

Analyse und Vorhersage der Interaktionen von
Wirkstoffen mit ihren Zielmolekülen zur
Tumorthherapie

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
Jessica Ahmed
aus Berlin

2011

Diese Arbeit wurde im Zeitraum Januar 2007 bis Februar 2011 unter der Leitung von PD Dr. Robert Preißner an der Charité - Universitätsmedizin Berlin angefertigt.

1. Gutachter: PD Dr. Robert Preißner (Charité, Berlin)
2. Gutachter: Prof. Dr. Burghardt Wittig (Freie Universität Berlin)

Disputation am 02.08.2011

Abstract

This work contributes to the integration of methods and knowledge in order to personalize and improve tumor therapy. It seeks to build a bridge between the different levels of genes, proteins, drugs and signaling pathways, which are usually considered separately. Small, therapeutically active molecules (potential drugs) can, and should be viewed from different perspectives, because 1) the chemical structure and the underlying fragments are important aspects of synthesis and preclinical evaluation and 2) their three-dimensional structure and the energetically favorable conformations are important for binding to the medically relevant targets. This has been implemented in the present work algorithmically.

To identify novel and appropriate targets, an analysis that considers the entire genome compared to our much less cancer susceptible relative, the chimpanzee, was performed. In this work tumor-related pseudogenes were found.

The concept of cellular fingerprints was developed and implemented in order to establish a link between experimental data on the chemosensitivity of cancer cell lines and their particular protein targets. By including the associated expression data, the rational selection of therapeutic agents for a specific type of tumor was possible. The final part of this work was also to develop predictions for new, innovative drugs. The identification of Notch-specific γ -secretase modulators for the treatment of multiple myeloma and renal failure is based on the methodology of analogy modeling developed in this group and presents the repositioning of a target. The use of Aprepitant as an apoptosis inducer demonstrates drug-repositioning. Both groups of compounds have been successfully patented.

Keywords:

tumor therapy, Notch-pathway, γ -secretase, angiogenesis, VEGFR-2, apoptosis, target prediction, compound prediction

Zusammenfassung

Diese Arbeit leistet einen Beitrag zur Integration von Methoden und Wissen, um eine individuelle und verbesserte Tumorthherapie zu ermöglichen. Dabei wurde versucht, eine Brücke zwischen den verschiedenen Ebenen: Gene, Proteine, Wirkstoffe und Signalwege, zu schlagen, die sonst oftmals separat betrachtet werden. Kleine, therapeutisch wirksame Moleküle, also potenzielle Arzneistoffe, können und müssen aus verschiedenen Blickwinkeln betrachtet werden, weil 1) die chemische Struktur und die zu Grunde liegenden Fragmente wichtig für Aspekte der Synthese und die präklinische Beurteilung sind und 2) ihre dreidimensionale Struktur und die energetisch günstigen Konformationen wichtig für die Bindung an die medizinisch relevanten Zielmoleküle (Targets) sind. Dies wurde in der vorliegenden Arbeit algorithmisch umgesetzt.

Um geeignete unter Umständen neue Targets zu identifizieren, ist ein größerer Blickwinkel empfehlenswert, der das gesamte Genom betrachtet, beispielsweise im Vergleich zu unserem weniger Tumor-anfälligen Verwandten, dem Schimpansen. Im Rahmen dieser Arbeit konnten auf diesem Weg Tumor-relevante Pseudogene aufgefunden werden.

Das Konzept der zellulären Fingerprints wurde entwickelt und implementiert, um eine Verbindung zwischen experimentellen Daten zur Chemosensitivität von Tumorzelllinien und ihrer Targetausstattung herzustellen. Durch Einbeziehung von zugehörigen Expressionsdaten wurde die rationale Auswahl von Therapeutika für einen bestimmten Typ von Tumorerkrankung ermöglicht. Gegenstand dieser Arbeit war es weiterhin, explizite Vorhersagen für neue, innovative Wirkstoffe zu erarbeiten. Die Identifizierung von Notch-spezifischen γ -Sekretase Modulatoren zur Therapie des Multiplen Myeloms und der Niereninsuffizienz beruhte auf der in dieser Arbeitsgruppe entwickelten Analogie-Modellierung und stellt eine Neupositionierung eines Targets dar. Die Verwendung von Aprepitant als Apoptose-Induktor stellt eine Neupositionierung eines Arzneistoffes dar. Beide Substanzgruppen wurden erfolgreich patentiert.

Schlagwörter:

Tumorthherapie, Notch-Signalweg, γ -Sekretase, Angiogenese, VEGFR-2, Apoptose, Targetvorhersage, Ligandenvorhersage

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Zielsetzung	4
1.3	Physikalische Aspekte	5
1.3.1	Protein-Liganden-Wechselwirkungen	5
1.3.2	Entropische Beiträge	7
1.3.3	Affinität und Spezifität	8
1.4	Biochemische Aspekte	10
1.4.1	Apoptose	10
1.4.2	Angiogenese und die Rolle der vaskulären Wachstumsrezeptoren	12
1.4.3	Der Notch-Signalweg und die Rolle der γ -Sekretase Proteine .	14
1.5	Wirkstoffentwicklung	18
1.5.1	Ziele und Wege der Wirkstoffentwicklung	18
1.5.2	Eigenschaften oral verfügbarer Arzneistoffe	22
1.5.2.1	Lipinskis 5er Regel	22
1.5.2.2	ADMETox	23
1.5.3	NCI 60 Screening Projekt	24
1.6	Genetische und evolutionäre Aspekte	25
1.7	Datenintegration	27
2	Material und Methoden	29
2.1	Analyse von Zielproteinen, Wirkstoffen und ihren Interaktionen . . .	29
2.1.1	Genom- und Proteinstrukturanalyse onkologisch relevanter Proteine von Mensch und Schimpanse	29
2.1.1.1	Sequenzvergleiche	33
2.1.2	Generierung und Assemblierung von Kleinstrukturfragmenten	33
2.1.3	Berechnung der Bindungstaschenpräferenzen von Fragmenten	37
2.2	Vorhersage von Zielproteinen und potenziellen Liganden	38

2.2.1	Homologie-Modellierung	38
2.2.2	<i>In silico</i> Screening	40
2.2.2.1	2D-Ähnlichkeit	41
2.2.2.2	3D-Ähnlichkeit	42
2.2.2.3	Protein-Liganden-Docking	43
2.2.3	ADMETox-Vorhersage	45
2.2.4	Analogie-Modellierung	47
2.2.5	Fingerprints zum schnellen Vergleich von Kleinstrukturen und Wirkprofilen	49
2.2.5.1	Struktureller Fingerprint zum Vergleich von Kleinstrukturen	49
2.2.5.2	Zellulärer Fingerprint zum Vergleich von Wirkprofilen	50
2.2.6	Target-Vorhersage	52
2.3	Integration von Zielproteinen, Wirkstoffen und ihren Interaktionen . .	53
2.3.1	Semi-automatisierte Wissensextraktion aus der Literatur . . .	54
2.3.2	Datenakquise für die Integration von Kleinstrukturen und Zielproteinen	55
2.3.3	Implementierung der Datenbanksysteme	59
3	Ergebnisse	60
3.1	Zielproteine	60
3.1.1	Evolutionäre Aspekte	60
3.1.1.1	Schimpanse vs. Mensch	61
3.2	Wirkstoffe	66
3.2.1	Membranproteine als Zielmolekül	66
3.2.1.1	Arbeitsmodell zur γ -Sekretase	66
3.2.1.2	VEGF-Rezeptor 2 als Targetprotein	74
3.3	Wirkstoff-Zielprotein-Interaktionen	81
3.3.1	Integration von Wirkstoff-Zielprotein-Interaktionen	81
3.3.1.1	Zellulärer Fingerprint zum Vergleich von Wirkprofilen	82
3.3.1.2	SuperTarget - Arzneistoff-Zielprotein-Interaktionen .	84
3.3.1.3	CancerResource (CaRe) - onkologische Datenbank mit integrierten Analyse- und Vorhersagemethoden .	90
3.3.2	SuperPred - Target-Vorhersage	98
3.3.3	FragmentStore - Fragmentierung und Analyse	101

4	Diskussion	106
4.1	Evolutionäre Aspekte	106
4.2	γ -Sekretase Modulatoren	109
4.3	VEGF-Rezeptor 2 als Zielprotein	112
4.4	Target-Vorhersage	114
4.5	Integration von Wirkstoff-Protein-Interaktionen	115
4.5.1	SuperTarget	115
4.5.2	CancerResource	117
4.6	FragmentStore - Fragmentanalysen	118
5	Ausblick	120
A	Verzeichnisse	121
	Abkürzungsverzeichnis	121
	Abbildungsverzeichnis	123
	Tabellenverzeichnis	125
	Literaturverzeichnis	127
B	Liste eigener Publikationen	148
	Peer-reviewed Publikationen	148
	Buchartikel	149
	Patente	150

Kapitel 1

Einleitung

1.1 Motivation

Im Jahr 2006 erkrankten in Deutschland nach Angaben des Robert-Koch-Instituts 426.800 Menschen, davon 229.200 Männer und 197.600 Frauen, an malignen Neubildungen, malignen soliden und hämoblastischen Tumoren [1]. Über 210.000 Todesfälle lassen sich 2006 auf diese onkologischen und hämatologischen Erkrankungen zurückführen. Diese Erkrankungen stellen immer noch die zweithäufigste Todesursache in Deutschland dar. Nach Schätzungen des Robert-Koch-Instituts werden für das Jahr 2010 rund 450.000 Neuerkrankungen erwartet [1]. Aktuell leben in Deutschland hochgerechnet 1,45 Millionen Menschen, bei denen in den letzten fünf Jahren eine onkologische oder hämatologische Erkrankung diagnostiziert wurde. Im Vergleich mit Zahlen aus dem Jahr 1990 bedeutet dies einen Zuwachs von 40% bei Frauen und 90% bei Männern, welcher auch unter anderem durch den demographischen Wandel begründet ist [1]. Aufgrund der stetigen Weiterentwicklung von Methoden zur Therapie haben sich die Überlebensprognosen der betroffenen Patienten in den letzten Jahrzehnten deutlich verbessert. Dennoch stellt die frühzeitige Erkennung und adäquate Therapie der Erkrankung eine erhebliche Herausforderung für die pharmazeutische und medizinische Wissenschaft dar.

Eine einheitliche Therapie dieser Erkrankungen ist nicht möglich, denn die Unterschiede in der Ausprägung sind mannigfaltig. Besonders auffällig sind geschlechtsspezifische Unterschiede, aber lokale, gewohnheitsabhängige oder ethnische Differenzen in der Inzidenz sind ebenfalls offensichtlich.

Obwohl onkologische Erkrankungen im gesamten Tierreich auftreten, gehört un-

ser engster genetischer Verwandter, der Schimpanse, zu den Tierarten, die praktisch nicht von Tumorerkrankungen betroffen sind [2, 3].

Vor allem Prostata- [4], Brustdrüsen- und Bronchialkarzinome sowie spontane Entartungen [3] entstehen kaum in Schimpansen-Populationen [5, 6, 7]. Es kann daher davon ausgegangen werden, dass die zwei Prozent genetischer Unterschied [8, 9] beim Menschen zu Problemen in der Auslösung der Apoptose- oder Reparatur-Mechanismen geführt haben. Im Verlauf der vorliegenden Arbeit wurde das Genom des Schimpansen verfügbar, so dass eine Analyse der Unterschiede der Ausstattung mit onkologisch relevanten Proteinen möglich wurde. Das Projekt zur Aufklärung des humanen Genoms [10] sollte es entsprechend der Erwartungen ermöglichen, viele Erkrankungen wie Karzinome ursächlich zu erkennen und damit kausal behandelbar zu machen. Allerdings stellte sich heraus, dass unser Verständnis der komplexen Regulation der Zellteilung und Apoptose so gering entwickelt ist, dass bisher kaum bekannt ist, welche Veränderungen an welchen Proteinen für die Entstehung und Ausprägung der Erkrankungen verantwortlich sind. Deshalb ist ein Bedarf entstanden, Veränderungen an medizinisch relevanten Zielmolekülen („Targets“, oft überexprimiert) mit zugehörigen Arzneistoffen (oft Inhibitoren) zu vernetzen. Dieser Aufgabe haben sich einige Arbeitsgruppen gestellt und es sind Datenbanken wie DrugBank [11], TTD [12] und CTD [13] entstanden. Dabei hat sich die Erkenntnis durchgesetzt, dass die Projektion der Zielproteine („Targets“) auf die metabolischen und Signal-Kaskaden zum Verständnis der Vorgänge erforderlich ist, ohne dass es bisher ein fundiertes Mapping zwischen Arzneistoffen, ihren Zielmolekülen und den Funktionseinheiten („Pathways“) gäbe.

Eine Analyse maligner Tumorerkrankungen zeigt, dass die starke Versorgung des Tumors mit Blutgefäßen ein typisches Merkmal ist, das mit ungünstiger Prognose verbunden ist, weshalb die Adressierung der Angiogenese bei einigen Tumoren die einzige Therapieoption darstellt. In dieser Arbeit wurde deshalb der vaskuläre endotheliale Wachstumsfaktor-Rezeptor (VEGFR) als interessantes Target für eine spezifische Therapie ausgewählt.

Die Adressierung neuer Targets in der Tumorthherapie wird bisher nur in geringem Umfang unternommen. Ein weiteres typisches Merkmal des Tumorwachstums ist die Fehlregulation der Proliferation. Die daran beteiligten Rezeptoren sind bekannt sowie die Erkenntnis, dass der Notch-Rezeptor dabei eine wichtige Rolle spielt. Die

Aktivierung des Notch-Signalweges wird endgültig ausgelöst, wenn die γ -Sekretase, eine Protease, den Notch-Rezeptor an geeigneter Stelle schneidet. Daher wurde die γ -Sekretase, eigentlich bekannt für ihre Rolle bei der Entstehung von neurotoxischen Varianten des β -Amyloids bei Morbus Alzheimer [14, 15], als interessantes Target für die Regulation der Apoptose identifiziert [16, 17]. Des Weiteren induziert die Inhibition des Notch-Schnitts auch in Zellen des Multiplen Myeloms Apoptose [18]. Eine spezifische Hemmung der γ -Sekretase bzgl. des Schnitts am Notch-Rezeptor wäre daher ein vielversprechendes Ziel der pharmazeutischen Wirkstoffentwicklung [19, 20].

Auf Grund der Relevanz der onkologischen und hämatologischen Erkrankungen wurden große experimentelle Programme aufgelegt, in denen ein Panel standardisierter Tumor-, Lymphom-, Melanom- und Leukämie-Zelllinien auf ihre Sensitivität bzgl. zehntausender chemischer Substanzen untersucht wurde [21]. Dabei wird deutlich, dass bestimmte Substanzen insgesamt recht zytotoxisch sind, während andere (interessantere) Substanzgruppen nur bestimmte Zelllinien an der Proliferation hindern. Dieses Muster von über 40.000 Substanzen und ihrer Wirkung auf etwa 100 Zelllinien liefert gleichzeitig ein Muster der Target-Ausstattung der jeweiligen Zelllinien, denn beispielsweise können nicht vorhandene bzw. nicht funktionelle Targets auch nicht inhibiert werden, was die entsprechende Resistenz gegen diese Substanzgruppe erklärt. Allerdings gibt es bisher kein Verfahren, um diese vorhandenen Daten zur schnellen, automatischen Suche nach individuell geeigneten Therapeutika einzusetzen.

Bisher sind die Erfolge der kombinatorischen Synthese bei der Generierung geeigneter Wirkstoffe relativ gering. Die meisten Wirkstoffe lassen sich auf Naturstoffe zurückführen, wobei eine Analyse zeigt, dass dies mit fehlender Stereoselektivität und fehlender Diversität zusammenhängt. Die chemischen Bausteine, die dabei Verwendung finden, werden als Fragmente bezeichnet. Eine umfassende Analyse des Vorkommens von Fragmenten in Naturstoffen, Metaboliten, Arzneistoffen oder toxischen Substanzen steht jedoch aus. Nur auf Grund einer solchen Datenbasis wäre eine rationale Auswahl geeigneter Fragmente für bestimmte Targets möglich.

Die Zusammenführung der genannten Ebenen in einer benutzerfreundlichen, Web-basierten Anwendung ist noch nicht vorhanden. Dabei soll eine Genom-basierte Analyse die onkologisch relevanten Targets liefern. Eine umfassende Analyse soll

möglichst viele Substanzen integrieren, die für die Targetadressierung geeignet sind. Um Kombinationstherapie rationaler gestalten zu können, ist es erforderlich, die entsprechenden Signalwege für die relevanten Targets zu integrieren. Mutationen an den Targets und ihre Auswirkungen sollen schnell auffindbar sein, die genetische Ausstattung der verschiedenen Tumore muss vergleichbar sein, denn nur durch die Integration dieser verschiedenen Informationsebenen können die zunehmend entstehenden Ergebnisse des „Next-Generation Sequencing“ von Tumorkranken wirklich Therapie-verbessernd verwendet werden.

1.2 Zielsetzung

Diese Arbeit verfolgte im Wesentlichen vier Ziele. Zunächst sollte eine methodische Infrastruktur geschaffen werden. Im zweiten Teil wurden konkrete Targets in ihrem Signalkontext und ihre mögliche Modulation zur Therapie der onkologischen und hämatologischen Erkrankungen behandelt. Im dritten Teil sollten neue Moleküle, also potenzielle Wirkstoffe, erarbeitet werden, die für die Therapie geeignet sind. Abschließend sollte die Integration der verschiedenen Daten und Methoden und die damit verbundene Veröffentlichung für weitere Analysen bereitgestellt werden.

1.3 Physikalische Aspekte

1.3.1 Protein-Liganden-Wechselwirkungen

Bei der Tumor-Therapie werden Wirkstoffe angewandt, die über die Bindung an bestimmte Zielproteine ihre Wirkung entfalten. Meistens besteht diese Wirkung darin, ein bestimmtes Protein zu aktivieren oder zu inhibieren und nachgeschaltete Signalkaskaden auszulösen oder zu unterbinden. Für die Ausbildung einer Protein-Liganden-Wechselwirkung müssen bestimmte Voraussetzungen erfüllt sein. Die Oberflächeneigenschaften der beiden Bindungspartner müssen zueinander kompatibel sein. Es gibt verschiedene Protein-Liganden-Wechselwirkungen, wobei grob zwischen kovalenten und nicht-kovalenten Bindungen unterschieden wird. Als kovalent wird eine Bindung bezeichnet, bei der jeweils mindestens ein Valenzelektron des einen Bindungspartners mit einem Valenzelektron des anderen Bindungspartners ein gemeinsames Elektronenpaar bildet. Handelt es sich um unpolare Bindungen, so wird diese Bindung als unpolare Atombindung bezeichnet, wobei die Elektronegativitätsdifferenz zwischen beiden Bindungspartnern gleich Null (max. 0.4) ist. Bei polaren Atombindungen wird das Elektronenpaar weitgehend nur dem Bindungspartner mit höherem Elektronegativitätswert zugeordnet. Die Bindung zwischen zwei elektrisch geladenen Atomen (Ionen) wird als ionische Bindung bezeichnet und die Elektronegativitätsdifferenz ist größer 1,7 [22]. Bei nicht-kovalenten Bindungen handelt es sich um Wechselwirkungen zwischen benachbarten Atomen, bei der keine chemische Bindung entsteht. Sie lassen sich in die drei Beiträge aufteilen: die elektrostatischen Wechselwirkungen (Coulomb-Potenzial), die Van der Waals Wechselwirkungen (Lennard-Jones-Potenzial) und die hydrophoben Wechselwirkungen. Die elektrostatischen Wechselwirkungen resultieren aus Partialladungen der Atome eines Moleküls, die durch die unterschiedlichen Elektronegativitäten der Atome entstehen. Entgegengesetzt geladene Atome von Molekülen weisen bei geringer werdender Distanz eine sehr starke anziehende Wechselwirkung auf, wobei im Gegensatz dazu gleich geladene Atome eine starke Abstoßung erzeugen (siehe Abbildung: 1.1). Das Coulomb-Potenzial wird meist mittels

$$P_{Coulomb}(r) = \frac{Z_a Z_b}{4\pi\epsilon_0 r}$$

bestimmt. Die Variablen Z_a und Z_b bezeichnen die Partialladungen der Teilchen

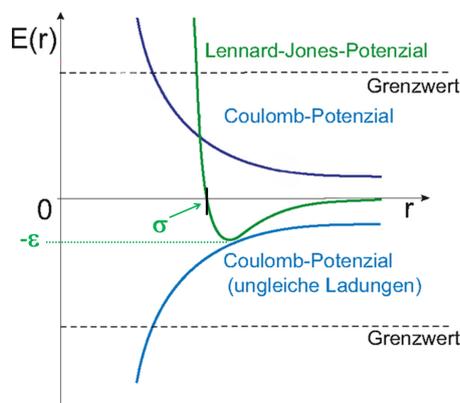
a und b , die Konstante ϵ_0 ist die elektrische Feldkonstante ($8,854 \cdot 10^{-12} \frac{As}{Vm}$) und r bezeichnet den Abstand zwischen den Partiaalladungen. Durch die Veränderung der Dielektrizitätskonstante wird das umgebene Medium berücksichtigt und die relative Permittivität verwendet. Für computergestützte Energieberechnungen beispielsweise bei Moleküldynamik-Simulationen oder Protein-Liganden-Docking Methoden wird, um ein Lösungsmittel zu berücksichtigen, der Wert der relativen Dielektrizitätskonstante ϵ_r zwischen 2 und 80 gewählt. Zu den elektrostatischen Wechselwirkungen gehören die Wasserstoffbrückenbindungen und die ionischen Wechselwirkungen. Die Wasserstoffbrückenbindung bezeichnet die polare Wechselwirkung, die zwischen dem partiell geladenen Wasserstoffatom, welches an ein elektronegativeres Atom gebunden ist, und einem elektronegativen Atom mit freiem Elektronenpaar besteht. Die Wasserstoffbrückenbindung ist im Vergleich zur ionischen Wechselwirkung eine gerichtete und schwächere Wechselwirkung. Ionische Wechselwirkungen bezeichnen die Anziehung zwischen entgegengesetzt geladenen Atomen (Ionen) und sind ungerichtete Coloumb-Wechselwirkungen zwischen Ladungen. Sie gehören zu den stärksten elektrostatischen Wechselwirkungen zwischen Atomen.

Van der Waals Kräfte entstehen zwischen nahe liegenden, ungeladenen Atomen. Es wird zwischen Dipol-Dipol Wechselwirkungen und Wechselwirkungen zwischen induzierten Dipolen unterschieden. Bei Dipol-Dipol-Wechselwirkungen bestehen permanent Partiaalladungen zwischen zwei Molekülen, während bei Dipol induzierten Dipol-Kräften durch einen Dipol ein Dipol in einem ungeladenen Molekül induziert wird. Bei induzierten Dipol-Kräften entstehen induzierte Partiaalladungen, die wiederum Dipole in Nachbarmolekülen induzieren. Es handelt sich um relativ schwache Wechselwirkungen, welche aufgrund polarer Eigenschaften der Atome entstehen. Die Energie der Van der Waals Wechselwirkung wird durch das Lennard-Jones-Potenzial beschrieben:

$$P_{Lennard-Jones}(r) = 4\epsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right)$$

Die Variable ϵ beschreibt die Tiefe des Potenzialminimums der beiden wechselwirkenden Teilchen und die Variable σ den Abstand beider Teilchen mit Potenzialwert Null. Es erfolgt damit eine Berechnung der Abweichung des aktuellen Abstands r zweier Teilchen von σ . Der E12-Term beschreibt dabei die Abstoßung der beiden

Teilchen und der E_6 -Term ihre Anziehung. Eine große Annäherung der beiden Teilchen führt dadurch zu einem exponentiellen Anstieg der Energie (Abbildung 1.1).



Aus: Klebe, Wirkstoffdesign, 2. Aufl. © Spektrum Akademischer Verlag GmbH, 2009

Abb. 1.1: Lennard-Jones- und Coulomb-Potenzial.

Aufgetragen ist das Potenzial $E(r)$ gegen die Distanz r . Das Lennard-Jones-Potenzial (grün) beschreibt näherungsweise die Wechselwirkung zwischen zwei ungeladenen, aneinander nicht gebundenen Atomen. Je weiter zwei Atome voneinander entfernt sind, desto geringer ist die Anziehungskraft und das Potenzial nähert sich der Null. Mit sich verringerndem Abstand zweier Atome steigt die Anziehungskraft und der Wert des Potenzials wird negativer. Je mehr sich die Wechselwirkung dem Potenzialminimum nähert, desto schneller sinkt der Potenzialwert. Je näher sich die Atome ab dem Potenzialminimum kommen, desto stärker ist die Abstoßung und das Potenzial nimmt positive Werte an. Das Coulomb-Potenzial (blau) beschreibt die elektrostatische Wechselwirkung zwischen kugelförmigen elektrischen Punktladungen (z.B. zwei Ionen). Bei zwei gleich geladenen Atomen und verringertem Abstand erhöht sich die Abstoßung und das Potenzial nimmt unendlich positive Werte an. Sind zwei Atome ungleich geladen und verringert sich der Abstand, erhöht sich die Anziehung und das Potenzial nimmt unendlich negative Werte an. *Aus Klebe (modifiziert) [23].*

Hydrophobe Wechselwirkungen basieren auf Entropieeffekten und entstehen zwischen unpolaren Gruppen des Proteins und hydrophoben Gruppen des Liganden, die sich in unmittelbarer Nähe befinden, wobei das Lösungsmittel verdrängt und damit die Entropie erhöht wird. Die hydrophoben Wechselwirkungen zeigen häufig den größten Beitrag zur Affinität des Proteins zu Liganden mit großem lipophilen Anteil [23].

1.3.2 Entropische Beiträge

Entropische Beiträge beeinflussen die Komplexbildung zwischen Liganden und Proteinen. Die Komplexbildung im biologischen System findet in der Zelle in wässriger Umgebung statt, was einen entscheidenden Einfluss auf die Bindungskonstanten und die Energie hat. Der Ligand ist von Wassermolekülen (Hydrathülle bzw. Hydratsphäre) umgeben und auch die Bindungstasche des Zielproteins ist mit Wassermolekülen

ausgefüllt. Diese Wassermoleküle und auch die umgebenen Moleküle müssen zunächst vom Liganden aus der Bindungsstelle verdrängt werden, damit es zu einer Komplexbildung zwischen Ligand und Protein kommen kann. Durch die Aufhebung der Wechselwirkungen zwischen den Wassermolekülen und der Tasche des Proteins können sich die Wassermoleküle frei bewegen, was die Anzahl der Freiheitsgrade und somit die Entropie erhöht. Die Bindung des Liganden an das Protein führt zu einer Erniedrigung der Entropie, da die Anzahl der Freiheitsgrade durch eine Erniedrigung der Rotations- und Translationsmöglichkeiten gesenkt wird. In den meisten Fällen überwiegt der Einfluss der Hydrathülle auf die Entropie und dessen Auflösung resultiert in einer höheren Entropiedifferenz, was zu einer Bevorzugung der Ligandenbindung führt.

$$\underbrace{\Delta G}_{\text{Freie Bindungsenthalpie}} = \Delta \underbrace{H}_{\text{Enthalpie}} - \underbrace{T}_{\text{Temperatur}} \Delta \underbrace{S}_{\text{Entropie}}$$

Eine Erhöhung des entropischen Beitrages führt zu einer Erniedrigung der freien Bindungsenthalpie. Je niedriger die freie Bindungsenthalpie ist, desto höher ist die Affinität.

Es gibt nicht nur Entropie-, sondern auch Enthalpie-getriebene Bindungen [23].

1.3.3 Affinität und Spezifität

Die Stärke der Wechselwirkung eines Protein-Liganden-Komplexes wird auch als Affinität bezeichnet. Sie ist experimentell messbar und wird mittels der Assoziationskonstante (Bindungskonstante) Ka , der Dissoziationskonstante (Kehrwert der Bindungskonstante) Kd oder als freie Bindungsenthalpie ΔG angegeben. Die Konstanten Ka und Kd beschreiben das Gleichgewicht zwischen gebundenen und ungebundenen Molekülen. Dabei beschreibt die Bindungskonstante die Bildung des Protein-Liganden-Komplexes und es gilt: je größer die Bindungskonstante ist, desto höher ist die Affinität eines Liganden zu einem Protein.

$$Ka = \frac{[Ligand \cdot Protein]}{[Ligand] \cdot [Protein]} \quad Kd = \frac{[Ligand] \cdot [Protein]}{[Ligand \cdot Protein]}$$

Im Gegensatz dazu gilt bei Angabe der Dissoziationskonstante: je kleiner der Wert,

desto höher die Affinität. Die Dissoziationskonstante beschreibt den Zerfall des Protein-Liganden-Komplexes. Die freie Bindungsenthalpie ΔG kann ebenfalls als Maß für die Affinität der Reaktionspartner interpretiert werden [23]. Es gilt:

$$\Delta G = RT \ln K,$$

wobei R der allgemeinen Gaskonstante ($8,314 \frac{J}{molK}$), T der absoluten Temperatur und K der Gleichgewichtskonstante entspricht. Hierbei gilt: je niedriger ΔG , desto höher die Affinität.

Biologische Maße für die Affinität

Neben den Konstanten können auch IC50-Werte als Maß für die inhibitorische Stärke der Protein-Liganden-Wechselwirkung angegeben werden. Der IC50-Wert gibt die Konzentration des Liganden an, bei der nur noch die Hälfte der Proteinaktivität vorhanden ist. Bei Zellversuchen werden außerdem bevorzugt der GI50-Wert (50%ige Wachstumshemmung), der LC50-Wert (Absterben der Hälfte der (Tumor-)Zellen) und der TGI-Wert (totale Wachstumshemmung) bestimmt, um die Wirkung eines Liganden zu messen.

Die Spezifität bezeichnet die Fähigkeit des Liganden selektiv nur ein bestimmtes Zielprotein oder eine bestimmte Zielproteingruppe zu binden. Oft zeigen Arzneistoffkandidaten, die eine hohe Zielprotein-Affinität besitzen, nur sehr niedrige Spezifität, da sie ebenfalls strukturell ähnliche Proteine adressieren, was zu unerwünschten Nebenwirkungen führen kann. Beispielsweise kann die Inhibition bestimmter Kinasen, wie der Tyrosinkinase-gekoppelte IGF-1-Rezeptor (Insulin-like Growth Factor 1), der in verschiedenen Signalwegen wie Zellwachstum und Apoptose eine entscheidende Rolle spielt, zur Apoptose-Induktion führen [24]. Die Verwendung hoch-affiner Inhibitoren, die zugleich verwandte Kinasen, wie den Insulinrezeptor adressieren, führt zur Ausbildung unerwünschter Phänotypen bzw. Krankheitsbildern, in diesem Fall zu einem diabetischen Phänotyp [25]. Daher wird bei der Wirkstoffentwicklung eine Balance zwischen Affinität und Spezifität angestrebt.

1.4 Biochemische Aspekte

1.4.1 Apoptose

In der onkologischen Arzneimittelforschung und -entwicklung ist ein Vergleich der Expressionsdaten des gesamten Genoms nicht zielführend, vielmehr ist eine spezifische Analyse der in der Malignom-Entstehung involvierten Signalwege und Proteine erforderlich. Ein während der Tumorbildung wichtiger Signalweg ist die Apoptose. Apoptose bezeichnet den programmierten Zelltod, der einer starken Regulation und Kontrolle unterliegt. Jede Zelle hat definierte Mechanismen, die in Gang gesetzt werden, sobald sie durch äußere oder innere Einflüsse geschädigt wird. Durch bestimmte physiologische Signale oder durch Schädigung der DNA kann Apoptose induziert werden. Täglich sterben mehrere Milliarden geschädigte oder gealterte Zellen des menschlichen Organismus durch den programmierten Zelltod [26]. Eine Deregulation der Apoptose spielt eine essentielle Rolle bei der Ausbildung von verschiedenen Krankheitsphänotypen. Eine verringerte Apoptose-Induzierung resultiert oftmals in der Entstehung von Tumoren [27], während eine verstärkte Apoptoseauslösung zu Krankheiten wie Morbus Alzheimer [28, 29] oder Chorea Huntington führen kann [30, 31]. Aus diesem Grund stellen die vollständige Aufklärung der Apoptose-Mechanismen und die Bestimmung aller in apoptotischen Signalwegen involvierten Proteine ein Forschungsgebiet zur Entwicklung neuer Therapiemöglichkeiten dar.

Die Aktivierung der Apoptose kann über einen extrinsischen und einen intrinsischen Initiationsweg ausgelöst werden und führt zu einem sich anschließenden Effektorweg, auch Caspase-Kaskade genannt. Abbildung 1.2 zeigt eine schematische Darstellung dieser Wege.

Während des extrinsischen Weges führt die Bindung von Signalen an sogenannte Todesrezeptoren auf der Zelloberfläche zur Einleitung der Apoptose. Zu den aktuell bekannten Todesrezeptoren gehören:

TNF-R1 (CD120a), CD95 (APO-1/Fas), DR3 (APO-3/LARD/TRAMP/WSL1), TRAIL-R1 (APO-2/DR4), TRAIL-R2 (DR5/KILLER/TRICK2) und DR6 [32].

Die Bindung des Homotrimeren FasL an drei CD95 resultiert in einer Trimerisierung des Todesrezeptors, welcher auf zytoplasmatischer Seite eine Todesdomäne aufweist, über die weitere Signalproteine mit der gleichen Domäne binden können. Die Rekrutierung des Adapterproteins FADD, welches neben der Todesdomäne auch eine Todeseffektordomäne besitzt, bindet an die Todesdomäne des CD95-Rezeptors.

Bei Einleitung der Apoptose durch den TNFR-1-Todesrezeptor spielt neben dem FADD-Protein auch das TRADD-Protein eine Rolle. Die Bindung des Todesliganden TNF-alpha an den Todesrezeptor TNFR-1 führt zunächst zur Rekrutierung und Bindung des Adapterproteins TRADD und anschließend zur Rekrutierung und Bindung des FADDs an das TRADD-Protein.

An die Todeseffektordomäne des FADD bindet nun die Initiator-Caspase 8, welche dadurch aktiviert wird und durch proteolytische Spaltung weiterer Caspasen die Caspase-Kaskade, also den Effektorweg, auslöst und so zur Apoptose führt. Durch die Bindung der Caspase 8 wird der Apoptose-induzierende Signalkomplex (DISC: Death-Inducing-Signalling Complex) vervollständigt und weitere Apoptoseschritte werden eingeleitet.

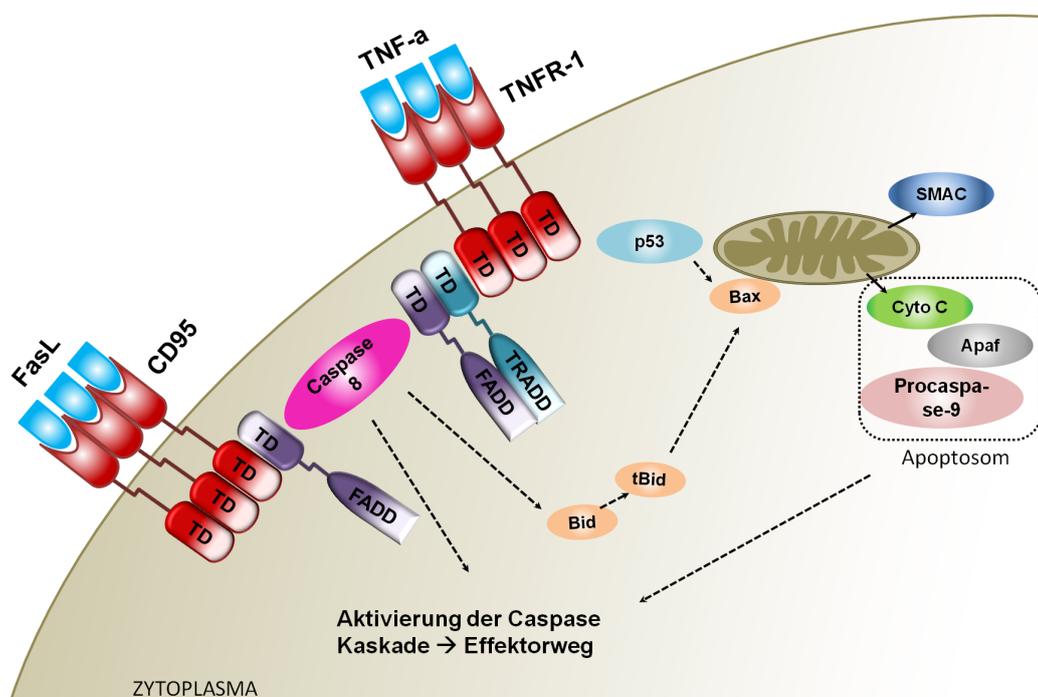


Abb. 1.2: Wege der Apoptose.

Dargestellt ist die Einleitung des extrinsischen Weges durch entweder den Todesrezeptor CD95 (Fas) oder TNFR1. Sie unterscheiden sich in der zusätzlichen Bindung des TRADD-Proteins vor der Rekrutierung des Adapterproteins FADD bei Einleitung durch TNFR1. Neben dem extrinsischen Weg ist auch der intrinsische Weg dargestellt, der durch p53-Aktivierung als Antwort auf DNA-Schäden ausgelöst wird. Beide Wege führen letztendlich zur Aktivierung der Caspase-Kaskade und zum Tod der Zelle.

Der intrinsische Weg wird durch innere Schäden, beispielsweise durch DNA-Schäden aufgrund fehlerhafter Replikation bei der Zellteilung, ausgelöst. Der intrinsische Weg wird durch die Proteine der Bcl-2 Familie reguliert. Dabei wird zwischen pro-

apoptotischen und anti-apoptotischen Bcl-2-Proteinen unterschieden, die die Apoptose entweder begünstigen oder inhibieren.

DNA-Schäden können zu einer Aktivierung des Tumorsuppressors p53 führen, was zu der Genexpression der pro-apoptotischen Bcl-2-Proteine wie Bax und Bak führt. Bax- und Bak-Proteine sind als Oligomere in der Lage Poren in der mitochondrialen Membran zu formen, die eine Freisetzung von Cytochrom C aus dem Mitochondrium ermöglichen. Neben Cytochrom C werden auch Faktoren wie Smac/DIABLO in das Zytosol ausgeschüttet. Das freigesetzte Cytochrom C bindet an Apaf-1 und führt zu einer Bindung des Komplexes an Procaspase 9 und bildet zusammen das Apoptosom. Die aktive Caspase 9 löst anschließend die Caspase-Kaskade aus.

Der extrinsische und intrinsische Signalweg sind über die Caspase 8 und dem Bid-Protein gekoppelt. Sollte aufgrund einer zu geringen Menge an Caspase 8 während des extrinsischen Weges die Apoptose nicht eingeleitet werden können, kann dieses Signal über den intrinsischen Weg verstärkt werden. Dabei spaltet die Caspase 8 das Bid-Protein zu tBid (truncated Bid), was zu einer Aktivierung der pro-apoptotischen Bcl-2-Proteine (z.B. Bax) führt und damit wieder zu einer Aktivierung der Caspase 9.

Bei dem Effektorweg werden die Effektor-Caspasen 3, 6 und 7 aktiviert, die anschließend zum programmierten Zelltod der Zelle führen. Die Caspasen aktivieren verschiedene Proteine, die z.B. für die Fragmentierung der DNA zuständig sind.

1.4.2 Angiogenese und die Rolle der vaskulären Wachstumsrezeptoren

Die Angiogenese umfasst den Prozess der Gewebsneubildung und die Entstehung und das Wachstum neuer Kapillargefäße während der Wundheilung, um das Gewebe mit lebenswichtigen Stoffen zu versorgen. Die vaskulären endothelialen Wachstumsfaktoren, kurz VEGF (Vascular Endothelial Growth Factor)-Proteine sind entscheidende Aktivatoren der Angiogenese, indem sie extrazellulär an die VEGF-Rezeptoren binden und zur Einleitung der Angiogenese führen. Dieser Prozess findet jedoch nicht nur während der Wundheilung statt, sondern auch während der Entwicklung eines malignen Tumors, der schon im frühen Stadium vermehrt VEGF sezerniert [33]. VEGF-Rezeptoren haben eine Tyrosinkinase-Domäne und sind die entscheidenden Mitspieler des Tumorwachstumsprozesses und der Versorgung maligner Tumore mit Sauerstoff und Nährstoffen und sind damit auch Zielproteine

bei der Entwicklung neuer pharmazeutischer Wirkstoffe für die Tumorthherapie. Bei dem VEGF-Rezeptor handelt es sich um ein Transmembranprotein mit einer ATP-Bindungsstelle im intrazellulären Bereich. Insgesamt gibt es drei verschiedene VEGF-Rezeptoren (VEGFR-1, VEGFR-2, VEGFR-3). VEGFR-1 ist bevorzugt in der hämatopoetischen Zellentwicklung, VEGFR-2 in der Regulation der vaskulären Endothelzellentwicklung und VEGFR-3 in der lymphatischen Endothelzellenentwicklung involviert [34]. Es wird angenommen, dass vor allem der VEGF-Rezeptor 2 bei der Angiogenese eine Rolle spielt [34].

Die Bindung eines Liganden führt zu einer Dimerisierung des Rezeptors, Bindung von ATP und dadurch zu einer Autophosphorylierung des Rezeptors und zur Auslösung verschiedener Signalkaskaden und Phosphorylierung weiterer wichtiger Signalproteine (siehe Abbildung 1.3).

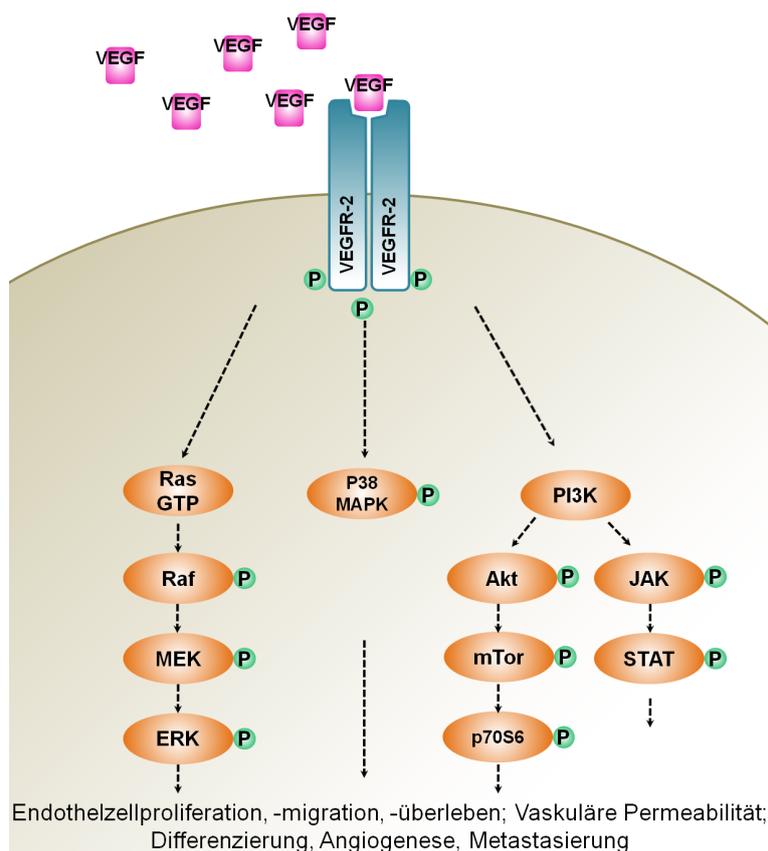


Abb. 1.3: Die Rolle des VEGF-Rezeptors 2.

Die Aktivierung des VEGF-Rezeptors 2 durch seinen Liganden führt zur Auslösung weiterer Signalkaskaden durch Aktivierung und Phosphorylierung von Signalproteinen. Das Ergebnis ist u.a. die Zell-Proliferation und Migration. Bei der Tumorentstehung wird im Rahmen der Angiogenese ein Kapillarsystem entwickelt, um den Tumor zu versorgen.

Die ausgelösten Signalkaskaden führen unter anderem zu Zellproliferation, Endothelzellenmigration und Angiogenese, also zum Blutgefäß-Wachstum. Da vor allem der VEGF-Rezeptor 2 als Schlüsselprotein in der Bildung der Kapillaren für die Versorgung des Tumors beteiligt ist, stellt er ein wichtiges Zielprotein für die Entwicklung neuer Therapien dar. Entscheidend ist hier die kompetitive Inhibition des Rezeptors durch Blockierung der intrazellulären ATP-Bindungsstelle des Rezeptors und damit die Verhinderung der Autophosphorylierung und Aktivierung der Signalkaskaden, die unter anderem zur Bildung Tumor versorgender Gefäße führen.

Die Blockierung der Angiogenese kann entweder durch VEGF- oder VEGFR-Antikörpern oder durch spezifische, niedermolekulare Inhibitoren der VEGFR Tyrosinkinase bewirkt werden. [35]. Anti-VEGF-Antikörper führen vergleichsweise nur zu leichten Nebenwirkungen [36] und befinden sich in klinischen Versuchen zur Behandlung solider Tumore [35]. Der monoklonale Antikörper Bevacizumab ist in Kombination mit Chemotherapeutika der erste Inhibitor dieser Klasse, der von der FDA zur Tumorthherapie zugelassen wurde [37, 38]. Bisher wurden einige Inhibitoren entwickelt, die sich zur Zeit in klinischen Phasen befinden [39]. Einer dieser VEGFR-2-Inhibitoren ist der von Novartis entwickelte Inhibitor Vatalanib (PTK787/ZK222584) [40]. Vatalanib adressiert die Tyrosinkinase von VEGFR-1,-2,-3, PDGFR und KIT [41] und ist zur Zeit in der klinischen Phase II/III für die Behandlung des kolorektalen Karzinoms [41] und des nicht-kleinzelligen Bronchiolalkarzinoms [42].

Im Laufe dieser Arbeit sollten bevorzugt VEGFR-2 Inhibitoren mit höherer Affinität identifiziert werden.

1.4.3 Der Notch-Signalweg und die Rolle der γ -Sekretase Proteine

Der Notch-Signalweg ist der im Laufe der Evolution am höchsten konservierte verglichen mit anderen Signalwegen und seine physiologische Rolle ist vielseitig. Der Notch-Signalweg ist u. a. verantwortlich für die Zelldifferenzierung, Aufrechterhaltung der Stammzellregulation, Zellentwicklungsregulation [43] und Zell-Zell-Kontaktinhibition [44]. Dabei kann die Aktivierung des Notch-Signalweges je nach Zelltyp und -umgebung zur Zellteilung oder zum Zelltod führen [45, 46]. Die Einlei-

tung des Signalweges erfolgt durch die extrazelluläre Bindung eines Liganden (z.B. Jagged) an den Notch-Rezeptor und die dadurch eingeleitete proteolytische Spaltung des Rezeptors durch die α - und γ -Sekretase (siehe Abbildung 1.4).

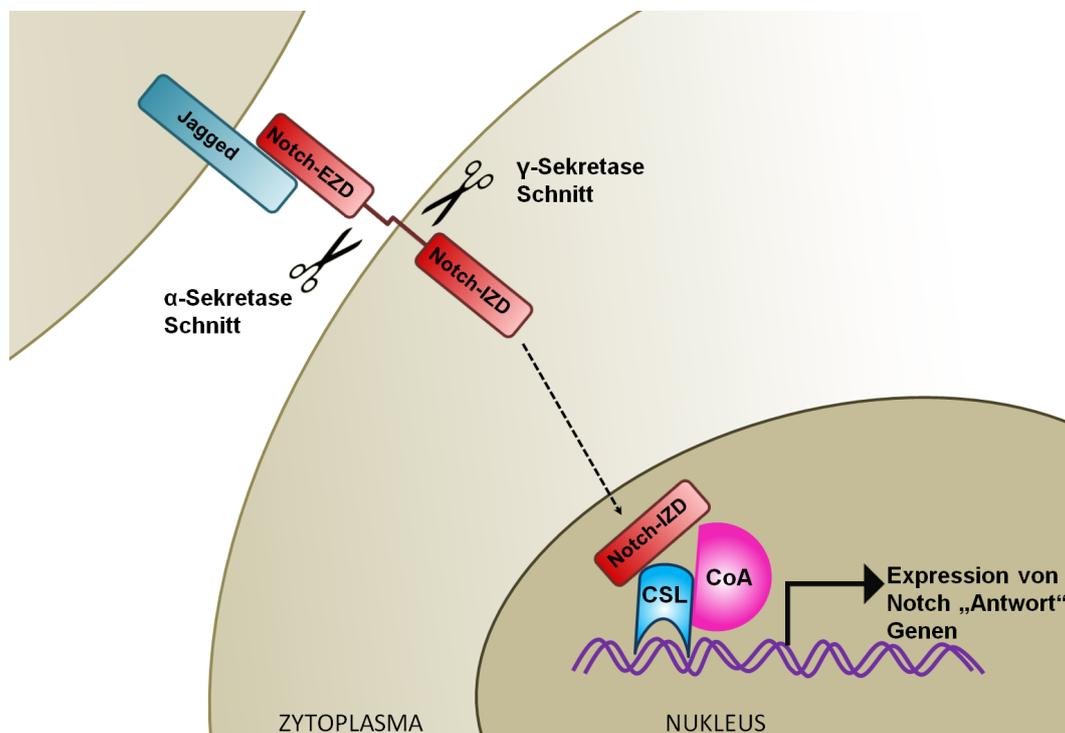


Abb. 1.4: Die Rolle der γ -Sekretase im Notch Signalweg.

