologischen Reaktionen. Daher ist diese Unterteilung für die Vorhersage der Interaktionen und den damit verbundenen Wirkmechanismus notwendig.
Um die Interaktionsdaten krankheitsspezifisch analysieren zu können, wurden diese Informationen auf Stoffwechselwege projiziert und interpretiert. Dadurch können in einem frühen Stadium unerwünschte Nebenwirkungen identifiziert werden. Diese Daten wurden im weiteren Verlauf der Dissertation zur Vorhersage von neuen Interaktionen bzw. zur Aufklärung von Nebenwirkungen bekannter Medikamente verwendet. Um die Vorhersagequalität zu verbessern, wurden verschiedene chemoinformatische Methoden miteinander kombiniert.
Mithilfe dieses integrativen Ansatzes ist es in dieser Arbeit gelungen, Nebenwirkungen durch

unerwünschte Interaktionen mit anderen Biomolekülen („Off-Targets“) aufzuklären bzw. neue Zielmoleküle zu identifizieren und deren genauen Bindungsmechanismus zu beschreiben.

Abstract

In the last years the number of publicly available bioactivity databases, such as ChEMBL, PubChem BioAssay and BindingDB, has raised awareness about the topics of data curation, quality and integrity. To increase the efficiency of drug development process, the vast information on chemical compounds presented in those databases need to be optimized.
However, the quality is a decisive criterion for the implementation of new *in-silico* algorithms. This thesis has devoted this issue and shows how various *in-silico* methods can be improved by using specific integrative data sources. Therefor information of the mentioned databases were integrated, validated and normalized to use them as data resource for various chemoinformatics methods. With this it is possible to improve the analysis of the interaction between small molecules with their biomolecules and elucidate their fundamental mechanism of actions.
In order to generate additional information to the databases, a specialized text mining pipeline has been developed. This offers the possibility to identify interactions between small molecules and their biomolecules. An embedded structure synonym mapping into the pipeline was programed. Further a manual validation step focusing on data integrity was placed. This is important to ensure that the database will be free of redundancies. The subsequent normalization of interaction data was carried out by several iterations to unify the existing structural information.
Based on the data generated by this pipeline, the underlying binding mechanism of the small structures has been identified and assigned. This is essential to distinguish between agonist and antagonist. In both cases, the same biomolecule is bound in various binding pockets and with different biological reactions. Therefore, this subdivision is necessary for the prediction of interaction and the associated mechanism of action.
In order to analyze the interaction data, this information has been projected to metabolic pathways. Thereby it is possible to interpret those interaction in a disease related content and undesirable side effects may be identified in an early stage. These data were used in following chapters of the thesis for predicting new interactions and the elucidation of side effects of known drugs. To improve the forecast quality, different chemoinformatics methods were combined.
This integrated approach was successfully used in this work for the identification of new drug targets and for the description of their exact binding mechanism. Finally it was possible to explain side effects due to unwanted interactions with other biomolecules („Off-Targets“).

Liste eigener Veröffentlichungen

Originalarbeiten (Peer-Reviewed); kumulativer Impact Factor (ResearchGate): 79,99

Publikation 1:

J. von Eichborn, M. Dunkel, B. O. Gohlke, S. C. Preissner, M. F. Hoffmann, J. M. Bauer, J. D. Armstrong, M. H. Schaefer, M. A. Andrade-Navarro, N. Le Novere, M. D. Croning, S. G. Grant, P. van Nierop, A. B. Smit, and R. Preissner.
SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations.
Nucleic Acids Res., 41(Database issue):D834–840, Jan 2013.
Eigener Anteil: Mitarbeit beim Erstellen des Datensatzes; Mitarbeit beim Erstellen der Weboberfläche; Erstellung der krankheitsspezifischen Netzwerke; Mitarbeit beim Verfassen der Publikation.

Publikation 2:

B. O. Gohlke, R. Preissner, and S. Preissner.
SuperPain–a resource on pain-relieving compounds targeting ion channels.
Nucleic Acids Res., 42(Database issue):D1107-1112, Jan 2014.
Eigener Anteil: Konzeption des Projekts; Durchführung und Analyse der *in-silico* Berechnungen. Dazu zählen Docking, Ähnlichkeitsanalyse sowie das Clustering der kleinen organischen Moleküle; Design und Programmierung der Website; Mitarbeit beim Verfassen der Publikation

Publikation 3:

S. Oertelt-Prigione, B. O. Gohlke, M. Dunkel, R. Preissner, and V. Regitz-Zagrosek.
GenderMedDB: an interactive database of sex and gender-specific medical literature.
Biol Sex Differ, 5:7, 2014.
Eigener Anteil: Mitarbeit bei der Durchführung des Text-Minings; Design und Programmierung der Website; Mitarbeit beim Verfassen der Publikation

Publikation 4:

J. Nickel*, B. O. Gohlke *, J. Erehman, P. Banerjee, W. W. Rong, A. Goede, M. Dunkel, and R. Preissner.
SuperPred: update on drug classification and target prediction.
Nucleic Acids Res., 42(Web Server issue):26-31, Jul 2014.
Eigener Anteil: Konzeption des Projekts; Design und Programmierung der Ähnlichkeitsanalyse; Design und Programmierung der Website; Verfassen der Publikation

Publikation 5:

P. Banerjee, J. Erehman, B. O. Gohlke, T. Wilhelm, R. Preissner, and M. Dunkel.
SuperNatural II–a database of natural products.
Nucleic Acids Res., 43(Database issue):D935-939, Jan 2015.
Eigener Anteil: Implementierung des Wirkmechanismus (MOA); Mitarbeit bei der Datenakquise; Analyse und Clustering der Naturstoffe; Mitarbeit beim Verfassen der Publikation

Publikation 6:

I. G. Metushi, A. Wriston, P. Banerjee, B. O. Gohlke, A. M. English, A. Lucas, C. Moore, J. Sidney, S. Buus, D. A. Ostrov, S. Mallal, E. Phillips, J. Shabanowitz, D. F. Hunt, R. Preissner, and B. Peters.
Acyclovir Has Low but Detectable Influence on HLA-B*57 : 01 Specificity without Inducing Hypersensitivity.
PLoS ONE, 10(5):e0124878, 2015.
Eigener Anteil: Konzeption der *in-silico* Pipeline; Durchführung und Analyse der *In-silico* Versuche; Mitarbeit beim Verfassen der Publikation

Publikation 7:

B. O. Gohlke, T. Overkamp, A. Richter, A. Richter, P. T. Daniel, B. Gillissen, and R. Preissner.
2D and 3D similarity landscape analysis identifies PARP as a novel off-target for the drug Vatalanib.
BMC Bioinformatics, 16:308, 2015.
Eigener Anteil: Konzeption des Projekts; Konzeption und Implementierung aller *in-silico* Experimente; Design und Programmierung der Ähnlichkeitsanalyse; Design und Programmierung der Website; Verfassen der Publikation

Publikation 8:

B. O. Gohlke, J. Nickel, R. Otto, M. Dunkel, and R. Preissner.
CancerResource-updated database of cancer-relevant proteins, mutations and interacting drugs.
Nucleic Acids Res., 44(D1):D932–937, Jan 2016.
Eigener Anteil: Konzeption des Projekts; Konzeption und Implementierung der *in-silico*

Experimente; Design und Programmierung der Ähnlichkeitsanalyse; Design und Programmierung der Website; Verfassen der Publikation

Publikation 9:

V. B. Siramshetty, J. Nickel, C. Omieczynski, B. O. Gohlke, M. N. Drwal, and R. Preissner.
WITHDRAWN-a resource for withdrawn and discontinued drugs.
Nucleic Acids Res., 44(D1):D1080–1086, Jan 2016.
Eigener Anteil: Mitarbeit bei der Konzeption des Projekts; Datenakquise; Alleinige Implementierung der Datenakquise für SNPs, sowie deren Analyse; Mitarbeit beim Verfassen der Publikation

Publikation 10:

S. Klaeger, B. O. Gohlke, J. Perrin, V. Gupta, S. Heinzlmeir, D. Helm, H. Qiao, G. Bergamini, H. Handa, M. M. Savitski, M. Bantscheff, G. Medard, R. Preissner, and B. Kuster.
Chemical Proteomics Reveals Ferrochelatase as a Common Off-target of Kinase Inhibitors.
ACS Chem. Biol., Feb 2016.
Eigener Anteil: Durchführung und Analyse der *in-silico* Experimente (Docking); Mitarbeit beim Verfassen der Publikation

Publikation 11:

M. Riehle, A. K. Buscher, B. O. Gohlke, M. Kassmann, M. Kolatsi-Joannou, J. H. Brasen, M. Nagel, J. U. Becker, P. Winyard, P. F. Hoyer, R. Preissner, D. Krautwurst, M. Gollasch, S. Weber, and C. Harteneck.
TRPC6 G757D Loss-of-Function Mutation Associates with FSGS.
J. Am. Soc. Nephrol., Feb 2016.
Eigener Anteil: Durchführung und Analyse der *in-silico* Experimente (Homologie-Modellierung, Erstellung der Mutagenesen); Mitarbeit beim Verfassen der Publikation

Kapitel 1

Einleitung

In diesem Abschnitt wird verdeutlicht, wie diese Arbeit in den Prozess des Wirkstoffentwurfs einzugliedern ist und wie sie in diesem Zusammenhang zu einem besseren Verständnis molekularer Wirkmechanismen von kleinen organischen Molekülen mit deren Biomolekülen beiträgt. Die Grundlagen der verwendeten Methoden werden genauer erörtert und in die übergeordnete Medikamentenentwicklungs-Pipeline eingeordnet.

1.1 Wirkstoffentwurf

Die Geschichte der Arznei- bzw. Wirkstoffentwicklung unterteilt sich grundlegend in vier Phasen: Volksmedizin, Tierversuche,molekulare und *in-vitro*-Testsysteme sowie die Strukturbiologie und theoretische Chemie. Die erste Phase beschreibt die Volksmedizin, welche sich mit Arzneistoffen aus der Natur, überwiegend pflanzenbasiert, befasst. Daran gliedert sich die organische Chemie an, wobei die systematische Suche nach synthetischen Arzneistoffen im Vordergrund steht. Da hierfür vorwiegend Tierexperimente durchgeführt werden, beschäftigen sich aktuelle Initiativen wie BB3R (Berlin Brandenburg Forschungsplatform; „reduce, refine, replace“) mit der Suche nach alternativen Lösungen. Ein Ziel ist es, Tierversuche durch molekulare und *in-vitro*-Testsysteme zu reduzieren und sie im Idealfall durch diese zu ersetzen. Informationen aus der Strukturbiologie und der theoretischen Chemie werden kombiniert, um die Entwicklung neuer Wirkstoffe zu optimieren.
Hauptanliegen des Wirkstoffentwurfs ist es, eine Substanz zu identifizieren, die nach Aufnahme im Organismus eine biochemische Wirkung hervorruft. Diese Substanz soll möglichst eine lindernde oder heilende Wirkung im Organismus bewirken. Dabei ist die gezielte Identifizierung von Wirkstoffen mit möglichst geringen Nebenwirkungen ein primäres Ziel.

1.2 Medikamentenentwicklungs-Pipeline

Die Entwicklung von neuen Medikamenten ist ein langwieriger und kostenintensiver Prozess, der sich grundsätzlich in präklinische und klinische Forschung unterteilen lässt [1]. Für die Initiation der Medikamentenentwicklungs-Pipeline wird im ersten Schritt eine Krankheit definiert, für die

ein neues Medikament entwickelt werden soll. Damit einhergehend erfolgt eine Datenakquise von bereits publiziertem Wissen. Hauptanliegen ist es, an der Krankheit beteiligte Zielmoleküle zu identifizieren, die als Angriffspunkte für neue, effektivere kleine organische Moleküle (Wirkstoffe) dienen könnten [2, 3].
In der präklinischen Forschung steht die Identifikation und Evaluierung eines neuen Wirkstoffs im Vordergrund. Dabei wird in einem mehrstufigen und iterativen Prozess eine Leitstruktur entwickelt, auf Basis derer weiterführende Experimente durchgeführt werden. Diese Leitstruktur wird anschließend als Wirkstoff-Kandidat weiter evaluiert und in Tiermodellen getestet.
Sofern die Wirksamkeit der getesteten Substanz nachgewiesen werden kann, wird diese in klinischen Studien auf Wirksamkeit und Unbedenklichkeit am Menschen getestet [4]. Wenn der Wirkstoff die Anforderungen der klinischen Phasen erfüllt, wird er in der Regel durch ein pharmazeutisches Unternehmen patentiert und vermarktet.
Eine schematische Darstellung der Medikamentenentwicklungs-Pipeline ist in Abbildung 1.1 dargestellt.

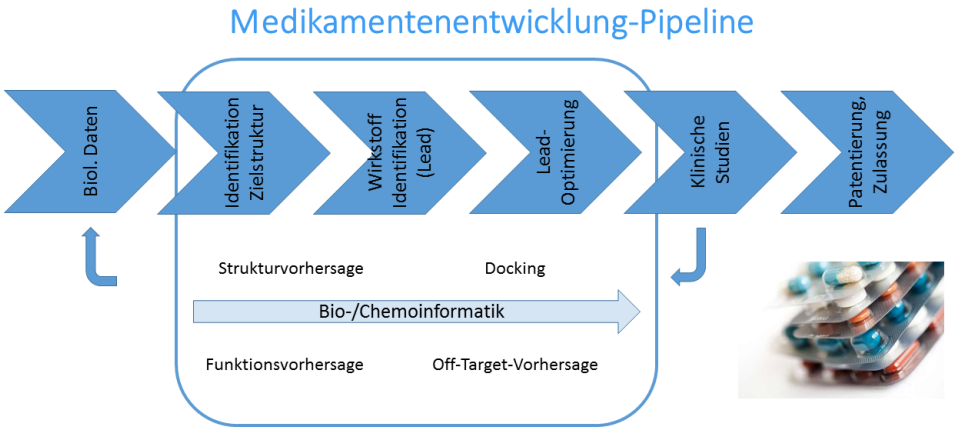


Abbildung 1.1: Schematische Darstellung der Pipeline von der Entwicklung neuer Medikamente bis zur Zulassung.

Die einzelnen Teilgebiete dieses Entwicklungsprozesses werden in den folgenden Abschnitten genauer beschrieben und eine Einordnung in die vorliegende Arbeit wird vorgenommen.

1.2.1 Chemoinformatik

Die Chemoinformatik ist eine Anwendungsgebiet der Informatik, das dazu dient, chemische Problemstellungen zu lösen [5]. Der Begriff „Chemoinformatik“ stammt aus den 1990er Jahren, wobei das Hauptanwendungsgebiet der Wirkstoffentwurf ist [6]. Die Chemoinformatik umfasst einen wichtigen Teil der präklinischen Forschung. Sie wurde maßgeblich durch F. K. Brown geprägt und wie folgt definiert:

> „Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization.“ (F. K. Brown, 1998)

Die Chemoinformatik umfasst drei Teilgebiete des präklinischen Entwurfs. Im ersten Schritt werden Daten akquiriert und evaluiert um, eines oder mehrere Zielmoleküle zu identifizieren. Darauf aufbauend wird gezielt nach kleinen organischen Molekülen gesucht, welche mit diesen Targets interagieren. Im dritten Schritt werden die Wirktstoffkandidaten optimiert, um beispielsweise Nebenwirkungen zu vermeiden oder zu verringern.
Eine große Schwierigkeit beim Arbeiten mit experimentellen Daten besteht darin, die Reproduzierbarkeit der Ergebnisse zu gewährleisten [7, 8]. Da dies leider nicht immer gegeben ist, stellt die Datenqualität stets ein wichtiges Kriterium für prädiktive *in-silico*-Modelle dar [9, 10]. Diese Problematik wird durch den zunehmenden technologischen Fortschritt bei der Generierung experimenteller Daten sowie ihrer technischen Erfassung und strukturierten Speicherung geprägt, da es bislang keine allgemeingültigen Richtlinien für die Durchführung dieser Experimente gibt. Damit stellt die Datenverarbeitung einen entscheidenden Anwendungszweig der Chemoinformatik dar. Der Schwerpunkt liegt dabei auf der Erfassung und Aufarbeitung der Daten, sowie der Analyse und Interpretation.
Um Bioaktivitäten zwischen kleinen organischen Molekülen mit deren Biomolekülen experimentell zu bestimmen, gibt es verschiedene Möglichkeiten. Die am häufigsten verwendeten Verfahren, sogenannte Assays, untergliedern sich hauptsächlich in direkte Bindungsassays und funktionelle Assays. Aufgrund dieser Differenzierung können die Assays zu unterschiedlichen Ergebnissen führen und müssen entsprechend separat aufgearbeitet werden. Aufgrund dieser Tatsache ist die Erstellung valider Datenquellen, unter Berücksichtigung verschiedener Selektionskriterien für die Arbeit mit Bioaktivitätsdaten unerlässlich [11].
Um anschließend ähnlichkeitsbasierte *in-silico*-Modelle zu entwickeln, muss zum Einen definiert werden, was „Ähnlichkeit von kleinen organischen Molekülen“ bedeutet, und zum Anderen müssen diese Methoden für große Datensätze konzipiert werden, da ansonsten virtuelle Hochdurchsatzverfahren (vHTS) nicht in angemessener Zeit angewendet werden können. Dies ist aufgrund von großen Strukturbibliotheken notwendig, um keine Informationen zu übersehen und somit den gesamten verfügbaren Bindungsraum auszuschöpfen.
Basierend auf dem Vorhandensein verschiedener Daten lässt sich die Chemoinformatik in zwei Anwendungsbereiche unterteilen: den liganden- und den strukturbasierten Wirkstoffentwurf (siehe Abbildung 1.2).

