

**Etablierung einer
Grid-Computing Umgebung und virtuelles
Screening an Cyclin-abhängigen Kinasen
von veterinär- und human-
relevanten Parasiten**

DISSERTATION

der Fakultät für Chemie und Pharmazie
der Eberhard-Karls-Universität Tübingen

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

2007

vorgelegt von
Carsten Beyer

Tag der mündlichen Prüfung:

23. Januar 2007

Dekan:

1. Berichterstatter

2. Berichterstatter

Prof. Dr. L. Wesemann

Priv. Doz. Dr. P. M. Selzer

Prof. Dr. M. Duszenko

Die vorliegende Arbeit wurde unter Anleitung von Herrn Priv. Doz. Dr. Paul M. Selzer bei der Intervet Innovation GmbH in Schwabenheim an der Selz in Zusammenarbeit mit der Universität Tübingen durchgeführt.

Mein besonderer Dank gilt Herrn Prof. Dr. Michael Duszenko und Herrn Priv. Doz. Dr. Paul M. Selzer. Ihnen danke ich für die Betreuung der Arbeit, die interessante Themenstellung, sowie die Bereitstellung des Arbeitsplatzes. Sie zeigten stets reges Interesse am Verlauf der Arbeit und standen mir jederzeit mit Rat beiseite.

Herrn Dr. Jörg Cramer möchte ich insbesondere für seine fachliche Betreuung und die vielen Hilfestellungen bei der Durchführung meiner Experimente danken. Seine ständige Diskussionsbereitschaft half mir stets, pragmatische Lösungen zu Problemen zu finden.

Bei der Firma Intervet Innovation GmbH in Schwabenheim möchte ich mich für die Bereitstellung des Arbeitsplatzes bedanken. Dies ermöglichte die Kooperation mit der Universität Tübingen, ohne die diese Arbeit nicht möglich gewesen wäre.

Besonders möchte ich mich bei Herrn Dr. Frank Oellien für die Hilfe bei Fragen der Tcl-Programmierung danken. Frau Maria L. Suarez Fernandez danke ich für die Bereitstellung der Messergebnisse der CDK-Inhibitoren. Frau Kristin Engels danke ich für die konstruktive Unterstützung bei der bioinformatischen Auswertung der Sequenzinformationen. Allen anderen Mitgliedern der BioChemInformatics möchte ich an dieser Stelle für das gute und kollegiale Arbeitsklima danken. Namentlich: Jörg Cramer, Kristin Engels, Maria L. Suarez-Fernandez, Anette Klinger, Andreas Krasky, Richard J. Marhöfer, Frank Oellien, Andreas Rohwer, Jörg Schröder, Hon Q. Tran und Paul M. Selzer.

Herrn Dr. Michael Podvinec danke ich für die Unterstützung bei der Portierung der Software Autodock.

Ebenso möchte ich mich bei Herrn Paul Zwionsek für die konstruktive Unterstützung bei der Administration des Grid-Computing Clusters bedanken.

Herrn Priv. Doz. Dr. Klaus-Jürgen Schleifer danke ich herzlich für das kritische Korrekturlesen dieser Arbeit und seine Diskussionsbereitschaft.

Bei Herrn Hans Peter Cramer und Herrn Peter Kurt Engels möchte ich mich an dieser Stelle herzlich für das Korrekturlesen bedanken.

Dem Boehringer Ingelheim Fonds und der Firma Intervet danke ich für die finanzielle Unterstützung, die es mir ermöglicht hat, am „Fifth European Workshop in Drug Design“ in Siena, Italien teilzunehmen.

Ganz besonderer Dank gilt meinen Eltern und meiner Familie, die mich auf meinem Weg stets unterstützt haben und mein Studium erst ermöglicht haben.

Meiner Lebensgefährtin Kristin Engels gilt nicht zuletzt besonderer Dank für ihre liebevolle Unterstützung, die das Leben in den letzten Monaten erheblich leichter machte. Die vielen fachlichen Diskussionen mit ihr trugen darüberhinaus zu einem wesentlichen Teil dieser Arbeit bei.

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Infektionskrankheiten der Haus- und Nutztiere.....	1
1.2	Human- und veterinärpathogene <i>Apicomplexa</i>	3
1.2.2	Der Lebenszyklus von <i>E. tenella</i>	6
1.3	<i>Target</i> -basierte Wirkstoffentwicklung.....	8
1.3.1	Computerunterstütztes rationales Wirkstoffdesign.....	10
1.3.1.1	Homologiemodellierung.....	11
1.3.1.2	Pharmakophormodellierung.....	14
1.3.1.3	Molekulares <i>Docking</i>	16
1.3.1.4	Quantifizierung von Protein-Ligand Wechselwirkungen.....	19
1.4	Computerinfrastrukturen für virtuelles Screening.....	20
1.4.1	<i>Grid-Computing</i>	20
1.4.1.1	<i>Grid-Computing</i> Software.....	22
1.5	Proteinkinasen als <i>Drug-Targets</i> des rationalen <i>Drug-Design</i>	22
1.5.1	Der eukaryontische Zellzyklus.....	25
1.5.1.1	Die Funktion der Cyclin-abhängigen Kinasen im Zellzyklus.....	26
1.5.1.2	Aufbau und Aktivierung der CDKs.....	27
1.5.2	CDKs als <i>Drug-Targets</i> in der Wirkstoffentwicklung.....	31
1.5.2.1	Das Bindetaschenmodell der CDKs nach Traxler.....	35
1.5.2.2	Inhibitoren der CDKs.....	38
1.6	Zielsetzung der Arbeit.....	41
2	Material und Methoden.....	43
2.1	Material.....	43
2.1.1	Hardware.....	43
2.1.2	Betriebssysteme.....	43
2.1.3	Software.....	43
2.1.4	Datenbanken.....	45
2.1.5	Programmiersprachen.....	46

2.6	Methoden.....	46
2.6.1	Sequenz-Alignments zum direkten Sequenzvergleich.....	46
2.6.2	Bestimmung orthologer Proteine der TgCRK3 in <i>E. tenella</i>	46
2.6.3	Homologiemodellierung der EtCRK2.....	47
2.6.4	Berechnung von EtCRK2-spezifischen Pharmakophormodellen..	47
2.6.5	Datenbanksuche mit der Software Catalyst.....	48
2.6.6	Vorbereitung der Substanzbibliothek.....	49
2.6.7	Docking und Scoring mit der Software Autodock.....	50
2.6.8	Virtuelles Screening mit der EtCRK2.....	52
3	Ergebnisse.....	53
3.1	Aufbau des <i>Grid-Computing Frameworks</i>	53
3.1.1	Aufbau des <i>Grid-Computing Clusters</i>	53
3.1.2	Programmierung.....	54
3.2	Validierung der Methode.....	58
3.2.1	Reproduzierbarkeit von Protein-Ligand-Komplexen.....	58
3.2.2	Evaluierung verschiedener Bewertungsfunktionen.....	61
3.2.3	Datensatzabhängigkeit der Anreicherungsexperimente.....	65
3.2.4	Nachprozessierung der <i>Docking</i> -Strukturen.....	68
3.3	Sequenzsuche und bioinformatische Tiefenanalyse.....	70
3.3.1	Sequenzsuche nach CDK-Orthologen in <i>E. tenella</i>	70
3.3.2	Proteinmotive.....	73
3.3.3	Röntgenkristallstrukturen der CDKs.....	73
3.4	Homologie-Modellierung der EtCRK2.....	75
3.4.1	Validierung des Homologiemodells.....	79
3.4.1.1	RMSD-Berechnung nach <i>Re-Docking</i> bekannter Inhibitoren.....	79
3.4.1.2	Anreicherungsexperimente mit <i>Docking</i> -Datensätzen.....	81
3.5	Pharmakophor-Analyse des Homologie-Modells.....	82
3.5.1	Validierung der Pharmakophor-Ensembles.....	88
3.6	<i>In silico</i> -Screening.....	89
3.6.1	Vorbereitung der Datenbanken.....	89

3.6.1.1	CDK-Datenbanken.....	89
3.6.1.2	Kinase-Datenbanken.....	90
3.6.1.3	<i>Intervet Chemicals Directory</i>	90
3.6.2	Screening.....	91
3.6.2.1	<i>Docking</i> der Datenbanken mit dem Modell der EtCRK2.....	91
3.6.3.2	Nachprozessierung der eingepassten Strukturkonformere.....	92
3.6.3.3	Visuelle Beurteilung der <i>Docking</i> -Experimente.....	93
3.7	Biochemische Testung der ausgewählten Substanzen.....	94
3.7.1	Bestimmung von IC ₅₀ -Werten der Inhibitoren.....	96
4	Diskussion.....	98
4.1	Evaluierung des <i>Grid-Computing Frameworks</i>	98
4.1.1	Validierung der <i>Docking</i> -Ergebnisse.....	100
4.2	Sequenzvergleiche.....	103
4.3	Homologiemodellierung.....	104
4.4	Pharmakophormodellierung.....	107
4.5	Virtuelles Screening.....	109
4.6	Auswahl von Substanzen für biochemische Experimente.....	110
4.7	Ausblick.....	111
5	Zusammenfassung.....	113
6	Literaturverzeichnis.....	115

Abkürzungen

Å	Ångström
AIDS	<i>Acquired Immune Deficiency Syndrom</i>
ATM	<i>Ataxia Telangiectasia Mutated (Kinase)</i>
ATP	Adenosintriphosphat
ATR	<i>Rad3-related (Kinase)</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
CDC	<i>Cell Division Cycle</i>
CDK	<i>Cyclin Dependent Kinase</i>
CIP	<i>CDK Inhibitory Protein</i>
CTX	<i>Gasteiger Cleartext Format</i>
DNS	Desoxyribonucleinsäure
EST	<i>Expressed Sequence Tag</i>
HTS	<i>High Throughput Screening</i>
IC	<i>Inhibitor Concentration</i>
ICD	<i>Intervet Chemicals Directory</i>
INK	<i>Inhibitors of Kinase</i>
kDa	Kilodalton
KIP	<i>Kinase Inhibitor Protein</i>
SDF	<i>Structure Data Format</i>
MYT	<i>Membrane-Associated Tyrosine- and Threonine-specific cdc2-inhibitory Kinase</i>
NCBI	<i>National Centre for Biotechnology Information</i>
nm	Nanometer
PDB	<i>Protein Data Bank</i>
PDF	<i>Probability Density Function</i>
PFPK	<i>Plasmodium falciparum Protein Kinase</i>
RINGO	<i>Rapid Inducer of G2/M Progression in Oocytes</i>
RMSD	<i>Root Mean Square Deviation</i>
SMILES	<i>Simplified Molecular Input Line Entry Specification</i>

Tcl	<i>Tool Command Language</i>
TgCRK	<i>Toxoplasma gondii CDK-related Kinase</i>
TR-FRET	<i>Time Resolved Fluorescence Resonance Energy Transfer</i>
WAF	<i>Wild-type p53 Activated Factor</i>

Aminosäuren

A, Ala	Alanin
V, Val	Valin
F, Phe	Phenylalanin
P, Pro	Prolin
M, Met	Methionin
I, Ile	Isoleucin
L, Leu	Leucin
D, Asp	Asparaginsäure
E, Glu	Glutaminsäure
K, Lys	Lysin
R, Arg	Arginin
S, Ser	Serin
T, Thr	Threonin
Y, Tyr	Tyrosin
H, His	Histidin
C, Cys	Cystein
N, Asn	Asparagin
Q, Gln	Glutamin
W, Trp	Tryptophan

1 Einleitung

1.1 Infektionskrankheiten der Haus- und Nutztiere

Erkrankungen von Haus- und Nutztieren erzeugen alljährlich weltweite wirtschaftliche Schäden in Milliardenhöhe. Dabei spielen Infektionskrankheiten von Nutztieren eine erhebliche Rolle. Zu den Infektionskrankheiten gehören Erkrankungen, die durch Viren, Bakterien, Pilze und Parasiten hervorgerufen werden [GRASSLY & FRASER, 2006]. Ein prominentes Beispiel für eine Viruserkrankung von Haus- und Nutztieren ist die Vogelgrippe, die sich derzeit in Asien und Europa stark verbreitet. Sie wird durch ein aviäres Influenzavirus verursacht, das auch auf den Menschen übertragbar ist und zum Tod führen kann [PERDUE & SWAYNE, 2005]. Dagegen erreichen die bakteriellen Erkrankungen eine geringere Aufmerksamkeit in den öffentlichen Medien, obwohl sie durch die Resistenzentstehung gegen gängige Antibiotika eine zunehmende Bedrohung für Mensch und Tier darstellen [HOGAN & KOLTER, 2002]. Ein verbreitetes Bakterium, ist das human- und veterinärpathogene Bakterium *Staphylococcus aureus*. Es verursacht u.a. die Mastitis, eine Infektion des Kuheuters [OTT, 1999]. Zugleich bereitet dieser Erreger in der Intensivmedizin große Probleme, da bereits Stämme dieses Erregers existieren, die gegen die Mehrzahl aller therapeutisch relevanten Antibiotika resistent sind [GOOSSENS, 2005].

Eine persistierende Bedrohung - nicht nur in der Tierhaltung - sind beispielsweise die Parasiten *Leishmania spec.*, *Ascaris spec.* und *Phlebotomus spec.* [PINK *et al.*, 2005]. Die Gruppe der Parasiten lässt sich in die Endo- und die Ektoparasiten unterteilen, wobei diese Unterteilung nicht der Systematik des Tierreiches entspricht, sondern vielmehr die unterschiedlichen Lebensräume eines Parasiten als Unterscheidungskriterium herangezogen werden. Während die Ektoparasiten den Wirt äußerlich befallen, dringen die Endoparasiten in den Wirt ein, vermehren sich dort und durchlaufen häufig komplizierte Generationswechsel, bis sie schliesslich weiterverbreitet werden [MEHLHORN,

2001]. Die Übertragungswege finden bisweilen über andere Ektoparasiten zu einem weiteren Wirt, wie das Beispiel der Leishmaniose zeigt. Leishmanien sind eukaryontische Einzeller (Protozoen) und befallen die Haut und innere Organe. *Leishmania spec.* ist im Mittelmeerraum weit verbreitet und stellt eine Gefährdung für Mensch und Tier dar. Die Übertragung der Leishmanien auf den Wirbeltierwirt findet durch die Sandfliege *Phlebotomus spec.* statt, die selbst ein blutsaugender Ektoparasit ist [MEHLHORN, 2001; MAROLI & KHOURY, 2004]. Weitere Endoparasiten sind zum Beispiel die Bandwürmer (*Cestodes*), die meist mit kontaminierter Nahrung aufgenommen werden und sich in inneren Organen und im Darm des Wirts vermehren [MEHLHORN, 2001].

In weiten Teilen der dritten Welt stellen parasitäre Erkrankungen eine ernste Bedrohung der dort lebenden Bevölkerung dar, da die Parasiten nicht nur die Nahrungsgrundlage der Menschen zerstören, sondern durch den Befall des Menschen schreckliche Erkrankungen verursachen können. Die Therapeutika, die heute gegen viele Parasiten eingesetzt werden, stammen häufig noch aus der ersten Hälfte des letzten Jahrhunderts [MEHLHORN, 2001]. Die Wirksamkeit dieser Medikamente ist häufig eingeschränkt und mit heftigen Nebenwirkungen verbunden. Zudem sind resistente Parasitenstämme entstanden, wie das Beispiel der Malaria und ihres Überträgers, der Mücke *Anopheles gambiae*, zeigt [KEENAN & WELSH, 2004]. Derzeit leben etwa 2,1 Milliarden Menschen in Malaria-Risikogebieten und es starben im Jahr 2002 weltweit 1,272 Millionen Menschen an der Malaria [PINK *et al.* 2005]. Diese Zahlen verdeutlichen die Notwendigkeit effektiver Medikamente.

1.2 Human- und veterinärpathogene *Apicomplexa*

Zur Klasse der *Apicomplexa* gehört u.a. der Parasit *Plasmodium falciparum*, welches der Erreger der Malaria ist. Ein weiterer Parasit aus der Klasse der *Apicomplexa* ist *Toxoplasma gondii*. *T. gondii* gilt als ein gefährlicher Parasit des Menschen, der weit verbreitet ist und starke Schäden am sich entwickelnden Embryo während einer Schwangerschaft verursacht. Weltweit sind etwa die Hälfte der Menschen mit *Toxoplasma* infiziert, wobei der Parasit vom Immunsystem des Wirts in aller Regel gut kontrolliert wird. Benötigen latent mit *T. gondii* infizierte Personen immunsupprimierende Medikamente oder leiden an einer Immunschwäche -wie etwa der Krankheit AIDS-, kann es zur Vermehrung von *T. gondii* im Organismus kommen, was zum Krankheitsbild der Toxoplasmose führt [McALLISTER, 2005]. Wirkstoffe für die Behandlung der Toxoplasmose sind u.a. Kombinationen aus Pyrimethamin und Sulfadiazin, sowie Atovaquon und Pyrimethamin.

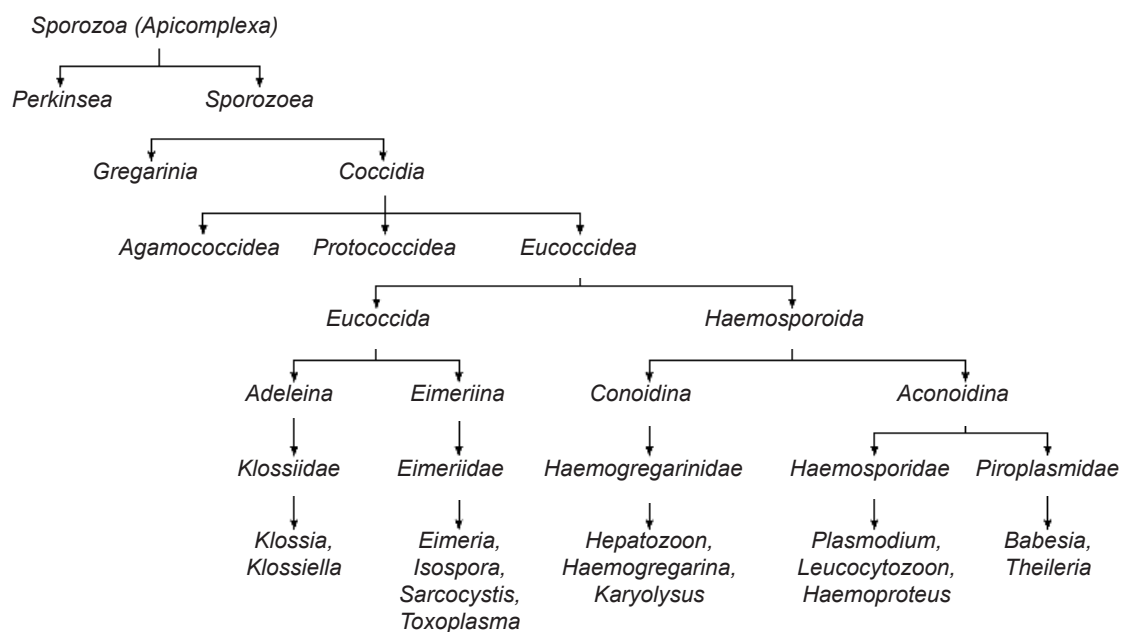


Abb. 1.1 Kladogramm der Apicomplexa [nach MEHLHORN, 2001]. Häufige human-pathogene Apicomplexa sind *Plasmodium spec.* und *Toxoplasma spec.*

T. gondii gehört zur Familie der *Eimeriidae*, die zur Ordnung der *Eucoccidia* zählt. Die Ordnung der *Eucoccidia* beinhaltet die Subordnungen *Adeleina* und *Eimeriina*. Innerhalb der Subordnung der *Eimeriina* sind *Eimeria*, *Isospora*, *Sarcocystis* und *Toxoplasma* in der Familie der *Eimeriidae* vereint (Abb. 1.1). Die bekannten Mitglieder der Familie der *Eimeriidae* sind Parasiten von Mensch und Tier [CAI *et al.*, 2003].

E. tenella ist ein weiteres Mitglied der *Eimeriidae*, das vor allem in der Geflügelhaltung weit verbreitet ist und dort große wirtschaftliche Schäden verursacht (Abb. 1.2). Innerhalb der Gattung *Eimeria* sind ca. 30 Arten bekannt, die u.a. Kühe, Schafe, Ziegen, Schweine, Pferde, Kaninchen, Hühner und Enten befallen können und zum Krankheitsbild der Kokzidiose führen. Symptome der akuten Kokzidiose sind starke, teils hämorrhagische Durchfälle. Dabei wird das Darmepithel großer Darmabschnitte zerstört, was zu erheblichen pathophysiologischen Störungen führt und schließlich zum Tod des Wirts führen kann. Häufig finden Koinfektionen mit weiteren Krankheitserregern statt [MEHLHORN, 2001]. Die unterschiedlichen *Eimeria*



Abb.1.2 Sporulierte Oocysten von *E. tenella* (Mit freundlicher Genehmigung von R. Hollingshausen, Intervet Innovation GmbH).

Arten sind streng wirts- und gewebespezifisch und deren Verbreitung erfolgt über kontaminierte Nahrung, die mit Infektionsstadien (Oocysten) verunreinigt ist. Unter natürlichen Bedingungen werden Säugetiere, im Gegensatz zu Vögeln, eher selten befallen. Allerdings bereitet *Eimeria* in der Massentierhaltung große Probleme, da hier die Verbreitung über kontaminierte Nahrung begünstigt wird. Der Schweregrad einer Infektion mit *Eimeria* hängt von der Anzahl der vom Wirt aufgenommenen Oocysten ab. Im Unterschied zu anderen Parasiten findet im Wirtsorganismus nur eine asexuelle Vermehrung statt, die auf eine begrenzte Anzahl limitiert ist. Dennoch bewirkt die Kokzidiose eine massive Zerstörung des intestinalen Epithels. Insbesondere in der Geflügelproduktion bereitet *Eimeria* sehr große Probleme und verursacht einen wirtschaftlichen Schaden in Höhe von ca. 300 Millionen Dollar pro Jahr [Interne Schätzung der Firma Intervet, 2005]. Eine wirtschaftliche Produktion von Geflügel ist daher nur mit einer begleitenden Chemoprophylaxe gegen *Eimeria* möglich. Ziel einer Chemoprophylaxe ist die Reduktion der Verseuchung durch Oocysten. Obwohl die Infektion selbstbegrenzend ist, ist eine Reinfektion möglich, sofern die verabreichten Wirkstoffe nicht gegen alle *Eimeria* Arten gerichtet sind oder Resistenzen auftreten.

Derzeit werden verschiedene Wege der Therapie und der Prophylaxe beschrieben. Eine Chemoprophylaxe wird durch eine kontinuierliche Gabe von Kokzidiostatika erreicht. Zu den Kokzidiostatika gehören die Wirkstoffe Clopidol, Decoquinat, Buquinolat und Methylbenzoquat. Diese Wirkstoffe behindern die Entwicklung verschiedener Entwicklungsstadien der Parasiten. Jedoch ist die Behandlung mit Kokzidiostatika reversibel, d.h. ein Absetzen der Therapie bewirkt die Fortsetzung des Entwicklungszyklus und eine fortschreitende Vermehrung der Parasiten. Durch den prophylaktischen Einsatz von Kokzidiostatika ist die Resistenzentwicklung gegen einige Wirkstoffe so weit vorangeschritten, dass ein Einsatz in der Geflügelproduktion kaum mehr Wirkung zeigt [WILLIAMS, 2006]. Im Gegensatz zu den Kokzidiostatika wirken die Kokzidiozide toxisch auf einzelne oder mehrere Entwicklungsstadien der Parasiten. Zu den Kokzidioziden gehören die Wirkstoffe Amprolium,

Diclazuril (Clinacox[®]) und Toltrazuril (Baycox[®]). Diese Wirkstoffe können zur Bekämpfung von Ausbrüchen der Kokzidiose eingesetzt werden. In der Tierhaltung werden häufig Futterzusatzmittel prophylaktisch verabreicht, die sowohl eine kokzidiostatische, als auch eine kokzidoizide Wirkung haben. Zu diesen Wirkstoffen zählen die Polyether-Antibiotika und die Ionophore [MEHLHORN, 2001]. Häufig werden zur Prophylaxe auch Wirkstoffkombinationen eingesetzt. Derzeit werden weitere Wirkstoffe erforscht um die Therapiemöglichkeiten zu erweitern [HAFEEZ *et al.*, 2006]. Insbesondere *E. tenella* dient hierbei als Zielorganismus, da *E. tenella* zum einen als Modellorganismus dient und zum anderen eine weite Verbreitung in der Geflügelproduktion hat.

1.2.2 Der Lebenszyklus von *E. tenella*

Der Lebenszyklus von *E. tenella* ist in Abb. 1.3 schematisch dargestellt. Nach der Aufnahme von sporulierten Oocysten (Abb. 1.3 [12-13]) schlüpfen die Sporozoiten im Dünndarm des Huhns aus den Sporocysten (Abb. 1.3 [1]) und penetrieren die Epithelzellen des Darms. Danach werden mehrkernige Schizonten in einer parasitophoren Vakuole gebildet. Aus den Schizonten gehen bewegliche Merozoiten hervor (Abb. 1.3 [2-5]), welche eine weitere Generation von Schizonten in anderen Wirtszellen hervorbringen können, oder zu männlichen und weiblichen Gamonten werden (Abb. 1.3 [6]). Die Gamonten entwickeln sich schließlich zu begeißelten mehrkernigen Mikrogameten (Abb. 1.3 [7]) und zu einkernigen Makrogameten (Abb. 1.3 [8]) weiter. Nach der Befruchtung entstehen aus den Zygoten neue Oocysten (Abb. 1.3 [9-10]). Nicht-sporulierte Oocysten werden mit dem Kot ausgeschieden, sodaß der Zyklus fortgesetzt werden kann. Die Sporulation der Oocysten außerhalb des Wirts ist temperaturabhängig und führt zur Bildung von 4 Sporocysten. Dadurch ist der Infektionszyklus geschlossen und es kann eine Neuinfektion stattfinden [Mehlhorn, 2001].

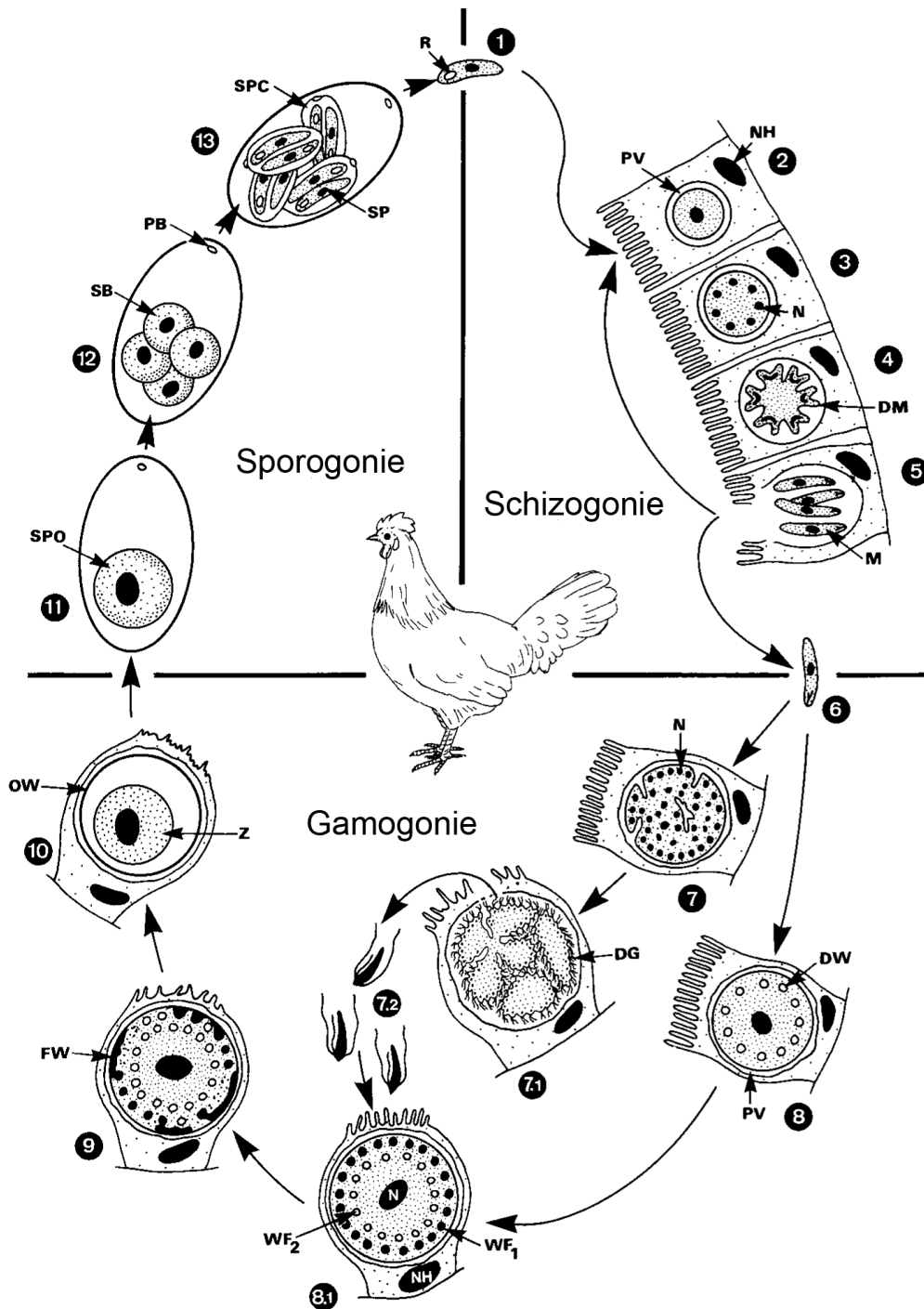


Abb. 1.3 Der Lebenszyklus von *Eimeria spec.* im Huhn [aus MEHLHORN, 2001]. (1) Nach der oralen Aufnahme der sporulierten Oocysten werden die Sporoziten freigesetzt und befallen den Dünndarm. Sie durchlaufen dort die Schizogonie sowie die Gamogonie bevor sie als unsporulierte Oocysten (10) mit dem Kot ausgeschieden werden. Sie durchlaufen dann die Sporogonie, bevor sie vom Wirt wieder aufgenommen werden.

Soll der Lebenszyklus von *Eimeria* effektiv unterbrochen und somit die weitere Verbreitung von infektiösen Sporozoiten verhindert werden, ist die Hemmung des Lebenszyklus des Parasiten von entscheidender Bedeutung. Ein mögliches Ziel einer Chemoprophylaxe ist deshalb die Unterbrechung des Zellzyklus des Parasiten.

1.3 **Target-basierte Wirkstoffentwicklung**

Die Geschichte der systematischen Wirkstoffforschung mit dem Ziel, neue Medikamente zu entwickeln, geht etwa auf die zweite Hälfte des 19. Jahrhunderts zurück. Seit dieser Zeit wurde die Wirkstofffindung sehr erfolgreich von der ungezielten direkten Testung (Screening) chemischer Substanzen in biologischen Systemen - meist Versuchstieren - bestimmt. Seit etwa den sechziger Jahren des letzten Jahrhunderts geht die Zahl der neuen Medikamente jedoch stetig zurück.

Durch die Sequenzierung ganzer Genome und die Generierung der dazugehörigen biologischen Information, was häufig unter den Begriffen *Genomics* und *Functional Genomics* zusammengefasst wird, hat sich der Forschungsansatz heute umgekehrt. Ausgangspunkt für die Entwicklung eines neuen Wirkstoffs ist dabei die Identifizierung eines Zielmoleküls (*Drug-Target*), das eine Schlüsselfunktion in einer Krankheit hat. Nachdem diese Funktion experimentell bestätigt wurde (*Drug-Target Validierung*), versucht man chemische Substanzen zu identifizieren, die das *Drug-Target* so beeinflussen, dass die entsprechende Krankheit gelindert oder geheilt werden kann (*Target-based Approach*). Die gezielte Hemmung eines Enzyms durch einen chemischen Inhibitor wäre ein mögliches Beispiel. Die Identifizierung solcher Substanzen wird häufig in automatisierten Tests mit hohem Durchsatz im *High Throughput* Screening (HTS) von sehr großen chemischen Substanzsammlungen an *in vitro* Testsystemen (z.B. Enzym-, Rezeptor- oder Zellkultur-Assays) durchgeführt [SELZER, 2005]. So identifizierte Leitstrukturen - chemische Substanzen die bereits eine gewünschte Wirkung zeigen, jedoch noch nicht

alle Eigenschaften eines wirklichen Arzneimittelkandidaten aufweisen - können dann chemisch variiert werden, um ihr Wirkstoffprofil am Testsystem zu optimieren. Erst wenn ein Mindestmaß geforderter Eigenschaften einer solchen Substanz erreicht wird (z.B. Wirkstärke, Selektivität, Wirkdauer, Bioverfügbarkeit, Toxizität), kann von einem Wirkstoffkandidaten gesprochen werden, der dann in biologischen Systemen getestet wird.

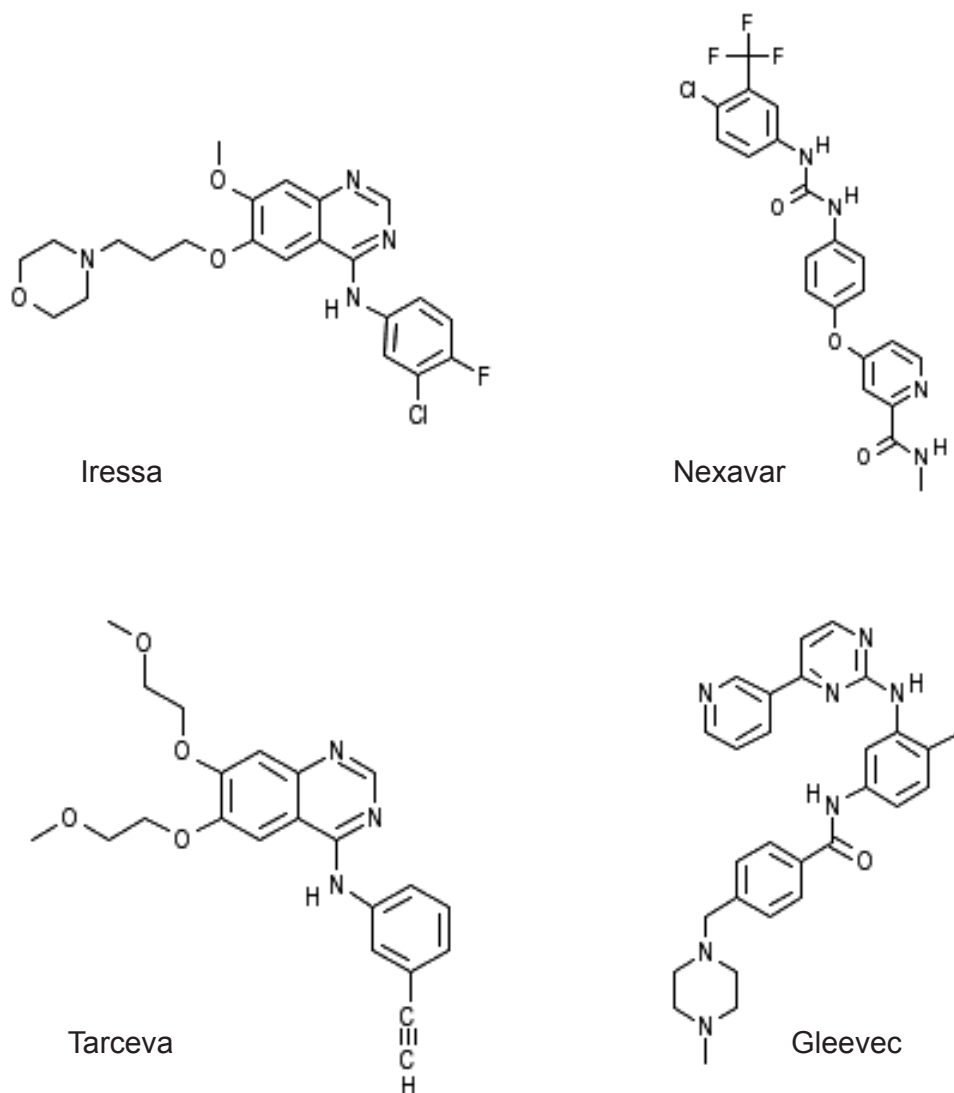


Abb. 1.4 Als Arzneimittel zugelassene Kinaseinhibitoren, die u.a. durch den **Target-Based Approach** entwickelt wurden. Insbesondere bei der Entwicklung von Gleevec wurden Methoden des *rational Drug-Design* angewendet, um die Selektivität zu verbessern [BUIJSMAN, 2004].

1.3.1 Computerunterstütztes rationales Wirkstoffdesign

Heute werden *Drug-Target* Proteine häufig mit Hilfe der Bioinformatik auffindig gemacht. Die Bioinformatik vereint die Speicherung, Organisation und Analyse von biologischen Daten [SELZER *et al.*, 2004]. Eine Möglichkeit der Identifizierung solcher Proteine ist die Sequenzsuche in öffentlichen Sequenzdatenbanken wie die des *National Center for Biological Information (NCBI)* oder des *Sanger Institute* [<http://www.ncbi.nlm.nih.gov/>; <http://www.sanger.ac.uk/>]. Dabei wird die Funktion eines Proteins anhand der Ähnlichkeit seiner Aminosäuresequenz mit den Sequenzen von Proteinen, deren Funktionen bereits bekannt sind, vorhergesagt. Über die Sequenzähnlichkeiten können Proteine in Familien mit ähnlicher Funktion eingeteilt werden [SELZER *et al.*, 2004]. Ein Beispiel hierfür sind die Proteasen, die die Spaltung von Peptidbindungen katalysieren. In Abb. 1.5 ist ein Sequenz-*Alignment* von Cathepsin B, einer Cysteinprotease des Menschen, und den homologen Proteinen der Maus und des Parasiten *T. gondii*, dargestellt. In allen Sequenzen liegen die Aminosäuren, die die makroskopische Proteinstruktur bestimmen, konserviert vor. Daraus lässt sich ableiten, daß die Proteine ähnliche Funktionen im jeweiligen Organismus haben müssen.

Im „*in silico*“ *Drug-Design* (computerunterstütztes Wirkstoffdesign) werden computerbasierte Methoden eingesetzt, um Lage und Affinität eines Liganden in einer Bindestelle abzuschätzen. Wichtige Methoden des computerunterstützten rationalen *Drug-Designs* sind u.a. das molekulare *Docking* in Proteinbindetaschen, die Homologiemodellierung von Proteinen und die Pharmakophormodellierung der Liganden eines *Drug-Target* Proteins. Aus den Strukturinformationen können dann Struktur- Aktivitätsbeziehungen abgeleitet werden.

Analog zum *High Throughput* Screening werden im virtuellen Screening Millionen von chemischen Substanzen auf die erforderlichen Eigenschaften hin untersucht [GOOD et al., 2000; KLEBE et al., 2000]. Zur effektiven Anwendung dieser Methode benötigt man eine entsprechende Computerinfrastruktur (*Compute-Cluster*), die es ermöglicht, große Rechenoperationen in einer vergleichsweise kurzen Zeit durchzuführen.

1.3.1.1 Homologiemodellierung

Die dreidimensionale Struktur von Proteinen kann durch Röntgenbeugungsmuster des kristallisierten Proteins dargestellt werden. So gewonnene Strukturinformationen sind in Proteinstrukturdatenbanken abgelegt. Die bekannteste Proteinstrukturdatenbank ist die *Protein Data Bank* (PDB) [BERMAN, 2002]. In dieser Datenbank sind derzeit etwa 33.500 Proteinstrukturen enthalten. Die Analyse so gewonnener Strukturdaten ist hilfreich, um beispielsweise die Funktionsweise eines Proteins und dessen Protein-Ligand Interaktionen zu verstehen. Insbesondere im strukturbasierten *Drug-Design* sind diese Informationen hilfreich, um die Bindungseigenschaften von Wirkstoffen an ein Protein zu verbessern oder um neue Liganden zu entwerfen. Häufig ist jedoch von einem *Drug-Target* Protein keine Struktur in einer Datenbank abgelegt, sodass man Methoden der Strukturvorhersage nutzen muss. Zu diesen Methoden gehört die Homologiemodellierung [SELZER et al., 2004]. Bei dieser Methode macht man sich die Abstammung der Proteine von gemeinsamen „Ursprungsmolekülen“ zunutze. So stammt die Proteinfamilie der Kinasen vermutlich von einem einzigen Protein ab, das im Laufe der Evolution entstanden ist. Diese Abstammung spiegelt sich in Gemeinsamkeiten in der Proteinsequenz, in ihrer Struktur und in der Funktion wieder. Da alle Kinasen die gleiche Reaktion katalysieren, sind die Faltung sowie die ATP-Bindestelle der Kinasen sehr stark konserviert. Diese Eigenschaften einer Proteinfamilie macht man sich bei der Homologiemodellierung zunutze, indem man eine Proteinstruktur verwendet, die experimentell bestimmt wurde.

```

          *           240           *           260           *
Cathepsin : ----FYNVDMSYLRKRLCGTFLG-----GPK---PPQRVMFTE
Toxopain  : VSLRFERYLSLKDAAKLMGTFLVNTKVEGFPTPKGMPLPAKEFENA

          280           *           300           *
Cathepsin : DLKLPASFDAREQWPQCP-TIKEIRDQGSCGSCWAFGAVEAISDR
Toxopain  : TEPVPAHFDARTAFPAACKDVVGHVRDQDCCGSCWAFASTEAFNDR

          320           *           340           *           360
Cathepsin : ICIHTNAHVSVEVSAEDLLTCC-GSMCG-DGCNGGYPAEAMNFWT
Toxopain  : LCI RSQGRMLMPLSAQHTTSCCNAIHCASF GCNNGQP GMAMRWFE

          *           380           *           400
Cathepsin : RKGLVSGGLYESH---VGC RPYSI P PCEHHVNGSRPPCTG---EG
Toxopain  : RKGVVTGGDFDALGKGTTCWPYEVFFCAHHAKAPFPDCDATLVPR

          *           420           *           440           *
Cathepsin : DTPKCSKICEPG-YSP---TYKQDKHYGNSYSVSNSEKDIMAEI
Toxopain  : KTPKCRKDCEEQAYADNVHPFDQDTHKATSAYSLSR-SRDDVVRDM

          460           *           480           *
Cathepsin : YKNGPVEGAFSVYSDFLLYKSGVYQHVTGEMMGGHAIIRILGWGVE
Toxopain  : MTHGPVSGAFMVYEDFLSYKSGVYKHVSGLPVGGHAIKIIGWGTE

          500           *           520           *           540
Cathepsin : NGTPYWLIVANSWNTDWDGNGFFKILRGQDHCGLIESEVVAGIIPRTD
Toxopain  : NGEELYWHAVNSWNTYWDGGQFKIAMG--QCGIDGEMVAGEAAMQ

          *           560           *
Cathepsin : QYWEKI----- : 339
Toxopain  : ETEGVVNGEEELPGQRAAGARAGAHAEEREM : 569

```

Abb. 1.5 Sequenz-Alignment von Toxopain und des humanen Cathepsin-B. Die Bereiche von Sequenzidentität sind gelb hinterlegt. Die für Cysteinproteasen charakteristischen Sequenzmotive sind analog zu Abb. 1.6 farbig hinterlegt.

