

# Wissensbasierte Analyse von Biomolekülen zur Vorhersage ihrer Wechselwirkungen

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

Stefan Günther

aus Gelsenkirchen

August, 2008

1. Gutachter: PD Dr. Robert Preißner (Charité Berlin)
2. Gutachter: Prof. Dr. Udo Heinemann (Freie Universität Berlin)

Disputation am 13.11.2008

## Abstract

Today, the sequences of most of all human genes are investigated, but the function of many of them is still not known or only roughly described. Interactions between macromolecules or metabolites form a molecular network that reflects the high complexity of cellular metabolism. A large fraction of those interactions is not yet identified and is subject to current research.

The present thesis describes methods of bioinformatics that are able to predict the interactions of biomolecules. For this purpose, the mutual recognition between proteins and other macromolecules as well as proteins and small molecules are analyzed. The presented computational methods of evaluation and analysis are applied to given experimental data. Publicly available data regarding the interactions between biomolecules increased conspicuously during the last two decades. Therefore, the extraction of relevant data from the flood of available information and its transfer to suitable databases is also described. It is shown, how new interactions between molecules can be derived automatically from literature. An important tool for this approach is the detection of similarities between different molecules. Several types of similarity comparisons are utilized. This includes not only the detection of similarities between gene or protein sequences but also the structural comparison of small molecules. The derivation of mutual interaction based on the 3D-structures of macromolecules is subject of the last part of this thesis. It is shown how the binding modes of macromolecules are predictable by the detection of local structural similarities between different protein interfaces.

### Keywords:

protein-ligand interaction, protein-protein interaction, protein-DNA/RNA interaction, structure-based modeling, drugs

## Zusammenfassung

Heute sind zwar die meisten der Gene des Menschen entschlüsselt, aber ein sehr großer Teil davon ist gar nicht oder nur in Ansätzen funktionell klassifiziert. Durch Wechselwirkungen von Proteinen mit anderen Makromolekülen oder Metaboliten formt sich ein molekulares Netzwerk, welches der hohen Komplexität des zellulären Stoffwechsels Rechnung trägt. Ein Großteil dieser Wechselwirkungen ist noch nicht bekannt und ist Gegenstand aktueller Forschung.

In der vorliegenden Arbeit wird beschrieben, wie sich mit Hilfe bioinformatischer Methoden Funktionen und Wechselwirkungen von Biomolekülen vorhersagen lassen. Dazu werden die Wechselwirkungen zwischen Proteinen und anderen Makromolekülen sowie zwischen Proteinen und Kleinstrukturen untersucht. Es kommen dabei Auswertungs- und Analysemethoden der Informatik zum tragen, die auf experimentelle Daten angewendet werden.

Die frei zur Verfügung stehende Datenmenge ist in den letzten zwei Jahrzehnten sehr stark angestiegen, ein Teil der Arbeit beschreibt daher, wie aus der Datenflut relevante Informationen extrahiert und in geeignete Datenbanksysteme überführt werden können. Es wird gezeigt, wie sich auf Grund von bekannten, in der Literatur beschriebenen Wechselwirkungen, neue Interaktionen zwischen Molekülen ableiten lassen. Ein wichtiges Werkzeug bei diesem Vorgehen ist die Aufdeckung von Ähnlichkeiten zwischen verschiedenen Molekülen. Das können Ähnlichkeiten zwischen Gen- oder Proteinsequenzen aber auch strukturelle Ähnlichkeiten zwischen Kleinstrukturen sein. Die Ableitung von Wechselwirkungen basierend auf den 3D-Strukturen von Makromolekülen ist Gegenstand des letzten Teils der Arbeit. Es wird gezeigt, wie sich die Bindungskonformationen zwischen Makromolekülen bestimmen lassen, indem lokale, strukturelle Ähnlichkeiten verschiedener Protein-Interaktionsflächen aufgedeckt werden.

### **Schlagwörter:**

Protein-Ligand Wechselwirkung, Protein-Protein Wechselwirkung, Protein-DNS/RNS Wechselwirkung, Strukturbasierte Modellierung, Medikamente



# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>  | <b>1</b>  |
| 1.1      | Molekülwechselwirkungen als Grundlage für die Funktion der Zelle . . . . .                               | 1         |
| 1.2      | Molekülwechselwirkungen in der Medizin . . . . .   | 3         |
| 1.3      | Messung der Wechselwirkungen zwischen Biomolekülen . . . . .   | 4         |
| 1.4      | Vorhersage der Wechselwirkungen zwischen Biomolekülen . . . . .  | 5         |
| 1.4.1    | Das Datamining - Auswertung der Daten . . . . .  | 6         |
| 1.4.2    | Vergleich von Kleinmolekülen . . . . .   | 7         |
| 1.4.3    | Untersuchung von Proteinstrukturen . . . . .   | 8         |
| 1.5      | Zielsetzung der Arbeit . . . . .   | 12        |
| <b>2</b> | <b>Methoden und Ergebnisse</b>   | <b>13</b> |
| 2.1      | Datenintegration am Beispiel von Wechselwirkungen zwischen Wirkstoffen und ihren Zielproteinen . . . . . | 13        |
| 2.1.1    | Beschreibung der Software “SuperTarget” . . . . .  | 13        |
| 2.1.2    | Beispielanwendung . . . . .  | 14        |
| 2.1.3    | Originalartikel: “SuperTarget and Matador: resources for exploring drug-target relationships” . . . . .  | 16        |
| 2.2      | Strukturbasierte Vorhersage der Wechselwirkungen von Kleinmolekülen . . . . .                            | 20        |
| 2.2.1    | 2D-Strukturvergleich mit Hilfe struktureller Fingerabdrücke . . . . .                                    | 20        |
| 2.2.2    | 2D-Strukturvergleich zur Vorhersage von Haptenen . . . . .   | 21        |
| 2.2.3    | Originalartikel: “Superhaptens and their potential role in contact dermatitis.” . . . . .                | 22        |
| 2.2.4    | Originalartikel: “SuperHapten: a comprehensive database for small immunogenic compounds” . . . . .       | 29        |
| 2.2.5    | Vorhersage der medizinischen Wirkung sowie der Zielproteine von Kleinmolekülen . . . . .                 | 34        |
| 2.2.6    | Originalartikel: “SuperPred: drug classification and target prediction” . . . . .                        | 35        |

|          |  |           |
|----------|--|-----------|
| 2.2.7    | 3D-Strukturvergleiche von gebundenen Wirkstoffen und generierten Konformeren . . . . .   | 40        |
| 2.2.8    | Originalartikel: "Representation of target-bound drugs by computed conformers" . . . . .   | 42        |
| 2.3      | Untersuchung der Wechselwirkungen zwischen Makromolekülen  | 53        |
| 2.3.1    | Vorhersage von Proteininteraktionen auf Grund der Faltung . . . . .  | 53        |
| 2.3.2    | Originalartikel: "Legionella pneumophila induces IFN-beta in lung epithelial cells via IPS-1 and IRF3, which also control bacterial replication" . . . . . | 55        |
| 2.3.3    | Analyse der strukturellen Flexibilität von Makromolekülen . . . . .  | 62        |
| 2.3.4    | Originalartikel: "Molecular flexibility in protein-DNA interactions" . . . . .   | 63        |
| 2.3.5    | Vorhersage der Protein-Protein Bindungskonformation  | 74        |
| 2.3.6    | Originalartikel: "ISEARCH-prediction of interactions using known interfaces" . . . . .   | 77        |
| <b>3</b> | <b>Diskussion</b>  | <b>83</b> |
| 3.1      | Datenintegration - eine Methode zur Bestimmung von Molekülwechselwirkungen . . . . .   | 83        |
| 3.2      | Strukturbasierte Ableitung der Bioaktivität von Kleinmolekülen   | 84        |
| 3.3      | Strukturbasierte Vorhersage von Protein-Protein-Interaktionen  | 86        |
| 3.4      | Vorhersage der Protein-Bindungskonformation . . . . .  | 88        |
| 3.5      | Schlussfolgerungen . . . . .   | 89        |

# Abbildungsverzeichnis

|           |   |    |
|-----------|---|----|
| Abb. 1.1: | Geschätzte Anzahl an Protein-Protein-Interaktionen (Interaktome) in verschiedenen eukaryotischen Organismen. . . .                        | 2  |
| Abb. 1.2: | Schema des Hefe-Zwei-Hybrid-Systems. . . . .  | 5  |
| Abb. 1.3: | Verschiedene Typen von Molekülstrukturmodellen des Wirkstoffs Diclofenac. . . . .   | 7  |
| Abb. 1.4: | Die erste dreidimensionale Struktur eines Proteins (Myoglobin), welches mit Hilfe der Röntgenkristallographie bestimmt wurde [1]. . . . . | 9  |
| Abb. 1.5: | Beispiele verschiedener Architekturen von Proteinen . . . .   | 10 |
| Abb. 1.6: | Schema einer Bindung nach dem Schlüssel-Schloss-Prinzip. .  | 11 |
| Abb. 2.1: | Schema der relationalen Datenbank von SuperTarget, einer Anwendung zur Analyse der Wirkmechanismen von Medikamenten. . . . .              | 14 |
| Abb. 2.2: | Struktur von Celecoxib gebunden an Carboanhydrase. . . .  | 40 |
| Abb. 3.1: | Beispiele für die Elektrophilie von Haptenen. . . . .   | 85 |

# Tabellenverzeichnis

|           |   |    |
|-----------|---|----|
| Tab. 2.1: | Trefferrate von “ZDock”, ein Programm zur Modellierung der Proteinbindungskonformation. . . . . | 75 |
| Tab. 2.2: | Trefferrate des Programms “ISEARCH” . . . . .   | 75 |

# Kapitel 1

## Einleitung

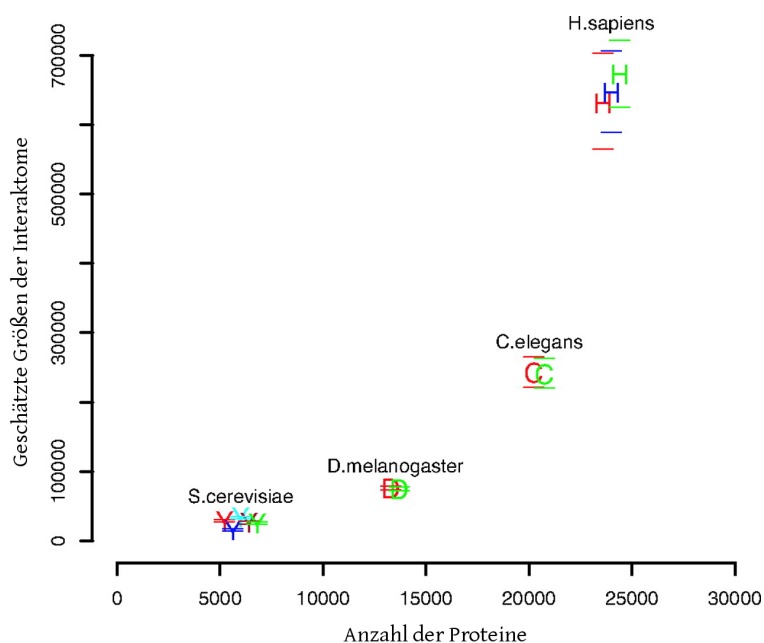
### 1.1 Molekülwechselwirkungen als Grundlage für die Funktion der Zelle

Neben dem genetischen Code des Menschen sind heute die Genome von etwa 750 weiteren Organismen komplett entschlüsselt <sup>1</sup>. Der genetische Code liefert die Rohdaten für das Verständnis des Lebens. Wir sind allerdings noch weit davon entfernt, diese Datenflut sinnvoll und vollständig interpretieren zu können. Die Ursache dafür liegt in der hohen Komplexität des molekularen Interaktionsnetzwerks begründet, welches dem zellulären Stoffwechsel zu Grunde liegt. Teilweise sind einzelne Teilnehmer in diesem Netzwerk erst seit kurzem bekannt geworden oder noch gar nicht identifiziert. So wurden beispielsweise 1993 erstmals microRNAs beschrieben [2], kurze Ribonukleinsäure (RNA)-Moleküle, die für die Genregulation verantwortlich sein können. Sie spielen eine entscheidende Rolle bei der menschlichen Entwicklung und deren Fehlfunktion ist mögliche Ursache zahlreicher Krankheiten [3]. Heute sind mehrere Hundert microRNAs im Menschen bekannt, aber über deren genaue Anzahl kann nur spekuliert werden [4].

Die meisten der Gene des Menschen sind jedoch mittlerweile identifiziert und teilweise auch sehr gut untersucht. Deren Übersetzung in Aminosäureketten, die Proteine, sind die Grundbausteine aller Zellen. Sie sind für die Struktur der Zelle verantwortlich, regeln den Transport von Metaboliten, sind die Hauptbestandteile von Signalwegen und katalysieren Tausende unterschiedlichster Reaktionen. Parameter wie z.B. die katalytische Reaktionsgeschwindigkeit einzelner Enzyme sind experimentell ermittelt worden und gehen heute ein in die noch junge Disziplin der Systembiologie, die ver-

---

<sup>1</sup>GenBank - Statistikseite (Juli 2008): <http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>



**Abb. 1.1:** Geschätzte Anzahl an Protein-Protein-Interaktionen (Interaktome) in verschiedenen eukaryotischen Organismen. Abbildung aus [8].

sucht biologische Organismen in ihrer Gesamtheit zu verstehen. Für die mathematische Modellierung eines biologischen Systems ist es notwendig, die meisten der relevanten beteiligten Moleküle, deren biochemischen Eigenschaften und Konzentrationen zu kennen. Zu diesem Zweck werden heute einfache Modellorganismen ausgewählt, welche von verschiedenen Arbeitsgruppen detailliert untersucht werden [5]. So ist beispielsweise der Kohlenstoffmetabolismus des Bakteriums *Escherichia coli* Gegenstand intensiver Forschung [6]. In dem weitaus komplexeren Organismus *Homo sapiens* sind nur wenige Stoffwechsel- oder Signalwege ähnlich gut beschrieben wie in *E. coli*. Aktuelle Untersuchungen gehen davon aus, dass heute weniger als 0,3% aller Protein-Protein-Interaktionen des Menschen bekannt sind [7]. Die Gesamtzahl aller Protein-Protein-Interaktionen (Interaktom) wird auf 650.000 geschätzt (vgl. Abb. 1.1). Die Identifizierung von Interaktionen zwischen Proteinen und anderen biologisch relevanten Molekülen, wie Nukleinsäuren oder Stoffwechselprodukten (Metabolite) ist ebenfalls Gegenstand aktueller Forschung und bildet die Grundlage für das Verständnis zellulärer Prozesse.

## 1.2 Molekülwechselwirkungen in der Medizin

Arzneistoffe müssen in der Lage sein, spezifisch mit bestimmten Molekülen wie Proteinen in Wechselwirkung zu treten, ohne dabei lebenswichtige Enzymsysteme zu beeinträchtigen. Bildlich kann man sich somit einen Arzneistoff als Schlüssel vorstellen, der nur in ein oder einige wenige Schlösser von vielen möglichen passen darf. Als Schlösser oder biologische Zielstrukturen (engl. Targets) für Arzneistoffe kommen Proteine, DNS, RNS und Membranlipide in Frage. Die Art und Spezifität der Interaktion wird durch die chemisch-physikalischen Eigenschaften der Wirkstoffe bestimmt. Durch die vielen möglichen Interaktionsteilnehmer ist es allerdings sehr aufwändig, einen Wirkstoff zu identifizieren, der durch seine chemischen Eigenschaften an ein gewünschtes Zielprotein bindet und dort seine spezifische Wirkung, wie z.B. die Inhibierung eines Enzyms, entfaltet.

