

**Bioinformatische Identifizierung und Charakterisierung  
von Zyklin-abhängigen Kinasen und Zyklinen  
des Parasiten *Eimeria tenella* und deren Nutzung im  
rationalen Wirkstoffdesign**

**Dissertation**

zur Erlangung des Grades

"Doktor der Naturwissenschaften"

am Fachbereich Biologie  
der Johannes Gutenberg-Universität  
in Mainz

**Kristin Engels**

geboren am 20.08.1978 in Mainz

Mainz, 2010

**Meinen Eltern**

*Logic is blind and often knows only its own past.*

-Archives from Genetics to Philosophy,  
compiled by the Sorceresses of Rossak

## **Vorbemerkung**

Fremdsprachliche Ausdrücke und Fachtermini, meist aus dem Angloamerikanischen sind durch kursive Schreibweise gekennzeichnet. Die Verwendung solcher Anglizismen konnte nicht immer vermieden werden. Bei gängigen Ausdrücken wie zum Beispiel „Software“ und „Hardware“ wurde aufgrund der besseren Lesbarkeit auf die kursive Kennzeichnung verzichtet.

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung .....</b>	<b>1</b>
1.1	Humane- und tierpathogene Parasiten der <i>Apicomplexa</i> .....	1
1.2	Der tierpathogene Parasit <i>Eimeria tenella</i> .....	3
1.3	Kokzidiose .....	5
1.3.1	Kokzidiose der Hühner .....	5
1.3.2	Therapie und Prophylaxe .....	7
1.4	Kinasen als Zielproteine in der Wirkstoffforschung .....	7
1.5	Zyclin-abhängige Kinasen .....	8
1.5.1	Funktion der Zyclin-abhängigen Kinasen im eukaryotischen Zellzyklus .....	8
1.5.2	Struktureller Aufbau und Aktivierung von Zyclin-abhängigen Kinasen .....	8
1.5.3	ATP-Bindetasche von Zyclin-abhängigen Kinasen nach Traxler .....	8
1.5.4	Aufbau der Zykline .....	8
1.5.5	Zyclin-abhängige Kinasen als Zielproteine zur Entwicklung neuer Antiparasitika .....	8
1.6	Bioinformatische Analyse zur Identifizierung von Zyclin-abhängigen Kinasen .....	8
1.7	Strukturbasiertes rationales Wirkstoffdesign .....	8
1.8	<i>Modelling</i> und computergestütztes Wirkstoffdesign .....	8
1.8.1	Homologiemodellierung .....	8
1.8.2	Molekulares <i>Docking</i> .....	8
1.8.3	Substruktur- und Ähnlichkeitssuchen .....	8
1.8.4	Virtuelles <i>Screening</i> .....	8
1.8.5	LIE-Simulationen zur Berechnung der freien Bindungsenergie .....	8
1.8.6	Leitstrukturoptimierung .....	8
1.9	Aufgabenstellung .....	8
<b>2</b>	<b>Material und Methoden .....</b>	<b>8</b>
2.1	Hardware und Software .....	8
2.1.1	Hardware .....	8
2.1.2	Software .....	8
2.2	Datenbanken .....	8
2.3	Methoden zum Sequenzvergleich .....	8
2.3.1	Sequenz- <i>Alignments</i> .....	8
2.4	Identifizierung von CDK-ähnlichen und Zyclin-ähnlichen Proteinen .....	8
2.5	Homologiemodellierung der EtMRK und der PfMRK .....	8
2.6	Substruktur- und Ähnlichkeitssuchen .....	8
2.7	Vorbereitung der Substanzdaten für das molekulare Docking .....	8

2.8	<i>Docking</i> .....	8
2.8.1	GOLD.....	8
2.8.2	PLANTS.....	8
2.8.3	Eingesetzte Bewertungsmethoden .....	8
2.8.3.1	RMSD-Wert.....	8
2.8.3.2	Spearman Korrelationskoeffizient ( $R_s$ ).....	8
2.8.3.3	AUC-Wert.....	8
2.8.3.4	Anreicherungsfaktor (EF) .....	8
2.9	<i>Docking</i> der erhaltenen Substanzen nach Substruktursuchen.....	8
2.10	Virtuelles <i>Screening</i> nach potentiellen EtMRK- und PfMRK-Inhibitoren.....	8
2.11	Linear Interaction Energy (LIE) Berechnungen .....	8
2.12	Generierung einer kombinatorischen Substanzdatenbank .....	8
2.13	Löslichkeitsvorhersagen.....	8
2.14	<i>In vitro</i> Messungen.....	8
<b>3</b>	<b>Ergebnisse.....</b>	<b>8</b>
3.1	Bioinformatische Sequenzanalyse von CDKs und Zyklinen von <i>E. tenella</i> .....	8
3.1.1	CDK-homologe Proteine in <i>E. tenella</i> .....	8
3.1.1.1	EtMRK .....	8
3.1.1.2	EtCRK1 .....	8
3.1.1.3	EtCRK3 .....	8
3.1.2	Zyklone von <i>E. tenella</i> .....	8
3.2	Optimierung und Anreicherung von EtCRK2-Inhibitoren .....	8
3.2.1	Substruktur- und Ähnlichkeitssuchen .....	8
3.2.2	Molekulares <i>Docking</i> zur Auswahl virtueller <i>Hits</i> .....	8
3.2.3	<i>In vitro</i> Testung der ausgewählten Substanzen.....	8
3.2.4	Struktur-Aktivitäts-Beziehungen und Analyse der Protein-Liganden-Bindung.....	8
3.2.5	Validierung der LIE-Methode und Berechnung der Gewichtungsfaktoren.....	8
3.2.6	Bindemodi-Vorhersage der Benzimidazol-Carbonitrile .....	8
3.2.7	<i>In silico</i> Optimierung der Benzimidazol-Carbonitrile .....	8
3.3	Identifizierung von EtMRK- und PfMRK-Inhibitoren .....	8
3.3.1	Homologiemodellierung der EtMRK und der PfMRK.....	8
3.3.2	Evaluierung der <i>Docking</i> -Software PLANTS.....	8
3.3.2.1	Re- <i>Docking</i> Experimente .....	8
3.3.2.2	Anreicherungsexperimente .....	8
3.3.3	Virtuelles <i>Screening</i> nach potentiellen Inhibitoren der EtMRK und der PfMRK ....	8

<b>4</b>	<b>Diskussion</b> .....	<b>8</b>
4.1	Zyclin-abhängige Kinasen als Zielproteine der Therapie von Protozoen- Infektionen.....	8
4.2	Bioinformatische Identifizierung und Charakterisierung von Zyclin-abhängigen Kinasen in <i>E. tenella</i> .....	8
4.3	Bioinformatische Identifizierung und Charakterisierung von Zyklinen in <i>E. tenella</i> .....	8
4.4	Strategien zur Hemmung von Kinasen.....	8
4.5	Leitstrukturidentifizierung von Inhibitoren der EtCRK2.....	8
4.6	Bindemodi-Vorhersage und Optimierung der Benzimidazole.....	8
4.7	Homologiemodellierung der EtMRK und der PfMRK.....	8
4.8	Validierung der Software PLANTS.....	8
4.9	Virtuelles <i>Screening</i> nach EtMRK- und PfMRK-Inhibitoren.....	8
4.10	Ausblick.....	8
<b>5</b>	<b>Zusammenfassung</b> .....	<b>8</b>
<b>6</b>	<b>Literaturverzeichnis</b> .....	<b>8</b>

## Abkürzungsverzeichnis

3D	Dreidimensional
Å	Ångström
ABR	Adenin-Binderegion
Acc.-Nr.	<i>Accession-Nummer</i>
ATP	Adenosintriphosphat
AUC	<i>Area Under Curve</i>
Bb	<i>Babesia bovis</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
BL2SEQ	<i>Blast two Sequences</i>
BLOSUM	<i>Blocks Substitutions Matrix</i>
CAK	<i>CDK-activating Kinase</i>
CAMK	<i>Ca<sup>2+</sup>/Calmodulin-dependent Protein Kinase</i>
cdc	<i>cell-division-cycle</i>
CDK	<i>Cyclin-dependent Kinase</i>
Ch	<i>Cryptosporidium hominis</i>
ChK1	<i>Checkpoint Kinase 1</i>
Cip	<i>CD-interacting Protein</i>
Cm	<i>Cryptosporidium muris</i>
Cp	<i>Cryptosporidium parvum</i>
CRK	<i>CDK-related Kinase</i>
Cyc	<i>Cyclin</i>
DNA	<i>Desoxyribonucleic Acid</i>
EF	<i>Enrichment Factor</i>
EGF-R	<i>Epidermal-Growth-Factor-Receptor</i>
E-value	<i>Expectation Value</i>
EST	<i>Expressed-Sequence-Tag</i>
Et	<i>Eimeria tenella</i>
Etcyc	<i>Eimeria tenella Cyclin</i>
GSK3	<i>Glycogensynthase-Kinase-3</i>
Hs	<i>Homo sapiens</i>
IC	<i>Inhibitor Concentration</i>
ICD	<i>Intervet Chemical Database</i>
ID	<i>Identifier</i>
INK4	<i>Inhibitor of CDK4</i>
InterPro	<i>Integrated Resource of Protein Domains and Functional Sites</i>
kDa	Kilodalton
MAP-Kinase	<i>Mitogen-activated Protein Kinase</i>
NCBI	<i>National Centre for Biotechnology Information</i>
MRK	<i>MO15-related Kinase</i>
NLS	<i>Nuclear Localization Signal</i>
PAM	<i>Position Accepted Mutation</i>
PBR	Phosphat-Binderegion
PDB	<i>Protein Data Bank</i>
PLP	<i>Piecewise Linear Potential</i>
Pb	<i>Plasmodium berghei</i>
Pcc	<i>Plasmodium chabaudi chabaudi</i>
Pf	<i>Plasmodium falciparum</i>



Pfam	<i>Protein Domain Families</i>
Pk	<i>Plasmodium knowlesi</i>
PK	Protein Kinase
Pv	<i>Plasmodium vivax</i>
Pyy	<i>Plasmodium yoelii yoelii</i>
QSPR	<i>Quantitative Structure Property Relationship</i>
RBR	Ribose-Binderegion
RINGO	<i>Rapid Inducer of G2/M progression in Oocytes</i>
RMSD	<i>Root Mean Square Deviation</i>
ROC	<i>Receiver Operating Characteristic</i>
Ta	<i>Theileria annulata</i>
Tp	<i>Theileria parva</i>
Tcl	<i>Tool Command Language</i>
Tg	<i>Toxoplasma gondii</i>
Tp	<i>Theileria parva</i>
TR-FRET	<i>Time Resolved Fluorescence Resonance Energy Transfer</i>

## Aminosäuren

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

# 1 Einleitung

## 1.1 Humane- und tierpathogene Parasiten der *Apicomplexa*

Die Sporentierchen oder *Sporozoen* (*Apicomplexa*, siehe Abbildung 1.1) sind obligate Endoparasiten von Wirbellosen und Wirbeltieren (LUCIUS & LOOS-FRANK, 2008). Sie durchlaufen einen charakteristischen Generationswechsel, bei dem sich geschlechtlich und ungeschlechtlich vermehrende Zellen abwechseln (MEHLHORN, 2008). Am Vorderpol der vegetativen Stadien besitzen die Sporozoen einen spezialisierten Plastiden – den Apikoplasten. Dieser ist namensgebend und charakteristisch für die Gruppe der *Apicomplexa* und wesentlich am Invasionsprozess sowie der Etablierung des Parasiten in der Wirtszelle beteiligt (MEHLHORN, 2008). Der Apikoplast besitzt ein eigenes Plastiden-Genom (~35 kb), welches dem pflanzlicher Plastiden ähnelt (FICHERA & ROOS, 1997). Dieses Genom kodiert nur ca. 30 Proteine (WILSON *et al.*, 1996). Für die Aufrechterhaltung der biosynthetischen Prozesse im Apikoplasten ist es daher notwendig, kernkodierte Proteine posttranslational in die Plastide zu transportieren. Man geht davon aus, dass ca. 600 verschiedene Proteine – circa 10 Prozent des Parasitenproteoms – in den Apikoplasten transloziert werden müssen (GARDNER *et al.*, 2002; ZUEGGE *et al.*, 2001). Die Translokation von Proteinen in den Apikoplasten erfordert zweiteilige Signale, die aus einem N-terminalen Signalpeptid für den Import in das ER-Lumen und einem Serin-/ Threoninreichen Transitpeptid bestehen (DEROCHER *et al.*, 2000; FOTH & MCFADDEN, 2003; WALLER *et al.*, 2000; YUNG *et al.*, 2001). Zudem finden im Apikoplasten für den Parasiten essentielle Prozesse, wie z. B. Fettsäuresynthese und Isoprenoid-Synthese statt. Aus diesem Grund ist der Apikoplast für die meisten *Apicomplexa* unabdingbar notwendig. Da die Stoffwechselwege dieses Organells bakteriellen Ursprungs, und damit sehr unterschiedlich vom Metabolismus des Wirtes sind (LUCIUS & LOOS-FRANK, 2008), bieten sie attraktive Angriffspunkte für die Bekämpfung von beispielsweise *Plasmodium falciparum* den Erreger der Krankheit *Malaria tropica* (FICHERA & ROOS, 1997; JOMAA *et al.*, 1999; ROOS *et al.*, 2002; WALLER *et al.*, 1998).

Der Stamm der *Apicomplexa* (Abbildung 1.1) umfasst mehrere bedeutende humanpathogene, als auch tierpathogene Parasiten (CAI *et al.*, 2003). Zu den humanpathogenen Parasiten zählen beispielsweise Parasiten aus den Gattungen *Toxoplasma*, *Cryptosporidia* und *Plasmodia*. Toxoplasmose wird durch den Einzeller *T. gondii* verursacht (MCALLISTER, 2005). Dieser

Parasit kann verschiedene Wirtsorganismen, wie z. B. Mäuse oder Schweine befallen, und entwickelt dabei infektiöse und widerstandsfähige Zustandsformen (Zysten) (MCALLISTER, 2005). Der Endwirt ist in der Regel die Katze, die die infektiösen Erreger durch Fressen eines infizierten Zwischenwirts (Maus) aufnimmt und nach einer Weiterentwicklung der Toxoplasmen, infektiöse Oozysten mit dem Kot ausscheidet (LUCIUS & LOOS-FRANK, 2008). Auf den Menschen übertragen werden die Erreger über den engen Kontakt zu Katzen und die orale Aufnahme der Oozysten (z. B. Katzenkot im Katzenstreu). Die Toxoplasmose verläuft bei den meisten Menschen symptomlos und beschwerdefrei, und ist meist nur problematisch bei einer Erstinfektion von Schwangeren, da der Erreger auf das Ungeborene übergehen und schwere Schäden in der Embryonalentwicklung verursachen kann. Toxoplasmose kann auch für Individuen mit einem gestörten Immunsystem gefährlich verlaufen, da diese keine ausreichende Immunität besitzen, um eine Vermehrung des Parasiten zu verhindern (MONTROYA & LIESENFELD, 2004).

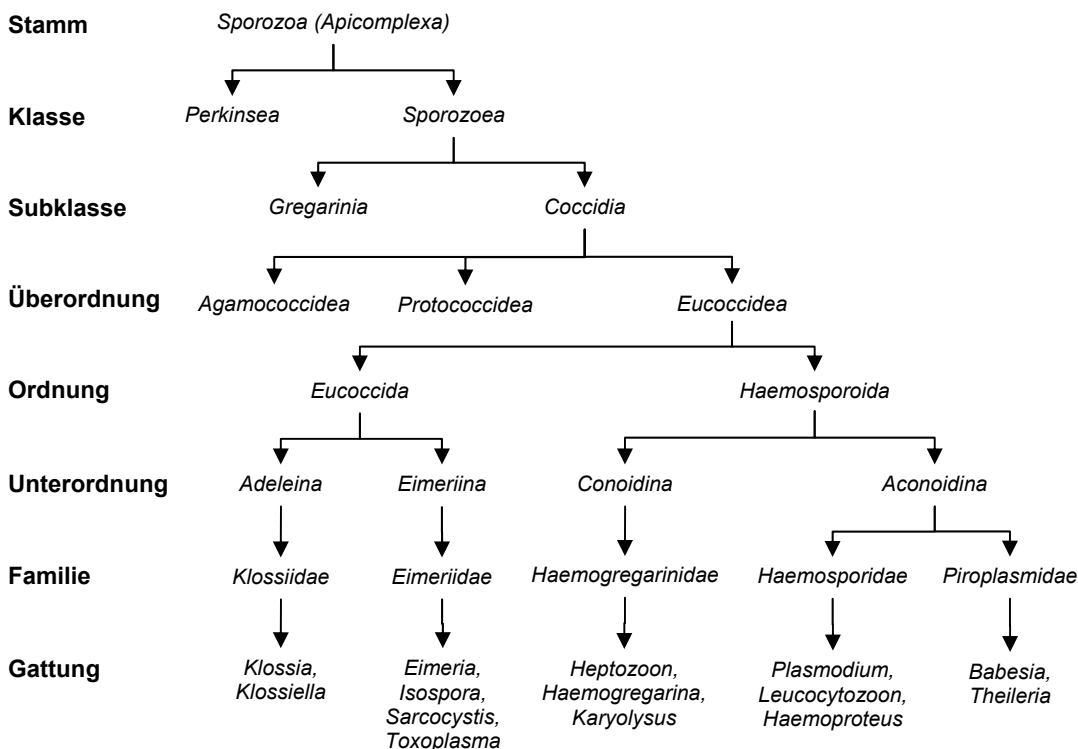


Abbildung 1.1: Ausschnitt aus dem Kladogramm der Apicomplexa (MEHLHORN, 2001).

Der Stamm der *Apicomplexa* umfasst mehrere bedeutende humanpathogene und tierpathogene Parasiten wie beispielsweise *Eimeria* spp., *Plasmodium* spp., *Babesia* spp. und *Theileria* spp.

Die Durchfallerkrankung Cryptosporidose – hervorgerufen durch eine Infektion mit *Cryptosporidium parvum* – ist bei gesunden Menschen ebenfalls folgenlos und heilt innerhalb einer Woche aus (CHALMERS & DAVIES, 2009). Zu den weitaus gefährlicheren Erkrankungen zählen Malaria Erkrankungen, die von den einzelligen Parasiten der Gattung *Plasmodium* hervorgerufen werden. Malaria – bekannt auch als Sumpffieber - ist eine meldepflichtige Krankheit und in den tropischen und subtropischen Ländern weit verbreitet. Speziell betroffen sind weite Teile Afrikas, Asiens und Südamerikas. Die *Plasmodien* werden durch den Stich der weiblichen Anopheles Mücke übertragen. Für den Menschen gefährlich sind die Erreger *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* und *Plasmodium knowlesi*, die verschiedene Formen der Malaria auslösen können. *Plasmodium falciparum* ist der klinisch bedeutsamste und bedrohlichste Erreger. Nach Angaben des Robert-Koch-Instituts in Berlin, sterben weltweit jährlich 1,5 bis 2,7 Millionen Menschen an Malaria, etwa die Hälfte von ihnen sind Kinder unter fünf Jahren (MURPHY & BREMAN, 2001). 90 % der Erkrankten leben auf dem afrikanischen Kontinent. Die Zahl der jährlichen Neuerkrankungen wird auf 300 bis 500 Millionen Fälle geschätzt (MURPHY & BREMAN, 2001). Diese Zahlen verdeutlichen die Notwendigkeit effektiver Therapiemöglichkeiten, sowie die Entwicklung neuer Antiparasitika.

## 1.2 Der tierpathogene Parasit *Eimeria tenella*

*Eimeria tenella* ist eine parasitisch lebende Unterart des Stammes der *Apicomplexa* (MEHLHORN, 2008). *Eimeria* Arten sind obligat intrazelluläre Protozoen, die in einer Viel-



**Abbildung 1.2: Sporulierte Oocyste von *E. tenella*.**

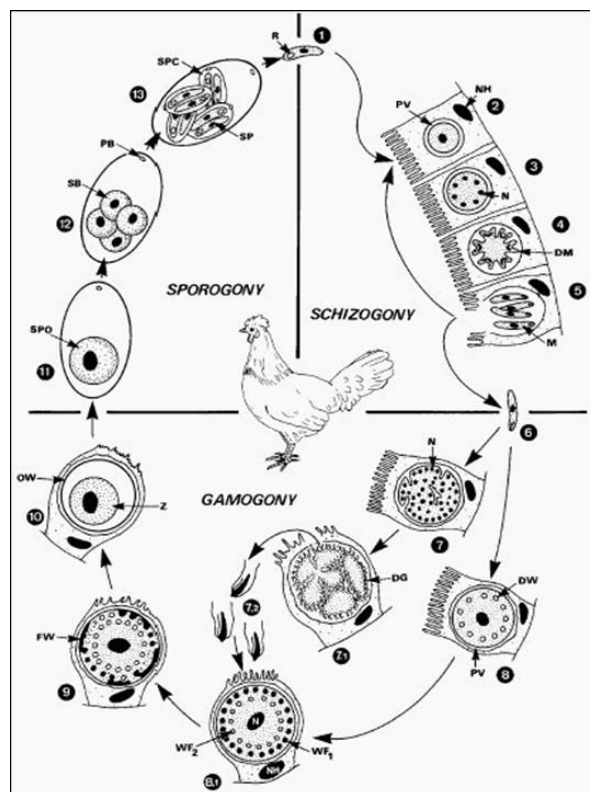
(Intervet Innovation GmbH).

zahl von verschiedenen Wirten wie beispielsweise Rindern, Schafen, Kaninchen und Geflügelarten vorkommen. Sie besiedeln den Darm oder darm-assoziierte Organe des Wirtstieres.

*Eimeria tenella* ist eine von neun bekannten *Eimeria*-Arten die das Haushuhn befallen (MEHLHORN, 2008) (siehe Abschnitt 1.3.1). Der Lebenszyklus von *Eimeria spp.* ist nicht nur wirtsspezifisch (d. h. *Eimeria spp.* die Hühner befallen, lassen sich z. B. nicht auf Gänse übertragen), sondern sie entwickeln sich auch stets artspezifisch in

Epithelzellen ganz bestimmter Darmabschnitte.

Der Lebenszyklus von *E. tenella* ist in Abbildung 1.3 schematisch dargestellt. Zunächst werden die sporulierten Oozysten vom Wirtsorganismus aufgenommen. Danach schlüpfen die Sporozoiten im Dünndarm aus den Sporozysten und invadieren die Epithelzellen des Darms. Es kommt zu einer mehrkernigen Schizontenbildung in einer parasitophoren Vakuole. Aus den Schizonten gehen bewegliche Merozoiten hervor, welche eine weitere Generation von Schizonten in anderen Wirtszellen hervorbringen können, oder zu männlichen und weiblichen Gamonten werden. Die Gamonten werden anschließend zu begeißelten, mehrkernigen Mikrogameten und zu einkernigen Makrogameten weiterentwickelt. Mikro- und Makrogameten verschmelzen in den Epithelzellen zur Zygote, woraus neue Oozysten entstehen, die mit dem Kot des Wirtstieres ausgeschieden werden, so dass der Zyklus erneut beginnt. Außerhalb des Wirtes findet die Sporulation der Oozysten, bei hoher Luftfeuchtigkeit und Anwesenheit von Sauerstoff, innerhalb von zwei Tagen statt. Eine Neuinfektion des Wirtstieres ist somit wieder möglich.



**Abbildung 1.3: Lebenszyklus von *Eimeria* spp. (Mehlhorn, 2001).**

(1) Nach der oralen Aufnahme der sporulierten Oozysten werden die Sporozoiten freigesetzt und befallen den Dünndarm. Sie durchlaufen dort die Schizogonie (2 bis 5) sowie die Gamogonie (6 bis 10) bevor sie als unsporulierte Oozysten (10) mit dem Kot ausgeschieden werden. Es folgt das Stadium der Sporogonie (11 bis 13) bevor sie vom Wirt wieder aufgenommen werden (1).

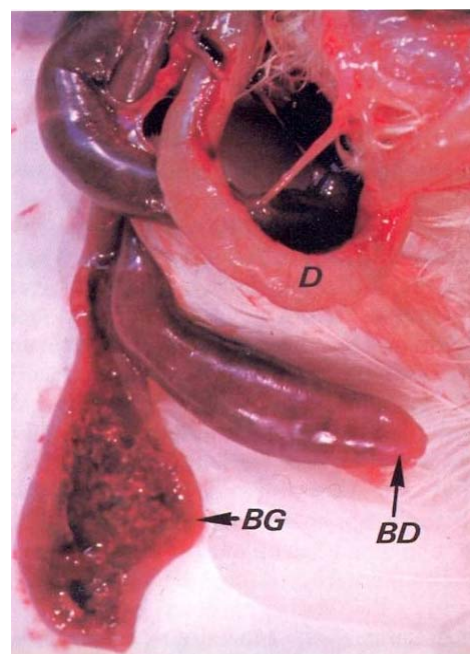
### 1.3 Kokzidiose

Als Kokzidiose bezeichnet man Erkrankungen durch Einzeller, die der Klasse der *Coccidia* angehören (Abbildung 1.1) und als Kokzidien bezeichnet werden (MEHLHORN, 2008). Diese Erkrankung zählt zu den Protozoeninfektionen. Die Kokzidien befallen vor allem den Darm, bei einigen Tieren auch Leber und Niere. Bei geringem Befall verläuft die Kokzidiose symptomlos. Sind jedoch ausgedehnte Darmschleimhautbezirke durch Kokzidien entzündet, kann die Nahrung nicht mehr richtig verdaut und ausgenutzt werden. Die an Kokzidiose erkrankten Tiere verlieren an Gewicht und leiden unter schleimigem bis blutigem Durchfall (MEHLHORN, 2008). Ist die Darmflora durch Kokzidien geschädigt, können sich auch bakterielle Infektionen ausbreiten, die mit starker Gasbildung im Darm einhergehen. Jungtiere erkranken am schwersten unter dem Befall von Kokzidien. Zu den wirtschaftlich bedeutsamen Kokzidiose-Erkrankungen zählen die Rinderkokzidiose, die Schafskokzidiose, die Kaninchenkokzidiose und die Geflügelkokzidiosen (MEHLHORN, 2008).

#### 1.3.1 Kokzidiose der Hühner

Die Kokzidiosen der Hühner sind ausschließlich Darmkokzidiosen (BOULD *et al.*, 2009) die durch *Eimerien* hervorgerufen werden. Kokzidiose ist weltweit verbreitet und tritt häufig in Mastflügelbeständen mit Masttierhaltung auf. Über kotverschmutztes Futter erfolgt die Infektion durch Aufnahme von sporulierten Oozysten. Kokzidienoozysten sind außerordentlich umweltstabil und können ein Jahr lang infektiös bleiben. Die Sporulation erfolgt bei Feuchtigkeit, Wärme und in Anwesenheit von Sauerstoff. Alle Altersstufen der Hühnervögel können befallen werden, da sich keine dauerhafte und vollständige oder nur artspezifische Immunität entwickelt (MEHLHORN *et al.*, 1993).

Zu den Parasiten der Hühner zählen insgesamt neun Arten, wovon *Eimeria tenella* und *Eimeria necatrix* stark pathogen sind (siehe Tabelle 1.1) (KUMMERFELD, 2005). *E. tenella* lebt intrazellulär in den Epithelzellen der Blinddärme von Hühnern



**Abbildung 1.4: Blutige Blinddärme des Hühnchens nach einer *E. tenella* Infektion**  
Abbildung aus (MEHLHORN *et al.*, 1993).

(siehe Abbildung 1.4) und kann hämorrhagische Enteritis auslösen, die zu einer hohen Sterblichkeitsrate junger Vögel und damit zu hohen ökonomischen Verlusten führt.

Tabelle 1.1: *Eimeria*-Arten des Huhnes sowie Lokalisationen und pathologische Veränderungen (verändert nach (KUMMERFELD, 2005)).

<i>Eimeria</i> spp.	Präpatenz [Tg]	Lokalisation	Mortalität	Makroskopische Läsionen	Durchfall
<i>E. tenella</i>	6	Zäkum	+++	Hämorrhagische Enteritis	blutig
<i>E. necatrix</i>	6	mittlerer Dünndarm	+++	Hämorrhagische Enteritis	schleimig, blutig
<i>E. maxima</i>	5	mittlerer Dünndarm	++	Darmerweiterung, verdickte Mukosa, Petechien	schleimig, rosa
<i>E. brunetti</i>	5	Ileum	++	Katarrhalisch-hämorrhagische Enteritis	schleimig, blutig, bröcklig
<i>E. acervulina</i>	4	vorderer Dünndarm	+	weißliche Querstreifung	cremeartig, schleimig
<i>E. mivati</i>	4	vorderer Dünndarm	+	weiße, runde Herde	cremeartig, schleimig
<i>E. mitis</i>	4	vorderer Dünndarm	+	Keine	-
<i>E. praecox</i>	4	vorderer Dünndarm	-	Keine	-
<i>E. hagani</i>	6	vorderer Dünndarm	-	Keine	-

Kokzidiosen führen speziell in der intensiven Tierhaltung zu großen Problemen. In der Hühnerzucht sind weltweit jährlich rund 40 Milliarden Hühner betroffen (DALLOUL & LILLEHOJ, 2006). Durch diese Erkrankung entstehen jährlich weltweit Verluste in einer geschätzten Höhe zwischen 500 Millionen £ (SHIRLEY *et al.*, 2007) und unfassbaren 3 Milliarden US\$ (DALLOUL & LILLEHOJ, 2006).

### 1.3.2 Therapie und Prophylaxe

Auch bei einer schweren Kokzidioseerkrankung wird meist nur symptomatisch therapiert. Die symptomatische Therapie umfasst eine ausreichende Flüssigkeitszufuhr - notfalls muss sogar eine Blutübertragung erfolgen - und die Gabe von Eisenpräparaten. Unterstützend wird die Gabe von Multivitaminpräparaten empfohlen. Gleichzeitig sollten auch die Hygienebedingungen verbessert werden. Hierbei spielt der Ausschluss von fäkaler Kontamination des Futters, Kontaktvermeidung des Trinkwassers mit dem Federkleid der Tiere, und Senkung der Luftfeuchtigkeit eine entscheidende Rolle. Zur Prophylaxe der Kokzidiose können auch Kokzidiostatika wie Sulfonamide, Decoquinat, Amprolium, Toltrazuril und Diclazuril eingesetzt werden (MEHLHORN *et al.*, 1993; MEHLHORN, 2008). Die ersten Kokzidiostatika wurden Ende der 30iger Jahre entwickelt, und ermöglichten die Kontrolle der Kokzidienpopulationen, was eine intensive Geflügelhaltung erlaubte. Aufgrund häufiger Resistenzbildung gegenüber diesen Wirkstoffen ist die Entwicklung neuer, effektiver und preiswerter Alternativen jedoch dringend erforderlich (KINNAIRD *et al.*, 2004; MEHLHORN, 2001; SHIRLEY *et al.*, 2007; STEPHEN *et al.*, 1997). Darüber hinaus existieren auch attenuierte Lebendvakzine zur Kokzidiose-Prophylaxe, wie beispielsweise Paracox® (WILLIAMS *et al.*, 1999). Lebendvakzinerung ist zwar effektiv, jedoch sehr teuer und ökonomisch nicht sinnvoll auf breiter Ebene in der intensiven Hühnerzucht einsetzbar (KINNAIRD *et al.*, 2004; MCDONALD & SHIRLEY, 2009; SHIRLEY, 1989).

### 1.4 Kinasen als Zielproteine in der Wirkstoffforschung

Von den rund 20.000 bis 25.000 Genen des menschlichen Genoms (IHGSC, 2004), kodieren 518 für Proteinkinasen (dies entspricht etwa 1,7% aller menschlichen Gene). Damit zählen sie zu einer der größten Proteinfamilien (SHEINERMAN *et al.*, 2005). Proteinkinasen fungieren als Phosphoryltransferasen, d. h. sie transferieren eine Phosphatgruppe von einem Donor (ATP) auf Seitenketten bestimmter Aminosäuren. Je nach Spezifität werden sie hierbei in zwei Gruppen eingeteilt. Die eine Gruppe phosphoryliert Proteine an Serin- und/oder Threoninresten, die andere an der Aminosäure Tyrosin (KREBS & BEAVO, 1979). Allerdings sind auch Proteinkinasen bekannt, die sowohl Serin und Threonin als auch Tyrosin phosphorylieren können (LINDBERG *et al.*, 1992). Durch den Phosphatgruppentransfer spielen sie eine wichtige Rolle in der Signaltransduktion der Zelle, in dem sie beispielsweise die Aktivität von Proteinen oder Transkriptionsfaktoren regulieren. Ist die Funktion der Kinasen nicht mehr gegeben, kommt es zu Erkrankungen. Mehr als 250 der Gene für Proteinkinasen des Men-



schen werden mit Krankheiten in Verbindung gebracht (KNUUTILA *et al.*, 1998), wie beispielsweise Krebserkrankungen (MALUMBRES & BARBACID, 2009), Adipositas (BAYS, 2004), Typ-I-Diabetes (BASTARRACHEA *et al.*, 2008), Bluthochdruck (COGOLLUDO *et al.*, 2005), Schizophrenie (KOROS & DORNER-CIOSSEK, 2007) und der Alzheimer Erkrankung (IQBAL *et al.*, 2003).

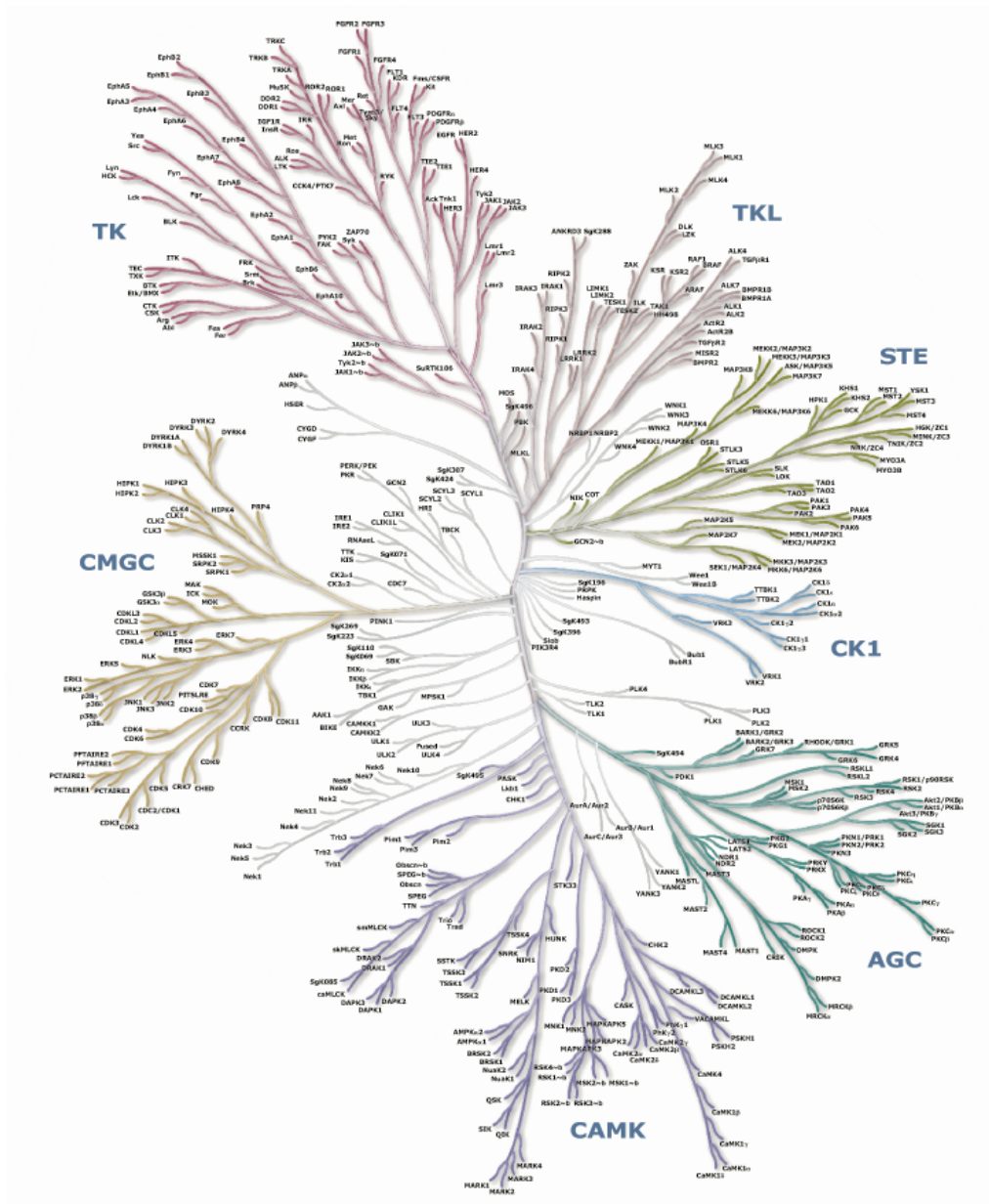
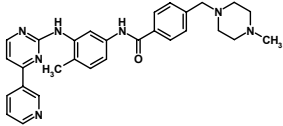
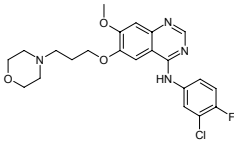
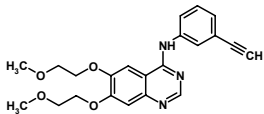
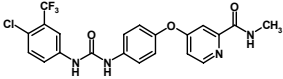
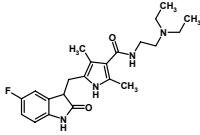
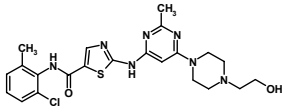
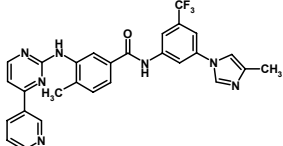
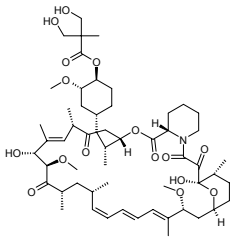
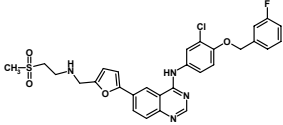


Abbildung 1.5: Dendrogramm des humanen Kinoms.

Familien humaner Kinasen lassen sich in einem Dendrogramm zusammengefasst (nach [MANNING *et al.*, 2002]). Je ähnlicher zwei Kinasesequenzen zueinander sind, desto dichter liegen sie im Dendrogramm zusammen, wie beispielsweise die humane CDK1 und CDK1 der CAMG-Familie.

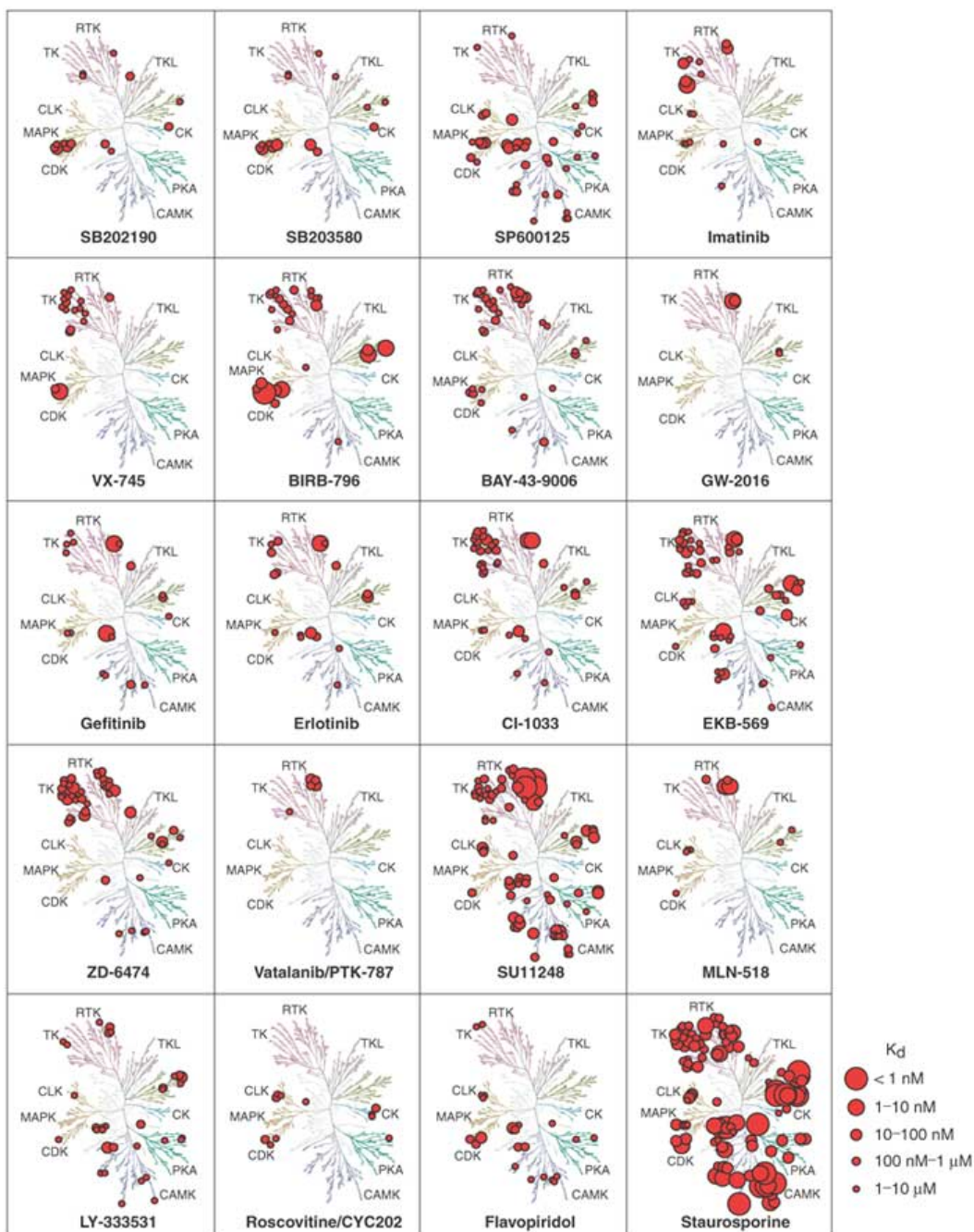
Tabelle 1.2: Auf dem Europäischen Markt zugelassene Kinase-Inhibitoren (EMEA, 2009).