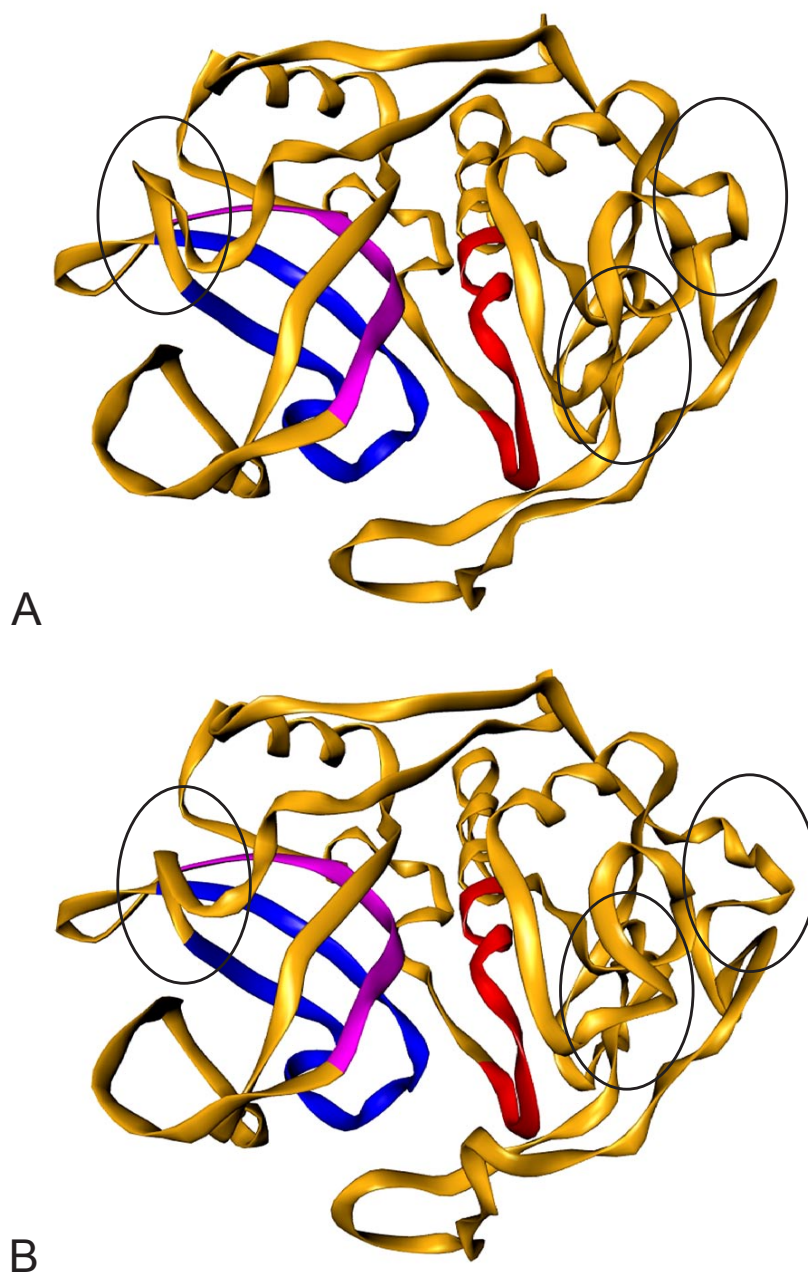


Abb. 1.6 Homologiemodellierung am Beispiel des Toxopains. Abb. 1.6 A ist eine röntgenkristallographisch erstellte Proteinstruktur des humanen Cathepsin B (PDB Acc. No.: 1CSB). Die Struktur wurde als Templat eines Homologiemodells von Toxopain aus *T. gondii* (Abb. 1.6 B) verwendet. Die für Cysteinproteasen charakteristischen Sequenzmotive sind farbig markiert. In Abb. 1.5 ist das dazugehörige Sequenz-*Alignment* dargestellt. Die konservierten Sequenzmotive sind sowohl in der Kristallstruktur als auch im Modell identisch lokalisiert. Die Unterschiede zwischen Templat und Modell liegen in verschiedenen Aminosäureseitenketten und in den Loop-Regionen (durch Kreise markiert). Das Modell wurde mit dem Homology-Modul mit der Software InsightII erstellt.

Nach einem Sequenz-*Alignment* der Sequenz, deren Proteinstruktur unbekannt ist, mit der zugehörigen Sequenz einer bekannten Struktur, können nun die Aminosäuren der unbekannt Sequenz auf die experimentell bekannte Struktur übertragen werden (Abb. 1.5). Auf diese Weise entsteht ein Proteinmodell der Sequenz, zu der keine experimentellen Strukturdaten vorliegen (Abb. 1.6). Diese Modelle sind um so exakter, je enger die Proteine miteinander verwandt sind [SELZER *et al.*, 2004].

1.3.1.2 Pharmakophormodellierung

Pharmakophore beschreiben die Eigenschaften von Molekülen, die für die Ligandbindung an ein *Drug-Target* Protein verantwortlich sind. Typischerweise handelt es sich um funktionelle Gruppen oder Strukturelemente. Häufig werden diese als dreidimensionale Modelle dargestellt, bei denen neben der Art der Interaktion auch die Lage zueinander beschrieben wird. Sie spiegeln im Idealfall die komplementären Gegebenheiten einer Bindetasche z.B. eines Proteins wieder [GOOD *et al.*, 2000]. Pharmakophormodelle lassen sich prinzipiell auf zwei Arten generieren: Eine Möglichkeit besteht darin, die bioaktive Konformation mehrerer Liganden, die in der selben Bindetasche binden, zu vergleichen und daraus Gemeinsamkeiten abzuleiten (qualitatives ligandbasiertes Pharmakophormodell).

Eine weitere Möglichkeit der Erzeugung solcher Modelle ist die Analyse einer Proteinbindetasche (proteinbasiertes Pharmakophormodell). Dabei werden die Eigenschaften der an einer Ligandbindung beteiligten Aminosäuren in einer dreidimensionalen Proteinstruktur betrachtet (Abb. 1.7). Diese kann experimentell ermittelt oder durch Homologiemodellierung generiert worden sein. Jede Aminosäure besitzt funktionelle Gruppen und atomare Eigenschaften, die dem zugehörigen Bindungsmuster eines Pharmakophors entsprechen. Ist in der Proteinbindetasche beispielsweise ein Wasserstoffbrückendonator vorhanden, so kann das Pharmakophormodell einen Wasserstoffbrückenakzeptor, der auf diesen Donor ausgerichtet ist, besitzen. Die-

ser Akzeptor wird im Modell soweit abstrahiert, dass er durch verschiedene Atomtypen realisiert werden kann. Während der Bindetaschenanalyse werden funktionelle Gruppen der Aminosäuren identifiziert, die für die Bindung eines Liganden essentiell sind. Daraus lässt sich dann ein Pharmakophormodell ableiten. Bindetaschenmodell und Pharmakophormodell verhalten sich dabei wie ein Negativabbild und Positivabbild der Bindetasche zueinander.

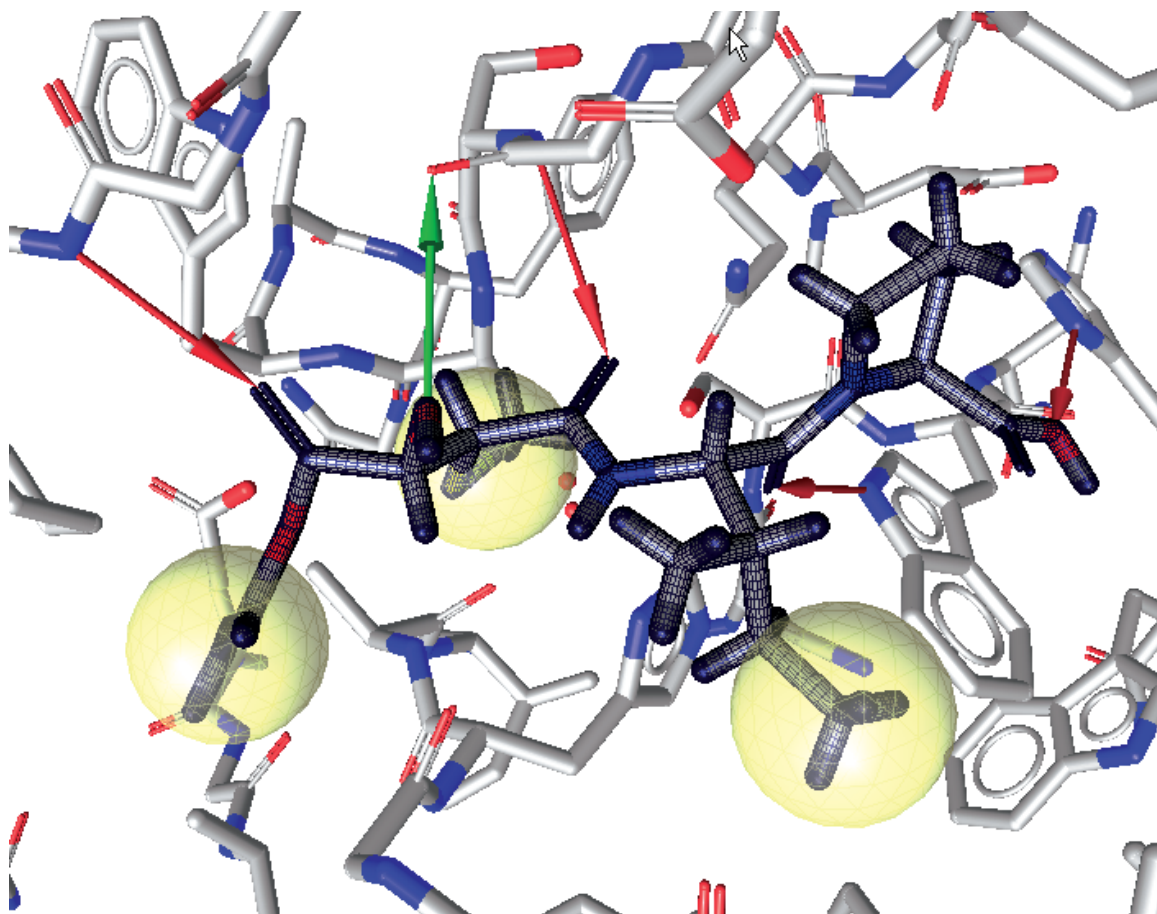


Abb. 1.7 Darstellung eines Pharmakophormodells basierend auf humanem **Cathepsin B** und dem Inhibitor **CA030**. Die zugrundeliegende Röntgenkristallstruktur 1CSB wurde aus der PDB entnommen. Der Inhibitor CA030 ist blau schraffiert dargestellt. Die gelben Kugeln sind lipophile Bereiche des Liganden. Die roten Pfeile markieren Wasserstoffbrücken, die auf einen Wasserstoffbrückenakzeptor zeigen. Die grünen Pfeile markieren Wasserstoffbrücken, die auf einen Wasserstoffbrückenakzeptor zeigen.

Durch die Kombinationen von ligandbasierten und proteinbasierten Pharmakophoren können einzelne Bindungsmodi eines Liganden sehr exakt reproduziert werden. Durch die Kombination mehrerer Bindungsmodi unterschiedlicher Liganden können darüberhinaus unbekannte Bindungsmodi abgeleitet werden.

1.3.1.3 Molekulares *Docking*

Molekulares *Docking* simuliert die Einpassung eines Liganden in die Bindetasche eines Makromoleküls [KLEBE et al., 2000]. Dieses kann ein Protein, aber beispielsweise auch eine Desoxyribonucleinsäure (DNS) sein.

Die Proteinbindetasche wird von den verschiedenen *Docking*-Programmen unterschiedlich beschrieben. Dabei kommen grundsätzlich zwei Verfahren zum Einsatz: Zum einen sind das die *Grid*-basierten Verfahren und zum anderen die Verfahren, die gerichtete Protein-Ligand Interaktionen betrachten. Bei den *Grid*-basierten Verfahren wird die Bindetasche in Gitterpunkte unterteilt (Abb. 1.8). Jeder dieser Gitterpunkte repräsentiert einen möglichen Interaktionspunkt. Je nach Entfernung zu einem Proteinatorm und dessen chemischen Eigenschaften, werden dem Gitterpunkt mögliche Interaktionsenergien und Interaktionstypen zugeordnet. Nun können Ligandatome auf diese Gitterpunkte gelegt werden und anhand des Atomtyps, welcher auf den Gitterpunkt projiziert wird, wird der Typ der Interaktion bestimmt. Anhand der zuvor berechneten Interaktionsenergie des Gitterpunktes kann nun die Interaktion des Ligandatoms mit dem Protein berechnet werden. Wird dies für die Gesamtheit der Atome eines Liganden durchgeführt, so lässt sich hieraus eine Aussage über die Intensität der Protein-Ligand Wechselwirkung ableiten [KITCHEN et al., 2004].

Einfachste *Docking* Methoden benutzen eine einzelne Ligandkonformation, die in eine Bindetasche eingepasst wird. Sehr viel aufwendiger ist die Berechnung der idealen Ligandkonformation in der Bindetasche. Sollen unterschiedliche Proteinkonformationen oder *induced-fit* Effekte des Proteins

berücksichtigt werden, ist die zuverlässige Berechnung mit den gängigen Methoden sehr schwierig und aufgrund der Komplexität der Modelle sehr rechenintensiv [KITCHEN *et al.*, 2004].

Für die Ermittlung einer idealen Ligandkonformation werden unterschiedliche Algorithmen verwendet. Grundsätzlich lassen sich hierfür drei Methoden unterscheiden: 1. fragmentbasierte Methoden, 2. zufallsbasierte Suche und 3. Simulationsmethoden.

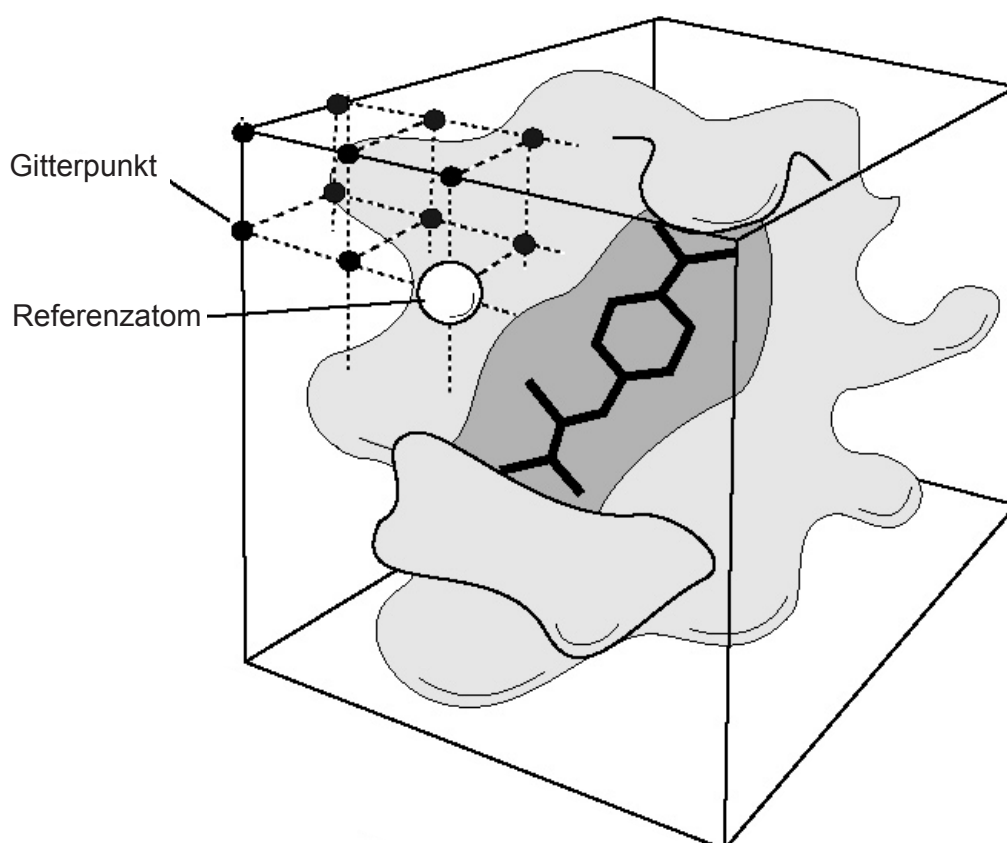


Abb. 1.8 Grid-basierte Betrachtung der Bindetasche eines Proteins. Die Bindetasche wird durch die Zerlegung in Gitterpunkte vereinfacht dargestellt. Ligandatome werden auf die einzelnen Gitterpunkte projiziert (Referenzatom). Jedem Gitterpunkt sind Interaktionsenergien zugeordnet. Eine Bewertungsfunktion berechnet anhand der Einzelbeiträge der Energien der Gitterpunkte eine Gesamtenergie der Interaktion [aus dem Autodock User's Guide, [HTTP://WWW.SCRIPPS.EDU/MB/OLSON/DOC/AUTODOCK/AD305/USING_AUTO DOCK_305.HTML](http://www.scripps.edu/mb/olson/doc/autodock/ad305/using_autodock_305.html)].

Die fragmentbasierte Methode wird von den weit verbreiteten Programmen FlexX und DOCK verwendet [<http://www.biosolveit.de>; <http://dock.compbio.ucsf.edu/>]. Dabei wird der Ligand in Fragmente zerlegt und ein Kernfragment in der Bindetasche plaziert. Ausgehend von diesem Fragment werden nun iterativ weitere Fragmente hinzugefügt, bis eine endgültige Konformation errechnet wurde.

Bei zufallsgesteuerten Methoden, die häufig auch als stochastische Methoden bezeichnet werden, werden zufällig Änderungen an einzelnen oder an Populationen von Ligandkonformationen vorgenommen und mit Hilfe einer Bewertungsfunktion optimiert. Häufig dienen zur Optimierung genetische Algorithmen, wie sie zum Beispiel von den *Docking*-Programmen Gold und Autodock verwendet werden [STAHL, 2000]. Dabei werden zunächst einige Startkonformationen eines Liganden generiert und in die Bindetasche eingepasst. Von den günstigsten Startkonformationen ausgehend, werden weitere, ähnliche Konformere generiert. In einem genetischen Algorithmus kommen die Prinzipien der Mutation, Selektion und Vererbung zum Einsatz. Übertragen auf ein kleines Molekül bedeutet das, dass jedem Bindungswinkel einer rotierbaren Bindung ein „Gen“ zugeordnet ist. Die „Gene“ des Moleküls sind auf ein „Chromosom“ verteilt. In der Population der Molekülkonformationen „überleben“ jeweils nur die Moleküle mit der günstigsten Konformation - also einem für die „Umwelt“ der Bindetasche optimierten Chromosom. Chromosomen, die für günstige Konformationen kodieren, werden mittels Mutation oder eines „Crossing Over“ mit einem anderen Chromosom weiter verändert. Dadurch werden die jeweils günstigsten Konformationen bevorzugt, was am Ende der Simulation zu einer optimierten Ligandstruktur führt [MORRIS *et al.*, 1998].

Die Simulationsmethoden hingegen bedienen sich der Methoden der Moleküldynamik, wobei die Newton'sche Bewegungsgleichung auf alle Atome des Moleküls angewendet wird. Dabei wird das Molekülverhalten unter simulierten Umweltbedingungen, wie z.B. der Umgebungstemperatur, berechnet [BOEHM & STAHL., 2002].

1.3.1.4 Quantifizierung von Protein-Ligand Wechselwirkungen

Die Intensität einer Wechselwirkung wird im *Docking* durch Bewertungsfunktionen (*Scoring Functions*) quantifiziert. Bewertungsfunktionen kommen dabei für zwei unterschiedliche Zwecke zum Einsatz. Zum einen dienen sie während des *Dockings* als Fitnessfunktion, wobei der *Fitness Score* darüber entscheidet, ob eine Ligandkonformation gut an ein Protein bindet, oder ob eine Bindung unwahrscheinlich ist. Zum anderen dienen die Bewertungsfunktionen nach dem *Docking* von unterschiedlichen Liganden zur Vorhersage der Bindungsaffinitäten der unterschiedlichen Liganden. Im Idealfall würde ein *Score* mit der Bindungsaffinität eines Liganden korrelieren. Da die Bindung eines Liganden an ein Protein ein sehr schwierig zu beschreibender physikalischer Prozess ist, beinhalten die meisten Bewertungsfunktionen Terme, die die Bindung stark vereinfacht betrachten. Dabei wird zwischen verschiedenen Bindungsarten unterschieden: Typische Interaktionen in Protein-Ligand Komplexen sind die Wasserstoffbrücken, ionische Interaktionen, van der Waals Interaktionen, Kation- π Wechselwirkungen, Metallkomplexierung und der entropische Energiegewinn durch den hydrophoben Effekt. Grundsätzlich wird zwischen kraftfeldbasierten, empirischen und wissensbasierten Verfahren unterschieden [BOEHM & STAHL, 2002]. Bei den kraftfeldbasierten Verfahren kommen meistens Terme für van der Waals und elektrostatische Interaktionen zum Einsatz. Diese werden anhand von Lennard-Jones Potentialfunktionen und Abwandlungen dieser Funktionen beschrieben [BOEHM & STAHL, 2002]. Die empirischen Bewertungsfunktionen beinhalten Terme, die voneinander unabhängige Bindungseigenschaften beschreiben. Diese Terme werden mittels Regressionsanalyse solange miteinander kombiniert, bis eine Korrelation zwischen gemessener Aktivität und dem resultierenden *Score* zustande kommt. Wissensbasierte Bewertungsfunktionen wurden entwickelt, um experimentell ermittelte Ligand-Proteinkomplexe zu reproduzieren. Die wissensbasierten Bewertungsfunktionen beschreiben einfache Interaktionstypen, wobei Atompaarinteraktionen in Abhängigkeit von

deren molekularer Umwelt fest definiert sind. Durch Summation dieser Interaktionswerte ergibt sich dabei der Scoringwert. Allen Bewertungsfunktionen gemeinsam ist die Vernachlässigung der Entropie als Einzelgröße [KITCHEN *et al.*, 2004; STAHL, 2000].

1.4 Computerinfrastrukturen für virtuelles Screening

Das *Docking* eines Liganden in seine Bindetasche ist ein sehr rechenintensiver Prozess [SCHULZ-GASCH & STAHL, 2003]. Deshalb werden im virtuellen Screening von großen Substanzdatenbanken in der Regel viele Computer eingesetzt, die in einem Verbund zusammengeschlossen sind. Solche Computerverbände werden entweder als *Cluster* oder als *Compute-Grid* bezeichnet. Dabei gibt es technisch unterschiedliche Möglichkeiten, um erforderliche Rechenkapazitäten zu bündeln.

1.4.1 Grid-Computing

Eine Möglichkeit, die Computer zu einem Verbundsystem zusammenzuschließen sind die *Linux Cluster* [BERNACKI *et al.*, 2005]. Eine andere Möglichkeit besteht darin, bereits anderweitig genutzte Computer zu einem *Compute-Grid* zusammenzuschließen [CHIEN *et al.*, 2002]. Durch den Zusammenschluss vieler Computer in einem Netzwerk ist es dadurch möglich, sehr hohe Rechenkapazitäten für eine Aufgabe zu bündeln. Dabei können Computer, die für alltägliche Büroarbeiten verwendet werden, zusätzlich für das virtuelle Screening genutzt werden (Abb. 1.9). Die durchschnittliche Benutzung eines Büro-Computer erstreckt sich auf ca. 2-5 % eines Arbeitstages. Die restliche Zeit wird der Computer nicht benutzt, d.h. er arbeitet im „Leerlauf“. Diese Zeit kann für Berechnungen mittels *Grid-Computing* genutzt werden, um z.B. virtuelles Screening durchzuführen [CHIEN *et al.*, 2002]. Der Anwender merkt von diesen Berechnungen in der Regel nichts, da die *Grid-Computing Software* automatisch die Belastung herunterregelt,

sobald der Benutzer den Computer für Büroaufgaben benötigt. Einige große *Grid-Computing* Projekte bedienen sich der Ressourcen sehr vieler Computer die im Internet miteinander vernetzt sind. Das wohl bekannteste Projekt ist das *Search for Extraterrestrial Intelligence at Home* Projekt, das zu den Pionierprojekten des *Grid-Computing* gehört [<http://setiathome.ssl.berkeley.edu>]. Einige biomedizinische Anwendungen des verteilten Rechnens über das Internet umfassen u.a. die Projekte *FightAids@Home* und *Folding@Home*, wobei molekulares *Docking* und *ab initio* Berechnungen von Proteinstrukturen zum Einsatz kommen [<http://fightaidsathome.scripps.edu>; <http://folding.stanford.edu>].

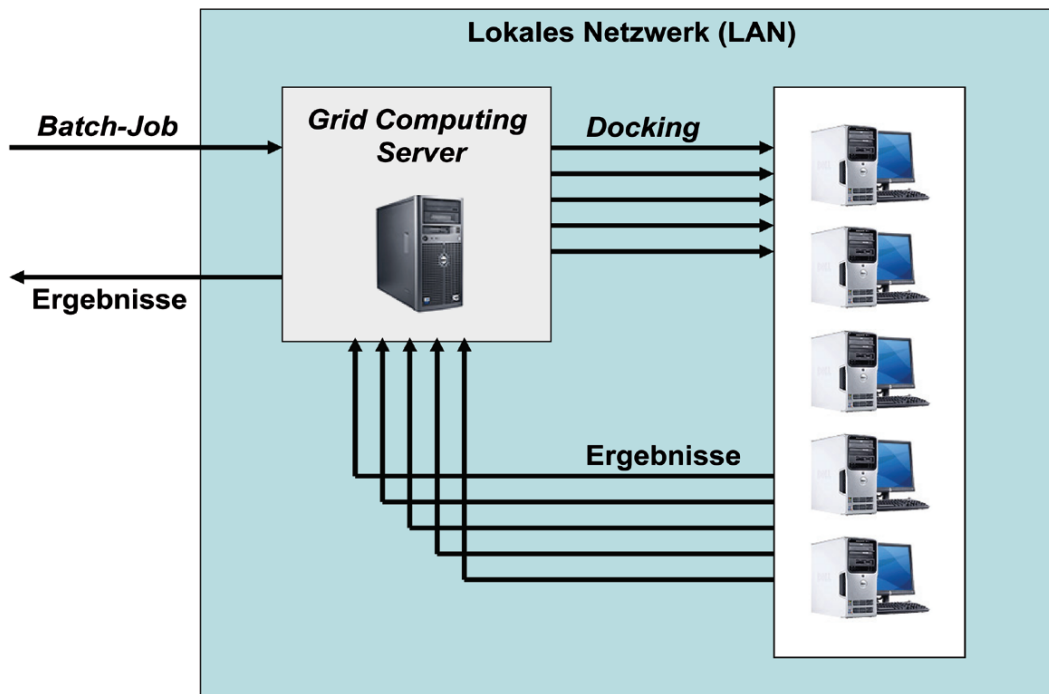


Abb. 1.9 Schema einer Grid Computing Umgebung. Eine computerintensive Fragestellung wird an den *Grid-Computing Server* zur Berechnung weitergegeben. Der Server verteilt die einzelnen Aufgaben an die einzelnen *Computer* zur Berechnung. Die Ergebnisse der Berechnungen werden auf dem *Server* gesammelt und nach Beendigung an den Benutzer weiter gegeben

1.4.1.1 *Grid-Computing Software*

Eine *Grid-Computing Software* soll den kompletten Ablauf der Bearbeitung einer Aufgabe auf vielen Computern organisieren. Dafür muss sie sowohl die Funktion eines *Queing-Systems* übernehmen, d.h. die sequenzielle Bearbeitung mehrerer Aufgaben steuern, als auch das *Cluster-Management* übernehmen, indem freie Ressourcen erkannt und zugewiesen werden. So muss der *Software* bekannt sein, ob beispielsweise ein Computer gerade in Benutzung durch einen Anwender ist oder ob eine Netzwerkverbindung unterbrochen ist. Die Aufgaben miteinander konkurrierender Prozesse werden automatisch auf die Computer verteilt. Gleichzeitig werden von der *Software* die für eine bestimmte Aufgabe benötigten Daten verteilt und die Ergebnisse der Rechnungen zusammengeführt. Darüber hinaus muss die *Software* Sicherheitsstandards in Netzwerken, wie z.B. ein *User-Management*, d.h. die Identifikation des Benutzers und die Bearbeitung von Rechenoperationen in dessen Benutzerkontext, berücksichtigen. Alle diese Aufgaben erfordern eine sehr genaue Einhaltung der beschriebenen Abläufe die nur durch spezielle *Software* gewährleistet wird. Beispiele einer *Grid-Computing Software* sind GridMP und Enfuzion 8.2 [<http://www.uniteddevices.com/>; <http://www.axceleon.com>].

1.5 Proteinkinasen als *Drug-Targets* des rationalen *Drug-Design*

Die Proteinkinasen stellen eine der größten Proteinfamilien in der belebten Natur dar [SHEINERMAN *et al.* 2005]. Das humane „Kinom“, das die Gesamtheit aller Proteinkinasen des Menschen beinhaltet, besteht aus 518 Vertretern (Abb. 1.10) [MANNING *et al.*, 2002]. Proteinkinasen spielen in vielen Signalkaskaden innerhalb der Zelle eine bedeutende Rolle als Signalvermittler. Proteinkinasen übertragen dabei die γ -Phosphatgruppe eines Adenosintriphosphatmoleküls (ATP) auf eine Hydroxygruppe eines Peptid- oder Proteinsubstrats. Meist werden die Substrate durch die Phosphorylierung aktiviert

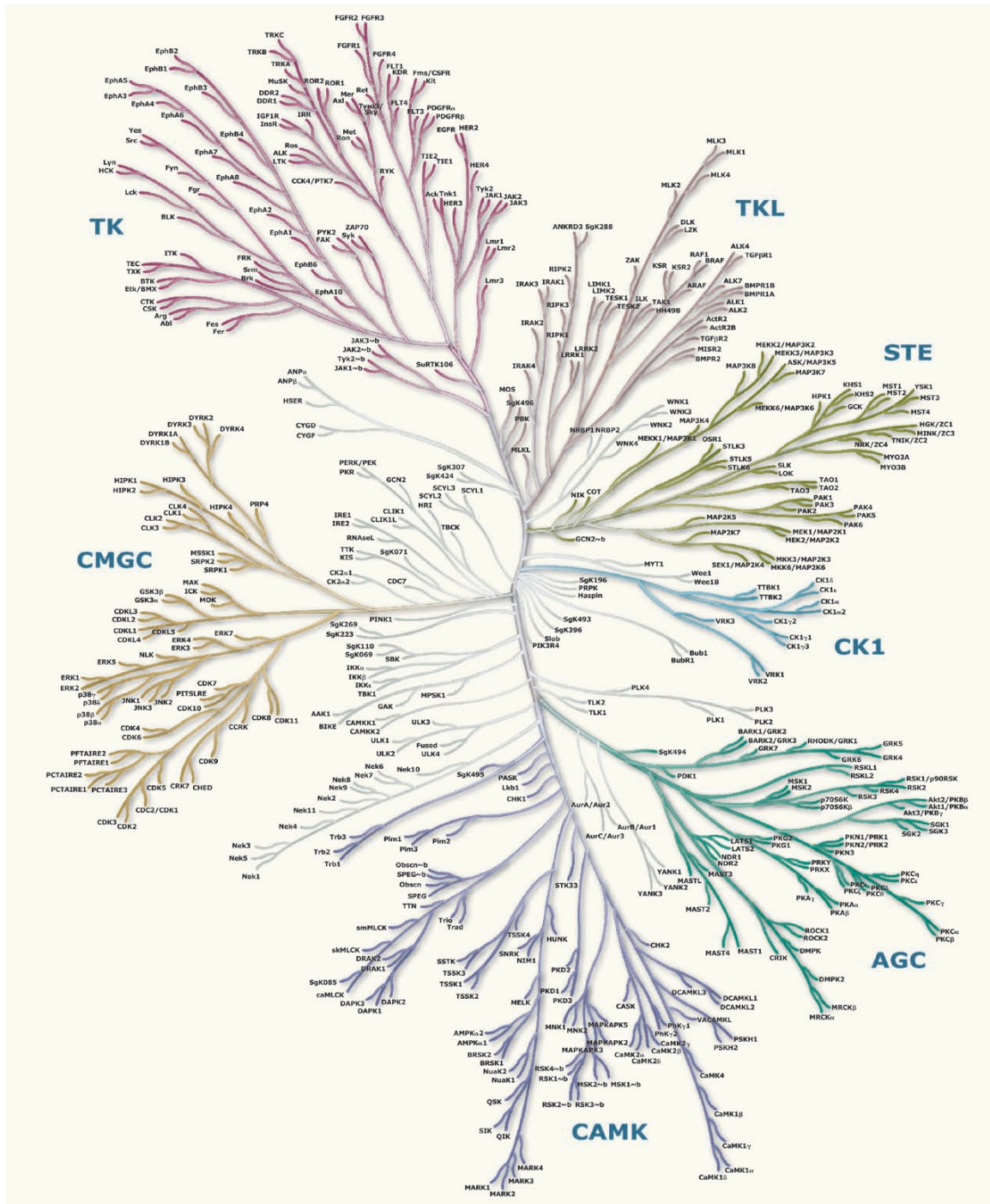


Abb. 1.10 Dendrogramm der Proteinkinasen des Menschen. Die baumartige Struktur des Dendrogramms zeigt die Sequenzähnlichkeit der Proteinkinasen des Menschen. Anhand der Sequenzähnlichkeit können die Proteinkinasen in Familien eingeteilt werden. Bspw. enthält die CMGC Familie die CDKs, MAP Kinasen, GSK3 und CLKs. [nach Manning et al., 2002]

oder inaktiviert. Durch die Aktivierung anderer Proteinkinasen entstehen Signalkaskaden, die eine sehr genaue Regulation zellulärer Prozesse ermöglichen. Proteinkinasen sind bei zellulären Prozessen wie dem Metabolismus, der Transkription und der Zellzykluskontrolle als Regulatoren von entscheidender Bedeutung [Bradham & McClay, 2006]. Störungen dieser Regulationsmechanismen sind die Ursachen einer Vielzahl von Krankheiten, wie z.B. Diabetes, Krebs und entzündliche Erkrankungen. So sind beispielsweise 164 der humanen Proteinkinasen mit verschiedenen Tumoren genetisch assoziiert [MANNING *et al.*, 2002]. Aus diesen Gründen wurde in den vergangenen Jahren die Forschung an Mitgliedern dieser Proteinfamilie stark intensiviert. In Abbildung 1.11 sind die wichtigsten Meilensteine der Proteinkinase-Forschung dargestellt. Die Methoden des rationalen *Drug-Design* ermöglichten die Erforschung von selektiven Proteinkinase-Inhibitoren.

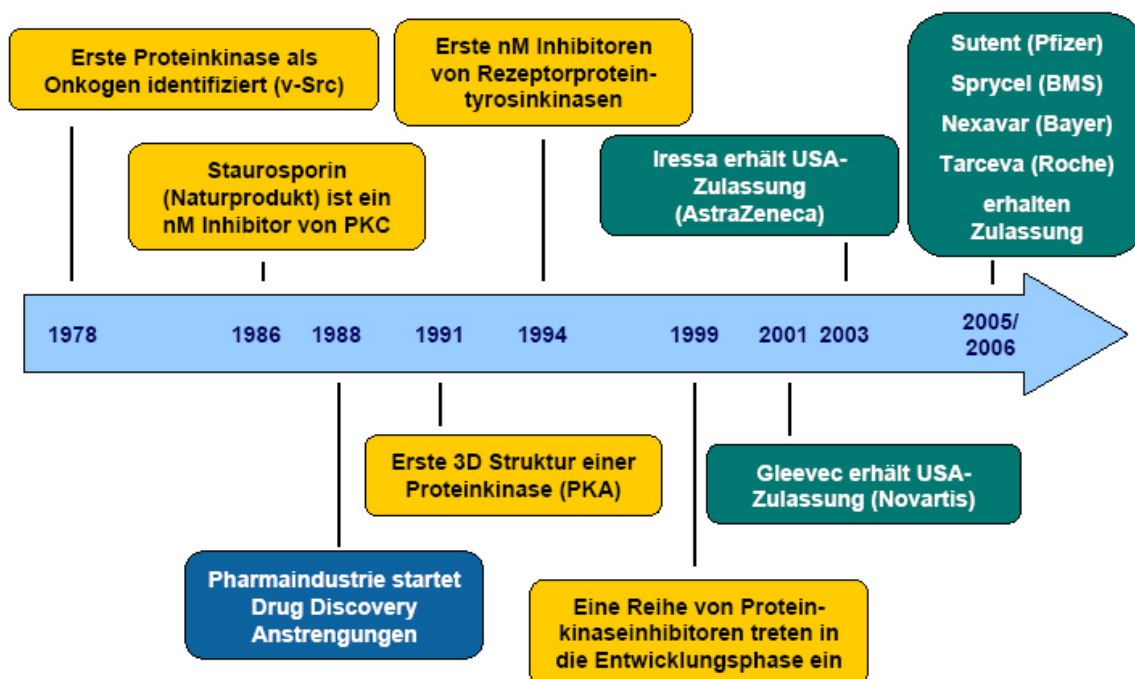


Abb. 1.11 Meilensteine der Proteinkinase-Forschung. Die Forschungsaktivitäten auf dem Gebiet der Proteinkinase-Forschung reichen bis in das Jahr 1978 zurück. Mittlerweile sind einige Kinaseinhibitoren als Arzneimittel zugelassen.

Insbesondere in der Onkologie bestehen große Erwartungen hinsichtlich der Therapiebarkeit von Tumoren durch neuartige Proteinkinase-Inhibitoren.

Die Proteinkinasen bilden in eukaryontischen Zellen die zweithäufigste Proteinklasse. Sie werden häufig nach der Art der modifizierten Gruppe (Serin, Threonin, Tyrosin), oder der phylogenetischen Gruppenzuordnung eingeteilt [MANNING *et al.*, 2002]. Die Phosphorylierung der Substrate der Proteinkinasen kann durch spezialisierte Protein-Phosphatasen wieder rückgängig gemacht werden. Häufig entstehen dadurch Regelkreise, die in verschiedenen Signalkaskaden vorkommen.

1.5.1 Der eukaryontische Zellzyklus

Der eukaryontische Zellzyklus beinhaltet eine Sammlung von hochgeordneten, gleich ablaufenden Veränderungen, die als Übergänge, Phasen oder als Stadien bezeichnet werden (Abb. 1.12). Hierzu zählen die Phasen G_1 (Gap 1), S (DNS Synthese), G_2 (Gap 2) und M (Mitose) [ELLEDGE, 1996; JOHNSON & WALKER, 1999]. Ruhende, d.h. nicht-teilungsaktive Zellen, befinden sich in der Ruhephase (G_0 -Phase). Beginnt ein neuer Zellzyklus erreicht die Zelle die erste Gap-Phase (G_1 -Phase). Während der darauffolgenden Synthesephase (S-Phase) wird das Genom der Zelle repliziert. Eine weitere Gap-Phase (G_2 -Phase) unterbricht den Zyklus bis zur darauffolgenden Mitose-Phase (M-Phase). Das Genom wird in der M-Phase auf die zwei neu entstehenden Tochterzellen verteilt. Während dieser Phase findet auch die eigentliche Zellteilung statt. Sobald die Zellen ihre Vermehrung unterbrochen haben, infolge von spezifischen, antimitogenen Signalen oder aufgrund des Fehlens von angemessenen mitogenen Signalen, beispielsweise durch Hormonsignale, verlassen diese Zellen den Zyklus und treten erneut in die Ruhephase (G_0) ein [MALUMBRES & BARBACID, 2005].

1.5.1.1 Die Funktion der Cyclin-abhängigen Kinasen im Zellzyklus

Die genaue Abfolge der verschiedenen Phasen des Zellzyklus ist von entscheidender Bedeutung für das Schicksal der Zelle, da Störungen zu einer unvollständigen DNS-Replikation und somit zu einer Veränderung des Genoms führen können. Um den reibungslosen Ablauf des Zellzyklus zu gewährleisten, haben die Zellen Kontrollpunkte entwickelt [MALUMBRES & BARBACID, 2005]. Die wichtigsten Regulatoren sind hierbei die Cyclin-abhängigen Kinasen (engl.: *cyclin-dependent kinase* = CDK). Die CDK gehören zur Gruppe der Serin-Threonin Proteinkinasen und haben ein Molekulargewicht von etwa 33 kDa. Im humanen Kinom sind 11 Gene von CDKs bekannt [MALUMBRES & BARBACID, 2005; MANNING *et al.*, 2002]. Zusätzlich existieren 9 weitere Gene mit einer Ähnlichkeit zu den CDK. Sie bilden eine Gruppe von homologen, hochkonservierten Proteinen, die in allen Eukaryonten zu finden sind. Die CDKs sind maßgeblich an der Regulation des Zellzyklus beteiligt.

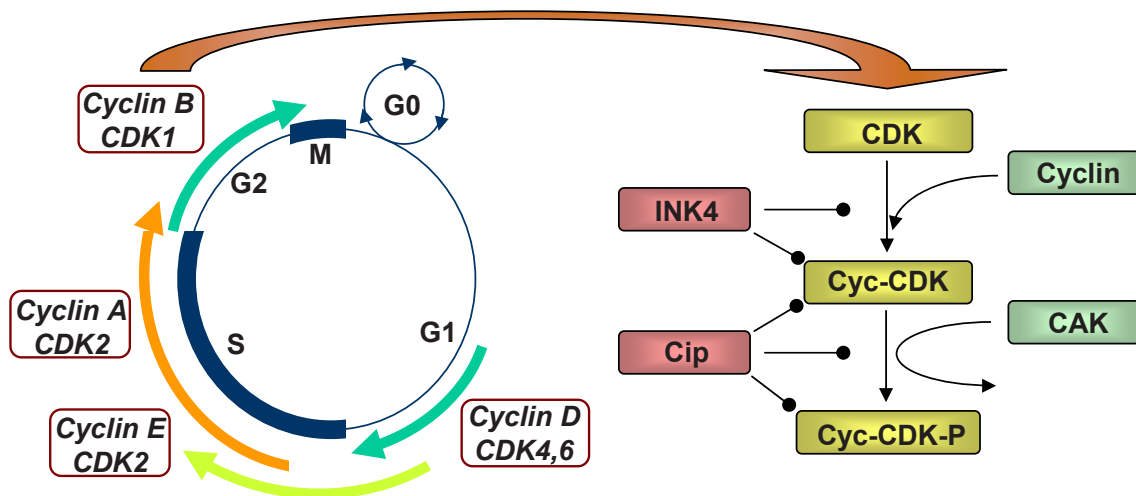


Abb. 1.12 Der Zellzyklus des Menschen. Die verschiedenen Phasen sind im inneren Kreis dargestellt. Dabei symbolisiert die G_0 -Phase, die Phase, in welcher eine Zelle ruht. Die farbigen Pfeile entlang des Zyklus symbolisieren die Aktivitätsfenster der jeweiligen CDK-Cyclin Heterodimere. Rechts neben dem Zyklus ist die Aktivierung und Inaktivierung durch Cycline, Cyclin-aktivierende Kinase (CAK) und durch endogene Inhibitoren schematisch dargestellt [nach BRANDON & TOOZE, 1999].

Sie agieren als Schalter, die den Übergang von der G₁- zu der S-Phase und der G₂ zu der M-Phase, sowie alle anderen Schritte innerhalb des Zellzyklus steuern (Abb. 1.12). Sie bilden Heterodimere mit den Cyclinen, wodurch sie allosterisch aktiviert werden. Die so aktivierten Proteinkinasen phosphorylieren u.a. Transkriptionsrepressoren, die dadurch inaktiviert werden. Die Phosphorylierung durch die CDKs erfolgt mittels des Phosphatdonors ATP und findet an Serin- und Threoninresten statt. In der Folge wird die Expression von Genen, die die nächste Phase des Zellzyklus steuern, möglich. Die CDK werden durch verschiedene Faktoren kontrolliert: zum einen durch die Bindung der Cycline, zum anderen aber auch durch inhibitorische Peptide der INK4-, CIP- und KIP- Familien (Abb. 1.12). Die Regulation der Aktivierung erfolgt durch die Synthese von Cyclinen und deren ubiquitinsteuerten Abbau am Proteasom sowie der Phosphorylierung und Dephosphorylierung bestimmter Aminosäuren der CDK. Auf diese Art entsteht ein Regelkreis, der die spezifischen Phasen des Zellzyklus kontrolliert [NOBLE *et al.*, 2004; ENDICOTT *et al.*, 1999; MORGAN, 1997]. Wird dieser Kreislauf z.B. durch spezifische CDK-Inhibitoren unterbrochen, ist die Zellvermehrung gestoppt.

1.5.1.2 Aufbau und Aktivierung der CDKs

In den folgenden Kapiteln wird die Benennung der Aminosäuren analog zur humanen CDK2 verwendet. Abweichende Notationen sind im Text entsprechend ausgewiesen.