Die Aktivierung des Notch-Signalweges erfolgt durch Bindung eines Liganden der Oberfläche einer Nachbarzelle an den Notch-Rezeptor und die anschließende proteolytische Spaltung durch die α - und die γ - Sekretase. Die intrazelluläre Domäne diffundiert in den Zellkern und agiert als Transkriptionsfaktor. Nach Shih et al. [43].

Der Notch-Rezeptor ist ein heterodimerer Transmembranrezeptor mit einer extra- und intrazellulären Domäne. Insgesamt gibt es vier Notch-Rezeptoren. Nach Abspaltung der intrazellulären Domäne (ID) des Notch-Rezeptors durch die γ -Sekretase und Diffusion der ID in den Zellkern interagiert die ID mit dem Regulatorprotein CSL. Dadurch werden weitere Koaktivatoren rekrutiert und ein aktiver Transkriptionsfaktor-Komplex geformt [43]. Die Folge ist die Transkription bzw. die Expression bestimmter Notch-Antwort-Zielproteine (z.B. Hes-Familie). Die γ -Sekretase ist eine Aspartat-Protease, da sie im aktiven Zentrum an Position 257 und 385 zwei Aspartat-Residuen besitzt. Es handelt sich um ein Transmembranprotein bestehend aus mindestens vier Untereinheiten. Die katalytische Untereinheit, Presenilin (PS1 oder PS2), besteht entsprechend vorangegangener Arbeiten entweder aus acht [47]

oder aus neun [48, 49] membrandurchspannenden Helices. Letzteres wird aufgrund der experimentellen Arbeit von Laudon et. al. [49] in dieser Arbeit angenommen (siehe Abbildung 1.5). Die Aspartatreste Asp257 und Asp385 befinden sich in den Helices sechs und acht. Des Weiteren formen das Nicastrin, der Stabilisator APH-1 und Pen-2 die weiteren Untereinheiten der γ -Sekretase.

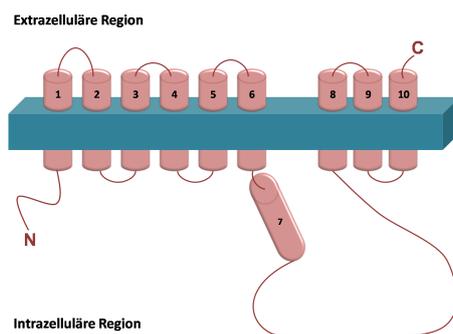


Abb. 1.5: Schematische Darstellung des Presenilins der γ -Sekretase.

Presenilin ist ein Transmembranprotein, welches aus zehn Helices besteht, wobei neun davon Membran-durchspannend sind. Die katalytischen Aspartatresiduen befinden sich in den Transmembrandomänen sechs und acht.

Eine Fehlregulation der Notch-Signaltransduktion und die konstitutive Aktivierung durch die γ -Sekretase bewirken die Ausbildung einiger Krankheitsbilder, was diesen Signalweg zu einem interessanten Forschungsgebiet der pharmazeutischen Wissenschaft macht. Zur Tumorentwicklung aufgrund des Notch-Rezeptors führen hauptsächlich genetische Veränderungen des Notch-Rezeptor Gens, die zu einer kontinuierlichen Aktivierung des Signalweges führen und zu einer erhöhten Konzentration der intrazellulären Domäne [43].

Vorangegangene experimentelle Arbeiten zeigten, dass bei verschiedenen Tumorerkrankungen eine Überexpression des Notch-Rezeptors vorhanden ist [50, 51, 52, 53]. Darüber hinaus wurde das onkogenetische Potenzial des Notch-Schnitts im Multiplen Myelom analysiert und die Erkenntnis gewonnen, dass die Behandlung der Zellen mit γ -Sekretase Inhibitoren, die spezifisch den Notch-Schnitt inhibieren, Apoptose in Multiplen Myelom Zellen induzieren [18, 54].

Desweiteren ist die γ -Sekretase an der Ausprägung der Niereninsuffizienz beteiligt [55]. Die Aktivierung des Notch-Signalweges trägt zu Nierenerkrankungen bei. Eine erhöhte Expression der intrazellulären Domäne des Notch-Rezeptors führt zur Apoptose-Induktion in Podozyten und damit langfristig zu Niereninsuffizienz [56].

Eine pharmakologische Inhibition mit spezifischen γ -Sekretase Inhibitoren verhindert das weitere Absterben der Podozyten [56].

Aktuell wird auch die Rolle der γ -Sekretase in der Entwicklung der Fibrose diskutiert, für die kein geeignetes Therapeutikum existiert, da Fibrose durch TGF β gefördert wird und aufgrund seines pleiotropen Effektes nicht als Zielprotein verwendbar ist [57]. Etwa 45% der Todesfälle in Entwicklungsländern sind auf fibrotische Erkrankungen zurückzuführen. Es wird vermutet, dass durch die Inhibition des Notch-Signalweges diese lebensbedrohliche Erkrankung behandelbar ist [58, 59].

Aufgrund der therapeutischen Relevanz des Notch-Signalweges in den oben genannten Krankheitsbildern stellt die γ -Sekretase ein Zielprotein bei der Entwicklung neuer Wirkstoffe dar. Bisher existieren jedoch in der medizinischen Anwendung kaum Protease-Inhibitoren, da diese die vorklinischen oder klinischen Phasen aufgrund der fehlenden Spezifität nicht überstanden haben oder sich noch in einer Entwicklungsphase befinden [60].

Die bisher entwickelten γ -Sekretase Inhibitoren führen oft zu sehr starken toxischen Nebenwirkungen, da sie ein breites Spektrum an γ -Sekretase Schnitten diverser Substrate blockieren.

Die γ -Sekretase spielte bisher in der Wirkstoffentwicklung vor allem als Target für die Therapie von Morbus Alzheimer eine Rolle [61]. Die Fragmentierung des Amyloid Precursor Proteins durch die β - und γ -Sekretase führt im humanen Gehirn zu einer vermehrten Bildung einer neurotoxischen Variante des β -Amyloids, das sich zu so genannten Plaques ansammelt und vereinigt [62]. Einige entwickelte Wirkstoffe, wie DAPT inhibieren die γ -Sekretase, jedoch führen sie zu starken Nebenwirkungen, die den Magen-Darm-Trakt, Thymus und die Milz betreffen, was auf eine Inhibition des Notch-Signalweges zurückzuführen ist [62]. Der Wirkstoff DAPT ist für den Notch-Signalweg um bis zu 1000-fach selektiver verglichen mit der Inhibition des Amyloid Precursor Schnitts [63].

Ziel dieser Arbeit war es Inhibitoren der γ -Sekretase zu identifizieren, die den Notch-Schnitt inhibieren. Die Blockierung des Amyloid Precursor Protein Schnitts sollte möglichst gering sein. Auch eine zu starke Inhibition des Notch-Signalweges, wie sie das DAPT Peptidomimetikum ausübt, sollte aufgrund der essentiellen Rolle des

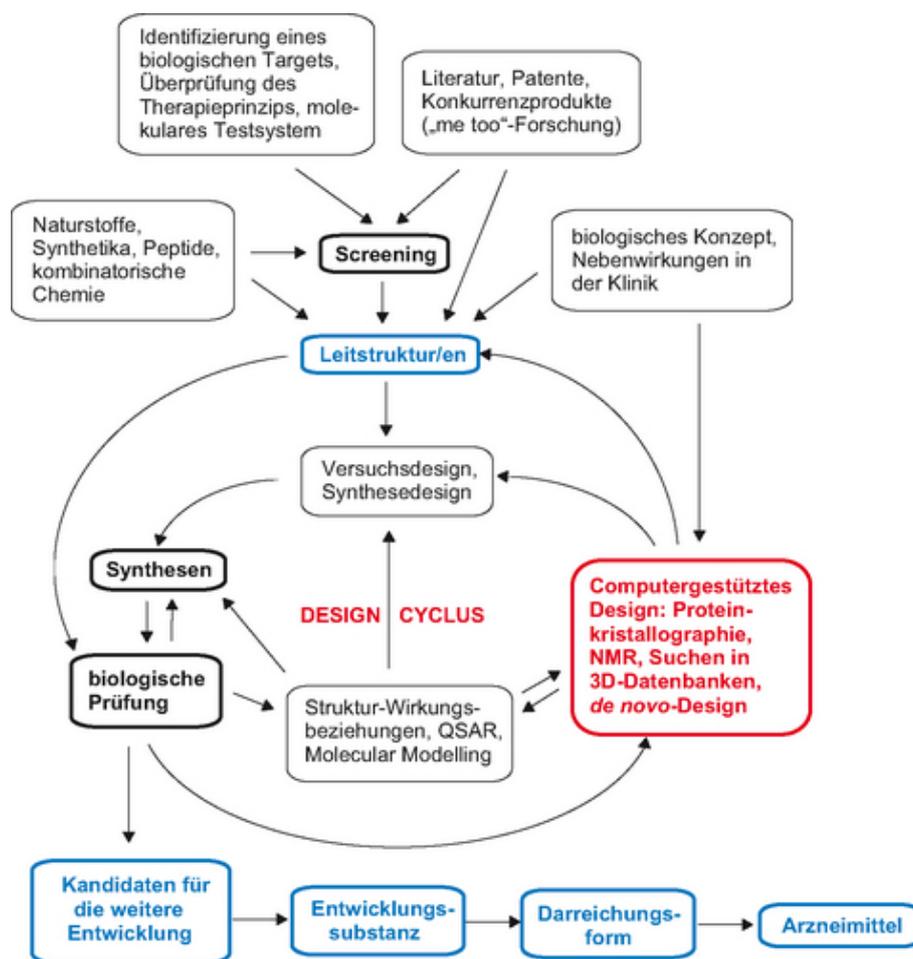
Notch-Signalweges in der Zelldifferenzierung vermieden werden. Eine vollständige Inhibition des Notch-Signalweges führt zu drastischen Nebenwirkungen.

1.5 Wirkstoffentwicklung

1.5.1 Ziele und Wege der Wirkstoffentwicklung

Im Jahr 2010 waren in Deutschland rund 8.500 verschiedene Präparate in der Roten Liste, dem deutschen Arzneimittelverzeichnis, eingetragen, wobei ca. 80% davon synthetisch hergestellte Substanzen ausmachen [64]. Bei der Wirkstoffentwicklung handelt es sich um einen langwierigen und teuren Prozess bis aus einer Leitstruktur ein Arzneistoff wird. Dabei muss der Wirkstoff mehrere klinische Phasen durchlaufen. Die Anforderungen an ein Arzneistoff sind hoch. Der Wirkstoff soll möglichst die Krankheitsursachen bekämpfen, die Symptome lindern und dabei die Nebenwirkungen möglichst gering halten. Das grundlegende Ziel der Arzneistoffentwicklung ist die Identifizierung der Wirkstoffe, die spezifisch an bestimmte Zielproteine binden, diese inhibieren oder aktivieren und damit eine Signalkaskade unterbrechen oder auslösen und für den menschlichen Organismus verträglich sind.

Die Analyse und das Verständnis der verantwortlichen Signalwege und die Identifizierung eines Zielproteins, das entscheidende Bedeutung für die Ausprägung eines Krankheitsbildes besitzt, stellen den ersten Schritt bei der Entwicklung neuer Therapeutika dar. Darauf folgt die Suche nach Wirkstoffen, die als Leitstrukturen (Substanzen mit bereits bekannter Aktivierungs- bzw. Inhibitionswirkung) fungieren und zu potenziellen Arzneistoffen optimiert werden können. Abbildung 1.6 zeigt schematisch den Entwicklungsweg einer Leitstruktur zum Arzneimittel.



Aus: Klebe, *Wirkstoffdesign*, 2. Aufl. © Spektrum Akademischer Verlag GmbH, 2009

Abb. 1.6: Der Weg zum Arzneimittel. Aus Klebe [23].

Bei der Suche nach Leitstrukturen werden verschiedene Kleinstrukturen wie Peptide, Naturstoffe, synthetische Substanzen etc. auf speziell entwickelte *in vitro* Testsysteme auf Inhibitions- bzw. Aktivierungswirkung überprüft [23]. Eine in der pharmazeutischen Industrie sehr häufig verwendete Methode ist das „High-Throughput-Screening“ (HTS) [65]. Hierbei werden tausende Kleinstrukturen gleichzeitig auf einem Testsystem, das sich in mehreren kleinen Reaktionsgefäßen (Wells) einer Platte befindet, getestet. Um die Affinitäten der Substanzen auf ein bestimmtes Protein oder eine Proteingruppe ermitteln zu können, finden in der pharmazeutischen Wirkstoffentwicklung neben HTS auch Methoden wie „Tethering“ [66, 67, 68] und Surface Plasmon Resonanz (SPR)-Spektroskopie [69, 70, 71] Anwendung. Vereinzelt werden die Affinitäten auch durch 3D-Struktur-Aufklärung des Komplexes mittels NMR-Spektroskopie (Nuclear Magnetic Resonance) [72, 73] und X-Ray

(Röntgenstrahlung)-Methoden [74] bestimmt.

Die identifizierten Leitstrukturen können anschließend in einem Design-Zyklus aus computergestützten Verfahren und biologischer *in vitro* Prüfung optimiert werden. Dieser Zyklus wird in den meisten Fällen mehrmals durchlaufen. Dabei werden u.a. die chemischen Eigenschaften von Kleinstrukturen, die für die biologische Aktivität verantwortlich sind, bestimmt (quantitative Strukturaktivitätsbeziehung (z.B. QSAR)), um optimierte Wirkstoffkandidaten zu identifizieren. Mit ADMETox- Analysen und weiteren empirischen Regeln wird außerdem versucht die Spezifität und die orale Bioverfügbarkeit zu erhöhen und die Toxizität zu minimieren. Anschließend werden die Strukturen wieder experimentell validiert und getestet. Das Resultat sind Arzneistoffkandidaten, die in die Weiterentwicklung gehen. Es folgen *in vivo*-Experimente und die klinischen Phasen.

In der strukturellen Bioinformatik und im Verlauf dieser Arbeit werden hauptsächlich computergestützte Verfahren angewendet, um potenzielle Wirkstoffkandidaten zu identifizieren und sie im Labor zu validieren. Hier wurden als Leitstrukturen bekannte Inhibitoren bzw. Aktivatoren in PubMed [75] identifiziert und mit Hilfe verschiedener Screening-Verfahren optimiert (siehe Kapitel 2.2.2). Des Weiteren wurden im Laufe dieser Arbeit nur Inhibitoren und keine Aktivatoren für spezifische Zielproteine gesucht. Insgesamt sind zwei Arten der Enzyminhibition möglich: die reversible Enzyminhibition (kompetitive, nicht-kompetitive, unkompetitive, partiell-kompetitive Hemmung etc.) und die irreversible Enzyminhibition. In dieser Arbeit wurden ausschließlich reversible kompetitiv inhibierende Substanzen für bestimmte Zielproteine gesucht, was bedeutet, dass die entwickelten potenziellen Wirkstoffkandidaten als Antagonisten mit höherer Affinität fungieren und mit den natürlichen Agonisten um die Bindungsstelle konkurrieren.

Es gibt vier verschiedene Ausgangspunkte bei der Entwicklung der Wirkstoffkandidaten in der strukturellen Bioinformatik. Grob kann zwischen Liganden-basiertem Wirkstoffdesign und Struktur-basiertem Wirkstoffdesign unterschieden werden, je nachdem ob eine Proteinstruktur vorhanden ist. Im ungünstigsten Fall sind weder 3D-Strukturen des Zielproteins in der Protein Data Bank (PDB) [76] noch schon bekannte Liganden, die als Leitstruktur dienen können, bekannt. Auch liegen keine Strukturen mit ähnlichen Eigenschaften vor, die als Template für Homologie-

Modellierungsverfahren genutzt werden können (siehe Kapitel 2.2.1). In diesem Fall können nur pharmazeutische, Liganden-basierte Methoden wie High-Throughput Assays und Kombinatorische Chemie [77, 78] angewendet werden. Ist keine 3D-Struktur vorhanden, jedoch Leitstrukturen, so können aus diesen Pharmakophore erstellt und *in silico* Screening Methoden angewendet werden, um weitere Leitstrukturen zu identifizieren und sie ggf. zu optimieren. Ist eine 3D-Proteinstruktur vorhanden, jedoch keine bekannten Liganden, so finden Struktur-basierte Methoden wie DIP (Datenbank of Interfaces in Proteins) [79] und *De Novo* Design [80, 81, 82] Anwendung, um eine Leitstruktur zu identifizieren, die später zu einem potenziellen Arzneistoff optimiert werden kann. Im günstigsten Fall sind sowohl mindestens eine 3D-Struktur des Zielproteins als auch schon bekannte Liganden vorhanden. Hier wird mit der Optimierung der Leitstruktur begonnen. Während des Struktur-basierten Wirkstoffdesigns können für die *in silico*-Optimierung der Leitstruktur zusätzlich Protein-Liganden-Docking-Verfahren [83, 84] genutzt werden, um die Affinitäten vorherzusagen und zu erklären.

Eine Methode, die eine immer größere Bedeutung in der Wirkstoffentwicklung erhält ist das Fragment-basierte Wirkstoffdesign [85]. Hierbei wird versucht durch Neukombination von Fragmenten eine optimal wirkende Substanz zu generieren. Die pharmazeutische Industrie besitzt große Wirkstoffbibliotheken für High-Throughput-Screening Assays, wobei eine Vielzahl an Kleinstrukturen nur mäßige Wirkungen auf das Testsystem gezeigt haben. Diese mittelmäßigen Affinitäts-Treffer können fragmentiert und die Fragmente anschließend neu kombiniert werden, um so Kandidaten mit höherer Affinität zu erhalten. Hier wird wieder zwischen theoretischen und experimentellen Methoden unterschieden. In der Bioinformatik werden einzelne Fragmente mit Docking Methoden in die Bindungstasche der Proteinstruktur positioniert und anschließend mit Hilfe von Linkern zusammengefügt. In der pharmazeutischen Industrie werden hier vor allem experimentelle Methoden wie das HTS angewendet und tausende Fragmente in einem Testsystem bzgl. ihrer Bindung überprüft [86]. Die Positionen der Fragmente lassen sich mit NMR-Spektroskopie ermitteln und anschließend zu größeren Liganden kombinieren (durch Linking, Growing oder Merging) [85]. Ein großer Vorteil des Fragment-basierten Wirkstoffdesigns ist die Verwendung einer im Vergleich mit anderen Methoden deutlich kleineren zu testenden Substanzbibliothek, um den chemischen Raum abzutasten. Beispielsweise reichen 1.000 Fragmente aus, um 190 Millionen Kleinstrukturen zu generieren.

Im Rahmen dieser Arbeit sollte eine für Fragmentanalysen geeignete Fragmentbibliothek generiert werden. Durch eine zusätzlich implementierte Assemblierungsmethode für Fragmente sollten Wirkstoffe durch chemisch synthetisierbare Verbindungen generiert und damit ein erster Schritt zur Automatisierung des Fragment-basierten Wirkstoffdesigns vollzogen werden.

1.5.2 Eigenschaften oral verfügbarer Arzneistoffe

Die *in vitro* Inhibition eines Proteins durch einen Liganden ist nur eine Voraussetzung dafür, dass aus einem potenziellen Kandidaten ein Arzneistoff wird. Es ist nicht selten, dass Inhibitoren für bestimmte Proteine mit z.B. *in silico*-Verfahren identifiziert werden, aber dann doch aufgrund verschiedener Eigenschaften nicht als Arzneistoff einsetzbar sind und in einer klinischen Phase ausselektiert werden müssen, weil sie z.B. im humanen System toxisch wirken. Nachdem die Inhibition eines Proteins durch einen neuen Liganden festgestellt wurde, muss gewährleistet werden, dass der Inhibitor bei oraler Anwendung an seinen Wirkort gelangt. Dafür muss der Wirkstoff u. a. persistent sein und dem sauren pH-Wert des Magens standhalten. Um schon vorab während des *in silico* Screenings die Kandidaten auszusondern, die nicht als Wirkstoff in Frage kommen, hilft die Anwendung empirischer Regeln.

1.5.2.1 Lipinskis 5er Regel

Eine oft angewandte Methode, um die orale Bioverfügbarkeit abzuschätzen ist die Berechnung der sogenannten Rule of Five [87] oder auch 5er Regel, die 1997 von Christopher Lipinski aufgestellt wurde. Die Regel beschreibt vier Bedingungen, die die orale Bioverfügbarkeit von Wirkstoffen bewerten bzw. abschätzen. Der Name „5er Regel“ bezieht sich auf die Zahlenwerte, die entweder fünf oder ein Vielfaches von fünf entsprechen. Ein potenzieller Arzneistoff besitzt demnach eine gute orale Bioverfügbarkeit, wenn nur maximal eine der folgenden Regeln verletzt wird:

1. Das Molekulargewicht der Substanz darf nicht mehr als 500 g/mol betragen
2. Die Substanz darf nicht mehr als 5 Wasserstoffbrücken-Donatoren besitzen
3. Die Substanz darf nicht mehr als 10 Wasserstoffbrücken-Akzeptoren besitzen
4. Der Verteilungskoeffizient darf nicht größer als 5 sein

Der Verteilungskoeffizient $\log P$ gibt hier an wie sich die Substanz zwischen Oktanol und Wasser, zwei unmischbaren Substanzen, verteilt und beschreibt damit das Verhältnis der Hydrophobie und Hydrophilie einer Substanz. Eine gute Wasserlöslichkeit ist für den Transport des Wirkstoffs durch die Blutbahn essentiell, jedoch hindert eine verminderte Fettlöslichkeit die Passage des Wirkstoffs durch die Membran. Daher ist das Verhältnis beider Löslichkeiten von entscheidender Bedeutung für das Ankommen des Wirkstoffs am Zielort. Auch die Größe und die Anzahl der Wasserstoffbrücken-Donatoren und -Akzeptoren können die Fähigkeit des Wirkstoffs die Membran zu passieren beeinflussen.

1.5.2.2 ADMETox

Eine weitere Möglichkeit zur Abschätzung der oralen Bioverfügbarkeit von Wirkstoffen ist die Anwendung der ADMETox-Analysen. ADMET [88] oder ADME-Tox ist in der pharmazeutischen Wissenschaft eine Abkürzung für die Aufnahme (**A**bsorption) in die Blutbahn, Verteilung (**D**istribution) im Organismus, Verstoffwechslung (**M**etabolismus), Ausscheidung (**E**xkretion) und **T**oxizität eines Wirkstoffs. Dabei sollen die ADME-Eigenschaften hoch und die Toxizität möglichst gering sein.

Die Absorption (Resorption) eines Wirkstoffs beschreibt dessen Fähigkeit vom menschlichen Körper aufgenommen zu werden und in den Blutstrom zu gelangen, was z.B. durch Unlöslichkeit des Wirkstoffs verhindert werden kann. Nach Eintritt der Substanz in die Blutbahn, beginnt dessen Verteilung (Distribution) und damit sein Transport in die verschiedenen Gewebe. Dabei kann die Verteilung durch Einflüsse wie pH-Wert, geringe Bindung der Substanz an Transporter, den sogenannten „Carrier“-Proteinen, und Löslichkeit des Wirkstoffs gestört werden.

Auch die Fähigkeit der Verstoffwechslung (Metabolismus), also des Abbaus der Wirkstoffe spielt eine wichtige Rolle bei der Entwicklung von potenziellen Arzneimitteln. Der Wirkstoff ist im Organismus verschiedenen Abbauprozessen in Leber, Schleimhäuten, Magen-Darm und Plasma ausgesetzt. Dabei ist es zum einen wichtig, dass der Wirkstoff abgebaut bzw. ausgeschieden werden kann und eine Akkumulation des Wirkstoffs im Organismus und die daraus resultierende toxische Wirkung und Nebenwirkung möglichst verhindert werden. Zum anderen kann ein zu schneller Abbau des Wirkstoffs durch Cytochrome P450 zu Wirkungsverlust führen. Die Ausscheidung (Exkretion) der, durch den Abbau entstandenen, Metabolite des Wirkstoffs geschieht überwiegend renal. Dieser kann jedoch z.B. auch über Haut,

Schleimhäute und Lunge ausgeschieden werden.

Ein Wirkstoff kann in vielerlei Hinsicht toxisch wirken. Neben der Hepatotoxizität, die durch Akkumulation eines Wirkstoffs in der Leber aufgrund des verhinderten Wirkstoff-Abbaus durch Cytochrome entstehen kann, kann ein Wirkstoff unter anderem auch phototoxisch oder zytotoxisch wirken.

Auch die Passage der Blut-Hirn-Schranke ist während der Wirkstoffentwicklung von entscheidender Bedeutung. In den meisten Fällen ist eine Durchdringung des Wirkstoffs der Blut-Hirn-Schranke unerwünscht, da die entstehenden Nebenwirkungen nicht im Zentralnervensystem (ZNS) auftreten sollen und somit zum Ausschluss der Wirkstoffe während der klinischen Phasen führen kann. Jedoch ist das Passieren der Blut-Hirn-Schranke bei der Behandlung neurodegenerativer Erkrankungen wie Alzheimer oder Depressionen, aber auch von Gehirntumoren durchaus wünschenswert.

Die frühzeitige Anwendung geeigneter ADMETox-Analyse-Verfahren während der Wirkstoffentwicklung hilft bei der Abschätzung, Bewertung und Selektion der Arzneistoffkandidaten. Dabei berechnen die meisten Verfahren die Lipinski- und „Polar Surface Area“ (PSA)-Eigenschaften [89]. PSA ist die Oberflächensumme über alle polaren Atome und ist entscheidend für die Membrangängigkeit einer Substanz. Wirkstoffe mit PSA-Werten größer als 140 \AA^2 passieren Zellmembranen in der Regel nur sehr schwer. Des Weiteren bieten die Programme verschiedene Testmodelle für die ADMETox-Eigenschaften an, denen Kleinstrukturen zugeführt und Vorhersagen getätigt werden können (siehe Kapitel 2.2.3).

Für die Abschätzung der Zytotoxizität entwickelte das NCI das *in vitro* Zelllinien Screening Projekt (IVCLSP), das im Folgenden näher erläutert werden soll.

1.5.3 NCI 60 Screening Projekt

Das *National Cancer Institute* startete 1990 das *in vitro* Zelllinien Screening Projekt (IVCLSP) [21], in dem Sensitivitätsprofile von 60 onkologischen Zelllinien gemessen wurden. Aktuell umfasst dieses Projekt GI50- (50 prozentige Wachstumshemmung), TGI- (totale Wachstumshemmung) und LC50-Werte (Absterben der Hälfte der Tumorzellen) für über 40.000 verschiedene Substanzen. Neben den Inhibitions-werten der verschiedenen Kleinstrukturen gegenüber den 60 Zelllinien sind auch Strukturinformationen, Expressionsdaten und Targetausstattung der Zelllinien in der NCI-Datenbank abgelegt. Die 60 verschiedenen Zelllinien beinhalten verschiede-

ne Leukämie-, Melanom-, Lungentumor-, Darmtumor-, Gehirntumor-, Eierstocktumor -, Brusttumor-, Prostata- und Nierentumor-Zelllinien. Die Kombination der Chemosensitivität der verschiedenen Zelllinien und die Analyse der Targetausstattung könnten neue und bessere Therapiemöglichkeiten offenbaren. Dies sollte auch im Rahmen dieser Arbeit realisiert werden.

1.6 Genetische und evolutionäre Aspekte

Mögliche Gründe für die Entstehung von malignen Tumorerkrankungen können in Umwelteinflüssen, Ernährungsweisen und anderen ökologischen Ursachen gefunden werden. Signifikante Veränderungen in der unterschiedlichen Ausprägung der Karzinom-Entstehung in Brust, Eierstock, Lunge, Magen sowie Darm- und Mastdarm wie sie zwischen Schimpanse (*Pan Troglodytes*) und Mensch (*Homo Sapiens*) beobachtet wurden [2], lassen sich nur bedingt mit den oben genannten Argumenten begründen. Dies lenkt den Blick auf intrinsische Faktoren wie Empfindlichkeit und Toleranz, die von genetischen Faktoren abhängen.

Genetische Variabilität führt zu einer Vielfalt der Tumorzinzidenz in verschiedenen Populationen. Japaner und nicht-hispanische Weiße besitzen das höchste Tumorrisiko, während die männlichen Bewohner Alaskas das niedrigste Tumorrisiko [90] zeigen. Prostata- [91] und Brust- [92] Karzinome treten sehr selten in der Bevölkerung der Eskimos auf. Allerdings zeigen Indianer und Aleuten ein signifikant erhöhtes Risiko an Prostatakarzinomen zu erkranken, verglichen mit den männlichen Einwohnern Alaskas [91]. Bekannte genetische Anomalien, welche für verschiedene Tumorerkrankungen verantwortlich sind, zeigt das BRCA1-Gen, welches Mutationen bei Patientinnen mit Brustdrüsenkarzinom (Mammakarzinom) aufweist. Das translatierte Protein wirkt als Tumorsuppressor und ist an Prozessen der DNA-Reparatur beteiligt. In den meisten Fällen führt ein DNA-Schaden entweder zur Auslösung der DNA-Reparatur-Mechanismen oder der Apoptose. Jedoch begünstigt eine Fehlregulation dieser Programme die Ausprägung maligner Tumorerkrankungen. Es wird allgemein angenommen, dass genetische Varianten zur Störung der Apoptose- und/oder DNA-Reparaturmechanismen und damit bevorzugt zur Malignom-Entstehung führen [93, 94]. Es ist bekannt, dass eine immer wiederkehrende Mutation des BRCA1-Gens bei Patientinnen mit Brustdrüsentumoren der chinesischen Bevölkerung auftritt [95]. Es existiert eine kurze Variante des Philadelphia-Chromosoms, die zur Transkription eines neuen Gens namens BCR-ABL führt und an der Entstehung

der Leukämie beteiligt ist [96].

Datenbanken wie OMIM speichern eine Vielzahl an SNPs (Single Nucleotide Polymorphismen) und Gene, die für die Ausprägung der malignen Tumoren, bösartigen Lymphomen und Leukämie verantwortlich sind [97].

Der Schimpanse (*Pan Troglodytes*) stellt mit einer Sequenzidentität von mehr als 98% den evolutionär nächsten Verwandten des Menschen (*Homo Sapiens*) dar [8, 9]. Da der Schimpanse dennoch im Vergleich zum Menschen kaum an Tumorerkrankungen leidet, liegt die Vermutung nahe, dass sich der Schimpanse im Laufe der Evolution eine Genausstattung angeeignet hat, welche die Tumorentstehung weitestgehend verhindert oder sich während der Evolution des Menschen die Genausstattung tumorbegünstigend entwickelt hat. Besonders Prostata [4], Brust- und Bronchialkarzinome und spontane Entartungen [3] treten selten in Schimpansen Populationen auf [5, 6, 7]. Die Ursache für die Beobachtung könnte in den Tumor-relevanten Signalwegen wie Apoptose oder DNA-Reparatur und deren Komponenten zu finden sein. Eine Analyse nach der Sequenzierung des Genoms zeigt, dass zwischen Mensch und Schimpanse etwa 35 Millionen einzelne Nukleotidveränderungen und etwa 5 Millionen Insertionen/Deletionen (Indels) auftraten [98]. Frühere Arbeiten haben gezeigt, dass Unterschiede auf Nukleotid- und Aminosäure-Ebene zwischen Schimpansen- und humanen Proteinen vorhanden sind, die bei der Malignom-Entstehung beteiligt sind [6]. Diese Ergebnisse unterstützen die Hypothese, dass die geringe Inzidenz maligner Tumorerkrankungen beim Schimpansen genetisch bedingt ist. Puente et al. untersuchte 333 onkologisch relevante Gene des Menschen und Schimpansen und identifizierte etwa 1.500 Unterschiede in der Aminosäure-Sequenz der Proteine [6]. Beispielsweise entdeckten sie an Position 73 in der kodierenden Sequenz des Tumorsuppressors p53 ein Arginin, das nur im humanen Genom vorzufinden ist. Andere Primaten wie Schimpanse, Gorillas und Madrills besitzen ein Prolin an dieser Stelle [9]. Darüber hinaus wurden in der Arbeit von Nielsen et al. rund 13.700 annotierte Gene des Schimpansen und Menschen analysiert und ihre positive Selektion berechnet [99]. In diesem Vergleich wurde entdeckt, dass mehrere onkologisch relevante Proteine, die an Apoptose, Tumorsuppression und Kontrolle des Zellzyklus beteiligt sind, positiv selektiert sind. Dies wird durch die Ergebnisse des Nature-Konsortiums [98], das die positive Selektion der Protein-Gruppen in der GeneOntology [100] analysierte, unterstützt. Hier sind die Apoptose und die Proteine, die den Zellzyklus kontrollieren ebenfalls positiv selektiert [98].

Um mögliche Ursachen für die unterschiedliche Tumorzinzidenz zu finden und vielmehr neue Möglichkeiten der Tumorthherapie bzw. neue mögliche Zielproteine abzuleiten, sollte in dieser Arbeit eine größere Datenmenge von etwa 500 Proteinen, die entweder in der Apoptose, DNA-Reparatur und im Zellzyklus eine Rolle spielen, betrachtet werden. Im Gegensatz zu früheren Studien sollte eine Analyse dieser Proteine auf verschiedenen Ebenen durchgeführt werden: chromosomale Organisation, Genstruktur, Modifikationen sowie strukturelle und funktionelle Veränderungen.

1.7 Datenintegration

In der onkologischen Forschung werden regelmäßig neue Erkenntnisse gewonnen, involvierte Signalwege aufgeklärt sowie neue Zielgene, Zielproteine und Wirkstoffe identifiziert. Der Großteil dieser Informationen und Daten kann in der Literatur-Datenbank PubMed, die aktuell rund 19 Millionen Publikationen speichert, gefunden werden. Jedoch stellt die Extraktion dieser Informationen und Zusammenhänge, um einen Überblick über den aktuellen Forschungsstand zu erhalten, eine Herausforderung dar. Die Notwendigkeit und vor allem die Nachfrage nach Datenbanken für die wissenschaftliche Gemeinschaft führte zu einem Zuwachs solcher Ressourcen, angefangen bei der UniProt-Datenbank [101], die aktuell über zehn Millionen Proteineinträge mit Informationen und Verlinkungen zu anderen Datenbanken wie KEGG, PDB und GeneOntology enthält. Die PDB enthält aktuell über 70.000 Proteinstrukturen, teilweise auch im Komplex mit Liganden [76]. Die KEGG-Datenbank ist eine der wichtigsten Datenquellen für Signalwege [102]. Neben Protein- und Signalwegdatenbanken gibt es auch eine große Anzahl an Wirkstoffdatenbanken, wie die DrugBank [11] oder die PubChem [103]. Alle diese Datenbanken sind häufig besuchte Seiten im Internet und als Hilfsmittel in der Forschung nicht mehr wegzudenken. Dennoch sind Verknüpfungen zwischen diesen Bereichen kaum vorhanden. Die DrugBank, TTD und CTD beinhalten mittlerweile viele Wirkstoff-Protein-Verbindungen [104]. Aber eine genaue Analyse zeigte, dass die Überlappung dieser Datenbanken marginal ist. Es fehlen oft die zugehörigen PubMed-Referenzen, Signalweginformationen und experimentelle Informationen, wie Expressionsdaten und auch Informationen zur Wirkung von Substanzen auf Zelllinien. Die CTP enthält neben direkten Wirkstoff-Protein-Interaktionen auch indirekte Interaktionen, wie beispielsweise die Beeinflussung von Wirkstoffen auf die Expression eines Proteins. Interaktionen und direkte Interaktionen bezeichnen im Laufe dieser Arbeit die

physikalische Wechselwirkung von Ligand und Protein. Bei indirekten Interaktionen gehen Protein und Ligand keine physikalische Wechselwirkung ein, können sich aber dennoch beeinflussen. Pharmakogene können aus der PharmGKB [105, 104] gewonnen werden. Hier werden u.a. Gene, die im Zusammenhang mit der Dynamik- und Kinetik-Veränderung von Wirkstoffen stehen, gespeichert. Die DTP-Ressource des *National Cancer Institutes* beinhaltet die Chemosensitivität von über 40.000 Substanzen auf 60 Zelllinien. Eine umfassende Integration dieser Datenquellen und zusätzlich implementierte Analyse-, Visualisierungs- und Vorhersageprogramme sind ein wichtiger Schritt für das Verständnis und die Analyse in der Onkologie.

Die Umsetzung einer solchen Integration war eine Herausforderung dieser Arbeit. Es sollte die Integration von chemischen, biologischen sowie medizinischen Informationen bezüglich onkologisch relevanten Zielproteinen und auf diese wirkenden Kleinstrukturen erfolgen. Um die große Datenmenge an onkologisch relevanten Targets und deren Expressionen, Mutationen, Signalwegen und die Effekte von onkologisch relevanten Kleinstrukturen auf onkologische Zelllinien organisieren zu können, sollte eine Datenstruktur erstellt werden, welche die verschiedenen genannten Themengebiete kombiniert und Analysemethoden bereitstellt. Während der Datenintegration sollten bevorzugt direkte Interaktionen betrachtet werden. Indirekte Interaktionen dienten nur zur Vervollständigung des Datensatzes und wurden als solche markiert. Desweiteren sollte die Kombination bereits vorhandener Expressionsdaten und Wirkprofile von Tumor-Zelllinien mit Daten bezüglich der genetischen Variabilität von onkologisch relevanten Genen von hochgeladenen Daten ermöglicht werden. Durch den Vergleich von tumorspezifischen Expressions- und Mutationsmustern mit denen der NCI-Zelllinien wurde die Erstellung patientenspezifisch optimierter Therapievor schläge angestrebt. Individuelle Patientendaten zu Target-Ausstattung und Mutationen des zu behandelnden Tumorgewebes sollten hochladbar sein, mit gespeicherten Daten verglichen und die bestmögliche Therapie in Form von Wirkstoffen ermittelt werden. Durch den fortwährenden Zuwachs an individualisierten Daten bezüglich der Target-Ausstattung, ihrer Expressionen und Mutationen können Ressourcen wie diese einen ersten Schritt in Richtung individualisierte Therapie darstellen.

Kapitel 2

Material und Methoden

2.1 Analyse von Zielproteinen, Wirkstoffen und ihren Interaktionen

Der Methodenteil zur Analyse von Zielproteinen, Wirkstoffen und deren Interaktionen besteht aus zwei Hauptteilen. Der erste Abschnitt beschreibt das Material und die Methoden, die für die Analyse des Unterschieds zwischen Mensch und Schimpanse auf Genom- und Proteinebene verwendet wurden, während sich der zweite Abschnitt mit den Materialien und Methoden für die Analyse von Kleinstrukturen in Hinblick auf ihre Fragmente und deren Bindungstaschenpräferenzen bezieht.

2.1.1 Genom- und Proteinstrukturanalyse onkologisch relevanter Proteine von Mensch und Schimpanse

Genomanalyse

Der Begriff „Genomanalyse“ bezeichnet hier Methoden zur Analyse der Erbanlagen einer Spezies. Durch die Sequenzierung des Genoms des Menschen (*Homo Sapiens*) und seines nächsten Verwandten des Schimpansen (*Pan Troglodytes*) sind grundlegende Daten bereitgestellt, deren Analyse einen Beitrag zum Verständnis der menschlichen Evolution leisten kann. Es gibt mehr als 40 Millionen genetische Unterschiede zwischen Mensch und Schimpanse, die meisten davon sind kleine Nukleotid-Substitutionen. Vom medizinischen Gesichtspunkt aus, ist die Ursache der unterschiedlichen Ausprägung bestimmter Krankheiten in verschiedenen Spezies für die Entdeckung eventueller neuer Zielproteine und Entwicklung neu-

er Therapien interessant und wichtig. Die Motivation der Genomanalyse und des Vergleichs von Schimpansen- und Humangenom basiert auf der Beobachtung, dass Schimpansen im Vergleich zu Menschen weniger häufig an onkologischen Krankheiten leiden. Um mögliche molekulare Ursachen dafür zu finden, wurden in dieser Arbeit die Unterschiede zwischen Mensch und Schimpanse in Bezug auf Apoptose und DNA-Reparatur auf verschiedenen Ebenen analysiert. Für die Genomanalyse wurden Gene betrachtet, die für die Expression von Proteinen verantwortlich sind, die im Zusammenhang mit der Tumorentstehung stehen und in diesen Signalwegen eine Rolle spielen. Insgesamt wurden 493 Proteine und deren kodierenden Gene selektiert, die in der DNA-Reparatur, der Apoptose und dem Zellzyklus eine Rolle spielen und für die verschiedenen Tumorinzidenzen verantwortlich sein könnten. Die Proteine wurden mit Hilfe der GeneOntology- [106, 98] und der KEGG-Datenbank identifiziert. Bei der Vervollständigung des Datensatzes half Prof. Dr. Peter Daniels Arbeitsgruppe. Die Gensequenzen stammten aus der Ensembl 42-44 [107].

Folgende Punkte wurden auf Genebene untersucht:

1. Chromosomen, Genorganisation und Genstruktur
2. Pseudogenisierung
3. SNPs, Insertionen und Deletionen

Chromosomen-Organisation und Genstruktur:

Die Darstellung der Chromosomen-Organisation und der Vergleich der Genlokalisierungen auf humanen und Schimpansen Chromosomen wurde mit dem Online-Programm Autograph [108] realisiert. Die Verteilung der einzelnen kodierenden Sequenzen (coding sequence = CDS) auf den Chromosomen wurde mit dem GenomeThreader Programm gewonnen [109], das von den ca. 500 Proteinsequenzen der UniProt-Datenbank die annotierte CDS im Schimpansen-Genom von GenBank mit allen Informationen identifiziert.

Für die Analyse der Introns und Exons wurden die onkologisch relevanten humanen Proteine aus der UniProt [101] selektiert und deren kodierenden Sequenzen auf dem Schimpansen- und Humangenom mittels des GenomeThreader Programms detektiert und markiert. Dieses Programm nutzt NCBI's RefSeq [110, 111], welches von GenBank [112, 113] abgeleitet ist. Die GenomeThreader Methode markiert die Exons und Introns, so dass eine automatisierte Bestimmung der Längen und Anzahlen möglich war.

Pseudogenisierung:

Ein weiterer, immer wichtiger werdender Aspekt ist die Analyse der Pseudogenisierung. Dafür wurde jedes annotierte Pseudogen in Bezug auf Frame-Shifts und frühzeitige Stop-Codons untersucht. Im Laufe dieser Analyse wurden die Pseudogene der RefSeq Annotation betrachtet.

SNPs, Insertionen und Deletionen:

Die Alignments der Exon-Nukleotidsequenzen stammten aus der Ensembl-Datenbank. Diese wurden automatisiert eingelesen und die Mutationen und Gaps, die in den Alignments markiert waren, gezählt und verglichen.

Positive Selektion:

Die Positive Selektion S_p beschreibt das Ereignis der Durchsetzung und Ausbreitung eines besser angepassten mutierten Gens oder einer Genregion in einer Population. Dabei führt eine Mutation in einer Gensequenz zur Veränderung des Phänotyps und zu einer Verbesserung einer Eigenschaft des Organismus. Die Abschätzung der positiven Selektion mit Hilfe von Verhältnistests hilft bei der Detektion von Genen, die sich unter Evolutionsdruck schneller entwickelt haben als andere. Hierbei gehen bevorzugt Punktmutationen in der Nukleotidsequenz ein, wobei zwischen synonymen („silent“) und nicht-synonymen („missense“) Mutationen unterschieden wird. Die Nukleotidsequenz eines Gens besteht aus Triplets, die jeweils für eine Aminosäure kodieren. Resultiert durch eine Substitution einer Triplet-Base ein Aminosäureaustausch, so liegt eine nicht-synonyme Mutation vor, andernfalls eine synonyme (siehe Abbildung 2.1). Substitutionen, die synonyme Mutationen produzieren repräsentieren nur eine Art konstantes Hintergrundgeräusch, während sich nicht-synonyme Veränderungen entweder für den Organismus als vorteilhaft oder während der Evolution als Nachteil erweisen, indem sie nicht langfristig akzeptiert werden. Ein hoher Quotient (>1) aus nicht-synonymer Mutationsrate (K_a) und synonyme Mutationsrate (K_s) weist auf einen starken Selektionsdruck hin, während ein kleiner Quotient auf eine Konservierung der Sequenz an dieser Stelle hindeutet. Beim Vergleich zweier Spezies sagt der Quotient jedoch nichts über die Schnelligkeit der Entfernung dieser von einem gemeinsamen Vorfahren aus. Ferner hilft es bei der Identifikation von Genen, die sich am stärksten entwickelt haben bzw. divergiert sind. Das Verhältnis von K_a und K_s wurde in dieser Arbeit folgendermaßen abgeschätzt:

$$S_p = \frac{d_A/d_S}{N_A/N_S},$$

wobei d_A der Anzahl der nicht-synonymen Mutationen und d_S der Anzahl der synonymen Mutationen entspricht. N_A ist die Anzahl der möglichen nicht-synonymen Mutationen und N_S die Anzahl der möglichen synonymen Substitutionen.

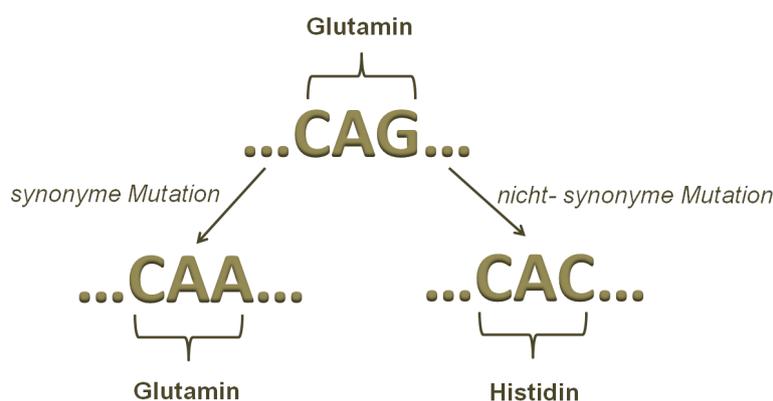


Abb. 2.1: Beispiel für eine synonyme und nicht-synonyme Mutation.

Strukturanalyse

Die Proteinsequenzen wurden aus der UniProt selektiert und mit Hilfe von ScanProsite nach funktionellen Aminosäuremotiven durchsucht [114].

Mutationen in der Gensequenz können zu einer Veränderung der Proteinfunktion führen, sofern sie sich in konservierten und wichtigen funktionellen Bereichen des Proteins befinden. Um die Auswirkung von Substitutionen auf die Proteinfunktion zwischen Mensch und Schimpanse zu analysieren, wurde das Programm PolyPhen [115] verwendet.

PolyPhen ist ein Online-Programm, das Aussagen über die Auswirkung einer Aminosäuresubstitution auf die Proteinstruktur und seine Funktion trifft und so eine Abschätzung der Schwere der Auswirkung einer Aminosäureveränderung für die Proteinstruktur ermöglicht. Dabei fließen bei der Erstellung der Vorhersage der Auswirkung neben der Proteinsequenz auch phylogenetische und strukturelle Informationen mit ein. Die zu überprüfende Proteinsequenz wird an den PolyPhen-Server übermittelt oder mit Angabe der UniProtID automatisch aus der Datenbank geladen.

2.1.1.1 Sequenzvergleiche

Sequenzvergleiche auf Protein- oder DNA-Ebene gehören zu den am häufigsten angewendeten Methoden in der Bioinformatik. Mit Hilfe von Alignierungs-Algorithmen können nicht nur zwei oder mehrere Sequenzen miteinander verglichen werden, sondern auch ganze Datenbanken, um z.B. nach verwandten Sequenzstücken zu suchen. Neben dem Sequenzvergleich einer Sequenz mit einer anderen, wurden die Algorithmen insofern erweitert und optimiert, dass auch multiple Vergleiche möglich sind, um z.B. Verwandtschaftsbeziehungen kenntlich zu machen. Zu den am häufigsten angewendeten Algorithmen gehören der Algorithmus von Needleman-Wunsch [116] und Smith-Waterman [117]. Für Multiple Sequenzalignierungs-Algorithmen sind vor allem ClustalW [118] und T-Coffee [119] zu nennen.

Für den manuellen Vergleich von onkologisch relevanten Genen von Mensch und Schimpanse wurde der in der NCBI [120, 121, 122] integrierte BLASTN [123] verwendet. Dabei wurden die menschlichen Gensequenzen jeweils gegen das Schimpansen-Genom aligniert, um neben der Lokalisation des Schimpansen-Gens auch Unterschiede wie Mutationen und Deletionen/Insertionen zu analysieren.