Wirkstoff	Wirkort	Hersteller	Jahr der Zulassung	Literatur
<p>Glivec® (Imatinib)</p> 	Bcr-Abl	Novartis	2001	(AYOUB <i>et al.</i> , 2005)
<p>Iressa® (Gefitinib)</p> 	EGFR	AstraZeneca	2004	(NORO & GEMMA, 2008)
<p>Tarceva® (Erlotinib)</p> 	EGFR, HER1	Roche	2005	(BARESCHINO <i>et al.</i> , 2007)
<p>Nexavar® (Sorafenib)</p> 	c-Raf, b-Raf, p38a, VEGF- 2, VEGF-3, PDGFβ, c-KIT	Bayer HealthCare	2006	(HAUSCHILD, 2007)
<p>Sutent® (Sunitinib)</p> 	VEGF 1-3, PDGFα, PDGFβ, c-KIT	Pfizer	2006	(CHOUHAN <i>et al.</i> , 2007)
<p>Sprycel® (Dasatinib)</p> 	Bcr-Abl, Src, PDGFβ, c-KIT	Bristol- Myers Squibb	2006	(BRAVE <i>et al.</i> , 2008)
<p>Tasigna® (Nilotinib)</p> 	Bcr-Abl, PDGF, c-KIT	Novartis	2007	(QUINTAS- CARDAMA & COR- TES, 2007)

Wirkstoff	Wirkort	Hersteller	Jahr der Zulassung	Literatur
<p>Torisel® (Temosirolimus)</p> 	mTor	Wyeth Pharmaceuticals	2007	(MALIZZIA & HSU, 2008)
<p>Tyverb® (Lapatinib)</p> 	EGFR, HER2	Glaxo-SmithKline	2008	[FRAMPTON, 2009]

Bei der Entwicklung von Proteinkinase-Inhibitoren sind prinzipiell vier verschiedene Arten der Inhibition denkbar: ATP-kompetitive Inhibitoren, Substrat-kompetitive Inhibitoren, Inhibitoren der Proteinkinase-Aktivierung und irreversible Inhibitoren [KLEBE, 2009]. In der Vergangenheit hat man sich nahezu ausschließlich auf eine kompetitive Verdrängung des ATPs von seiner Bindestelle fokussiert. Diese ist innerhalb der Proteinkinasen ganz besonders hoch konserviert, so dass sich gute Strategien für das Einpassen von Inhibitoren entwickeln ließen. Allerdings hat die große Ähnlichkeit der ATP-Bindestellen den Nachteil, dass es schwierig ist, einen Inhibitor zu entwickeln, der selektiv die Bindetasche einer ganz bestimmten Kinase blockiert. Dies wurde lange Zeit als unmöglich erachtet. Darüber hinaus sind wichtige Signaltransduktionswege häufig redundant, so dass der Ausfall eines Weges über einen zweiten Weg kompensiert werden kann [KLEBE, 2009]. Ein einzelner Weg muss somit nicht zwangsläufig essentiell sein. Die Entwicklung kompetitiver Inhibitoren der ATP-Bindung wird zudem dadurch erschwert, dass der Wirkstoff mit der im millimolaren Bereich in der Zelle vorhandenen ATP-Konzentration [EGLLEN & REISINE, 2009] konkurrieren muss, und die meisten Proteinkinasen  $K_m$ -Werte für ATP im hohen mikromolaren Bereich besitzen [KLEBE, 2009]. Der erste ATP-Kompetitor Imatinib (Glivec, Novartis), der zur Therapie der Chronischen Myeloischen Leukämie eingesetzt wird, diente als Beweis, dass die Entwicklung solcher Wirkstoffe zwar schwierig, aber dennoch möglich ist [DRUKER, 2002]. Weitere Beispiele selektiver Kinase-Inhibitoren finden sich in Abbildung 1.6, wie beispielsweise der

Wirkstoff Gefitinib (Iressa<sup>®</sup>) und der experimentelle Inhibitor GW-2016 [FABIAN *et al.*, 2005]. Heutzutage werden insgesamt neun Inhibitoren humaner Proteinkinasen für verschiedene Indikationen vermarktet (siehe Tabelle 1.6) [EMEA, 2009; JOHNSON, 2009].



**Abbildung 1.6: Drug-Target-Selektivität einiger Kinase-Inhibitoren.**

Auf den einzelnen Abbildungen ist jeweils ein Dendrogramm humaner Kinasen dargestellt. Die roten Kreise markieren die Wirkstärke eines Inhibitors auf die jeweilige Kinase. Aus der Anzahl der Kreise lässt sich die Selektivität der einzelnen Inhibitoren ableiten. Staurosporin ist beispielsweise ein unspezifischer Kinase-Inhibitor, während GW-2016 nur auf einzelne Kinasen wirkt. Prinzipiell ist es somit möglich einen selektiven Kinase-Inhibitor zu finden (nach [FABIAN *et al.*, 2005]).

## 1.5 Zyklin-abhängige Kinasen

Zyklin-abhängige Kinasen (CDKs = *Cyclin-Dependent Kinases*) bilden eine Gruppe von homologen, hoch konservierten Proteinkinasen, die in allen Eukaryonten zu finden sind. Einige dieser Proteine sind maßgeblich an der Kontrolle des Zellzyklus beteiligt. Hierbei aktivieren sie eine Vielzahl von anderen Proteinen durch Phosphorylierung, die zu bestimmten Zeitpunkten im Zellzyklus benötigt werden. CDKs bilden Komplexe mit den zugehörigen Zyklinen, wodurch sie allosterisch aktiviert werden. Die Zyklin-abhängigen Kinasen sind nur in Verbindung mit ihren zugehörigen Zyklinen aktiv. Dadurch kann die Aktivität der CDKs in der Zelle zeitlich gesteuert und reguliert werden. Die Zykline selbst werden hierbei durch zellzyklusabhängige Konzentrationsänderungen reguliert.

### 1.5.1 Funktion der Zyklin-abhängigen Kinasen im eukaryotischen Zellzyklus

Die Zyklin-abhängigen Kinasen besitzen eine wichtige Funktion in der Regulation des eukaryotischen Zellzyklus. Der eukaryotische Zellzyklus besteht aus einer Abfolge hoch geordneter, gleich ablaufender Veränderungen, die als Übergänge, Phasen oder als Stadien bezeichnet werden. Hierzu zählen die *Gap1*-Phase (G1), die DNA-Synthese-Phase (S), die *Gap2*-Phase (G2) und die Mitose-Phase (M) [ELLEDDGE, 1996; JOHNSON & WALKER, 1999] (Abbildung 1.7). Der Zellzyklus beginnt mit der G1-Phase. Im Anschluss folgt die S-Phase, in der die DNA der Zelle repliziert wird. Danach werden die G2-Phase und die M-Phase durchlaufen. Die ersten drei Phasen werden auch als Interphase zusammengefasst. Während der Mitose erfolgt die eigentliche Zellteilung, und die Aufteilung der DNA auf die zwei neu entstehenden Tochterzellen. Sobald die Zellen infolge spezifischer, antimitogener Signale oder aufgrund des Fehlens angemessener, mitogener Signale, ihre Vermehrung gestoppt haben, verlassen sie den Zyklus und treten in eine Ruhe-Phase (G0) ein [MALUMBRES & BARBACID, 2005]. Die genaue Abfolge der verschiedenen Phasen des Zellzyklus ist von entscheidender Bedeutung für die Zelle, da Störungen zu einer unvollständigen DNA-Replikation, und somit zu einer Veränderung des Genoms führen können. Kontrollpunkte, die beispielsweise durch die Zyklin-abhängigen Kinasen gesteuert sind, gewähren daher den reibungslosen Ablauf des Zellzyklus.

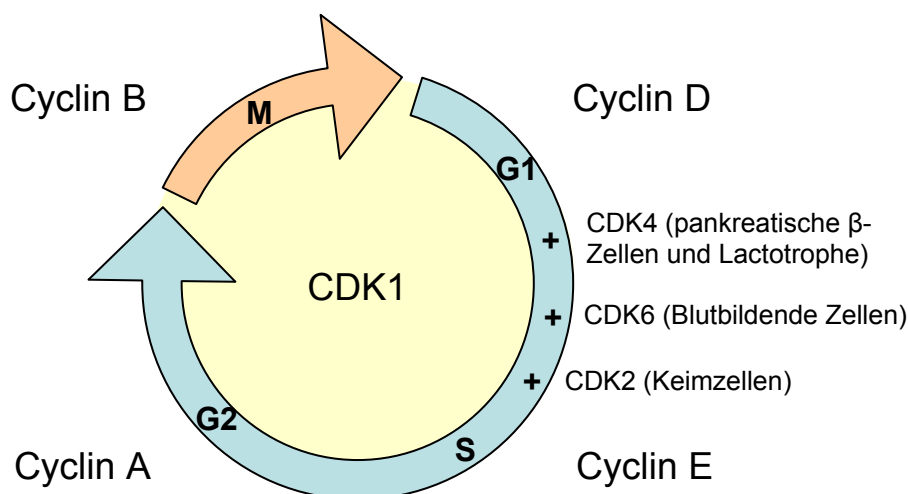


Abbildung 1.7: Schema des eukaryotischen Zellzyklus und die Regulation durch Zyklone sowie durch Zyklin-abhängige Kinasen (nach [Malumbres&Barbacid,2009]).

Der Zellzyklus gliedert sich in die beiden Zwischen-Phasen G1 und G2, die Synthese-Phase S und die Mitose-Phase M. Reguliert wird dieser Zyklus mit Hilfe von Zyklin-abhängigen Kinasen sowie durch deren Aktivatoren (Zyklone). Die CDK-Zyklin-Komplexe steuern die Übergänge zwischen den einzelnen Phasen. Die Kinasen CDK2, CDK4 sowie CDK6 sind essenziell für die Proliferation von spezialisierten Zellen [MALUMBRES & BARBACID, 2009].

Im humanen Kinom, das alle bekannten und mutmaßlichen Kinasen umfasst, sind 11 CDK Proteine bekannt. Jedoch sind nur einige der CDK Proteine in der direkten Zellzyklus-Regulation involviert (CDK1, CDK2, CDK4 und CDK6). Die CDK-Proteine agieren als Schalter, die den Übergang von der G1 zur S1 Phase und der G2 zur M Phase, sowie alle anderen Schritte innerhalb des Zellzyklus steuern. Sie bilden Heterodimere mit den Zyklonen, wodurch sie allosterisch aktiviert werden. Beim Menschen sind 29 Zyklone bekannt, jedoch agieren nur 9 von ihnen im Zellzyklus (Zyklone: A1, A2, B1, B2, D1, D2, D3, E1, E2) [MALUMBRES & BARBACID, 2005]. Die so aktivierten Kinasen phosphorylieren Transkriptionsrepressoren, die dadurch inaktiviert werden. Die Phosphorylierung durch die CDK-Proteine 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. Kontrolliert werden die CDKs durch verschiedene Faktoren: zum einen durch die Bindung der Zyklone, zum anderen aber auch durch inhibitorische Untereinheiten, wie beispielsweise der Cip- oder INK4-Proteinfamilie. Auf diese Art entsteht ein Regelkreis, der die spezifischen Phasen des Zellzyklus kontrolliert [MORGAN, 1997]. Wird dieser Kreislauf, z.B. durch spezifische CDK-Inhibitoren unterbrochen, ist die Zellvermehrung gehemmt.

### 1.5.2 Struktureller Aufbau und Aktivierung von Zyklin-abhängigen Kinasen

Die Faltungstopologie der Proteinkinasen ist stark konserviert [BUIJSMAN, 2004], und sie zeigen eine hohe Sequenzähnlichkeit untereinander [BRANDEN & TOOZE, 1999]. Nach Hanks und Hunter [HANKS & HUNTER, 1995] lassen sich alle Proteinkinase-Strukturen in elf Subdomänen einteilen (Abbildung 1.8).

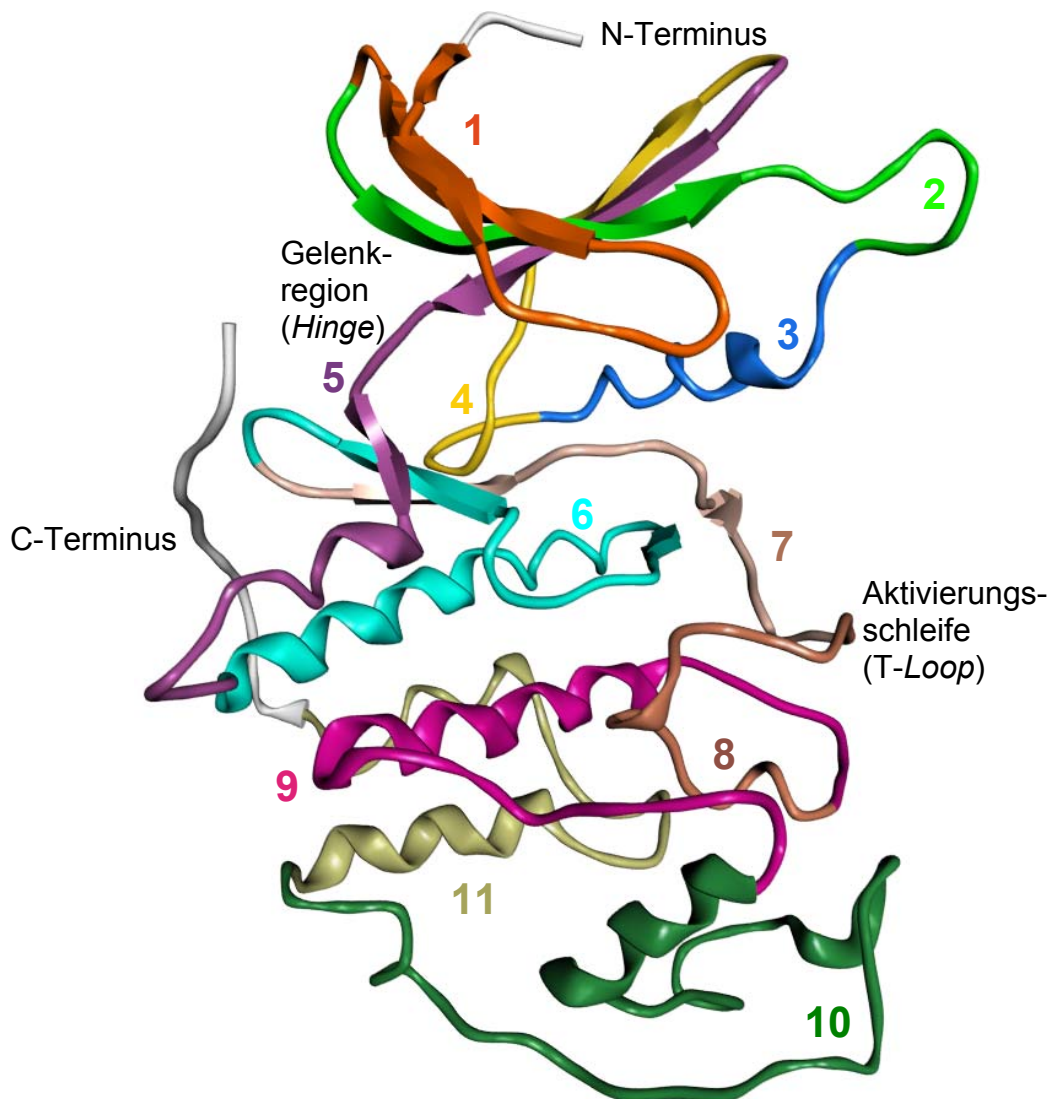


Abbildung 1.8: Struktureller Aufbau von Proteinkinasen.

Farbige Hervorhebung der elf Subdomänen von Proteinkinasen (nach Hanks und Hunter [HANKS & HUNTER, 1995]) am Beispiel Zyklin-abhängiger Kinasen, PDB-ID: 1JST.

Auch für den CDK2-ZyklinA-Komplex sind detaillierte Strukturinformationen bekannt [JEFFREY *et al.*, 1995]. Der strukturelle Aufbau der CDK2 zeigt dasselbe Grundgerüst, wie bei anderen eukaryotischen Proteinkinasen. Dieser wurde zum ersten Mal für die cAMP-abhängige Proteinkinase [KNIGHTON *et al.*, 1991] charakterisiert. Die CDK2-Struktur be-

steht aus einer kleineren N-terminalen Domäne von 85 Aminosäuren, die aus  $\beta$ -Faltblatt-Strukturen aufgebaut ist, sowie einer größeren C-terminalen Domäne. Sie umfasst 213 Aminosäuren und besitzt einen hohen  $\alpha$ -helikalen Anteil [MORGAN, 1997; PAVLETICH, 1999].

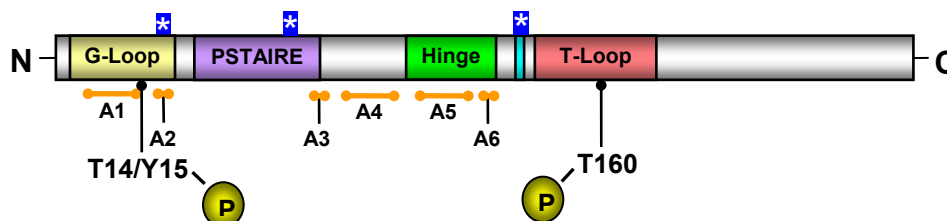


Abbildung 1.9: Charakteristische Sequenzmotive von humanen CDK Proteinen.

Zu den charakteristischen Sequenzmotiven zählen PSTAIRE-Motiv (violett), G-Loop (gelb), Hinge-Region (grün) und Aktivierungs-Loop (rot). Das Zyklin-Bindemotiv (PSTAIRE-Motiv) variiert in der Sequenzabfolge in humanen CDK Proteinen. Die Aminosäuren des katalytischen Zentrums K32, E50 und D144 (EtCRK2 Nummerierung) sind durch „\*\*\*“ markiert. A1 bis A6 kennzeichnen die ATP-Bindungsstellen. Die Phosphorylierungsstellen T14, Y15 und T160 sind ebenfalls bei humanen CDK Proteinen hoch konserviert.

Charakteristisch für alle CDK Proteine ist eine konservierte Zyklin-Bindedomäne (PSTAIRE, CDK2), die zur einzigen  $\alpha$ -Helix (C-Helix) der N-terminalen Domäne gehört und entscheidend für die Zyklinbindung ist (siehe Abbildung 1.9). An der Verbindungsstelle der beiden Domänen (auch als Hinge-Region bezeichnet [MCINNES & FISCHER, 2005]), liegt eine tiefe Spalte. Dort befindet sich die hydrophobe ATP-Bindetasche. Diese ist, ebenso wie die Substratbindetasche, in der inaktiven Form durch eine Schleife (T-Loop) blockiert. Der T-Loop (Aktivierungsschleife) enthält die Aminosäure Thr160 (Nummerierung nach humaner CDK2), deren Phosphorylierung zu einer drastischen Aktivitätssteigerung der CDK2 führt [JEFFREY *et al.*, 1995; MORGAN, 1997]. Die Bindung des Zyklins an die CDK bewirkt starke Veränderungen der Tertiär-Struktur der Proteine, und ist entscheidend für die Substratspezifität. Hierbei wird die Substratbindestelle demaskiert und die katalytischen Reste (Lys33, Glu51 und Asp145, Nummerierung nach humaner CDK2) in geeigneter Weise für den Phosphatgruppentransfer vom ATP auf das Substrat angeordnet. Dies beinhaltet die Bewegung des T-Loops von der ATP- und Substratbindestelle weg, Rotation der C-Helix und Änderung der Position des Glycin-reichen Loops. Diese Umkonfigurierung der CDK hat zur Folge, dass inhibitorische Teile aus dem katalytischen Zentrum entfernt (T-Loop), sowie regulatorische Bereiche einer Phosphorylierung zugänglich gemacht werden (PSTAIRE-Helix) [JEFFREY *et al.*, 1995; MORGAN, 1997]. In Abbildung 1.10 ist die Bewegung der Aktivie-



rungsschleife, zwischen inaktiver und aktiver CDK2-Struktur, deutlich zu erkennen [HOLM *et al.*, 2008; NICULESCU-DUVAZ *et al.*, 2007].

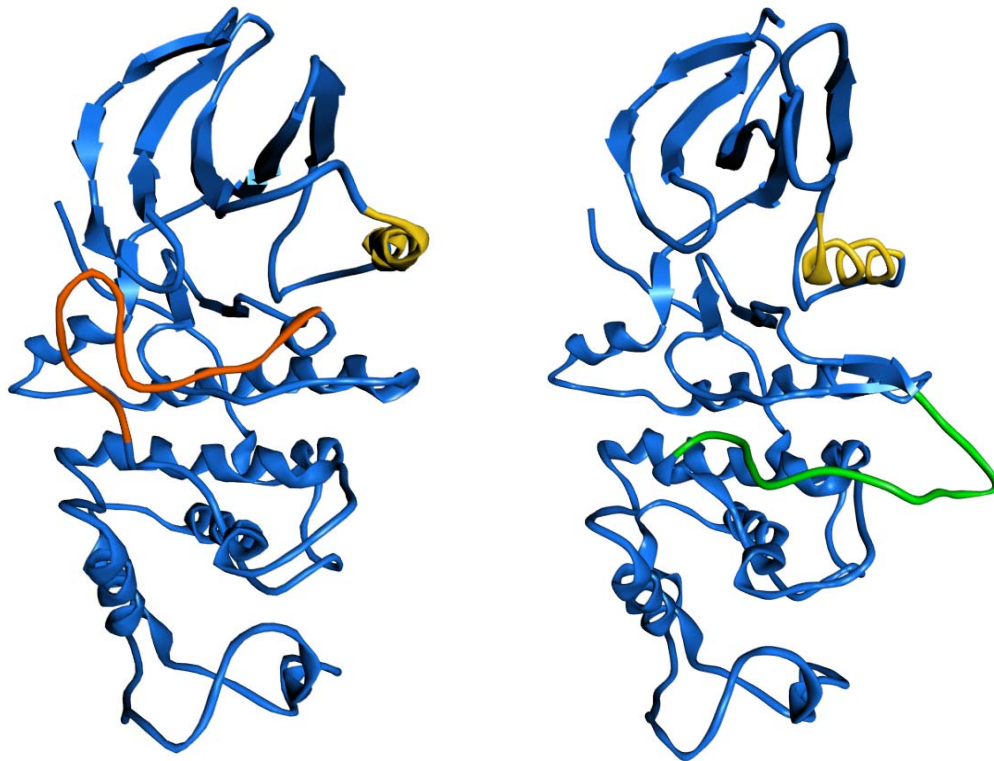


Abbildung 1.10: Vergleich der inaktiven und der aktiven CDK Struktur.

Deutlich zu erkennen ist die Bewegung des Aktivierungs-Loops (rot = inaktiv; grün = aktiv) sowie der C-Helix (gelb).

### 1.5.3 ATP-Bindetasche von Zyklin-abhängigen Kinasen nach Traxler

In Anlehnung an das ATP-Bindetaschenmodell nach Traxler [TRAXLER *et al.*, 1999] für ATP-kompetitive Kinase-Inhibitoren (siehe Abbildung 1.11), besteht die ATP-Bindetasche aus fünf verschiedenen Subtaschen, welche unterschiedliche physikochemische Eigenschaften aufweisen. Zu diesen fünf Subtaschen zählen a) die Adenin-Binderegion (ABR), b) die Ribosebindetasche (RBR), c) die Phosphat-Binderegion (PBR), d) die Binderegion I (BR-I) und e) die Binderegion II (BR-II) (siehe Abbildung 1.11). Alle ATP-kompetitiven Kinase-Inhibitoren binden in der hydrophoben ABR, und interagieren mit der *Hinge*-Region über Wasserstoffbrücken. Die RBR hingegen ist hydrophil und wird oft durch lösliche Gruppen adressiert. Diese Region ist weniger konserviert und beinhaltet spezifische Aminosäuren, welche für eine Erhöhung der Selektivität ausgenutzt werden können [BUIJSMAN, 2004; TRAXLER *et al.*, 1999]. Die Tasche der BR-I befindet sich in Verlängerung des N6-

Stickstoffs des ATPs, und ist nicht in der Bindung des ATPs involviert. Diese Region ist nicht konserviert und wird genutzt, um sowohl die Selektivität als auch die Aktivität von Inhibitoren zu erhöhen [BUIJSMAN, 2004]. Der Zugang zu dieser Tasche wird durch die sogenannten „Gatekeeper“-Aminosäure kontrolliert [BUIJSMAN, 2004; GORRE *et al.*, 2001]. Ein Zugang zu dieser Binderegion ist nur möglich, wenn an dieser Stelle eine sehr kleine Aminosäure (*Gatekeeper*) vorliegt. Genauso wie BR-I ist auch BR-II nicht an der ATP-Bindung beteiligt. Diese Subtasche kann ebenfalls zu Erhöhung der Selektivität als auch der Aktivität von Inhibitoren genutzt werden [BUIJSMAN, 2004]. Die hydrophobe Tasche PBR ist dem Lösungsmittel zugewandt und scheint für die Affinität von Inhibitoren eine eher untergeordnete, allenfalls modulierende Bedeutung zu haben [BUIJSMAN, 2004; HOLM *et al.*, 2008].

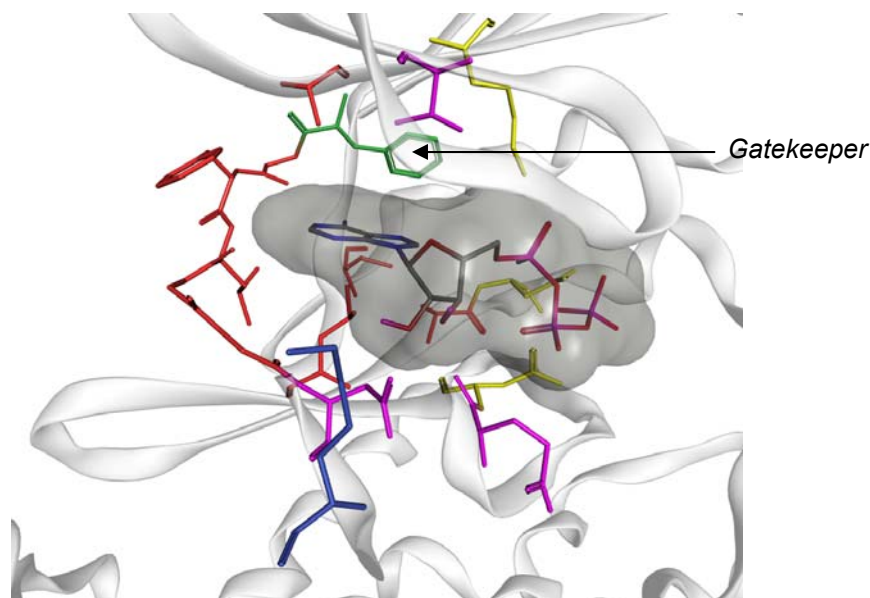


Abbildung 1.11: ATP-Bindetasche nach Traxler.

Die Bindetasche lässt sich in 5 verschiedene Subtaschen aufteilen: A: ABR = rot; B: RBR = violett; PBR = gelb; BR-I = grün; BR-II = blau.

#### 1.5.4 Aufbau der Zykline

Zykline gehören zu einer sehr diversen Proteinfamilie. Ihre Größe variiert von 35 bis 90 kDa. Auf der primären Sequenzebene sind sie nur sehr gering konserviert. Die dennoch vorhandene Sequenzähnlichkeit konzentriert sich auf einen Aminosäureabschnitt, bestehend aus ca. 100 Aminosäuren, der als Zyklin-Box bekannt ist. Dieser Abschnitt ist notwendig zur CDK-Bindung [KOBAYASHI *et al.*, 1992; LEES & HARLOW, 1993]. Abgesehen von dem großen Sequenzunterschied besitzen jedoch alle Zykline eine ähnliche Grundstruktur. Sie umfassen

alle im Kernbereich 10 Helices, die in 2 Helixbündeln aus jeweils 5 Helices ( $H^1 - H^5$  und  $H^1' - H^5'$ ) organisiert sind, wobei die Zyklin-Box die ersten 5 Helices umfasst (siehe Abbildung 1.12). Diese Kerndomäne ist beispielsweise bei Zyklin H und Zyklin A besonders stark konserviert. Dies legt die Vermutung nahe, dass diese Zykline ein rigides Templat zur Kinaseerkennung besitzen. Außerhalb dieses konservierten Kernbereichs besitzen alle Zykline zudem eine N- und C-terminale Struktur-Domäne, die jedoch in Position und Sekundärstruktur variieren kann. Sie dient möglicherweise zur Stabilisierung und Positionsfixierung der Kernregionen [MORGAN, 1997].

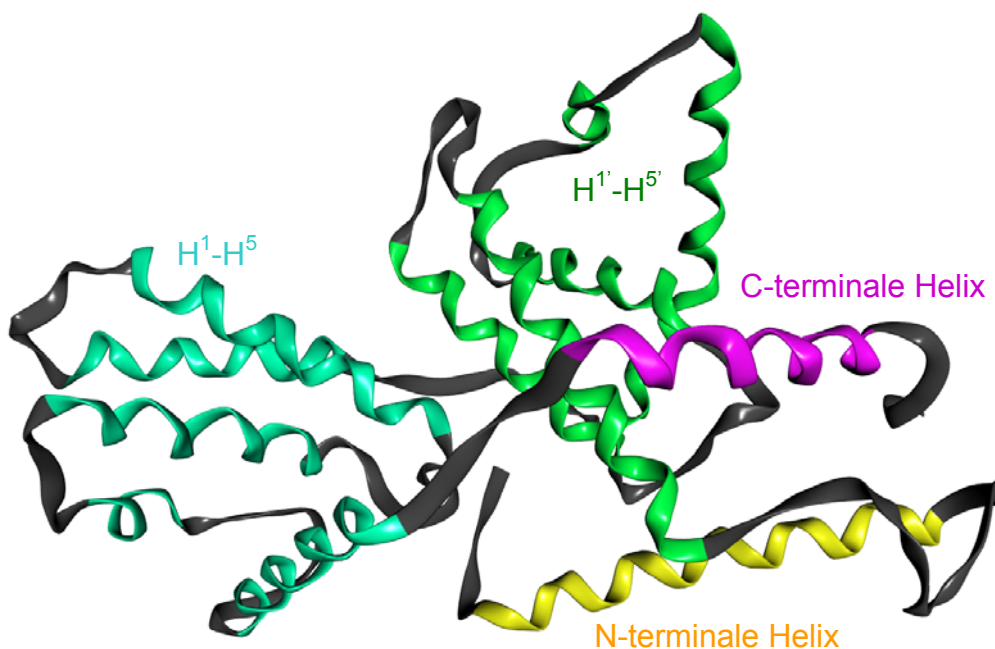


Abbildung 1.12: Dreidimensionale Struktur von Zyklin H (PDB-Accession-Nr.: 1JKW)

Dargestellt ist die Tertiärstruktur des humanen Zyklin H Proteins. Die N-terminale Helix ist in gelb, die C-terminale Helix in violett eingefärbt. Die Helices  $H^1-H^5$  sind in cyan, die Helices  $H^1'-H^5'$  in grün eingefärbt.

Von dem humanen Zyklin H existieren derzeit zwei Kristallstrukturen aus dem Jahr 1997 [PDB-Accession-Nummer: 1JKW und 1KXU] [ANDERSEN *et al.*, 1997; KIM & CHO, 1997]. Die Strukturen besitzen jeweils eine Auflösung von 2,6 Å. Zyklin H zeigt eine typische Zyklin-Faltung. Die Grundstruktur umfasst eine ähnliche Kerndomäne, bestehend aus einem Paar von jeweils 5 ähnlichen, helikalen Domänen. Diese zentrale Faltung wird von zwei Helices - der N-terminalen sowie C-terminalen Helix flankiert - die in Position und Sekundärstruktur, innerhalb der unterschiedlichen Klassen der Zykline, variieren können (siehe Abbildung 1.12) [ANDERSEN *et al.*, 1997; KVAAL *et al.*, 2002; MORGAN, 1997].

### 1.5.5 Zyklin-abhängige Kinasen als Zielproteine zur Entwicklung neuer Antiparasitika

Die prinzipielle Eignung von CDKs als Zielproteine für Antiparasitika wurde bereits in der Literatur gezeigt [DOERIG, 2004; HASSAN *et al.*, 2001; MOTTRAM *et al.*, 1996]. Die Vermehrung der Protozoen *Plasmodium falciparum*, *Leishmania mexicana* und *Trypanosoma brucei* konnte durch CDK-Inhibitoren, die für humane CDKs entwickelt wurden, unterbunden werden [GRANT *et al.*, 2004; HARMSE *et al.*, 2001; WOODARD *et al.*, 2003]. Vorversuche mit solchen Inhibitoren an *E. tenella* haben zu ähnlichen Ergebnissen geführt [BEYER, 2007]. Hierzu wurde der bekannte CDK-Inhibitor Flavopiridol [ZHAI *et al.*, 2002] in einem Schizonten-Reifungs-Assay getestet. Diese Messungen wurden von Joachim Hofmann, Anja Heckerroth und Britta von Oepen (Intervet Innovation GmbH, Schwabenheim) durchgeführt. Die Schizonten-Reifung von *E. tenella* konnte mit Flavopiridol-Konzentrationen von 150 nM und 300 nM inhibiert werden. Inhibitorkonzentrationen unter 80 nM zeigten keine Wirkung. Konzentrationen oberhalb 600 nM waren jedoch giftig für die Wirtszelle. Die für Flavopiridol bei einer Konzentration von 150 nM festgestellte vollständige Inhibierung der Schizonten-Reifung, korreliert sehr gut mit der Inhibierungskonstante von 11 nM im Enzym-Assay einer Zyklin-abhängigen Kinase von *E. tenella* (EtCRK2). Der Enzym-Assay wurde von María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim) durchgeführt.

Tabelle 1.3: Ergebnisse der chemischen Validierung mit dem bekannten CDK-Inhibitor Flavopiridol.

Flavopiridol [nM]					
		EtCRK2		HsCDK2	
Enzyme-Assay	IC50	Ki	IC50	Ki	
	33 ± 10	11 ± 3	36 ± 10	19 ± 3	
Schizonten Entwicklung					
Schizonten-Reifungs-Assay	100% Inhibierung	keine Inhibierung		Zell-Toxizität	
	150	<80		>600	
	300				

Daten von María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim; Enzym-Assay) sowie Joachim Hofmann, Anja Heckerroth und Britta von Oepen (Intervet Innovation GmbH, Schwabenheim; Schizonten-reifungstest) [unpublizierte Daten]. Anzahl der Durchführungen im Enzyme-Assay: n = 3-4 sowie im Schizonten-Reifungs-Assay: n = 2.

## 1.6 Bioinformatische Analyse zur Identifizierung von Zyklin-abhängigen Kinasen

Unter homologen Proteinen versteht man Proteine, die von einem gemeinsamen Vorfahren abstammen, egal ob das Merkmal verändert oder unverändert von dem Träger übernommen wurde [SELZER *et al.*, 2008]. Häufig ist es mit *in silico* Methoden jedoch nicht möglich, eine gemeinsame Vorläufersequenz zu identifizieren. Bei Vorliegen signifikanter Übereinstimmung zwischen den Sequenzen, geht man daher häufig von homologen Proteinen aus.

Die Identifizierung homologer oder ähnlicher Proteine, wie beispielsweise CDK Proteine, in biologischen Datenbanken, beruht auf dem sukzessiven, paarweisen Sequenzvergleich der Suchsequenz mit den in der Datenbank gespeicherten Proteinsequenzen. Einer der meist verwendeten Algorithmen, zur Sequenzsuche in biologischen Datenbanken, ist der BLAST-Algorithmus (*Basic Local Alignment Search Tool*) [ALTSCHUL *et al.*, 1990]. Es handelt sich hierbei um einen heuristischen *Alignment*-Algorithmus, der auf einer Abfolge von Näherungen basiert [SELZER *et al.*, 2008]. Der BLAST-Algorithmus macht sich die Kenntnis von Sequenz- und *Alignment*-Statistik zunutze. Mit dieser Technik kann, im Gegensatz zu dynamischen Algorithmen wie z. B. dem Needleman-Wunsch-Algorithmus [NEEDLEMAN & WUNSCH, 1970], der Rechenzeitbedarf und damit der Gesamtzeitbedarf einer Datenbanksuche stark gesenkt werden.

Der BLAST-Algorithmus [ALTSCHUL *et al.*, 1990] bildet die Basis eines Programmpakets, das aus sechs verschiedenen Programmen besteht. Fünf Programme dieses Pakets dienen der Datenbanksuche und repräsentieren die verschiedenen denkbaren Kombinationen aus der Art der Abfragesequenz (Nukleotid bzw. Protein) und der Datenbank [SELZER *et al.*, 2008]. Das sechste Programm „*bl2seq*“, dient dem paarweisen Sequenzvergleich zweier einzelner Sequenzen. Die Ermittlung homologer Proteine ist für die Identifizierung potentieller Zielproteine zur Wirkstoffentwicklung von großer Bedeutung. Anhand ihres Vorkommens in verschiedenen Organismen ist es möglich, eine Abschätzung des Wirkspektrums der zu entwickelnden Medikamente vorzunehmen.

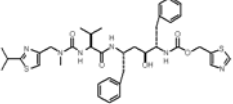
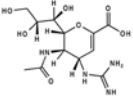
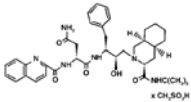
## 1.7 Strukturbasiertes rationales Wirkstoffdesign

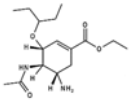
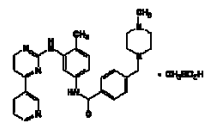
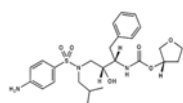
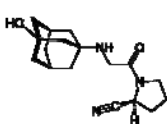
Unter computerbasiertem (*in silico*) rationalen Wirkstoffdesign versteht man die Anwendung verschiedener mathematischer Modelle und Methoden, zur Bewertung der Interaktion eines Zielproteins mit einer chemischen Substanz [SELZER, 2005]. Das Ziel ist dabei die Entwick-

lung eines optimierten Wirkstoffs. Arzneimittel, die unter maßgeblicher Beteiligung des Wirkstoffdesigns entwickelt wurden, konnten bereits erfolgreich im Markt platziert werden. Neuere prominente Beispiele dafür sind Glivec<sup>®</sup> [AYOUB *et al.*, 2005], ein Medikament zur Behandlung der Chronischen Myeloiden Leukämie, Glavus<sup>®</sup> zur Behandlung von *Diabetes Mellitus* Typ 2 [SCHEEN & PAQUOT, 2009], Tamiflu<sup>®</sup> und Relenza<sup>®</sup> [MACDONALD, 2000] zur Behandlung von Influenza, sowie Agenerase<sup>®</sup> [BROPHY *et al.*, 2000], Invirase<sup>®</sup> [REGAZZI *et al.*, 1999] und Norvir<sup>®</sup> [ARVIEUX & TRIBUT, 2005], drei HIV Protease Inhibitoren (siehe Tabelle 1.4).

Das strukturbasierte, rationale Wirkstoffdesign beinhaltet im Wesentlichen die Methoden virtuelles *Screening* [KITCHEN *et al.*, 2004], molekulare Leitstrukturoptimierung [DEAN, 1994] und das *de novo* Design [SCHNEIDER & FECHNER, 2005]. Grundlage dieser verschiedenen Aspekte des rationalen Wirkstoffdesigns ist stets die Vorhersage der Affinität des Liganden zum entsprechenden Rezeptor. Die Affinitätsvorhersage beruht auf der Bewertung der molekularen Wechselwirkungen zwischen dem Protein, d.h. der Aminosäuren im aktiven Zentrum mit dem gebundenen Liganden. Bei dieser Vorhersage bedient man sich struktureller, topologischer und physikalisch-chemischer Merkmale, die mittels computerbasierter Methoden berechnet werden können.

Tabelle 1.4: Beispiele für einige Medikamente die unter Beteiligung von rationalem Wirkstoffdesign entwickelt wurden.

