

# Bioinformatische Datenintegration und Analyse der Eigenschaften molekularer Interaktionen

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

Joachim von Eichborn

aus Bonn

2012

Diese Arbeit wurde im Zeitraum von Januar 2009 bis September 2012 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 12.12.2012

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b> .....	<b>ii</b>
<b>Zusammenfassung</b> .....	<b>iv</b>
<b>Abstract</b> .....	<b>v</b>
<b>Eigene Publikationen</b> .....	<b>vi</b>
<b>Kapitel 1 Einleitung</b> .....	<b>1</b>
1.1 Protein-Protein-Interaktionen .....	2
1.2 Protein-Wirkstoff-Interaktionen .....	6
1.3 Zielsetzung der Arbeit .....	8
<b>Kapitel 2 Methoden und Ergebnisse</b> .....	<b>10</b>
2.1 Protein-Protein-Interaktionen .....	10
2.1.1 Erstellen einer Bibliothek von Makromolekül-Interaktionsflächen .....	10
Originalartikel: "JAIL: a Structure-Based Interface Library for Macromolecules" .....	12
2.1.2 Analyse von Protein-Interaktionsflächen .....	16
Originalartikel: "Structural Features and Evolution of Protein-Protein Interactions" .....	18
2.1.3 Erstellen wissensbasierter Interaktionspotentiale .....	28
Originalartikel: "Development of Immune-Specific Interaction Potentials and Their Application in the Multi-Agent-System VacImm" .....	30
2.2 Protein-Wirkstoff-Interaktionen .....	41
2.2.1 Erstellen einer Ressource für Protein-Wirkstoff-Interaktionen .....	41
Originalartikel: "PROMISCUOUS: a Database for Network-Based Drug- Repositioning" .....	43
2.2.2 Analyse promiskuitiver und nicht-promiskuitiver Wirkstoffe .....	50
Eingereichtes Manuskript: "Polypharmacology of Drugs: Analysing Molecular and Network Properties" .....	52
2.2.3 Interaktive Visualisierung komplexer Netzwerke .....	61

Originalartikel: "Cobweb: a Java Applet for Network Exploration and Visualisation" .....	63
2.2.4 Berücksichtigung quantitativer Daten bei Protein-Wirkstoff-Interaktionen .....	65
Originalartikel: "SuperTarget Goes Quantitative: Update on Drug-Target Interactions" .....	66
<b>Kapitel 3 Diskussion .....</b>	<b>71</b>
3.1 Bereitstellen und Auswerten struktureller Informationen über Protein-Protein-Interaktionen .....	71
3.1.1 Akquisition struktureller Informationen über Makromolekül-Interaktionsflächen .....	71
3.1.2 Strukturelle und evolutionäre Eigenschaften von Protein-Interaktionsflächen .....	72
3.1.3 Anwendung wissensbasierter Interaktionspotentiale zur Bindungsvorhersage .....	73
3.2 Konzeption und Auswertung molekularer Interaktionsnetzwerke.....	74
3.2.1 Datenintegration zur Generierung eines komplexen molekularen Interaktionsnetzwerks .....	74
3.2.2 Promiskuität von Wirkstoffen.....	76
3.2.3 Interaktive Netzwerkdarstellung hochgradig verknüpfter Daten.....	78
3.3 Fazit und Ausblick.....	79
<b>Verzeichnisse .....</b>	<b>81</b>
Literaturverzeichnis .....	81
Abbildungsverzeichnis .....	85
Abkürzungsverzeichnis.....	86
<b>Anhänge .....</b>	<b>87</b>
Anhang A Danksagung .....	87
Anhang B Lebenslauf .....	88
<b>Selbständigkeitserklärung.....</b>	<b>90</b>

# Zusammenfassung

Mit der Veröffentlichung der Sequenz des humanen Genoms im Jahr 2001 wurde die theoretische Grundlage menschlichen Lebens bekannt. Dadurch konnten die in dem Genom kodierten Anweisungen, nach denen die Ribonukleinsäuren und Proteine des menschlichen Körpers zusammengesetzt werden, studiert werden. Die Hoffnung, dass sich daraus die Funktionsweise des gesamten Organismus ableiten ließe, wurde aber enttäuscht. Dies liegt unter anderem daran, dass Proteine nicht unabhängig von ihrer Umgebung aktiv, sondern Teil eines komplexen molekularen Netzwerks sind. In diesem bestehen vielfältige Interaktionen zwischen Proteinen, anderen Makromolekülen und Kleinstrukturen. Diese Wechselwirkungen ermöglichen überhaupt erst die komplexen Vorgänge, die in biologischen Organismen stattfinden.

Die vorliegende Arbeit untersucht dieses Interaktionsnetzwerk, sowohl auf molekularer Ebene als auch mit Hilfe netzwerkbasierter Ansätze. Dazu werden zunächst geeignete Datensätze für die Analyse von Protein-Protein-Interaktionen einerseits und von Protein-Wirkstoff-Interaktionen andererseits erstellt. Die Protein-Protein-Interaktionen werden auf Charakteristika untersucht, die Hinweise geben können, welche Faktoren entscheidend für die Interaktionen sind. Die dabei gewonnenen Erkenntnisse werden zum Beispiel zur Vorhersage von molekularen Wechselwirkungen verwendet. Die Protein-Wirkstoff-Interaktionen werden als Teil des molekularen Interaktionsnetzwerks von einem netzwerkbasierten Standpunkt aus analysiert. Dabei steht vor allem die Frage im Vordergrund, inwiefern promiskuitive und nicht-promiskuitive Wirkstoffe beziehungsweise deren Zielproteine sich unterscheiden. Promiskuität von Wirkstoffen kann zu unerwünschten Nebenwirkungen führen, daher sind diese Unterschiede als mögliche Ursachen der Promiskuität bei der Entwicklung neuer Wirkstoffe von großem Interesse.

Diese Arbeit leistet einen Beitrag zu einem umfassenden Verständnis molekularer Interaktionen. Bei der Erstellung der dafür nötigen Datensätze werden verschiedenartige Daten aus unterschiedlichen Quellen miteinander verknüpft. Daher ist ein wichtiger Aspekt der Arbeit, wie relevante Informationen erschlossen und integriert werden können. In diesem Zusammenhang spielt auch eine zweckmäßige Präsentation und Archivierung der gewonnenen Daten eine Rolle. Diese Aspekte werden im Zuge des technologischen Fortschritts in den Lebenswissenschaften eine immer größere Rolle spielen, da umfangreiche Datenmengen innerhalb kürzester Zeit erhoben werden können.

# Abstract

Publication of the preliminary sequence of the human genome in 2001 revealed the theoretical foundations of human life. Since then, the genetic instructions used to build ribonucleic acids and proteins in the human body can be studied. However, hopes that this will unveil how the whole organism works were disappointed. Part of the problem is that proteins are not active independent of their environment, but as parts of a complex molecular network. Within this network many interactions occur between proteins, other macromolecules and small compounds. These interactions enable the organism to fulfil a wide variety of complex tasks with a rather limited set of proteins.

This thesis analyses this interaction network, both on a molecular level as well as by applying network-based approaches. For this purpose, well suited datasets are created in order to analyse protein-protein interactions on the one hand and interactions between proteins and active substances on the other hand. The protein-protein interactions are examined for characteristics that can indicate which features are crucial for them. The knowledge gained is used for example to predict molecular interactions. The interactions between proteins and active substances are analysed as parts of the molecular interaction network using network-based approaches. The key question tackled in this context is how promiscuous and non-promiscuous compounds and their respective targets differ. Promiscuity of active substances can lead to adverse reactions; therefore, these differences that may cause promiscuity are of great interest in drug design.

This work contributes to a comprehensive understanding of molecular interactions. In order to create the necessary datasets data from heterogeneous sources are combined. Therefore, one important aspect of this work is how relevant information can be accessed and integrated. In this context presenting and storing the data in a convenient way is important as well. These aspects are becoming increasingly important due to the ongoing technological advancement in the life sciences that allows for obtaining greater datasets with less effort.

# Eigene Publikationen

## Veröffentlichte Arbeiten:

### Publikation 1:

GÜNTHER S, VON EICHBORN J, MAY P, PREISSNER R

JAIL: a structure-based interface library for macromolecules.

*Nucleic Acids Research* 37, D338-341, 2009

Eigener Beitrag: Mitarbeit bei der Konzeption des Projekts; Mitarbeit beim Erstellen des Datenbankschemas; Mitarbeit bei der Entwicklung der Programme zur Generierung des Datensatzes; Automatisierung der Prozesse; Entwicklung der Weboberfläche

### Publikation 2:

VON EICHBORN J, GÜNTHER S, PREISSNER R

Structural features and evolution of protein-protein interactions.

*Genome Informatics* 22, 1-10, 2010

Eigener Beitrag: Konzeption des Projekts; Durchführung der Versuche; Analyse der Ergebnisse; Verfassen der Publikation

### Publikation 3:

VON EICHBORN J, MURGUEITIO M S, DUNKEL M, KOERNER S, BOURNE P E, PREISSNER R

PROMISCUOUS: a database for network-based drug-repositioning.

*Nucleic Acids Research* 39, D1060-D1066, 2011

Eigener Beitrag: Konzeption des Projektes; Erstellen des Datenbankschemas; Aufsetzen eines Systems zur manuellen Validierung der Ergebnisse des automatischen Textminings; Mitarbeit bei dieser manuellen Validierung; Vervollständigen des Datensatzes mit Daten aus anderen Quellen / Metadaten; Verfassen der Publikation

**Publikation 4:**

VON EICHBORN J, BOURNE P E, PREISSNER R

Cobweb: a Java applet for network exploration and visualisation.

*Bioinformatics* 27,1725-1726, 2011

Eigener Beitrag: Konzeption des Projekts; Programmierung der Applikation; Erstellen der Web-  
oberfläche; Verfassen der Publikation

**Publikation 5:**

WOELKE A L, VON EICHBORN J, MURGUEITIO M S, WORTH C L, CASTIGLIONE F, PREISSNER R

Development of Immune-Specific Interaction Potentials and Their Application in the Multi-Agent-  
System Vacclmm.

*PLoS ONE* 6, e23257, 2011

Eigener Beitrag: Mitarbeit bei der Konzeption des Projekts; Mitarbeit bei der Programmierung des  
agentenbasierten Modells; Mitarbeit bei der Durchführung der Versuche; Mitarbeit bei der Analy-  
se der Ergebnisse; Mitarbeit beim Verfassen der Publikation

**Publikation 6:**

HECKER N, AHMED J, VON EICHBORN J, DUNKEL M, MACHA K, ECKERT A, GILSON M K, BOURNE P E,  
PREISSNER R

SuperTarget goes quantitative: update on drug-target interactions.

*Nucleic Acids Research* 40, D1113-D1117, 2012

Eigener Beitrag: Mitarbeit beim Erstellen des Datensatzes; Erstellen der Weboberfläche

.....

**Eingereichte Manuskripte:**



**Eingereichtes Manuskript 1:**

von Eichborn J, Dunkel M, Gohlke B O, Preissner S C, Hoffmann M, Bauer J M J, Armstrong J D, Schaefer M H, Andrade-Navarro M A, LeNovere N, Croning M D R, Grant S G N, van Nierop P, Smit A B, Preissner R

SynSysNet: Integration of experimental data on synaptic protein-protein interactions with drug-target relations

*Nucleic Acids Research* (submitted)

Eigener Beitrag: Mitarbeit beim Erstellen des Datensatzes; Erstellen der Weboberfläche; Erstellen der Visualisierung; Mitarbeit beim Verfassen der Publikation

**Eingereichtes Manuskript 2:**

von Eichborn J, Woelke A L, Castiglione F, Preissner R

VacImm: Simulating peptide vaccination in cancer therapy

*BMC Bioinformatics* (submitted)

Eigener Beitrag: Konzeption des Projekts; Mitarbeit bei der Programmierung des agentenbasierten Modells; Mitarbeit bei der Durchführung der Versuche; Mitarbeit bei der Analyse der Ergebnisse; Erstellen des Webservers; Verfassen der Publikation

**Eingereichtes Manuskript 3:**

von Eichborn J, Preissner R

Polypharmacology of Drugs: Analysing Molecular and Network Properties

*PLoS ONE* (submitted)

Eigener Beitrag: Konzeption des Projekts; Durchführen der Versuche; Analyse der Ergebnisse; Verfassen der Publikation

# Kapitel 1 Einleitung

Gregor Mendel entdeckte schon Mitte des 19. Jahrhunderts bei Versuchen mit Pflanzen grundlegende Regeln der Vererbung [1]. Er konnte zeigen, dass einzelne Merkmale eines Lebewesens unabhängig voneinander vererbt werden können, womit er die Grundlage der heutigen Genetik geschaffen hat. Außerdem konnte er bereits dominante und rezessive Merkmale unterscheiden. Damit war ein entscheidender Grundstein für das Verständnis der Vorgänge der Vererbung und der Determination des Phänotyps durch den Genotyp gelegt.

Nachdem die von ihm aufgestellten Mendelschen Regeln zunächst wenig Beachtung fanden, wurden sie Anfang des 20. Jahrhunderts wiederentdeckt [2]. In den folgenden Jahren wurde die Theorie der Vererbung weiter ausgearbeitet; jedoch war die Erbinformation ein abstrakter Begriff, der keine physische Repräsentation hatte. Es herrschte sogar Uneinigkeit darüber, ob eine solche Repräsentation überhaupt existiert. Erst 1944 gelang es Avery *et al.* zu zeigen, dass die Desoxyribonukleinsäure (DNA) der Träger der Erbinformation ist [3].