2.1.2 Generierung und Assemblierung von Kleinstrukturfragmenten

Im Rahmen dieser Arbeit bezeichnet „Fragmentierung“ die Zerlegung von Kleinstrukturen in kleinere Untereinheiten, die für eine genauere Analyse von Wirkstoffen von Nutzen sind. Ein essentieller Schritt des Fragment-basierten Wirkstoffdesigns ist die Erstellung einer adäquaten Fragmentbibliothek, deren Fragmente gemäß der Bindungstasche eines Zielproteins zu potenziellen neuen Wirkstoffkandidaten erweitert und kombiniert werden. Eine Fragmentbibliothek unterliegt hohen Ansprüchen. Sie soll den aus mehr als 10^{60} Kleinstrukturen bestehenden chemischen Raum weitestgehend abdecken. Für die Erstellung einer geeigneten Fragmentbibliothek wurden in dieser Arbeit verschiedene Substanzbibliotheken fragmentiert und analysiert. Die Fragmentierung der Substanzbibliotheken erfolgte durch verschiedene Schneidestrategien, die im Folgenden erklärt werden. Zunächst mussten jedoch die charakteristischen Eigenschaften für ein Fragment definiert werden. Ein Fragment musste folgende Eigenschaften besitzen und durfte nur gegen eine dieser Regeln verstoßen:

1. Das Molekulargewicht betrug maximal 300 g/mol

2. Das Fragment besaß maximal 3 Wasserstoffbrücken-Donatoren
3. Das Fragment besaß maximal 3 Wasserstoffbrücken-Akzeptoren
4. Der Verteilungskoeffizient betrug maximal 3
5. Das Fragment besaß mindestens 3 Atome
6. Der PSA-Wert (Polar Surface Area) betrug maximal 60 \AA^2
7. Das Fragment besaß maximal 3 rotierbare Bindungen

Die Regeln eins bis vier werden auch als „Rule of Three“ (für Fragmente) [124] bezeichnet und wurden von der bekannten „Rule of Five“ (für Wirkstoffe) [87] abgeleitet. Die anderen Eigenschaften stammten von Congreve et al. [124].

Generierung des zu fragmentierenden Substanzdatensatzes:

Die Erstellung des Fragmentdatensatzes erfolgte durch die Fragmentierung von drei verschiedenen Klassen von Kleinstruktur-Datenbanken (therapeutisch angewandte Wirkstoffe, Metabolite, toxische Kleinstrukturen) mittels drei verschiedener Fragmentierungsstrategien. Der Arzneistoffdatensatz bestand aus der Fragmentierung von verschiedenen Wirkstoffen aus den folgenden Datenbanken:

1. SuperDrug Wirkstoffe (~2.300)
2. KEGG Wirkstoffe (~6.000)
3. zugelassene Wirkstoffe der DrugBank (~1.300)
4. WDI Wirkstoffe (~6.000)
5. CMC Wirkstoffe (~8.000)

Nach Aussortierung der Duplikate entstand ein Arzneistoff-Datensatz von ca. 16.100 Kleinstrukturen. Für den metabolischen Substanzdatensatz wurden die über 13.000 Kleinstrukturen der KEGG-Metabolite fragmentiert. Der dritte Datensatz setzte sich aus den giftigsten Substanzen der SuperToxic-Datenbank [125] zusammen. Dabei galt eine Substanz als sehr giftig, wenn sie in einem beliebigen Organismus einen LD50-Wert (Lethale Dosis) von kleiner als 25 mg/kg besitzt unabhängig von der Art der Verabreichung. Da auch *in vitro* Daten in der SuperToxic-Datenbank gespeichert sind, wurde hierfür ein nlog(LC50) -Grenzwert von ≥ 9 (Konzentration ist

also $\leq 10^{-9} \text{ mol/l}$) gewählt. Insgesamt umfasste der Datensatz über 2.300 toxische Substanzen.

Fragmentierungsstrategien:

Die erste Fragmentierungsmethode erstellte eine Fragmentbibliothek bestehend aus Fragmenten aus einfachen Bausteinen, die mittels einfacher chemischer Verbindungen zu größeren Molekülen verknüpft und neu kombiniert werden können. Hierbei wurden die Kleinstrukturen mittels der Recap-Regeln [126] rekursiv fragmentiert. Dabei trennen die Recap-Regeln elf chemische Bindungen (siehe Abbildung 2.2). Eine weitere Fragmentierungsstrategie war das nicht-rekursive Herausschneiden von

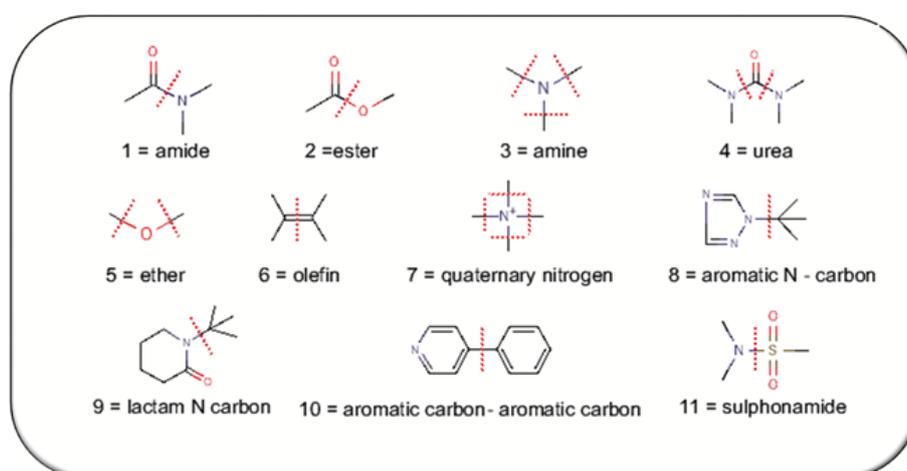


Abb. 2.2: Recap-Regeln zur Fragmentierung.

Die Recap-Regeln trennen elf chemische Bindungen auf. Aus *Lewell et al.* [126].

Linkern, d.h. von langkettigen Atomen zwischen zwei Ringstrukturen. Dieser Algorithmus wurde für statistische Analysen bevorzugt, da bei dieser Methode vollständige Fragmente entstehen, d.h. Fragmente mit den funktionellen Gruppen. Diese funktionellen Gruppen wurden bei dem rekursiven Recap-Schnitt teilweise abgeschnitten. Eine dritte Fragmentierungsstrategie war die rekursive Fragmentierung der Strukturen an ihren rotierbaren Bindungen. Verglichen mit den zuvor beschriebenen Methoden entsteht nach Anwendung der letzten Strategie der umfangreichste Fragmentdatensatz. Abbildung 2.3 zeigt exemplarisch die Anwendung der drei Fragmentierungsstrategien an einem Beispiel. Alle Fragmentierungsalgorithmen wurden in der Programmiersprache Python implementiert.

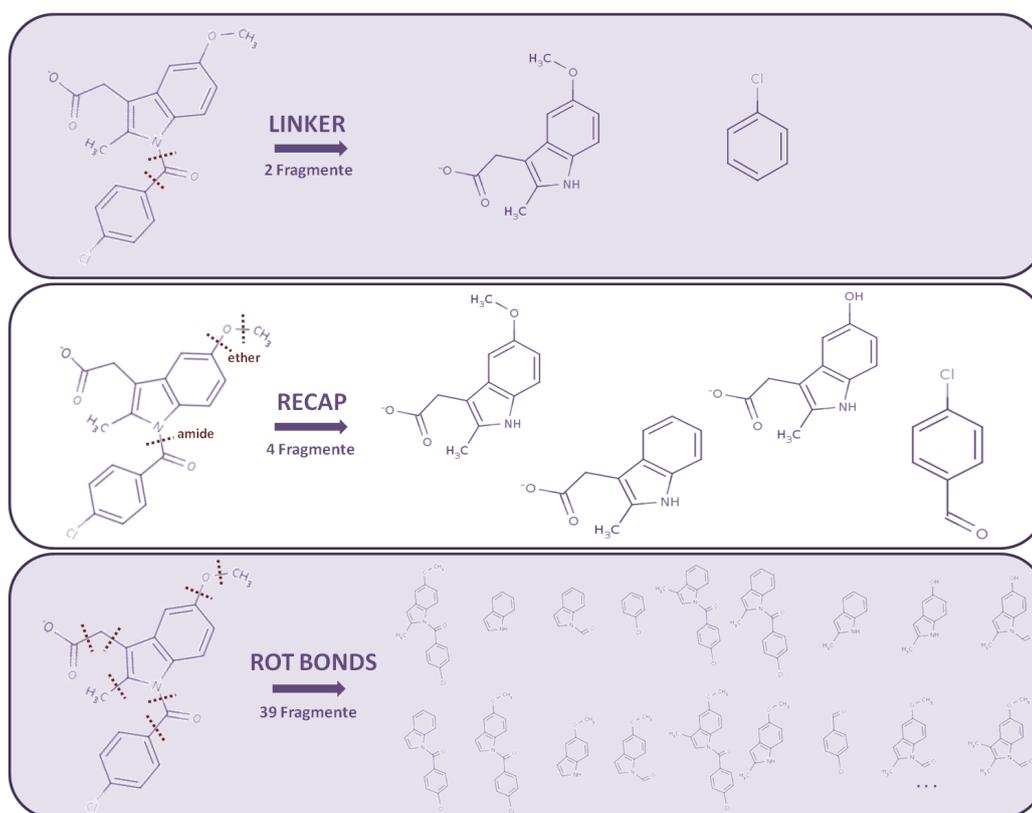


Abb. 2.3: Fragmentierungsstrategien angewandt an einem Beispiel.

Mit der Linker-Fragmentierung wird nur die Kettenstruktur zwischen den beiden Ringen herausgeschnitten. Es entstehen zwei Fragmente. Die rekursive Fragmentierung nach Recap generiert vier verschiedene Fragmente. Die rekursive Fragmentierung nach den rotierbaren Bindungen liefert die größte Anzahl an Fragmenten (insgesamt 39), wobei nur ein Teil visualisiert ist. Alle generierten und dargestellten Fragmente verstoßen gegen maximal eine Regel der „Rule of Three“. *Aus Ahmed et al.* [127].

Fragment-Assemblierung:

Für die Implementierung einer Fragment-Assemblierungsmethode zur Generierung von Kleinstrukturen aus mindestens zwei Fragmenten wurde die Programmiersprache Python verwendet. Zur Bestimmung der Schnittstellen, an denen die Fragmente verknüpft werden sollten, wurden erneut die Recap-Regeln angewandt. Nach Identifizierung der benötigten Enden für eine bestimmte chemische Verbindung konnten die Fragmente an dieser Stelle verknüpft werden. Lag beispielsweise ein Fragment mit einer freien Carbonylgruppe und ein weiteres mit einem freien Sauerstoffatom vor, wurden diese Fragmente so verbunden, dass eine Ester-Bindung entstand. Mit dieser Assemblierungsmethode wurden alle möglichen Kleinstruktur-Kombinationen

aus maximal drei Fragmenten erstellt, wobei die aktuelle Struktur nur dann um ein weiteres Fragment erweitert wurde, wenn dadurch nur gegen maximal eine der Lipinski 5er Regel verstoßen wurde. Die Aussortierung der identischen Moleküle mit InChi-Vergleichen, lieferte eine Kleinstrukturliste von SMILES, die mit Hilfe von verschiedenen Programmen visualisiert oder als Moleküldatei abgespeichert werden konnte.

2.1.3 Berechnung der Bindungstaschenpräferenzen von Fragmenten

Als Bindungstaschenpräferenz (BTP) wird in dieser Arbeit die bevorzugte Aminosäurezusammensetzung der Bindungstasche des bindenden Fragments bezeichnet. Die Präferenz setzte sich im Detail aus der Aminosäurezusammensetzung in Abstand von 5 Å des Fragments und der Zusammensetzung der Aminosäuren (AS) der Proteinoberfläche zusammen:

$$BTP = \frac{\frac{\#AAS_{BT}}{\#AS_{BT}}}{\frac{\#AAS_{PO}}{\#AS_{PO}}},$$

wobei $\#AAS_{BT}$ der Anzahl der aktuellen Aminosäure in der Bindungstasche, $\#AS_{BT}$ der Gesamtanzahl der Aminosäuren in der Bindungstasche, $\#AAS_{PO}$ der Anzahl der aktuellen Aminosäure auf der Proteinoberfläche und $\#AS_{PO}$ der Gesamtanzahl der AS auf der Proteinoberfläche entspricht. Für die Berechnung der Aminosäuren der Proteinoberfläche wurde das Programm DSSP (<http://swift.cmbi.ru.nl/gv/dssp/>) verwendet [128]. Bei dieser Methode wandert eine Kugel mit bestimmtem Durchmesser über die Proteinstruktur und deklariert alle mit der Kugel in Kontakt tretenden Aminosäuren als Proteinoberflächenaminosäuren.

Es wurde davon ausgegangen, dass Fragmente eines Wirkstoffs spezifischere Aminosäurenpräferenzen besitzen als die Kleinstruktur. Desweiteren sind diese Präferenzen signifikanter, weil viele Fragmente in den Bindungstaschen verschiedener Proteine vorkommen und in die Berechnung mit eingehen, während die meisten Liganden nur in einem Protein oder ähnlicher Proteingruppe gebunden sind. Für die Implementierung wurde die Skript-Sprache Python und das Visualisierungsprogramm PyMol verwendet.

2.2 Vorhersage von Zielproteinen und potenziellen Liganden

Der Methodenteil zur Vorhersage von Zielproteinen und von potenziellen Liganden besteht aus drei Teilen. Im ersten Abschnitt werden die Verfahren erläutert, die für die Computer- und Struktur-basierte Wirkstoffentwicklung notwendig waren. Der zweite Abschnitt behandelt so genannte Fingerprints, die den schnellen Vergleich von Kleinstrukturen und deren Wirkprofile möglich machten. Im letzten Abschnitt werden die Materialien und Methoden für die Zielproteinvorhersage basierend auf physikochemischen Eigenschaften und strukturellen Ähnlichkeiten von Kleinstrukturen beschrieben.

2.2.1 Homologie-Modellierung

Für die Durchführung von Struktur-basiertem Wirkstoffdesign ist eine 3D-Struktur des Zielproteins essentiell. Ist keine vollständige PDB-Struktur des Targetproteins erhältlich, so kann diese im Falle des Vorhandenseins einer geeigneten Template-Struktur modelliert werden. Voraussetzung für eine Homologie-Modellierung ist die Präsenz einer Template-Struktur in der PDB mit mindestens 30% Sequenzidentität. Dafür wird in einem ersten Schritt die Sequenz des zu modellierenden Proteins gegen die PDB aligniert, um ein geeignetes Template zu identifizieren. Die PDB stellt für den Vergleich von einzelnen Proteinsequenzen gegen die Struktur-Datenbank eigene Alignierungs-Funktionen zur Verfügung, jedoch liefert eine lokale Funktion einen Geschwindigkeitsvorteil. Im Laufe dieser Arbeit wurde für das automatisierte Alignieren von Proteinsequenzen gegen die PDB eine lokale Python-Funktion verwendet. Die Struktursequenzen stellt die PDB als Textdatei zur Verfügung, welche nach Speicherung in eine lokale Datenbank automatisiert mit den zu modellierenden Proteinsequenzen verglichen wurde. Im nächsten Schritt wurde das Sequenzalignment des Templates und des Targets optimiert. Anschließend mussten die unterschiedlichen Aminosäuren der Template-Proteinstruktur ausgehend vom Alignment ausgetauscht, die Deletionen gelöscht und die Insertionen eingefügt werden. Die Modellierung fand entweder mit Hilfe mehrerer Programme manuell oder automatisiert statt. Der SwissPDBViewer ist ein Visualisierungsprogramm für 3D-Proteinstrukturen aus der PDB und bietet mehrere Methoden und Funktionen auch für die Homologie-Modellierung an. Mit Hilfe des SwissPDBViewers sind Aminosäuresubstitutionen

und Deletionen manuell durchführbar. Für das Einfügen von längeren Insertionen in eine Proteinstruktur wurde das Online-Programm SuperLooper [129] genutzt. Das SuperLooper Online-Programm beinhaltet eine Datenbank mit mehreren Millionen 3D-Linkerstrukturen verschiedener Länge, die in eine Struktur eingepasst werden können. Eine anschließende Energieminimierung der modellierten Struktur mit Hilfe des Visualisierungsprogramms SwissPDBViewer lieferte die energetisch günstigste Konformation des Strukturmodells. Die Änderung von Torsionswinkeln führen zur Vermeidung von zu engen Kontakten und zum Erhalt eines energiebezogenen lokalen Minimums.

Ein weiteres Programm zur manuellen Homologie-Modellierung ist das Online Programm Swissmodeller [130]. Dieses Programm benötigt nur ein Alignment der Templatessequenz, einen PDB-Datenbank-Identifikator und die zu modellierende Proteinsequenz. Das Swissmodeller-Programm lädt die Template-Struktur aus der PDB, passt die Proteinstruktursequenz ausgehend von dem Alignment der zu modellierenden Proteinstruktur an und führt anschließend die Minimierung durch. Die stereochemischen Qualitäten der modellierten Proteinstrukturen wurden mit dem Programm Procheck validiert [131].

Für das automatisierte Erstellen mehrerer Strukturmodelle wurde das Modeller-Programm [132, 133] verwendet. Dieses Programm benötigt Alignments im PIR-Format (siehe Abbildung 2.4).

```
>P1;1bdma
structureX:1bdm.pdb:  0 :A:+318 :A:undefined:undefined:-1.00:-1.00
MKAPVRVAVTGAAGQIGYSLLFRIAAGEMLGKQDPVILQLLEIPQAMKALEGVMELEDCAFPLLAGLEATDDPD
VAFKDADYALLVGAAPRL-----QVNGKIFTEQGRALAEVAKKDVKVLVVGPNANTNALIAYKNAPGLNPR
NFTAMTRLDHNRAKAQLAKKTGTGVDRIIRRMVWGNHSSIMFPDLFHAED----GRPALELVDMEWYKVFIP
VAQRGAIIQARGASSAASAANAIEHIRDVALGTPEGDWVSMVAVPSQ--GEYGIPEGIVYSFPVTAK-DGAYRV
VEGLEINEFARKRMEITAQELLDEMEQVKAL---GLI*

>P1;TVLDH
sequence:TVLDH:      : :      : :: 0.00: 0.00
MSEAAHVLITGAAGQIGYILSHWIASGELYG-DRQVYLHLLDIPPAMNRLTALTMELEDCAFPHLAGFVATTDPK
AAFKDIDCAFLVASMPLKPGQVRADLISSNSVIFKNTGEYLSKWAKPSVKVLVIGNPDNTNCEIAMLHAKNLKPE
NFSSL SMLDQNRAYYEVASKLGVDVKDVHDIIVWGNHGESMVADLTQATFTKEGKTQKVVDVLDHDYVFDTFKK
IGHRAWIDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNPYGIKPGVVFSPFCNVDEKGIHV
VEGFKVNDWLRKLDLDFTEKDLFHEKEIALNHLAQGG*
```

Abb. 2.4: Beispiel für ein vom Modeller-Programm als Input benötigtes Alignment im PIR-Format. Die alignierten Sequenzen werden in zwei Blöcken hintereinander als Input angegeben. *Aus: <http://www.salilab.org/modeller/documentation.html>.*

Die Alignments wurden automatisiert für alle zu modellierenden Proteine und einer PDB-Proteinsequenz mit höchster Sequenzidentität erstellt. Dafür wurden in einem ersten Schritt die PDB-Sequenzen heruntergeladen und mittels dem BLAST-Algorithmus für Proteine, der in Biopython zur Verfügung gestellt wird, erstellt und in die für den Modeller gewünschte Form gebracht. Anschließend wurde das

Modeller-Programm automatisiert mit den Alignments angewendet. Neben den Mutationen, Deletionen und Insertionen wird auch eine Energieminimierung der modellierten Proteinstruktur durchgeführt. Eine Qualitätskontrolle wird ebenfalls von dem Modeller-Programm mitgeliefert.

2.2.2 *In silico* Screening

In silico Screening bezeichnet in der Arzneistoffforschung die Suche nach zu einer Leitstruktur ähnlichen Kleinstrukturen in großen Substanzbibliotheken unter ausschließlicher Verwendung von computergestützten Verfahren. *In silico* Screening Methoden werden zur Leitstruktursuche und -optimierung verwendet. Im Laufe dieser Arbeit wurden parallel 2D- und 3D- Ähnlichkeitsverfahren durchgeführt. Ähnlichkeitssuche ist eine bewährte Methode zur Identifizierung von Kleinstrukturen mit hoher Ähnlichkeit zu einer Leitstruktur und ähnlichen Eigenschaften. Mittels 2D- und 3D-Screening Algorithmen wurden Kleinstruktur-Datenbanken durchsucht, die mehrere Millionen Substanzen enthalten. Eine anschließende Bewertung und Selektion der Screening-Hits erfolgte mit verschiedenen Filterkriterien. Neben der Lipinski 5er Regel und ADMETox-Analysen wurden auch Protein-Liganden-Docking Analysen zur Bewertung einbezogen. Nach *in vitro* Test Validierung der resultierenden Substanzen und potenziellen Wirkstoffkandidaten dienten die positiv getesteten Substanzen als Leitstrukturen für einen neuen *in silico* Screening Durchlauf (siehe Abbildung 2.5).

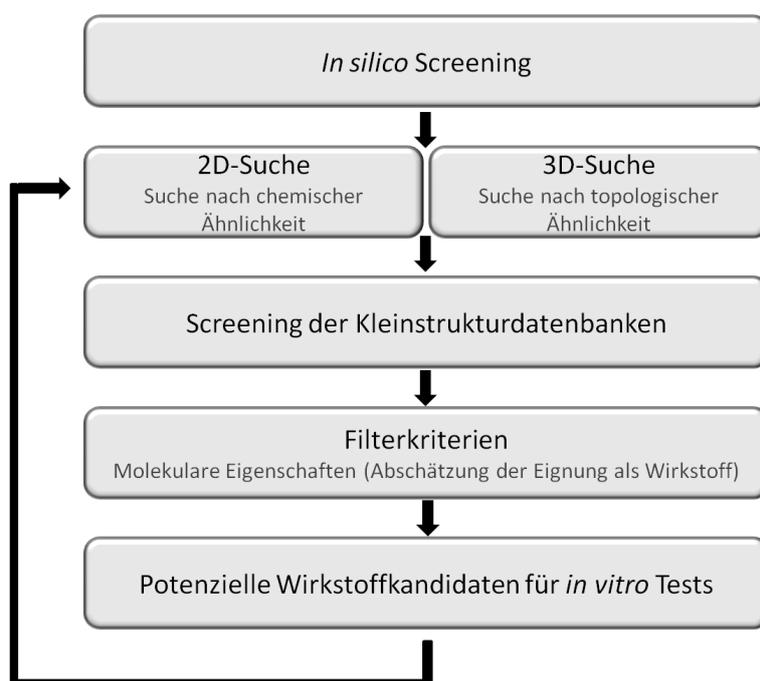


Abb. 2.5: *In silico* Screening.

Ausgehend von einer Leitstruktur werden bei einem *in silico* Screening-Durchlauf parallel 2D- und 3D-Ähnlichkeitssuchen in großen Kleinstruktur-Datenbanken durchgeführt. Mit Hilfe der 2D-Suche können Strukturen mit ähnlichen chemischen Eigenschaften identifiziert werden. Um auch Strukturen mit gleicher Topologie, aber anderen chemischen Grundgerüsten zu identifizieren, wird zusätzlich eine 3D-Ähnlichkeitssuche verwendet. Dadurch können so genannte „Scaffold-Hopper“ detektiert werden. Anschließend werden die Treffer beider Suchen nach Eigenschaften gefiltert, mit Docking-Methoden bewertet und die resultierenden Kleinstrukturen experimentell getestet. Die positiven Treffer können anschließend wieder als Leitstruktur für das *in silico* Screening verwendet werden, um die Kleinstrukturen zu optimieren und weitere potenzielle Binder zu finden.

2.2.2.1 2D-Ähnlichkeit

Für die Identifikation von potenziellen bzw. potenziell affineren Wirkstoffkandidaten als die Leitstruktur mit ähnlicher chemischer Zusammensetzung wurde eine Inhouse-Datenbank mit ca. vier Millionen Kleinstrukturen durchsucht. Hierfür wurde von den zu vergleichenden Substanzen ein struktureller Fingerprint, wie in Kapitel 2.2.5.1 beschrieben, generiert. Potenzielle Wirkstoffkandidaten, die in den nächsten Entwicklungsschritt involviert wurden, besaßen eine 2D-Ähnlichkeit von mindestens 85% zur Leitstruktur. Die Verwendung von 2D-Ähnlichkeitssuchen ermöglicht die Abtastung des chemischen Raums und die Detektion von chemisch ähnlichen Substanzen, jedoch lassen sich keine Aussagen über die räumliche Ähnlichkeit der Kleinstrukturen treffen. Dafür wurde in der Arbeitsgruppe die 3D-Ähnlichkeitssuche entwickelt.

2.2.2.2 3D-Ähnlichkeit

Für die Analyse der räumlichen Ähnlichkeit zweier Kleinstrukturen dient die 3D-Ähnlichkeitssuche, die im Folgenden beschrieben wird. Neben Kleinstrukturen mit ähnlicher chemischer Zusammensetzung stellen auch Kleinstrukturen mit ähnlicher dreidimensionaler Raumstruktur potenziell bessere Inhibitoren dar. Diese so genannten „Scaffold-Hopper“ besitzen ähnliche dreidimensionale Grundgerüste wie die Leitstruktur und sind ebenfalls für eine Positionierung in der Bindungstasche eines Zielproteins geeignet. Die Verwendung eines in der Arbeitsgruppe entwickelten 3D-Überlagerungs-Algorithmus [134] zeigte schon in vorangegangenen Arbeiten Erfolge bei der Suche nach neuen Anti-Tumor- [135] und neuen TSE-Wirkstoffen [136]. Die Suche basiert auf dem Vergleich einer Leitstruktur mit Kleinstrukturen und seinen mit dem MedChemExplorer von Accelrys [137] vorberechneten Konformeren aus der Datenbank. Die verwendete Inhouse-Datenbank beinhaltet mehr als 140 Millionen vorberechnete Konformere von ca. vier Millionen verschiedenen Substanzen. Der schnelle 3D-Vergleich der Kleinstrukturen basiert auf der Berechnung der jeweiligen Mittelschwerpunkte der auf die Kleinstrukturen gelegten Ebenen (siehe Abbildung 2.6). Die anschließende Überlagerung der Ebenen bei festem Mittelschwerpunkt resultiert in vier verschiedenen Überlagerungsmöglichkeiten, wobei Atome in unmittelbarer Nähe nach W.Kabsch überlagert werden [138]. Die anschließende Verfeinerung der Überlagerung dient der Optimierung des folgenden Scores:

$$score = \%ueberlagerteAtome \cdot e^{(-rmsd)}$$

Der Score basiert auf der Multiplikation des Prozentsatzes der überlagerten Atome mit dem negativen RMSD (Root-Mean-Square Deviation). Je höher die Anzahl der überlagerten Atome und je kleiner der RMSD, desto höher ist der Score und desto besser ist die Überlagerung.

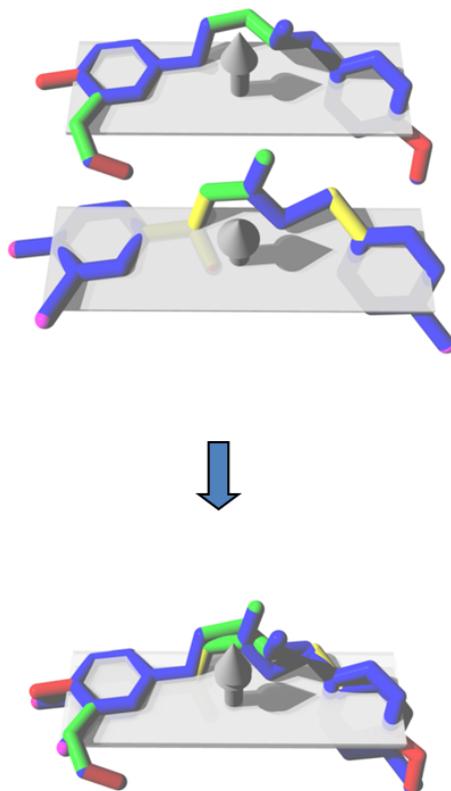


Abb. 2.6: Prinzip der 3D-Überlagerung.

Für den 3D-Vergleich wird in die zu vergleichenden Kleinstrukturen je eine Ebene gelegt, die das Trägheitsmoment repräsentiert und der Schwerpunkt berechnet. Die langen und schmalen Seiten der Ebenen werden überlagert, wobei es vier Möglichkeiten gibt. Die Atome, die sich in der Nähe befinden werden nach W. Kabsch überlagert und ein Score berechnet. Dieser Score berechnet sich aus der Anzahl der überlagerten Atome und dem RMSD. *Aus Ahmed et al.* [139].

2.2.2.3 Protein-Liganden-Docking

Protein-Liganden-Docking beschreibt die automatisierte Positionierung eines Liganden in die Bindungstasche einer Proteinstruktur. Docking lässt sich anwenden, um den Bindungsmodus eines Liganden in der Bindungsstelle einer bekannten Zielproteinstruktur zu erforschen. Es gibt verschiedene Docking-Programme, die auf verschiedenen Algorithmen basieren. Im Laufe dieser Arbeit fand das Docking-Programm Gold Anwendung [140], welches auf einem genetischen Algorithmus basiert (siehe Abbildung 2.7) und drei verschiedene Scoring-Funktionen bereitstellt. Hier wird die Bindungstasche auf einem Gitter dargestellt, wobei auf jedem Gitterpunkt Atom-Wechselwirkungspotenziale abgebildet werden. Das Gitter stellt den Suchraum für die zu positionierenden Liganden dar. Beim Docking werden nur nicht-kovalente Wechselwirkungen betrachtet. Darüber hinaus wird der Ligand bei Docking mit

Gold flexibel gehalten, indem verschiedene Konformere generiert werden. Zusätzlich können maximal zehn Aminosäurereste der Bindungstasche angewählt werden, die ebenfalls flexibel gehalten werden.

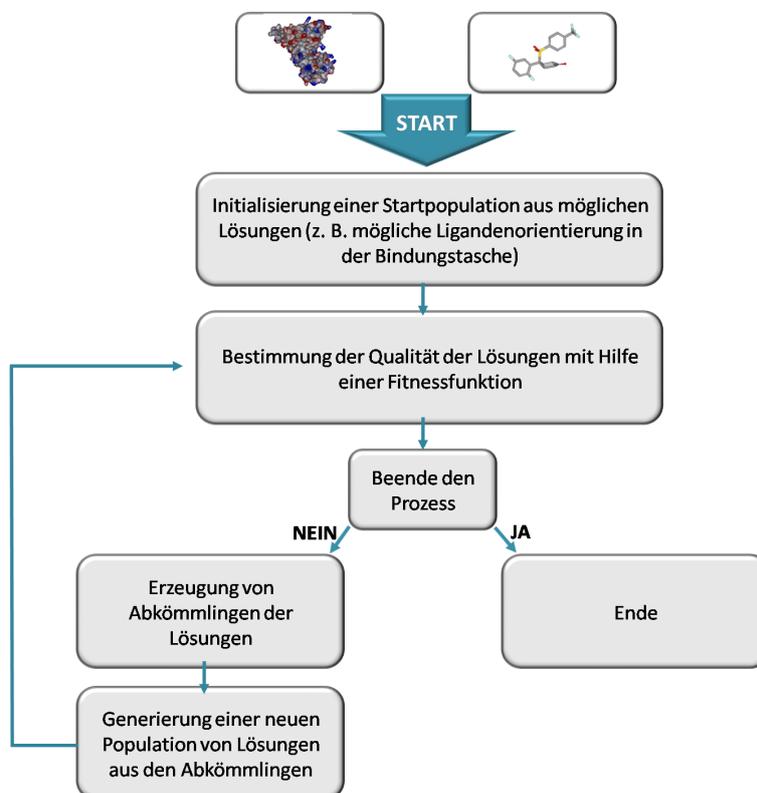


Abb. 2.7: Prinzip des genetischen Algorithmus beim Protein-Liganden Docking.

Der genetische Algorithmus ist ein Verfahren, das den Prozess der Evolution simuliert, indem eine Kollektion von Daten verändert wird. Ausgehend von dem Zielprotein und dem Liganden wird zunächst eine Startpopulation erstellt, die verschiedene Konformationen des Liganden oder verschiedene Orientierungen des Liganden in der Bindungstasche beinhalten kann. Diese werden in einem nächsten Schritt mittels einer Fitnessfunktion bewertet und die besten Ergebnisse selektiert. Diese können ausgegeben oder in einen weiteren Zyklus als „Eltern“ fungieren, aus denen dann eine neue Population generiert wird, womit die nächste Generationsschleife beginnt.

Im Rahmen dieser Arbeit diente das Docking mit Gold zur Bewertung von potenziellen Liganden, die mit dem *in silico* Screening identifiziert wurden. Des Weiteren diente es zur Analyse der Bindungsmodi der verschiedenen Kleinstrukturen.

Docking von Liganden in die Bindungstasche der γ -Sekretase:

Für das Docking von Kleinstrukturen in die γ -Sekretase wurde die Bindungstasche definiert als den Proteinbereich 10 Å um das Sauerstoffatom eines katalytischen Aspartatrestes (AtomID: 3862). Für jeden Liganden wurden 20 Konformere (Standard:

10) berechnet und das Standard-Abbruchkriterium deaktiviert, so dass nicht bei einem ersten gefundenen besten Hit abgebrochen wurde. Die Populationsgröße wurde auf 100 gesetzt. Ansonsten galten die Standardeinstellungen. Es wurde sowohl der ChemScore als auch der GoldScore verwendet.

2.2.3 ADMETox-Vorhersage

Für die ADMETox-Analysen fand im Laufe dieser Arbeit das Accelrys-Softwarepaket Verwendung (<http://accelrys.com/products/>). Unter verschiedenen anderen Analyse-Programmen enthält das Paket drei ADMETox-Vorhersage Protokolle: „ADMET-Deskriptors“, „Toxicity Prediction (Extensible)“, „Toxicity Prediction (TOPKAT)“, wobei das erste Protokoll hauptsächlich ADME-Eigenschaften und die letzten beiden Protokolle toxikologische Eigenschaften vorhersagt. Im Laufe dieser Arbeit wurden alle drei Protokolle verwendet.

Vorhersage mit dem Protokoll „ADMET-Deskriptors“:

Die Berechnung bzw. Abschätzung folgender Eigenschaften werden durch dieses Protokoll realisiert, wobei weitere Informationen und die genaue Beschreibung der Modellsysteme auf der offiziellen Accelrys Webseite (<http://accelrys.com>) und in der angegebenen Referenz zu finden sind:

- **Wasserlöslichkeit bei 25°C** [141]
- **Penetration der Blut-Hirn-Schranke nach oraler Einnahme** [142]
- **CYP2D6 Bindung und Inhibition** [143]
- **Hepatotoxizität** (Lebertoxizität durch nicht mögliche Metabolisierung der Substanz und Anreicherung in der Leber) [144]
- **Intestinale Absorption** [145, 142]
- **Plasma Protein Bindung** für die Fähigkeit Carrier-Proteine zu binden [146]
- **AlogP98** (Atombasierte Berechnung von logP)
- **PSA-2D**

Die Wasserlöslichkeit, die Fähigkeit der Substanz zur Penetration der Blut-Hirn-Schranke, intestinale Absorption, AlogP98, PSA und die Plasma-Protein Bindung

sind entscheidende Eigenschaften für die Aufnahme und Verteilung des Wirkstoffs im menschlichen Organismus. Die Inhibition des Cytochrom CYP2D6 kann zum veränderten Wirkstoff-Metabolismus führen. Zusätzlich wird die Lebertoxizität, die auch eine Toxizitätsvorhersage darstellt, abgeschätzt.

Vorhersage mit den Protokollen „Toxicity Prediction (Extensible)“ und „Toxicity Prediction (TOPKAT)“:

Für die Toxizitätsvorhersage können mit dem zweiten und dem dritten von Accelrys zur Verfügung gestellten Protokollen die Testsubstanzen verschiedenen Modelltestsystemen zugeführt werden, um eine Abschätzung des jeweiligen toxischen Potenzials zu erhalten. Beide Protokolle nutzen ähnliche Modelle, jedoch sind die Ausgaben unterschiedlich detailliert. Der Vollständigkeit halber wurden beide Toxizitätsprotokolle verwendet. Es werden folgende Modelle bereitgestellt:

- NTP Nagetier Karzinogenität
- FDA Nagetier Karzinogenität
- “Weight of Evidence“ Karzinogenität
- Karzinogenität Potenzial TD50
- Ames Mutagenität (schnelle Erkennung von karzinogenen Substanzen)
- **Entwickelndes Toxizitätspotenzial**
- **Ratte Oral LD50**
- **Ratte maximal tolerierte Dosis**
- Ratte inhalativ LC50
- Ratte chronisch LOAEL
- Hautreizung
- Hautsensibilisierung
- Okularreizung
- Sauerstoffabhängige Abbaubarkeit

- Fathead Minnow LC50
- Wasserflöhe EC50

Die im Rahmen dieser Arbeit zu testenden Substanzen wurden nur den markierten Modellen zugeführt.

2.2.4 Analogie-Modellierung

Die Analogie-Modellierung [139] stellt eine Kombination verschiedener Methoden für die Entwicklung neuer potenzieller Liganden dar und ermöglicht Struktur-basierte Wirkstoffentwicklung ohne Präsenz einer PDB-Proteinstruktur mit mindestens 30% Sequenzähnlichkeit und damit ohne die Möglichkeit eine Homologie-Modellierung durchführen zu können. Sie beinhaltet Methoden zur Proteinstrukturvorhersage und Ähnlichkeitssuchen für die Ligandensuche. Der erste Schritt bestand in der Detektion einer geeigneten PDB-Template Struktur mit ähnlicher Faltung. Im Laufe dieser Arbeit wurde für die Bindungstasche des Zielproteins eine andere PDB-Struktur verwendet als für den Rest des Proteins. Zuerst wird jedoch mit Hilfe von Faltungserkennung das Template für die Modellierung der vollständigen Struktur gesucht. Für die Faltungserkennung wird ein sequenzbasiertes Eigenschaftsprofil des zu modellierenden Zielproteins erstellt, welches in die Strukturen der Proteine in der PDB gefädelt („Threading“) wird. Die Proteinstruktur mit dem ähnlichsten Profil zum zu modellierenden Zielprotein wird als Template verwendet. Im zweiten Schritt wird die Bindungstasche modelliert. Hier kommen Bindungstaschen von Proteinen in Frage, die zu einer ähnlichen Proteinklasse gehören, beispielsweise besitzen die β - und die γ -Sekretase jeweils eine katalytische Bindungstasche mit zwei Aspartat-Resten, daher wäre es von Vorteil, wenn eine der beiden Strukturen in der PDB vorliegen würde, um die Bindungstasche der anderen Struktur mit dieser als Template modellieren zu können. Anschließend wurde die Geometrie der Bindungstasche auf die der Templatestruktur des Restproteins übertragen. Im nächsten Schritt wurden neue potenzielle Inhibitoren ausgehend von einer Leitstruktur mit 2D- und 3D-Ähnlichkeitssuchen (2.2.2) identifiziert, experimentell validiert und die Bindungstasche optimiert. Dieser Zyklus kann mehrmals durchlaufen werden.

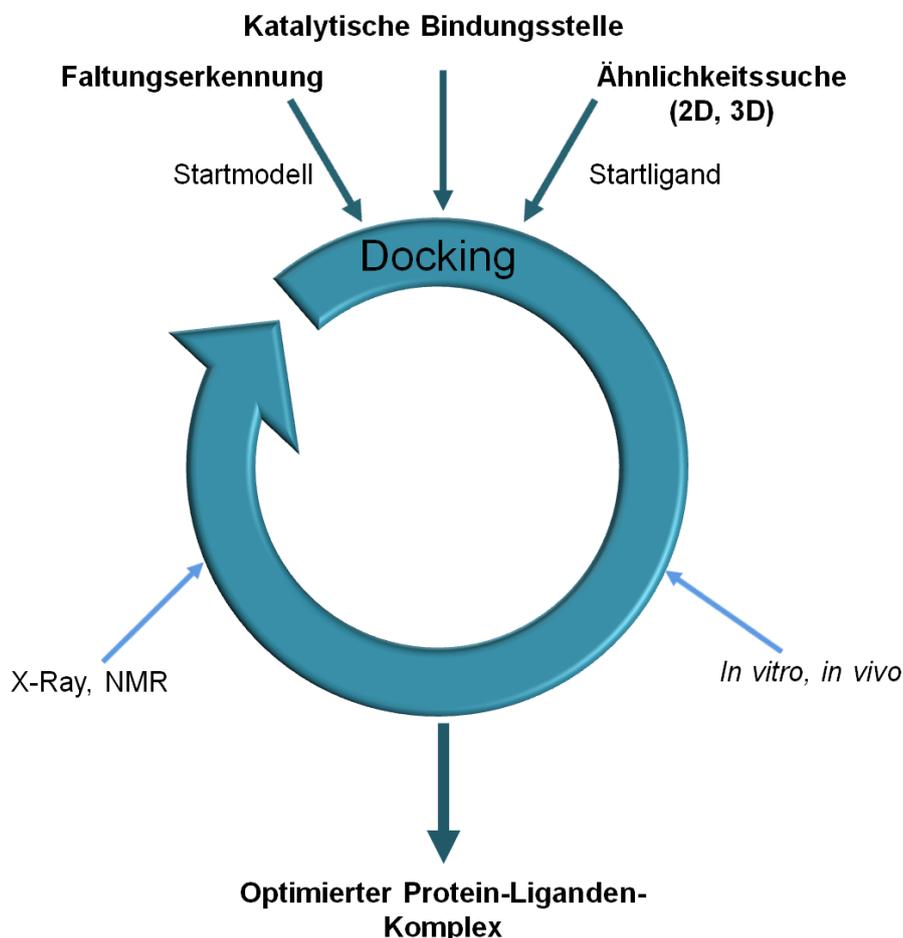


Abb. 2.8: Schematische Darstellung der Analogie-Modellierung.

Das Analogie-Modell ist eine Kombination verschiedener Methoden zur Erstellung einer Zielproteinstruktur und Bestimmung neuer Kleinstrukturen für dessen Inhibition. Sind Strukturen mit ähnlicher struktureller Ausstattung zu dem Zielprotein vorhanden, kann für die Strukturvorhersage eine Kombination von Faltungsvorhersage und ligandenunterstützter Bindungsstellenmodellierung verwendet werden. Ausgehend von einer Leitstruktur können verschiedene Screeningmethoden wie 2D- und 3D-Ähnlichkeitssuchen zur Vorhersage von potenziellen Kleinstrukturen genutzt werden, die bei der Verfeinerung der Bindungsstelle des Zielproteins von Nutzen sein können. Auch *in vitro* und *in vivo* Experimente sowie X-Ray- und NMR-Verfahren können bei der Verfeinerung und der Validierung der Zielproteinstruktur und der potenziellen Liganden helfen. *In vitro* Tests können schnell Nichtbinder identifizieren und aussortieren, während Röntgenkristallstrukturanalyse und NMR-Titration die Bindungsmodi von Liganden verifizieren. Das Ergebnis der Analogie-Modellierung sind optimierte Zielproteinstrukturen und aktive Liganden. In dieser Arbeit wurden die letzten Schritte: *in vivo*-Tests, NMR und X-Ray-Untersuchungen ausgelassen. Aus Ahmed et al. (modifiziert) [139].

2.2.5 Fingerprints zum schnellen Vergleich von Kleinstrukturen und Wirkprofilen

Fingerprints sind Bitvektoren bestimmter Länge, die bestimmte Eigenschaften kodieren und einen schnellen Vergleich von mehreren tausend Daten ermöglichen.

2.2.5.1 Struktureller Fingerprint zum Vergleich von Kleinstrukturen

Bei dem strukturellen Fingerprint handelt es sich in der Regel um einen Bitvektor der Länge 1024, der verschiedene chemische Eigenschaften eines Moleküls abstrakt repräsentiert. Mit Hilfe des strukturellen Fingerprints lassen sich Substanzen einer Datenbank in Hinblick auf ihre chemische Zusammensetzung innerhalb weniger Sekunden vergleichen. Dafür werden für alle Substanzen in jeder Datenbank, die im Laufe dieser Arbeit erstellt wurden, vorab der strukturelle Fingerprint generiert und abgelegt.

Die Generierung eines Fingerprints eines Moleküls lässt sich mit Hilfe verschiedener Methoden durchführen. Daylight (<http://www.daylight.com/>), CDK für Java oder Openbabel für Python bieten die Methode zum einfachen Anwenden an. Das Prinzip ist jedoch bei allen Methoden gleich.

In einem ersten Schritt werden ausgehend von jedem Atom des Moleküls alle möglichen Pattern bis zu einer maximalen Länge (in der Regel sieben) generiert. D.h. bei einer Länge von sieben Atomen werden in einem ersten Schritt für jedes Atom alle Pfade der Länge null (nur das Atom selbst) generiert, dann der Länge eins bis zu der Länge sieben. Die Pattern repräsentieren dabei die möglichen Pfade durch die Nachbaratome des Moleküls. Die so generierten Pfade werden anschließend mittels einer Hash-Funktion in eine Pseudo-Zufallszahl umgewandelt und mit einem logischen ODER zu dem strukturellen Fingerprint addiert.

Um die Fingerprints und damit die Kleinstrukturen miteinander vergleichen zu können, gibt es verschiedene Vergleichskoeffizienten, die angewendet werden können. Darunter fällt neben dem Tanimoto Koeffizient, der Euklidische Koeffizient, der Pearson-Koeffizient und der Hamman-Koeffizient. Im Laufe dieser Arbeit wurde der Tanimoto-Koeffizient verwendet, da sich dieser schon in vorangegangenen Arbeiten bewährt hat [136]. Dieser Koeffizient kann einen Wert zwischen null und eins annehmen, wobei gilt: je ähnlicher, desto größer der Wert. Der Tanimoto-Koeffizient T wird folgendermaßen berechnet:

$$T = \frac{N_a \cdot N_b}{N_{ab}},$$

wobei N_{ab} der Anzahl der Bits entspricht, die für jeweils beide Kleinstrukturen gesetzt sind, N_a entspricht der Anzahl der Bits, die nur in dem Bitvektor der ersten Kleinstruktur gesetzt sind und N_b ist die Anzahl der Bits, die nur in dem Bitvektor der zweiten Kleinstruktur gesetzt sind.

In der Arbeit wurde der Fingerprint von CDK oder Openbabel verwendet. Wichtig dabei war jedoch, dass die zu vergleichenden Fingerprints der Moleküle mit Hilfe der gleichen Methode aus der gleichen Version generiert wurden, damit die Fingerprints mit dem Tanimoto-Koeffizienten vergleichbar waren.

2.2.5.2 Zellulärer Fingerprint zum Vergleich von Wirkprofilen

Der zelluläre Fingerprint ermöglicht den schnellen Vergleich der Wirkprofile von Kleinstrukturen. Bei dem zellulären Fingerprint [147] handelt es sich um einen Bitvektor der Länge 4.800, der das Wirkprofil einer Substanz auf die 60 NCI-Zelllinien beschreibt. Für jeden der 60 Zelllinien sind 80 Bits des Fingerprints reserviert. Abbildung 2.9 zeigt beispielhaft ein von der NCI-Datenbank generiertes „Mean“-Diagramm des Anti-Tumor Wirkstoffs Paclitaxel. Die absoluten GI50-Werte wurden in dem zellulären Fingerprint kodiert. In einem ersten Schritt wurden die NCI-Substanzen ausgenommen, die nur eine geringe Zelltyp-Spezifität aufweisen. Die übrig gebliebenen Daten wurden mittels der Z-Normalisierung normalisiert, um die Daten vergleichbar und die Unterschiede in Hinblick auf die Zelllinien-Spezifität sichtbar zu machen. Anschließend wurden die Wirkprofile in dem zellulären Fingerprint kodiert. Auch hier wurde für den Vergleich der Tanimoto-Koeffizient verwendet. Der zelluläre Fingerprint beschreibt die Effekte einer Kleinstruktur auf 60 verschiedenen Tumor-, Lymphom-, Melanom- und Leukämie-Zelllinien.

Es wurde davon ausgegangen, dass Strukturähnlichkeiten von über 85% zu einer ähnlichen Funktionsweise der Kleinstrukturen führen. Unter dieser Annahme stellt sich die Frage, ob eine Korrelation zwischen struktureller und zellulärer Ähnlichkeit vorhanden ist. Dafür wurde eine Gruppierung (Cluster) der NCI-Kleinstrukturen mit ähnlichem zellulären Effekt in Hinblick auf deren strukturellen Ähnlichkeiten durchgeführt.