Wirkstoff	Zielprotein	Hersteller	Krankheit	Zulassung In Europa
Norvir <sup>®</sup> (Ritonavir) 	Cytochrom P450 CYP 3A4	Abbott GmbH & Co. KG	HIV	1996
Relenza <sup>®</sup> (Zanamivir) 	Virale Neuraminidase	GlaxoSmith- Kline GmbH & Co. KG	Influenza	1999
Invirase <sup>®</sup> (Saquinavir) 	Cytochrom- P450- Enzymsystem	Hoffmann- La Roche AG	HIV	1996

Wirkstoff	Zielprotein	Hersteller	Krankheit	Zulassung In Europa
Tamiflu® (Oseltamivir) 	Virale Neuraminidase	Hoffmann- La Roche AG	Influenza	2002
Glivec® 	Abelson Tyro- sin-kinase (ABL)	Novartis AG	Chronischen Myeloiden Leukämie	2001
Agenerase® (Amprenavir) 	Virale Protease HIV-1	GlaxoSmith- Kline GmbH & Co. KG	HIV	2000
Glavus® (Vildagliptin) 	Dipeptidyl- Peptidase-4 (DPP4)	Novartis AG	Diabetes mellitus Typ 2	2007

## 1.8 Modelling und computergestütztes Wirkstoffdesign

Das rationale Wirkstoffdesign wird heutzutage durch leistungsfähige Computer unterstützt [DHALIWAL & CHEN, 2009]. Diese sind in der Lage, Eigenschaften von Molekülen zu modellieren. In den letzten Jahren hat das *Molecular Modelling*, besonders das Modellieren der Ligand-Rezeptor-Wechselwirkung, in der Arzneimittelforschung zunehmend an Bedeutung gewonnen [JIANG, 2008]. Die eigentliche molekulare Modellierung dient hierbei vorwiegend zur strukturellen Modifikation von Leitstrukturen, während beim struktur- und computergestützten Design sowohl neue Leitstrukturen identifiziert als auch modifiziert werden können [BARRIL & SOLIVA, 2006; TAFT *et al.*, 2008]. Neben dem *Modelling* tragen vor allem quantitative Strukturwirkungsbeziehungen zum Verständnis der Zusammenhänge zwischen der chemischen Struktur und ihrer Wirkung bei [CLARK, 2009]. In den folgenden Kapiteln soll nun näher auf die Methoden des *Modellings* und computergestützten Wirkstoffdesigns eingegangen werden. In Abbildung 1.13 ist ein typischer Prozess der Leitstrukturidentifizierung dargestellt.

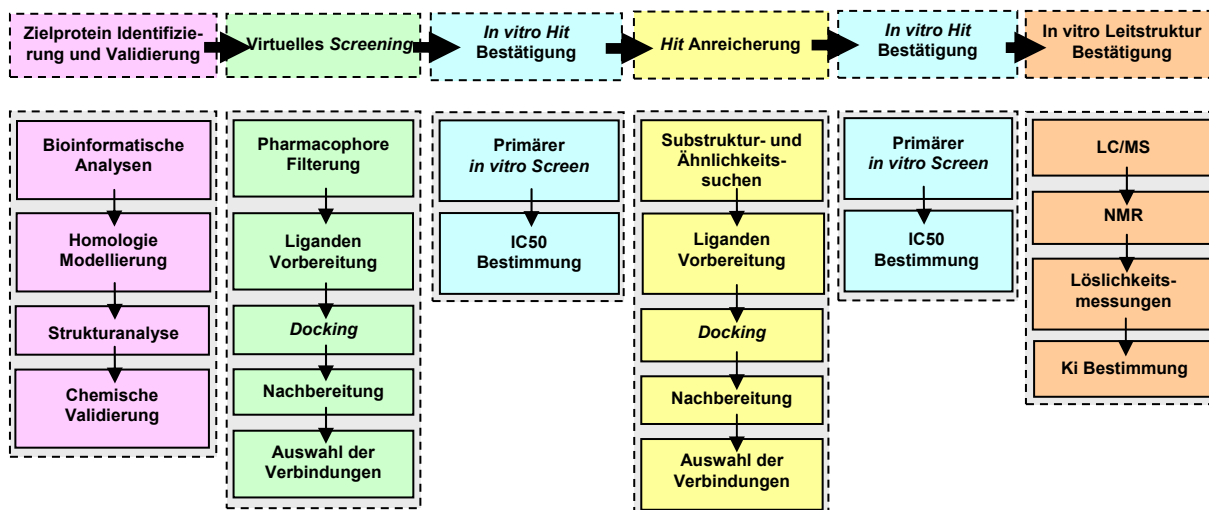


Abbildung 1.13: Workflow zur Identifizierung potentieller Leitstrukturen.

Dargestellt sind die sechs Hauptschritte im Leitstrukturidentifizierungs-Prozess. Dieser Prozess ist ein kombinierter Ansatz aus *in silico* sowie *in vitro* Arbeitsschritten.

### 1.8.1 Homologiemodellierung

Grundlage vieler Methoden des rationalen Wirkstoffdesigns, wie beispielsweise des molekularen *Dockings*, ist die Kenntnis der dreidimensionalen, molekularen Struktur (3D-Struktur) des Zielproteins [BARRIL & SOLIVA, 2006]. Oftmals ist diese Struktur jedoch nicht bekannt und kann, aufgrund der Aufwendigkeit des experimentellen Verfahrens, zur Ableitung molekularer 3D-Strukturen biologischer Makromoleküle, nicht einfach bestimmt werden [TAFT *et al.*, 2008]. Eine Möglichkeit, diese Methoden dennoch auch auf Zielproteine unbekannter 3D-Struktur anzuwenden, bietet das *Homology Protein Modelling* [CAVASOTTO & PHATAK, 2009]. Grundlage dieses Verfahrens ist die Annahme, dass homologe Proteine einer Proteinfamilie, die ähnliche Aminosäuresequenzen aufweisen, üblicherweise auch eine ähnliche Proteinfaltung zeigen. Ist die molekulare 3D-Struktur eines oder mehrerer Vertreter einer Proteinfamilie bekannt, können diese Referenzstrukturen (*Template*) als Schablone für die Ableitung eines molekularen Struktur-Modells eines homologen Proteins unbekannter Struktur genutzt werden [XIANG, 2006].



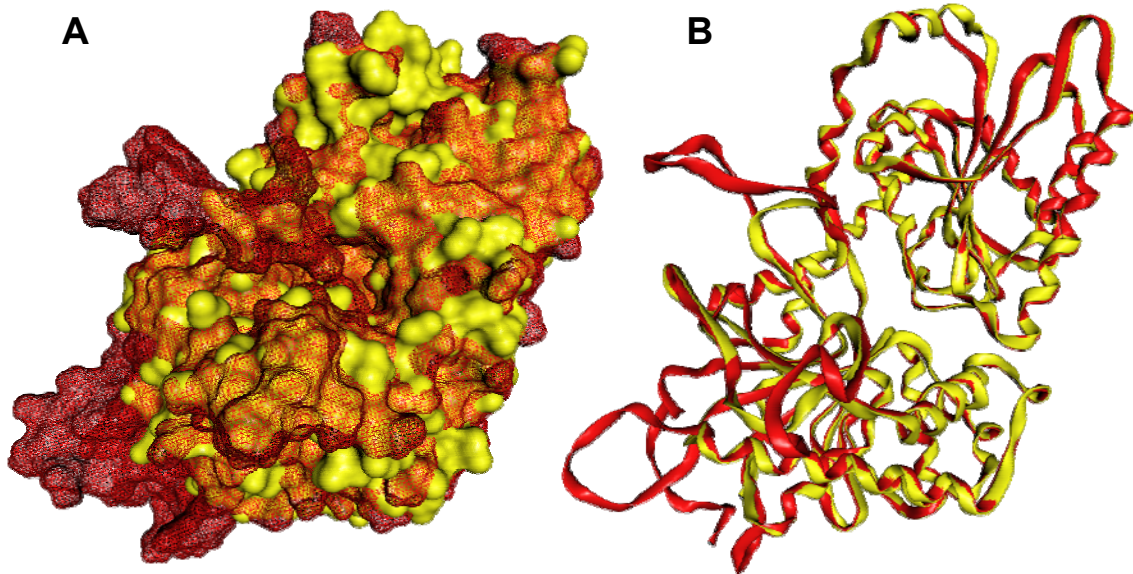


Abbildung 1.14: Überlagerung der Homologiemodelle der Trehalose-6-Phosphat-Synthase (T6PS) von *Drosophila melanogaster* und *Ctenocephalides felis* mit der *Escherichia coli* T6PS Kristallstruktur 1UQT.

A: Oberflächendarstellung des *D. melanogaster* Homologiemodells (rot) und der *E. coli* Kristallstruktur 1UQT (gelb). B: *Ribbon*-Darstellung des *C. felis* Homologiemodells (rot) und der *E. coli* Kristallstruktur 1UQT (gelb). Zum Vergleich wurden die mittels Molcad [BRICKMANN, 1995; BRICKMANN *et al.*, 2000] berechneten Oberflächen übereinander gelegt.

Am Anfang der Homologiemodellierung steht zunächst das *Alignment* der Aminosäuresequenzen des Proteins mit unbekannter Struktur, und eines Proteins mit bekannter Struktur [QU *et al.*, 2009]. Je höher die Sequenzidentität hierbei ist, desto besser sind auch die berechneten Homologiemodelle [XIANG, 2006]. Enthält das *Alignment* größere Lückenbereiche, werden Abschnitte des bekannten Proteins in das Modell eingefügt. Bei nicht übereinstimmenden Aminosäuren an den einzelnen Positionen wird versucht, die Konformationswinkel der Seitenketten entsprechend der Originalstruktur beizubehalten. Diese weniger konservierten Bereiche bilden oft Schleifen (*Loops*) und finden sich interessanterweise meist an der Oberfläche des Proteins. Sie zeigen einen vergleichsweise geringen Einfluss auf die strukturell konservierten Regionen im Innern des Proteins, wo meist auch das aktive Zentrum lokalisiert ist [SELZER *et al.*, 2008]. Im Anschluss erfolgt eine Bewertung des Modells, die Kollisionen zwischen Atomen berücksichtigt. Eventuell auftretende Atomkollisionen werden manuell entfernt [BÖHM *et al.*, 2002; FORSTER, 2002; SELZER *et al.*, 2008]. Auf diese Weise entsteht ein Proteinmodell der Sequenz, zu der keine experimentellen Strukturdaten vorliegen (siehe Abbildung 1.14).

### 1.8.2 Molekulares *Docking*

Als *Docking* bezeichnet man das computergestützte Einpassen kleiner chemischer Strukturen in die Bindetasche einer molekularen 3D-Proteinstruktur [BÖHM *et al.*, 2002; KLEBE, 2009; KROEMER, 2007]. Das Ziel ist dabei das Auffinden von Bindemodi möglichst geringer Energie. Dazu wird zunächst in einer rein qualitativen Betrachtung der Ligand in der Bindetasche platziert [CAVASOTTO & ORRY, 2007]. Die richtige Größe und Gestalt des Liganden sind, entsprechend des Schlüssel-Schloss-Prinzips [FISCHER, 1894], die in diesem Schritt wichtigsten Kriterien.

Nach dem *Docking* erfolgt eine quantitative Betrachtung. Diese wird als *Scoring* bezeichnet [LEACH *et al.*, 2006]. Hierbei wird die freie Bindungsenthalpie  $\Delta G$ , für die Bindung zwischen Ligand und Protein, mit Hilfe mathematischer Modelle abgeschätzt. Die Abschätzung von  $\Delta G$  beruht dabei auf der Bewertung der Wechselwirkungen zwischen funktionellen Gruppen des Liganden und des Proteins [STAHL, 2000]. Insbesondere finden hierbei Wasserstoffbrückenbindungen, ionische Wechselwirkungen, Metallkomplexe, Kation- $\pi$ -Wechselwirkungen,  $\pi$ - $\pi$ -Wechselwirkungen und hydrophobe Wechselwirkungen Berücksichtigung. Der hydrophobe Effekt ist im Wesentlichen entropischer Natur, d.h. es handelt sich nicht um eine direkte Wechselwirkung [KLEBE, 2009]. Vielmehr beruht der entropische Beitrag darauf, dass sich für das Ligand-Protein-Wasser-System die Zahl der Freiheitsgrade aufgrund der Ligand-Assoziation an das Protein ändert [BÖHM *et al.*, 2002; KITCHEN *et al.*, 2004; KLEBE, 2009].

### 1.8.3 Substruktur- und Ähnlichkeitssuchen

Um Substanzen zu identifizieren, die strukturell ähnlich zu bereits gefundenen aktiven Inhibitoren sind, jedoch eventuell eine bessere inhibitorische Aktivität aufweisen, dienen Ähnlichkeits- und Substruktursuchen. Bei Ähnlichkeitssuchen wird ein Satz von Charakteristika eines Ausgangsmoleküls, mit den Molekülen in einer Datenbank verglichen [WILLETT, 2003]. Daraus resultieren Moleküle, die dem Ausgangsmolekül chemisch ähnlich sind, aber durchaus einer anderen Strukturklasse angehören können.

Die Suche nach Strukturfragmenten dagegen bezeichnet man als Substruktursuchen [QUINTUS *et al.*, 2009]. Bei dieser Art der Suche erhält man Moleküle, welche das Ausgangsmolekül als Bestandteil ihrer Struktur enthalten, jedoch verschiedene Substituenten besitzen.

#### 1.8.4 Virtuelles *Screening*

Virtuelles *Screening* [BÖHM & SCHNEIDER, 2000; MCINNES, 2007] bezeichnet als Oberbegriff alle Methoden der computergestützten Suche in großen Substanzdatenbanken, die im Hochdurchsatz-Verfahren durchgeführt werden [WALTERS *et al.*, 1998]. In dieser Arbeit wurde *Docking* als Bewertungsmethode für das virtuelle *Screening* eingesetzt. Das virtuelle *Screening* ähnelt damit einem *High Throughput Screening* (labortechnische Hochdurchsatz-Reihenuntersuchung) [KLEBE, 2006; SCHNEIDER & FECHNER, 2005]. Ziel des virtuellen *Screenings* ist es, aus einer großen Substanzdatenbank diejenigen Substanzen herauszufiltern, die als potentielle Wirkstoffe am untersuchten Zielprotein eine biologische Funktion besitzen [WALTERS *et al.*, 1998]. Im Anschluss können dann die interessanten, potentiellen Wirkstoffe in verschiedenen Prüf-/Optimierungsrunden im Labor weiter optimiert werden. Die Methode erlaubt es auch synthetisch schwer zugängliche Substanzen im Hochdurchsatzverfahren auf ihre Wirkung am Zielprotein zu untersuchen [MCINNES, 2007; WALTERS *et al.*, 1998]. Zusätzlich eröffnet sie die Möglichkeit eines effektiven Tests kombinatorischer Substanzbibliotheken. Die Anzahl der zu synthetisierenden Substanzen kann durch das virtuelle *Screening* stark gesenkt werden. Dadurch werden die Kosten in einem Projekt gesenkt.

Da *Docking* einen hohen Rechenaufwand bedingt, sind virtuelle *Screening* Studien großer Substanzdatenbanken auf Basis von *Docking* sehr zeitaufwendig [DHALIWAL & CHEN, 2009; KROEMER, 2007]. Pro Struktur ist mit einem Zeitaufwand von einigen Minuten zu rechnen, so dass sich der Zeitaufwand für ein virtuelles *Screening* einer großen Substanzdatenbank leicht zu einer beträchtlichen Größe summiert. Zudem umfassen virtuelle Substanzdatenbanken eine deutlich größere Anzahl an Strukturen, als die zugehörige reale Substanzdatenbank, da alle Modifikationen die in der Realität auftreten, als eigenständige Strukturen in der virtuellen Substanzdatenbank zu erfassen sind. Zur effektiven Bearbeitung des virtuellen *Screenings* ist daher meist der Einsatz einer leistungsfähigen, parallelen Computerplattform notwendig [DHALIWAL & CHEN, 2009]. Zur Evaluierung des virtuellen *Screenings*, sowie Auswahl geeigneter *Docking*-Parameter, wird zunächst versucht, die Orientierung eines Liganden durch Einpassen in den vom Liganden befreiten Proteinkristall zu reproduzieren. Bei diesem als *Re-Docking* bezeichneten Vorgehen können optimale Einstellung der *Docking* Parameter für das anschließende virtuelle *Screening* ermittelt, sowie die beste *Scoring*-Funktion für diese *Docking*-Studie bestimmt werden. Darüber hinaus erlaubt das *Re-Docking*, des bekannten Liganden in die Homologiemodelle, die weitere Validierung der Modelle. Die

Güte der Homologiemodelle wird dabei über die Berechnung von RMSD-Werten [COHEN & STERNBERG, 1980] der Ligandatomen beurteilt.

### 1.8.5 LIE-Simulationen zur Berechnung der freien Bindungsenergie

Die *Linear Interaction Energy* Methode (LIE) dient der Berechnung von Bindungsaffinitäten [AQVIST & MARELIUS, 2001]. Diese semiempirische Simulationsmethode basiert auf Moleküldynamik- oder Monte-Carlo-generierten Ensembles von Konfigurationen [FOLOPPE & HUBBARD, 2006]. In dieser Arbeit wurde ausschließlich die Moleküldynamiksimulation zur Generierung von Ensembles verwendet. Die Moleküldynamik ist eine semiempirische Computersimulation der Bewegung von Atomen und Molekülen auf der Basis von Kraftfeldern [KLEBE, 2009]. Bei der Ermittlung der Affinitäten nach der LIE-Methode wird versucht, sinnvolle Informationen über die Bindungsenergie aus Simulationen der physikalisch relevanten Zustände (frei und gebunden) des Liganden abzuleiten [AQVIST *et al.*, 2002]. Die Idee hierbei ist, dass die freie Bindungsenergie des Liganden als Änderung seiner freien Energie betrachtet wird, wenn er aus der wässrigen Lösung (freier Zustand) in seine solvatisierte Bindungstasche (gebundener Zustand) überführt wird [DE AMORIM *et al.*, 2008]. Aus diesem Grund betrachtet die LIE-Methode nur die beiden Endpunkte im Ligandenbindungsprozess (freier Zustand sowie gebundener Zustand). Für die molekülmechanischen Berechnungen der Bindungsprozessendpunkte müssen zwei LIE-Simulationen durchgeführt werden. Eine Simulation für den gebundenen Liganden sowie eine weitere für den Liganden in wässriger Lösung (siehe Abbildung 1.15). Aus den molekülmechanischen Berechnungen erhält man jeweils für die beiden Zustände die elektrostatischen sowie die van der Waals Wechselwirkungsbeiträge.

Zur Generierung eines LIE-Modells werden außerdem experimentelle Bindungsaffinitäten mit den Energietermen der Molekülmechanik kombiniert, um eine Bewertungsfunktion zur Vorhersage von Bindungsenergien zu erhalten. Es lässt sich somit folgende allgemeine Näherung für die Bestimmung der freien Bindungsenergie aufstellen (siehe Gleichung 1.1) [AQVIST *et al.*, 2002]:

$$\Delta G = \alpha \Delta \langle U_{l-s}^{VDW} \rangle + \beta \Delta \langle U_{l-s}^{el} \rangle + \gamma \quad 1.1$$

Hierbei ist „ $\Delta \langle U_{l-s}^{VDW} \rangle$ “ gleich der Differenz der Ensemblemittelwerte der VDW Wechselwirkungen im gebundenen (l) und ungebundenen Zustand (s) des Liganden. Entsprechend repräsentiert „ $\Delta \langle U_{l-s}^{el} \rangle$ “ die elektrostatischen Wechselwirkungen zwischen Ligand und Umgebung.

Die Gewichtungsfaktoren  $\alpha$ ,  $\beta$  und  $\gamma$  werden im Anschluss durch lineare Regression an experimentell bestimmten Bindungsenergien angepasst.

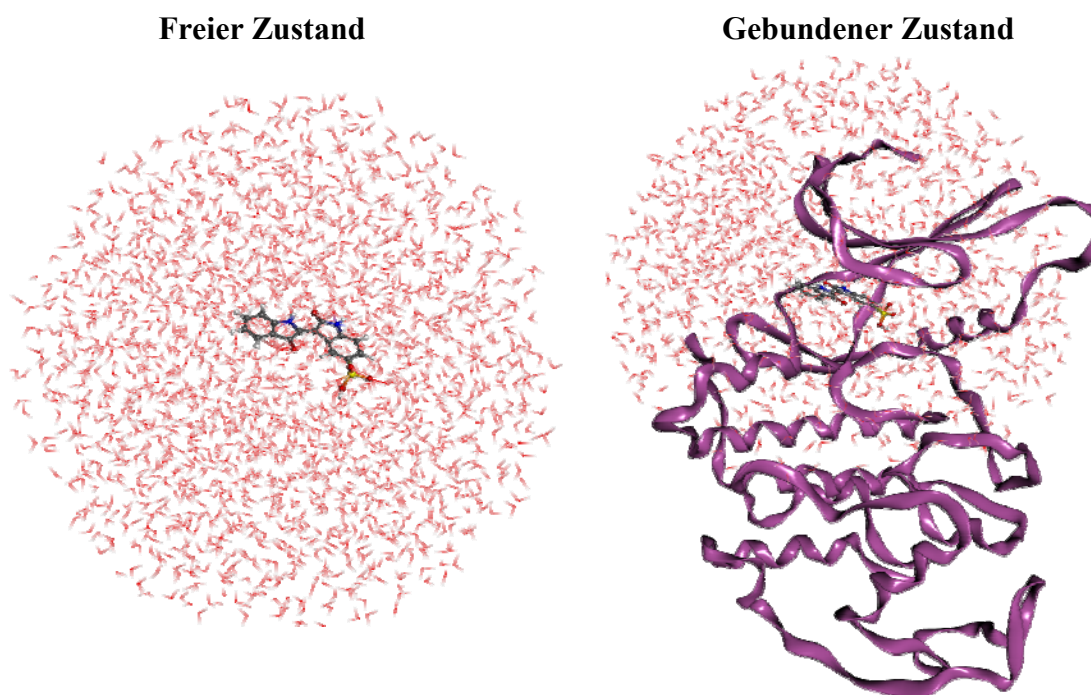


Abbildung 1.15: Linear Interaction Energy Methode

Dargestellt sind die notwendigen Simulationen für die Bestimmung der freien Bindungsenergie des Liganden: zum einen der Ligand in Lösung und zum anderen der Ligand gebunden an den voll solvatisierten Rezeptor.

Für die Qualität der Ergebnisse bei der Simulation mit MOE [<http://www.chemcomp.com/>], spielt das *Setup* und die Güte der Kraftfeldparameter für Ligand, Solvent und Rezeptor eine entscheidende Rolle [ALMLOF *et al.*, 2004].

### 1.8.6 Leitstrukturoptimierung

Die Optimierung von Leitstrukturen erfolgt meist in einem langwierigen, iterativen Prozess [NWAKA & RIDLEY, 2003; RAYMOND *et al.*, 2009]. Sie beinhaltet neben der Verbesserung der Bindungsaffinität zu einem Zielprotein, auch eine Optimierung weiterer Parameter wie beispielsweise Löslichkeit, Toxizität, Spezifität, Stabilität und Wirkdauer. Leitstrukturen werden über die gezielte Änderung der Gestalt und der Form, d. h. der dreidimensionalen Struktur und/oder physikochemischen Eigenschaften optimiert. Computermethoden wie molekulares *Docking*, QSAR-Methoden oder dem *De-novo*-Design unterstützen hierbei den iterativen Prozess und sind eng mit Struktursynthese und experimenteller Strukturbestimmung verwoben.

Grobe Struktur-Wirkungsbeziehungen lassen sich anhand von Bindungsaffinitäten potentieller Inhibitoren aufstellen. Essentielle Wechselwirkungen zwischen Liganden und Protein können durch das Einpassen der Liganden in die Bindetasche des Rezeptors (*Docking*) ermittelt werden. So lassen sich beispielsweise zusätzliche Bereiche in der Bindetasche identifizieren, die von bekannten Liganden noch nicht ausgenutzt werden. Oder es ergeben sich so Ideen für die Vereinfachung von Liganden, um beispielsweise die Löslichkeit, die Lipophilie oder die Transport- und Verteilungseigenschaften (ADME-Eigenschaften) zu optimieren [SCHWARDT *et al.*, 2003]. Hierbei ist jedoch die Kenntnis des tatsächlichen Bindemodus für die zu optimierende Leitstruktur entscheidend, um verlässliche Aussagen über die Struktur-Wirkungsbeziehungen treffen zu können.

Eine grundlegende Rolle, in der Optimierung von Inhibitoren zyklin-abhängiger Kinasen, spielt die Optimierung hingehend der Selektivität der Inhibitoren gegenüber anderen Kinasen, bzw. gegenüber Kinasen des Wirtsorganismus. Kinasen sind im Allgemeinen sehr hoch konserviert, speziell die ATP-Bindetasche ist bis auf wenige Aminosäureaustausche identisch [KLEBE, 2009; KUBINYI & MÜLLER, 2004; NOBLE *et al.*, 2004]. Dies macht es schwierig selektive Kinase-Inhibitoren zu entwickeln. Jedoch kann bereits ein einziger Austausch in der ATP-Bindetasche genutzt werden, um beispielsweise eine zusätzliche Binderegion zu adressieren (Glivec, ABL-Kinase) [KUBINYI & MÜLLER, 2004]. Weiterhin spielt oft auch die Löslichkeit von Kinase-Inhibitoren eine große Rolle im Leitstrukturoptimierungsprozess [JAUTELAT *et al.*, 2005; KHIRE *et al.*, 2004; OSLOB *et al.*, 2009].

## 1.9 Aufgabenstellung

Die rasante Weiterentwicklung von Computerhardware und Software erlaubt es heute, sehr große Datenmengen bio- und chemoinformatisch innerhalb kurzer Zeit zu verarbeiten. Computergestützte Methoden des virtuellen *Screenings* und des rationalen *Drug Designs* werden mittlerweile routinemäßig in der Agrar-, Tiergesundheits- und humanen Pharmaforschung eingesetzt. Ein wichtiger Bereich der veterinärmedizinischen Wirkstoffforschung sind die Parasitosen von Haus- und Nutztieren. *Eimeria tenella* ist ein parasitisch lebender Protozoe des Stammes der *Apicomplexa*. Dieser tierpathogene Parasit verursacht Kokzidiose, eine Hühnererkrankung die weltweit zu hohen Verlusten in der Geflügelproduktion führt. In der vorliegenden Arbeit sollte die Wirkstofffindung von chemotherapeutischen Antikokzidien durch moderne Methoden der Bio- und Chemieinformatik durchgeführt werden. In vorange-

gangenen Experimenten wurden CDK-Proteine als geeignete Zielproteine in *E. tenella* bereits chemisch validiert (siehe Tabelle 1.3). Außerdem konnten bereits erste Inhibitoren der EtCRK2 – einer CDK aus *E. tenella* – in einem virtuellen *Screening* identifiziert werden [BEYER, 2007].

Die vorliegende Arbeit gliedert sich in die folgenden drei Abschnitte: 1. Bioinformatische Identifizierung und Charakterisierung von CDK Proteinen und deren Aktivatoren (Zyklone) aus *E. tenella*. 2. Anreicherung und Optimierung von EtCRK2-Inhibitoren sowie 3. Virtuelles *Screening* mit der *Docking*-Software PLANTS nach PfMRK- und EtMRK-Inhibitoren.

Ziel im ersten Abschnitt dieser Arbeit war eine bioinformatische Tiefenanalyse von CDK-homologen sowie Zyklin-ähnlichen Proteinen von *E. tenella*. Hierbei sollte zunächst eine Charakterisierung dieser homologen CDKs in Bezug auf Funktion und Struktur stattfinden. Diese Charakterisierung erfolgte anhand der Sammlung und Bewertung von Sequenzinformation. CDKs sind innerhalb der *Apicomplexa* für *Plasmodium falciparum*, dessen Genom vollständig sequenziert ist, am besten untersucht. Daher wurde die Suche mit den bekannten CDKs und Zyklonen von *P. falciparum* durchgeführt. Darüber hinaus sollten die identifizierten Zyklone den CDK Proteinen aus *E. tenella* als potentielle Aktivatoren zugeordnet werden.

Das zweite Ziel dieser Arbeit war die Anreicherung von Inhibitoren, die bereits eine Aktivität auf EtCRK2 gezeigt haben. Diese Inhibitoren stammten aus einem früheren virtuellen *Screening* [BEYER, 2007] und konnten biochemisch bestätigt werden. Insgesamt waren zu Beginn dieser Arbeit elf Substanzen bekannt, die einen IC50-Wert von kleiner 10  $\mu$ M zeigten. Zudem war es Ziel dieser Arbeit, die Benzimidazol-Carbonitrile, die *in vitro* als Leitstrukturen der EtCRK2 bestätigt werden konnten, *in silico* zu optimieren. Hierfür sollte der Bindemodus dieser Strukturklasse im Detail analysiert werden, um im Anschluss geeignete Modifikationen dieser Verbindungen vorschlagen zu können.

Im dritten Abschnitt dieser Arbeit sollte ein virtuelles *Screening* nach potentiellen Inhibitoren der PfMRK und EtMRK durchgeführt werden. Diese Proteine sind ebenfalls Zyklin-abhängige Kinasen der Organismen *P. falciparum* und *E. tenella*. Hierzu sollte die *Docking*-Software PLANTS (Protein-Ligand-ANT-System) verwendet und evaluiert werden. Nach Methodevaluierung sollte zudem, geeignete Substrukturen aus einer Strukturdatenbank als potentielle *Hits* ausgewählt werden, um eine Liste geeigneter Inhibitoren humaner Kinasen zu erhalten, die in weiteren Schritten für die parasitären Zielproteine optimiert werden können.

## 2 Material und Methoden

### 2.1 Hardware und Software

#### 2.1.1 Hardware

- Linux-Cluster: Sun Fire V20z,  
2x AMD Opteron 250 (32 *Nodes*),  
Betriebssystem: SUSE Linux 9.0
- PC: Dell OptiPlex GX110, x86  
Betriebssystem: Microsoft Windows XP (SP1)
- Unix-Computer: SGI Origin 3200, SGI O2,  
SGI Octane 2, SGI Fuel  
Betriebssystem: Irix 6.5

#### 2.1.2 Software

- BioEdit 5.0.9, Sequenz-*Alignment* Editor  
North Carolina State University, USA, [HALL, 2009]
- ClustalW v1.4, Programm zur Erstellung multipler, globaler Sequenz-*Alignments*  
European Molecular Biology Laboratory, Heidelberg, [THOMPSON *et al.*, 1994]
- Cerius2 v.4.10, Chemoinformatisches Softwarepaket  
Accelrys Inc., San Diego, USA, [<http://www.accelrys.com/>]
- Corina 3.3, Konformergenerator  
Molecular Networks GmbH, Erlangen [<http://www.mol-net.com>]
- GeneDoc 2.6.001, Sequenz-*Alignment* Editor  
Pittsburgh Supercomputing Center, USA, [NICHOLAS *et al.*, 1997]
- GOLD 2.2 bis 4.0, Software zum Protein-Ligand-*Docking*  
Cambridge Crystallographic Data Centre, UK, [JONES *et al.*, 1995]



- Insight II, Chemoinformatisches Softwarepaket  
Accelrys Inc., San Diego, USA, [<http://www.accelrys.com/>]
- ISIS/Base, Datenbanksystem für chemische Substanzen  
MDL Information Systems Inc., San Ramon, USA, [<http://www.mdl.com>]
- Metalife Trinity1.4, Bioinformatisches Programmpaket  
Metalife AG, Winden, [<http://www.metalife.de/>]
- MOE, Chemoinformatisches Softwarepaket  
Tripos Inc., St. Louis, USA, [<http://www.chemcomp.com/>]
- Molcad, Software zur dreidimensionalen Visualisierung von Molekülen  
Tripos Inc., St. Louis, USA, [<http://www.tripos.com>]  
[BRICKMANN, 1995; BRICKMANN *et al.*, 2000]
- PipelinePilot, Chemoinformatisches Pipelinetool  
Accelrys Inc., San Diego, USA, [<http://www.accelrys.com/>]
- Volsurf+, Softwarepaket zur Berechnung von Deskriptoren  
Molecular Discovery Ltd., ITALY, [<http://www.moldiscovery.com>]

## 2.2 Datenbanken

Tabelle 2.1: Verwendete Datenbanken.

Datenbanken	URL	Referenz
BindingDB	<a href="http://www.bindingdb.org">http://www.bindingdb.org</a>	[CHEN <i>et al.</i> , 2001]
DUD	<a href="http://dud.docking.org/">http://dud.docking.org/</a>	[HUANG <i>et al.</i> , 2006]
GenBank	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>	[BENSON <i>et al.</i> , 1993]
GenScan	<a href="http://genes.mit.edu/GENSCAN.html">http://genes.mit.edu/GENSCAN.html</a>	[BURGE & KARLIN, 1997]
PubMed	<a href="http://www.pubmed.gov/">http://www.pubmed.gov/</a>	[WHEELER <i>et al.</i> , 2004]

Datenbanken	URL	Referenz
PROSITE	<a href="http://us.expasy.org/prosite/">http://us.expasy.org/prosite/</a>	[HULO <i>et al.</i> , 2006]
Pfam	<a href="http://www.sanger.ac.uk/Software/Pfam/index.shtml">http://www.sanger.ac.uk/Software/Pfam/index.shtml</a>	[BATEMAN <i>et al.</i> , 2004]
InterPro	<a href="http://www.ebi.ac.uk/interpro/">http://www.ebi.ac.uk/interpro/</a>	[MULDER <i>et al.</i> , 2005]
Swiss-Prot	<a href="http://us.expasy.org/sprot">http://us.expasy.org/sprot</a>	[BOECKMANN <i>et al.</i> , 2003]
PDB	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>	[BERMAN <i>et al.</i> , 2003]

## 2.3 Methoden zum Sequenzvergleich

### 2.3.1 Sequenz-Alignments

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 *Metalife Trinity 1.4* implementierten Programm *ClustalW* [THOMPSON *et al.*, 1994], unter Verwendung der Standardeinstellungen, durchgeführt. Hierzu wurden die Sequenzen in ein Sequenzset im Modul *Metalife-Predictor* der Software *Metalife Trinity 1.4* [Metalife AG, Winden, Deutschland] importiert. *ClustalW* erstellt aus multiplen FASTA-Dateien globale, multiple *Alignments*. Die Ergebnisse wurden, zur Visualisierung und graphischen Reproduktion, in das Programm *GeneDoc* [NICHOLAS *et al.*, 1997] importiert.

### 2.4 Identifizierung von CDK-ähnlichen und Zyklin-ähnlichen Proteinen

Alle Sequenzdaten wurden vom National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) oder vom Sanger Institute (<http://www.sanger.ac.uk/>) bezogen. Zur Identifizierung von CDK-ähnlichen Proteinen, sowie der entsprechenden Zykline aus *E. tenella*, wurden rekursive BLAST-suchen (BLASTN, TBLASTX oder TBLASTN) [KORF, 2003] auf dem OmniBlast Server des *E. tenella* Sequenzierungsprojekts ([http://www.sanger.ac.uk/cgi-bin/blast/submitblast/e\\_tenella/omni](http://www.sanger.ac.uk/cgi-bin/blast/submitblast/e_tenella/omni)) durchgeführt. Diese BLAST Datenbank beinhaltet unter anderem die vollständigen, jedoch noch nicht publizierten Sequenz-Daten des

*E. tenella* Genoms. Rekursive BLAST-Suchen wurden mit bereits bekannten CDK-ähnlichen und Zyklin-ähnlichen Proteinen von *P. falciparum* durchgeführt. Außerdem wurden zuvor identifizierte Sequenzen von anderen Organismen aus dem Stamm der *Apicomplexa*, wie beispielsweise *Plasmodium yoelii yoelii*, *Cryptosporidium parvum*, *Cryptosporidium hominis*, *Theileria parva*, *Theileria annulata* und *Toxoplasma gondii* [ENGELS, 2005], zur weiteren Verifizierung der mutmaßlichen Genstruktur mittels TBLASTN herangezogen. Der Vergleich der bekannten Proteine erlaubte neben einer genauen Charakterisierung, wie beispielsweise der Ermittlung von Sequenzmotiven und strukturellen Eigenschaften, auch die Berechnung des Konservierungsgrades dieser Proteine, innerhalb der einzelnen Ordnungen *Coccidia*, *Haemosporida* und *Piroplasmida*. Die identifizierten Sequenzen aus *E. tenella*, mit einem *E-value* [KORF, 2003] kleiner  $10^{-3}$  und einem *Bit-Score* größer als 100, wurden als signifikant erachtet.

Da es sich bei den so ermittelten neuen Sequenzen um genomische Sequenzen handelt, mussten im Nachgang noch Exon- und Intronbereiche vorhergesagt werden. Dazu wurde der *GenScan*-Server des *Massachusetts Institute of Technology* verwendet. Das Programm *GenScan* [BURGE & KARLIN, 1997] basiert auf einem Hidden Markov Modell [EDDY, 2004], das trainiert wurde, die Exon-/Intronstruktur eukaryotischer Gene aus genomischen Sequenzen zu extrahieren. Diese Vorhersage verlief nicht fehlerfrei, so dass die Exon-/Introngrenzen manuell, mittels der GT-AG-Regel [MOUNT, 1982] angepasst werden mussten. Ein möglicher Grund für die fehlerhafte Vorhersage, ist die Länge der eingesetzten genomischen Sequenzen, die teilweise mehrere tausend Basenpaare umfasste. Eine korrekte Vorhersage der Exon/Intron-Struktur wird, mit zunehmender Länge und Komplexität der Testsequenz, immer unwahrscheinlicher [KANTER, 2005].

## 2.5 Homologiemodellierung der EtMRK und der PfMRK

Alle Homologiemodelle der PfMRK und der EtMRK wurden mit dem Softwarepaket MOE erstellt. Die Modelle wurden mittels Standardeinstellungen erstellt, soweit nicht anders angegeben. Die Sequenzen von *Target* und *Template* wurden zunächst in MOE, mittels der *blosum62* Substitutionsmatrix [HENIKOFF & HENIKOFF, 1993], aligniert. Als Templatstruktur wurde eine humane CDK2 Kristallstruktur [ANDERSON *et al.*, 2003] (PDB-ID = 1OIR) verwendet. Diese CDK2 Struktur liegt in der inaktiven Konformation vor, d.h. ohne gebundenes Zyklin und ohne Phosphorylierung im Aktivierungs-*Loop*. Aufgrund der hohen Auflö-

sung von 2,1 Å, wurde diese Struktur als Templat für die Modellbildung eingesetzt. Die Auswahl der Templatstruktur erfolgte anhand verschiedener Parameter, wie beispielsweise isotroper Debye-Waller-Temperaturfaktoren (B-Faktoren) [GAO & PENG, 1999] der Templatstrukturen, Ausrichtung der Seitenketten, sowie Berechnung der mittleren quadratischen Abweichung der einzelnen Atomlagen (*Root Mean Square Deviation*, RMSD-Werte) [COHEN & STERNBERG, 1980] zwischen *Template*- und *Target*-Struktur. Es wurden jeweils für EtMRK und PfMRK zehn Homologiemodelle mit dem höchstmöglichen Optimierungsgrad generiert. Jedes Modell wurde zunächst grob minimiert, zur Optimierung der sterischen Interaktionen, und dann jeweils das beste Zwischenprodukt noch einmal etwas feiner minimiert. Zur Minimierung wurde das MMFF94 Kraftfeld [HALGREN, 1996] verwendet. Einige Seitenketten wurden gegebenenfalls manuell optimiert. Hierfür wurde beispielsweise ein geeigneteres Rotamer benutzt. Dies diente der Optimierung der Bindungsmodi einzelner Referenzliganden. Die Visualisierung der finalen Modelle erfolgte mit dem Programm Molcad [BRICKMANN, 1995; BRICKMANN *et al.*, 2000]. Für die Generierung der Ramachandran-Plots [RAMACHANDRAN *et al.*, 1963] wurde ebenfalls das Softwarepaket MOE verwendet.

## 2.6 Substruktur- und Ähnlichkeitssuchen

Um Substanzen zu identifizieren, die strukturell ähnlich zu bereits bekannten aktiven Inhibitoren sind [BEYER, 2007], aber eventuell eine bessere inhibitorische Aktivität aufweisen, wurden Ähnlichkeits- und Substruktursuchen durchgeführt. Hierfür wurde zum einen das chemische Datenbanksystem ISIS/Base verwendet und zum anderen die PipelinePilot Software. Bei Ähnlichkeitssuchen wird ein Satz von Charakteristika eines Ausgangsmoleküls mit den Molekülen in einer Datenbank verglichen. Zur Beschreibung der chemischen Strukturen wurden die in PipelinePilot implementierten *Fingerprints* [Accelrys Inc., San Diego, USA] ECFP\_4 und FCFP\_4 [ROGERS *et al.*, 2005] berechnet. Es wurde ein Tanimoto-Index [JACCARD, 1912; WILLETT, 2006] von 70% als Schwellenwert für die Ähnlichkeitssuchen verwendet. Es resultierten hieraus 2.331 Substanzen, die dem Ausgangsmolekül chemisch ähnelten, teilweise aber auch einer anderen Strukturklasse angehörten.

Die Suche nach Strukturfragmenten ergab 42.099 Substanzen. Bei dieser Art der Suche erhält man Moleküle, welche das Ausgangsmolekül als Bestandteil ihrer Struktur enthalten, aber verschiedene Substituenten tragen.