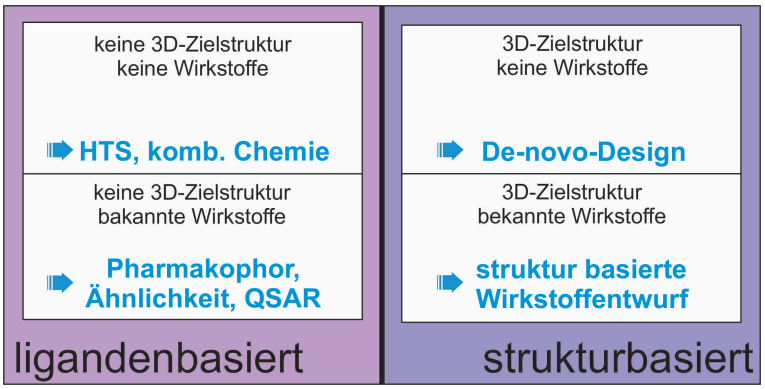


Abbildung 1.2: Anwendungsgebiete Chemoinformatik - Unterteilung in struktur- und liganden-basierte Wirkstoffentwurf unter Berücksichtigung der Datengrundlage

Bei dem ligandenbasierten Entwurf werden Algorithmen verwendet, welche die Ähnlichkeit kleiner organischer Moleküle miteinander vergleichen. Diese Algorithmen folgen dem zentralen Dogma der Chemoinformatik welches besagt, dass ähnliche Strukturen ähnliche Eigenschaften und Bindungspartner teilen [12]. Sofern keine Interaktionsdaten vorhanden sind, wird kombinatorische Chemie mit Hochdurchsatzverfahren kombiniert, um neue Biomoleküle oder neue Substanzen zu identifizieren.
Für den strukturbasierten Wirkstoffentwurf muss eine aufgeklärte dreidimensionale Struktur (Röntgenkristallographie (X-Ray), Kernspinresonanzspektroskopie (NMR), Kryo-Elektronenmikroskopie) oder ein Homologiemodell (siehe Abschnitt 1.3) vorhanden sein. Basierend auf diesen Informationen können beim „De-Novo-Design“ neue Leitstrukturen identifiziert werden. Aufgrund des Vorhandenseins von Proteinstruktur sowie potentiellen oder bekannten Wirkstoffen, können virtuelle Hochdurchsatzverfahren durchgeführt werden. Hierzu zählen vor allem Docking und virtuelles Screening von großen Datenbanken mit kleinen organischen Molekülen [13, 14]. Nach der erfolgreichen Identifikation einer Leitstruktur durch die genannten Methoden wird der Wirkstoffkandidat in weiteren Untersuchungen hinsichtlich verschiedener Eigenschaften optimiert, was als Lead-Optimierung bezeichnet wird (siehe Abschnitt 1.2.2).
In der vorliegenden Arbeit wurden verschiedene Fingerprints verwendet, um die strukturelle Ähnlichkeit von Verbindungen zu beschreiben und miteinander zu vergleichen. Des Weiteren wurde die dreidimensionale Ausrichtung (Flexibilität) von kleinen organischen Molekülen in den Ähnlichkeitsvergleich integriert. Dabei wurden verschiedene energetisch günstige Orientierungen der Verbindungen berechnet und anschließend miteinander abgeglichen, wodurch die Flexibilität der Moleküle simuliert wird.

1.2.2 Präklinische Forschung

Die präklinische Forschung untergliedert sich in drei Teilgebiete: *in-silico* (bioinformatisch/chemoinformatisch), *in-vitro* (außerhalb eines lebenden Organismus) und *in-vivo* (im lebenden Organismus). Der chemoinformatische Bereich umfasst die computergestützte

Analyse bekannter biologischer Daten, welche für die Entwicklung einer Therapie notwendig sind (siehe Abschnitt 1.2.1). Liegen diese Informationen nicht vor, werden metabolische Netzwerke analysiert, um mögliche Zielmoleküle zu identifizieren. Anschließend wird nach einem kleinen organischen Molekül gesucht, welches mit dem Zielmolekül interagiert und den gewünschten Effekt erzeugt. Die so identifizierten Moleküle („Hits“) werden durch *in-vitro*- und *in-vivo* Versuche validiert und ausgeweitet. Dies ist notwendig, um eine größtmögliche Zuverlässigkeit von neuen Medikamenten im prä-klinischen Bereich zu erreichen. Sofern die so gewonnene Substanz gute Wirkungseigenschaften zeigt, wird sie als sogenannte Leitstruktur („Lead“) verwendet. In einem iterativen Verfahren wird versucht, den Wirkungsgrad der Leitstruktur zu erhöhen sowie potenzielle Nebenwirkungen zu verringern. Dieser Prozess wird als Lead-Optimierung bezeichnet [15, 16, 10].
Ein wichtiger Aspekt ist dabei die Promiskuitivität der Leitstruktur. Sie wird auf mögliche Interaktionen mit toxikologisch relevanten Zielmolekülen untersucht („Off-Targets“), welche in einem iterativen Prozess eliminiert werden [17]. Des Weiteren werden Pharmakokinetik und Pharmakodynamik des Wirkstoffkandidaten untersucht. Hauptaspekt der Pharmakokinetik ist die Frage, wie der Organismus auf dem Wirkstoff reagiert, diesen verarbeitet und im Körper verteilt. Dies spiegelt sich in dem Term ADME(T) (Absorption, Distribution, Metabolism, Excretion, (Toxicity)) wider. Im Gegensatz hierzu ist die Pharmakodynamik auf die Auswirkungen des Wirkstoff auf den Organismus fokussiert. Dabei werden die Dosis, die Wirkung des Wirkstoffkandidaten auf das Zielmolekül (Rezeptor) sowie Wechselwirkungen mit anderen Medikamenten im Detail untersucht. Nach erfolgreicher Identifizierung und Optimierung wird der Wirkstoffkandidat in klinischen Studien genauer untersucht.
Als pharmakologischer Parameter wird häufig die Bioverfügbarkeit betrachtet, welche den Anteil des Wirkstoffes beschreibt, der unverändert im systemischen Kreislauf zur Verfügung steht [18]. Hierdurch wird deutlich, wie viel von dem Wirkstoff aufgenommen wurde und wie viel am Wirkungsort zur Verfügung steht [19].

1.2.3 Klinische Forschung

Nach erfolgreicher Optimierung wird der Wirkstoff in den klinischen Phasen weiter getestet. In der ersten klinischen Phase wird das potenzielle Medikament an gesunden Probanden untersucht. Diese Studien umfassen normalerweise zwischen 20 und 100 Teilnehmer. Ziel dieser Untersuchung ist es, Nebenwirkungen und Verträglichkeit zu evaluieren, sowie die geeignete Dosierung zu bestimmen. Sofern keine gravierenden Nebenwirkungen aufgetreten sind, erfolgt die weitere Prüfung in Phase II. Hier wird das Medikament an einer größeren Kohorte von 100 bis 300 Probanden auf seine Wirksamkeit, die biologische Aktivität und den daraus resultierenden therapeutischen Effekt untersucht [20]. Des Weiteren erfolgt eine genaue Analyse von Nebenwirkungen. Es ist die kritischste Phase in der Entwicklung eines Medikaments, da bei ungünstigem Risiko-Nutzen-Verhältnis das Studienprogramm an dieser Stelle abgebrochen werden muss.
Die klinische Phase III wird in der Regel an verschiedenen Standorten (multizentrisch)

durchgeführt. Daraus ergibt sich eine Kohortengröße von 300 bis 3.000 Probanden. In dieser Phase werden die Effektivität und der damit verbundene klinische Impact analysiert und ausgewertet. Durch die Größe der Population und umfangreiche Anforderungen ist dies mit Abstand die teuerste und zeitaufwendigste Studie bei der Entwicklung eines neuen Medikaments.
Sofern das neue Medikament in allen Phasen einen guten klinischen Effekt gezeigt hat, wird es in der Regel patentiert und durch ein pharmazeutisches Unternehmen vermarktet. Dieses kann nun, geschützt durch das Patent, mit dem Medikament die Kosten für seine Entwicklung ausgleichen und Profit erwirtschaften.

1.3 Identifizierung potenzieller Zielmoleküle

Das humane Genom ist ca. 3,27 Milliarden Basenpaaren lang und wurde 2001 erstmals komplett sequenziert [21]. Darin enthalten sind ungefähr 22.500 Gene [22]. Diese Zahl entspricht einer Schätzung verschiedener Konsortien wie NCBI, Ensembl und UCSC Genome Browser sowie den daraus identisch annotierten Proteinen durch das CCDS-Projekt (Consensus Coding Sequence Project) [23]. Ein Gen kann hierbei für verschiedene Proteine kodieren.
Ihrer Schätzung nach kodieren etwa 10-14% der humanen Gene für potenzielle Zielmoleküle von kleinen organischen Molekülen [24]. Des Weiteren sind nur solche Gene bei der Wirkstoffentwicklung von Interesse, deren Genprodukte mit einer Erkrankung assoziiert sind. In einer ersten genaueren Untersuchung beschrieben Hopkins & Groom im Jahr 2002 ein Genprodukt als „druggable“, sofern dieses in der Lage ist, wirkstoffähnliche Moleküle zu binden. Daraus ergibt sich eine Gesamtzahl an potenziellen Zielmolekülen von ungefähr 1.500 Strukturen (siehe Abbildung 1.3).

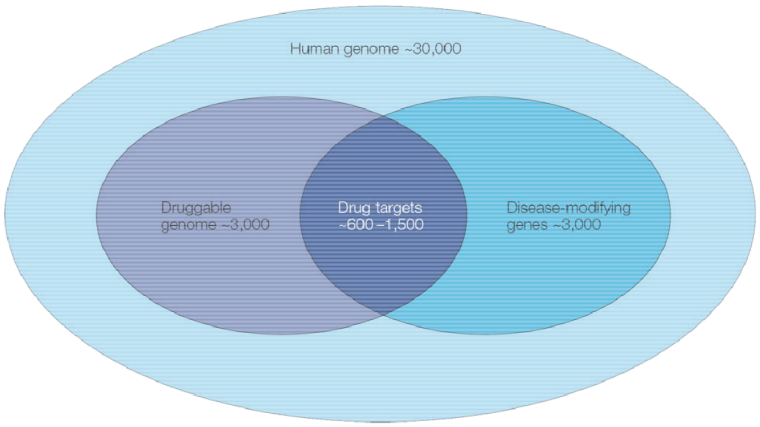


Abbildung 1.3: Visualisierung potenzieller Zielmoleküle für den Wirkstoffentwurf, basierend auf den Schätzungen von Hopkins & Groom [24].

Um diese Gene und Genprodukte krankheitsspezifisch analysieren zu können, müssen die Daten im Kontext zur jeweiligen Erkrankung interpretiert werden. Hierzu werden metabolische Netzwerke verwendet, welche den aktuellen Wissensstand für verschiedene Organismen (vor allem aber für Homo sapiens) enthalten. Eine bekannte Datenquellen hierfür ist die „Kyoto

Encyclopedia of Genes and Genomes“ (KEGG) [3]. Um ein Biomolekül als potenzielles Zielmolekül zu identifizieren, wird eine differenzielle Analyse von krankem und gesundem Gewebe durchgeführt um krankheitsassoziierte, regulierte Gene zu identifizieren. Zur experimentellen Verifizierung werden Methoden der Transkriptomik (RNA-Sequencing etc) und Proteomik (Flüssigchromatographie mit Massenspektrometrie-Kopplung (HPLC-MS) oder Isotopenkodierte Tags [25, 26, 27] verwendet. Dabei ist zu beachten, dass die mRNA-Menge häufig nicht mit der Protein-Menge korreliert.
Durch das Verwenden der genannten Methoden ist es möglich darzustellen, ob das potenzielle Zielmolekül eine zentrale Rolle in den krankheitsassoziierten Stoffwechselwegen spielt und somit durch ein kleines organisches Molekül therapeutisch beeinflusst werden könnte [28]. Ist der Test positiv, wird das so identifizierte Zielmolekül validiert und weitere Informationen werden gesammelt.
Voraussetzung für die Verwendung von strukturbasierten *in-silico*-Methoden ist eine dreidimensionale Struktur des Zielmoleküls. Alternativ kann, basierend auf homologen dreidimensionalen Strukturen, ein Homologiemodell des Zielmoleküls erstellt werden. Hierfür gibt es verschiedene Bibliotheken und Webserver, wie zum Beispiel Swissmodel oder Itasser, die ein automatisiertes Homologiemodell von der Zielsequenz erstellen [29, 30]. Hier lassen sich verschiedene Einstellungen wählen, um beispielsweise das Alignment zu verbessern oder multiple Referenz-Strukturen zu verwenden. Ein häufig verwendetes Softwaretool ist „Modeller“, welches auf der Programmiersprache Python basiert. Dieses Tool bietet eine freie Konfiguration aller Parameter an, die bei der Erstellung eines Homologiemodells verwendet werden können [31].
Der normale Workflow zur Erstellung eines Homologiemodells beginnt mit einer Blast-Suche [32] gegen eine Protein-Datenbank [33]. Sofern ähnliche Strukturen mit Kristallstruktur gefunden werden, wird ein multiples Sequenzalignment erstellt. Basierend auf diesem Alignment wird das Homologiemodell angefertigt. Hierbei ist ein qualitativ hochwertiges Alignment Voraussetzung für die Erstellung eines zuverlässigen Modells. Da bestimmte Regionen wie Transmembranhelizes oder Coild-Coil-Strukturen konserviert sind, wird das Alignment meistens an diese Regionen optimiert. Eine weitere „Qualitätskontrolle“ des Modells erfolgt durch den Ramachandran-Plot. Dieser stellt die Phi- und Psi-Winkel der einzelnen Aminosäuren in einem 2D-Plot dar. Da für verschiedene Sekundärstrukturelemente wie α-Helizes oder β-Faltblatt bestimmte Winkel bevorzugt werden, lässt sich anhand dieses Plots die Energie/Qualität des Modells einfach analysieren. Um mit diesem Modell praxisnah arbeiten zu können, muss es experimentell validiert werden. Hierzu eignen sich zum Beispiel Mutagenesen, mit deren Hilfe der zelluläre mit dem vorhergesagten Effekt verglichen werden kann.
Sofern bekannte Interaktionspartner für das Zielmolekül existieren, kann die Validierung durch Docking oder vHTS erfolgen. Basierend auf den Ergebnissen wird überprüft, ob durch das Homologiemodell Binder von Nicht-Bindern unterschieden werden können. Sofern dies gewährleistet ist, werden neue kleine organische Moleküle vorgeschlagen und anhand experimenteller Bindungsassays überprüft. Die dadurch ermittelten Strukturen können anschließend als neue Wirkstoffe/Wirkstoffkandidaten genutzt werden (siehe Abschnitt 1.4).

1.4 *In-silico*-Vorhersage des Verhaltens von Biomolekülen

Die Vorhersage von neuen Biomolekülen bzw. die Aufklärung von möglichen Nebenwirkungen („Off-Targets“) von Wirkstoffen ist ein wichtiges Werkzeug der Chemoinformatik. Beide Bereiche sind von entscheidender Bedeutung für die Entwicklung neuer Medikamente, da durch die Verwendung zuverlässiger Algorithmen sowohl Zeit als auch Kosten eingespart werden können. Je früher unerwünschte „Off-Targets“ identifiziert werden, desto besser können weitere Substanzen auf hohe Wirksamkeit mit möglichst geringen Nebenwirkungen getestet werden. Ein Hauptaugenmerk liegt dabei auf der Identifizierung von Interaktionen mit Zielmolekülen, durch die toxikologische Reaktionen im Körper ausgelöst werden können.
In der Chemoinformatik ist die Vorhersage von Biomolekülen auf verschiedenen Datengrundlagen anwendbar. Im Folgenden wird die Vorhersage basierend auf ähnlichen Strukturen genauer erläutert. Für diese Algorithmen unterscheidet man zwei verschiedenen Arten von Molekül-Bibliotheken: Zielstruktur-Bibliotheken („targeted libraries“), welche meist auf Proteinfamilien ausgerichtet sind, und fokussierte Bibliotheken, welche alle Wechselwirkungen von kleinen organischen Molekülen abdecken. Letztere sind von großer Bedeutung bei der Aufklärung von Nebenwirkungen und der Identifizierung von neuen Zielmolekülen.
Der Wirkstoffraum („chem space“) von kleinen organischen Molekülen ist aufgrund von Polymeren nahezu unendlich, deshalb müssen potenzielle Substanzen eingegrenzt werden [35]. Eine erste Einschränkung erfolgt in der Regel durch die „Rule of Five“ von Lipinski [36]:

- Molekulargewicht $< 500\ g/mol$
- Donatoren von Wasserstoffbrückenbindungen ≤ 5
- Akzeptoren von Wasserstoffbrückenbindungen ≤ 10
- Verteilungskoeffizienten ($logP$) zwischen Oktanol und Wasser ≤ 5

Sie wird häufig verwendet, da die meisten Medikamente oral verabreicht werden. Allerdings schränkt diese Regel den Wirkstoffraum auf „Wirkstoff-ähnliche“ („druglike“) Moleküle ein, weshalb sie oftmals nicht ausschließlich für orale Wirkstoffe, sondern als genereller Filter verwendet wird.
Solch ein reduzierter Wirkstoffraum wird als Grundlage für die *in-silico*-Berechnungen verwendet. Dabei wird für die gefilterten kleinen organischen Moleküle nach bekannten Interaktionen gesucht. Aufgrund von verschiedenen biologischen Experimenten müssen diese Daten integriert, normalisiert und überprüft werden [9]. Dieser Schritt ist von entscheidender Bedeutung, da er die Datengrundlage für die computergestützten Modelle bildet.
In den vergangenen Jahren wurde der Wirkstoffraum für bestimmte Wirkstoffgruppen drastisch reduziert, da nur bereits zugelassene Medikamente unter dem Aspekt möglicher neuer Indikationen betrachtet wurden. Informationen bezüglich der Indikation werden durch die WHO (World Health Organisation) bereitgestellt und sind durch das „Anatomisch-therapeutisch-chemische Klassifikationssystem“ (ATC) kodiert [37]. Dieses System unterscheidet im ersten Level fünfzehn verschiedene anatomische Gruppen. Für die Indikation

der Medikamente ist mindestens das zweite Level notwendig, welches die therapeutischen Hauptgruppen beschreibt. Dieser Prozess wird als „Drug-Repositioning“ bezeichnet. Ein großer Vorteil dieses Ansatzes ist, dass für diese Medikamente verschiedene Experimente nicht noch einmal durchgeführt werden müssen, wodurch viel Zeit im Entwicklungsprozess eingespart werden kann.
Nachdem der Wirkstoffraum definiert wurde, wird anhand dieses Datensatzes ein entsprechendes prädiktives Modell erstellt, trainiert und anschließend die Qualität der Vorhersage validiert. Für diesen Validierungsschritt können zum Beispiel „Dictionaries of useful decoys“ (DUDs) verwendet werden [38]. In diesen Datenquellen sind für Zielmoleküle aktive und inaktive kleine organische Moleküle zusammengetragen worden. Dadurch lassen sich Spezifität, Sensitivität sowie Genauigkeit der implementierten Methode berechnen und die Qualität des Modells beurteilen. Anschließend müssen diese *in-silico*-Ergebnisse in weiterführenden *in-vitro*- und *in-vivo*-Experimenten untersucht und gegebenenfalls in klinische Studien überführt werden.
Vor allem im akademischen Bereich gibt es seit einigen Jahren verschiedene Initiativen, welche durch kombinierte Methoden eine Reduktion von Tierexperimente anstreben [39, 17]. Aktuell werden in diese Netzwerke vor allem das Potenzial sowie die Umsetzbarkeit dieser Methoden evaluieren [40, 41].