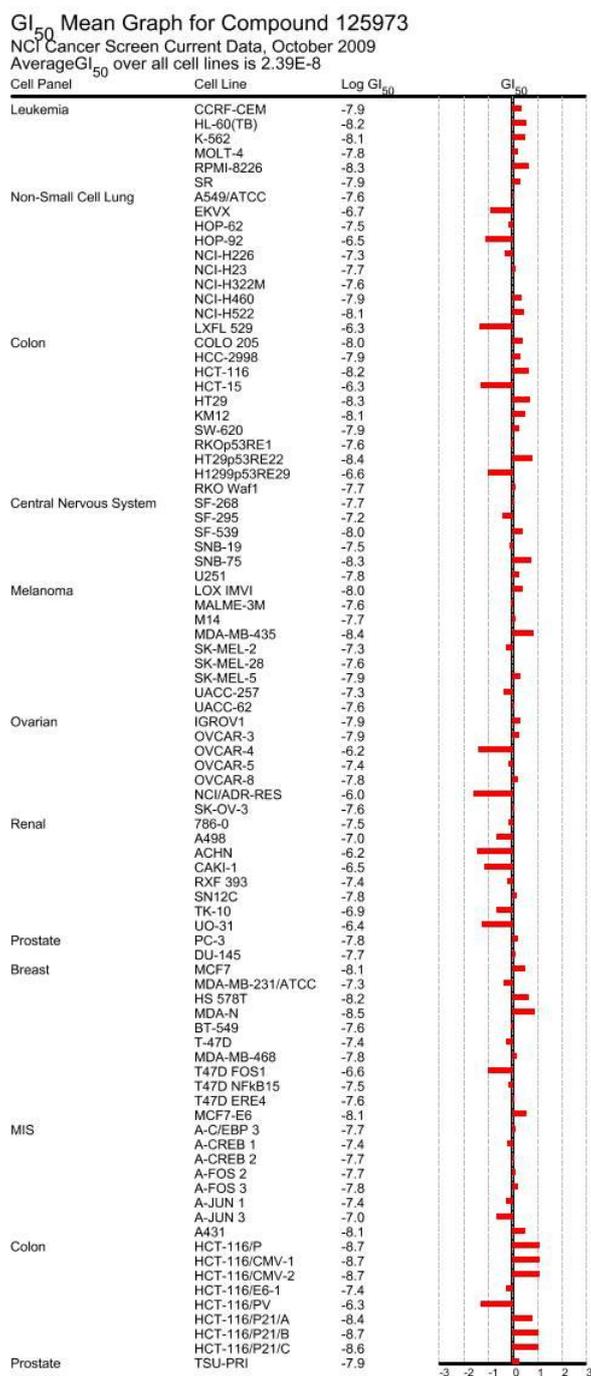


Abb. 2.9: „Mean“-Diagramm für Paclitaxel aus der NCI-Ressource [21]. Die erste Spalte beinhaltet den Zelltyp, die zweite Spalte die Zelllinie und die dritte Spalte den absoluten logarithmierten GI₅₀-Wert. In der letzten Spalte ist der GI₅₀-Wert der einzelnen Zelllinien im Vergleich zum Mittelwert der GI₅₀-Werte über die Gesamtheit aller Zelllinien angegeben.

2.2.6 Target-Vorhersage

Die „Target-Vorhersage“ beschreibt die Suche nach den von Kleinstrukturen adressierten Zielproteinen. In diesem Projekt basierte die Vorhersage auf strukturellen Ähnlichkeiten von Kleinstrukturen zu ATC-klassifizierten (Anatomisch-Therapeutisch-Chemische Klassifikation der WHO) Wirkstoffen. Zusätzlich wurden physikochemische Eigenschaftsprofile berechnet. Anschließend sollte die Targetvorhersage als Webserver mit dem Namen „SuperPred“ publiziert werden [148].

Generierung des Datensatzes:

Die Generierung des Datensatzes für die Targetvorhersage begann mit der Berechnung des strukturellen Fingerprints der 2.500 ATC-klassifizierten Wirkstoffe aus der SuperDrug-Datenbank und dem anschließenden Vergleich mit den Arzneistoffeinträgen der SuperTarget-Datenbank [149], die zusätzlich zu den Substanzinformationen auch Zielproteininformationen speichert. Insgesamt konnten für über die Hälfte der SuperDrug Kleinstrukturen Targetinformationen gefunden werden. Neben den 2.500 Arzneistoffen wurde die Datenbank um weitere 4.000 Substanzen vergrößert, indem strukturell sehr ähnliche Substanzen mit aufgenommen wurden. Für diese Substanzen wurden sowohl die ATC-Klasse als auch die Zielproteine vorhergesagt.

Vorhersage:

Die Vorhersage basierte auf strukturellen Ähnlichkeitsvergleichen. Eine Vorhersage einer ATC-Gruppe einer Kleinstruktur mit unbekanntem Target liefert die ATC-Gruppe und die Zielproteine des ähnlichsten Wirkstoffs. Für die strukturelle Ähnlichkeit wurde die in Kapitel 2.2.2.1 beschriebene Methode verwendet. Es wurde angenommen, dass Strukturen mit mindestens 80% chemischer Strukturähnlichkeit ähnliche Wirkungsweisen besitzen und die ATC-Zugehörigkeit vorhergesagt werden kann. Des Weiteren soll ein physikochemisches Eigenschaftsprofil der Substanzen dabei helfen die orale Bioverfügbarkeit des Stoffs abzuschätzen als auch den vorhergesagten Tanimoto-Score in seiner Aussage zu unterstützen. Dabei bestand die Hypothese, dass Verbindungen häufig ähnliche biophysikalische Aktivität aufweisen, wenn sie auch ähnliche physikalisch-chemische Eigenschaften besitzen.

Das Eigenschaftsprofil einer Substanz besteht hier aus den Eigenschaften der Lipinski 5er Regel. Neben dem Eigenschaftsprofil werden auch Eigenschaftsprofilgrenzwerte angegeben. Diese Grenzwerte sind für jede der Lipinski 5er Regel-Eigenschaften

der niedrigste und höchste Wert der Substanzen einer ATC-Klasse. In SuperPred wird die medizinische Wirkung und der Indikationsbereich mit Hilfe des ATC-Codes ausgedrückt.

Reduzierter Datensatz für die Vorhersagebewertung:

Für die statistische Evaluierung der Vorhersage wurde die Leave-one-out-Methode und ein Subdatensatz aus 1.035 Wirkstoffen verwendet. Bei der Auswahl der einzelnen Wirkstoffe für den Datensatz wurde folgendermaßen vorgegangen:

1. Entfernung der Wirkstoffe mit mehr als einer Indikation
2. jede ATC-Gruppe besteht aus mindestens 3 Molekülen
3. für die Beseitigung von Ausreißern, wurden Substanzen innerhalb eines ATC-Codes entfernt, die um mehr als das 1,5-fache vom durchschnittlichen Tanimoto-Wert abwichen, nachdem jeder Wirkstoff gegen jedes Gruppenmitglied verglichen wurde.
4. Zusammenführung von ATC-Klassen mit ähnlichen Indikationen. Häufig adressieren Arzneistoffe, die zur gleichen Indikationsgruppe gehören auch dasselbe Zielprotein, besitzen jedoch nicht den gleichen ATC-Code, da eine unterschiedliche Applikation vorliegt. Z.B. werden Kortikoide oral verabreicht, intestinal und auch tropfisch. Da jedoch der ATC-Code auch die Applikation beinhaltet, können einige ATC-Codes zusammengefasst werden.

Anschließend wurde systematisch eine Kleinstruktur aus dem Datensatz entfernt, um für diese die ATC-Gruppe vorherzusagen. Das wurde für alle Wirkstoffe des Subdatensatzes durchgeführt, um die Vorhersagegenauigkeit zu evaluieren.

2.3 Integration von Zielproteinen, Wirkstoffen und ihren Interaktionen

Im letzten Abschnitt des Methodenteils werden die Methoden zur Gewinnung der Kleinstruktur-Zielprotein-Interaktionen aus Publikationen sowie die Materialien und Methoden für die Implementierung der Datenbanken und Webserver beschrieben.

2.3.1 Semi-automatisierte Wissensextraktion aus der Literatur

Mit „Textmining“ wird allgemein die automatische Gewinnung von Wissen aus der Literatur oder Textbausteinen mit Hilfe von Computer-gestützten Algorithmen bezeichnet. Im Rahmen dieser Arbeit wurde mit Hilfe dieser Methodik Kleinstruktur-Zielmolekül-Interaktionen, die in PubMed-Publikationen enthalten sind, aufgefunden und in einer Datenbank abgelegt und für weitere Analysen bereitgestellt. Hierbei wurden die über 19 Millionen öffentlich verfügbaren PubMed-Zusammenfassungen (Abstracts) im XML-Format heruntergeladen und mit Hilfe von Lucene ein PubMed-Index über die einzelnen Worte der Abstracts erstellt. Der nächste Schritt bestand in der Generierung einer Namens- und Synonymliste potenzieller Zielproteine und Wirkstoffe, nach denen in dem Index gesucht werden sollte. Die Zielproteinnamen stammten dabei bevorzugt aus der UniProt. Für die Substanznamen wurden die freien Datenbanken PubChem, DrugBank und SuperDrug [150] und teilweise die kommerziellen Datenbanken CMC (Comprehensive Medicinal Chemicals) und der WDI (World Drug Index) extrahiert. Für die Suche nach onkologisch relevanten Interaktionen mittels Textmining wurden nur Wirkstoffe in die Liste einbezogen, die z.B. als „anti-neoplastic“, „cancer“ gekennzeichnet oder in die ATC-Klasse L0 abgelegt sind. Für die Suche von Interaktionen auf dem PubMed-Index und um die Wahrscheinlichkeit einer echten Kleinstruktur-Proteininteraktion zu erhöhen, wurde neben der Namensliste noch eine zweite Liste mit Schlagworten, wie z.B. „inhibition“ und „binds“ generiert. Die Einträge beider Listen wurden anschließend in dem erstellten PubMed-Index gesucht und es entstand eine Ergebnisliste aus Substanz-Zielprotein-Interaktionen, die anschließend manuell validiert wurde, um die falsch-positiven Treffer auszusondern. Dafür wurde eine Website erstellt, die potenzielle Kleinstruktur-Protein-Interaktionen angibt, wobei die gewünschten MeSH-Terms, das Protein und die Kleinstruktur im Abstract farblich markiert waren. Ein möglicher Treffer einer Wirkstoff-Zielprotein-Interaktion wurde jedoch nur angezeigt, wenn in einem Abstract mindestens ein Wirkstoff und ein Zielprotein gefunden wurden.

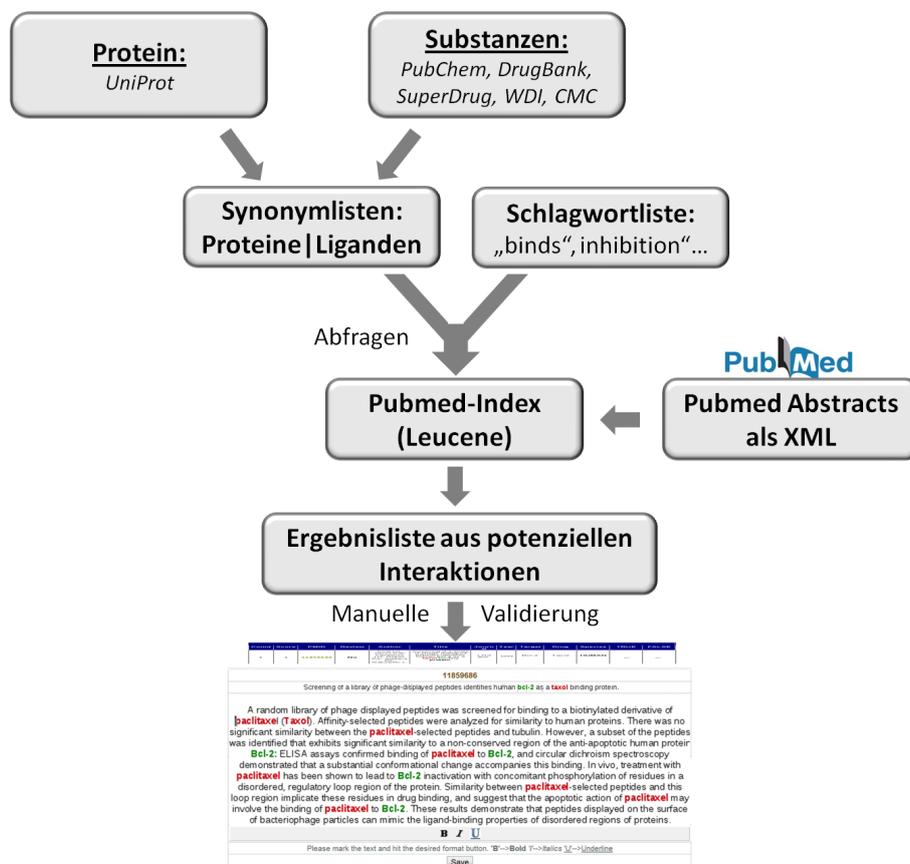


Abb. 2.10: Textmining Fließdiagramm.

2.3.2 Datenakquise für die Integration von Kleinstrukturen und Zielproteinen

Für die Integration von Kleinstrukturen und deren Zielproteine sind verschiedene Methoden bzw. Datenbanken angewandt worden.

SuperTarget - Integration von Arzneistoffen und ihren Zielmolekülen

Zielproteine:

Die UniProt-Datenbank beinhaltet über drei Millionen verschiedene Proteine, die jeweils ein potenzielles Zielprotein darstellen. Diese Proteine wurden heruntergeladen und die Informationen wie IDs, Sequenz, Organismus, Synonyme und PDB-ID selektiert. Die dreidimensionalen Proteinstrukturen wurden aus der PDB selektiert.

Arzneistoffe:

Für die Wirkstoffe wurde die SuperDrug-Datenbank extrahiert, die ca. 2.500 ATC-klassifizierte Wirkstoffe enthält. Hierbei wurden alle Informationen wie CAS-Nummer, Name, Synonyme und Struktur selektiert. Für die PDB-Kleinstruktur ID wurde die SuperLigands-Datenbank [151] verwendet. Dabei halfen Ähnlichkeitssuchen ausgehend von den Wirkstoffen der SuperDrug in der SuperLigands-Datenbank nach identischen Treffern. Weitere Synonyme wurden mit einer Ähnlichkeitssuche in der PubChem-Datenbank erworben.

Wirkstoff-Zielprotein-Interaktionen:

Neben der durch Textmining Methoden identifizierten Interaktionen wurden auch die in verschiedenen Datenbanken annotierten Interaktions-Informationen integriert. Hier spielten vor allem die Datenbanken TTD (Therapeutic Target Database) und die DrugBank eine Rolle. Die TTD beinhaltet neben den Interaktionen auch die zugehörigen Literaturreferenzen, während diese bei der DrugBank nur vereinzelt vorliegen. Hier wurde in der SuperTarget auf die DrugBank verwiesen. Während die Informationen aus der TTD einzeln abgefragt werden mussten, konnte die DrugBank im Textformat heruntergeladen werden.

Signalwege:

Für die Signalwege wurden die ca. 200 Signalwege aus der KEGG-Datenbank extrahiert. Die Targetmarkierungen in den Signalwegen wurden mit Hilfe von SOAP (Simple Object Access Protocol) durchgeführt. Dieses Protokoll erlaubt den Datenaustausch zwischen Systemen. Dabei wurden die markierten Signalwege generiert, indem ein von der KEGG zur Verfügung gestellter Server mit der KEGG-SignalwegsID und der zu markierenden TargetID befragt wurde. Das Resultat ist der angefragte Signalweg mit mindestens einem markierten Target.

Ontologien:

Die Ontologie-Begriffe stammten aus der GeneOntology-Datenbank (GO). Die GO-Datenbank wurde exportiert und die Verlinkung zu den Zielproteinen erstellt. Dafür wurden bei dem Herunterladen der Targets aus der UniProt auch die dazugehörigen GO-Terme und Links selektiert.

Cytochrome:

Die Cytochrome P450 stammten aus der Arbeitsgruppe Flockhart der klinischen Physiologie der Indiana University, die eine Tabelle bereitstellen mit Cytochrome P450, deren Substrate und Inhibitoren

(<http://medicine.iupui.edu/clinpharm/ddis/table.asp>) und viele Studien geleitet haben [152]. Auch die DrugBank-Datenbank beinhaltet Cytochrom P450 Informationen, die ausgelesen wurden. Die Proteinsequenzen stammten aus der UniProt.

Nebenwirkungen:

Die Nebenwirkungen stammten aus dem „Canadian Adverse Drug Reaction Information System“ (CADRIS). Es handelt sich hierbei um eine Kollektion von Nebenwirkungen verschiedener Wirkstoffe, die bei Patienten beobachtet wurden. Dabei werden neben den beobachteten Nebenwirkungen weitere Angaben wie Alter und Geschlecht zur Person angegeben. Diese Daten konnten auf der „Health Canada“ Website (http://www.hc-sc.gc.ca/dhpmpps/medeff/advers-react-neg/fs-if/cadris-2_e.html) exportiert und integriert werden.

CancerResource - Integration von onkologisch relevanten Kleinstrukturen und ihren Zielmolekülen:

Zielproteine:

Der Datensatz für Proteine, die bei der Tumorentstehung eine Rolle spielen und damit potenzielle Targets bei der Entwicklung neuer Therapien darstellen, stammt aus verschiedenen Ressourcen wie UniProt und Ensembl. Die Proteine aus dem Schimpansen-Mensch-Vergleich (Kapitel 2.1.1) lieferte 493 „Startzielproteine“. Nach Bestimmung der für die Tumorentstehung essentiellen Signalwege aus der KEGG-Datenbank, wurden die in den Signalwegen mitwirkenden Proteine als zusätzliche potenzielle onkologische Targets definiert. Auch Proteine mit gebundenen onkologisch relevanten Wirkstoffen in der PDB-Datenbank stellten Zielproteine dar. Zur Vervollständigung des Datensatzes trug die Textmining-Methode bei.

Kleinstrukturen:

Im Laufe dieser Arbeit umfasst der Begriff onkologisch relevante Kleinstrukturen, alle Substanzen mit Anti-Tumor Effekt. Dabei muss es sich bei der Substanz nicht um einen schon entwickelten und in der Tumorthherapie angewandten Wirkstoff handeln. Den größten Anteil der Kleinstrukturen lieferte die DTP (Developmental The-

rapeutics Program) des NCI. Die offizielle Website ermöglicht das Herunterladen von über 40.000 Substanzen mit inhibitorischem Effekt auf mindestens eine der 60 NCI-Zelllinien. Die Selektion der Substanzen mit ATC-Klassifikation L0 (Anti-neoplastische Substanzen) aus der SuperDrug lieferte weitere onkologisch relevante Kleinstrukturen. Darüber hinaus wurden die kommerziellen Datenbanken CMC und WDI nach Schlagwörtern in der Klassifikation durchsucht. Zu diesen Schlagwörtern fielen u.a. „anti-neoplastic“, „cancer“, „tumor“, „cell-cycle“, „angiogenesis“ etc. Eine anschließende Ähnlichkeitssuche in der PubChem nach identischen Strukturen lieferte die endgültigen Kleinstrukturen für die Liste.

Kleinstruktur-Zielprotein-Interaktionen:

Die onkologisch relevanten Interaktionen stammten neben der Literatursuche mit Textmining-Methoden auch aus wichtigen Interaktionsdatenbanken wie TTD und DrugBank. Weitere direkte, aber auch indirekte Interaktionen wurden aus der CTD gewonnen. Hier sind auch Substanzen abgelegt, die die Target-Expression beeinflussen. PharmGKB liefert zusätzlich Informationen über Gene, die die Dynamik und Kinetik von Wirkstoffen beeinflussen.

Signalwege:

Für die Signalwege, die bei der Tumorentstehung eine Rolle spielen, wurde die KEGG-Datenbank manuell durchsucht. Dabei wurden insgesamt 44 Signalwege identifiziert, die im Zusammenhang mit der Tumorentstehung stehen. Für die Markierung der Zielproteine wurde SOAP verwendet.

Expressionsdaten:

Die Expressionsdaten wurden für die ca. 40.000 NCI-Substanzen ebenfalls aus der DTP der NCI extrahiert, verarbeitet und für den Vergleich mit externen Daten normalisiert. Dabei wurde eine Median-Normalisierung für alle Affymetrix U133A „Probe Sets“ für jede der einzelnen NCI60-Zelllinien durchgeführt. Anschließend wurde die relative Häufigkeit über den Zelllinien Expressionsvektor für jede Affymetrix „Probe Set“ berechnet. Gene, die für multiple Gene standen wurden ausselektiert. Der Durchschnitt wurde berechnet, wenn ein multiples ProbeSet nach Ensembl zu einem Gen gehört [[153] Anhang].

Die Connectivity Map [154] Informationen wurden integriert, um funktionelle Verbindungen zwischen Krankheiten und Wirkstoffwirkung zu entdecken. In diesem

Projekt sind die Einflüsse von mehr als 200 Substanzen auf fünf verschiedenen onkologischen und hämatologischen Zelllinien, die auch im NCI60-Screening Projekt vorliegen, bestimmt und untersucht [155].

2.3.3 Implementierung der Datenbanksysteme

Datenbanksystem für Arzneistoff-Zielprotein-Interaktionen:

Die Webseiten für die Arzneistoff-Zielprotein-Interaktions-Datenbank ist mit JavaScript, Ajax und HTML implementiert. Das JavaScript bietet die Möglichkeit Eingaben sofort zu verarbeiten und die Webseite dementsprechend anzupassen ohne die vollständige Webseite neu laden zu müssen. Die zugrundeliegende Datenbank wurde mit MySQL erstellt. Der Webzugriff wurde durch einen Tomcat Webserver ermöglicht.

Weitere Datenbanksysteme und Webserver:

Die erstellten relationalen Datenbanksysteme für den Targetvorhersage-Webserver, die Fragment-Datenbank und die Datenbank mit Tumor-relevanten Proteinen und Wirkstoffen basieren jeweils auf MySQL. Um die Zugänglichkeit der Informationen zu gewährleisten, wurden für die Webseitenerstellung die Programmiersprachen PHP5, Ajax und HTML verwendet. Ziel war es, einen dynamischen und modernen Webauftritt zu generieren und eine optimale Verknüpfung zur Datenbank zu ermöglichen. PHP (Hypertext Preprocessor) ist eine serverseitig interpretierte Skriptsprache, welche die dynamische Generierung von HTML (Hyper Text Markup Language) Code und die Einbindung der MySQL-Datenbank erlaubt. Durch die Verwendung von Ajax (Asynchronous JavaScript and XML) wird eine komfortable Interaktion zwischen Datenbank und User ermöglicht. Im Rechenzentrum der Charité wurde ein virtueller Server konfiguriert.

Für die Berechnung von chemischen Eigenschaften wurde das freizugängliche Paket MyChem/OpenBabel installiert und verwendet. Der Webzugriff wird durch einen Apache Webserver 2.2. ermöglicht. Der Benutzer besitzt die Möglichkeit Kleinstrukturen hochzuladen bzw. diese zu zeichnen. Hierfür wurde in den Webseiten das Programm MarvinSketch integriert. Für die dreidimensionale Visualisierung der Moleküle in den Datenbank- und Webserver-Systemen wurde Jmol verwendet.

Kapitel 3

Ergebnisse

3.1 Zielproteine

3.1.1 Evolutionäre Aspekte

Das Schimpansen-Genom besteht aus 24 Chromosomen, während das humane Genom nur 23 Chromosomen besitzt [8]. Dabei liegen im Schimpansen-Genom zwei separate Chromosomen 2a und 2b vor, wobei das humane Genom ein fusioniertes Chromosom 2 besitzt. Darüber hinaus entstanden während der Evolution neun perizentrische Inversionen, wobei das Zentromer eingeschlossen ist [8]. Eine dieser Inversionen erfolgte auf Chromosom 18 und ist in Abbildung 3.1 dargestellt.

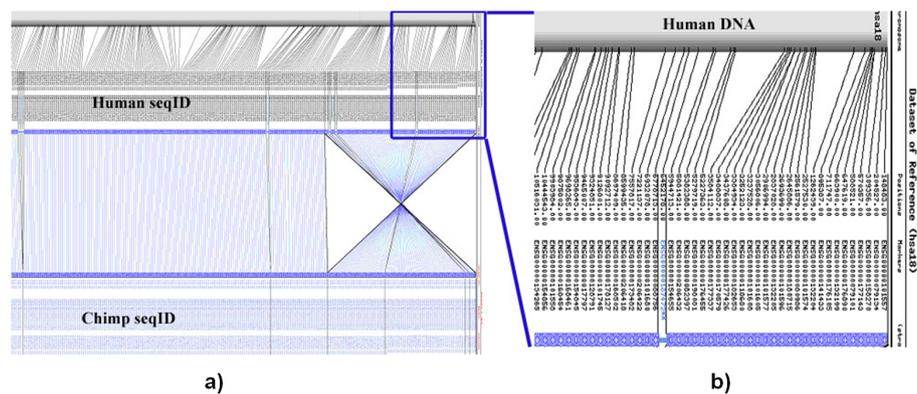


Abb. 3.1: Vergleich des humanen und des Schimpansen-Chromosoms 18.

Der Vergleich zwischen den humanen und orthologen Genen des Schimpansen-Genoms wurde mit dem Programm AutoGraph [108] erstellt. a) oben: die humanen Gene auf Chromosom 18; unten: die orthologen Gene des Schimpansen auf Chromosom 18. Deutlich zu sehen ist die perizentrische Inversion. b) zeigt die Vergrößerung eines Ausschnitts des humanen Chromosoms 18. *Aus Ahmed et al. [156].*

Neben diesen bekannten und offensichtlichen Veränderungen zwischen humaner und Schimpansen-Genstruktur sollten in dieser Arbeit weitere evolutionäre Veränderungen des Genoms analysiert werden, die für eine veränderte Tumorzinzidenz zwischen beiden Spezies verantwortlich sein könnten. Die Analyse basierte im Jahr 2007 auf einen Datensatz verschiedener Proteine und Gene, die in den verschiedenen KEGG-Signalwegen der DNA-Reparatur, Apoptose und des Zellzyklus eine Rolle spielen. Darüber hinaus vervollständigte die Kooperationsarbeitsgruppe von Prof. Dr. Peter Daniel den Datensatz, so dass ein Analyse-Set von 493 verschiedenen Proteinen entstand.

3.1.1.1 Schimpanse vs. Mensch

Chromosomen-Organisation und Genstruktur

Die Lokalisationen der annotierten, kodierenden Sequenzen der 493 Proteine von denen vermutet wurde, dass sie für die unterschiedliche Tumoranfälligkeit verantwortlich sind, waren über die Chromosomen des Human- und Schimpansen-Genoms verteilt aufzufinden. Die Verteilung der Gene auf den Chromosomen war bei beiden Spezies jedoch nicht identisch (siehe Abbildung: 3.2).

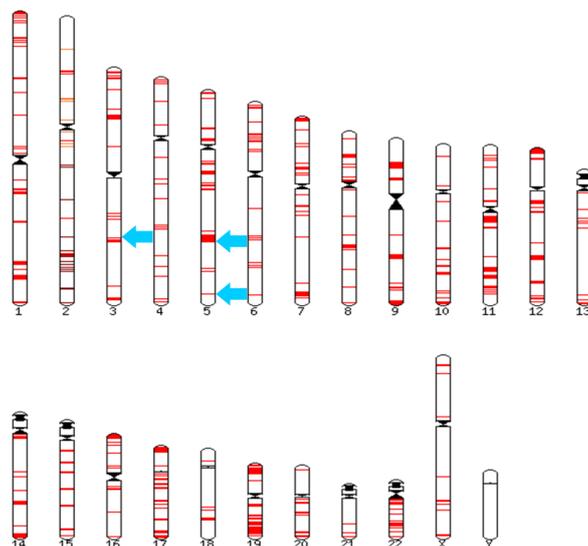


Abb. 3.2: Lokalisationen der CDS der Proteine auf den Chromosomen.

Die rot-markierten Abschnitte definieren die Gene, welche für eines der 493 Proteine transkribiert werden und sich bei beiden Spezies auf den gleichen Chromosomen befanden. Die blau-angezeigten kodierenden Sequenzen befanden sich bei beiden Organismen auf unterschiedlichen Chromosomen. Aus Ahmed et al. [156].

Im Einzelnen sind die unterschiedlichen Genlokationen in Tabelle 3.1 aufge-

führt. Es existierte eine annotierte kodierende Sequenz (CDS) für das „Ring Finger Protein 7“ in den jeweiligen Genomen. Jedoch befand sich das orthologe Gen im Schimpansen-Genom auf Chromosom 7 und nicht wie im humanen Genom auf Chromosom 3. Es existierte jedoch auch auf dem Schimpansen-Chromosom 3 eine Region, die für dieses Protein kodieren könnte. Sie war jedoch nicht als kodierende Sequenz für das Protein annotiert. Es gab keine ähnliche Region auf Chromosom 7 des humanen Genoms. Auch für die anderen beiden Proteine aus der Tabelle konnten keine ähnlichen kodierenden Regionen auf den korrespondierenden Chromosomen identifiziert werden.

Protein	UniProt Name	Chromo (HS)	Chromo (PT)
Ring Finger Protein 7	RBX2	3	7
Replication Protein C	PP2AA	5	X
Nucleophomin	NPM	5	16

Tab. 3.1: Proteine, deren kodierende Sequenz bei Mensch und Schimpanse auf verschiedenen Chromosomen annotiert waren. *Aus Ahmed et al. [156].*

HS: Homo Sapiens; PT: Pan Troglodytes

Für die Analyse der Genstruktur wurden die Exon- und Intron-Struktur und Längen der für die Proteine kodierenden Gene betrachtet. Der erste Schritt bestand in der Bestimmung der Anzahlen der Exons, die für die Transkription der mRNAs der 493 Proteine verantwortlich sind. Insgesamt wurden 300 Gene in die Analyse der Genstruktur mit einbezogen, da für diese eine annotierte humane CDS und eine nahezu 100% sequenzähnliche annotierte Schimpansen-CDS identifiziert werden konnte. Die Analyse ergab, dass ca. 5% der 300 mRNAs in beiden Spezies durch verschiedene Anzahlen an Exons kodiert wurden. Ein Drittel der betroffenen mRNAs wurde im menschlichen Organismus von mindestens einem zusätzlichen Exon kodiert, während zwei Drittel der mRNAs von mindestens einem zusätzlichen Exon im Schimpansen-Genom, verglichen mit dem humanen Genom, kodiert wurden. Um einen groben Überblick über die Intron Struktur zu erhalten, wurden die Intronlängen der humanen und Schimpansen-Gene bestimmt und verglichen. 84 Gene besaßen mindestens ein Intron mit mindestens 50 Basenpaare Längenunterschied zwischen den Spezies. Insgesamt beinhalteten die Schimpansen-Gene längere Introns verglichen mit den humanen Genen (siehe Abbildung: 3.3).

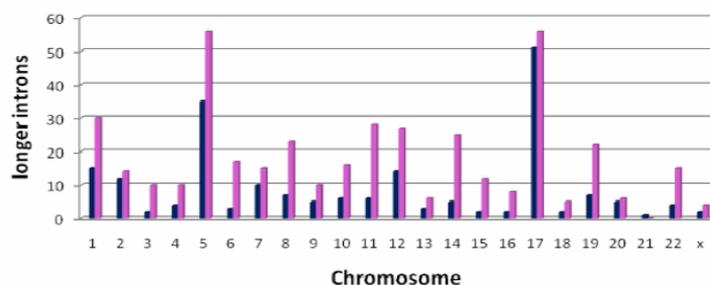


Abb. 3.3: Lokalisation der 84 Gene mit längeren Introns auf den Chromosomen. Im Durchschnitt besaßen die Gene des Schimpansen längere Introns als die humanen Gene. Blau: Anzahl der längeren Introns pro humanem Chromosom; Magenta: Anzahl der längeren Introns pro Schimpansen-Chromosom. Aus Ahmed et al. [156].

SNPs und Positive Selektion

Im Durchschnitt wies der Gen-Sequenzdatensatz, beim Vergleich der humanen und Schimpansen-Gene, neun Nukleotidänderungen pro Gen auf.

Um Gene zu identifizieren, die sich schneller entwickelten als andere wurde der Ka/Ks -Wert der 493 Proteine ermittelt. Dabei besaßen positiv selektierte Proteine einen Ka/Ks -Wert größer 1. Insgesamt waren 13 Proteine, einschließlich des BRCA1-Gens, positiv selektiert.

Protein	UniProt Name	Ks/Ka -Wert
Thioredoxin-like protein p46	TXND5	3,13
Immediate early protein GLY96	IEX1	2,13
Poly[ADP-ribose]polymerase 3	PARP3	2,06
dUTP pyrophosphatase	DUT	1,70
TNFR superfamily member 10D	TR10D	1,67
TNFR superfamily member 18	TNR18	1,43
Breast cancer type 1 susceptibility protein	BRCA1	1,28
TNFR superfamily 10A	TR10A	1,21
Caspase-5	CASP5	1,21
VEGF A	VEGFA	1,17
Myc proto-oncogene protein	MYC	1,14
Serpin B9	SPB9	1,02
Transcription factor E2F1	E2F1	1,00

Tab. 3.2: Proteine mit einem Ka/Ks -Wert größer 1 und damit potenziell positiv selektiert sind. Aus Ahmed et al. [156].

Pseudogenisierung

Ein weiterer wichtiger Prozess ist die Pseudogenisierung von Genen bzw. Genberei-

chen. Die Analyse zeigte, dass 49 der 493 Gene mögliche Pseudogene im Schimpansen-Genom darstellen, während sie aktive Gene im Humangenom sind. Eines dieser Pseudogene stellte der Transkriptionsfaktor E2F3 dar, welcher in der Kontrolle des Zellzyklus involviert ist. Das Protein war in der RefSeq Annotation als Pseudogen annotiert (Stand: 2007). Die Ensembl Datenbank hingegen listete dieses Gen als funktionierendes Gen auf, was durch eine Insertion eines Cysteins erwirkt wurde.

Proteinsequenz- und Strukturanalyse:

Es wurden acht Substitutionen pro Protein identifiziert. Insgesamt fanden während der Evolution in den ca. 500 Proteinen 4.111 einzelne Aminosäure-Substitutionen und 288 Insertionen/Deletionen statt. Die Proteinsequenzidentität beträgt bei den onkologisch relevanten Proteinen 96%. In Tabelle 3.3 sind Proteine aufgelistet, in denen Pattern von bestimmten Domänen durch eine Mutation verändert wurden. Beispielsweise, ist das BH2-Motif des Bcl-2 Proteins B2La1, welches die Apoptose inhibiert, durch eine Mutation verändert.

UniProt Name	Motiv	Human	Pan
IFNA2	Interferon	X	
MIF	Macrophage migration inhibitory factor	X	
SSR3	Microbodies C-terminal targeting signal		X
TXND5	Thioredoxin family active site	X	
VEFFA	Platelet-derived growth factor family signature	X	
B2LA1	BH2 motif	X	
BRCA2	ATP binding motif	X	
ERCC3	Helicase ATP-binding type domain profile	X	
ERN1	SER/THR protein kinase activve-site signature	X	
TBB2C	Tubulin beta mRNA autoregulation signal	X	

Tab. 3.3: Motive, die in der Schimpansen- oder humanen Proteinsequenz identifiziert wurden. X: Dieses Motiv besaß nur das Protein des markierten Organismus. *Aus Ahmed et al. [156].*

Insgesamt existierten 209 Proteinstrukturen der 493 Proteine in der PDB für die dadurch eine Analyse der Substitutionsauswirkungen auf die Funktion der Struktur möglich war. Für 59 Proteine lag eine Substitution in einer Protein-Protein Interaktionsstelle und für 10 Proteine in einer Liganden-Bindungstasche vor. Ein Beispiel für eine Veränderung einer Protein-Bindungstasche durch eine Mutation ist die Topoisomerase II alpha. Die Proteinstruktur der ATPase-Domäne der Topoisomerase mit dem PDB-Code 1ZXM ist in der PDB vorhanden. In diesem Fall verändert sich die Bindungstasche bzgl. zwei Aminosäuren. Die Topoisomerase II alpha ist ein

Schlüsselenzym, das bei der DNA-Replikation eine Rolle spielt und zusätzlich in der Entstehung des Brustdrüsenkarzinoms involviert ist.

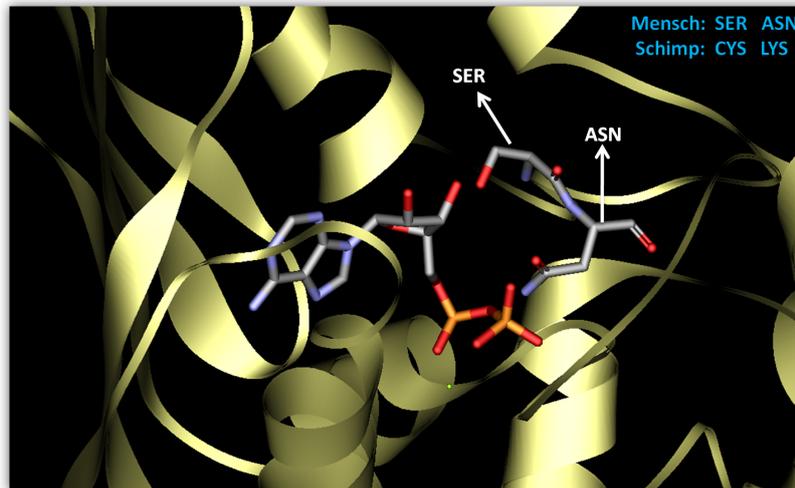


Abb. 3.4: ATPase Domäne der humanen Topoisomerase II alpha mit gebundenem ATP Molekül.

Die humane ATPase Domäne unterscheidet sich in zwei Aminosäuren von der des Schimpansen. Während bei der humanen Domäne ein Serin und ein Asparagin an der Bindung des Liganden beteiligt sind, befindet sich beim Schimpansen an diesen Positionen ein Cystein und ein Lysin. *Aus Ahmed et al. [156].*

Zusammenfassung

Die folgende Zusammenfassung bezieht sich nur auf die Analyse der ca. 500 Gene bzw. Proteine.

Gene und Genstruktur:

- im Durchschnitt neun Mutationen pro Gen
- 49 mögliche Pseudogene im Schimpansen-Genom, aber aktive Gene im humanen Genom
- 5% der 300 mRNAs wurden in Mensch und Schimpanse von unterschiedlichen Anzahlen an Exons kodiert
- 84 humane Gene besaßen unterschiedliche Intronlängen verglichen mit den orthologen Schimpansen Genen

Proteinstruktur- und funktion:

- acht Aminosäuremutationen pro Protein
- 209 Proteinstrukturen vorhanden in der PDB
- 59 Proteine wiesen mögliche Schäden in Protein-Protein-Interaktionsstellen auf
- 10 Proteine wiesen Veränderungen in der Protein-Liganden-Bindungsstelle auf

3.2 Wirkstoffe

3.2.1 Membranproteine als Zielmolekül

3.2.1.1 Arbeitsmodell zur γ -Sekretase

Aufgrund der entscheidenden Rolle der γ -Sekretase in der Entstehung von Tumoren und Niereninsuffizienz ist die Suche nach spezifischen potenziellen Wirkstoffen, die bevorzugt die Inhibition des Notch-Schnitts durch die γ -Sekretase bewirken, von hoher Relevanz in der Entwicklung neuer Therapiemöglichkeiten. Hierfür wurden Inhibitoren der γ -Sekretase mit der in Kapitel 2.2.4 beschriebenen Analogie-Modellierung entwickelt.

Modellierung des Presenilins (PS1) der γ -Sekretase-Struktur:

Die γ -Sekretase ist ein Proteinkomplex, bestehend aus den folgenden vier Untereinheiten: der katalytischen Domäne Presenilin (PS1 oder PS2), Nicastrin, Aph-1 und Pen-2. Für die Hydrolyse der Substrat-Peptidbindung sind zwei aktive Aspartatreste in der Bindungstasche des Presenilins verantwortlich.

Die Anwendung des Struktur-basierten Wirkstoffdesigns bedingt das Vorhandensein einer 3D-Struktur, die jedoch zum Zeitpunkt der Projektbearbeitung nicht in der PDB vorhanden war. Daher bestand der erste Schritt in der Identifizierung eines Templates mit mindestens zehn Helices für die Modellierung der katalytischen Untereinheit Presenilin-1 mit Hilfe der Faltungserkennung.

Nach der Templatestruktursuche stellte sich das Translokon mit der PDB-ID 2YXR als geeignetes Template für die Modellierung der Presenilin-Struktur heraus. Das Translokon ist ein Membranproteinkomplex, welcher während der Biosynthese für den Protein-Transport in das endoplasmatische Retikulum (ER) und für den Export

fehlgefalteter Proteine aus dem ER zum Abbau verantwortlich ist [157]. Durch die Verwendung der Bindungstaschen-Struktur der β -Sekretase und deren Einfügen in die Ausrichtung des Translokons sollte eine optimierte Modellierung der Bindungstasche des Presenilins gewährleistet werden, da es sich bei beiden Sekretasen um Proteasen mit zwei Aspartatresten in der katalytischen Bindungstasche handelt. Im Rahmen der Modellierung wurde anschließend die Sequenz durch die der Presenilinsequenz ersetzt.

Abschließend folgte eine Überführung der dreidimensionalen Struktur des Presenilins in die energetisch günstigste Konformation durch Energieminimierung. Abbildung 3.5 zeigt die in der Arbeitsgruppe vorhergesagte bzw. modellierte Presenilin-1-Struktur.

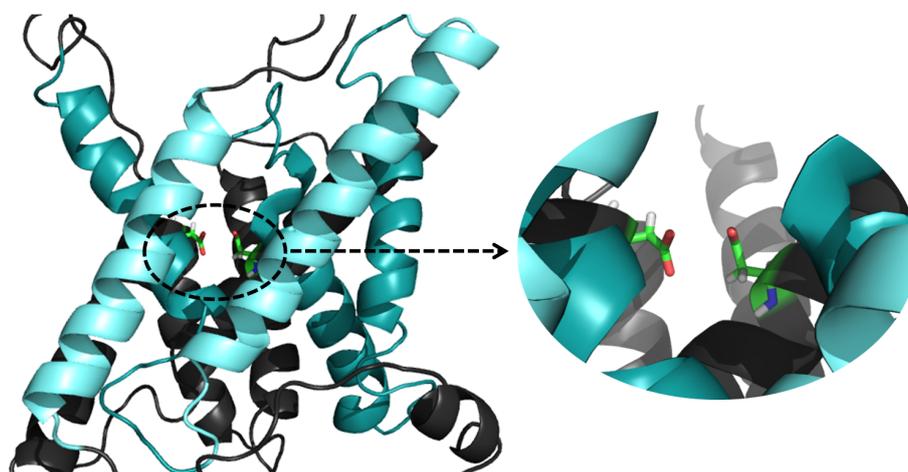


Abb. 3.5: Modellierter Struktur des Presenilin-1.

Presenilin-1, wobei die zwei katalytischen Aspartat-Seitenketten in der Bindungstasche in Stick-Darstellung markiert sind (Asp257, Asp385).

Inhibitorensuche:

Nach der Modellierung der Struktur folgte der erste Zyklus der Inhibitorensuche, beginnend mit der Suche nach bekannten γ -Sekretase Inhibitoren, die als Leitstruktur Verwendung finden konnten. Eine manuelle Literatursuche in PubMed und Patentdatenbanken lieferte zwölf verschiedene Inhibitoren, die mit Hilfe der in Kapitel 2.2.2 beschriebenen *in silico* Screening Verfahren optimiert werden sollten. Abbildung 3.6 zeigt eine Liste der zwölf Leitstrukturen, wobei unter anderem das DAPT, ein bekannter Sekretase Inhibitor, zu finden ist.

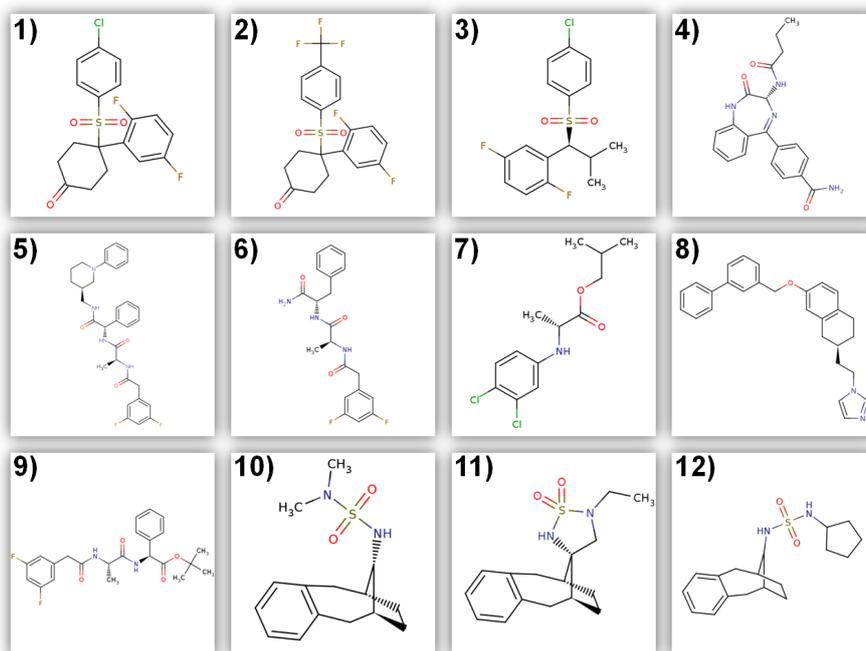


Abb. 3.6: Bekannte γ -Sekretase Inhibitoren, die als Leitstrukturen dienen [158, 159, 160, 161].

Die Kleinstruktur 9) zeigt den bekanntesten Sekretase Inhibitor DAPT.

Ausgehend von zwölf Leitstrukturen folgte eine 2D-Suche in der Substanzdatenbank nach chemisch ähnlichen Strukturen mit einem Tanimoto-Koeffizienten von mindestens 0.85. Zur Identifikation von „Scaffold-Hoppern“ führte die Ausführung einer zusätzlichen 3D-Suche. Ein Beispiel für die Notwendigkeit einer zusätzlichen 3D-Ähnlichkeitssuche ist in Abbildung 3.7 dargestellt. Dabei handelt es sich um eine Kleinstruktur, die aufgrund eines zu niedrigen 2D-Scores ausgesiekt wurde, jedoch durch einen hohen 3D-Score doch in die Liste der potenziellen Inhibitoren mit aufgenommen wurde. Eine spätere experimentelle Validierung belegte, dass es sich bei dieser Kleinstruktur um einen neuen potenziellen Arzneistoff zur Inhibition der γ -Sekretase handelt. Dieses Ergebnis zeigt, dass das 2D-Screening, welches bevorzugt während der Wirkstoffsuche verwendet wird, potenzielle Wirkstoffkandidaten aussiektiert und eine Kombination aus 2D- und 3D-Ähnlichkeitssuchen von Vorteil ist. Insgesamt ergab der erste *in silico* Screening-Durchlauf eine Inhibitorenliste bestehend aus mehreren hundert Kleinstrukturen.

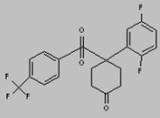
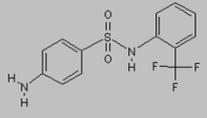
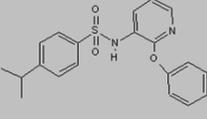
Ausgangs-Substanz	"Hit"	3D score	2D score
	 RJF02082	90,9	87,4
<small>g_inhib_sulphon_ex41_rh02015</small>	 RH02015	96,4	76,9

Abb. 3.7: Beispiel des Sekretase Screenings nach 2D- und 3D-Ähnlichkeitssuche.

Die Abbildung zeigt die zweite Ausgangssubstanz die als Leitstruktur verwendet wurde und zwei Beispieltreffer mit dem 2D- und dem 3D-Score. Bei ausschließlicher Verwendung des 2D-Screenings können potenzielle Kleinstrukturen mit anderer Topologie verloren gehen (hier: zweite Substanz). Diese Scaffold-Hopper werden durch das 3D-Screening identifiziert. Dieser Hit zeigte auch bei der experimentellen Validierung ein gutes Inhibitionsergebnis.

Die identifizierten potenziellen γ -Sekretase Inhibitoren wurden anschließend nach der Lipinski 5er Regel bzgl. der oralen Bioverfügbarkeit gefiltert und in die modellierete Presenilin Bindungstasche positioniert (gedockt), um eine reduzierte Liste experimentell testen zu lassen. Die Liste bestand aus ca. 25 verschiedenen Kleinstrukturen.

Mit Hilfe der experimentell positiv validierten Inhibitoren wurde die Bindungstasche optimiert und ein neuer *in silico* Screening Zyklus mit den wirksamsten Inhibitoren als Leitstruktur eingeleitet. Der zweite *in silico* Screening Durchlauf ermittelte noch weitere 15 Wirkstoffkandidaten für die experimentelle Testung. Insgesamt zeigten 12 Kleinstrukturen eine Wachstumsinhibition im mikromolaren Bereich.

Mit Hilfe der Analogie-Modellierung wurde eine Proteinstruktur der katalytischen Bindungstasche der γ -Sekretase und potenzielle Inhibitoren vorhergesagt. Die experimentell bestätigten Inhibitoren besaßen jeweils ein ähnliches Gerüst: symmetrisch, drei aromatische Ringsysteme, ein polares und ein apolares Ende. Abbildung 3.8 zeigt sechs experimentell bestätigte und damit potenzielle Arzneistoffkandidaten.

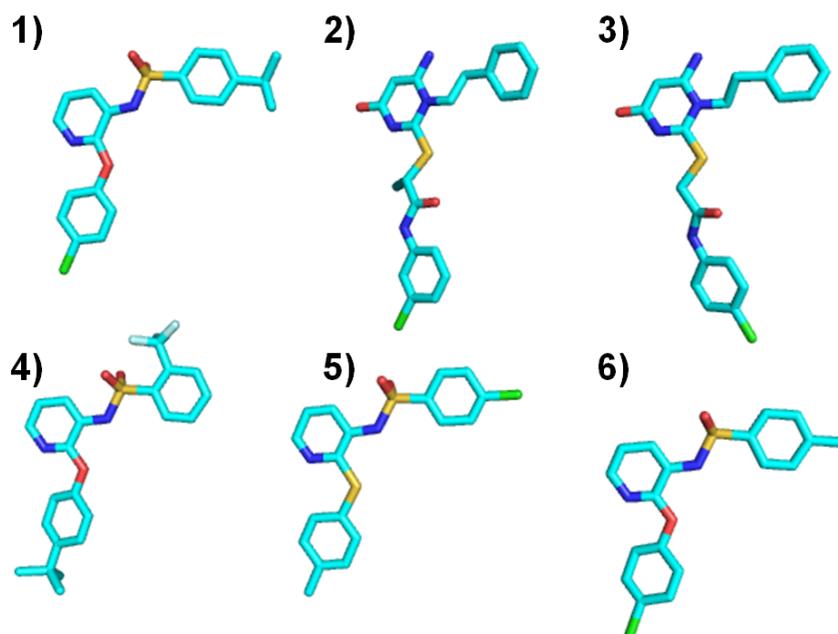


Abb. 3.8: γ -Sekretase Inhibitoren.