Sehr häufig sind Nebenwirkungen eines Medikamentes darauf zurückzuführen, dass neben dem eigentlichen Target weitere Proteine mit dem Wirkstoff interagieren und dadurch in ihrer Funktion unerwünscht verändert werden. Ein bekanntes Beispiel ist der Wirkstoff Acetylsalicylsäure (Aspirin). Der Wirkstoff inhibiert die Enzyme Cyclooxygenase-1 (COX-1) und Cyclooxygenase-2 (COX-2). COX-2 ist verantwortlich für die Bildung von hormonähnlichen Entzündungsstoffen, den Prostaglandinen. Durch die Inhibierung kommt es daher zu einer gewünschten Entzündungshemmung. Die Inhibierung von COX-1 führt jedoch ebenfalls zu einer Unterdrückung von Prostaglandinen, die für den natürlichen Schutz des Magen-Darm-Traktes und der Nierenfunktion Bedeutung haben. Als häufige Nebenwirkungen von Aspirin treten daher Magengeschwüre oder Nierenfunktionsverschlechterungen auf. Eine Weiterentwicklung des Wirkstoffs führte zu selektiven COX-2-Hemmern (Coxibe), welche nicht mehr die für Aspirin typischen Nebenwirkungen aufwiesen. Die hohen Erwartungen an diese Substanzklasse konnten allerdings nicht erfüllt werden, da die Einnahme von Coxiben ein erhöhtes Risiko für kardiovaskuläre Erkrankungen mit sich bringt. Medikamente, wie z.B. Vioxx, mussten auf Grund dieses Risikos wieder vom Markt genommen werden [9].

Das Wissen über die Wechselwirkungen von Medikamenten mit den Zielstrukturen und anderen wirkungsrelevanten Proteinen (engl. off-Target) ist daher ein wichtiges Ziel der Wirkstoff-Forschung. Dabei treten in den einzelnen Entwicklungsschritten eines Wirkstoffs unterschiedliche Fragestellungen auf: Welcher Wirkstoff ist geeignet, die Funktion eines gegebenen Targets zu verändern? Welche off-Targets gibt es, bzw. wie spezifisch ist der Wirk-

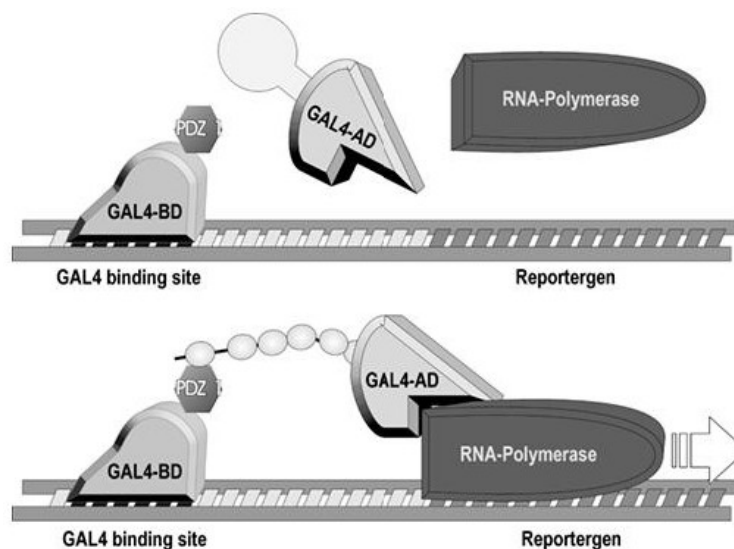
stoff? Wie lässt sich gegebenenfalls die Spezifität des Wirkstoffs erhöhen? Solche Fragestellungen sind meistens sehr schwierig zu beantworten und erfordern heute sehr aufwändige *in vitro* und *in vivo* Untersuchungen. Theoretische Analysen, die helfen, teure Labor-Experimente und Tierversuche einzuschränken, sind daher von hohem wirtschaftlichem, wissenschaftlichem und ethischem Nutzen.

Medikamente und viele andere Kleinstrukturen können aber auch als Haptene wirken und durch ihre Wechselwirkungen mit Antikörpern oder T-Zellen gesundheitsschädigend wirken. Etwa zehn Prozent der Nebenwirkungen von Medikamenten werden auf diesen Effekt zurückgeführt [10]. Haptene sind kleine Moleküle, die eine Immunreaktion auslösen können, wenn sie an spezielle Trägerproteine binden [11]. Dadurch können sie eine Reihe von Autoimmunkrankheiten, wie Kontaktdermatitis oder Asthma, auslösen. Haptene treten in unterschiedlichsten Bereichen der Umwelt auf und können sowohl natürliche Substanzen sein als auch anthropogene organische Verbindungen, die in der Natur nicht vorkommen. Durch die EU-Verordnung "REACH" sind die Hersteller und Importeure von Chemikalien seit Juni 2007 verpflichtet, deren gesundheitliche Unbedenklichkeit für die Verbraucher nachzuweisen (<http://www.reach-info.de>). Methoden, die eine mögliche immunogene Wirkung vorhersagen können und das gesundheitliche Risiko neu synthetisierter Substanzen abschätzen können, würden daher sehr dabei helfen, eine Vielzahl heutiger Laboranalysen zu ersetzen.

### 1.3 Messung der Wechselwirkungen zwischen Biomolekülen

Zur experimentellen Messung von molekularen Interaktionen gibt es heute verschiedene Verfahren, die je nach Art der zu untersuchenden Moleküle und der verwendeten Reaktionsumgebung (*in vivo/in vitro*) mehr oder weniger aufwändig sind.

*In vitro* Messungen zwischen Proteinen und Kleinmolekülen, die bei der Wirkstoffentwicklung eine wichtige Rolle spielen, werden heute vielfach mit Hilfe der Oberflächenplasmonenresonanz-Spektroskopie durchgeführt. Dazu werden mittels polarisierten Lichtes die freien Elektronen an einem Sensorchip immobilisierter Moleküle angeregt. Der zu testende Reaktionspartner wird mittels eines Fließinjektionssystems zugeführt und verändert bei der Bindung die Resonanz der Elektronen und damit die Reflexion des Laserlichtes. Durch die Messung der Veränderung kann die Interaktion qualitativ (Spezifität des Wirkstoffs) und quantitativ (Affinität des Wirkstoffs) gemes-



**Abb. 1.2:** Schema des Hefe-Zwei-Hybrid-Systems. Abbildung aus Wikipedia.

sen werden.

Für die qualitative Messung von Protein-Protein Wechselwirkungen wird häufig das sogenannte Hefe-Zwei-Hybrid-System (engl. Yeast Two-Hybrid System) verwendet. Dazu werden zwei Teile (Domänen) eines Transkriptionsfaktors (GAL4) der Hefe an die zu messenden interagierenden Proteine gekoppelt. Durch die Interaktion der Proteine werden die beiden Domänen wieder zusammengeführt und der Transkriptionsfaktor wieder funktionsfähig. Die Aktivierung löst die Transkription eines Reportergens aus, welches wiederum detektiert werden kann (vgl. Abb. 1.2).

Neben diesen beiden Labormethoden gibt es viele weitere unterschiedliche experimentelle Verfahren, um Wechselwirkungen zwischen Biomolekülen zu messen. Sie besitzen jedoch alle die Gemeinsamkeit, dass sie ein modernes Labor benötigen und die Durchführung teuer und zeitaufwändig ist. Eine Vorhersage eines Teils dieser Ergebnisse mit Hilfe eines theoretischen Modells, würde daher viel Zeit und Kosten einsparen.

## 1.4 Vorhersage der Wechselwirkungen zwischen Biomolekülen

Die interdisziplinäre Forschungsrichtung Bioinformatik löst Probleme aus den Lebenswissenschaften mit theoretischen Methoden. Einen ungeahnten Boom



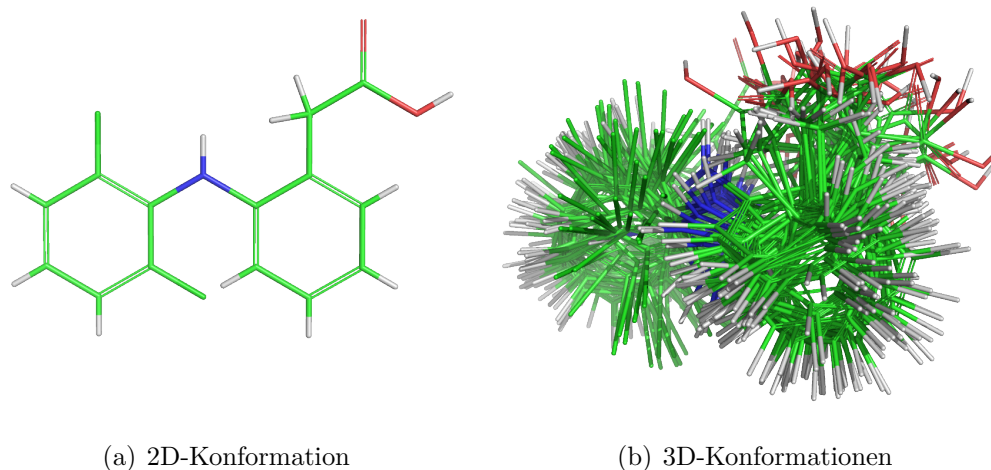
erlebte die Bioinformatik in den 90er Jahren durch die Entwicklung neuer Algorithmen, die für die Entschlüsselung des menschlichen Genoms benötigt wurden. Heute ist sie ein unverzichtbares Werkzeug der Biologie, der Medizin und einiger verwandter Forschungsgebiete. Mit bioinformatischen Methoden lassen sich auch Vorhersagen zu physiologischen Prozessen in der Zelle treffen. Dabei werden vorhandene experimentell gewonnene Daten verwendet (Lernmenge), mit deren Hilfe eine auf bestimmte Zwecke ausgerichtete vereinfachende Beschreibung der Wirklichkeit (Modell) erstellt wird. In einem zweiten Schritt wird mit Hilfe einer Testmenge geprüft in wie weit sich mit Hilfe des Modells Prognosen über Ereignisse ableiten lassen, die nicht in der Lernmenge enthalten waren (Validierung). Bei ausreichender Vorhersagegüte lassen sich im letzten Schritt echte Vorhersagen mit abschätzbarer Fehlerquote über biologische Prozesse treffen.

Wissensbasiert (engl. knowledge-based) bedeutet in diesem Zusammenhang, dass bei einer Vorhersage Ähnlichkeiten zu einem bereits beobachteten Ereignis genutzt werden. Soll beispielsweise vorhergesagt werden, ob zwei Moleküle miteinander interagieren, so könnten gemessene Bindungsaffinitäten zweier sehr ähnlicher Moleküle zur Vorhersage genutzt werden, falls solche Daten vorhanden sind. Solche Methoden haben manchmal den Nachteil, dass sie nicht immer die Ursache für ein eintretendes Ereignis aufzeigen, sondern nur auf frühere Beobachtungen verweisen können. Es muss jedoch bei der Methodenentwicklung analysiert werden, worin die Ähnlichkeiten zwischen den Beobachtungen bestehen. So muss z.B. bei der Molekülinteraktion bestimmt werden, welche vergleichbaren chemischen Eigenschaften für ein ähnliches Interaktionsverhalten wichtig sind.

Die biologischen Sachgebiete mit denen sich wissensbasierte Vorhersagemethoden befassen, sind so vielfältig, wie das Spektrum der biologischen Laborergebnisse. In der vorliegenden Arbeit werden unterschiedliche Arten von Molekülinteraktionen und Vorhersagemethoden verwendet, die im folgenden kurz vorgestellt werden.

### **1.4.1 Das Datamining - Auswertung der Daten**

Ein derzeitiges Problem biologischer Labordaten besteht darin, dass sie verstreut in vielen Datenbanken oder Fachartikeln gespeichert bzw. beschrieben sind. Durch die Verknüpfung dieser Daten lassen sich neue Wirkungsbeziehungen herleiten, die sich nicht durch die Betrachtung der Einzelergebnisse ableiten lassen (engl. Datamining). In Kapitel 2.1 wird gezeigt, wie sich Wechselwirkungen zwischen Medikamenten und Proteinen bzw. Nukleinsäuren ableiten lassen, die noch nicht in der Literatur beschrieben sind. Hierfür wurden mehrere Millionen Zusammenfassungen von Fachartikeln automati-



**Abb. 1.3:** Verschiedene Typen von Molekülstrukturmodellen des Wirkstoffs Diclofenac. Visualisiert mit PyMol.

siert durchsucht und Beschreibungen von Wechselwirkungen zwischen Medikamenten und deren Zielproteinen extrahiert. Die Wechselwirkungen wurden in eine Datenbank überführt, mit weiteren Datenbanken verknüpft und mit Analysewerkzeugen ausgestattet. Das Ergebnis ist eine frei zugängliche, umfassende Web-Applikation mit deren Hilfe sich die Wirkungen von Medikamenten im Menschen aufzeigen und analysieren lassen. Durch Ähnlichkeitsvergleiche zwischen den Medikamenten bzw. Proteine, können darüberhinaus auch neue Wechselwirkungen prognostiziert werden.

### 1.4.2 Vergleich von Kleinmolekülen

Ein häufig verwendetes Werkzeug, um Aussagen über die Wirkung eines Biomoleküls treffen zu können, ist die Ableitung des biologischen Effektes über zweidimensionale (2D) Ähnlichkeit des Moleküls zu besser untersuchten Substanzen. Kapitel 2.2.2 zeigt diesen Zusammenhang am Beispiel von Haptenen. Durch den Vergleich von gut untersuchten Haptenen mit Millionen frei bestellbarer Substanzen lässt sich abschätzen, welche dieser Substanzen womöglich immunreaktiv sind und daher mit besonderer Vorsicht verwendet werden sollten. Das Ergebnis ist eine frei zu verwendende Internetdatenbank, in der die geprüften und die potentiellen Haptene beschrieben sind.

Auf ähnlicher Weise zeigt Kapitel 2.2.6 an Hand einer Internetanwendung, wie sich die Interaktionspartner und damit die medizinische Wirkung vieler Substanzen über deren chemischen Eigenschaften vorhersagen lassen.

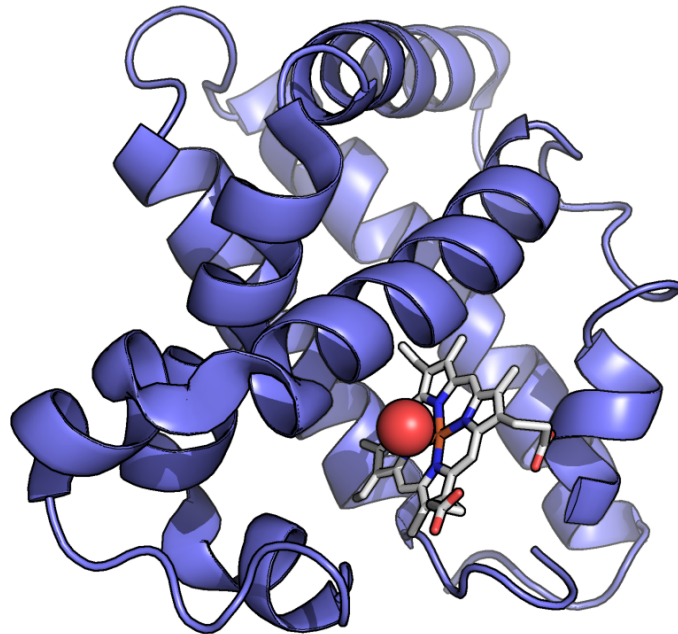
Soll eine mögliche medizinische Applikation einer Substanz bestimmt werden, so wird in einem ersten Schritt überprüft, ob die Substanz allgemeinen chemischen Eigenschaften genügt, die für Wirkstoffe typisch sind. So kommen beispielsweise Moleküle mit geringer Lipophilie bzw. geringem logP-Wert nur selten als Wirkstoffe in Betracht, da sie kaum membrangängig sind [12]. In einem zweiten Schritt wird ermittelt, ob die Substanz zweidimensionale Ähnlichkeiten zu bekannten Wirkstoffen besitzt. Falls solche Ähnlichkeiten detektiert werden, können die möglichen Interaktionspartner mit hinreichender Genauigkeit vorhergesagt werden.

Die Abstraktion eines Biomoleküls in eine planare 2D-Struktur ist für einen schnellen Ähnlichkeitsvergleich zwischen Molekülen sehr zweckmäßig. Die Wirkung des Moleküls wird aber durch die Interaktion mit Makromolekülen und somit durch die 3D-Struktur des Wirkstoffs bestimmt. Wirkstoffe und andere Kleinstrukturen besitzen meistens mehr oder weniger viele drehbare Bindungen, die in einer unterschiedlich großen Anzahl möglicher dreidimensionaler Konformationen resultieren können (vgl. Abb. 1.3). In einem statischen Modell, welches die Interaktion mit einem Protein simuliert oder vorhersagt, ist die 3D-Struktur des Moleküls abhängig von der Bindungstasche. Die in Frage kommenden verschiedenen Konformationen des Moleküls müssen daher im Modell berücksichtigt werden. Kapitel 2.2.7 stellt eine umfassende Analyse vor, die beschreibt, wie viele statische Konformere generiert werden müssen, um die an das Zielprotein gebundene Konformation mit hinreichender Genauigkeit abbilden zu können.