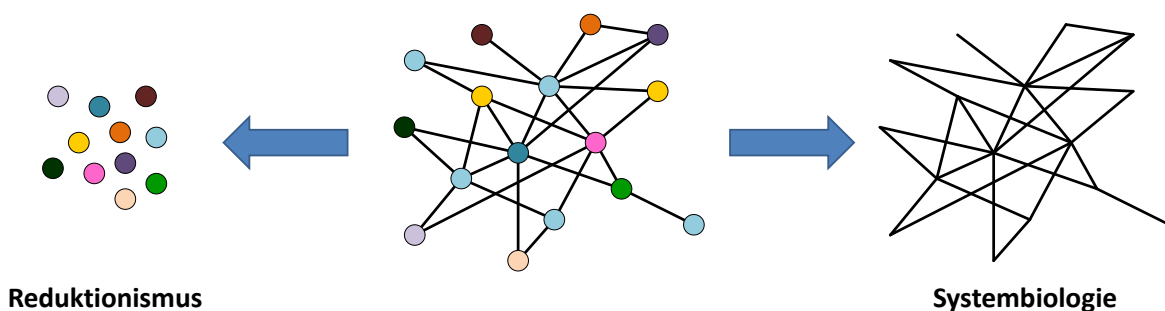
Mit der DNA war somit der Ort bekannt, an dem die Erbinformation einer Zelle gespeichert ist. Zu dieser Zeit wurde davon ausgegangen, dass ein Gen ein Enzym kodiert. Deshalb dachte man die Funktionsweise des ganzen Organismus verstehen zu können, wenn man seine DNA auslesen könnte. Man versprach sich dadurch unter anderem in die Lage versetzt zu werden, bisher kaum therapierbare Krankheiten heilen zu können. Im Zuge biotechnologischer Fortschritte wurde es Ende der 70er Jahre des letzten Jahrhunderts möglich, DNA zu sequenzieren [4]. Im Jahr 1990 formierte sich daraufhin ein internationales Konsortium, um im Rahmen des Humangenomprojekts das komplette menschliche Genom zu sequenzieren; 1998 startete ein privatwirtschaftliches Projekt mit demselben Ziel. Im Jahr 2001 fanden die Projekte ihre Höhepunkte in der gleichzeitigen Vorstellung der jeweils erstellten vorläufigen Version der Sequenz des menschlichen Genoms [5, 6].

Mittlerweile sind die Genome von mehr als 1150 Organismen vollständig sequenziert. Allerdings hat sich die Hoffnung, dadurch die biologischen Vorgänge in den Organismen verstehen zu können, als zu optimistisch herausgestellt. Dies liegt daran, dass die genetischen Daten nur die Grundlage für ein Verständnis des Lebens liefern. Zum einen lassen sich die genetischen Informationen nicht einfach in Proteine übersetzen, der Zusammenhang zwischen Genen und Proteinen ist komplexer. Ein Gen kann ein Polypeptid kodieren, durch alternatives Spleißen aber auch mehrere verschiedene Polypeptide. Ebenso kann es sein, dass ein Gen zwar in eine Ribonukleinsäure

(RNA) transkribiert wird, diese aber nicht zu einem Polypeptid translatiert wird. Diese RNAs können dann zum Beispiel entweder als Ribozym selber enzymatisch aktiv sein oder als miRNA die Genexpression regulieren. Zum anderen sind selbst für bekannte Proteine nicht immer auch ihre Aufgaben bekannt. Daher sind viele Genome zwar ausgelesen, aber bei weitem noch nicht dechiffriert.

Aber selbst wenn allen Proteinen Funktionen zugewiesen werden könnten, könnte daraus noch nicht die Funktion des gesamten Organismus abgeleitet werden [7]. Dies liegt daran, dass Proteine nicht komplett eigenständige Einheiten sind, die unabhängig von ihrer Umgebung aktiv sind. Vielmehr besteht innerhalb einer Zelle ein hochkomplexes molekulares Netzwerk, in dem Proteine, Kleinstrukturen, miRNAs und andere Faktoren miteinander interagieren und sich gegenseitig regulieren. Nur durch die hohe Anzahl an Interaktionen innerhalb dieses Netzwerks ist es möglich, mit einem relativ beschränkten Satz an Proteinen die vielfältigen Aufgaben des zellulären Stoffwechsels zu bewältigen.

Die Erkenntnis, dass es nicht reicht, einzelne Elemente isoliert zu untersuchen um zu verstehen, wie ein gesamter Organismus funktioniert, führte zu dem Aufschwung der Systembiologie. Diese begreift biologische Vorgänge als komplexe Systeme, deren Ganzes mehr als die Summe der Teile ist. Sie stellt daher nicht die einzelnen Komponenten des molekularen Netzwerks in den Mittelpunkt, sondern die Interaktionen, die zwischen diesen Elementen bestehen (Abbildung 1.1) [8, 9]. Dabei ist es das Ziel der Systembiologie, Modelle zu erstellen, die die Vorgänge in biologischen Systemen realistisch abbilden. Um ein solches Modell erstellen zu können, müssen sowohl die beteiligten Komponenten als auch die Interaktionen zwischen diesen Komponenten bekannt sein.



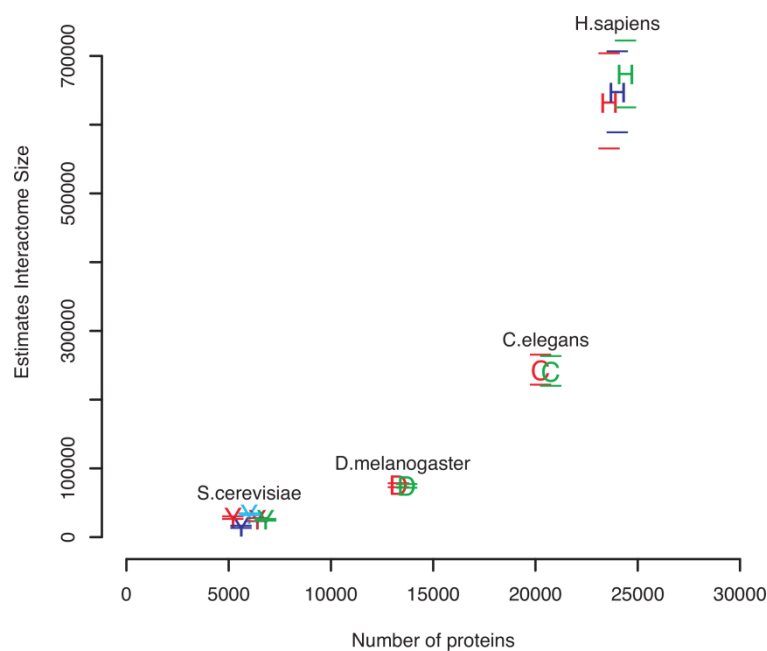
**Abbildung 1.1:** Illustration des Grundgedankens der Systembiologie. Anstatt das Netzwerk auf seine einzelnen Komponenten zu reduzieren und diese zu untersuchen, werden die Interaktionen im Netzwerk in den Mittelpunkt gestellt. (Abbildung modifiziert nach [10])

## 1.1 Protein-Protein-Interaktionen

Proteine erfüllen in Zellen vielfältige, teilweise sehr unterschiedliche Aufgaben. Sie bestimmen die Struktur der Zelle, regulieren Stoffwechselvorgänge, transportieren andere Moleküle und kataly-

sieren als Enzyme zahlreiche Reaktionen, um nur einige Funktionen zu nennen. Allerdings können Proteine ihre Aufgaben nicht als alleinstehende Komponenten erfüllen, sondern nur als Teile eines komplexen Netzwerks miteinander interagierender Proteine. Daher sind diese Interaktionen zwischen Proteinen von entscheidender Bedeutung für die Funktionsweise jedes Organismus.

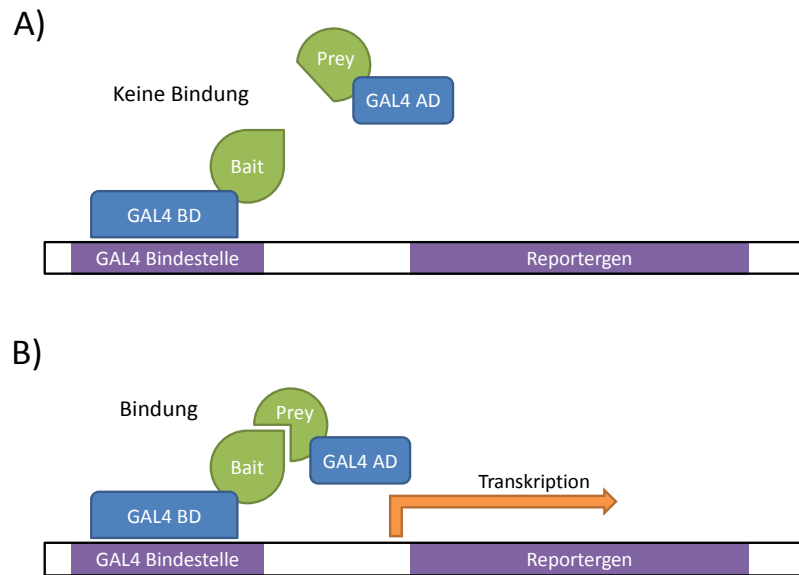
Einen Eindruck von der Komplexität des Protein-Protein-Interaktionsnetzwerks in Zellen vermittelt eine Studie aus dem Jahr 2008. Darin wird die Gesamtmenge der Protein-Protein-Interaktionen im Menschen, das so genannte Interaktom, auf circa 650.000 Interaktionen zwischen 25.000 Proteinen geschätzt (Abbildung 1.2) [11]. Diese Zahlen verdeutlichen zum einen, dass Interaktionen zwischen Proteinen sehr häufig vorkommen. Im Schnitt interagiert ein Protein der Schätzung zu Folge mit 26 anderen Proteinen. Zum anderen zeigen sie aber auch, wie spezifisch Protein-Protein-Interaktionen sind; theoretisch wären weitaus mehr als 650.000 Interaktionen zwischen den Proteinen möglich. Allerdings ist nur ein Bruchteil der im Interaktom vermuteten Interaktionen bekannt, laut einer anderen Arbeit aus dem Jahr 2008 waren es zu diesem Zeitpunkt weniger als 0,3 % [12].



**Abbildung 1.2:** Geschätzte Größe der Protein-Protein-Interaktionsnetzwerke für vier eukaryotische Organismen. Die horizontalen Striche geben die Größe des 95 % Konfidenzintervalls an. Die unterschiedlichen Farben der Buchstaben stehen für unterschiedliche Datensätze, die für die Schätzungen verwendet wurden. **Abbildung übernommen aus [11].**

Die Zahl der bekannten Interaktionen ist deshalb so gering, weil Protein-Protein-Interaktionen relativ aufwändig mit experimentellen Verfahren, wie zum Beispiel dem Hefe-Zwei-Hybrid-System (Yeast-Two-Hybrid, Abbildung 1.3), einzeln nachgewiesen werden müssen. Bei diesem Verfahren wird an zwei Proteine (bezeichnet als Bait und Prey) jeweils eine Domäne eines Transkriptionsfaktors gehängt. Jede Domäne für sich ist dabei nicht in der Lage, die Transkription einzuleiten, son-

dern nur beide Domänen gemeinsam. Wenn die beiden Proteine miteinander interagieren, kommen dadurch die angehängten Domänen des Transkriptionsfaktors in räumliche Nähe zueinander und können die Transkription einleiten. Anhand des Vorkommens des Transkriptionsprodukts kann schließlich ermittelt werden, ob die Proteine interagieren oder nicht. Diese experimentelle Methode ist somit in der Lage festzustellen, ob eine Interaktion zwischen zwei Proteinen stattfindet oder nicht. Sie liefert aber keine Informationen darüber, welche Bereiche der Proteinoberflächen sich bei der Interaktion miteinander in Kontakt befinden.

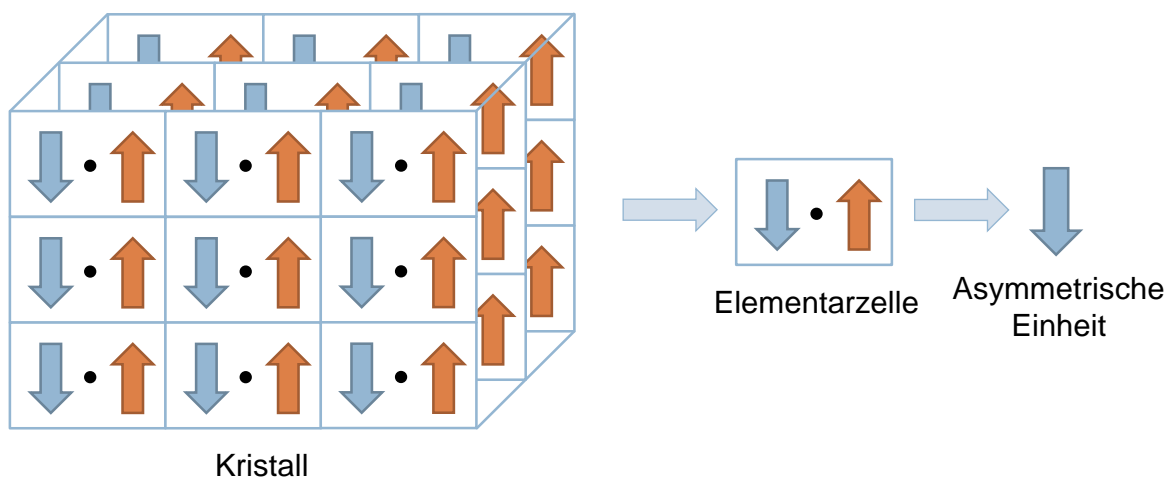


**Abbildung 1.3:** Schema des Hefe-Zwei-Hybrid-Systems am Beispiel des Transkriptionsfaktors GAL4. Die Bindedomäne (BD) des Transkriptionsfaktors bindet an die Bindestelle auf der DNA. A) Bait- und Prey-Protein interagieren nicht miteinander, dadurch kommt die Aktivierungsdomäne (AD) des Transkriptionsfaktors nicht in räumliche Nähe der Bindedomäne, es findet deshalb keine Transkription des Reportergens statt. B) Bait- und Prey-Protein interagieren miteinander, dadurch gerät die Aktivierungsdomäne in räumliche Nähe der Bindedomäne. Der so entstehende Komplex leitet die Transkription ein, das transkribierte Reportergen kann zum Nachweis der Bindung benutzt werden.