Sechs potenzielle Arzneistoffkandidaten mit experimentell nachgewiesener inhibitorischer Wirkung. Sie besitzen alle ein polares und ein apolares Ende. Grün: Chlor, eisblau: Fluor

ADMETox-Analysen:

Während der ADMETox-Analysen wiesen nicht alle der vorhergesagten Eigenschaften auf eine gute orale Bioverfügbarkeit der Substanzen hin. Alle Substanzen verstoßen maximal gegen eine der Lipinski 5er Regel (siehe Tabelle 3.4), was auf eine gute orale Bioverfügbarkeit hinweist.

Inhibitor	MW	H-don.	H-acc.	clogP	PSA
Inh 1	402	1	5	6,6	76,67
Inh 2	428	2	7	4,4	115
Inh 3	414	2	7	4,1	115
Inh 4	450	1	5	7,1	76,67
Inh 5	390	1	5	6,1	92,74
Inh 6	374	1	5	5,7	76,67

Tab. 3.4: Lipinski-Eigenschaften und die „Polar Surface Area“ (PSA).

Die Inhibitoren brechen jeweils nur eine Lipinski-Regel. Der berechnete clogP-Wert, der fünf nicht überschreiten sollte, ist hier oft zu hoch.

Vier Substanzen besitzen einen erhöhten berechneten logP-Wert, was auf eine mäßige Wasserlöslichkeit hindeutet. Tabelle 3.5 zeigt die ADMET-Eigenschaften der sechs Wirkstoffkandidaten, die mit dem Accelrys „ADMET-Descriptors“-Protokoll

abgeschätzt wurden. Für alle sechs Inhibitoren wurde eine gute bis moderate humane intestinale Absorptionsfähigkeit und sehr gute Fähigkeit Plasma-Proteine zu binden vorhergesagt. Auch die Vorhersage der Fähigkeit die Blut-Hirn-Schranke zu passieren (siehe Abbildung 3.9) ist für fünf der Substanzen hoch bis medium, was jedoch während der klinischen Phasen der Wirkstoffentwicklung oft zum Ausscheiden potenzieller Substanzen führt. Nur für Substanz vier war keine Abschätzung möglich.

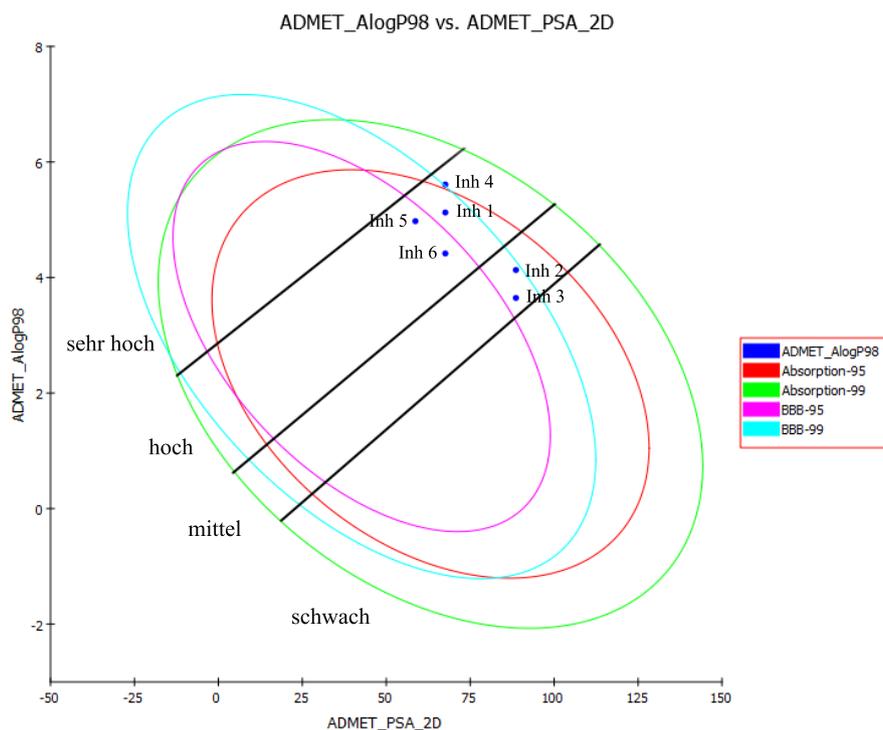


Abb. 3.9: Mit der Accelrys Software erstellte intestinale Absorptions-Vorhersage der sechs Inhibitoren nach oraler Administration aufgetragen auf die AlogP98 gegen PSA-2D Ebene.

Neben der, mit Hilfe der Accelrys Software vorhergesagten, AlogP98- und PSA-2D-Werte der sechs potenziellen Inhibitoren sind in der Ebene auch die 99% und 95% Konfidenzellipsen der intestinalen Absorption und der Blut-Hirn-Schranke Penetration visualisiert.

Die rote und grüne Konfidenzellipse definieren den Bereich, in der die gut-absorbierbaren Substanzen erwartet werden. 95% der gut-absorbierbaren Strukturen werden im Bereich der roten Konfidenzellipse und 99% innerhalb der grünen erwartet. In diesem Modell ist eine gut-absorbierbare Substanz eine, die zu mindestens 90% in den Blutstrom absorbiert werden kann. Die magentafarbene Konfidenzellipse definiert den Bereich, in dem 95% der Substanzen erwartet werden, welche die Blut-Hirn-Schranke nicht, kaum, gut oder sehr gut durchdringen können. Die türkisfarbene Ellipse deckt 99% der Substanzen ab. Die schwarzen Linien gelten nur für die Ellipsen der Blut-Hirn-Schranke und definieren den Bereich, in denen die sehr hohen; hohen; mittleren und schwachen Blut-Hirn-Schranke durchdringenden Substanzen erwartet werden. Für alle Substanzen außerhalb der türkisfarbenen Ellipse können keine Blut-Hirn-Schranke Penetrationswerte vorhergesagt werden. Inhibitor eins, fünf und sechs besitzen die höchsten vorhergesagten Absorptions- und Blut-Hirn-Schranke-Penetrationswerte. Inhibitor vier ist hingegen nicht bestimmbar (außerhalb der türkisfarbenen Ellipse).

Für die Inhibitoren fünf und sechs wird außerdem eine 50%ige Wahrscheinlichkeit vorhergesagt das Cytochrom P450 2D6 zu inhibieren. Die Hepatotoxizitätswerte sind sehr hoch, was bedeutet, dass alle Strukturen mit höherer Wahrscheinlichkeit Leber-toxisch wirken, jedoch wird für alle Substanzen bei oraler Verabreichung keine Toxizität vorhergesagt. Alle Substanzen besitzen einen hohen vorhergesagten LD50-Wert (siehe auch Tabelle 3.6).

Inhibitor	BBB-Level	Absorp.	Löslich.	Hepatotox	Cyp-Inh	PPB-Level
Inh 1	hoch	gut	möglich	tox (0,5)	nein	>95%
Inh 2	med	gut	ja	tox (0,9)	nein	>90%
Inh 3	med	gut	ja	tox (0,9)	nein	>95%
Inh 4	undefiniert	moderate	möglich	tox (0,5)	nein	>95%
Inh 5	hoch	gut	möglich	tox (0,7)	ja (0,5)	>95%
Inh 6	hoch	gut	ja	tox (0,8)	ja(0,5)	>95%

Tab. 3.5: Mit der Accelrys Software vorhergesagten ADMET-Eigenschaften für die sechs potenziellen Sekretase Inhibitoren. Vorhergesagt mit Accelrys „Admet-Descriptors“. BBB = Blood Brain Barrier; PPB = Plasma Protein Binding.

Inhibitor	LD50 Log (1/Moles)	LD50 in $\frac{mg}{kg}$	max. toler. Dosis in $\frac{mg}{kg}$
Inh 1	3,144	289	96
Inh 2	2,273	2300	142,3
Inh 3	0,741	10000	581,8
Inh 4	4,006	420,5	1200
Inh 5	2,735	718,9	69,4
Inh 6	2,601	939,6	150,7

Tab. 3.6: Die vorhergesagten LD50-Werte und die vorhergesagten maximal tolerierten Dosen für das Modell: Ratte oral.

Alle vorhergesagten Werte liegen über 25 mg/kg und sind damit unbedenklich. Vorhergesagt mit Accelrys „TOPKAT“

Die experimentelle Validierung erfolgte in der Kooperationsgruppe der Frau Dr. Jundt [162]. In der Publikation von Schwarzer et al. werden die in der AG Jundt durchgeführten experimentellen Validierungstest auf Multiplen Myelom Zellen beschrieben [162].

Abbildung 3.10 visualisiert die Positionierung des experimentell wirksamsten Liganden in der katalytischen Bindungstasche des Presenilins der γ -Sekretase, wobei eine Wasserstoffbrückenbindung zwischen einem Stickstoffatom des Liganden mit dem Sauerstoffatom eines katalytischen Aspartatrestes als auch neun hydrophobe Wechselwirkungen zwischen den Bindungspartnern ausgebildet werden (siehe Abbildung 3.11).

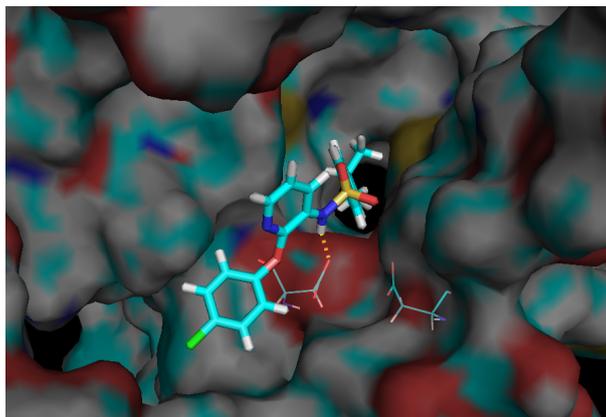


Abb. 3.10: Docking des Inhibitors mit dem niedrigsten experimentell gemessenen GI50-Wert in die modellierte Bindungstasche des Presenilins.

Gelb markiert ist eine Wasserstoffbrückenbindung zwischen dem Stickstoffatom des Liganden und dem Sauerstoffatom eines katalytischen Aspartats (Asp385).

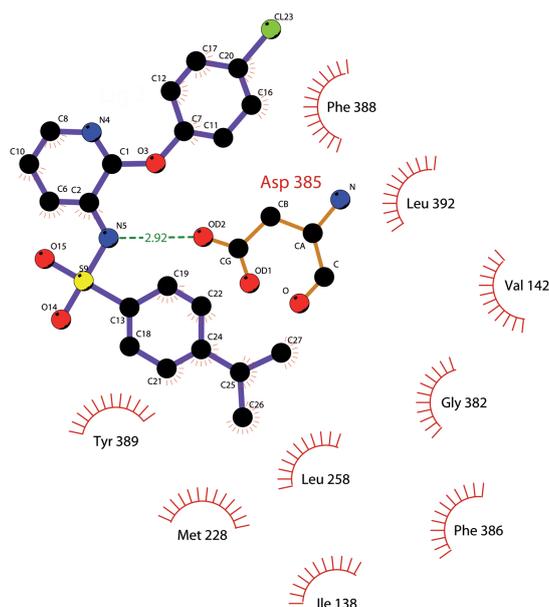


Abb. 3.11: Planare Visualisierung der Docking Pose des wirksamsten Liganden in die Bindungstasche des Presenilins mit Hilfe der Ligplot-Software [163].

Gelb: Wasserstoffbrückenbindung mit dem Abstand 2,9 Å und neun hydrophobe Wechselwirkungen.

Weitere experimentelle Untersuchungen konnten belegen, dass der affinste entwickelte γ -Sekretase Inhibitor zu keiner Inhibition der β -Sekretase führt was auf eine hohe Spezifität des Liganden auf die Adressierung der γ -Sekretase hin deutet. Eine Konzentration von $1.0E^{-05}$ M zeigte keine inhibitorische Wirkung in einem β -Sekretase

Assay, welches von der Firma Cerep (<http://www.cerep.fr/Cerep/Users/index.asp>) durchgeführt wurde. Die identifizierten γ -Sekretase Modulatoren wurden erfolgreich patentiert. Das Patent wurde von der Firma „Fresenius Medical Care“ international angemeldet.

3.2.1.2 VEGF-Rezeptor 2 als Targetprotein

Die Blockierung des Angiogenese Signalweges durch die Inhibition des VEGF- Rezeptors 2 stellt einen wichtigen Schritt in der Tumorthherapie dar. Speziell die Unterdrückung der Versorgung des Tumors mit essentiellen Nährstoffen induziert eine Reduktion des Tumors. Für die Durchführung des Struktur-basierten Wirkstoffdesigns lagen Proteinstrukturen der ATP-Bindungstasche des VEGFR-2 Proteins in der PDB vor. Jedoch lagen, vor allem in der Bindungstasche, einige fehlende Sequenzen und damit sogenannte „Gaps“ vor, die mit Homologie-Modellierungsmethoden vervollständigt werden mussten. Für die Struktur wurde zunächst die PDB-Struktur mit der ID 1YWN gewählt. Jedoch wurde 2008/9 eine verbesserte Struktur mit der PDB-ID 3EWH veröffentlicht, die keine Lücken in der Bindungstasche aufwies. Abbildung 3.12 zeigt das Alignment der in der PDB-vorhandenen Struktursequenz (3EWH) und der Original-Sequenz des aktiven Zentrums aus der UniProt (VGFR2_HUMAN). Auch diese Struktur wurde mit Hilfe der Homologie- Modellierungsmethoden vervollständigt. Hierfür wurden zwei kurzen Gaps durch jeweils einen Linker geschlossen und Threonin an der Stelle 916 durch Valin mutiert. Abbildung 3.13 zeigt die modellierte katalytische Untereinheit des VEGF-Rezeptors 2 mit markierter Bindungstasche für das biologische Substrat ATP.

```

3EWH Chain A, Crystal Structure Of The Vegfr2 Kinase Domain In Complex
With A Pyridyl-Pyrimidine Benzimidazole Inhibitor
Length=314

Score = 609 bits (1571), Expect = 5e-175, Method: Compositional matrix adjust
Identities = 300/352 (86%), Positives = 300/352 (86%), Gaps = 50/352 (14%)

Query 820 LPYDASKWEFPRDRCLKGKPLGRGAFGQVIEADAFGIDKTAICRTVAVKMLKEGATHSEH 879
LPYDASKWEFPRDRCLKGKPLGRGAFGQVIEADAFGIDKTAICRTVA KMLKEGATHSEH
3EWH 820 LPYDASKWEFPRDRCLKGKPLGRGAFGQVIEADAFGID----CRTVAAAKMLKEGATHSEH 879

Query 880 RALMSELKILIHIGHHLNVNLLGACTKPGGPLMVIVEFCKFGNLSTYLRSKRNEFVFPYK 939
RALMSELKILIHIGHHLNVNLLGACTKPGGPLMVI EFCKFGNLSTYLRSKRNEFVFPYK
3EWH 880 RALMSELKILIHIGHHLNVNLLGACTKPGGPLMVITEFCKFGNLSTYLRSKRNEFVFPYK 939

Query 940 TKGARFRQKDYVGAIPVDLKRRLDSITSSQSSASSGFVEEKSLSDVVEEEAAPEDLYKDF 999
APEDLYKDF
3EWH 940 -----APEDLYKDF 999

Query 1000 LTLEHLICYSFQVAKGMEFLASRCKIHRDLAARNILLSEKNVVKICDFGLARDIYKDPDY 1059
LTLEHLICYSFQVAKGMEFLASRCKIHRDLAARNILLSEKNVVKICDFGLARDIYKDPDY
3EWH 1000 LTLEHLICYSFQVAKGMEFLASRCKIHRDLAARNILLSEKNVVKICDFGLAR-----PDY 1059

Query 1060 VRKGDARLPLKWMAPETIFDRVYTIQSDVWSFGVLLWEIFSLGASPYPGVKIDEEFCRRL 1119
VRKGDARLPLKWMAPETIFDRVYTIQSDVWSFGVLLWEIFSLGASPYPGVKIDEEFCRRL
3EWH 1060 VRKGDARLPLKWMAPETIFDRVYTIQSDVWSFGVLLWEIFSLGASPYPGVKIDEEFCRRL 1119

Query 1120 KEGTRMRAPDYTTPEMYQTMDCWHGEPQRSRPTFSELVEHLGNLLQANAQQD 1171
KEGTRMRAPDYTTPEMYQTMDCWHGEPQRSRPTFSELVEHLGNLLQANAQQD
3EWH 1120 KEGTRMRAPDYTTPEMYQTMDCWHGEPQRSRPTFSELVEHLGNLLQANAQQD 1171
    
```

Abb. 3.12: Alignment der kristallisierten VEGFR-2 Sequenz aus der PDB (3EWH) und der Proteinsequenz aus der UniProt (VGFR2_HUMAN).

Gelb: Aminosäuren der Bindungstasche des Rezeptors, wobei die Bindungstasche als folgenden Bereich definiert wurde: 6 Å um den co-kristallisierten Liganden.

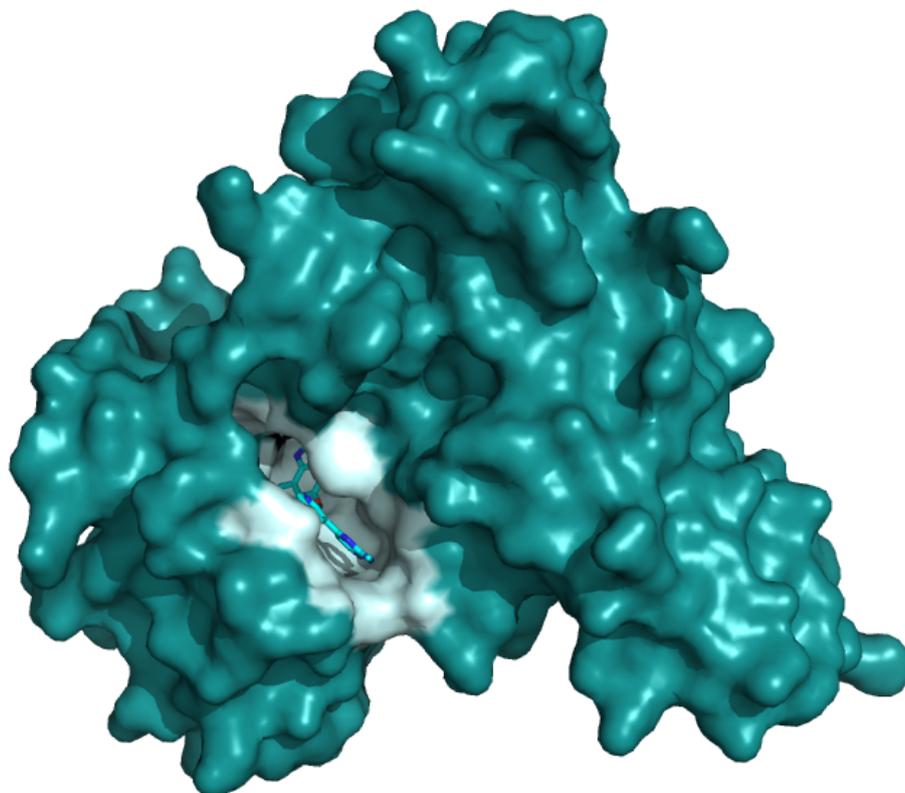


Abb. 3.13: Durch Homologie-Modellierung vervollständigte katalytische Untereinheit des VEGF-Rezeptors 2 (PDB: 3EWH).

Die Bindungstasche ist als Bereich 6 Å um den co-kristallisierten Liganden definiert.

Inhibitorensuche:

Der erste Schritt der Identifizierung neuer potenzieller Inhibitoren für den VEGF-Rezeptor 2 bestand in der Identifizierung einer geeigneten Leitstruktur. Nach Literaturrecherchen wurde der publizierte VEGF-Rezeptor Inhibitor Vatalanib als Leitstruktur verwendet (siehe Abbildung 3.14).

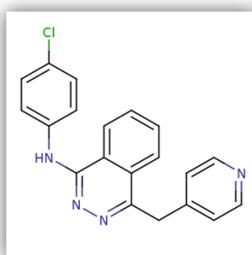


Abb. 3.14: Vatalanib.

Diente als Leitstruktur für die Identifizierung und Optimierung wirksamerer potenzieller Wirkstoffe durch Bindung an VEGFR-2.

Die parallele Anwendung der 2D- und 3D-Ähnlichkeitssuchen ergaben ca. 60 verschiedene Kleinstrukturen, wobei die Liste nach der Filterung bzgl. der oralen Bioverfügbarkeit mit Hilfe der Lipinski 5er Regel auf 21 Substanzen gekürzt wurde. Insgesamt testete die Kooperationsgruppe von Dr. Höpfner zwölf verschiedene Substanzen bzgl. ihrer Wirkung auf Zellwachstum der Endothelzelllinie EA-HY 926. Abbildung 3.15 zeigt die zwölf getesteten Substanzen.

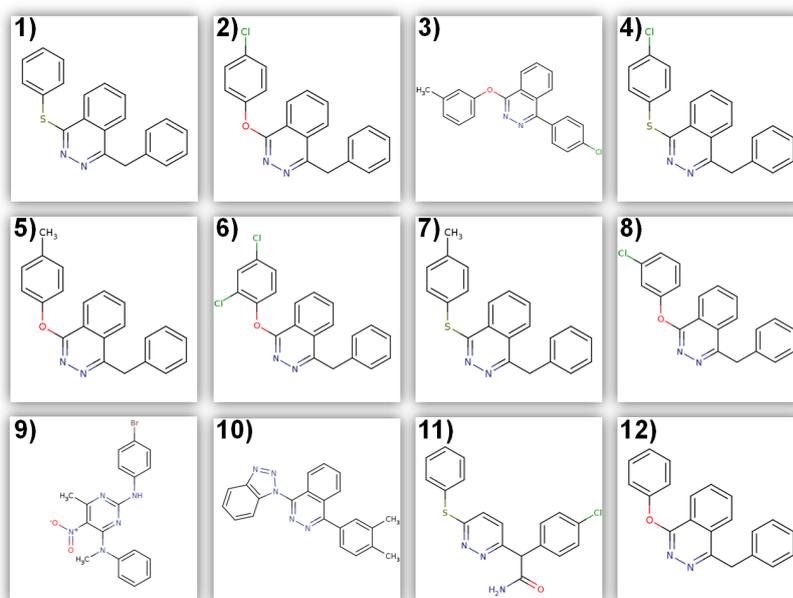


Abb. 3.15: Potenzielle Wirkstoffkandidaten nach der Anwendung des 2D- und 3D-Screenings

Die durch das *in silico* Screening ermittelten zwölf Substanzen wurden experimentell validiert und IC50-Werte bestimmt.

Abbildung 3.16.a) zeigt die Kleinstruktur mit dem niedrigsten IC50-Wert und damit stärkstem inhibitorischen Effekt während des Zellwachstumsexperiments. Diese Kleinstruktur diente anschließend als Leitstruktur für einen weiteren *in silico*-Screening Durchlauf, wobei drei weitere Kleinstrukturen identifiziert wurden (siehe Abbildung 3.16.b)) für die auch eine experimentelle Validierung durchgeführt wurde.

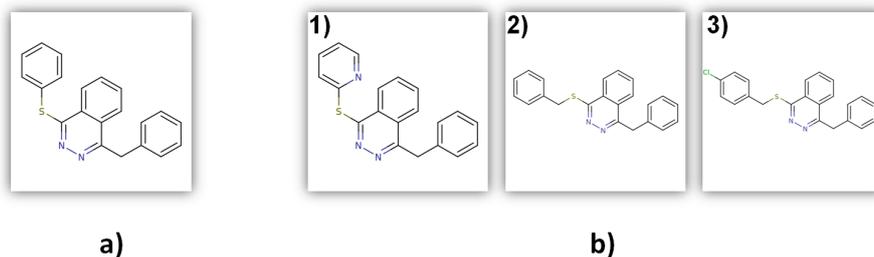


Abb. 3.16: Wirksamster Inhibitor nach experimenteller Validierung diente als neue Leitstruktur.

a) zeigt den Inhibitor, der als neue Leitstruktur für den nächsten *in silico* Screening-Durchlauf fungierte b) dadurch wurden drei weitere Wirkstoffkandidaten identifiziert.

Insgesamt konnte von der Arbeitsgruppe Dr. Höpfner für sieben Kleinstrukturen experimentell eine Zellwachstumsinhibition nachgewiesen und ein IC₅₀-Wert ermittelt werden. Die genauen Werte und weitere von der AG Höpfner durchgeführten Validierungsexperimente sind in der Publikation von Nitzsche et al. [34] beschrieben. Für einige der Substanzen wurde ein niedrigerer IC₅₀-Wert ermittelt als für die Leitstruktur Vatalanib.

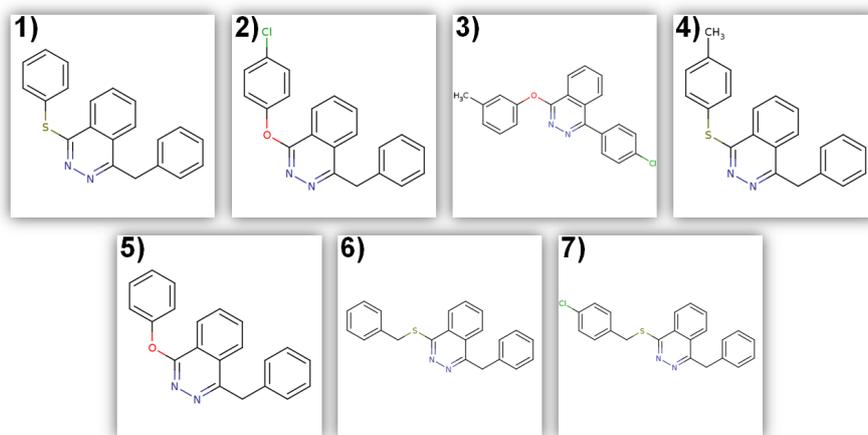


Abb. 3.17: Potenzielle VEGFR-2-Arzneistoffkandidaten nach experimenteller Validierung.

ADMETox-Analysen:

Alle sieben Wirkstoffkandidaten zeigen bei Anwendung der Lipinski 5er Regel eine gute orale Bioverfügbarkeit (siehe Tabelle: 3.7). Ähnlich wie schon bei der Suche nach potenziellen γ -Sekretase Inhibitoren ist der berechnete clogP-Wert oft zu hoch.

Inhibitor	MW	H-don.	H-acc.	clogP	PSA
Inh 1	328	0	3	5,4	51
Inh 2	346	0	3	5,6	35
Inh 3	346	0	3	6	35
Inh 4	342	0	3	5,7	51
Inh 5	312	0	3	5	35
Inh 6	342	0	3	5,5	51
Inh 7	376	0	3	6,2	51

Tab. 3.7: Lipinski-Eigenschaften und die „Polar Surface Area“ (PSA). Die Inhibitoren verstoßen jeweils gegen nur eine Lipinski-Regel. Der berechnete logP ist meistens minimal zu hoch.

Des Weiteren werden für alle Inhibitoren eine sehr gute Fähigkeit die Blut-Hirn-Schranke zu passieren vorhergesagt (siehe Abbildung 3.18). Die Fähigkeiten der Absorption, Bindung an „Carrier“-Proteine und seiner Verteilung im Organismus werden für die Substanzen von gut bis moderat eingeschätzt (siehe Tabelle 3.8). Wie schon bei den γ -Sekretase Inhibitoren werden auch für die potenziellen VEGFR-2-Inhibitoren hohe Lebertoxizitäts-Wahrscheinlichkeiten als auch Cytochrom-Inhibitions-Wahrscheinlichkeiten vorhergesagt. Der vorhergesagte LD50-Wert jedoch ist hoch, was auf eine hohe tolerierte orale Dosis bei Ratten hinweist (siehe Tabelle 3.9).

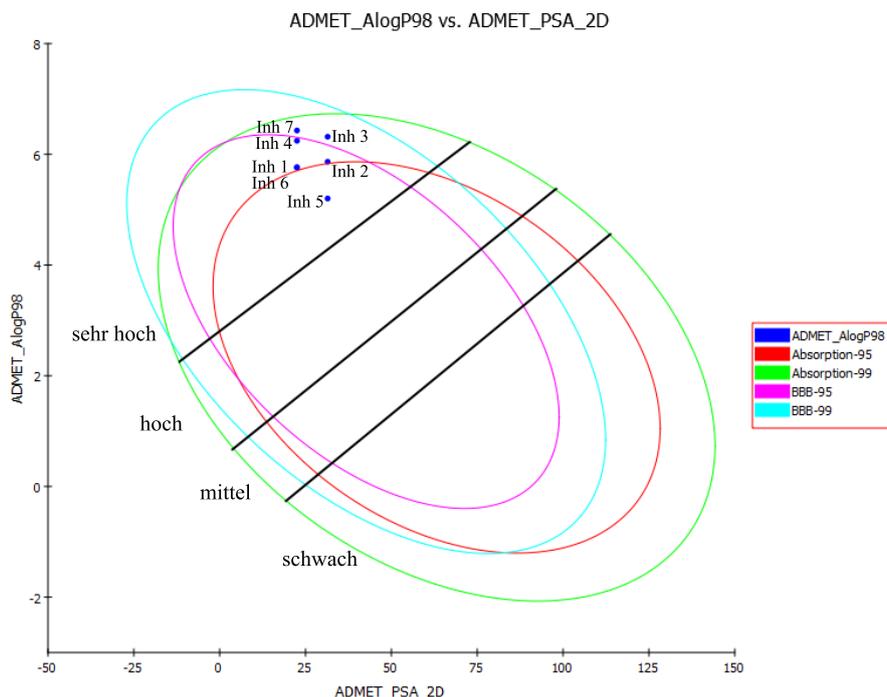


Abb. 3.18: Mit der Accelrys Software erstellte intestinale Absorptions-Vorhersage der sieben potenziellen VEGFR-2 Inhibitoren nach oraler Administration aufgetragen auf die AlogP98 gegen PSA-2D Ebene.

Neben der, mit Hilfe der Accelrys Software vorhergesagten, AlogP98- und PSA-2D-Werte der sieben potenziellen Inhibitoren sind in der Ebene auch die 99% und 95% Konfidenzellipsen der intestinalen Absorption und der Blut-Hirn-Schranke Penetration visualisiert.

Die rote und grüne Konfidenzellipse definieren den Bereich, in dem die gut-absorbierbaren Substanzen erwartet werden. 95% der gut-absorbierbaren Strukturen werden in dem Bereich der roten Konfidenzellipse und 99% innerhalb der grünen erwartet. In diesem Modell ist eine gut absorbierbare Substanz eine, die zu mindestens 90% in den Blutstrom absorbiert werden kann. Die magenta-farbene Konfidenzellipse definiert den Bereich, in den 95% der Substanzen erwartet, welche die Blut-Hirn-Schranke nicht, kaum, gut oder sehr gut durchdringen können. Die türkis-farbene Ellipse deckt 99% der Substanzen ab. Die schwarzen Linien gelten nur für die Ellipsen der Blut-Hirn-Schranke und definieren den Bereich, in denen die sehr hohen; hohen; mittleren und schwachen Blut-Hirn-Schranke durchdringenden Substanzen erwartet werden. Für alle Substanzen außerhalb der türkis-farbenen Ellipse können keine Blut-Hirn-Schranke-Penetrationswerte vorhergesagt werden. Für alle VEGFR-2 Inhibitoren wurde eine sehr hohe Fähigkeit vorhergesagt, die Blut-Hirn-Schranke zu durchqueren. Inhibitor 5 weist auf eine gut Absorptionfähigkeit hin. Für die anderen Substanzen wird eine moderate Absorptionfähigkeit vorhergesagt.

Inhibitor	BBB-Level	Absorp.	Löslich.	Hepatotox	Cyp-Inh	PPB-Level
Inh 1	sehr hoch	moderate	möglich	tox (0,9)	ja (0,7)	>95%
Inh 2	sehr hoch	moderate	möglich	tox (0,9)	ja (0,7)	>95%
Inh 3	sehr hoch	moderate	möglich	tox (0,9)	ja (0,7)	>95%
Inh 4	sehr hoch	moderate	möglich	tox (0,9)	ja (0,6)	>95%
Inh 5	sehr hoch	gut	möglich	tox (0,9)	ja (0,7)	>95%
Inh 6	sehr hoch	moderate	möglich	tox (0,9)	ja(0,7)	>95%
Inh 7	sehr hoch	moderate	möglich	tox (0,9)	ja(0,6)	>95%

Tab. 3.8: Mit der Accelrys Software vorhergesagte ADMET-Eigenschaften für die sieben potenziellen VEGFR-2 Inhibitoren. Vorhergesagt mit Accelrys „Admet-Descriptors“. BBB = Blood Brain Barrier; PPB = Plasma Protein Binding.

Inhibitor	LD50 Log (1/Moles)	LD50 in $\frac{mg}{kg}$	max. toler. Dosis in $\frac{mg}{kg}$
Inh 1	2,694	665,0	99
Inh 2	2,915	422,2	42,6
Inh 3	3,211	213,4	26,6
Inh 4	2,776	573,8	95,8
Inh 5	2,582	818,2	54,1
Inh 6	1,715	6600	315,7
Inh 7	2,120	2900	224,1

Tab. 3.9: Die vorhergesagten LD50-Werte und die maximal tolerierten Dosen für das Modell: Ratte oral.

Alle vorhergesagten Werte liegen über 25 mg/kg und sind damit unbedenklich.

3.3 Wirkstoff-Zielprotein-Interaktionen

3.3.1 Integration von Wirkstoff-Zielprotein-Interaktionen

Aufgrund starker Nebenwirkungen scheidet die Entwicklung neuer Arzneistoffe oftmals schon in den ersten klinischen Phasen. Daher ist eine Analyse der Wechselwirkungen zwischen Zielprotein und Substanz sowie der resultierenden Nebenwirkungen durch eine hohe Affinität und niedrige Spezifität hilfreich, um schon vorab Wirkstoffkandidaten auszuschließen. In den letzten Forschungsjahren häufte sich ein immer größer werdendes Wissen über Zielproteine, deren Funktion in Signalwegen und deren Adressierung verschiedener Substanzen an. Die Integration experimenteller Daten, Literaturwissen über Wirkstoffe, deren Metabolismus, Zielproteine, Ontologien und Signalwege stellt einen wichtigen Schritt für das Verständnis der Wirkungsweisen der Substanzen dar. Eine solche Integration ist mit der Entwicklung der SuperTarget-Datenbank (Kapitel 3.3.1.2.) realisiert worden. Für onkologisch spezifische Daten wurde die CancerResource (CaRe) entwickelt, die neben den oben

genannten Bereichen auch Expressionsdaten und vergleichbare Wirkprofile von über 40.000 Substanzen auf 60 Zelllinien enthält (Kapitel 3.3.1.3.). Um Wirkprofile und die Zytotoxizität verschiedener Kleinstrukturen auf verschiedenen Zelllinien analysieren und Aussagen über mögliche Zielproteine treffen zu können, ist ein Vergleich derer von essentieller Bedeutung. Der Vergleich wurde im Laufe dieser Arbeit mit Hilfe eines zellulären Fingerprints ermöglicht.

3.3.1.1 Zellulärer Fingerprint zum Vergleich von Wirkprofilen

Bei dem zellulären Fingerprint [147] handelt es sich um einen 4.800 Bits langen Bitvektor, der das Wirkprofil einer Substanz und seinen Anti-Tumor Effekt auf die NCI60-Zelllinien repräsentiert (siehe Kapitel 2.2.5.2).

Es ist beschrieben, dass Strukturähnlichkeiten von über 85% mit 30% iger Wahrscheinlichkeit auch zu einer ähnlichen Funktionsweise der Kleinstrukturen führen [164]. Es stellt sich die Frage, ob eine Korrelation zwischen struktureller und zellulärer Ähnlichkeit vorhanden ist, also ob strukturell ähnliche Substanzen auch ähnliche Inhibitionen auf den verschiedenen Zelllinien zeigen und damit auf ähnliche Weise wirken. Dafür wurde eine Clusterung (Gruppierung) der NCI60-Kleinstrukturen mit ähnlichem zellulären Effekt in Hinblick auf deren strukturellen Ähnlichkeiten durchgeführt (siehe Abbildung 3.19). Insgesamt ist zu erkennen, dass sich viele strukturell ähnliche Substanzen in dem Datensatz befinden. Der Großteil der Kleinstrukturen findet sich in drei größeren Clustern (Gruppen), wobei jedes dieser Cluster einem Scaffold, also einem Kleinstruktur-Gerüst, entspricht. Cluster 1 beinhaltet verschiedene Colchicine, Cluster 2 verschiedene Taxane und Cluster 3 Quinolinone. Abbildung 3.20 zeigt ein weiteres Beispiel, in dem für das 7-Chlorcamptothecin strukturell und zellulär ähnliche Substanzen ermittelt wurden. Alle vier Substanzen besitzen den Baustein Camptothecin.

Die ersten Analysen zeigten, dass die Aufstellung von Hypothesen bzgl. ähnlicher Wirkprofile basierend auf strukturellen Eigenschaften möglich ist. Strukturell ähnliche Substanzen können ähnliche Wirkprofile besitzen (siehe Abbildung 3.20) und oft strukturell ähnliche Substanzen auch ähnliche Wirkprofile haben. Jedoch können auch strukturell unterschiedliche Substanzen ähnliche Wirkprofile besitzen. Das zeigt, dass der zelluläre Fingerprint auf die Targetspezifität hinweist, ohne dass der strukturelle Fingerprint eindeutig zu interpretieren ist.

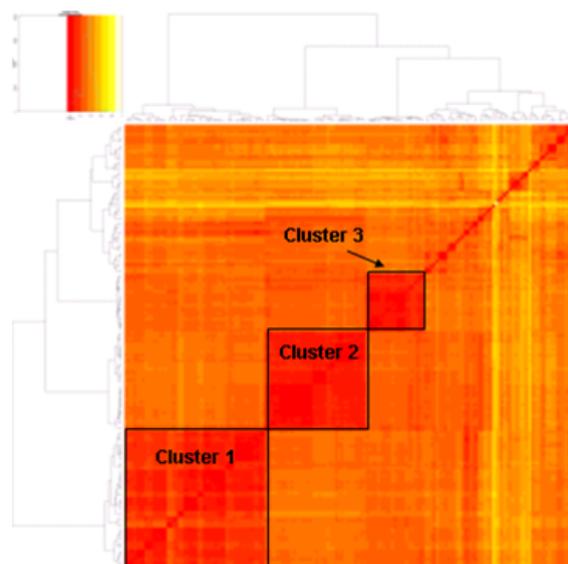


Abb. 3.19: Gruppierung (Clustering) der NCI60-Substanzen mit ähnlichem zellulären Effekt bzgl. ihrer strukturellen Ähnlichkeiten.

Die drei großen Cluster beinhalten jeweils Substanzen mit ähnlichen Grundgerüsten. Es gilt: je dunkler die Farbe desto ähnlicher die Substanzen. *Aus Günther et al. [147].*

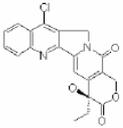
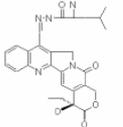
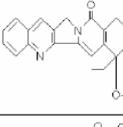
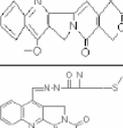
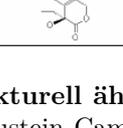
NCS-ID	Cell. Tanimoto	Struct. Tanimoto	Structure
249910	100	100	
681634	98.8	87.5	
606986	98.3	87.7	
295501	98.2	96.2	
681635	98.2	86.4	

Abb. 3.20: 7-Chlorcamptothecin und vier strukturell ähnliche Substanzen mit ähnlichem Wirkprofil. Alle Strukturen besitzen den Baustein Camptothecin, welcher als Topoisomerase I Inhibitor und damit als Tumorthapeutikum Verwendung findet. *Aus Günther et al. [147].*

Die zellulären Fingerprints helfen, neue Zielproteine zu identifizieren und die Wirkungsweise der Wirkstoffe zu erklären. Es ist möglich für Substanzen mit strukturellen Ähnlichkeiten auch auf ein zellulär ähnliches Wirkungsverhalten zu schließen und umgekehrt. Darüber hinaus erleichtern bekannte Kleinstrukturprofile die Auswahl von sensitiven Zelllinien für die *in vitro* Testung von ähnlichen Kleinstrukturen. Die Kombination aus strukturellem und zellulärem Fingerprint führt zur Identifizierung von neuen „Scaffold Hoppfern“. Das Resultat sind Strukturen mit verschiedenen Topologien aber ähnlicher Wirkung.

Zur Validierung dieser Methodik ist die experimentelle Bestätigung vorhergesagter Wirkprofile erforderlich.

3.3.1.2 SuperTarget - Arzneistoff-Zielprotein-Interaktionen

Um die molekularen Mechanismen der Arzneistoffe und ihren Zielproteinen besser verstehen und Nebenwirkungen erklären zu können, war ein erster Schritt die Erstellung einer Datenstruktur, die neben Arzneistoffen, auch deren Zielproteine und Signalwege beinhaltet und miteinander verknüpft. Es entstand die Ressource SuperTarget (<http://bioinf-tomcat.charite.de/supertarget/>), die derzeit über 3.000.000 potenzielle Zielproteine, 2.400 Arzneistoffe und ca. 200 Signalwege beinhaltet. Aus diesen Daten konnten mittels automatisierter Methoden und manueller Validierung über 7.200 Arzneistoff-Zielprotein-Verknüpfungen mit dazugehöriger PubMed-Referenz, 215.000 Zielprotein-Signalweg-Verknüpfungen und 1.900 Arzneistoff-Signalweg-Verknüpfungen gewonnen werden. Neben den diversen Verknüpfungen sind auch Cytochrom Informationen gespeichert. Hierbei wurde vor allem der Abbau der Arzneistoffe durch die Cytochrome P450 betrachtet. Insgesamt enthält die SuperTarget ca. 40 verschiedene Cytochrome P450 und die Verknüpfung zu den von ihnen abgebauten Arzneistoffen. Für die Ontologien wurden 20.000 Ontologie-Begriffe der Gene Ontology extrahiert (siehe Abbildung 3.21). Die SuperTarget stellt jedoch nicht nur eine Datenansammlung dar, sondern ist vielmehr ein effektives Analyse-Werkzeug mit integrierten Such- und Analyseprogrammen.

Die SuperTarget besteht aus fünf verschiedenen Menüpunkten: „*Drugs*“, „*Targets*“, „*Pathways*“, „*Ontologies*“ und „*Cyp450s*“ in denen jeweils einfache Anfragen über verschiedene Bezeichner wie Namen und anderen Identifikatoren (PDB, UniProt, Kegg, etc.) durchführbar sind. Zusätzlich bietet die SuperTarget auf der Startseite eine Volltextsuche „*Targle*“, die eine Schlagwortsuche über die vollständige Datenbank erlaubt. Bei dieser Suche werden jeweils alle Bereiche nach den Schlagworten

durchsucht und eine Ergebnisseite mit allen gefundenen Treffern bezüglich Arzneistoffen, Zielproteinen und Signalwegen ausgegeben. Interessante Zwischenergebnisse lassen sich in einem sogenannten „Clipboard“ speichern, um anschließend mit Hilfe des „Queryboards“ und logischen UND und ODER Verknüpfungen kompliziertere Anfragen zu stellen. Im Anwendungsbeispiel wird dieser Vorgang näher beschrieben (siehe Seite 88).

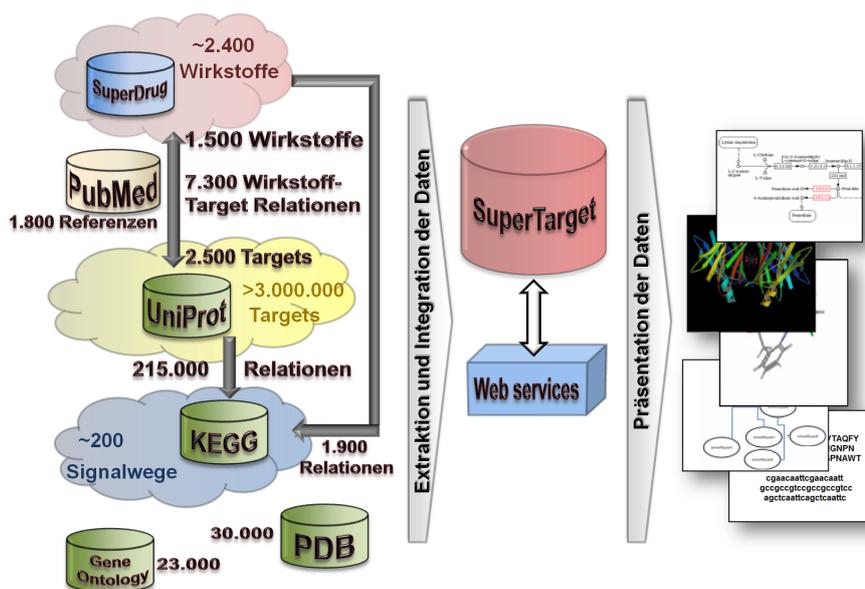


Abb. 3.21: Architektur der SuperTarget Datenbank.

Die SuperTarget Datenbank beinhaltet als potenzielle Targets die UniProt-Proteine und für die Wirkstoffe die SuperDrug-Kleinstrukturen. Zusätzlich werden neben den KEGG-Signalwegen noch 23.000 verschiedene GO-Begriffe und 30.000 Verlinkungen zu PDB-Proteinstrukturen gespeichert. Bisher stellt die SuperTarget mehr als 7.000 manuell validierte Arzneistoff-Target-Interaktionen zur Verfügung. Alle Daten sind in SuperTarget miteinander verknüpft und lassen sich mit Hilfe der in SuperTarget angebotenen Werkzeuge visualisieren und analysieren. Aus Günther et al. (modifiziert) [149].

Pharmazeutische Wirkstoffe - „Drugs“

SuperTarget bietet die Möglichkeit mit Hilfe des Wirkstoffnamens, der CAS-Nummer, des PDB-Liganden Identifikators und des ATC-Codes nach Wirkstoffen, die aus der SuperDrug stammten zu suchen. Aktuell umfasst die Datenbank 2.400 WHO ATC-klassifizierte Substanzen. Zusätzlich hilft ein implementierter ATC-Baum, das „Browsen“ durch die Klassen und die Selektion der Wirkstoffe einer Klassifikationsgruppe. Die Ergebnisliste liefert neben der CAS-Nummer, die Synonyme und, wenn vorhanden, die 3D Struktur als Link auf die SuperLigand Datenbank (enthält alle PDB-Liganden, die an mindestens ein Protein gebunden sind) und/oder SuperDrug-

Datenbank. Auch die Information, wie viele Verknüpfungen zu Zielproteinen, Signalwegen, Ontologien und Cytochrome als auch die Protein-Liganden-Referenz werden nach der Wirkstoffsuche angegeben. Darüber hinaus sind die zu den Arzneistoffen beobachteten und dokumentierten Nebenwirkungen gespeichert.

Als Analysemethode ist eine Strukturähnlichkeitssuche implementiert, die es ermöglicht Ähnlichkeiten zwischen Arzneistoffen, die mittels Tanimoto-Koeffizienten angegeben werden, zu bestimmen. Bei Anwendung der Suche nach ähnlichen Wirkstoffen, erhält der Nutzer neben der Liste der ähnlichen Substanzen auch die zugehörigen Zielprotein-Informationen.

Der pharmazeutische Wirkstoff Paclitaxel, ein Chemotherapeutikum, welcher zur Behandlung maligner Tumore verwendet wird, ist in der SuperTarget unter diesem Namen und weiteren Synonymen zu finden. Neben verschiedenen Synonymen und Identifikationsnummern, werden auch die beschriebenen Nebenwirkungen aufgelistet. Insgesamt wurden die folgenden 19 Nebenwirkungen beobachtet: Anaphylaktoide Reaktionen, Kurzatmigkeit, Bronchospasmus, Herzstillstand, Brustschmerzen, Tod, Benommenheit, Dysphagie, Dysphonie, Erythema, Extravasation, Gesichtsoedeme, Hautrötungen, Bluthochdruck, pulmonale Ödeme, Blässe, Tachykardie und kribblige Haut.

Zusätzlich wird hier angegeben, dass die SuperTarget drei Zielproteine, 18 Signalwege, zwei Ontologien und zwei Cytochrome zu dem Wirkstoff speichert. Auch die Referenzen zu Wirkstoff-Zielprotein-Beziehungen werden hier angegeben. Details zu den von Paclitaxel adressierten Proteinen, den Signalwegen, den Cytochromen und den zugehörigen Ontologien sind, ausgehend vom Wirkstoff Paclitaxel, in der jeweiligen Kategorie zu finden.

Eine anschließende Ähnlichkeitssuche lieferte die zehn ähnlichsten Kleinstrukturen, wobei die Treffer mit der höchsten Ähnlichkeit weitere Taxole wie Docetaxel darstellen.