### 1.4.3 Untersuchung von Proteinstrukturen

Durch die Röntgenstrukturanalyse gelang es 1959 erstmals eine komplette dreidimensionale Struktur eines Proteins aufzuklären [13]. 1962 erhielten Max Ferdinand Perutz zusammen mit John Cowdery Kendrew den Nobelpreis für die Bestimmung der Molekülstruktur des Hämoglobins und des Myoglobins (Abb. 1.4). Heute sind über 50.000 Proteinstrukturen aufgeklärt und in der Protein Datenbank PDB [14] beschrieben. Die Struktur eines weiteren großen Anteils aller Proteine lässt sich ausgehend von den bekannten Strukturen modellieren, wenn hinreichende Sequenzähnlichkeiten zwischen diesen vorliegen [15].

Die Funktion eines Proteins wird von den chemischen Eigenschaften des gefalteten Proteins diktiert, daher kann mit Hilfe der Proteinstruktur theoretisch auch bestimmt werden, welche Moleküle mit dem Protein interagieren. Allerdings ist eine reine strukturbasierte Vorhersage der Bindungspartner sehr aufwändig, da die Anzahl der in Frage kommenden Moleküle meistens sehr groß ist und die Simulation der molekularen Dynamik einer einzelnen

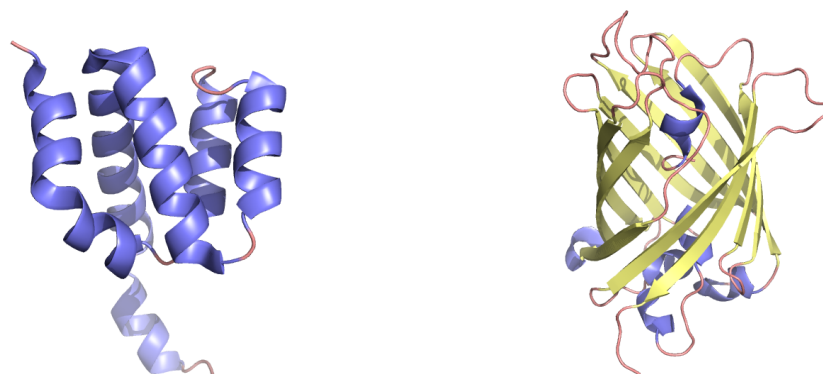


**Abb. 1.4:** Die erste dreidimensionale Struktur eines Proteins (Myoglobin), welches mit Hilfe der Röntgenkristallographie bestimmt wurde [1]. Visualisiert mit PyMol: einkettiges Protein - blau; eisenhaltiges Protoporphyrin - Stabdarstellung; gebundenes  $OH^-$  - rot. Abbildung ohne Wasserstoffe.

Interaktion bereits extrem rechenaufwändig ist. In Einzelfällen konnten jedoch die Bindungspartner von Proteinen mit Hilfe der Struktur vorhergesagt werden, wie z.B. die kürzlich geglückte Aufklärung der Interaktion des Proteins Tm0936 mit dem Metaboliten S-inosylhomocysteine in dem Bakterium *Thermotoga maritima* [16].

### Vorhersage anhand der Proteinarchitektur

Strukturinformationen können allerdings kombiniert mit weiteren Informationen zur Funktionsbeschreibung genutzt werden. Es ist z.B. teilweise bekannt, welche Architekturen bzw. Domänen (vgl. Abb. 1.5) häufig miteinander interagieren. Daher lassen sich mit dem Wissen der Domänenstruktur potentielle Bindungspartner aus einer großen Menge herausfiltern. Kapitel 2.3.1 zeigt ein Beispiel für die Suche nach einem Bindungspartner von IPS-1. Das Protein leitet ein Signal weiter, welches bei einer intrazellulären Infizierung mit Bakterien ausgelöst wird und dann eine Immunabwehrreaktion hervorruft. Der dafür notwendige direkte Signalgeber für IPS-1 ist bisher



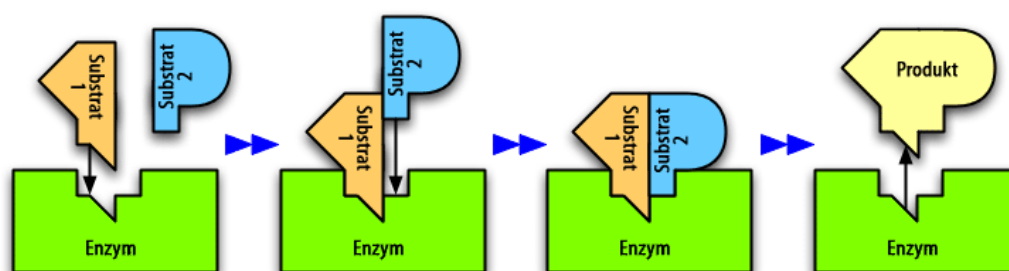
(a) CARD-Domäne. Protein Nod1 (PDB-ID: 2NSN) (b) GFP-Domäne. Grün fluoreszierendes Protein (PDB-ID: 1H6R)

**Abb. 1.5:** Beispiele verschiedener Architekturen bzw. Domänen von Proteinen. Die Domäne entscheidet häufig über mögliche Interaktionspartner der Proteine. Proteine visualisiert mit PyMol.

nicht bekannt. Es wird vermutet, dass ein intrazellulärer Rezeptor die bakterielle DNS erkennt und dieses Signal direkt oder indirekt an IPS-1 weitergibt. Für IPS-1 wurde bereits eine andere Proteinwechselwirkung mit dem Protein MDA5 beschrieben [17]. Dieses Protein bindet an IPS-1 mit Hilfe einer CARD-Domäne (Abb. 1.5a). Für die im Ergebnisteil beschriebene Suche nach dem Signalprotein kamen daher besonders Kandidatengene in Frage, welche eine ähnliche Architektur wie MDA5 aufweisen.

### Analyse der Flexibilität von Makromolekülen

Das oben genannte Schlüssel-Schloss-Prinzip, welches die Wechselwirkungen zwischen Proteinen und Substraten beschreibt, wurde schon 1894 von Emil Fischer beschrieben [18] und ist heute noch immer sehr gut geeignet die Bindungsprinzipien vieler Reaktionen zu erklären (vgl. Abb. 1.6). Es vernachlässigt allerdings bei der Bindung auftretende strukturelle Veränderungen der beteiligten Moleküle. Gerade für geometrisch basierte Verfahren der Interaktionsvorhersage stellen diese Veränderungen ein großes Problem dar. Eine strukturbasierte Vorhersage einer Interaktion basiert normalerweise auf den ungebundenen Formen der Moleküle. Die während der Bindung auftretenden Veränderungen der Konformationen müssen daher im Modell vorhergesagt werden. Je nach Modellierung der an der Reaktion beteiligten Molekültypen müssen unterschiedliche Formen und Ausmaße der strukturellen Veränderungen modelliert werden.



**Abb. 1.6:** Schema einer Bindung nach dem Schlüssel-Schloss-Prinzip: Ein Enzym katalysiert die Reaktion von zwei Substraten, um ein neues Produkt zu erstellen. Abbildung aus Wikipedia.

Kapitel 2.3.3 untersucht u.a. die strukturellen Veränderungen die bei der Bindung zwischen Protein und DNS auftreten können. Es werden die globalen Veränderungen der Makromoleküle an Hand des Rückgrats (engl. Backbone) der Proteine und der DNS untersucht, sowie die lokalen Konformationsänderungen der Seitenketten auf der Oberfläche des Proteins bzw. den Basen der DNS. Auf Grund unterschiedlich auftretender Typen von Konformationsänderungen können die Proteine verschiedenen Interaktionsmodellen wie z.B. dem “induced fit”-Prinzip [19] oder der “Equilibrium Hypothese” [20], zugeordnet werden.

### Strukturbasierte Vorhersage von Molekülinteraktionen

Die strukturbasierte Vorhersage von Interaktionen zwischen Proteinen mit einer tolerierbaren Fehlerquote gelingt heute nur, wenn neben den eigentlichen Strukturmodellen weitere Informationen zur Bindung vorliegen. So könnte beispielsweise geprüft werden, ob Ähnlichkeiten zwischen gegebenen ungebundenen Strukturmodellen und bereits co-kristallisierten Proteinen bestehen. Da zu Komplexen zusammengelagerte Proteine nur sehr schwer zu kristallisieren sind, gibt es heute nur etwa 600 verschiedene Proteinkomplexe in der Datenbank PDB [21]. Auf Grund der niedrigen Anzahl können diese bisher nur selten zur Vorhersage von Proteininteraktionen genutzt werden.

Die Bestimmung des richtigen Bindungsmodus zweier Proteine (engl. Docking), wenn die Interaktion bereits experimentell nachgewiesen wurde, bleibt ebenfalls nicht trivial. Mit diesem Teilproblem zur Interaktionsvorhersage befassen sich zahlreiche Untersuchungen und Modellierungsmethoden. Es gibt heute verschiedene öffentlich zugängliche Programme, die Modelle eines Proteinkomplexes generieren, wenn die beiden Einzelstrukturen im ungebundenen Zustand vorgegeben werden [22, 23, 24, 25]. Solche Programme

arbeiten fast alle *ab initio*, d.h. es werden außer den Strukturen keine zusätzliche Informationen verwendet, um den Proteinkomplex zu modellieren. Sie besitzen allerdings auch eine sehr hohe Fehlerrate bzw. kann der richtige Bindungsmodus häufig nur mit einer hohen Unschärfe bestimmt werden.

Kapitel 2.3.5 stellt einen neuen wissensbasierten Ansatz zur Lösung des Protein-Protein-Docking-Problems vor. Die Methode basiert auf der Hypothese, dass auch strukturell unterschiedliche Protein-Protein-Komplexe dennoch strukturell ähnliche Interaktionsflächen ausbilden können. Diese Beobachtung ist noch relativ neu, kann aber in der vorliegenden, wie auch in einer weiteren Arbeit [26] deutlich bestätigt werden. Dadurch lassen sich auf der Oberfläche der ungebundenen Proteine Ähnlichkeiten zu bereits kristallisierten Protein-Interaktionsflächen detektieren und analog dieser Ähnlichkeiten der richtige Bindungsmodus modellieren. Im Ergebnisteil wird die Modellierung der Bindungskonformation des Ribonuklease-Inhibitor und der Ribonuklease (PDB-ID: 1DFJ) auf Grundlage zweier interagierender Domänen der Fruktose-2,6-bisphosphatase (PDB-ID: 1FBT) demonstriert. Verglichen mit dem häufig verwendeten Modellierungsprogramm “ZDock” [22] erreicht die vorgestellte Methode zur Generierung von Kandidatenmodellen ähnlich gute Ergebnisse. Weitere mögliche Verbesserungen der Methode werden im Schlussteil diskutiert.

## 1.5 Zielsetzung der Arbeit

Insbesondere in der menschlichen Zelle ist zur Zeit nur ein Bruchteil aller molekularer Wechselwirkungen bekannt. Dennoch sind zahlreiche experimentelle Versuchsergebnisse öffentlich zugänglich, die genutzt werden können, um fundierte Prognosen über mögliche Wechselwirkungen aufstellen zu können.

In der vorliegenden Arbeit sollten bioinformatische Methoden entwickelt werden, die es erlauben mit Hilfe vorhandener experimentell bestimmter Daten neue Molekülwechselwirkungen vorherzusagen. Dabei sollten sowohl Klein- als auch Makromoleküle betrachtet werden. Die Gesamtbetrachtung umfasst sowohl die Extraktion relevanter Daten aus der Vielzahl der zur Verfügung stehenden Literatur, eine geeignete Speicherung der verschiedenen Datentypen, sowie die Erstellung und Auswertung verschiedener Verfahren, die zur Wechselwirkungsvorhersage geeignet sind. Ein besonderer Schwerpunkt sollte dabei auf solchen Methoden liegen, welche die räumliche atomare Zusammensetzung der Biomoleküle berücksichtigen.

# Kapitel 2

## Methoden und Ergebnisse

### 2.1 Datenintegration am Beispiel von Wechselwirkungen zwischen Wirkstoffen und ihren Zielproteinen

#### 2.1.1 Beschreibung der Software “SuperTarget”

Das Wissen über die Wirkungsmechanismen und Zielproteine von Medikamenten hat in den letzten zwei Jahrzehnten stark zugenommen. Gleichwohl ist man bei vielen Wirkstoffen weit davon entfernt, die molekularen Prozesse des Wirkeffektes zu verstehen. Einige Medikamente, wie Diloxanid, Niclosamid oder Ambroxol werden heute verabreicht, ohne dass deren Wirkort bzw. das Primärzielprotein bekannt wäre [27]. Ein weiteres Problem besteht darin, dass das vorhandene Wissen über die Wirkung von Medikamenten verstreut in Tausenden von wissenschaftlichen Artikeln und Lehrbüchern beschrieben ist. Dadurch ist es sehr aufwändig, vergleichende Analysen über eine Vielzahl von Wirkstoffen durchzuführen.

Um diesen Sachverhalt zu verbessern, wurde eine Internetanwendung (“SuperTarget”) erstellt, in der 7300 Interaktionen zwischen Medikamenten und deren Zielmoleküle beschrieben sind [28]. Zur Datengewinnung wurden halbautomatisch die Zusammenfassungen von etwa 15 Millionen wissenschaftlichen Artikeln durchsucht. Potentielle Interaktionen zwischen Wirkstoffen und deren Zielproteinen wurden in einem zweiten Schritt manuell auf deren Richtigkeit und Relevanz überprüft und dann gegebenenfalls in die Datenbank aufgenommen.

Dieser Basisdatensatz wurde verknüpft mit zahlreichen weiterführenden Informationen, wie z.B. den Nebenwirkungen der Medikamente, den involvierten Stoffwechselwegen, die für den Abbau verantwortlichen Cytochrome,



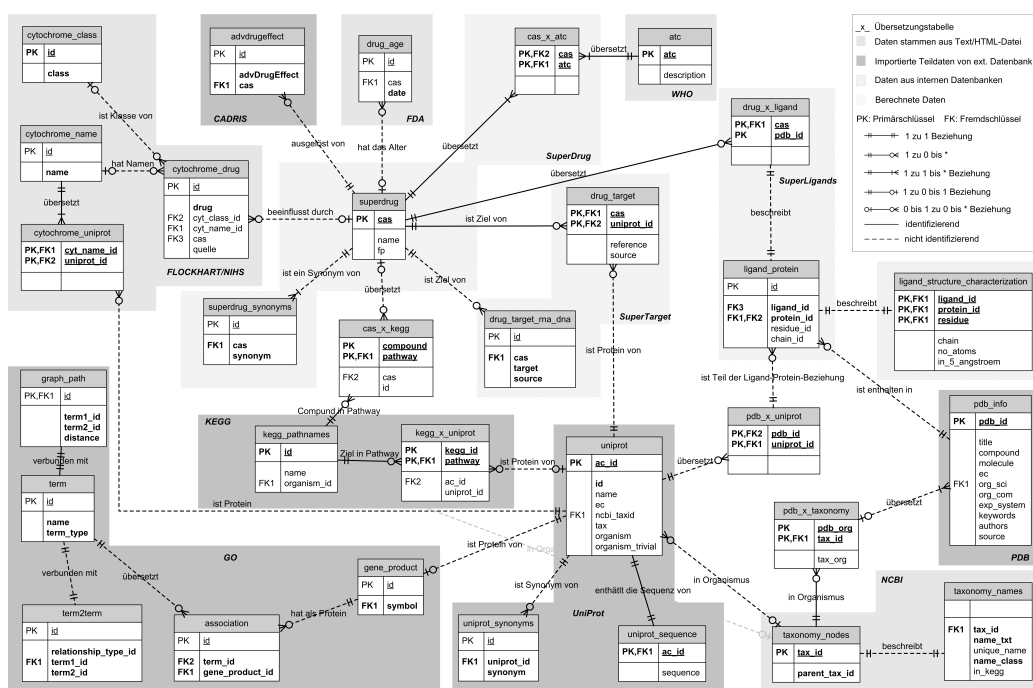


Abb. 2.1: Schema der relationalen Datenbank von SuperTarget, einer Anwendung zur Analyse der Wirkmechanismen von Medikamenten.

die therapeutische WHO-Klassifikation der Wirkstoffe [29] sowie den Sequenzen und Strukturen aller geprüften und potentiellen Zielproteine. Die Gesamtinformationen wurden in eine relationale Datenbank überführt, welche etwa 30 verschiedene Tabellen umfasst (vgl. Abb. 2.1).