Diese Information kann beispielsweise gewonnen werden, indem die dreidimensionale (3D) Struktur des Proteinkomplexes untersucht wird. Um die 3D-Struktur eines Proteinkomplexes auf experimentellem Wege zu ermitteln sind jedoch noch aufwändigere und langwierigere Experimente notwendig. Die am häufigsten zur Aufklärung von Proteinstrukturen eingesetzte Technik ist die Röntgenkristallographie, bei der die Struktur mit Hilfe der Beugungsmuster von Röntgenstrahlen bestimmt wird. Dazu wird das zu untersuchende Protein zunächst aufgereinigt, um anschließend einen Kristall dieses Proteins zu züchten. Wenn es gelingt, einen brauchbaren Proteinkristall zu gewinnen, so besteht dieser aus vielen identischen so genannten Elementarzellen (Abbildung 1.4). Jede Elementarzelle beinhaltet jeweils eine oder mehrere asymmetrische Einheiten. Eine asymmetrische Einheit ist die kleinste Einheit einer Elementarzelle, aus der sich durch Symmetrieeoperationen der komplette Inhalt der Elementarzelle und damit auch der komplette Kristall rekonstruieren lassen. Im einfachsten Falle enthält die asymmetrische Einheit genau eine Kopie des

untersuchten Proteins und ist somit identisch mit der biologischen Einheit, die dem funktionalen Protein im Organismus entspricht. Allerdings kann die asymmetrische Einheit auch aus mehreren Kopien des Proteins oder nur aus Fragmenten des Proteins bestehen.

Der gezüchtete Kristall wird mit Röntgenstrahlen beschossen, dabei lenken die enthaltenen Atome die Strahlen ab und erzeugen so ein spezifisches Beugungsmuster. Dieses Beugungsmuster wird aufgezeichnet; mit Hilfe von Fourier-Transformationen kann hieraus anschließend die Verteilung der Elektronendichte in den Elementarzellen berechnet werden. Anhand dieser Karte der Elektronendichten kann schließlich ein Modell der Struktur der Asymmetrischen Einheit erstellt werden.



**Abbildung 1.4:** Schematische Skizze des Aufbaus eines Proteinkristalls. Der Kristall lässt sich durch Translationsoperationen aus der Elementarzelle rekonstruieren. Der Inhalt der Elementarzelle wiederum kann durch Symmetrieeoperationen auf die asymmetrische Einheit reduziert werden. Somit kann ausgehend von der asymmetrischen Einheit die komplette Kristallstruktur berechnet werden. Abbildung modifiziert nach [www.rcsb.org/pdb/101/static101.do?p=education\\_discussion/Looking-at-Structures/bioassembly\\_tutorial.html](http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/bioassembly_tutorial.html).

Die Herstellung eines Proteinkristalls erfordert viel Erfahrung des Experimentators und kann viel Zeit in Anspruch nehmen. Proteinkomplexe zu kristallisieren ist oft besonders schwierig, da sie meist heterogener und instabiler als ungebundene Proteine sind. Daher können auf diesem Weg nur in beschränktem Umfang Informationen über Komplexkonformationen von Proteinen gewonnen werden. Eine Alternative hierzu ist, die 3D-Strukturen der einzelnen Proteine im ungebundenen Zustand zu ermitteln. Anschließend kann am Rechner ein Modell erstellt werden, wie die Proteine aneinander binden. Dieser Ansatz wird beim Protein-Protein-Docking verfolgt [13–15].

Um beim Protein-Protein Docking bestimmen zu können, wie zwei Proteine miteinander interagieren, ist ein Verständnis der Eigenschaften von Protein-Interaktionsflächen notwendig. Eine Grundvoraussetzung für eine Interaktion ist, dass die Interaktionsflächen beider Proteine geometrisch komplementär zueinander sind. Darüber hinaus können bei der Suche nach der richtigen

Bindestelle noch weitere Kriterien wie z.B. Hydrophobizität und elektrostatische Eigenschaften berücksichtigt werden [14].

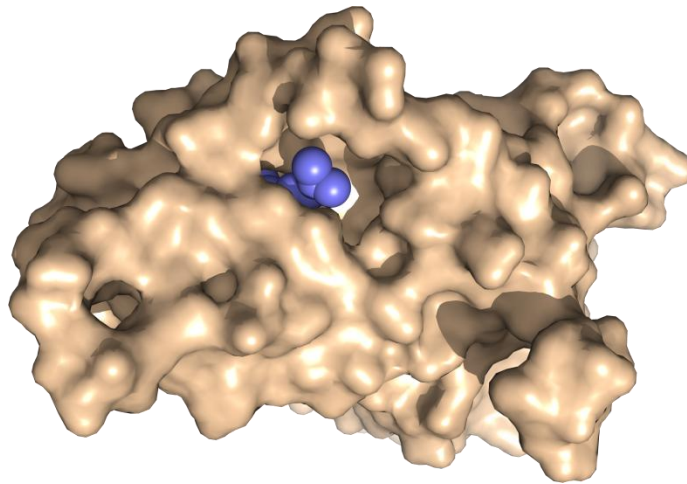
Um die Eigenschaften von Protein-Interaktionsflächen analysieren zu können, wurde im Rahmen dieser Arbeit eine Bibliothek bekannter Interaktionsflächen aufgebaut. Diese in Abschnitt 2.1.1 vorgestellte Bibliothek beinhaltet einen umfassenden Datensatz an Interaktionsflächen verschiedenen Typs, basierend auf allen in der Protein Data Bank (PDB) [16] vorhandenen 3D-Strukturen. Dieser Datensatz wurde analysiert und die Interaktionsflächen auf spezifische Eigenschaften hin untersucht. In der in Abschnitt 2.1.2 vorgestellten Analyse wurden insbesondere transiente und obligate Interaktionen miteinander verglichen.

Eine praktische Anwendung des gewonnen Wissens über Protein-Interaktionsflächen wird in Abschnitt 2.1.3 vorgestellt. Die ermittelten Bindungspotentiale werden hierbei dazu benutzt, molekulare Bindungen vorherzusagen.

## 1.2 Protein-Wirkstoff-Interaktionen

Neben Protein-Protein-Interaktionen spielen Interaktionen zwischen Proteinen und Kleinstrukturen eine wichtige Rolle im molekularen Interaktionsnetzwerk. Die Kleinstrukturen nehmen dabei vielfältige Funktionen wahr, beispielsweise können sie als Botenstoff zur Signaltransduktion oder als Substrat für eine enzymatische Reaktion dienen. Speziell ausgewählte Kleinstrukturen können auch dazu benutzt werden, als Arzneistoff gezielt bestimmte Proteine zu aktivieren oder inhibieren und dadurch eine therapeutische Wirkung zu erzielen. Klassischerweise werden diese Interaktionen von Arzneistoffen und Proteinen als hochspezifisch und hochselektiv angesehen. Dieser Sichtweise zufolge interagiert ein Arzneistoff spezifisch mit nur einem Protein, seinem Zielprotein. Diese Art der Interaktion wird häufig als Schlüssel-Schloss-Prinzip bezeichnet, da ein „Schlüssel“ (der Arzneistoff) nur spezifisch zu einem „Schloss“ (dem Zielprotein) passt (Abbildung 1.5).

In der Realität finden diese Interaktionen aber nicht immer so spezifisch statt. Ein bekanntes Beispiel hierfür ist das weit verbreitete Arzneimittel Aspirin, das eine entzündungshemmende und schmerzstillende Wirkung hat [17: S. 235]. Sein Wirkstoff Acetylsalicylsäure inhibiert die Enzyme Cyclooxygenase 1 und Cyclooxygenase 2; beide Enzyme katalysieren die gleiche, entscheidende Reaktion in der Prostaglandinsynthese. Prostaglandine und daraus abgeleitete Thromboxane erfüllen vielfältige Aufgaben im menschlichen Körper. Sie sind unter anderem bei Entzündungsreaktionen von Bedeutung; deshalb erzielt Acetylsalicylsäure seinen therapeutischen Effekt durch die Inhibition der Prostaglandinsynthese durch die genannten Cyclooxygenasen. Allerdings sind



**Abbildung 1.5:** Spezifische Bindung des Arzneistoffs Tamoxifen (blau) an sein Zielprotein, den humanen Estrogenrezeptor (gelb). Tamoxifen bindet an den Rezeptor, indem es die Form seines Liganden Estrogen nachahmt. Durch die Bindung wird der Rezeptor in eine inaktive Konformation gebracht und somit inhibiert. Der Großteil des Tamoxifens ist nicht sichtbar, da er tief in die Bindetasche des Proteins hineinragt. (PDB-Id 3ERT, visualisiert mit PyMOL [18]).

Prostaglandine und Thromboxane auch bei der Sekretion des Magensaftes beteiligt. Daher kommt es durch die Einnahme von Acetylsalicylsäure zu unerwünschten Nebenwirkungen wie zum Beispiel Magenblutungen. Diese ließen sich aber durchaus vermeiden: Cyclooxygenase 1 und Cyclooxygenase 2 katalysieren zwar die gleiche Reaktion, aber sie unterscheiden sich in ihren Expressionsprofilen. Cyclooxygenase 1 wird konstitutiv exprimiert, Cyclooxygenase 2 hingegen wird als Reaktion auf Entzündungen und Verletzungen lokal verstärkt exprimiert. Ein spezifischer Hemmer der Cyclooxygenase 2 könnte deshalb einen Großteil des therapeutischen Effekts der Acetylsalicylsäure erreichen und gleichzeitig die Nebenwirkungen deutlich abschwächen.

Proteine wie die Cyclooxygenase 1, die nicht die eigentlichen Zielproteine (engl. Targets) eines Arzneistoffs sind, aber dennoch von ihm adressiert werden, werden als Off-Targets bezeichnet. Untersuchungen haben gezeigt, dass ein Großteil der Arzneistoffe mehr als ein Protein adressiert [19]. Dieses als promiskuitiv bezeichnete Verhalten der Arzneistoffe führt häufig zu unerwünschten Nebenwirkungen, wie bei dem im vorigen Absatz aufgeführten Beispiel Acetylsalicylsäure zu Magenblutungen. In einer systematischen Studie wurde gezeigt, dass das Wissen über Nebenwirkungen von Arzneimitteln dazu benutzt werden kann, neue Indikationsgebiete für bekannte Arzneimittel zu erschließen [20]. Bekannte Arzneistoffe für neue Indikationsgebiete anzuwenden ist attraktiv, da die Stoffe für ihr ursprüngliches Anwendungsgebiet schon ausführlich getestet wurden. Somit sind weniger aufwändige, langwierige und potentiell gefährliche klinische Studien nötig. Ein bekanntes Beispiel hierfür ist Sildenafil, das ursprünglich als Arzneistoff gegen Bluthochdruck und Angina pectoris konzipiert war [21: S. 94]. Während der klinischen Studien wurde festgestellt, dass Sildenafil gegen Angina pectoris nicht wirksam war. Allerdings zeigte sich, dass der Wirkstoff bei erektiler Dysfunktion half. Daraufhin wurde er unter dem Namen Viagra zunächst



zur Behandlung der erektilen Dysfunktion zugelassen; mittlerweile ist er unter der Bezeichnung Revatio auch als Medikament gegen pulmonal-arterielle Hypertonie erhältlich.

Um die promiskuitive Natur von Wirkstoffen zu untersuchen ist es wichtig, auf ein möglichst umfangreiches Wissen über Protein-Wirkstoff- und Protein-Protein-Interaktionen zurückgreifen zu können. Es gibt zwar viele Daten zu Protein-Wirkstoff-Interaktionen, allerdings sind diese über viele Quellen verstreut. Viele Informationen stehen obendrein nur in Textform in wissenschaftlichen Publikationen und sind daher nicht direkt abrufbar.

Daher wurde eine umfassende Datenbank der relevanten Interaktionen erstellt, um eine Analyse des promiskuitiven Verhaltens von Wirkstoffen zu ermöglichen; diese Datenbank wird in Abschnitt 2.2.1 vorgestellt. Die gewonnenen Daten wurden anschließend analysiert, wobei das Augenmerk besonders auf die Unterschiede von promiskuitiven und nicht-promiskuitiven Wirkstoffen gerichtet war (Abschnitt 2.2.2).

Die Daten über Protein-Protein- und Protein-Wirkstoff-Interaktionen bilden, wie bereits beschrieben, ein großes molekulares Netzwerk. Um dieser Netzwerkstruktur der Daten gerecht zu werden, wurde eine interaktive Visualisierung der Daten entwickelt. Diese ermöglicht die Darstellung der Interaktionen als Netzwerk an Stelle von Tabellen mit Interaktionsinformationen. Das in Abschnitt 2.2.3 vorgestellte Programm ist frei verfügbar und kann auch von Dritten für andere Projekte eingesetzt werden.

Ein weiterer Schritt in der Analyse der Protein-Wirkstoff-Interaktionen ist die Berücksichtigung quantitativer Daten. Um eine solche zu ermöglichen, wurde eine Protein-Wirkstoff Datenbank erstellt, in der Bindungsaffinitäten mit einbezogen sind (Abschnitt 2.2.4).

## **1.3 Zielsetzung der Arbeit**

Das Hauptanliegen der vorliegenden Arbeit ist es, zum Verständnis der Eigenschaften molekularer Interaktionen beizutragen. Deswegen werden zum einen Protein-Protein-Interaktionen und zum anderen Interaktionen zwischen Proteinen und Wirkstoffen behandelt.

In beiden Teilen sollten bioinformatische Methoden entwickelt werden, die es erlauben, Informationen über die genannten Interaktionsarten zu gewinnen und zu verarbeiten. Dies erfolgt entweder auf Grundlage bekannter 3D-Strukturen oder über die halbautomatisierte Analyse wissenschaftlicher Veröffentlichungen. Die ermittelten Informationen sollten in Datenbanken strukturiert abgelegt werden, so dass ein komfortabler Zugriff auf die Daten möglich ist. Dabei sollten die Daten über Protein-Wirkstoff-Interaktionen in den zellulären Kontext der Proteine eingebettet

werden, um die Effekte der Interaktionen im zellulären Zusammenhang verstehen zu können. Um die Daten der Öffentlichkeit zugänglich zu machen, sollten jeweils Weboberflächen erstellt werden, die freien Zugang zu den Informationen gewähren. Aufbauend auf den Daten sollten Eigenschaften der Interaktionen analysiert werden, um das in den Datensammlungen enthaltene Wissen nutzbar zu machen.