Die CDKs sind Proteine mit Proteinkinaseaktivität. Die Faltungstopologie der Proteinkinasen ist stark konserviert und sie zeigen eine hohe Sequenzähnlichkeit untereinander [BUIJSMAN, 2004; BRANDON & TOOZE, 1999]. Die erste Charakterisierung der Struktur einer Proteinkinase wurde von der cAMP-abhängigen Proteinkinase PKA durchgeführt [KNIGHTON *et al.*, 1991].

Mittlerweile sind detaillierte Strukturinformationen für den CDK2-Cyclin A Komplex bekannt [JEFFREY *et al.*, 1995]. Der strukturelle Aufbau der CDK2 zeigt dasselbe Grundgerüst, wie bei anderen eukaryontischen Proteinkinasen.

Die CDK2 Struktur besteht aus einer kleineren N-terminalen Domäne von 85 Aminosäuren, die überwiegend aus 5 β -Faltblatt-Strukturen aufgebaut ist, sowie einer größeren C-terminalen Domäne (Abb. 1.13). Diese umfasst 213 Aminosäuren und besitzt einen hohen α -helikalen Anteil [BUIJSMAN, 2004; MORGAN, 1997; PAVLETICH, 1999]. Charakteristisch für CDKs ist ein konserviertes Sequenzmotiv (PSTAIRE-Box), das zur einzigen α -Helix (C-Helix) der N-terminalen Domäne gehört und entscheidend für die Cyclinbindung ist. Beide Domänen sind über die *Hinge*-Region miteinander verbunden und zwischen diesen liegt eine tiefe Spalte. Dort befindet sich die ATP-Bindetasche. Diese wird von zwei Schleifen bedeckt, dem glycinreichen *Loop* (*G-Loop*) und dem *T-Loop*. Dieser bedeckt ebenso die Substratbindetasche, die dadurch in der inaktiven Form blockiert wird. Im *T-Loop* liegt die Aminosäure Thr160, deren Phosphorylierung zu einer drastischen Aktivitätssteigerung der CDK2 führt [BUIJSMAN, 2004; JEFFREY *et al.*, 1995; MORGAN, 1997]. Die CDKs werden durch die Bindung eines Cyclinmoleküls aktiviert (Abb. 1.15). Von den Cyclinen sind im humanen Genom bislang 29 unterschiedliche Typen identifiziert worden, die alle eine konservierte Domäne von 150 Aminosäuren besitzen. Diese Domäne bezeichnet man als Cyclin-Box. Sie beinhaltet fünf helikale Regionen und ist an der Bindung an die CDK beteiligt [NOBLE *et al.*, 1997]. Jede CDK bildet mit bestimmten Cyclinen Heterodimere. CDK1 bindet bevorzugt die Cycline A und B, wogegen CDK2 A- und E-Cycline bevorzugt [MALUMBRES & BARBACID, 2005].

Durch die unterschiedlichen Cyclin-CDK Heterodimere, werden schließlich unterschiedliche Substrate phosphoryliert. Zusätzlich werden in den verschiedenen Phasen des Zellzyklus unterschiedliche CDKs aktiviert, bzw. deaktiviert [MALUMBRES & BARBACID, 2005]. Eine Deaktivierung erfolgt in der Regel an einem konservierten Threonin- oder Tyrosinrest, in CDK2 sind dies die Aminosäuren Thr14 und Tyr15. Die deaktivierende Phosphorylierung ist unabdingbar für einen normalen Zellzyklus und tritt häufig an Zellzyklus-Kontrollpunkten (*Checkpoints*) auf. Ein solcher *Checkpoint* ist beispielsweise der *DNA-Damage Checkpoint* am Ende der G_2 -Phase des Zellzyklus. Die

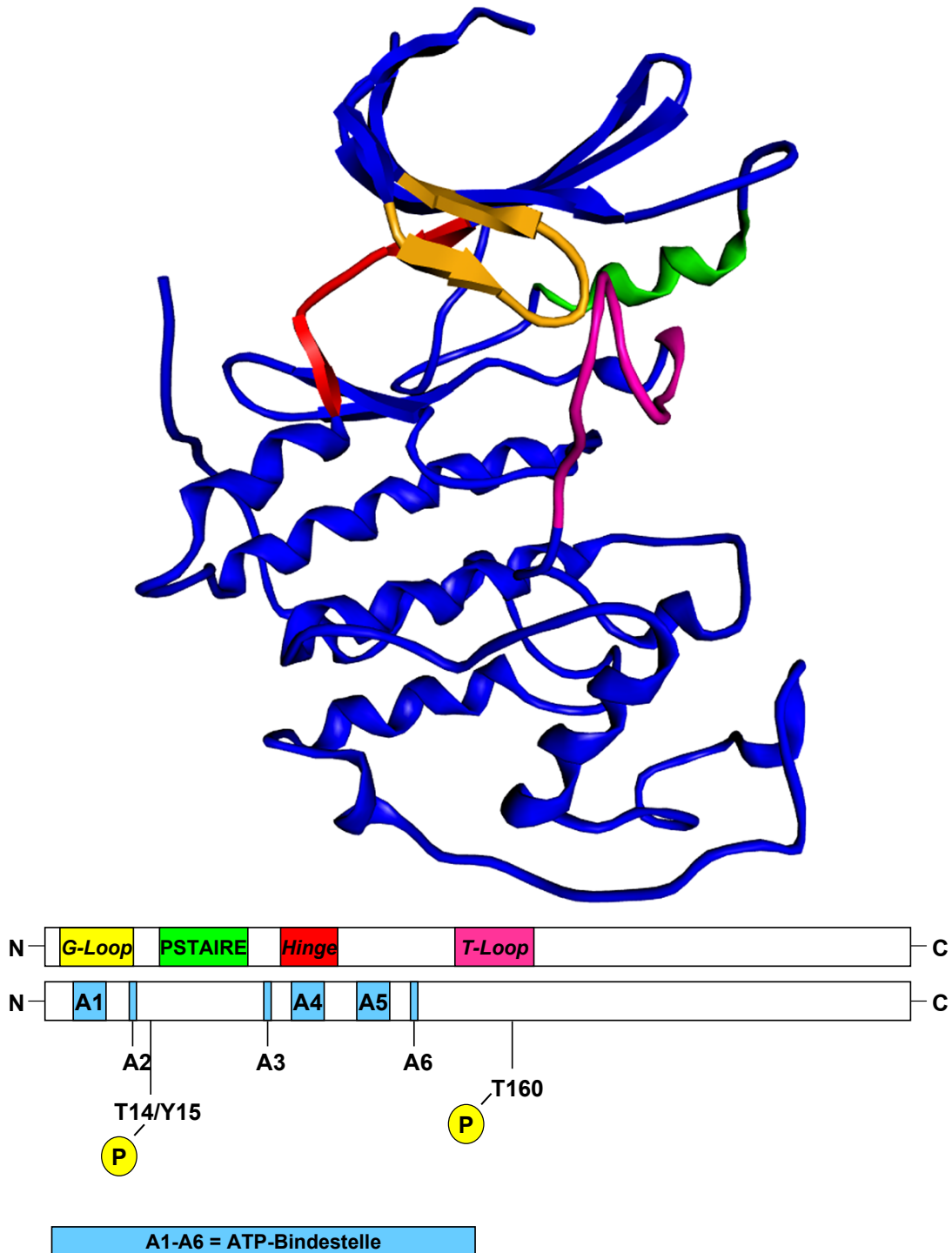


Abb. 1.13 Tertiär- und Sekundärstruktur-Merkmale einer CDK. Die *Hinge*-Region, welche die N-terminale mit der C-terminalen Domäne verbindet, ist in Rot dargestellt. Die C-Helix, an die die Cycline binden, ist grün eingefärbt. Der Aktivierungs-Loop mit dem Thr160 ist in Magenta dargestellt. Der glycinreiche Loop (G-Loop) ist gelb eingefärbt.

deaktivierende Phosphorylierung der CDK1 wird dabei durch die Proteinkinasen WEE1 und MYT1 bewirkt. Die Aktivierung der CDK1 erfordert eine Dephosphorylierung der beteiligten Aminosäuren. Dies geschieht durch die Phosphatase CDC25, von der drei Isoformen in Säugerzellen vorhanden sind. Der G_2 *DNS Damage Checkpoint* arretiert den Zellzyklus überwiegend durch die deaktivierende Phosphorylierung. Bei einer Beschädigung der DNS werden durch die Proteinkinasen ATM bzw. ATR die Proteinkinasen CHK1 bzw. CHK2 aktiviert. Diese phosphorylieren die CDC25C an der Aminosäure Ser216. Dadurch wird die Phosphatase inaktiviert, weshalb die CDK1 im inaktiven Zustand verbleibt. Bei der CDK2 ist die Situation nicht so eindeutig. Dort bestehen mehrere, parallel auftretende Wege der Deaktivierung, z.B. durch den Inhibitor p21^{CIP1/WAF1}. Neuere Untersuchungen haben ergeben, dass eine deaktivierende Phosphorylierung bei der CDK1 essentiell für einen normalen Zellzyklus ist, jedoch bei der Regulation der CDK2 eine untergeordnete Rolle spielt [CHOW *et al.* 2003].

Ein weiterer wichtiger Schritt in der Aktivierung einer CDK ist die Phosphorylierung des T-Loops, die erst nach der Bindung des Cyclins möglich wird (Abb. 1.14). Dieser *Loop* besitzt typischerweise ein Threonin, das an seiner Hydroxylgruppe phosphoryliert wird [JOHNSON *et al.*, 1996]. Die Bindung des Cyclins an die CDK bewirkt starke Veränderungen der Tertiärstruktur des Proteins und ist entscheidend für die Substratspezifität (Abb. 1.15). Diese Veränderungen ermöglichen sowohl dem ATP, als auch dem Substrat den freien Zugang zur katalytischen Spalte [ELLEDDGE, 1996; LE ROCH *et al.*, 2000; MORGAN, 1997]. Die Aktivierung durch das Cyclin bewirkt außerdem die spezifische Orientierung der Phosphatkette des ATPs, die dadurch eine Phosphatabspaltung und eine Übertragung der Phosphatgruppe auf ein Substrat erlaubt [MORGAN, 1997]. Durch die Rotation der C-Helix wird die Aminosäure Glu51 in die Nähe der Aminosäure Lys33 gebracht, wodurch beide Reste ein Ionenpaar bilden. Dadurch wird die Seitenkette der Aminosäure Lys33 bewegt, sodass es die negativ geladenen Sauerstoffatome der α - und β - Phosphatgruppen des ATP so positioniert, dass das γ - Phosphat auf die Hy-

droxygruppe des Serin oder Threonin des Substrats übertragen werden kann (Abb. 1.16) [BUIJSMAN, 2004]. Die Konformationsänderung des T-Loops bewirkt eine Rotation des in Proteinkinasen konservierten DFG Motivs, das die der Aminosäuresequenz Asp-Phe-Gly beschreibt. Die Aminosäure Asp145 des DFG Motivs interagiert mit einem von zwei Magnesiumionen, welche die β - und γ - Phosphatgruppen des ATP chelatisieren. Das zweite Magnesiumion interagiert zusätzlich mit dem Asn123. Während der Katalyse ist das Substrat über Wasserstoffbrücken mit dem Asp127 und dem Lys129 koordiniert. Durch die Anordnung des Magnesiumions, der γ -Phosphatgruppe und des Substrats wird ein nukleophiler Angriff auf das Substrat möglich (COOK *et al.* 2002; BUIJSMAN, 2004).

1.5.2 CDKs als *Drug-Targets* in der Wirkstoffentwicklung

In den vergangenen Jahren wurde sehr intensiv auf dem Gebiet des Zellzyklus und der CDKs geforscht, da diese als *Drug-Targets* in der Bekämpfung von proliferativen Erkrankungen wie z.B. Krebs angesehen werden. Jede sich teilende Zelle ist auf einen funktionierenden Zellzyklus angewiesen, gesunde Zellen ebenso wie maligne Tumorzellen. In Tumorzellen ist der Zellzyklus häufig dereguliert, was zur Folge hat, dass sich die Tumorzellen unkontrolliert teilen. Gelingt es, die CDKs zu inhibieren, ist ein Fortschreiten der Tumorentstehung unmöglich. Gleichzeitig ist ruhendes - nicht proliferierendes Gewebe - von dieser Maßnahme nicht betroffen. Der sequenzielle Einsatz von dem CDK-Inhibitor Flavopiridol in Kombination mit den Chemotherapeutika aus der Gruppe der Taxane zeigte eine synergistische Wirkung in klinischen Studien der Phase I. Es wäre demnach denkbar, die Dosen einer Chemotherapie zu senken, ohne den Therapieerfolg zu beeinträchtigen [SHAPIRO, 2004; 1999]. Häufig sind auch die CDKs in Tumoren überexprimiert, so dass zu erwarten ist, mit entsprechenden Inhibitoren genau diese Zellen im Wachstum zu hemmen, was gleichzeitig den Erfolg von Bestrahlungen oder operativen Maßnahmen erhöhen könnte [SHAPIRO & HARPER, 1999].

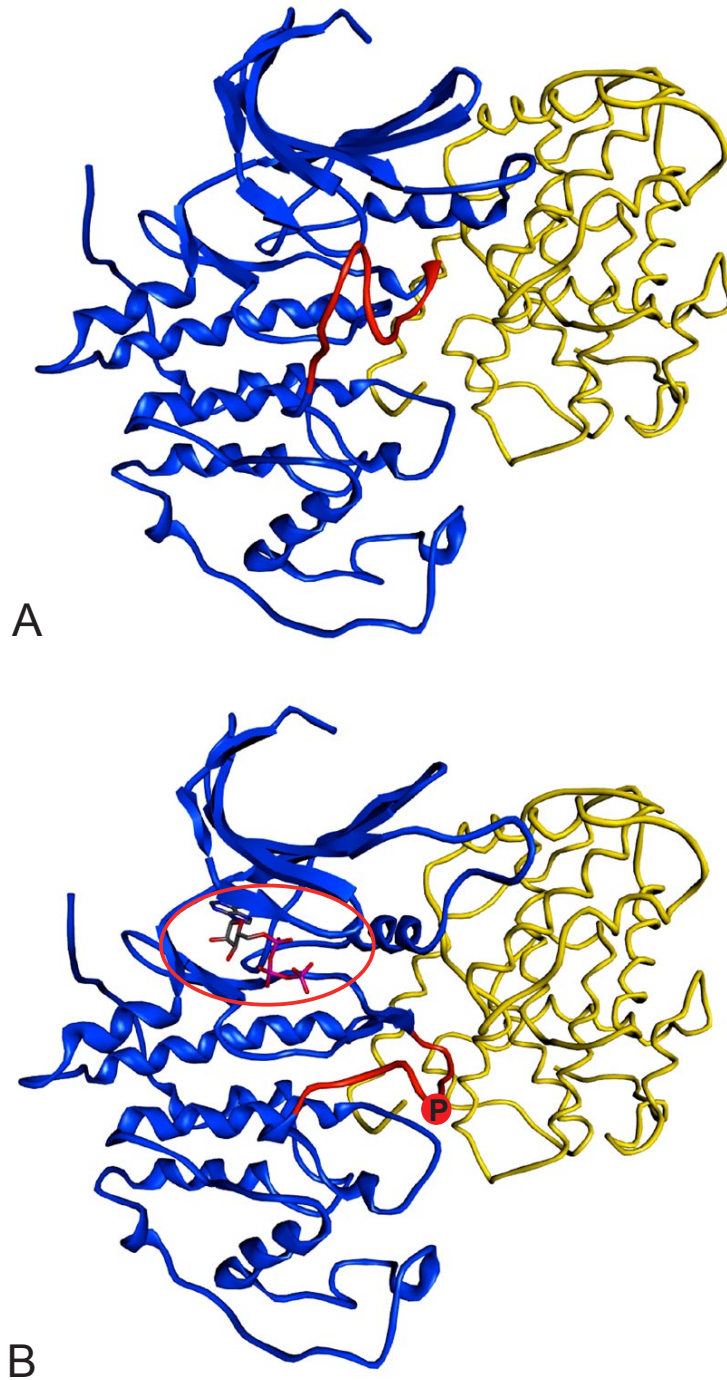


Abb. 1.14 Konformationsänderung der CDK (blau) nach Bindung eines Cyclins (gelb) und Phosphorylierung des Aktivierungs-Loops (rot). Abb 1.14 A zeigt die CDK im inaktiven Zustand kurz nach der Bindung eines Cyclins. Der Aktivierungs-Loop verdeckt die ATP Bindestelle. Nach Phosphorylierung der Aminosäure Thr160 ist der Aktivierungs-Loop heruntergeklappt und die ATP Bindestelle für das ATP (roter Kreis) zugänglich (Abb 1.14 B) [JOHNSON *et al.*, 1996].

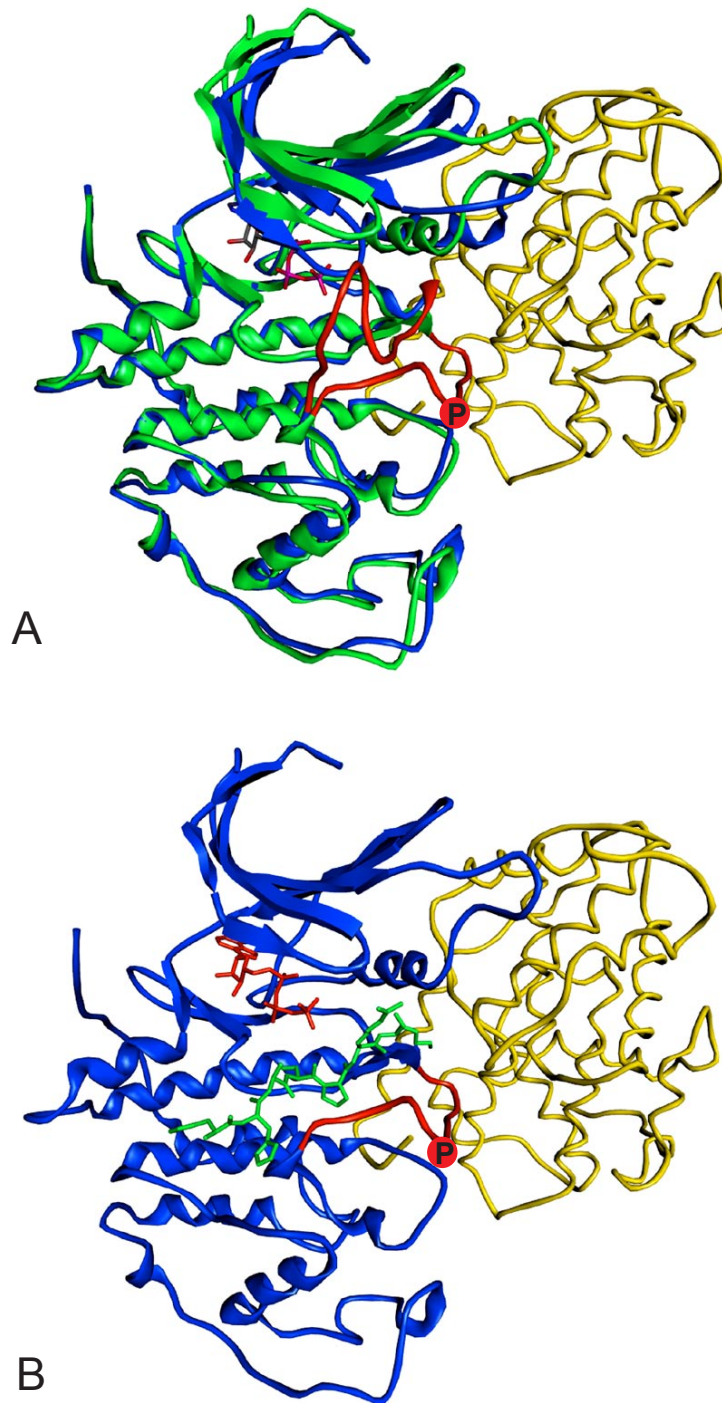


Abb. 1.15 Konformationsänderung der CDK (blau) nach Bindung eines Cyclins (gelb) und Phosphorylierung des Aktivierungs-Loops (rot). Abb 1.15 A zeigt die Konformationsänderungen der inaktiven CDK (blau) und der aktiven CDK (grün) im Vergleich als Überlagerung. Die Phosphorylierung eines Substrates (grün) ist in Abb. 1.15 B gezeigt.

Ebenso wie die Krebszellen besitzt jeder eukaryontische Parasit einen Zellzyklus, der dem des Menschen ähnlich ist. So konnten bereits einige CDKs in parasitären Protozoen identifiziert werden [DOERIG *et al.*, 2000]. Die prinzipielle Eignung von CDKs als *Drug-Target* Proteine in Parasiten wurde bereits mehrfach in der Literatur gezeigt [WOODARD *et al.*, 2003; HOLTON *et al.*, 2003; GOMEZ *et al.*, 2001]. Die Vermehrung der parasitären Protozoen *P. falciparum* und *Leishmania mexicana* - dem Erreger der Leishmaniose - konnte durch CDK-Inhibitoren, die humane CDKs inhibieren, ebenfalls unterbunden werden (HARMSE *et al.*, 2001, GRANT *et al.*, 2004). In Experimenten mit dem CDK-Inhibitor Flavopiridol konnte die Entwicklung und die Infektiosität von *E. tenella* *in vitro* gehemmt werden [Hofmann *et al.*, unveröffentlichte Da-

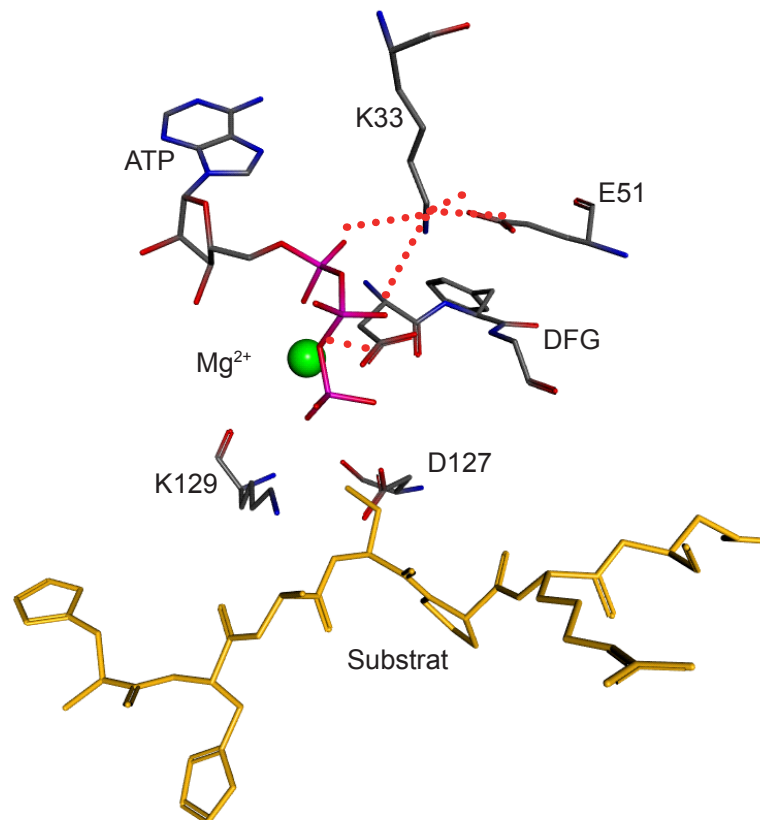


Abb. 1.16 Räumliche Anordnung der an der Katalyse beteiligten Aminosäurereste. Die Aminosäuren Lys33 (K33), Glu51 (E51), Asp145 (D145, DFG) koordinieren die Phosphatreste des ATP so, dass eine Übertragung der γ -Phosphatgruppe auf das Substrat ermöglicht wird. Ionische Wechselwirkungen zwischen den Aminosäuren Glu51 (E51), Lys33 (K33), Mg^{2+} und den Phosphatgruppen des ATP sind als rot gepunktete Linien dargestellt.

ten]. Die Sequenzähnlichkeiten von CDKs der Parasiten und die Inhibierung durch die gleichen Wirkstoffe lassen den Schluss zu, dass es sich bei der Inhibierung um einen ähnlichen Mechanismus handelt. Geringfügige Unterschiede der ATP Bindetaschen von CDKs der Vertebraten und der Protozoen, sollten darüberhinaus die Entwicklung selektiver Wirkstoffe erlauben [HOLTON *et al.*, 2003]. Im Analogieschluss sollte das Prinzip der Inhibierung einer CDK auch bei dem Parasiten *E. tenella* funktionieren. Eine Inhibierung von einer CDK von *E. tenella* sollte zur Folge haben, dass der Zellzyklus von *E. tenella* unterbunden wird, eine Zellvermehrung nicht mehr stattfinden kann und somit der Lebenszyklus des Parasiten unterbrochen wird.

1.5.2.1 Das Bindetaschenmodell der CDKs nach Traxler

Die Bindetasche einer CDK lässt sich in definierte Bereiche, in denen die Liganden binden, einteilen. Ein Bindetaschenmodell der CDK2 teilt die Bindetasche in die Bereiche ein, die von den zugehörigen Substrukturen des ATP besetzt sind (Abb. 1.17) [TRAXLER *et al.*, 1999]. In direkter Nähe zur Hinge-Region befindet sich die Adenin-Binderegion, die von allen bekannten Inhibitoren besetzt ist und an den *Gatekeeper* (Phe80; Abb. 1.17, F80) angrenzt. Einige Inhibitoren besetzen darüberhinaus auch die Ribose-Binderegion (z.B. Staurosporin; Abb. 1.18). Die Phosphat-Binderegion wird nur selten von Inhibitoren besetzt. Die Interaktion des ATP mit der Aminosäure Lys33 (Abb. 1.17, K33) wird von einigen Inhibitoren imitiert, z.B. vom Indirubin-5-sulphonat, das eine Ionenpaarbildung mit seiner Sulphonatgruppe zur positiv geladenen Aminosäure Lys33 eingeht (Abb. 1.18). Sehr wichtig für die Entwicklung selektiver Inhibitoren der CDKs ist die „Selektivitätstasche“, die zusätzlich von einigen Inhibitoren belegt wird. Beispielsweise wird diese Tasche von dem Inhibitor Indirubin-5-sulphonat und dem Phenylmethylsulpho-
namidrest des Inhibitors GW-9499 belegt [FURET, 2003; BUIJSMAN, 2004].

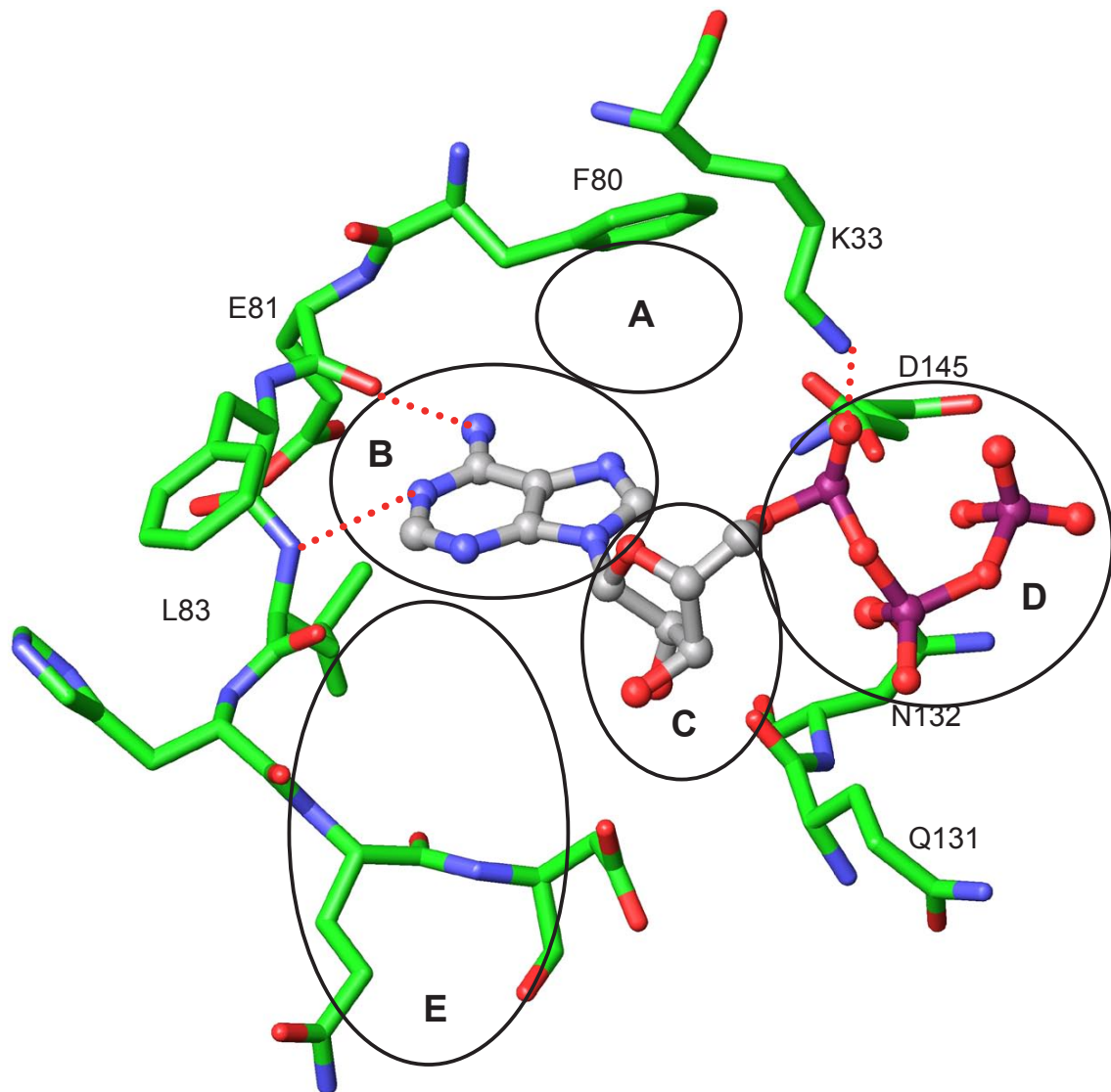


Abb. 1.17 Modell der Bindetasche von CDK2 [nach TRAXLER *et al.*, 1999]. Die Bindetasche lässt sich in 5 Binderegionen untergliedern: A: „Gatekeeper“-Region; B: Adenin-Binderegion, die an die *Hinge*-Region des Proteins angrenzt; C: Ribose-Binderegion; D. Phosphat-Binderegion; E. „Selektivitätstasche“, die dem Cytosol zugewandt ist. Alle Binderegionen können von Substrukturen verschiedener Inhibitoren besetzt werden (Vgl. Abb. 1.20). Insbesondere die Selektivitätstasche kann für das Design spezifischer Inhibitoren genutzt werden, da in dieser Tasche besonders häufig strukturelle Unterschiede in verschiedenen Proteinkinasen auftreten.

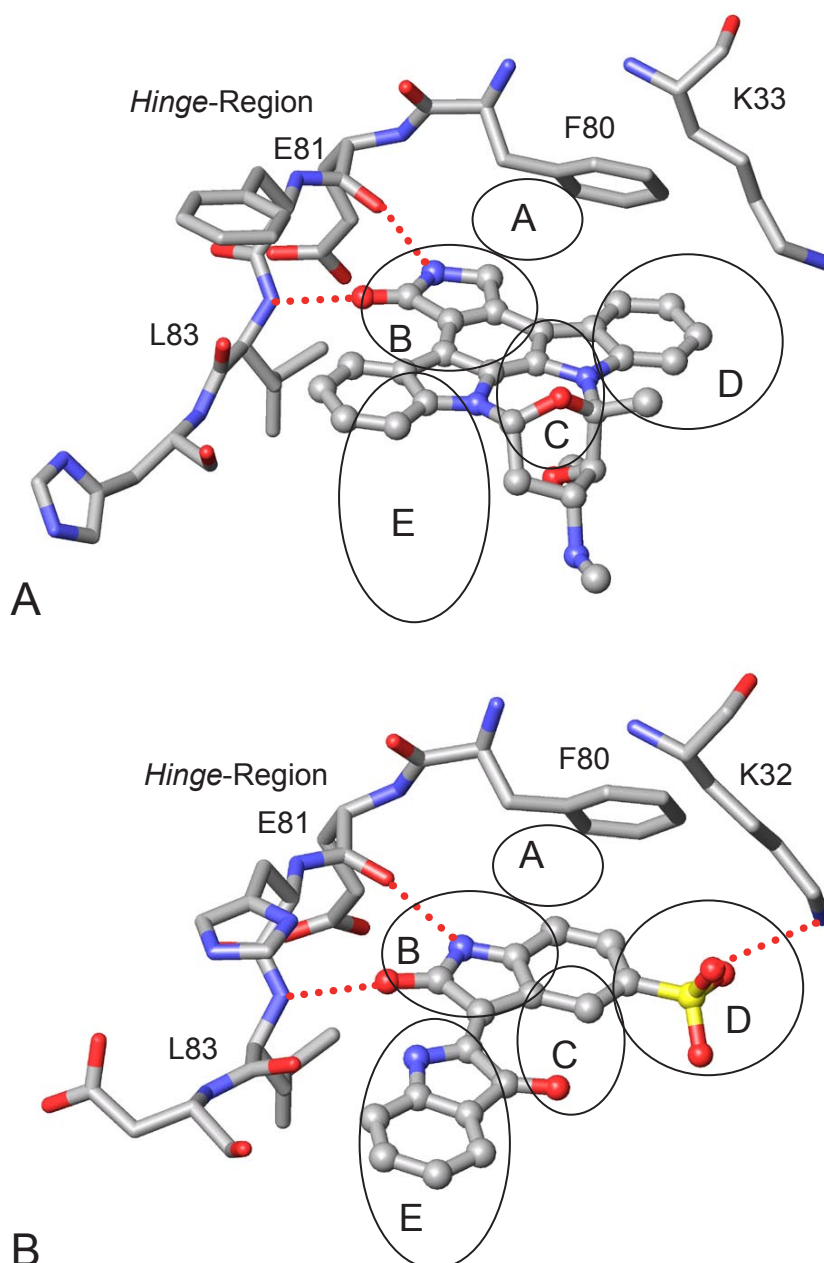


Abb. 1.18 Darstellung der Bindungsmodi von Staurosporin und Indirubin-5-sulphonat am Beispiel von CDK2 und PfPK5. Die Abbildungen wurden anhand von experimentell ermittelten Protein-Ligand Komplexen erstellt. Die Kreise bezeichnen die Binderegionen analog zu Abb. 1.17. Abb. 1.18 A zeigt die Bindung des Staurosporin. Das Staurosporin besetzt sowohl die Adenintasche (B), als auch die Ribosetasche (C) und die Selektivitätstasche (E) (PDB Acc.No.: 1AQ1). Abb. 1.18 B zeigt das Indirubin-5-sulphonat, das über drei Wasserstoffbrücken mit der *Hinge*-Region der PfPK5, einer CDK aus *P. falciparum*, interagiert und die Sulphonatgruppe zum Lys32 (homolog zu Lys33 in humaner CDK2) hin über eine ionische Interaktion orientiert (PDB Acc. No.: 1V00).

1.5.2.2 Inhibitoren der CDKs

In Abb. 1.19 sind einige Inhibitoren humaner CDKs dargestellt. Alle bekannten ATP-kompetitiven Inhibitoren der CDKs besitzen ein planares Ringsystem, das über Wasserstoffbrücken mit der *Hinge*-Region der Proteinkinasen interagiert. Meist finden sich hydrophobe Regionen des Liganden im Bereich des „*Gatekeeper*“ Phe80 (humane CDK2) und am *G-Loop* der CDKs [BUIJSMAN, 2004; FURET, 2003]. Die größte Gruppe der Inhibitoren der CDKs bilden die Purinderivate. Zu den Purinderivaten gehören das Purvalanol B, das (R-)Roscovitin, Olomoucine und NU2058. Zum Vergleich wurde das ATP mit abgebildet. Im Unterschied zum ATP erfolgt die Bindung der Purinderivate unter einer Rotation des Purinsystems um etwa 160° relativ zum Purin des ATP [FURET, 2003; BUIJSMAN, 2004].

Weitere Inhibitoren sind die Naturstoffderivate Flavopiridol, 10Z-Hymenialdisine, Staurosporin und das Indirubin-5-sulphonat [FURET, 2003; NOBLE & ENDICOTT, 1999; GRAY *et al.*, 1999]. Der Inhibitor Flavopiridol, der alle humanen CDKs gleichermaßen inhibiert, wurde bereits in klinischen Studien am Menschen getestet, konnte allerdings die Zulassungskriterien bisher nicht erfüllen [SENDEROWICZ, 1999; FISCHER & GIANELLA-BORRADORI, 2005]. Staurosporine sind im allgemeinen unselektive Proteinkinase-Inhibitoren, dennoch werden Derivate des Staurosporins als Proteinkinase-Inhibitoren in verschiedenen klinischen Studien geprüft [BUIJSMAN, 2004]. Indirubine sind selektive Inhibitoren der CDKs, die auch die PfPK5 - eine CDK aus *P. falciparum* - inhibieren [HOLTON *et al.*, 2003].

Alle in Abb. 1.19 dargestellten Inhibitoren sind ATP-kompetitive Inhibitoren. Diese müssen mit hohen ATP Konzentrationen in der Zelle konkurrieren und benötigen deshalb eine relativ starke Bindungsaffinität. Aus den Bindungsmodi der bekannten ATP-kompetitiven Inhibitoren der CDKs lässt sich ein Bindetaschen-Modell für das rationale *Drug-Design* von CDK-Inhibitoren ableiten.

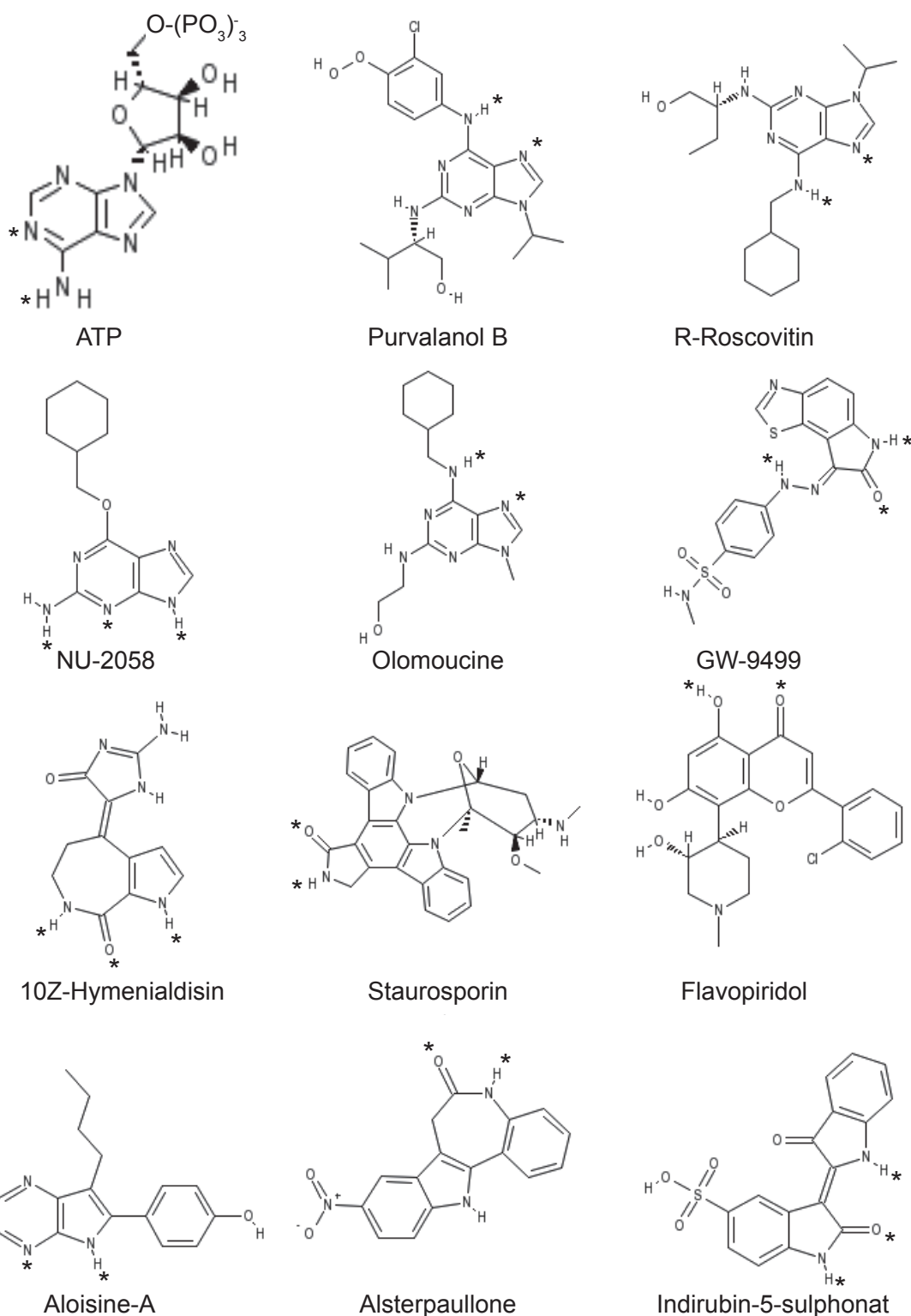


Abb. 1.19 Strukturformeln von ATP und bekannten Inhibitoren humaner CDKs. Die Wasserstoffbrückenbindungen zur Hinge-Region sind durch Sterne markiert [nach Furet, 2003]

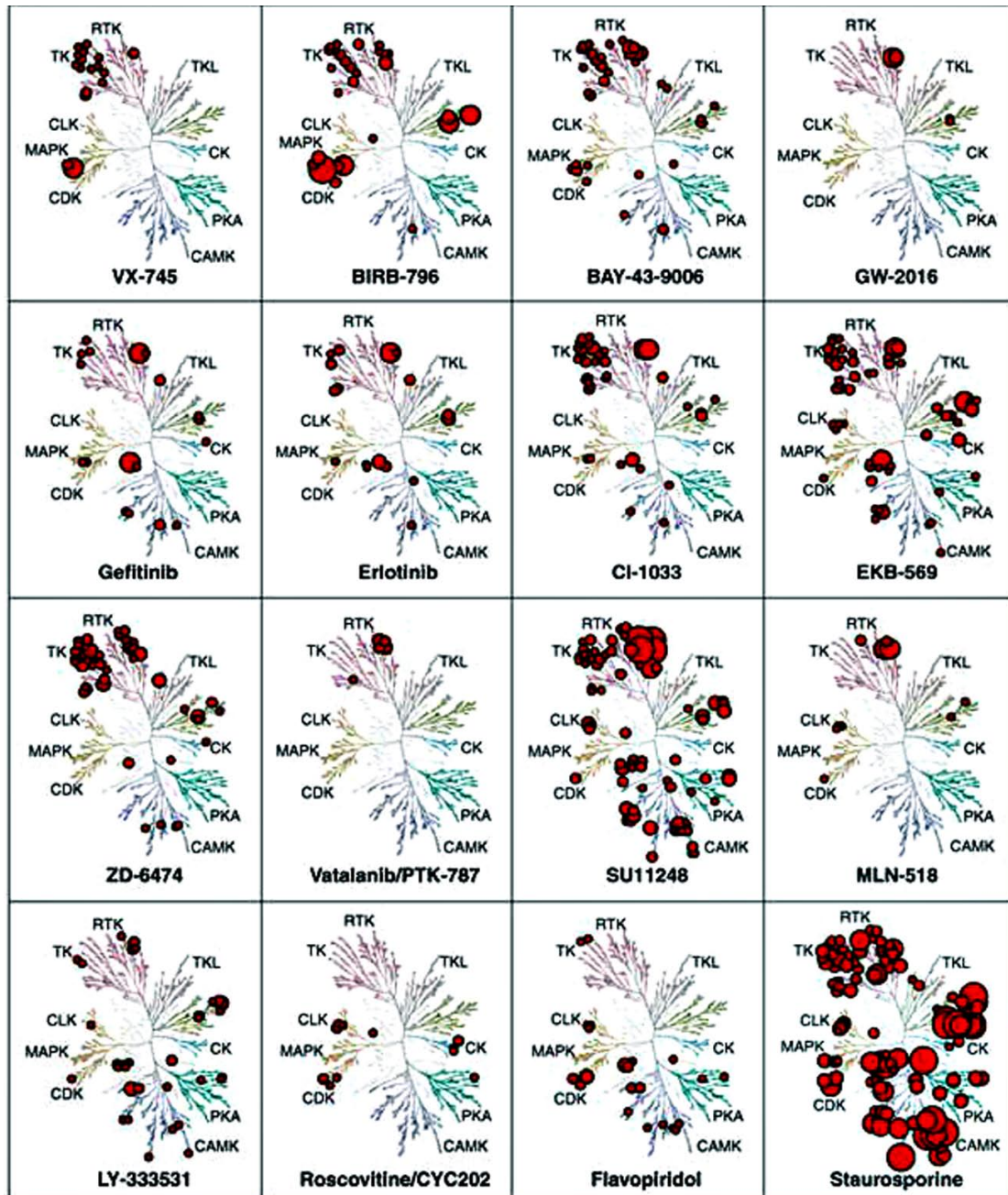


Abb. 1.20 Drug-Target-Selektivität einiger Kinase-Inhibitoren. Auf den Einzelabbildungen ist jeweils ein Dendrogramm der Kinasen des Menschen abgebildet. Die Größe der roten Kreise markiert die Wirkstärke eines Inhibitors an einer bestimmten Kinase. Aus der Anzahl der Kreise lässt sich die Selektivität des Inhibitors ablesen. Staurosporin ist bspw. ein Inhibitor, der sehr unspezifisch ist. Einige Inhibitoren wirken nur auf einzelnen Proteinkinase-Familien, wie bspw. der Inhibitor GW-2016. Die Abbildung demonstriert, dass es möglich ist, selektive Kinase-Inhibitoren zu finden [nach FABIAN *et al.*, 2005].