Durch die Erweiterung mit unterschiedlichen Analysewerkzeugen, wie dem strukturellen Ähnlichkeitsvergleich zwischen den Wirkstoffen, sowie den sequentiellen Vergleich der Zielproteine, ist es möglich Informationen zu gewinnen, die auch Vorhersagen zu möglichen Molekülinteraktionen erlauben, die noch nicht in der Datenbank enthalten sind.

## 2.1.2 Beispielanwendung

Zahlreiche Nebenwirkungen von Medikamenten lassen sich erklären durch die Interaktionen mit sogenannten off-Targets, Proteine, die den eigentlichen Zielproteinen häufig strukturell ähnlich sind und daher ebenfalls mit dem Medikament interagieren. Durch die Zusammenführung der vorhandenen Informationen lassen sich fundierte Prognosen über mögliche off-Targets erstellen.

So lassen sich beispielsweise mögliche off-Targets für den Wirkstoff Me-

mantin bestimmen. Memantin ist ein Wirkstoff, der in der Behandlung von moderater bis schwerer Demenz vom Alzheimer-Typ eingesetzt wird. Als Antagonist des Glutamaterezeptors sorgt der Wirkstoff für eine erniedrigte Calcium-Leitfähigkeit des Rezeptors und verändert dadurch die Signalgebung von Nervenzellen. Als Nebenwirkungen werden motorische Unruhe, Kopfschmerzen, Müdigkeit, Verwirrtheit, Halluzinationen, Verstopfung, anormaler Gang, Schwindel, Übelkeit, Erbrechen und erhöhte Krampfbereitschaft angegeben [30]. Wird eine Ähnlichkeitssuche mit Memantin als Ausgangsmolekül gestartet (Methodenbeschreibung im folgenden Kapitel), so zeigt die Ergebnisliste einen strukturell ähnlichen Wirkstoff namens Amantadin, das unter anderem in der Parkinson-Therapie eingesetzt wird. Es wirkt agonistisch an Dopamin-2-Rezeptoren im Gehirn. Für Amantadin ist jedoch auch eine Wechselwirkung mit Glutamaterezeptoren beschrieben. Es liegt daher nahe, dass Memantin ebenfalls eine mehr oder weniger starke Wechselwirkung mit dem Dopamin-2-Rezeptor ausbildet und eine Aktivierung des Rezeptors auslösen kann. Der Dopaminrezeptor ist wie der Glutamaterezeptor in den Transport von Neurotransmittern involviert, allerdings vornehmlich Kalium-Ionen. Die Aktivierung durch Memantin könnte einen Teil der oben aufgezählten Nebenwirkungen des Wirkstoffs erklären. Es ist beispielsweise bekannt, dass das Erbrechen durch eine übermäßige Aktivierung von Dopamin-2-Rezeptoren ausgelöst werden kann [30].

### **2.1.3 Originalartikel: “SuperTarget and Matador: resources for exploring drug-target relationships”**

GÜNTHER, STEFAN; KUHN, MICHAEL; DUNKEL, MATHIAS; CAMPILLOS, MONICA; SENGER, CHRISTIAN; PETSALAKI, EVANGELIA; AHMED, JESSICA; URDIALES, EDUARDO G.; GEWIESS, ANDREAS; JENSEN, LARS J.; SCHNEIDER, REINHARD; SKOBLO, ROMAN; RUSSELL, ROBERT B.; BOURNE, PHILIP E.; BORK, PEER; PREISSNER, ROBERT

Nucleic Acids Res 36 (2008), S. D919-D922

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1093/nar/gkm862>

## 2.2 Strukturbasierte Vorhersage der Wechselwirkungen von Kleinmolekülen

### 2.2.1 2D-Strukturvergleich mit Hilfe struktureller Fingerabdrücke

Eine effektive Methode, die Bioaktivität einer Substanz abschätzen zu können, ist der Vergleich zu chemisch ähnlich aufgebauten Substanzen, über die mehr Informationen bekannt sind. Die chemische Ähnlichkeit lässt sich auf verschiedene Weise messen. Ein Vergleich einer Molekülstruktur mit mehreren Millionen Kandidatenmolekülen, deren Strukturen in einer Datenbank gespeichert sind, erfordert ein sehr schnelles Vergleichsverfahren. Eine derzeit häufig benutzte Methode ist die Übersetzung der 2D-Struktur in einen sogenannten strukturellen Fingerabdruck. Dazu werden in einem ersten Schritt alle möglichen Pfade bis zu einer definierten Länge bestimmt, die durch den Graph eines Moleküls gelegt werden können. In einem zweiten Schritt wird für jeden vorhandenen Pfad ein bestimmtes Bit in einem Booleschen Feld gesetzt. Die Bestimmung des zu setzenden Bits erfolgt dabei mit Hilfe einer Hash-Funktion. Auf diesem Weg bekommt zwar jeder Pfad eindeutig ein Bit zugewiesen, jedoch kann ein Bit für viele verschiedene Pfade kodieren. Ein struktureller Fingerabdruck ist daher mit einem gewissen Informationsverlust behaftet und kann nicht mehr ohne weiteres in die ursprüngliche Molekülstruktur zurückübersetzt werden. Die Kombination der auftretenden Pfade ist jedoch eindeutig genug, um das Molekül biochemisch sehr genau zu klassifizieren. Für die in diesem Kapitel durchgeführten Analysen wurden strukturelle Fingerabdrücke verwendet, die mit Hilfe des "Chemistry Development Kit" [31] berechnet wurden.

Die Ähnlichkeit zweier Boolescher Felder kann durch den Tanimoto-Koeffizienten bestimmt werden, der sich als Ähnlichkeitsmaß für Moleküle sehr gut eignet [32] und wie folgt berechnet wird:

$$T = \frac{N_{ab}}{N_a + N_b - N_{ab}}$$

mit  $N_{ab}$  = Anzahl gesetzter Bits in beiden Feldern,  $N_a$  = Anzahl gesetzter Bits in Feld a,  $N_b$  = Anzahl gesetzter Bits in Feld b. Der Koeffizient liegt zwischen Null (keinerlei Ähnlichkeit zwischen den Feldern) und Eins (identische Felder). Ein Tanimoto-Koeffizient größer 0,85 kann als Indikator dafür angesehen werden, dass zwei Moleküle eine ähnliche biologische Aktivität besitzen [33, 34].

### 2.2.2 2D-Strukturvergleich zur Vorhersage von Haptenen

Haptene sind kleine Moleküle, die eine Immunreaktion auslösen können, wenn sie an Trägerproteinen gebunden sind. Bekannte Haptene sind anthropogen hergestellte oder natürliche Substanzen, die eine Reihe von Autoimmunkrankheiten wie Kontaktdermatitis oder Asthma auslösen können. Eine besondere Gruppe der Haptene sind die sogenannten Superhaptene, da sie außergewöhnlich häufig und teilweise sehr starke Immunreaktionen auslösen können. Ein im Zuge dieser Arbeit entstandener Übersichtsartikel beschreibt Haptene im Zusammenhang mit Kontaktdermatitis [35]. Es werden die bekanntesten Haptene und Superhaptene genannt, deren Vorkommen und Wirkmechanismen beschrieben sowie derzeit bekannte Therapiemöglichkeiten aufgezählt.

Eine der größten Datenbanken, die Haptene beschreibt, ist die Datenbank "SuperHapten", die in der Arbeitsgruppe "Strukturelle Bioinformatik" der Charité erstellt wurde [36]. Sie umfasst zur Zeit etwa 2000 Moleküle, für die eine Bindung an Antikörper oder T-Zell-Rezeptoren beschrieben wurde. Um diesen Datensatz zusammenzutragen, wurden die Zusammenfassungen von mehreren Millionen Artikeln automatisiert auf das Vorhandensein bestimmter Schlagwörter durchsucht und gefiltert. Die verbliebene Menge wurde manuell auf eine Beschreibung von spezifischen Haptenen geprüft. Beschriebene Hapten-Wechselwirkungen wurden in eine Datenbank überführt und mit weiteren Informationen, wie beispielsweise den Molekülstrukturinformationen der Haptene, den Literaturquellen oder den Bestellinformationen zu spezifisch bindenden Antikörpern erweitert.

Dieser Basisdatensatz diente für einen strukturellen Ähnlichkeitsvergleich mit mehreren Millionen bestellbarer Substanzen, deren Strukturinformationen in einer hausinternen Datenbank abgelegt sind. Die Suche resultierte in 5000 weiteren Substanzen, die strukturell sehr ähnlich zu den geprüften Haptenen sind (Tanimoto-Koeffizient  $> 0,9$ ) und eventuell eine Immunreaktion auslösen können, da sie eine Kreuzreaktivität auslösen. Die potentiellen Haptene wurden ebenfalls in die Datenbank aufgenommen und befinden sich dort in einer eigenen abrufbaren Rubrik.

### **2.2.3 Originalartikel: “Superhaptens and their potential role in contact dermatitis.”**

GÜNTHER, STEFAN; WEISER, ARMIN A.; AHMED, JESSICA; PREISSNER, ROBERT

Drug Discovery Today: Therapeutic Strategies 4 (2007), S. 11-17

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1016/j.ddstr.2007.08.003>

#### **2.2.4 Originalartikel: “SuperHapten: a comprehensive database for small immunogenic compounds”**

GÜNTHER, STEFAN; HEMPEL, DOROTHEA; DUNKEL, MATHIAS; ROTHER, KRISTIAN; PREISSNER, ROBERT

Nucleic Acids Res 35 (2007), S. D906-D910

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1093/nar/gk1849>

### 2.2.5 Vorhersage der medizinischen Wirkung sowie der Zielproteine von Kleinmolekülen

Zur Vorhersage, welche Moleküle sich womöglich als Medikamente eignen und mit welchen Zielproteinen sie interagieren, lassen sich ebenfalls Struktur-Wirkungs-Beziehungen verwenden. Die Internetanwendung "SuperPred" [37] erlaubt es vom Anwender definierte Moleküle auf deren mögliche medizinische Wirkung *in silico* zu prüfen. Dazu wird in einem ersten Schritt die definierte Molekülstruktur auf vier Eigenschaften hin geprüft, die bei Medikamenten gehäuft auftreten, die sogenannten "Lipinski's Rule of Five" [12]:

- Nicht mehr als 5 Wasserstoffbrücken-Donatoren
- Nicht mehr als 10 Wasserstoffbrücken-Akzeptoren
- Molekulargewicht kleiner 500g/mol
- Verteilungskoeffizienten ( $\log P$ ) kleiner 5

Wenn mehr als eine dieser Regeln gebrochen wird, ist es unwahrscheinlich, dass die zu prüfende Substanz sich als Medikament eignet.

Im zweiten Schritt wird nach strukturellen Ähnlichkeiten zu bekannten Medikamenten gesucht. Werden solche gefunden (Tanimoto-Koeffizient  $> 0,85$ ), so kann mit hoher Wahrscheinlichkeit die medizinische Indikation und die Zielproteine der Substanz vorhergesagt werden. Der Anwendung zu Grunde liegt ein Datensatz mit den Strukturinformationen der meisten der weltweit eingesetzten Wirkstoffe, beschrieben in der Datenbank "SuperDrug" [38] und die Informationen über die Zielproteine der Medikamente beschrieben in "SuperTarget" [28]. Der Vergleich des vom Anwender definierten Moleküls mit 6500 verfügbaren Wirkstoffen wird mit Hilfe struktureller Fingerabdrücke und der Berechnung des Tanimoto-Koeffizienten durchgeführt. Die Gesamtzeit für eine Vorhersage beträgt durchschnittlich etwa drei Sekunden.



### **2.2.6 Originalartikel: “SuperPred: drug classification and target prediction”**

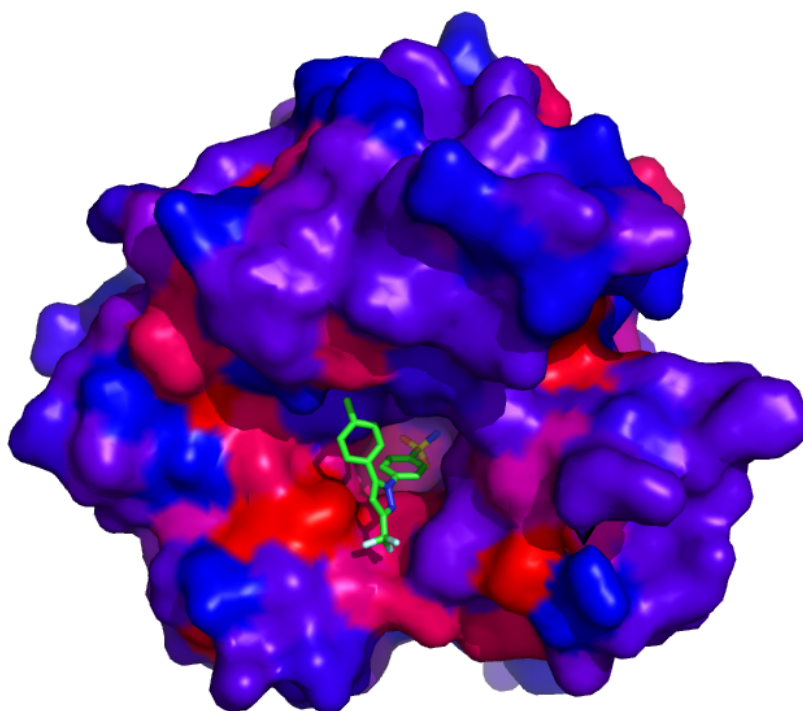
DUNKEL, MATHIAS; GÜNTHER, STEFAN; AHMED, JESSICA ; WITTIG, BURGHARDT; PREISSNER, ROBERT

Nucleic Acids Res 36 (2008), S. W55-W59

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1093/nar/gkn307>

### 2.2.7 3D-Strukturvergleiche von gebundenen Wirkstoffen und generierten Konformeren

Im Gegensatz zu dem strukturellen Fingerabdruck eines Moleküls, wird in einem 3D-Modell die räumliche Verteilung der Atome besonders gewichtet. Der Vergleich zweier Moleküle auf Basis der 3D-Struktur ist dadurch in der Regel wesentlich aufwändiger und langsamer. Die 3D-Struktur bestimmt jedoch, welche Atome des Kleinmoleküls mit dem Zielprotein bei der Bindung in Wechselwirkung treten (vgl. Abb. 2.2). Eine Einbeziehung der 3D-Struktur kann daher helfen, wenn beispielsweise zu einer Bindungstasche eines Zielproteins ein passender Wirkstoff gefunden werden soll. In diesem Fall können die 3D-Eigenschaften der Bindungstasche in ein Modell übertragen werden (Pharmakophor) und zum Durchsuchen von Molekül-Bibliotheken genutzt werden. Geeignete Kandidatenmoleküle stimmen mit ihren 3D-Eigenschaften gut mit dem Pharmakophor überein.



**Abb. 2.2:** Struktur von Celecoxib gebunden an Carboanhydrase (PDB-ID: 1OQ5). Der Wirkstoff wurde als selektiver Inhibitor der Cyclooxygenase-2 bekannt. Später wurde auch eine Inhibierung der Carboanhydrase festgestellt [39]. Kristallstruktur visualisiert mit Py-Mol.

Bei diesem Vorgehen muss jedoch beachtet werden, dass ein Kandidatenmolekül auf Grund seiner drehbaren Bindungen flexibel sein und daher unterschiedliche Konformationen annehmen kann. Wie viele diskrete Konformationen ein Molekül annimmt, ist abhängig von der Anzahl der lokalen Energieminima, die es im gesamten Konformationsraum des Moleküls gibt.