## Kapitel 2 Methoden und Ergebnisse

### 2.1 Protein-Protein-Interaktionen

#### 2.1.1 Erstellen einer Bibliothek von Makromolekül-Interaktionsflächen

Die Grundlage der strukturbasierten Analyse von Protein-Interaktionsflächen sind dreidimensionale Strukturen von Proteinen. Diese können mit experimentellen Methoden wie der Röntgenstrukturanalyse oder der Kernmagnetresonanzspektroskopie ermittelt werden. Mittlerweile sind über 80.000 experimentell bestimmte 3D-Strukturen in der Proteinstrukturdatenbank PDB [16] hinterlegt.

Dieser Datensatz beinhaltet aber zwei Limitierungen: Zum einen besitzt er eine relativ hohe Redundanz, da Proteinstrukturen häufig mehrfach ermittelt werden, zum Beispiel um die Bindung verschiedener Liganden an das gleiche Protein zu untersuchen. Zum anderen beschreiben die meisten 3D-Strukturen nur einzelne Proteine oder sogar nur Teile von Proteinen, da es experimentell deutlich schwieriger ist, die Struktur von Proteinkomplexen zu bestimmen. Deshalb ist der nutzbare Datensatz an Protein-Protein-Komplexstrukturen viel kleiner als der Gesamtdatensatz.

Diese Limitierungen können umgangen werden, indem auch Interaktionsflächen zwischen Proteindomänen betrachtet werden. Proteindomänen sind globuläre Untereinheiten innerhalb eines Proteins, die relativ unabhängig vom Rest des Proteins sind. So behalten sie häufig ihre Struktur auch dann bei, wenn sie vom Rest des Proteins getrennt werden. Außerdem erfüllen verschiedene Domänen innerhalb eines Proteins meist verschiedene Aufgaben: so kann zum Beispiel eine Domäne für die Bindung einer Kleinstruktur zuständig sein, und eine andere für die Interaktion mit anderen Proteinen. Größere Proteine bestehen in der Regel aus mehreren Domänen; diese Domänen besitzen dann untereinander Interaktionsflächen, die ähnliche Eigenschaften aufweisen wie Interaktionsflächen zwischen Proteinen [22].

Die hier vorgestellte Bibliothek von Makromolekül-Interaktionsflächen JAIL [23] basiert auf allen in der PDB vorhandenen 3D-Strukturen. Innerhalb dieser Strukturen wurde nach Interaktionsflächen gesucht. Dabei wurden neben Interaktionsflächen zwischen Proteindomänen auch Interaktionsflächen zwischen einzelnen Aminosäureketten eines Proteins und zwischen Proteinen und Nukleinsäureketten berücksichtigt. Außerdem wurden gegebenenfalls zusätzliche Interaktionen in den biologischen Einheiten aufgenommen, falls diese von den asymmetrischen Einheiten abwei-

chen. Somit stellt die Bibliothek einen umfassenden Datensatz über Interaktionsflächen verschiedenen Typs in bekannten 3D-Strukturen zur Verfügung.

Um die Daten der Öffentlichkeit zur Verfügung zu stellen, wurde eine Weboberfläche erstellt. Diese bietet eine Vielzahl von Möglichkeiten, um nach Interaktionsflächen zu suchen. Neben der Suche nach Interaktionen anhand bestimmter Stichwörter oder Kennungen, wie der Uniprot accession number, ist es auch möglich, alle Interaktionen zwischen bestimmten Domärentypen anzuzeigen. Zu jeder 3D-Struktur werden Metadaten angezeigt, zum Beispiel die Methode, mit der die 3D-Struktur ermittelt wurde und, soweit vorhanden, strukturelle Klassifikationen der enthaltenen Domänen. Daneben werden die gefundenen Interaktionsflächen aufgelistet, sofern welche vorhanden sind. Die Interaktionsflächen lassen sich entweder direkt auf der Seite anzeigen oder im PDB-Format herunterladen. Zusätzlich ist es möglich, repräsentative, nicht-redundante Datensätze von Interaktionsflächen zusammenzustellen und gesammelt herunterzuladen.

## **Originalartikel: "JAIL: a Structure-Based Interface Library for Macromolecules"**

GÜNTHER S, VON EICHBORN J, MAY P, PREISSNER R

*Nucleic Acids Research* 37, D338-341, 2009

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:

<http://dx.doi.org/10.1093/nar/gkn599>

## 2.1.2 Analyse von Protein-Interaktionsflächen

Die Analyse bekannter Protein-Interaktionsflächen kann Einblicke in die Mechanismen molekularer Erkennung bieten. Wenn Gesetzmäßigkeiten erkannt werden, können diese zum Beispiel bei der Vorhersage der Bindungskonformation beim Protein-Protein-Docking helfen. Daher wurden die Interaktionsflächen aus der im vorigen Abschnitt vorgestellten Bibliothek verwendet, um sie im Hinblick auf strukturelle und evolutionäre Eigenschaften zu analysieren. Es wurden zwei distinkte Datensätze gebildet, die beide auf Domänen-Interaktionen basieren. Der erste Datensatz besteht aus transienten Interaktionen zwischen Domänen verschiedener Proteine, wohingegen im zweiten nur obligate Interaktionen zwischen Domänen eines Multidomänenproteins enthalten sind.

Beide Datensätze wurden in Bezug auf die physikochemischen Eigenschaften und die evolutionäre Konservierung der Interaktionsflächen untersucht. Dabei wurden besonders die Eigenschaften der Residuen analysiert, die sich an den Interaktionsflächen gegenüberstehen, die also während der Interaktion miteinander in Kontakt kommen.

Bei der Analyse der physikochemischen Eigenschaften der gegenüberstehenden Residuen konnten Präferenzen beobachtet werden, die sowohl bei transienten wie auch bei obligaten Interaktionen zuträfen. Werden die verschiedenen Aminosäuren in basische, saure, polare ungeladene und nicht polare eingeteilt, so ergeben sich klare Kontaktmuster: nicht polare Residuen stehen bevorzugt mit anderen nicht polaren in Kontakt, und basische bevorzugt mit sauren. Dahingegen kommen Kontakte zwischen zwei sauren oder zwei basischen Aminosäuren unterdurchschnittlich häufig vor. Neben diesem generellen Verhalten gibt es noch weitere auffällige Präferenzen, die sich durch besondere Eigenschaften der beteiligten Residuen erklären lassen. So stehen die aromatischen Aminosäuren Tyrosin, Phenylalanin und Tryptophan bevorzugt in Kontakt mit anderen aromatischen Aminosäuren. Diese individuellen Präferenzen sind bei obligaten und transienten Interaktionen unterschiedlich stark ausgeprägt.

Der zweite Teil der Analyse untersucht die evolutionären Eigenschaften der Interaktionsflächen. Basierend auf denselben Datensätzen wurde zunächst festgestellt, dass die Kontaktflächen im Durchschnitt stärker konserviert sind als der Rest der Proteinoberflächen. Diese Unterschiede in der Konservierung sind bei obligaten Interaktionen stärker ausgeprägt als bei transienten. Anschließend wurde die Stärke der Konservierung bei einander an den Kontaktflächen gegenüberstehenden Residuen verglichen. Dabei zeigte sich, dass Residuen präferiert in Kontakt mit anderen Residuen stehen, die ähnlich stark konserviert sind. Das trifft in besonderem Maße auf besonders stark und besonders schwach konservierte Residuen zu, ist aber bei allen Konservierungsstärken

zu beobachten. Dieses Muster tritt sowohl bei transienten als auch bei obligaten Interaktionen auf, ist bei obligaten aber ausgeprägter; dies wird im Abschnitt 3.1.2 dieser Arbeit diskutiert.

## **Originalartikel: "Structural Features and Evolution of Protein-Protein Interactions"**

VON EICHBORN J, GÜNTHER S, PREISSNER R

*Genome Informatics* 22, 1-10, 2010

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:

[http://dx.doi.org/10.1142/9781848165786\\_0001](http://dx.doi.org/10.1142/9781848165786_0001)



### 2.1.3 Erstellen wissensbasierter Interaktionspotentiale

Das Wissen um die Eigenschaften von Interaktionsflächen kann dazu benutzt werden, anhand bekannter Interaktionen Interaktionspotentiale zu ermitteln und damit zu berechnen, ob zwei Proteine in einer gegebenen Komplexkonformation miteinander interagieren. Diese Möglichkeit wurde im Rahmen eines agentenbasierten Modells zur Simulation der Peptidvakzinierung gegen Tumoren genutzt.

Konventionelle Methoden der Krebstherapie verursachen häufig schwere Nebenwirkungen; außerdem gelingt es oft nicht, den Tumor vollständig zu eradizieren. Deshalb besteht großes Interesse an Verbesserungen oder Alternativen wie der Immuntherapie. Sie basiert auf der Erkenntnis, dass das Immunsystem in der Lage ist, einen Tumor wirksam zu bekämpfen und zu vernichten, wenn es ihn erkennt [24]. Deshalb ist es das Ziel der Immuntherapie, das Immunsystem so zu stimulieren, dass eine Immunantwort gegen die Tumorzellen erfolgt. Dazu werden dem Patienten bestimmte Peptide injiziert, die das Immunsystems dazu anregen, gegen diese Antigene spezifische Effektor- und Gedächtniszellen zu bilden. Obwohl die Möglichkeiten der Immuntherapie zur Bekämpfung von Tumoren in den letzten Jahren immer besser untersucht wurden [25–27], sind viele Fragen, wie zum Beispiel der Dosierung und des optimalen zeitlichen Ablaufs der Vakzination noch offen. Diese in Experimenten zu klären ist sehr zeitaufwändig und beansprucht obendrein beachtliche finanzielle und personelle Ressourcen. Daher ist es sehr hilfreich, wenn die Experimente durch rechnergestützte Simulationen unterstützt werden können. Das hier vorgestellte agentenbasierten Modell soll diesen Weg eröffnen.

Als Simulationsmethode wurde die agentenbasierte Modellierung gewählt, da sie es erlaubt, viele autonome Einheiten zu simulieren, die miteinander interagieren können. Dabei können verschiedene Einheitentypen existieren, die jeweils bestimmte Eigenschaften und Verhaltensregeln haben. Diese Verhaltensregeln bestimmen, wie sich die Einheiten in diskreten Zeitschritten durch den simulierten Raum bewegen und miteinander interagieren [28].

Das neu entwickelte Modell VacImm basiert auf den etablierten C-ImmSim/ImmSim-Modellen zur Modellierung des Immunsystems [29, 30]. In diesen Modellen werden die verschiedenen Rezeptoren jeweils durch Ketten von Nullen und Einsen (Bitstrings) dargestellt, die keinen direkten Bezug zu den Aminosäuren in der biologischen Realität haben. Die spezifischen Interaktionen des adaptiven Immunsystems werden dabei über die Komplementarität der Bitstrings der Rezeptoren mit denen ihrer Interaktionspartner simuliert.

Als entscheidende Erweiterung wurde in Vacclmm daher die Fähigkeit eingebaut, diese Interaktionen basierend auf den tatsächlichen Aminosäuresequenzen der antigenpräsentierenden Proteine, der Rezeptoren und der Antigene vorherzusagen. Für diese Vorhersage wurden Potentiale für die Interaktionen zwischen den antigenpräsentierenden MHC (major histocompatibility complex)-Proteinen und den T-Zell-Rezeptoren beziehungsweise zwischen den Antigenen und den B-Zell-Rezeptoren des Immunsystems basierend auf bekannten 3D-Strukturen gebildet. Diese neuen Interaktionspotentiale sind zuverlässig in der Lage, reale Komplexe der Rezeptoren des Immunsystems mit ihren Interaktionspartnern von randomisierten Komplexen zu unterscheiden, daher sind sie für die Interaktionsvorhersage der Rezeptoren geeignet. Somit können bei der Simulation der Peptidvakzinierung mit Vacclmm die tatsächlichen Peptidsequenzen an Stelle artifizierlicher Bitstrings benutzt werden, um Interaktionen zu berechnen.

**Originalartikel: "Development of Immune-Specific Interaction Potentials and Their Application in the Multi-Agent-System Vacclmm"**

WOELKE A L, VON EICHBORN J, MURGUEITIO M S, WORTH C L, CASTIGLIONE F, PREISSNER R

*PLoS ONE* 6, e23257, 2011

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:

<http://dx.doi.org/10.1371/journal.pone.0023257>

## 2.2 Protein-Wirkstoff-Interaktionen

### 2.2.1 Erstellen einer Ressource für Protein-Wirkstoff-Interaktionen

In der Einleitung wurde bereits betont, wie wichtig es ist, die Wirkung von Wirkstoffen in ihrem zellulären Kontext zu sehen. Dafür ist es entscheidend, einen möglichst umfassenden Datensatz von Protein-Wirkstoff-Interaktionen zur Verfügung zu haben. Die Menge an Informationen zu Protein-Wirkstoff-Interaktionen nimmt zwar stetig zu, allerdings sind diese Informationen häufig nicht direkt verfügbar. Zum einen sind sie über sehr viele Quellen verteilt und zum anderen finden sie sich häufig nur in Textform in Publikationen. Diese Daten sind deshalb nicht automatisiert zugänglich und können nicht ohne weiteres verwendet werden.

Um dem abzuweichen wurde der hier vorgestellte PROMISCUOUS Datensatz [31] erstellt. Er umfasst 21.500 Interaktionen zwischen Proteinen und Wirkstoffen. Diese Daten wurden sowohl halbautomatisch aus Originalveröffentlichungen gewonnen, als auch aus den Datenbanken DrugBank [32], SuperTarget [33] und SuperCyp [34] übernommen.

Um den Kontext dieser Informationen mit abzubilden, wurden sie mit einer Reihe weiterer Daten verknüpft. Dies umfasst zum Beispiel Protein-Protein-Interaktionen und Informationen über Nebenwirkungen von Wirkstoffen als weitere Interaktionstypen. Wo vorhanden, wurden den Einträgen weitere Metadaten zugeordnet, so zum Beispiel ATC (Anatomisch Therapeutisch Chemische)-Klassifizierungen der Weltgesundheitsorganisation (WHO) [35] für Wirkstoffe und EC-Nummern [36] für Proteine. Um die in der Datenbank enthaltenen Proteine und Wirkstoffe in ihrem zellulären Kontext zu zeigen, wurden die Proteine außerdem Signal- und Stoffwechselwegen aus der KEGG Datenbank [37] zugeordnet und auf diesen abgebildet.