- K_d
- < 1 nm
 - 1-10 nm
 - 10-100 nm
 - 100 nm-1 μ m
 - 1-10 μ m

In Abbildung 1.20 sind die Selektivitäten einiger Proteinkinase-Inhibitoren für die Proteinkinase-Familien des Menschen dargestellt [FABIAN *et al.*, 2005]. Die Abbildung demonstriert die Möglichkeit der Entwicklung von relativ selektiven Proteinkinase-Inhibitoren, wie beispielsweise des Wirkstoffs Gefitinib (Iressa[®]) und des experimentellen Inhibitors GW-2016.

1.6 Zielsetzung der Arbeit

Computermethoden erfahren eine zunehmende Verbreitung in der Wirkstoffforschung. Die steigende Rechenleistung von Computern ermöglicht dabei immer aufwendigere und genauere Berechnungen. Dennoch reicht die Rechenleistung eines einzelnen Computers meist für Anwendungen in der Wirkstoffforschung nicht aus. Daher müssen die Rechenoperationen parallelisiert auf vielen Computern durchgeführt werden.

Die computerunterstützte Wirkstoffforschung wird heute in der Pharma-, Agrar- und veterinärmedizinischen Forschung routinemäßig durchgeführt. Ein wichtiger Bereich der veterinärmedizinischen Wirkstoffforschung sind die Parasitosen von Haus- und Nutztieren. Ein Parasit, der in der Geflügelzucht weltweit große wirtschaftliche Schäden verursacht, ist der Protozoe *E. tenella*. Aus vorangegangenen Experimenten war bekannt, dass der Parasit *E. tenella* durch den CDK-Inhibitor Flavopiridol in der Entwicklung und der Infektiosität begrenzt werden kann.

In der vorliegenden Arbeit sollte eine *Grid-Computing* Infrastruktur für die Wirkstoffforschung etabliert werden, die die parallele Berechnung vieler Einzel-Operationen ermöglicht. Von besonderer Bedeutung war die Implementierung eines *Docking*-Algorithmus, der die Bewertung von Protein-Ligand-Wechselwirkungen ermöglicht. Die großen Datenmengen, die bei diesen Berechnungen anfallen, sollten anschließend mit üblicherweise verwendeter Software weiter bearbeitet werden können. Die Anwendbarkeit der *Grid-Computing* Methode sollte zudem mit bekannten Daten aus der Literatur überprüft werden.

Ein weiteres Ziel war die Suche nach neuartigen Inhibitoren von CDKs des Parasiten *E. tenella*. Dafür sollten Modelle auf der Grundlage von Experimentaldaten der humanen CDKs erstellt werden. Zunächst sollten Homologiemodelle basierend auf Röntgenkristallstrukturen von anderen CDKs generiert werden. Darüberhinaus sollten anhand von bekannten CDK-Inhibitoren mögliche Struktur-Aktivitätsbeziehungen von Inhibitoren einer CDK aus *E. tenella* abgeleitet werden. Die Struktur-Aktivitätsbeziehungen sollten der Interpretation der Daten und der Selektion von potenziellen Inhibitoren der CDK dienen.

2 Material und Methoden

2.1 Material

2.1.1 Hardware

SGI: SGI Origin 3200, SGI O2, SGI Octane 2, SGI Fuel

Dell: Dell Optiplex GX260, Dell Optiplex GX240

2.1.2 Betriebssysteme

Unix: Irix 6.5 (SGI Computer)

Linux: RedHat 9.0 (Dell Computer)
SuSE 9.0 Professional (Dell Computer)

Windows: Windows XP Professional (Dell Computer)

2.1.3 Software

Cactvs Toolkit 3.x: Ein chemoinformatisches Toolkit. Xemistry GmbH, Lahntal. [[HTTP://WWW.XEMISTRY.DE](http://www.xemistry.de)].

Catalyst (Versionen 4.9 - 4.11):
Software zum Erstellen von Pharmakophormodellen und für die Strukturdatenbanksuche. Accelrys Inc., San Diego, USA [[HTTP://WWW.ACCELRY.COM](http://www.accelrys.com)].

Cerius2 (Versionen 4.9 und 4.10):
Chemoinformatisches Software-Paket. Accelrys Inc., San Diego,
USA [[HTTP://WWW.ACCELRY.COM](http://www.accelrys.com)].

Accelrys Inc., San Diego, USA.
Corina 3.3: Konformergenerator für kleine Moleküle
Molecular Networks GmbH, Erlangen [[HTTP://WWW.MOL-NET.COM](http://www.mol-net.com)].

GeneDoc (Version 2.6.001): Sequenz-*Alignment* Editor
Pittsburgh Supercomputing Center, USA [NICHOLAS *et al.*, 1997].

Insight II (Versionen 2000 und 2005): Software für Molekulares
Modelling von Proteinen. Accelrys Inc., San Diego, USA
[[HTTP://WWW.ACCELRY.COM](http://www.accelrys.com)].

Lasergene: Software zur Sequenzanalyse.
DNASTAR Inc., Madison, USA [[HTTP://WWW.DNASTAR.COM](http://www.dnastar.com)].

Maestro: Software zur Visualisierung und Manipulation von
Molekülen. Schrödinger Inc., New York, USA
[[HTTP://WWW.SCHRODINGER.COM](http://www.schrodinger.com)].

Metalife Trinity 1.4: Bioinformatisches Programmpaket
Metalife AG, Winden [[HTTP://WWW.METALIFE.DE](http://www.metalife.de)].

Modeller: Software-Modul zum Erstellen von Homologie-
modellen. Accelrys Inc., San Diego, USA [SALI *et al.*, 1995].

Molcad: Software zur dreidimensionalen Visualisierung von
Molekülen. Tripos Inc., St. Louis, USA
[BRICKMANN *et al.*, 1995, 2000].

Petra 3.22: Software zur Berechnung von Moleküleigenschaften
Molecular Networks GmbH, Erlangen [[HTTP:WWW.MOL-NET.COM](http://www.mol-net.com)].

Vector NTI: Bioinformatisches Programmpaket. Invitrogen Ltd.,
Paisley, UK. [[HTTP://WWW.INVITROGEN.COM](http://www.invitrogen.com)].

2.1.4 Datenbanken

Datenbank	URL	Referenz
GenBank	http://www.ncbi.nlm.nih.gov/	[BENSON <i>et al.</i> , 2004]
PubMed	http://www.pubmed.gov/	[WEEHLER <i>et al.</i> , 2004]
PROSITE	http://us.expasy.org/prosite/	[HULO <i>et al.</i> , 2003]
InterPro	http://www.ebi.ac.uk/interpro/	[MULDER <i>et al.</i> , 2005]
PDB	http://www.rcsb.org/pdb/	[BERMAN <i>et al.</i> , 2002]
ProDom	http://www.toulouse.inra.fr/	[BRU <i>et al.</i> , 2004]
<i>E. tenella</i> Genome Project	http://www.sanger.ac.uk/Projects/E_tenella/	[CHAPMAN <i>et al.</i> , 2003]

2.1.5 Programmiersprachen

Tcl/Tk: Tool Command Language [OUSTERHOUT *et al.*, 1994]. Verwendung im Zusammenhang mit CACTVS, einem chemischen Datenbankmanagementsystem als Erweiterung der Tcl-Skriptsprache durch chemische Daten- und Strukturmanipulationskommandos [IHLENFELDT *et al.*, 2002].

2.2 Methoden

2.2.1 Sequenz-Alignments zum direkten Sequenzvergleich

Sequenz-Alignments dienen der Detektion und Darstellung konservierter sowie diverser Bereiche zwischen zwei (paarweises *Alignment*) oder mehreren Sequenzen (multiples *Alignment*). Die vorliegenden multiplen *Alignments* zur Erstellung der Bioinformatischen Tiefenanalyse wurden ausschließlich mit dem in der Software Lasergene implementierten ClustalW Algorithmus unter Verwendung der Standardeinstellungen durchgeführt [THOMPSON *et al.*, 1994]. Hierzu wurden die Sequenzen in ein Sequenz-Set im Softwaremodul MegAlign importiert. ClustalW erstellt aus mehreren Sequenzen globale multiple *Alignments*. Diese wurden in das Programm GeneDoc importiert, um die *Alignments* graphisch darzustellen [NICHOLAS *et al.*, 1997].

2.2.2 Bestimmung orthologer Proteine der TgCRK3 in *E. tenella*

Zur Suche orthologer Proteine der *Toxoplasma gondii* CRK3 wurde mit der CRK3 Proteinsequenz eine tBLASTn-Suche gegen eine Sequenzdatenbank von *E. tenella* durchgeführt [ALTSCHUL *et al.*, 1990; BENSON *et al.*, 2004]. Die so identifizierten Sequenzen wurden anschließend jeweils als Suchsequenz für die BLASTp-Datenbanksuche gegen eine nicht-redundante Proteindatenbank (GenBank) eingesetzt [KORF *et al.*, 2003]. Wurden die Sequenzen von

Cyclin-abhängigen Kinasen anderer Organismen dabei wieder als beste Treffer gefunden (Best-Best-Hit) wurden die zugehörigen Proteine als ortholog bezeichnet [BECKSTETTE *et al.*, 2004].

2.2.3 Homologiemodellierung der EtCRK2

Zur Erstellung der Homologiemodelle von EtCRK2 wurde das Homology-Modul der Software Insight II verwendet [FISER & SALI, 2000; GREER, 1990]. Als Templatstruktur wurde eine Kristallstruktur der humanen CDK2 (PDB *Acc. No.*: 1OIR) eingesetzt. In der PDB-Datenbank sind weitere Kristallstrukturen von Cyclin-abhängigen Kinasen verschiedener Organismen abgelegt. Aufgrund ihrer geringeren Auflösungen und anderer struktureller Eigenschaften, wie beispielsweise höherer isotroper Debye-Waller-Temperaturfaktoren (B-Faktoren) und der Ausrichtungen der Seitenketten wurden diese Strukturen jedoch nicht als Template für die Modellbildung eingesetzt [GAO & PENG, 1999; GARCIA-SOSA *et al.*, 2003]. Das Homologiemodell wurde auf Basis eines multiplen *Alignments* der Sequenzen der bekannten Kristallstruktur 1OIR aus *Homo sapiens* und der EtCRK2 mit dem Align123 Algorithmus erstellt [HTTP://WWW.ACCELRY.COM/; THOMPSON *et al.*, 1994]. Es wurden jeweils vier Modellvorschläge erstellt. Die Ramachandran-Plots der Homologiemodelle wurden mit der Software Procheck generiert [LASKOWSKI, 1993].

2.2.4 Berechnung von EtCRK2-spezifischen Pharmakophormodellen

Die Protein-Ligand-basierten Pharmakophormodelle wurden mit der Software Cerius2 erstellt. Hierbei wurden insbesondere die Interaktionsmuster, welche in der Literatur für verschiedene Cyclin-abhängige Kinasen beschrieben sind, berücksichtigt. In einem ersten Schritt wurde mit Hilfe des Moduls „Structure Based Focusing“ eine proteinbasierte Interaktionskarte erstellt, welche die vom Protein ausgehenden, möglichen Wasserstoffbrückenwech-

selwirkungen sowie Bereiche möglicher lipophiler Wechselwirkungen beinhaltet. Danach wurden die einzelnen Interaktionsbereiche geprüft und manuell diejenigen entfernt, die für die Auswahl der Liganden nicht relevant erschienen. Im anschließenden Schritt wurden die einzelnen Interaktionsbereiche mittels Cluster-Analyse in Gruppen zusammengefaßt, sodass nach der Analyse nur noch jeweils eine Interaktion pro Cluster bestehen blieb. Die Interaktionsgruppen wurden in verschiedenen Dateien abgespeichert, sodass eine weitere Bearbeitung möglich war. In der weiteren Bearbeitung wurden einzelne Interaktionscluster für die Datenbanksuche mit Catalyst verändert. Dabei wurde die dem Lys32 der EtCRK2 zugeordnete Wasserstoffbrückenakzeptor-Funktion zu der Pharmakophor-Eigenschaft „negativ ionisierbar“ geändert. Die dem Glu88 zugeordnete Wasserstoffbrückendonorfunktion wurde zu der Eigenschaft „positiv ionisierbar“ verändert. Aus diesen Interaktionsgruppen wurden anschließend alle möglichen Kombinationen von Pharmakophoren ermittelt, die immer 4 Interaktionen beinhalten, wovon mindestens eine Wasserstoffbrückendonor-, eine Wasserstoffbrückenakzeptor- und eine hydrophobe Interaktion in jedem Modell erhalten blieb. Die resultierenden Modelle wurden anschließend selektiert, sodass in allen Modellen die dem Leu82 zugeordnete Wasserstoffbrückenakzeptor-Funktion vorhanden war.

2.2.5 Datenbanksuche mit der Software Catalyst

Mit der Software Catalyst wurden mehrere Datenbanken von kommerziell erhältlichen Substanzen erstellt. Die Software Catalyst generiert dafür für jedes Molekül, das in einer Datenbank vorhanden ist, eine Anzahl von Konformeren. Für jedes Konformer werden alle möglichen Funktionen (Wasserstoffbrückendonor, Wasserstoffbrückenakzeptor, hydrophobe Wechselwirkungen usw.), die einer chemischen Gruppe zugeordnet sind, in der zugehörigen Catalyst Datenbank abgelegt. Abweichend von den Herstellerangaben, wurde die in der Smiles-Notation vorgegebene Stereochemie eines

Moleküls beibehalten und es wurden maximal 150 Konformere für ein Molekül generiert. Für die Datenbanksuchen wurden die mittels Cerius2 erstellten Pharmakophormodelle mit Catalyst modifiziert. Dabei wurden die einzelnen Pharmakophorhypothesen zu einem kombinierten Modell der Einzelhypothesen zusammengefaßt, um die Datenbanksuche effizienter zu gestalten. Die Suche wurde mit der Option „best Flexible“ durchgeführt.

2.2.6 Vorbereitung der Substanzbibliothek

Chemische Verbindungen können verschiedene Zustände hinsichtlich ihrer strukturellen Eigenschaften (Konformere), ihrer elektrochemischen Eigenschaften (Ionisierungszustände) und ihres Tautomerisierungszustandes (Tautomere) annehmen. Die verschiedenen Konformere einer chemischen Verbindung können im flexiblen *Docking* von Molekülen automatisch betrachtet werden. Die möglichen Ionisierungszustände und Tautomere müssen in der Regel jedoch vorher berechnet und als individuelle Strukturen abgespeichert werden. Die Anzahl chemischer Verbindungen einer Substanzdatenbank kann daher in der *in silico* Repräsentation leicht eine Größenordnung höher sein. Die für das *Docking* verwendeten Datenbanken wurden unmittelbar vor dem *Docking* gefiltert und verschiedene Ionisierungszustände elektronisch generiert. Mit einem Eigenschafts-Filter wurden zuerst Moleküle aussortiert, deren Molekulargewicht größer als 450 Da oder die mehr als 8 rotierbare Bindungen oder andere Atome als C, N, O, S, P, H, F, Cl, Br, I enthielten. Auf die Betrachtung von unterschiedlichen Tautomerzuständen wurde im Rahmen dieser Arbeit verzichtet, da für ein Massenscreening dieser Daten keine Rechenkapazitäten zur Verfügung standen. Stattdessen wurde das in der Strukturdatenbank vorhandene Tautomer verwendet. Die Aufbereitung der Strukturdatenbanken wurde mit Tcl-Skripten unter Einsatz der Programmierbibliothek Cactvs durchgeführt [IHLENFELDT *et al.*, 2002; OELLIEN *et al.*, 2005].

Für die so vorbereiteten Strukturen wurden mit der Software Corina Start-Konformere berechnet. Die Strukturen wurden ihres Ionisierungszustandes entsprechend mit Wasserstoffatomen abgesättigt. Für die ionisierten Konformere wurden atomare Partialladungen mit der Software Petra berechnet. Das Ausgabeformat von der Software Petra ist das Gasteiger-Cleartext-Format (CTX-Datei). Die CTX-Dateien wurden dann für die Verwendung im *Grid-Computing Framework* bereitgehalten.

2.2.7 Docking und Scoring mit der Software Autodock

Als *Docking*-Software wurde die Software Autodock 3.05 verwendet [MORRIS *et al.* 1996]. Autodock verwendet u.a. einen genetischen Algorithmus zur Generierung von Konformeren einer Struktur in der Proteinbindetasche. Bei einem genetischen Algorithmus wird der Prozeß der Evolution mittels Mutation, Rekombination und Selektion von Individuen einer Population simuliert. Die Individuen entsprechen dabei den jeweiligen Ligand-Konformationen. Eine Population ist demnach die Menge aller Ligand-Konformationen einer Struktur. Die Fitness (*Scoring*-Wert) eines Individuums bzw. einer Ligand-Konformation, wird mit einer Bewertungsfunktion ermittelt. Diese enthält folgende fünf Bewertungskomponenten: 1. Bindungsenergie der Wasserstoffbrücken zwischen Protein und Ligand (intermolekulare Wasserstoffbrücken), 2. intermolekulare van-der-Waals-Energie (vdw) zwischen Protein und Ligand, 3. intramolekulare vdw-Energie des Liganden, 4. Energie der Torsionsspannung des Liganden und 5. Bindungsenergie der intramolekularen Wasserstoffbrücken im Liganden. Das Protein wird bei diesem Verfahren starr gehalten. Ergebnis des *Dockings* ist eine Datei im Autodock-Format, das die Atomkoordinaten des Liganden und die *Scoring*-Werte enthält.

Die folgenden Autodock-Parametereinstellungen wurden für das Hochdurchsatz-*Docking* verwendet:

```
tran0    random
quat0    random
dihe0    random
tstep    0.5
qstep    5.0
dstep    5.0
outlev    1
rmstol    0.5
extnrg    1000.0
e0max    0.0 10000
ga_pop_size 50
ga_num_evals 100000
ga_num_generations 27000
ga_elitism 1
ga_mutation_rate 0.02
ga_crossover_rate 0.8
ga_window_size 10
ga_cauchy_alpha 0.0
ga_cauchy_beta 1.0
sw_max_its 300
sw_max_succ 4
sw_max_fail 4
sw_rho 1.0
sw_lb_rho 0.01
ls_search_freq 0.06
ga_run 10
```

Für die genauere Berechnung von Ligand-Posen im Rahmen der Validierung wurden die Werte der Parameter `ga_num_evals` und `ga_num_generations` auf 300.000 und 100.000 erhöht.

2.2.8 Virtuelles Screening mit der EtCRK2

Mit der *Docking*-Software Autodock wurde eine Substanzbibliothek aus ca. 450 000 kommerziell erhältlichen Substanzen, die ca. 530 000 *in silico*-Substanzen entsprechen, auf einem *Compute-Grid Cluster* untersucht. Die so gewonnenen Rohdaten wurden in einer Post-Prozessierung weiter analysiert. Die Post-Prozessierung umfasst das Aussortieren gedockter Konformere die bestimmte Kriterien nicht erfüllen, ein *Re-Scoring* und die automatisierte Extraktion der bestbewerteten Strukturen. Für die Extraktion der bestbewerteten Strukturen und ihre Orientierungen wurde unter Verwendung der Programmierbibliothek Cactvs ein Tcl-basiertes Skript erstellt [IHLENFELDT *et al.*, 2002]. Das Skript kann durch den Benutzer parameterisiert werden und erlaubt folgende Auswertungen:

1. Ausgabe der jeweils besten Orientierung für jeden Liganden und jede Bewertungsfunktion
2. Ausgabe der jeweils besten Strukturen für jede Bewertungsfunktion.

Die aus diesem Verfahren resultierenden Strukturen wurden dann anhand des *Scoring*-Wertes analysiert und bewertet. Davon wurden die ca. 4000 Strukturen mit den günstigsten Bewertungen optisch in der Bindetasche begutachtet. Dabei wurden ähnliche Bindungsmodi, wie sie für andere Inhibitoren von CDKs in der Literatur beschrieben sind, als Maß einer günstigen Interaktion mit der EtCRK2 angesehen. Die so gefundenen Substanzen wurden für biochemische Experimente ausgewählt.

3 Ergebnisse

3.1 Aufbau des *Grid-Computing Frameworks*

Das *Docking* von Strukturdatenbanken ist ein rechenintensiver Vorgang [SCHULZ-GASCH & STAHL, 2003]. Sollen große Strukturdatenbanken in kurzer Zeit in eine Bindetasche eingepasst werden, ist es daher erforderlich, die Rechenoperationen auf viele Computer zu verteilen. Eine Möglichkeit dafür ist das *Grid-Computing* [CHIEN *et al.*, 2003; HERNANDEZ *et al.*, 2005]. Dabei werden viele Computer parallel für das *Docking* vieler Strukturen genutzt (Kap. 1.4). Im Rahmen dieser Arbeit wurde ein *Docking*-Verfahren mit einer *Grid-Computing*-Methode kombiniert [MORRIS *et al.*, 1998; ROGERS *et al.*, 2006].

3.1.1 Aufbau des *Grid-Computing Clusters*

Abb. 3.1 stellt den Aufbau des *Grid-Computing* Testsystems schematisch dar. Zentrale Steuerfunktionen übernimmt der *Grid-Computing* Server [[HTTP://WWW.AXCELEON.COM](http://www.axceleon.com)]. Der Server steuert die Vergabe der einzelnen *Dockings* an die *Client*-Computer. Die *Client*-Computer sind über zwei Netzwerk-*Switches* mit dem Server verbunden. Sechs der *Client*-Computer werden täglich für Büroanwendungen genutzt. Zehn weitere Computer sind für Testzwecke zusätzlich in das *Grid*-Netzwerk integriert. Ein zweiter Server stellt eine Oracle Datenbank zur Verfügung, auf der der *Grid-Computing* Server Daten ablegt, die für eine vollständige Auswertung und Vervollständigung der Ergebnisse der *Dockings* notwendig sind [[HTTP://WWW.ORACLE.COM](http://www.oracle.com)]. Die Server sind von den *Client*-Computern mittels eines lokalen Netzwerks erreichbar [[HTTP://WWW.NETSARANG.COM](http://www.netsarang.com)].

3.1.2 Programmierung

Kernelement des *Grid-Computing Frameworks* und der Vorbereitung der Liganddateien für das *Docking*, ist ein Tcl-Skript (Cactus-Skript) [OELLIEN *et al.*, 2006; IHLENFELDT *et al.*, 1996; [HTTP://WWW.XEMISTRY.COM](http://www.xemistry.com)]. Die Strukturinformationen werden im Gasteiger *clear-text-format* (CTX-Datei) gespeichert und innerhalb des *Frameworks* ausgelesen. Danach werden sie in das Autodock Ligandenformat konvertiert (Abb. 3.2) [GASTEIGER *et al.*,

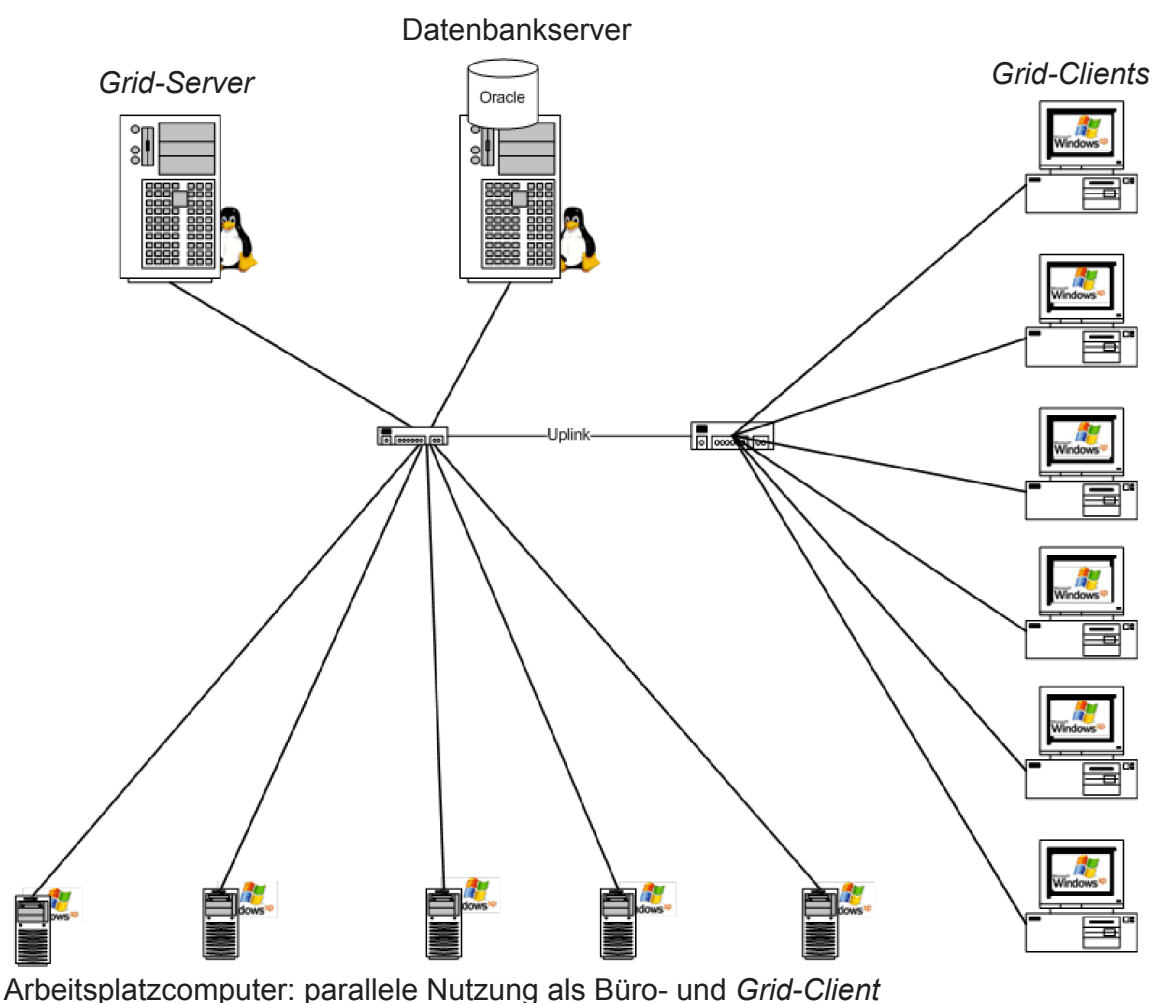


Abb. 3.1 Schema des *Grid-Computing Frameworks*. Kernbestandteile des Systems sind der *Grid-Server* und der *Datenbankserver*. Der *Grid-Server* steuert den gesamten Prozess und speichert alle benötigten Daten temporär auf dem *Datenbankserver*. Die einzelnen *Docking* Berechnungen werden auf die *Client-Computer* verteilt und sind über zwei Netzwerk-Switches mit den Servern verbunden.

1994]. Die Software Autodock verwendet proprietäre Dateiformate, die sie zu anderer Software inkompatibel macht. Auch ist die Software nicht für das *Docking* vieler Strukturen in einem Vorgang ausgelegt. Durch Anpassungen innerhalb des *Grid-Computing Frameworks* konnten einige der Nachteile behoben werden.

Das Autodock Liganddateiformat beschreibt die Liganden anhand einer baumartigen Struktur ihrer durch rotierbare Bindungen getrennten Molekülfragmente. Ausgehend von einem zentralen Bestandteil des Moleküls werden auf diese Weise „Äste“ und „Zweige“ des Moleküls aufgebaut [AUTODOCK USER'S GUIDE, [HTTP://WWW.SCRIPPS.EDU/MB/OLSON/DOC/AUTODOCK/AD305/USING_AUTO-DOCK_305.HTML](http://www.scripps.edu/mb/olson/doc/autodock/ad305/using_autodock_305.html)]. Die dafür erforderliche Routine wurde in das *Framework*, das das *Grid-Computing* steuert, implementiert. Die Informationen zu Atomladung und Konnektivität müssen parallel dazu in der Datenbank gespeichert werden, um für eine spätere Ergebnisausgabe und Auswertung zur Verfügung

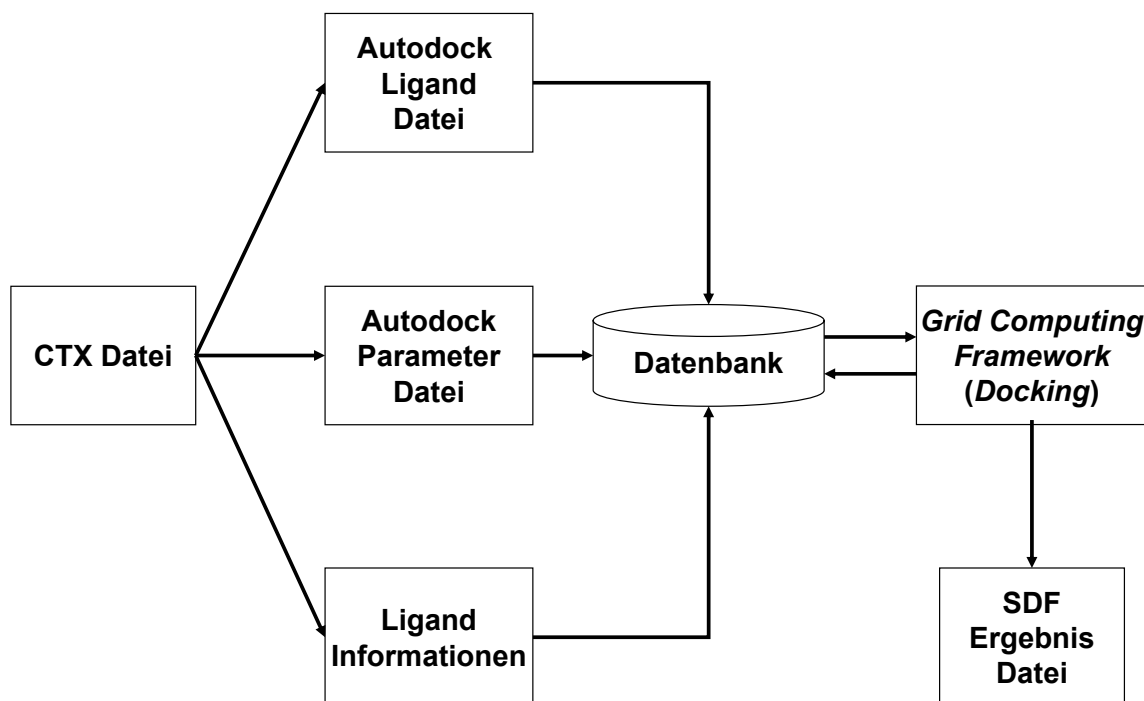


Abb. 3.2 Flussdiagramm des *Docking*-Prozesses im *Grid-Computing Framework*. Ausgehend von einer CTX-Datei werden Ligand- und Parameterdateien generiert und zusammen mit weiteren Ligand-Informationen in der Datenbank gespeichert. Das *Grid-Computing Framework* steuert dabei die *Grid-Computing* Software. Die Ergebnisse der *Dockings* werden in einer SDF-Datei gespeichert.

zu stehen. Darüberhinaus werden alle weiteren Informationen, die über das jeweilige Molekülensemble in der CTX-Datei enthalten sind, in der Datenbank abgelegt. Die standardmäßigen Autodock Atomtypen wurden um zwei Atomtypen erweitert: 1. um nicht-polarisierten Stickstoff und 2. um nicht-polarisierten Wasserstoff. Beide Atominformationen erwiesen sich für verbesserte *Docking*-Ergebnisse notwendig.

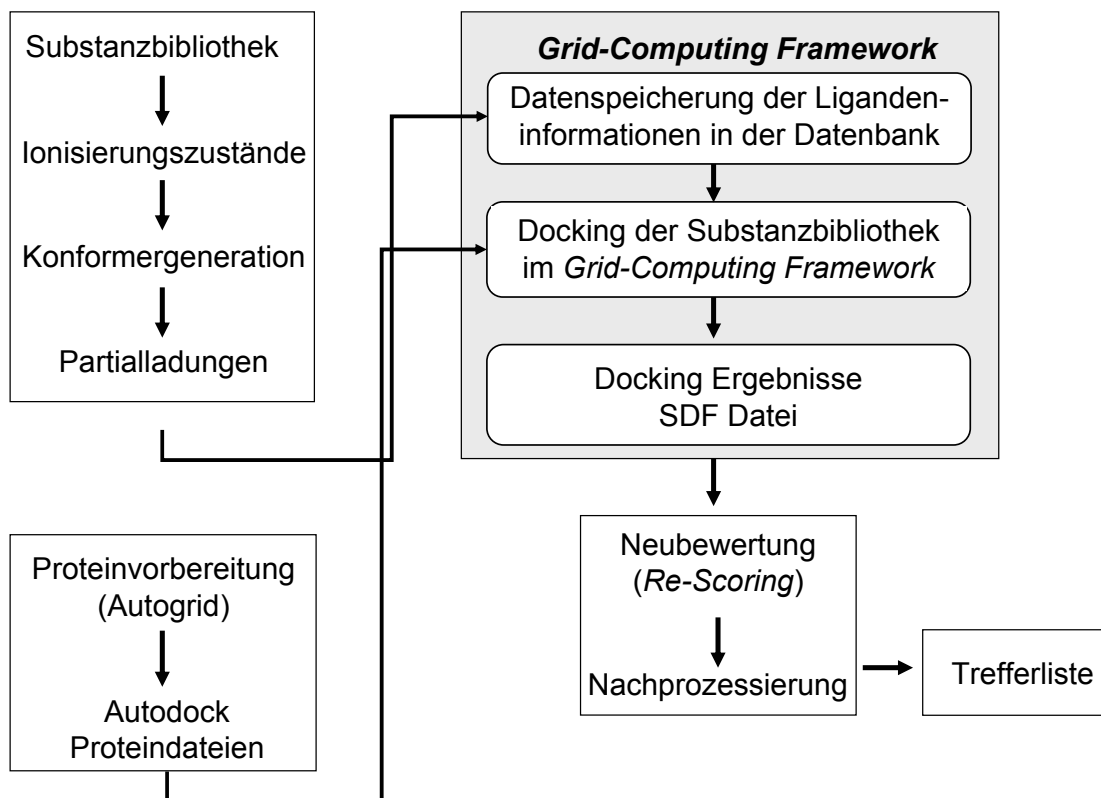


Abb. 3.3 Flussdiagramm des *Docking*-Prozesses. Die Substanzbibliothek wird für das *Docking* vorbereitet. Dafür werden Ionisierungszustände und eine Startkonformation der Strukturen berechnet. Die Ligandatome werden anschließend mit Partialladungen versehen. Die Proteindateien werden mit der Software Autodocktools und Autogrid vorbereitet. Die so aufbereiteten Dateien werden an das *Grid-Computing Framework* übergeben. Das *Framework* steuert den *Grid-Computing* Prozess und die *Dockings*. Die Liganden in der Ergebnisdatei werden anschließend neu bewertet (*Re-scored*) und nachprozessiert. Die so gefundenen Strukturen werden danach visuell untersucht.

In einem weiteren Schritt werden die für den jeweiligen Liganden benötigten Autodock *Docking*-Parameter-Dateien generiert. Diese Dateien werden ebenfalls in der Datenbank abgelegt. Anschließend werden alle Dateien, die für eine Initialisierung des *Grid-Computings* benötigt werden, von dem *Framework* bereitgestellt.

Das *Framework* überwacht den kompletten Prozess des *Grid-Computings* und des *Dockings* und wertet die von den *Client-Computern* eintreffenden Ergebnisdateien automatisch aus. Dabei werden aus den Atomkoordinaten der Liganden und den in der Datenbank abgelegten Strukturinformationen die ursprünglichen Molekülbindungen und die Formalladungen der Atome der Strukturen wiederhergestellt. Diese Informationen werden zusammen mit den Bindungsstärke-Bewertungen (*Docking-Scores*) im MDL Struktur-Daten-Format (MDL-SDF-Datei) gespeichert [[HTTP://WWW.MDLI.COM](http://www.mdli.com)]. Das ermöglicht die Weiterbearbeitung mit den gängigen Softwarepaketen (Abb. 3.3).

Das *Framework* überwacht den Prozess des *Grid-Computings* so lange, bis die CTX-Datei vollständig abgearbeitet ist. Autodock ist als eigenständige *Docking*-Applikation nicht geeignet, sequenziell mehrere Strukturen nacheinander automatisch zu docken und eine gemeinsame Datei, die die *Docking*-Ergebnisse enthält, zu erstellen. Durch die Speicherung von Informationen in der Datenbank und die sequenzielle Bearbeitung der CTX-Datei, wurde es möglich, große Datensätze parallel zu prozessieren. Weitere Funktionen des Systems beinhalten die Identifizierung von Benutzern und deren automatische Anmeldung an der Datenbank. Eine Benutzerverwaltung erlaubt eine gleichzeitige Verwendung des *Grid-Computing Frameworks* von mehreren Benutzern.

3.2 Validierung der Methode

3.2.1 Reproduzierbarkeit von Protein-Ligand-Komplexen

Da bisher keine Erfahrungen mit einer Windows-Version von Autodock vorlagen, wurde diese anhand von ausgewählten CDK-Inhibitor-Komplexen aus der PDB validiert [BERMAN *et al.*, 2002]. Die Strukturen wurden nach verschiedenen Kriterien ausgewählt: Einerseits mussten sie eine Auflösung von

PDB Acc. No.	RMSD
1AQ1	1,0
1DI8	1,7
1E1V	1,1
1GZ8	1,4
1H01	4,3
1JSV	1,8
1JVP	0,8
1KE9	2,7
1OIQ	4,1
1OIR	3,6
1OIU	2,9
1PXI	0,6
1R78	3,1
1URW	4,0
1V0O	0,4

Tab. 3.1 RMSD Werte der Ligand-Posen nach einem *Re-Docking* in die Proteinbindetasche. Es ist jeweils der niedrigste RMSD Wert einer Ligandkonformation (Ligand-Pose) eines Referenzliganden dargestellt.

mindestens 0,25 nm (2,5 Å) besitzen und andererseits mit einem Inhibitor co-kristallisiert vorliegen. Heterodimere CDK-Cyclin Komplexe und Komplexe mit Peptidinhibitoren, sowie Substraten wurden ebenfalls entfernt. Die Aminosäuren in der Bindetasche sollten vollständig kristallographisch aufgelöst sein.

Danach wurden die co-kristallisierten Liganden aus der Proteinbindetasche entfernt und für ein *Re-Docking* in die Bindetasche des Proteins präpariert. Die Bindungen der Ligandatome wurden neu gezeichnet und eine Startkonformation berechnet [TETKO *et al.*, 2005; SADOWSKI, 1997]. Mit diesen Liganden wurde ein *Re-Docking* in die jeweilige Bindetasche durchgeführt (Abb. 3.4). Als Maß der Genauigkeit des *Dockings* wurden *root mean square deviations* (RMSD) der Schweratome des eingepassten Liganden und der Ursprungskonformation des Liganden berechnet (Tab. 3.1).

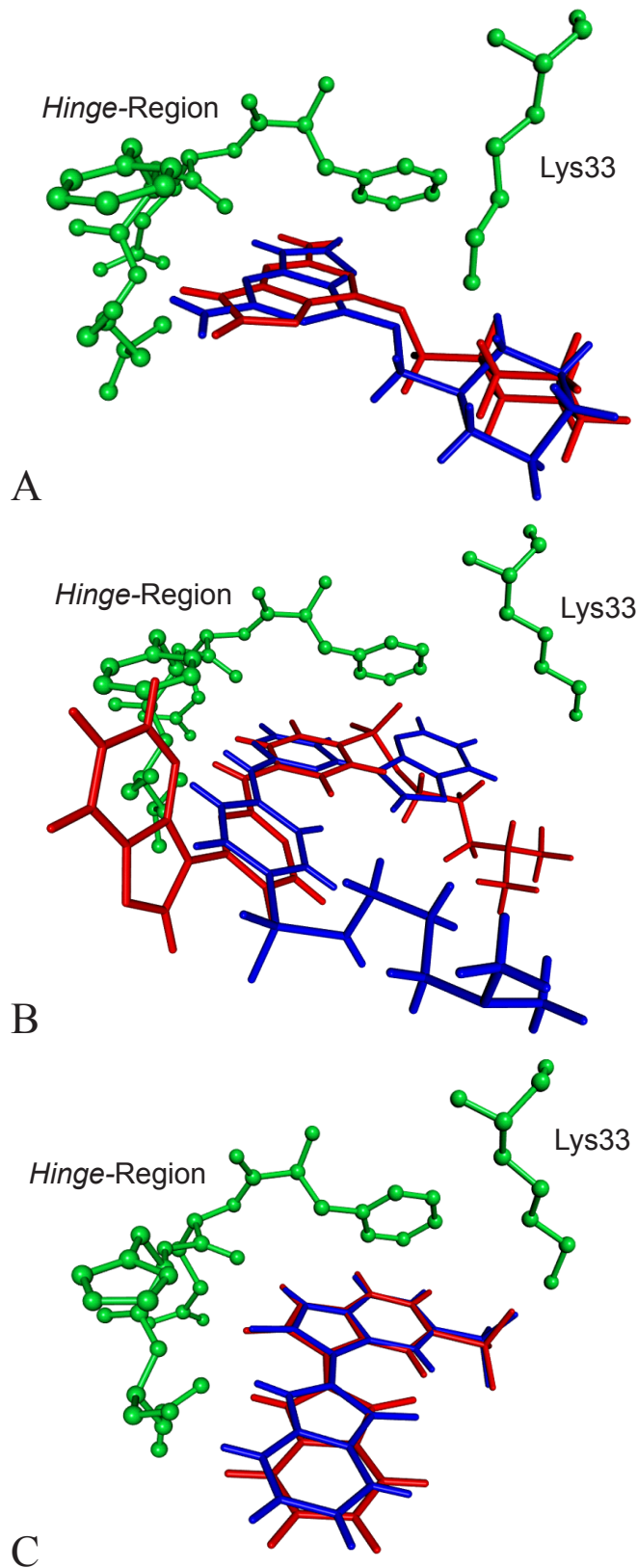


Abb. 3.4 Vergleich der jeweils besten Pose (rotes Molekül) eines Liganden mit der zugehörigen experimentell ermittelten Konformation (blaues Molekül). Abb. 3.4. A: PDB Acc. Nr. 1E1V, RMSD: 0,11 nm (1,1 Å). Abb. 3.4. B: PDB Acc. Nr. 1URW, RMSD: 0,4 nm (4,0 Å). Abb. 3.4. C: PDB Acc. Nr. 1V0O, RMSD: 0,04 nm (0,4 Å). In Abhängigkeit von der verwendeten Proteinstruktur und dem verwendeten Liganden sind die *Dockings* von unterschiedlicher Qualität. Ein Grund kann eine fehlerhafte Proteinstruktur sein, ein weiterer Grund kann eine ungenügende Konformer-Generation während des *Dockings* sein.