Um die Frage zu klären, wieviele unterschiedliche Konformationsmodelle (Konformere) generiert werden müssen, um die an das Zielprotein gebundene Konformation möglichst gut abzubilden, wurde eine umfassende Untersuchung an co-kristallisierten Wirkstoffen durchgeführt [40]. Die Konformere wurden mit dem Programm “Catalyst” (Firma Accelrys<sup>1</sup>) mit verschiedenen Einstellungen generiert und pro Satz der räumliche Abstand zu den am Zielprotein gebundenen Konformationen berechnet.

Als eine vereinfachende Regel konnte unter anderen gezeigt werden, dass für durchschnittlich große Wirkstoffe (5-6 drehbare Bindungen) etwa 100 generierte Konformere notwendig sind, um die gebundene Konformation mit ausreichender Genauigkeit abbilden zu können. Ebenso konnte gezeigt werden, dass die meisten der Wirkstoffe verschiedene bioaktive Konformationen einnehmen, wenn sie an unterschiedliche Proteine gebunden sind. Bei gleichen oder sehr ähnlichen Zielproteinen sind die Konformationen, die das Molekül in gebundenem Zustand annimmt, strukturell nahezu identisch untereinander. Im gebundenen Zustand ist der Konformationsraum des Wirkstoffs daher stark eingeschränkt.

---

<sup>1</sup>Accelrys, Catalyst - Internetseite: <http://accelrys.com/products/scitegic/component-collections/catalyst.html>

### **2.2.8 Originalartikel: “Representation of target-bound drugs by computed conformers”**

GÜNTHER, STEFAN; SENGER, CHRISTIAN; MICHALSKY, ELKE; GOEDE, ANDREAN; PREISSNER, ROBERT

BMC Bioinformatics 7 (2006), S. 293

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1186/1471-2105-7-293>

## 2.3 Untersuchung der Wechselwirkungen zwischen Makromolekülen

### 2.3.1 Vorhersage von Proteininteraktionen auf Grund der Faltung

Ein Protein kann aus einem oder mehreren Modulen (Domänen) bestehen, welche jeweils eine bestimmte Architektur aufweisen. Unter der Architektur einer Domäne versteht man die Art und Weise, wie sich eine Aminosäurekette zusammenlagert. Die Anzahl der verschiedenen Domänen ist wesentlich kleiner als die Unterschiede zwischen den Proteinsequenzen aller Organismen es vermuten lassen würden. In den letzten 8 Jahren hat die Anzahl neu entdeckter Architekturen pro Jahr immer weiter abgenommen, obwohl die Anzahl der bekannten Proteinstrukturen in der gleichen Zeit stark zugenommen hat<sup>2</sup>. Es lässt sich daher vermuten, dass schon jetzt der größte Teil aller Domärentypen strukturell aufgeklärt wurde.

Manche Protein-Protein-Wechselwirkungen sind häufig mit bestimmten Domänenarchitekturen assoziiert, in diesen Fällen lässt sich mit dem Wissen über die Domäne eines Proteins auch auf mögliche Bindungspartner schließen. Die Domäne eines Proteins lässt sich auch dann bestimmen, wenn nur sehr weit entfernte homologe Proteine strukturell aufgeklärt sind. Vor diesem Hintergrund wurden im Rahmen der vorliegenden Arbeit Kandidatengene bestimmt, die für ein bestimmtes, noch nicht identifiziertes Signalprotein kodieren.

Der bakterielle Krankheitserreger *Legionella pneumophila* kann eine schwere Lungenkrankheit, die sogenannte Legionärskrankheit auslösen. In der Abteilung Infektionsbiologie der Charité Berlin konnte gezeigt werden, dass bei einer Infektion mit dem Bakterium das Protein IPS-1 eine wichtige Rolle bei der Auslösung einer spezifischen Immunabwehr spielt. Es wird angenommen, dass ein Rezeptor existiert, der intrazelluläre Bakterien-DNS erkennt und das Signal direkt oder indirekt an IPS-1 weitergibt. Im Rahmen der vorliegenden Arbeit wurden Kandidatenproteine ermittelt, die durch ihre Faltung als potentielle Interaktionspartner von IPS-1 in Frage kommen.

Das Protein MDA5 bindet virale RNS und interagiert ebenfalls mit IPS-1 [17]. Der Teil, der mit IPS-1 interagiert, weist eine CARD Domäne auf (vgl. Abb. 1.5(a)). Unter der Annahme, dass das gesuchte Signalprotein eine ähnliche Architektur besitzt, wurden im ersten Schritt alle Proteine identifiziert, die höchstwahrscheinlich eine CARD-Domäne enthalten. Die Domäneninfor-

---

<sup>2</sup>PDB - Statistikseite (Juli 2008): <http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=fold-cath>

mationen wurden aus der Datenbank Pfam [41] entnommen, die auf der Basis von Hidden Markov Ketten [42] multiple Sequenzalignments vorberechnet zur Verfügung stellt. Dadurch lassen sich Domänenstrukturen von Proteinen bestimmen, zu denen nur sehr entfernt Verwandte strukturell bekannt sind. Im zweiten Schritt wurden Proteine aus der Auswahl entfernt, die auf Grund der vorhandenen funktionellen Beschreibungen oder auf Grund der extrazellulären Lokalisation nicht für die gesuchte Funktion in Frage kommen. Eine Auswahl von 10 verbliebenden Proteinen, wie z.B. Nod5 und Nod27, wurde am Ende in der oben genannten Arbeitsgruppe auf die hypothetische Signalprotein-Funktion getestet. Obwohl keines der getesteten Proteine die gesuchte Funktion aufwies, wurden die Ergebnisse als Hinweis für nachfolgende Arbeiten veröffentlicht [43]. Warum der gesuchte Rezeptor nicht gefunden wurde, wird in Kapitel 3.3 diskutiert.

### **2.3.2 Originalartikel: “Legionella pneumophila induces IFNbeta in lung epithelial cells via IPS-1 and IRF3, which also control bacterial replication”**

OPITZ, BASTIAN; VINZING, MAYA; LAAK, VINCENT VAN; SCHMECK, BERND; HEINE, GUIDO; GÜNTHER, STEFAN; PREISSNER, ROBERT; SLEVOGT, HORTENSE; N'GUESSAN, PHILIPPE D.; EITEL, JULIA; GOLDMANN, TORSTEN; FLIEGER, ANTJE; SUTTORP, NORBERT; HIPPENSTIEL, STEFAN

J Biol Chem 281 (2006), S. 36173-36179

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1074/jbc.M604638200>

### 2.3.3 Analyse der strukturellen Flexibilität von Makromolekülen

Proteine sind nicht starr, sondern flexibel. Insbesondere an der Oberfläche von Proteinen können die flexiblen Seitenketten unterschiedliche Konformationen annehmen. In der Regel werden Molekülkomplexmodelle aus den ungebundenen Strukturen erstellt. In dem resultierenden Modell müssen daher strukturelle Veränderungen, die bei der Bindung auftreten können, berücksichtigt werden.

Bei den Wechselwirkungen eines Proteins mit der DNS, kommt der Bindungsstelle eines Proteins eine besondere Rolle zu, da sie je nach Funktion des Proteins die DNS mehr oder weniger sequenzspezifisch erkennt. Bei den sehr spezifisch bindenden Transkriptionsfaktoren lassen sich zum Teil einzelne Aminosäuren der Bindungsstelle identifizieren, die für die spezifische Erkennung der passenden Basensequenz verantwortlich sind [44].

Eine im Rahmen der Promotion durchgeführte Analyse zeigt die bei der Bindung auftretenden strukturellen Veränderungen von DNS-bindenden Proteinen [45]. Die Konformationsveränderungen der Proteine wurden einerseits im Bereich der Bindungsstelle und andererseits der Gesamtstruktur betrachtet. Dazu wurden mehrfach vorhandene Kristallstrukturen von vier Transkriptionsfaktoren und drei Enzymen in den gebundenen und den ungebundenen Konformationen miteinander verglichen und auf Grund der Art der festgestellten strukturellen Veränderung klassifiziert.

Alle untersuchten Protein-Bindungsstellen lassen sich dem "induced-fit" Modell zuordnen, d.h. bei den Bindungsstellen der Proteine gibt es mehr oder weniger starke strukturelle Veränderungen, die sich durch die Bindung ereignen. Im DNS-freien Zustand ist eine Bindungsstelle sehr flexibel. Wird eine freie Bindungsstelle des gleichen Proteins in unterschiedlichen Kristallstrukturen verglichen, so findet man immer deutliche strukturelle Unterschiede. Die Konformationen der Bindungsstelle im gebundenen Zustand sind hingegen nahezu identisch.

Die globalen Strukturveränderungen, die über das Rückgrat (engl. backbone) des Proteins auftreten können, werden häufig mit anderen Modellen besser beschrieben als dem induced-fit Prinzip. So lässt sich in vielen Fällen die gebundene Konformation auch im DNS-freien Zustand wiederfinden, jedoch als eine unter verschiedenen auftretenden Konformationsvarianten. In der Veröffentlichung werden diejenigen unterschiedlichen Interaktionsmodelle vorgestellt, welche die beobachteten Konformationsveränderungen am besten beschreiben.



### **2.3.4 Originalartikel: “Molecular flexibility in protein-DNA interactions”**

GÜNTHER, STEFAN; ROTHER, KRISTIAN; FRÖMMEL, CORNELIUS

Biosystems 85 (2006), S. 126-136

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1016/j.biosystems.2005.12.007>

### 2.3.5 Vorhersage der Protein-Protein Bindungskonformation

Sind die Strukturen zweier Proteine bekannt, so ist eine strukturbasierte Vorhersage einer gegenseitigen funktionellen Wechselwirkung, derzeitig kaum möglich. Eine Ursache dafür liegt in der Schwierigkeit, die richtige Bindungskonformation bzw. den Bindungsmodus zweier Proteine aus der Struktur zu bestimmen. Der Bindungsmodus bestimmt, auf welche Art und Weise zwei Proteine geometrisch in Wechselwirkung treten.

Zwar gibt es einige Methoden, die Modelle zu den Bindungskonformationen von Proteinen berechnen [22, 23, 24, 25], sie sind aber noch mit hohen Fehlerraten behaftet [46]. Eines der bekanntesten Programme (ZDock, [22]) schafft es bei einem sehr häufig benutzten Testdatensatz [47] in nur 80% der Fälle den richtigen Bindungsmodus als einen unter 100 Vorschlägen zu modellieren (Tab. 2.1).

Diese Größe muss allerdings in Relation zu Millionen von unterschiedlichen Möglichkeiten gesehen werden, wie sich zwei Proteine, die sich an den Oberflächen berühren, im Modell anordnen lassen. ZDock, wie auch die meisten der anderen heute verwendeten Verfahren, beruhen auf einer sogenannten “Fast-Fourier-Transformation”, einer Prozedur, die es erlaubt, die hohe Anzahl von Möglichkeiten auf geometrisch sinnvolle und hinreichend unterschiedliche Modelle zu reduzieren [48]. ZDock verbleiben nach dieser Filterung noch durchschnittlich 54.000 unterschiedliche Modelle, die im folgenden Schritt durch eine Bewertungsfunktion nach verschiedenen Kriterien geordnet werden [49]. Dieses Verfahren ist allerdings auf Proteine beschränkt, deren strukturelle Veränderungen, die durch die Wechselwirkungen entstehen, nur gering sind. Bei sehr flexiblen Proteinen können die richtigen Bindungskonformationen auf diese Art und Weise häufig gar nicht oder nur ungenau bestimmt werden (vgl. Tab. 2.1, “Hohe Flexibilität”).

Die im Rahmen dieser Arbeit entwickelte Methode ISEARCH [50] beruht auf einem wissensbasierten Prinzip. Es basiert auf der Beobachtung, dass die Fläche, die sich bei der Bindung zweier Proteine ausbildet, häufig Ähnlichkeiten zu bereits kristallisierten Bindungsflächen aufweist [26]. Diese Ähnlichkeiten können auch dann bestehen, wenn zwischen den Proteinen kaum oder gar keine sequentielle Verwandtschaft besteht.

Um ein Proteinkomplexmodell zu generieren, werden im ersten Schritt die Oberflächen der ungebundenen Proteine auf Gemeinsamkeiten zu bereits bekannten Bindungsflächen hin untersucht. Hierfür eignet sich ein Algorithmus, der 2003 an der Charité entwickelt wurde und strukturelle Gemeinsamkeiten zwischen unterschiedlich großen Molekülen detektieren kann [51]. Werden solche Ähnlichkeiten gefunden, kann im nächsten Schritt das Modell

**Tab. 2.1:** Trefferrate des Programms “ZDock” bei einem Testset mit 59 Einzelfällen [47]. Insgesamt werden bei ZDock durchschnittlich etwa 54.000 Modelle pro Testfall erstellt. Danach werden die Modelle nach verschiedenen Kriterien bewertet und sortiert. Die Zahlen geben an in wievielen der 59 Fälle unter den besten 100 bzw. 1,000 und 3,600 Kandidaten mindestens ein akzeptables Modell generiert wurde. Die Werte in Klammern geben die Anzahl der guten Modelle an. Ein akzeptables Modell besitzt einen Interface-RMSD (iRMSD) kleiner 5 Å zwischen dem Modell und dem kristallisiertem Proteinkomplex. Ein gutes Modell besitzt einen iRMSD kleiner 2.5 Å.

| Interaktionstyp    | Anzahl Testfälle | Trefferzahl |         |         |
|--------------------|------------------|-------------|---------|---------|
|                    |                  | 100         | 1.000   | 3.600   |
| Enzym-Inhibitor    | 22               | 22 (20)     | 22 (21) | 22 (22) |
| Antikörper-Antigen | 19               | 17 (11)     | 19 (17) | 19 (17) |
| Hohe Flexibilität  | 7                | 2 (0)       | 3 (0)   | 6 (0)   |
| Andere             | 11               | 7 (4)       | 10 (6)  | 10 (8)  |
| Gesamt             | 59               | 48 (35)     | 54 (44) | 57 (47) |

**Tab. 2.2:** Trefferrate des Programms “ISEARCH” bei dem Benchmark 1.0 [47]. Insgesamt wurden durch ISEARCH pro Testfall 4.900 bis 11.000 Komplexmodelle erstellt. Diese wurden nicht gefiltert oder sortiert. Die Zahlen geben an, in wievielen der 59 Fälle mindestens ein akzeptables bzw. gutes Modell generiert wurde. Akzeptables Modell: iRMSD < 5.0, gutes Modell: iRMSD < 2.5.

| Interaktionstyp    | Anzahl Testfälle | Anzahl der Modelle<br>4900 - 11.200 |
|--------------------|------------------|-------------------------------------|
| Enzym-Inhibitor    | 22               | 22 (14)                             |
| Antikörper-Antigen | 19               | 18 (2)                              |
| Hohe Flexibilität  | 7                | 6 (1)                               |
| Andere             | 11               | 10 (5)                              |
| Gesamt             | 59               | 56 (32)                             |

des Proteinkomplexes analog zu den detektierten ähnlichen Bindungsflächen modelliert werden.

Für die Generierung der Modelle wurden etwa 8.500 nicht redundante Bindungsflächen zwischen interagierenden Domänen in der PDB identifiziert und in eine Datenbank überführt. Diese dienen der Suche nach Ähnlichkeiten auf den Oberflächen der ungebundenen Einzelproteine. Je nach Einstellung des Überlagerungsalgorithmus können auch sehr schwache Ähnlichkeiten zwischen verschiedenen Bindungsflächen detektiert werden. In diesem Falle erhöht sich allerdings die Anzahl der Modelle. Im Extremfall könnte jede der wechselseitigen Bindungsflächen aus der Datenbank zwei unterschiedliche Protein-Protein Komplexmodelle erzeugen. Bei 8.500 wechselseitigen Bindungsflächen wäre somit die Erstellung von 17.000 verschiedenen Modellen möglich. Tabelle 2.2 zeigt einen vorläufigen Durchlauf von ISEARCH, in dem etwa 8.000 Modelle pro Testfall aus dem Benchmark 1.0 erstellt wurden. Verschiedene Punkte, welche die Trefferrate von ISEARCH weiter erhöhen könnten, werden in Kapitel 3.4 diskutiert.