1.5 Personalisierte Medizin

In der Fachliteratur werden verschiedene Bezeichnungen für „Personalisierte Medizin“ (PM) verwendet. Dazu gehören unter anderem „individualisierte Medizin“ oder „Präzisionsmedizin“. Die biologische Bedeutung wird unter Fachleuten kontrovers diskutiert. Die vorliegende Arbeit befasst sich nicht mit der Konnotation dieser Begriffe, weshalb sie unter folgendem Aspekt gleichbedeutend verwendet werden:
„Das Ziel der personalisierten, individualisierten oder Präzisionsmedizin ist es, die Behandlung von Patienten unter Einbeziehung derer genetischer sowie molekularbiologischer Ausstattung anzupassen.“ [42]. Dies ergibt sich aus der finalen Definition von Scheidgen et al., welche „personalisierte Medizin“ wie folgt definieren:

> „PM seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics. “ (Scheidgen et al., 2013)

Für aktuelle klinische Studien werden verschieden große Patientenpopulationen getestet. Auswahlkriterien sind phänotypische Eigenschaften wie Alter, Gewicht, Begleiterkrankungen, Alkoholkonsum, Ernährung oder Raucherstatus. Allerdings kommt es häufig vor, dass einige Patienten von dem neuen Medikament nicht profitieren oder unerwünschte Nebenwirkungen auftreten. Der Hauptgrund hierfür liegt in der genetischen Variabilität der Patienten. Diese Problematik wird durch die personalisierte Medizin aufgegriffen und durch zwei Fachgebiete, die Pharmakogenomik und die Pharmakogenetik, abgebildet.
Die Pharmakogenomik analysiert den genomischen Einfluss auf die Wirksamkeit von Substanzen.

Dabei erfolgt eine DNA-basierte Genotypisierung mit dem Ziel, populationsspezifische Arzneistoffe zu entwickeln sowie Nebenwirkungen und Wirkungsgrad in unterschiedlichen Populationen zu analysieren. Anwendungsgebiet der Pharmakogenetik ist die Identifizierung und Analyse von genetischen Variationen unter Berücksichtigung der Wirksamkeit der Arzneistoffe bei Individuen [43]. Mithilfe dieser zwei Fachgebiete soll in den klinischen Studien sichergestellt werden, dass nur Patienten rekrutiert werden, die, basierend auf ihren genetischen Informationen wahrscheinlich von dem Arzneistoff profitieren. Ziel ist es, in den klinischen Phasen (siehe Abschnitt 1.2.3) Kosten einzusparen und die Medikamente später kostengünstiger zu vermarkten. Bevor solche Medikamente verschrieben werden können, müssen die Patienten genetisch analysiert werden. Nur so kann gewährleistet werden, dass die bestmögliche Therapie ausgewählt wird. Daran kann man erkennen, dass in der PM viele Chancen liegen. Diese sind jedoch mit neuen klinischen Analysen verbunden, aus denen ethische, rechtliche und soziale Fragen resultieren. Beispielsweise lassen sich durch weiterführende Analysen potenzielle Risiken für weitere Erkrankungen identifizieren. Dies kann dazu führen, dass Patienten psychologisch unter Druck geraten und Krankenkassen ihr Beitragsmodell umstellen. Ein solches Wahrscheinlichkeitsmodell wurde bei dem Projekt „23andMe“ eine Zeitlang verwendet, jedoch aufgrund eines Gerichtsurteils nach einiger Zeit eingestellt [44].
Ein wichtiger Aspekt beim Umgang mit personenbezogenen Proben liegt darüber hinaus im Datenschutz und der damit verbundenen Wahrung der Vertraulichkeit. Dies verdeutlicht, dass für personalisierte Medizin verschiedene Dimensionen berücksichtigt werden müssen, um sie in der Praxis zu etablieren.

1.6 Zielsetzung dieser Arbeit

Ziel der vorliegenden Arbeit ist es, die Mechanismen molekularer Interaktionen unter Verwendung und Kombination verschiedener chemoinformatischer Methoden aufzuklären und deren Effekt vorherzusagen. Diese Thematik ist für die Entwicklung neuer Medikamente essenziell, um krankheitsspezifische und nebenwirkungsarme Medikamente zielgerichtet und kostengünstig zu entwickeln. Außerdem werden Verknüpfungen mit dem daran angegliederten *in-vitro*/*in-vivo*-Schwerpunktbereich aufgezeigt (Abschnitt: 1.2). Die vorliegende Arbeit kann somit in den Bereich der präklinischen „Medikamentenentwicklungs-Pipeline“ (Chemoinformatik) eingegliedert werden.
Es wird erörtert, wie wichtig valide und präzise Datenquellen für die Verwendung von prädiktiven *in-silico* Methoden sind und worin ihr Potenzial für weiterführende Experimente besteht. Viele existierende Datenquellen sind redundant und müssen somit prozessiert werden. Der ersten Teil dieser Arbeit beschäftigt sich mit der Erstellung spezifischer integrativer Datenquellen, welche in Zusammenarbeit mit entsprechenden Experten des jeweiligen Fachgebietes konstruiert wurden. Um diese Daten anderen Forschern zur Verfügung zu stellen, wurden sie in Form von dynamischen Websites veröffentlicht.
Basierend auf den so generierten Daten wurde der Wirkmechanismus der kleinen organischen Strukturen identifiziert und zugewiesen. Um die Interaktionsdaten krankheitsspezifisch analysieren zu können, wurden diese Informationen auf Stoffwechselwege projiziert und interpretiert. Diese Daten sind in vorliegender Dissertation zur Vorhersage von neuen Interaktionen und zur Aufklärung von Nebenwirkungen bekannter Medikamente verwendet worden. Um die Vorhersagequalität zu verbessern, wurden verschiedene chemoinformatische Methoden miteinander kombiniert. Mit diesem integrativen Ansatz wurde es möglich, potenzielle Nebenwirkungen durch unerwünschte Interaktionen mit anderen Biomolekülen („Off-Targets“) durch *in-silico*-Modelle zu identifizieren und zu beschreiben.
Ein weiterer Aspekt bei der Reduzierung von Nebenwirkungen stellt der translationale Aspekt der Kombination von Chemoinformatik und Genomik dar. Eine solche aufgearbeitete Datenquelle stellt einen wichtigen Aspekt der personalisierten Medizin dar, da sowohl die generellen Wirkmechanismen als auch der Genotyp des Individuums betrachtet werden. Durch diese Kombination können Variation in Genen durch chemoinformatische Methoden dargestellt, analysiert und individuell ausgewertet werden.

Kapitel 2

Methoden und Ergebnisse

In diesem Kapitel werden die Ergebnisse der vorliegenden Arbeit unter Beschreibung der verwendeten Methoden aufgezeigt sowie die Originalarbeiten in der publizierten Form integriert. Dabei wird in jedem Unterkapitel jeweils eine Originalarbeit beschrieben und auf den eigens bearbeiteten Anteil fokussiert.

2.1 Chemoinformatische Datenintegration/Aufbereitung

Um die Möglichkeiten der chemoinformatischen Datenintegration aufzuzeigen, werden in diesem Abschnitt verschiedene Anwendungsgebiete vorgestellt, für welche akkurate, Hand- validierte Datenquellen erschaffen wurden. Diese Datenquellen wurden durch Textmining komplettiert und mit experimentellen Daten aus verschiedenen Datenbanken zusammengefügt. Um einen präzisen Datensatz zu erstellen, erfolgte eine Standardisierung und Normalisierung der integrierten Informationen.
In den verschiedenen Anwendungsgebieten fehlte die akkurate Schaffung einer Datenquelle zur validen Aufarbeitung und Integration von Interaktionsdaten. Die genaue Analyse und Aufarbeitung heterogener Daten ist der Schlüssel zur Identifikation und Aufklärung von Interaktionen, da die Genauigkeit eines Modells immer von der Qualität der zu Grunde gelegten Daten abhängt [9].

2.1.1 Erstellung einer wissensbasierten Datenquelle für schmerzlindernde kleine organische Moleküle unter Berücksichtigung von Ionenkanälen

Originalarbeit: *SuperPain–a resource on pain-relieving compounds targeting ion channels.*

Ein sehr häufiger Grund für ärztliche Konsultationen sind unangenehme schmerzhafte Sinneswahrnehmungen. Dieser Schmerz geht oftmals mit potenziellen oder realen Gewebeschäden einher, was die Lebensqualität drastisch beeinflussen kann. Die Schmerztherapie ist multifaktoriell, weshalb ein großer Bedarf nach neuen Medikamenten und Zielmolekülen besteht.
In der Originalarbeit SuperPain, wurde eine integrative Datenquelle für

schmerzlindernde/-stimulierende kleine organische Moleküle erstellt. Dabei wurden Interaktionen kleiner organischer Moleküle mit Ionenkanälen analysiert. In Kooperation mit Experten im Bereich Schmerzforschung wurden folgende Ionenkanäle ausgewählt: TRPV1, TRPM8, TRPA1, TREK1, TRESK, hERG, ASIC, P2X und allgemein spannungsabhängige Natriumkanäle. Die Datenquelle wurde mit Interaktionsdaten verschiedener Datenbanken (Drugbank [2], BindingDB [45], Chembl [46] und Supertarget [47]) gefüllt. Da die Datenqualität in den einzelnen Datenbanken stark variiert, wurden alle Daten in einem iterativen Prozess standardisiert und in SuperPain integriert. Der chemische Bindungsraum der Ionenkanäle wurde durch ein Clustering aller aktiven Liganden definiert. Hierfür wurde ein Algorithmus basierend auf dem k-means-Algorithmus implementiert. Dadurch ist es möglich, den Bindungsraum der Kanäle zu klassifizieren und zielgerichtet potenzielle Inhibitoren zu bestimmen. Die Identifikation dieser Moleküle erfolgte durch die Verwendung von chemischen Deskriptoren, sogenannten molekularen Fingerabdrücken. Dabei wird die zweidimensionale Ähnlichkeit zweier Liganden bestimmt. Hierfür wurden OpenBabel Fingerprint 2 (FP2) und Fingerprint 4 (FP4) verwendet. Mit FP2 können kleine organische Moleküle verglichen werden, indem lineare Segmente eines Fragments mit einer Länge von bis zu sieben Atomen verbunden werden. Für jedes Fragment werden Informationen über Atome, Bindungen und das Vorhandenseins von Ringsystemen abgespeichert und in einem Bit-Vektor kodiert. Dieser Bit-Vektor wird durch FP4 komplettiert. Mit diesem Fingerprint werden ausschließlich die funktionellen Gruppen der Moleküle betrachtet. Die Bit-Vektoren werden anschließend durch den Tanimoto-Koeffizienten verglichen und die Ähnlichkeit bestimmt. Um zwischen ähnlichen und unähnlichen Strukturen unterscheiden zu können, wurde für die Tanimoto-Werte ein Schwellenwert von $0,8$ verwendet [48].
Im nächsten Schritt wurden die bekannten Inhibitoren von TRPV1, P2X und ASIC in die 3D-Strukturen gedockt und damit deren Bindungsmechanismus durch eine interaktive Visualisierung auf der Website bereitgestellt. Zusätzlich konnte mit SuperPain gezeigt werden, dass verschiedene Schmerztherapeutika mehrere Ionenkanäle inhibieren („multitarget drugs“). Diese Tatsache muss in weiterführenden Studien genau untersucht werden und stellt somit einen weiteren wichtigen Aspekt beim Entwurf neuer schmerzlindernder Medikamente dar.

Originalarbeit: *SuperPain–a resource on pain-relieving compounds targeting ion channels.*

Björn-O. Gohlke, Robert Preissner and Saskia Preissner

Nucleic Acids Research, 2014, Vol. 42

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1093/nar/gkt1176

2.1.2 Integrative Datenbank für Naturstoffe zur Verwendung im Wirkstoffentwurf

Originalarbeit: *SuperNatural II–a database of natural products*

Naturstoffe spielen eine entscheidende Rolle bei der Entwicklung von Wirkstoffen, da sie an ca. 64% dieser Prozesse beteiligt sind [49]. Viele topologische Pharmakophor-Muster von Naturstoffen stimmen mit denen von kommerziell erwerblichen Medikamenten überein. Daher ist es nicht überraschend, dass verfügbare Medikamente wie Lovastatin, Paclitaxel, Penicillin oder Silibinin direkt oder indirekt von Naturstoffen abgeleitet wurden [71]. Ein besseres Verständnis der spezifischen physikalisch-chemischen und strukturellen Eigenschaften von Naturstoffen ist notwendig, um die Entwicklung von Arzneimitteln voranzutreiben.
Basierend auf diesen Informationen ist es wichtig, Substrukturen von Naturstoffen zu betrachten und diese für den Wirkstoffentwurf einfach zugänglich bereit zu stellen. Diese Informationen wurden in der Originalarbeit SuperNatural II zusammengefügt und physikochemische Eigenschaften für mehr als 320.000 Naturstoffe zusammengetragen. Um effektiv mit Naturstoffen arbeiten zu können ist es notwendig, den genauen Wirkmechanismus („mechanism of action“) zu verstehen, vergleichbar zu machen und auf Stoffwechselwege zu projizieren. Aufgrund der großen Anzahl an organischen Molekülen wurden die Naturstoffe in Gruppen unterteilt („clustern“).
Die Daten wurden in SuperNatural II bearbeitet, auftretende Probleme gelöst und in übersichtlicher Form auf einer Website zusammengetragen (`http://bioinformatics.charite.de/supernatural/`). Dafür stand die Aufbereitung und Vereinheitlichung der Strukturdateien im Fokus. Kernpunkte der Datenintegration waren:

- Entfernung von unvollständigen Strukturdateien
- Normalisierung von Atomladungen
- Entfernung von Ionen, Salzen und Wasser
- Löschen von Duplikaten

Im Anschluss an diesen Integrationsprozess erfolgte die Zusammenfassung der Strukturen in Clustern. Dazu wurde eine angepasste Version des DBSCAN-Algorithmus (Density-based spatial Clustering of Application with Noise) implementiert [50]. Kernidee bei dieser Implementierung war es, nur Strukturen in einem Cluster aufzunehmen, für welche die interne Ähnlichkeit aller Strukturen mindestens $0,8$ beträgt (siehe Abbildung 2.1). Dabei wird ein Cluster aus sehr ähnlichen Strukturen gebildet, die mit hoher Wahrscheinlichkeit einen ähnlichen Bindungsmechanismus aufweisen.

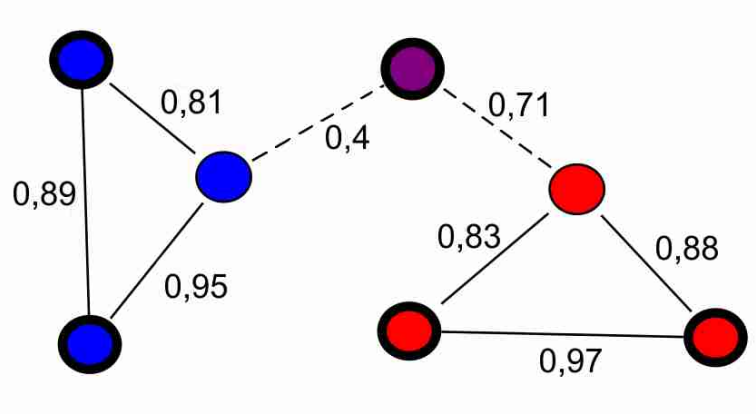


Abbildung 2.1: Visualisierung der Funktionalität des eigens implementierten DBSCAN-Algorithmus in SuperNatural II. Der lila Kreis wurde nicht in den blauen oder roten Cluster mit aufgenommen, da die Ähnlichkeit zu mindestens einer der Strukturen bereits in dem Cluster enthaltener Moleküle kleiner als $0,8$ war. Die minimale interne Ähnlichkeit jedes Moleküls jedes Clusters wurde auf $\geq 0,8$ festgelegt.