Die RMSD-Werte sind ein Maß der Abweichung von der eingepassten Struktur zu der experimentell ermittelten Struktur. Als ein ausreichend genaues *Docking* wurde ein RMSD-Wert von maximal 0,2 nm (2 Å) angenommen [FERRARA *et al.*, 2004]. In den Experimenten wurden 15 Protein-Ligand-Komplexe verwendet. Davon konnten 8 Ligand-Konformationen mit einem RMSD von unter 0,2 nm (2 Å) reproduziert werden (Tab. 3.1). Für diese Experimente wurden **Autodock-Parameter verwendet, die eine schnelle Berechnung für einen einzelnen Liganden erlauben** (Kap. 2.6.7). Diese Parameter beeinträchtigen jedoch die Genauigkeit des *Dockings* von Liganden mit vielen Rotationsfreiheitsgraden [SOTRIFFER & DRAMBURG, 2005].

Um zu überprüfen, ob die Protein-Ligand-Komplexe, die nicht erfolgreich reproduziert werden konnten, generell für das *Re-Docking* geeignet sind, wurden weitere Versuche mit anderen Parametern durchgeführt. Diese Parameter wurden so gewählt, dass ein ausreichend hohes Maß an Genauigkeit erreicht werden sollte (Kap. 2.6.7). Von den 7 Protein-Ligand Komplexen konnten auf diese Weise 3 Komplexe mit einem RMSD Wert von unter 0,2 nm (2 Å) reproduziert werden. Aufgrund der sehr langen Prozessierung der *Dockings* wurden diese Bedingungen nicht im virtuellen Screening verwendet.

Das erfolgreiche *Re-Docking* mit diesen Bedingungen lässt jedoch den Schluss zu, dass vier der verwendeten Proteinstrukturen generell mit der verwendeten Software nicht erfolgreich reproduziert werden können. Dies könnte auf Fehler in den Proteinstrukturen zurückzuführen sein [SOTRIFFER & DRAMBURG, 2005]. Die gewählten Parameter stellen jedoch offensichtlich einen guten Kompromiss zwischen der Genauigkeit und der Geschwindigkeit des *Dockings* dar.

3.2.2 Evaluierung verschiedener Bewertungsfunktionen

Die 15 Proteinstrukturen aus Kap. 3.2.1 bildeten darüberhinaus den Datensatz für Anreicherungsexperimente mit bekannten CDK-Inhibitoren. Für diese Experimente wurden 2 zusätzliche Datensätze generiert: Ein Datensatz enthielt die Liganden, die aus den 15 Proteinstrukturen entfernt wurden. Dieser wurde mit 29 Inhibitoren der CDKs erweitert, die in der Literatur beschrieben sind (CDK-Inhibitoren). Von diesen Liganden liegen lediglich Aktivitätsbestimmungen vor und sie wurden so gewählt, daß sie die humane CDK2 mit einem IC_{50} -Wert, der in einem nanomolaren Bereich liegt, inhibieren. Für die Experimente wurde eine Auswahl verschiedener Strukturklassen gewählt.

Der zweite Datensatz wurde aus kommerziell erhältlichen Strukturdatenbanken zusammengestellt (*Kinase focussed library* [Asinex; Enamine; ChemDiv; LifeChemicals; Otava]), die für einen Einsatz im Screening mit Proteinkinasen ausgerichtet worden sind (Kinase-Datensatz). Von diesem Datensatz wurden bekannte Inhibitoren und speziell für das Screening von CDK-Inhibitoren ausgerichtete Strukturen entfernt. Dieser Datensatz sollte im Idealfall in Bezug auf seine chemischen Eigenschaften möglichst ähnlich zu den bekannten Inhibitoren der CDKs sein, aber keine tatsächlichen Inhibitoren der CDKs enthalten. Beide Datensätze wurden vereint und ergaben zusammen genau 1000 Do-

Bewertungsfunktion	Anzahl der Liganden
Autodock	5
Chemscore	6
D-Score	4
G-Score	3
PMF (Sybyl)	4
Jain	6
Ligscore 1	3
Ligscore 2	2
Ludi	3
PLP 1	5
PLP 2	5
PMF (Cerius 2)	5

Tab. 3.2 Anzahl der Liganden, die mit verschiedenen Bewertungsfunktionen mit einem RMSD unter 0,2 nm (2 Å) identifiziert werden konnten. Dabei konnten mit den Bewertungsfunktionen Chemscore und Jain am häufigsten korrekt eingepasste Konformationen eines Liganden ermittelt werden.

cking-Strukturen. Mit diesem Datensatz wurden virtuelle Screenings an den 15 Proteinstrukturen durchgeführt. Anschließend wurden die in die Bindetaschen eingepassten Strukturen mit verschiedenen Bewertungsfunktionen erneut bewertet. Mit diesen Daten konnte überprüft werden, wie verlässlich die unterschiedlichen Bewertungsfunktionen in Kombination mit dem verwendeten *Docking*-Algorithmus geeignet sind, um Strukturen von Inhibitoren von unwirksamen Substanzen zu unterscheiden [FERRERA *et al.*, 2004].

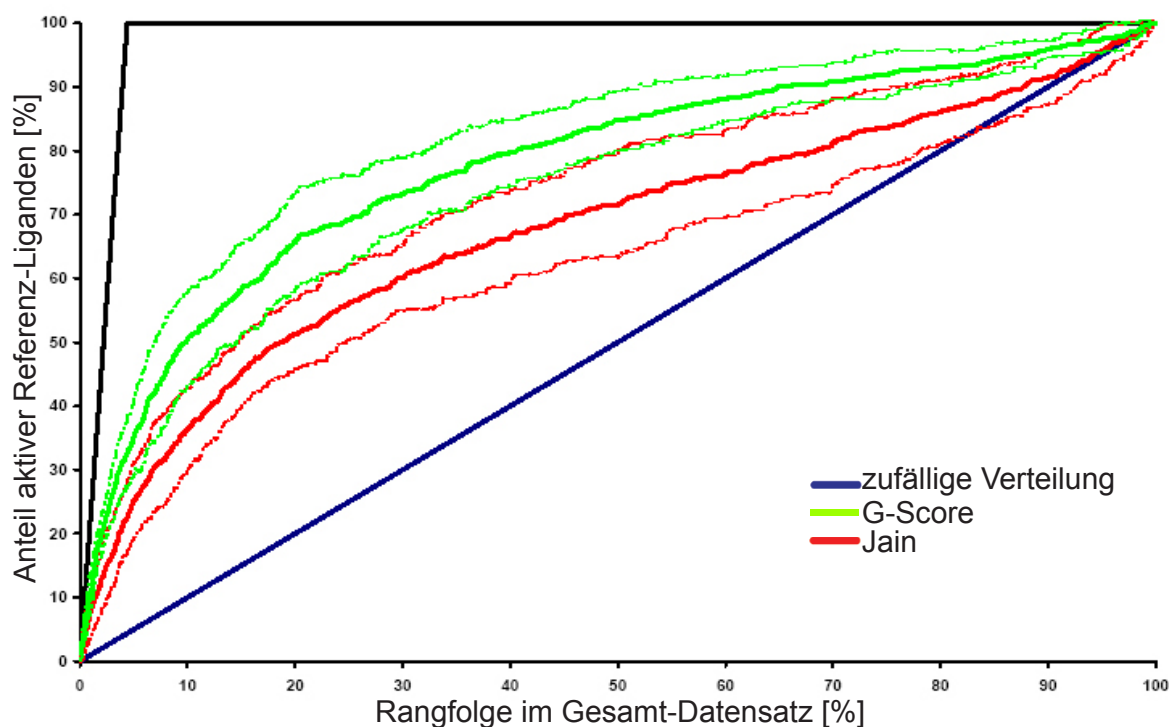


Abb. 3.5 Vergleich der Anreicherungskurven verschiedener Bewertungsfunktionen für die Ermittlung der besten eingepassten Ligandkonformation (erste Bewertung). Die Kurven wurden mit den Mittelwerten der *Scoring*-Werte der 15 Proteinstrukturen ermittelt. Zum Vergleich sind die Konformations-bestimmenden Bewertungsfunktionen (erste Bewertungsfunktion) gewählt, die eine maximale Anreicherung (G-Score: grüne Kurve) oder eine akkurate Ermittlung der korrekten Ligandkonformation (Jain: rote Kurve) ermöglichen. Die gestrichelten Linien stellen die jeweiligen Standardabweichungen dar.

Ausgehend von der Annahme, dass eine hinreichend genaue Platzierung von bindenden Molekülen in einem Datensatz von nicht-bindenden Molekülen nur mit einer korrekt vorhergesagten Ligand-Konformation mit der richtigen Orientierung in der Bindetasche (Ligand-Pose) möglich ist, wurde eine Zwei-stufen-Bewertung eingeführt [YOON *et al.*, 2004].

Im *Docking* wurden für jede Struktur 10 Ligand-Posen in der Bindetasche generiert. Danach wurde in der ersten Stufe mit einer Bewertungsfunktion die Ligand-Pose des Liganden ermittelt, die -ein korrektes *Docking* voraus-

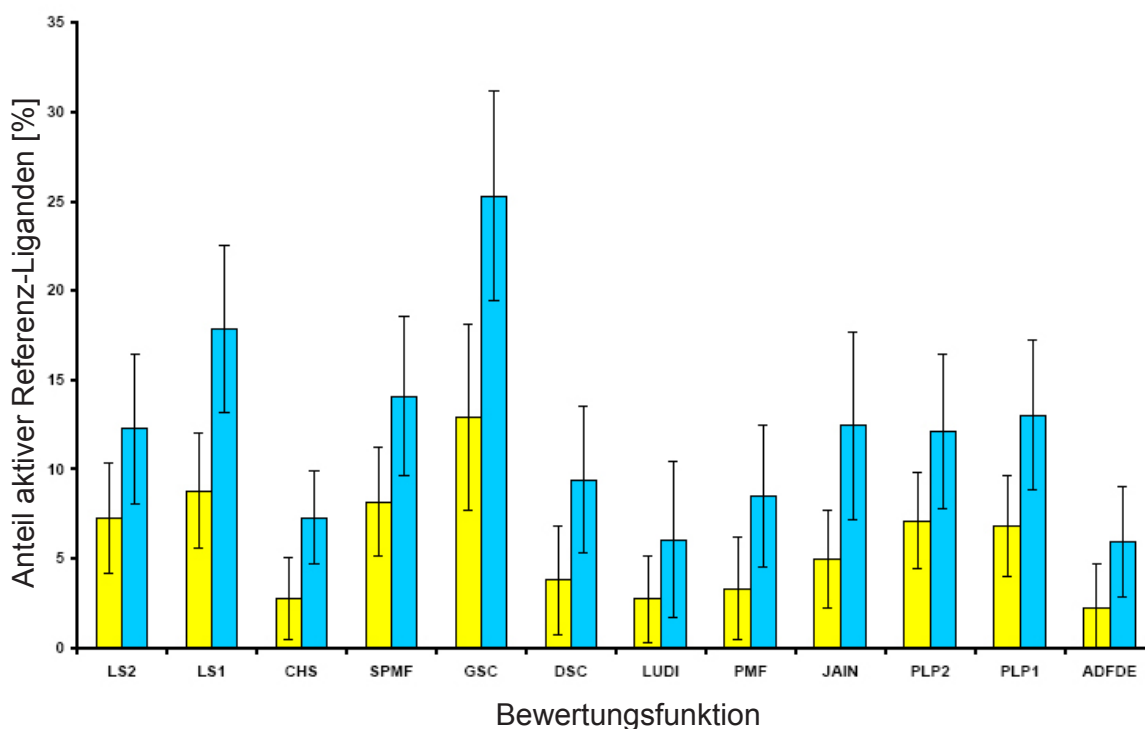


Abb. 3.6 Anreicherungen nach der Rangfolgen-Bewertung. Die Bewertungsfunktion Jain wurde für die Bestimmung der günstigsten Ligand-Konformation verwendet (1. Bewertungsschritt). Auf der Ordinate ist die jeweilige Rangfolgen-Bewertungsfunktion dargestellt. Die Abszisse bezeichnet die prozentuale Anteile von bekannten Inhibitoren in einer Rangfolge. Die gelben Balken beziehen sich auf die bestbewerteten 2 % der Rangfolge, die blauen Balken beziehen sich auf die bestbewerteten 5 % der Rangfolge. Sowohl im Bereich der bestbewerteten 2 %, als auch im Bereich der bestbewerteten 5 % der Rangfolge, übertraf die Bewertungsfunktion G-Score die anderen Bewertungsfunktionen hinsichtlich der Anreicherung von bekannten Inhibitoren. LS2: Ligscore2, LS1: Ligscore1, CHS: Chemscore, SPMF: PMF (Sybyl), GSC: G-Score, DSC: D-Score, LUDI: Ludi, PMF: PMF, JAIN: Jain, PLP2: PLP2, PLP1: PLP1, ADFDE: Autodock

gesetzt- der bioaktiven Konformation in der Bindetasche entspricht. Bei den hier verwendeten Proteinstrukturen erwiesen sich die Bewertungsfunktionen Jain und Chemscore als geeignet, um die beste Ligand-Pose zu ermitteln (Tab. 3.2). Im Kontrast zu diesem Ergebnis konnte mit dem Einsatz der Bewertungsfunktion G-Score zwar eine höhere Anreicherung erzielt werden, was jedoch zu falsch vorhergesagten Ligand-Posen führte (Abb. 3.5).

Von den verbleibenden einzelnen Konformeren jeder Struktur wird mit einer zweiten Bewertungsfunktion eine Rangfolge ermittelt. Abb. 3.6 zeigt die Anreicherungsraten der bekannten CDK-Inhibitoren in den oberen 2 %, bzw. 5 % der Rangfolge als Mittelwerte aller verwendeten Proteinstrukturen. Da keine Bewertungsfunktion für alle Proteinbindetaschen unterschiedlicher Proteinfamilien gleichermaßen geeignet ist, mussten mehrere Bewertungsfunktionen erprobt werden. Dabei waren starke Unterschiede zu beobachten, die die Problematik der Bewertungsfunktionen offenbaren. In Kombination mit den Bewertungsfunktionen Jain oder Chemscore als diejenigen Bewertungsfunktionen, die die bioaktive Konformation korrekt vorhersagen, erwies sich die Bewertungsfunktion G-Score für die nachfolgende Rangfolgen-Bewertung als besonders geeignet. Erwartungsgemäß waren zusätzlich Unterschiede in Abhängigkeit von der verwendeten Proteinstruktur erkennbar.

Wurde mit der Bewertungsfunktion Jain im ersten Schritt und G-Score im zweiten Schritt bewertet, so zeigten die zugrundeliegenden Proteinstrukturen deutliche Unterschiede hinsichtlich der Anreicherung von bekannten Inhibitoren. Abb 3.9 stellt die Anreicherungen von Referenz-Liganden in Abhängigkeit der verwendeten Proteinstrukturen graphisch dar. Insbesondere die Strukturen mit den PDB *Acc. No.* 1E1V und 1OIR zeigten mit dem verwendeten Datensatz sehr gute Anreicherungen.

3.2.3 Datensatzabhängigkeit der Anreicherungsexperimente

Ein weiterer Aspekt in der Beurteilung von Anreicherungsexperimenten ist der zugrundeliegende Datensatz. Um zu überprüfen, inwieweit die Anreicherung von dem verwendeten Datensatz abhängt, wurden 3 zusätzliche Datensätze generiert. Die bekannten Inhibitoren waren in allen Datensätzen identisch. Die Strukturen von Substanzen unbekannter Aktivität wurden für die zusätzlichen Datensätze aus unterschiedlichen Datenbanken entnommen.

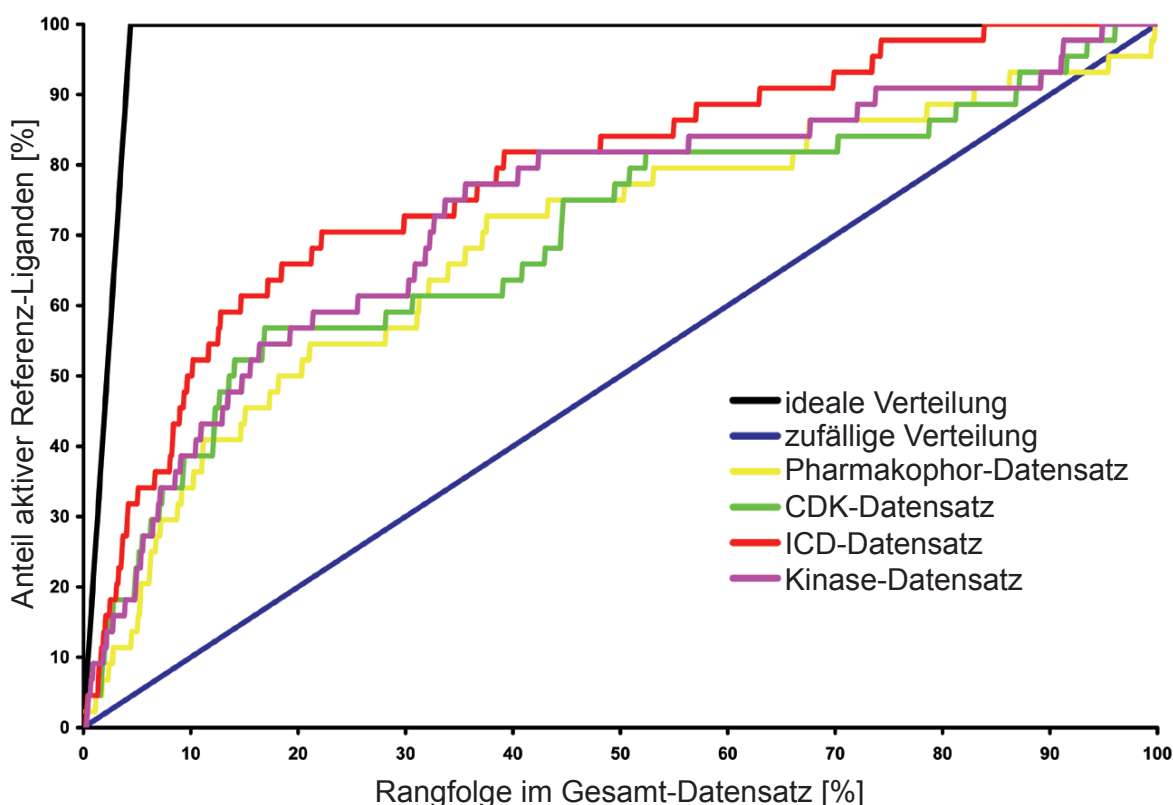


Abb. 3.7 Abhängigkeit der Anreicherung aktiver Referenz-Liganden von dem zugrundeliegenden Test-Datensatz. Mit einem Datensatz der aus bekannten CDK-Inhibitoren und zufällig gewählten Strukturen besteht, lassen sich die höchsten Anreicherungen erzielen. Mit den Datensätzen, deren unbekannte Strukturen auf das *Drug-Target* optimiert sind, sind geringere Anreicherungen möglich. Daher ist die Anreicherung ein Maß der erfolgreichen Optimierung einer Strukturdatenbank. Offensichtlich hat die Optimierung mit den Pharmakophormodellen die besten Resultate erzielt.

Einer dieser Datensätze beinhaltete eine zufällige Auswahl von 956 Strukturen, wie sie in der Strukturdatenbank der Intervet Innovation GmbH (*Intervet Chemicals Directory*, ICD) vorkommen (ICD-Datensatz). Ein weiterer Datensatz bestand ebenfalls aus 956 Strukturen, die speziell für den Einsatz im Screening potenzieller Inhibitoren von CDKs bestimmt sind (CDK-Datensatz). Als dritter Datensatz wurde ein Datensatz mit 956 Strukturen verwendet, wie sie durch die Datenbanksuchen in Kap. 3.6.1.3 mit den Pharmakophormodellen herausgefunden wurden (Pharmakophor-Datensatz). Alle Datensätze wurden zufällig aus den größeren Datenbanken zusammengestellt. Zum Vergleich wurden die Anreicherungen mit dem Kinase-Datensatz aus Kap. 3.2.2 herangezogen. Erwartungsgemäß konnten Unterschiede der Anreicherungen mit den verschiedenen Datensätzen beobachtet werden. Abb. 3.7 zeigt die Anreicherungskurven der unterschiedlichen Datensätze im Vergleich. Alle Anreicherungen wurden mit der optimalen Kombination von Jain als erster und G-Score als zweiter Bewertungsfunktion ermittelt (Kap. 3.2.2). Die Anreicherungskurve des Kinase-Datensatzes nimmt dabei einen ähnlichen Verlauf, wie ihn die Anreicherungskurve des ICD-Datensatzes aufweist. Die Anreicherungskurve des CDK-Datensatzes nimmt im Ursprung der Kurve einen ähnlichen Verlauf wie die des Kinase-Datensatzes und des ICD-Datensatzes, weist jedoch im weiteren Verlauf eine niedrigere Anreicherungsrate auf.

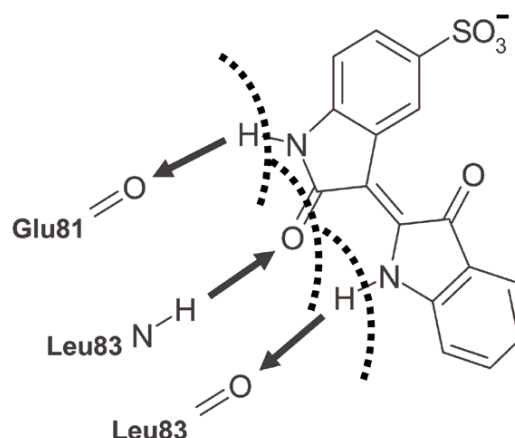


Abb. 3.8 Schema der Funktionsweise des Interaktionsfilters. Als Grundbedingung für das Passieren des Filters muss ein Wasserstoffbrückenakzeptor einer Ligand-Pose in Reichweite des Wasserstoffbrückendonors des Leu83 der Hinge-Region vorhanden sein. Zusätzlich muss mindestens ein Wasserstoffbrückendonor in Reichweite des Glu81 oder des Carbonylsauerstoffs des Leu83 sein. Als Mindestreichweite wurde 30 nm (3 Å) festgelegt.

Einzig die Anreicherungskurve des Pharmakophor-Datensatzes wies einen allgemein flacheren Kurvenverlauf im Bereich der oberen 30 % der Scoringwerte als die anderen Datensätze auf.

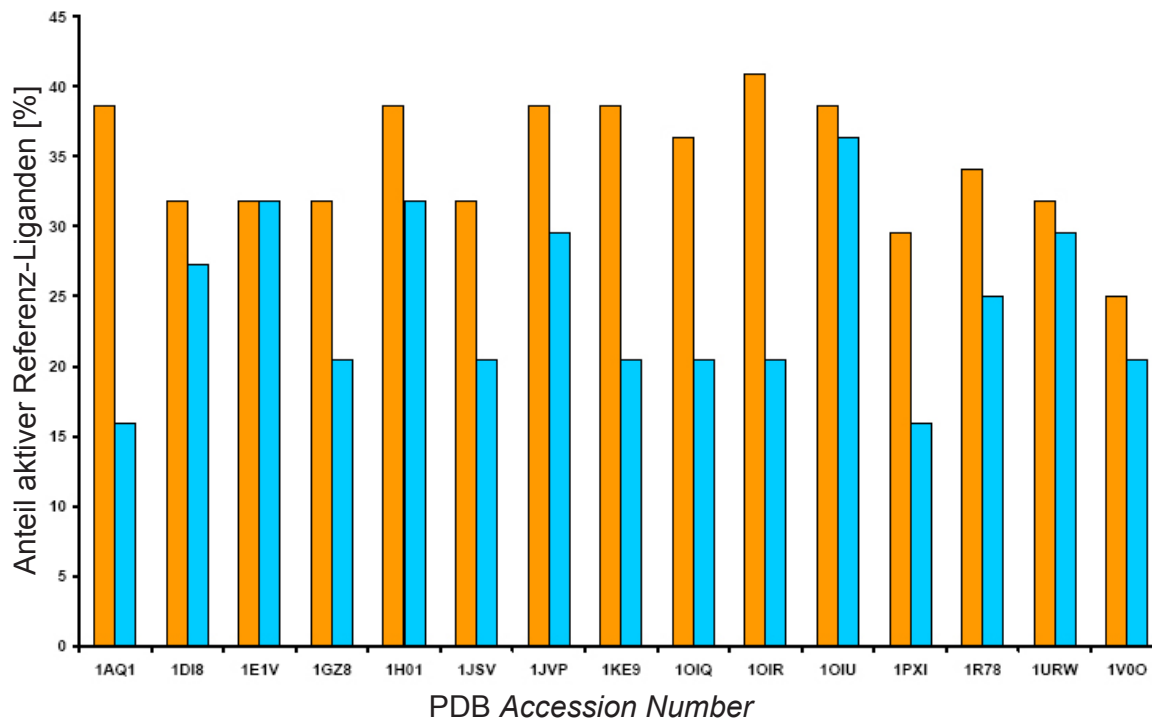


Abb. 3.9 Prozentuale Anreicherung im *Docking* bekannter Inhibitoren unter Betrachtung verschiedener Proteinstrukturen. Für die Betrachtung wurden die besten 5% des Kinase-Datensatzes verwendet. Die orange gefärbten Balken stellen die Anreicherung aktiver Referenz-Liganden unter Nutzung des Interaktionsfilters dar. Die hellblauen Balken stellen die Anreicherung aktiver Referenz-Liganden ohne Nutzung des Interaktionsfilters dar.

3.2.4 Nachprozessierung der *Docking*-Strukturen

Für eine weitere Verbesserung der Anreicherungen von Referenz-Liganden wurde über den in Kap. 3.2.3 beschriebenen Ansatz hinaus eine weitere Methode der Nachprozessierung erprobt. Dazu wurde ein Cactvs-Skript erstellt, das die eingepassten Ligand-Posen heraussucht, die potenziell mit der *Hinge*-Region des Proteins interagieren können. Dabei wird eine Kugel mit einem Radius von 0,3 nm (3 Å) um die Donor- und Akzeptoratome der *Hinge*-Region angenommen. Befindet sich ein passendes Akzeptor- oder Donoratom einer Ligand-Pose in einer der Kugeln, wird diese Ligand-Pose beibehalten.

Bewertungsfunktion	Anzahl der Liganden
Autodock	5
Chemscore	6
D-Score	5
G-Score	4
PMF (Sybyl)	4
Jain	6
Ligscore 1	4
Ligscore 2	3
Ludi	5
PLP 1	4
PLP 2	3
PMF (Cerius2)	5

Tab. 3.3 Anzahl der eingepassten CDK-Inhibitoren mit einem RMSD Wert unter 0,2 nm (2 Å), die mit verschiedenen Bewertungsfunktion und unter Anwendung des Interaktionsfilters gefunden wurden. Von den 15 bekannten Protein-Ligand Komplexen konnten maximal 6 mit einem akzeptablen RMSD-Wert gefunden werden.

Dadurch wurde die Anzahl der Ligand-Posen der *Docking*-Datensätze auf ca. 5 % des Ursprungs-Datensatzes reduziert. Die verbleibenden Ligand-Posen der Einzel-Strukturen wurden anschließend dem in Kapitel 3.2.2 beschriebenen Nach-Bewertungsprozess unterzogen. Abb. 3.8 zeigt schematisch die Funktionsweise des Interaktionsfilters. Der Einsatz dieses Filters ermöglicht deutlich höhere Anreicherungsraten von bioaktiven Ligand-Konformationen (Abb. 3.9).

Zusätzlich wurde dieser Filter durch vergleichende RMSD-Bestimmungen erprobt, um zu gewährleisten, dass die bekannten Inhibitoren den Filter in einer bioaktiven Konformation passieren. Nach Einsatz des Filters wurden die Liganden der

PDB-Strukturen 1DI8 und 1JSV nicht mehr korrekt vorhergesagt (Tab. 3.3). Eine Ursache dafür ist die Eigenschaft dieser Inhibitoren, lediglich eine der Wasserstoffbrückenbindungen zu den Aminosäuren Leu83 und Glu81 herzustellen. Alle anderen Ligand-Posen wurden nach wie vor korrekt aus den *Docking*-Daten extrahiert. Nach Anwendung des Filters waren alle Bewertungsfunktionen besser geeignet, zwischen den korrekt eingepassten Liganden und den fehlerhaften *Dockings* zu unterscheiden. In Kombination mit diesem Filter waren wiederum die Bewertungsfunktionen Jain und Chemscore besser geeignet, alle korrekt eingepassten CDK-Inhibitoren unter den 10 angebotenen Lösung herauszufinden.

3.3 Sequenzsuche und bioinformatische Tiefenanalyse

3.3.1 Sequenzsuche nach CDK-Orthologen in *E. tenella*

Für die Sequenzsuche wurde eine cDNA Sequenz einer CDK aus *Toxoplasma gondii* (TgCRK2 ; *Acc. No.*: AJ001085), verwendet. Mit der Sequenz wurde eine BLAST-Suche auf verschiedenen *E. tenella* Datenbanken durchgeführt ([HTTP://WWW.SANGER.AC.UK/CGI-BIN/BLAST/SUBMITBLAST/E_TENELLA/OMNI](http://www.sanger.ac.uk/cgi-bin/blast/submitblast/e_tenella/omni)) [CHAPMAN *et al.*, 2004]. Dabei konnten zwei unterschiedliche Sequenzen gefunden werden.

Eine dieser Sequenzen ist eine genomische Contig-Sequenz (Contig 2257242.c000902127), die mittels Sequenz-Alignments in Exon- und Intronbereiche unterteilt werden konnte. Aus den Exons konnte eine Peptidsequenz hergeleitet werden. Sequenz-Alignments mit anderen CDKs deuteten darauf hin, dass der C-Terminus der Peptidsequenz nicht vollständig ist. Ein weiterer Hinweis

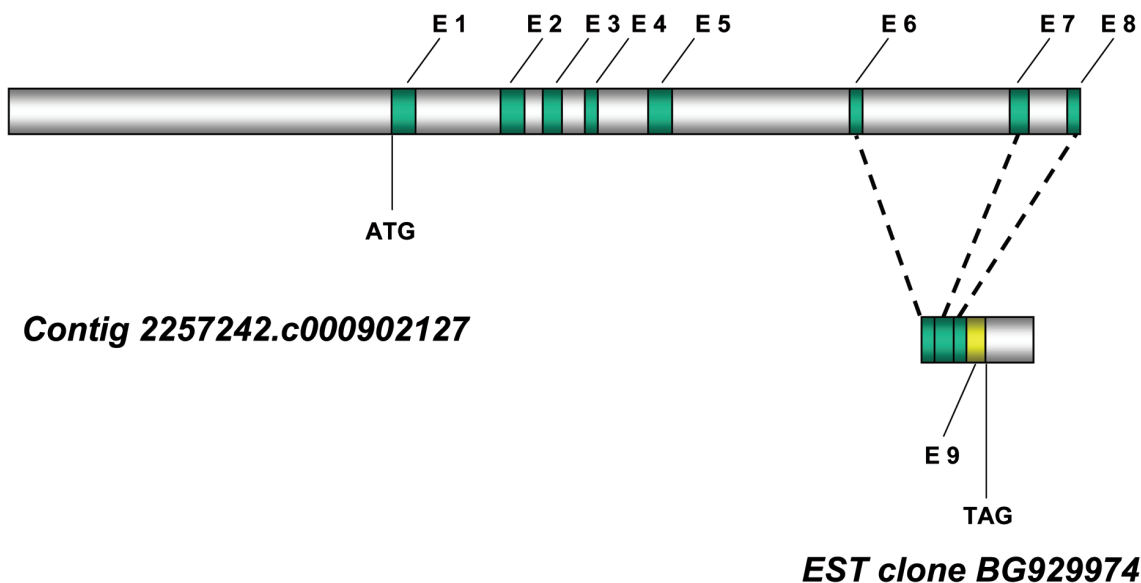


Abb. 3.10 Schema der durch bioinformatische Sequenzsuche gefundenen CDK-Sequenzen aus *E. tenella*. Die genomische Contigsequenz konnte durch Sequenzvergleiche mit bekannten CDKs in Exon- (E1- E8) und Intronbereiche unterteilt werden. Die am C-Terminus unvollständige Contigsequenz konnte durch eine EST-Sequenz komplettiert werden (E9).

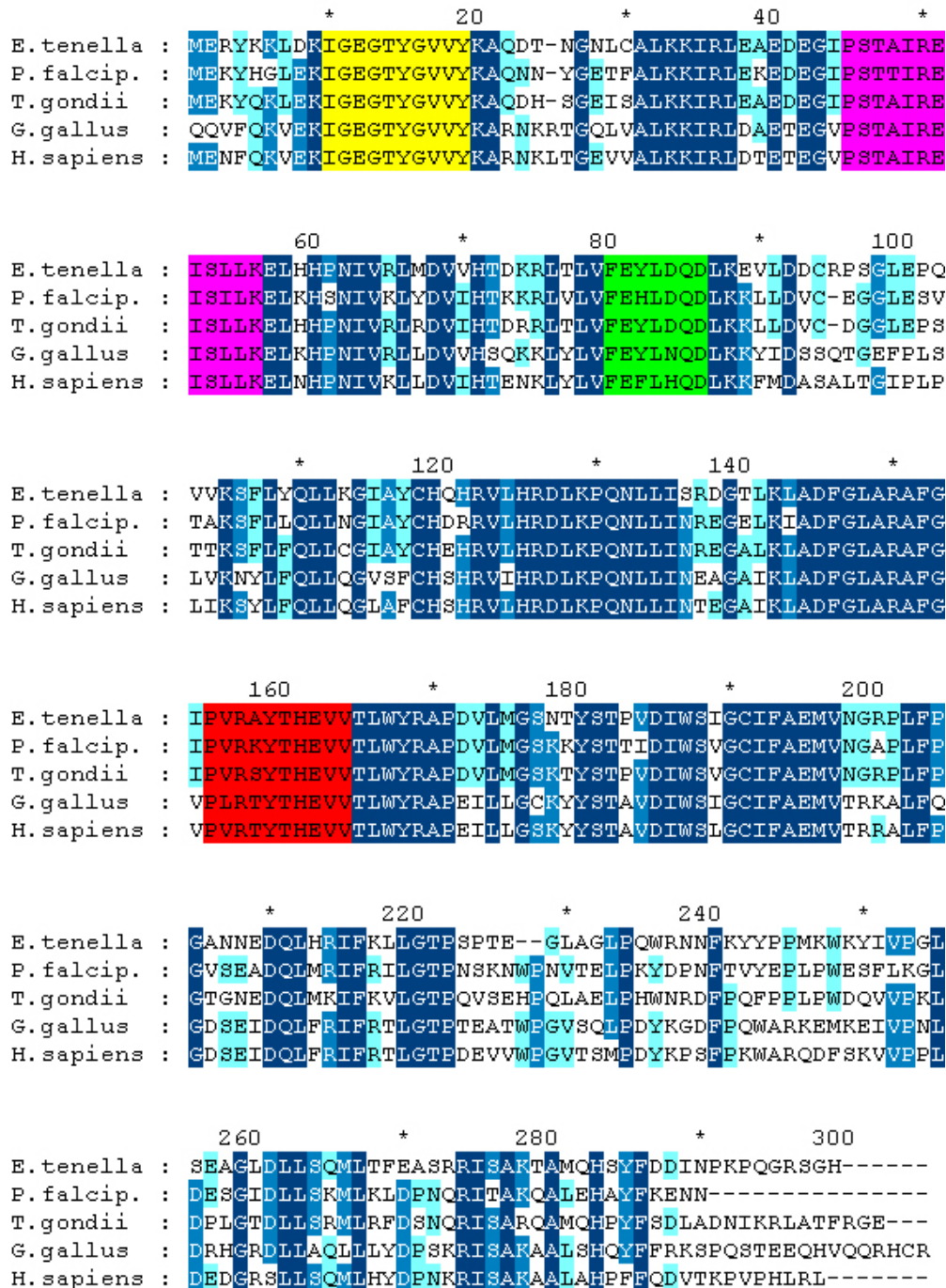


Abb. 3.11 Multiples Sequenz-Alignment von der gefundenen CDK und CDKs aus verschiedenen Organismen. In dem multiplen Sequenz-Alignment sind die Proteine EtCRK2 von *E. tenella*, TgCRK2 aus *T. gondii*, Pfpk5 aus *P. falciparum*, CDK3 aus *G. gallus* und des CDK2 aus *H. sapiens* dargestellt. Die für diese Proteinkinasen wichtigen strukturellen Regionen sind in unterschiedlichen Farben gekennzeichnet (T-Loop, rot; C-Helix, violett; G-Loop, gelb; Hinge-Region, grün). Diese Bereiche sind in allen CDK-Homologen stark konserviert.

dafür war das Fehlen eines Stop-Codons im Leserahmen. Die zweite Sequenz war eine *Expressed-Sequence-Tag* (EST) Sequenz, die am 5`- Ende, also dem N-Terminus der translatierten Sequenz unvollständig war. Beide Sequenzstücke konnten aneinandergesetzt werden, sodass eine Überlappung von ca. 100 Basen resultierte (Abb. 3.10). Durch Sequenz-Alignments und manuelles Herausschneiden der Intronsequenzen konnte eine Peptidsequenz dieser putativen CDK aus *E. tenella* rekonstruiert werden. Diese Sequenz wurde später von KINNAIRD *et al.* mit biochemischen Methoden bestätigt [KINNAIRD *et al.*, 2004]. Die Autoren benannten das Protein EtCRK2.

Um die Sequenz auf eine Ähnlichkeit zu anderen CDKs in verschiedenen Organismen zu überprüfen, wurde eine BLAST-Suche gegen eine nicht-redundante Proteindatenbank (GeneBank) mit der translatierten Sequenz von *E. tenella* durchgeführt [BENSON *et al.*, 2004]. Homologe Proteine konnten dabei nicht nur innerhalb der Gruppe der *Apicomplexa*, sondern beispielsweise auch in verschiedenen Wirbeltieren, Pilzen, Würmern und Pflanzen gefunden werden. Unter anderem konnten orthologe Sequenzen von *P. falciparum*, *T. gondii*, *Gallus gallus* (Haushuhn) und *Homo sapiens* gefunden werden (Tab. 3.4). Dabei erwiesen sich die Sequenzen als relativ gut konserviert. Die Sequenzidentitäten von verschiedenen Organismen reichten von 59 % - 79 %. Zur Untersuchung des Vorkommens von CDKs in weiteren, weniger gut charakterisierten Organismen, wurde eine BLAST-Suche mit der CDK Sequenz aus *E. tenella* gegen die dbEST-Datenbank durchgeführt. Die Datenbank dbEST enthält Sequenzinformationen aus *Expressed-Sequence-Tag* (EST)

Organismus	<i>E. tenella</i>	<i>P. falciparum</i>	<i>T. gondii</i>	<i>G. gallus</i>	<i>H. sapiens</i>
<i>E. tenella</i>		67	76	60	59
<i>P. falciparum</i>			73	62	62
<i>T. gondii</i>				59	59
<i>G. gallus</i>					79

Tab. 3.4 Sequenzidentität [%] zwischen CDK-Proteinen verschiedener Organismen. Mit Hilfe des BLAST-Algorithmus wurden Sequenzvergleiche zwischen CDK-Proteinen aus verschiedenen Organismen durchgeführt. In der Tabelle sind die Sequenzidentitäten der CDK2-Sequenz (As 1-296) von *E. tenella* mit den orthologen Proteinsequenzen aus *P. falciparum*, *T. gondii*, *G. gallus* und *H. sapiens* dargestellt.

Projekten [SELZER *et al.*, 2004]. Die BLAST-Analyse ergab Treffer für EST-Sequenzen von *Bos taurus* (Hausrind), *Drosophila melanogaster* (Fruchtfliege), *G. gallus*, *H. sapiens*, *Mus musculus* (Maus), *Sus scrofa* (Wildschwein), *Xenopus laevis* (Krallenfrosch) und einigen Pflanzenarten. Zwischenzeitlich wurden in *E. tenella* zwei weitere, putative CDKs mittels bioinformatischer Methoden identifiziert [ENGELS *et al.*, unveröffentlichte Daten].

3.3.2 Proteinmotive

Für die Tiefenanalyse wurden die Proteinsequenzen der CDK einer Proteinmotiveanalyse unterzogen. Dazu wurde die Sequenz der EtCRK2 mit dem Programm Interproscan der Datenbank Interpro analysiert [MULDER *et al.*, 2005]. Der Interproscan ermittelte eine Serin-/ Threonin-Proteinkinase Domäne, sowie eine katalytisch aktive Kinase-Domäne. Ähnliche Ergebnisse wurden mit den Proteinen aus *P. falciparum* (PfPK5) und *T. gondii* (TgCRK2) erzielt. Die Sequenzen besitzen mehrere stark konservierte Regionen im aktiven Zentrum. Hierzu zählen die PSTAIRE-Box (C-Helix), die *Hinge*-Region, der *T-Loop* und ein Glycin-reiches Sequenzmotiv (*G-Loop*) (Abb. 3.11). Eine zentrale Rolle spielt hierbei die konservierte Aminosäure Asp145, die für die katalytische Aktivität von entscheidender Bedeutung ist. Diese gehört zum hochkonservierten DFG-Motiv, das die Aminosäuresequenz Asp-Phe-Gly beschreibt und eine wichtige Rolle bei der Aktivierung der CDKs spielt (Kap. 1.5.1.2).