### **2.3.6 Originalartikel: “ISEARCH-prediction of interactions using known interfaces”**

GÜNTHER, STEFAN; MAY, PATRICK; HOPPE, ANDREAS ; FRÖMMEL, CORNELIUS; PREISSNER, ROBERT

Proteins 69 (2007), S. 839-844

Der Originalartikel ist unter folgender Adresse im Internet abrufbar:  
<http://dx.doi.org/10.1002/prot.21746>

# Kapitel 3

## Diskussion

### 3.1 Datenintegration - eine Methode zur Bestimmung von Molekülwechselwirkungen

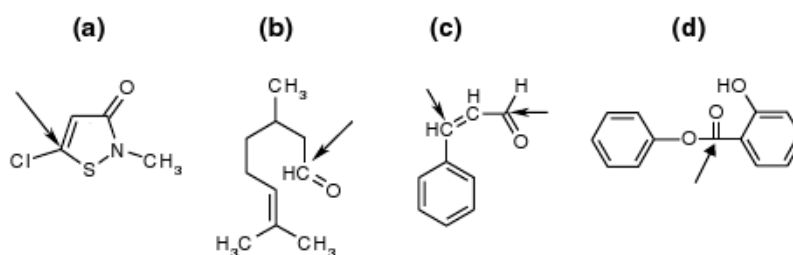
Moderne Labormethoden, wie z.B. das Testen der biologischen Aktivität von Kleinmolekülen im Hochdurchsatzverfahren, produzieren eine große Menge von Daten, die bei der Identifizierung von Wechselwirkungen zwischen Biomolekülen genutzt werden können. Für die Auswertung dieser Datenmengen existieren heute schon viele etablierte Methoden der Bioinformatik. Defizite gibt es allerdings noch bei der Zusammenführung der verschiedenen Daten, der Datenintegration. Für die Identifizierung von spezifischen Wechselwirkungen zwischen Wirkstoffen und Proteinen, ist es notwendig alle relevanten Informationen zu den in Frage kommenden Molekülen miteinander zu verbinden, um eine globale Sicht auf den Wirkmechanismus zu ermöglichen [52]. Dazu zählen auch phänotypische Informationen, wie beobachtete Nebenwirkungen von Medikamenten. Sie lassen sich nutzen, um auf zusätzliche Proteine zu schließen, die möglicherweise mit dem Wirkstoff in Wechselwirkung treten [53]. Eine Integration verschiedener Datenquellen wird jedoch häufig dadurch erschwert, dass sie in verschiedenen Formaten vorliegen, wie beispielsweise in Textform - verteilt auf Millionen unterschiedlicher Artikel.

Die im Ergebnisteil vorgestellte Webapplikation SuperTarget (Kapitel 2.1) ist ein Beispiel für die Integration verschiedener Datenquellen, durch deren Hilfe sich die Wirkmechanismen von Medikamenten analysieren lassen. Hierzu wurden in der Literatur beschriebene Wechselwirkungen zwischen Wirkstoffen und ihren Zielproteinen in eine Datenbank überführt und mit zahlreichen weiteren Informationen, wie zum Beispiel den involvierten

Stoffwechselwegen oder den Sequenzen der Zielproteine verknüpft. Es gibt jedoch eine ganze Reihe von weiteren Daten, die bisher noch nicht einbezogen wurden, aber dennoch medizinisch relevant sind. Das können beispielsweise bekannte und häufig auftretende Mutationen in den P450-Cytochromen sein. Diese Enzyme bauen Wirkstoffe im Körper ab und eine veränderte Funktion ist häufig die Ursache für Medikamentenunverträglichkeiten [54]. Außerdem könnte die Identifikation neuer Zielproteine für Wirkstoffe erleichtert werden. In SuperTarget sind alle bekannten Proteine des Menschen aufgenommen, unabhängig davon, ob sie mit Wirkstoffen interagieren oder nicht. In dieser Menge sind viele potentielle Zielproteine enthalten, die bisher als solche noch nicht identifiziert wurden. Als „druggability“ bezeichnet man, wie geeignet ein Protein als potentielles Zielprotein für Wirkstoffe ist. Kriterien, wie eine erhöhte Expression eines Proteins bei einem Krankheitsverlauf oder eine organspezifische Expression, erhöhen die druggability eines Proteins [55]. Proteinstrukturinformationen lassen sich ebenfalls zu diesem Zweck nutzen. So lässt sich beispielsweise abschätzen, ob Bindungsstellen am Protein vorhanden sind, welche für eine spezifische Wechselwirkung mit wirkstoffähnlichen Substanzen in Frage kommen [56]. Eine Einbeziehung solcher Informationen in SuperTarget würde die Identifizierung neuer Zielproteine weiter unterstützen.

## 3.2 Strukturbasierte Ableitung der Bioaktivität von Kleinmolekülen

In Kapitel 2.2 wird gezeigt, wie sich mit Hilfe der Struktur von Kleinmolekülen auf deren bioaktive Aktivität schließen lässt. Im ersten Beispiel wird die immunogene Wirkung einer Vielzahl bestellbarer Moleküle vorhergesagt, die in vielen Bereichen der Umwelt eingesetzt werden. Im zweiten Fall wird vorhergesagt, welche Substanzen sich als Medikamente eignen könnten und welche deren mögliche Zielstrukturen sind. Das Grundverfahren ist in beiden Methoden ähnlich: es wird mit Hilfe eines Strukturvergleiches auf vorhandenes Wissen zu ähnlichen, besser untersuchten, Substanzen zurückgegriffen. Sind solche ähnlichen Substanzen bekannt, so liefert das Verfahren gute Ergebnisse. Weitere Verbesserungen lassen sich dadurch erzielen, dass der zum Strukturvergleich verwendete Fingerabdruck auf diejenigen chemischen Eigenschaften optimiert wird, welche für die zu prüfende biologische Wirkung besonders relevant sind. Haptene, die Immunreaktionen auslösen können, haben beispielsweise oft bestimmte chemische Eigenschaften, wie stark polarierte Bindungen (vgl. Abb. 3.1). Solche Eigenschaften ließen sich in einem



**Abb. 3.1:** Beispiele für die Elektrophilie von Haptenen. Elektrophile Zentren sind mit Pfeilen angegeben. (a) 5-chloro-2-methyl-4-isothiazolin-3-one (MCI), (b) Citronellal, (c) Cinnamaldehyd und (d) Phenyl-Salicylat. Abbildung aus [57].

strukturellen Fingerabdruck besonders gewichten, dadurch könnten potentielle Haptene in einer großen Masse von Kandidatenmolekülen eindeutiger identifiziert werden.

Stehen besser untersuchte ähnliche Substanzen nicht zur Verfügung, kann die Bioaktivität von Molekülen durch andere theoretische Verfahren abgeschätzt werden. Im Falle der Vorhersage einer medizinischen Wirkung von Molekülen können empirische Größen wie z.B. die Anzahl der Wasserstoffdonatoren bei typischen Wirkstoffen erste wichtige Hinweise liefern [12]. Eine weitere Option ist die Einbeziehung der Struktur vorgegebener Zielproteine. Dieses Verfahren wird oft gewählt, wenn Substanzen identifiziert werden sollen, die als potentielle Inhibitoren oder Aktivatoren für ein gewünschtes Zielprotein in Frage kommen. Ausgehend von einer Bindungstasche, welche in ein Pharmakophormodell transformiert wurde, lassen sich Kandidatenmoleküle im Modell testen, die als mögliche Bindungspartner in Frage kommen. Bei einem solchen Vorgehen werden in der Regel Konformationen der Kandidatenstrukturen generiert, um die dreidimensionale Struktur des Moleküls abbilden zu können.

Im letzten Teil von Kapitel 2.2 wurde analysiert auf welche Art und Weise Konformere generiert werden müssen, um eine möglichst gute Abbildung der am Protein gebundenen Konformation zu erhalten. Es kann gezeigt werden, dass die bioaktive Konformation eines Moleküls direkt von der Zielstruktur abhängig ist und sich diesem mehr oder weniger flexibel anpasst. Werden Tausende von Moleküle in einer Bibliothek abgelegt, die nach dreidimensionalen Eigenschaften durchsucht werden sollen, müssen daher viele in Frage kommende 3D-Konformere der Moleküle erzeugt werden. In der Arbeit konnte auch gezeigt werden, dass bei einer durchschnittlich großen Anzahl drehbarer Bindungen, die in Kleinmolekülen auftritt, bis zu 100 verschiedene Rotamere ausreichen, um eine strukturell sehr ähnliche Abbildung zur bioaktiven Kon-



formation zu erhalten ( $\text{RMSD} < 1 \text{ \AA}$ ). Eine Anzahl, die von heutigen Rechnern schnell genug bearbeitet werden kann, um Pharmakophorsuchen für Hunderttausende von Rotameren zu ermöglichen. Firmen wie z.B. Accelrys bieten Programme zur Wirkstoffidentifizierung an, die auf einem solchen Verfahren beruhen [58]. Datenbanken unterschiedlicher Substanzklassen, die in der Arbeitsgruppe “Strukturelle Bioinformatik” der Charité entwickelt wurden, stellen die vorberechneten Konformationen der Biomoleküle frei zum Herunterladen zur Verfügung [38, 59].

Es gibt bereits viele Beispiele, die zeigen, wie durch die Identifizierung von Bindungspartnern durch Pharmakophore aufwändige Laboruntersuchungen vermieden werden konnten [60, 61, 62]. Allerdings erfordert diese Art der Identifizierung von Bindungspartnern die gelöste Kristallstruktur des Zielproteins. Zahlreiche wichtige potentielle Zielproteine für Wirkstoffe, wie z.B. ein großer Anteil der Transmembranproteine, sind heute in atomarer Auflösung noch nicht bekannt, deren strukturelle Aufklärung ist aber Gegenstand aktueller Forschung. Methoden zur Proteinkristallisation [63] sowie der Proteinstrukturvorhersage [64] werden laufend verbessert. Die Einbeziehung dieser zusätzlichen Strukturinformationen bei der Wirkstoffentwicklung, wird daher von zunehmender Bedeutung sein [65].

### 3.3 Strukturbasierte Vorhersage von Protein-Protein-Interaktionen

Die Vorhersage von Proteininteraktionen mit zufriedenstellender Vorhersagegüte funktioniert dann sehr gut, wenn eine Interaktion verwandter Proteine bereits beschrieben wurde [66]. Allerdings muss dabei beachtet werden, dass die gemessenen Wechselwirkungen, auf denen die Vorhersage basiert, nicht zwangsläufig den *in vivo* Bedingungen entsprechen müssen. Es ist beispielsweise bekannt, dass Labormessungen, die auf dem Hefe-Zwei-Hybrid-System beruhen, mit sehr hohen Fehlerraten behaftet sind und daher mit weiteren experimentellen Befunden überprüft werden sollten [67].

Wichtige Hinweise auf Bindungspartner können auch weit entfernte homologe Proteine liefern, da sie in der Regel die gleiche Architektur aufweisen und diese häufig eine wichtige Voraussetzung für die gegenseitige Erkennung der Proteine ist. Das gilt allerdings nicht für alle Architekturen in gleichem Maße. So besitzen etwa 10% aller Enzyme eine  $\alpha/\beta$ -Tonnen-Domäne und es wird vermutet, dass alle diese verschiedenen Proteine auf einen gemeinsamen Vorläufer zurückzuführen sind [68]. Im Laufe der Evolution hat eine divergente Entwicklung dieser Proteine stattgefunden, die dazu geführt hat, dass sie sich

funktionell unabhängig voneinander entwickelt haben und unterschiedlichste Bindungspartner besitzen. Andere Architekturen geben deutliche Hinweise auf mögliche Bindungspartner, so ist z.B. eine bZIP-Domäne typischerweise mit der DNS-Wechselwirkung und der Funktion eines Transkriptionsfaktors verbunden.

Kapitel 2.3.1 schildert ein Beispiel für die Suche eines Bindungspartners von IPS-1, welcher das Signal des Vorhandenseins zellfremder DNS weitergibt. IPS-1 interagiert mit einem bekannten Protein (MDA5), welches eine CARD Domäne besitzt. Es wurde daher ein Signalprotein gesucht, welches eine CARD Domäne aufweist und entweder direkt an die DNS bindet oder über ein Adapterprotein dieses Signal entgegennimmt.

Das gesuchte Protein konnte im Rahmen des Projektes nicht gefunden werden, dennoch lieferte die Überprüfung der ermittelten Kandidatenproteine wertvolle Hinweise für weitere Untersuchungen [43]. Für die missglückte Identifizierung gibt es unterschiedliche Erklärungen. So kann es beispielsweise sein, dass das gesuchte Signalprotein nicht wie angenommen über eine CARD Domäne mit IPS-1 interagiert. Durch die bekannte Wechselwirkung zwischen IPS-1 und MDA5 [17] sind Wechselwirkungen zwischen IPS-1 und weiteren Domänenarchitekturen natürlich nicht ausgeschlossen. Eine andere Erklärung ist, dass das gesuchte Protein trotz seiner CARD-Domäne nicht in der ersten Kandidatenliste enthalten war. Über Sequenzalignments lassen sich nur Proteine mit einer CARD-Domäne identifizieren, zu denen es strukturell bekannte homologe Proteine gibt. Ist keine solche Verwandtschaft vorhanden, kann die Architektur eines Proteins zur Zeit nur sehr ungenau vorhergesagt werden. Es ist daher davon auszugehen, dass noch viele weitere Proteine mit CARD-Domäne existieren, die aber noch nicht identifiziert werden konnten. Darunter könnte sich auch das gesuchte Signalprotein befinden. Darüberhinaus wurden nicht alle Proteine mit CARD-Domäne getestet, sondern nur die vielversprechendsten Kandidaten. Diese Auswahl basierte auf verschiedenen Faktoren, wie dem Vorkommen des Proteins im Zytosol oder einer ungeklärten Funktion. Die Pfam Datenbank gibt insgesamt 65 unterschiedliche Kandidatensequenzen an, die im Menschen vorkommen und eine CARD-Domäne aufweisen<sup>1</sup>. Einige von ihnen werden zur Zeit noch in der Abteilung Infektionsbiologie der Charité Berlin untersucht.

---

<sup>1</sup>Pfam, CARD-Domäne: <http://pfam.janelia.org/family?acc=PF00619>

### 3.4 Vorhersage der Protein-Bindungskonformation

Durch eine Methode, die eine Wechselwirkung aufgrund der Atommodelle zweier Proteine richtig vorhersagt, ließen sich viele neue Protein-Protein-Interaktionen ableiten. Vorerst bleibt es jedoch ein Ziel der strukturellen Bioinformatik die Bindungskonformation zweier Proteine hinreichend genau vorherzusagen, wenn eine Wechselwirkung dieser Proteine bereits beschrieben wurde.

Ein großes Problem bei der Modellierung von Wechselwirkungen ist die strukturelle Flexibilität der Proteine. Bei der gegenseitigen Wechselwirkung der Proteine treten strukturelle Veränderungen sowohl in der globalen Gesamtstruktur des Proteinrückgrats auf, als auch lokal bei den interagierenden Seitenketten. Kapitel 2.3.3 zeigt deutlich die verschiedenen Klassen der auftretenden strukturellen Veränderungen am Beispiel der Interaktion zwischen Proteinen und der DNS.

Bei der Modellierung eines Protein-Protein-Komplexes basierend auf den beiden Einzelproteinen, wird häufig die Flexibilität der Makromoleküle vernachlässigt, um die Komplexität des Modells zu reduzieren. Solche sogenannten starren Anlagerungsmethoden (engl. “rigid-docking-methods”) scheitern häufig bei der Vorhersage, wenn die strukturellen Veränderungen, die bei der Bindung auftreten, sehr groß sind (vgl. Tabelle 2.1). Erste Modellierungsmethoden, die eine Flexibilität der Proteine einkalkulieren, basieren auf der Darstellung der Einzelstrukturen durch ein Ensemble von starren Konformationen, die entweder mit verschiedenen Methoden modelliert [69, 70, 71] oder mit Hilfe der NMR-Spektroskopie bestimmt wurden [72].