Zielproteine - „Targets“

Für Analysezwecke ausgehend von einem Zielprotein bietet die SuperTarget die Suche mit Hilfe des Namens, des PDB-Identifikators, der EC-Klassifikation und des UniProt Identifikators des Zielproteins. Aktuell beinhaltet die Datenbank über 3.000.000 potenzielle Zielproteine, wobei auch „House-Keeping“-Proteine beinhaltet sind. Zusätzlich sind für jedes Protein, wenn vorhanden, die Synonyme, Organismus

und EC-Klassifikation gespeichert. Die vorhandenen Verknüpfungsinformationen zu den Signalwegen, Arzneistoffen und Ontologien werden auch angegeben. Sequenzähnlichkeitsvergleiche ermöglichen die Suche nach ähnlichen Proteinen. Für die 3D-Proteinstruktur-Information existiert eine Verlinkung zu den entsprechenden PDB-Einträgen.

Als humane Zielproteine für Paclitaxel wurde das Apoptose-Regulator-Protein Bcl-2 (P10415), die Tubulin beta-1 Kette (Q9H4B7) und die Tubulin-Tyrosin Ligase in SuperTarget gespeichert (Q8NG68), wobei nur für das Bcl-2 Proteinstrukturen vorhanden waren.

Signalwege - „Pathways“

Ausgehend von einem Signalweg bietet die SuperTarget die Möglichkeit sich einen von ca. 200 KEGG-Signalwegen darstellen zu lassen. Neben den Verknüpfungen werden auch die vollständigen KEGG-Signalwege angezeigt und angefragte Proteine visuell hervorgehoben. Hier werden zusätzlich die zugehörigen gespeicherten Wirkstoffe angezeigt.

SuperTarget speichert für Paclitaxel 17 humane Signalwege. In diesen Signalwegen spielt also eines der drei oben genannten Zielproteine des Paclitaxels eine Rolle: C21-Steroid Hormon Metabolismus, Tryptophan Metabolismus, Arachidonsäure Metabolismus, Linolsäure Metabolismus, Retinol Metabolismus, Metabolismus des Xenobiotikas über Cytochrome P450, Wirkstoff Metabolismus, Apoptose, Focal Adhesion, Gap Junction, Neurotrophin Signalweg, Amyotrophe Lateralsklerose, Pathogene Escherichia Coli Infektion, Kolorektaler Karzinom, Prostata Karzinom, Bronchialkarzinom und weitere onkologisch relevante Signalwege.

Ontologien - „Ontologies“

Um Proteine zu identifizieren bzw. Proteingruppen aus speziellen Ontologien zu erhalten, bietet die Ressource verschiedene Möglichkeiten. Neben der Suche nach Schlagworten, die sich auf die molekulare Funktionen der Proteine bezieht, bietet die SuperTarget noch die Möglichkeit nach der zellulären Komponente und nach dem biologischen Prozess in dem das Protein bzw. die Proteine eine Rolle spielen zu suchen. Darüber hinaus wird ein implementierter Ontologie-Baum zur Verfügung gestellt, der das „Browsen“ innerhalb der verschiedenen Ontologien erlaubt.

Die Zielproteine des Paclitaxels spielen in vielen Ontologien eine Rolle. Neben B-Zell Proliferation und verschiedenen Apoptose-Ontologie-Begriffen werden noch Zellkomponenten wie Membran des endoplasmatischen Retikulums, Kernmembran, Cytoplasma, Mikrotubulin und molekulare Funktionen wie BH3-Bindung, Protease Bindung, GTP-Bindung und ATP-Bindung gespeichert.

Cytochrome - „Cyp450s“

Bei der Suche nach Cytochrom P450 Informationen, kann entweder eines der 40 angebotenen Cytochrome ausgewählt und alle Informationen dazu wie UniProt-Identifikator, Substrat, Initiatoren und Inhibitoren des Cytochroms erhalten werden oder es können alle Informationen zu allen 40 Cytochromen gleichzeitig ausgegeben werden.

Paclitaxel wird vom Cytochrom 2C8 und 3A4 metabolisiert. Der Initiator des Cytochroms 2C8 ist Rifampin und die Inhibitoren sind Gemfibrozil, Montelukast und Trimethoprim. Für das Cytochrom 3A4 waren zum Zeitpunkt der Projektbearbeitung keine weiteren Informationen erhältlich.

Anwendungsbeispiel

Neben einfachen Suchanfragen lassen sich auch komplexere Fragestellungen formulieren und an die Datenbank stellen. Dabei können in jeder Kategorie Daten gesammelt und verwendet werden, um in einer anderen Kategorie komplexere Anfragen zu stellen. Im Folgenden soll ein Beispiel gezeigt werden (siehe Abbildung 3.22). Hier wird nach einem Arzneistoff gesucht, der Bcl-2 Proteine adressiert. Ein Arzneistoff wäre z.B. das Paclitaxel. Ein mutiertes Cytochrom und die dadurch verhinderte Verstoffwechslung des Wirkstoffs kann zu verstärkten Nebenwirkungen als auch zu erhöhter Toxizität führen. Eine Fragestellung könnte daher sein, einen alternativen Wirkstoff zu identifizieren. Dabei wird folgendermaßen vorgegangen:

1. Suche nach Paclitaxel (Taxol) unter dem Menüpunkt *Drugs*
2. Das Ergebnis (Taxol) wird im *Clipboard* gespeichert, um es für die nächste Suche zu verwenden
3. Durch Verschieben des Taxol Eintrages in das *Queryboard* unter dem Menüpunkt *Cytochromes*, wird das Cytochrom ausgegeben, welches Taxol abbaut: 2C8

4. Verschieben des Cytochromes 2C8 ins *Clipboard*
5. Durch Verschieben des Taxol-Eintrages aus dem *Clipboard* in das *Queryboard* unter dem Menüpunkt *Targets*, liefert das Target, welches Taxol adressiert: Bcl-2
6. Verschieben des Targets Bcl-2 ins *Clipboard*
7. Um nun einen alternativen Wirkstoff zu erhalten, werden das Bcl-2-Protein und das Cytochrom aus dem *Clipboard* in das *Queryboard* verschoben. Die kombinierte Suche wird im Menüpunkt *Drugs* durchgeführt. Um einen Arzneistoff zu finden, der Bcl-2 adressiert jedoch nicht vom Cytochrom 2C8 abgebaut wird, wird das NOT neben dem 2C8 angeklickt und die Anfrage gestartet. Das Ergebnis ist: Docetaxel

Nach diesem Schema lassen sich verknüpfte Anfragen in allen Kategorien erstellen.

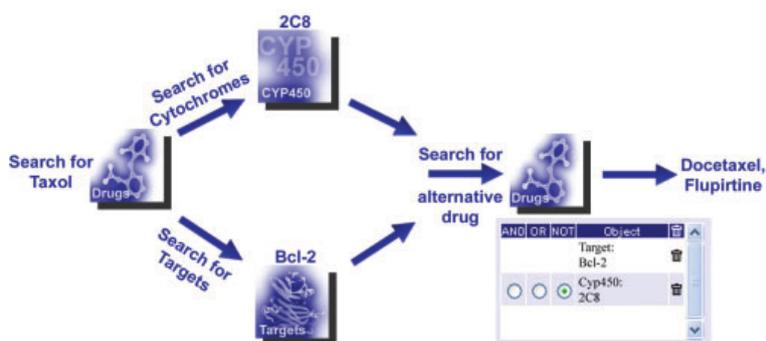


Abb. 3.22: Anwendungsbeispiel für die SuperTarget.

Vorgehensweise bei der Suche nach einem alternativen Wirkstoff zu Taxol, der Bcl-2 adressiert, jedoch nicht von Cyp 2C8 metabolisiert wird. Aus Günther et al. [149].

Aprepitant

Durch eine Analyse der Daten aus SuperTarget/Matador bzgl. Nebenwirkungsähnlichkeiten hat die Kooperationsgruppe von Dr. Peer Bork aus Heidelberg neue Arzneistoff-Zielprotein-Interaktionen identifiziert [165]. Dabei wurde auch die antiapoptotische Wirkung des Antiemetikums Aprepitant entdeckt, was nach der Vorhersage von Campillos et al. ähnliche Targets wie der chemotherapeutische Wirkstoff Pemetrexed [166] besitzt. Pemetrexed inhibiert die drei Enzyme Thymidilate Synthase, Dihydrofolate Reduktase und die Glycinamide Ribonukleotid Formyltransferase. Es handelt sich hierbei um ein Beispiel zur Repositionierung eines Arzneistoffs.

3.3.1.3 CancerResource (CaRe) - onkologische Datenbank mit integrierten Analyse- und Vorhersagemethoden

Um die große Datenmenge an onkologisch relevanten Targets und deren Expressionen, Signalwegen und die Effekte von onkologisch relevanten Kleinstrukturen auf onkologischen Zelllinien organisieren und analysieren zu können, wurde eine Ressource mit dem Namen CancerResource (CaRe) bereitgestellt, welche die verschiedenen genannten Themengebiete kombiniert. Sie ist unter folgender URL erreichbar:

http://bioinf – data.charite.de/cancerresource/. Aktuell beinhaltet diese über 50.000 onkologisch relevante Substanzen, über 2.400 onkologisch relevante Zielproteine mit ihren insgesamt 44 Signalwegen und mehr als 10.000 Kleinstruktur-Zielmolekül-Beziehungen (hauptsächlich Interaktionen). Die Kleinstruktur-Zielmolekül-Beziehungen wurden durch Integration der vorhandenen Ressourcen und durch Textmining-Methoden, die anschließend manuell validiert wurden generiert. Für jede Kleinstruktur und jedes Zielprotein sind umfangreiche Informationen sowie die Verknüpfung zu anderen Datenbanken wie PubChem, UniProt, GeneOntology, KEGG, NCI und PDB gespeichert. Außerdem sind für ca. 40.000 Kleinstrukturen Wirkprofile, welche die NCI's GI50-Werte einer Kleinstruktur auf die 60 NCI-Tumorzelllinien repräsentieren, enthalten. Darüber hinaus beinhaltet die CaRe Expressionsdaten für die 2.400 onkologisch relevanten Targets in den verschiedenen 60 Zelllinien. Des Weiteren ist es möglich, Daten wie Kleinstrukturen, Expressionsdaten und GI50-Werte hochzuladen und weitere Informationen bzgl. der anderen Bereiche zu erhalten. Neben der Kombination der verschiedenen Bereiche wie Targets, Expressionsdaten, Wirkstoffe, Wirkprofile und Signalwege bietet die CaRe auch Analyse- und Vergleichsmethoden. Zur Verbesserung der Funktionalität der CancerResource als onkologisches Web-basiertes Analyseprogramm wurden auch verschiedene experimentelle Daten mit aufgenommen. Indirekte Wechselwirkungen der Wirkstoffe auf Zielgene, die mittels Microarray-Untersuchungen oder Untersuchungen auf den NCI60-Zelllinien gewonnen wurden, können sofort mit den Textmining-Ergebnissen für direkte Interaktionen verglichen werden. CaRe bietet eine Reihe an Programmen zur Analyse der gespeicherten Daten an. Sie bietet Informationen für die Entwicklung neuer Leitstrukturen, Substanzwirkungen auf Proteine, Identifikation neuer Zielproteine für die Abschätzung und Vorhersage der Chemosensitivität der Substanzen bzgl. der 60 NCI-Zelllinien. Ein weiterer Aspekt ist die Identifizierung und Visualisierung der Wirkstoffe, die verschiedene Zielproteine oder Proteingruppen adressieren, als auch der einzelnen Zielproteine, die von einer größeren Anzahl an Wirkstoffen adressiert

werden. Die CaRe soll Wissenschaftlern dabei helfen, die Wirkung von Kleinstrukturen besser zu verstehen und neue Zielproteine zu identifizieren. Des Weiteren sollen hochgeladene Expressionsdaten mit gespeicherten Daten verglichen und potenzielle Wirkstoffkandidaten vorgeschlagen werden. Unter Verwendung dieser in CaRe zur Verfügung gestellten Daten und Analysemöglichkeiten soll die onkologische Ressource Wissenschaftler dabei unterstützen, Arbeitsschritte zu verfolgen, die in verschiedene Richtungen gehen. Abbildung 3.23 zeigt die Architektur der Ressource.

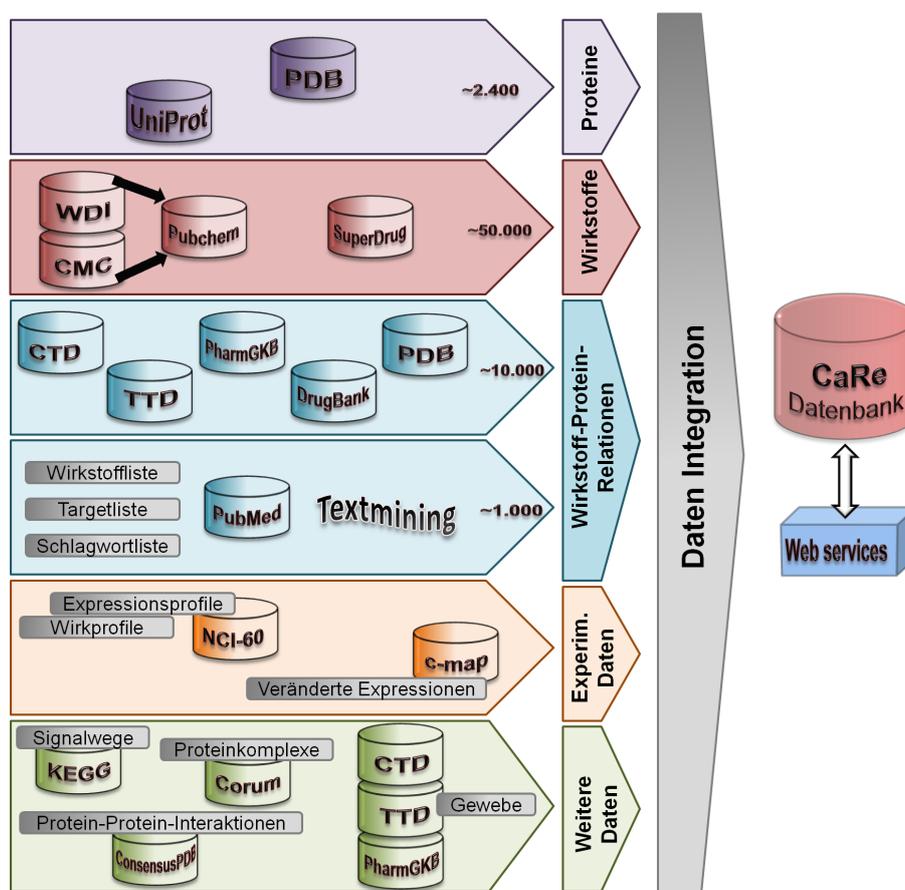


Abb. 3.23: Architektur der CaRe-Datenbank.

Onkologisch relevante Kleinstrukturen

Die CaRe beinhaltet aktuell rund 50.000 Kleinstrukturen mit Anti-Tumor Effekt, die aus den folgenden Quellen stammten: Developmental Therapeutic Program (DTP), The National Cancer Institut (NCI), SuperTarget, SuperDrug und PubChem. Für einen Wirkstoff erhält ein Nutzer viele zusätzliche Informationen, die in drei Teilen angeordnet sind. Zuerst werden die Signalwege angezeigt, in denen der Wirkstoff Einfluss ausübt. Im zweiten Teil werden die Wirkstoff-Zielprotein-Beziehungen, wobei es

sich hauptsächlich um Interaktionen handelt, mit PubMed-Referenz und Quelle angegeben. Darüber hinaus werden Informationen wie CAS-Nummer, NSC-Nummer, ATC-Code, PubChem-Id und Synonyme angegeben.

Onkologisch relevante Zielproteine

Die onkologisch relevanten Zielproteine stammten aus der UniProt- und Ensembl-Datenbank. Für jedes Protein erhält der Nutzer neben dem Namen noch folgende Informationen über das Protein:

1. Interaktionen zuzüglich Quelle und PubMed-Referenz
2. indirekte Interaktionen zuzüglich Quelle
3. Verlinkungen zur OMIM-Datenbank (Online Mendelian Inheritance in Men) [167], die Mutationsinformationen für das entsprechende Protein enthält
4. Protein-Protein-Interaktionen, die aus der Consensus-Path-Datenbank [168] stammten
5. NCI Expressionsdaten
6. Connectivity Map [12] Expressionsdaten, die eine Verknüpfung zwischen Wirkstoffwirkungsweise und genetisches Material und Krankheiten herstellt
7. Synonyme
8. Verlinkungen zu Datenbanken wie UniProt, iHop [169], Pfam [170], Ensembl [171], Reactome [172], Systers [173]

Neben den PDB-Strukturen der Zielproteine, wurden Proteinstrukturen von ca. 250 Proteinen modelliert, da zwar keine Struktur in der PDB verfügbar war, jedoch ein geeignetes Template.

Kleinstruktur-Zielprotein-Beziehungen

Aktuell enthält die CaRe über 10.800 nicht-redundante Kleinstruktur-Zielprotein-Beziehungen. Über 6.000 sind dabei mit Tumoren, Lymphomen und Leukämie assoziiert. Die onkologisch relevanten Interaktionen stammen aus wichtigen Interaktionsdatenbanken wie TTD [12] und teils aus der CTD [13]. Darüber hinaus sind alle Interaktionen der DrugBank integriert. Auch die in der PDB vorhandenen

Liganden-Protein-Interaktionen wurden integriert. Um den Datensatz zu vervollständigen sind mittels Textmining-Methoden und manueller Validierung 1.000 Interaktionen dazu gekommen (siehe Tabelle: 3.10). Indirekte Interaktionen stammen aus der PharmaGKB. Interessanterweise zeigte eine Analyse, dass die Überlappung der Relationseinträge der einzelnen Datenbanken marginal ist und daher die Integration und die Erstellung einer Ressource mit der Gesamtinformation notwendig war.

Datenbank	K-Z-B redundant	K-Z-B unique
CancerResource; Datenintegration	11,535	10,824
CancerResource; Textmining	1,122	992
TTD	282	163
CTB	3,875	3,748
DrugBank	4,949	4,763
PharmaGKB	1,307	1,158

Tab. 3.10: Zahlen der redundanten und unigen Kleinstruktur-Zielprotein Relationen in CaRe. K-Z-B: Kleinstruktur-Zielprotein-Beziehungen. Aus Ahmed et al. [153].

Onkologisch relevante Signalwege

Insgesamt 44 verschiedene onkologisch relevante Signalwege sind aus der KEGG-Datenbank selektiert und in CaRe integriert. Dabei werden alle Zielproteine zu denen Wirkstoff-Informationen in der CaRe gespeichert sind, markiert (siehe Abbildung 3.24). Durch Klicken auf das Protein in dem Signalweg erhält der Nutzer zusätzlich Informationen zu den Wirkstoffen, die auf das Target wirken. Eine Verlinkung zu den detaillierten Targetinformationen ist hier ebenfalls zu finden. Des Weiteren wird eine Wirkstoff-Zielprotein-Beziehungs-Matrix generiert, in der alle Proteine und Liganden gegeneinander aufgetragen sind und, wenn eine Relation vorhanden ist, die Zelle mit einem Punkt markiert. Abbildung 3.24 zeigt den Apoptose-Signalweg mit den markierten Proteinen für die Wirkstoff-Zielprotein-Beziehungen, die in CaRe gespeichert sind. Für nur neun Proteine sind keine Wirkstoffinformationen in CaRe gespeichert.

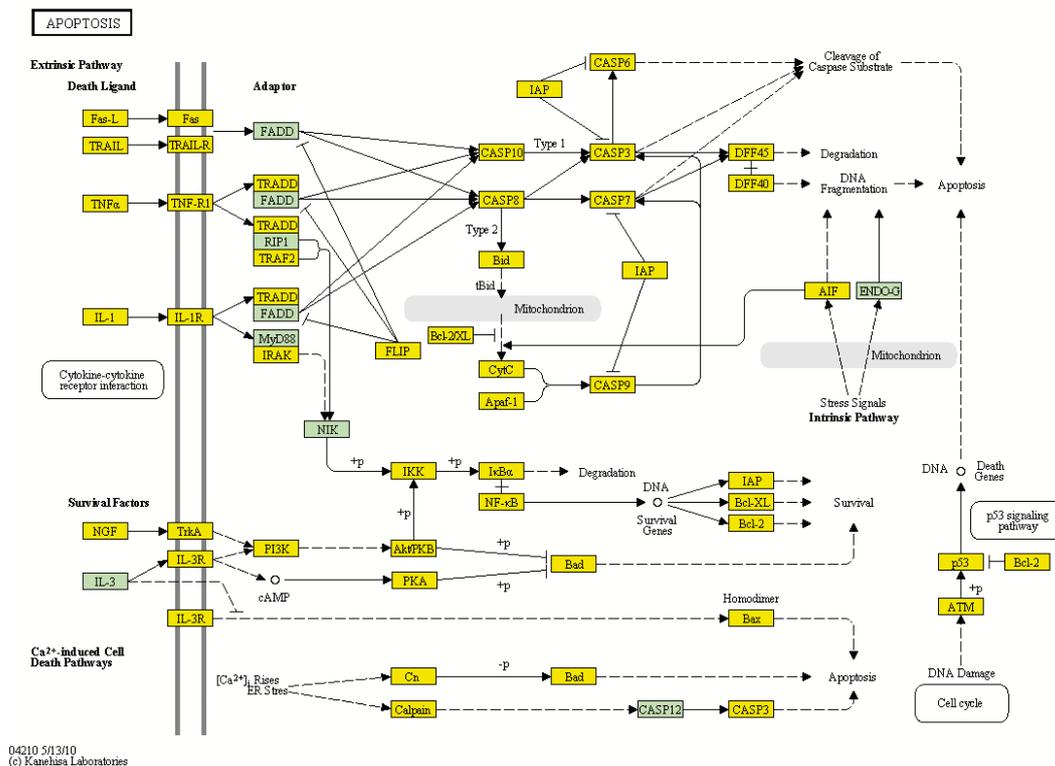


Abb. 3.24: Apoptose Signalweg mit markierten Proteinen für die Wirkstoff Informationen in CaRe gespeichert sind.

Zelllinien und Wirkprofile

Insgesamt sind für über 40.000 Substanzen die Chemosensitivität auf die NCI60-Zelllinien gespeichert. Diese Profile sind in sogenannte zelluläre Fingerprints übersetzt worden, die den schnellen Vergleich und die Visualisierung der Wirkprofile verschiedener Kleinstrukturen oder Kleinstrukturgruppen möglich macht. Die Integration dieses Datensatzes ermöglicht die Suche nach ähnlich wirkenden Substanzen. Dabei kann zum einen nur das Wirkprofil einer Substanz erfragt werden oder auch ähnliche Wirkprofile zu einer hochgeladenen Substanz ermittelt werden. Darüber hinaus gibt es die Möglichkeit, manuell gemessene GI50-Werte einzutragen und diese mit den gespeicherten Zelliniendaten zu vergleichen. Abbildung 3.25 zeigt den Vergleich der zellulären Fingerprints zweier Taxole. Auch Anfragen zu den wirksamsten Substanzen einer Zelllinie sind möglich.

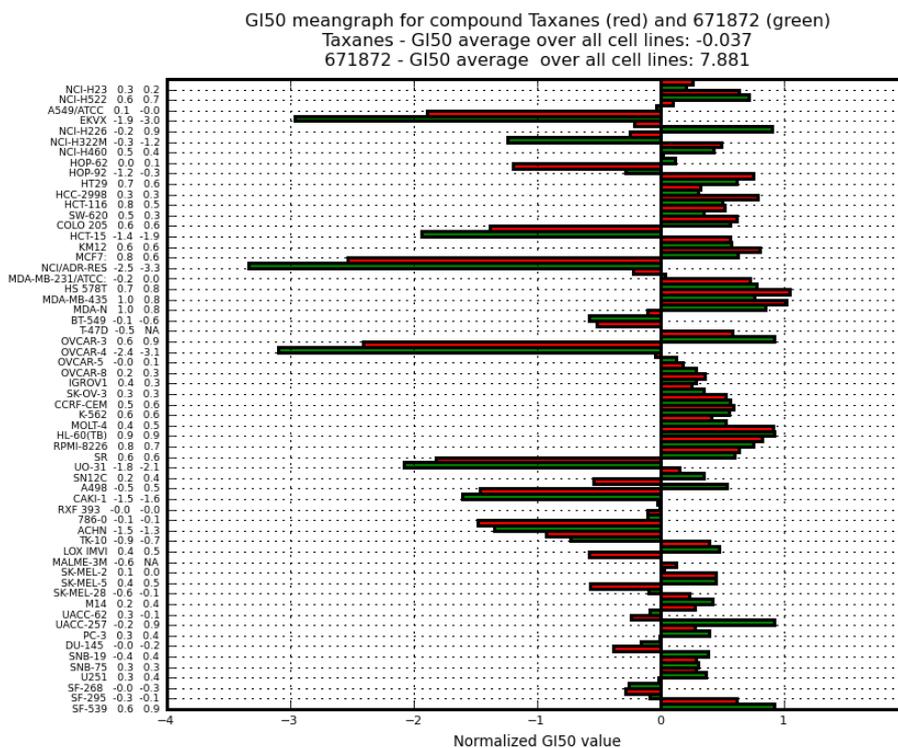


Abb. 3.25: Wirkprofile zweier Taxane.

Expressionsdaten

Die Expressionsdaten sind für die NCI60-Zelllinien aus der DTP extrahiert worden, prozessiert und damit vergleichbar für externe hochgeladene Daten. Darüber hinaus sind auch Veränderungen der Expressionsdaten durch Wirkstoffe in CaRe gespeichert, die aus der Connectivity Map [154] integriert wurden.

Die Integration der Expressionsdaten und der Chemosensitivität der Wirkstoffe (Wirkprofile), ermöglicht den Vergleich der benutzerdefinierten Expressionsdaten und die Ausgabe der möglicherweise wirkenden Substanzen für die hochgeladenen Expressionsdaten. Der Nutzer kann einzelne experimentelle Daten gegen die Expressionsdaten der NCI vergleichen.

Anwendungsbeispiel

Die CancerResource (CaRe) ermöglicht die Suche nach Wirkstoffen ausgehend von Expressionsdaten. Besitzt der Nutzer beispielsweise individuelle Expressionsdaten können diese in CaRe hochgeladen werden. Diese werden anschließend prozessiert und der Pearson Korrelationskoeffizient der hochgeladenen Daten und der Expressionen der 60 Zelllinien berechnet und die ähnlichste Zelllinie bestimmt. Die Ausgabe

liefert zusätzlich die Substanzen mit der größten Inhibitionswirkung auf die ähnlichste Zelllinie bzw. eine Matrix mit Targets und Inhibitoren (siehe Abbildung 3.26).

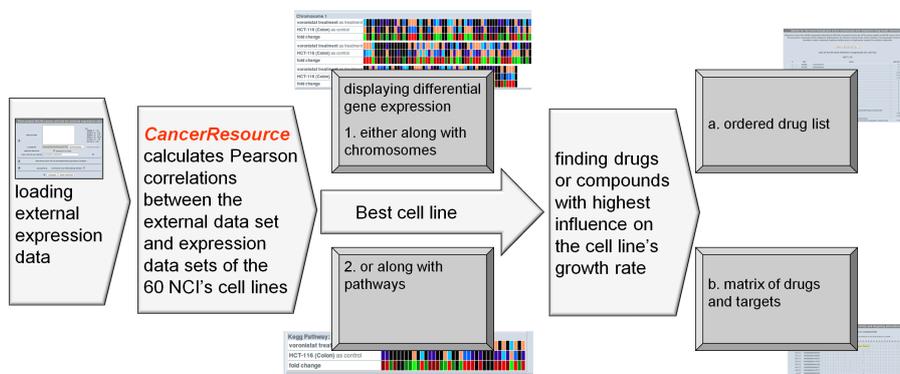


Abb. 3.26: Anwendungsbeispiel der CaRe-Datenbank.

Individuelle Daten können hochgeladen und mit gespeicherten Daten bzgl. der Expressionsprofile verglichen werden. Das Ergebnis liefert die ähnlichste Zelllinie, hochregulierte Targets und potenzielle Wirkstoffe. *Aus Ahmed et al. [153].*

Vergleich der Anwendungsmöglichkeiten der CancerResource und anderen ähnlichen Datenbanken:

Im Folgenden werden wichtige Merkmale und Eigenschaften der CaRe aufgelistet, wobei Tabelle 3.11 andere Datenbanken bezüglich der Eigenschaften mit CaRe vergleichend darstellt.

1. Zelluläre Fingerprints beinhalten Wirkprofile der 60 NCI-Zelllinien auf ca. 40.000 Substanzen
2. Bestimmung der aktuellen wirksamsten Substanz auf einer der Zelllinien mit Hilfe der zellulären Fingerprintdaten. Diese Art der Analyse könnte einen ersten Schritt in Richtung individualisierter Therapie darstellen.
3. Pre-prozessierte Genexpressionsdaten für die 60 Zelllinien
4. Selektion der ähnlichsten NCI60-Zelllinie nach Vergleich mit benutzerdefinierten Expressionsdaten
5. Visualisierung der Expressionsdaten mit einer farblichen Matrix für selektierte Gene

6. Vergleich der Expressionsdaten mit externen, benutzerdefinierten somatischen onkologischen Daten, um Ähnlichkeiten zwischen Zelllinien zu bestimmen oder für die Erkennung von unterschiedlich exprimierten Genen
7. Visualisierung der unterschiedlichen Genexpressionen, die durch Anwendung verschiedener selektiver Substanzen begründet sind (Connectivity Map Daten)
8. Targetgene und Substanzen sind in onkologisch relevante Signalwege der KEGG-Datenbank projiziert
9. Die Ressource projiziert Expressionsdaten auf onkologisch relevante KEGG Signalwege und berechnet eine Abschätzung der Überrepräsentation in jedem einzelnen Signalweg
10. Substanz-Zielprotein-Matrizen zeigen einzelne Proteine, die von mehreren Wirkstoffen adressiert werden als auch Substanzen, die mehrere Zielproteine binden
11. ermöglicht den Vergleich der direkten und indirekten Substanz-Effekte, die durch Textmining-Methoden oder durch Microarray Daten bestimmt wurden

	CTD	TTD	PharmaGKB	DrugBank	NCI	CaRe
1	-	-	-	-	+	+
2	-	-	-	-	-	+
3	-	-	-	-	+	+
4	-	-	-	-	(-)	+
5	-	-	-	-	-	+
6	-	-	-	-	-	+
7	-	-	-	-	-	+
8	++	+	++	+	-	+
9	-	-	-	-	-	+
10	-	-	-	-	-	+
11	-	-	-	-	-	+

Tab. 3.11: Vergleich der Eigenschaften der Datenbanken und CaRe.

Des Weiteren soll die Datenbank um vorhandene Daten von Zelllinienexperimenten und Tumorpatienten von verschiedenen Kooperationspartnern erweitert werden.

3.3.2 SuperPred - Target-Vorhersage

SuperPred ist ein Webservice mit integrierter Targetvorhersage basierend auf der Annahme, dass strukturell ähnliche Kleinstrukturen auch in dieselbe Kategorie des Anatomisch-Therapeutisch-Chemischen-Klassifikationssystems (ATC) eingeordnet und ähnliche Targets besitzen oder zumindest Hypothesen über die adressierten Proteine aufgestellt werden können. Für die Interpretation der Vorhersage gehen auch physikochemische Eigenschaften ein. Der Webserver ist frei zugänglich und unter <http://bioinformatics.charite.de/superpred/> nutzbar. Der Webserver basiert auf eine Datenbank bestehend aus ca. 2.500 ATC-klassifizierten Wirkstoffen, ca. 4.000 vorhergesagten Wirkstoffen und, wenn vorhanden (für mehr als die Hälfte), die zugehörigen Targetinformationen. Die Wirkstoffe mit vorhergesagter ATC-Klassifikation sind bei der Nutzung des Servers farblich hervorgehoben.

Hochgeladene oder in SuperPred gezeichnete Kleinstrukturen werden in einen strukturellen Fingerprint übersetzt und mit den Einträgen der Datenbank verglichen. Der ATC-Code und die potenziellen Zielproteine des strukturell ähnlichsten Wirkstoffs werden für die hochgeladene Kleinstruktur als ATC-Code und die adressierten Zielproteine angenommen. Auch das physikochemische Eigenschaftsprofil der angefragten Substanz, die Grenzprofile der ATC-Klasse und die Anzahl der Lipinski 5er Regel-Verstöße werden angegeben. Diese Informationen können bei der Interpretation der Vorhersage hilfreich sein. Neben den physikochemischen Eigenschaften erhält der Nutzer weiterführende Informationen in Form von Links zu den Signalwegen, in denen die Zielproteine involviert sind.

SuperPred bietet dem Nutzer drei verschiedene Möglichkeiten seine Kleinstruktur hochzuladen und klassifizieren zu lassen:

1. Hochladen einer Struktur über ein Plugin (Marvin)
2. Zeichnen der zu untersuchenden Struktur mit Hilfe des Plugins
3. Eingabe des SMILES-Codes und anschließender Import in das Plugin

Als Ergebnis erhält der Nutzer eine nach dem Tanimoto-Koeffizienten sortierte Tabelle mit den ähnlichsten Kleinstrukturen der Datenbank, das Eigenschaftsprofil der angefragten Substanz und die obere und untere Eigenschaftsprofilgrenze der potenziellen ATC-Klasse. Für jeden Treffer werden sowohl der ATC-Code als auch die

von dem ähnlichsten Treffer adressierten Zielproteine angegeben. Auch die physikochemischen Eigenschaften der Substanzen sind anzeigbar. Des Weiteren werden die Tanimoto-Koeffizienten farblich hervorgehoben, wobei die jeweilige Farbe angibt, wie wahrscheinlich die Vorhersage ist. Die Farbskala reicht von grün bis rot, wobei grün für eine hohe und rot für eine geringe Wahrscheinlichkeit einer korrekten Vorhersage kodiert. Die detaillierte Tabelle wird bei jedem Treffer als Legende angegeben.

Neben der Substanzklassifikation ermöglicht SuperPred auch die Targetsuche mittels Volltextsuche für folgende Eigenschaften:

1. Targetname
2. Kleinstrukturname
3. Accession-Nummer oder Genname
4. ATC Code

Das Ergebnis sind die bindenden Wirkstoffe und die Targetinformation. Des Weiteren ist hier eine Ähnlichkeitssuche möglich, welche die ähnlichsten 10, 20 oder 30 Kleinstrukturen ausgibt.

Mit der Implementierung eines ATC-Baumes stellt der Webserver eine weitere Suchmöglichkeit zur Verfügung. Der Baum ermöglicht das „Browsen“ in der Datenbank anhand der ATC-Klassen.

Statistische Validierung der Vorhersage-Genauigkeit

Um die Robustheit und die Genauigkeit der Methode zu bestimmen wurde die in Kapitel 2.2.6 beschriebene „Leave-one-out“-Methode verwendet. Für die statistische Auswertung wurde eine Untergruppe bestehend aus 1.035 Arzneistoffen verwendet. Abbildung 3.27 zeigt die Anzahl der Arzneistoffe, die durch Vorhersage der entsprechenden ATC-Codes richtig klassifiziert wurden. Die Treffer wurden bzgl. des Tanimoto-Scores aufsteigend sortiert.

Die Aggregation ähnlicher ATC-Codes ergibt eine Genauigkeit von etwa 70%. Mit steigendem Tanimoto-Wert, nimmt die Zuverlässigkeit der Prognosen zu.

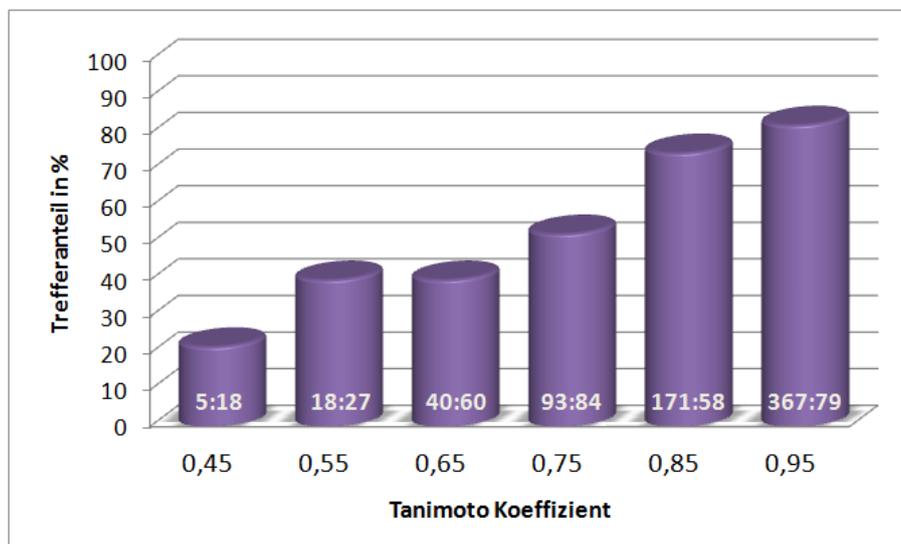


Abb. 3.27: Ergebnis der Vorhersage-Genauigkeit mit der Leave-One-Out-Methode. Die Grafik visualisiert die prozentualen Vorhersagetreffer bezüglich verschiedener Tanimoto-Koeffizienten. Die Zahlen in den Balken stehen für: Anzahl der positiven Treffer / Anzahl der falsch positiven Treffer bei dem entsprechendem Tanimoto-Wert. Die Höhe der Balken entspricht dem Verhältnis. Zu erkennen ist, dass die Vorhersage-Genauigkeit mit steigendem Tanimoto-Wert zunimmt.

Anwendungsbeispiel

Das „National Cancer Institute“ beinhaltet eine große Anzahl an auf Tumorzelllinien wirkenden Screening-Substanzen, jedoch mit unbekanntem Zielprotein-Informationen. Ein Beispiel stellt die NCI Substanz 600221 (600220) dar. Die Anwendung des SuperPred-Servers mit dem NCI-Wirkstoff als Ausgangspunkt, liefert mit einer 92% igen Ähnlichkeit Paclitaxel, einem antineoplastischen Arzneistoff, welcher unter anderem Tubulin adressiert. Die Tabelle 3.12 zeigt die berechneten physikochemischen Eigenschaftsprofile. Das Eigenschaftsprofil der Substanz 600221 liegt nicht innerhalb des durch die Grenzeigenschaftsprofile bestimmten Bereichs. Außerdem verstößt der Wirkstoff NSC600221 gegen drei der Lipinski 5er Regeln, was auf eine schlechte orale Bioverfügbarkeit der Substanz schließen lässt. Das gilt auch für das Chemotherapeutikum Paclitaxel.

Substanzname	H-acc	H-don	mol. Gewicht	logP	Verstöße d. Lipinski 5er
NSC600221	18	4	953	28	3
Paclitaxel	15	4	853	28	3
obere Grenze L01CD	15	5	853	28	3
untere Grenze L01CD	15	4	807	17,6	3

Tab. 3.12: Physikochemische Eigenschaftsprofile der Substanzen NSC600221 und Paclitaxel und die obere und untere Eigenschaftsprofilgrenze der ATC-Gruppe L01CD.

3.3.3 FragmentStore - Fragmentierung und Analyse

FragmentStore

Die Ressource FragmentStore ist unter

http://bioinf-applied.charite.de/fragment_store/ erreichbar und besteht aus über 200.000 verschiedenen Fragmenten, die durch die Fragmentierung der mehr als 13.000 KEGG-Metaboliten, 2.200 toxischen Kleinstrukturen aus der SuperToxic und 16.000 Arzneistoffen verschiedener Arzneistoff-Datenbanken mit Hilfe dreier verschiedener Fragmentierungsstrategien (siehe Kapitel 2.1.3) generiert wurden. Über 150.000 verschiedene Fragmente entstanden durch die rekursive Fragmentierung an den rotierbaren Bindungen, wobei diese Strategie Fragmente generiert, die sich teilweise nur marginal unterscheiden. Zwei weitere angewandte Fragmentierungsstrategien nach „Linkern“ und nach „Recap-Regeln“ lieferten einen Datensatz bestehend aus 35.000 unterschiedlichen Fragmenten. Dieser Datensatz weist eine höhere Diversität zwischen den einzelnen Fragmenten auf und wird als Analysedatensatz vorgeschlagen.

Abbildung 3.28 zeigt die durchschnittlichen Eigenschaften der Kleinstrukturen der drei Substanz-Klassen:

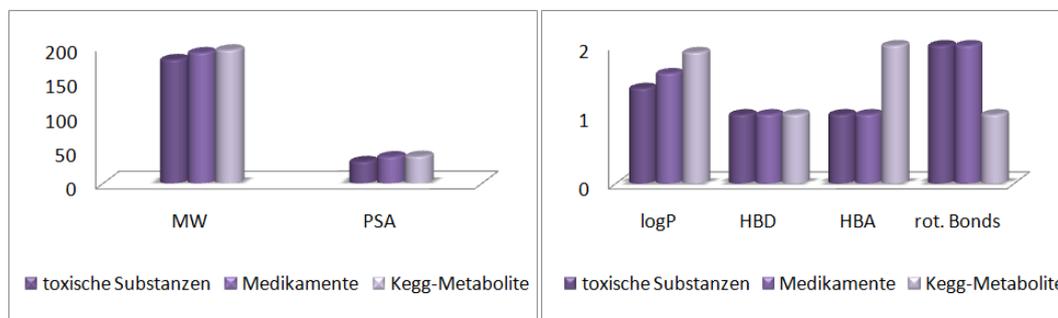


Abb. 3.28: Mittelwerte der Eigenschaften der 35.000 FragmentStore Fragmente, generiert mit der Recap- und Linker-Schnittweise. MW: Molekulargewicht in g/mol; PSA: Polar Surface Area in Å²; HBD: Wasserstoffbrücken-Donatoren; HBA: Wasserstoffbrücken-Akzeptoren; rot. Bonds: Anzahl der rotierbaren Bindungen.

Für alle Fragmente des FragmentStores sind folgende Informationen gespeichert und in der Detailansicht ausgegeben:

- Molekulargewicht
- clogP
- Anzahl der Wasserstoffbrücken-Donatoren und -Akzeptoren

- Anzahl der drehbaren Bindungen
- die „Polar Surface Area“ (PSA)
- Ligandenname und Verlinkung zu dem ursprünglichen Datenbankeintrag, von dem das Fragment ein Teil ist
- Aminosäure-Präferenz als Histogramm, wenn das Fragment Teil mindestens einer in der PDB vorhandenen Kleinstruktur ist

Darüber hinaus wird die Struktur des Fragments visualisiert. Ist das Fragment Teil einer PDB-Kleinstruktur, stellt der FragmentStore eine drei-dimensionale Visualisierung des Fragments in der Bindungstasche des Zielproteins zur Verfügung. Bei der Anwesenheit von mindestens zwei unterschiedlichen Proteinstrukturen, sind die Fragmente und die Bindungstaschen überlagert und visualisiert. Diese Visualisierung erlaubt außerdem das An- und Ausschalten von interessanten Aminosäuren der Bindungstasche des Fragments und erleichtert eine visuelle Analyse.

Darüber hinaus bietet der FragmentStore noch weitere Suchmöglichkeiten:

- Fragmente können anhand der Namen ihres Wirkstoffs selektiert werden
- Fragmente können anhand der Regel selektiert werden, nach denen sie generiert wurden, nach ihren physikochemischen Eigenschaften oder anhand der Zugehörigkeit der ursprünglichen Substanz zu Metaboliten, toxischen Substanzen oder Arzneistoffe
- es ist möglich Fragmente aufgrund ihrer Bindungstaschenpräferenz zu selektieren
- Ein Enzymklassifikationsbaum (EC-Baum) kann durchsucht werden und die Fragmente der Liganden werden ausgegeben
- eine weitere Browsingmöglichkeit bietet der Scop-Baum
- strukturelle Ähnlichkeitssuchen können durchgeführt werden, indem ein Molekül gezeichnet oder hochgeladen wird

Darüber hinaus bietet der FragmentStore die Möglichkeit, drei Fragmente auszuwählen, für die mittels des Fragmentassemblierungs Programms eine Kleinstrukturbibliothek erstellt wird, die der Lipinski 5er Regel entspricht. Die Generierung der

Bibliothek für maximal drei Fragmente dauert maximal 20 Minuten und wird als Email an den Nutzer versendet. Der Nutzer erhält eine Datei mit SMILES, die er in ein entsprechendes Programm laden kann, um die generierten Moleküle zu visualisieren.

Anwendungsbeispiel

Bei der Suche nach potenziellen Wirkstoffen für eine Bindungstasche eines bestimmten Zielproteins, stellt der erste Schritt die Charakterisierung der Bindungstasche dar. Die Aminosäurezusammensetzung als auch die Eigenschaftsprofile der Bindungstasche können hierbei eine Rolle spielen, um anschließend nach Fragmenten zu suchen, die spezifisch Teile der Bindungstasche adressieren können. Die Fragmente, die bestimmte Bindungstaschen präferieren, sind in der FragmentStore-Ressource zu finden. Beinhaltet die Bindungstasche beispielsweise bevorzugt hydrophobe Aminosäuren wie Methionin, erlaubt die Ressource die Suche nach Fragmenten, die bestimmte Bindungstaschen adressieren. In diesem Beispiel sucht der Nutzer nach Fragmenten, die bevorzugt in hydrophoben Bindungstaschen binden. Neben den physikochemischen Eigenschaften der Fragmente, die bevorzugt Aminosäuren wie Methionin binden, erhält der Nutzer auch einen Präferenz-Plot. Abbildung 3.29 zeigt ein Beispiel ausgehend von der Suche nach Fragmenten mit methioninreicher Aminosäurepräferenz. Darüber hinaus liefert die Ressource eine Visualisierung überlagert Fragmente und dessen Bindungstaschen, in denen das Fragment in der Ursprungskleinstruktur vorkommt.

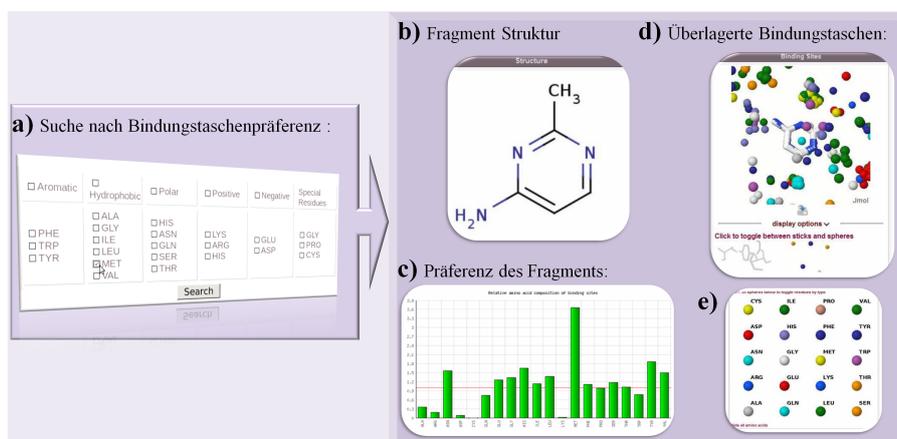


Abb. 3.29: Anwendungsbeispiel des FragmentStores. a) Formular zur Suche nach Fragmenten mit bestimmten Bindungstaschenpräferenzen (BTP) b) Struktur des Fragments mit der gewünschten BTP c) BTP als Histogramm d) überlagerte Fragmente inkl. der Bindungstaschen (BT) e) Formular zur Aktivierung/Inaktivierung der Aminosäuren in der BT. Aus Ahmed et al. [127].

Erste Analysen

In der folgenden Analyse wurden nur Fragmente betrachtet, die nach Fragmentierung der SuperDrug, der KEGG-Metabolite und der toxischen Kleinstrukturen der SuperToxic bei Anwendung der Linker-Schnittweise entstanden sind. Die Überlappung der Ligandendatensätze ist in Abbildung 3.30 gezeigt. Wie erwartet, ist die Anzahl der überlappenden Liganden zwischen Metaboliten und Arzneistoffen größer als die zwischen Metaboliten und toxischen Liganden. Ursprünglich lag auch eine Überlappung zwischen toxischen Liganden und Wirkstoffen vor, jedoch wurden diese doppelten Einträge aus beiden Datensätzen entfernt, um Fragmente zu identifizieren, die aufgrund ihrer toxischen Wirkung für stärkere Nebenwirkungen der Wirkstoffe verantwortlich sind. Nach Fragmentierung der Liganden und Analyse der Fragmente konnte festgestellt werden, dass nur eine geringe Anzahl an gemeinsamen Fragmenten zwischen den drei Klassen vorliegen. Wie erwartet sind weniger überlappende Fragmente zwischen den toxischen und metabolischen Fragmenten im Vergleich zu den gemeinsamen Fragmenten der metabolischen und medizinisch relevanten Fragmente entstanden. Interessanterweise teilen sich die toxischen und metabolischen Substanzen einige Fragmente, obwohl zuvor keine identischen Liganden in den jeweiligen Datensätzen vorhanden waren. In Abbildung 3.30 wird ein Fragment dargestellt, das sowohl in den toxischen Fragmenten als auch im medizinisch relevanten Wirkstoff-Fragmentdatensatz zu finden ist. Eine genaue Analyse zeigt, dass das Fragment kein Fragment in metabolischen Substanzen ist. Das

Fragment ist Teil eines Chemotherapeutikums, dem Prednimustine [174] und Teil mehrerer toxischer Kleinstrukturen, wie dem 4'-(di-2chloroethylamino)-4-hydroxy-3-methyldiphenylamine. Diese Kleinstruktur hat einen LD50-Wert von 1.43mg/kg (i.p) in Ratten und ist damit sehr toxisch [175].