3.3.3 Röntgenkristallstrukturen der CDKs

Von *P. falciparum* sind derzeit 4 Röntgenkristallstrukturen eines CDK-Homologen (PfPK5) in der PDB hinterlegt [PDB Acc. No.: 1OB3, 1V0B, 1V0O und 1V0P; HOLTON *et al.* 2004]. In der Regel sind die CDK Strukturen Holostrukturen, dh. sie wurden mit verschiedenen Inhibitoren und teilweise mit Peptidketten der Substrate co-kristallisiert. Teilweise liegen die Strukturen

als Heterodimer mit einem gebundenem Cyclin vor und zeigen die aktivierte Form der Proteinkinase. Vollständig aktivierte Proteinkinasen, die an der Aminosäure Thr160 phosphoryliert sind, liegen ebenfalls als PDB Einträge vor [PDB Acc. No.: 1H1P, 1H1Q, 1H1R und 1H1S, DAVIES *et al.*, 2002]. Die ältesten in der PDB abgelegten Strukturen der CDK2 von *H. sapiens* wurden 1996 veröffentlicht [PDB Acc. No.: 1HCK und 1HCL, SCHULZE-GAHMEN *et al.* 1995]. Diese Strukturen besitzen eine Auflösung von 0,18 nm (1,8 Å) und 0,19 nm (1,9 Å) (1HCK und 1HCL). In 1HCK ist der natürliche Ligand ATP

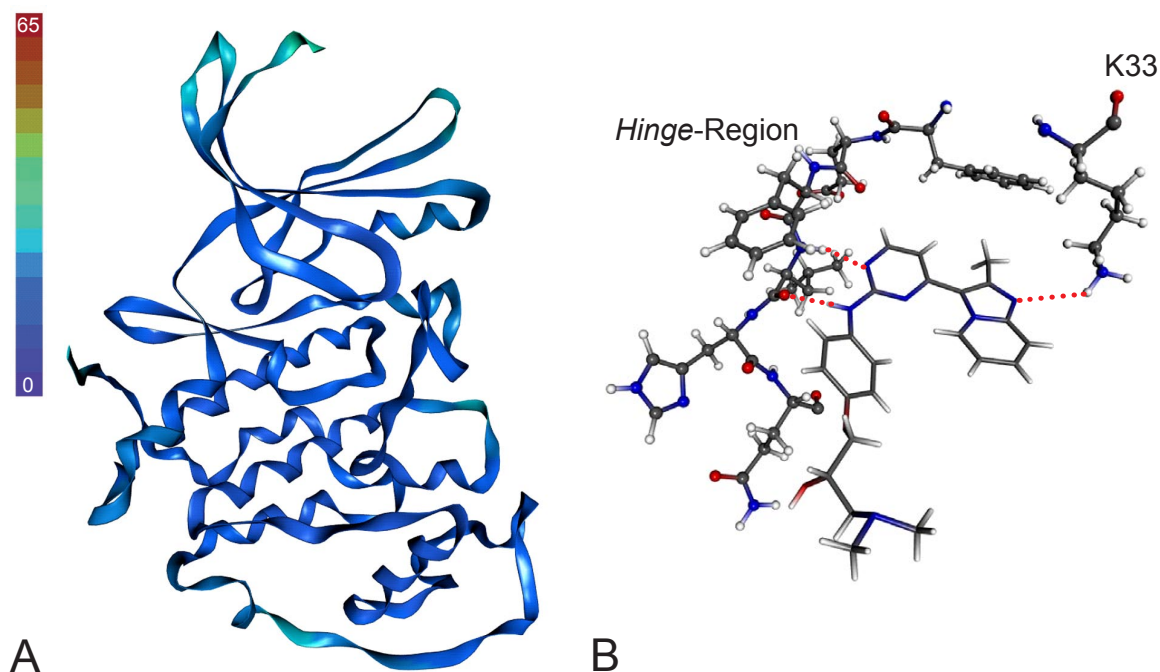


Abb. 3.12 Ribbon-Darstellung einer röntgenkristallographisch ermittelten Proteinstruktur der humanen CDK2. Die Strukturen sind nach Bereichen mit unterschiedlichen Debye-Waller Temperaturfaktoren (B-Faktoren) eingefärbt. Ein hoher Temperaturfaktor spiegelt eine große Beweglichkeit der Atome wieder [BREITENLECHNER *et al.*, 2005]. Hohe Temperaturfaktoren erzeugen eine Unschärfe in den Strukturdaten. Diese Unschärfe kann ein Modell erheblich beeinträchtigen. Bereiche blauer Färbung zeigen Aminosäuren mit niedrigen Temperaturfaktoren, rote Bereiche Aminosäuren mit erhöhten Temperaturfaktoren. Abb. 3.12 A: PDB Acc. Nr. 1OIR. In Abb. 3.12 B ist der Ligand aus 1OIR zusammen mit der *Hinge*-Region und dem Lys33 dargestellt. Salzbrücken- und Wasserstoffbrückenbindungen sind durch gestrichelte Linien dargestellt. Nicht aufgelöste Bereiche der Proteinstrukturen wurden interpoliert dargestellt. Das Lys33 liegt in einer der aktiven Kinase entsprechenden Rotamerkonformation vor. Dies ermöglicht eine Salzbrücke mit einigen Inhibitoren, an deren Affinität diese einen hohen Beitrag hat.

enthalten. Aus der Literatur ist bekannt, dass bereits 1993 eine CDK2 Struktur kristallisiert wurde, diese ist jedoch nicht in der PDB hinterlegt. Sowohl die Strukturen 1HCK und 1HCL von 1996, als auch die ältere CDK2-Struktur aus dem Jahre 1993 wurden nicht im Komplex mit Cyclin kristallisiert. Für das strukturbasierte Wirkstoffdesign sind insbesondere die Strukturen interessant, die mit einem Inhibitor komplexiert sind. Diese Strukturen weisen diejenigen Merkmale einer Proteinkonformation auf, die durch die Interaktion mit dem Liganden zustandekommen (*Induced Fit*) [ENDICOTT *et al.*, 1999; NOBLE *et al.*, 2004]. Diese Strukturmerkmale ermöglichen die gezielte Entwicklung neuer Inhibitoren mit einer für die Proteinbindetasche optimierten Struktur. In Abb. 3.12 ist die Röntgenkristallstruktur von CDK2 dargestellt (PDB Acc. No.: 1OIR).

3.4 Homologie-Modellierung der EtCRK2

Ausgehend von der Templatstruktur 1OIR wurde ein Homologiemodell der EtCRK2 erstellt. Die Struktur ist eine Röntgenkristallstruktur von der inaktiven CDK2 aus *H. sapiens*. Die Auswahl erfolgte aufgrund der relativ hohen Auflösung von 0,21 nm (2.1 Å), sowie der niedrigen Temperaturfaktoren der Aminosäurereste in der Bindetasche (Abb. 3.12) [BREITENLECHNER *et al.*, 2005; NISSINK *et al.*, 2002]. Darüberhinaus konnten mit dieser Struktur in den *in silico* Anreicherungsexperimenten sehr gute Ergebnisse erzielt werden. Eine weitere, für das *Drug-Design* wichtige Eigenschaft der Struktur war die Ausrichtung des Lys33, das die ionische Wechselwirkung mit potenziellen Inhibitoren ermöglicht. Die Sequenzen von 1OIR und der EtCRK2 wurden in einem Sequenz-*Alignment* aneinander angepasst und anschließend wurden vier Proteinmodelle generiert [HTTP://WWW.ACCELRY.COM]. Diese Modelle wurden anschließend mittels Analyse der einzelnen Aminosäuren betrachtet. Dabei wurde im ersten Schritt eine Analyse der Aminosäuren durchgeführt, die nicht dem Ramachandran-*Plot* entsprechen. In einem Ramachandran-*Plot* werden die φ - und ψ - Winkel der Aminosäuren gegeneinander aufgetragen.

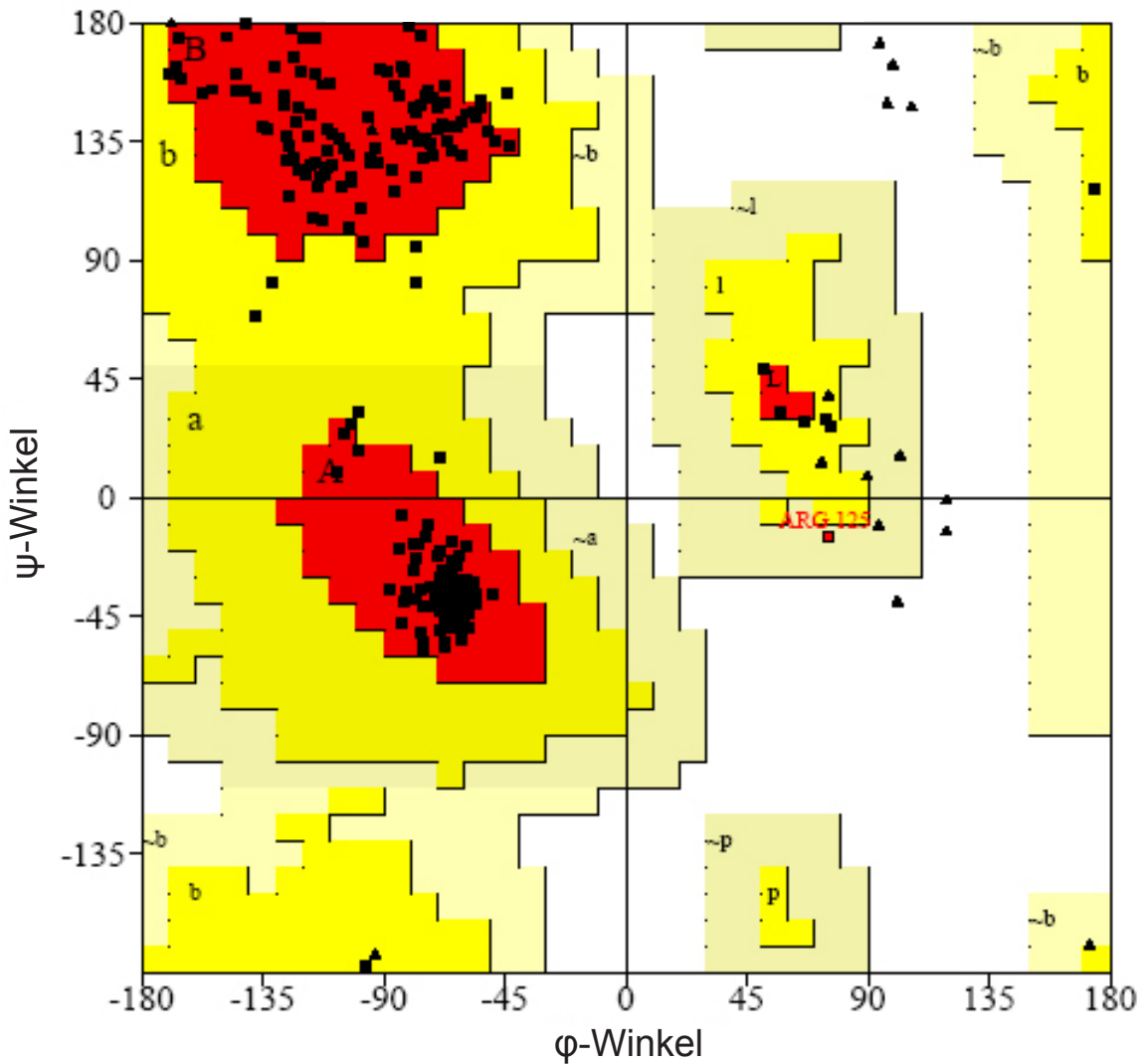


Abb. 3.13 Ramachandran-Plot des Homologiemodells der EtCRK2. Bereiche favorisierter Bindungswinkel sind rot hinterlegt. In gelb sind die Bereiche markiert, in denen erlaubte Bindungswinkel liegen. In weißen Regionen sind Aminosäuren positioniert, in denen Aminosäuren mit ungünstigen Bindungswinkel vorkommen. Eine Ausnahme von diesen Kriterien bildet die Aminosäure Glycin (\blacktriangle), die aufgrund ihrer Flexibilität auch ungünstige Bindungswinkel erlaubt. Insgesamt zeigt der Ramachandran-Plot eine überwiegend günstig bewertete Proteinstruktur an. Einzig die Aminosäure Arg125 besitzt ungünstige Bindungswinkel. Diese Konformation ist ebenfalls in der homologen Aminosäure des Templates vorhanden.

In dem *Plot* gibt es favorisierte und erlaubte Bereiche, die auf bestimmte Sekundärstrukturelemente hinweisen. Aminosäuren, die sich in nicht-favorisierten und erlaubten Bereichen aufhalten, sollten überprüft und mit der Templatstruktur verglichen werden. War die Aminosäure im Templat ebenfalls nicht in günstigen Bereichen des Ramachandran-*Plots*, so wurde dieser Fehler als unkritisch eingestuft. Obwohl solche Aminosäuren oftmals in einem energetisch ungünstigen Zustand vorliegen, sind diese Zustände häufig für die Funktion des Proteins notwendig. Mögliche Fehler außerhalb der Bindungsregion wurden weitgehend außer acht gelassen, sofern der überwiegende Anteil der Aminosäuren des Modells innerhalb zulässiger Bereiche des Ramachandran-*Plots* war (Abb. 3.13). Analog dazu wurden Bindungslängen und Bindungswinkel der Aminosäuren-Seitenketten untersucht. Traten Abweichungen dieser Werte zwischen den Modell-Aminosäuren und den favorisierten Bereichen auf, so wurden diese Aminosäuren näher betrachtet und verschiedene Seitenkettenrotamere aus einer Rotamer-Datenbank als Alternativen erprobt. Dabei wurden mögliche Interaktionen der Seitenketten, wie z.B. Wasserstoffbrücken, berücksichtigt. Im letzten Schritt wurde die Qualität des Modells mit einer Wahrscheinlichkeits-Dichte-Funktional (PDF) Methode untersucht. Dabei werden stereochemische Beschränkungen, wie Bindungslängen, Bindungswinkel, Diederwinkel, van der Waals Abstoßungen und Disulfidbrücken berücksichtigt. Die zugrundeliegenden Parameter entstammen der Fachliteratur und experimentellen Daten und sind in der Software Modeller implementiert [BROOKS *et al.*, 1983; SALI & OVERINGTON, 1994]. Das Homologiemodell wird während der Bewertung mittels der PDF-Methode mit den zugrundeliegenden Daten verglichen. Als Ergebnis werden Wahrscheinlichkeiten der einzelnen Parameter für jede Aminosäure ausgegeben.

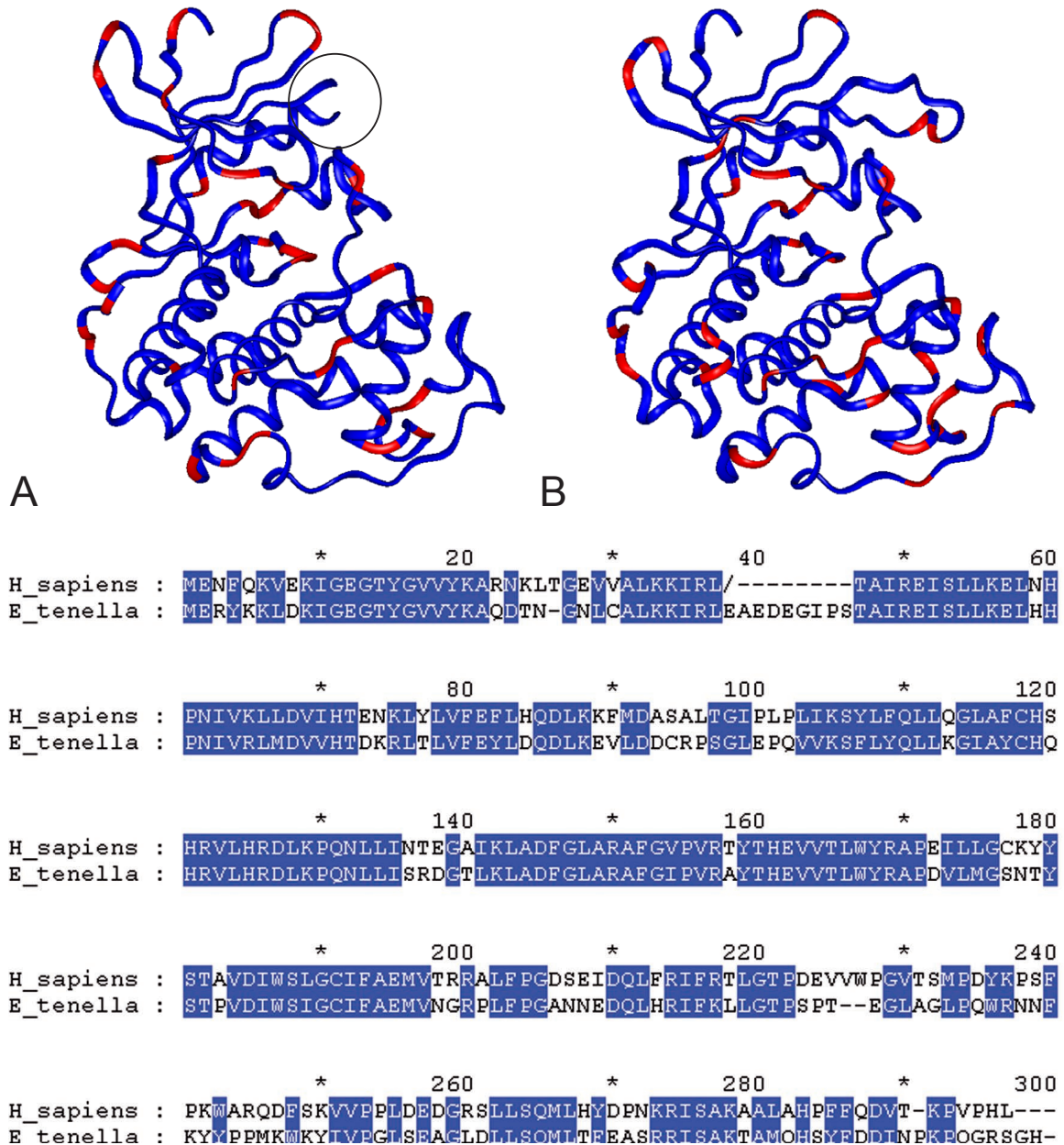


Abb. 3.14 Die Abbildungen 3.14 A und 1.14 B zeigen die Aminosäuren mit ungünstigen Bindungswinkeln (rot markiert) des Templat-Proteins (A) und des Homologiemodells (B). Die Lücke im Bereich der PSTAIRE Helix ist auf einen in der Röntgenkristallstruktur nicht aufgelösten Bereich des Templats zurückzuführen (Kreis). Darunter ist das zugehörige Sequenzalignment dargestellt.

Aus den Einzelbeiträgen wurden für jede Aminosäure Basis-PDF-Werte berechnet. Diese wurden farblich dargestellt und auf die Ribbondarstellung des Modells übertragen. Ein überdurchschnittlich hoher Basis-PDF-Wert ist dabei ein Anhaltspunkt für eine ungünstige Aminosäurekonformation des Modells. Nach Analyse der Modelle, wurde das in Abb. 3.14 dargestellte Modell für das virtuelle Screening ausgewählt. Abweichungen der Bindetaschen des Modells und des Templats waren dabei lediglich geringfügig.

3.4.1 Validierung des Homologiemodells

Das Homologiemodell konnte mittels *Docking* überprüft werden. Bekannte Inhibitoren der humanen CDKs dienten dabei als Referenz. Diese wurden in die Proteinbindetasche eingepasst und RMSD-Werte berechnet. Darüberhinaus konnten mit diesem Modell Anreicherungsexperimente mit den Datensätzen, die in Kap. 3.2.2 und 3.2.3 verwendet wurden, durchgeführt werden.

3.4.1.1 RMSD-Berechnung nach *Re-Docking* bekannter Inhibitoren

Um RMSD-Werte bekannter Liganden zu berechnen, wurden dieselben Liganden, wie sie in Kap. 3.2.1 verwendet wurden, herangezogen. Die verschiedenen Röntgenkristallstrukturen der humanen CDK2 wurden mittels eines strukturellen *Alignments* mit dem Homologiemodell angeglichen und die zugehörigen Liganden im selben Koordinatensystem wie das Homologiemodell gesondert abgespeichert. Diese konnten anschließend manuell in die Bindetasche des Homologiemodells eingepasst werden. Dieselben Liganden wurden einem *Re-Docking* zugeführt. Die Liganden, die dem *Re-Docking* unterzogen wurden, konnten nun mit den manuell eingepassten Liganden verglichen werden. Für einen objektiven Vergleich der durch *Docking* gefundenen Pose des Liganden mit der Ligand-Struktur, wie sie in der PDB vorkommt, wurden RMSD-Werte von den manuell eingepassten und den durch *Docking* eingepassten Liganden berechnet. Die RMSD-Werte sind in Tab. 3.5

aufgeführt. Durch Vergleich der RMSD Werte mit den RMSD Werten, die für die einzelnen PDB-Strukturen ermittelt wurden, war es nun möglich, Aussagen hinsichtlich der *Docking*-Qualität des Homologiemodells zu treffen. Obwohl die Liganden für humane CDKs adressieren, kann davon ausgegangen werden, dass dieselben Liganden gleichfalls an die EtCRK2 binden. Daher sind die *Re-Dockings* in das Homologiemodell ein Indikator für die Qualität des Modells. Insgesamt zeigte sich, dass das Homologiemodell sehr gut geeignet ist, die meisten Liganden mit einem akzeptablen RMSD-Wert, zu reproduzieren. Einzig die Liganden der PDB-Strukturen 1H01, 1JSV, 1JVP und 1URW zeigten relativ schlechte RMSD-Werte. Dies lässt sich u.a. auf eine hohe Flexibilität der Liganden, sowie die spezifische Beschaffenheit der Bindetasche des Homologiemodells zurückführen. Vergleicht man die Werte in Tab. 3.5 mit den Werten in Tab. 3.1 so fällt auf, dass 6 von 15 Strukturen mit einem RMSD-Wert, der kleiner als 0,25 nm (2,5 Å) ist, reproduziert werden können. Dagegen können 9 der Ligand-Konformationen gut mit den jeweils zugehörigen Proteinstrukturen reproduziert werden (Tab. 3.1). Ein Grund für diesen Unterschied liegt in der Flexibilität der Proteinstrukturen, wobei die verschiedenen Protein-Konformationen, an denen Liganden unterschiedlich binden, nicht durch ein einzelnes Homologiemodell repräsentiert werden können.

PDB Accession Number	RMSD
1AQ1	3,1
1DI8	2,0
1E1V	2,0
1GZ8	1,4
1H01	5,1
1JSV	6,2
1JVP	5,3
1KE9	1,5
1OIQ	3,8
1OIR	3,3
1OIU	3,7
1PXI	3,4
1R78	2,5
1URW	4,8
1V0O	1,3

Tab. 3.5 RMSD-Werte der besten Ligand-Pose nach *Re-Docking* in das Homologiemodell. Das Homologiemodell erlaubt ein korrektes *Re-Docking* von 6 der 15 untersuchten CDK-Inhibitoren.

3.4.1.2 Anreicherungsexperimente mit *Docking*-Datensätzen

Für eine Eignung des Homologiemodells im *in silico*-Screening sollte das Modell in der Lage sein, zwischen bekannten Inhibitoren humaner CDKs und Strukturen unbekannter Aktivität an CDKs zu unterscheiden. Deshalb wurden mit dem Homologiemodell *Docking*-Experimente mit den Datensätzen aus Kap. 3.2.2 durchgeführt.

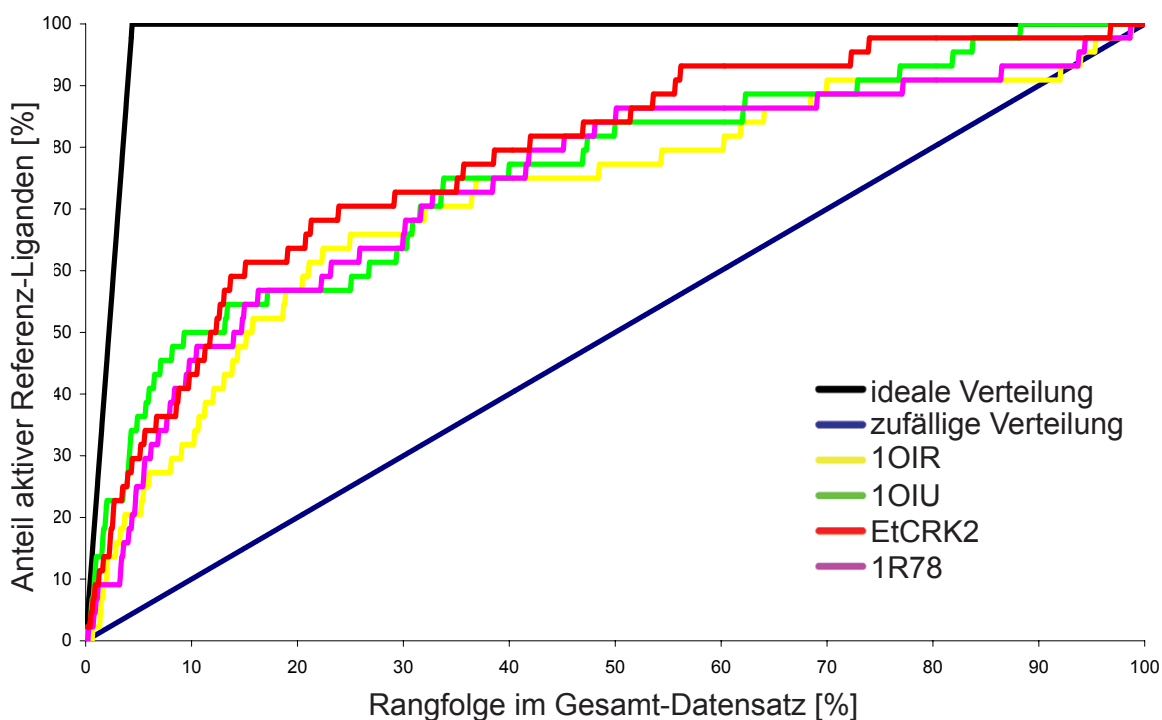


Abb. 3.15 Anreicherungskurven der CDK2 Inhibitoren nach Docking in verschiedene Röntgenkristallstrukturen und in das Homologiemodell der EtCRK2. Für die rote Kurve wurde das Homologiemodell der EtCRK2 verwendet. Für die grüne Kurve wurde die PDB-Struktur 1OIU verwendet. Die violette Kurve wurde mit der PDB-Struktur 1R78 erstellt. Die gelbe Kurve wurde mit der PDB-Struktur 1OIR erstellt. Die blaue Linie stellt eine zufällige Verteilung der Referenz-Liganden dar. In schwarz ist eine ideale Rangfolge der Referenz-Liganden dargestellt, d.h. alle Referenz-Liganden belegen die ersten Plätze in der Rangfolge.

Für die Anreicherungsexperimente wurde die Kombination aus Jain und G-Score, wie sie in Kap. 3.2.2 für eine optimale Zweistufen-Bewertung identifiziert wurde, verwendet. Abb. 3.15 zeigt die zugehörigen Anreicherungskurven des Homologiemodells im Vergleich mit den Anreicherungskurven der verschiedenen PDB-Strukturen unter Verwendung der Bewertungsfunktionen Jain und G-Score.

Bei diesen Experimenten zeigt das Homologiemodell eine besonders gute Anreicherung der CDK-Inhibitoren. Zum Vergleich sind in der Abbildung zusätzlich die Anreicherungskurven der PDB-Strukturen enthalten, die ebenfalls sehr gute Anreicherungen bewirken. Beim Vergleich fällt auf, dass die Anreicherungskurven ähnlich verlaufen. Im Bereich der besten Positionen (0,1 % - 2 %) der Rangfolge sind die PDB-Struktur 1OIU und das Homologiemodell vergleichbar gut geeignet, zwischen Inhibitoren und Strukturen unbekannter Aktivität zu unterscheiden. Betrachtet man den gesamten Verlauf der Kurven ist das Homologiemodell vergleichbar mit den besten PDB-Strukturen. Deshalb kann davon ausgegangen werden, dass das Homologiemodell für den Einsatz im virtuellen Screening geeignet ist.

3.5 Pharmakophor-Analyse des Homologie-Modells

Die Bindetasche des Homologiemodells wurde mit der Software Cerius2 analysiert. Dabei werden 3 Wechselwirkungstypen von der Software untersucht. Dazu gehören Wasserstoffbrückendonoren und -akzeptoren, sowie hydrophobe Regionen (Abb. 3.16). Nach dem *Clustering* der Wechselwirkungsvektoren kann daraus ein Pharmakophormodell abgeleitet werden [[HTTP://WWW.ACCELRY.COM](http://www.accelrys.com)]. Vergleicht man die Bindetasche von CDK3 des Huhns mit der Bindetasche der EtCRK2, so findet man weitgehende Übereinstimmungen. Diese lassen bei dieser Betrachtungsweise wenig Möglichkeiten zu, das Modell hinsichtlich der Suche nach selektiven Inhibitoren der EtCRK2 zu optimieren. Lediglich zwei Positionen weisen offensichtliche Unterschiede zwischen beiden Spezies auf.

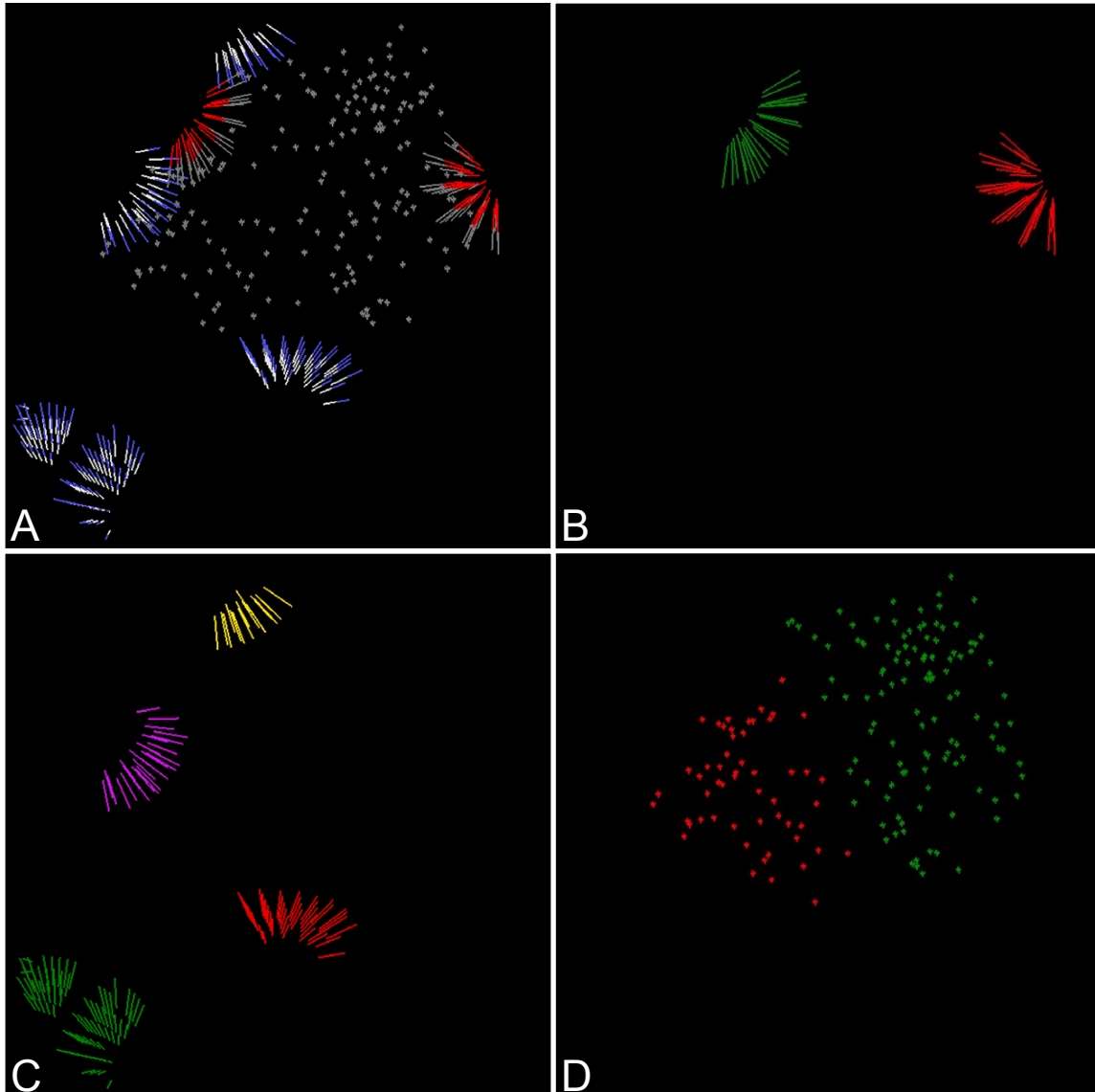


Abb. 3.16 Bindetaschenanalyse des Homologiemodells. In Abb. 3.16 A sind alle Interaktionen, die für das Pharmakophormodell ausgewählt wurden, dargestellt. Die blau-weißen Cluster repräsentieren Wasserstoffbrückenakzeptoren, die rot-weißen Cluster Wasserstoffbrückendonoren. Die grauen Punkte repräsentieren Bereiche lipophiler Wechselwirkungen. In Abb. 3.16 B sind die Cluster der Wasserstoffbrückendonoren (Leu82: grün, Lys32: rot) farblich markiert. In Abb. 3.16 C sind die Wasserstoffbrückenakzeptoren (Glu80: gelb, Leu82: violett, Asp85: rot, Glu88: grün) dargestellt. In Abb. 3.16 D sind die Cluster lipophiler Interaktionen farblich unterschiedlich markiert.

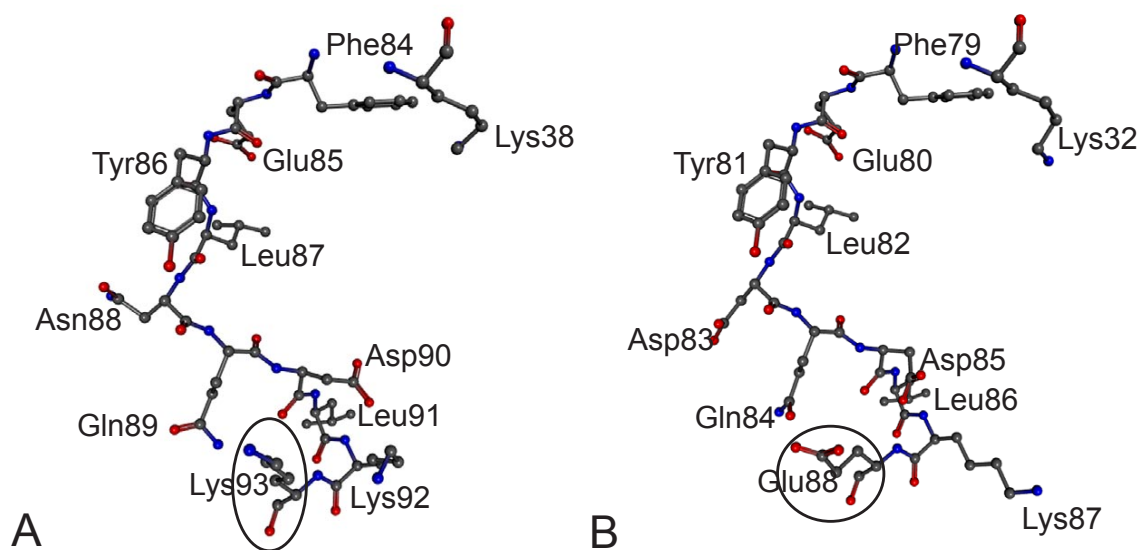


Abb. 3.17 Für die Bindetaschenanalyse herangezogene Unterschiede der Bindetaschen der Homologiemodelle von *G. gallus* CDK3 (A) und *EtCRK2* (B). Durch Deletionen in der Sequenz der *EtCRK2* sind die Aminosäuren unterschiedlich nummeriert. Die positiv geladene Aminosäure Lys93 (*G. gallus*, Abb. 3.17 A) ist gegen das negativ geladene Glu88 (*E. tenella*, Abb. 3.17 B) ausgetauscht (Kreise). Ebenfalls ist die negativ geladene Aminosäure Asp83 (*E. tenella*) gegen die neutrale Aminosäure Asn88 (*G. gallus*) ausgetauscht, jedoch ist dieser Austausch nicht für das Design selektiverer Inhibitoren nutzbar, da die Aminosäure an dieser Position mit dem Peptidrückgrat in die Bindetasche zeigt.

Zum einen ist dies ein Austausch von Asparagin zu Aspartat an den homologen Positionen Asn88 (Huhn) und Asp83 (*Eimeria*) und ein Austausch des Lys93 des Huhns gegen ein Glutamat an Position 88 bei der *EtCRK2* (Abb. 3.17). Für den Einsatz in einem selektiven Pharmakophormodell erschien lediglich der Lysin-Glutamat Austausch von Relevanz, da die Seitenkette der Aminosäure Asp83 aus der Bindetasche des Homologiemodells herausweist, was eine Zugänglichkeit für das Ligand-Design unmöglich macht.

Für den Einsatz als *in silico*-Selektionsmethode von Strukturen wurde eine Reihe von Pharmakophormodellen auf Grundlage dieser Bindetaschenanalyse generiert. Die Modelle wurden auf Wechselwirkungen beschränkt, die für Inhibitoren humaner CDKs in der Literatur beschrieben sind.

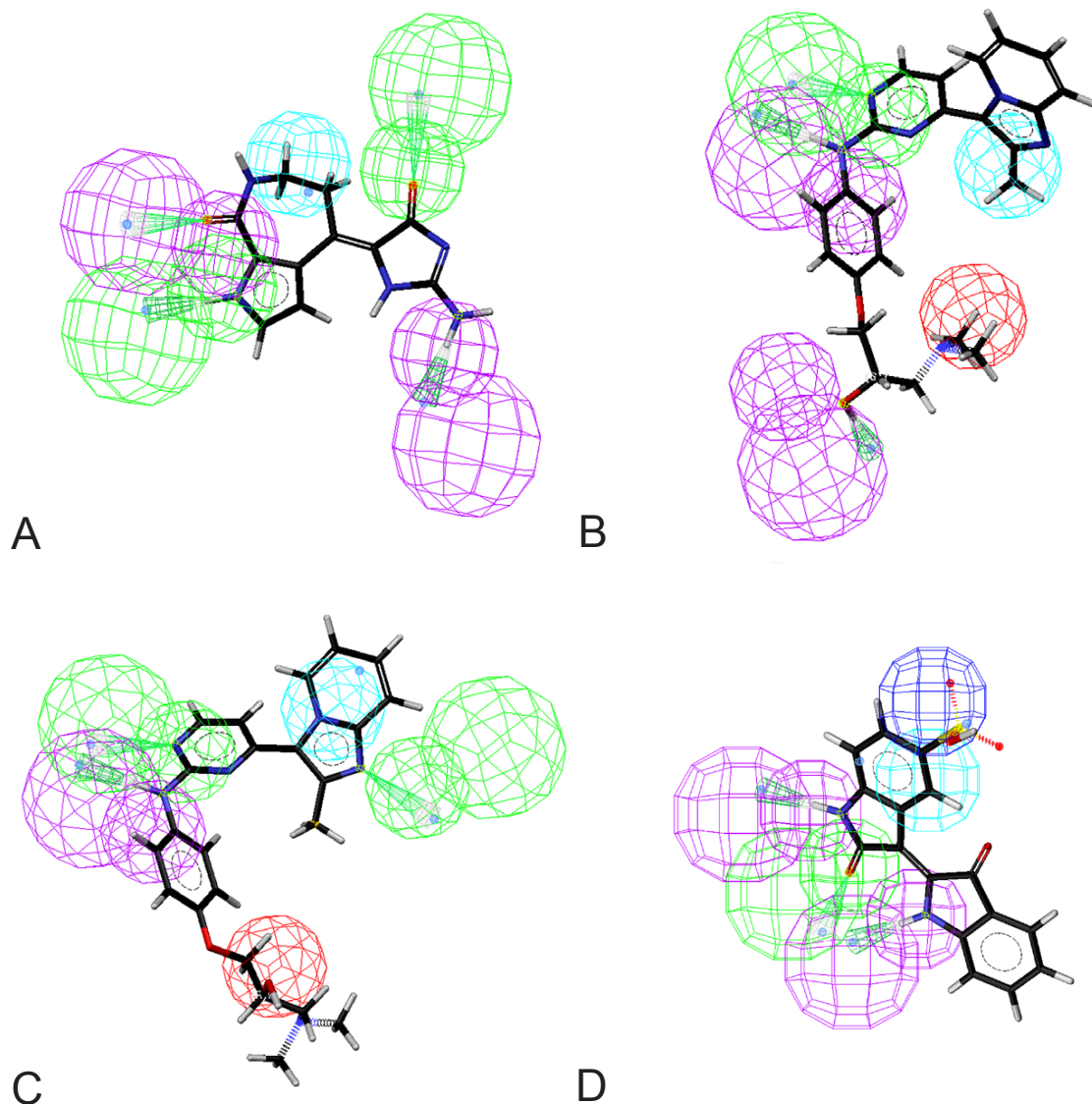


Abb. 3.18 Beispiele für Mitglieder der Ensembles von Pharmakophormodellen, die aus der Bindetaschenanalyse abgeleitet wurden. Die grünen Sphären sind Bereiche, in denen ein Wasserstoffbrückenakzeptor für eine Ligandbindung günstig ist. Die violetten Sphären stellen Wasserstoffbrückendonoren dar. Dunkelblaue Kugeln sind Bereiche eines Liganden, die negativ ionisierbar sind. Rote Kugeln stellen positiv ionisierbare Bereiche dar. Lipophile Bereiche sind durch türkisfarbene Kugeln dargestellt. Abb. 3.18 A-D stellen Repräsentanten der Pharmakophor-Ensembles 1-4 dar.

Datensatz	Hypothese	bekannte Inhibitoren [%]	unbekannte Substanzen [%]	Anreicherungsfaktor
ICD-Datensatz	Ensemble 1	63,3	10,1	6,3
	Ensemble 2	27,3	2,4	11,4
	Ensemble 3	25,0	37,3	0,7
	Ensemble 4	0,0	1,0	0,0
Kinase-Datensatz	Ensemble 1	65,9	10,0	6,6
	Ensemble 2	25,0	2,7	9,3
	Ensemble 3	25,0	4,0	6,3
	Ensemble 4	0,0	0,5	0,0
CDK-Datensatz	Ensemble 1	63,3	6,4	9,9
	Ensemble 2	25,0	0,9	27,8
	Ensemble 3	25,0	1,7	14,7
	Ensemble 4	0,0	0,2	0,0
Pharmakophor-Datensatz	Ensemble 1	65,9	64,1	1,0
	Ensemble 2	27,3	17,3	1,6
	Ensemble 3	27,3	24,8	1,1
	Ensemble 4	0,0	7,9	0,0

Tab. 3.6 Anreicherungsfaktoren bekannter Inhibitoren der humanen CDK2 in den unterschiedlichen Testdatensätzen mit verschiedenen Pharmakophor-Ensembles. In der 3. Spalte ist die prozentuale Anzahl von Inhibitoren dargestellt, die mit den verschiedenen Pharmakophor-Ensembles wiedergefunden wurden. In der dritten Spalte ist der Anreicherungsfaktor dargestellt. Er ergibt sich aus:

$$\frac{[\text{Prozent gefundene bekannte Inhibitoren}]}{[\text{Prozent gefundene unbekannte Substanzen}]} = \text{Anreicherungsfaktor}$$

Um die Suche nicht zu stark einzuschränken, wurden die Modelle so gewählt, dass sie jeweils genau 4 Eigenschaftsvektoren besitzen. Jedes dieser Modelle beinhaltet mindestens eine Wechselwirkung, die das NH des Leu82 als Wasserstoffbrückendonator adressiert. Eine weitere Bedingung ist eine Wasserstoffbrücke entweder zu dem Sauerstoff des Leu82 oder zu dem Sauerstoff des Glu80. Als dritte Eigenschaft muss mindestens einer der beiden hydrophoben Cluster von dem Molekül abgedeckt werden. Alle weiteren Kombinationsmöglichkeiten wurden iterativ von der Software generiert. Die resultierenden Einzel-Hypothesen wurden in 4 verschiedenen Pharmakophor-Ensembles zusammengefasst, um die Datenbanksuche zu vereinfachen (Abb. 3.18).

In den einzelnen Pharmakophor-Ensembles sind mehrere Hypothesen sequenziell miteinander verknüpft. Bei einer Datenbanksuche können so Strukturen auf alle Hypothesen geprüft werden. Pharmakophor-Ensemble 1 beinhaltet dabei die in einem Ensemble vorhandenen Pharmakophormodelle, welche Wasserstoffbrücken zu den Aminosäuren Asp85 und Lys32 ermöglichen (Abb. 3.17). Das Pharmakophor-Ensemble 2 gleicht weitgehend dem Pharmakophor-Ensemble 1, jedoch sind die Interaktionen mit dem Asp85 mit der Pharmakophor-Eigenschaft „positiv ionisierbar“ anstelle von Wasserstoffbrückendonoren dargestellt. In dem Pharmakophor-Ensemble 3 sind analog dazu die Interaktionen mit dem Glu88 als „positiv ionisierbar“ dargestellt. Das Pharmakophor-Ensemble 4 stellt die Interaktion mit dem Lys33 mit der Eigenschaft „negativ ionisierbar“ anstelle eines Wasserstoffbrückenakzeptors dar. Alle Pharmakophor-Ensembles verfügen über identische Wasserstoffbrücken-Vektoren zur *Hinge*-Region (Abb. 3.18). Die Pharmakophor-Ensembles wurden aus allen Kombinationsmöglichkeiten von Interaktionen zusammengestellt, wobei mindestens zwei Interaktionen mit der *Hinge*-Region enthalten sein mussten. Die Hypothesen des Pharmakophor-Ensembles 1 wurden darüberhinaus auf diejenigen Hypothesen beschränkt, die in dem Pharmakophor-Ensemble möglichst viele bekannte Inhibitoren des Test-Datensatzes, die mit einem Anbieterkatalog (Maybridge Library) vermischt wurden, wiederfinden [[HTTP://WWW.MAYBRIDGE.COM](http://www.maybridge.com)]. Einzelhypothesen, die

dabei keine zusätzliche Anreicherung erbrachten, wurden verworfen. Dabei erwies sich das Pharmakophor-Ensemble 1 als besonders selektiv bezüglich einer hohen Anreicherung der CDK-Inhibitoren. Es wurden mit dieser Hypothese 1999 von 59652 (3,35 %) Maybridge-Strukturen und 29 von 44 (65,91 %) Strukturen des Test-Datensatzes gefunden, was darauf hindeutet, dass das Pharmakophor-Ensemble 2 sehr selektiv für die Anreicherung von CDK2 Inhibitoren ist (Tab. 3.6).