Für den Ähnlichkeitsvergleich wurden die Strukturen in molekulare Fingerprints überführt und in der Datenbank als Bit-Vektoren abgespeichert. Dadurch kann der Benutzer seine eigenen Strukturen mit der Datenbank vergleichen und erhält als Ergebnis eine Auflistung der strukturell ähnlichsten Naturstoffe. Für die Identifizierung des Bindungsmechanismus der Naturstoffe wurden Interaktionen der Datenbank Supertarget [47] extrahiert und anhand eines Tanimoto-Vergleiches auf die Naturstoffe projiziert (Tanimoto $> 0,8$). Um pharmazeutische Verwendungszwecke zu ermitteln, wurden für jeden Naturstoff potenzielle Zielmoleküle vorhergesagt.
Damit umfasst SuperNatural II ein breites Spektrum an Naturstoffen mit einer großen chemischen Vielfalt. Durch verschiedene biologische und pharmakologische Eigenschaften bietet diese Datenbank eine einzigartige Quelle für den Einstieg in den virtuellen Wirkstoffentwurf, die Metabolomik und dem damit verbundene Design neuer Wirkstoffe.

Originalarbeit: *SuperNatural II–a database of natural products*

Priyanka Banerjee, Jevgeni Erehman1, Björn-Oliver Gohlke, Thomas Wilhelm, Robert Preissner and Mathias Dunkel

Nucleic Acids Research, 2015, Vol. 43

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1093/nar/gku886

2.1.3 Datenintegration experimenteller Interaktionsdaten synaptischer Proteine

Originalarbeit: *SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations.*

Neuronale Synapsen sind grundlegende Strukturen, welche Nervenzellen im Gehirn miteinander verbinden. Sie sind verantwortlich für die Kommunikation und Informationsverarbeitung. Dadurch werden sehr komplexe und dynamische Informationsnetzwerke aufgespannt, in denen Proteine eine wichtige Rolle spielen. Durch die stetige Weiterentwicklung im Bereich der Proteomics sowie des *Yeast-Two-Hybrid*-Systems (Hefe-Zwei-Hybrid-Systems) lassen sich Interaktionen zwischen kleinen organischen Molekülen mit deren Zielmolekülen besser und mit größerer Genauigkeit auflösen. Da diese Interaktionen durch verschiedene Methoden experimentell bestimmt werden, ist ein Vergleich dieser Daten nicht trivial. Hierfür müssen die Interaktions-Werte standardisiert und normalisiert werden, bevor sie in eine Datenbank integriert werden können. Dieser Schritt ist essenziell, um die Interaktionen in Bezug auf das menschliche Verhalten und Krankheiten zu verstehen und mit diesem Wissen neue Therapien entwickeln zu können.
Dafür wurde SynSysNet entwickelt und ein Ensemble von 1.000 synapsenspezifischen Proteinen durch eine Expertenkommission (European Union Seventh Framework Programme SYNSYS) zusammengestellt. Für diese Proteine wurden verschiedene chemoinformatische Eigenschaften bestimmt, Interaktionsdaten integriert und auf Stoffwechselwege projiziert. Um die Bindungsmechanismen aufzuklären, wurden Homologiemodelle mit Modeller 9v10 erstellt. Hierfür wurden nur die nicht redundanten PDB-Strukturen mit einer Sequenzidentität von 95% verwendet. Die Homologiemodelle wurden anhand der ähnlichsten PDB-Struktur erstellt und durch den „Discrete Optimized Protein Energy“ (DOPE)-Score validiert [51].
SynSysNet umfasst 46.000 kleine organische Moleküle, welche mit den synaptischen Proteinen interagieren. 750 dieser Moleküle konnten als zugelassene Medikamente (mit ATC-Code) identifiziert werden. Die Klasse mit den am meisten interagierenden Medikamenten stellt das Nervensystem (Klasse N: $>$60%) dar. Weitere Klassen sind das Herz-Kreislauf-System (Klasse C: 40%), Sinnesorgane (Klasse S: 40%) und das Atmungssystem (Klasse R: 28%). Des Weiteren wurden die Interaktionen genauer untersucht und das gesamte Interaktionsnetzwerk visualisiert. Dabei war deutlich zu erkennen, dass Medikamente gleicher ATC-Klassen ähnliche Interaktionsnetzwerke aufweisen (siehe Abbildung 1 der Publikation SynSysNet).

Originalarbeit: *SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations.*

Joachim von Eichborn, Mathias Dunkel, Björn-O. Gohlke, et al.

Nucleic Acids Research, 2013, Vol. 41

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1093/nar/gks1040

2.1.4 Erstellung einer interaktiven Datenbank geschlechtsspezifischer medizinischer Literatur

Originalarbeit: *GenderMedDB: an interactive database of sex and gender-specific medical literature.*

Die Suche nach geschlechtsspezifischer Literatur ist kompliziert und benötigt einen speziell hierfür erstellten und validierten Algorithmus. Grundlage für einen solchen Algorithmus ist die Definition eines geschlechter- und krankheitsspezifischen Wörterbuches. Diese Begriffsdefinition wurde in enger Zusammenarbeit mit dem Institut für Geschlechterforschung in der Medizin (GiM) der Charité ausgearbeitet. Die Eingruppierung erfolgt hierbei durch eine iterative Pipeline von Befehlen, in welcher in erster Linie geschlechtsspezifische Bezeichnungen in einem krankheitsassoziierten Kontext in der Fachliteratur vorkommen müssen. Um dies in einem automatisierten und regelmäßigen Turnus durchzuführen, wurde eine Text-Mining-Pipeline implementiert. Dafür wurden alle Publikationen, die in PubMed/Medline frei zugänglich sind, heruntergeladen und im XML-Format strukturiert erfasst. Mithilfe der Softwarepakete Lucene [52] und LingPipe [53] werden anschließend alle Publikationen indiziert und relevante Informationen markiert. Diese Daten werden dann vollautomatisch durch GenderMedST analysiert und mittels einer Website durch Experten validiert und aufgearbeitet. GenderMedDB bietet die Vorteile einer fachspezifischen Suchmaschine, visualisiert diese und stellt sie der Gender-Medizin-Community in einer aufbereiteten Form frei zur Verfügung.
Um hoch qualitatives Arbeiten mit diesen Daten zu ermöglichen, wurde eine vollständig dynamische Website erstellt. Durch unterschiedliche Visualisierungen wurden die Daten weiter aufgearbeitet. Aufgrund der Einteilung von Publikationen nach Krankheiten ist es möglich auf kohortenbasierte klinische Ergebnisse zuzugreifen und geschlechtsspezifische Analysen zu generieren. Dadurch können zum Beispiel Informationen bezüglich der geschlechterspezifischen Wirkung von Medikamenten abgerufen werden, sofern diese publiziert wurden.

Originalarbeit: *GenderMedDB: an interactive database of sex and gender-specific medical literature.*

Sabine Oertelt-Prigione, Björn-Oliver Gohlke, Mathias Dunkel, Robert Preissner and Vera Regitz-Zagrosek

Biology of Sex Differences 2014, 5:7

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1186/2042-6410-5-7

2.2 Wirkstoffentwurf

Beim Wirkstoffentwurf unterscheidet man maßgeblich zwischen zwei verschiedenen Ansätzen. Der struktur- und ligandenbasierten Entwurf wurde bereits auf Seite 4 in Abbildung 1.2 beschrieben. Basierend auf der Datengrundlage kann häufig nur eine begrenzte Anzahl an chemoinformatischen Methoden verwendet werden. Die Anwendung dieser Verfahren und deren Validierung führt zu einem besseren Verständnis der Interaktionen von kleinen organischen Molekülen mit deren Biomolekülen.

2.2.1 Homologiemodell-basierte Mutagenese-Analyse

Originalarbeit: *TRPC6 G757D Loss-of-Function Mutation Associates with FSGS.*

Als „Fokal segmentale Glomerulosklerose“ (FSGS) wird eine Gruppe von chronischen Erkrankungen der Niere (CDK) bezeichnet, die schließlich zu chronischen Nierenversagen führen (ESRD). Die Ursachen von FSGS sind vielfältig. Angeborene Formen können auf Mutationen in verschiedenen Proteinen zurückgeführt werden. Ein wichtiges Gen in diesem Zusammenhang ist der „Transiente Rezeptor Potential Kationenkanal, Unterfamilie C, Mitglied 6“ (TRPC6), was für einen nicht-selektiven Kationenkanal kodiert. Die meisten dieser Mutationen bewirken einen Gain-of-Function-Phänotyp, der zu einem Calcium-getriggerten podocyten Zelltod führt. Die ursächlichen molekularen Wirkmechanismen sind bislang nicht aufgeklärt.
Die molekulare Wirkung von krankheitsbedingten Mutationen wurde mit Hilfe der Entwicklung eines Homologiemodells untersucht und experimentell validiert . Hierfür wurden 19 humane FSGS-bezogene TRPC6-Mutationen charakterisiert und auf Funktionalität untersucht.
Die Erstellung des Homologiemodells war aufgrund der durchschnittlichen Sequenzähnlichkeit und der großen Anzahl verschiedener Domänen nur durch einen iterativen Prozess durchführbar. Die Modellierung wurde durch eine BLAST-Suche gegen die Protein-Datenbank PDB [54] initiiert. Dabei konnten mehrere ähnliche Strukturen mit einem signifikant kleinen $e - value$ identifiziert werden. Ein Großteil dieser Strukturen konnte auf die Ankyrin-Domäne von TRPC6 gemappt werden. Um die gesamte Sequenz dieser Domänen abbilden zu können, wurden sechs ähnliche Strukturen mit überlappenden Aminosäuren ausgewählt. Dies wurde als Grundvoraussetzung für die richtige Rekonstruktion der dreidimensionalen Struktur angenommen. Für die Modellierung der Transmembranhelizes und der C-terminalen Region wurde als Templatestruktur der „spannungsabhängige Kaliumkanal, Unterfamilie A, Mitglied 2“ (Kv1.2, PDB-Code: 2R9R) verwendet. Da die Struktur der Transmembranhelizes stark konserviert ist, wurde das Alignment für diese Regionen manuell optimiert. Die N-terminale Region von TRPC6 wurde anhand des HCN-Kanals (PDB:1q43) rekonstruiert. Für die Rekonstruktion der tetramerischen Struktur von TRPC6 wurde die Cryo-EM-Struktur von TRPV1 verwendet. Das resultierende Homologiemodell wurde hinsichtlich struktureller Störungen bereinigt, Seitenketten wurden optimiert und die Ordnung der Bindungen angepasst. Im Ramachandran-Plot wurden die Diederwinkel ϕ (Phi) und ψ (Psi) gegeneinander aufgetragen. Viele Kombinationen von Winkeln in Polypeptidketten sind verboten bzw. energetisch ungünstig. Die Diederwinkel des

Homologiemodells liegen in den präferierten Regionen für Sekundärstukturelemente (siehe Abbildung 2.2).

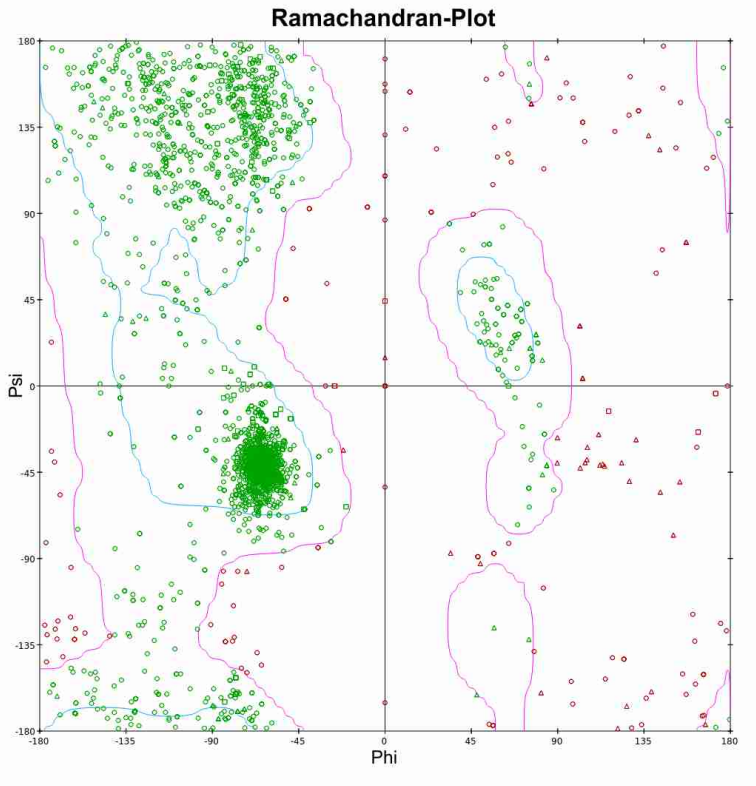


Abbildung 2.2: Visualisierung der Diederwinkel des Homologiemodells zur Überprüfung der Qualität. Ca. 97 % aller Winkel sind in den bevorzugten Regionen. Speziell für das Protein-Protein-Interface, wo die Mutation G757D lokalisiert ist, sind alle Winkel in den bevorzugten Regionen

Basierend auf diesem Modell wurde die Mutation G757D in einer Domäne lokalisiert, welche als Grenzfläche der TRPC6-Untereinheit zuzuordnen ist. Eine Koexpression von Wildtyp-TRPC6 und TRPC6 G757D imitiert Heterozygotie. Dies wurde bereits in Patientenkohorten beobachtet und liess auf eine dominant negative Wirkung von TRPC6 G757D schliessen. Es kann somit angenommen werden, dass der Verlust der TRPC6-Funktion als zusätzliches Konzept der erblich bedingten FSGS anzusehen ist und liefert so molekulare Einblicke in den Mechanismus, der verantwortlich für den Loss-of-Function-Phänotyp TRPC6 G757D beim Menschen ist. Nur durch die Erstellung des Homologiemodells konnte die dreidimensionale Position dieser Mutation aufgeklärt werden. Basierend darauf war es möglich, Mutationen zu erstellen, die die Funktionalität von TRPC6 wiederherstellen konnten („Rescue-Mutanten“) [73].

Originalarbeit: *TRPC6 G757D Loss-of-Function Mutation Associates with FSGS.*

Marc Riehle, Anja K. Büscher, Björn-Oliver Gohlke et al.

J Am Soc Nephrol. 2016 Sep;27(9):2771-83

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1681/ASN.2015030318

2.2.2 Chemoinformatische Aufklärung eines Bindungsmechanismus

Originalarbeit: *Chemical Proteomics Reveals Ferrochelatase as a Common Off-target of Kinase Inhibitors.*

Bei der Entwicklung von Arzneistoffen in der Onkologie werden häufig Proteinkinasen inhibiert, die an Signaltransduktionsprozessen beteiligt sind. Aufgrund der ATC-Bindungstasche von Proteinkinasen sind viele Kinase-Inhibitoren eher unspezifische Binder. Das damit verbundene promiskuitive Verhalten dieser Wirkstoffe kann unerwünschte und toxische Nebenwirkungen hervorrufen. Daher ist es wichtig, den kompletten Bindungsraum von (potenziellen) Wirkstoffe aufzuklären, um die biologische Wirkung im Organismus zu verstehen und vorhersagen zu können.
Das Enzym Ferrochelatase (FECH) katalysiert die Umwandlung von Protoporphyrin IX in Häm und wurde vor kurzem als Off-Target des BRAF-Inhibitors Vemurafenib identifiziert. Das könnte der Grund für das Auftreten einer Phototoxizität bei Melanom-Patienten sein.
Diese Fragestellung wurde in der vorliegenden Arbeit aufgegriffen und in einer Metaanalyse von 226 klinischen Kinase-Inhibitoren genauer untersucht. Es konnte gezeigt werden, dass 29 der untersuchten Kinase-Inhibitoren ebenfalls eine Interaktion mit FECH im nanomolar oder niedrigen mikromolar Bereich aufweisen. Weiterführende Experimente zeigten, dass mit Vemurafenib, Linsitinib, Neratinib und MK-2461 eine Reduktion des Häm-Levels in K562 Zellen erfolgt [55]. Dies ist durch die Bindung der Moleküle in die Häm-Bindungstasche und die damit verbundene Reduktion der FECH-Aktivität zu erklären.
Die genaue Bindungsmechanismus wurde anhand von Docking-Studien untersucht. Aufgrund der existierenden Kristallstruktur von FECH (PDB: 3w1w) und der ko-kristallisierten Cholat-Moleküle wurde eine iterative Pipeline implementiert, um den Bindungsmodus dieser Moleküle zu reproduzieren. Dabei sind die Cholat-Moleküle in dreifacher Ausführung in der Häm-Bindungstasche lokalisiert (siehe Abbildung 2.3).

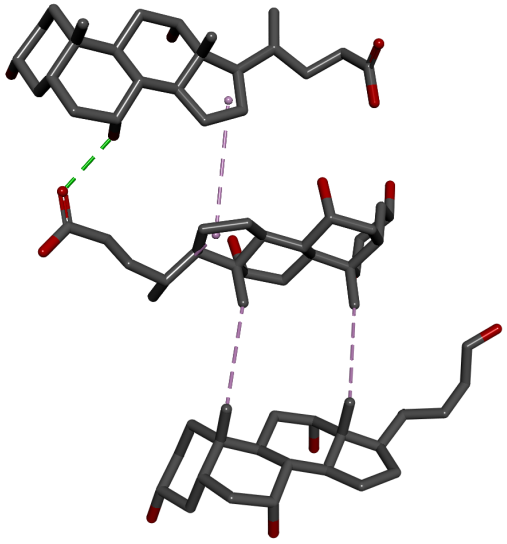


Abbildung 2.3: Visualisierung des Bindungsmechanismus der ko-kristallisierten Cholat-Moleküle in FECH (PDB: 3w1w).