PROMISCUOUS stellt somit eine umfassende Datenbasis über Protein-Wirkstoff-Beziehungen in ihrem zellulären Kontext zur Verfügung. Die Daten sind über eine öffentlich zugängliche Weboberfläche zu erreichen. Diese bietet umfassende Möglichkeiten, um nach bestimmtem Wirkstoffen, Proteinen oder Signal- und Stoffwechselwegen zu suchen. Auf den entsprechenden Ergebnisseiten werden detaillierte Informationen und Metadaten zu den Suchergebnissen angezeigt, wie z.B. ATC-Klassifizierungen und bekannte Nebenwirkungen für Wirkstoffe. Auf den Ergebnisseiten für Signal- und Stoffwechselwege werden die in der Datenbank vorhandenen Proteine hervorgehoben und Informationen zu den jeweils bekannten Wirkstoffen angezeigt. Der Fokus von PROMISCUOUS liegt auf der Vermittlung des molekularen Netzwerks, in das jedes Element eingebettet ist. Deshalb wurde eine spezielle Anwendung entwickelt, um dieses Netzwerk zu visualisie-

ren. Diese Anwendung, die in Abschnitt 2.2.3 im Detail vorgestellt wird, ermöglicht eine schrittweise Exploration des Kontextes eines Wirkstoffes oder Proteins. Dazu können ausgehend von einem Subnetzwerk gezielt neue Teile des Netzwerks nachgeladen werden. Dieser Mechanismus wird in Abschnitt 2.2.3 genauer erläutert.

## **Originalartikel: "PROMISCUOUS: a Database for Network-Based Drug-Repositioning"**

VON EICHBORN J, MURGUEITIO M S, DUNKEL M, KOERNER S, BOURNE P E, PREISSNER R

*Nucleic Acids Research* 39, D1060-D1066, 2011

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:

<http://dx.doi.org/10.1093/nar/gkq1037>

## 2.2.2 Analyse promiskuitiver und nicht-promiskuitiver Wirkstoffe

Wie in der Einleitung geschildert, wurde in den letzten Jahren festgestellt, dass Wirkstoffe häufig nicht so spezifisch wirken, wie lange Zeit gedacht wurde. Das beobachtete promiskuitive Verhalten der Wirkstoffe kann zu Nebenwirkungen führen, die schlimmstenfalls den Einsatz der Wirkstoffe unmöglich machen können. Vor diesem Hintergrund wurde der im vorigen Abschnitt vorgestellte PROMISCUOUS Datensatz genutzt, um mögliche Ursachen der Promiskuität von Wirkstoffen zu untersuchen.

Dazu wurden die Wirkstoffe anhand der in dem Datensatz vorhandenen Protein-Wirkstoff-Interaktionen in promiskuitive und nicht-promiskuitive Wirkstoffe eingeteilt. Als nicht-promiskuitiv wurden Wirkstoffe mit nur einem bekannten Zielprotein klassifiziert, als promiskuitiv solche mit mehr als 3 bekannten Zielproteinen. Zusätzlich wurde aus der Datenbank DrugBank [32] eine Liste vom Markt zurückgezogener Wirkstoffe extrahiert, diese Wirkstoffe wurden zu einer dritten Gruppe zusammen gefasst.

Die so gebildeten Gruppen wurden zunächst im Hinblick auf die molekularen Eigenschaften der enthaltenen Wirkstoffe untersucht. Es zeigte sich, dass promiskuitive und nicht-promiskuitive Wirkstoffe ähnliche physikochemische Eigenschaften und Atomzusammensetzungen aufweisen; größere Unterschiede gibt es vor allem in der atomaren Zusammensetzung bei den Halogenen Brom, Chlor und Fluor sowie bei Schwefel. Zurückgezogene Wirkstoffe hingegen unterscheiden sich stärker von den beiden anderen Gruppen, besonders in Hinblick auf den Oktanol-Wasser-Verteilungskoeffizienten Log P und die Größe der polaren Oberfläche. Aber auch die Atomzusammensetzung unterscheidet sich hier stärker.

Die Unterschiede in der atomaren Zusammensetzung der Wirkstoffe legen nahe, dass sie sich auch in den vorhandenen funktionellen Gruppen unterscheiden. Daher wurden mehr als 300 verschiedene funktionelle Gruppen daraufhin untersucht, wie häufig sie in den Wirkstoffen auftreten. Am häufigsten kommen aromatische Ringe vor, die in ungefähr 65 % der promiskuitiven und nicht-promiskuitiven Wirkstoffe und sogar in 90 % der zurückgezogenen Wirkstoffe vorhanden sind. Auffällig waren die Häufigkeiten von Amid- und Amin-Gruppen, da Amide bevorzugt in nicht-promiskuitiven und Amine bevorzugt in promiskuitiven Wirkstoffen auftreten. Ihre Rolle in Protein-Wirkstoff Bindungen wurde genauer untersucht, dabei wurde festgestellt, dass Amide häufiger Wasserstoffbrückenbindungen zum Rückgrat des Proteins ausbilden als Amine.

Neben den molekularen Eigenschaften der Wirkstoffe wurden auch Netzwerkeigenschaften der Zielproteine im humanen Protein-Protein-Interaktionsnetzwerk untersucht. Dazu wurden Zentralitätsmaße für Zielproteine von nicht-promiskuitiven Wirkstoffen, von promiskuitiven Wirkstoffen und von zurückgezogenen Wirkstoffen sowie für Proteine, für die keine Wirkstoffe bekannt sind, berechnet. Die Ergebnisse zeigten, dass die Zielproteine von Wirkstoffen im Durchschnitt erheblich zentraler im Netzwerk positioniert sind als Proteine, die nicht von Wirkstoffen adressiert werden. Die Zielproteine nicht-promiskuitiver Wirkstoffe sind dabei nochmals zentraler als die promiskuitiver Wirkstoffe.

Des Weiteren wurde untersucht, ob sich verschiedene Klassen von Wirkstoffen in den Eigenschaften ihrer Zielproteine unterscheiden; dazu wurden die Wirkstoffe anhand ihrer ATC Klassifizierung gruppiert. Die Untersuchung ergab klare Präferenzen einiger Wirkstoffgruppen im Hinblick auf die Zentralität der Zielproteine im humanen Protein-Protein-Interaktionsnetzwerk.



## **Eingereichtes Manuskript: "Polypharmacology of Drugs: Analysing Molecular and Network Properties"**

VON EICHBORN J, PREISSNER R

Der vollständige Text ist in der beigelegten Datei „Anlage\_Analyse.pdf“ enthalten.

### 2.2.3 Interaktive Visualisierung komplexer Netzwerke

Mit der in Abschnitt 2.2.1 vorgestellten Datenbank PROMISCUOUS existiert eine Ressource mit komplexen, hochgradig miteinander verknüpften Informationen über Proteine, Wirkstoffe und Nebenwirkungen. Diese Daten können als Netzwerk begriffen werden, in dem die verschiedenen Entitäten (Proteine, Wirkstoffe und Nebenwirkungen) als Punkte dargestellt werden. Wenn zwischen zwei Punkten eine Beziehung besteht, es sich zum Beispiel um miteinander interagierende Proteine handelt, dann werden sie durch eine Linie miteinander verbunden. Die Punkte in dieser Netzwerkdarstellung werden als Knoten bezeichnet und die Linien zwischen ihnen als Kanten.

Die Darstellung eines kompletten Datensatzes als Netzwerk kann für kleine Datensätze übersichtlich gelingen. Bei einem großen Datensatz, wie dem von PROMISCUOUS mit vielen tausend Knoten und Kanten, wäre hingegen ein undurchschaubares Gewirr von Kanten das Ergebnis. Um dieses Problem bei der Visualisierung der in PROMISCUOUS gespeicherten Daten zu umgehen, wurde das Programm Cobweb [38] entwickelt. Dieses erlaubt, den Datensatz interaktiv zu visualisieren.

Dazu kann der Anwender nach einem bestimmten Eintrag in der Datenbank suchen und eine Netzwerkvisualisierung, ausgehend von diesem Eintrag, starten. Daraufhin wird zunächst nur der gewählte Eintrag und seine direkte Nachbarschaft angezeigt, das heißt ein Knoten, der den gefundenen Eintrag selbst repräsentiert, und andere Knoten, die über Kanten direkt mit dem Ursprungsknoten verbunden sind. In dieser Ansicht kann der Benutzer dann jeden beliebigen Knoten auswählen, und wiederum dessen direkte Nachbarschaft einblenden. Auf diese Weise kann, ausgehend von einem Eintrag, das gesamte Netzwerk Schritt für Schritt erkundet werden.

Die Anordnung des Netzwerks wird mit Hilfe eines Partikelsystems dynamisch berechnet. In dem Partikelsystem stoßen sich alle Knoten paarweise voneinander ab; Kanten haben eine vordefinierte Ruhelänge, die sich in gewissem Umfang stauchen oder strecken lässt. Aus dem Zusammenspiel der abstoßenden Kräfte zwischen allen Knoten und dem Bestreben der Kanten, ihre Ruhelänge zu bewahren, ergibt sich die Anordnung der Knoten des Netzwerks. Die Positionen der Knoten werden kontinuierlich neu berechnet, so dass immer eine optimale Darstellung des Netzwerks erreicht wird. Der Benutzer kann Knoten verschieben, hinzufügen oder löschen, und die Kräfte im Partikelsystem sorgen dafür, dass das so veränderte Netzwerk wiederum zu einer optimalen Anordnung findet. Die Knoten verändern ihre Positionen dabei schrittweise, so dass die Verschiebungen in der Netzwerkanordnung für den Benutzer nachvollziehbar bleiben und sich Knoten nicht unvermittelt an anderen Positionen befinden.

Diese Visualisierung eignet sich besonders dazu, in Internetseiten eingebettet zu werden. Um weitere Teile des Netzwerks darzustellen, greift sie auf Informationen zu, die auf dem Webserver gespeichert sind. Dabei wird jeweils nur der Teil des Netzwerks an den Benutzer übermittelt, der angezeigt werden soll. Dadurch wird das Herunterladen großer Datenmengen und somit das Auftreten langer Ladezeiten vermieden. Deshalb liegt die besondere Stärke der Visualisierung in der schnellen Anzeige kleiner Subnetzwerke auf Internetseiten, die anschließend vom Anwender weiter erforscht werden können.

Das Problem, komplexe Netzwerkdaten zu visualisieren, tritt in vielen Bereichen auf. Damit Cobweb möglichst universell einsetzbar ist, unterstützt es die Formate XGML und GraphML um die Netzwerkdaten einzulesen und auszugeben. Beide Formate sind XML-basierend und weit verbreitet, um Netzwerkdaten zu speichern. Dadurch wird es für Anwender nicht nur einfacher, Cobweb für ihre Zwecke einzusetzen, sondern es erlaubt auch den Datenaustausch mit weit verbreiteten Programmen wie Cytoscape [39] und Gephi [40].

## **Originalartikel: "Cobweb: a Java Applet for Network Exploration and Visualisation"**

VON EICHBORN J, BOURNE P E, PREISSNER R

*Bioinformatics* 27,1725-1726, 2011

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:

<http://dx.doi.org/10.1093/bioinformatics/btr195>

## 2.2.4 Berücksichtigung quantitativer Daten bei Protein-Wirkstoff-Interaktionen

Für die Analyse von Protein-Wirkstoff Interaktionen ist nicht nur die qualitative Angabe, ob eine Interaktion stattfindet, von Interesse, sondern auch die quantitative, also wie stark eine Interaktion ist. Eine geringe Affinität eines Wirkstoffes zu seinem Zielprotein kann eine höhere Dosierung des Wirkstoffes zum Erreichen eines pharmazeutischen Effekts nötig machen; mit steigender Dosierung wiederum steigt die Gefahr von Nebenwirkungen. Deshalb wurden bei der Aktualisierung der Protein-Wirkstoff-Interaktionsdatenbank SuperTarget [33] in Kooperation mit den Autoren der BindingDB [41] quantitative Daten eingebunden. Der daraus resultierende Datensatz [42] beinhaltet Interaktionen von knapp 200.000 Kleinstrukturen mit über 6.000 Proteinen. Zu den meisten Interaktionen werden Referenzen in der Literatur und Bindungsaffinitäten angegeben, die Angabe der Bindungsstärken erfolgt dabei anhand ihrer IC50-, Kd-, Ki- oder EC50-Werte. Außerdem enthält der Datensatz viele zusätzliche Daten wie Protein-Protein-Interaktionen, Zuordnungen der Proteine zu Stoffwechsel- und Signalwegen und zu Einträgen der Gene Ontologie [43].

Ähnlich wie bei PROMISCUOUS wurde auch für SuperTarget eine umfangreiche Weboberfläche erstellt, mit deren Hilfe der Datenbestand durchsucht und Ergebnisse angezeigt werden können. Die Suche ist nicht nur über den Wirkstoff oder das Protein möglich, sondern auch anhand des entsprechenden Signal- oder Stoffwechselweges, der Genontologie oder des beteiligten Cytochroms. Außerdem gibt es die Möglichkeit, eine erweiterte Suche durchzuführen, bei der sich vielfältige Eigenschaften kombinieren lassen. So ist zum Beispiel die Suche nach Kleinstrukturen möglich, die mit einem bestimmten Protein interagieren und deren Molekulargewichte in einem bestimmten Bereich liegen. Die Suchergebnisse umfassen neben den reinen Interaktionsinformationen, gegebenenfalls mit Literaturreferenzen und Angabe der Affinitäten, Metadaten zu den beteiligten Elementen. Dies sind für Wirkstoffe zum Beispiel physikochemische Eigenschaften und Synonyme.

## **Originalartikel: "SuperTarget Goes Quantitative: Update on Drug-Target Interactions"**

HECKER N, AHMED J, VON EICHBORN J, DUNKEL M, MACHA K, ECKERT A, GILSON M K, BOURNE P E, PREISSNER R

*Nucleic Acids Research* 40, D1113-D1117, 2012

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:

<http://dx.doi.org/10.1093/nar/gkr912>

## Kapitel 3 Diskussion

### 3.1 Bereitstellen und Auswerten struktureller Informationen über Protein-Protein-Interaktionen

#### 3.1.1 Akquisition struktureller Informationen über Makromolekül-Interaktionsflächen

Der Bestand an bekannten 3D-Strukturen von Proteinen ist innerhalb der letzten Jahre rasant gewachsen; die Zahl der neu aufgeklärten Strukturen pro Jahr steigt seit zwanzig Jahren kontinuierlich an. Im letzten Jahr lag diese Marke bei nahezu 8.100 neuen 3D-Strukturen; der Trend wird voraussichtlich anhalten. Diese gewaltige Datenmenge beinhaltet viele Interaktionen zwischen Proteinen und Protein-Domänen, die allerdings nicht sofort zugänglich sind. Denn die Interaktionen sind zwar in den 3D-Strukturdaten implizit enthalten, aber nicht explizit ausgezeichnet.