## 2.7 Vorbereitung der Substanzdaten für das molekulare Docking

Chemische Verbindungen können in der Realität verschiedene Zustände, hinsichtlich ihrer strukturellen Eigenschaften (Konformere), ihrer elektrochemischen Eigenschaften (Ionisierungszustände) und ihres Tautomerisierungszustandes (Tautomere) [OELLIEN *et al.*, 2006], annehmen. Die verschiedenen Konformere einer chemischen Verbindung können im molekularen *Docking* automatisch betrachtet werden. Mögliche Ionisierungszustände und Tautomere müssen jedoch vorher berechnet und als individuelle Strukturen abgespeichert werden. Die Aufbereitung der Substanzdatenbank wurde mit Tcl-Skripten, unter Einsatz der Programmierbibliothek Cactvs [IHLENFELDT *et al.*, 2002; OELLIEN *et al.*, 2005], durchgeführt. Für die so vorbereiteten Strukturen wurden dann im Anschluss Startkonformere mit der Software Corina berechnet. Zudem wurden die Strukturen entsprechend ihres Ionisierungszustandes mit Wasserstoffatomen abgesättigt.

## 2.8 Docking

### 2.8.1 GOLD

Als Programm zum Durchführen des *Dockings*, zur Bewertung der potentiellen EtCRK2 [KINNAIRD *et al.*, 2004] Inhibitoren, wurde GOLD v4.0 (*Genetic Optimisation for Ligand Docking*) [JONES *et al.*, 1995] verwendet. GOLD benutzt einen genetischen Algorithmus zum Einpassen der chemischen Strukturen in die Proteinbindetasche. Bei einem genetischen Algorithmus wird der Prozess der Evolution mittels Mutation, Rekombination und Selektion von Individuen einer Population simuliert. Die Individuen entsprechen hierbei den jeweiligen Ligandenkonformationen. Eine Population ist demnach die Menge aller Ligandenkonfigurationen. Die Fitness (*Scoring*-Wert) eines Individuums bzw. einer Ligandenkonformation, wird mittels Optimierungskriterien, wie z. B. der Wechselwirkungsenergie zwischen einem Ligand und seinem Protein, berechnet. Das Protein wird bei diesem Verfahren starr gehalten, nur die Positionen der Wasserstoffatome können sich ändern. Der Ligand hingegen wird flexibel behandelt. Für das *Docking* können in GOLD verschiedene Arten von *Scoring*-Funktionen ausgewählt werden. Neben den bereits implementierten *Scoring*-Funktionen GOLD-Score [JONES *et al.*, 1995] und Chem-Score [ELDRIDGE *et al.*, 1997] ist es möglich, benutzerdefinierte Funktionen zu nutzen. Für das so durchgeführte *Docking* wurde die GOLD-Score Fitnessfunktion verwendet. 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 (siehe Gleichung 2.1).

$$GOLDScore = 1,375 \cdot VDW_{\text{int ermolekular}} - HB_{\text{int ermolekular}} - SE_{\text{int ramolekular}} \quad 2.1$$

Als Parametereinstellung für den genetischen Algorithmus wurden die Standardeinstellungen verwendet, die eine sehr genaue Vorhersage gewährleisten. Ein Nachteil ist jedoch der hohe Zeitverbrauch bei dieser Einstellung [JONES *et al.*, 1995].

Die GOLD-Parametereinstellungen für das molekulare *Docking* sind in Tabelle 2.2 aufgeführt.

Tabelle 2.2: Verwendete GOLD-Parametereinstellungen.

Einstellungen	Parameterbezeichnung in GOLD	Wert
Automatic Settings	search efficiency	200% for optimization
		10% for screening
Population	popsiz	auto
	select_pressure	auto
	n_islands	auto
	maxops	auto
	niche_siz	auto
Genetic Operators	pt_crosswt	auto
	allele_mutatewt	auto
	migratewt	auto
Flood Fill	radius	12
	do_cavity	1
	floodfill_atom_no	0

Einstellungen	Parameterbezeichnung in GOLD	Wert
Data Files	param_file	default
	set_ligand_atom_files	1
	set_protein_atom_files	1
	tordist_file	default
Flags	display	0
	internal_ligand_h_bonds	1
	n_ligand_bumps	0
	flip_free_corners	1
	flip_amide_bonds	1
	flip_planar_n	1
	flip_pyramidal_n	1
	rotate_carboxylic_oh	flip
use_tordist	1	
Termination	early_termination	0
Constraints	force_constraints	0
Fitness Function	initial_virtual_pt_match_max	3.0
	relative_ligand_energy	0
	start_vdw_linear_cutoff	6

### 2.8.2 PLANTS

Das *Docking*-Programm PLANTS (*Protein Ligand Ant System*) [KORB *et al.*, 2006; KORB *et al.*, 2009] diente zum virtuellen *Screen* nach potentiellen EtMRK- und PfMRK-Inhibitoren. Der verwendete Algorithmus in diesem Programm basiert auf einem stochastischen Optimierungsalgorithmus - dem „Ameisenalgorithmus“ [DORIGO *et al.*, 1996]. Dieser Algorithmus lehnt sich an das Verhalten von Ameisen an, die in der Lage sind, den kürzesten Weg zwi-

schen ihrem Nest und einer Futterquelle zu finden. Hierfür nutzen sie indirekte Kommunikation in Form von Pheromon-markierten Pfaden. Auf das molekulare *Docking* angewendet, findet die artifizielle Ameisenkolonie das Energieminimum einer Ligandenkonformation in der aktiven Tasche des Proteins mittels verschiedener „Duftspuren“. Hierbei imitieren die artifiziellen Ameisen das Verhalten von realen Ameisen. Sie markieren niederenergetische Ligandenkonformationen mit Pheromonen. Die Information des artifiziellen Pheromonpfades wird dann in darauf folgenden Iterationen so modifiziert, dass niederenergetische Ligandenkonformationen mit einer höheren Wahrscheinlichkeit generiert werden.

Für die *Docking*-Studien mit PLANTS wurde die  $PLANTS_{CHEMPLP}$  *Scoring*-Funktion (PLANTS-Score) [KORB *et al.*, 2009] verwendet (siehe Gleichung 2.2). Diese beinhaltet einige Terme aus der Chem-Score-Implementierung in GOLD wie winkelabhängige Terme für Wasserstoffbrückenbindungen ( $f_{hb}$ ,  $f_{hb-ch}$ ,  $f_{hb-CHO}$ ) und Metallbindungen ( $f_{met}$ ,  $f_{met-coord}$ ,  $f_{met-ch}$ ,  $f_{met-coord-ch}$ ). Zudem enthält sie folgende vier eigene Bewertungskomponenten: 1. Stückweises lineares Potential (PLP = *Piecewise Linear Potential*) [Gelhaar *et al.*, 2005], 2. empirisches Potential für Schweratome zur Verhinderung von internen Ligandenkollisionen, 3. Energie der Torsionsspannung des Liganden und 4. Bindungsenergie der intramolekularen Wasserstoffbrücken im Liganden (siehe Gleichung 2.2).

$$f_{PLANTS_{CHEMPLP}} = f_{PLP} + f_{hb} + f_{hb-ch} + f_{hb-CHO} + f_{met} + f_{met-coord} + f_{met-ch} + f_{met-coord-ch} + f_{clash} + f_{tors} + c_{site} \quad 2.2$$

Es wurden folgende Parametereinstellungen für das molekulare *Docking* verwendet (siehe Tabelle 2.3).

Tabelle 2.3: Verwendete PLANTS-Parametereinstellungen.

Einstellungen	Parameterbezeichnung in PLANTS	Values
search settings	aco_ants	50
	aco_evap	0.25
	aco_sigma	0.25



Einstellungen	Parameterbezeichnung in PLANTS	Values
input	protein_file	EtMRK_OPTI.mol2
	ligand_file	BindingDB3D_V2000_cor.mol2
output	output_dir	s025a50p025
write single mol2 files	write_multi_mol2	1
binding site definition	bindingsite_center	-2.079 32.8943 11.2825
	bindingsite_radius	14
scoring function	scoring_function	chemplp
cluster algorithm	cluster_structures	1
	cluster_rmsd	1.0

### 2.8.3 Eingesetzte Bewertungsmethoden

#### 2.8.3.1 RMSD-Wert

Zum Vergleich der Qualität der einzelnen *Docking*- oder *Screening*-Experimente wurde der RMSD-Wert ( $RMSD = \text{Root-Mean-Square Deviation}$ ) der einzelnen Schweratome herangezogen [COHEN & STERNBERG, 1980]. Dieser gibt den durchschnittlichen Abstand zwischen der berechneten Ligandenposition und der experimentell gefundenen Ligandenposition an (siehe Gleichung 2.3).

$$RMSD = \sqrt{\frac{\sum_i d_i^2}{n}} \quad 2.3$$

Wobei  $d_i$  die Differenz zwischen der Position des Schweratoms  $i$  des Liganden aus der PDB und der Position des Atoms  $i$  des gedockten Liganden darstellen. Der Anteil der Liganden mit einem RMSD kleiner oder gleich 2 Å wird als korrekte Ligandenpose gewertet.

### 2.8.3.2 Spearman Korrelationskoeffizient ( $R_s$ )

Die Korrelationsanalyse zwischen *Scoring* und dem RMSD-Wert wurden nach Spearman durchgeführt [WISSELER, 1905]. Bei dieser Art Analyse wurde untersucht, inwieweit die Variablen *Scoring*- und *RMSD*-Werte in Beziehung zueinander stehen.

$$R_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n-1)} \quad 2.4$$

Der hierbei berechnete Spearman Korrelationskoeffizient  $R_s$  kann zwischen -1 und 1 variieren (-1 = inverse Korrelation, 0 = keine Korrelation, 1 = Korrelation) (siehe Gleichung 2.4).

### 2.8.3.3 AUC-Wert

Um zu überprüfen, ob die verwendeten *Docking*-Tools eine Anreicherung von aktiven Substanzen abbilden, wurde die Fläche unter der ROC-Kurve (ROC = *Receiver Operating Characteristic*) berechnet. In der ROC-Kurve wird die Rate der „Falsch-Positiven“ (inaktive Liganden) mit hohem *Docking-Score* zur Rate der „Richtig-Positiven“ (aktive Liganden) aufgetragen. Die berechnete Fläche unter der ROC-Kurve wird als AUC-Wert (AUC = *Area Under Curve*) bezeichnet [DUDA *et al.*, 2001]. Dieser kann Werte zwischen null und eins annehmen. Ein schlechtes Modell würde eine flache Kurve in der Gegend der Winkelhalbierenden mit einem AUC-Wert von ca. 0,5 liefern.

### 2.8.3.4 Anreicherungsfaktor (EF)

Der *Docking*-Anreicherungsfaktor (EF = *Enrichment Factor*) gibt Auskunft über die Fähigkeit, innerhalb von *Docking*-Experimenten, die „Richtig-Positiven“ gegenüber dem Gesamtdatensatz herauszufiltern, im Vergleich zu einer willkürlichen Auswahl (siehe Gleichung 2.5) [CAVASOTTO *et al.*, 2005].

$$EF = \frac{Hits_{sampled}^{x\%}}{N_{sampled}^{x\%}} \cdot \frac{N_{total}}{Hits_{total}} \quad 2.5$$

Zur Bewertung der Anreicherungen wurden drei verschiedene Anreicherungs-Faktoren berechnet (EF1 = Anreicherung innerhalb der oberen 1% der Datenbank, EF2 = 2% der Datenbank, EF3 = 5% der Datenbank).

## 2.9 Docking der erhaltenen Substanzen nach Substruktursuchen

Mit der *Docking*-Software GOLD wurde die erhaltene Substanzbibliothek, aus den Substruktur- und Ähnlichkeitssuchen, die 44.430 Substanzen enthielt, auf einem Linux-Cluster untersucht. Die so gewonnenen Rohdaten wurden dann in einer Post-Prozessierung weiter analysiert. Die Post-Prozessierung umfasste 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 ein Tcl-basiertes Cactvs-Skript verwendet wie bereits von Beyer beschrieben [BEYER, 2007].

Die aus diesem Verfahren resultierenden Strukturen wurden dann anhand des *Scoring*-Wertes analysiert und bewertet. Davon wurden die ca. 200 Strukturen für jede Strukturklasse mit den günstigsten Bewertungen, in der Bindetasche visuell 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.

## 2.10 Virtuelles Screening nach potentiellen EtMRK- und PfMRK-Inhibitoren

Für die virtuelle Suche nach potentiellen EtMRK- und PfMRK-Inhibitoren, wurde die BindingDB Datenbank in einem virtuellen *Screening* geprüft. Diese öffentlich zugängliche Datenbank enthält ca. 21.000 experimentell bestimmte Bindungsaffinitäten für Protein-Ligand-Komplexe von 110 Zielproteinen mit verschiedenen Isoformen und Mutanten sowie ca. 11.000 Verbindungen. Für das virtuelle *Screening* wurden die 5.461 Kinase-Inhibitoren mit der *Docking*-Software PLANTS jeweils in das aktive Zentrum der EtMRK und der PfMRK eingepasst. Pro Ligand wurden jeweils zehn *Docking*-Posen generiert. Die Bewertung der Posen, sowie das finale *Ranking*, wurden mit der PLANTS<sub>CHEMPLP</sub> *Scoring*-Funktion (PLANTS-Score) durchgeführt. Danach wurde für die zwei Prozent besten Strukturen der finalen Datenbank diverse Murcko *Assemblies* [BEMIS & MURCKO, 1996] mittels der Software PipelinePilot generiert, um geeignete *Hit-Cluster* für die EtMRK und die PfMRK auszuwählen.

## 2.11 Linear Interaction Energy (LIE) Berechnungen

Alle LIE-Simulationen [AQVIST & MARELIUS, 2001] wurden mit dem Softwarepaket MOE, unter Verwendung des MMFF94 Kraftfeldes [HALGREN, 1996] durchgeführt, ähnlich wie bereits bei English *et al.* beschrieben [ENGLISH, 2007]. Zur Generierung der Ensembles wurden Moleküldynamiksimulationen verwendet. Für alle Simulationen wurde eine sphärische Simulationsbox mit einem Radius von 12 Å benutzt (Abbildung 2.1). Ligand- bzw. Ligand-Proteinkomplexe wurden innerhalb der Simulationsbox mit Wassermolekülen in *Bulkdichte* solvatisiert. Als Modell für das Wasser wurde TIP3P benutzt [CARLSON *et al.*, 1985]. Atome außerhalb der Wassersphäre wurden starr gehalten. Um die einzelnen Wassermoleküle daran zu hindern, in das Vakuum abzuwandern, wurden alle Sauerstoffe in der äußeren Wassersphäre, innerhalb von 3 Å (Abbildung 2.1), mit einer harmonischen Rückstellkraft von 50 kcal mol<sup>-1</sup> · Å<sup>-2</sup> belegt.

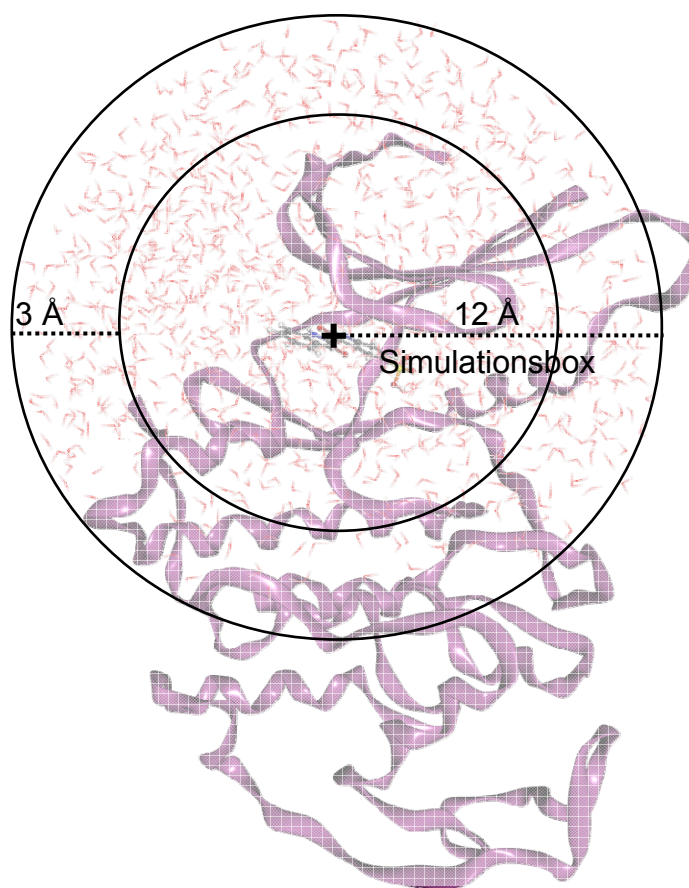


Abbildung 2.1: Simulationsbox des Ligand-Proteinkomplex

Dargestellt ist die sphärische Simulationsbox mit einem Radius von 12 Å. Alle Atome die außerhalb dieser Box liegen, wurden starr gehalten. Wassermoleküle die sich in der äußeren Kugelschale mit einer Schichtdicke von 3 Å befinden, wurden durch eine harmonische Rückstellkraft belegt.

Die Simulationssysteme wurden in einem mehrstufigen Prozess energieminiert. Für den freien Zustand wurde eine Zweiphasen-Minimierung durchgeführt. In der ersten Minimierungsphase wurden die Schweratome mit harmonischen Rückstellkräften von  $20 \text{ kcal mol}^{-1} \cdot \text{Å}^{-2}$  belegt. In der zweiten Minimierungsphase wurden dann alle Restriktionen aufgehoben. Nacheinander wurden die Methoden *Steepest Descent* (SD) [CURRY, 1944], *Conjugate Gradient* (CG) [PAYNE *et al.*, 1992] und *Truncated Newton* (TN) [SCHLICK & FOGELSON, 1992] in den beiden Minimierungsphasen angewendet. Die Abbruchgradienten wurden jeweils auf  $1.000$ ,  $100$  und  $1 \text{ kcal mol}^{-1} \cdot \text{Å}^{-1}$  festgelegt.

Für den gebundenen Zustand wurden die Gradienten in der TN Methode abschnittsweise in einer Vierphasen-Minimierung reduziert. In der ersten Minimierungsphase wurden alle Schweratome fixiert und ein TN Gradient von  $100 \text{ kcal mol}^{-1} \cdot \text{Å}^{-1}$  verwendet. Danach wurden alle Schweratome des Rezeptors mit harmonischen Rückstellkräften von  $500 \text{ kcal mol}^{-1} \cdot \text{Å}^{-2}$  belegt. In einem weiteren Minimierungsschritt wurden alle Rezeptoratome, die in einem Radius von  $15 \text{ Å}$  zu dem Liganden liegen, als sogenannte „Binderegion“ definiert. Alle Schweratome die sich in dieser Region befinden, wurden mit einer Kraft von  $100 \text{ kcal mol}^{-1} \cdot \text{Å}^{-2}$  festgehalten. Außerhalb dieser Region betrug die Restriktionskraft für Schweratome  $500 \text{ kcal mol}^{-1} \cdot \text{Å}^{-2}$ . Für diese Region wurde ein TN Gradient von  $10 \text{ kcal mol}^{-1} \cdot \text{Å}^{-1}$  verwendet. Abschließend wurden alle Schweratome, sowohl in der Binderegion als auch außerhalb dieser Region, mit einer Kraft von  $20$  bzw.  $100 \text{ kcal mol}^{-1} \cdot \text{Å}^{-2}$  festgehalten. Hierfür wurde ein TN Gradient von  $1 \text{ kcal mol}^{-1} \cdot \text{Å}^{-1}$  verwendet.

Die Moleküldynamiksimulation wurde im kanonischen Ensemble im Anschluss an die zuvor beschriebenen Energieminimierungen durchgeführt. Bei diesem Ensemble werden die Teilchenanzahl ( $N$ ), das Volumen ( $V$ ) und die Temperatur ( $T$ ) konstant gehalten. Es wurden folgende Parametereinstellungen für die Moleküldynamik verwendet (siehe Tabelle 2.4).

Die relativen Bindungsenergien wurden auf Basis der erhaltenen Trajektorien, mittels der LIE-Methode gemäß Gleichung 1.1 in MOE berechnet. Die experimentellen  $K_i$ -Werte wurden in die relativen Bindungsenergien  $\Delta G$  mittels Gleichung 2.6 umgerechnet.

$$\Delta G = -R \cdot T \cdot \ln K_i \quad 2.6$$

Tabelle 2.4: Verwendete Einstellungen der Moleküldynamiksimulation in MOE.

<b>Parameter- bezeichnung in MOE</b>	<b>Werte</b>	<b>Beschreibung</b>
dt	0.0015	time step (ps)
sample	1000	sample period (ps)
T0	298.0	initial velocities temperature
QT	1.0	temperature response time (ps)
QP	2.0	pressure response time (ps)
waterConstraints	'rigid'	['none', 'bonds', 'rigid']
bondConstraints	'light'	['none', 'light', 'all']
constraintTol	1.0e-9	tolerance for holonomic constraints
savePosition	0	save positions to database
saveVelocity	0	save velocities to database
saveAcceleration	0	save accelerations to database
ensemble	'NVT'	'NVT', 'NPT', 'NVE', 'NPH'
en_sample	0.005	time for energy sampling
ht	0.0	heating time (ps)
et	1.5	'production' time for equilibration (ps)
prodt	5.5	production time at equilibrium temperature (ps)
ct	0.0	cooling time (ps)
hT	50.0	initial temperature (K)
eT	298.0	equilibrium temperature (K)
cT	298.0	cooling target temperature (K)
hP	101.0	initial pressure (kPa)
eP	101.0	equilibrium pressure (kPa)
cP	101.0	cooling target pressure (kPa)
T_inc	100.0	temperature increment up to eT (K)

## 2.12 Generierung einer kombinatorischen Substanzdatenbank

Um die potentiellen Leitstrukturen der Benzimidazol-Carbonitrile aus dem virtuellen *Screening* weiter zu modifizieren, wurde eine kombinatorische Substanzdatenbank generiert. Zu diesem Zweck wurde ein Cactvs-Skript erstellt, das zunächst Substituenten an aromatischen Ringsystemen von bekannten Kinase-Inhibitoren erkennt sowie abtrennt. Diese Fragmente wurden gesondert in einer Substituentendatenbank abgelegt. Die bekannten Kinase-Inhibitoren stammten von Kinase-Katalogen einiger kommerzieller Anbieter. Fragmente mit einem Molekulargewicht von kleiner 180 g/mol, wurden im Anschluss als Substituenten an die Grundstruktur an den Positionen R1 bis R7 angefügt (Abbildung 2.2). Hierdurch wurden insgesamt 12.356 Strukturen der Klasse der Benzimidazol-Carbonitrile generiert. Die chemische Zugänglichkeit wurde für die Generierung der kombinatorischen Substanzbibliothek nicht berücksichtigt.

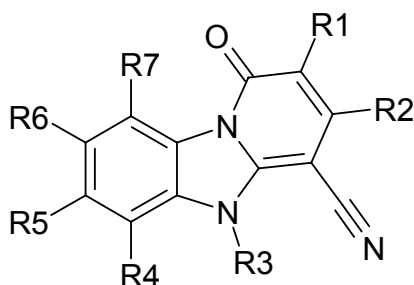


Abbildung 2.2: Substitutionsmöglichkeiten der Benzimidazol-Carbonitrile

Zur Generierung einer kombinatorischen Substanzbibliothek, wurde an der Benzimidazol-Grundstruktur an den Positionen R1 bis R7 substituiert.

Die so erhaltene kombinatorische Substanzdatenbank wurde dann durch molekulares *Docking* bewertet und interessante Liganden selektiert. Für die so erhaltenen Verbindungen wurden Löslichkeitsvorhersagen (siehe Abschnitt 2.13) durchgeführt. Substanzen mit einem guten Löslichkeitsprofil wurden beibehalten und Bindungsenergien mittels der LIE-Methode berechnet (siehe Abschnitt 2.11).

## 2.13 Löslichkeitsvorhersagen

Zur Vorhersage von Löslichkeiten einzelner Verbindungen wurde das thermodynamische Löslichkeitsmodell des Programmpakets Volsurf+ [<http://www.moldiscovery.com>] verwendet. Volsurf+ benutzt ein QSPR-Modell (QSPR = *Quantitative Structure Property Relationship*). Das Modell basiert auf ca. 1.100 chemischen Strukturen. Volsurf+ sagt die thermody-

namische Löslichkeit  $\log S$  in Wasser bei 298 K vorher, wobei  $S$  in mol/l angegeben wird. Der Standardfehler der Vorhersage wird von den Autoren mit etwa  $\pm 0,7$  logarithmischen Einheiten angegeben [<http://www.moldiscovery.com>]. Somit kann die Methode als Filterungsmethode für die virtuelle Substanzdatenbank eingesetzt werden.

## 2.14 *In vitro* Messungen

Die Untersuchungen der mittels Substruktur- und Ähnlichkeitssuche identifizierten Substanzen auf biologische Aktivität, wurden von María Luísa Suárez Fernández (Intervet Innovation GmbH, Schwabenheim) durchgeführt. Dabei wurde die Inhibition der katalytischen Aktivität der Enzyme in Anwesenheit der verschiedenen Substanzen gemessen. Für die Messungen wurde der „*TR-FRET Screening Express Kit*“ von Molecular Devices verwendet [<http://www.Moleculardevices.com>]. Für die Versuche wurden die EtCRK2, EtCYC3a und das RINGO-Protein [FERBY *et al.*, 1999] des Krallenfroschs *Xenopus laevis* in einem *Escherichia coli* Expressionssystem rekombinant exprimiert und aufgereinigt. Das Protein RINGO bindet unspezifisch an verschiedene CDKs und ist wie EtCYC3a ein Aktivator der EtCRK2. Die aktivierte EtCRK2 wurde nach Herstellervorgaben in dem Test eingesetzt. Zudem wurden für einige ausgewählte Substanzen enzymkinetische Messungen durchgeführt.

Für die identifizierten EtCRK2-Inhibitoren der zweiten Generation wurden außerdem die Löslichkeiten in Wasser mittels Nephelometrie bestimmt. Zusätzlich wurden diese Substanzen mittels LC/MS- und NMR-Messungen einer Qualitätskontrolle unterzogen. Diese Messungen wurden von Liane Walther (Nephelometrie), Stefan Derschum und Christian Schorn (LC/MS- und NMR-Messungen) bei Intervet Innovation GmbH, Schwabenheim durchgeführt.



## 3 Ergebnisse

### 3.1 Bioinformatische Sequenzanalyse von CDKs und Zyklinen von *E. tenella*

Von der *E. tenella* Genomsequenz wurde im Frühjahr 2007 ein erster Entwurf (*Draft Assembly*) fertig gestellt, der jedoch noch nicht vollständig annotiert wurde. Die Sequenzdaten wurden am Sanger Institut, im Rahmen eines *E. tenella* Genom-Sequenzierungsprojektes, generiert und sind auf einem öffentlich zugänglichen FTP-Server frei verfügbar hinterlegt [<ftp://ftp.sanger.ac.uk/pub/pathogens/Eimeria/tenella>]. Für diese BLAST-Analysen wurden bereits bekannte CDK- sowie Zyklin-ähnliche Proteinsequenzen von Organismen der *Apicomplexa*, als Abfragesequenzen verwendet.

#### 3.1.1 CDK-homologe Proteine in *E. tenella*

Zu Beginn dieser Arbeit war nur eine CDK von *E. tenella* bekannt (EtCRK2; CRK = *CDK-Related Kinase*). EtCRK2 wurde bereits von einer kooperierenden Arbeitsgruppe kloniert, exprimiert und biochemisch charakterisiert [KINNAIRD *et al.*, 2004]. Sie gilt als ein homologes Protein des humanen CDK2 Proteins (HsCDK2) [KINNAIRD *et al.*, 2004]. Auf der Aminosäuresequenzebene besteht eine Ähnlichkeit von 68% zur HsCDK2 sowie zu der humanen CDK1 (HsCDK1). Zu anderen CDK-ähnlichen Proteinen der *Apicomplexa*, wie beispielsweise der PfPK5 (PfPK5 = *Plasmodium falciparum* Protein Kinase 5) [ROSS-MACDONALD *et al.*, 1994], besteht eine Sequenzähnlichkeit zwischen 75% und 80% [KINNAIRD *et al.*, 2004].

Anders als CDKs von Vertebraten, die über die gesamte Sequenzlänge hoch konserviert vorliegen [MALUMBRES & BARBACID, 2005], variiert die Aminosäureabfolge der CDK Sequenzen der *Apicomplexa*, insbesondere am C-Terminus. Bekannte CDKs der *Apicomplexa* weisen Sequenzidentitäten von weniger als 40% in den 100 C-terminalen Aminosäuren auf. Neben den teilweise großen Sequenzunterschieden am C-Terminus, sind auch die Zyklin-Bindemotive (PSTAIRE-Motiv) in ihrer Sequenzabfolge sehr verschieden. Zudem zeigen diese Proteine oft N-terminale und C-terminale Verlängerungen. Eine Proteinmotivanalyse auf Basis der Motivdatenbanken Pfam und PROSITE [BATEMAN *et al.*, 2004; HULO *et al.*, 2006] ergab, dass weitere für Kinasen typische Sequenzmotive in allen CRK Proteinen der

*Apicomplexa* enthalten sind. Sie besitzen mehrere stark konservierte Regionen im aktiven Zentrum. Hierzu zählen die *Hinge*-Region, der *T-Loop* und ein Glycin-reiches Sequenzmotiv (*G-Loop*) (siehe Abbildung 1.9 und Abbildung 3.2). Hierbei zeigte sich, dass der Grad der Konservierung von Aminosäuren in funktionellen Bereichen der Enzyme weitaus stärker ausgeprägt ist. Eine zentrale Rolle spielt zudem die konservierte Aminosäure D144 (Nummerierung nach der Aminosäuresequenz der EtCRK2), die für die katalytische Aktivität von entscheidender Bedeutung ist. Sie ist Bestandteil der katalytischen Zentrums K32, E50 und D144 (EtCRK2) [BRANDEN & TOOZE, 1999]. Diese Aminosäuren liegen im multiplen *Alignment* der CRK-Proteinsequenzen der *Apicomplexa* an identischen Positionen vor.

Zur Identifizierung von weiteren CRK-Sequenzen in *E. tenella*, wurden Suchen mit dem BLASTX-Programm in der genomischen Sequenzdatenbank von *E. tenella* des Sanger Institutes [<http://www.sanger.ac.uk>] durchgeführt. Für diese BLAST-Analysen wurden die sieben bereits bekannten CRKs von *P. falciparum* - PfPK5, PfPK6, PfMRK, PfCRK1, PfCRK3, PfCRK4 und PfCRK5 [BRACCHI-RICARD *et al.*, 2000; DOERIG *et al.*, 1995; DOERIG *et al.*, 2002; DOERIG, 2004; LI *et al.*, 1996; ROSS-MACDONALD *et al.*, 1994] – als Abfragesequenzen verwendet. Hierbei konnten neben der bereits bekannten EtCRK2 [KINNAIRD *et al.*, 2004], die homolog zu PfPK5 ist, drei unbekannte CRKs von *E. tenella* vorhergesagt werden. Sie wurden in Anlehnung an die Namen der entsprechenden Proteine von *P. falciparum* (PfMRK, PfCRK1 und PfCRK3) EtMRK, EtCRK1 und EtCRK3 genannt. Die Ähnlichkeitsbeziehungen der CDK-ähnlichen Proteine von *P. falciparum*, zu den neu identifizierten Proteinen in *E. tenella*, sind in Tabelle 3.1 und Abbildung 3.1 dargestellt.

Lediglich für die EtMRK konnte eine vollständige Sequenz identifiziert werden, während für EtCRK1 und EtCRK3 nur Teilsequenzen gefunden wurden (siehe Tabelle 3.1 und Abbildung 3.1). Die Datenbanksuchen mit den Abfragesequenzen PfPK6, PfCRK4 und PfCRK5 ergaben keine signifikanten Treffer. Aufgrund der im Vergleich zu anderen CDK-Sequenzen der *Apicomplexa* sehr geringen Sequenzähnlichkeit, sind diese Ergebnisse nicht ungewöhnlich. Zudem ist allgemein bekannt, dass die Kinasen PfPK6, PfCRK4 sowie PfCRK5 ähnliche Identitätswerte zu den CDKs, als auch zu den MAP Kinasen (MAPK = *Mitogen-Activated Protein Kinase*) aufweisen [DOERIG *et al.*, 2002; WARD *et al.*, 2004], weshalb sie auch als Hybridproteine vorhergesagt werden [GEYER *et al.*, 2005]. Diese Tatsache zeigt sich auch in BLASTP Analysen der PfCRK4 und der PfCRK5, die sowohl Sequenzmotive von CDKs, als auch von MAPKs aufweisen [WARD *et al.*, 2004]. Die PfPK6 nimmt hierbei eine Sonderstel-

lung ein, da sie offensichtlich die einzige CDK-ähnliche Sequenz in *P. falciparum* ist, die kein Zyklinbindemotiv besitzt [CHEN *et al.*, 2006], somit könnte sie auch als eine Zyklin-unabhängige CDK angesehen werden, was biochemisch bestätigt wurde [BRACCHIRICARD *et al.*, 2000].

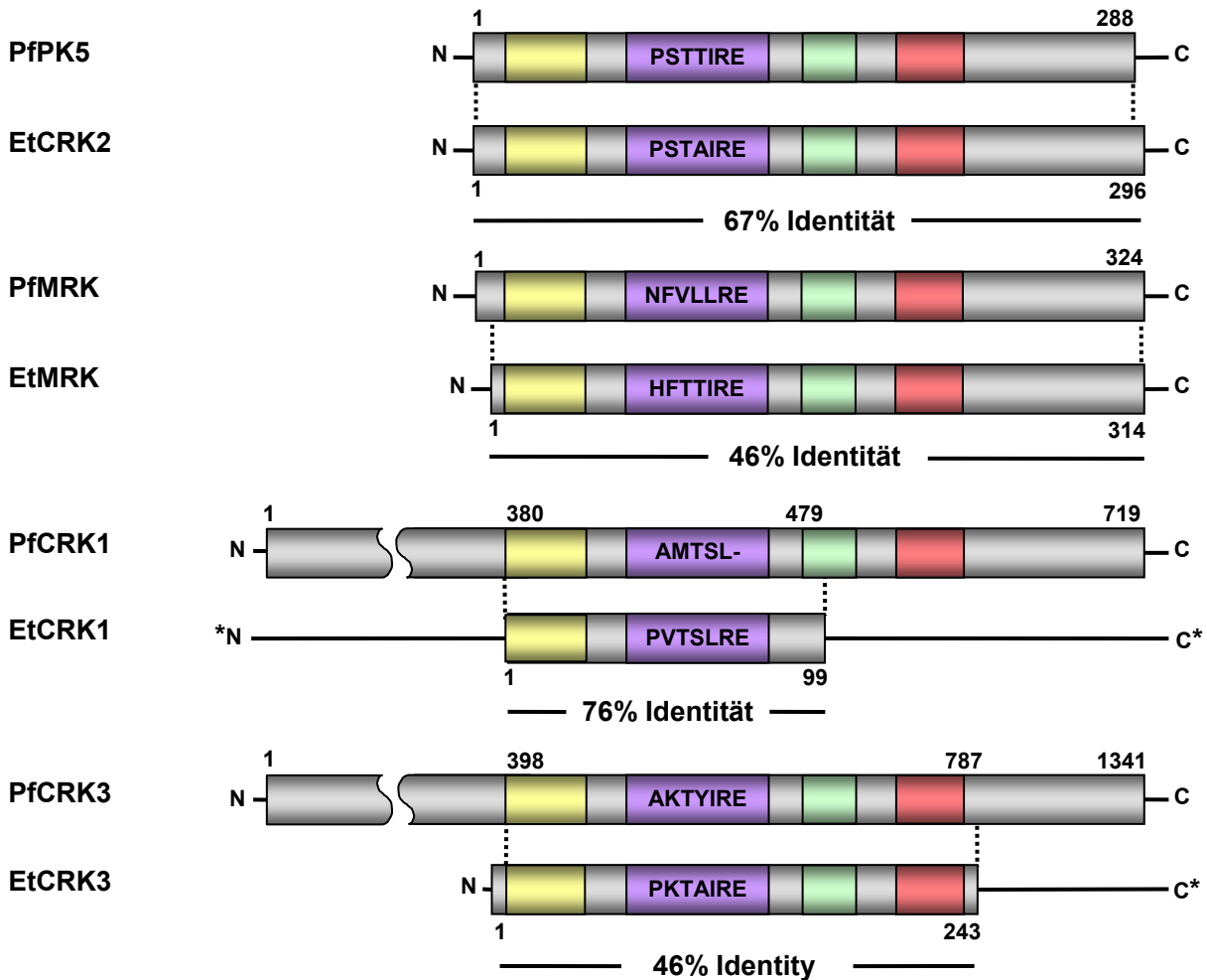


Abbildung 3.1: Schematischer Sequenzvergleich von CDK-ähnlichen Proteinen von *P. falciparum* und *E. tenella*. Neben den bereits bekannten Sequenzen von *P. falciparum* und der EtCRK2 von *E. tenella*, wurden mittels BLAST-Suchen drei neue CDK-ähnliche Sequenzen für *E. tenella* identifiziert. Die Abbildung zeigt die drei vorhergesagten Proteinsequenzen sowie EtCRK2. Die CDK-ähnlichen Sequenzen sind jeweils als Paare zu den homologen Sequenzen in *P. falciparum* abgebildet. Für die EtCRK1 und die EtCRK3 konnten jedoch nur unvollständige Proteinsequenzen vorhergesagt werden. Die unvollständigen Sequenzabschnitte sind mit einem „\*“ markiert. Das charakteristische Zyklin-Bindemotiv zeigt Variationen in den CDK-ähnlichen Proteinsequenzen (violett). Weitere charakteristische Sequenzmotive sind: G-Loop (gelb), Hinge (grün) und Aktivierungs-Loop (rot).

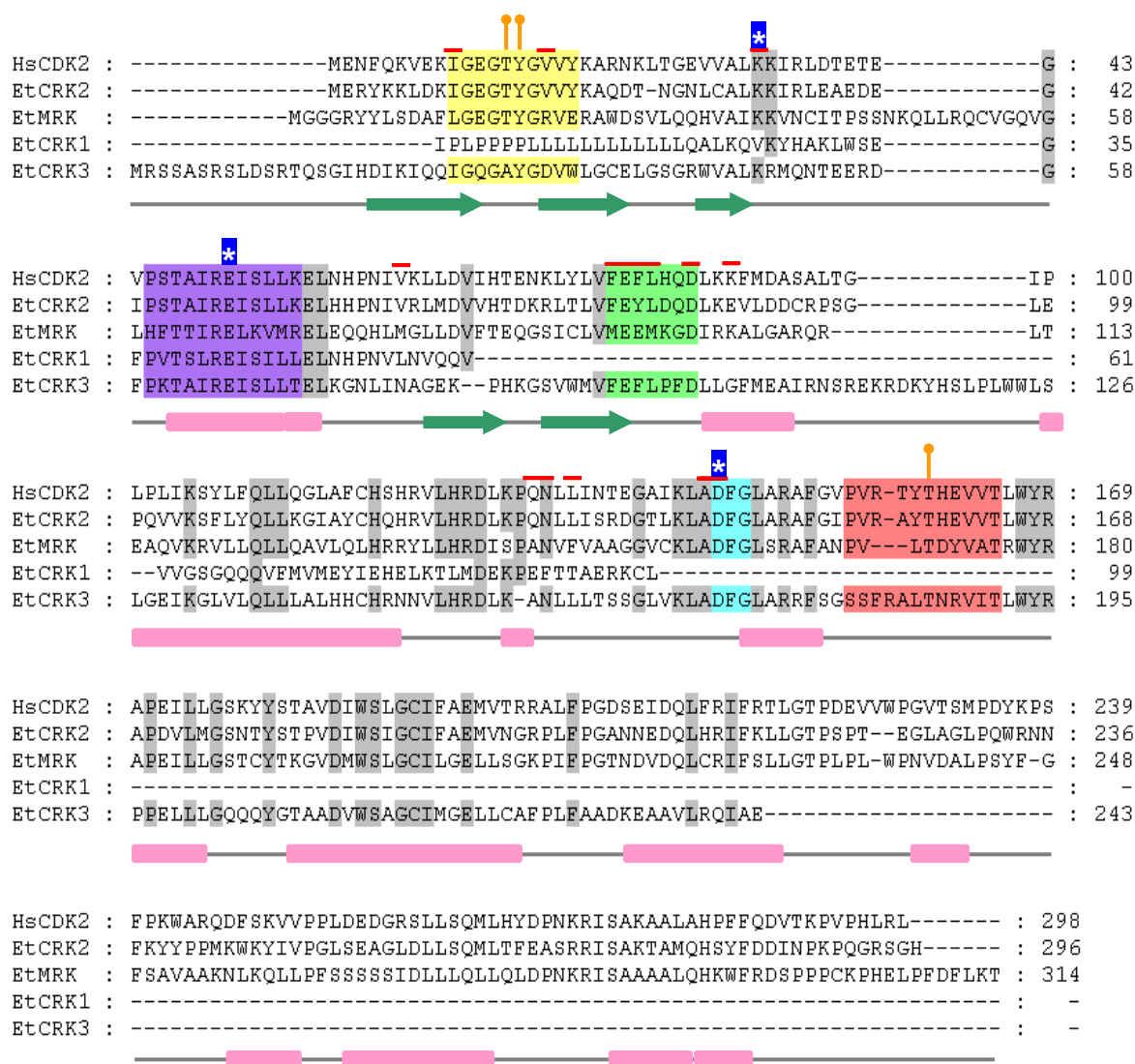


Abbildung 3.2: Multiples Sequenzalignment von HsCDK2 und CDK-Homologen von *E. tenella*.