3.5.1 Validierung der Pharmakophor-Ensembles

Die für das virtuelle Screening verwendeten Pharmakophor-Ensembles wurden mit den bekannten Inhibitoren humaner CDKs und verschiedenen Datensätzen mit Strukturen von Substanzen unbekannter Aktivität erprobt. Von den aktiven Inhibitoren des Test-Datensatzes wurden insgesamt 68,2 % durch die verwendeten Hypothesen erkannt.

Die weitere Validierung der Modelle erfolgte unter Zuhilfenahme von verschiedenen Datensätzen (Tab. 3.6). Dieselben Datensätze wurden auch für die Anreicherungsexperimente im *Docking* benutzt (Kap. 3.2.3). Als Kontrolle wurde zusätzlich ein Pharmakophor-Datensatz eingesetzt. Die Datensätze wurden zunächst in das SMILES-Zeichenketten-Format konvertiert und aus diesen wurden anschliessend Pharmakophor-Datenbanken generiert. Dabei wurden Moleküle, die bestimmten chemischen Regeln nicht entsprechen oder chemisch falsch in der Ursprungs-Datenbank dargestellt waren, von der Software Catalyst automatisch entfernt, was eine Reduktion der Anzahl der Strukturen in den resultierenden Pharmakophor-Datenbanken zur Folge hatte. Von dem Datensatz der CDK Inhibitoren wurden 100 % in das Catalyst-Datenbankformat überführt. Der Kinase-Datensatz wurde zu 85,5 %, der CDK-Datensatz zu 95,9 %, der ICD-Datensatz zu 93,5 % und der Pharmakophor-Datensatz zu 100 % in das Catalyst Format konvertiert. Der Pharmakophor-Datensatz wurde als Positiv-Kontrolle generiert. Wider Erwarten konnten aus diesem Datensatz mit den Pharmakophor-Hypothesen lediglich 96,1

% der Strukturen nochmals gefunden werden. Dies lässt sich wahrscheinlich darauf zurückführen, dass die Konformer-Generierung während der Datenbank-Konvertierung nicht die identischen Konformere hervorbrachte, wie sie in der vorher durchsuchten Datenbank vorhanden waren. Dies würde ein Auffinden der Liganden unmöglich machen.

3.6 *In silico*-Screening

3.6.1 Vorbereitung der Datenbanken

Etliche Strukturdatenbanken werden von kommerziellen Anbietern in Katalogen angeboten. Teilweise werden diese Kataloge für einen bestimmten Zweck zusammengefasst. Aus diesen Strukturdatenbanken können dann die einzelnen Substanzen selektiert und anschließend ihre Aktivität am *Drug-Target* in einem biochemischen Test bestimmt werden. Beispielsweise werden solche Kataloge für den Einsatz an bestimmten *Drug-Target* Familien ausgerichtet (*focussed library*).

3.6.1.1 CDK-Datenbanken

Von einigen Anbietern waren bereits Datenbanken verfügbar, die für den Einsatz im Screening von CDK-Inhibitoren ausgerichtet sind. Von diesen Datenbanken wurden bestimmte Strukturen bereits vor den weiteren Schritten aussortiert. Zu diesem Zweck wurden verschiedene Filter angewendet. Insgesamt konnten aus den ursprünglichen Händlerkatalogen 98.088 Strukturen in einer Datenbank zusammengefasst werden. Nach Herausfiltern von Strukturen mit unerwünschten Gruppen, beispielsweise mit reaktiven chemischen Eigenschaften, resultierten 96.915 Strukturen. Diese wurden mit einem Eigenschaftsfilter, der wirkstoffähnliche Strukturen selektiert, beschränkt. Von den verbleibenden 85899 Strukturen wurden anschließend Ionisierungszustände berechnet, woraus 91.547 *Docking*-Strukturen resultierten (Abb. 3.21).

3.6.1.2 Kinase-Datenbanken

Mit den Kinase-Katalogen der Anbieter wurde wie in Kap. 3.6.1.1 beschrieben, verfahren. Ausgehend von 226.373 Strukturen, blieben nach den beiden Filterschritten 225.757 Strukturen übrig. Von diesen wurden die bereits in der CDK-Datenbank vorhandenen Strukturen entfernt, sodass 140.474 Strukturen übrig blieben, deren Ionisierungszustände berechnet wurden. Daraus resultierten 148.926 *Docking*-Strukturen.

3.6.1.3 Intervet Chemicals Directory

Die Datenbank, die für das Screening nach Inhibitoren der EtCRK2 verwendet wurde, enthielt 3.703.478 Strukturen aus unterschiedlichen Händlerkatalogen. Die Datenbank setzte sich aus den Datenbanken unterschiedlicher Anbieter und einer *in-house* Datenbank der Intervet Innovation GmbH zusammen. In einem ersten Schritt wurde die Datenbank um unerwünschte Molekülgruppen, die z.B. reaktiv oder toxisch sein können, bereinigt (Abb. 3.21). Daraus resultierte eine Anzahl von 3.630.541 Molekülen in der verbleibenden Datenbank. Die aus diesem Filterprozess resultierenden Strukturen wurden in eine Pharmakophor-Datenbank umgewandelt. Diese wurde mit den in Kap. 3.5 beschriebenen Pharmakophor-Ensembles durchsucht. Die gefundenen Strukturen wurden auf Duplikate, die in dem resultierenden Datensatz vorhanden waren, untersucht. Nach Abzug der Duplikate blieben 474.977 Strukturen übrig. Weiterhin wurden von den resultierenden Datensätzen Strukturen entfernt, die in den anderen, bereits von den Anbietern vorbereiteten, Kinase-Datenbanken enthalten waren. Dies ergab eine Gesamtmenge von 467.727 Molekülen. Nach Einsatz des Wirkstoff-Eigenschaftsfilters verblieben von dem Pharmakophor-Datensatz 230.357 Strukturen, die nach Berechnung von Ionisierungszuständen 289.369 *Docking*-Strukturen ergaben (Abb. 3.21).

3.6.2 Screening

3.6.2.1 Docking der Datenbanken mit dem Modell der EtCRK2

Das Hochdurchsatz-*Docking* wurde auf 16 *Grid-Computing Client*-Computern durchgeführt. Insgesamt wurden 456.730 Einzel-Strukturen (529.842 ionisierte *Docking*-Strukturen) in die Bindetasche des Homologiemodells eingepasst. Der Vorgang dauerte ca. 40 Tage, was einem Durchsatz von ca. 32 Strukturen pro Stunde auf einem einzelnen Computer entspricht. Die Fehlerrate der Strukturbearbeitung des *Dockings* lag unter 0,1 %.

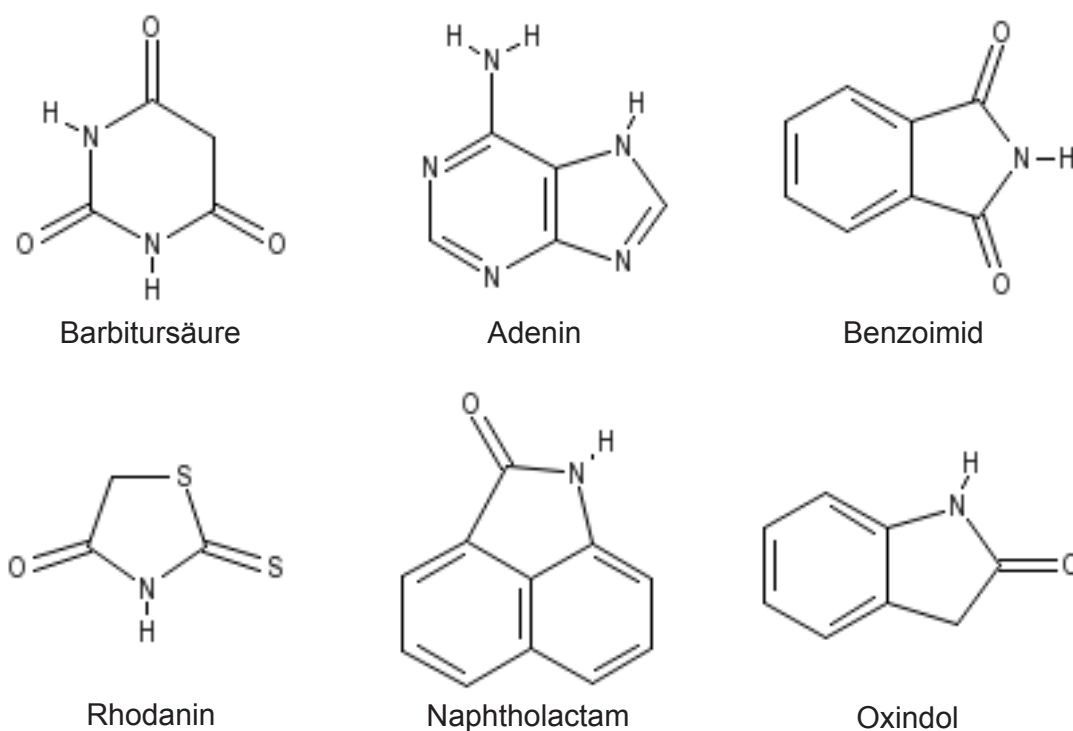


Abb. 3.19. Die sechs häufigsten, durch den *in-silico* Ansatz selektierten Strukturklassen. Die hier gezeigten Strukturklassen bilden etwa die Hälfte der selektierten Strukturen ab. Die Barbitursäure, das Adenin und das Benzoimid sind als Beispiele der Strukturklassen der Barbiturate, der Purine und der Phthalimide dargestellt. Die weiteren hier abgebildeten Strukturen sind die Bestandteile der gleichnamigen Strukturklassen.

3.6.3.2 Nachprozessierung der eingepassten Strukturkonformere

Die in die Bindetasche eingepassten Strukturen wurden mittels des in Kap. 3.7 beschriebenen Interaktionsfilters nachprozessiert. Von den CDK-Datenbanken passierten insgesamt 14.723 Konformere, die 7.025 Einzel-Strukturen entsprechen, den Filter. Die verbleibenden Strukturen wurden dann einer Nach-Bewertung mit verschiedenen Bewertungsfunktionen (*Re-Scoring*) unterzogen. Nach dem *Re-Scoring* wurde in einem ersten Schritt von jeder Struktur eine mögliche bioaktive Konformation ermittelt. Dies geschah mit der Bewertungsfunktion Jain, die für diesen Zweck geeignet erschien (Kap. 3.2.2). Die verbleibenden Einzel-Strukturen wurden dann mittels der Bewertungsfunktion G-Score bewertet und anhand dieser Bewertung in einer Rangfolge geordnet. Diese Bewertungsfunktion wies die höchsten Anreicherungen an einem Testdatensatz auf (Kap. 3.2.2).

Mit den Kinase-Datensätzen wurde ebenso verfahren. Dabei wurden zunächst 32.443 Konformere herausgefiltert, die 13.894 Einzel-Strukturen entsprechen. Die ICD konnte auf 150.813 Konformere reduziert werden, was 62.622 Einzel-Strukturen entspricht. Zusätzlich wurden Interaktionsfilter verwendet, die die Unterschiede der EtCRK2 und der CDK3 des Huhns berücksichtigen. Diese sollten genutzt werden, um selektive Inhibitoren der EtCRK2 zu finden. Dabei wurden Moleküle herausgefiltert, die zusätzlich zu der *Hinge*-Interaktion noch eine positive Ladung in räumlicher Nähe zum Glu88 der EtCRK2 besitzen. An der analogen Position besitzt die CDK3 des Huhns die Aminosäure Lysin. Dieser Austausch bewirkt eine Ladungsumkehr der Aminosäure.

3.6.3.3 Visuelle Beurteilung der *Docking*-Experimente

Von den so selektierten Strukturen wurden von jedem Datensatz die 500 - 1.000 best-bewerteten Strukturen betrachtet, um geeignete Ligand-Posen zu selektieren. Diese so gefundenen Strukturen wurden in einer Datei gespeichert. Durch den Einsatz der verschiedenen Filter konnten darüberhinaus auch Strukturen gefunden werden, die die EtCRK2-selektiven Aminosäuren adressieren. Durch diese Vorgehensweise wurden 195 Einzel-Substanzen ausgewählt und für biochemische Experimente bestellt.

Es fiel dabei auf, dass sich die Strukturklassen der Naphtholactame, der Phthalimide, der Rhodanine, der Purine, der Barbiturate und der Oxindole stark in dieser Auswahl häuften (Abb. 3.19). Insgesamt deckten diese 6 Strukturklassen rund die Hälfte der ausgewählten Substanzen ab. Von den Purinen, den Phthalimiden, den Oxindolen und den Naphtholactamen war seit längerem bekannt, dass einige Vertreter dieser Strukturklassen Inhibitoren von CDKs sind [ROUTIER *et al.*, 2005; GRAY *et al.*, 1999; HARDCASTLE *et al.*, 2004; BRAMSON *et al.*, 2001; LUK *et al.*, 2004; Liu *et al.*, 2003]. Die Rhodanine sind u. a. als Inhibitoren der Glycogensynthase-Kinase GSK-3 beschrieben worden, während für die Barbiturate u. a. eine Hemmung der Proteinkinase C Aktivität beschrieben wurde [DESHMUKH *et al.*, 1989]. Die Häufung von Strukturklassen mit bekannter Kinaseaktivität deutet darauf hin, dass während des virtuellen Screenings eine Anreicherung von CDK-Inhibitoren stattgefunden hat. Die Kinasen Glykogensynthase-Kinase und die CDKs sind tatsächlich eng verwandt [MANNING *et al.*, 2002]. Da die Bindetaschen der Kinasen allgemein sehr ähnlich sind, ist denkbar, dass die Strukturklassen, deren Wirksamkeit an anderen Kinasen nachgewiesen wurde, auch CDK-inhibierende Vertreter beitzten. Das zusätzliche Vorkommen nicht so stark populierter Strukturklassen in der Strukturauswahl lässt darauf hoffen, für Kinasen bislang gänzlich unbekannte Inhibitor-Klassen zu finden. Biochemische Experimente werden weitere Anhaltspunkte dafür liefern, wie die Ligandbindung an die EtCRK2 erfolgt.

3.7 Biochemische Testung der ausgewählten Substanzen

Von den 195 bestellten Substanzen standen lediglich 139 zur Verfügung. Deren Aktivität wurde zusammen mit 11 bekannten CDK-Inhibitoren in einem EtCRK2-Enzymtest bestimmt. Für die Messungen wurde der „TR-FRET Screening Express Kit“ von Molecular Devices verwendet [HTTP://WWW.MOLECULARDEVICES.COM]. Für die Versuche wurden die EtCRK2 und das RINGO-Protein des Krallenfroschs *Xenopus laevis* in einem *Escherichia coli* Expressionssystem rekombinant exprimiert und aufgereinigt. Das Protein RINGO bindet unspezifisch an verschiedene CDKs und ist ein starker Aktivator der EtCRK2 [SUAREZ-FERNANDEZ, unpublizierte Daten; MERCKX *et al.*, 2003].

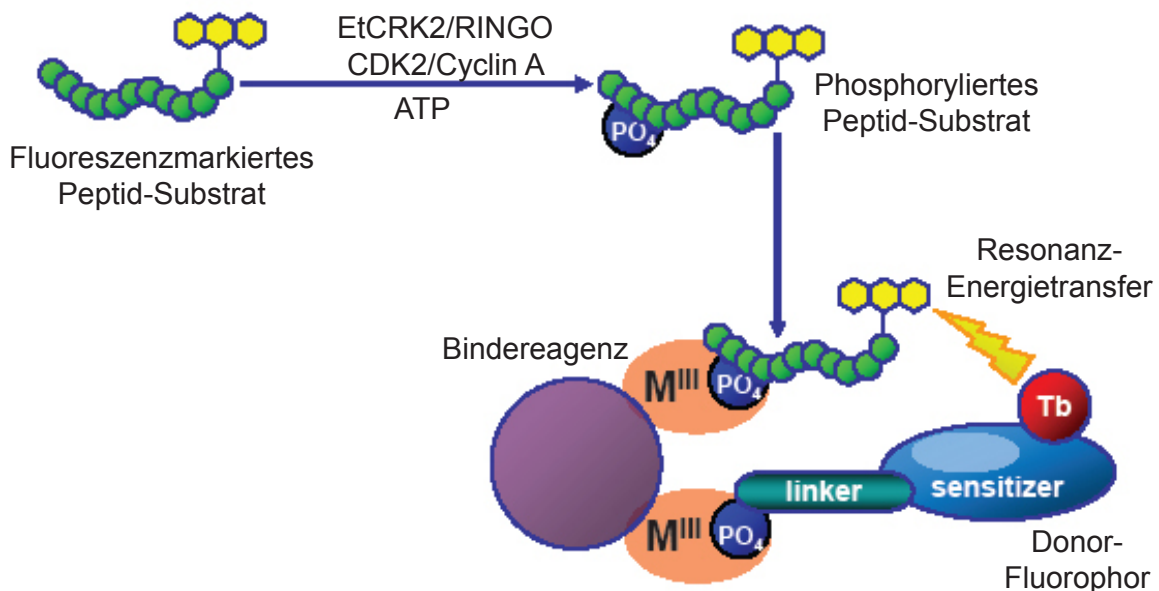


Abb. 3.20. Funktionsprinzip des verwendeten Testsystems für die Aktivitätsbestimmungen der Testsubstanzen. Durch den EtCRK2-RINGO-Komplex wird ein fluoreszenzmarkiertes Substratpeptid phosphoryliert. Das Bindereagenz bindet das phosphorylierte Substrat und ein Donor-Fluorophor (Tb). Durch die Nähe des Donor-Fluorophors zum fluoreszenzmarkierten Substratmolekül kommt es zum Resonanz-Energie-Transfer. Dabei wird die Energie eines angeregten Donor-Fluorophors nicht in Form von Fluoreszenz abgegeben, sondern strahlungslos über Dipol-Dipol-Wechselwirkungen auf einen fluoreszierenden Akzeptor übertragen. Die entstehende Fluoreszenz kann daraufhin detektiert werden. Durch diesen Verstärkungsmechanismus können auch sehr schwache Signale detektiert werden [HTTP://WWW.MOLECULARDEVICES.COM].

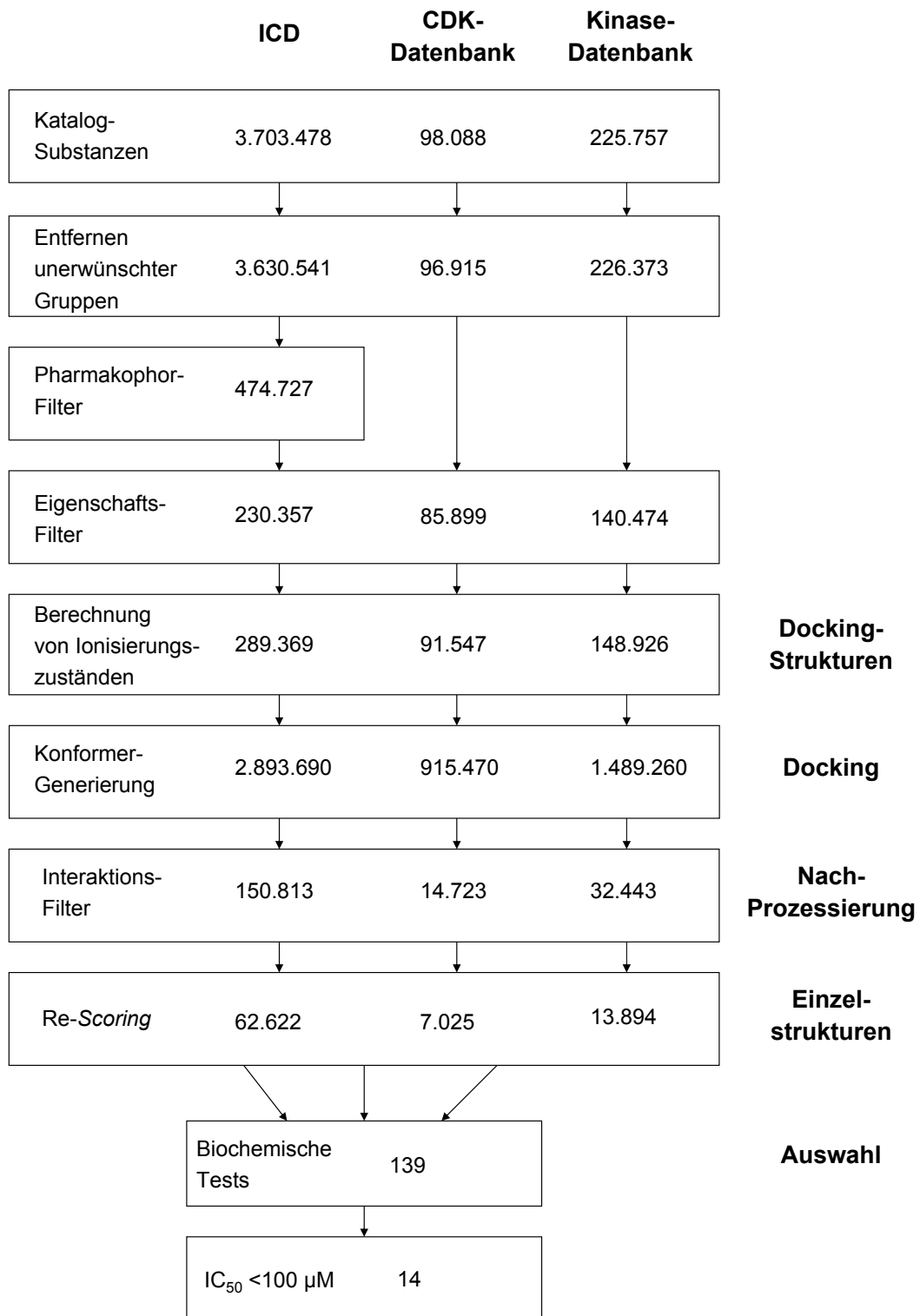


Abb. 3.21. Schema der verschiedenen Filterschritte im *in-silico*-Screening. Von den ca. 4 Mio. Strukturen wurden letztendlich 139 Substanzen mit der EtCRK2 getestet. Davon besaßen 14 Substanzen einen IC₅₀-Wert von weniger als 100 µM.

Die mit dem RINGO-Protein aktivierte EtCRK2 wurde nach Herstellervorgaben in den Tests eingesetzt (Abb. 3.20). Die Konzentration der Testsubstanzen in einem ersten Test betrug 30 μM . Eine Inhibition von 30 % bei dieser Konzentration wurde als *Hit* bewertet, d.h. diese Substanzen zeigten eine *Drug-Target*-Aktivität und wurden in einem zweiten Test näher untersucht.

3.7.1 Bestimmung von IC_{50} -Werten der Inhibitoren

Von den 139 getesteten Substanzen fielen 39 aktive Substanzen auf. Das entspricht einer *Hit*-Rate von 28 %. In einem HTS-*Assay* fallen üblicherweise 0,1 - 0,5 % der gemessenen Substanzen als *Hit* auf [POLGAR *et al.*, 2005]. Jedoch können diese Werte nur bedingt verglichen werden, da häufig unterschiedliche Kriterien für einen *Hit*-Substanz gelten. Auch ist die *Hit*-Rate sehr stark von dem *Drug-Target* Protein und der verwendeten Substanz-Kollektion abhängig.

Von den *Hit*-Substanzen wurden IC_{50} -Werte bestimmt. Der IC_{50} -Wert entspricht der Konzentration, bei der das Enzym zu 50 % inhibiert ist. Sie werden allgemein als Maß der Wirkstärke eines Inhibitors verwendet. Da bei der Bestimmung von IC_{50} -Werten jedoch die Kinetik der Enzymreaktion nicht berücksichtigt wird, sind diese Werte im allgemeinen nicht vergleichbar, wenn unterschiedliche Messmethoden eingesetzt wurden.

In den Tests konnten für die Vertreter der in Kap. 3.6 beschriebenen Strukturklassen IC_{50} -Werte zwischen 5 μM und 100 μM gemessen werden. Insgesamt konnten von den 39 aktiven Substanzen 14 Substanzen mit einer Aktivität von unter 100 μM gemessen werden. Die so bestätigten *Hits* (*confirmed hits*) ergeben eine *confirmed-Hit*-Rate von 10,07 %.

Referenz-Substanz	IC ₅₀ CDK2 [μM]	IC ₅₀ EtCRK2 [μM]
Alsterpauillon	8	0,2
CVT-313	4	24
9-Cyanopauillon	>200	>200
10Z-Hymenialdisin	0,6	0,5
Aloisin	34	93
Indirubin-5-SO ₃ ⁻	1,5	0,6
Purvalanol A	0,7	0,9
Roscovitin	7	20

Struktur-Nummer	IC ₅₀ CDK2 [μM]	IC ₅₀ EtCRK2 [μM]
00062021	5,31	4,37
00136679	75,53	51,16
00197412	82,25	90,44
00241415	26,44	29,68
00251930	59,63	80,21
00251945	80,11	96,02
00251997	44,27	34,73
00252002	24,73	78,27
00252017	43,48	39,68
00252020	9,52	27,24
00252034	51,41	23,29

Tab. 3.7 IC₅₀-Werte verschiedener Inhibitoren der CDKs. In der oberen Tabelle sind die Werte der Referenz-Substanzen dargestellt. Auffallend sind die unterschiedlichen Aktivitäten für humane CDK2 und EtCRK2. In der unteren Tabelle sind die Werte der gefundenen Inhibitoren dargestellt. Insbesondere die Strukturen 00252020 und 00252002 zeigten deutliche Unterschiede hinsichtlich der Inhibierung der Enzyme.

4. Diskussion

4.1 Evaluierung des *Grid-Computing Frameworks*

Wurden in früheren Jahren Leitstrukturen nach dem Zufallsprinzip gefunden, wird dies heute mit Methoden versucht, die auf einem biochemischen Verständnis der zugrundeliegenden Wirkmechanismen beruhen [SELZER *et al.*, 2000, 2004; SELZER, 2005]. Dabei werden unterschiedliche Methoden miteinander kombiniert, was häufig durch wirtschaftliche und technische Belange notwendig ist. Eine etablierte Methode ist die computerunterstützte Leitstruktursuche [ALVAREZ *et al.*, 2004; LYNE *et al.* 2002]. Dabei werden Ligand- und Proteinstruktur-basierte Verfahren eingesetzt [BISSANTZ *et al.*, 2000, 2003; ZUCCOTTO *et al.* 2003; GEYER *et al.*, 2005]. Diese sind auf heute üblicher Computerhardware effizient und kostengünstig anwendbar. Eine gebräuchliche Methode der Leitstruktursuche ist das virtuelle Screening von großen Substanzdatenbanken mittels *Docking* [KAIRYS *et al.*, 2006]. Einige Beispiele belegen die Effizienz dieser Methode [CUMMINGS *et al.*, 2005]. Diese Methode erfordert nach wie vor relativ hohe Rechenkapazitäten, wenn sie für Millionen Substanzen durchgeführt werden soll [SCHULZ-GASCH & STAHL, 2003]. *Grid-Computing* ist eine Möglichkeit, nicht genutzte Rechenkapazitäten für die Leitstruktursuche bereitzustellen [HERNANDEZ *et al.*, 2005; CHIEN *et al.*, 2002; STEVENS *et al.*, 2003].

Ein Aspekt der vorliegenden Arbeit war die Etablierung einer heterogenen *Grid-Computing*-Infrastruktur und deren Evaluierung für die Anwendung im virtuellen Screening. Im Rahmen dieser Arbeit konnten so nicht genutzte Rechenkapazitäten für ein virtuelles Screening genutzt werden. Dabei konnten hohe Anzahlen von Strukturen, in einer mit anderen Methoden vergleichbaren Geschwindigkeit, prozessiert werden. Häufig werden für die Bereitstellung hoher Rechenkapazitäten dedizierte Systeme eingesetzt, wie z.B. die *Linux-Cluster*. Die Methode des *Grid-Computing* findet dagegen eine weniger breite Anwendung. Ursprünglich wurde das *Grid-Computing* im Rahmen

wissenschaftlicher Berechnungen bei der Auswertung von Radioteleskop-Signalen eingesetzt [[HTTP://SETIATHOME.BERKELEY.EDU/](http://setiathome.berkeley.edu/)]. Mittlerweile kommt *Grid-Computing* auch in einigen biomedizinische Anwendungen zum Einsatz. Zwei Vertreter solcher über das Internet gesteuerten Anwendungen sind das FightAids@Home Projekt, sowie das *United Devices Cancer Research Project* [[HTTP://FIGHTAIDSATHOME.SCRIPPS.EDU/](http://fightaidsathome.scripps.edu/); [HTTP://WWW.GRID.ORG/PROJECTS/CANCER/](http://www.grid.org/projects/cancer/)]. Beide Projekte haben das Ziel, neue Leitstrukturen für unterschiedliche Indikationen zu finden. Bisher konnte die *Grid-Computing* Methode in der Agrar- und Pharmaindustrie nur einen geringen Stellenwert erlangen. Lediglich die Novartis AG setzt die *Grid-Computing* Methode kommerziell für die Leitstruktursuche ein. Weitere Anwendungen der *Grid-Computing* Methode in der Leitstruktursuche konnten in der Literatur nicht gefunden werden.

Für das Molekulare *Docking* von etwa 450.000 Strukturen mit 16 Computern als *Grid-Computing Clients* wurden im Rahmen dieser Arbeit lediglich ca. 40 Tage benötigt. Ein Einsatz mehrerer hundert Computer, wie sie häufig in Unternehmensnetzwerken vorhanden sind, ermöglicht demzufolge noch komplexere Berechnungen in einer wesentlich kürzeren Zeit. Eine Einschränkung des *Grid-Computings* in Anwendungen des wissenschaftlichen Rechnens dürfte die weite Verbreitung des *Windows*-Betriebssystems auf den Computern in Unternehmensnetzwerken darstellen. Viele Software-Applikationen für wissenschaftliches Rechnen sind dagegen an UNIX- und Linux-Betriebssysteme gebunden. Eine heterogene Infrastruktur, wie sie im Rahmen dieser Arbeit etabliert wurde, erfordert jedoch die Anpassung der Software-Applikation an verschiedene Betriebssysteme. Die Software Autodock liegt ebenfalls in einer UNIX-Version vor, weshalb sie auf die *Windows*-Plattform portiert werden musste. Die Unterschiede der verwendeten Betriebssysteme können jedoch die Ergebnisse der Berechnungen maßgeblich beeinflussen. Daher wurde die portierte *Windows*-Version der Software Autodock intensiv im *in silico*-Screening erprobt.

4.1.1 Validierung der *Docking*-Ergebnisse

Eine Validierung der *Docking*-Ergebnisse war aufgrund der Portierung der Software Autodock erforderlich. Zusätzlich zur Portierung wurden einige Modifikationen an der Generierung der Autodock Struktur- und Parameterdateien vorgenommen. Diese Modifikationen beinhalten u.a. die Ergänzung von Parametern für Stickstoff- und Wasserstoffatome, die Wasserstoffbrückenbindungen erzeugen können. Es wurde davon ausgegangen, dass diese Änderungen einen positiven Einfluss auf die *Docking*-Ergebnisse haben. Aus diesem Grund wurden 15 Röntgenkristallstrukturen von Protein-Ligandkomplexen für einen Vergleich der *Dockings* mit Experimentaldaten herangezogen. Von den 15 Protein- Ligandkomplexen konnten 11 mit einem RMSD-Wert von unter 0,2 nm (2 Å) reproduziert werden. Vergleicht man die mit der modifizierten Version der Software Autodock erzielten Ergebnisse mit den in der Literatur beschriebenen Ergebnissen verschiedener *Docking*-Software, konnten vergleichbar gute Ergebnisse erzielt werden [DE GRAAF *et al.*, 2005; ROCKEY *et al.*, 2002].

Im Rahmen der Validierung der *Docking*-Ergebnisse wurden verschiedene Optimierungsmethoden erprobt [SOTRIFFER & DRAMBURG., 2005]. Anhand des *Re-Dockings* experimentell ermittelter Protein-Ligandkomplexe, konnte eine geeignete Kombination unterschiedlicher Optimierungsparameter bestimmt werden. In einem ersten Schritt wurden dabei jene *Docking*-Parameter ermittelt, die eine zuverlässige Reproduzierbarkeit der bekannten Protein-Ligandkomplexe bei einem minimalen Rechenaufwand ermöglichten (Kap. 2.6.7). In den nachfolgenden Experimenten wurden verschiedene Bewertungsfunktionen auf deren Anwendbarkeit für die korrekte Bewertung von CDK-Inhibitoren untersucht. In Anreicherungsexperimenten fiel auf, dass die Wahl der Bewertungsfunktion einen entscheidenden Einfluss hat. Prinzipiell sollte jede Bewertungsfunktion die korrekte Ligandkonformation in der Bindetasche ermitteln und verschiedene Liganden entsprechend ihrer Affinität bewerten. Aus verschiedenen Publikationen war jedoch bekannt, dass die Bewertungs-

funktionen in der Regel zu ungenau sind, um jeden Protein-Ligandkomplex zuverlässig zu bewerten [YOON *et al.*, 2004]. Selbst die Bewertungsfunktionen, die einander ähnliche Terme besitzen, beispielsweise für die Bewertung von Wasserstoffbrücken, bewerten die Protein-Ligandkomplexe häufig sehr unterschiedlich [BISSANTZ *et al.*, 2000].

Daher galt es, eine Kombination von Bewertungsfunktionen zu finden, die eine Unterscheidung zwischen CDK-Inhibitoren und nicht-bindenden Substanzen ermöglicht. Diese konnten mit Jain und Chemscore Bewertungsfunktionen identifiziert werden, welche die Konformationen der eingepassten CDK-Inhibitoren gut voneinander diskriminieren. Zu einem ähnlichen Ergebnis kamen die Autoren einer *Docking*-Studie [COTESTA *et al.*, 2005]. Da die Bewertungsfunktionen in Abhängigkeit von der Beschaffenheit der untersuchten Bindestelle unterschiedliche Resultate liefern können, musste hinterfragt werden, ob Jain und Chemscore für die nachfolgende Rangfolgen-Bewertung geeignet waren. Es wurde daher ein 2-Stufen-Bewertungsschema angewendet, mit dem eine Anreicherung von Einzel-Konformationen der eingepassten Strukturen erzielt werden sollte [YOON *et al.*, 2004]. In dem Anreicherungs-Schritt waren es nicht die Bewertungsfunktionen Jain und Chemscore, sondern die Bewertungsfunktion G-Score, die die besten Resultate lieferte. Interessanterweise ist G-Score ursprünglich als „*Docking*“ Bewertungsfunktion entwickelt worden und nicht für die Vorhersage von Bindungsaffinitäten. Dennoch konnte sie in publizierten Untersuchungen gute Resultate in Anreicherungsexperimenten liefern [WANG *et al.*, 2003].

Erschwerend kam hinzu, dass die Anreicherungsexperimente durch die Zusammensetzung der verwendeten Datensätze eventuell verfälscht wurden. Die Anreicherung bekannter Liganden sollte steigen, wenn die inaktiven Substanzen im Datensatz den aktiven Substanzen strukturell wenig ähnlich sind. Dies ist beispielsweise der Fall, wenn die bekannten Liganden ein höheres Molekulargewicht oder eine andere Anzahl funktioneller Gruppen besitzen, als die nicht-bindenden Substanzen. Umgekehrt sollte die Anreicherung sinken, wenn viele unbekannte Liganden im Datensatz enthalten sind

und mit den bekannten Liganden um die höheren Ränge in der Rangfolge kompetitieren [MUEGGE *et al.*, 2004]. Aus diesem Grund wurde für die Anreicherungsexperimente, die einer Optimierung der Methode dienten, ein Datensatz mit weitgehend ähnlichen Eigenschaften gewählt. Erwartungsgemäß waren unter Verwendung dieses Datensatzes niedrigere Anreicherungsraten der CDK-Inhibitoren im Vergleich mit anderen Datensätzen zu beobachten. Dennoch wurde dieser Datensatz für die weiteren Experimente verwendet, da mit diesem die diskriminierenden Eigenschaften der Bewertungsfunktionen genauer beurteilt werden konnten.

Eine Analyse der Bindungsmodi der in der Literatur beschriebenen CDK-Inhibitoren ließ charakteristische Wechselwirkungen der Inhibitoren mit der *Hinge*-Region der CDKs erkennen [TRAXLER *et al.*, 1999]. Durch den Einsatz eines Interaktionsfilters, der sich dieses Muster zunutze machte, wurden zum einen fälschlicherweise hoch bewertete Substanzen heraussortiert, zum anderen wurden durch diesen Interaktionsfilter auch viele falsch eingepasste CDK-Inhibitoren entfernt. Beides verringerte zwar die Anreicherungsraten der CDK-Inhibitoren, jedoch stieg dadurch die Qualität der Ergebnisse, d.h. die Rate der korrekt eingepassten CDK-Inhibitoren. Dies konnte unter anderem an den RMSD Werten der so gefilterten Ligandkonformationen gezeigt werden. Die Aminosäurezusammensetzung der Proteinkinase-Bindetaschen ist sehr hoch konserviert, sodass die hier verwendeten Methoden auch auf andere Proteinkinasen übertragbar sein sollten. Dies wird auch durch die breite Wirksamkeit der Mehrzahl der bekannten Proteinkinaseinhibitoren auf die unterschiedlichen Proteinkinasen belegt [FABIAN *et al.*, 2005].

4.2 Sequenzvergleiche

Die Sequenzsuche nach CDKs in *E. tenella* wurde mittels Sequenz-Alignments bekannter CDKs durchgeführt. Erschwert wurde die Suche durch die Unvollständigkeit der Sequenzen des *E. tenella* Genoms [[HTTP://WWW.SANGER.AC.UK/PROJECTS/E_TENELLA/](http://www.sanger.ac.uk/projects/e_tenella/)]. Darüberhinaus ist die Ähnlichkeit der CDKs zu anderen Proteinkinasen, wie z.B. den MAP-Kinasen relativ hoch. Dies erschwerte eine klare Abgrenzung in einigen Fällen. Eine genaue Zuordnung parasitärer CDKs zu den Subtypen humaner CDKs ist daher auf Sequenzebene nicht möglich. Die im Rahmen dieser Arbeit gefundene Proteinsequenz der EtCRK2 besitzt eine Identität von ca. 60% sowohl zur Sequenz der humanen CDK1 als auch zur Sequenz der humanen CDK2. Daraus lassen sich daher keine Rückschlüsse auf deren Funktion ziehen, da die Ähnlichkeit zu weiteren CDKs des Menschen ebenfalls sehr hoch ist. Parallel zu der vorliegenden Arbeit wurde mittels biochemischer Experimente die selbe Sequenz der EtCRK2 entdeckt [KINNAIRD *et al.*, 2004]. Die biochemischen Experimente lassen den Schluss zu, dass es sich tatsächlich um eine CDK von *E. tenella* handelt. Die von Kinnaird *et al.* gezeigte Aktivierung der EtCRK2 durch humanes Cyclin A könnte bedeuten, dass es sich um ein Orthologes von CDK2 handeln könnte. Dies muss jedoch durch weitere biochemische und genetische Studien manifestiert werden. Daher wäre es notwendig, weitere CDKs und die dazugehörigen Cycline aus *E. tenella* zu isolieren und in Funktionsstudien, beispielsweise in *knock-out*-Mutagenese Experimenten, zu untersuchen.

4.3 Homologiemodellierung

Etliche Merkmale der Struktur der humanen CDK2 ließen sich auch in der EtCRK2-Sequenz nachweisen. Das Vorkommen dieser Merkmale lässt den Schluss zu, dass sich ein aussagekräftiges Homologiemodell auf der Grundlage der Kristallstrukturen der CDK2 des Menschen generieren lässt. Homologiemodelle sind in anderen Fällen bereits erfolgreich in der Wirkstofffindung und -entwicklung eingesetzt worden [KEENAN & WELSH, 2004; HILLISCH *et al.*, 2004].

Bereits bei der Homologiemodellierung wurden einige strukturelle Merkmale der CDKs dafür benutzt, das Homologiemodell für die Suche nach selektiven Inhibitoren zu optimieren. Eines dieser Merkmale ist der große strukturelle Unterschied zwischen der aktiven und der inaktiven Form der CDKs [DAVIES *et al.*, 2001; HUSE & KURIYAN, 2002]. Es existieren in der Literatur unterschiedliche Meinungen darüber, welcher Aktivierungszustand der Proteinkinase besser geeignet ist, um selektive Inhibitoren zu finden.

Zum einen wird die inaktive Konformation der Proteinkinase als der thermodynamisch stabilere Zustand angesehen, wodurch der Protein-Ligand Komplex zusätzlich stabilisiert werden könnte. Darüberhinaus ist die Bindetasche der inaktiven Konformation im Vergleich zu anderen Proteinkinasen weniger strukturell konserviert [NOBLE *et al.*, 2004]. Beispiele an anderen Proteinkinasen zeigen diese Effekte auf eindrucksvolle Weise. So bewirkt die Bindung des Wirkstoffes Imatinib (Gleevec®) eine Konformationsänderung der betroffenen Proteinkinase, sodass diese im inaktiven Zustand verharrt [NOBLE *et al.*, 2004]. Gleichzeitig führt der Wirkstoff Imatinib in den meisten anderen Proteinkinasen diese Konformationsänderung nicht herbei. Dadurch ist eine hohe Selektivität gewährleistet und der Wirkstoff besitzt ein günstigeres toxikologisches Profil als andere, breiter wirksame Proteinkinaseinhibitoren [FABIAN *et al.*, 2005]. Andererseits ist in der aktiven Konformation der CDK die Bindetasche größer und damit in der Lage, unterschiedlichere Liganden aufzunehmen [THOMAS *et al.*, 2006].

Um die Vorteile beider Ansätze zu kombinieren, wurden für die Templatauswahl für ein Homologiemodell der EtCRK2 ausschließlich Proteinstrukturen der CDKs gewählt, die im inaktiven Zustand mit einem Inhibitor co-kristallisiert sind. Gleichzeitig sind dies auch die häufigsten Proteinstrukturen der CDKs, die in der PDB vorhanden sind. Das erleichtert ein Anpassen des Homologiemodells an Experimentaldaten erheblich. Bei der Auswahl der Template wurde darauf geachtet, dass die Struktur detailliert aufgelöst ist. Bei den *Docking*-Studien an den Templatstrukturen konnte ein offensichtlicher Zusammenhang zwischen einem erfolgreichen *Docking* und der Seitenkettenflexibilität der Aminosäuren in der Bindetasche beobachtet werden. Die Seitenkettenflexibilität wird indirekt über die Debye-Waller Temperaturfaktoren (B-Faktoren) dargestellt [GAO & PENG, 1999]. Sind die Temperaturfaktoren der Atome einer Aminosäure hoch, d.h. größer als ca. 30 \AA^2 , ist von einer relativ hohen Beweglichkeit der Aminosäure auszugehen. Häufig gehen hohe Temperaturfaktoren mit einer schlechten Auflösung der Gesamtstruktur einher. Dadurch ist es schwierig, die exakten Koordinaten von Atomen der Struktur zu bestimmen. Das macht eine genaue Vorhersage der korrekten Ligand-Pose schwierig, da bereits in der Proteinstruktur-Vorhersage Ungenauigkeiten existieren.

Von den *Docking*-Ergebnissen ausgehend, wurden die Proteinstrukturen ausgewählt, die eine hohe Anreicherung bekannter Inhibitoren der humanen CDK2 ermöglichen. Dies waren ausschließlich Strukturen mit niedrigen B-Faktoren der Bindetaschen-Atome. Die so selektierten Proteinstrukturen wurden dann näher untersucht. Dabei wurden Aspekte der Proteinflexibilität der CDKs mit in Betracht gezogen.