Die in dieser Arbeit vorgestellte Datenbank JAIL (Abschnitt 2.1.1) soll diese Lücke schließen. Sie stellt eine umfassende Sammlung von derzeit mehr als 260.000 Interaktionsflächen in allen in der PDB vorhandenen 3D-Strukturen zur Verfügung. Zum Zeitpunkt ihrer Veröffentlichung im Jahr 2009 enthielt sie ungefähr 184.000 Interaktionsflächen; dieser Zuwachs um circa 80.000 Interaktionsflächen verdeutlicht, wie wichtig es ist, solche Ressourcen fortlaufend zu aktualisieren. JAIL berücksichtigt Interaktionen zwischen verschiedenen Protein-Domänen, zwischen verschiedenen Aminosäureketten sowie zwischen Aminosäureketten und Nukleinsäureketten. Zusätzlich werden Interaktionen innerhalb der biologischen Einheiten berücksichtigt, sofern diese von den entsprechenden asymmetrischen Einheiten abweichen.

JAIL unterscheidet sich durch diese Vollständigkeit der berücksichtigten Interaktionen von anderen Protein-Protein-Interaktionsdatenbanken wie SCOPPI [44], SNAPPI [45] oder PIBASE [46]. Außerdem wird JAIL im Gegensatz zu anderen Datenbanken alle drei Monate aktualisiert, so dass sie immer einen aktuellen Stand der bekannten Interaktionsflächen repräsentiert. Für Benutzer wird diese Menge an Daten über verschiedene Suchmethoden und nicht-redundante, repräsentative Teildatensätze zum Herunterladen zugänglich gemacht.

Die in JAIL gespeicherten Daten wurden auch direkt für einen speziellen Protein-Protein-Docking Ansatz namens ISEARCH [47] verwendet, der wissenschaftsbasiert auf Grundlage bekannter Interaktionsflächen arbeitet. Für den Erfolg dieser Methode ist es entscheidend, dass ein möglichst umfassender Datensatz an bekannten Interaktionsflächen zur Verfügung steht. ISEARCH mit JAIL als

Interaktionsflächen-Bibliothek wurde benutzt, um eine Komplexstruktur im Rahmen von CAPRI 2009 vorherzusagen. CAPRI (Critical Assessment of Prediction of Interactions) ist eine Veranstaltung, die seit dem Jahr 2000 wiederholt stattfindet und deren Zweck es ist, die Leistungsfähigkeit aktueller Methoden des Protein-Protein-Dockings zu evaluieren [48]. Zur Evaluierung der Methoden wird die 3D-Struktur eines neu aufgeklärten Proteinkomplexes nicht veröffentlicht, sondern nur die Einzelstrukturen der beiden beteiligten Proteine. Daraufhin können alle Teilnehmer je zehn Modelle der Komplexkonformationen einsenden; nach Ablauf der Teilnahmefrist werden diese mit der zuvor unveröffentlichten Struktur des Komplexes verglichen. Dadurch wird ein unverfälschter Eindruck der Leistungsfähigkeit der verschiedenen Methoden gewonnen. Mit der Kombination aus ISEARCH und JAIL gelang es, für den betrachteten Komplex zwei Vorhersagen der Komplexkonformation zu generieren, die als korrekt akzeptiert wurden [49].

### 3.1.2 Strukturelle und evolutionäre Eigenschaften von Protein-Interaktionsflächen

Die in der JAIL Datenbank vorhandenen Informationen können auf vielfältige Weise benutzt werden. Im vorigen Abschnitt wurde die Anwendung der gespeicherten Interaktionsflächen zur Vorhersage der Komplexkonformation zweier interagierender Proteine angesprochen. Ein anderer Verwendungszweck wurde in Abschnitt 2.1.2 vorgestellt. Dort wurden die Interaktionsflächen untersucht, hauptsächlich in Hinblick auf mögliche Unterschiede zwischen transienten und obligaten Interaktionen. Bei Protein-Protein-Interaktionen kommt es nicht nur darauf an, ob eine Interaktion stattfindet oder nicht, sondern ebenso, wie dauerhaft diese Interaktion ist. Es gibt obligate Proteinkomplexe, die sich direkt nach der Synthese der beteiligten Proteine bilden und für die ganze Lebensdauer der Proteine bestehen bleiben. Es gibt aber auch transiente Interaktionen, die nur kurzfristig zustande kommen. Bei diesen ist es oft von entscheidender Bedeutung, dass die Interaktionen reversibel sind, um die korrekte Funktion der Proteine zu gewährleisten. Ein Beispiel dafür sind heterotrimere G-Proteine, die in vielen Signalwegen vorkommen. Diese sind aus drei Untereinheiten aufgebaut; bei Aktivierung des Proteins dissoziiert die  $\alpha$ -Untereinheit von dem Komplex der  $\beta$ - und  $\gamma$ -Untereinheiten. Für die Aktivierbarkeit des Proteins ist also die transiente Natur der Interaktion der Untereinheiten entscheidend. Da die Unterschiede zwischen transienten und obligaten Interaktionen somit von entscheidender Bedeutung im Organismus sind, ist das Interesse an den physikochemischen Eigenschaften der verschiedenen Interaktionstypen dementsprechend groß.

Die Analyse der Interaktionsflächen ergab, dass obligate Interaktionsflächen eine geringere Mutationsrate haben als transiente. Eine mögliche Erklärung für diese Entdeckung ist die bereits von



Mintseris und Weng geäußerte Vermutung, dass Proteine dadurch an obligaten Interaktionsstellen korrelierte Mutationen ausbilden können, also gemeinsam evolvieren [50]. Außerdem wurde beobachtet, dass Residuen an Interaktionsflächen bevorzugt mit anderen Residuen in Kontakt stehen, die denselben Grad an Konservierung haben; dies gilt insbesondere für besonders stark und besonders schwach konservierte Residuen. Es bestehen also bevorzugt Kontakte zwischen Residuen der interagierenden Proteine, bei denen beide Residuen stark oder beide schwach konserviert sind. Dies bekräftigt die "Hot-Spot"-Hypothese, nach der einige wenige Kontakte zwischen Residuen der beiden Interaktionspartner in Interaktionsflächen besonders wichtig für das Zustandekommen der Interaktion sind [51]. Diese besonders wichtigen Residuen sind dementsprechend besonders stark konserviert.

Es wurden auch Unterschiede in der Aminosäure-Zusammensetzung und der Konservierung der Interaktionsflächen im Vergleich zu der restlichen Oberfläche der Proteine festgestellt. So zeigten zum Beispiel die aromatischen Aminosäuren Phenylalanin, Tyrosin und Tryptophan eine deutliche Präferenz für Interaktionsflächen im Vergleich zu der restlichen Proteinoberfläche. Diese Informationen können dazu genutzt werden, um auf Proteinoberflächen nach möglichen Interaktionsflächen zu suchen.

Die betrachteten Interaktionsflächen wurden ferner daraufhin untersucht, welche Aminosäuren der beiden Interaktionspartner bevorzugt in Kontakt miteinander stehen. Beispielsweise haben aromatische Aminosäuren eine Präferenz dafür, mit anderen aromatischen Aminosäuren zu interagieren, vermutlich weil dadurch energetisch vorteilhafte Stapel von aromatischen Ringen ausgebildet werden können; das erklärt auch die im vorigen Absatz erwähnte Präferenz der aromatischen Aminosäuren für Interaktionsflächen. Anhand der beobachteten Kontakte zwischen Residuen der Interaktionspartner wurden Interaktionspotentiale berechnet, die für jeden möglichen Kontakt zwischen zwei Residuen die Information beinhalten, wie wahrscheinlich dieser Kontakt ist. Mit Hilfe dieser Potentiale können mögliche Interaktionen darauf getestet werden, wie wahrscheinlich ihr Zustandekommen ist.

### **3.1.3 Anwendung wissensbasierter Interaktionspotentiale zur Bindungsvorhersage**

Ein analoges Verfahren wurde in der im Abschnitt 2.1.3 vorgestellten Arbeit dazu genutzt, um im Rahmen des agentenbasierten Modells Vacclmm Interaktionspotentiale für Rezeptoren des Immunsystems mit ihren Liganden zu bestimmen. Die spezifisch für die jeweiligen Rezeptoren berechneten Interaktionspotentiale erwiesen sich dabei als zuverlässiger im Unterscheiden realer

und randomisierter Interaktionen als allgemeinere Potentiale, wie zum Beispiel das von Miyazawa und Jernigan [52].

Vacclmm kann dazu dienen, Einblicke in die entscheidenden Mechanismen und Zusammenhänge der Peptidvakzinierung zu gewinnen, ohne aufwändige und langwierige Laborversuche durchführen zu müssen. Ein kritischer Punkt bei einem solchen Modell ist, wie empfindlich es auf Fehler in der Bindungsvorhersage der Rezeptor-Liganden-Komplexe reagiert. Auch wenn die verwendeten Interaktionspotentiale, wie bereits geschildert, zuverlässig in der Lage waren, reale von randomisierten Komplexen zu unterscheiden, so ist doch mit gelegentlichen falschen Vorhersagen zu rechnen. Deshalb wurde untersucht, wie stark sich fehlerhafte Bindungsvorhersagen auf das Gesamtergebnis der Simulation auswirken. Die Analyse ergab, dass Veränderungen der Interaktionspotentiale die Ergebnisse der Simulation zwar beeinflussen, ein qualitativ anderes Ergebnis jedoch erst ab Abweichungen der Bindungsstärken von 10 bis 15 % zu erwarten ist. Damit ist das Modell gegenüber kleinen Fehlern bei der Berechnung der Bindungsstärken robust.

Vacclmm ist ein weiterer Schritt in der Entwicklung rechnergestützter Modelle zur Simulation des Immunsystems, der im Gegensatz zu seinen Vorgängern wichtige Erweiterungen wie die Berücksichtigung der realen Peptidsequenzen beinhaltet. Allerdings ist das menschliche Immunsystem hochkomplex. Daher können Simulationen mit den heutigen technischen Mitteln nicht alle möglichen Komponenten, die einen Einfluss haben könnten, berücksichtigen. Auch in Vacclmm werden einige Faktoren derzeit noch nicht beachtet. Ein wichtiger Punkt, der in Zukunft hinzugefügt werden soll, ist die Möglichkeit der Mutation von Tumorzellen. Dadurch könnte simuliert werden, wie einzelne, mutierte Tumorzellen der Erkennung und Vernichtung durch das Immunsystem entgegen können. Durch den modularen Aufbau des Programms ist das Hinzufügen solcher neuer Eigenschaften mit vergleichsweise geringem Aufwand möglich, so dass das Modell erweitert und aktuell gehalten werden kann.

## **3.2 Konzeption und Auswertung molekularer Interaktionsnetzwerke**

### **3.2.1 Datenintegration zur Generierung eines komplexen molekularen Interaktionsnetzwerks**

Ähnlich wie bei Protein-Protein-Interaktionen gibt es auch bei Protein-Kleinstruktur-Interaktionen einen stetig wachsenden Bestand an Informationen. Durch die Entwicklung moderner Hochdurchsatzverfahren ist die Analyse großer Bibliotheken von Kleinstrukturen im Hinblick auf

ihre Bindung an bestimmte Proteine durchführbar geworden. Allerdings werden die Ergebnisse dieser Analysen, wie bereits in Abschnitt 2.2.1 erwähnt, kaum zentral gesammelt; gerade Untersuchungen über die biologische Aktivität von Wirkstoffen werden häufig nur als Artikel in Fachzeitschriften veröffentlicht. Deshalb wurde die in Abschnitt 2.2.1 vorgestellte Datenbank PROMISCUOS entwickelt, die einen umfassenden Datensatz zu Protein-Wirkstoff-Interaktionen und deren zelluläres Umfeld bereit stellt. Dieser Datensatz ergibt ein komplexes Netzwerk verschiedenartiger Informationen, die zueinander in Beziehung gesetzt sind. In der in Abschnitt 2.2.4 vorgestellten Datenbank SuperTarget gehören zu diesem Netzwerk zusätzlich noch quantitative Daten über die Bindungsstärke von Kleinstrukturen an Proteine.

Die miteinander verknüpften Daten ergeben einen Einblick in die molekularen Zusammenhänge, in denen sich die einzelnen Proteine und Kleinstrukturen befinden. Allerdings gibt es noch viele weitere Informationen, die das Wissen um die beteiligten Komponenten erweitern könnten. Für Proteine wären zum Beispiel Informationen über häufige Mutationen und deren Auswirkungen interessant. Außerdem sind auch die bisher enthaltenen Informationen noch bei weitem nicht vollständig. Da fortwährend neue Interaktionen experimentell getestet werden, müssen solche Datenbanken aktuell gehalten werden, zum Beispiel indem neu erscheinende Artikel regelmäßig auf neue Interaktionen untersucht werden. Darüber hinaus könnten weitere Datenquellen, wie beispielsweise die Therapeutic Target Database [53] mit aufgenommen werden. Das Hinzufügen weiterer Datenquellen dient gleich zwei Zwecken: zum einen werden noch mehr Interaktionen abgedeckt, da sich die verschiedenen Datenbestände nur teilweise überschneiden. Zum anderen können gerade diese Überschneidungen genutzt werden, um ein Vertrauensmaß für einzelne Einträge zu definieren.