$\alpha$ -Helices und  $\beta$ -Faltblätter sind unter den Sequenzen als rosa Blöcke, bzw. als grüne Pfeile dargestellt. Zu den charakteristischen Sequenzmotiven zählen PSTAIRE-Motiv (violett), G-Loop (gelb), Hinge-region (grün), DFG-Motiv (blau) und Aktivierungs-Loop (rot). Das Zyklin-Bindemotiv (PSTAIRE-Motiv) variiert in der Sequenzabfolge in CRK-ähnlichen Proteinen der *Apicomplexa*. Die Aminosäuren des katalytische Zentrums K32, E50 und D144 (EtCRK2 Nummerierung) sind durch „\*“ markiert. Aminosäuren die durch „-“ kennzeichnen die ATP-Bindungsstellen. Die Phosphorylierungsstellen T14, Y15 und T160 sind ebenfalls bei den meisten CRK2 ähnlichen Proteinen hoch konserviert und durch „☞“ markiert. Aminosäuren, die konserviert vorliegen, sind grau unterlegt.

Tabelle 3.1: Vergleich der bekannten CDK-ähnlichen Sequenzen von *P. falciparum* und *E. tenella* sowie den Aktivatoren.

<i>Plasmodium falciparum</i>		<i>Eimeria tenella</i>		
Gene / Protein	Aktivator	Protein	Identität zu <i>P. falciparum</i> Protein	Aktivator
PfPK5 [ROSS-MACDONALD <i>et al.</i> , 1994]	cycH, p25, PfCYC1, PfCYC3, cycA, RINGO [LE ROCH <i>et al.</i> , 2000; MERCKX <i>et al.</i> , 2003]	EtCRK2 [KINNAIRD <i>et al.</i> , 2004]	67%	Cyclin A [KINNAIRD <i>et al.</i> , 2004], RINGO, EtCYC3a [SUÁREZ FERNÁNDEZ, unpublizierte Daten]
PfPK6 [BRACCHI-RICARD <i>et al.</i> , 2000]	Cyclin-independent [DOERIG <i>et al.</i> , 2002]			
PfMRK [LI <i>et al.</i> , 1996]	cycH, PfCYC1 [LE ROCH <i>et al.</i> , 2000; WATERS <i>et al.</i> , 2000]	<i>EtMRK</i>	42%	
PfCRK-1 [DOERIG <i>et al.</i> , 1995]		<i>EtCRK1</i>	76%*	
PfCRK-3 [DOERIG <i>et al.</i> , 2002]		<i>EtCRK3</i>	46%*	
Pfcrk-4 [DOERIG <i>et al.</i> , 2002]				
Pfcrk-5 [WARD <i>et al.</i> , 2004]				

Suchen mit dem BLAST-Algorithmus ermöglichten die Identifikation von homologen CDK-Proteinen (EtMRK, EtCRK1 und EtCRK3). EtCRK1 sowie EtCRK3 konnten bisher nur unvollständig identifiziert werden. Für die Teilsequenzen sind die angegebenen Identitäten mit einem „\*“ markiert. Als Aktivatoren der EtCRK2 dienen Cyclin A [KINNAIRD *et al.*, 2004] sowie RINGO und EtCYC3a [SUÁREZ FERNÁNDEZ, unpublizierte Daten, Intervet Innovation GmbH, Schwabenheim].

### 3.1.1.1 EtMRK

EtMRK weist eine große N-terminale Verlängerung von 43 Aminosäuren (bezeichnet im Folgenden als EtMRK\_lang) auf. Viele andere homologe Proteinsequenzen, innerhalb des Stammes der *Apicomplexa*, besitzen einen deutlich kürzeren N-Terminus. Eine ähnliche, jedoch noch längere Extension ist für die TaCRK3 aus *Theileria annulata* bekannt [KINNAIRD *et al.*, 2001]. Im Falle der TaCRK3 beinhaltet diese N-terminale Sequenz ein mutmaßliches Kern-Lokalisierungssignal (NLS = *Nuclear Localization Signal*) [KINNAIRD *et al.*, 2001; MEIER & BLOBEL, 1990]. Für die EtMRK\_lang konnte mit dem Software-Tool predictNLS kein NLS-Konsensusmotiv vorhergesagt werden. Zudem wurden in dieser Verlängerung kei-

ne weiteren, bekannten Sequenz-Motive gefunden, was vermuten lässt, dass ein Sequenzierungsfehler vorliegt. Die Nukleotidsequenzanalyse ergab im Leserahmen +1 ein „ACG“-Triplet an Position 139 der Etmrk. Falls innerhalb des „ACG“ ein Sequenzierungsfehler vorliegt und man stattdessen ein „ATG“ annimmt, erhält man ein Kozak-Konsensus-Motiv. Das vermeintliche Startcodon der EtMRK\_lang-Sequenz ist hingegen nicht in einem Kozak-Motiv enthalten. Aus diesem Grund wurde für alle weiteren Arbeiten die kürzere Nukleotidsequenz angenommen. Die vorhergesagte EtMRK Proteinsequenz resultierte so in einer vergleichbaren Länge und Aminosäureabfolge zu anderen homologen MRK-Sequenzen innerhalb der *Apicomplexa* (siehe Abbildung 3.3). Die EtMRK-Proteinsequenz ist zu 46% identisch mit der verwandten *P. falciparum* PfMRK-Proteinsequenz. Die Analyse des spezifischen Zyklin-Bindemotives ergab die Aminosäureabfolge „HFTTIRE“ für die EtMRK. Dieses Motiv ist für alle MRK-ähnlichen Proteine weitgehend konserviert (siehe Tabelle 3.2). Alle anderen Kinase-Motive wie *G-Loop*, *Hinge-Region*, *T-Loop* und *DFG-Motiv* konnten ebenfalls in der EtMRK identifiziert werden.

Aus der Literatur ist, im Gegensatz zu CDK-Proteinen von Säugern, nur wenig über die Funktion homologer MRK Proteine der *Apicomplexa* bekannt [CANDURI *et al.*, 2007; CHEN *et al.*, 2006; KINNAIRD *et al.*, 2001]. Beispielsweise scheint PfMRK eine andere Regulationsfunktion zu besitzen, als die ähnlichen Säugerproteine [CHEN *et al.*, 2006; KALDIS, 1999]. Die PfMRK besitzt, im Vergleich zu CDK-Proteinen von Säugern, die höchste Identität (38%) zum humanen CDK7 Protein (HsCDK7; [gi|4502743]). HsCDK7 besitzt eine duale Funktion [FISHER, 2005]. Einerseits ist HsCDK7 in der Transkriptionsregulation – durch Phosphorylierung der C-terminalen Domäne der RNA-Polymerase II – involviert. Andererseits agiert HsCDK7 als CDK-Aktivierungskinese (CAK) [FISHER, 2005]. CAK Proteine phosphorylieren spezifische Threonin-Reste im Aktivierungs-*Loop* von Kinasen, was zu einer vollständigen Aktivierung der Kinasen führt [SCLAFANI, 1996]. Die duale Funktion der HsCDK7 wird wiederum durch verschiedene Effektoren reguliert. Das humane MAT1 Protein („menage à trois“) im Komplex mit dem humanen Zyklin H überführt die HsCDK7 in den „Transkriptionsregulations-Modus“, während das Zyklin H allein die HsCDK7 in den „CAK-Modus“ überführt [CHEN *et al.*, 2006]. Für das PfMRK-Protein konnte bisher keine ähnliche duale Funktion wie bei der humanen CDK7 biochemisch nachgewiesen werden. Chen *et al.* konnten zwar zeigen, dass ein homologes MAT1 Protein (PfMAT1) in *P. falciparum* existiert, welches als trimerer Komplex PfMRK/PfCYC1/PfMAT1 die Phosphorylierung der C-terminalen Domäne der RNA-Polymerase II vollzieht. Dennoch wurde die Funktion einer

CAK für die PfMRK, mit PfPK5 als ein mögliches Substrat, bisher biochemisch nicht bestätigt [CHEN *et al.*, 2006].

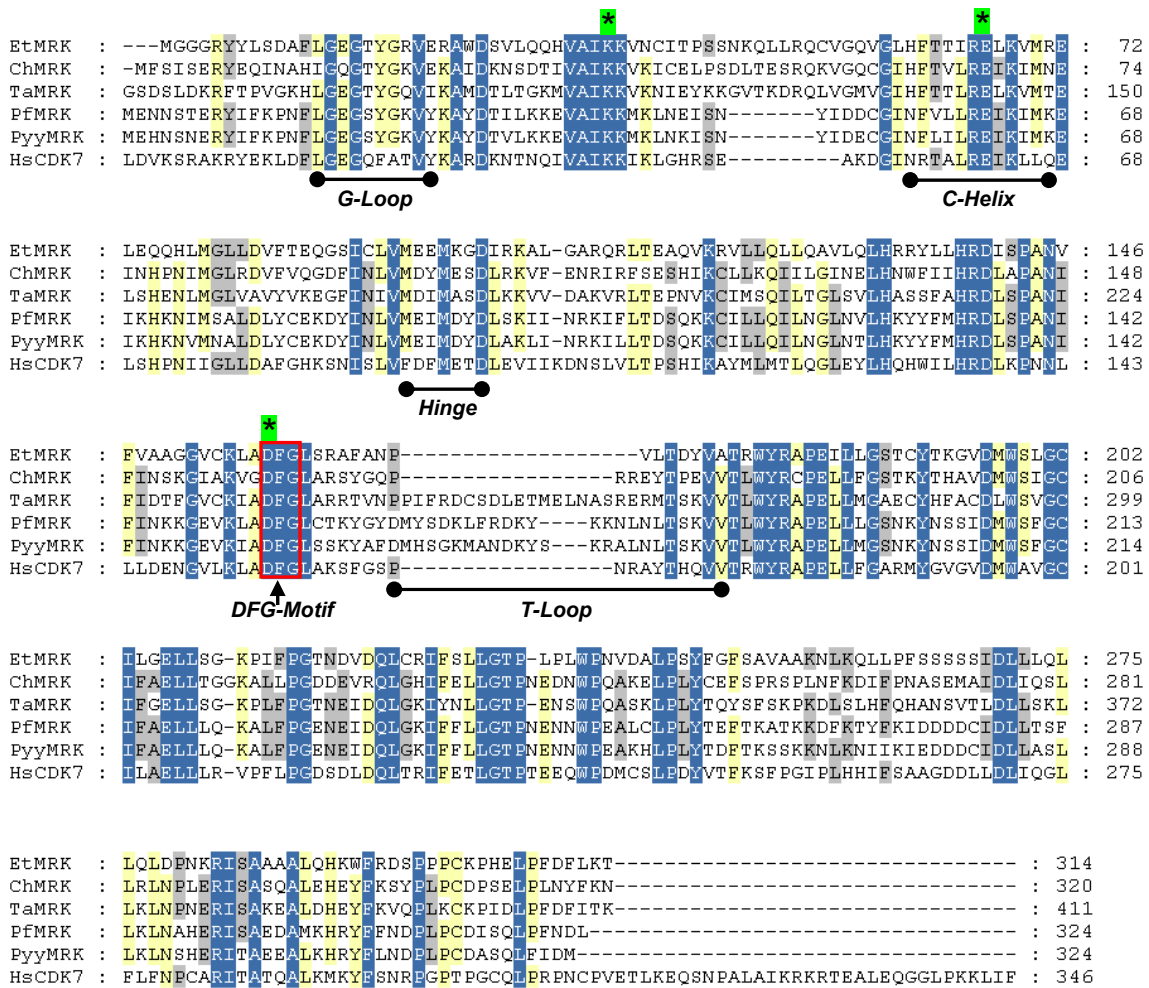


Abbildung 3.3: Sequenzalignment von MRK-ähnlichen Proteinen der Apicomplexa sowie HsCDK7.

Aligniert wurden die MRK Proteinsequenzen von *Cryptosporidium hominis* (ChMRK), *Theileria annulata* (TaMRK), *Plasmodium falciparum* (PfMRK), *Plasmodium yoelii yoelii* (PyyMRK) sowie die humane CDK7 Proteinsequenz (HsCDK7). Die wichtigen Sequenzmotive sind im *Alignment* markiert. Aminosäuren, die in allen Proteinsequenzen konserviert vorliegen, sind blau unterlegt. Gelb unterlegte Aminosäuren sind in fünf Sequenzen, sowie grau unterlegte Aminosäuren sind in vier Sequenzen konserviert.

Um zu untersuchen, ob das EtMRK Protein ebenfalls an der Transkriptionsregulation der RNA-Polymerase II beteiligt ist, wurde die *E. tenella* Genomsequenz nach einer zu PfMAT1 bzw. einer zu HsMAT1 homologen Sequenz durchsucht. Eine BLASTX-Suche mit PfMAT1 in der *E. tenella* Genomsequenz identifizierte eine homologe EtMAT1. Dies lässt vermuten, dass das Protein EtMRK ebenfalls an der Transkriptionsregulation der RNA Polymerase II beteiligt ist. Allerdings widerspricht das Fehlen eines NLS Motivs dieser Hypothese, da Proteine, die an der Transkriptionsregulation beteiligt sind, mit Hilfe des NLS Signals in den Zellkern

transloziert werden müssen. Die humane CDK7 Sequenz weist im Gegensatz zur EtMRK Sequenz ein NLS-Signal auf.

Die Frage, ob EtMRK ebenfalls eine CAK Funktion besitzt, kann nicht allein durch Sequenzvergleiche vorhergesagt werden und sollte daher biochemisch untersucht werden. Jedoch gibt es einige Hinweise die gegen eine solche Funktion sprechen: (I) Es konnte kein CAK-Bindemotiv in den potentiellen Substraten EtCRK2 [KINNAIRD *et al.*, 2004] und EtCRK3 identifiziert werden. (II) Die Phosphorylierung des Aktivierungs-*Loops* der EtCRK2 und der PfPK5 kann auch durch Autophosphorylierung erfolgen. Für die PfPK5 wurde bislang noch nicht bestätigt, ob diese Autophosphorylierung die einzige Aktivierungsmethode ist [CHEN *et al.*, 2006; KINNAIRD *et al.*, 2004].

Tabelle 3.2: MRK-ähnliche Proteine der Apicomplexa sowie deren Zyklin-Bindedomänen.

Organismus	Name	NCBI Sequenz-ID	Zyklin- Bindedomäne	Identität zu PfMRK [%]
<i>Plasmodium falciparum</i>	PfMRK	XP_001347426	NFVLLRE	100
<i>Plasmodium berghei</i>	PbMRK	XP_766302	NFLILRE	91
<i>Plasmodium knowlesi</i>	PkMRK	XP_002258785	NFVLLRE	88
<i>Plasmodium vivax</i>	PvMRK	XP_001614455	NFVLLRE	87
<i>Plasmodium yoelii yoelii</i>	PyyMRK	XP_725555	NFLILRE	85
<i>Cryptosporidium hominis</i>	ChMRK	XP_665899	HFTVLRE	50
<i>Cryptosporidium muris</i>	CmMRK	XP_002142071	HFTVLRE	49
<i>Cryptosporidium parvum</i>	CpMRK	XP_001388367	HFTVLRE	51
<i>Babesia bovis</i>	BbMRK	XP_001610945	HFTALRE	47
<i>Theileria annulata</i>	TaMRK	XP_678603	HFTTLRE	49
<i>Theileria parva</i>	TpMRK	XP_766302	HFTTLRE	48
<i>Toxoplasma gondii</i>	TgMRK	XP_002365709	HFTTVRE	39
<i>Eimeria tenella</i>	EtMRK	-	HFTTIRE	43



### 3.1.1.2 EtCRK1

Wie bereits in Abschnitt 3.1.1 beschrieben, konnte die EtCRK1 bislang nur unvollständig vorhergesagt werden. Für beide Termini der EtCRK1 Aminosäuresequenzen lagen keine Daten in den Datenbanken vor. Die EtCRK1 Proteinteilsequenz besteht aus 99 Aminosäuren. Sie beinhaltet das Zyklinbinde-Motiv „PVTSLRE“, welches innerhalb der homologen CRK1-Proteinsequenzen der *Apicomplexa* weitgehend konserviert vorliegt (siehe Tabelle 3.3). Andere spezifische Kinase-Sequenzmotive konnten aufgrund der unvollständigen Sequenz nicht identifiziert werden. Die bioinformatisch identifizierte EtCRK1-Teilsequenz ist zu 76% identisch zur bekannten PfCRK1 Protein-Sequenz. Die Konservierung am N- und am C-Terminus der CRK1-homologen Proteine ist im Vergleich zum mittleren Sequenzabschnitt deutlich geringer, wodurch die Identifizierung der terminalen Sequenzabschnitte erschwert ist.

Tabelle 3.3: : CRK1-ähnliche Proteine der *Apicomplexa* sowie deren Zyklin-Bindedomänen.

Organismus	Name	NCBI Sequenz-ID	Zyklin- Bindedomäne	Identität zu PfCRK1 [%]
<i>Plasmodium falciparum</i>	PfCRK1	XP_001351486	AMTSLRE	100
<i>Plasmodium chabaudi chabaudi</i>	PccCRK1	XP_742956	AITSLRE	76
<i>Plasmodium knowlesi</i>	PkCRK1	XP_002258234	AITSLRE	81
<i>Plasmodium vivax</i>	PvCRK1	XP_001615088	AITSLRE	77
<i>Plasmodium yoelii yoelii</i>	PyyCRK1	XP_731313	AITSLRE	77
<i>Babesia bovis</i>	BbCRK1	XP_001609961	PVSYLRE	49
<i>Theileria annulata</i>	TaCRK1	XP_953202	PLTYLRE	50
<i>Theileria parva</i>	TpCRK1	XP_764186	PLTNLRE	54
<i>Toxoplasma gondii</i>	TgCRK1	XP_002370291	PVTSLRE	39
<i>Eimeria tenella</i>	EtCRK1	-	PVTSLRE	76

### 3.1.1.3 EtCRK3

EtCRK3 ist der HsCDK1 (36% Sequenzidentität) sowie der PfCRK3 ähnlich (46% Sequenzidentität). Diese Aminosäuresequenz konnte ebenfalls nur unvollständig vorhergesagt werden. Während für den N-Terminus ein Startcodon gefunden wurde, konnte für den C-Terminus kein Startcodon identifiziert werden. Die EtCRK3 Proteinsequenz besteht aus 243 Aminosäuren. Das Zyklin-Bindemotiv ist für alle CRK3-ähnlichen Proteine weitgehend konserviert (siehe Tabelle 3.4). Die verbleibenden Kinase-spezifischen Sequenzmotive konnten ebenfalls in der EtCRK3 Proteinsequenz identifiziert werden. Einige der homologen CRK3 Sequenzen anderer *Apicomplexa* besitzen eine Verlängerung des N-Terminus, wie er beispielsweise bei der PfCRK3 Proteinsequenz vorliegt. Die N-terminalen Verlängerungen weisen eine geringe Sequenzidentität zueinander auf. Diese N-terminale Verlängerung konnte jedoch nicht für die EtCRK3 Proteinsequenz identifiziert werden.

Ein weiterer signifikanter Unterschied der EtCRK3 und PfCRK3 findet sich im mittleren Bereich der Proteinsequenz, zwischen den Domänen der C-Helix und der *Hinge*-Region. Die PfCRK3 Proteinsequenz weist an dieser Stelle eine sehr große Insertion von etwa 200 Aminosäuren (siehe Abbildung 3.4) auf. Ähnliche Insertionen werden auch bei CpCRK3 und TpCRK3 gefunden, jedoch nicht bei HsCDK1 (Abbildung 3.4). Aus diesem Grund scheint die Sequenzkomposition von EtCRK3 ähnlicher zur HsCDK1 als zu der PfCRK3 zu sein.

Die Aminosäuren Thr14 und Tyr15 der HsCDK1 sind an der Inaktivierung des CDK1/Zyklin Komplexes beteiligt und charakteristisch für CDK1 ähnliche Proteine. Die Inaktivierung erfolgt durch Phosphorylierung. Während Tyr15 bei allen CRK3-ähnlichen Sequenzen der *Apicomplexa* konserviert vorliegt, ist Thr14 in der EtCRK3 Sequenz durch ein Alanin ausgetauscht (siehe Abbildung 3.4). Dies lässt vermuten, dass die Inaktivierung dieses Proteins möglicherweise auf eine andere Art reguliert wird. Neben der Inaktivierungsregulation ist auch die genaue Funktion der CpCRK3 sowie aller CRK3 homologen Proteine bislang nicht bekannt [CANDURI *et al.*, 2007].

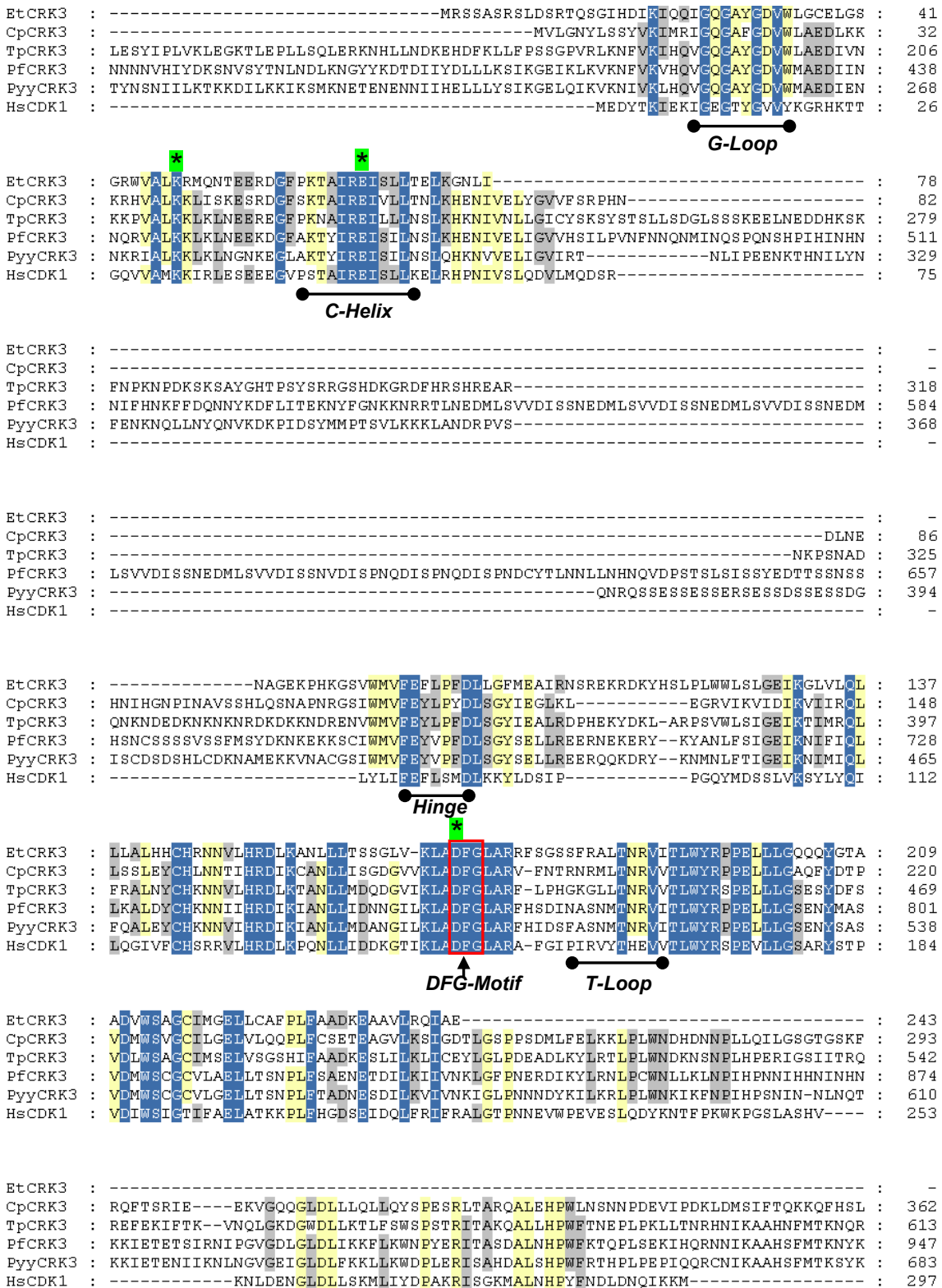


Abbildung 3.4: Sequenzalignment von CRK3-ähnlichen Proteinen der Apicomplexa sowie HsCDK1.

Im Alignment sind wichtige Sequenzmotive markiert. Aminosäuren, die in allen Proteinsequenzen konserviert vorliegen, sind blau unterlegt. Gelb unterlegte Aminosäuren sind in fünf Sequenzen, grau unterlegte Aminosäuren in vier Sequenzen konserviert.

Tabelle 3.4: CRK3-ähnliche Proteine der Apicomplexa sowie deren Zyklin-Bindedomänen.

Organismus	Name	NCBI Sequenz-ID	Zyklin- Bindedomäne	Identität zu PfCRK3 [%]
<i>Plasmodium falciparum</i>	PfCRK3	XP_001351461	AKTYIRE	100
<i>Plasmodium berghei</i>	PbCRK3	XP_677288	AKTYIRE	73
<i>Plasmodium chabaudi</i> <i>chabaudi</i>	PccCRK3	XP_745291	AKTYIRE	73
<i>Plasmodium knowlesi</i>	PkCRK3	XP_002258203	AKTYIRE	50
<i>Plasmodium vivax</i>	PvCRK3	XP_001615034	AKTYIRE	69
<i>Plasmodium yoelii yoelii</i>	PyyCRK3	XP_724274	AKTYIRE	39
<i>Cryptosporidium hominis</i>	ChCRK3	XP_668166	SKTAIRE	45
<i>Cryptosporidium muris</i>	CmCRK3	XP_002140712	SKTAIRE	44
<i>Cryptosporidium parvum</i>	CpCRK3	XP_001388312	SKTAIRE	44
<i>Babesia bovis</i>	BbCRK3	XP_001608847	PKSAIRE	46
<i>Theileria annulata</i>	TaCRK3	XP_954840	PKNAIRE	50
<i>Theileria parva</i>	TpCRK3	XP_763158	PKNAIRE	50
<i>Toxoplasma gondii</i>	TgCRK3	XP_002367260	PRTAIRE	45
<i>Eimeria tenella</i>	EtCRK3	-	PKTAIRE	46

### 3.1.2 Zykline von *E. tenella*

Die Suche nach Zyklinen mit Ähnlichkeiten zu bereits bekannten Zyklinen aus *P. falciparum*, wurden mittels BLAST-Suchen in einer Genom-Datenbank von *E. tenella* durchgeführt. Es konnten vier Sequenzen identifiziert, sowie die entsprechenden Proteinsequenzen (EtCYC1, EtCYC3A, EtCYC3B und EtCYC4; Abbildung 3.5 sowie Abbildung 3.6) vorhergesagt werden. Diese Proteinsequenzen besitzen Ähnlichkeiten zu den Zyklinen PfCYC1, PfCYC3 und PfCYC4 von *P. falciparum* (siehe Abbildung 3.5 und Abbildung 3.6). In einem Aktivierungs-Assay mit EtCYC3a sowie EtCRK2 konnte María L. Suárez Fernández (Intervet Innovation

GmbH, Schwabenheim) zeigen, dass EtCYC3a ein geeigneter Aktivator der EtCRK2 ist. EtCYC3a ist zu 22% zu PfCYC3 und zu 17% zu HsCYCA identisch.

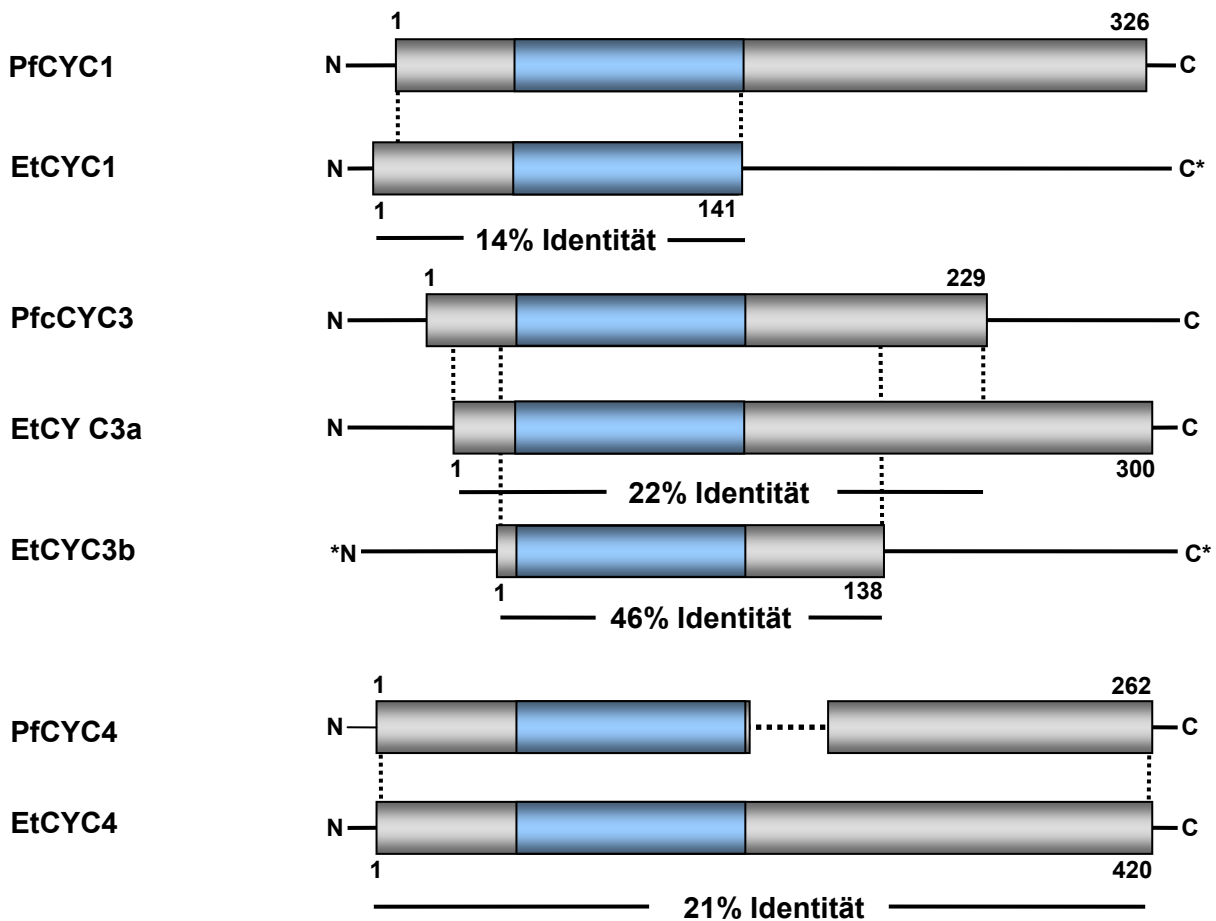


Abbildung 3.5: Schematischer Vergleich von Zyklin-ähnlichen Proteinen von *P. falciparum* und *E. tenella*.

Mit den bekannten Zyklinen von *P. falciparum*, konnten mittels BLAST-Suchen vier Zyklin-ähnliche Sequenzen von *E. tenella* identifiziert werden. Die gefundenen Zyklin-ähnlichen Sequenzen sind jeweils als Paare zu ihren homologen Sequenzen in *P. falciparum* dargestellt. Für EtCYC1 und EtCYC3b (mit einem „\*“ markiert) konnten jedoch nur unvollständige Proteinsequenzen hergeleitet werden. Das charakteristische Zyklin-Box Motiv ist blau markiert.

Die Suche mit der Abfragesequenz PfCYC1 von *P. falciparum* ergab keine ähnlichen Sequenzen in *E. tenella*. Aufgrund der geringen Sequenzidentität der Zyklone der einzelnen Klassen zueinander, wurde die Suche mit der zu PfCYC1 homologen Sequenz aus *T. gondii* (Zyklin H) durchgeführt. Da die Organismen *T. gondii* und *E. tenella* beide der Klasse der *Coccidia* angehören, kann mit Konservierung der Sequenzen gerechnet werden. Aus dieser Suche resultierte eine, in der Literatur bislang nicht beschriebene, Zyklin-ähnliche Sequenz von *E. tenella* (siehe Abbildung 3.5; EtCYC1). Der korrekte Leserahmen wurde mittels paar-

weisem *Alignment* bestimmt und die resultierende Nukleotidsequenz in eine Proteinsequenz translatiert. Dazu wurde der Algorithmus *transeq* verwendet [RICE *et al.*, 2000]. Die translatierte Protein-Sequenz besitzt 115 Aminosäuren und ist zu 14% mit der Proteinsequenz PfCYC1 aus *P. falciparum* identisch (siehe Abbildung 3.5). Eine *InterproScan*-Analyse bestätigt das Vorliegen einer Zyklin-ähnlichen Domäne. Die translatierte Proteinsequenz EtCYC1 ist jedoch sowohl am N-Terminus, als auch am C-Terminus der Aminosäuresequenz unvollständig.

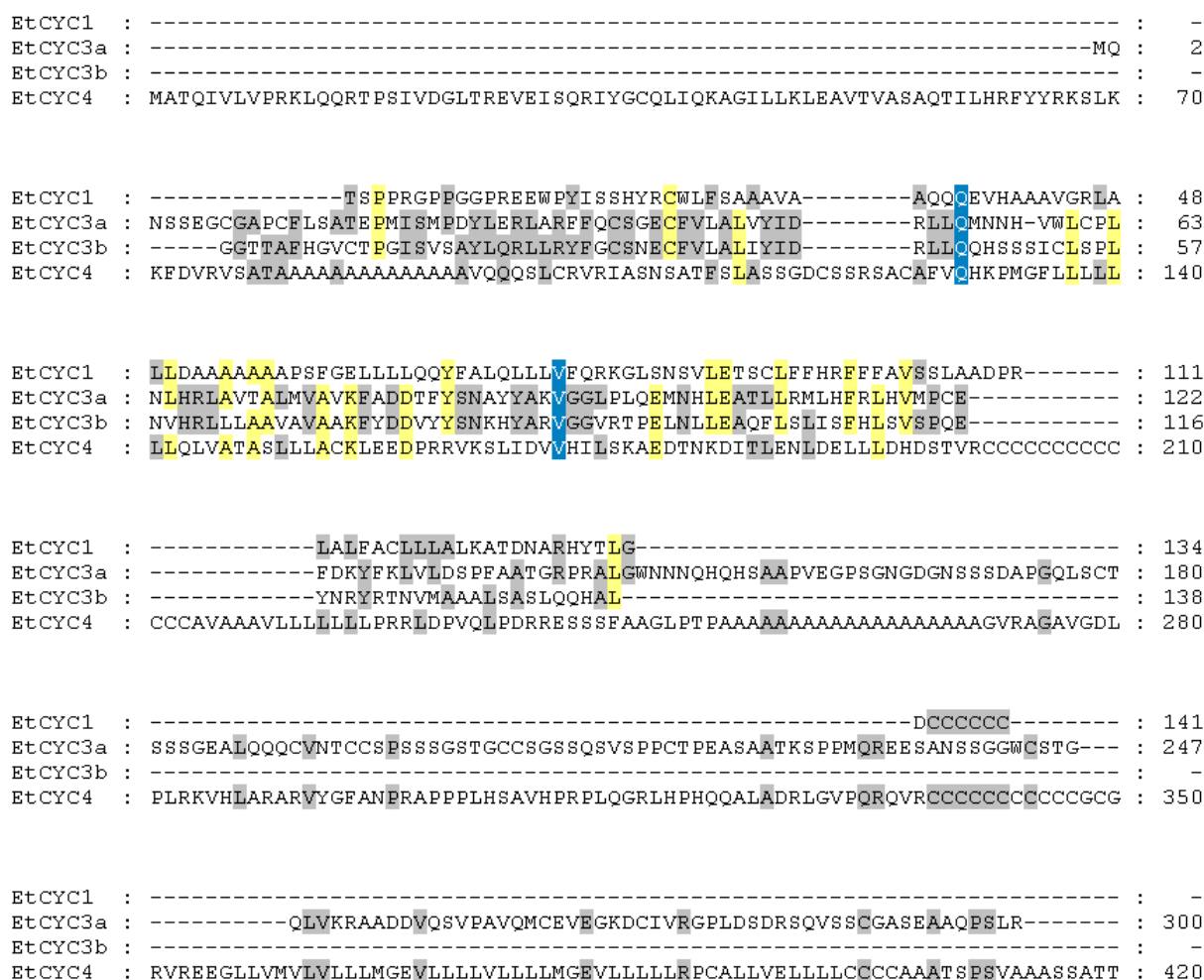


Abbildung 3.6: Sequenzalignment von Zyklin-ähnlichen Proteinen von *E. tenella*.

BLAST-Suchen ergaben vier Zyklin-ähnliche Sequenzen in *E. tenella*. Hier dargestellt sind die vorhergesagten Proteinsequenzen. Für die EtCYC1 und die EtCYC3b konnten jedoch nur unvollständige Sequenzen identifiziert werden. Blau unterlegte Aminosäuren sind in vier Sequenzen, gelb unterlegte Aminosäuren in drei Sequenzen sowie grau unterlegte Aminosäuren in zwei Sequenzen konserviert.

Durch Sequenzsuche mit der Abfragesequenz PfCYC2, konnte keine Sequenz in *E. tenella* ermittelt werden, jedoch identifizierte die Sequenz PfCYC3 zwei ähnliche Sequenzen. Die

Sequenzsuche mit der Abfragesequenz PfCYC4 ergab eine neue Sequenz in *E. tenella*. Für die mittels der Abfragesequenzen PfCYC3 und PfCYC4 gefundenen Sequenzen (Etcyc3a, Etcyc3b und Etcyc4), konnte durch einen direkten Sequenzvergleich der korrekte Leserahmen identifiziert werden. Für die Sequenzen von Etcyc3a und Etcyc4 konnten keine Introns in der genomischen Sequenz identifiziert werden, was jedoch das Vorhandensein von Introns nicht vollständig ausschließt. Abbildung 3.6 zeigt ein multiples *Alignment* von Zyklin-ähnlichen Proteinen von *E. tenella*.

## 3.2 Optimierung und Anreicherung von EtCRK2-Inhibitoren

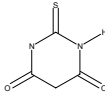
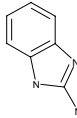
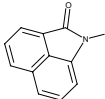

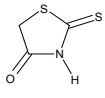
In der Arbeit von Beyer [BEYER, 2007] wurden in einem initialen, virtuellen *Screening* an EtCRK2 bereits einige *Hit-Compounds* identifiziert. Diese wurden anschließend durch *in vitro* Testung bestätigt. Ziel dieser Arbeit war es, die identifizierten *Hits* zu *Hit-Cluster* anzureichern, sowie die Potenz dieser *Compounds* durch gezielte Optimierung zu erhöhen.

### 3.2.1 Substruktur- und Ähnlichkeitssuchen

Zur Anreicherung potentieller Inhibitoren der EtCRK2 wurden Substruktur- und Ähnlichkeitssuchen durchgeführt. Als Ausgangssubstanzen dienten elf im virtuellen *Screening* gefundene Inhibitoren, die einen IC<sub>50</sub>-Wert von kleiner 100 µM im Enzym-*Assay* zeigten. Für die Struktursuchen wurden zunächst geeignete Substrukturen ausgewählt (siehe Tabelle 3.5). Die Auswahl fand anhand der Grundstruktur (*Scaffold*) statt, die im molekularen *Docking* eine charakteristische Wasserstoffbrücken-Interaktion mit der *Hinge*-Region der EtCRK2 zeigte (*Warhead*) (siehe Abschnitt 1.5.3). Insgesamt konnten so 44.351 Substanzen aus der Intervet Chemical Database (ICD) selektiert werden, welche die gewünschten Substrukturen zeigten.

Die Ähnlichkeitssuchen wurden mittels eines PipelinePilot *Workflows*, wie in Abschnitt 2.6 beschrieben, durchgeführt. Es konnten 2.331 ähnliche Verbindungen mit einem Tanimoto-Ähnlichkeitskoeffizienten von mindestens 70% zur Ausgangsstruktur identifiziert werden. Die Nutzung einer Ähnlichkeitssuche hat gegenüber der Substruktursuche den Vorteil, dass Substanzen gefunden werden, die von ihrer Substruktur zur Templatstruktur abweichen können, jedoch in ihren chemischen Eigenschaften ähnlich sind. Beide Arten von Struktursuchen können somit wertvolle Zusatzinformationen über die Struktur-Aktivitäts-Beziehungen liefern.

Tabelle 3.5: Substruktursuchen zur Anreicherung von EtCRK2-Inhibitoren.

Strukturklasse	Substruktur	Anzahl gefundener Strukturen in der ICD
Barbiturate		10.492
Benzimidazole		10.616
Naphtolactame		394
Oxindole		17.513
Rhodanine		3.084

Mit den hier gezeigten Strukturklassen wurden Substruktursuchen in der Intervet Chemical Database (ICD) durchgeführt. Die Auswahl geeigneter Strukturklassen fand anhand der Grundstruktur statt, die mit der *Hinge*-Region über Wasserstoffbrücken interagiert.