Der angenommene Bindungsmechanismus der Inhibitoren Vemurafenib, Neratinib, Linsitinib, MK-2461, und CUDC-101 konnte durch *in-silico*-Docking bestätigt werden. Vergleicht man dieses Bindungsverhalten mit Dabrafenib, einem „Nicht-Binder“ unterscheidet sich dieser um 90° [77].
Aufgrund dieser Daten wurde empfohlen, ein FECH-Assay in der präklinischen Entwicklung neuer Kinase-Inhibitoren aufzunehmen, da etwa 13% aller Kinase-Inhibitoren auch FECH-Inhibitoren sind [77]. Dadurch würde sichergestellt werden, dass bei der präklinischen toxikologischen Untersuchung eine mögliche Lichtempfindlichkeit durch den Verlust von FECH-Aktivität bei Patienten frühzeitig erkannt und durch den iterativen Prozess der Lead-Optimierung behoben wird.

Originalarbeit: *Chemical Proteomics Reveals Ferrochelatase as a Common Off-target of Kinase Inhibitors.*

Klaeger S, Gohlke B, Perrin J, Gupta V, Heinzlmeir S, Helm D, Qiao H, Bergamini G, Handa H, Savitski MM, Bantscheff M, Médard G, Preissner R, Kuster B

ACS Chem Biol. 2016 May 20;11(5):1245-54

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1021/acschembio.5b01063

2.2.3 Ähnlichkeitsanalysen zur Identifikation von Nebenwirkungen („Off-targets“)

2.2.3.1 Identifikation neuer Zielmoleküle durch Analyse von Ähnlichkeitslandschaften

Originalarbeit: *2D and 3D similarity landscape analysis identifies PARP as a novel off-target for the drug Vatalanib.*

Die zugrunde liegende Studie befasst sich mit der Aufklärung von Off-Targets unter Einbeziehung struktureller Ähnlichkeit kleiner organischer Moleküle. Hierfür wurden zwölf Biomoleküle genauer untersucht, die in der Krebstherapie Zielmoleküle von Arzneistoffen sind. Für diese Zielmoleküle wurde in der SuperTarget-Datenbank nach bekannten kleinen organischen Molekülen gesucht, welche eine Affinität $< 10\mu M$ aufweisen [47]. Hierfür wurde die strukturelle Ähnlichkeit basierend auf verschiedenen Fingerprints untersucht. Aufgrund verschiedener Implementierungen der Fingerprints wurden OpenBabel-Fingerprints sowie Konnektivitäts-Fingerprints (Extended-connectivity Fingerprints; ECFP) berechnet. Die Berechnung der ECFP erfolgte mit dem ChemAxon-Toolkit und wurde durch eine Python-Implementation automatisiert. Da solche Deskriptoren nur die Anordnung der Atome beschreiben, jedoch nicht deren räumliche Ausrichtung, wurde die Ähnlichkeit zudem durch einen Überlagerungs-Algorithmus (Superposition) komplementiert. Hierfür wurde der Konformationsraum jedes Moleküls durch eine Vielzahl an Konformationen abgebildet. Des Weiteren wurde jede Konformation eines Moleküls mit denen eines zweiten Moleküls verglichen. Für diesen Vergleich wurden die Koordinatensysteme normalisiert und überlagert. Anschließend wurden die Trägheitshauptachsen geschätzt und ausgerichtet. Dadurch kam es zu einer starken Reduktion der Anzahl an möglichen Überlagerungen (Orientierungen) zweier Moleküle. Für jede dieser Ausrichtungen wurden jeweils zwei Atome mit möglichst geringem Abstand einander zugeordnet. Aufgrund der chemischen Aktivität von Atomen wurde eine Abstandsobergrenze für die Zuordnung gewählt. Infolge dessen konnten nicht alle Atome eines Moleküls einem Atom des zweiten Moleküls (Konformationen) zugeordnet werden. Da alle Konformationspaare miteinander verglichen wurden, wurde das Paar mit dem geringsten RMSD-Wert ausgewählt. Die Formel zur Berechnung ist im Folgenden dargestellt:

$$rmsd(M) = \sqrt{\frac{1}{n}\sum dist(a_i^M, b_i^M)^2} \tag{2.1}$$

Formel 2.2.1: Formel zur Berechnung der RMSD-Werte

Durch Kombination der beschriebenen 2D- und 3D-Methoden wurden die einzelnen krebsrelevanten Biomoleküle paarweise miteinander verglichen und kombiniert, um ähnliche Strukturen zu identifizieren. Dabei wurden nur Strukturen betrachtet, für welche die 2D-Ähnlichkeit (Tanimoto) $< 0,6$ berechnet wurde und der 3D-RMSD-Wert im 5% Quantil aller RMSD-Werte lag (RMSD $< 0,215$). Damit wurde sichergestellt, dass die beschriebenen Ähnlichkeiten nicht im Hintergrundrauschen des Datensatzes liegen und schaffte so einen

Zugewinn gegenüber etablierten Datenbanken [82]. Diese Herangehensweise war nur möglich, da die Verteilung der RMSD-Werte einer Normalverteilung folgt. Um dies zu demonstrieren, wurde die Verteilung für die Biomoleküle „Vascular endothelial Growth Factor Receptor“ (VEGFR) und „Poly(ADP-ribose)-Polymerase“ (PARP) visualisiert (siehe Abbildung 2.4).

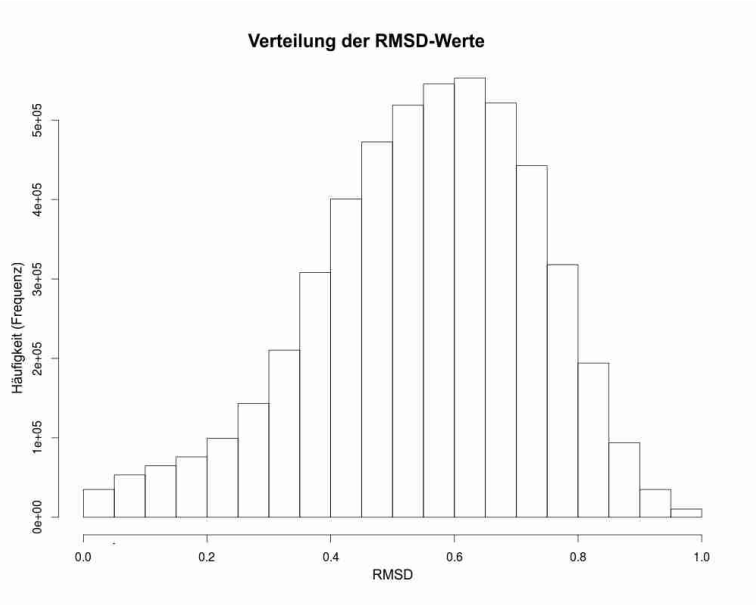


Abbildung 2.4: Verteilung der RMSD-Werte für den Vergleich von VEGFR und PARP

Durch diese Kombination chemoinformatischer Methoden konnte für das bekannte Medikament Vatalanib, PARP als weiteres Zielmolekül identifiziert und experimentell validiert werden [82].

Originalarbeit: *2D and 3D similarity landscape analysis identifies PARP as a novel off-target for the drug Vatalanib.*

Gohlke BO, Overkamp T, Richter A, Richter A, Daniel PT, Gillissen B, Preissner R

BMC Bioinformatics. 2015 Sep 24;16:308

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1186/s12859-015-0730-x

2.2.3.2 Identifizierung Immunvermittelter Nebenwirkungen (IM-ADRs)

Originalarbeit: *Acyclovir Has Low but Detectable Influence on HLA-B*57 : 01 Specificity without Inducing Hypersensitivity.*

Immunvermittelte Nebenwirkungen von Medikamenten (IM-ADRs) können mit dem „Humanen Leukozyten Antigen“ (HLA) Klasse I und II assoziiert sein und zum Tode führen. Eine solche Reaktion zeichnet sich durch das so genannte Überempfindlichkeitssyndrom aus, das sich in Fieber, Hautausschlag sowie Beteiligung verschiedener innerer Organe äußert. Eine schnelle Diagnose und ein sofortiges Absetzen und Ausscheiden des Medikaments ist notwendig, um lebensbedrohliche Auswirkungen zu verhindern.
Das Abacavir-induzierte Überempfindlichkeitssyndrom ist eine cytotoxische T-Zell-abhängige ($CD8^+$) immunvermittelte Nebenwirkung, welche ausschließlich durch HLA-B*57 vermittelt wird. Der Grund hierfür ist die Bindung von Abacavir in die Peptidbindungstasche, wodurch das repräsentierende Muster der Bindungstasche des genannten *Humanen Leukozyten Antiges* verändert wird. Dies resultiert in einer erhöhten Affinität für „Selbst-Peptide“ sowie veränderten Repertoire-gebundenen Peptiden mit der Gefahr lebensbedrohlicher Folgen.
Basierend auf dieser Tatsache wurde der Bindungsraum um Abacavir genauer untersucht und nach ähnlich agierenden Strukturen gesucht. Hierfür wurde eine Pipeline verschiedener chemoinformatischer Methoden entwickelt. Ein einschränkendes Kriterium für die Auswahl der Strukturen stellte ihre Bestellbarkeit dar. Daher wurde die Zinc-Datenbank als Datenquelle ausgewählt [56], da diese ebenfalls Händlerinformationen für die enthaltenen Strukturen beinhaltet. Basierend auf dem Dogma der Chemoinformatik, dass ähnliche Strukturen ähnliche chemische Eigenschaften aufweisen, wurden die Strukturen basierend auf ihrer 2D-Ähnlichkeit zu Abacavir gefiltert. Aufgrund des bekannten Bindungsmechanismus von Abacavir konnten die Strukturen hingehend ihrer 3D-Ähnlichkeit untersucht und gefiltert werden. Auf Basis dieser Überlagerung wurden an der Bindung beteiligte Atome als „Contrains“ gesetzt und für weitere Filterkriterien verwendet (Pharmakophor). Anhand dieser Aneinanderreihung verschiedener Filterkriterien konnten 54 ähnliche Strukturen identifiziert werden. Für diese Strukturen wurde eine weiterführende Docking-Studie durchgeführt, um deren genauen Bindungsmechanismus aufzuklären. Die vielversprechendsten sieben Strukturen wurden in weiterführenden *in-vitro*-Experimenten validiert. Dabei konnte Acyclovir als Interaktionspartner von HLA-B*57 : 01 durch ein direkten Bindungsassay identifiziert werden. In weiteren Experimenten ließ sich keine Acyclovir-spezifische Antwort in einkernigen Zellen des Blutes nachweisen (Peripheral Blood Mononuclear Cell; PBMC). Das heißt, dass durch die Acyclovir-gebundene Veränderung der Bindungsstelle das Repertoire gebundener Peptide nicht so stark beeinflusst wird, dass IM-ADRs ausgelöst werden. Das würde bedeuten, dass ein bestimmtes Level an Bindungsaffinität erreicht werden muss, um IM-ADRs auszulösen [83].
Am Beispiel von Abacavir und Acyclovir wurde die Aufnahme von *in-vitro* präklinischen Screening-Strategien vorgeschlagen, um die Wahrscheinlichkeit für die Entwicklung klinisch relevanter IM-ADRs zu reduzieren.

Originalarbeit: ***Acyclovir Has Low but Detectable Influence on HLA-B*** $^{\star}57:01$ ***Specificity without Inducing Hypersensitivity.***

Imir G. Metushi, Amanda Wriston, Priyanka Banerjee, Bjoern Oliver Gohlke et al.

PLoS One. 2015 May 29;10(5):e0124878

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1371/journal.pone.0124878

2.2.4 Ensemble-basierte Zielmolekül-Vorhersage und Drug-Repositioning

Originalarbeit: *SuperPred: update on drug classification and target prediction.*

Die Ähnlichkeit von Wirkstoffen kann zur Aufklärung von Interaktionen hinsichtlich unterschiedlicher Fragestellungen beitragen. Dabei unterscheidet man vor allem die Identifikation von neuen Wirkstoffen sowie die Aufklärung von Interaktionen kleiner organischer Moleküle mit deren Zielmolekülen. So können zum Beispiel Nebenwirkungen oder neue Anwendungsgebiete für bereits erhältliche Arzneistoffe („Drug-Repositioning“) identifiziert werden. Für beide Fragestellungen wurde ein Algorithmus implementiert, um die Vorhersagerate zu optimieren.
Um Arzneistoffe zu klassifizieren, wurde der durch die WHO bereitgestellte ATC-Baum verwendet. Der in Abschnitt 1.4 beschriebene Aufbau dieser Klassifizierung wurde durch eine Baumstruktur implementiert und auf der frei zugänglichen Website (`http://prediction.charite.de/index.php?site=atc_tree`) veröffentlicht. Um in diesem Zusammenhang wirkstoffähnliche Strukturen zu identifizieren, wurde die „Rule-of-Five“ implementiert und dem Benutzer als Ergebnis ausgegeben.
So können neue Strukturen mit bekannten Wirkstoffen verglichen werden um mögliche Anwendungsgebiete zu erkennen. Dafür müssen potenzielle Interaktionen dieser kleinen organischen Moleküle mit Biomolekülen identifiziert und vorhergesagt werden. Viele öffentlich zugängliche Datenbanken verwenden hierfür ausschließlich 2D-Ähnlichkeitsvergleiche. Dabei wird die Ähnlichkeit der neuen Struktur mit der Datenbank abgeglichen. Für die Vorhersage reicht die Ähnlichkeit über einen definierten Schwellenwert zu einem Molekül der Zielstruktur aus, um als potenzielles Zielmolekül erkannt zu werden. Diese Methode hat Vor- und Nachteile. Ein großer Vorteil ist, dass diese Methode sehr schnell auf große Datenmengen angewendet werden kann und oftmals gute Ergebnisse liefert. Sofern die experimentellen Daten nicht valide aufgearbeitet wurden, kann die Vorhersage in zufällige Ergebnisse resultieren. Um diese Analysen zu optimieren wurde hier eine Methode implementiert, in der das gesamte Ensemble an bekannten Bindern eines Zielmoleküls in die Vorhersage mit einbezogen wurde [81]. Dies basiert auf der von Keiser et al. publizierten Methode: „Similarity ensemble approach“ (SEA). Die Implementierung wurde wie folgend durchgeführt:

- Summation aller Tanimoto-Werte eines Targets, für welche der Tanimoto $> 0,45$ ist ($Raw-Score$)
- Normalisierung des $Raw - Scores$ durch das Teilen durch die Anzahl an Bindern
- Berechnung des $Z - Scores$, um die Spezifität der Vorhersage zu definieren
- Berechnung des $E - Values$; beschreibt die Anzahl an Zielmolekülen, mit denen die unbekannte Struktur durch Zufall interagiert (je kleiner der E-Value, desto signifikanter ist die Vorhersage).
- $E - Value$ und $Z - Score$ verhalten sich konträr

Berechnung des $Z - scores$:

$$Z_A = \frac{(\frac{rawscore_A}{N_A} - \mu)exp(0,335ln(N_A))}{\sigma}$$

Formel 2.2.2: Formel zur Berechnung des Z-Scores

für die gilt:

N_A: Anzahl bekannter Binder für eine Zielstruktur

$rawscore$: Raw-Score

μ und σ: Beschreiben das Grundrauschen der Datenbank

$0,335$: Gewichtungsfaktor

Da einige Bindungstaschen sehr unspezifisch sind, lässt sich die Ähnlichkeit bekannter Binder mit dem Hintergrundrauschen der Datenbank vergleichen. Um diese Problematik zu umgehen, wurde ein gewichteter $Z - Score$ eingeführt, auf welchem die Vorhersage basiert. Die Berechnung ist in der folgenden Formel abgebildet:

$$\lambda_A = exp(0,335ln\frac{rawscore_{AA}}{N_{AA}})\ _{(Gewichtungsfaktor)}$$
$$Z_{\lambda_A} = Z - Score * \lambda_A\ _{(GewichteterZ-Score)}$$

Formel 2.2.3: Formel zur Berechnung des gewichteten Z-Score unter Berücksichtigung des Gewichtungsfaktors zur Unterscheidung unspezifischer Bindungstaschen mit dem Hintergrundrauschen der Datenbank.

Aufgrund der Kombination verschiedener Methoden ist die Vorhersagequalität von SuperPred II besser als vergleichbarer Methoden [81]. Die Vorhersage-Raten sind in der vorliegenden Originalarbeit genauer aufgezeigt und visualisiert.

Originalarbeit: ***SuperPred: update on drug classification and target prediction.***

Nickel J, Gohlke BO, Erehman J, Banerjee P, Rong WW, Goede A, Dunkel M, Preissner R

Nucleic Acids Res. 2014 Jul;42

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1093/nar/gku477

2.2.5 Identifizierung und Analyse zurückgezogener Medikamente

Originalarbeit: *WITHDRAWN-a resource for withdrawn and discontinued drugs.*

Der Entzug der Zulassung von Medikamenten kann verschiedene Gründe haben. Am schwerwiegendsten sind Todesfälle bei Patienten sowie schwere, lebenseinschränkende Nebenwirkungen. Allerdings können Medikamente auch aufgrund von mangelnder Wirksamkeit, Herstellungsproblemen sowie aus regulatorischen oder geschäftlichen Gründen zurückgezogen werden.
Für die Entwicklung neuer Medikamente sind vor allem die Substanzen von Interesse, welche aufgrund von Nebenwirkungen/Todesfällen zurückgezogen wurden. Aus diesem Grund wurden solche Wirkstoffe hinsichtlich des Grundes katalogisiert und klassifiziert [72]. Im Zentrum der Untersuchung stand die Analyse der Toxizitäten dieser Arzneistoffe. Die Klassifizierung erfolgte auf Grund der in der Fachliteratur angegebenen Problematiken. Sofern möglich, wurden diese Gründe mit Interaktionen zu bestimmten Zielmolekülen assoziiert.
So konnte beispielsweise für das Medikament Sibutramin gezeigt werden, dass die Off-Target-Interaktionen mit dem Androgenrezeptor Alpha 2B (ADRA2B) mit einer erhöhten Noradrenalinfreisetzung einhergeht, was wiederum zu erhöhtem Blutdruck und unerwünschten kardiovaskulären Ereignissen führt.
Diese Nebenwirkungen treten jedoch nicht bei allen Patienten auf. Ein ursächlicher Mechanismus für die erhöhte Konzentration von Sibutramin ist die verlangsamte Verstoffwechselung des Medikamentes. Dies ist auf eine Punktmutation (SNP) zurückzuführen [57], die durch Zhang et al. [58] mit den genannten Nebenwirkungen assoziiert wurde. Die Identifikation und Assoziation von SNPs in Hinblick auf bestimmte Krankheitsbilder erfordert Zugang zu Patientenmaterial und der Möglichkeit der Daten-Evaluation. In der vorliegenden Arbeit wurden für identifizierte Zielmoleküle, welche an toxikologischen Reaktionen beteiligt sind, SNPs identifiziert und in die Datenbank aufgenommen. Die Extraktion erfolgte mit der Datenbank dbSNP [59], wodurch ebenfalls populationsspezifische Informationen aufgenommen werden konnten.
So konnte gezeigt werden, dass die individuelle Betrachtung von Populationen bzw. einzelnen Patienten eine wichtige Rolle bei der Vorhersage von möglichen Nebenwirkungen spielt. Um diese individualisierte oder personalisierte Medizin in der Praxis flächendeckend umzusetzen, sind jedoch unterschiedliche Voraussetzungen notwendig (siehe Abschnitt 1.5).