Die in Kapitel 2.3.5 vorgestellte Methode beruht darauf, dass potentielle Bindungsstellen auf der Oberfläche der Einzelproteine detektiert werden können, indem sie mit bereits bekannten Bindungsstellen aus anderen Proteinen verglichen werden. Durch diese neue Herangehensweise scheint die Methode auch Stellen auf der Oberfläche identifizieren zu können, die bis zu einem gewissen Grad strukturell verändert sind. Sechs von sieben Komplexen, die sich durch große strukturelle Veränderungen bei der Anlagerung der Einzelproteine auszeichnen, konnten im Modell mindestens akzeptabel nachgebildet werden (vgl. Tab. 2.2).

Die vorgestellte Methode beschreibt eine vielversprechende Herangehensweise, die allerdings noch weiterentwickelt werden muss. Das kann auf verschiedenen Ebenen geschehen. Der verwendete Algorithmus NeedleHaystack [51] wurde ursprünglich entwickelt um Kleinstrukturbindungsstellen zu identifizieren. Durch eine Anpassung an Proteinbindungsstellen könnte der Algo-

rithmus optimiert und dadurch vorhandene Ähnlichkeiten besser identifiziert werden.

Die akzeptablen oder guten Modelle befinden sich unter einer großen Menge von generierten falschen Komplexmodellen. Es muss daher untersucht werden, welche Merkmale für eine Ähnlichkeit zwischen den Bindungsstellen ausschlaggebend sind, um falsche von richtigen Modellen besser unterscheiden zu können.

Bisher wurde keinerlei qualitative Sortierung der generierten Modellkomplexe implementiert. Eine Methode die guten Modelle in einer großen Menge von Kandidaten zu identifizieren, besteht in der Abschätzung der Qualität der modellierten Proteinbindung. Diese ließe sich anhand verschiedener Eigenschaften abschätzen, wie z.B. die Ladung der interagierenden Aminosäuren, die Art und Anzahl der im Modell auftretenden direkten Atomkontakte, die evolutionäre Konservierung der interagierenden Aminosäuren oder der Abstand der Schwerpunkte der Einzelproteine im Modell des Komplexes. Viele unterschiedliche Parameter sind bekannt, die mit der Qualität eines generierten Protein-Protein-Komplex Modells korrelieren [73]. In wieweit es mit Hilfe solcher Beschreibungen gelingt, das richtige Modell aus der Menge von Kandidatenkomplexen zu identifizieren, muss erst noch gezeigt werden, die neue Modellierungsmethode hat jedoch schon jetzt internationales Interesse geweckt [74].

### 3.5 Schlussfolgerungen

Zahlreiche Informationsquellen und Labordaten stehen heute zur Verfügung mit Hilfe derer sich fundierte Prognosen über neue Wechselwirkungen zwischen Biomolekülen aufstellen lassen. Methoden der Informatik helfen diese Daten zu integrieren und übergreifend auszuwerten. Auf diese Weise lassen sich Informationen über Wechselwirkungen von gut untersuchten Molekülen auf ähnliche andere übertragen. In der vorliegenden Arbeit wurde dieses Vorgehen gewählt, um Haptene zu identifizieren sowie die Wechselwirkungen zwischen Wirkstoffen und medizinisch relevanten Proteinen zu bestimmen.

Eine weitere Methode ist die Überprüfung von Wechselwirkungen zwischen Kleinmolekülen und Makromolekülen mit Hilfe eines 3D-Modells. Da mehr oder weniger stark ausgeprägte strukturelle Veränderungen bei der Wechselwirkung von Biomolekülen auftreten, ist es wichtig die Konformationsänderungen im Modell zu berücksichtigen. Im Falle von Kleinstrukturen gelingt dieses, durch eine begrenzte Anzahl von strukturell unterschiedlichen vorberechneten Konformeren.

Für die strukturbasierte Vorhersage spezifischer Protein-Protein-Wechsel-

wirkungen kann die Domänenstruktur der Proteine wichtige Hinweise liefern. Über Sequenzalignments kann die Architektur des Proteins auch dann bestimmt werden, wenn nur homologe Proteine in atomarer Auflösung bekannt sind. Durch den Vergleich mit bekannten Domäneninteraktionen, lassen sich neue Wechselwirkungen ableiten.

Eine Vorhersage der spezifischen Wechselwirkung zwischen zwei Proteinen, welche allein auf den atomaren Strukturmodellen beruht, bleibt ein Ziel der strukturellen Bioinformatik. Ein Teilproblem ist die Vorhersage der richtigen Bindungskonformation. Unterschiedliche Protein-Protein Wechselwirkungen können Ähnlichkeiten im direkten lokalen Bereich der Interaktionsflächen aufweisen. Im Rahmen der vorliegenden Arbeit konnte eine erste Anwendung zur Vorhersage der Protein-Bindungskonformation entwickelt werden, welche auf diesem Prinzip beruht.

# Literaturverzeichnis

- [1] WATSON, H. C.: Stereochemistry of the protein myoglobin. In: *Prog Stereochem* 4 (1969), S. 299—303
- [2] LEE, R. C. ; FEINBAUM, R. L. ; AMBROS, V.: The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. In: *Cell* 75 (1993), Dec, Nr. 5, S. 843–854
- [3] STEFANI, Giovanni ; SLACK, Frank J.: Small non-coding RNAs in animal development. In: *Nat Rev Mol Cell Biol* 9 (2008), Mar, Nr. 3, S. 219–230
- [4] BENTWICH, Isaac: Identifying human microRNAs. In: *Curr Top Microbiol Immunol* 320 (2008), S. 257–269
- [5] SAUER, Uwe ; HEINEMANN, Matthias ; ZAMBONI, Nicola: Genetics. Getting closer to the whole picture. In: *Science* 316 (2007), Apr, Nr. 5824, S. 550–551
- [6] SAUER, Uwe: Metabolic networks in motion: <sup>13</sup>C-based flux analysis. In: *Mol Syst Biol* 2 (2006), S. 62
- [7] AMARAL, Luis A N.: A truer measure of our ignorance. In: *Proc Natl Acad Sci U S A* 105 (2008), May, Nr. 19, S. 6795–6796
- [8] STUMPF, Michael P H. ; THORNE, Thomas ; SILVA, Eric de ; STEWART, Ronald ; AN, Hyeong J. ; LAPPE, Michael ; WIUF, Carsten: Estimating the size of the human interactome. In: *Proc Natl Acad Sci U S A* 105 (2008), May, Nr. 19, S. 6959–6964
- [9] STRAND, Vibeke: Are COX-2 inhibitors preferable to non-selective non-steroidal anti-inflammatory drugs in patients with risk of cardiovascular events taking low-dose aspirin? In: *Lancet* 370 (2007), Dec, Nr. 9605, 2138–2151. [http://dx.doi.org/10.1016/S0140-6736\(07\)61909-6](http://dx.doi.org/10.1016/S0140-6736(07)61909-6). – DOI 10.1016/S0140-6736(07)61909-6

- [10] JU, Cynthia ; POHL, Lance R.: Tolerogenic role of Kupffer cells in immune-mediated adverse drug reactions. In: *Toxicology* 209 (2005), Apr, Nr. 2, S. 109–112
- [11] SINGH, K. V. ; KAUR, Jasdeep ; VARSHNEY, Grish C. ; RAJE, Manoj ; SURI, C. R.: Synthesis and characterization of hapten-protein conjugates for antibody production against small molecules. In: *Bioconjug Chem* 15 (2004), Nr. 1, S. 168–173
- [12] LIPINSKI, C. A. ; LOMBARDO, F. ; DOMINY, B. W. ; FEENEY, P. J.: Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. In: *Adv Drug Deliv Rev* 46 (2001), Mar, Nr. 1-3, S. 3–26
- [13] KENDREW, J. C.: Structure and function in myoglobin and other proteins. In: *Fed Proc* 18 (1959), Jul, Nr. 2, Part 1, S. 740–751
- [14] BERMAN, H. M. ; WESTBROOK, J. ; FENG, Z. ; GILLILAND, G. ; BHAT, T. N. ; WEISSIG, H. ; SHINDYALOV, I. N. ; BOURNE, P. E.: The Protein Data Bank. In: *Nucleic Acids Res* 28 (2000), Jan, Nr. 1, S. 235–242
- [15] SCHUELER-FURMAN, Ora ; WANG, Chu ; BRADLEY, Phil ; MISURA, Kira ; BAKER, David: Progress in modeling of protein structures and interactions. In: *Science* 310 (2005), Oct, Nr. 5748, 638–642. <http://dx.doi.org/10.1126/science.1112160>. – DOI 10.1126/science.1112160
- [16] HERMANN, Johannes C. ; MARTI-ARBONA, Ricardo ; FEDOROV, Alexander A. ; FEDOROV, Elena ; ALMO, Steven C. ; SHOICHET, Brian K. ; RAUSHEL, Frank M.: Structure-based activity prediction for an enzyme of unknown function. In: *Nature* 448 (2007), Aug, Nr. 7155, S. 775–779
- [17] KAWAI, Taro ; TAKAHASHI, Ken ; SATO, Shintaro ; COBAN, Cevayir ; KUMAR, Himanshu ; KATO, Hiroki ; ISHII, Ken J. ; TAKEUCHI, Osamu ; AKIRA, Shizuo: IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. In: *Nat Immunol* 6 (2005), Oct, Nr. 10, S. 981–988
- [18] FISCHER, Emil: Einfluss der configuration auf die Wirkung der Enzyme. In: *Chem Ber* 27 (1894), S. 2985–2993
- [19] KOSHLAND, D. E.: Application of a Theory of Enzyme Specificity to Protein Synthesis. In: *Proc Natl Acad Sci U S A* 44 (1958), Feb, Nr. 2, S. 98–104

- [20] MONOD, J. ; WYMAN, J. ; CHANGEUX, J. P.: On the nature of allosteric transitions: a plausible model. In: *J Mol Biol* 12 (1965), May, S. 88–118
- [21] GAO, Ying ; DOUGUET, Dominique ; TOVCHIGRECHKO, Andrey ; VAKSER, Ilya A.: DOCKGROUND system of databases for protein recognition studies: unbound structures for docking. In: *Proteins* 69 (2007), Dec, Nr. 4, S. 845–851
- [22] CHEN, Rong ; LI, Li ; WENG, Zhiping: ZDOCK: an initial-stage protein-docking algorithm. In: *Proteins* 52 (2003), Jul, Nr. 1, S. 80–87
- [23] DOMINGUEZ, Cyril ; BOELEN, Rolf ; BONVIN, Alexandre M J J.: HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. In: *J Am Chem Soc* 125 (2003), Feb, Nr. 7, S. 1731–1737
- [24] TOVCHIGRECHKO, Andrey ; VAKSER, Ilya A.: GRAMM-X public web server for protein-protein docking. In: *Nucleic Acids Res* 34 (2006), Jul, Nr. Web Server issue, S. W310–W314
- [25] COMEAU, Stephen R. ; GATCHELL, David W. ; VAJDA, Sandor ; CAMACHO, Carlos J.: ClusPro: an automated docking and discrimination method for the prediction of protein complexes. In: *Bioinformatics* 20 (2004), Jan, Nr. 1, S. 45–50
- [26] SHULMAN-PELEG, Alexandra ; SHATSKY, Maxim ; NUSSINOV, Ruth ; WOLFSON, Haim J.: Spatial chemical conservation of hot spot interactions in protein-protein complexes. In: *BMC Biol* 5 (2007), S. 43
- [27] IMMING, Peter ; SINNING, Christian ; MEYER, Achim: Drugs, their targets and the nature and number of drug targets. In: *Nat Rev Drug Discov* 5 (2006), Oct, Nr. 10, S. 821–834
- [28] GÜNTHER, Stefan ; KUHN, Michael ; DUNKEL, Mathias ; CAMPILLOS, Monica ; SENGER, Christian ; PETSALAKI, Evangelia ; AHMED, Jessica ; URDIALES, Eduardo G. ; GEWIESS, Andreas ; JENSEN, Lars J. ; SCHNEIDER, Reinhard ; SKOBLO, Roman ; RUSSELL, Robert B. ; BOURNE, Philip E. ; BORK, Peer ; PREISSNER, Robert: SuperTarget and Matador: resources for exploring drug-target relationships. In: *Nucleic Acids Res* 36 (2008), S. D919–D922
- [29] COMMITTEE, W. H. O. E.: The selection and use of essential medicines. In: *World Health Organ Tech Rep Ser* (2007), Nr. 946, S. 1–162, back cover



- [30] COMMITTEE, Joint F. (Hrsg.): *British National Formulary (BNF)*. Bd. ISBN 0-85369-584-9. 47. London: British Medical Association and the Royal Pharmaceutical Society of Great Britain., 2004
- [31] STEINBECK, Christoph ; HOPPE, Christian ; KUHN, Stefan ; FLORIS, Matteo ; GUHA, Rajarshi ; WILLIGHAGEN, Egon L.: Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. In: *Curr Pharm Des* 12 (2006), Nr. 17, S. 2111–2120
- [32] WHITTLE, Martin ; GILLET, Valerie J. ; WILLETT, Peter ; ALEX, Alexander ; LOESEL, Jens: Enhancing the effectiveness of virtual screening by fusing nearest neighbor lists: a comparison of similarity coefficients. In: *J Chem Inf Comput Sci* 44 (2004), Nr. 5, S. 1840–1848
- [33] DELANEY, J. S.: Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. In: *Mol Divers* 1 (1996), Aug, Nr. 4, S. 217–222
- [34] MARTIN, Yvonne C. ; KOFRON, James L. ; TRAPHAGEN, Linda M.: Do structurally similar molecules have similar biological activity? In: *J Med Chem* 45 (2002), Sep, Nr. 19, S. 4350–4358
- [35] GÜNTHER, Stefan ; WEISER, Armin A. ; AHMED, Jessica ; PREISSNER, Robert: Superhaptens and their potential role in contact dermatitis. In: *Drug Discovery Today: Therapeutic Strategies* 4 (2007), S. 11–17
- [36] GÜNTHER, Stefan ; HEMPEL, Dorothea ; DUNKEL, Mathias ; ROTHER, Kristian ; PREISSNER, Robert: SuperHapten: a comprehensive database for small immunogenic compounds. In: *Nucleic Acids Res* 35 (2007), S. D906–D910
- [37] DUNKEL, Mathias ; GÜNTHER, Stefan ; AHMED, Jessica ; WITTIG, Burghardt ; PREISSNER, Robert: SuperPred: drug classification and target prediction. In: *Nucleic Acids Res* 36 (2008), S. W55–W59
- [38] GOEDE, Andrean ; DUNKEL, Mathias ; MESTER, Nina ; FROMMEL, Cornelius ; PREISSNER, Robert: SuperDrug: a conformational drug database. In: *Bioinformatics* 21 (2005), May, Nr. 9, S. 1751–1753
- [39] DOGNÉ, Jean-Michel ; THIRY, Anne ; PRATICO, Domenico ; MASEREEL, Bernard ; SUPURAN, Claudiu T.: Dual carbonic anhydrase–cyclooxygenase-2 inhibitors. In: *Curr Top Med Chem* 7 (2007), Nr. 9, S. 885–891

- [40] GÜNTHER, Stefan ; SENGER, Christian ; MICHALSKY, Elke ; GOEDE, Andrean ; PREISSNER, Robert: Representation of target-bound drugs by computed conformers: implications for conformational libraries. In: *BMC Bioinformatics* 7 (2006), S. 293
- [41] SONNHAMMER, E. L. ; EDDY, S. R. ; DURBIN, R.: Pfam: a comprehensive database of protein domain families based on seed alignments. In: *Proteins* 28 (1997), Jul, Nr. 3, S. 405–420
- [42] KROGH, A. ; BROWN, M. ; MIAN, I. S. ; SJÖLANDER, K. ; HAUSSLER, D.: Hidden Markov models in computational biology. Applications to protein modeling. In: *J Mol Biol* 235 (1994), Feb, Nr. 5, S. 1501–1531
- [43] OPITZ, Bastian ; VINZING, Maya ; LAAK, Vincent van ; SCHMECK, Bernd ; HEINE, Guido ; GÜNTHER, Stefan ; PREISSNER, Robert ; SLEVOGT, Hortense ; N’GUESSAN, Philippe D. ; EITEL, Julia ; GOLDMANN, Torsten ; FLIEGER, Antje ; SUTTORP, Norbert ; HIPPENSTIEL, Stefan: Legionella pneumophila induces IFNbeta in lung epithelial cells via IPS-1 and IRF3, which also control bacterial replication. In: *J Biol Chem* 281 (2006), S. 36173–36179
- [44] KHARE, Dheeraj ; ZIEGELIN, Günter ; LANKA, Erich ; HEINEMANN, Udo: Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. In: *Nat Struct Mol Biol* 11 (2004), Jul, Nr. 7, S. 656–663
- [45] GÜNTHER, Stefan ; ROTHER, Kristian ; FRÖMMEL, Cornelius: Molecular flexibility in protein-DNA interactions. In: *Biosystems* 85 (2006), S. 126–136
- [46] LENSINK, Marc F. ; MÉNDEZ, Raúl ; WODAK, Shoshana J.: Docking and scoring protein complexes: CAPRI 3rd Edition. In: *Proteins* 69 (2007), Dec, Nr. 4, S. 704–718
- [47] CHEN, Rong ; MINTSERIS, Julian ; JANIN, Joël ; WENG, Zhiping: A protein-protein docking benchmark. In: *Proteins* 52 (2003), Jul, Nr. 1, S. 88–91
- [48] RITCHIE, David W. ; KOZAKOV, Dima ; VAJDA, Sandor: Accelerating and Focusing Protein-Protein Docking Correlations Using Multi-Dimensional Rotational FFT Generating Functions. In: *Bioinformatics* (2008), Jun