### 3.2.2 Molekulares *Docking* zur Auswahl virtueller *Hits*

Die mittels Substruktur- bzw. Ähnlichkeitssuchen gefundenen Verbindungen wurden in Struktur-*Cluster* zusammengefasst. Die Anzahl dieser Strukturen war jedoch zu groß, um diese Substanzen im *in vitro* Test zu untersuchen. Daher wurden diese Struktur-*Cluster* mittels molekularem *Docking* weiter fokussiert. Dafür mussten zunächst die wahrscheinlichsten Bindungsmodi der zugrunde liegenden *Scaffolds* ermittelt werden. Aus jeweils zehn möglichen Posen einer Struktur wurde ein gemeinsamer Bindungsmodus für einen *Scaffold* ermittelt.

Da fast alle bekannten Kinase-Inhibitoren (ähnlich dem ATP) über mindestens zwei Wasserstoffbrückenbindungen mit der *Hinge*-Region (Glu80 und Leu82, EtCRK2-Nummerierung) des Proteins interagieren (siehe Abbildung 3.7), wurde ein Interaktionsfilter benutzt, um Posen mit diesem Bindemodus zu identifizieren [BEYER, 2007]. In dem hierfür verwendeten Cactvs-Skript [BEYER, 2007] wurde dazu eine Kugel mit einem Radius von 3 Å um die Donor- und Akzeptoratome der *Hinge*-Region angenommen. Befindet sich ein passendes Gegen-



stück (Akzeptor- oder Donoratom) einer Ligandenpose im Radius dieser Kugel, wurde die Liganden-Pose beibehalten.

Von den Posen, die einen wahrscheinlichen Bindungsmodus repräsentieren, wurde mittels einer Zweistufen-Bewertung [YOON & WELSH, 2004] eine Rangfolge erstellt. Anhand dieser Rangfolge, wurde die Strukturauswahl fokussiert. Die optimale Zweistufen-Bewertung wurde bereits von Beyer beschrieben [BEYER, 2007]. Dabei kamen zwei verschiedene Bewertungsfunktionen zum Einsatz. In einer ersten Bewertung (*Jain-Scoringfunktion*, [JAIN, 1996]) wurde zunächst die Pose einer Struktur ausgewählt, die der wahrscheinlichsten bioaktiven Konformation entspricht. In einer zweiten Bewertung (*G-Score*, [JONES *et al.*, 1997]) wurden die verbleibenden Posen verschiedener Strukturen miteinander verglichen. Hierbei wurde die endgültige Rangfolge ermittelt.

Um Wirkstoffe zu finden, die selektiv die EtCRK2 inhibieren, sollten diese die negativ geladene Aminosäure Glu88 adressieren. An der analogen Position besitzt die CDK3 des Huhns die positiv geladene Aminosäure Lys89 (siehe Abbildung 3.7). Liganden mit einer funktionellen Gruppe, die diese Ladungsumkehr adressieren, sollten dadurch selektiv an die EtCRK2 binden. Für die Selektion entsprechender Strukturen wurde wiederum ein Interaktionsfilter verwendet [BEYER, 2007].

Durch das *Docking*, die 2-Stufen-Bewertung und die Interaktionsfilter konnten aus den durch Struktursuchen resultierenden Substanzen (insgesamt 46.682), 6.589 Einzelsubstanzen selektiert werden (siehe Tabelle 3.6). Von diesen wurden die 200 Liganden mit der besten Bewertung jeder einzelnen Strukturklasse betrachtet, um geeignete Substanzen auszuwählen. Die so gefundenen Strukturen wurden in einer Datei gespeichert. Durch diese Vorgehensweise konnten insgesamt 505 virtuelle *Hits* der zweiten Generation selektiert werden. Da diese Anzahl an *Compounds* für biochemische Experimente zu groß erschien, wurden nur Substanzen aus den Strukturklassen der Benzimidazole, Naphtolactame und Rhodanine für biochemische Experimente bestellt, da die initialen *Hit-Compounds* dieser Klassen die besten Aktivitätswerte zeigten. Es wurden somit 289 Substanzen zur *in vitro* Testung bestellt.

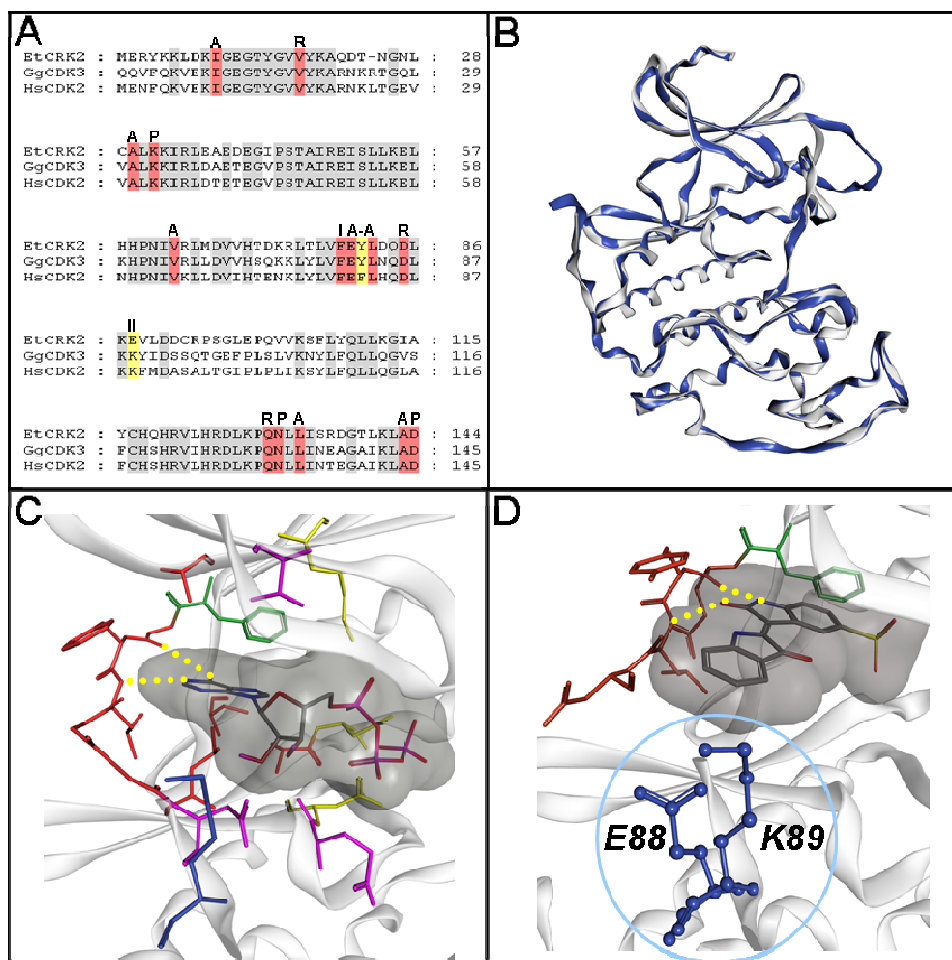


Abbildung 3.7: Primär- und Tertiärstruktur der EtCRK2 und deren ATP-Bindetasche.

A) Sequenzalignment zwischen HsCDK2 und EtCRK2 sowie der CDK3 des Huhns (GgCDK3). Dieses *Alignment* diente als Basis für die Erstellung des Homologiemodells. Grau hinterlegte Aminosäuren signalisieren 100%ige Übereinstimmung. Rot (konservierte) und gelb (Aminosäureaustausch) eingefärbte Aminosäuren der ATP-Bindetasche. Aminosäuren, die mit Buchstaben gekennzeichnet sind, gehören zu der Adeninbinderegion (A), der Ribosebinderegion (R) und der Phosphatbinderegion (P). Aminosäuren, die zu den beiden Selektivitätstaschen gehören, sind mit römischen Ziffern I bzw. II gekennzeichnet.

B) Homologiemodell der EtCRK2 von *E. tenella* (blau) in Ribbon-Darstellung [BEYER, 2007]. Zum Vergleich wurde die humane Templatstruktur von HsCDK2 (PDB-ID: 1OIR; grau) mit der EtCRK2-Modellstruktur überlagert (blau).

C) ATP-Bindetasche der HsCDK2 mit co-kristallisiertem ATP. ATP bindet über zwei Wasserstoffbrücken an die *Hinge*-Region (gelb-gepunktete Linien). Aminosäuren die zur Adenin-Binderegion (rot), Ribose-Binderegion (magenta), Phosphat-Binderegion (gelb) oder den Selektivitätstaschen (blau) gehören, sind farbig markiert.

D) Wichtiger Aminosäureaustausch im Bereich der ATP-Bindetasche zwischen der EtCRK2 und der GgCDK3 Tertiärstruktur. Der Ladungsumkehr von der negativ geladenen Aminosäure Glu88 (EtCRK2) zur positiv geladenen Aminosäure Lys89 (GgCDK3), könnte adressiert werden, um für die EtCRK2 selektive Inhibitoren zu erhalten. In der ATP-Bindetasche befindet sich Indirubin-5-Sulfonat [DAVIES *et al.*, 2001].

Tabelle 3.6: Screening nach potentiellen EtCRK2-Inhibitoren.

Selektionsschritte der <i>Hit</i> -Anreicherung	Anzahl der Verbindungen
Struktursuchen	46.682
<i>Docking</i>	46.069
Nachprozessierung	6.589
<i>Compound</i> -Auswahl	505
Bestellte Substanzen	289
Biochemische Tests	159
Aktive Verbindungen	6

Hier sind die einzelnen Selektionsschritte der *Hit*-Anreicherung dargestellt.

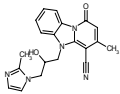
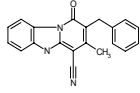
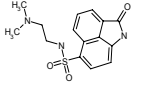

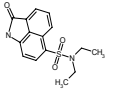
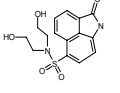
### 3.2.3 *In vitro* Testung der ausgewählten Substanzen

Von den bestellten Substanzen (insgesamt 289) der Klassen der Benzimidazole, Naphtolactame und Rhodanine konnten 159 tatsächlich geliefert werden. Die Aktivität dieser Substanzen wurde in einem EtCRK2-Enzymtest bestimmt. Für die Messungen wurde der *TR-FRET Screening Express Kit* von Molecular Devices [<http://www.Moleculardevices.com>] verwendet (siehe Abschnitt 2.14). Sechs der 159 Substanzen wiesen einen IC50-Wert kleiner als 100  $\mu\text{M}$  auf (siehe Tabelle 3.7). Für alle aktiven Substanzen wurden die Löslichkeiten in Wasser nephelometrisch bestimmt (Liane Walter, Intervet Innovation GmbH, Schwabenheim). Substanzen die eine Löslichkeit größer als 500  $\mu\text{M}$  zeigten, wurden mittels LC/MS- und NMR-Messungen einer Qualitätskontrolle unterzogen (Stefan Derschum und Christian Schorn, Intervet Innovation GmbH, Schwabenheim).

Um aufzuklären ob es sich bei den Substanzen tatsächlich um ATP-kompetitive Inhibitoren handelt, wurden von María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim) Enzym-kinetische Messungen durchgeführt. Die Ergebnisse dieser Messungen bestätigten die

Annahme, dass es sich um ATP-kompetitive Inhibitoren handelt. Substanzen wurden nur als Leitstrukturen definiert, wenn sie für die EtCRK2 einen IC<sub>50</sub>-Wert sowie einen K<sub>i</sub>-Wert von kleiner 100  $\mu$ M zeigten und sie eine Wasserlöslichkeit größer 500  $\mu$ M besaßen sowie ihre chemische Struktur mit LC/MS- und NMR-Messungen bestätigt werden konnten. Die identifizierten Leitstrukturen sind in Tabelle 3.8 aufgeführt.

Tabelle 3.7: *In vitro* Experimente der identifizierten EtCRK2-Inhibitoren der zweiten Generation.






Struktur	Klasse	EtCRK2 IC <sub>50</sub> [ $\mu$ M]	Löslichkeit in Wasser [ $\mu$ M]
	Benzimidazol- Carbonitril	49	>500
	Benzimidazol- Carbonitril	62	>500
	Naphtolactam	56	125
	Naphtolactam	65	125
	Naphtolactam	41	>500
	Naphtolactam	53	>500

Unpublizierte Ergebnisse des Enzym-Assays von María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim).

Die interessanteste Klasse der EtCRK2-Inhibitoren sind die Benzimidazol-Carbonitrile, da sie bislang in der Literatur nur von Badawey, Kappe und El Hawash [BADAWEY & KAPPE, 1999; EL HAWASH *et al.*, 1999] sowie in einem Patent [ARBISER & AMBLARD, 2006] in Zusammenhang mit der antineoplastischen Wirkung dieser Klasse beschrieben wurden. Alle anderen Strukturklassen sind schon längere Zeit als Inhibitoren der CDKs oder Glycogen-synthasen (GSK) bekannt [BRAMSON *et al.*, 2001; GRAY *et al.*, 1999; LIU *et al.*, 2003]. Die Familien der CDK und der GSK sind eng miteinander verwandt [MANNING *et al.*,

2002], daher ist es naheliegend, dass ähnliche Strukturtypen in beiden Proteinfamilien eine Aktivität zeigen.

Tabelle 3.8: *In vitro* Experimente der identifizierten Leitstrukturen der EtCRK2.

Struktur	Klasse	IC50 [µM]			Ki[µM]		Löslichkeit in Wasser [µM]
		HsCDK2	EtCRK2	PfPK5	HsCDK2	EtCRK2	
	Oxindol	0,23 ± 0,11	0,66 ± 0,19	2,70 ± 0,78	0,08 ± 0,01	0,17 ± 0,03	>500
	Rhodanin-Oxindol	26 ± 19	23 ± 14	24 ± 7	17 ± 8	16 ± 9	>500
	Naphtolactam	12 ± 7	36 ± 8	>200	2 ± 1	6 ± 2	>500
	Benzimidazol-Carbonitril	41 ± 13	61 ± 27	>200	41 ± 13	83 ± 29	>500
	Benzimidazol	35 ± 3	15 ± 3	>200	8 ± 2	8 ± 2	>500

Es konnten vier Leitstrukturen von Inhibitoren der EtCRK2 bestätigt werden. Als Referenz diente ein Oxindol (Indirubin-5-Sulfonat [DAVIES *et al.*, 2001]) welches ein Inhibitor der humanen CDK2 ist. Der Enzym-Assay wurde von María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim) durchgeführt. Anzahl der Durchführungen im Enzyme-Assay: n = 3-4.

### 3.2.4 Struktur-Aktivitäts-Beziehungen und Analyse der Protein-Liganden-Bindung

Zur chemischen Optimierung der identifizierten Strukturen wurden *Docking*-Experimente zur Ableitung von Struktur-Aktivitätsbeziehungen sowie zur Vorhersage der Bindemodi durchgeführt. Hierzu wurden alle durch *in vitro* Testung bestätigten EtCRK2-Inhibitoren in die ATP-Bindetasche mittels *Docking* eingepasst und jeweils 100 Liganden-Posen generiert. Für die Strukturklassen der Oxindole, Rhodanine und Naphtolactame konnte in *Docking*-Experimenten ein eindeutiger Bindemodus abgeleitet werden. Diese Bindemodi entsprachen solchen, die für diese Strukturklassen bereits in der Literatur beschrieben wurden [DUBININA *et al.*, 2007]. Sie binden alle über mindestens eine Wasserstoffbrückeninterakti-

on zur *Hinge*-Region. Für die Klasse der Benzimidazol-Carbonitrile konnte jedoch der Bindemodus nicht eindeutig vorhergesagt werden.

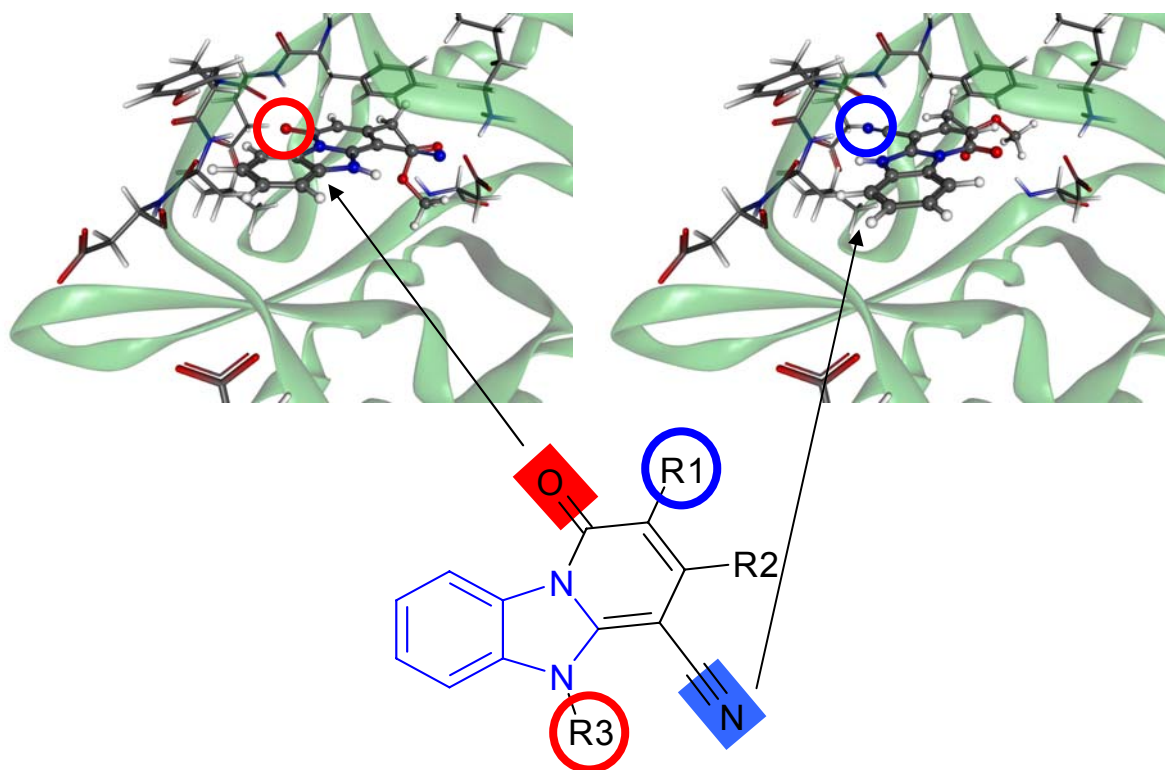


Abbildung 3.8: Analyse der Bindemodi der Benzimidazol-Carbonitrile.

Hier sind die zwei möglichen Bindemodi der Benzimidazol-Carbonitrile dargestellt. Eine Substitution an Position R1 resultiert im molekularen *Docking* in einer Pose der Benzimidazol-Carbonitrile, die über den Nitril-Stickstoff an die *Hinge*-Region binden (blau). Substitutionen an Position R3 erzeugen *Docking*-Posen, bei denen die Benzimidazol-Carbonitrile eine Wasserstoffbrücke über den Carbonyl-Sauerstoff zur *Hinge*-Region bilden (grün).

Die Bindemodi der Benzimidazole variieren in *Docking*-Analysen abhängig vom Substitutionsmuster an Position R1, R2 und R3 (siehe Abbildung 3.8). Eine Substitution an Position R1 und R3 lieferte im molekularen *Docking* eindeutige Bindemodi. Dabei dirigiert ein Substituent an Position R1 den Liganden so, dass er unter Ausbildung einer Wasserstoffbrücke über den Nitril-Stickstoff an die *Hinge*-Region bindet. Eine Substitution an Position R3 bewirkt eine Rotation des Liganden um  $180^\circ$  entlang der horizontalen Achse in der Papierebene (siehe Tabelle 3.8). Dadurch bindet der Ligand über den Carbonyl-Sauerstoff unter Ausbildung einer Wasserstoffbrücke an die *Hinge*-Region. Substitutionen an R2 führten zu beiden möglichen Bindemodi. Bei 100 generierten Liganden-Posen mit Substituent an R2 wurden beide Binde-

modi annähernd gleichmäßig verteilt gefunden. Die *Scoring*-Werte unterschieden sich nicht signifikant. Es war mittels *Docking* nicht möglich die Bevorzugung eines Bindemodus nachzuvollziehen. Zur genauen Untersuchung des Bindeverhaltens wurden deshalb Freie-Energie-Simulationen, speziell die LIE-Methode [AQVIST & MARELIUS, 2001], eingesetzt. Diese ist in der Lage, die freie Bindungsenergie für Rezeptor-Ligand-Komplexe zu berechnen. In der LIE-Methode wird der Protein-Ligand-Komplex vollständig flexibel betrachtet, sowie im Bereich der Ligand-Bindetasche solvatisiert. Dadurch ist es möglich, Entropiebeiträge, die im *Docking* naturgemäß vernachlässigt werden, zu berücksichtigen. Aufgrund der Vereinfachungen, die der LIE-Methode zugrunde liegen, ist es nötig, die Ergebnisse über die Gewichtungsfaktoren  $\alpha$ ,  $\beta$  und  $\gamma$  an experimentell bestimmten Werten anzupassen.

### 3.2.5 Validierung der LIE-Methode und Berechnung der Gewichtungsfaktoren

Zur Validierung der LIE-Methode wurde die Kristallstruktur der humanen CDK2 (PDB-ID: 1OIR) herangezogen. Die Ligandendatensätze wurden auf Basis von veröffentlichten Bindungsenergie-daten erstellt. Hierzu wurden zwei Liganddatensätze verwendet. Der eine Datensatz stammt von Kolb *et al.* [BRAMSON *et al.*, 2001; GIBSON *et al.*, 2002; KOLB *et al.*, 2008] der andere Datensatz enthält Liganden aus einem Patent [CHONG WESLEY *et al.*, 2003]. Hierbei wurden zunächst die Strukturen der beiden Testdatensätze mittels *Docking* in die Kristallstruktur eingepasst, um eine Ausgangskonformation eines möglichen Protein-Ligand-Komplexes zu erhalten. Sowohl für den Patentdatensatz als auch für den Datensatz von Kolb *et al.* konnte eine gute Korrelation zwischen experimentell bestimmten Bindungsenergien und mittels der LIE-Methode berechneten Energien erzielt werden. Nach Kreuz-Validierung mit der *Leave-One-Out* Methode [TOMINAGA & JORGENSEN, 2004] wurde ein  $Q^2$  von 0,71 für den Patentdatensatz und ein  $Q^2$  von 0,79 für den Datensatz von Kolb *et al.* erreicht. Die durchschnittliche Abweichung von den experimentell ermittelten Bindungsenergien betrug 0,41 kcal/mol für den Patentdatensatz und 0,86 kcal/mol für den Datensatz von Kolb *et al.* Für die LIE-Methode wird eine durchschnittliche Abweichung zwischen experimentell ermitteltem und berechnetem  $\Delta G$ -Wert von etwa einem kcal/mol angenommen [FOLOPPE & HUBBARD, 2006].

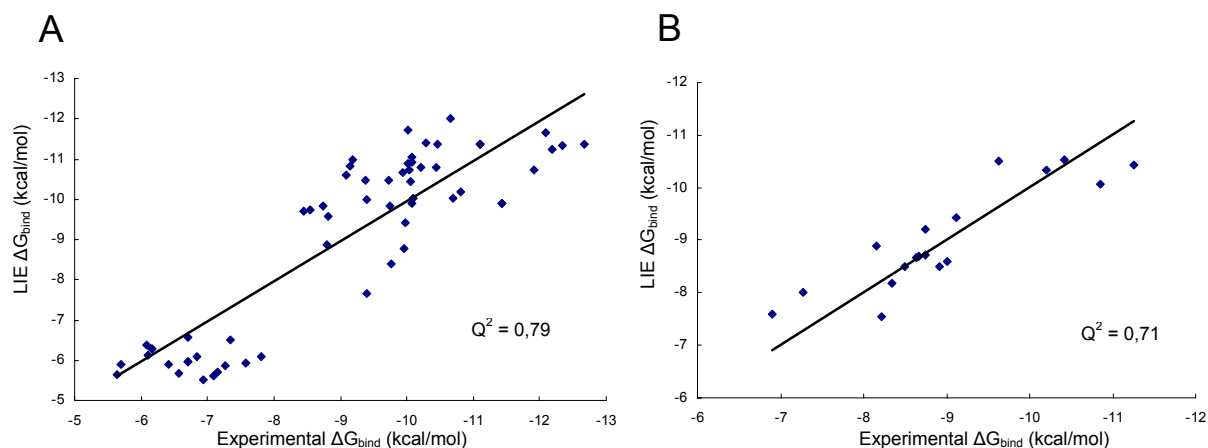


Abbildung 3.9: Scatter-Plots der Simulationsergebnisse von CDK-Inhibitoren mit der LIE-Methode.

A) LIE-Simulationen für den Datensatz von Kolb *et al.* [BRAMSON *et al.*, 2001; GIBSON *et al.*, 2002; KOLB *et al.*, 2008].

B) LIE-Simulationen für den Patentdatensatz [CHONG WESLEY *et al.*, 2003].

Zur Anpassung der Gewichtungsfaktoren diente das Homologiemodell der EtCRK2, sowie bekannte CDK-Inhibitoren für die *in vitro*  $K_i$ -Werte mit der EtCRK2 gemessen wurden. Das LIE-Modell ergab ein  $Q^2$  von 0,87 in der Kreuz-Validierung und eine durchschnittliche Abweichung von 0,38 kcal/mol von den experimentell ermittelten Bindungsenergien. Es wurden die Koeffizienten  $\alpha = 0,80$ ,  $\beta = 0,30$  und  $\gamma = -7,62$  für das EtCRK2-LIE-Modell ermittelt.

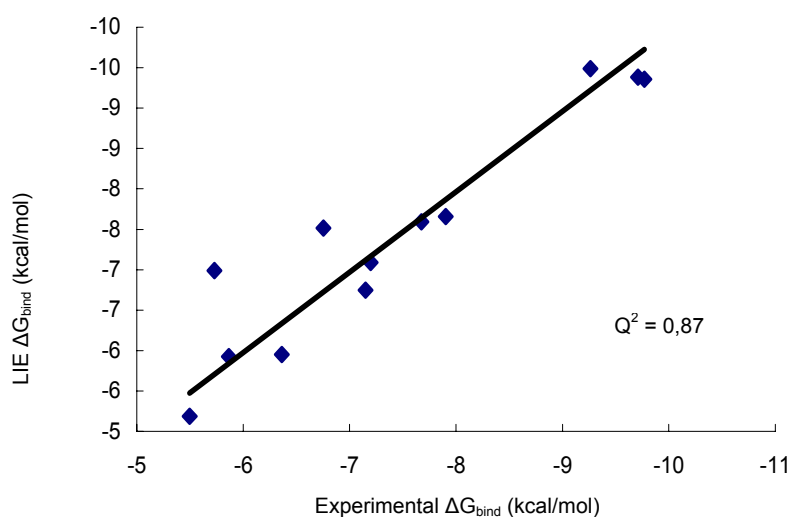


Abbildung 3.10: Scatter-Plots der LIE-Simulationen für die EtCRK2.

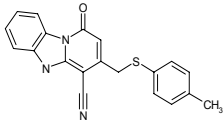
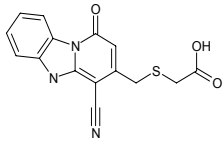
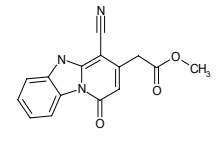
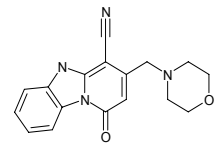
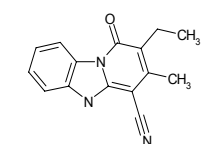
Für das EtCRK2 Homologiemodell mit bekannten CDK2-Inhibitoren, konnte ein prädiktives LIE-Modell mit einem  $Q^2$  von 0,87 erstellt werden.



### 3.2.6 Bindemodi-Vorhersage der Benzimidazol-Carbonitrile

Für die Benzimidazol-Carbonitrile mit Substitution an Position R2 wurden LIE-Simulationen durchgeführt. Hierfür wurden pro Liganden-Pose zwei LIE-Simulationen erstellt. Die Bindungsenergien wurden für jede *Docking*-Pose (einmal über den Carbonyl-Sauerstoff über eine Wasserstoffbrücke gebunden an die *Hinge*-Region und einmal über den Nitril-Stickstoff gebunden) berechnet. Die beiden Bindungsmodi eines Liganden wurden anhand des jeweils besten *Docking-Scores* ausgewählt.

Tabelle 3.9: Bindemodi-Vorhersage der Benzimidazol-Carbonitrile mittels LIE-Methode.

Struktur	Bindemodus	Anzahl Docking-Posen	$\Delta G_{cal}$ [kcal/mol]
	O	55	-7,95
	N	45	-8,41
	O	55	-6,93
	N	45	-9,48
	O	22	-4,55
	N	78	-6,07
	O	25	-4,31
	N	75	-5,80
	O	93	-3,33
	N	7	-2,87

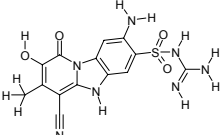
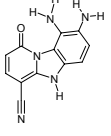
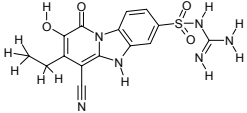
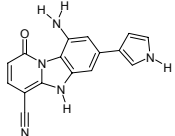
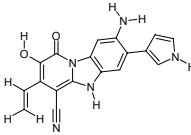
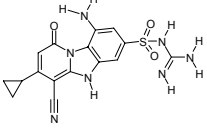
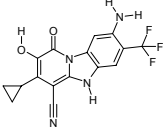
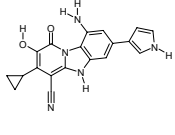
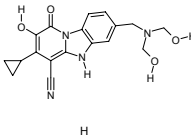
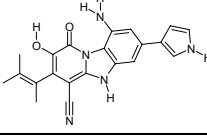
Aus den LIE-Simulationen leiten sich leichte Unterschiede in den Bindungsenergien für die jeweiligen Bindemodi ab. Diese liegen jedoch meist in der Größenordnung des Standardfehlers von einem kcal/mol.

Die LIE-Simulationen ergaben leichte Unterschiede in den Bindungsenergien für die beiden Bindemodi einzelner Liganden. Diese Unterschiede sind jedoch zumeist in der Größenordnung des Standardfehlers von einem kcal/mol. Drei Fälle bewegen sich in einer Größenordnung außerhalb des angenommenen Fehlers, jedoch auch nicht in einer Größenordnung, die es erlaubt, eine Orientierung sicher vorherzusagen. Die Bindung über den Nitril-Stickstoff zur *Hinge*-Region scheint energetisch leicht bevorzugt, dennoch sind wahrscheinlich beide Orientierungen möglich.

### 3.2.7 *In silico* Optimierung der Benzimidazol-Carbonitrile

Zur *in silico* Optimierung der Benzimidazol-Carbonitrile wurde eine kombinatorische Substanzbibliothek für diese Strukturklasse generiert. Hierzu wurde an den Positionen R1 bis R7 substituiert (siehe Abschnitt 2.12 sowie Abbildung 2.2). Die Generierung der kombinatorischen Substanzdatenbank wurde mittels der Software Cactvs, wie in Kapitel 2.12 beschrieben, durchgeführt. Hierdurch wurden insgesamt 12.356 Strukturen mit der Grundstruktur der Benzimidazol-Carbonitrile generiert. Die Synthetisierbarkeit wurde für die *in silico* Generierung der kombinatorischen Substanzbibliothek nicht berücksichtigt. Daher handelte es sich um eine rein virtuelle Bibliothek. Die Verbindungen dieser virtuellen Datenbank wurden in das EtCRK2-Homologiemodell mittels *Docking* eingepasst. Es wurden entsprechend dem in Abschnitt 3.2.2 beschriebenen Vorgehen jeweils 10 Posen generiert und insgesamt 100 geeignete Substanzen durch 2-Stufen-Bewertung, Interaktionsfilter sowie visueller Begutachtung ausgewählt. Als weiteres Auswahlkriterium wurde die Löslichkeit der Substanzen – berechnet mit Volsurf+ – herangezogen. Nach Löslichkeitsanalysen verblieben 20 Verbindungen in der finalen Auswahl. Mit diesen Strukturen wurden im Anschluss Bindungsenergien mittels der LIE-Methode berechnet. Tabelle 3.10 zeigt die finale Auswahl potentieller Inhibitoren sowie deren berechnete  $\Delta G$ -Werte bzw.  $K_i$ -Werte. Alle optimierten Strukturen der finalen Auswahl der Benzimidazol-Carbonitrile, zeigten im *Docking* eine einheitliche Liganden-Pose. Diese Strukturen bilden in den *Docking*-Experimenten über den Carbonyl-Stickstoff eine Wasserstoffbrücke zur *Hinge*-Region aus (siehe Abbildung 3.11).

Tabelle 3.10: Optimierte Benzimidazol-Carbonitrile mit vorhergesagten Ki-Werten.

Optimierte Struktur	$\Delta G$ [kcal/mol]	Ki [nM]
	-11,93	1,8
	-10,94	9,6
	-10,61	16,9
	-10,28	29,0
	-10,07	41,6
	-10,00	47,3
	-9,50	109,9
	-9,20	181,5
	-9,12	209,3
	-8,84	334,0

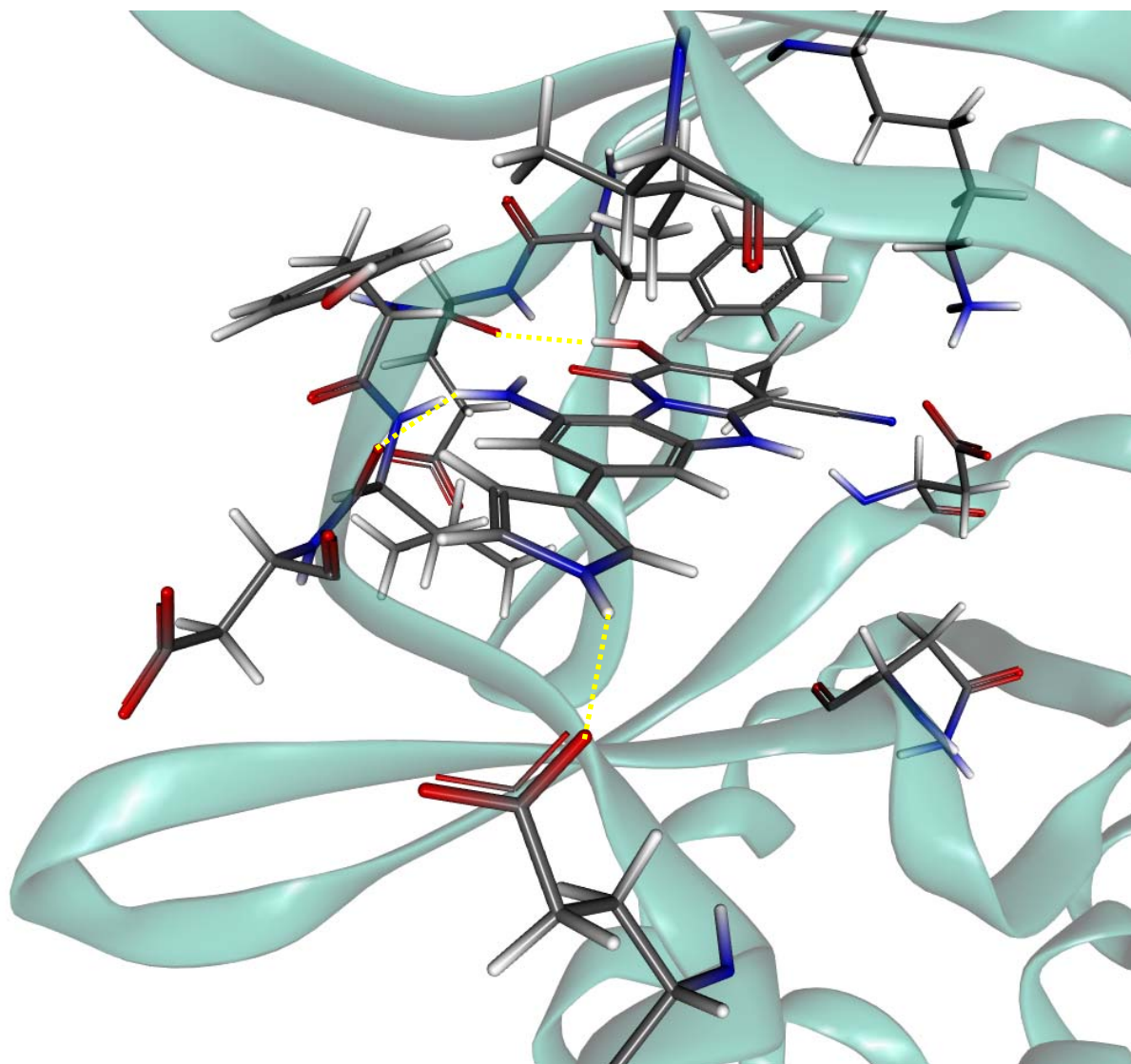


Abbildung 3.11: Optimierte Benzimidazol-Carbonitril-Struktur.

Die optimierte Struktur bildet zwei Wasserstoffbrücken zur *Hinge*-Region aus sowie eine Wasserstoffbrücke zum Glu88 und adressiert somit auch die Selektivitätstasche.

### 3.3 Identifizierung von EtMRK- und PfMRK-Inhibitoren

Ein weiteres Zielprotein zur Entwicklung neuer Antiparasitika sind die MRK-homologen Kinasen (MRK = *MO15 Related Kinases*) der *Apicomplexa*. Diese Proteine gehören ebenfalls zu der Familie der Zyklin-abhängigen Kinasen und wurden sowohl für die EtMRK als auch für die PfMRK als geeignete Zielproteine validiert (siehe Abschnitt 1.5.5). Ein Ziel dieser Arbeit war es, mittels virtuellen *Screenings* potentielle Inhibitoren der PfMRK und EtMRK zu identifizieren. Hierzu wurden Homologiemodelle der EtMRK und der PfMRK generiert und potentielle Inhibitoren mit der *Docking*-Software PLANTS [KORB *et al.*, 2006; KORB *et al.*,

2009] eingepasst. Als *Screening*-Datenbank wurde die BindingDB [www.bindingdb.org] verwendet. Die BindingDB enthält bekannte Strukturen sowie Aktivitätsdaten von Inhibitoren humaner Kinasen [LIU *et al.*, 2007], die auf diese Weise für den Einsatz als Antiparasitika optimiert werden könnten.

### 3.3.1 Homologiemodellierung der EtMRK und der PfMRK

Für das virtuelle *Screening* nach potentiellen EtMRK- und PfMRK-Inhibitoren wurden Homologiemodelle dieser Proteine erstellt. Zum Zeitpunkt der Arbeiten (Stand: Juli 2008) sind in der Proteinstrukturdatenbank PDB [BERMAN *et al.*, 2003] fünf Kristallstrukturen von CDK-ähnlichen Proteinen der *Apicomplexa* (eine von *Cryptosporidium parvum* sowie vier von *P. falciparum*) verzeichnet, die als potentielle *Template* für die Modellierung dienen könnten. Diese Proteine sind homolog zu HsCDK2. Des Weiteren waren 191 Kristallstrukturen der HsCDK2 und eine Kristallstruktur der HsCDK7 in der PDB abgelegt. Somit standen insgesamt 197 Templatstrukturen für die Homologiemodellierung der EtMRK und der PfMRK zur Verfügung. Jedoch sollte für die Modellierung eine Kristallstruktur als Templat gewählt werden, die eine möglichst hohe Sequenzidentität zu der *Target*-Proteinsequenz aufweist. In Tabelle 3.11 sind mögliche Templatstrukturen mit den jeweiligen Sequenzidentitäten zu der EtMRK und der PfMRK angegeben.

Die in der PDB abgelegten HsCDK2 Kristallstrukturen liegen zum Teil im aktiven Komplex als Heterodimer, d. h. gebunden mit einem Zyklin vor, und sind teilweise, mittels Phosphorylierung an Thr160 im Aktivierungs-*Loop*, vollständig aktiviert. Für das strukturbasierte Wirkstoffdesign sind insbesondere Strukturen interessant, die mit einem Inhibitor co-kristallisiert sind. Diese Strukturen weisen eine Proteinkonformation auf, die durch die Interaktion mit dem Liganden zustande kommen (*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, und eignen sich daher besonders gut als Templatstruktur für die Homologiemodellierung sowie auch für die Analyse von Protein-Ligand-Wechselwirkungen. Darüber hinaus sollte eine Templatstruktur eine möglichst hohe Auflösung und möglichst geringe isotrope Debye-Waller-Temperaturfaktoren [GAO & PENG, 1999] der Aminosäurenreste der Ligand-Bindetasche aufweisen. Unter Berücksichtigung aller dieser Gütefaktoren wurde die HsCDK2 Struktur mit der PDB-ID 1OIR als Templatstruktur zur Erstellung der Modelle ausgewählt. Sie besitzt eine relativ hohe Auflösung von 2,1 Å, sowie sehr geringe Temperaturfaktoren der Aminosäurereste der Bindetasche

[BREITENLECHNER *et al.*, 2005]; [NISSINK *et al.*, 2002]. Außerdem zeigt diese Kristallstruktur eine geeignete Konformation des Aminosäurerestes Lys33 (HsCDK2-Nummerierung), die eine ionische Wechselwirkung mit potentiellen Inhibitoren erlaubt. Anreicherungsexperimente bekannter Liganden mittels *Docking* zeigen gute Ergebnisse mit dieser Kristallstruktur [BEYER, 2007].