Originalarbeit: *WITHDRAWN-a resource for withdrawn and discontinued drugs.*

Siramshetty VB, Nickel J, Omieczynski C, Gohlke BO, Drwal MN, Preissner R

Nucleic Acids Res. 2016 Jan 4;44(D1):D1080-6

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1093/nar/gkv1192

2.3 Disziplin-übergreifende Wissensdatenbank - ein Schritt in Richtung personalisierter Medizin

Originalarbeit: *CancerResource-updated database of cancer-relevant proteins, mutations and interacting drugs.*

Basierend auf Daten der Weltgesundheitsorganisation (WHO) ist Krebs eine der häufigsten Todesursachen weltweit. Im Jahr 2012 starben ungefähr 8,2 Millionen Menschen an einer onkologischen Erkrankung. Aufgrund dieser Tatsache ist es von essenzieller Bedeutung, bereits bekanntes Wissen über die verschiedenen Krebs-Entitäten zu bündeln. Um dies zu realisieren und die Erkenntnisse der Krebsforschung in aufbereiteter Form zur Verfügung zu stellen, wurde CancerResource entwickelt und als Website frei zugänglich publiziert [63].
Einen wichtigen Aspekt stellt die Datenakquise dar. Es wurden verfügbare Informationen zu mehr als 2.000 Zelllinien gesammelt und hinsichtlich genomischer sowie chemogenetischer Informationen untersucht. Im Bereich der genomischen Analyse wurden sowohl Genexpressions- als auch Mutationsprofile für die einzelnen Zelllinien erstellt und normalisiert. Durch die Möglichkeit des eigenen Daten-Uploads ist es dem Benutzer möglich, für seine spezifische Probe nach bekannten Mustern zu suchen. Durch die Kombination mit der Chemoinformatik (Chemogenetik) und der Berechnung von Aktivitätsprofilen („cellular fingerprint“) ist es somit möglich, das potenziell wirkungsvollste Medikament für die hochgeladenen Daten zu identifizieren.
Eine weitere Herausforderung bei der Therapie von Krebspatienten ist das Auftreten von Resistenzen gegen bestimmte Arzneistoffe. Diese Resistenzen lassen sich häufig auf einzelne Alterationen in bestimmten Genen zurückführen. Daher ist die Suche nach alternativen Arzneistoffen ein wichtiger Aspekt in der Krebstherapie. Hierfür kann CancerResource verwendet werden [63]. Zudem lassen sich ebenfalls durch einen einfachen Suchmechanismus strukturell ähnliche Arzneistoffe ermitteln. Ein weiterer Aspekt, der zur Resistenzentwicklung führen kann, wird durch den Expressionsstatus einzelner Gene beschrieben. In diesem Kontext ist es notwendig, involvierte Stoffwechselwege genauer zu untersuchen. Sofern beispielsweise eine Zielstruktur eines Arzneistoffes in herunterregulierter Form vorliegt, kann man durch die Projektion auf die Signalwege nach weiteren potenziellen Zielstrukturen für alternative Arzneistoffe suchen.
Aufgrund dieser Tatsache kann CancerResource als ein erster Schritt in Richtung personalisierter Medizin angesehen werden, da verschiedene Grundvoraussetzungen für die Betrachtung individueller Datensätze gegeben sind und hierfür die vermutlich wirkungsvollsten Arzneistoffe ausgegeben werden.

Originalarbeit: *CancerResource-updated database of cancer-relevant proteins, mutations and interacting drugs.*

Gohlke BO, Nickel J, Otto R, Dunkel M, Preissner R

Nucleic Acids Res. 2016 Jan 4;44(D1):D932-7

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:
http://dx.doi.org/10.1093/nar/gkv1283

Kapitel 3

Diskussion

3.1 Implementierung nicht redundanter valider Datenquellen

Die Datenaufbereitung experimentell bestimmter Interaktionen ist von großer Bedeutung für die Verwendung von *in-silico* Methoden. Bioaktivitäts-Daten können mit Hilfe unterschiedlicher Methoden gewonnen werden. Hierbei wird zwischen Bindungsassays, funktionalen Assays und ADMET-Assays unterschieden. Für die Vorhersage von Interaktionen kleiner organischer Moleküle mit deren Biomolekülen eignen sich vor allem Bindungsassays, da hier eine direkte Interaktion bestimmt werden kann. Um die Interaktionsstärke zu beschreiben, werden in der Regel IC_{50}, K_i und K_d-Werte bestimmt. Allerdings ist es schwer, diese unterschiedlichen Informationen in eine gemeinsame Einheit zu überführen [60]. Aufgrund der Komplexität bei der Messung solcher Interaktionen ist es jedoch unerlässlich, alle verfügbaren Informationen zu betrachten und zielgerichtet auszuwerten.
Ein weiterer Aspekt bei *in-vitro* und *in-vivo* Experimenten ist die Reproduzierbarkeit. Begley & Ellis untersuchten die Reproduzierbarkeit präklinischer Ergebnisse [61]. Basierend auf der Zitationsrate wurde festgestellt, dass in ca. 50% weiterführender Experimente keine Reproduktion von publizierten Studien möglich war. Das Arbeiten mit solch wiedersprüchlichen Daten ist daher äußerst kompliziert [62]. Daher ist die Entwicklung geeigneter Vorverarbeitungsalgorithmen von Bioaktivitätsdaten nötig. Aufgrund der Komplexität einzelner Forschungsschwerpunkte empfiehlt sich eine Kooperation zwischen Forschung und Datenaufbereitung. Dies ist zum Beispiel von Vorteil, um für die verschiedenen Forschungsschwerpunkte bzw. Krankheiten ein Ensemble an Biomolekülen zu definieren, die mögliche Zielmoleküle von kleinen organischen Molekülen sein könnten (siehe Abschnitt 1.3). Als Beispiel hierfür ist die Originalarbeit CancerResource [63] zu nennen, in welcher krebsrelevante Zielmoleküle vorausgewählt und nur diese in die Datenbank aufgenommen wurden.
Zur Datenaufbereitung müssen in erster Linie die Inhalte verschiedener Datenbanken analysiert und in eine gemeinsame Datenquelle integriert werden. In den letzten Jahren ist eine Vielzahl an Interaktionsdatenbanken veröffentlicht (z.B. ChEMBL [46], PubChem BioAssay [64], Wombat [65] und BindingDB [45]) worden. Anzumerken hierbei ist, dass sich diese Datenbanken sowohl in

der Datenqualität und -integrität, als auch bei den beinhaltenden Strukturen stark unterscheiden. In einer Studie von Tiikkainen et al. wurden ChEMBL, Liceptor und Wombat miteinander verglichen [66]. Es konnte gezeigt werden, dass nur 1,5% (73.076) aller Strukturen in allen Datenbanken enthalten sind (siehe Abbildung 3.1).

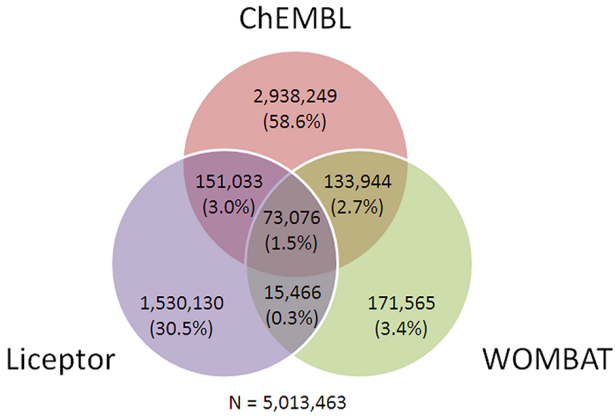
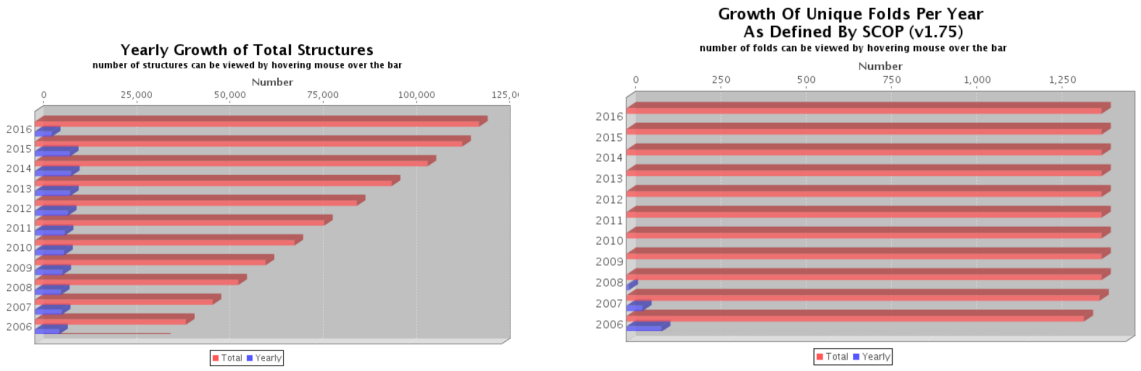


Abbildung 3.1: Überlappung von Bioaktivitätsdaten in ChEMBL, Liceptor und Wombat [66].

Des Weiteren wurde in dieser Publikation eine Fehlerwahrscheinlichkeit für die jeweiligen Datenbanken bestimmt. Trotz der geringen Überlappung gehen Tiikainen et al. von einer Fehlerwahrscheinlichkeit von 5-7% in den einzelnen Datenbanken für enthaltene Liganden aus [66]. Weitere Studien beschäftigen sich mit der Datenqualität von Bioaktivitätsdatenbanken. Zusammengenommen kann festgehalten werden, dass diese Datenbanken eine wertvolle Quelle für weiterführende Methoden im Bereich computerbasiertem Wirkstoffentwurf darstellen [67]. Allerdings müssen die Daten hinsichtlich der Anwendungsgebiete aufgearbeitet werden, um in korrekten Vorhersagen zu resultieren [68].
Die vorliegende Arbeit hatte zum Ziel, diese Problematik aufzugreifen. Durch Kooperationen mit Experten unterschiedlicher Schwerpunkte konnten Originalarbeiten wie SuperPain [69], SynSysNet [70], SuperNatural II [71] und Withdrawn [72] veröffentlicht werden. Um in den Projekten übergreifend den Bindungsraum bekannter Liganden zu analysieren, wurden verschiedene Datenquellen integriert. Dafür wurden alle Liganden normalisiert, bei Identität zusammengefügt und deren Bioaktivitätswerte/-einheiten analysiert. Entsprechende Filterkriterien sind in den betreffenden Publikationen genauer beschrieben. Durch diese Herangehensweise sind spezifische, integrative Datenquellen entstanden, in denen das Wissen interdisziplinärer Felder miteinander kombiniert wurde. Um die integrierten Daten kontextbasiert auswerten zu können, wurden molekularbiologische Daten bezüglich der Annotation von Interaktionen auf Stoffwechselwege, der Identifikation des Wirkmechanismus oder dem Clustering von Liganden den Datenbanken zugefügt. Die so entstandenen integrativen Datenquellen stellen somit eine valide Grundlage dar, um weiterführende *in-silico* Analysen und Vorhersagen durchzuführen.

3.2 *In-silico*-Strukturvorhersage/-analyse

Die Aufklärung dreidimensionaler Strukturen von Proteinen ist stark vorangeschritten. In den vergangenen Jahren wurden pro Jahr ca. 10.000 Strukturen bzw. Komplexe aufgeklärt und in die Proteindatenbank PDB aufgenommen [33], so dass mittlerweile ca. 120.000 Einträge vorhanden sind. Dennoch stagniert die Anzahl unterschiedlicher Proteinfaltungen („folds“). Beide Statistiken sind von der PDB-Website extrahiert und in der folgenden Abbildung zusammengefasst:



(a) (b)

Abbildung 3.2: Ausschnitt aus der Statistik der PDB vom 24 Mai 2016. (a) PDB Statistik über die Aufnahme neuer Strukturen im Kombination mit der gesamten Anzahl an Strukturen. (b) PDB Statistik über die Identifikation neuer Faltungen im Kombination mit der gesamten Anzahl an Faltungen.

Das Vorhandensein von dreidimensionalen Proteinstrukturen kann als Grundvoraussetzung zur Aufklärung von Bindungsmechanismen sowie für die Verwendung verschiedener chemoinformatischer Methoden im Bereich des strukturierten Wirkstoffdesigns angesehen werden. Sofern keine Proteinstrukturen vorliegen, gibt es verschiedene *in-silico*-Methoden, um die 3D-Strukturen vorherzusagen. Dabei unterscheidet man drei unterschiedliche Ansätze: „Ab-initio-Techniken“ (Verwendung der Sequenz und von Potentialen), „Fold recognition“ (Vorhersage von Proteinfaltungen) und „Homologiemodellierung“ (Vorhersage der Struktur, basierend auf homologen Strukturen).
Da die Anzahl möglicher Faltungen in den letzten Jahren stagnierte, wird vor allem Homologiemodellierung bei geeigneter Sequenzähnlichkeit für die Erstellung von Proteinmodellen verwendet. Diese Modelle müssen jedoch durch molekularbiologische Experimente validiert werden (siehe Abschnitt 1.3). Dies ist notwendig um die Funktionalität des Modells zu gewährleisten. Anschließend kann sowohl mit der aufgeklärten Proteinstruktur als auch mit dem Modell gearbeitet werden. In der vorliegenden Arbeit wurde ein Homologiemodell von TRPC6 (siehe Abschnitt 2.2.1) erstellt. Dieses Modell wurde mit der Mutation G757D validiert und es konnte ein Loss-of-Function des Proteins nachgewiesen werden [73]. Anhand des Homologiemodells wurde die Mutation in dem Interface zweier Untereinheiten des Proteins lokalisiert. Durch die Mutation kommt es zu Ladungsproblemen in diesem Interface, wodurch vermutlich der Funktionsverlust zustande kommt. Um das Interface genauer zu untersuchen und

die Funktionalität des Proteins wieder herzustellen („Rescue-Mutanten"), wurden verschiedene Kombinationen von Mutationen basierend auf dem Modell erstellt. Die Kombinationen dieser Mutationen wurde visualisiert und mit der prozentualen Aktivität assoziiert [73]). Ohne diese Modellierung der 3D-Struktur wäre die Analyse in dieser Art und Weise nicht möglich gewesen.
Im strukturbasierten Wirkstoffentwurf ist das Vorhandensein von 3D-Strukturen für die Verwendung verschiedener Methoden essentiell, da diese Strukturen vor allem für Docking, virtuelle Hochdurchsatzmethoden und Pharmakophor-Modellierung benötigt werden. Docking-Studien werden durchgeführt, um unbekannte Bindungsmechanismen zu identifizieren und vorherzusagen [74, 75]. Basierend auf diesen Informationen können wiederum Algorithmen entwickelt werden, um zwischen Bindern und Nicht-Bindern (Decoy Sets) zu unterscheiden [38]. Dabei gibt es eine Vielzahl an unterschiedlichen Bindungsmechanismen. Je größer die Bindungstasche dabei ist, desto unspezifischer wird sie. Stornaiuolo et al. beschreiben beispielsweise einen Bindungsmechanismus, bei dem dasselbe Molekül in dreifacher Ausfertigung in der Bindungstasche gebunden ist [76].
Basierend auf diesem Wissen und dem Vorhandensein eines vergleichbaren Bindungsverhaltens des ko-kristallisierten Cholats bei Ferrochelatase (siehe Abbildung 2.3) wurden in der vorliegenden Arbeit Off-Target-Effekte von Kinase-Inhibitoren aufgrund der Interaktion mit Ferrochelatase (Fech) anhand von Docking-Studien genauer untersucht.
Durch die gewonnenen Ergebnisse konnte für verschiedene Inhibitoren ein vergleichbarer Bindungsmechanismus aufgezeigt werden. Für einzelne Strukturen, die keine Interaktion mit Fech eingehen, wurde ein alternativer Bindungsmechanismus (90° gedreht) gefunden. Nur durch das Vorhandensein der 3D-Struktur konnten diese *in-silico*-Studien durchgeführt und durch Kläger et al. publiziert werden [77].
Anhand der vorliegenden publizierten Daten lässt sich das Potenzial des strukturbasierten Wirkstoffentwurfs erkennen, welches vor allem durch das Vorhandensein einer 3D-Proteinstruktur ausgeschöpft werden konnte. Es konnte gezeigt werden, dass die Kombination von Methoden ein wirkungsvoller Mechanismus ist um computergestützte Vorhersagemodelle zu generieren. Mit Hilfe dieser Modelle könnte der Prozess des Wirkstoffdesigns stark beschleunigt werden, indem zu einem frühen Zeitpunkt unerwünschte Wechselwirkungen (wie zum Beispiel im Fall von Fech) identifiziert werden.