Die Flexibilität der CDKs wirkt sich überwiegend auf die Konformation der *Loops* und einzelner Aminosäureseitenketten aus [DAVIES *et al.*, 2001]. Diese Konformationsänderungen können wiederum eine Folge des Aktivierungszustandes der CDKs sein [JOHNSON *et al.*, 2002]. Da für diese Arbeit nur Proteinstrukturen der inaktiven CDKs herangezogen wurden, konnte die Betrachtung von Konformationszuständen auf den Aktivierungs-*Loop*, den

Glycin-reichen *Loop* und die Aminosäure Lys33 beschränkt werden. Nach Bindung der bekannten Inhibitoren konnten keine gravierenden Einflüsse des Aktivierungsloops und des Glycin-reichen Loops auf die Anreicherungsraten der CDK-Inhibitoren festgestellt werden. Jedoch binden die unterschiedlichen CDK-Inhibitoren verschieden stark an die jeweilige Konformation. Eine Möglichkeit, die Proteinflexibilität im *Docking* zu berücksichtigen ist die Verwendung mehrerer Konformationen der CDK. In dieser Arbeit wurde jedoch ein Kompromiss eingegangen, um die Anzahl der Proteinkonformationen im *Docking* nicht weiter zu erhöhen. Das hätte eine drastische Erhöhung des Rechenaufwands zur Folge gehabt. Daher wurden für die Suche nach einer geeigneten Templatstruktur folgende Kriterien berücksichtigt: Mit der Templatstruktur sollten sich die Bindemodi der meisten bekannten CDK-Inhibitoren korrekt reproduzieren lassen. Darüberhinaus galt ein besonderes Augenmerk der Konformation der Aminosäure-Seitenkette des Lys33, die eine direkte Interaktion mit einzelnen Liganden ermöglichen sollte, aber die Ligandbindung von anderen Liganden nicht aus sterischen Gründen verhindern darf. Das bedeutet, dass das Volumen der Bindetasche ausreichen muss, um relativ große Liganden aufzunehmen. Alle diese Kriterien wurden von der PDB-Struktur 1OIR erfüllt.

Bei der weiteren Analyse des Modells fiel auf, dass die Aminosäure Lys89 (humane CDK2) an der homologen Position gegen ein Glutamat (EtCRK2, Glu88) ausgetauscht ist. Aus Untersuchungen an den Proteinkristallstrukturen 1CKP und 1H1S war bekannt, dass einige Inhibitoren mit dieser Aminosäure interagieren, was zu einer Affinitätssteigerung und zum Teil zu einer höheren Selektivität der untersuchten Liganden führte [ANDERSON *et al.*, 2003; DAVIES *et al.*, 2001; GRAY *et al.*, 1999]. Daraus lässt sich ableiten, dass, obwohl die homologe Aminosäure Glu88 am Rand der Bindetasche platziert ist, diese mit in die Entwicklung von selektiven Inhibitoren einbezogen werden kann. Durch den Austausch der positiv geladenen Aminosäure Lysin mit der negativ geladenen Aminosäure Glutamat ist es möglich, einen Liganden mit einer komplementären Eigenschaft zu finden. Es ist bekannt, dass ionische Wech-

selwirkungen besonders stark sind [BOEHM *et al.*, 1996; BOEHM & STAHL, 2002]. Ein Sequenzvergleich mit der Sequenz von CDK3 des Huhns zeigt, dass auch die CDK3 des Huhns ein Lysin an dieser Position besitzt. Aus diesem Grund wurde das entsprechende Pharmakophor in die Pharmakophormodelle der EtCRK2 einbezogen.

4.4 Pharmakophormodellierung

Vergleiche der für diese Arbeit generierten Pharmakophormodelle der EtCRK2 mit Pharmakophormodellen von CDKs anderer *Apicomplexa* zeigten große Übereinstimmungen [BHATTACHARJEE *et al.*, 2005; GEYER *et al.*, 2005]. Die in der Literatur beschriebenen Pharmakophormodelle von Pfmrk -einer CDK aus *P. falciparum*- konnten erfolgreich für eine Datenbanksuche nach neuen Inhibitoren der Pfmrk verwendet werden. Die Generierung der Pharmakophormodelle für *P. falciparum* wurde mit einem ligandbasierten Ansatz durchgeführt. Dafür waren bekannte Inhibitoren der Pfmrk verwendet worden [WOODARD *et al.*, 2003]. Für EtCRK2 sind bislang keine Inhibitoren bekannt. Deshalb wurden die Pharmakophormodelle von dem Homologiemodell der EtCRK2 abgeleitet. Durch den Vergleich mit bekannten Bindungsmodi von anderen CDKs können auf diese Weise sehr exakte Pharmakophormodelle erzeugt werden. Allerdings zeigte sich während der Generierung der Pharmakophormodelle der EtCRK2, dass diese Modelle lediglich als Filter von Substanzdatenbanken geeignet sind, da eine relativ hohe Anzahl von Substanzen aus einer Suche mit diesen Pharmakophormodellen resultierte. Für die Arbeiten an Pfmrk wurde die Pharmakophorsuche durch eine weitere „Shape“ Restriktion eingeschränkt. Dabei werden Sphären um die Ligandatome generiert, innerhalb derer die auf ein Pharmakophormodell passenden Atome eines möglichen Liganden ebenfalls hineinpassen müssen. Im Rahmen dieser Arbeit wurde auf eine zusätzliche Restriktion verzichtet, da der Schritt des anschließenden *Docking* eine genauere Diskriminierung ermöglichen sollte. Darüberhinaus ist die Anwendung von Pharmakophor-

Ensembles besser geeignet, um neuartige Strukturgerüste von Inhibitoren zu entdecken. Durch eine *Shape*-Restriktion könnte diese Möglichkeit stark eingeschränkt werden.

Untersuchungen mit Testdatensätzen und den Pharmakophor-Ensembles zeigten erwartungsgemäß sehr unterschiedliche Anreicherungsraten. Dabei kann ein hoher Anreicherungsfaktor als ein Maß einer hohen Selektivität der Hypothese angesehen werden. Das Ensemble 2 ermöglichte zwar nicht die höchsten Anreicherungen, jedoch waren die Anreicherungsfaktoren dieser Hypothese bei allen verwendeten Datensätzen am höchsten. Dies könnte durch die Einbeziehung der Interaktionen der Aminosäuren Lys32 („Wasserstoffbrückenakzeptor“ im Pharmakophormodell) und Asp85 („positiv ionisierbar“ im Pharmakophormodell) hervorgerufen werden. Dies stellt ein Selektivitätskriterium für CDK-Inhibitoren dar. Vergleicht man das Ensemble 1 mit den anderen Ensembles, so fällt auf, dass Ensemble 1 die höchsten Anreicherungsraten ermöglicht. Die Ursache dafür könnte in dem Pharmakophor (Wasserstoffbrücken-Donor) der Ligand-Asp85 Interaktion liegen, die eine Anpassung an häufiger vorkommender Strukturen erlaubt.

Mit den Pharmakophor-Ensembles wurden Anreicherungsexperimente mit verschiedenen Datensätzen durchgeführt. Dabei wurden immer identische Strukturen bekannter CDK-Inhibitoren verwendet und mit Strukturen unbekannter Aktivität vermischt. Überraschenderweise konnte im CDK-Datensatz die geringste Anzahl von Strukturen unbekannter Aktivität gefunden werden. Hier waren demzufolge die höchsten Anreicherungsfaktoren von bekannten CDK-Inhibitoren zu beobachten. Von einem Datensatz, der neben den bekannten CDK-Inhibitoren viele unbekannte CDK-Inhibitoren enthält, wäre eigentlich das Gegenteil zu erwarten gewesen [MUEGGE *et al.*, 2004]. Dies könnte unterschiedliche Gründe haben. Zum einen ist denkbar, dass die Substanzen durch verschiedene chemische Optimierungen eine größere Anzahl rotierbarer Bindungen -und somit mehr konformationelle Freiheitsgrade besitzen. Dadurch ist womöglich der Konformationsraum nur ungenügend durch den Konformergenerator der Software Catalyst abgebildet worden.

Wahrscheinlicher ist, dass die Strukturen des Datensatzes tatsächlich nicht gut mit den Pharmakophormodellen übereinstimmen. Dies könnte methodische Ursachen der Ausrichtung der zugrundeliegenden *CDK-focussed library* des Datensatzes haben. Zu einem ähnlichen Ergebnis kamen bereits andere Autoren [MUEGGE *et al.*, 2004].

Mit Hilfe der Pharmakophor-Ensembles wurde ein Test-Datensatz generiert. Mittels *Docking*-Experimenten an dem Homologiemodell konnte die Qualität des Pharmakophor-Filters evaluiert werden. Dabei ergab sich folgendes Bild: Einerseits zeigt der Kurvenverlauf der Anreicherungskurve in den ersten 20% des Pharmakophor-Datensatzes die geringste Steigung. Dies spricht für eine hohe Anzahl von unbekanntem CDK-Inhibitoren in dem Datensatz, die mit den bekannten CDK-Inhibitoren um die hohen Bewertungs-Ränge konkurrieren. Im Gegensatz dazu enthalten die kommerziell vorbereiteten *Target*-spezifischen Datenbanken offensichtlich weniger Strukturen, die den Pharmakophormodellen entsprechen, da hier relativ gute Anreicherungen von CDK-Inhibitoren erreicht werden konnten.

4.5 Virtuelles Screening

Für die Suche nach neuartigen Inhibitoren von der parasitären EtCRK2 wurde die im Rahmen der Evaluierung des *Grid-Computing Frameworks* etablierte Methode des Screenings humaner CDK-Inhibitoren verwendet. Die Anwendung der Methode geschah unter anderem unter der Annahme, dass die Bindetaschen parasitärer CDKs zu denen humaner CDKs so ähnlich sind, dass die Methodik übertragbar ist. Diese Ähnlichkeit konnte anhand von Sequenz-*Alignments* und Strukturvergleichen gezeigt werden. Aus Experimenten an verschiedenen *Apicomplexa* war zudem bekannt, dass CDK-Inhibitoren des Menschen auch das Wachstum von verschiedenen Protozoen hemmen [BHATTACHARJEE *et al.*, 2005; HARMSE *et al.*, 2001, GEYER *et al.*, 2005; WOODARD *et al.*, 2003].

4.6 Auswahl von Substanzen für biochemische Experimente

Nach dem erfolgreichen Screening nach Inhibitoren der EtCRK2 wurden 195 Substanzen für biochemische Tests ausgewählt. Insbesondere fiel auf, dass sich die Strukturklassen der Naphtholactame, der Phthalimide, der Rhodanine, der Purine, der Barbiturate und der Oxindole stark in dieser Auswahl häuften. Insgesamt decken diese 6 Strukturklassen rund die Hälfte der ausgewählten Substanzen ab. Von den Purinen, den Phthalimiden, den Naphtholactamen und den Oxindolen ist seit längerem bekannt, dass einige Vertreter dieser Strukturklassen Inhibitoren humaner CDKs sind [ROUTIER *et al.*, 2005; GRAY *et al.*, 1999; HARDCASTLE *et al.*, 2004; BRAMSON *et al.*, 2001; LUK *et al.*, 2004; LIU *et al.*, 2003]. Die Rhodanine sind als Inhibitoren der Glycogensynthase GSK-3 beschrieben worden, während für die Naphtholactame und die Barbiturate eine Hemmung der Proteinkinase-C beschrieben wurde [IRIE *et al.*, 2005; DESHMUKH *et al.*, 1989]. Die Kinasen Glykogensynthase und die CDKs sind tatsächlich eng verwandt [MANNING *et al.*, 2002]. Die Häufung von Strukturklassen mit bekannter Proteinkinaseaktivität deutet darauf hin, daß während des virtuellen Screenings eine Anreicherung von CDK-Inhibitoren stattgefunden hat.

Die Aktivitätsbestimmung von Vertretern der jeweiligen Strukturklassen konnte zeigen, dass die *in silico*-Methode zu einer hohen Anreicherung von CDK-Inhibitoren geführt hat. Das Vorhandensein von Inhibitoren mit einer relativ hohen Affinität demonstriert, dass die Nutzung eines Homologiemodells vollkommen ausreichend sein kann, um damit erfolgreich Wirkstoffe zu finden. Der in dieser Arbeit beschrittene Weg, zunächst relativ unspezifische Filter einzusetzen, um die Datenmenge zu reduzieren, ermöglichte in Kombination mit dem Hochdurchsatz-*Docking* das Auffinden neuer potenzieller CDK-Inhibitoren.

4.7 Ausblick

Die EtCRK2 weist eine Reihe von Unterschieden zu Vertebraten-CDKs auf. Dies ist eine wichtige Voraussetzung bei der Entwicklung neuer Chemotherapeutika, die spezifisch das Enzym des Parasiten inhibieren sollen ohne die Enzyme des Wirtsorganismus zu beeinträchtigen. Da die EtCRK2 mittlerweile kloniert und die Aminosäuresequenz bekannt ist, konnte eine erste biochemische Charakterisierung der EtCRK2 bereits durchgeführt werden [KINNAIRD *et al.*, 2004].

Derzeit werden bekannte CDK-Inhibitoren und die in der vorliegenden Arbeit gefundenen Inhibitoren der EtCRK2 dahingehend untersucht, weitere Informationen über die molekularen Grundlagen der Unterschiede der Vertebraten-CDKs und der EtCRK2 zu liefern. Die biochemischen Experimente können dazu beitragen, spezifische Merkmale von Inhibitoren der EtCRK2 zu identifizieren, die für die Struktur-Wirkungsbeziehung verantwortlich sind. Mit diesen Ergebnissen ließe sich zum einen das Homologiemodell der EtCRK2 so optimieren, dass es mit den Experimentaldaten noch besser korreliert, zum anderen ließen sich damit Moleküle entwerfen, die eine höhere Selektivität und Affinität für die EtCRK2 besitzen.

Der Bindungsmodus der gefundenen Inhibitoren lässt sich allerdings erst durch ein Experiment endgültig bestätigen. Diese Bestätigung kann über die Röntgenstruktur-Aufklärung eines co-kristallisierten Inhibitors mit der EtCRK2 erfolgen. Dennoch erfährt man auf diese Weise nicht, welche weiteren Kinasen möglicherweise ebenfalls durch diesen Inhibitor gehemmt werden. Dafür müssen weitere biochemische Experimente mit einer Reihe anderer Kinasen durchgeführt werden. Zudem ist bisher kein Cyclin aus *E. tenella* gefunden worden, dessen Funktion im Komplex mit der EtCRK2 bestätigt worden ist. Auch die Substrate der EtCRK2 sind unbekannt. In unserer Arbeitsgruppe wurden bereits einige bioinformatische Untersuchungen durchgeführt, die weitere mögliche CDKs aus *E. tenella* identifiziert haben. Gleichzeitig wurden mögliche Cycline identifiziert, von denen eines bereits kloniert wurde.

Erste Expressionsversuche mit diesen Proteinen werden bereits durchgeführt. Damit sollte es möglich sein, die funktionell aktiven CDK-Cyclin-Komplexe des Zellzyklus von *E. tenella* zu identifizieren. Funktionsstudien an *E. tenella* müssen darüberhinaus klären, welche Rolle die EtCRK2 im Zellzyklus von *E. tenella* hat. Diese Arbeiten sollten zudem eine Optimierung von Inhibitoren der EtCRK2 ermöglichen, die letztendlich den Einsatz zur kurativen, bzw. prophylaktischen Anwendung am Huhn erlaubt.

5 Zusammenfassung

Die computerunterstützte Wirkstoffforschung ist eine etablierte Methode im Entwicklungsprozess neuer Wirksubstanzen. Obwohl sich die Rechenleistung von Computern alle 5 Jahre verdoppelt hat, genügen einzelne Computer in der Regel nicht, die immer aufwendigeren Berechnungen durchzuführen. Daher werden häufig viele Computer in Netzwerken zusammengefasst, um die Berechnungen zu parallelisieren. Eine Methode, Computernetzwerke für die Wirkstoffsuche zu verwenden, ist das *Grid-Computing*. Solche Computernetzwerke existieren in Unternehmen und Universitäten hauptsächlich für die Nutzung von Büroanwendungen. Im Gegensatz zu anderen Infrastrukturen, beispielsweise den *Linux-Clustern*, werden im *Grid-Computing* die bereits vorhandenen Computer zusätzlich für komplexe Berechnungen verwendet. Im Rahmen der vorliegenden Arbeit wurde eine *Grid-Computing* Infrastruktur in ein bestehendes Computernetzwerk implementiert. Die Methode des *Grid-Computing* wurde mit einer *Docking*-Software kombiniert. Mit diesem *Framework* konnte gezeigt werden, dass die Methode des *Grid-Computing* für das *in-silico*-Screening im Rahmen der Wirkstoffforschung genutzt werden kann.

Ein Zielorganismus in der veterinärmedizinischen Forschung ist der Parasit *Eimeria tenella*. Dieser Protozoe befällt u.a. den Darm des Huhns, was zum Krankheitsbild der Kokzidiose führt. Symptome der akuten Kokzidiose sind starke, teils hämorrhagische Durchfälle. Dabei wird das Darmepithel großer Darmabschnitte zerstört, was zu erheblichen pathophysiologischen Störungen führt und schließlich zum Tod des Huhns führen kann. Durch Infektionen mit *E. tenella* werden weltweit große Schäden in der kommerziellen Geflügelzucht verursacht. Die zunehmende Resistenzentstehung gegen die gängigen kokzidiostatischen und kokzidioziden Wirkstoffe macht die Entwicklung von neuen Wirkstoffen zur kurativen und prophylaktischen Behandlung der Kokzidiose notwendig.

Aus vorangegangenen Arbeiten war bekannt, dass der Proteinkinase-Inhibitor Flavopiridol auch die Vermehrung von *E. tenella* in Kultur inhibiert. Flavopiridol ist ein Inhibitor der Cyclin-abhängigen Kinasen (CDK) und befindet sich derzeit in klinischen Studien zur Behandlung verschiedener humaner Krebserkrankungen. Daher erscheint es naheliegend, dass die Wirkung von Flavopiridol auf die Vermehrung von *E. tenella* auf die Inhibierung von CDKs von *E. tenella* zurückzuführen ist. Deshalb können die CDKs von *E. tenella* als geeignete Zielmoleküle für die Entwicklung einer Chemotherapie der Kokzidiose angesehen werden.

Bisher waren keine Informationen über CDKs dieses Parasiten bekannt. Jedoch wurde im Jahr 2002 mit der systematischen Sequenzierung des Genoms von *E. tenella* begonnen, sodass zumindest teilweise Sequenzinformationen dieses Parasiten über das Internet verfügbar sind. Die Anwendung von bioinformatischen Methoden ermöglichte es, die EtCRK2, eine putative CDK von *E. tenella*, zu identifizieren. Die genomischen Sequenzinformationen konnten in eine Aminosäuresequenz translatiert werden. Mit dieser Sequenz konnte ein Homologiemodell der Proteinstruktur der EtCRK2 generiert werden. Dieses Modell wurde in einem *in-silico*-Screening von potenziellen Inhibitoren mittels des *Grid-Computing* Frameworks verwendet. Darüberhinaus konnten anhand des Modells Unterschiede zwischen Vertebraten-CDKs und der EtCRK2 gefunden werden, die eine selektive Inhibierung der EtCRK2 möglich erscheinen lassen. Im *in-silico*-Screening konnten einige Inhibitoren der EtCRK2 gefunden werden, die zu Strukturklassen gehören, die bisher nicht als Inhibitoren anderer CDKs beschrieben sind. Die Ergebnisse der biochemischen Experimente zeigen deutliche Unterschiede in der Inhibierung von einer humanen CDK und der EtCRK2. Diese Ergebnisse lassen hoffen, auf der Grundlage der gefundenen Inhibitoren einen neuen Wirkstoff für die Kokzidiosetherapie zu finden.

6 Literaturverzeichnis

- Altschul S F, Gish W, Miller W, Myers E W, Lipman D J. Basic local alignment search tool. *J Mol Biol* 1990; (215): 403-410.
- Alvarez J C. High-throughput docking as a source of novel drug leads. *Current Opinion in Chemical Biology* 2004; (8): 365-370.
- Anderson M, Beattie J F, Breault G A, Breed J, Byth K F, Culshaw J D, Ellston R P, Green S, Minshull C A, Norman R A, Pauptit R A, Stanway J, Thomas A P, Jewsbury P J. Imidazo[1,2-a]pyridines: a potent and selective class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation. *Bioorg Med Chem Lett* 2003; (13): 3021-3026.
- Beckstette, M.; Mailänder, J. T.; Marhöfer, R. J.; Sczyrba, A.; Ohlebusch, E.; Giegerich, R.; Selzer, P. M. Yearbook Bioinformatics. Ralf Hofestädt (Ed.) *Journal of Integrative Bioinformatics*. 2004; 1, 79-94.
- Benson D A, Karsch-Mizrachi I, Lipman D J, Ostell J, Wheeler D L. GenBank: update. *Nucleic Acids Res* 2004; (32): D23-D26.
- Berman H M, Battistuz T, Bhat T N, Bluhm W F, Bourne P E, Burkhardt K, Feng Z, Gilliland G L, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook J D, Zardecki C. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 2002; (58): 899-907.
- Bernacki K, Kalyanaraman C, Jacobson M P. Virtual ligand screening against Escherichia coli dihydrofolate reductase: improving docking enrichment using physics-based methods. *J Biomol Screen* 2005; (10): 675-681.

- Bhattacharjee A K, Geyer J A, Woodard C L, Kathcart A K, Nichols D A, Prigge S T, Li Z, Mott B T, Waters N C. A Three-Dimensional in Silico Pharmacophore Model for Inhibition of Plasmodium falciparum Cyclin-Dependent Kinases and Discovery of Different Classes of Novel Pfmrk Specific Inhibitors. *J Med Chem* 2004; (47): 5418-5426.
- Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* 2000; (43): 4759-4767.
- Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* 2003; (50): 5-25.
- Boehm HJ, Stahl M. The use of scoring functions in drug discovery applications. *Reviews in Computational Chemistry* 18, 41-87. 2002.
- Bradham C, McClay D R. p38 MAPK in development and cancer. *Cell Cycle* 2006; (5): 824-828.
- Bramson H N, Corona J, Davis S T, Dickerson S H, Edelstein M, Frye S V, Gampe R T, Jr., Harris P A, Hassell A, Holmes W D, Hunter R N, Lackey K E, Lovejoy B, Luzzio M J, Montana V, Rocque W J, Rusnak D, Shewchuk L, Veal J M, Walker D H, Kuyper L F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J Med Chem* 2001; (44): 4339-4358.
- Brandon, C.; Tooze, J. Introduction to Protein Structure. *Garland Publishing Inc.* New York & London 1999.
- Breitenlechner C B, Bossemeyer D, Engh R A. Crystallography for protein kinase drug design: PKA and SRC case studies. *Biochim Biophys Acta* 2005; (1754): 38-49.

- Brickmann, J.; Goetze, T.; Heiden, W.; Moeckel, G.; Reiling, S.; Vollhardt, H.; Zachmann, C.-D. Interactive Visualization of Molecular Scenarios with MOLCAD/SYBYL. *Addison-Wesley-Verlag* 1995; 83-97.
- Brickmann, J.; Exner, T. E.; Keil, M.; Marhöfer, R. J. Molecular Graphics – Trends and Perspectives. *J. Mol. Mod.* 2000; 6, 328-340.
- Brooks B R, Bruccoleri R E, Olafson B D, States D J, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 1983; (4): 187-217.
- Buijsman, R. Structural Aspects of Kinases and Their Inhibitors. In: Chemogenomics in Drug Discovery. Kubinyi, H.; Müller, G. (Ed) *Wiley-VCH Verlag GmbH & Co. KGaA*, Weinheim. 2004; pp. 191-219
- Cai X, Fuller A L, McDougald L R, Zhu G. Apicoplast genome of the coccidian *Eimeria tenella*. *Gene* 2003; (321): 39-46.
- Chapman T. Drug discovery: the leading edge. *Nature* 2004; (430): 109-115.
- Chien A, Foster I, Goddette D. Grid technologies empowering drug discovery. *Drug Discov Today* 2002; (7): S176-S180.
- Chow J P, Siu W Y, Ho H T, Ma K H, Ho C C, Poon R Y. Differential contribution of inhibitory phosphorylation of CDC2 and CDK2 for unperturbed cell cycle control and DNA integrity checkpoints. *J Biol Chem* 2003; (278): 40815-40828.
- Cotesta S, Giordanetto F, Trosset J Y, Crivori P, Kroemer R T, Stouten P F, Vulpetti A. Virtual screening to enrich a compound collection with CDK2 inhibitors using docking, scoring, and composite scoring models. *Proteins* 2005; (60): 629-643.
- Cummings M D, DesJarlais R L, Gibbs A C, Mohan V, Jaeger E P. Comparison of automated docking programs as virtual screening tools. *J Med Chem* 2005; (48): 962-976.

- Davies T G, Tunnah P, Meijer L, Marko D, Eisenbrand G, Endicott J A, Noble M E. Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/indirubin-5-sulphonate. *Structure (Camb)* 2001; (9): 389-397.
- Davies T G, Bentley J, Arris C E, Boyle F T, Curtin N J, Endicott J A, Gibson A E, Golding B T, Griffin R J, Hardcastle I R, Jewsbury P, Johnson L N, Mesguiche V, Newell D R, Noble M E, Tucker J A, Wang L, Whitfield H J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat Struct Biol* 2002; (9): 745-749.
- Davies T G, Pratt D J, Endicott J A, Johnson L N, Noble M E. Structure-based design of cyclin-dependent kinase inhibitors. *Pharmacol Ther* 2002; (93): 125-133.
- de Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen N P. Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J Med Chem* 2005; (48): 2308-2318.
- Deshmukh D S, Kuizon S, Chauhan V P, Brockerhoff H. Effect of barbiturates on polyphosphoinositide biosynthesis and protein kinase C activity in synaptosomes. *Neuropharmacology* 1989; (28): 1317-1323.
- Doerig C, Chakrabarti D, Kappes B, Matthews K. The cell cycle in protozoan parasites. *Prog Cell Cycle Res* 2000; (4): 163-183.
- Doerig C, Endicott J, Chakrabarti D. Cyclin-dependent kinase homologues of Plasmodium falciparum. *Int J Parasitol* 2002; (32): 1575-1585.
- Doerig C, Meijer L, Mottram J C. Protein kinases as drug targets in parasitic protozoa. *Trends Parasitol* 2002; (18): 366-371.
- Elledge S J. Cell cycle checkpoints: preventing an identity crisis. *Science* 1996; (274): 1664-1672.

- Endicott J A, Noble M E, Tucker J A. Cyclin-dependent kinases: inhibition and substrate recognition. *Curr Opin Struct Biol* 1999; (9): 738-744.
- Fabian M A, Biggs W H, III, Treiber D K, Atteridge C E, Azimioara M D, Benedetti M G, Carter T A, Ciceri P, Edeen P T, Floyd M, Ford J M, Galvin M, Gerlach J L, Grotzfeld R M, Herrgard S, Insko D E, Insko M A, Lai A G, Lelias J M, Mehta S A, Milanov Z V, Velasco A M, Wodicka L M, Patel H K, Zarrinkar P P, Lockhart D J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* 2005; (23): 329-336.
- Ferrara P, Gohlke H, Price D J, Klebe G, Brooks C L, III. Assessing scoring functions for protein-ligand interactions. *J Med Chem* 2004; (47): 3032-3047.
- Fischer P M, Gianella-Borradori A. Recent progress in the discovery and development of cyclin-dependent kinase inhibitors. *Expert Opin Investig Drugs* 2005; (14): 457-477.
- Fiser A, Do R K, Sali A. Modeling of loops in protein structures. *Protein Sci* 2000; (9): 1753-1773.
- Fradera X, Mestres J. Guided docking approaches to structure-based design and screening. *Curr Top Med Chem* 2004; (4): 687-700.
- Furet P. X-ray crystallographic studies of CDK2, a basis for cyclin-dependent kinase inhibitor design in anti-cancer drug research. *Curr Med Chem Anti -Canc Agents* 2003; (3): 15-23.
- Gao, H. X.; Peng, L. M. Parameterization of the temperature dependence of the Debye-Waller factors. *Acta. Crystallogr. A.* 1999; 55, 926-932.
- Garcia-Sosa A T, Mancera R L, Dean P M. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *J Mol Model* (Online) 2003; (9): 172-182.

- Gasteiger J, Li X, Uschold A. The beauty of molecular surfaces as revealed by self-organizing neural networks. *J Mol Graph* 1994; (12): 90-97.
- Geyer J A, Prigge S T, Waters N C. Targeting malaria with specific CDK inhibitors. *Biochim Biophys Acta* 2005.
- Gomez E B, Santori M I, Laria S, Engel J C, Swindle J, Eisen H, Szankasi P, Tellez-Inon M T. Characterization of the *Trypanosoma cruzi* Cdc2p-related protein kinase 1 and identification of three novel associating cyclins. *Mol Biochem Parasitol* 2001; (113): 97-108.
- Good A. C., Mason J. S., Pickett S. D. Pharmacophore Pattern Application in Virtual Screening, Library Design and QSAR. In: Virtual Screening for Bioactive Molecules. Böhm H.-J., Schneider G., Mannhold R., Kubinyi H., Timmerman H.. (Ed) *Wiley-VCH Verlag GmbH & Co. KGaA*, Weinheim. 2000; pp. 131-156
- Goossens H. European status of resistance in nosocomial infections. *Chemotherapy* 2005; (51): 177-181.
- Grassly N C, Fraser C. Seasonal infectious disease epidemiology. *Proc Biol Sci* 2006; (273): 2541-2550.
- Graves A P, Brenk R, Shoichet B K. Decoys for docking. *J Med Chem* 2005; (48): 3714-3728.
- Gray N, Detivaud L, Doerig C, Meijer L. ATP-site directed inhibitors of cyclin-dependent kinases. *Curr Med Chem* 1999; (6): 859-875.
- Greer J. Comparative modeling methods: application to the family of the mammalian serine proteases. *Proteins* 1990; (7): 317-334.
- Hafeez M A, Akhtar M, Ayaz M M. Some observations on the adaptation of *Eimeria tenella* (local isolates) sporozoites on chicken embryos through chorioallantoic membrane. *J Vet Sci* 2006; (7): 59-61.

- Hardcastle I R, Arris C E, Bentley J, Boyle F T, Chen Y, Curtin N J, Endicott J A, Gibson A E, Golding B T, Griffin R J, Jewsbury P, Menyerol J, Mesguiche V, Newell D R, Noble M E, Pratt D J, Wang L Z, Whitfield H J. N2-substituted O6-cyclohexylmethylguanine derivatives: potent inhibitors of cyclin-dependent kinases 1 and 2. *J Med Chem* 2004; (47): 3710-3722.
- Harmse L, van Zyl R, Gray N, Schultz P, LeClerc S, Meijer L, Doerig C, Havlik I. Structure-activity relationships and inhibitory effects of various purine derivatives on the in vitro growth of *Plasmodium falciparum*. *Biochem Pharmacol* 2001; (62): 341-348.
- Hernandez V, Blanquer I. The Grid as a healthcare provision tool. *Methods Inf Med* 2005; (44): 144-148.
- Hillisch A, Pineda L F, Hilgenfeld R. Utility of homology models in the drug discovery process. *Drug Discov Today* 2004; (9): 659-669.
- Hogan D, Kolter R. Why are bacteria refractory to antimicrobials? *Curr Opin Microbiol* 2002; (5): 472-477.
- Holton S, Merckx A, Burgess D, Doerig C, Noble M, Endicott J. Structures of *P. falciparum* PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition. *Structure (Camb)* 2003; (11): 1329-1337.
- Huse M, Kuriyan J. The conformational plasticity of protein kinases. *Cell* 2002; (109): 275-282.
- Ihlenfeldt W D, Gasteiger J. Beyond the hyperactive molecule: search, salvage and visualization of chemical information from the Internet. *Pac Symp Biocomput* 1996; 384-395.
- Ihlenfeldt W D, Voigt J H, Bienfait B, Oellien F, Nicklaus M C. Enhanced CACTVS browser of the Open NCI Database. *J Chem Inf Comput Sci* 2002; (42): 46-57.

- Jeffrey P D, Russo A A, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich N P. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 1995; (376): 313-320.
- Johnson D G, Walker C L. Cyclins and cell cycle checkpoints. *Annu Rev Pharmacol Toxicol* 1999; (39): 295-312.
- Johnson L N, Noble M E, Owen D J. Active and inactive protein kinases: structural basis for regulation. *Cell* 1996; (85): 149-158.
- Johnson L N, De Moliner E, Brown N R, Song H, Barford D, Endicott J A, Noble M E. Structural studies with inhibitors of the cell cycle regulatory kinase cyclin-dependent protein kinase 2. *Pharmacol Ther* 2002; (93): 113-124.
- Kairys V, Fernandes M X, Gilson M K. Screening drug-like compounds by docking to homology models: a systematic study. *J Chem Inf Model* 2006; (46): 365-379.
- Keenan S M, Welsh W J. Characteristics of the Plasmodium falciparum PK5 ATP-binding site: implications for the design of novel antimalarial agents. *Journal of Molecular Graphics and Modelling* 2004; (22): 241-247.
- Kim K S, Kimball S D, Misra R N, Rawlins D B, Hunt J T, Xiao H Y, Lu S, Qian L, Han W C, Shan W, Mitt T, Cai Z W, Poss M A, Zhu H, Sack J S, Tokarski J S, Chang C Y, Pavletich N, Kamath A, Humphreys W G, Marathe P, Bursuker I, Kellar K A, Roongta U, Batorsky R, Mulheron J G, Bol D, Fairchild C R, Lee F Y, Webster K R. Discovery of aminothiazole inhibitors of cyclin-dependent kinase 2: synthesis, X-ray crystallographic analysis, and biological activities. *J Med Chem* 2002; (45): 3905-3927.
- Kinnaird J H, Bumstead J M, Mann D J, Ryan R, Shirley M W, Shiels B R, Tomley F M. EtCRK2, a cyclin-dependent kinase gene expressed during the sexual and asexual phases of the Eimeria tenella life cycle. *Int J Parasitol* 2004; (34): 683-692.

- Kitchen D B, Decornez H, Furr J R, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004; (3): 935-949.
- Klebe G., Grädler U., Grüneberg S., Krämer O., Gohlke H. Understanding Receptor-Ligand Interactions as a Prerequisite for Virtual Screening. In: Virtual Screening for Bioactive Molecules. Böhm H.-J., Schneider G., Mannhold R., Kubinyi H., Timmerman H. (Ed) *Wiley-VCH Verlag GmbH & Co. KGaA*, Weinheim. 2000; pp. 207-226
- Knighton D R, Zheng J H, Ten Eyck L F, Ashford V A, Xuong N H, Taylor S S, Sowadski J M. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 1991; (253): 407-414.
- Korf, I.; Yandell, M.; Bedell, J. BLAST. *Sebastopol.: O'Reilly & Associates* 2003.
- Laskowski R A, Moss D S, Thornton J M. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 1993; (231): 1049-1067.
- Le Roch K, Sestier C, Dorin D, Waters N, Kappes B, Chakrabarti D, Meijer L, Doerig C. Activation of a Plasmodium falciparum cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a P. falciparum cyclin homologue. *J Biol Chem* 2000; (275): 8952-8958.
- Liu J J, Dermatakis A, Lukacs C, Konzelmann F, Chen Y, Schutt A, Simcox M E, Luk K C. 3,5,6-Trisubstituted naphthostyrils as CDK2 inhibitors. *Bioorg Med Chem Lett* 2003; (15):2465-2468
- Luk K C, Simcox M E, Schutt A, Rowan K, Thompson T, Chen Y, Kammlott U, DePinto W, Dunten P, Dermatakis A. A new series of potent oxindole inhibitors of CDK2. *Bioorg Med Chem Lett* 2004; (14): 913-917.
- Lyne P D. Structure-based virtual screening: an overview. *Drug Discov Today* 2002; (7): 1047-1055.

- Malumbres M, Barbacid M. Mammalian cyclin-dependent kinases. *Trends Biochem Sci* 2005; (30): 630-641.
- Manning G, Whyte D B, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* 2002; (298): 1912-1934.
- Maroli M, Khoury C. [Prevention and control of leishmaniasis vectors: current approaches]. *Parassitologia* 2004; (46): 211-215.
- McAllister M M. A decade of discoveries in veterinary protozoology changes our concept of „subclinical“ toxoplasmosis. *Vet Parasitol* 2005; (132): 241-247.
- Mehlhorn, H. Encyclopedic reference of parasitology. *Springer*, Berlin, Heidelberg, New York. 2001
- Merckx A, Le Roch K, Nivez M P, Dorin D, Alano P, Gutierrez G J, Nebreda A R, Goldring D, Whittle C, Patterson S, Chakrabarti D, Doerig C. Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite *Plasmodium falciparum*. *J Biol Chem* 2003; (278): 39839-39850.
- Morgan D O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* 1997; (13): 261-291.
- Morris G M, Goodsell D S, Huey R, Olson A J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* 1996; (10): 293-304.
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* 19[14], 1639-1662. 1998.
- Muegge I, Enyedy I J. Virtual screening for kinase targets. *Curr Med Chem* 2004; (11): 693-707.

- Mulder N J, Apweiler R, Attwood T K, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, Copley R, Courcelle E, Das U, Durbin R, Fleischmann W, Gough J, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McDowall J, Mitchell A, Nikolskaya A N, Orchard S, Pagni M, Ponting C P, Quevillon E, Selengut J, Sigrist C J, Silventoinen V, Studholme D J, Vaughan R, Wu C H. InterPro, progress and status in 2005. *Nucleic Acids Res* 2005; (33): D201-D205.
- Nicholas, K. B.; Nicholas H. B.; Deerfield, D. W. GeneDoc: Analysis and Visualization of Genetic Variation. *Embnew. News.* 1997; 4, 14..
- Nissink J W, Murray C, Hartshorn M, Verdonk M L, Cole J C, Taylor R. A new test set for validating predictions of protein-ligand interaction. *Proteins* 2002; (49): 457-471.
- Noble M E, Endicott J A, Brown N R, Johnson L N. The cyclin box fold: protein recognition in cell-cycle and transcription control. *Trends Biochem Sci* 1997; (22): 482-487.
- Noble M E, Endicott J A. Chemical inhibitors of cyclin-dependent kinases: insights into design from X-ray crystallographic studies. *Pharmacol Ther* 1999; (82): 269-278.
- Noble M E, Endicott J A, Johnson L N. Protein kinase inhibitors: insights into drug design from structure. *Science* 2004; (303): 1800-1805.
- Oellien F, Ihlenfeldt W D, Gasteiger J. InfVis--platform-independent visual data mining of multidimensional chemical data sets. *J Chem Inf Model* 2005; (45): 1456-1467.
- Oellien, F.; Cramer, J.; Beyer, C.; Ihlenfeldt, W.-D.; Selzer, P. M. The Impact of Tautomer Forms on Pharmacophor-based Virtual Screening. *J Chem Inf Model* 2006; (46):2342-2354.

- Ott S L, Wells S J, Wagner B A. Herd-level economic losses associated with Johne's disease on US dairy operations. *Prev Vet Med* 1999; (40): 179-192.
- Pavletich N P. Mechanisms of cyclin-dependent kinase regulation: structures of cdks, their cyclin activators, and cip and INK4 inhibitors. *Journal of Molecular Biology* 1999; (287): 821-828.
- Perdue M L, Swayne D E. Public health risk from avian influenza viruses. *Avian Dis* 2005; (49): 317-327.
- Pink R, Hudson A, Mouries M A, Bendig M. Opportunities and challenges in antiparasitic drug discovery. *Nat Rev Drug Discov* 2005; (4): 727-740.
- Polgar T, Baki A, Szendrei DI, Keseru GM. Comparative Virtual and Experimental High-Throughput Screening for Glycogen Synthase Kinase-3a Inhibitors. *J Med Chem* 2005; (48):7946-59.
- Rockey W M, Elcock A H. Progress toward virtual screening for drug side effects. *Proteins* 2002; (48): 664-671.
- Rogers J P, Beuscher AE, Flajolet M, McAvoy T, Nairn a C, Olson A J, Greengard P. Discovery of protein phosphatase 2C inhibitors by virtual screening. *J Med Chem* 2006; (49):1658-1667
- Routier S, Peixoto P, Merour J Y, Coudert G, Dias N, Bailly C, Pierre A, Leonce S, Caignard D H. Synthesis and biological evaluation of novel naphthocarbazoles as potential anticancer agents. *J Med Chem* 2005; (48): 1401-1413.
- Sadowski J. A hybrid approach for addressing ring flexibility in 3D database searching. *J Comput Aided Mol Des* 1997; (11): 53-60.
- Sali A, Overington J P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci* 1994; (3): 1582-1596.

- Schulz-Gasch T, Stahl M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model* (Online) 2003; (9): 47-57.
- Schulze-Gahmen U, Brandsen J, Jones H D, Morgan D O, Meijer L, Vesely J, Kim S H. Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins* 1995; (22): 378-391.
- Selzer P M, Brutsche S, Wiesner P, Schmid P, Mullner H. Target-based drug discovery for the development of novel antiinfectives. *Int J Med Microbiol* 2000; (290): 191-201.
- Selzer, P. M; Marhöfer, R. J.; Rohwer, A. Angewandte Bioinformatik. *Springer Verlag*, Berlin, Heidelberg, New York 2004.
- Selzer, P. M. Structure-Based-Rational-Drug-Design: Neue Wege der modernen Wirkstoffentwicklung. In: Lucius, R.; Hiepe, T.; Gottstein, B. (Hrsg) Grundzüge der allgemeinen Parasitologie. *Parey*, Berlin 2005.
- Senderowicz A M. Flavopiridol: the First Cyclin-Dependent Kinase Inhibitor in Human Clinical Trials. *Inv New Drugs* 1999; (17):313-320.
- Shapiro G I, Harper J W. Anticancer drug targets: cell cycle and checkpoint control. *J Clin Invest* 1999; (104): 1645-1653.
- Shapiro G I. Preclinical and clinical development of the cyclin-dependent kinase inhibitor flavopiridol. *Clin Cancer Res* 2004; (10): 4270s-4275s.
- Sheinerman F B, Giraud E, Laoui A. High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J Mol Biol* 2005; (352): 1134-1156.
- Sottriffer C A, Dramburg I. „In situ cross-docking“ to simultaneously address multiple targets. *J Med Chem* 2005; (48): 3122-3125.

- Stahl M., Structure-Based Library Design . In: Virtual Screening for Bioactive Molecules. Böhm H.-J., Schneider G., Mannhold R., Kubinyi H., Timmerman H. (Ed) *Wiley-VCH Verlag GmbH & Co. KGaA*, Weinheim. 2000; pp. 229-259
- Stevens R D, Robinson A J, Goble C A. myGrid: personalised bioinformatics on the information grid. *Bioinformatics* 2003; (19 Suppl 1): i302-i304.
- Tetko I V. Computing chemistry on the web. *Drug Discovery Today* 2005; (10): 1497-1500.
- Tetsu O, McCormick F. Proliferation of cancer cells despite CDK2 inhibition. *Cancer Cell* 2003; (3): 233-245.
- Thomas M P, McInnes C, Fischer P M. Protein structures in virtual screening: a case study with CDK2. *J Med Chem* 2006; (49): 92-104.
- Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; (22): 4673-4680.
- Traxler P, Green J, Mett H, Sequin U, Furet P. Use of a pharmacophore model for the design of EGFR tyrosine kinase inhibitors: isoflavones and 3-phenyl-4(1H)-quinolones. *J Med Chem* 1999; (42): 1018-1026.
- Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* 2003; (46): 2287-2303.
- Williams RB. Tracing the emerge of drug-resistance in coccidia (*Eimeria* spp.) of commercial broiler flocks medicated with decoquinate for the first time in the United Kingdom. *Vet Parasitol* 2006; (135):1-14
- Woodard C L, Li Z, Kathcart A K, Terrell J, Gerena L, Lopez-Sanchez M, Kyle D E, Bhattacharjee A K, Nichols D A, Ellis W, Prigge S T, Geyer J A, Waters N C. Oxindole-based compounds are selective inhibitors of

Plasmodium falciparum cyclin dependent protein kinases. *J Med Chem* 2003; (46): 3877-3882.

Yoon S, Welsh W J. Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring. *J Chem Inf Comput Sci* 2004; (44): 88-96.

Zuccotto F. Pharmacophore features distributions in different classes of compounds. *J Chem Inf Comput Sci* 2003; (43): 1542-1552.

Meine akademischen Lehrer

Zoologie	A. Dorresteijn, G. Eisenbeis, A. Fischer, K. Honomichl, U. Hoeger, J. Markl, G. Wegener
Genetik	M. Blessing, E. Gateff, T. Hankeln, E. R. Schmidt, G. Technau, A. Winterpacht
Immunologie	E. Rüde, A. B. Reske-Kunz
Mikrobiologie	H. König, G. Unden
Botanik	G. Rothe, J.W. Kadereit, W. Wernicke, A. Wild
Biochemie	P. M. Selzer, M. Duszenko
Physik	E. W. Otten
Mathematik	H. J. Schuh
Chemie	G. Kiel, K. K. Unger