Tabelle 3.11: Mögliche Templatstrukturen zur Homologiemodellierung der EtMRK und der PfMRK

Organismus	EtMRK		PfMRK		PDB-ID	Auflösung [Å]
	Identität	Ähnlichkeit	Identität	Ähnlichkeit		
	[%]	[%]	[%]	[%]		
CpCDK2	35	55	39	56	2QKR	2,6
PfPK5	36	51	42	53	1V00	1,9
HsCDK2	38	59	39	52	1OIR	1,9
HsCDK7	38	62	40	52	1UA2	3,0

Zur Generierung der Homologiemodelle für die EtMRK und die PfMRK wurde die humane CDK2 Struktur (PDB-Acc.-Nr.: 1OIR), trotz der etwas geringeren Sequenzähnlichkeit im Vergleich zur humanen CDK7 verwendet, da diese Struktur eine sehr gute Auflösung besitzt.

Zu Beginn der Homologiemodellierung wurden die Sequenzen von 1OIR und der EtMRK sowie der PfMRK aligniert und dann jeweils zehn Modelle mit der Software MOE generiert. Die Modelle wurden anschließend anhand einer Protein-Geometrie-Analyse in MOE bewertet. Dabei wurde zunächst ein Ramachandran-Plot für beide Modelle erstellt (siehe Abbildung 3.12). In diesem Plot werden die  $\Phi$ - und  $\Psi$ - Winkel der Aminosäuren gegeneinander aufgetragen. Dieser Plot ermöglicht eine einfache graphische Aussage über favorisierte und erlaubte Bereiche einzelner Aminosäuren. Aminosäuren, die sich in nicht-favorisierten Bereichen befinden, sollten genauer überprüft werden. Während beim EtMRK-Modell alle Aminosäuren günstige Torsionswinkel und somit niedrige Energien besitzen, finden sich beim PfMRK-Modell vier Aminosäuren mit ungünstigen Torsionswinkeln (siehe Abbildung 3.12). Diese liegen jedoch außerhalb der Proteinbindetasche und wurden daher nicht weiter berücksichtigt.

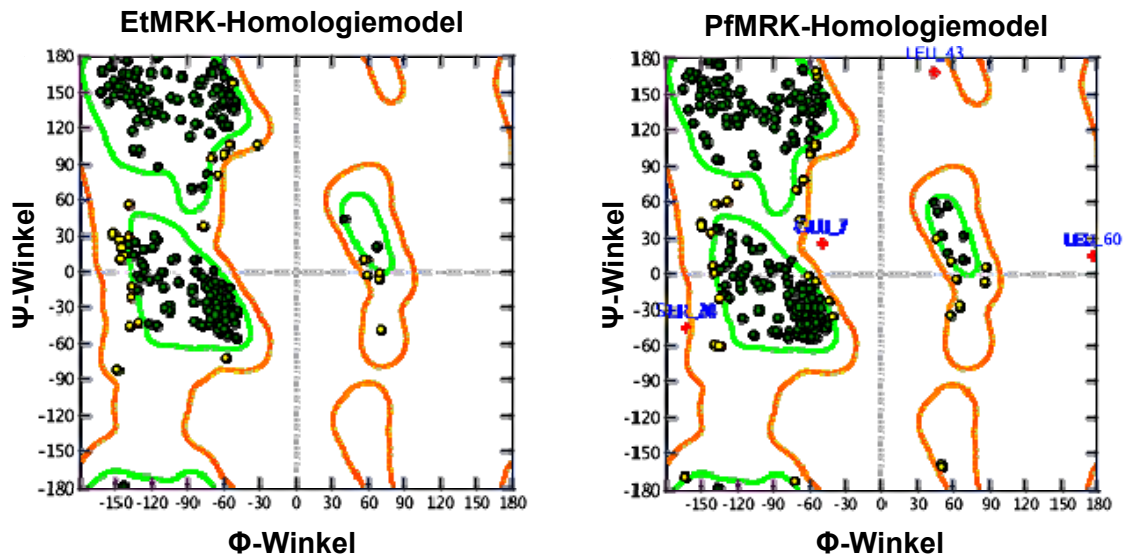


Abbildung 3.12: Ramachandran-Plots der Homologiemodelle der EtMRK und der PfMRK.

Bereiche favorisierter Torsionswinkel sind grün umrandet. Gelb umrandet sind Bereiche, deren Aminosäuren erlaubte Torsionswinkel besitzen. In weißen Regionen (d. h. ohne farbige Umrandung) sind Aminosäuren positioniert, in denen Aminosäuren mit ungünstigen Bindungswinkeln vorkommen.

Analog zu den  $\Phi$ - und  $\Psi$ -Winkeln wurden auch die Bindungslängen und die Bindungswinkel der Aminosäureseitenketten überprüft. Kam es hier zu Abweichungen von Standardwerten wurden diese Aminosäuren näher betrachtet und mittels einer Energieminimierung optimiert. Gegebenenfalls wurden geeignetere Rotamere einzelner Seitenketten aus einer Rotamer-Datenbanken verwendet. Hierbei wurde auf mögliche Interaktionen der Seitenketten (beispielsweise Wasserstoffbrücken) mit benachbarten Aminosäuren geachtet. Im letzten Schritt wurden mögliche Atomkollisionen untersucht und Kontaktenergien berechnet. In beiden Modellen wurden keine Atomkollisionen detektiert und die Kontaktenergien der einzelnen Atome lagen innerhalb der Normwerte.

Die Güte der so erstellten Homologiemodelle der EtMRK sowie der PfMRK wurde mittels *Docking* überprüft. Hierzu wurden bekannte Oxindol-Inhibitoren der PfMRK [WOODARD *et al.*, 2003] unter Verwendung der *Docking*-Software GOLD in die Modelle eingepasst. Die *Docking*-Experimente lieferten Liganden-Posen mit einem Bindungsmodus, wie er für Oxindol-Inhibitoren zu erwarten wäre [BRAMSON *et al.*, 2001; WOODARD *et al.*, 2003] und bestätigten so die Güte der Modelle (siehe Abbildung 3.13).

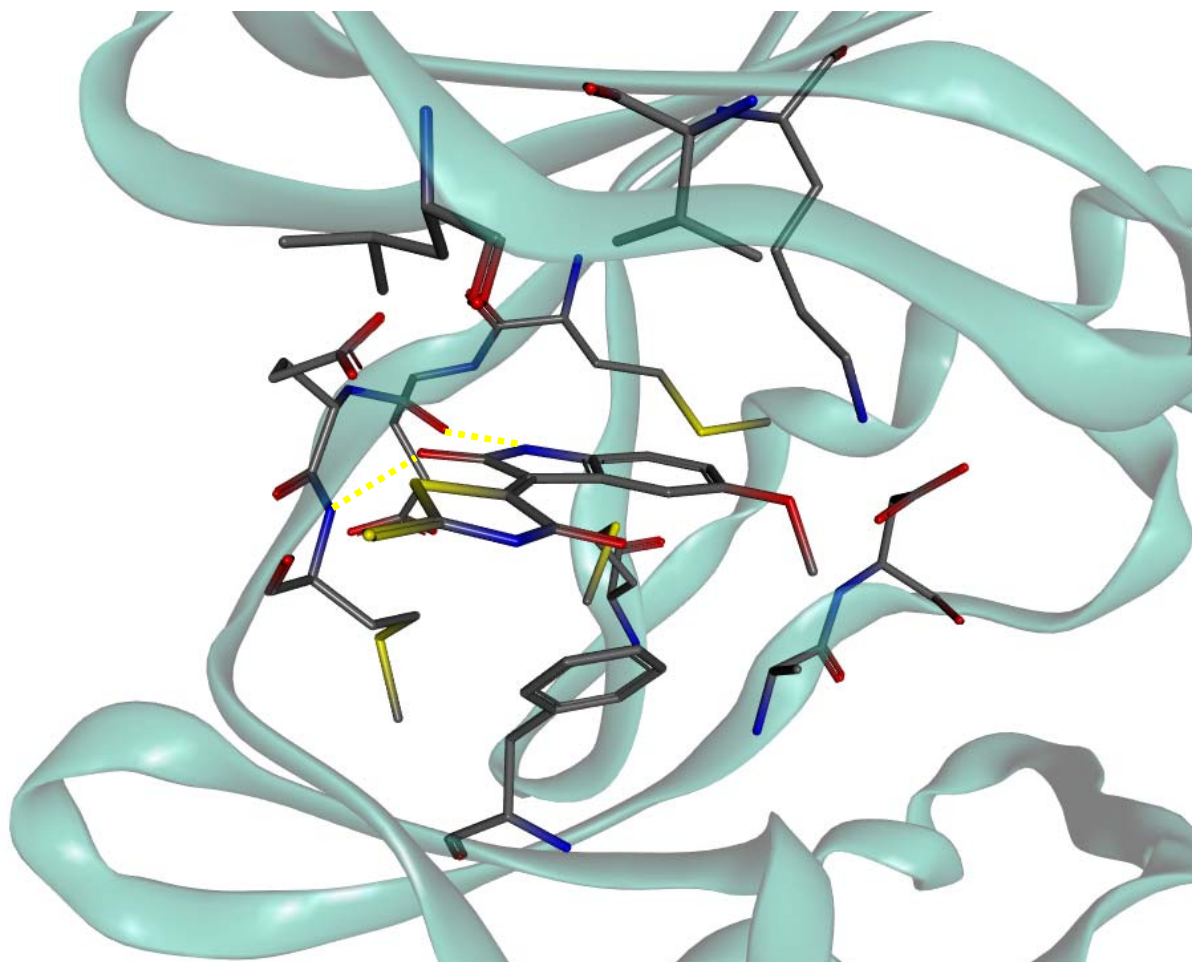


Abbildung 3.13: Bindungsmodus eines bekannten PfMRK-Inhibitors.

Der bekannte Oxindole-Inhibitor [WOODARD *et al.*, 2003] bindet über typische Wasserstoffbrückenbindungen zur Hinge-Region.

### 3.3.2 Evaluierung der Docking-Software PLANTS

Die Docking-Software PLANTS ermöglicht es, eine virtuelle Screening-Datenbank sehr schnell nach potentiellen Liganden zu durchsuchen und wurde daher zum virtuellen Screening nach Inhibitoren der EtMRK sowie PfMRK eingesetzt [KORB *et al.*, 2006; KORB *et al.*, 2009]. Da bislang nur wenige virtuelle Screenings mit dieser Docking-Software durchgeführt wurden, musste die Software zunächst hinreichend evaluiert werden. Sie wurde daher in einigen Docking-Experimenten mit der Docking-Software GOLD [JONES *et al.*, 1995] verglichen.



### 3.3.2.1 Re-Docking Experimente

Zur Evaluierung der *Docking*-Software PLANTS wurden zunächst Re-Docking Experimente durchgeführt und mit der Standard-Dockingsoftware GOLD verglichen. Für ein Re-Docking Experiment benötigt man geeignete Kristallstrukturen mit co-kristallisierten Liganden. Hierzu wurden 15 Kristallstrukturen der humanen CDK2 verwendet, die eine Auflösung kleiner 2,5 Å besitzen, mit einem Inhibitor co-kristallisiert sind und deren Bindetaschen-Aminosäuren vollständig kirystallographisch aufgelöst sind. Für das Re-Docking wurden die Inhibitoren zunächst aus der Proteinbindetasche entfernt und neue Startkonformationen berechnet [SADOWSKI, 1997; TETKO, 2005]. Für den Vergleich zwischen Ursprungskonformation der Liganden und Konformation der Liganden nach dem *Docking*, wurden *root mean square deviation* (RMSD) Werte der Schweratome der Liganden berechnet und dienten als Maß für die Qualität der Posen-Generierung des *Docking*-Algorithmus. Je kleiner der resultierende RMSD-Wert, desto genauer reproduziert die *Docking*-Software die Liganden-Pose. Als ein ausreichend genaues *Docking* wurde ein RMSD-Wert von maximal 2 Å angenommen. Für diese Experimente wurden die Einstellungen der PLANTS- und der GOLD-Software so gewählt, dass sie eine möglichst schnelle Berechnung der einzelnen Liganden erlauben (*Screening* Modus). Für jeden Liganden wurden 25 verschiedene Posen generiert und im Anschluss daran Mittelwerte der berechneten RMSD-Werte gebildet (siehe Abbildung 3.14). Die Erfolgswahrscheinlichkeit, eine Liganden-Pose mit einem Durchschnitts-RMSD von kleiner 2 Å, verglichen mit der Konformation der Kristallstruktur, zu generieren, lag bei 27% für die PLANTS-Software sowie bei 53% für die GOLD-Software. Die relativ hohe Standardabweichung der Durchschnitts-RMSD-Werte der einzelnen Protein-Ligand Komplexe liegt unter anderem darin begründet, dass in einigen Experimenten sehr unterschiedliche Posen desselben Liganden vorhergesagt wurden, welche dementsprechend hohe RMSD-Werte aufwiesen. Einige Inhibitoren zeigten zudem *Docking*-Posen die um 180° rotiert vorlagen. Unterschiedliche Bindemodi in *Docking*-Experimenten, finden sich für CDK-Inhibitoren häufig, da sie im *Docking* mit der *Hinge*-Region über verschiedene Wasserstoffbrücken-Muster in Interaktion treten können [WU *et al.*, 2003] und bei einer Drehung um 180° sogar das typische Interaktions-Muster mit der *Hinge*-Region beibehalten.

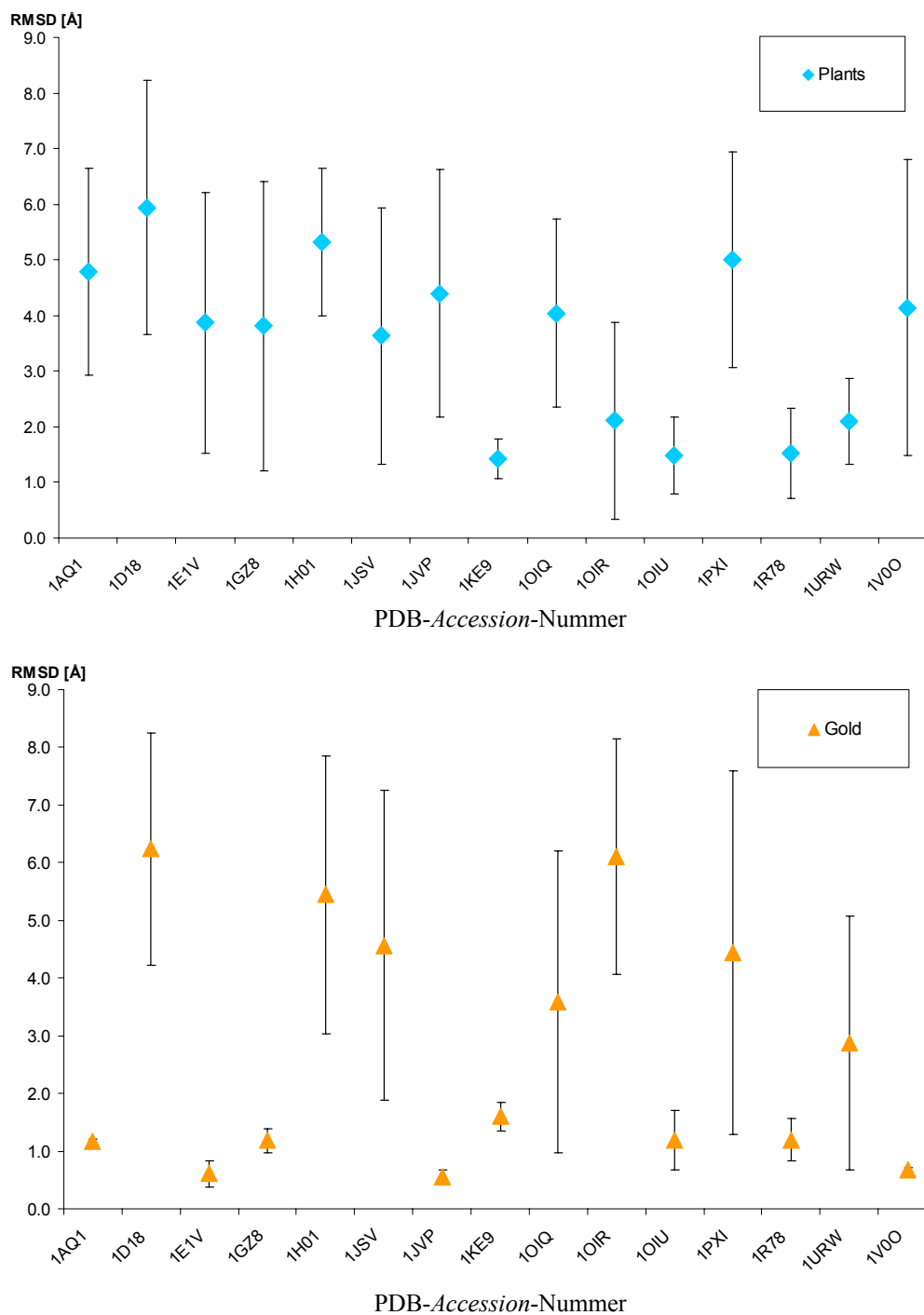


Abbildung 3.14: Durchschnitts-RMSD-Werte der Liganden-Posen nach einem Re-Docking mit der Software *PLANTS* und *GOLD*.

Die Beiden Grafiken veranschaulichen deutlich, dass die *Docking*-Software *GOLD* häufiger die bioaktive Konformation reproduzieren konnte.

In dem Balkendiagramm (Abbildung 3.15) repräsentieren die Balken die Anzahl der richtig generierten *Docking*-Posen. Die *GOLD*-Software war in 66% aller generierten Posen in der Lage, die Protein-Ligand-Komplexe erfolgreich zu reproduzieren. Eine ähnlich gute *Perfor-*

*mance* wurde auch von Kellenberger *et al.* beschrieben [KELLENBERGER *et al.*, 2004]. PLANTS erzielte nur in 33% aller generierten Posen einen Durchschnitts-RMSD von kleiner 2 Å, verglichen mit der Konformation der Kristallstruktur. Für ein virtuelles *Screening* ist es nicht notwendig, dass alle generierten Liganden-Posen eines einzelnen Liganden der bioaktiven Konformation entsprechen. Wichtig ist, dass die verwendete Bewertungsfunktion zwischen „richtig“ und „falsch“ generierten *Docking*-Posen diskriminieren kann [CROSS *et al.*, 2009]. Zudem wurden *Docking*-Parameter verwendet, die eine schnelle Berechnung der Liganden-Posen erlauben. Für die Kristallstruktur 1H01 findet PLANTS in 25 generierten *Docking*-Posen keine korrekte Pose. Die Software GOLD war hingegen nicht in der Lage eine korrekte Liganden-Pose der Kristallstruktur 1OIR zu generieren. Zu einem ähnlichen Ergebnis kam Beyer unter Verwendung der Software AutoDock 3.0 [BEYER, 2007]. Dies könnte auf Fehler in den Proteinstrukturen zurückzuführen sein [SOTRIFFER & DRAMBURG, 2005].

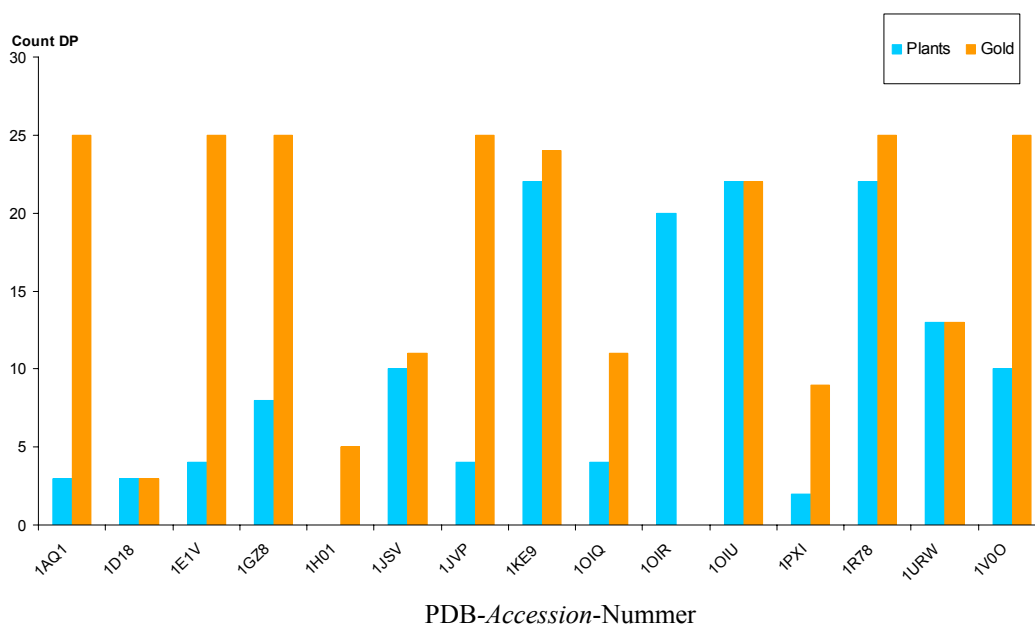


Abbildung 3.15: Anzahl der richtig reproduzierten Liganden-Posen nach Re-Docking Experimenten mit der Software PLANTS und GOLD.

Die *Docking*-Software GOLD findet im Gegensatz zu PLANTS häufiger korrekte Liganden-Posen, die der Kristallstruktur entsprechen. Für sechs der 15 Kristallstrukturen wurde stets die korrekte Liganden-Pose ermittelt.

Neben der Generierung einer korrekten Liganden-Pose, wurden die beiden *Docking*-Algorithmen von GOLD und PLANTS auch in Bezug auf ihre Bewertungsfunktionen getestet. Hierbei sollte überprüft werden, ob die Güte der ermittelten Liganden-Posen auch mit der internen Bewertungsfunktion der *Docking*-Software korreliert. Dazu wurde das finale *Ran-*

king via RMSD-Wert und via *Scoring*-Wert miteinander verglichen und eine Spearman Korrelation [WISSLER, 1905] durchgeführt. Bei dieser Art von Korrelationsanalyse wird untersucht, inwieweit zwei Variablen (z. B. *Scoring*- und RMSD-Wert) in Beziehung zueinander stehen. Der hierbei berechnete Spearman Korrelationskoeffizient  $R_s$  kann zwischen -1 und 1 variieren (-1 = inverse Korrelation, 0 = keine Korrelation, 1 = Korrelation).  $R_s$  sollte möglichst nahe eins liegen. Dies würde bedeuten, dass richtige Liganden-Posen einen besseren *Docking-Score* erhalten als solche mit nicht korrekten Posen. Bei einem negativen  $R_s$  hingegen würden schlechtere Liganden-Posen besser bewertet werden als korrekte Posen. Die Abbildung 3.16 zeigt, dass die interne Bewertungsfunktion von PLANTS gut geeignet ist, um eine gute Liganden-Pose zu identifizieren. Für sieben der 15 Re-*Scoring*-Experimente errechnete sich ein Korrelationskoeffizient  $R_s$  von über 0,5. Der Durchschnitts- $R_s$  über alle Experimente lag für PLANTS bei 0,32. Die *Docking*-Software GOLD schnitt hier mit einem Durchschnitts-  $R_s$  von 0,08 deutlich schlechter ab und zeigte im Durchschnitt keine Korrelation zwischen dem *Docking-Score* und dem RMSD-Wert. In elf von 15 Re-*Docking*-Experimenten war  $R_s$  von PLANTS höher als bei GOLD und nur in zwei Re-*Docking*-Experimenten (für die Kristallstrukturen 1JSV und 1PXI) schnitt GOLD besser ab.

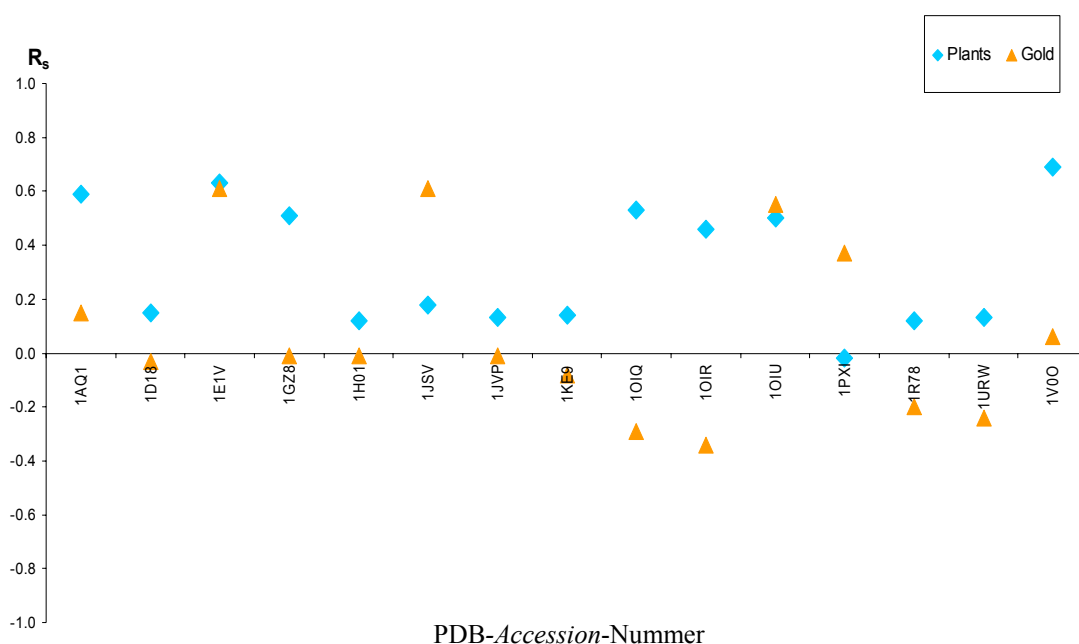


Abbildung 3.16: Vergleich zwischen dem RMSD-Wert und der Anreicherung von aktiven Substanzen.

Die Grafik zeigt, dass die *Scoring*-Funktion von PLANTS häufiger eine korrekte *Docking*-Pose identifiziert. Die korrekten *Docking*-Posen (d. h. Posen mit einem RMSD von kleiner 2 Å) bekommen einen besseren *Docking-Score*.

Entscheidend für ein virtuelles *Screening* ist, dass die Liganden-Pose mit dem besten *Scoring*-Wert, einen RMSD-Wert von kleiner 2 Å besitzt um Posen, die nicht der bioaktiven Konformation entsprechen, zu entfernen. Die GOLD-Software war für 73% der Protein-Ligand-Komplexe geeignet, eine ähnliche bioaktive Konformation zu selektieren. PLANTS erzielte in 93% der Fälle eine Selektion der bioaktiven Konformation (siehe Tabelle 3.12).

Tabelle 3.12: RMSD-Werte der Liganden-Posen mit dem besten GOLD-Score sowie PLANTS-Score nach einem *Re-Docking* in die Proteinbindetasche.

PDB-ID	GOLD-Score	PLANTS-Score
	RMSD	RMSD
1AQ1	1,2	0,9
1D18	7,1*	1,2
1E1V	0,4	0,6
1GZ8	1,0	0,8
1H01	2,1*	4,4*
1JSV	1,5	1,1
1JVP	0,6	0,5
1KE9	1,5	1,5
1OIQ	0,7	0,7
1OIR	8,3*	1,7
1OIU	0,8	0,8
1PXI	0,3	0,7
1R78	1,4	1,5
1URW	2,9*	2,0
1V0O	0,7	0,5

Angegeben ist der RMSD-Wert der Liganden-Pose eines Referenz-Liganden, die von den einzelnen *Scoring*-Funktionen am besten bewertet wurde.

### 3.3.2.2 Anreicherungsexperimente

Zur weiteren Evaluierung der *Software* GOLD und PLANTS wurden Anreicherungsexperimente von bekannten Inhibitoren der humanen CDK2 durchgeführt. Prinzipiell sollte eine *Docking-Software* im *in silico Screening* geeignet sein, zwischen Liganden eines Proteins und solchen, die wahrscheinlich nicht binden zu unterscheiden. Deshalb wurden *Docking-Experimente* mit einem speziell für solche Fragestellungen entwickelten Testdatensatz (siehe 2.2; DUD-Testdatensatz, [HUANG *et al.*, 2006]) mit dem CDK2-Protein durchgeführt. Dieser Datensatz enthält 72 CDK-Inhibitoren und insgesamt 2.074 inaktive Verbindungen. Die Inaktivität für diese Proteinklasse wurde *in silico* vorhergesagt und wurde nicht *in vitro* bestätigt. Als Proteinstruktur für das *Docking* diente wiederum eine Struktur der *humanen* CDK2 (PDB-Acc.-Nr.: 1OIR). Die Einstellungen der Parameter von PLANTS und GOLD entsprachen denen eines virtuellen *Screening Setups*.

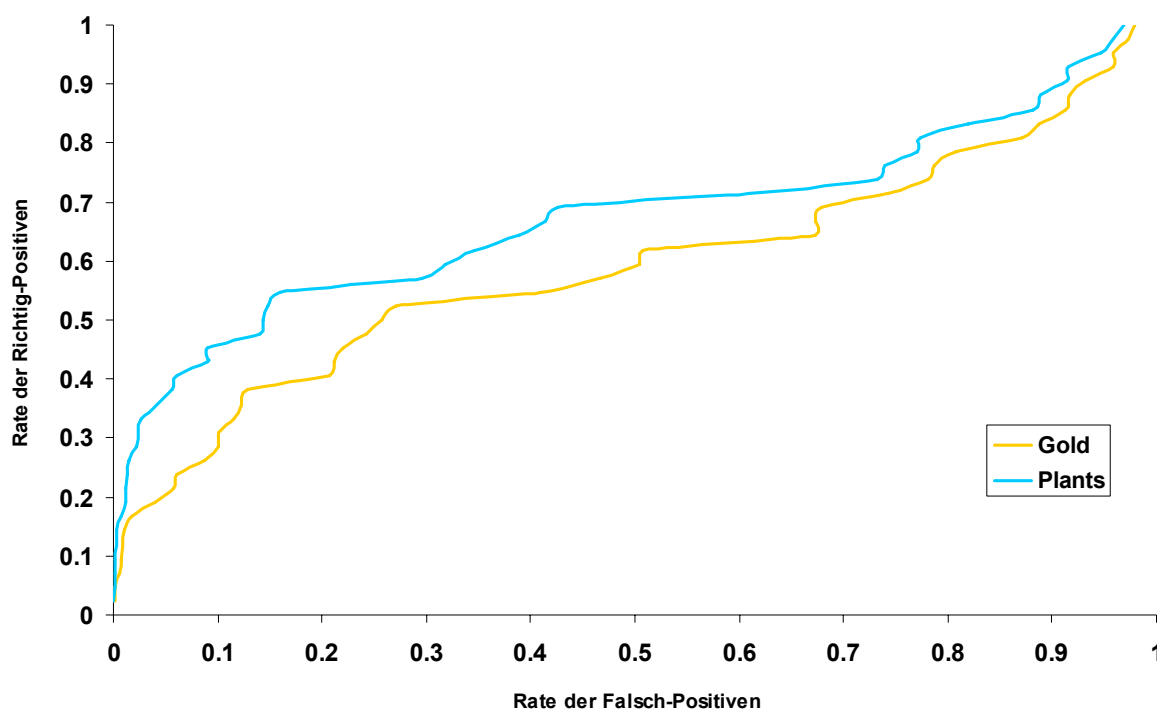


Abbildung 3.17: ROC-Kurve

Die beiden generierten ROC-Kurven für das PLANTS- und GOLD-*Docking* zeigen, dass die *Scoring-Funktion* PLANTS-Score im Gegensatz zu GOLD-Score besser in der Lage war, zwischen „Richtig-Positiven“ (d.h. CDK-Inhibitoren) von „Falsch-Positiven“ (*Decoys*) zu diskriminieren.

Zur Auswertung des Anreicherungs-Experimentes diente die ROC-Kurve (ROC = *Receiver Operating Characteristic*). In dieser wird die Rate der „Falsch-Positiven“ (inaktive Liganden)

mit hohem *Docking-Score* zur Rate der „Richtig-Positiven“ (aktive Liganden) aufgetragen. Abbildung 3.17 zeigt den Vergleich zwischen den beiden ROC-Kurven für das PLANTS- und das GOLD-*Docking*. Eine gute Anreicherung von aktiven Substanzen konnte mit der PLANTS-Software erzielt werden. Die GOLD-Software schnitt bei den durchgeführten *Docking*-Experimenten schlechter ab. Ein Maß für die Güte des Tests ist die Fläche unter der ROC-Kurve (AUC = *Area Under Curve*). Die Fläche kann Werte zwischen 0 und 1 annehmen, wobei ein höherer Wert eine höhere Güte anzeigt. Mit PLANTS konnte ein besserer AUC-Wert von 0,69 im Vergleich zu GOLD 0,58 erzielt werden. Ähnliche AUC-Werte wurden für die Software GOLD bereits publiziert [Cambridge Crystallographic Data Centre; <http://www.ccdc.cam.ac.uk/>].

### 3.3.3 Virtuelles *Screening* nach potentiellen Inhibitoren der EtMRK und der PfMRK

Für das Hochdurchsatz-*Docking* mit der Software PLANTS wurden zuvor Kinase-Inhibitoren aus der BindingDB selektiert, die einen IC<sub>50</sub>-Wert von unter 10 µM zeigten. Dieser Schritt wurde durchgeführt, um einen Testdatensatz geeigneter Inhibitoren humaner Kinasen zu erhalten. Diese Inhibitoren sollten für parasitäre Zielproteine optimiert werden können, da die ATP-Bindetaschen humaner und parasitärer Kinasen sehr ähnlich sind. Im Hochdurchsatz-*Docking* wurden insgesamt 2.139 Kinase-Inhibitoren der BindingDB Substanzdatenbank in die aktiven Zentren der Homologiemodelle EtMRK und PfMRK eingepasst und mittels der PLANTS-*Scoringfunktion* bewertet.

Mit dem Homologiemodell der EtMRK konnten im virtuellen *Screening* hauptsächlich Inhibitoren der humanen CDK7 und der humanen Protein Kinase B (PKB) angereichert werden. Zudem fand eine Anreicherung von Inhibitoren der humanen CDK1, CDK2 und CDK4 statt. In Abbildung 3.18 sind die Anreicherungsfaktoren (EF = *Enrichment Factor*) der bekannten Kinase-Inhibitoren in den oberen 1% (EF1), 2% (EF2) bzw. 5% (EF5) der finalen mit PLANTS-*Score* bewerteten Datenbank dargestellt. Im Fall der PfMRK konnten hauptsächlich Inhibitoren der *Epidermal-Growth-Factor-Receptor* Kinase (EGF-R), der 3-Phosphoinositide-abhängigen Protein Kinase 1 (PDK1) und der PKB in den oberen zwei Prozent der finalen Datenbank angereichert werden. In den oberen 5 Prozent der Datenbank fanden sich zudem auch Inhibitoren der Thymidin-Kinase und Ca<sup>2+</sup>/Calmodulin-dependent Protein Kinase (CAMK).

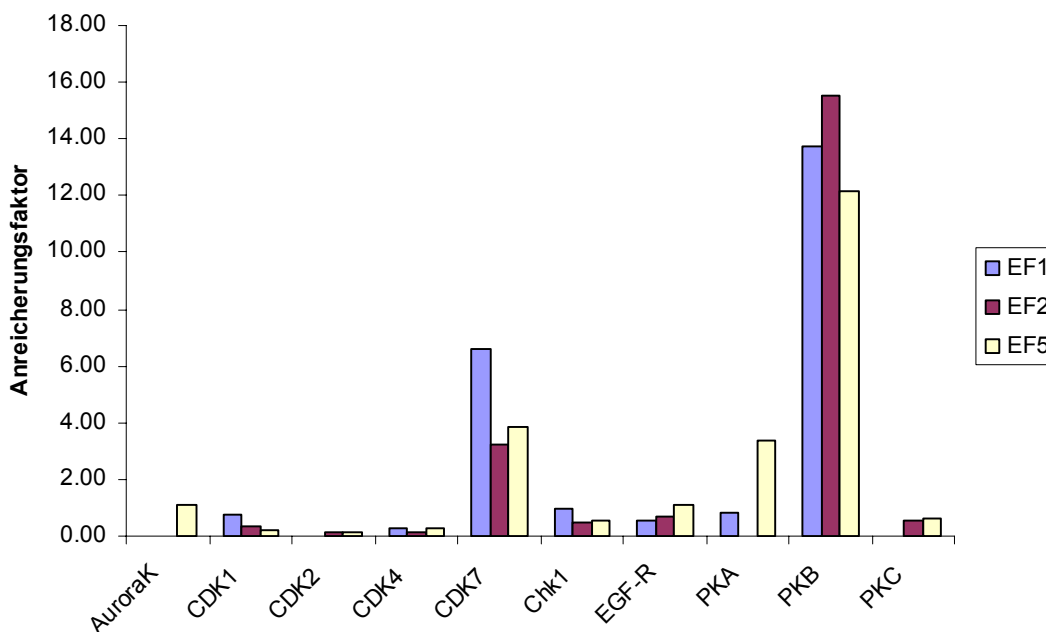


Abbildung 3.18: Anreicherungen der jeweiligen Kinase-Inhibitoren für die EtMRK.

Die Bewertungsfunktion PLANTS-Score wurde für das finale Ranking der Datenbank verwendet. Auf der Ordinate sind die jeweiligen Kinase-Inhibitoren aufgetragen. Die Abszisse bezeichnet die jeweiligen Anreicherungs-faktoren (EF), in den oberen 1% (EF1), 2% (EF2) und 5% (EF5) der finalen Datenbank.

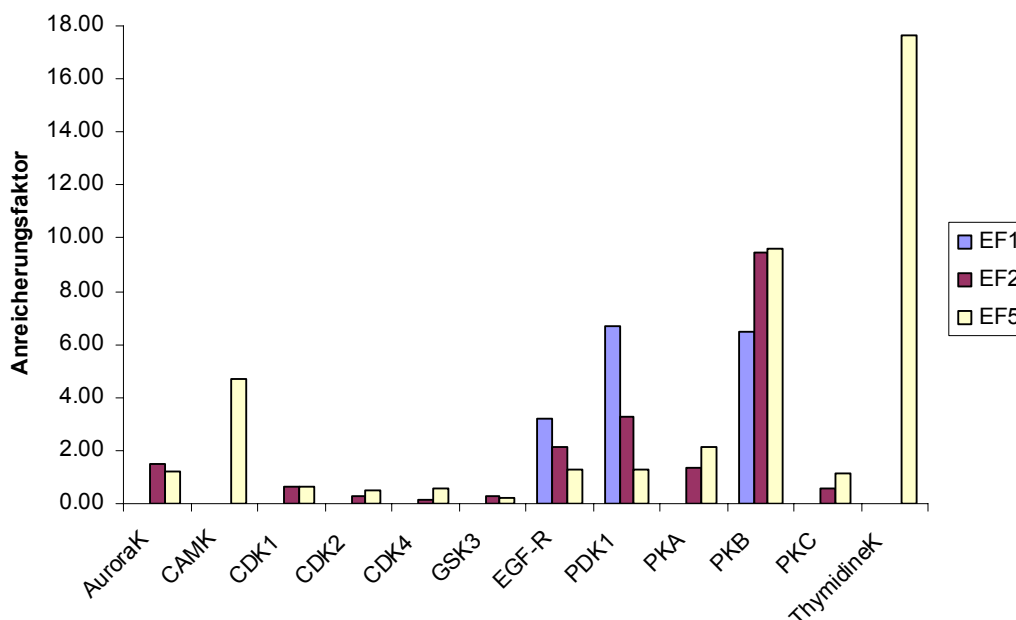


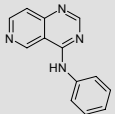
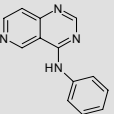
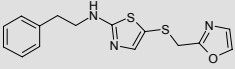
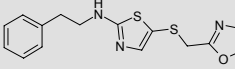
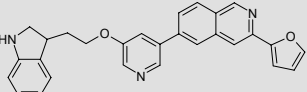
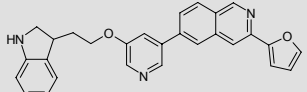
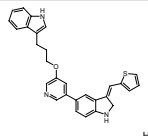
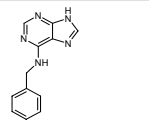
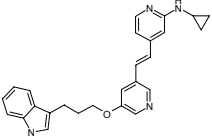
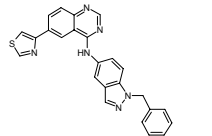
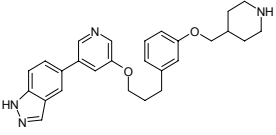
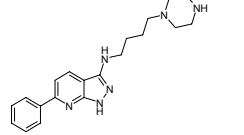
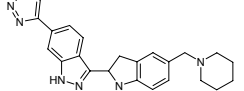
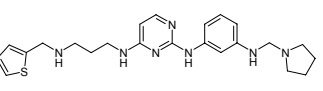
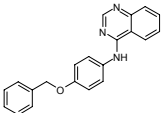
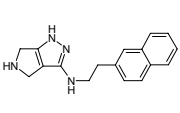
Abbildung 3.19: Anreicherungen für die PfMRK der jeweiligen Kinase-Inhibitoren.

Die Bewertungsfunktion PLANTS-Score wurde für das finale Ranking der Datenbank verwendet. Die Abszisse bezeichnet die jeweilige Anreicherung, in den oberen 1% (EF1), 2% (EF2) und 5% (EF5) der finalen Datenbank.