Neben der Anzahl der Datenbanken, die eine bestimmte Interaktion nennen, könnte das Vertrauensmaß auch berücksichtigen, wie viele Literaturbelege es für eine Interaktion gibt und welche Technik für ihren Nachweis angewendet wurde. Eine solche Bewertung der Interaktionen ist wünschenswert, da bei der Erstellung von Interaktionsdatenbanken viele Fehlerquellen vorhanden sind. Zum einen können die experimentellen Daten an sich fehlerhaft sein, entweder durch experimentelle Fehler oder eine falsche Planung der Studien. Beispielsweise könnte eine *in vitro* nachgewiesene Bindung physiologisch irrelevant sein, wenn die beiden Bindungspartner *in vivo* in verschiedenen Zellkompartimenten vorliegen. Zum anderen können bei der Erfassung der in Fachzeitschriften veröffentlichten Artikel Fehler auftreten. Da die Daten nicht einfach maschinenlesbar sind, müssen die Einträge entweder manuell erfasst oder mittels informatischer Methoden der Textanalyse extrahiert werden. Um umfassende Datenbanken zu erstellen, müssen sehr große Datenmengen untersucht werden, daher ist ein rein manueller Ansatz meist nicht zu leisten. Die

Texte werden deswegen in der Regel mit Hilfe informatischer Methoden analysiert; dieser Vorgang wird als Text Mining bezeichnet. Algorithmen zum Text Mining sind heutzutage in der Lage, Satzglieder zu erkennen und dadurch die Syntax der Sätze zu verstehen. Allerdings sind die Grammatiken der menschlichen Sprachen so komplex, dass die Algorithmen nicht immer richtige Ergebnisse liefern. Zudem kommen gerade in den hier betrachteten biologischen Arbeiten viele Abkürzungen und Synonyme vor, die die Arbeit der Algorithmen zusätzlich erschweren.

Ein sinnvoller Ansatz zur Auswertung der veröffentlichten Arbeiten ist daher die Erstellung leistungsfähiger Algorithmen, die den Großteil der Arbeit automatisch erledigen. Dabei sollte der Algorithmus zu jedem Ergebnis vermerken, wie sicher er sich ist, dass es korrekt ist. So könnten unsichere Ergebnisse gezielt manuell überprüft und der Algorithmus gegebenenfalls angepasst werden. Außerdem müssen die übrigen Ergebnisse stichprobenartig kontrolliert werden, um eine hohe Qualität der automatischen Erkennung sicher zu stellen.

### 3.2.2 Promiskuität von Wirkstoffen

Wie im vorigen Abschnitt erwähnt, erlauben umfassende Datensammlungen zu Protein-Wirkstoff-Interaktionen die Analyse dieser Beziehungen. In der in Abschnitt 2.2.2 vorgestellten Arbeit wurde der PROMISCUOUS Datensatz genutzt, um Unterschiede zwischen promiskuitiven, nicht-promiskuitiven und zurückgezogenen Wirkstoffen, sowie zwischen deren Zielproteinen zu untersuchen.

Die Analyse zeigte Unterschiede in den physikochemischen Eigenschaften der promiskuitiven, der nicht-promiskuitiven und der zurückgezogenen Wirkstoffe. Diese Unterschiede sind beim Vergleich der zurückgezogenen Wirkstoffe mit den anderen Gruppen am stärksten, insbesondere die polare Oberfläche sowie der Oktanol-Wasser-Verteilungskoeffizient  $\log P$  unterscheiden sich hier deutlich. Beide Werte sind für die pharmakologische Wirkung der Stoffe von Bedeutung: die geringere polare Oberfläche führt dazu, dass die Stoffe leichter die Blut-Hirn-Schranke überwinden und somit unerwünschte Wirkungen im Gehirn ausüben können. Der höhere  $\log P$  Wert besagt, dass die Stoffe lipophiler als die der anderen Gruppen sind. Dadurch lösen sich diese Stoffe schwerer im Blut und können also weniger leicht im Körper zur Verfügung gestellt werden.

Bei der Untersuchung der Häufigkeit von funktionellen Gruppen in den verschiedenen Gruppen von Wirkstoffen wurden teilweise erhebliche Unterschiede festgestellt. Beispielsweise kommen Amine bevorzugt in promiskuitiven Wirkstoffen vor; das ist eine interessante Ergänzung zu früheren Beobachtungen, wonach Stickstoff-Heterozyklen häufig in promiskuitiven Molekülen vorkommen [54]. Bemerkenswerterweise zeigen Amide eine genau entgegengesetzte Präferenz. Die

Untersuchung der Interaktionen der beiden funktionellen Gruppen mit ihren Zielproteinen ergab, dass Amide häufiger Wasserstoffbrückenbindungen zum Proteinerückgrat ausbilden als Amine.

Neben den Wirkstoffen selbst wurden auch ihre Zielproteine untersucht. Die Analyse fand in Bezug auf deren Stellung im humanen Protein-Protein-Interaktionsnetzwerk statt. Dabei zeigte sich, dass Zielproteine von Wirkstoffen in der Regel deutlich zentraler im Netzwerk gelegen sind, als Proteine, zu denen keine Wirkstoff-Interaktionen bekannt sind. Die Zielproteine nicht-promiskuitiver Wirkstoffe sind dabei wiederum zentraler als die promiskuitiver Wirkstoffe. Daraus folgt zum einen, dass Proteine, die in der Peripherie des Netzwerks liegen, als Zielproteine weniger attraktiv sind; vermutlich würde das Adressieren eines solchen Proteins häufig keine ausreichende Auswirkung auf das Gesamtnetzwerk haben, um einen therapeutischen Effekt zu erzielen. Zum anderen werden besonders zentrale Proteine bevorzugt von nicht-promiskuitiven Wirkstoffen adressiert. Eine Erklärung hierfür ist, dass diese Proteine großen Einfluss auf das restliche Netzwerk haben; deswegen müssen sie sehr gezielt adressiert werden und dürfen nicht versehentlich von promiskuitiven Wirkstoffen mit beeinflusst werden.

Außerdem wurde untersucht, ob sich die Zielproteine verschiedener Wirkstoffklassen in ihren Netzwerkeigenschaften unterscheiden. Dazu wurden ATC klassifizierte Wirkstoffe anhand der Organe oder Systeme, in denen sie aktiv sind, gruppiert und ihre Zielproteine miteinander verglichen. Hier zeigten sich deutliche Unterschiede zwischen einigen Wirkstoffklassen, so dass anhand der Zentralitäten der Zielproteine Gruppen identifiziert werden konnten, die bevorzugt zentralere beziehungsweise weniger zentrale Proteine adressieren. Solche Präferenzen können zum Beispiel in der Wirkstoffentwicklung berücksichtigt werden, um die Eignung möglicher Zielproteine abzuschätzen.

Die Frage nach den Gründen und Auswirkungen promiskuitiven Verhaltens von Wirkstoffen hat in den letzten Jahren vermehrt Beachtung gefunden, seitdem entdeckt wurde, dass dieses Verhalten häufiger auftritt als zuvor angenommen. Die oben vorgestellten Ergebnisse liefern Einblicke in die Zusammenhänge zwischen molekularen Eigenschaften und Promiskuität der Wirkstoffe einerseits und zwischen Promiskuität der Wirkstoffe und Bedeutung der Zielproteine im Interaktionsnetzwerk andererseits. Solche Informationen können bei der Entwicklung neuer Wirkstoffe eingesetzt werden, um die Gefahr unerwünschter Nebenwirkungen zu verringern und bei der Auswahl geeigneter Zielproteine zu helfen.

### 3.2.3 Interaktive Netzwerkdarstellung hochgradig verknüpfter Daten

Um hochgradig miteinander verknüpfte Datensätze darzustellen, bietet sich die Visualisierung als Netzwerk an Stelle von Tabellen mit in Beziehung stehenden Elementen an. Deswegen hat sich gerade auch im Bereich der Systembiologie die netzwerkbasierende Visualisierung etabliert; so sind sowohl in Fachartikeln wie auch auf diesbezüglichen Internetseiten häufig Netzwerke abgebildet. Allerdings sind diese Netzwerke meistens statisch, daher sind sie in der Regel entweder so groß, dass die einzelnen Komponenten in der Darstellung nicht mehr zu erkennen sind, oder die Darstellungen beschränken sich auf einen kleinen Ausschnitt des Gesamtnetzwerks.

Diese Nachteile lassen sich mit einer interaktiven Visualisierung vermeiden, wie sie in Abschnitt 2.2.3 dieser Arbeit vorgestellt wurde. Diese präsentiert zu Beginn nur einen ausgewählten, kleinen Teil des Netzwerks; auf Wunsch des Anwenders können anschließend weitere Teile des Netzwerks nachgeladen werden. Dadurch kann sich der Nutzer genau den Teil des Gesamtnetzwerks anzeigen lassen, der für ihn von Interesse ist. Diese Visualisierung wird bereits in mehreren Projekten verwendet, unter anderem Estrella [55], PROMISCUOUS [31] und SuperTarget [42]. Die Implementierung in andere Projekte, wie zum Beispiel die BindingDB [41], findet derzeit statt.

Der genaue Mechanismus, um ein Netzwerk schrittweise zu erforschen, wurde bereits in Abschnitt 2.2.3 erläutert. Wichtig ist dabei, dass die Knoten immer automatisch zu einer optimalen Anordnung finden, um das Netzwerk übersichtlich darzustellen. Dabei wechselwirken abstoßende Kräfte zwischen den Knoten und die vorhandenen Kanten, so dass eine überwiegend kreisförmige Anordnung der Knoten entsteht. Bei gerichteten Netzwerken können die Kantenrichtungen mit berücksichtigt werden, dadurch ergibt sich eine vertikale Anordnung, ähnlich wie in Darstellungen biologischer Signalwege.

In Zukunft werden noch weitere Optionen zur unterschiedlichen Darstellung von Netzwerken möglich sein. Gegenwärtig wird bereits an einer Darstellungsweise gearbeitet, in der einzelne Zellkompartimente getrennt voneinander dargestellt werden. Dadurch wird die Netzwerkdarstellung noch enger an bekannte Darstellungen biologischer Stoffwechselwege angelehnt. Außerdem wird es in Zukunft möglich sein, Knoten nach Belieben zu Metaknoten zusammen zu fassen. Dadurch wird besonders in komplexen Netzwerken eine bessere Übersicht ermöglicht.

Bei der Erweiterung der Visualisierung darf aber nicht außer Acht gelassen werden, dass ihr primäres Ziel die unkomplizierte und schnelle Darstellung von Netzwerken direkt auf Internetseiten ist. Es ist erklärtermaßen nicht das Ziel, den Funktionsumfang komplexer Programme wie

Cytoscape [39], das aufwändige Netzwerkanalysen mittels einer Plug-In Architektur unterstützt, anzustreben.

### 3.3 Fazit und Ausblick

In dieser Arbeit wurden Eigenschaften molekularer Interaktionen untersucht, um Einblicke in das komplexe molekulare Interaktionsnetzwerk von Proteinen und Kleinstrukturen zu gewinnen.

Im ersten Teil der Arbeit standen Protein-Protein-Interaktionen im Mittelpunkt. Ein umfassender Datensatz an Protein-Protein-Interaktionsflächen wurde zusammengestellt, um deren Charakteristika zu untersuchen. Dabei wurden strukturelle und evolutionäre Eigenschaften der Interaktionsflächen analysiert. Die gewonnenen Informationen wurden benutzt, um Komplexkonformationen von Proteinen vorherzusagen und um mögliche Protein-Interaktionen innerhalb eines agentenbasierten Modells zu bewerten.

Im zweiten Teil wurden Interaktionen zwischen Proteinen und Wirkstoffen behandelt. Auch hier bestand der erste Schritt darin, eine umfassende Datengrundlage für die folgenden Analysen zusammen zu stellen. Auf dieser Basis wurde insbesondere promiskuitives Verhalten von Wirkstoffen untersucht. Da dieses zu Nebenwirkungen der Wirkstoffe führen kann, können solche Informationen in Zukunft bei der Wirkstoffentwicklung hilfreich sein.

Um die für die Analysen notwendigen Daten zusammen zu stellen, mussten Daten aus verschiedenen Quellen integriert werden. Dazu wurden verschiedene Ansätze entwickelt, deren Anwendung in der Arbeit vorgestellt wird. Bei der Integration von Daten stellen sich dabei zwei Hauptprobleme: zum einen können die Daten in vielen Quellen und nicht maschinenlesbar vorliegen, so dass sie mühevoll ermittelt werden müssen, wie es zum Beispiel bei den Protein-Wirkstoff-Interaktionsinformationen in Abschnitt 2.2.1 nötig war. Zum anderen sind selbst bereits vorhandene Datensammlungen mit umfangreichen Informationen teilweise nur schwierig miteinander kombinierbar, etwa wenn sie verschiedene Kennungen für die enthaltenen Elemente benutzen.

Die Integration von Daten wird in den nächsten Jahren eine noch entscheidendere Rolle spielen, wenn mit immer fortgeschritteneren experimentellen Methoden immer größere Datenmengen erhoben werden können. Beispielweise arbeitet ein internationales Projekt daran, 2.500 humane Genomsequenzen auszulesen [56]; ein Projekt, das vor einigen Jahren noch nicht umzusetzen gewesen wäre. Die erhobenen Daten dieses Projekts beinhalten detaillierte Informationen zu den ermittelten Genomen, wie zum Beispiel über Einzelnukleotid-Polymorphismen (single nucleotide polymorphism, SNP), die den größten Teil der genetischen Unterschiede zwischen einzelnen Indi-

viduen darstellen. Die Verknüpfung dieser Informationen mit dem Wissen über molekulare Interaktionsnetzwerke, wie es Gegenstand dieser Arbeit ist, könnte zum Beispiel erklären, warum bestimmte Menschen Unverträglichkeiten gegenüber bestimmten Stoffen aufweisen. Solche Erkenntnisse sind ein wichtiger Schritt in Richtung einer personalisierten Medizin, in der der Genotyp eines Patienten berücksichtigt wird, um eine optimale Therapie zu ermöglichen.