- [49] PIERCE, Brian ; TONG, Weiwei ; WENG, Zhiping: M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. In: *Bioinformatics* 21 (2005), Apr, Nr. 8, S. 1472–1478
- [50] GÜNTHER, Stefan ; MAY, Patrick ; HOPPE, Andreas ; FRÖMMEL, Cornelius ; PREISSNER, Robert: Docking without docking: ISEARCH-prediction of interactions using known interfaces. In: *Proteins* 69 (2007), S. 839–844
- [51] HOPPE, Andreas ; FRÖMMEL, Cornelius: a program for the rapid recognition of local structures in large set of atomic coordinates. In: *J Appl Cryst* 36 (2003), S. 1090–1097
- [52] KUHN, Michael ; CAMPILLOS, Mónica ; GONZÁLEZ, Paula ; JENSEN, Lars J. ; BORK, Peer: Large-scale prediction of drug-target relationships. In: *FEBS Lett* 582 (2008), Apr, Nr. 8, S. 1283–1290
- [53] CAMPILLOS, Monica ; KUHN, Michael ; GAVIN, Anne-Claude ; JENSEN, Lars J. ; BORK, Peer: Drug target identification using side-effect similarity. In: *Science* 321 (2008), Jul, Nr. 5886, S. 263–266
- [54] JAIN, Kewal K.: Applications of AmpliChip CYP450. In: *Mol Diagn* 9 (2005), Nr. 3, S. 119–127
- [55] SCHNEIDER, Martin: A rational approach to maximize success rate in target discovery. In: *Arch Pharm (Weinheim)* 337 (2004), Dec, Nr. 12, S. 625–633
- [56] CHENG, Alan C. ; COLEMAN, Ryan G. ; SMYTH, Kathleen T. ; CAO, Qing ; SOULARD, Patricia ; CAFFREY, Daniel R. ; SALZBERG, Anna C. ; HUANG, Enoch S.: Structure-based maximal affinity model predicts small-molecule druggability. In: *Nat Biotechnol* 25 (2007), Jan, Nr. 1, 71–75. <http://dx.doi.org/10.1038/nbt1273>. – DOI 10.1038/nbt1273
- [57] DIVKOVIC, Maja ; PEASE, Camilla K. ; GERBERICK, G. F. ; BASKETTER, David A.: Hapten-protein binding: from theory to practical application in the in vitro prediction of skin sensitization. In: *Contact Dermatitis* 53 (2005), Oct, Nr. 4, S. 189–200
- [58] LI, Jiabo ; EHLERS, Tedman ; SUTTER, Jon ; VARMA-O'BRIEN, Shikha ; KIRCHMAIR, Johannes: CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. In: *J Chem Inf Model* 47 (2007), Nr. 5, 1923–1932. <http://dx.doi.org/10.1021/ci700136x>. – DOI 10.1021/ci700136x

- [59] DUNKEL, Mathias ; FULLBECK, Melanie ; NEUMANN, Stefanie ; PREISSNER, Robert: SuperNatural: a searchable database of available natural compounds. In: *Nucleic Acids Res* 34 (2006), D678–D683. <http://dx.doi.org/10.1093/nar/gkj132>. – DOI 10.1093/nar/gkj132
- [60] RECANATINI, Maurizio ; CAVALLI, Andrea ; MASETTI, Matteo: In silico modelling–pharmacophores and hERG channel models. In: *Novartis Found Symp* 266 (2005), S. 171–81; discussion 181–5
- [61] KLABUNDE, Thomas ; HESSLER, Gerhard: Drug design strategies for targeting G-protein-coupled receptors. In: *Chembiochem* 3 (2002), Oct, Nr. 10, S. 928–944
- [62] SANTO, Roberto D. ; FERMEGLIA, Maurizio ; FERRONE, Marco ; PANENI, Maria S. ; COSTI, Roberta ; ARTICO, Marino ; ROUX, Alessandra ; GABRIELE, Mirko ; TARDIF, Keith D. ; SIDDIQUI, Aleem ; PRICL, Sabrina: Simple but highly effective three-dimensional chemical-feature-based pharmacophore model for diketo acid derivatives as hepatitis C virus RNA-dependent RNA polymerase inhibitors. In: *J Med Chem* 48 (2005), Oct, Nr. 20, S. 6304–6314
- [63] NIESEN, Frank H. ; KOCH, Anja ; LENSKI, Ulf ; HARTTIG, Ulrich ; ROSKE, Yvette ; HEINEMANN, Udo ; HOFMANN, Klaus P.: An approach to quality management in structural biology: biophysical selection of proteins for successful crystallization. In: *J Struct Biol* 162 (2008), Jun, Nr. 3, 451–459. <http://dx.doi.org/10.1016/j.jsb.2008.03.007>. – DOI 10.1016/j.jsb.2008.03.007
- [64] QIAN, Bin ; RAMAN, Srivatsan ; DAS, Rhiju ; BRADLEY, Philip ; MCCOY, Airlie J. ; READ, Randy J. ; BAKER, David: High-resolution structure prediction and the crystallographic phase problem. In: *Nature* 450 (2007), Nov, Nr. 7167, 259–264. <http://dx.doi.org/10.1038/nature06249>. – DOI 10.1038/nature06249
- [65] ANDERSON, Amy C.: The process of structure-based drug design. In: *Chem Biol* 10 (2003), Sep, Nr. 9, S. 787–797
- [66] MERING, Christian von ; JENSEN, Lars J. ; SNEL, Berend ; HOOPER, Sean D. ; KRUPP, Markus ; FOGlierINI, Mathilde ; JOUFFRE, Nelly ; HUYNEN, Martijn A. ; BORK, Peer: STRING: known and predicted protein-protein associations, integrated and transferred across organisms. In: *Nucleic Acids Res* 33 (2005), Jan, Nr. Database issue, S. D433–D437

- [67] IOSSIFOV, Ivan ; KRAUTHAMMER, Michael ; FRIEDMAN, Carol ; HATZIVASSILOGLOU, Vasileios ; BADER, Joel S. ; WHITE, Kevin P. ; RZHETSKY, Andrey: Probabilistic inference of molecular networks from noisy data sources. In: *Bioinformatics* 20 (2004), May, Nr. 8, S. 1205–1213
- [68] REARDON, D. ; FARBER, G. K.: The structure and evolution of alpha/beta barrel proteins. In: *FASEB J* 9 (1995), Apr, Nr. 7, S. 497–503
- [69] GRÜNBERG, Raik ; LECKNER, Johan ; NILGES, Michael: Complementarity of structure ensembles in protein-protein binding. In: *Structure* 12 (2004), Dec, Nr. 12, S. 2125–2136
- [70] MUSTARD, Diana ; RITCHIE, David W.: Docking essential dynamics eigenstructures. In: *Proteins* 60 (2005), Aug, Nr. 2, S. 269–274
- [71] MAY, Andreas ; ZACHARIAS, Martin: Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. In: *Proteins* 70 (2008), Feb, Nr. 3, S. 794–809
- [72] CHAUDHURY, Sidhartha ; GRAY, Jeffrey J.: Conformer Selection and Induced Fit in Flexible Backbone Protein-Protein Docking Using Computational and NMR Ensembles. In: *J Mol Biol* (2008), May
- [73] HALPERIN, Inbal ; MA, Buyong ; WOLFSON, Haim ; NUSSINOV, Ruth: Principles of docking: An overview of search algorithms and a guide to scoring functions. In: *Proteins* 47 (2002), Jun, Nr. 4, S. 409–443
- [74] JANIN, Joël ; WODAK, Shoshana: The third CAPRI assessment meeting Toronto, Canada, April 20-21, 2007. In: *Structure* 15 (2007), Jul, Nr. 7, S. 755–759

# Abkürzungen

| Abkürzung | Erklärung  |
|-----------|--|
| DNS       | Desoxyribonukleinsäure   |
| RNS       | Ribonukleinsäure   |
| COX       | Cyclooxygenase   |
| REACH     | Registration, Evaluation, Authorisation and Restriction of Chemicals |
| 2D        | zweidimensional  |
| 3D        | dreidimensional  |
| PDB       | Protein Data Bank  |
| IPS-1     | Interferon-beta promoter stimulator protein 1                        |
| MDA5      | Melanoma differentiation-associated protein 5                        |
| CARD      | Caspase recruitment domain   |
| Nod       | Nucleotide-binding oligomerization domain-containing protein         |
| GFP       | Green fluorescent protein  |
| WHO       | World Health Organization  |
| Pfam      | Database of protein domain families                                  |
| RMSD      | Root Mean Square Deviation   |

# Danksagung

An dieser Stelle möchte ich mich bei einigen Menschen bedanken, durch deren Unterstützung die Erstellung der vorliegenden Arbeit nicht möglich gewesen wäre.

Zunächst danke ich Herrn PD Dr. Preißner, der mir die Gelegenheit gab, in seiner Arbeitsgruppe an der Lösung zahlreicher interessanter Fragestellungen der Bioinformatik mitzuarbeiten.

Des Weiteren danke ich Herrn Prof. Frömmel, dem es zu verdanken ist, dass meine Begeisterung für die Welt der Proteinstrukturen geweckt wurde.

Einen besonderen Dank verdienen meine lieben bestehenden oder Ex-Arbeitskollegen: Jessica Ahmed, Ulrike Schmidt, Elke Michalsky, Melanie Füllbeck, Ines Jäger, Julia Hoßbach, Swantje Struck, Mathias Dunkel, Raphael Bauer, Peter Hildebrand, Dominic Jansen, Björn Grüning und Kristian Rother. Sie haben dafür gesorgt, dass ich mich immer sehr wohl in der Arbeitsgruppe gefühlt habe auch wenn mal wieder eine der berühmten “Deadlines” vor der Tür stand.

Ferner danke ich Jan Saam, Andreas Hoppe, Christoph Gille, Ralph Kuban und Herrn Prof. Holzhütter für die angenehme Zusammenarbeit im Institut für Biochemie der Charité. Gleiches gilt für Bastian Opitz aus der Infektiologie der Charité.

Patrick May, Stephan Lorenzen und Armin Weiser verdanke ich viele lustige Abende an denen wir nicht nur wissenschaftliche Themen erörtert haben. In diesem Zusammenhang danke ich auch dem Team vom Anna Koschke. Meiner Freundin Lena danke ich dafür, dass sie den Stellenwert der Wissenschaft so häufig für mich ins rechte Licht gerückt hat.

Abschließend danke ich herzlichst meiner Familie für ihre Unterstützung in allen Lebenslagen.

# Eigene Publikationen

GÜNTHER, Stefan; ROTHER, Kristian; FRÖMMEL, Cornelius: Molecular flexibility in protein-DNA interactions. In: *Biosystems* 85 (2006), S. 126–136

GÜNTHER, Stefan; SENGER, Christian; MICHALSKY, Elke; GOEDE, Andrean; PREISSNER, Robert: Representation of target-bound drugs by computed conformers: implications for conformational libraries. In: *BMC Bioinformatics* 7 (2006), S. 293

OPITZ, Bastian; VINZING, Maya; LAAK, Vincent van; SCHMECK, Bernd; HEINE, Guido; GÜNTHER, Stefan; PREISSNER, Robert; SLEVOGT, Hortense; N'GUESSAN, Philippe D.; EITEL, Julia; GOLDMANN, Torsten; FLIEGER, Antje; SUTTORP, Norbert; HIPPENSTIEL, Stefan: Legionella pneumophila induces IFNbeta in lung epithelial cells via IPS-1 and IRF3, which also control bacterial replication. In: *J Biol Chem* 281 (2006), S. 36173–36179

AHMED, Jessica; GÜNTHER, Stefan; MÖLLER, Friedrich; PREISSNER, Robert: A Structural Genomics Approach to the Regulation of Apoptosis: Chimp vs. Human. In: *Genome Inform* 18 (2007), S. 22–34

GÜNTHER, Stefan; HEMPEL, Dorothea; DUNKEL, Mathias; ROTHER, Kristian; PREISSNER, Robert: SuperHapten: a comprehensive database for small immunogenic compounds. In: *Nucleic Acids Res* 35 (2007), S. D906–D910

GÜNTHER, Stefan; MAY, Patrick; HOPPE, Andreas ; FRÖMMEL, Cornelius; PREISSNER, Robert: Docking without docking: ISEARCH–prediction of interactions using known interfaces. In: *Proteins* 69 (2007), S. 839–844

GÜNTHER, Stefan; NEUMANN, Stefanie; AHMED, Jessica; PREISSNER, Robert: Cellular Fingerprints: A Novel Concept for the Integration of Experimental Data and Compound-Target-Pathway Relations. In: *LNBI 4643* (2007), S. 167–70



GÜNTHER, Stefan; WEISER, Armin A.; AHMED, Jessica; PREISSNER, Robert: Superhaptens and their potential role in contact dermatitis. In: *Drug Discovery Today: Therapeutic Strategies* 4 (2007), S. 11–17

DUNKEL, Mathias; GÜNTHER, Stefan; AHMED, Jessica ; WITTIG, Burghardt; PREISSNER, Robert: SuperPred: drug classification and target prediction. In: *Nucleic Acids Res* 36 (2008), S. W55–W59

GÜNTHER, Stefan; KUHN, Michael; DUNKEL, Mathias; CAMPILLOS, Monica; SENGER, Christian; PETSALAKI, Evangelia; AHMED, Jessica; URDIALES, Eduardo G.; GEWIESS, Andreas; JENSEN, Lars J.; SCHNEIDER, Reinhard; SKOBLO, Roman; RUSSELL, Robert B.; BOURNE, Philip E.; BORK, Peer; PREISSNER, Robert: SuperTarget and Matador: resources for exploring drug-target relationships. In: *Nucleic Acids Res* 36 (2008), S. D919–D922

HILDEBRAND, Peter W.; GÜNTHER, Stefan; GOEDE, Andrean; FORREST, Lucy; FRÖMMEL, Cornelius; PREISSNER, Robert: Hydrogen-bonding and packing features of membrane proteins: functional implications. In: *Biophys J* 94 (2008), S. 1945–1953

.....

BAUER, Raphael A.; GÜNTHER, Stefan; JANSEN, Dominic; HEEGER, Carolin; THABEN, Paul F.; PREISSNER, Robert: SuperSite: Dictionary of metabolite and drug binding sites in proteins. In: *Nucleic Acids Res (submitted)* (2008)

GÜNTHER, Stefan; EICHBORN, Joachim von; PREISSNER, Robert: JAIL: a structure-based interface library for macromolecules. In: *Nucleic Acids Res (submitted)* (2008)

# Selbständigkeitserklärung

Diese Arbeit entstand in der Zeit zwischen März 2005 und November 2006 am Institut für Biochemie (Charité, Berlin) und von Dezember 2006 bis Juli 2008 am Institut für Molekularbiologie und Bioinformatik (Charité, Berlin) in der Arbeitsgruppe strukturelle Bioinformatik (AG-Leiter: Herr PD Dr. Preißner). Die Arbeit wurde selbständig verfasst und alle Hilfsmittel wurden entsprechend aufgeführt.

Stefan Günther  
2. August 2008