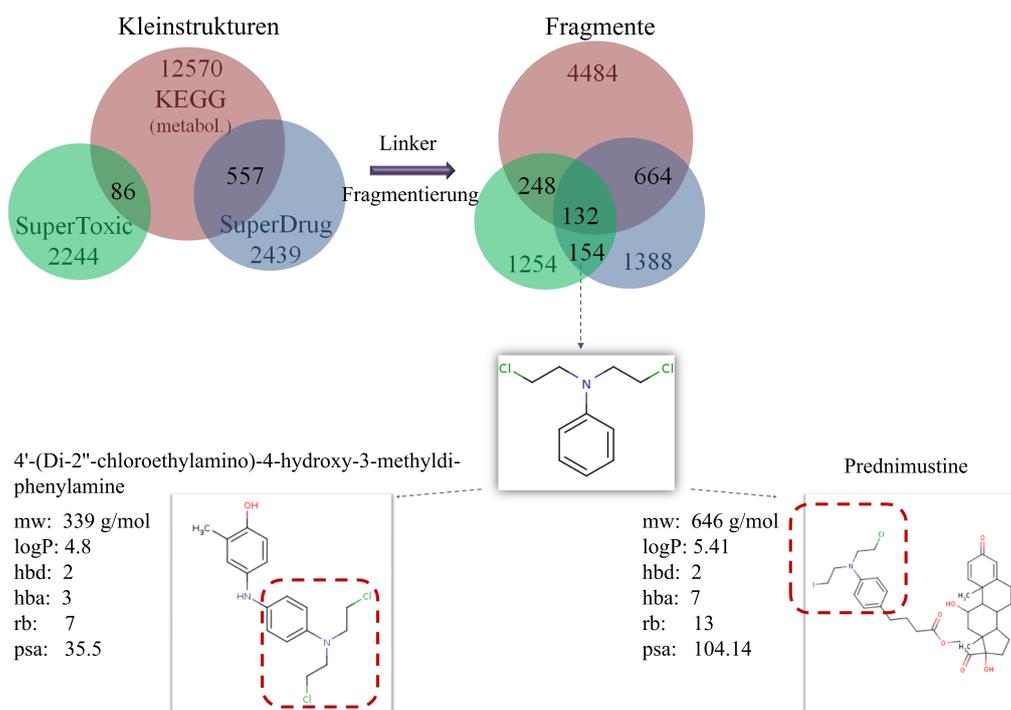


Abb. 3.30: Vergleich der gemeinsamen Strukturen der metabolischen, toxischen und Wirkstoffdaten auf Kleinstruktur und Fragmentebene.

Oben: Darstellung der Schnittmengen der Kleinstrukturen und Fragmente. Mitte: Beispiel eines Fragments welches sowohl in toxischen als auch in pharmazeutischen Substanzen vorkommt. Unten: Prednimustine (rechts) aus der SuperDrug und eine Kleinstruktur aus der SuperToxic (links). Aus Ahmed et al. [127].

Kapitel 4

Diskussion

4.1 Evolutionäre Aspekte

Die Beobachtung, dass Primaten kaum an onkologischen Erkrankungen leiden wurde in einigen Studien beschrieben [3, 2]. Mögliche Erklärungen für die unterschiedliche Inzidenz der malignen Tumoren zwischen Schimpanse und Mensch könnten durch die kürzere Lebenserwartung, Infektionen oder fehlende medizinische Versorgung, Ernährung und gesündere äußere Umstände gegeben sein [176]. Es ist jedoch wahrscheinlich, dass die Ursache nicht nur in ökologischen, sondern auch in genetischen Faktoren zu suchen und zu finden ist.

In dieser Arbeit konnten auf verschiedenen Analyse-Ebenen Unterschiede zwischen Schimpansen-Genen und humanen Genen gefunden werden. Daraus resultiert eine veränderte Biosynthese der Proteine, die in aus Sicht der Onkologie relevanten Signalwege wie Apoptose, DNA-Reparatur und Zell-Zyklus-Regulation eine Rolle spielen. Diese Unterschiede könnten Hinweise auf die unterschiedliche Tumorzinzenz liefern. Es ist allerdings auch nicht ausgeschlossen, dass es sich um Sequenzierungs- und Annotationsfehler handelt. Das muss Gegenstand weiterer Untersuchungen sein. Der Analysedatensatz bezog sich auf die Ensembl-, GeneBank- und UniProt-Eintragungen von Januar 2007. Im Durchschnitt wiesen die verglichenen Gene neun Substitutionen pro Gen auf. Puente et al. zeigte, dass das für Brustdrüsenkarzinom verantwortliche Gen und das p53-Gen im Schimpansen-Genom im Vergleich zum humanen Genom verändert sind [6]. In dieser Arbeit wurden die Nukleotidsubstitutionen nicht weiter analysiert. Sie wurden hauptsächlich für die Berechnung der positiven Selektion ermittelt. Insgesamt wiesen 13 Gene, unter anderem das BRCA1-

Gen, eine positive Selektion auf. Die positive Selektion des BRCA1-Gens und seine Entwicklung wurden schon in vorangegangenen Arbeiten beschrieben [99, 8]. Die Analyse von Nielsen et al. von über 13.000 annotierten Schimpansen- und humanen Genen zeigt, dass bevorzugt Gene der Apoptose, der Zellzyklus-Kontrolle und der Tumorsuppression auf eine positive Selektion hinweisen [99], während die Arbeit von Puente et al. bei direktem Vergleich der 333 onkologisch relevanten Gene eine hohe Konservierung feststellte [6].

Einige Gene beider Spezies setzten sich aus unterschiedlichen Exon-Anzahlen und unterschiedlichen Intron-Längen zusammen. Die betroffenen Gene des Schimpansen wiesen im Durchschnitt eine größere Anzahl an Exons und längere Introns auf als die orthologen humanen Gene. Der Unterschied in der Anzahl der Exons kann zu verschiedenen Splice-Varianten und damit zu einer größeren Varianz an Proteinen führen. Die Introns hingegen kodieren für regulatorische RNAs, was zu der Vermutung führt, dass das Schimpansen-Genom eine größere Anzahl dieser regulatorischen RNAs besitzt, die möglicherweise auch in der Regulation der Apoptose oder bei der Replikation der Apoptose-relevanten Gene involviert sein könnten. Selektive siRNAs könnten die Möglichkeit bieten, deren Auftreten in Tumorzelllinien zu analysieren. Weis et al. identifizierte eine schnellere Evolution der nicht-kodierenden Sequenzen der DNA-Reparatur Gene des Schimpansen verglichen mit den humanen orthologen Genen [177]. Darüber hinaus wurden von Weis et al. auch die Gene weiterer Primaten wie die des Makaken einbezogen.

Neben möglichen Sequenzierungsfehlern könnten auch allgemeine Gen-Annotationsfehler eine Erklärung für die gefundene unterschiedliche Genstruktur sein. In dem Fall eines Annotationsfehlers handelt es sich bei dem in der GenBank annotierten Gen nicht um das Gen, welches für die Transkription des Proteins verantwortlich ist, sondern lediglich um ein Duplikat an anderer Stelle. Auch die Pseudogenisierung und damit Inaktivierung des E2F3 Gens im Schimpansen-Genom könnte auf einen Sequenzierungsfehler zurückzuführen sein. Die Ensembl Datenbank speicherte im Gegensatz zur GenBank ein nicht pseudogenisiertes E2F3 Gen, was entweder ein Artefakt in der Ensembl Datenbank oder in der GenBank sein könnte. Hier sind weitere experimentelle Validierungen notwendig, um zu prüfen, welche dieser Aussagen richtig ist. Darüber hinaus werden die Sequenzdatensätze des Menschen- und insbesondere des Schimpansen-Genoms regelmäßig verbessert und aktualisiert und die Inhalte der Ressourcen beinhalten schrittweise die gleichen Sequenzen, so dass

eine Analyse zu einem späteren Zeitpunkt mehr Klarheit bringt. Bisher wurde eine Analyse der Pseudogenisierung für das Chromosom 22 durchgeführt [178]. Watanabe et al. identifizierten für ca. acht Gene neben mehreren Duplikaten auch bis zu hundert Pseudogene auf Chromosom 22. Beispielsweise besitzt das Gen HSA21q 284 kodierende Sequenzen und 98 Pseudogene.

In einer Vergleichsanalyse des Schimpansen- und Humanproteoms von Glazko et al. wiesen die Proteine im Durchschnitt zwei Aminosäure-Substitutionen pro Protein [179] auf, während in dieser Arbeit im reduzierten Apoptose Datensatz die Proteine ganze acht Veränderungen pro Protein aufwiesen. Es wurden für rund 500 Proteine 4.111 Aminosäureveränderungen identifiziert, während Puente et al. nur 1.542 Aminosäureveränderungen in 333 verschiedenen onkologisch relevanten Proteinen identifiziert haben. Diese Differenz könnte entweder mit einem unterschiedlichen Proteindatensatz oder mit der Aktualität des Datensatzes zum Zeitpunkt der Analyse zusammenhängen.

Für über 200 Proteine konnten die Proteinstrukturen analysiert werden. Zehn Proteine besaßen in der Bindungstasche eine Aminosäureveränderung, was die Bindungsstärke zu einem natürlichen Liganden beeinflussen könnte. Beispielsweise könnte die negative Ladung der Phosphatgruppe und die positive Ladung des Lysins in der Bindungstasche der Topoisomerase II alpha des Schimpansen eine stärkere Bindung des ATPs ermöglichen.

In einem nächsten Schritt müssten die Apoptose- und die DNA-Reparatur-Signalwege rekonstruiert werden, um den Einfluss der Veränderung der Proteine analysieren zu können.

Eine weitere sehr interessante Arbeit von Weis et al. analysierte die Zellantwort bzgl. DNA-Schäden und stellte fest, dass die Zellen des Schimpansen einige DNA-Schäden schneller reparieren als die humanen Zellen [177]. Dieser Befund rechtfertigt im Nachhinein die Hinzunahme der in der DNA-Reparatur involvierten Gene in die durchgeführte Analyse.

Die Analyse der onkologisch relevanten Gene des Schimpansen und Menschen auf den verschiedenen Ebenen lieferte im Vergleich zur Arbeit von Watanabe et al., die im Detail Chromosom 22 beider Spezies vergleicht, einen Überblick und in erster Linie Hinweise, die für die unterschiedliche Tumorentstehung verantwortlich sein könnten. Dennoch motivieren die interessanten Ergebnisse dazu, den einen oder anderen Bereich näher zu beleuchten. Vor allem die Analyse der Genstruktur und

Pseudogenisierung ist in anderen Arbeiten nur marginal bis gar nicht untersucht worden. Auch weitere Untersuchungen der Auswirkung der Nukleotidsubstitutionen auf die Funktion des Proteins könnte die Antwort auf die unterschiedlichen Tumorzinidenzen beider Spezies liefern. Die hier präsentierten Resultate bestätigten vorangegangene Arbeiten. Die Analysen müssen jedoch fortgesetzt werden, weil sich nur so genaue Aussagen treffen lassen können. Die Wiederholung der Analysen mit einem aktuellen Sequenzdatensatz ist sinnvoll, da sich vor allem die Sequenzierung des Schimpansen-Genoms zum Zeitpunkt der Analyse noch in der Entwicklung befand. Auch die Berücksichtigung weiterer Primatengenome könnte im Rahmen einer vergleichenden Untersuchung mehr Aufschluss geben.

4.2 γ -Sekretase Modulatoren

Die Suche nach neuen Wirkstoffkandidaten ist eine der schwierigsten und herausforderndsten Aufgaben in der bioinformatischen Wissenschaft. Struktur-basierte Wirkstoffentwicklung entwickelte sich zu einem festen Bestandteil der Wirkstoffentwicklungs-Pipeline. Voraussetzung für Struktur-basiertes Design ist jedoch eine 3D-Proteinstruktur. Diese ist oftmals nicht vorhanden, so dass die Methoden in diesen Fällen auf Ähnlichkeitssuchen in großen Substanz-Bibliotheken beschränkt sind. Hier ist ein erfolgreiches Beispiel, wie mit Hilfe von über 50.000 experimentell bestimmten Proteinstrukturen und die Anwendung der Analogie- und Homologie-Modellierung dennoch die Struktur-basierte Wirkstoffentwicklung ermöglicht wurde. Aktuell ist eine große Anzahl von Strukturfaltungen bekannt, während die Anzahl der Neuentdeckungen bisher unbekannter Faltungstypen drastisch gesunken ist. Dadurch ist es für die meisten Proteine, inklusive Membranproteine, möglich, Homologe mit bekannter Struktur oder zumindest Proteine mit ähnlicher Faltung zu identifizieren. Die Faltungsvorhersage oder „Threading“, die vielmehr Strukturmodelle von mittlerer Qualität liefert, müssen mit Liganden-unterstützten Modellierungsmethoden kombiniert werden, um die Qualität der vorhergesagten Strukturen, die dann in Protein-Liganden-Docking-Versuchen mit potenziellen Wirkstoffen genutzt werden, zu verbessern. Mit Hilfe der Analogie-Modellierung konnten neue potenzielle Wirkstoffkandidaten, die die γ -Sekretase inhibieren, identifiziert werden. Dabei bestand der erste Schritt in der Modellierung der Presenilin-Struktur. An dieser Stelle muss hervorgehoben werden, dass es sich bei der modellierten dreidimensionalen Struktur des Presenilins um eine Annäherung bzw. einen Vorschlag handelt. Die genaue Struk-

tur lässt sich nur mit Hilfe der experimentellen Strukturaufklärung entschlüsseln. Im April 2010 wurde von V. Dötsch eine erste Proteinstruktur der CTF-Untereinheit des Presenilins mit der ID 2KR6 in der PDB veröffentlicht [180]. Die aufgeklärte Struktur des CT-Fragments beginnt mit dem Methionin an Position 292 und beinhaltet u.a. die vollständigen Helices acht bis zehn, einschließlich des katalytischen Aspartatrestes an Position 385. Daher sollte in einem nächsten Schritt in die modellierte Struktur die aufgeklärte CTF-Untereinheit eingebaut werden und das Verhalten bzgl. der Bindungstaschenkonformation nach Energieminimierung und Moleküldynamik Simulationen analysiert werden. Das Einfügen der aufgeklärten Teilstruktur führt zu einer optimierten modellierten Bindungstasche. Während der Modellierung wurde hauptsächlich Wert auf die genaue Modellierung der Bindungstasche gelegt, da diese für die Generierung neuer Inhibitoren erforderlich war, jedoch würde die Information der neuen Proteinstruktur bei einer Optimierung dieser helfen können. Das Ergebnis der *in silico* Optimierung mit Hilfe der modellierten Bindungstasche und die anschließende experimentelle Validierung bestätigt jedoch erst mal das Modell der Presenilin-1 Bindungstasche.

Die von der AG Dr. Jundt durchgeführten experimentellen Validierungen bestätigten die Wirkung der vorgeschlagenen Inhibitoren auf Zellen des Multiplen Myeloms [162].

ADMETox-Analysen:

Die ADMETox-Analysen helfen bei der Abschätzung der oralen Bioverfügbarkeit der Kleinstrukturen. Jedoch sollten die Eigenschaften nicht als Ausschlusskriterium angewendet werden. Speziell die Toxizitätswahrscheinlichkeit ist schwer abzuschätzen. Nach dieser Analyse besaßen alle Kleinstrukturen eine hohe Lebertoxizitätswahrscheinlichkeit und hätten daher ausgeschlossen werden müssen. Für den im Experiment wirksamsten Inhibitor 1 wurde eine Toxizitätswahrscheinlichkeit von über 0.5 vorhergesagt, wobei die vorhergesagten LD50-Werte für alle Substanzen größer als 25 mg/kg waren und damit zumindest auf dieser Ebene nicht toxisch wirken sollen. Inhibitor 1 besaß sogar den niedrigsten vorhergesagten LD50-Wert verglichen mit den Werten der anderen Substanzen. Die ersten experimentellen Voranalysen zeigten für Inhibitor 1 einen LD50-Wert von größer 800 mg/kg, was bedeutet, dass die Toxizität sehr niedrig ist (Persönliche Mitteilung: R. Preißner, F. Jundt). Für zwei Substanzen wurden auch eine höhere Wahrscheinlichkeit der Cytochrom Inhibition vorhergesagt. Da sie bei 0.5 liegt, ist davon auszugehen, dass auch

nur eine experimentelle Validierung Aufschluss über die Cytochrom-Inhibition geben kann. Auch die Vorhersage der Penetrationsfähigkeit der Substanzen der Blut-Hirn-Schranke und ihre Auswirkungen auf Therapieoptionen muss experimentell validiert werden. Jedoch wäre eine Passage der Blut-Hirn-Schranke in einigen Fällen, z.B. bei der Behandlung der Gehirntumore, eher wünschenswert bzw. erforderlich. Bei der Behandlung der Niereninsuffizienz und Fibrose hingegen wäre dieser Effekt eher zu vermeiden.

Für den Ausschluss der Substanzen wurde daher nur die Lipinski 5er Regel verwendet und die ADMETox-Analysen trugen zur Bewertung der Substanzen bei und gaben Hinweise auf durchzuführende experimentelle Untersuchungen.

Die bisher entwickelten γ -Sekretase Inhibitoren führten oft zu sehr starken toxischen Nebenwirkungen, da sie ein breites Spektrum an γ -Sekretase Schnitten diverser Substrate blockieren. Dennoch ist die Inhibition als Therapieoption nicht wegzudenken und stellt nach wie vor ein Forschungsgebiet bei der Entwicklung neuer Therapie-Möglichkeiten dar. Weitere potenzielle γ -Sekretase Wirkstoffe, die den Notch-Schnitt inhibieren, wurden entwickelt. Dazu gehört das MRK-003 [181], das bereits vielversprechende Ergebnisse in Mausmodellen zeigte [182, 183]. Einige potenzielle γ -Sekretase Inhibitoren zeigten synergistischen Effekt mit anderen Chemotherapeutika [184, 183]. Auch die in dieser Arbeit identifizierten Substanzen können zur Behandlung des Multiplen Myeloms mit Proteasom-Inhibitoren kombiniert werden [162].

Im Laufe der Bearbeitung des Projekts nahm das Wissen um die Rolle des Notch-Schnitts durch die γ -Sekretase für die Entwicklung verschiedener Krankheitsbilder zu. Neben der Behandlung des Multiplen Myeloms gibt es auch Hinweise darauf, dass die Inhibition des Notch-Schnitts auch eine Therapiemöglichkeit bei Niereninsuffizienz und Fibrose darstellt [55, 56, 57]. Erste experimentelle Analysen des in dieser Arbeit identifizierten wirksamsten Wirkstoffs zur Behandlung des Multiplen Myeloms [162] weisen auch auf eine Inhibition fibrotischer Prozesse und eine Besserung chemisch induzierter Niereninsuffizienz bei Mäusen hin (Persönliche Mitteilung: R. Preißner, S. Bachmann). Ein Hinweis auf inhibierten APP-Abbau wurde nicht gefunden (Persönliche Mitteilung: R. Preißner, H. Esselmann).

Die erfolgreiche Anwendung der Analogie-Modellierung für die Suche nach neuen γ -Sekretase Inhibitoren ermutigt zur Anwendung dieser Methode in weiteren Studien mit fehlenden Strukturinformationen. Die identifizierten Wirkstoffe scheinen spezifisch den Notch-Schnitt zu modulieren, ohne ihn vollständig zu verhindern und damit schließlich weniger starke Nebenwirkungen zu bewirken. Ob es sich jedoch wirklich um die Blockierung der modellierten Presenilin Bindungsstelle handelt oder die Inhibition durch eine andere noch unbekannte Bindungstasche der γ -Sekretase oder sogar durch ein anderes in der Signalkaskade der Sekretase vorgeschaltetes Target ausgelöst wurde, kann hier nicht abschließend festgestellt werden. Weitere Analysen, γ -Sekretase-Assay Experimente sowie die Proteinstruktur der Sekretase, können zur Aufklärung führen. Unabhängig davon besitzen die Wirkstoffe großes Potenzial für die Weiterentwicklung.

4.3 VEGF-Rezeptor 2 als Zielprotein

In den letzten Jahren haben sich Forscher die Aufgabe gestellt, affinere und spezifischere VEGF-Rezeptor Inhibitoren zu identifizieren, um die Angiogenese zu hemmen. Dabei sind Substanzen wie Vatalanib der Firma Novartis, Vandetanib von AstraZeneca und Axitinib von Pfizer in der klinischen Phase III [39].

Da die Wirkungen der Inhibitoren nur mäßigen Effekt zeigten, wurden sie mit konventionellen Therapien (Zytostatika) kombiniert [37]. Ein neuer Therapieansatz stellt nicht mehr die vollständige Blockierung der Angiogenese durch Inhibitoren oder monoklonale Antikörper dar, sondern vielmehr die Wiederherstellung eines normalen Blutgefäßsystems, wodurch die Zugänglichkeit erhöht und eine effektivere Wirkung durch Chemotherapeutika erreicht werden kann [37].

Das im Laufe dieser Arbeit verwendete *in silico* Screening lieferte neue potenzielle Wirkstoffkandidaten für die Behandlung der Angiogenese. Interessanterweise zeigte vor allem die Kleinstruktur mit einem Schwefelatom zwischen zwei Ringen die größte Wirkung bei den Zellwachstums-Inhibitionsexperimenten, so dass diese als Leitstruktur für weitere Optimierungsschritte mit 2D- und 3D- Ähnlichkeitssuchen verwendet wurde. So wurde eine weitere Liste an zu testenden potenziellen Wirkstoffen mit einer Schwefel-Komponente generiert. Bei genauer Analyse ist eine große Ähnlichkeit zwischen diesen Inhibitoren und einem PARP-Inhibitor zu erkennen [185]. Generell besteht zwischen dem VEGFR-Liganden Vatalanib und dem

PARP-Liganden eine große strukturelle Ähnlichkeit. Durch die experimentellen Validierungen der Arbeitsgruppe von Dr. Höpfner ist nicht geklärt, ob der inhibitorische Effekt auf die Inhibition der VEGF-Rezeptoren zurückzuführen ist. Die Beobachtung könnte auch das Resultat einer PARP-Inhibition sein. Rajesh et al. zeigt, dass die Inhibition von PARP, die von VEGF-induzierte Proliferation, Migration und die Angiogenese hemmt [186], was die Vermutung unterstützt, dass es sich hier auch um einen PARP-Inhibitor handeln könnte.

Es gibt nun also folgende vier mögliche Hypothesen:

1. die entwickelten Inhibitoren adressieren den VEGFR-2
2. es handelt sich bei den Resultaten der Zellwachstumsversuche mit den Inhibitoren (zumindest mit der Schwefel-Komponente) in erster Linie um einen PARP-Effekt und nicht vornehmlich um eine auf VEGFR-2 induzierte Wirkung
3. VEGFR-2 und PARP werden durch den Inhibitor adressiert
4. die Liganden adressieren ein anderes Protein des Angiogenese-Signalweges

Die endgültige Belegung einer der aufgestellten Hypothesen kann nur mit Proteinspezifischen Kinase-Assays gewährleistet werden. Interessant jedoch bleibt, dass die Veränderung einzelner Atome einer Kleinstruktur hier zur Adressierung verschiedener Enzymklassen führen könnte. In einem weiteren Analyseschritt sollen daher die in Abbildung 4.1 dargestellten Wirkstoffe in Kinase Assays bzgl. VEGFR-2 und PARP Inhibition geprüft werden.

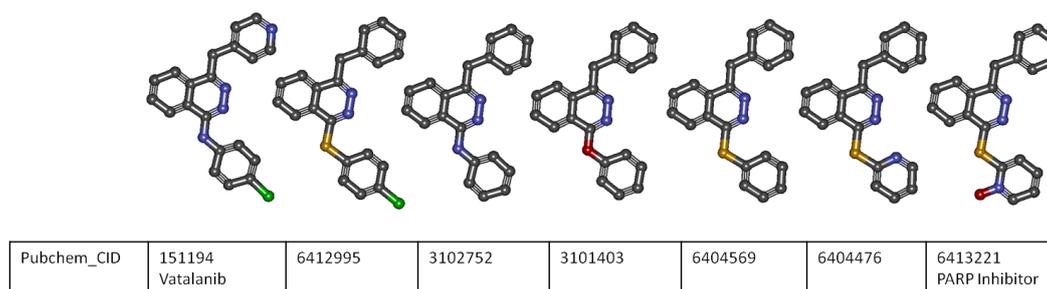


Abb. 4.1: Atomweise vom VEGFR-Inhibitor (Vatalanib) zum PARP-Inhibitor
Visualisiert sind bestellbare Substanzen, die auf Enzym-Assays bzgl. ihrer PARP und VEGF-Inhibition getestet werden sollen.

Erste Voranalysen mit Hilfe eines PARP Assays der Firma Trevigen zeigten eine Inhibition durch den identifizierten Inhibitor 1 (Persönliche Mitteilung: Julia Hoss-

bach). Die weiteren Assay Untersuchungen folgen nach Abschluss dieser Arbeit. Die VEGFR-2 Inhibitionsuche stellt ein Gegenbeispiel dafür dar, dass ähnliche Substanzen auch das gleiche Zielprotein adressieren. Es muss in Zukunft darum gehen, Substanzen mit indikationsgeeigneten Targetprofilen auszuwählen. Dennoch zeigen die Substanzen gute und teilweise auch bessere inhibitorische Wirkung als die Leitstruktur Vatalanib. Die Inhibitoren stellen daher nach wie vor potenzielle Arzneistoffkandidaten dar, wobei das Zielprotein noch nicht eindeutig identifiziert zu sein scheint.

ADMETox:

Ähnlich wie im Falle der γ -Sekretase Substanzen, werden für die Liganden eine hohe Hepatotoxizitäts- und eine hohe Cytochrom Inhibitions-Wahrscheinlichkeit vorhergesagt. Beim Vergleich der Eigenschaften mit denen des bekannten VEGFR Inhibitors Vatalanib, ist festzustellen, dass auch diese Substanz eine hohe Hepatotoxizitäts- und Cytochrom-Inhibitions-Wahrscheinlichkeit aufweist, dennoch ist der Wirkstoff schon weit in den klinischen Studien fortgeschritten. Das bedeutet, dass auch unsere Substanzen nicht ausgeschlossen werden durften.

4.4 Target-Vorhersage

Der SuperPred Webserver wurde entwickelt, um die medizinische Indikation chemischer Substanzen mit unbekanntem Zielprotein vorherzusagen. Die Kombination aus physikochemischen Eigenschaften und struktureller Ähnlichkeit bietet die Möglichkeit, neue biologisch aktive Substanzen und neue Targets für Kleinstrukturen zu detektieren. Jedoch lassen sich vor allem bei strukturell sehr ähnlichen Substanzen (Tanimoto > 0.95) eine zu 80% richtige Vorhersage treffen. Insgesamt liegt die Vorhersagegenauigkeit bei ca. 70%.

Beispielsweise konnte für die NCI-Substanz 600221 (oder 600220) eine ähnliche Wirkung wie Paclitaxel vorhergesagt werden. Leider jedoch fiel das Eigenschaftsprofil der Substanz 600221 nicht in die Eigenschaftsprofilgrenzwerte für die ATC-Gruppe L01CD. Das Molekulargewicht als auch die Anzahl der Wasserstoffakzeptoren ist größer als der maximale Wert für alle in dieser ATC-Gruppe klassifizierten Wirkstoffe. Da jedoch die Eigenschaftsprofilwerte nur minimal über der oberen Grenze und die strukturelle Ähnlichkeit bei 92% liegen, ist dennoch davon auszugehen, dass

die Substanzen ähnliche Zielproteine besitzen. Jedoch kann nur eine experimentelle Testung Klarheit über diese Vorhersage liefern.

Bisher gibt es keine Target-Vorhersage-Webserver, bei denen für eine Substanz ein Zielprotein vorhergesagt wird. Im letzten Jahr wurden hauptsächlich Target-Vorhersage-Algorithmen entwickelt, die Targets der mikroRNAs vorhersagen [187, 188]. MikroRNAs regulieren die Proteinbiosynthese, indem sie an die mRNA der Zielgene binden und entweder zur Translations-Inhibition oder zum RNA-Abbau führen.

Klipp et al. konzentrieren sich auf eine biochemische, Netzwerk-basierte Wirkstoff-Zielprotein Vorhersage [189]. Sie beschreiben Netze und Wege, in denen die Komponenten durch physikalische Wechselwirkungen oder biochemische Prozesse in Verbindung stehen. Es werden vor allem mathematische Modellierungsansätze diskutiert, um mögliche Zielproteine zu identifizieren [189].

Keiser et al. untersuchten über 3.500 FDA-bestätigte Arzneistoffe bzgl. hunderter Zielproteine mit Ligandeninformationen [190]. Dabei gingen wie bei SuperPred bevorzugt chemische Ähnlichkeitsvergleiche in die Analysen mit ein. Sie identifizierten dabei über 20 neue Wirkstoff-Zielproteine. Die Resultate der Arbeit von Keiser et al. unterstützen die auf chemische Ähnlichkeitsvergleiche basierte Methodik.

Eine weitere Anwendung bietet die SuperPred in Hinblick auf das „Drug-Repositioning“. Dabei können für Wirkstoffe mit bekannten Targetinformationen, aufgrund der hohen Ähnlichkeit zu anderen Substanzen, weitere bisher unbekannte potenzielle Zielproteine identifiziert werden.

Darüber hinaus soll der Webserver dementsprechend erweitert werden, Nebenwirkungen vorherzusagen, die durch Off-Target-Treffer hervorgerufen werden. Als Target wird hierbei das gewünschte adressierte Zielprotein bezeichnet, während die Off-Targets zusätzliche, nicht erwünschte Zielproteine darstellen. Die Adressierung der Off-Targets führt durch Modulation anderer Signalwege zu unerwünschten Effekten und somit zu Nebenwirkungen.

4.5 Integration von Wirkstoff-Protein-Interaktionen

4.5.1 SuperTarget

Die SuperTarget Datenbank stellt eine komplexe Datenstruktur dar, die es dem Nutzer ermöglicht, viele interessante und komplexe Fragestellungen zu formulieren und

zu analysieren, die es bis dahin noch nicht gegeben hat. Datenbanken wie DrugBank und TTD (Therapeutic Target Database) besitzen auch Arzneistoff-Target Informationen, wobei DrugBank aktuell 2.600 und die TTD 1.000 Relationen anbietet. Es fehlen jedoch die Kreuz-Verknüpfungen zu den Signalwegen und weitere Informationen und Verlinkungen, wie sie in der SuperTarget verfügbar und abrufbar sind. Die TTD beinhaltet z.B. keine Verlinkungen zu der PubChem, ChemDB, CAS Registry und UniProt-Datenbank, was viele Analysen erschwert. Es fehlen damit Informationen über die chemische Struktur des Arzneistoffs, die physikochemischen Eigenschaften, die adressierten Zielproteinsequenzen und 3D Struktur oder die Information des biologischen Signalweges, der vom Arzneistoff beeinflusst wird. Die SuperTarget Datenbank stellt nicht nur verknüpfte Informationen der Wirkstoffe und ihren Zielproteinen und Signalwegen dar, sondern vielmehr auch eine Methodensammlung zur Analyse verschiedenster Fragestellungen. Allein die Möglichkeit komplexere Anfragen an die Datenbank zu stellen, um beispielsweise alternative Wirkstoffe oder Wirkstoffe, die in bestimmten Signaltransduktionen eine Rolle spielen und von einem bestimmten Cytochrom abgebaut werden, macht die SuperTarget zu einem besonders effektiven Werkzeug. Mit Hilfe der Daten der SuperTarget konnte ein bisher als Antiemetikum eingesetzter Arzneistoff erfolgreich „repositioniert“ werden.

Des Weiteren wäre es von entscheidender Bedeutung neben den pharmazeutisch angewandten Wirkstoffen auch Wirkstoffe in die Datenbank zu integrieren, die zwar eine Wirkung gezeigt haben, sich jedoch nicht zu Arzneistoffen entwickeln ließen oder wieder zurückgezogen wurden. Diese Informationen würden nicht nur den Datensatz vervollständigen, sondern auch noch ausführlichere Analysen möglich machen. In der CancerResource sollte unter anderem diese Aufgabe bewältigt werden.

Die SuperTarget Datenbank ist noch unvollständig und muss regelmäßig aktualisiert werden. Zu etwa 900 Arzneistoffen aus der SuperTarget gibt es noch keine Targetinformationen. Für nur 2.500 Targets gibt es Arzneistoffinformationen, jedoch muss hier auch beachtet werden, dass nicht jedes der 3.000.000 Proteine der UniProt, die in SuperTarget integriert wurden auch ein im Augenblick genutztes medizinisches Zielprotein darstellt. Die Inhibition einiger („House-keeping“) Proteine führt zum Zusammenbruch der physiologischen Zellprozesse und wäre damit zytotoxisch.

4.5.2 CancerResource

Viele onkologisch und hämatologisch relevante Informationen liegen teilweise in diversen, auf unterschiedliche Aspekte spezialisierten, Datenbanken vor. Jedoch beschränken sich die meisten Datenbanken auf spezifische Tumorarten, genetische Informationen oder anderen Tumor-relevanten Aspekten. Andere bilden diverse Wirkstoff-Target-Beziehungen aus verschiedenen Organismen und Bereichen ab. Notwendig wäre jedoch eine umfassende Quelle, welche die verschiedenen Gesichtspunkte (Targets, Signalwege, Expressionsdaten, Wirkprofile etc.) miteinander verknüpft und gleichzeitig Methoden zur Analyse und Vorhersage bereitstellt. Dazu wurde die CaRe Datenstruktur entwickelt. Durch die Integration und Verknüpfung verschiedener relevanter Bereiche ermöglicht CaRe die Identifizierung neuer Targets, die Analyse und den Vergleich individualisierter Daten, Suche der ähnlichsten Zelllinien und die Identifizierung der wirksamsten Substanzen für diese Daten.

Eine genaue Analyse der Kleinstrukturen der Datenbanken TTD, CTP, DrugBank und PharmGKB zeigte nur eine marginale Überlappung (Daten nicht gezeigt) der durch verschiedene Methoden gewonnenen direkten und indirekten Interaktionsdaten zwischen den Datenbanken. Aus diesem Grund ist eine Integration und Zusammenführung dieser Datenbanken unumgänglich. Ein separater Textmining-Durchlauf vervollständigte den Datensatz um ca. 1.000 Interaktionen. Neben den in der Therapie angewandten Wirkstoffen, die in der SuperTarget gespeichert sind, speichert die CancerResource alle Substanzen, die eine Wirkung auf ein onkologisch relevantes Target oder eine Zelllinie zeigen konnten.

Des Weiteren ist die Kombination bereits vorhandener Expressionsdaten und Wirkprofile der Tumor-Zelllinien mit Daten bzgl. der genetischen Variabilität onkologisch relevanter Gene der Patienten möglich. Durch den Vergleich der Tumor-spezifischen Expressions-/Mutationsmuster mit denen der NCI-Zelllinien wird angestrebt, patientenspezifisch optimierte Therapievorschlüsse zu erstellen. Das „Cancer Genome Project“ des Sanger Instituts stellt für Wissenschaftler eine systematische Charakterisierung der genetischen Ausstattung der Tumorzelllinien zur Verfügung (<http://www.sanger.ac.uk/genetics/CGP/>). Es ermöglicht unter anderem das Browsen in den Zelllinien und Mutationsanalysen [191]. Die NCI60-Zelllinien sind ebenfalls enthalten. Jedoch bietet die Datenbank keine Möglichkeit die Zelllinien mit eigenen Daten zu vergleichen. Es sind auch keine Verknüpfungen zu Signalwegen vorhanden.

In CaRe sind individuelle Daten hochladbar und werden mit gespeicherten Daten

verglichen und die bestmöglichen Wirkstoffe ermittelt. Durch den fortwährenden Zuwachs an individualisierten Daten bzgl. der Target-Ausstattung, ihrer Expressionen und Mutationen, können Ressourcen wie CaRe einen ersten Schritt in Richtung individualisierte Therapie darstellen.

Kürzlich hat sich ein internationales Konsortium („International Cancer Genome Consortium“ (ICGC) (<http://www.icgc.org/>)) [192] der Aufgabe gestellt, Genome von 50 verschiedenen Tumoren zu sequenzieren. Dieses Projekt stellt einen entscheidenden Schritt in Richtung individualisierter Therapie dar. Eine Kombination der Charakterisierung der Tumoren mit einer zusätzlichen Analyse der Chemosensitivität hilft, individuelle Daten zu vergleichen und potenzielle Wirkstoffe vorzuschlagen.

Die Anzahl der Informationen der Substanz- und Target-Interaktionen steigt stetig, da gerade die Entwicklung neuer Therapien einen Schwerpunkt in der Forschung darstellt. Dadurch wächst auch der Informationsgehalt der CaRe konstant.

4.6 FragmentStore - Fragmentanalysen

Der FragmentStore bietet Informationen zu Fragmenten von Substanzen, Metaboliten und toxischen Kleinstrukturen. Diese Daten sollen Wissenschaftlern der Wirkstoffentwicklung zum einen geeignete Fragmentbibliotheken liefern und zum anderen auch die virtuelle Synthese von Substanzen aus gewählten Fragmenten ermöglichen. Zusätzlich stellt der FragmentStore Daten bereit, deren Analyse hilfreich für das Verständnis einiger Wirkstoff-Nebenwirkungen sein kann. Beispielsweise könnte das Vorkommen gleicher Fragmente in verschiedenen Arzneistoffen auf ähnliche Zielproteine hindeuten. Das Auftreten gleicher Fragmente in Wirkstoffen und toxischen Substanzen hingegen könnte ein Hinweis auf Nebenwirkungen geben. Die systematische, automatisierte Synthese von Substanzbibliotheken ausgehend von drei Fragmenten führt im Durchschnitt zu 10.000 Substanzen, die den chemischen Raum für die Adressierung eines spezifischen Zielproteins abtasten.

Erste Analysen:

Das Venn-Diagramm der Fragmente zeigt 55 gleiche Fragmente, die jeweils aus allen drei Substanz-Datensätzen mit der „Linker“-Schnittweise generiert werden kann-

ten. Bei diesen Fragmenten handelt es sich bevorzugt um kleine Fragmente, die wahrscheinlich keine ausschlaggebende Bindungstaschenpräferenz und damit auch nur verminderte Zielproteinspezifität besitzen. Der einfache Benzolring gehört beispielsweise zu dieser Gruppe. Interessant sind in dieser Analyse hauptsächlich die gemeinsamen Fragmente zwischen Wirkstoffen und toxischen Substanzen, da diese Fragmente einen Hinweis auf den toxischen Effekt der Arzneistoffe geben oder evtl. entstehende Nebenwirkungen erklären können. Beispielsweise könnte ein Fragment für die Spezifität einer Substanz für bestimmte Proteine oder Proteingruppen verantwortlich sein. Eine Analyse der Fragmente der metabolischen, toxischen und Arzneistoff-Substanzen wurde bisher nicht durchgeführt. Jedoch haben Khanna et al. die Eigenschaften ähnlicher drei Datensätze bzgl. chemischer und physikochemischer Eigenschaften hin untersucht und ein ähnliches Venn-Diagramm erstellt [193]. Die Anzahl der überlappenden Substanzen zwischen den Kleinstrukturen war ähnlich der in dieser Arbeit. Jedoch untersuchte Khanna et al. die Substanzen nicht auf Fragmentebene.

Weitere detaillierte Analysen der einzelnen Schnittmengen werden folgen. Hierbei sollen auch die zugehörigen Targetinformationen mit eingeschlossen werden, um Fragmente zu identifizieren, die spezifisch bestimmte Enzyme binden.

In einem nächsten Schritt müssen die Fragmente bzgl. ihrer Tauglichkeit für die Generierung einer optimierten Fragment-Bibliothek für das Fragment-basierte Wirkstoffdesign analysiert werden. Wang et al. beispielsweise haben Wirkstoffdatensätze und deren Fragmente in Bezug auf deren Potenzial als „Building Block“ zu fungieren, analysiert [194].

Bisher ist die Nutzung von Fragmentbibliotheken zum Wirkstoffdesign hauptsächlich im Bereich der industriellen Forschung zu finden. Hier werden bevorzugt Fragmentbibliotheken verwendet, die aus HTS-Hits resultierenden Strukturen mit mittlerer Affinität generiert wurden. Daher ist ein Vergleich des FragmentStores mit anderen Bibliotheken nicht möglich. Die Generierung der Fragmente als auch die integrierte Assemblierung der Fragmente soll ein erster Schritt zur Automatisierung des Fragment-basierten Wirkstoffdesigns darstellen. Die Aufgabe eine Fragment-basierte Wirkstoffentwicklungs-Pipeline zu entwickeln wird über diese Arbeit hinaus bearbeitet.

Kapitel 5

Ausblick

Diese Arbeit hat gezeigt, dass ein Genom-Vergleich interessante Aspekte wie funktionelle Pseudogene aufwerfen kann, deren Rolle bei der Tumorentwicklung im Einzelfall zu untersuchen bleibt. Die im Rahmen dieser Arbeit entwickelte Analogie-Modellierung hat in Zukunft ein breites Anwendungsspektrum, denn viele Membranproteine sind besonders interessante Targets, aber ihre experimentelle Strukturaufklärung bleibt voraussichtlich sehr aufwändig. Die Analyse der Fragmente von Arzneistoffen, Metaboliten und toxischen Substanzen hat gezeigt, dass diese neue Betrachtungsweise sowohl für systembiologische als auch für medizinal-chemische Betrachtungen sehr wertvoll sein wird. Darüber hinaus wurden Datenbanken und Webserver entwickelt, die eine Integration verschiedener Informationen über Tumor(zellen) leistet und so einen ersten Schritt in Richtung individualisierte Medizin ermöglichen. Mit der Zunahme individueller Informationen zur Targetausstattung der Patienten und ihrer Tumore (Next-Generation Sequencing) kommt der Weiterentwicklung solcher Ressourcen eine wachsende Bedeutung zu.

Anhang A

Verzeichnisse

Abkürzungsverzeichnis

Abkürzung	Erklärung
ADMET	Absorption, Distribution, Metabolismus, Exkretion, Toxizität
APP	Amyloid Precursor Protein
ATC	Anatomisch-Therapeutisch-Chemische Classification
BBB	Blood Brain Barrier
BTP	Bindungstaschenpräferenz
CaRe	CancerResource
CMC	Chemistry, Manufacturing and Control
CDS	Coding Sequence
CTD	Comparative Toxicogenomics Database
DTP	Developmental Therapeutics Program
GI50	Growth Inhibition
GO	GeneOntology
IC50	Inhibitory Concentration
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC50	Lethal Concentration
LD50	Lethal Dose
NCI	National Cancer Institute
PDB	Protein Data Bank
PharmGKB	Pharmacogenetics Knowledge Base
PPB	Plasma Protein Binding
PSA	Polar Surface Area

Abkürzung	Erklärung
Recap	Retrosynthetic combinatorial analysis procedure
RMSD	Root Mean Square Deviation
SNP	Single Nucleotide Polymorphism
TGI	Tumor Growth Inhibition
TTD	Therapeutic Target Database
VEGFR	Vascular Endothelial Growth Factor Receptor
WDI	World Drug Index
WHO	World Health Organization

Abbildungsverzeichnis

Abb. 1.1:	Lennard-Jones- und Coulomb-Potenzial aus Klebe.	7
Abb. 1.2:	Wege der Apoptose.	11
Abb. 1.3:	Die Rolle des VEGFR-2	13
Abb. 1.4:	Die Rolle der γ -Sekretase im Notch Signalweg.	15
Abb. 1.5:	Schematische Darstellung des Presenilins der γ -Sekretase	16
Abb. 1.6:	Der Weg zum Arzneimittel aus Klebe.	19
Abb. 2.1:	Beispiel für eine synonyme und nicht-synonyme Mutation.	32
Abb. 2.2:	Recap-Regeln zur Fragmentierung.	35
Abb. 2.3:	Fragmentierungsstrategien angewandt an einem Beispiel.	36
Abb. 2.4:	Alignment für das Modellerprogramm.	39
Abb. 2.5:	<i>In silico</i> Screenings.	41
Abb. 2.6:	Prinzip der 3D-Überlagerung.	43
Abb. 2.7:	Prinzip des genetischen Algorithmus beim Protein-Liganden Docking.	44
Abb. 2.8:	Schematische Darstellung der Analogie-Modellierung.	48
Abb. 2.9:	„Mean“-Diagramm für Paclitaxel aus der NCI-Ressource (DTP).	51
Abb. 2.10:	Textmining Schema.	55
Abb. 3.1:	Vergleich des Menschen- und des Schimpansen-Chromosoms 18.	60
Abb. 3.2:	Lokalisationen der CDS der Proteine auf den Chromosomen.	61
Abb. 3.3:	Lokalisation der 84 Gene mit längeren Introns auf den Chromosomen.	63
Abb. 3.4:	ATPase Domäne der humanen Topoisomerase II alpha (1ZXM) mit gebundenem ATP Molekül.	65
Abb. 3.5:	Modellierte Struktur des Presenilin-1.	67
Abb. 3.6:	Bekannte γ -Sekretase Inhibitoren, die als Leitstrukturen dienen.	68
Abb. 3.7:	Beispiel des Sekretase Screenings.	69
Abb. 3.8:	γ -Sekretase Inhibitoren	70

Abb. 3.9: Mit der Accelrys Software erstellte intestinale Absorptions-Vorhersage der sechs potenziellen γ -Sekretase Inhibitoren aufgetragen auf der AlogP98 gegen PSA-2D Ebene.	71
Abb. 3.10: Docking Pose des Liganden mit der stärksten Inhibitionswirkung auf die γ -Sekretase	73
Abb. 3.11: Planare Visualisierung der Docking Pose des affinsten Liganden in die Bindungstasche des Presenilins.	73
Abb. 3.12: Alignment der kristallisierten VEGFR-2 Sequenz und der Proteinsequenz aus der UniProt.	75
Abb. 3.13: Durch Homologie-Modellierung vervollständigte katalytische Untereinheit der des VEGFR-2.	76
Abb. 3.14: Vatalanib.	76
Abb. 3.15: Potenzielle Wirkstoffkandidaten nach der Anwendung des 2D- und 3D-Screenings	77
Abb. 3.16: Wirksamster Inhibitor nach experimenteller Validierung diente als neue Leitstruktur.	78
Abb. 3.17: Potenzielle VEGFR-2-Arzneistoffkandidaten nach experimenteller Validierung.	78
Abb. 3.18: Mit der Accelrys Software erstellte intestinale Absorptions-Vorhersage der sieben potenziellen VEGFR-2 Inhibitoren nach oraler Administration aufgetragen auf die AlogP98 gegen PSA-2D Ebene.	80
Abb. 3.19: Gruppierung (Clustering) der NCI60-Substanzen mit ähnlichem zellulären Effect.	83
Abb. 3.20: Cluster aus 7-Chlorocamptothecin und vier strukturell ähnlichen Substanzen mit ähnlichem Wirkprofil.	83
Abb. 3.21: Architektur der SuperTarget Datenbank.	85
Abb. 3.22: Anwendungsbeispiel für die SuperTarget	89
Abb. 3.23: Architektur der CaRe-Datenbank.	91
Abb. 3.24: Apoptose Signalweg mit markierten Protein-Wirkstoff-Beziehungen.	94
Abb. 3.25: Zellulärer Fingerprint zweier Taxane.	95
Abb. 3.26: Anwendungsbeispiel der CaRe-Datenbank ausgehend von externen Expressionsdaten.	96
Abb. 3.27: Statistik SuperPred	100
Abb. 3.28: Mittelwerte der Eigenschaften der 35.000 FragmentStore Fragmente, generiert mit der Recap- und Linker-Schnittweise.	101

Abb. 3.29: Anwendungsbeispiel des FragmentStores.	104
Abb. 3.30: Vergleich der gemeinsamen Strukturen der metabolischen, toxischen und Wirkstoffdaten auf Kleinstruktur und Fragmentebene.	105
Abb. 4.1: Atomweise vom Vatalanib zur PARP-Substanz.	113

Tabellenverzeichnis

Tab. 3.1: Proteine, deren kodierende Sequenz bei Mensch und Schimpanse auf verschiedenen Chromosomen annotiert waren. <i>Aus Ahmed et al.</i> [156]. HS: Homo Sapiens; PT: Pan Troglodytes	62
Tab. 3.2: Proteine mit einem Ka/Ks -Wert größer 1 und damit potenziell positiv selektiert sind. <i>Aus Ahmed et al.</i> [156].	63
Tab. 3.3: Motive, die in der Schimpansen- oder humanen Proteinsequenz identifiziert wurden. X: Dieses Motiv besaß nur das Protein des markierten Organismus. <i>Aus Ahmed et al.</i> [156].	64
Tab. 3.4: Lipinski-Eigenschaften und die „Polar Surface Area“ (PSA). Die Inhibitoren brechen jeweils nur eine Lipinski-Regel. Der berechnete clogP-Wert, der fünf nicht überschreiten sollte, ist hier oft zu hoch.	70
Tab. 3.5: Mit der Accelrys Software vorhergesagten ADMET-Eigenschaften für die sechs potenziellen Sekretase Inhibitoren. Vorhergesagt mit Accelrys „Admet-Descriptors“. BBB = Blood Brain Barrier; PPB = Plasma Protein Binding.	72
Tab. 3.6: Die vorhergesagten LD50-Werte und die vorhergesagten maximal tolerierten Dosen für das Modell: Ratte oral. Alle vorhergesagten Werte liegen über 25 mg/kg und sind damit unbedenklich. Vorhergesagt mit Accelrys „TOPKAT“	72
Tab. 3.7: Lipinski-Eigenschaften und die „Polar Surface Area“ (PSA). Die Inhibitoren verstoßen jeweils gegen nur eine Lipinski-Regel. Der berechnete logP ist meistens minimal zu hoch.	79

Tab. 3.8:	Mit der Accelrys Software vorhergesagte ADMET-Eigenschaften für die sieben potenziellen VEGFR-2 Inhibitoren. Vorhergesagt mit Accelrys „Admet-Descriptors“. BBB = Blood Brain Barrier; PPB = Plasma Protein Binding.	81
Tab. 3.9:	Die vorhergesagten LD50-Werte und die maximal tolerierten Dosen für das Modell: Ratte oral. Alle vorhergesagten Werte liegen über 25 mg/kg und sind damit unbedenklich.	81
Tab. 3.10:	Zahlen der redundanten und unigen Kleinstruktur-Zielprotein Relationen in CaRe. K-Z-B: Kleinstruktur-Zielprotein-Beziehungen. <i>Aus Ahmed et al.</i> [153].	93
Tab. 3.11:	Vergleich der Eigenschaften der Datenbanken und CaRe.	97
Tab. 3.12:	Physikochemische Eigenschaftsprofile der Substanzen NSC600221 und Paclitaxel und die obere und untere Eigenschaftsprofilgrenze der ATC-Gruppe L01CD.	100

Literaturverzeichnis

- [1] Gesellschaft für epidemiologischen Krebsregister in Deutschland e.V. Robert Koch-Institut Berlin. *Krebs in Deutschland 2005/2006 Häufigkeiten und Trends*, volume 7. Ausgabe. 2010. ISBN 978-3-89606-207-9.
- [2] A. Varki. A chimpanzee genome project is a biomedical imperative. *Genome Res*, 10(8):1065–1070, Aug 2000.
- [3] D. S. Beniashvili. An overview of the world literature on spontaneous tumors in nonhuman primates. *J Med Primatol*, 18(6):423–437, 1989.
- [4] D. J. Waters, W. A. Sakr, D. W. Hayden, C. M. Lang, L. McKinney, G. P. Murphy, R. Radinsky, R. Ramoner, R. C. Richardson, and D. J. Tindall. Workgroup 4: spontaneous prostate carcinoma in dogs and nonhuman primates. *Prostate*, 36(1):64–67, Jun 1998.
- [5] H. M. McClure. Tumors in nonhuman primates: observations during a six-year period in the Yerkes primate center colony. *Am J Phys Anthropol*, 38(2):425–429, Mar 1973.
- [6] X. S. Puente, G. Velasco, A. Gutiérrez-Fernández, J. Bertranpetit, M.-C. King, and C. López-Otín. Comparative analysis of cancer genes in the human and chimpanzee genomes. *BMC Genomics*, 7:15, 2006.
- [7] H. R. Seibold and R. H. Wolf. Neoplasms and proliferative lesions in 1065 nonhuman primate necropsies. *Lab Anim Sci*, 23(4):533–539, Aug 1973.
- [8] Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055):69–87, Sep 2005.