# Verzeichnisse

## Literaturverzeichnis

- [1] MENDEL, G.: Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn* 4:3–47, 1866
- [2] FISHER, R.A.: Has Mendel's work been rediscovered? *Annals of Science* 1(2):115–137, 1936
- [3] AVERY, O.T.; MACLEOD, C.M.; MCCARTY, M.: Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *The Journal of Experimental Medicine* 79(2):137–158, 1944
- [4] SANGER, F.; NICKLEN, S.; COULSON, A.R.: DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74(12):5463–5467, 1977
- [5] LANDER, E.S.; LINTON, L.M.; BIRREN, B.; NUSBAUM, C.; ZODY, M.C.; BALDWIN, J.; DEVON, K.; DEWAR, K.; DOYLE, M. ET AL.: Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921, 2001
- [6] VENTER, J.C.; ADAMS, M.D.; MYERS, E.W.; LI, P.W.; MURAL, R.J.; SUTTON, G.G.; SMITH, H.O.; YANDELL, M.; EVANS, C.A. ET AL.: The sequence of the human genome. *Science* 291(5507):1304–1351, 2001
- [7] CORNISH-BOWDEN, A.; CÁRDENAS, M.L.: Complex networks of interactions connect genes to phenotypes. *Trends in Biochemical Sciences* 26(8):463–465, 2001
- [8] SCHUSTER, S.; EILS, R.; PRANK, K.: 5th International Conference on Systems Biology (ICSB 2004), Heidelberg, October 9-13, 2004. *Bio Systems* 83(2-3):71–265, 2006
- [9] KITANO, H.: Systems biology: a brief overview. *Science* 295(5560):1662–1664, 2002
- [10] AHN, A.C.; TEWARI, M.; POON, C.-S.; PHILLIPS, R.S.: The clinical applications of a systems approach. *PLoS Medicine* 3(7):e209, 2006
- [11] STUMPF, M.P.H.; THORNE, T.; DE SILVA, E.; STEWART, R.; AN, H.J.; LAPPE, M.; WIUF, C.: Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences of the United States of America* 105(19):6959–6964, 2008
- [12] AMARAL, L.A.N.: A truer measure of our ignorance. *Proceedings of the National Academy of Sciences of the United States of America* 105(19):6795–6796, 2008
- [13] VAJDA, S.; KOZAKOV, D.: Convergence and combination of methods in protein-protein docking. *Current Opinion in Structural Biology* 19(2):164–170, 2009

- [14] MOREIRA, I.S.; FERNANDES, P.A.; RAMOS, M.J.: Protein-protein docking dealing with the unknown. *Journal of Computational Chemistry* 31(2):317–342, 2010
- [15] GROSDIDIER, S.; SOLERNOU, A.; PÉREZ-CANO, L.; PONS, C.; FERNÁNDEZ-RECIO, J.: Present and future challenges and limitations in protein-protein docking. *Proteins* 78(1):95–108, 2009
- [16] BERMAN, H.M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T.N.; WEISSIG, H.; SHINDYALOV, I.N.; BOURNE, P.E.: The Protein Data Bank. *Nucleic Acids Research* 28(1):235–242, 2000
- [17] AKTORIES, K.; FÖRSTERMANN, U.; HOFMANN, F.B.; STARKE, K.: *Allgemeine und spezielle Pharmakologie und Toxikologie*. 9. Auflage. München: Elsevier GmbH, 2005
- [18] DELANO, W.L.: The PyMOL Molecular Graphics System, Schrödinger, LLC, 2002
- [19] YILDIRIM, M.A.; GOH, K.-I.; CUSICK, M.E.; BARABÁSI, A.-L.; VIDAL, M.: Drug-target network. *Nature Biotechnology* 25(10):1119–1126, 2007
- [20] CAMPILLOS, M.; KUHN, M.; GAVIN, A.-C.; JENSEN, L.J.; BORK, P.: Drug target identification using side-effect similarity. *Science* 321(5886):263–266, 2008
- [21] COREY, E.J.; CZAKÓ, B.; KÜRTI, L.: *Molecules and Medicine*. Hoboken, New Jersey: John Wiley & Sons Inc., 2007
- [22] TUNCBAG, N.; GURSOY, A.; GUNAY, E.; NUSSINOV, R.; KESKIN, O.: Architectures and functional coverage of protein-protein interfaces. *Journal of Molecular Biology* 381(3):785–802, 2008
- [23] GÜNTHER, S.; VON EICHBORN, J.; MAY, P.; PREISSNER, R.: JAIL: a structure-based interface library for macromolecules. *Nucleic Acids Research* 37(Database issue):D338–341, 2009
- [24] FINN, O.J.: Cancer immunology. *The New England Journal of Medicine* 358(25):2704–2715, 2008
- [25] PALUCKA, K.; UENO, H.; BANCHEREAU, J.: Recent developments in cancer vaccines. *Journal of Immunology* 186(3):1325–1331, 2011
- [26] DOUGAN, M.; DRANOFF, G.: Immune therapy for cancer. *Annual Review of Immunology* 27:83–117, 2009
- [27] JENQ, R.R.; VAN DEN BRINK, M.R.M.: Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. *Nature Reviews Cancer* 10(3):213–221, 2010
- [28] WOELKE, A.L.; MURGUEITIO, M.S.; PREISSNER, R.: Theoretical modeling techniques and their impact on tumor immunology. *Clinical & Developmental Immunology* 2010:271794, 2010
- [29] CELADA, F.; SEIDEN, P.E.: A computer model of cellular interactions in the immune system. *Immunology Today* 13(2):56–62, 1992
- [30] CASTIGLIONE, F.; TOSCHI, F.; BERNASCHI, M.; SUCCI, S.; BENEDETTI, R.; FALINI, B.; LISO, A.: Computational modeling of the immune response to tumor antigens. *Journal of Theoretical Biology* 237(4):390–400, 2005

- [31] VON EICHBORN, J.; MURGUEITIO, M.S.; DUNKEL, M.; KOERNER, S.; BOURNE, P.E.; PREISSNER, R.: PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Research* 39(Database issue):D1060–1066, 2011
- [32] WISHART, D.S.; KNOX, C.; GUO, A.C.; CHENG, D.; SHRIVASTAVA, S.; TZUR, D.; GAUTAM, B.; HASSANALI, M.: DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research* 36(Database issue):D901–906, 2008
- [33] GÜNTHER, S.; KUHN, M.; DUNKEL, M.; CAMPILLOS, M.; SENGER, C.; PETSALAKI, E.; AHMED, J.; URDIALES, E.G.; GEWIESS, A. ET AL.: SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Research* 36(Database issue):D919–922, 2008
- [34] PREISSNER, S.; KROLL, K.; DUNKEL, M.; SENGER, C.; GOLDSOBEL, G.; KUZMAN, D.; GÜNTHER, S.; WINNENBURG, R.; SCHROEDER, M. ET AL.: SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Research* 38(Database issue):D237–243, 2010
- [35] WHO COLLABORATING CENTRE FOR DRUG STATISTICS METHODOLOGY: *Guidelines for ATC classification and DDD assignment 2012*. 15. Auflage. Oslo, 2011
- [36] CHANG, A.; SCHEER, M.; GROTE, A.; SCHOMBURG, I.; SCHOMBURG, D.: BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research* 37(Database issue):D588–592, 2009
- [37] KANEHISA, M.; GOTO, S.; HATTORI, M.; AOKI-KINOSHITA, K.F.; ITOH, M.; KAWASHIMA, S.; KATAYAMA, T.; ARAKI, M.; HIRAKAWA, M.: From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* 34(Database issue):D354–357, 2006
- [38] VON EICHBORN, J.; BOURNE, P.E.; PREISSNER, R.: Cobweb: a Java applet for network exploration and visualisation. *Bioinformatics* 27(12):1725–1726, 2011
- [39] SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N.S.; WANG, J.T.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11):2498–2504, 2003
- [40] BASTIAN, M.; HEYMANN, S.; MATHIEU, J.: Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*, 2009
- [41] LIU, T.; LIN, Y.; WEN, X.; JORISSEN, R.N.; GILSON, M.K.: BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research* 35(Database issue):D198–201, 2007
- [42] HECKER, N.; AHMED, J.; VON EICHBORN, J.; DUNKEL, M.; MACHA, K.; ECKERT, A.; GILSON, M.K.; BOURNE, P.E.; PREISSNER, R.: SuperTarget goes quantitative: update on drug-target interactions. *Nucleic Acids Research* 40(Database issue):D1113–1117, 2012
- [43] ASHBURNER, M.; BALL, C.A.; BLAKE, J.A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J.M.; DAVIS, A.P.; DOLINSKI, K.; DWIGHT, S.S. ET AL.: Gene ontology: tool for the unification of biology. *Nature Genetics* 25(1):25–29, 2000
- [44] WINTER, C.; HENSCHL, A.; KIM, W.K.; SCHROEDER, M.: SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Research* 34(Database issue):D310–314, 2006

- [45] JEFFERSON, E.R.; WALSH, T.P.; ROBERTS, T.J.; BARTON, G.J.: SNAPPI-DB: a database and API of Structures, iNterfaces and Alignments for Protein-Protein Interactions. *Nucleic Acids Research* 35(Database issue):D580–589, 2007
- [46] DAVIS, F.P.; SALI, A.: PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics* 21(9):1901–1907, 2005
- [47] GÜNTHER, S.; MAY, P.; HOPPE, A.; FRÖMMEL, C.; PREISSNER, R.: Docking without docking: ISEARCH--prediction of interactions using known interfaces. *Proteins* 69(4):839–844, 2007
- [48] JANIN, J.: Assessing predictions of protein-protein interaction: the CAPRI experiment. *Protein Science* 14(2):278–283, 2005
- [49] LENSINK, M.F.; WODAK, S.J.: Docking and scoring protein interactions: CAPRI 2009. *Proteins* 78(15):3073–3084, 2010
- [50] MINTSERIS, J.; WENG, Z.: Structure, function, and evolution of transient and obligate protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* 102(31):10930–10935, 2005
- [51] CLACKSON, T.; WELLS, J.A.: A hot spot of binding energy in a hormone-receptor interface. *Science* 267(5196):383–386, 1995
- [52] MIYAZAWA, S.; JERNIGAN, R.L.: An empirical energy potential with a reference state for protein fold and sequence recognition. *Proteins* 36(3):357–369, 1999
- [53] ZHU, F.; HAN, B.; KUMAR, P.; LIU, X.; MA, X.; WEI, X.; HUANG, L.; GUO, Y.; HAN, L. ET AL.: Update of TTD: Therapeutic Target Database. *Nucleic Acids Research* 38(Database issue):D787–791, 2010
- [54] CRISMAN, T.J.; PARKER, C.N.; JENKINS, J.L.; SCHEIBER, J.; THOMA, M.; KANG, Z.B.; KIM, R.; BENDER, A.; NETTLES, J.H. ET AL.: Understanding false positives in reporter gene assays: in silico chemogenomics approaches to prioritize cell-based HTS data. *Journal of Chemical Information and Modeling* 47(4):1319–1327, 2007
- [55] SÁNCHEZ CLAROS, C.; TRAMONTANO, A.: Detecting Mutually Exclusive Interactions in Protein-Protein Interaction Maps. *PLoS ONE* 7(6):e38765, 2012
- [56] THE 1000 GENOMES PROJECT CONSORTIUM: A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061–1073, 2010

---

## Abbildungsverzeichnis

Abbildung 1.1:	Illustration des Grundgedankens der Systembiologie. ....	2
Abbildung 1.2:	Geschätzte Größe der Protein-Protein-Interaktionsnetzwerke für vier eukaryotische Organismen. ....	3
Abbildung 1.3:	Schema des Hefe-Zwei-Hybrid-Systems am Beispiel des Transkriptionsfaktors GAL4. ....	4
Abbildung 1.4:	Schematische Skizze des Aufbaus eines Proteinkristalls. ....	5
Abbildung 1.5:	Spezifische Bindung des Arzneistoffs Tamoxifen (blau) an sein Zielprotein, den humanen Estrogenrezeptor (gelb). ....	7

## Abkürzungsverzeichnis

<b>Abkürzung</b>	<b>Bedeutung</b>
3D	Dreidimensional
ATC	Anatomisch Therapeutisch Chemisch
CAPRI	Critical Assessment of Prediction of Interactions
DNA	Desoxyribonukleinsäure
EC	Enzyme Commission
GraphML	Graph Markup Language
KEGG	Kyoto Encyclopedia of Genes and Genomes
MHC	Major Histocompatibility Complex
PDB	Protein Data Bank
RNA	Ribonucleinsäure
SNP	Single Nucleotide Polymorphism
WHO	World Health Organisation
XGMML	Extensible Graph Markup and Modeling Language
XML	Extensible Markup Language

# Anhänge

## Anhang A Danksagung

An dieser Stelle möchte ich einigen Menschen Dank sagen, deren Unterstützung beim Erstellen dieser Arbeit wichtig war.

Zuerst bedanke ich mich bei PD Dr. Robert Preißner dafür, dass er mir die Möglichkeit gegeben hat, diese Arbeit in seiner Arbeitsgruppe anzufertigen. Des Weiteren danke ich Prof. Dr. Udo Heinemann für die Übernahme der Zweitgutachterschaft meiner Dissertation.

Dank verdienen auch meine aktuellen und ehemaligen Arbeitskollegen; zum Teil für die Zusammenarbeit in Projekten und an Veröffentlichungen, zum Teil auch einfach nur für nette Gespräche und spannende Diskussionen: Stefan Günther, Mathias Dunkel, Anna Lena Wölke, Manuela Murgueitio, Catherine Worth, Roza Parol-Kryger, Robert Adams, Elke Michalsky, Hellmuth Meyer, Nikolai Hecker, Björn Gohlke, Jessica Ahmed, Ulrike Schmidt, Raphael Bauer, Julia Hossbach, Ines Jäger, Paul Thaben, Thomas Meinel, Robert Lehmann und Karel Mácha.

Ich danke außerdem den Graduiertenschulen, deren Mitglied ich sein durfte: dem internationalen Graduiertenkolleg „Genomics and Systems Biology of Molecular Networks“ und dem Graduiertenkolleg „Computational Systems Biology“. Sie haben es mir ermöglicht, an internationalen Konferenzen und anderen Lehrangeboten teilzunehmen, und den Großteil dieser Arbeit finanziert.

Abschließend möchte ich meiner Familie danken, die mich immer unterstützt hat: Vielen Dank an meine Eltern, meine Geschwister, meine Schwiegereltern und ganz besonders an meine Frau Lena.

## **Anhang B Lebenslauf**

Der Lebenslauf wird aus Gründen des Datenschutzes nicht veröffentlicht.



# Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Dissertationsschrift selbständig verfasst und ausschließlich die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, 19. September 2012

Joachim von Eichborn