3.3 Vorhersage von Interaktionen kleiner organischer Moleküle mit deren Biomolekülen

Die Vorhersage von Interaktionen kleiner organischer Moleküle mit deren Biomolekülen ist ein wichtiger Bestandteil der Medikamentenentwicklungs-Pipeline und gliedert sich in den Bereich der Chemoinformatik ein. Für die Entwicklung von *in-silico*-Methoden zur Vorhersage von Interaktionen ist eine valide, nicht redundante Datengrundlagen notwendig. Dafür müssen entsprechende Algorithmen entwickelt werden, um diese Informationen aufzubereiten und aus verschiedenen Datenquellen zu integrieren (siehe Abschnitt 3.1). In entsprechenden Publikationen wurden unterschiedliche Algorithmen wie SEA (Similarity Ensemble Approach) [78], Baysian-Modelle zur Identifikation von Nebenwirkungen [79] oder das Mapping des pharmakologischen Raumes [80] zur Vorhersage von Interaktionen verwendet und deren Potential von *in-silico*-Methoden aufgezeigt.
Daran angliedernd konnte in der vorliegenden Arbeit drei Originalarbeiten veröffentlicht werden, in welchen durch die Kombination von chemoinformatischer Methoden Zielmoleküle vorhergesagt wurden. Diese Vorhersagen basieren auf der Verknüpfung verschiedener Ähnlichkeitsalgorithmen, wodurch verschiedene Funktionen identifiziert werden konnten. So konnten beispielsweise zwei unterschiedliche Vorhersage-Ziele miteinander kombiniert (SuperPred [81]) werden. Zum Einen wird die Indikation von kleinen organischen Molekülen basierend auf der durch die WHO klassifizierten Arzneistoffliste (ATC-Baum) vorhergesagt, zum Anderen mögliche Interaktionen mit Zielmolekülen berechnet. Für die ATC-Vorhersage wurde eine Pipeline von Deskriptoren verwendet, welche die Ähnlichkeit von kleinen organischen Molekülen auf verschiedenen Ebenen beschreibt. Diese Pipeline kombiniert die 2D-, Fragment- und 3D-Ähnlichkeit von kleinen organischen Molekülen unter Verwendung geeigneter Schwellenwerte für die Vorhersage. Im Vergleich zu der ersten Version von „SuperPred“ sowie weiteren Studien ist die Vohersagerate durch die implementierte Pipeline die erfolgreichste [81], da geeignete chemoinformatische Methoden mit dem Vorhandensein von Bioaktivitätsdaten in den verschiedenen Datenbanken sowie einer akkuraten Aufarbeitung kombiniert wurden. Die Vorhersage von möglichen Interaktionen wurde in Bezug zu der publizierten Methode „Similarity Ensemble Approach“ [78] implementiert und an den Datensatz angepasst.
Bei Durchführung der Studien bezüglich des Medikaments Vatalanib konnte PARP als neues Zielmolekül identifiziert wurde [82]. Dabei wurden zwei Ähnlichkeitsalgorithmen miteinander kombiniert, was für die Identifikation notwendig war, da durch den 2D-Algorithmus keine Ähnlichkeit aufgezeigt wurde. In diesem speziellen Fall lag dies an dem Austausch eines Atoms, welches zwei Ringsysteme miteinander verknüpft. Durch diesen Austausch haben verschiedene 2D-Methoden (ECFP, OpenBabel FP2/4 sowie MACCS Fingerprints) nur eine sehr geringe Ähnlichkeit identifizieren können. Nur durch die Zuhilfenahme des dreidimensionalen Vergleichs der Moleküle konnte die Ähnlichkeit gezeigt werden.
Im Fall des Medikaments Acyclovir stand das Filtern von möglichen Kandidaten zur Analyse von Off-Target-Interaktion mit HLA im Vordergrund [83]. Acyclovir wurde aufgrund der Ähnlichkeit zu einem vorher mit HLA-interagierenden Medikament (Abacavir) identifiziert.

Aufgrund der Komplexität der Versuche war ein umfassendes Testen an Kandidaten nicht möglich. Daher wurden sieben mögliche kleine organische Moleküle mit einer Ähnlichkeit zu Abacavir identifiziert. Für die Auswahl dieser Moleküle wurde eine komplexe chemoinformatische Pipeline implementiert, in der Wechselwirkungen mit dem Zielmolekül am wahrscheinlichsten sind. Durch diese Kombination wurde Acyclovir als weiterer Inhibitor identifiziert.
So konnten im Rahmen von zwei Publikationen Empfehlungen für die Entwicklung neuer Wirkstoffe ausgesprochen werden, um die gefundenen Off-Target Effekte routinemäßig zu überprüfen [83, 77]. Dadurch könnte möglichen Nebenwirkungen zu einem späteren Zeitpunkt der Pipeline vorgebeugt werden.
Anhand dieser Studien lässt sich erkennen, dass die Vorhersage von Zielmolekülen für eine Vielzahl von Anwendungsgebieten genutzt werden kann und diese einen äußerst wichtigen Teil der Medikamentenentwicklungs-Pipeline abdecken. Allerdings ist es notwendig, die Vorhersageraten noch zu verbessern. Dieser Optimierungsprozess muss auf verschiedenen Ebenen wie der Datenerhebung (experimentelle Bestimmung), der Datenerfassung, der Normalisierung bis hin zur Analyse und Auswertung geschehen. Da die meisten *in-silico*-Methoden auf experimentell bestimmten Daten basieren, stellt dieser Schritt in Kombination mit der chemoinformatischen Aufbereitung der Messwerte den Schlüssel zur Verbesserung der Vorhersagequalität dar. Hierbei muss die Reproduzierbarkeit von Daten, wie z.B. die Bestimmung von Bioaktivitätswerten, gewährleistet sein. Eine Möglichkeit der Qualitätssicherung ist die sogenannte „Gute Herstellungspraxis“ (GMP). Darin werden Richtlinien zur Qualitätssicherung, zu Projektabläufen und den Umwelteinflüssen wie der Temperatur für die Experimente festgehalten [84]. Allerdings ist dies mit hohen Kosten verbunden und somit für viele Forschungseinrichtungen schwer umsetzbar.

Kapitel 4

Fazit und Ausblick

In der vorliegenden Arbeit wurden verschiedene chemoinformatische Methoden verwendet, um die Komplexität der Aufklärung von Nebenwirkungen in Wirkmechanismen und in dreidimensionale Strukturen aufzuklären. Ein wichtiger Bestandteil dieser Arbeit war die Aufbereitung experimentell bestimmter Bioaktivitätsdaten. Durch Kooperationen mit Wissenschaftlern der verschiedenen Anwendungsgebiete gelang es, integrative Datenquellen zu erstellen. Es konnten verschiedene Algorithmen implementiert werden, um diese Daten aufzubereiten und in weiteren Projekten für die Vorhersagen zu verwenden. Durch diese Tatsache und die Kombination verschiedener Methoden ist es im Bereich der Vorhersage von Zielmolekülen gelungen, schwere Nebenwirkungen bestimmter Medikamente zu identifizieren. Daran anknüpfend wurden Empfehlungen ausgesprochen, diese Off-Targets in die routinemäßige Prüfung bei der Entwicklung neuer Medikamente mit einzubeziehen [77, 83].
Ein Ziel für den zukünftigen Entwurf neuer Wirkstoffe könnte zudem sein, Kernstrukturen von zurückgezogenen Medikamenten zu identifizieren [72] und mit denen von neuen Leitstrukturen abzugleichen. Damit könnten es möglich sein, potenzielle Nebenwirkungen bereits in einem frühen Stadium zu erkennen, experimentell zu überprüfen und gegebenenfalls die Leitstruktur zu modifizieren.
Weiterhin wurde aufgezeigt, wie vorteilhaft die Aufklärung der dreidimensionalen Struktur von Biomolekülen für die Verwendung chemoinformatischer Methoden ist. Aufgrund der aktuellen Datengrundlagen im Bereich aufgeklärter 3D-Strukturen (siehe Abbildung 3.2 (a)) ist es gelungen ein Homologiemodell zu generieren und basierend darauf den Einfluss einzelner Mutagenesen zu analysieren um Rescue-Mutanten zur Herstellung der Grundfunktionalität zu erstellen.
Durch die Aufklärung des Bindungsmechanismus der Protoporpherin-Bindungstasche von Ferrochelatase konnten Binder (Vemurafenib) und Nicht-Binder (Dabrafenib) voneinander unterschieden werden. Aufgrund von Docking-Experimenten konnte festgestellt werden, dass Vemurafenib in einer dreifach-gestackten Version in die Protoporpherin Bindungstasche bindet. Hingegen wurde für Dabrafenib ein um 90° gedrehter Bindungsmechanismus identifiziert. Diese Aufteilung zwischen Binder und Nicht-Binder war nur durch das Vorhandensein einer 3D-Struktur und dem darauf durchgeführten Docking-Experiment möglich [77].
In einem weiteren Teil der vorliegenden Arbeit wurde das Wissen über molekulare

Interaktionen von kleinen organischen Molekülen mit Biomolekülen hinsichtlich genomischer Mutationsinformationen (SNPs) verbunden. In bereits publizierten Studien konnte gezeigt werden, dass solche SNPs für eine Unverträglichkeit und im schlimmsten Fall für den Tod verantwortlich sein können. Diese Erkenntnisse sind ein essenzieller Schritt in Richtung personalisierter Medizin. Allerdings macht diese Tatsache die Verknüpfung verschiedener Forschungsschwerpunkte notwendig. Es ist essentiell personenspezifische Informationen wie zum Beispiel Genmutationen und -expressionen zu messen, zu annotieren und auszuwerten, sowie auf molekularer Ebene den Einfluss auf die Bindung des Medikaments oder die Faltung von Proteinen vorherzusagen.
In einem weiteren innovativen Schritt wurde die Datenbank CancerResource entwickelt. Hierfür wurden Biomoleküle bestimmt, welche in der Krebstherapie relevant sind. Für diese Biomoleküle wurden krebsspezifische Bioaktivitätsdaten integriert und mit genomischen Daten komplettiert. Hierzu wurden Expressions- und Mutationsdaten von über 2.000 Zelllinien aufgenommen und normalisiert. Die Kombination dieser beiden Themengebiete liefert eine erste Datengrundlage um krebsrelevante Informationen individualisiert auszuwerten. Dabei müssen Informationen integriert werden, welche die Punktmutationen mit Krankheiten in Verbindung bringen und dies muss klinisch überprüft wurden. Nur durch die Aufarbeitung solcher Informationen und die Aufnahme valider und klinisch überprüfter Daten ist eine zielgerichtete Interpretation möglich. Hierfür bietet zum Beispiel PharmGKB diverse klinisch relevante Informationen an [85].
In weiterführenden Projekten sollen diese Daten mit molekularen Interaktionsnetzwerken kombiniert werden. Für die Therapie verschiedener Krankheiten werden oftmals mehrere Medikamente verordnet. Dies birgt weitere Probleme, da die Wirkstoffe miteinander interagieren und möglicherweise weitere Nebenwirkungen auftreten können. Ziel ist es, durch die Kombination dieser Daten frühzeitig bekannte Nebenwirkungen zu identifizieren und durch alternative Medikamente zu umgehen. Dadurch könnten Patienten individuell von angepassten Therapieen profitieren.
Individualisierte Tumortherapie wird aktuell in der Praxis von verschiedenen Konsortien, wie zum Beispiel die INFORM-Studie [86] des Deutschen Krebsforschungszentrums, umgesetzt. Dabei werden personenspezifische Therapievorschläge für krebskranke Kinder ausgearbeitet. Solche Studien sind jedoch aktuell eher als experimentell anzusehen und keine standardmäßige Therapie bei Patienten.
In der Zukunft müssen auf verschiedenen Ebenen grundlegende Entscheidungen, Verpflichtungen und Gesetze geschaffen werden, um einen sicheren Umgang mit sensiblen personenbezogenen Daten zu gewährleisten und individualisierte Therapiemaßnahmen umzusetzen.

Abbildungsverzeichnis

Formelverzeichnis

Abkürzungsverzeichnis

2D	Zweidimensional
3R	Reduce, Refine, Replace of animal experiments
3D	Dreidimensional
ADME(T)	Absorption, Distribution, Metabolism, Excretion, (Toxicity)
ADRA2B	Androgenrezeptor Alpha 2B
ATC	Anatomisch-therapeutisch-chemische Klassifikationssystem
ATP	Adenosintriphosphat
BB3R	Berlin-Brandenburger Forschungsplattform BB3R
bzw	Beziehungsweise
CCDS	Consensus Coding Sequence Project
CDK	Chronischen Erkrankungen der Niere
DBSCAN	Density-based spatial Clustering of Application with Noise
DOPE	Discrete Optimized Protein Energy
DUDs	Dictionaries of useful decoys
ECFP	Extended-connectivity Fingerprints
ESRD	End Stage Renal Disease (chronischen Nierenversagen)
FECH	Ferrochelatase
FP	Fingerprint
FSGS	Fokal segmentale Glomerulosklerose
GiM	Geschlechterforschung in der Medizin
GMP	Gute Herstellungspraxis
HLA	Humanen Leukozyten Antigen
HPLC-MS	Flüssigchromatographie mit Massenspektrometrie-Kopplung
ICAT	Isotopenkodierte Tags
IM-ADRs	Immunvermittelte Nebenwirkungen (Immune mediated adverse drug reactions)
KEGG	Kyoto Encyclopedia of Genes and Genomes
Kv1.2	Spannungsabhängige Kaliumkanal, Unterfamilie A, Mitglied 2
NMR	Kernspinresonanzspektroskopie
PARP	Poly(ADP-ribose)-Polymerase
PM	Personalisierte Medizin
PDB	Proteindatenbank
RMSD	Root mean square deviation

SEA	Similarity ensemble approach
SNP(s)	Single nucleotide polymorphisms (Einzelnukleotid-Polymorphismus)
TRPC-6	Transiente Rezeptor Potential Kationenkanal, Unterfamilie C, Mitglied 6
VEGF	Vascular Endothelial Growth Factor
vHTS	virtuelle Hochdurchsatzverfahren (virtual high throuthput screening)
WHO	Weltgesundheitsorganisation (world health organisation)
WSBB	Wasserstoffbrückenbindungen
WW	Wechselwirkung
X-Ray	Röntgenkristallographie

Literaturverzeichnis

[1] L. Zhang, M. Pfister, and B. Meibohm, "Concepts and challenges in quantitative pharmacology and model-based drug development," *AAPS J*, vol. 10, pp. 552–559, Dec 2008.

[2] D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, and M. Hassanali, "DrugBank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Res.*, vol. 36, pp. D901–906, Jan 2008.

[3] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes and genomes," *Nucleic acids research*, vol. 28, no. 1, pp. 27–30, 2000.

[4] G. M. Keseru and G. M. Makara, "Hit discovery and hit-to-lead approaches," *Drug Discov. Today*, vol. 11, pp. 741–748, Aug 2006.

[5] F. K. Brown *et al.*, "Chemoinformatics: what is it and how does it impact drug discovery," *Annual reports in medicinal chemistry*, vol. 33, pp. 375–384, 1998.

[6] F. Brown, "Editorial opinion: chemoinformatics-a ten year update.," *Current opinion in drug discovery & development*, vol. 8, no. 3, pp. 298–302, 2005.

[7] L. P. Freedman and M. C. Gibson, "The impact of preclinical irreproducibility on drug development," *Clin. Pharmacol. Ther.*, vol. 97, pp. 16–18, Jan 2015.

[8] L. R. Jager and J. T. Leek, "An estimate of the science-wise false discovery rate and application to the top medical literature," *Biostatistics*, vol. 15, pp. 1–12, Jan 2014.

[9] Y. Hu and J. Bajorath, "Learning from 'big data': compounds and targets," *Drug Discov. Today*, vol. 19, pp. 357–360, Apr 2014.

[10] F. Schmidt, H. Matter, G. Hessler, and A. Czich, "Predictive in silico off-target profiling in drug discovery," *Future Med Chem*, vol. 6, pp. 295–317, Mar 2014.

[11] A. K. Ghose, T. Herbertz, J. M. Salvino, and J. P. Mallamo, "Knowledge-based chemoinformatic approaches to drug discovery," *Drug Discov. Today*, vol. 11, pp. 1107–1114, Dec 2006.

[12] M. A. Johnson and G. M. Maggiora, *Concepts and applications of molecular similarity*. Wiley, 1990.

[13] C. N. Parker and S. K. Schreyer, "Application of chemoinformatics to high-throughput screening: practical considerations," *Methods Mol. Biol.*, vol. 275, pp. 85–110, 2004.

[14] P. Kolb, R. S. Ferreira, J. J. Irwin, and B. K. Shoichet, "Docking and chemoinformatic screens for new ligands and targets," *Curr. Opin. Biotechnol.*, vol. 20, pp. 429–436, Aug 2009.

[15] K. C. Cheng, W. A. Korfmacher, R. E. White, and F. G. Njoroge, "Lead Optimization in Discovery Drug Metabolism and Pharmacokinetics/Case study: The Hepatitis C Virus (HCV) Protease Inhibitor SCH 503034," *Perspect Medicin Chem*, vol. 1, pp. 1–9, 2008.

[16] K. Liszewski, "Drug discovery: Successful lead optimization strategies," *Genet. Eng. Biotechonol. News*, vol. 26, p. 14, 2006.

[17] C. Merlot, "Computational toxicology–a tool for early safety evaluation," *Drug Discov. Today*, vol. 15, pp. 16–22, Jan 2010.