- [9] H. Kehrer-Sawatzki and D. N. Cooper. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum Mutat*, 28(2):99–130, Feb 2007.
- [10] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, and W. FitzHugh et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, Feb 2001.
- [11] D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, and M. Hassanali. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*, 36(Database issue):D901–D906, Jan 2008.
- [12] F. Zhu, B. Han, P. Kumar, X. Liu, X. Ma, X. Wei, L. Huang, Y. Guo, L. Han, C. Zheng, and Y. Chen. Update of TTD: Therapeutic Target Database. *Nucleic Acids Res*, 38(Database issue):D787–D791, Jan 2010.
- [13] A. P. Davis, B. L. King, S. Mockus, C. G. Murphy, C. Saraceni-Richards, M. Rosenstein, T. Wiegers, and C. J. Mattingly. The Comparative Toxicogenomics Database: update 2011. *Nucleic Acids Res*, Sep 2010.
- [14] O. Berezovska, M. Frosch, P. McLean, R. Knowles, E. Koo, D. Kang, J. Shen, F. M. Lu, S. E. Lux, S. Tonegawa, and B. T. Hyman. The Alzheimer-related gene presenilin 1 facilitates notch 1 in primary mammalian neurons. *Brain Res Mol Brain Res*, 69(2):273–280, Jun 1999.
- [15] D. Beher and M. S. Shearman. Gamma-secretase inhibition. *Biochem Soc Trans*, 30(4):534–537, Aug 2002.
- [16] E. J. Allenspach, I. Maillard, J. C. Aster, and W. S. Pear. Notch signaling in cancer. *Cancer Biol Ther*, 1(5):466–476, 2002.
- [17] B. J. Collins, W. Kleeberger, and D. W. Ball. Notch in lung development and lung cancer. *Semin Cancer Biol*, 14(5):357–364, Oct 2004.
- [18] Y. Nefedova, D. M. Sullivan, S. C. Bolick, W. S. Dalton, and D. I. Gabrilovich. Inhibition of Notch signaling induces apoptosis of myeloma cells and enhances sensitivity to chemotherapy. *Blood*, 111(4):2220–2229, Feb 2008.

- [19] H. Al-Husaini, D. Subramanyam, M. J. Reedijk, and S. S. Sridhar. Notch Signaling Pathway as a Therapeutic Target in Breast Cancer. *Mol Cancer Ther*, Oct 2010.
- [20] L. Miele, H. Miao, and B. J. Nickoloff. NOTCH signaling as a novel cancer therapeutic target. *Curr Cancer Drug Targets*, 6(4):313–323, Jun 2006.
- [21] R. H. Shoemaker. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer*, 6(10):813–823, Oct 2006.
- [22] U. Müller C. E. Mortimer. *Das Basiswissen der Chemie.*, volume 6. Thieme Georg Verlag, 1996.
- [23] G. Klebe. *Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen.* Spektrum, 2009.
- [24] C. Bähr and B. Groner. The IGF-1 receptor and its contributions to metastatic tumor growth—novel approaches to the inhibition of IGF-1R function. *Growth Factors*, 23(1):1–14, Mar 2005.
- [25] D. H. Wirtz. *Strukturbasierte Entwicklung von Kinaseinhibitoren.* PhD thesis, Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln, 2009.
- [26] S. Nagata. Apoptosis and autoimmune diseases. *Ann N Y Acad Sci*, 1209:10–16, Oct 2010.
- [27] N. Leung, C. Turbide, B. Balachandra, V. Marcus, and N. Beauchemin. Intestinal tumor progression is promoted by decreased apoptosis and dysregulated Wnt signaling in Ceacam1^{-/-} mice. *Oncogene*, 27(36):4943–4953, Aug 2008.
- [28] A. Sawa, G. W. Wiegand, J. Cooper, R. L. Margolis, A. H. Sharp, J. F. Lawler, J. T. Greenamyre, S. H. Snyder, and C. A. Ross. Increased apoptosis of Huntington disease lymphoblasts associated with repeat length-dependent mitochondrial depolarization. *Nat Med*, 5(10):1194–1198, Oct 1999.
- [29] A. V. F. F. Teles, T. R. Rosenstock, C. S. Okuno, G. S. Lopes, C. R. A. Bertoncini, and S. S. Smali. Increase in bax expression and apoptosis are associated in Huntington’s disease progression. *Neurosci Lett*, 438(1):59–63, Jun 2008.

- [30] A. Eckert, C. A. Marques, U. Keil, K. Schüssel, and W. E. Müller. Increased apoptotic cell death in sporadic and genetic Alzheimer's disease. *Ann N Y Acad Sci*, 1010:604–609, Dec 2003.
- [31] S. Shimohama. Apoptosis in Alzheimer's disease—an update. *Apoptosis*, 5(1):9–16, Feb 2000.
- [32] A. Krueger. *Das zelluläre FLICE-inhibitorische Protein (c-FLIP): Mechanismen der Regulation CD95-vermittelter Apoptose in der T-Zell-Immunantwort*. PhD thesis, Universität Hannover, 2002.
- [33] A. Yuan, C. J. Yu, W. J. Chen, F. Y. Lin, S. H. Kuo, K. T. Luh, and P. C. Yang. Correlation of total VEGF mRNA and protein expression with histologic type, tumor angiogenesis, patient survival and timing of relapse in non-small-cell lung cancer. *Int J Cancer*, 89(6):475–483, Nov 2000.
- [34] B. Nitzsche, C. Gloesenkamp, M. Schrader, M. Ocker, R. Preissner, M. Lein, A. Zakrzewicz, B. Hoffmann, and M. Höpfner. Novel compounds with anti-angiogenic and antiproliferative potency for growth control of testicular germ cell tumours. *Br J Cancer*, 103(1):18–28, Jun 2010.
- [35] A. Bikfalvi and R. Bicknell. Recent advances in angiogenesis, anti-angiogenesis and vascular targeting. *Trends Pharmacol Sci*, 23(12):576–582, Dec 2002.
- [36] H. M. Fraser, S. E. Dickson, S. F. Lunn, C. Wulff, K. D. Morris, V. A. Carroll, and R. Bicknell. Suppression of luteal angiogenesis in the primate after neutralization of vascular endothelial growth factor. *Endocrinology*, 141(3):995–1000, Mar 2000.
- [37] D. J. Hicklin and L. M. Ellis. Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. *J Clin Oncol*, 23(5):1011–1027, Feb 2005.
- [38] C. G. Willett, Y. Boucher, E. di Tomaso, D. G. Duda, L. L. Munn, R. T. Tong, D. C. Chung, D. V. Sahani, S. P. Kalva, S. V. Kozin, M. Mino, K. S. Cohen, D. T. Scadden, A. C. Hartford, A. J. Fischman, J. W. Clark, D. P. Ryan, A. X. Zhu, L. S. Blaszkowsky, H. X. Chen, P. C. Shellito, G. Y. Lauwers, and R. K. Jain. Direct evidence that the VEGF-specific antibody bevacizumab

- has antivasular effects in human rectal cancer. *Nat Med*, 10(2):145–147, Feb 2004.
- [39] S. P. Ivy, J. Y. Wick, and B. M. Kaufman. An overview of small-molecule inhibitors of VEGFR signaling. *Nat Rev Clin Oncol*, 6(10):569–579, Oct 2009.
- [40] G. Bold, K. H. Altmann, J. Frei, M. Lang, P. W. Manley, P. Traxler, B. Wietfeld, J. Brügggen, E. Buchdunger, R. Cozens, S. Ferrari, P. Furet, F. Hofmann, G. Martiny-Baron, J. Mestan, J. Rösel, M. Sills, D. Stover, F. Acemoglu, E. Boss, R. Emmenegger, L. Lässer, E. Masso, R. Roth, C. Schlachter, and W. Vetterli. New anilinophthalazines as potent and orally well absorbed inhibitors of the VEGF receptor tyrosine kinases useful as antagonists of tumor-driven angiogenesis. *J Med Chem*, 43(12):2310–2323, Jun 2000.
- [41] D. Pytel, T. Sliwinski, T. Poplawski, D. Ferriola, and I. Majsterek. Tyrosine kinase blockers: new hope for successful cancer therapy. *Anticancer Agents Med Chem*, 9(1):66–76, Jan 2009.
- [42] G. Scagliotti and R. Govindan. Targeting angiogenesis with multitargeted tyrosine kinase inhibitors in the treatment of non-small cell lung cancer. *Oncologist*, 15(5):436–446, 2010.
- [43] I.-M. Shih and T.-L. Wang. Notch signaling, gamma-secretase inhibitors, and cancer therapy. *Cancer Res*, 67(5):1879–1882, Mar 2007.
- [44] M. Nosedá, L. Chang, G. McLean, J. E. Grim, B. E. Clurman, L. L. Smith, and A. Karsan. Notch activation induces endothelial cell cycle arrest and participates in contact inhibition: role of p21Cip1 repression. *Mol Cell Biol*, 24(20):8813–8822, Oct 2004.
- [45] V. Bolós, J. Grego-Bessa, and J. L. de la Pompa. Notch signaling in development and cancer. *Endocr Rev*, 28(3):339–363, May 2007.
- [46] I. Schreiner. *Expression von Notch-Rezeptoren und ihren Liganden in der humanen Plazenta*. PhD thesis, Justus-Liebig-Universität Gießen, 2010.
- [47] W. Xia and M. S. Wolfe. Intramembrane proteolysis by presenilin and presenilin-like proteases. *J Cell Sci*, 116(Pt 14):2839–2844, Jul 2003.

- [48] D. Spasic, A. Tolia, K. Dillen, V. Baert, B. De Strooper, S. Vrijens, and W. Annaert. Presenilin-1 maintains a nine-transmembrane topology throughout the secretory pathway. *J Biol Chem*, 281(36):26569–26577, Sep 2006.
- [49] H. Laudon, E. M. Hansson, K. Melén, A. Bergman, M. R. Farmery, B. Winblad, U. Lendahl, G. von Heijne, and J. Näslund. A nine-transmembrane domain topology for presenilin 1. *J Biol Chem*, 280(42):35352–35360, Oct 2005.
- [50] C. Hu, A. Diévert, M. Lupien, E. Calvo, G. Tremblay, and P. Jolicœur. Overexpression of activated murine Notch1 and Notch3 in transgenic mice blocks mammary gland development and induces mammary tumors. *Am J Pathol*, 168(3):973–990, Mar 2006.
- [51] T. P. Dang, A. F. Gazdar, A. K. Virmani, T. Sepetavec, K. R. Hande, J. D. Minna, J. R. Roberts, and D. P. Carbone. Chromosome 19 translocation, overexpression of Notch3, and human lung cancer. *J Natl Cancer Inst*, 92(16):1355–1357, Aug 2000.
- [52] D. Bellavia, M. Mecarozzi, A. F. Campese, P. Grazioli, A. Gulino, and I. Screpanti. Notch and Ikaros: not only converging players in T cell leukemia. *Cell Cycle*, 6(22):2730–2734, Nov 2007.
- [53] S. Mittal, D. Subramanyam, D. Dey, R. V. Kumar, and A. Rangarajan. Cooperation of Notch and Ras/MAPK signaling pathways in human breast carcinogenesis. *Mol Cancer*, 8:128, 2009.
- [54] F. Jundt, K. Schulze Pröbsting, I. Anagnostopoulos, G. Muehlinghaus, M. Chatterjee, S. Mathas, R. C. Bargou, R. Manz, H. Stein, and B. Dörken. Jagged1-induced Notch signaling drives proliferation of multiple myeloma cells. *Blood*, 103(9):3511–3515, May 2004.
- [55] M. Kretzler and L. Allred. Notch inhibition reverses kidney failure. *Nat Med*, 14(3):246–247, Mar 2008.
- [56] T. Niranjana, B. Bielez, A. Gruenwald, M. P. Ponda, J. B. Kopp, D. B. Thomas, and K. Susztak. The Notch pathway in podocytes plays a role in the development of glomerular disease. *Nat Med*, 14(3):290–298, Mar 2008.

- [57] A. Leask. Targeting the jagged/notch pathway: a new treatment for fibrosis? *J Cell Commun Signal*, 4(4):197–198, Dec 2010.
- [58] K. C. Nyhan, N. Faherty, G. Murray, L. B. Cooley, C. Godson, J. K. Crean, and D. P. Brazil. Jagged/Notch signalling is required for a subset of TGF β 1 responses in human kidney epithelial cells. *Biochim Biophys Acta*, 1803(12):1386–1395, Dec 2010.
- [59] N. Kaviani, A. Servettaz, C. Mongaret, A. Wang, C. Nicco, C. Chéreau, P. Grange, V. Vuiblet, P. Birembaut, M.-D. Diebold, B. Weill, N. Dupin, and F. Batteux. Targeting ADAM-17/notch signaling abrogates the development of systemic sclerosis in a murine model. *Arthritis Rheum*, 62(11):3477–3487, Nov 2010.
- [60] J. Eder, U. Hommel, F. Cumin, B. Martoglio, and B. Gerhartz. Aspartic proteases in drug discovery. *Curr Pharm Des*, 13(3):271–285, 2007.
- [61] B. P. Imbimbo. Therapeutic potential of gamma-secretase inhibitors and modulators. *Curr Top Med Chem*, 8(1):54–61, 2008.
- [62] Y. Tian, C. J. Crump, and Y.-M. Li. Dual role of alpha-secretase cleavage in the regulation of gamma-secretase activity for amyloid production. *J Biol Chem*, 285(42):32549–32556, Oct 2010.
- [63] N. Höttecke. *Entwicklung und Synthese von gamma-Sekretase-Modulatoren zur Behandlung von Morbus Alzheimer*. PhD thesis, Technischen Universität Darmstadt, 2010.
- [64] *Rote Liste 2010 Buchausgabe*. 2010. ISBN-13:978-3-939192-40-4.
- [65] S. Fox, S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, R. Khoury, and M. Biro. High-throughput screening: update on practices and success. *J Biomol Screen*, 11(7):864–869, Oct 2006.
- [66] D. A. Erlanson, J. A. Wells, and A. C. Braisted. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct*, 33:199–223, 2004.
- [67] T. O’Brien, B. T. Fahr, M. M. Sopko, J. W. Lam, N. D. Waal, B. C. Raimundo, H. E. Purkey, P. Pham, and M. J. Romanowski. Structural analysis of caspase-

- 1 inhibitors derived from Tethering. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 61(Pt 5):451–458, May 2005.
- [68] I. C. Choong, W. Lew, D. Lee, P. Pham, M. T. Burdett, J. W. Lam, C. Wiesmann, T. N. Luong, B. Fahr, W. L. DeLano, R. S. McDowell, D. A. Allen, D. A. Erlanson, E. M. Gordon, and T. O'Brien. Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J Med Chem*, 45(23):5005–5022, Nov 2002.
- [69] Y.-P. Chang, M.-J. Tseng, and Y.-H. Chu. Using surface plasmon resonance to directly measure slow binding of low-molecular mass inhibitors to a VanX chip. *Anal Biochem*, 359(1):63–71, Dec 2006.
- [70] W. Huber and F. Mueller. Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology. *Curr Pharm Des*, 12(31):3999–4021, 2006.
- [71] D.G. Myszka and R.L. Rich. Implementing surface plasmon resonance biosensors in drug discovery. *Pharm Sci Technolo Today*, 3(9):310–317, Sep 2000.
- [72] C. Dalvit, M. Fasolini, M. Flocco, S. Knapp, P. Pevarello, and M. Veronesi. NMR-Based screening with competition water-ligand observed via gradient spectroscopy experiments: detection of high-affinity ligands. *J Med Chem*, 45(12):2610–2614, Jun 2002.
- [73] A. Chen and M. J. Shapiro. Affinity NMR. *Anal Chem*, 71(19):669A–675A, Oct 1999.
- [74] R. S. Rowland. Using X-ray crystallography in drug discovery. *Curr Opin Drug Discov Devel*, 5(4):613–619, Jul 2002.
- [75] A. Heinrichs. Pubmed central. *Trends Mol Med*, 7(3):102, Mar 2001.
- [76] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res*, 28(1):235–242, Jan 2000.
- [77] P. H. H. Hermkens and G. Müller. The impact of combinatorial chemistry on drug discovery. *Ernst Schering Res Found Workshop*, (42):201–220, 2003.

- [78] P. Edwards. Combinatorial chemistry. *Drug Discov Today*, 8(21):1004–1005, Nov 2003.
- [79] R. Preissner, A. Goede, and C. Frömmel. Dictionary of interfaces in proteins (DIP). Data bank of complementary molecular surface patches. *J Mol Biol*, 280(3):535–550, Jul 1998.
- [80] X. Dong, Z. Zhang, R. Wen, J. Shen, X. Shen, and H. Jiang. Structure-based de novo design, synthesis, and biological evaluation of the indole-based PPARgamma ligands (I). *Bioorg Med Chem Lett*, 16(22):5913–5916, Nov 2006.
- [81] V. Frecer, B. Ho, and J. L. Ding. De novo design of potent antimicrobial peptides. *Antimicrob Agents Chemother*, 48(9):3349–3357, Sep 2004.
- [82] H. Briem. De novo design methods. *Ernst Schering Res Found Workshop*, (42):153–166, 2003.
- [83] M. L. Verdonk, P. N. Mortenson, R. J. Hall, M. J. Hartshorn, and C. W. Murray. Protein-ligand docking against non-native protein conformers. *J Chem Inf Model*, 48(11):2214–2225, Nov 2008.
- [84] S.-Y. Huang, S. Z. Grinter, and X. Zou. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys*, 12(40):12899–12908, Oct 2010.
- [85] M. Fischer and R. E. Hubbard. Fragment-based ligand discovery. *Mol Interv*, 9(1):22–30, Feb 2009.
- [86] P. J. Hajduk and J. Greer. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov*, 6(3):211–219, Mar 2007.
- [87] C. A. Lipinski. Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discov Today*, 8(1):12–16, Jan 2003.
- [88] J. Hodgson. ADMET—turning chemicals into drugs. *Nat Biotechnol*, 19(8):722–726, Aug 2001.
- [89] B. Néron, H. Ménager, C. Maufrais, N. Joly, J. Maupetit, S. Letort, S. Carrere, P. Tuffery, and C. Letondal. Mobylye: a new full web bioinformatics framework. *Bioinformatics*, 25(22):3005–3011, Nov 2009.

- [90] B. A. Miller, S. M. Scoppa, and E. J. Feuer. Racial/ethnic patterns in lifetime and age-conditional risk estimates for selected cancers. *Cancer*, 106(3):670–682, Feb 2006.
- [91] O. B. Snyder, J. J. Kelly, and A. P. Lanier. Prostate cancer in Alaska Native men, 1969-2003. *Int J Circumpolar Health*, 65(1):8–17, Feb 2006.
- [92] N. S. Wampler, T. L. Lash, R. A. Silliman, and T. C. Heeren. Breast cancer survival of American Indian/Alaska Native women, 1973-1996. *Soz Präventivmed*, 50(4):230–237, 2005.
- [93] J. H. Houtgraaf, J. Versmissen, and W. J. van der Giessen. A concise review of DNA damage checkpoints and repair in mammalian cells. *Cardiovasc Revasc Med*, 7(3):165–172, 2006.
- [94] M. Nakanishi, M. Shimada, and H. Niida. Genetic instability in cancer cells by impaired cell cycle checkpoints. *Cancer Sci*, 97(10):984–989, Oct 2006.
- [95] W.-F. Li, Z. Hu, B. Zhang, M.-Z. Cao, Y.-S. Wang, X.-Y. Liu, L. Zhao, Y.-B. Liu, W.-T. Yuan, Z.-Z. Shen, W. H., and Z.-M. Shao. [BRCA1 1100delAT is a recurrent mutation in Chinese women with familial breast cancer]. *Zhonghua Yi Xue Za Zhi*, 87(2):76–80, Jan 2007.
- [96] G. Balatzenko, J. Grudeva-Popova, T. Krustev, D. Marichkov, and M. Stefanova. Philadelphia variant, t(5;9;22)(q13;q34;q11), in a case with chronic myeloid leukemia. *J BUON*, 8(1):65–67, 2003.
- [97] C. D. Bajdik, B. Kuo, S. Rusaw, S. Jones, and A. Brooks-Wilson. CGMIM: automated text-mining of Online Mendelian Inheritance in Man (OMIM) to identify genetically-associated cancers and candidate genes. *BMC Bioinformatics*, 6:78, 2005.
- [98] Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res*, 11(8):1425–1433, Aug 2001.
- [99] R. Nielsen, C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton, M. J. Hubisz, A. Fledel-Alon, D. M. Tanenbaum, D. Civello, T. J. White, J. J. Sninsky, M. D. Adams, and M. Cargill. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*, 3(6):e170, Jun 2005.

- [100] E. Camon, D. Barrell, V. Lee, E. Dimmer, and R. Apweiler. The Gene Ontology Annotation (GOA) Database—an integrated resource of GO annotations to the UniProt Knowledgebase. *In Silico Biol*, 4(1):5–6, 2004.
- [101] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O’Donovan, N. Redaschi, and L.-S. Yeh. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*, 32(Database issue):D115–D119, Jan 2004.
- [102] M. Kanehisa. The KEGG database. *Novartis Found Symp*, 247:91–101; discussion 101–3, 119–28, 244–52, 2002.
- [103] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*, 37(Web Server issue):W623–W633, Jul 2009.
- [104] M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman, and T. E. Klein. PharmGKB: the Pharmacogenetics Knowledge Base. *Nucleic Acids Res*, 30(1):163–165, Jan 2002.
- [105] A. E. Hodge, R. B. Altman, and T. E. Klein. The PharmGKB: integration, aggregation, and annotation of pharmacogenomic data and knowledge. *Clin Pharmacol Ther*, 81(1):21–24, Jan 2007.
- [106] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May 2000.
- [107] T. J. P. Hubbard, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokoinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel,

- S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. Ensembl 2007. *Nucleic Acids Res*, 35(Database issue):D610–D617, Jan 2007.
- [108] T. Derrien, C. André, F. Galibert, and C. Hitte. AutoGRAPH: an interactive web server for automating and visualizing comparative genome maps. *Bioinformatics*, 23(4):498–499, Feb 2007.
- [109] M. E. Sparks S. Kurtz G. Gremme, V. Brendel. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, 47, 2005.
- [110] K. D. Pruitt, T. Tatusova, and D. R. Maglott. NCBI Reference Sequence project: update and current status. *Nucleic Acids Res*, 31(1):34–37, Jan 2003.
- [111] K. D. Pruitt, T. Tatusova, W. Klimke, and D. R. Maglott. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res*, 37(Database issue):D32–D36, Jan 2009.
- [112] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. GenBank. *Nucleic Acids Res*, 38(Database issue):D46–D51, Jan 2010.
- [113] C. Burks, J. W. Fickett, W. B. Goad, M. Kanehisa, F. I. Lewitter, W. P. Rindone, C. D. Swindell, C. S. Tung, and H. S. Bilofsky. The GenBank nucleic acid sequence database. *Comput Appl Biosci*, 1(4):225–233, Dec 1985.
- [114] E. de Castro, C. J. A. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*, 34(Web Server issue):W362–W365, Jul 2006.
- [115] V. Ramensky, P. Bork, and S. Sunyaev. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 30(17):3894–3900, Sep 2002.
- [116] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48(3):443–453, Mar 1970.

- [117] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *J Mol Biol*, 147(1):195–197, Mar 1981.
- [118] K. Fukami-Kobayashi and N. Saito. [How to make good use of CLUSTALW]. *Tanpakushitsu Kakusan Koso*, 47(9):1237–1239, Jul 2002.
- [119] C. Notredame, D. G. Higgins, and J. Heringa. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, 302(1):205–217, Sep 2000.
- [120] A. D. Baxevanis. Searching the NCBI databases using Entrez. *Curr Protoc Bioinformatics*, Chapter 1:Unit 1.3, Mar 2006.
- [121] R. Edgar, M. Domrachev, and A. E. Lash. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, 30(1):207–210, Jan 2002.
- [122] J. P. Jenuth. The NCBI. Publicly available tools and resources on the Web. *Methods Mol Biol*, 132:301–312, 2000.
- [123] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden. NCBI BLAST: a better web interface. *Nucleic Acids Res*, 36(Web Server issue):W5–W9, Jul 2008.
- [124] M. Congreve, R. Carr, C. Murray, and H. Jhoti. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today*, 8(19):876–877, Oct 2003.
- [125] U. Schmidt, S. Struck, B. Gruening, J. Hossbach, I. S. Jaeger, R. Parol, U. Lindquist, E. Teuscher, and R. Preissner. SuperToxic: a comprehensive database of toxic compounds. *Nucleic Acids Res*, 37(Database issue):D295–D299, Jan 2009.
- [126] X. Q. Lewell, D. B. Judd, S. P. Watson, and M. M. Hann. RECAP—retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci*, 38(3):511–522, 1998.
- [127] J. Ahmed, C. L. Worth, P. Thaben, C. Matzig, C. Blasse, M. Dunkel, and R. Preissner. FragmentStore—a comprehensive database of fragments linking metabolites, toxic molecules and drugs. *Nucleic Acids Res*, Oct 2010.

- [128] R. W. Hooft, C. Sander, M. Scharf, and G. Vriend. The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. *Comput Appl Biosci*, 12(6):525–529, Dec 1996.
- [129] P. W. Hildebrand, A. Goede, R. A. Bauer, B. Gruening, J. Ismer, E. Michalsky, and R. Preissner. SuperLooper—a prediction server for the modeling of loops in globular and membrane proteins. *Nucleic Acids Res*, 37(Web Server issue):W571–W574, Jul 2009.
- [130] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics*, 22(2):195–201, Jan 2006.
- [131] R. A. Laskowski, M. W. MacArthur, D.S. Moss, and J.M. Thornton. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2):283–291, Apr 1993.
- [132] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*, Chapter 2:Unit 2.9, Nov 2007.
- [133] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, and A. Sali. Protein structure modeling with MODELLER. *Methods Mol Biol*, 426:145–159, 2008.
- [134] O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L. A. Müller, S. Y. Rhee, and M. Stitt. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*, 37(6):914–939, Mar 2004.
- [135] M. Füllbeck, X. Huang, R. Dumdey, C. Frommel, W. Dubiel, and R. Preissner. Novel curcumin- and emodin-related compounds identified by in silico 2D/3D conformer screening induce apoptosis in tumor cells. *BMC Cancer*, 5:97, 2005.
- [136] S. Lorenzen, M. Dunkel, and R. Preissner. In silico screening of drug databases for TSE inhibitors. *Biosystems*, 80(2):117–122, May 2005.
- [137] A. Smellie, R. Stanton, R. Henne, and S. Teig. Conformational analysis by intersection: CONAN. *J Comput Chem*, 24(1):10–20, Jan 2003.

- [138] W. Kabsch, H. Kabsch, and D. Eisenberg. Packing in a new crystalline form of glutamine synthetase from *Escherichia coli*. *J Mol Biol*, 100(3):283–291, Jan 1976.
- [139] J. Ahmed, J. Hossbach, P. Wrede, and R. Preissner. *Development of specific gamma secretase inhibitors*. Handbook of Research on Systems Biology Applications in Medicine (ISBN: 978-1-60566-076-9), 2008.
- [140] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor. Improved protein-ligand docking using GOLD. *Proteins*, 52(4):609–623, Sep 2003.
- [141] A. Cheng and K. M. Merz. Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships. *J Med Chem*, 46(17):3572–3580, Aug 2003.
- [142] W. J. Egan and G. Lauri. Prediction of intestinal permeability. *Adv Drug Deliv Rev*, 54(3):273–289, Mar 2002.
- [143] R. G. Susnow and S. L. Dixon. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. *J Chem Inf Comput Sci*, 43(4):1308–1315, 2003.
- [144] A. Cheng and S. L. Dixon. In silico models for the prediction of dose-dependent human hepatotoxicity. *J Comput Aided Mol Des*, 17(12):811–823, Dec 2003.
- [145] W. J. Egan, K. M. Merz, and J. J. Baldwin. Prediction of drug absorption using multivariate statistics. *J Med Chem*, 43(21):3867–3877, Oct 2000.
- [146] S. L. Dixon and K. M. Merz. One-dimensional molecular representations and similarity calculations: methodology and validation. *J Med Chem*, 44(23):3795–3809, Nov 2001.
- [147] S. Gunther, S. Neumann, J. Ahmed, and R. Preissner. Cellular Fingerprints: A Novel Concept for the Integration of Experimental Data and Compound-Target-Pathway Relations. pages LNBI 4643: 167–70, 2007.
- [148] M. Dunkel, S. Günther, J. Ahmed, B. Wittig, and R. Preissner. SuperPred: drug classification and target prediction. *Nucleic Acids Res*, 36(Web Server issue):W55–W59, Jul 2008.

- [149] S. Günther, M. Kuhn, M. Dunkel, M. Campillos, C. Senger, E. Petsalaki, J. Ahmed, E. G. Urdiales, A. Gewiess, L. J. Jensen, R. Schneider, R. Skoblo, R. B. Russell, P. E. Bourne, P. Bork, and R. Preissner. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res*, 36(Database issue):D919–D922, Jan 2008.
- [150] A. Goede, M. Dunkel, N. Mester, C. Frommel, and R. Preissner. SuperDrug: a conformational drug database. *Bioinformatics*, 21(9):1751–1753, May 2005.
- [151] E. Michalsky, M. Dunkel, A. Goede, and R. Preissner. SuperLigands - a database of ligand structures derived from the Protein Data Bank. *BMC Bioinformatics*, 6:122, 2005.
- [152] J. W. Ko, Z. Desta, N. V. Soukhova, T. Tracy, and D. A. Flockhart. In vitro inhibition of the cytochrome P450 (CYP450) system by the antiplatelet drug ticlopidine: potent effect on CYP2C19 and CYP2D6. *Br J Clin Pharmacol*, 49(4):343–351, Apr 2000.
- [153] J. Ahmed, T. Meinel, M. Dunkel, M. S. Murgueitio, R. Adams, C. Blasse, A. Eckert, S. Preissner, and R. Preissner. CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res*, Oct 2010.
- [154] J. Lamb, E. D. Crawford, D. P., J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, Sep 2006.
- [155] T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res*, 39(Database issue):D1005–D1010, Jan 2011.
- [156] J. Ahmed, S. Günther, F. Müller, and R. Preissner. A structural genomics approach to the regulation of apoptosis: chimp vs. human. *Genome Inform*, 18:22–34, 2007.

- [157] Y.-N. Wang, H. Yamaguchi, L. Huo, Y. Du, H.-J. Lee, H.-H. Lee, H. Wang, J.-M. Hsu, and M.-C. Hung. The translocon SEC61beta localized in the inner nuclear membrane transports membrane-embedded EGF receptor to the nucleus. *J Biol Chem*, Oct 2010.
- [158] T. Harrison, P.J. Oakley, M.R. Teall, S. William, and B.J. Williams. Sulphones which modulate the action of gamma secretase, Pub. No.: WO 02/081433 A1, October 2002.
- [159] I. J. Collins, J. C. Hannam, T. Harrison, S. J. Lewis, A. Madin, T. J. Sparey, and B. J. Williams. Sulfamides as gamma-secretase inhibitors, Pub. No.: WO 02/36555 A1, May 2002.
- [160] M. R. Teall. Gamma secretase inhibitors., Pub. No.: US 6,448,229 B2, September 2002.
- [161] J. L. Castro Pineiro, I. Churcher, K. Dinnell, T. Harrison, S. Kerrad, A.J. Nadin, P.J. Oakley, A.P. Owens, D.E. Shaw, M.R. Teall, S. William, and B.J. Williams. Sulphones which modulate the action of gamma secretase, Pub. No.: WO 02/081435 A1, October 2002.
- [162] R. Schwarzer, M. Kaiser, O. Acikgoez, U. Heider, S. Mathas, R. Preissner, O. Sezer, B. Doerken, and F. Jundt. Notch inhibition blocks multiple myeloma cell-induced osteoclast activation. *Leukemia*, 22(12):2273–2277, Dec 2008.
- [163] A. C. Wallace, R. A. Laskowski, and J. M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, 8(2):127–134, Feb 1995.
- [164] Y. C. Martin, J. L. Kofron, and L. M. Traphagen. Do structurally similar molecules have similar biological activity? *J Med Chem*, 45(19):4350–4358, Sep 2002.
- [165] M. Campillos, M. Kuhn, A.-C. Gavin, L. J. Jensen, and P. Bork. Drug target identification using side-effect similarity. *Science*, 321(5886):263–266, Jul 2008.
- [166] Z. k. Wang, Y. Hu, H. Zhao, and C. Fu. Thymidylate synthase expression and therapeutic effect analysis of pemetrexed in advanced lung adenocarcinoma. *Nan Fang Yi Ke Da Xue Xue Bao*, 30(5):978–980, May 2010.

- [167] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*, 33(Database issue):D514–D517, Jan 2005.
- [168] A. Kamburov, C. Wierling, H. Lehrach, and R. Herwig. ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res*, 37(Database issue):D623–D628, Jan 2009.
- [169] J. M. Fernández, R. Hoffmann, and A. Valencia. iHOP web services. *Nucleic Acids Res*, 35(Web Server issue):W21–W26, Jul 2007.
- [170] R. D. Finn, J. Mistry, J. Tate, P. Coghill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, and A. Bateman. The Pfam protein families database. *Nucleic Acids Res*, 38(Database issue):D211–D222, Jan 2010.
- [171] P. Flicek, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, T. Eyre, S. Fitzgerald, J. Fernandez-Banet, S. Gräf, S. Haider, M. Hammond, R. Holland, K. L. Howe, K. Howe, N. Johnson, A. Jenkinson, A. Kähäri, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, K. Megy, P. Meidl, B. Overduin, A. Parker, B. Pritchard, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. J. Vilella, J. Vogel, S. White, M. Wood, E. Birney, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, T. J P Hubbard, A. Kasprzyk, G. Proctor, J. Smith, A. Ureta-Vidal, and S. Searle. Ensembl 2008. *Nucleic Acids Res*, 36(Database issue):D707–D714, Jan 2008.
- [172] I. Vastrik, P. D’Eustachio, E. Schmidt, G. Joshi-Tope, G. Gopinath, D. Croft, B. de Bono, M. Gillespie, B. Jassal, S. Lewis, L. Matthews, G. Wu, E. Birney, and L. Stein. Reactome: a knowledge base of biologic pathways and processes. *Genome Biol*, 8(3):R39, 2007.
- [173] T. Meinel, A. Krause, H. Luz, M. Vingron, and E. Staub. The SYSTERS Protein Family Database in 2005. *Nucleic Acids Res*, 33(Database issue):D226–D229, Jan 2005.

- [174] M. Unterhalt, R. Herrmann, M. Tiemann, R. Parwaresch, H. Stein, L. Trümper, M. Nahler, M. Reuss-Borst, C. Tirier, A. Neubauer, M. Freund, E. D. Kreuser, H. Dietzfelbinger, H. Bodenstein, A. Engert, R. Stauder, H. Eimermacher, K. Landys, and W. Hiddemann. Prednimustine, mitoxantrone (PmM) vs cyclophosphamide, vincristine, prednisone (COP) for the treatment of advanced low-grade non-Hodgkin's lymphoma. German Low-Grade Lymphoma Study Group. *Leukemia*, 10(5):836–843, May 1996.
- [175] W. C. Ross. Aryl-2-Halogenoalkylarylamines. XXI. The Design of Agents to Exploit Differences in Cellular Oxidation-Reduction Potentials. *Biochem Pharmacol*, 13:969–982, Jul 1964.
- [176] C. E. Finch. Evolution in health and medicine Sackler colloquium: Evolution of the human lifespan and diseases of aging: roles of infection, inflammation, and nutrition. *Proc Natl Acad Sci U S A*, 107 Suppl 1:1718–1724, Jan 2010.
- [177] E. Weis, D. Galetzka, H. Herlyn, E. Schneider, and T. Haaf. Humans and chimpanzees differ in their cellular response to DNA damage and non-coding sequence elements of DNA repair-associated genes. *Cytogenet Genome Res*, 122(2):92–102, 2008.
- [178] H. Watanabe, A. Fujiyama, M. Hattori, T. D. Taylor, A. Toyoda, Y. Kuroki, H. Noguchi, A. BenKahla, H. Lehrach, R. Sudbrak, M. Kube, S. Taenzer, P. Galgoczy, M. Platzer, M. Scharfe, G. Nordsiek, H. Blöcker, I. Hellmann, P. Khaitovich, S. Pääbo, R. Reinhardt, H-J. Zheng, X-L. Zhang, G-F. Zhu, B-F. Wang, G. Fu, S-X. Ren, G-P. Zhao, Z. Chen, Y-S. Lee, J-E. Cheong, S-H. Choi, K-M. Wu, T-T. Liu, K-J. Hsiao, S-F. Tsai, C-G. Kim, S. Oota, T. Kitano, Y. Kohara, N. Saitou, H-S. Park, S-Y. Wang, M-L. Yaspo, and Y. Sakaki. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature*, 429(6990):382–388, May 2004.
- [179] G. Glazko, V. Veeramachaneni, M. Nei, and W. Makalowski. Eighty percent of proteins are different between humans and chimpanzees. *Gene*, 346:215–219, Feb 2005.
- [180] S. Sobhanifar, B. Schneider, F. Löhr, D. Gottstein, T. Ikeya, K. Mlynarczyk, W. Pulawski, U. Ghoshdastider, M. Kolinski, S. Filipek, P. Güntert, F. Bernhard, and V. Dötsch. Structural investigation of the C-terminal catalytic

- fragment of presenilin 1. *Proc Natl Acad Sci U S A*, 107(21):9644–9649, May 2010.
- [181] J. Konishi, K. S. Kawaguchi, H. Vo, N. Haruki, A. Gonzalez, D. P. Carbone, and T. P. Dang. Gamma-secretase inhibitor prevents Notch3 activation and reduces proliferation in human lung cancers. *Cancer Res*, 67(17):8051–8057, Sep 2007.
- [182] K. Cullion, K. M. Draheim, N. Hermance, J. Tammam, V. M. Sharma, C. Ware, G. Nikov, V. Krishnamoorthy, P. K. Majumder, and M. A. Kelliher. Targeting the Notch1 and mTOR pathways in a mouse T-ALL model. *Blood*, 113(24):6172–6181, Jun 2009.
- [183] E. Jorissen and B. De Strooper. Gamma-secretase and the intramembrane proteolysis of Notch. *Curr Top Dev Biol*, 92:201–230, 2010.
- [184] R. D. Meng, C. C. Shelton, Y.-M. Li, L.-X. Qin, D. Notterman, P. B. Paty, and G. K. Schwartz. gamma-Secretase inhibitors abrogate oxaliplatin-induced activation of the Notch-1 signaling pathway in colon cancer cells resulting in enhanced chemosensitivity. *Cancer Res*, 69(2):573–582, Jan 2009.
- [185] T. Melese, E. L. Perkins, E. Yeh, and D. Sun. Parp inhibitors, Pub. No.: WO 02/44157 A2, June 2002.
- [186] M. Rajesh, P. Mukhopadhyay, G. Godlewski, S. Batkai, G. Hasko, L. Liaudet, and P. Pacher. Poly(ADP-ribose)polymerase inhibition decreases angiogenesis. *Biochem Biophys Res Commun*, 350(4):1056–1062, Dec 2006.
- [187] N. D. Mendes, A. T. Freitas, and M-F. Sagot. Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Res*, 37(8):2419–2433, May 2009.
- [188] D. P. Bartel. MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2):215–233, Jan 2009.
- [189] E. Klipp, R. C. Wade, and U. Kummer. Biochemical network-based drug-target prediction. *Curr Opin Biotechnol*, 21(4):511–516, Aug 2010.

- [190] M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijjer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. H. Thomas, D. D. Edwards, B. K. Shoichet, and B. L. Roth. Predicting new molecular targets for known drugs. *Nature*, 462(7270):175–181, Nov 2009.
- [191] S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, and P. A. Futreal. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*, 39(Database issue):D945–D950, Jan 2011.
- [192] International Cancer Genome Consortium, T. J. Hudson, W. Anderson, A. Artez, A. D. Barker, C. Bell, R. R. Bernabé, and M. K. Bhan et al. International network of cancer genome projects. *Nature*, 464(7291):993–998, Apr 2010.
- [193] V. Khanna and S. Ranganathan. Physicochemical property space distribution among human metabolites, drugs and toxins. *BMC Bioinformatics*, 10 Suppl 15:S10, 2009.
- [194] J. Wang and T. Hou. Drug and drug candidate building block analysis. *J Chem Inf Model*, 50(1):55–67, Jan 2010.

Anhang B

Liste eigener Publikationen

Peer-reviewed Publikationen

1. **Ahmed J.**, Worth C. L., Thaben P., Matzig C., Blasse C., Dunkel M. and Preissner R.
FragmentStore - a comprehensive database of fragments linking metabolites, toxic molecules and drugs
Nucleic Acids Res, 39: D1049-54, 2010
2. **Ahmed J.**, Meinel T., Dunkel M., Mueller M. S., Adams R., Blasse C., Eckert A., Preissner S., and Preissner R.
CancerResource - a comprehensive database of cancer-relevant proteins and interacting compounds
Nucleic Acids Res, 39: D960-7, 2010
3. **Ahmed J.**, Preissner S., Dunkel M., Worth C. L., Eckert A. and Preissner R.
SuperSweet - a resource on natural and artificial sweetening agents
Nucleic Acids Res, 39: D377-82, 2010
4. Meinel T., Mueller M. S., **Ahmed J.**, Yildiriman R., Dunkel M., Herwig R. and Preissner R.
SOAP/WSDL-Based Web Services for Biomedicine: Demonstrating the Technique with the CancerResource
IFMBE Proceedings, Volume 29, Part 5, 835-838, 2010
5. Schmidt U., **Ahmed J.**, Michalsky E., Hoepfner M. and Preissner R.
Comparative VEGF receptor tyrosine kinase modeling for the deve-

- lopment of highly specific inhibitors of tumor angiogenesis.**
Genome Inform 20: 243-251, 2008
6. Dunkel M., Gunther S., **Ahmed J.**, Wittig B. and Preissner R.
SuperPred: drug classification and target prediction.
Nucleic Acids Res, 36: W55-9, 2008
7. Gunther S., Kuhn M., Dunkel M., Campillos M., Senger C., Petsalaki E.,
Ahmed J., Urdiales E. G., Gewiess A., Jensen L. J., Schneider R., Skoblo R.,
Russell R. B., Bourne P. E., Bork P. and Preissner R.
SuperTarget and Matador: resources for exploring drug-target relationships.
Nucleic Acids Res, 36: D919-22, 2008
8. **Ahmed J.**, Gunther S., Moller F. and Preissner R.
A structural genomics approach to the regulation of apoptosis: chimp vs. human.
Genome Inform 18: 22-34, 2007
9. Günther S., Weiser A. A., **Ahmed J.** and Preissner R.
Superhaptens and their potential role in contact dermatitis.
Drug Discovery Today 4(1): 11-17, 2007
10. Gunther S., Neumann S., **Ahmed J.** and Preissner R.
Cellular Fingerprints: A Novel Concept for the Integration of Experimental Data and Compound-Target-Pathway Relations.
LNBI 4643: 167-70, 2007

Buchartikel

- Ahmed J.**, Hossbach J., Wrede P. and Preissner R.
Development of specific gamma secretase inhibitors. Handbook of Research on Systems Biology Applications in Medicine (ISBN: 978-1-60566-076-9): 423-437. (2008)

Patente

Kuhn M., Campillos M., Bork P., Jensen L. J., Gavin A. C., Luciani R., Preissner R., Fan H., Hossbach J., Jing H., **Ahmed J.**

Patent: Use of Aprepitant and derivatives thereof for the treatment of cancer.

U.S. patent. Pub. No.: WO/2009/124756

International Application No.: PCT/EP2009/002621 (2009).

Priority Data: 61/043,299 08.04.2008 US

Acikgöz Ö., **Ahmed J.**, Dörken B., Frömmel C., Goede A., Jundt F., Kunze R. and Preissner R. Patent: Pharmaceutical compositions comprising gamma secretase modulators.

EU patent. Pub. No.: WO/2009/146875

International Application No.: PCT/EP2009/003915 (2009)

Priority Data: 08 010 126.4 03.06.2008 EP

Danksagung

Mein besonderer Dank gilt meinem Supervisor PD Dr. Robert Preißner für die Möglichkeit diese Arbeit in der Strukturellen Bioinformatik anzufertigen und für die große Unterstützung. Einen großen Dank auch an Herrn Prof. Dr. Burghardt Wittig für das Übernehmen der Zweitkorrektur und die Möglichkeit an der Freien Universität Berlin zu promovieren. Vor allem aber möchte ich meiner Kollegin und Freundin Ulrike Schmidt danken für ihr offenes Ohr und ihr Engagement beim Korrekturlesen. Außerdem danke ich allen in der Arbeit genannten Kooperationspartnern und folgenden Kollegen, die mit mir an interessanten Projekten gearbeitet haben, mich beim Korrekturlesen unterstützten oder einfach nur da waren: Stefan Günther, Matthias Dunkel, Andrean Goede, Peter Hildebrand, Catherine Worth, Christian Senger, Björn Grüning, Corinna Blasse, Andreas Eckert, Christian Scholz, Christian Matzig, Thomas Meinel, Anna-Lena Wölke, Manuela Murgueitio, Robert Adams, Elke Michalsky, Joachim von Eichborn und Raphael Bauer.

Ich danke der International Research Training Group (IRTG for Genomics and Systems Biology of Molecular Networks) für die vielen Lernangebote und die Möglichkeiten an Konferenzen teilzunehmen, um den aktuellen Stand der Forschung nicht aus den Augen zu verlieren.

Zu guter Letzt möchte ich die Gelegenheit nutzen mich bei ganz lieben Menschen zu bedanken, die mich privat unterstützt und immer ermutigt haben. Vielen lieben Dank an meine Mom, Mirjana Ahmed, meinem Dad, Jamal Uddin Ahmed, meinem Bruder, David Uddin Ahmed und meinen engsten Vertrauten Christian Schimmel und Daniel Schermer!

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorgelegte Dissertationsschrift mit dem Thema Analyse und Vorhersage der Interaktionen von Wirkstoffen mit ihren Zielmolekülen zur Tumorthherapie selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.