[18] P. Langguth, G. Fricker, and H. Wunderli-Allenspach, *Biopharmazie*. Wiley-VCH Verlag GmbH and Co. KGaA, 2012. S. 467.

[19] R. Voigt and A. Fahr, *Pharmazeutische Technologie für Studium und Beruf*. Deutscher Apotheker Verlag, 11 ed., 1995. S. 223.

[20] L. M. Friedman, C. Furberg, D. L. DeMets, D. M. Reboussin, and C. B. Granger, *Fundamentals of clinical trials*, vol. 4. Springer, 2010.

[21] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, *et al.*, "The sequence of the human genome," *science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[22] M. Pertea and S. L. Salzberg, "Between a chicken and a grape: estimating the number of human genes," *Genome Biol.*, vol. 11, no. 5, p. 206, 2010.

[23] K. D. Pruitt, J. Harrow, R. A. Harte, C. Wallin, M. Diekhans, D. R. Maglott, S. Searle, C. M. Farrell, J. E. Loveland, B. J. Ruef, E. Hart, M. M. Suner, M. J. Landrum, B. Aken, S. Ayling, R. Baertsch, J. Fernandez-Banet, J. L. Cherry, V. Curwen, M. Dicuccio, M. Kellis, J. Lee, M. F. Lin, M. Schuster, A. Shkeda, C. Amid, G. Brown, O. Dukhanina, A. Frankish, J. Hart, B. L. Maidak, J. Mudge, M. R. Murphy, T. Murphy, J. Rajan, B. Rajput, L. D. Riddick, C. Snow, C. Steward, D. Webb, J. A. Weber, L. Wilming, W. Wu, E. Birney, D. Haussler, T. Hubbard, J. Ostell, R. Durbin, and D. Lipman, "The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes," *Genome Res.*, vol. 19, pp. 1316–1323, Jul 2009.

[24] A. L. Hopkins and C. R. Groom, "The druggable genome," *Nature reviews Drug discovery*, vol. 1, no. 9, pp. 727–730, 2002.

[25] I. Finoulst, M. Pinkse, W. Van Dongen, and P. Verhaert, "Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices," *J. Biomed. Biotechnol.*, vol. 2011, p. 245291, 2011.

[26] S. Bajad and V. Shulaev, "Lc-ms-based metabolomics," *Metabolic Profiling: Methods and Protocols*, pp. 213–228, 2011.

[27] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold, "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags," *Nature biotechnology*, vol. 17, no. 10, pp. 994–999, 1999.

[28] M. Rask-Andersen, S. Masuram, and H. B. Schiöth, "The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication," *Annual review of pharmacology and toxicology*, vol. 54, pp. 9–26, 2014.

[29] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli, and T. Schwede, "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information," *Nucleic Acids Res.*, vol. 42, pp. W252–258, Jul 2014.

[30] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER Suite: protein structure and function prediction," *Nat. Methods*, vol. 12, pp. 7–8, Jan 2015.

[31] B. Webb and A. Sali, "Comparative Protein Structure Modeling Using MODELLER," *Curr Protoc Bioinformatics*, vol. 47, pp. 1–32, 2014.

[32] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.*, vol. 25, pp. 3389–3402, Sep 1997.

[33] H. M. Berman, G. J. Kleywegt, H. Nakamura, and J. L. Markley, "The protein data bank archive as an open data resource," *Journal of computer-aided molecular design*, vol. 28, no. 10, pp. 1009–1014, 2014.

[34] M. R. Fielden and K. L. Kolaja, "The role of early in vivo toxicity testing in drug discovery toxicology," *Expert Opin Drug Saf*, vol. 7, pp. 107–110, Mar 2008.

[35] C. Lipinski and A. Hopkins, "Navigating chemical space for biology and medicine," *Nature*, vol. 432, pp. 855–861, Dec 2004.

[36] C. A. Lipinski, "Lead- and drug-like compounds: the rule-of-five revolution," *Drug Discov Today Technol*, vol. 1, pp. 337–341, Dec 2004.

[37] W. H. Organization *et al.*, *The Selection and Use of Essential Medicines: Report of the WHO Expert Committee...(including the... Model List of Essential Medicines).* World Health Organization, 2003.

[38] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking," *J. Med. Chem.*, vol. 55, pp. 6582–6594, Jul 2012.

[39] T. Sahota, M. Danhof, and O. Della Pasqua, "Pharmacology-based toxicity assessment: towards quantitative risk prediction in humans," *Mutagenesis*, Mar 2016.

[40] M. Halder, A. Kienzler, M. Whelan, and A. Worth, *EURL ECVAM Strategy to replace, reduce and refine the use of fish in aquatic toxicity and bioaccumulation testing*. Publications Office, 2014.

[41] L. Farcal, F. Busquet, S. Coecke, I. Hristescu, C. Chesné, C. Pellevoisin, A. Orasanu, Z. Diaconeasa, A. Oros, A. Pintea, *et al.*, "Finding opportunities in the area of alternative methods to animal testing for romania and inauguration of the romanian center for alternative test methods (rocam).," *Altex*, vol. 32, no. 4, pp. 392–393, 2015.

[42] S. Schleidgen, C. Klingler, T. Bertram, W. H. Rogowski, and G. Marckmann, "What is personalized medicine: sharpening a vague term based on a systematic literature review," *BMC Med Ethics*, vol. 14, p. 55, 2013.

[43] E. M. Meslin, E. J. Thomson, and J. T. Boyer, "The ethical, legal, and social implications research program at the national human genome research institute," *Kennedy Institute of Ethics Journal*, vol. 7, no. 3, pp. 291–298, 1997.

[44] G. J. Annas and S. Elias, "23andme and the fda," *New England Journal of Medicine*, vol. 370, no. 11, pp. 985–988, 2014.

[45] M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, and J. Chong, "BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology," *Nucleic Acids Res.*, vol. 44, pp. D1045–1053, Jan 2016.

[46] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Kruger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, and J. P. Overington, "The ChEMBL bioactivity database: an update," *Nucleic Acids Res.*, vol. 42, pp. D1083–1090, Jan 2014.

[47] N. Hecker, J. Ahmed, J. von Eichborn, M. Dunkel, K. Macha, A. Eckert, M. K. Gilson, P. E. Bourne, and R. Preissner, "SuperTarget goes quantitative: update on drug-target interactions," *Nucleic Acids Res.*, vol. 40, pp. D1113–1117, Jan 2012.

[48] M. J. Keiser, J. J. Irwin, and B. K. Shoichet, "The chemical basis of pharmacology," *Biochemistry*, vol. 49, no. 48, pp. 10267–10276, 2010.

[49] D. J. Newman and G. M. Cragg, "Natural products as sources of new drugs over the 30 years from 1981 to 2010," *J. Nat. Prod.*, vol. 75, pp. 311–335, Mar 2012.

[50] S. H. Yue, P. Li, J. D. Guo, and S. G. Zhou, "Using Greedy algorithm: DBSCAN revisited II," *J. Zhejiang Univ. Sci.*, vol. 5, pp. 1405–1412, Nov 2004.

[51] A. Mathur, A. S. Vidyarthi, *et al.*, "Swift modeller v2. 0: a platform-independent gui for homology modeling," *Journal of molecular modeling*, vol. 18, no. 7, pp. 3021–3023, 2012.

[52] M. McCandless, E. Hatcher, and O. Gospodnetic, *Lucene in Action: Covers Apache Lucene 3.0*. Manning Publications Co., 2010.

[53] B. Baldwin and B. Carpenter, "Lingpipe," *Available from World Wide Web: http://alias-i.com/lingpipe*, 2003.

[54] H. M. Berman, G. J. Kleywegt, H. Nakamura, and J. L. Markley, "The Protein Data Bank archive as an open data resource," *J. Comput. Aided Mol. Des.*, vol. 28, pp. 1009–1014, Oct 2014.

[55] C. B. Lozzio and B. B. Lozzio, "Human chronic myelogenous leukemia cell-line with positive philadelphia chromosome," *Blood*, vol. 45, no. 3, pp. 321–334, 1975.

[56] J. J. Irwin and B. K. Shoichet, "Zinc-a free database of commercially available compounds for virtual screening," *Journal of chemical information and modeling*, vol. 45, no. 1, pp. 177–182, 2005.

[57] S. K. Bae, S. Cao, K. A. Seo, H. Kim, M. J. Kim, J. H. Shon, K. H. Liu, H. H. Zhou, and J. G. Shin, "Cytochrome P450 2B6 catalyzes the formation of pharmacologically active sibutramine (N-1-[1-(4-chlorophenyl)cyclobutyl]-3-methylbutyl-N,N-dimethylamine) metabolites in human liver microsomes," *Drug Metab. Dispos.*, vol. 36, pp. 1679–1688, Aug 2008.

[58] W. Zhang, M. W. Roederer, W. Q. Chen, L. Fan, and H. H. Zhou, "Pharmacogenetics of drugs withdrawn from the market," *Pharmacogenomics*, vol. 13, pp. 223–231, Jan 2012.

[59] E. M. Smigielski, K. Sirotkin, M. Ward, and S. T. Sherry, "dbSNP: a database of single nucleotide polymorphisms," *Nucleic Acids Res.*, vol. 28, pp. 352–355, Jan 2000.

[60] Y. Hu and J. Bajorath, "Growth of ligand–target interaction data in chembl is associated with increasing and activity measurement-dependent compound promiscuity," *Journal of chemical information and modeling*, vol. 52, no. 10, pp. 2550–2558, 2012.

[61] C. G. Begley and L. M. Ellis, "Drug development: Raise standards for preclinical cancer research," *Nature*, vol. 483, no. 7391, pp. 531–533, 2012.

[62] L. Freedman and M. Gibson, "The impact of preclinical irreproducibility on drug development," *Clinical Pharmacology & Therapeutics*, vol. 97, no. 1, pp. 16–18, 2015.

[63] B. O. Gohlke, J. Nickel, R. Otto, M. Dunkel, and R. Preissner, "CancerResource-updated database of cancer-relevant proteins, mutations and interacting drugs," *Nucleic Acids Res.*, vol. 44, pp. D932–937, Jan 2016.

[64] Y. Wang, T. Suzek, J. Zhang, J. Wang, S. He, T. Cheng, B. A. Shoemaker, A. Gindulyte, and S. H. Bryant, "Pubchem bioassay: 2014 update," *Nucleic acids research*, p. gkt978, 2013.

[65] M. Olah, M. Mracec, L. Ostopovici, R. Rad, A. Bora, N. Hadaruga, I. Olah, M. Banda, Z. Simon, M. Mracec, *et al.*, "Wombat: world of molecular bioactivity," *Chemoinformatics in drug discovery*, vol. 223239, 2004.

[66] P. Tiikkainen, L. Bellis, Y. Light, and L. Franke, "Estimating error rates in bioactivity databases," *Journal of chemical information and modeling*, vol. 53, no. 10, pp. 2499–2505, 2013.

[67] D. Fourches, E. Muratov, and A. Tropsha, "Trust, but verify: on the importance of chemical structure curation in cheminformatics and qsar modeling research," *Journal of chemical information and modeling*, vol. 50, no. 7, pp. 1189–1204, 2010.

[68] G. Papadatos, A. Gaulton, A. Hersey, and J. P. Overington, "Activity, assay and target data curation and quality in the chembl database," *Journal of computer-aided molecular design*, vol. 29, no. 9, pp. 885–896, 2015.

[69] B. O. Gohlke, R. Preissner, and S. Preissner, "SuperPain–a resource on pain-relieving compounds targeting ion channels," *Nucleic Acids Res.*, vol. 42, pp. D1107–1112, Jan 2014.

[70] J. von Eichborn, M. Dunkel, B. O. Gohlke, S. C. Preissner, M. F. Hoffmann, J. M. Bauer, J. D. Armstrong, M. H. Schaefer, M. A. Andrade-Navarro, N. Le Novere, M. D. Croning, S. G. Grant, P. van Nierop, A. B. Smit, and R. Preissner, "SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations," *Nucleic Acids Res.*, vol. 41, pp. D834–840, Jan 2013.

[71] P. Banerjee, J. Erehman, B. O. Gohlke, T. Wilhelm, R. Preissner, and M. Dunkel, "Super Natural II–a database of natural products," *Nucleic Acids Res.*, vol. 43, pp. D935–939, Jan 2015.

[72] V. B. Siramshetty, J. Nickel, C. Omieczynski, B. O. Gohlke, M. N. Drwal, and R. Preissner, "WITHDRAWN-a resource for withdrawn and discontinued drugs," *Nucleic Acids Res.*, vol. 44, pp. D1080–1086, Jan 2016.

[73] M. Riehle, A. K. Buscher, B. O. Gohlke, M. Kassmann, M. Kolatsi-Joannou, J. H. Brasen, M. Nagel, J. U. Becker, P. Winyard, P. F. Hoyer, R. Preissner, D. Krautwurst, M. Gollasch, S. Weber, and C. Harteneck, "TRPC6 G757D Loss-of-Function Mutation Associates with FSGS," *J. Am. Soc. Nephrol.*, Feb 2016.

[74] J. Huang, W. Hua, J. Li, and Z. Hua, "Molecular docking to explore the possible binding mode of potential inhibitors of thioredoxin glutathione reductase," *Mol Med Rep*, vol. 12, pp. 5787–5795, Oct 2015.

[75] X. Zhang, M. Li, Y. Wang, and Y. Zhao, "Insight into the binding mode of a novel lsd1 inhibitor by molecular docking and molecular dynamics simulations," *Journal of Receptors and Signal Transduction*, vol. 35, no. 5, pp. 363–369, 2015.

[76] M. Stornaiuolo, G. E. De Kloe, P. Rucktooa, A. Fish, R. van Elk, E. S. Edink, D. Bertrand, A. B. Smit, I. J. de Esch, and T. K. Sixma, "Assembly of a π–π stack of ligands in the binding site of an acetylcholine-binding protein," *Nature communications*, vol. 4, p. 1875, 2013.

[77] S. Klaeger, B. Gohlke, J. Perrin, V. Gupta, S. Heinzlmeir, D. Helm, H. Qiao, G. Bergamini, H. Handa, M. M. Savitski, M. Bantscheff, G. Medard, R. Preissner, and B. Kuster, "Chemical Proteomics Reveals Ferrochelatase as a Common Off-target of Kinase Inhibitors," *ACS Chem. Biol.*, Feb 2016.

[78] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet, "Relating protein pharmacology by ligand chemistry," *Nature biotechnology*, vol. 25, no. 2, pp. 197–206, 2007.

[79] A. Bender, J. Scheiber, M. Glick, J. W. Davies, K. Azzaoui, J. Hamon, L. Urban, S. Whitebread, and J. L. Jenkins, "Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure," *ChemMedChem*, vol. 2, no. 6, pp. 861–873, 2007.

[80] G. V. Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason, and A. L. Hopkins, "Global mapping of pharmacological space," *Nature biotechnology*, vol. 24, no. 7, pp. 805–815, 2006.

[81] J. Nickel, B. O. Gohlke, J. Erehman, P. Banerjee, W. W. Rong, A. Goede, M. Dunkel, and R. Preissner, "SuperPred: update on drug classification and target prediction," *Nucleic Acids Res.*, vol. 42, pp. 26–31, Jul 2014.

[82] B. O. Gohlke, T. Overkamp, A. Richter, A. Richter, P. T. Daniel, B. Gillissen, and R. Preissner, "2D and 3D similarity landscape analysis identifies PARP as a novel off-target for the drug Vatalanib," *BMC Bioinformatics*, vol. 16, p. 308, 2015.

[83] I. G. Metushi, A. Wriston, P. Banerjee, B. O. Gohlke, A. M. English, A. Lucas, C. Moore, J. Sidney, S. Buus, D. A. Ostrov, S. Mallal, E. Phillips, J. Shabanowitz, D. F. Hunt, R. Preissner, and B. Peters, "Acyclovir Has Low but Detectable Influence on HLA-B*57:01 Specificity without Inducing Hypersensitivity," *PLoS ONE*, vol. 10, no. 5, p. e0124878, 2015.

[84] V. EudraLex, "Good manufacturing practice (gmp) guidelines." `http://ec.europa.eu/health/documents/eudralex/vol-4/index_en.htm`, 4.

[85] M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman, and T. E. Klein, "Pharmgkb: the pharmacogenetics knowledge base," *Nucleic acids research*, vol. 30, no. 1, pp. 163–165, 2002.

[86] INFORM-Konsortium, "Individualized therapy for relapsed malignancies in childhood." `https://www.dkfz.de/de/inform/index.html`, 4.

Danksagungen

An dieser Stelle möchte ich die Gelegenheit nutzen, mich bei einigen Menschen zu bedanken, deren Unterstützung bei der Erstellung dieser Arbeit wichtig war.

Zuerst bedanke ich mich bei PD Dr. Robert Preißner dafür, dass er mir die Möglichkeit gegeben hat, die vorliegende Arbeit in seiner Arbeitsgruppe anzufertigen. Ein großes Dankeschön möchte ich Prof. Dr. Udo Heinemann für die Übernahme der Zweitgutachterschaft meiner Dissertation aussprechen.

Zudem möchte ich allen Kooperationspartner danken, mit denen ich zusammenarbeiten durfte und großartige Diskussionen führen konnte. Die daraus entstandenen multidisziplinären Publikationen zeigen die Ergebnisse und das Potential solch voranbringender Kooperationen.

Außerdem möchte ich dem Deutschen Krebsforschungszentrum (DKFZ) danken, deren finanzielle Unterstützung ein Großteil dieser Arbeit abgedeckt hat.

Ebenfalls möchte ich meiner Familie danken, die mich immer unterstützt hat: Vielen Dank an meine Eltern, meine Geschwister, meine Korrekturleser (Sabrina, Sebo und Jan) und ganz besonders an meine Freundin Tina.