Um geeignete Strukturklassen als virtuelle *Hits* auszuwählen, wurde für die zwei Prozent der Strukturen mit dem höchsten PLANTS-Score diverse *Murcko Assemblies* [BEMIS & MURCKO, 1996] generiert. Die so selektierten Substanzen dienten als *Hit-Cluster* der parasitären Zielproteine, die gegebenenfalls im Anschluss noch weiter optimiert werden können, sofern sich ihre Aktivität *in vitro* bestätigen lässt. Für jedes Zielprotein wurden acht Strukturen ausgewählt. Drei dieser acht Strukturen wurden sowohl für die EtCRK2 als auch für die PfMRK identifiziert (siehe Tabelle 3.13).

Tabelle 3.13: *In silico Hits* der EtMRK und der PfMRK

EtMRK	PfMRK
	
	
	
	
	
	
	
	

Es konnten sieben *in silico Hits* der EtMRK sowie der PfMRK durch das virtuelle *Screening* identifiziert werden. Drei *in silico Hits* wurden für beide Zielproteine identifiziert und sind daher grau unterlegt.

## 4 Diskussion

### 4.1 Zyklin-abhängige Kinasen als Zielproteine der Therapie von Protozoen-Infektionen

Zyklin-abhängige Kinasen sind Zielproteine in der Bekämpfung von Krebserkrankungen, da sie eine Schlüsselfunktion in der Regulation des Zellzyklus besitzen [MALUMBRES & BARBACID, 2009]. Der reibungslose Ablauf des Zellzyklus ist für jede sich teilende Zelle – auch für Tumorzellen – essentiell [JOHANSSON & PERSSON, 2008]. In Tumorzellen ist der Zellzyklus dereguliert [BERTHET & KALDIS, 2007]. Die Folge ist ein unkontrolliertes Zellwachstum [DEL SAL, 1997]. Eine gezielte Inhibierung von Zellzyklus-relevanten CDKs unterbindet ein Fortschreiten der Tumorbildung und hat nur geringen Einfluss auf ruhendes – nicht proliferierendes Gewebe [BERTHET & KALDIS, 2007; MONACO & VALLANO, 2003]. Krebszellen und parasitisch lebende Protozoen ähneln einander [KLINKERT & HEUSSLER, 2006]. Beide besitzen eine vergleichbar hohe Zellteilungsrate [DOERIG, 2004]. Diese Gemeinsamkeit inspirierte die Auswahl von Zyklin-abhängigen Kinasen als Zielproteine in parasitisch lebenden Protozoen. Die Eignung dieser Zielproteine konnte in Protozoen bereits mehrfach gezeigt werden [DOERIG, 2004; HAMMARTON *et al.*, 2003; NAULA *et al.*, 2005]. Analog sollte die Inhibierung von CDKs aus *E. tenella* den Zellzyklus unterbinden und damit den Lebenszyklus des Parasiten unterbrechen. In *in vitro* Experimenten konnte dies nachgewiesen werden. Hierfür wurde der Effekt des bekannten CDK-Inhibitors Flavopiridol auf die Inhibierung der Schizontenentwicklung getestet (siehe Abschnitt 1.5.5, Tabelle 1.3) [BEYER, 2007]. Die Ergebnisse dieser Studie validieren CDK-ähnliche Proteine chemisch als Zielproteine in *E. tenella*.

### 4.2 Bioinformatische Identifizierung und Charakterisierung von Zyklin-abhängigen Kinasen in *E. tenella*

Eine Voraussetzung für das rationale Wirkstoffdesign von Inhibitoren Zyklin-abhängiger Kinasen des Parasiten *E. tenella* war die bioinformatische Identifizierung dieser Proteine. Im Vergleich zur Situation bei Säugetieren ergab die bioinformatische Analyse für *E. tenella* eine deutlich geringere Anzahl von CDK Proteinen. Es wurden neben der bereits bekannten

EtCRK2 drei weitere CDK-ähnliche Sequenzen identifiziert. Im Menschen sind elf CDK Proteine bekannt. Hiervon besitzen vier CDK Proteine eine wichtige Funktion in der Regulation des Zellzyklus (HsCDK1, HsCDK2, HsCDK4 und HsCDK6). Für Hefen ist bekannt, dass die Zellzyklusregulation nur durch eine einzige CDK (*cdc2p/CDC28* equivalent zur humanen CDK1) mit verschiedenen Zyklinpartnern gesteuert wird [MORGAN, 1997]. Es ist somit anzunehmen, dass die CDK Proteine in Invertebraten mehrere Aufgaben in der Zellzyklusregulation übernehmen, und dass bei einem Funktionsverlust gravierende Folgen für den Organismus entstehen. Durch Sequenzvergleiche konnten in *E. tenella* nur zwei CDK Proteine identifiziert werden (EtCRK2 und EtCRK3), die vermutlich eine Rolle im Zellzyklus spielen. Hierbei ist EtCRK2 homolog zur HsCDK1 sowie HsCDK2. Dagegen besitzt EtCRK3 die höchste Identität zur HsCDK1. Die Vermutung, dass die EtCRK2 eine zentrale Funktion im Zellzyklus von *E. tenella* besitzt, wird durch Versuche mit anderen CDK homologen Proteinen parasitärer Protozoen gestützt [DOERIG *et al.*, 2000]. Es konnte gezeigt werden, dass die Vermehrung der Parasiten *Plasmodium falciparum*, *Leishmania mexicana* und *Trypanosoma brucei* durch CDK-Inhibitoren, die für humane CDKs entwickelt wurden, unterbunden werden konnte [GRANT *et al.*, 2004; HARMSE *et al.*, 2001; WOODARD *et al.*, 2003].

Die BLAST-Suche in Datenbanken des Sanger Institutes, ergab die bislang nicht beschriebenen Sequenzen EtMRK, EtCRK1 und EtCRK3. Dass es sich bei den gefundenen Sequenzen tatsächlich um Zyklin-abhängige Kinasen handelt, bestätigt die hohe Sequenzidentität sowohl zu EtCRK2 als auch zu PfPK5 sowie zu anderen homologen CDK-Sequenzen von *P. falciparum*. Ein multiples Sequenz-Alignment mit der humanen CDK2 Sequenz zeigte außerdem, dass alle drei Sequenzen die konservierten Domänen *G-Loop*, *PSTAIRE-Helix*, *Hinge-Region* und *T-Loop* besitzen. Die Sequenzen waren für EtCRK3 sowie für EtCRK1 an beiden Termini unvollständig, so dass für beide Sequenzen keine vollständigen Proteinsequenzen abgeleitet werden konnten. Für die EtMRK konnte jedoch eine vollständige Proteinsequenz vorhergesagt werden. Sie wies eine große N-terminale Verlängerung von 43 Aminosäuren auf. Aufgrund der Tatsache, dass die homologen Proteinsequenzen von Organismen der *Apicomplexa* einen deutlich kürzeren N-Terminus besitzen [LI *et al.*, 1996] sowie das angenommene Startcodon nicht in einem Kozak-Motiv [LEWIN, 2008] enthalten ist, wurde ein Sequenzierungsfehler angenommen. Die Nukleotidsequenzanalyse ergab im Leserahmen +1 ein „ACG“-Triplet an Position 139 der Etmrk. Nimmt man „ATG“ anstatt „ACG“ an, erhält man ein Kozak-Konsensus-Motiv. Die vorhergesagte EtMRK Proteinsequenz resultierte so in einer vergleichbaren Länge und Aminosäureabfolge zu anderen homologen MRK-Sequenzen inner-

halb der *Apicomplexa*. Eine experimentelle Prüfung dieser Hypothese wurde nicht durchgeführt.

Eine Vorhersage der Funktion, der im Rahmen dieser Arbeit identifizierten CDK-ähnlichen Proteine von *E. tenella*, ist mittels Sequenzvergleich zu Sequenzen gut charakterisierter humaner CDKs nicht vollständig möglich. EtMRK besitzt eine hohe Sequenzidentität zur humanen CDK7 (HsCDK7). Daher wäre eine orthologe Funktion zur HsCDK7 denkbar. Für diese ist eine duale Funktion bestätigt worden [FISHER, 2005]. HsCDK7 ist zum einen an der Transkriptionsregulation der RNA-Polymerase II beteiligt und zum anderen besitzt sie eine Funktion als CDK-Aktivierungskinese (CAK). Die Sequenzanalyse der EtMRK lieferte Hinweise, die gegen eine zur HsCDK7 orthologe Funktion sprechen. Es konnte kein Kern-Lokalisierungssignal (NLS) [MEIER & BLOBEL, 1990] in der EtMRK Sequenz vorhergesagt werden. Proteine die an der Transkriptionsregulation beteiligt sind, müssen in den Kern transportiert werden und benötigen hierfür ein NLS, das den Transport in den Zellkern gewährleistet [LANGE *et al.*, 2009]. Eine CAK-Funktion der EtMRK ist aufgrund des Fehlens von CAK-Bindemotiven in den potentiellen Substraten EtCRK2 und EtCRK3 ebenfalls fraglich. Zur vollständigen Aktivierung der EtCRK2 sowie der EtCRK3 muss die Aminosäure Threonin im Aktivierungs-*Loop* dieser Kinasen phosphoryliert werden. Dies kann auch durch eine Autophosphorylierung erfolgen und wurde für die EtCRK2 bereits nachgewiesen [KINNAIRD *et al.*, 2004]. Daher wäre ein CAK-Enzym zur vollständigen Aktivierung nicht zwingend notwendig. Diese Vermutung wird dadurch unterstützt, dass weder die EtCRK2 noch die EtCRK3 ein CAK-Bindemotiv besitzen.

Die EtCRK3 Proteinsequenz ist der Proteinsequenz der HsCDK1 ähnlich. Die Inaktivierung der HsCDK1 wird durch Phosphorylierung an den Aminosäuren Thr14 sowie Tyr15 reguliert und ist für den CDK1/Zyclin-Komplex charakteristisch [MALUMBRES & BARBACID, 2005]. Während Tyrosin in allen CRK3-ähnlichen Proteinsequenzen der *Apicomplexa* konserviert vorliegt, liegt an der homologen Position des Thr14 ein Alanin vor. Dies lässt vermuten, dass die mögliche Inaktivierung durch einen bisher noch unbekanntem Mechanismus reguliert wird.

### 4.3 Bioinformatische Identifizierung und Charakterisierung von Zyklinen in *E. tenella*

Ein weiteres Ziel dieser Arbeit war die Identifizierung Zyklin-ähnlicher Sequenzen in *E. tenella*. Die Auswahl der Suchsequenzen für die BLAST-Suche nach nicht identifizierten Zyklinen, die möglicherweise in der Lage sind, die Zyklin-abhängige Kinase EtCRK2 aus *E. tenella* zu aktivieren, erfolgte auf Basis bekannter Zykline aus *P. falciparum* (PfCYC1, PfCYC2, PfCYC3 und PfCYC4) [MERCKX *et al.*, 2003]. Wie MERCKX *et al.* zeigen konnten, sind die Proteine PfCYC1 und PfCYC3 in der Lage, das PfPK5 Protein zu aktivieren [MERCKX *et al.*, 2003]. Analog zur Situation in *P. falciparum* ist EtCYC1 sowie EtCYC3 vermutlich ebenfalls geeignet EtCRK2 zu aktivieren. Darüber hinaus ist auch RINGO [FERBY *et al.*, 1999], ein Protein aus *Xenopus laevis*, das keine Sequenzähnlichkeit mit den Zyklinen besitzt, jedoch strukturelle Ähnlichkeiten aufweist, in der Lage, PfPK5 zu aktivieren [MERCKX *et al.*, 2003]. María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim) konnte zeigen, dass RINGO geeignet ist, die EtCRK2 zu aktivieren. Die mit der BLAST-Suche zu PfCYC1 und PfCYC3 gefundenen Zykline (EtCYC1, EtCYC3a sowie EtCYC3b), sollten aus diesem Grund ebenfalls in der Lage sein, die anderen CDK-homologen Proteine zu aktivieren. Dies konnte von María L. Suárez Fernández *in vitro* für die EtCYC3a gezeigt werden und bestätigt so die angenommene These [María L. Suárez Fernández, unpublizierte Daten]. *In vitro* Experimente für die EtCYC1 sowie die EtCYC3b stehen allerdings noch aus.

Die Sequenz-Konservierung der Zykline innerhalb der beiden Familien *Haemosporidae* (*P. falciparum*) und *Eimeriidae* (*E. tenella*) ist sehr gering. Daher führte erst die BLAST-Suche mit der homologen TgCYC1 Proteinsequenz [KVAAL *et al.*, 2002] aus *T. gondii* zur Identifizierung einer ähnlichen Zyklin-Sequenz in *E. tenella*. Diese Sequenz aus *T. gondii* besitzt jedoch interessanterweise eine hohe Sequenzähnlichkeit zum humanen Zyklin H [KVAAL *et al.*, 2002], das auch in der Lage ist, PfPK5 zu aktivieren [MERCKX *et al.*, 2003]. Darüber hinaus bindet Zyklin H an die HsCDK2, allerdings ohne diese zu aktivieren [ANDERSEN *et al.*, 1997]. Die Zyklin-Box von TgCYC1 wiederum besitzt die höchste Ähnlichkeit zu PfCYC1, welches ebenfalls eine hohe Sequenzähnlichkeit zum humanen Zyklin H besitzt [LE ROCH *et al.*, 2000]. Das in dieser Tiefenanalyse ermittelte neue, in der Literatur bislang nicht beschriebene, homologe Zyklin könnte somit tatsächlich eine Aktivierungsfunktion für die EtCRK2 besitzen.

Die mit der Suchsequenz PfCYC3 durchgeführte BLAST-Suche ergab zwei homologe Sequenzen in *E. tenella* (EtCYC3a und EtCYC3b). EtCYC3a, EtCYC3b sowie PfCYC3 besitzen die höchste Sequenzähnlichkeit zu TgCYC2 aus *T. gondii* [HAMMARTON *et al.*, 2003]. TgCYC2 zeigt keine Sequenzähnlichkeit zu den humanen G1-Phase regulierenden Zyklinen D und E, ist aber in der Lage, die im Zellzyklus relevante TgCRK3 zu aktivieren. TaCRK3 ist homolog zu EtCRK2. Daher liegt die Vermutung nahe, dass auch EtCYC3a als auch EtCYC3b geeignet sind, EtCRK2 zu aktivieren. Wie bereits beschrieben, konnte diese These für die EtCYC3b *in vitro* bestätigt werden.

Die Abfragesequenzen PfCYC2 und PfCYC4 sind bislang nicht detailliert beschrieben. Für PfCYC2 ist die genaue Funktion bislang nicht bekannt und es konnte auch keine homologe Sequenz im Genom von *E. tenella* identifiziert werden. Für PfCYC4 vermutet man eine Funktion als Aktivierungspartner der PfCRK1 in *P. falciparum*. In *E. tenella* konnte ein zu PfCRK1 homologes Protein nachgewiesen werden (EtCRK1). Analog zu *P. falciparum* kann daher vermutlich die EtCRK1 ebenfalls durch das Zyklin EtCYC4 aktiviert werden.

In eukaryotischen Systemen ist die Kontrolle des Zellzyklus sehr gut untersucht. Es ist bekannt, dass die Anzahl der Zykline wesentlich höher ist, als die Anzahl der katalytischen Untereinheiten [MALUMBRES & BARBACID, 2005; SATYANARAYANA & KALDIS, 2009]. Es wurde auch nachgewiesen, dass diese von verschiedenen Zyklinen aktiviert werden [HOCHEGGER *et al.*, 2008; MALUMBRES, 2005]. In *P. falciparum* hingegen wurde nur eine deutlich geringere Anzahl von Zyklinen sowie CDKs gefunden [DOERIG *et al.*, 2002; MERCKX *et al.*, 2003]. In dieser Arbeit ist es durch *in silico* Methoden gelungen, drei bislang nicht annotierte CDK-ähnliche Proteine und vier Zykline in *E. tenella* zu identifizieren. Es wurde somit eine vergleichbare Anzahl von CDKs sowie Aktivatoren wie in anderen Organismen der *Apicomplexa* identifiziert. Es ist denkbar, dass noch weitere homologe Proteine in *E. tenella* existieren, da mittels *Best-Best-Hit* bei einer BLAST-Suche nicht zwangsläufig alle Homologen identifiziert werden. Dies liegt darin begründet, dass der beste Treffer einer BLAST-Suche aufgrund des zugrunde liegenden mathematischen Modells und des heuristischen Ansatzes, nicht zwangsläufig auch die Sequenz mit der geringsten genetischen Distanz ist [KOSKI & GOLDING, 2001]. Im Speziellen ist eine Sequenzhomologie-Analyse bei den sehr diversen Zyklinen schwierig [MERCKX *et al.*, 2003], da sie generell eine sehr geringe Sequenzähnlichkeit zueinander aufweisen [MORGAN, 1997]. Die Vorhersage der genauen Funktion von CDKs sowie den Zyklinen von *E. tenella* sollte – ähnlich wie zum Teil für *P.*

*falciparum* bereits durchgeführt [DOERIG *et al.*, 2002; MERCKX *et al.*, 2003] – durch biochemische Experimente belegt werden.

#### 4.4 Strategien zur Hemmung von Kinasen

Für eine Hemmung von Kinasen bieten sich grundsätzlich vier verschiedene Strategien an: eine Blockierung der Substratbindung, die Verdrängung von ATP, eine allosterische Regulation sowie eine Irreversible Inhibierung [KUBINYI & MÜLLER, 2004]. Im ersten Fall muss die Ausbildung eines Protein-Protein-Kontaktes verhindert werden, da Kinasen andere Proteine als Substrate erkennen und binden [KLEBE, 2009]. Aufgrund der Größe der sich ausbildenden Wechselwirkungsflächen erscheint diese Strategie äußerst schwierig, jedoch sind einige Beispiele aus der Literatur bekannt, in denen Inhibitoren die Substratbindetasche adressieren [ORZAEZ *et al.*, 2009]. Der Vorteil einer Blockierung der Substratbindung liegt in der geringen Konservierung im Bereich der Substrat-Bindetasche, was es ermöglicht, Selektivität zu erreichen [KUBINYI & MÜLLER, 2004]. Eine Verdrängung des ATP aus seiner Bindetasche beispielsweise führt aufgrund der hohen Konservierung häufig zu Selektivitätsproblemen, und könnte somit umgangen werden [ORZAEZ *et al.*, 2009]. Geringfügige Unterschiede der ATP-Bindetaschen von CDKs der Vertebraten und der Protozoen (ausführlicher diskutiert in Abschnitt 4.5), erlauben jedoch die Entwicklung von selektiven ATP-kompetitiven Inhibitoren.

#### 4.5 Leitstrukturidentifizierung von Inhibitoren der EtCRK2

Für die Suche nach neuen Inhibitoren die das EtCRK2 Protein adressieren wurde bereits ein *virtuelles Screening* durchgeführt [BEYER, 2007]. Das von Beyer verwendete *Screening*-Protokoll ist ähnlich zu vielen anderen Studien [MUTHAS *et al.*, 2008; OELLIEN *et al.*, 2009; PEROLA, 2006]. Es umfasste eine 3D-Pharmakophorfilterung, molekulares *Docking* und *Scoring* sowie verschiedene Nachprozessierungsschritte. Hierbei konnten elf EtCRK2-Inhibitoren identifiziert werden [BEYER, 2007].

Ein typischer Schritt für das sich anschließende rationale Wirkstoffdesign ist die Anreicherung von *Hit-Clustern* [SCHNECKE & BOSTROM, 2006]. Hierzu wurden Substruktur- und Ähnlichkeitssuchen auf Basis dieser Inhibitoren durchgeführt. Die eingesetzte Methode basiert auf der Annahme, dass strukturell ähnliche Verbindungen auch eine ähnliche biologische Aktivität zeigen [QUINTUS *et al.*, 2009]. Als Schlüsseleigenschaft für die Substruktursuchen

wurde der *Scaffold* in den bereits aktiven Substanzen bestimmt, der an die *Hinge*-Region bindet und somit essentiell für die Inhibitorbindung ist. Durch *Docking*-Experimente wurden die so identifizierten Substanzen bewertet und ausgewählt. Für sechs von 139 virtuellen *Hits* konnte *in vitro* Aktivität bestätigt werden. Im kombinierten *in silico* und *in vitro Workflow* konnten insgesamt vier potentielle Leitstrukturen der EtCRK2 identifiziert werden. Die interessanteste Klasse der EtCRK2-Inhibitoren sind die Benzimidazol-Carbonitrile, da sie bislang in der Literatur nur von Badawey, Kappe und El Hawash [BADAWEY & KAPPE, 1999; EL HAWASH *et al.*, 1999] sowie in einem Patent [ARBISER & AMBLARD, 2006] in Zusammenhang mit Kinasen beschrieben wurden. Aus diesem Grund wurde diese *in silico* weiter optimiert und Strukturen für eine chemische Synthese vorgeschlagen. Alle anderen Strukturklassen (Naphtolactame, Rhodanin und Oxindole) sind schon längere Zeit als Inhibitoren der CDKs oder Glycogensynthasen (GSK) bekannt [BRAMSON *et al.*, 2001; GRAY *et al.*, 1999; LIU *et al.*, 2003]. Die Familien der CDK und der GSK sind eng miteinander verwandt [MANNING *et al.*, 2002], daher ist es naheliegend, dass ähnliche Strukturtypen in beiden Proteinfamilien eine Aktivität zeigen. Die *in silico* Methoden waren somit in der Lage, interessante Verbindungen zu identifizieren.

Keine der vier identifizierten Leitstrukturen der EtCRK2 zeigte Selektivität für die parasitäre Kinase. Die Ergebnisse der biochemischen Experimente zeigen jedoch deutliche Unterschiede in der Aktivität gegenüber HsCDK2 und EtCRK2. Die Leitstrukturen wiesen niedrigere IC50-Werte und Ki-Werte auf die HsCDK2 als auf die EtCRK2 auf (siehe z.B. Naphtolactam, Tabelle 3.7, Tabelle 3.8 im Abschnitt 3.2.3). Diese Ergebnisse geben einen Hinweis darauf, dass es signifikante Unterschiede zwischen der HsCDK2 sowie der GgCDK3 und parasitischen CDK2 Proteinen geben muss, und diese somit auch von Inhibitoren adressiert werden können.

Obwohl CDK Proteine von Säugetieren eine hohe Sequenzidentität zueinander aufweisen, konnten selektive Inhibitoren für einzelne CDK-Subklassen identifiziert werden [KITAGAWA *et al.*, 1993; MEIJER *et al.*, 1997]. Hierbei kann ein einzelner Aminosäureaustausch ausreichen, damit ein Inhibitor selektiv eine CDK-Subklasse adressiert [PRATT *et al.*, 2006]. Zum Beispiel konnte ein Aminosäureaustausch vom Lys89 zu Thr89 (HsCDK2 Nummerierung) zwischen HsCDK2 und HsCDK4 ausgenutzt werden, um eine Selektivität des CINK4 Inhibitors zu erreichen [PRATT *et al.*, 2006; SONI *et al.*, 2001]. An derselben Position findet sich zwischen der EtCRK2 und der GgCDK3 ebenfalls ein Aminosäureaustausch



vom Glu88 (EtCRK2) zu Lys89 (GgCDK3). Daher ist anzunehmen, dass dasselbe Rational auch für die Entwicklung von selektiven EtCRK2-Inhibitoren angewendet werden kann. Es wird vermutet, dass ein Ladungsaustausch zwischen dem negativ geladenen Glutamat und dem positiv geladenen Lysin adressiert werden kann, um die Selektivität gegen die EtCRK2 zu erhöhen. Hierfür sind weitere Optimierungszyklen notwendig, um einen selektiven EtCRK2-Inhibitor zu entwickeln, der diesen Ladungsaustausch adressiert.

#### 4.6 Bindemodi-Vorhersage und Optimierung der Benzimidazole

Um die Benzimidazol-Carbonitrile strukturbasiert weiter zu optimieren, war es zunächst notwendig, den Bindemodus dieser Strukturen aufzuklären. Zudem gibt der Bindemodus wichtige Hinweise darüber, wie die Aktivität sowie die Selektivität oder auch die Eigenschaften (ADME Profil) dieser Inhibitoren verbessert werden könnten. Durch *Docking*-Analysen konnte der Bindemodus der Benzimidazol-Carbonitrile mit einer Substitution an Position R2 (siehe Abschnitt 3.2.4) nicht eindeutig identifiziert werden. Hierbei wurden zwei mögliche Bindemodi identifiziert, mit vergleichbaren *Scoring*-Werten, die beide als bioaktive Konformation möglich erschienen. Im dem einen Bindemodus binden die Benzimidazol-Carbonitrile über den Nitril-Stickstoff an die *Hinge*-Region unter Ausbildung einer Wasserstoffbrücke, im anderen über den Carbonyl-Sauerstoff. Es ist bekannt, dass *Scoring*-Funktionen nur zur schnellen Bewertung von Liganden in einem virtuellen *Screening* dienen und daher die Ligandenbindung nur grob abschätzen. Sie besitzen meist nicht ausreichend negative Beiträge, die durch polare Nichtübereinstimmungen und unvorteilhafte Desolvatation sowie auch durch sterische Kollisionen zustande kommen [WARREN *et al.*, 2006]. Um zu untersuchen, wie der wahrscheinliche Bindemodus der Benzimidazol-Carbonitrile ist, wurden Berechnungen durchgeführt, die im Gegensatz zu *Docking*-Analysen auch den entropischen Beitrag der Bindungsenergie berücksichtigen [BARRIL & SOLIVA, 2006; DE AMORIM *et al.*, 2008]. Zur genauen Untersuchung des Bindeverhaltens wurden Freie-Energie-Simulationen, speziell die LIE-Methode [AQVIST & MARELIUS, 2001] eingesetzt, um auch die negativen Beiträge einer Ligandenbindung stärker zu berücksichtigen [AQVIST & MARELIUS, 2001; NERVALL *et al.*, 2008; STJERNSCHANTZ *et al.*, 2006].

LIE-Simulationen ergaben für drei Verbindungen der Benzimidazol-Carbonitrile eine leichte Bevorzugung für den Bindemodus, in dem die Liganden über die Nitrilgruppe zur *Hinge*-Region binden. Jedoch ist die Differenz der Bindungsenergien nicht in einer Größenordnung,

die es erlaubt, sicher eine Orientierung vorherzusagen. Für zwei Verbindungen konnte keine eindeutige Aussage getroffen werden, da die Unterschiede zwischen den Bindungsenergien der beiden Bindemodi in der Größenordnung des Standardfehlers liegen. Es ist denkbar, dass beide Bindemodi möglich sind. Eine genaue Aussage über den Bindemodus dieser Strukturen der Benzimidazol-Carbonitrile kann somit nur eine Aufklärung der Kristallstruktur des Protein-Ligand-Komplexes mit einer ausreichend hohen Auflösung liefern.

Die weitere Optimierung der Benzimidazol-Carbonitrile legte den Bindemodus fest. Alle optimierten Strukturen zeigten im *Docking* eine einheitliche Liganden-Pose. Diese Strukturen binden in den *Docking*-Experimenten über den Carbonyl-Stickstoff über eine Wasserstoffbrücke zur *Hinge*-Region.

#### 4.7 Homologiemodellierung der EtMRK und der PfMRK

Ausgangspunkt für ein virtuelles *Screening* nach Inhibitoren waren Homologiemodelle der EtMRK und der PfMRK, da keine Kristallstrukturen dieser Proteine in öffentlichen Datenbanken verfügbar waren. Eine Analyse der MRK Proteine verschiedener Organismen auf Sequenzebene ergab einen niedrigen bis mittleren Konservierungsgrad (zwischen 31% und 38% Sequenzidentität). Die Güte der generierten Modelle sollte mit diesen Sequenzidentitäten ausreichend für ein molekulares *Docking* sein [KRYSHTAFOVYCH *et al.*, 2007; MOULT *et al.*, 2001]. Die Identitätswerte sind zudem auf die Gesamtsequenzlänge bezogen. Betrachtet man die ATP-Bindetasche separat, ist die Konservierung deutlich höher (ca. 60% Sequenzidentität). Zudem liegen auch die Sequenzidentitäten der strukturell konservierten Regionen (SCRs = *Structurally Conserved Regions*) weit höher (ca. 81%).

Für ein virtuelles *Screening* mit dem *Docking*-Ansatz ist es wichtig, bei nicht vorhandener Kristallstruktur ein gutes Homologiemodell auf Basis einer geeigneten Templatstruktur zu generieren. Als Templat dienen Proteinstrukturen mit hoher Auflösung sowie niedrigen Temperaturfaktoren [QU *et al.*, 2009]. Idealerweise sind in den Templatstrukturen co-kristallisierte Inhibitoren vorhanden. Dadurch ist es möglich, die unterschiedlichen Interaktionen, die verschiedene Liganden mit der Templatstruktur eingehen können, abzuleiten und diese Informationen für das virtuelle *Screening* mit dem Homologiemodell zu nutzen. Zudem bewirkt eine co-Kristallisation eines Proteins mit einem Inhibitor, dass die Aminosäuren in der Bindetasche in Konformationen vorliegen, die für die Liganden-Bindung geeignet sind.

Die geeigneten Konformationen der Aminosäuren werden somit vom Templat auf das Homologiemodell übertragen.

Die Homologiemodellierung erfolgte unter Nutzung einer humanen CDK2-Templatstruktur (PDB-ID: 1OIR, [ANDERSON *et al.*, 2003]). Eine weitere in der PDB-Datenbank verzeichnete humane CDK7-Kristallstruktur (PDB-Acc.-Nr.: 1UA2) wurde nicht als Templatstruktur für die Homologiemodellbildung eingesetzt. Die HsCDK7 besitzt zwar eine höhere Sequenzidentität zu PfMRK und EtMRK, ist aber aufgrund der geringeren Auflösung (3 Å) wenig geeignet.

Für die Homologiemodellierung wurde mit der Struktur 1OIR eine inaktive Form der Kinase verwendet. Zwischen aktiver und inaktiver Form gibt es große strukturelle Unterschiede. Noble *et al.* gehen davon aus, dass die strukturelle Konservierung bei der inaktiven Form geringer ist als bei der aktiven Form [NOBLE *et al.*, 2004]. So kann beispielsweise die inaktive Form eines Enzyms durch einen allosterischen Mechanismus stabilisiert werden. Dies erleichtert die Entwicklung eines selektiven Inhibitors, indem dieser eine allosterische Region der Bindetasche adressiert. Ein Beispiel hierfür ist der Wirkstoff Imatinib (Gleevec) [DRUKER, 2002], der eine Konformationsänderung der Abl Kinase [AYOUB *et al.*, 2005] bewirkt und dadurch den inaktiven Zustand stabilisiert [NOBLE *et al.*, 2004]. Diese Konformationsänderung wird bei anderen Kinasen nicht herbeigeführt [KUBINYI & MÜLLER, 2004]. Jedoch können auch geringere Konformationsänderungen der Bindetasche einen energetisch günstigen Beitrag zur Ligandbindung beisteuern. Hierdurch wäre eine höhere Selektivität gegenüber anderen Kinasen möglich [JOHNSON, 2009]. Dadurch besitzt der Wirkstoff ein günstigeres toxikologisches Profil als andere, breit wirksame Proteinkinase-Inhibitoren [FABIAN 2005].

Die Güte des PfMRK Homologiemodelles konnte mittels *Docking* überprüft werden. Hierzu wurden bekannte PfMRK Oxindol-Inhibitoren mit der *Docking*-Software GOLD in die ATP-Bindetasche eingepasst. Hierbei wurden plausible Liganden-Posen generiert, was die Güte des generierten Homologiemodells bestätigt.

## 4.8 Validierung der Software PLANTS

Ein Grund für die Auswahl von PLANTS war die Effizienz dieser Software. Literaturbeispiele belegen, dass der implementierte *Docking*-Algorithmus in der Lage ist, eine Substanzdatenbank sehr schnell nach potentiellen Substanzen zu durchsuchen [KORB *et al.*, 2006;

KORB *et al.*, 2009]. Eine Validierung der Software PLANTS war aufgrund der Verfügbarkeit nur weniger Literaturbeispiele zu virtuellen *Screens* notwendig.

Zur Methodvalidierung des PLANTS-*Dockings* wurden verschiedene Versuche durchgeführt, die die Güte des verwendeten Programms überprüfen sollten. Als Referenzprogramm diente die *Docking*-Software GOLD, die schon oft hinsichtlich ihrer *Performance* getestet wurde und hierbei gute Ergebnisse erzielte [ONODERA *et al.*, 2007; SOUSA *et al.*, 2006]. Anhand von *Re-Docking* Experimenten wurde überprüft, ob die Software PLANTS ebenfalls gute *Docking*-Posen generieren konnte, d. h. die experimentell ermittelte Protein-Ligand-Konformation reproduziert. Hierbei zeigte sich, dass es sowohl für PLANTS als auch für GOLD möglich war, für die meisten der 15 Kristallstrukturen, die Liganden-Posen mit einem RMSD-Wert von  $< 2 \text{ \AA}$  zu ermitteln, obwohl *Docking*-Parameter verwendet wurden, die schnelle Berechnungen erlauben. Für ein virtuelles *Screening*, in dem große Datenbanken nach virtuellen *Hits* durchsucht werden, ist ein minimaler Rechenaufwand ein entscheidender Faktor. Der in PLANTS verwendete Algorithmus erlaubt ein effizienteres *Docking* und eine genauere Posen-Bewertung als GOLD. Dies wurde ebenfalls von anderen Autoren beschrieben [KORB *et al.*, 2006].

In zwei der 30 *Re-Docking*-Experimente konnte die Protein-Liganden-Konformation nicht reproduziert werden. Für das *Re-Docking* in die Kristallstruktur 1H01 war PLANTS nicht in der Lage, eine korrekte Pose zu generieren. Die Software GOLD reproduzierte für den Liganden der Kristallstruktur 1OIR keine korrekte Liganden-Pose. Ein möglicher Grund dafür könnten Fehler in den Proteinstrukturen sein [SOTRIFTER & DRAMBURG, 2005].

In nachfolgenden Experimenten wurden die internen Bewertungsfunktionen PLANTS-*Score* und GOLD-*Score* auf deren Anwendbarkeit für die korrekte Bewertung von Liganden-Posen untersucht. Die Bewertung der *Docking*-Lösungen anhand des berechneten *Scores* ergab eine gute Korrelation zwischen GOLD-*Score* und dem errechneten RMSD-Wert. Die *Scoring*-Funktion PLANTS-*Score* korrelierte jedoch deutlich besser mit den RMSD-Werten und konnte besser zwischen korrekten und falschen Liganden-Posen diskriminieren als die *Scoring*-Funktion GOLD-*Score*. Diese Diskriminierung ist für ein virtuelles *Screening* entscheidend, da hier oft nur die bestbewertete Pose weiter betrachtet wird [CROSS *et al.*, 2009].

Bei der Untersuchung des finalen *Rankings* von aktiven zu inaktiven Verbindungen, war die *Scoring*-Funktion PLANTS-*Score* auch besser in der Lage zwischen aktiven und inaktiven

Verbindungen zu diskriminieren als die *Scoring*-Funktion *GOLD-Score*. Alle diese Ergebnisse führten zu dem Schluss, dass die PLANTS Software für ein virtuelles *Screening* nach Inhibitoren Zyklin-abhängiger Kinasen valide ist. Die *Performance* der Software GOLD und PLANTS kann für andere Proteinklassen variieren.

#### 4.9 Virtuelles *Screening* nach EtMRK- und PfMRK-Inhibitoren

Nach Überprüfung der Eignung von CDK-ähnlichen Proteinen als Zielproteine zur Entwicklung neuer Antiparasitika, wurde mit Hilfe chemoinformatischer Methoden eine Suche nach potentiellen Inhibitoren in einer großen Substanzbibliothek durchgeführt. Für die Suche nach neuen Leitstrukturen wurde das virtuelle *Screening* mittels molekularen *Dockings* als Methode eingesetzt, welche häufig Anwendung findet [BÖHM & SCHNEIDER, 2000; CAVASOTTO & ORRY, 2007; MCINNES, 2007; OELLIEN *et al.*, 2009]. Für das virtuelle *Screening* nach potentiellen EtMRK- und PfMRK-Inhibitoren wurden Struktursuchen in einer Substanzdatenbank durchgeführt, die Verbindungen enthält, welche bereits eine Aktivität auf verschiedene Kinasen gezeigt haben [CHEN *et al.*, 2001; LIU *et al.*, 2007]. Die Auswahl dieser Kinase-fokussierten Bibliothek erhöht die Chance, EtMRK- sowie PfMRK-Inhibitoren zu identifizieren. Aufgrund der sehr ähnlichen ATP-Bindetaschen der EtMRK und der PfMRK, wurden drei Strukturen im virtuellen *Screening* identifiziert, die prinzipiell in der Lage sind beide Proteine zu inhibieren. Zudem ist anzunehmen, dass die Oxindol-basierten PfMRK-Inhibitoren auch die EtMRK inhibieren könnten, da auch zwei der identifizierten EtCRK2 Leitstrukturen auf das homologe Protein PfPK5 eine Aktivität zeigten (siehe Abschnitt 3.2.3).

#### 4.10 Ausblick

Die Sequenzsuche nach CDK sowie Zyklin homologen Proteinen von *E. tenella* ergab mehrere neue CDKs sowie Zykline. Im weiteren Projektverlauf wäre es wünschenswert alle identifizierten Proteine zu klonieren und zu exprimieren. Ein weiteres Anschlussexperiment wäre die Durchführung von Aktivierungs-*Assays* der CRKs mit den bekannten Zyklinen, um die jeweiligen Aktivierungspartner der CRKs von *E. tenella* zu identifizieren. Ein weiteres Ziel wäre eine umfassende biochemische Charakterisierung der CDK- sowie Zyklin-ähnlichen Proteine in *E. tenella*, um beispielsweise die Substrate der verschiedenen CDK-Zyklin-Komplexe zu identifizieren. Daraus ließe sich möglicherweise auf deren Funktion im Parasi-

ten schließen. Darüber hinaus könnte die Identifizierung weiterer, weniger verwandter CDKs von *E. tenella* mittels molekularbiologischer Methoden erfolgen.

Eine virtuelle Leitstrukturoptimierung der Benzimidazol-Carbonitrile wurde bereits durchgeführt. Im weiteren Projektverlauf sollte die chemische *Optimierung* für alle identifizierten Leitstrukturen (Benzimidazole, Benzimidazole-Carbonitrile, Rhodanin-Oxindole und Naphtolactame) durchgeführt werden, sowie diese optimierten Leitstrukturen wiederum experimentell, das heißt *in vitro* am Zielprotein getestet werden. Mit Hilfe der so gewonnenen Informationen über die Aktivität der einzelnen Strukturen, könnten dann die Pharmakophormodelle sowie die *Docking*-Parameter verfeinert werden sowie auch die LIE-Parameter angepasst werden.

Hypothesen über mögliche Bindungsmodi der Liganden in der ATP-Bindetasche können mit Hilfe von Homologiemodellen meist nur bedingt abgeleitet werden [CAVASOTTO & PHATAK, 2009]. Aus diesem Grund sind eine bzw. verschiedene EtCRK2-Inhibitor Komplexe zu empfehlen. Zudem wären auch Strukturaufklärungen der humanen CDK2 mit den identifizierten Leitstrukturen als co-kristallisierte Liganden wünschenswert. Diese Ergebnisse sollten wichtige Hinweise für die experimentelle Vorgehensweise liefern.

Darüber hinaus könnte eine chemische Optimierung der ermittelten Leitstrukturen erster Generation unter Berücksichtigung der Parameter Toxizität, Selektivität und Löslichkeit durchgeführt werden. Anschließend sollte wiederum ein *in vitro* Test der resultierenden Leitstrukturen zweiter Generation erfolgen. Dieser *in vitro* / *in silico* Optimierungszyklus muss solange durchlaufen werden, bis eine optimale Wirkung am Zielprotein erreicht ist. Es wäre denkbar, einen gleichartigen Optimierungszyklus anzuschließen, in dem die resultierenden Leitstrukturen in der Zellkultur am Parasiten getestet werden können. Dieser Zyklus sollte wiederum solange durchlaufen werden, bis optimale Wirkung am Parasiten erreicht ist. Der angestrebte Abschluss in diesem Projekt wäre ein Wirkstoffkandidat, der den späten Phasen der Medikamentenentwicklung zugeführt werden kann. Diese Phasen beinhalten unter anderem die zulasungsrelevanten Tierversuche der Entwicklung von Tierarzneimitteln.

Gleiches gilt für die identifizierten Strukturen aus dem initialen virtuellen *Screening* nach potentiellen EtMRK- und PfMRK-Inhibitoren. Diese Verbindungen müssten zunächst *in vitro* als *Hits* bestätigt werden. Dazu sollte das identifizierte EtMRK Protein kloniert und expriert werden. Erste Expressionsversuche mit diesem Protein werden in unserer Arbeitsgruppe

bereits durchgeführt. Funktionsstudien, wie z.B. transgene Überexpression oder *knock out* Mutagenese [LEWIN, 2008] an *E. tenella* sollten darüber hinaus klären, welche Rolle die EtCRK2 sowie die EtMRK im Zellzyklus von *E. tenella* spielt.

## 5 Zusammenfassung

Parasiten der *Apicomplexa* umfassen sowohl mehrere humanpathogene, als auch tierpathogene Protozoen. Beispiele für wichtige Vertreter human- und tierpathogener Parasiten sind *Plasmodium falciparum* und *Eimeria tenella*. Durch *P. falciparum* wird die tropische Infektionskrankheit *Malaria tropica* hervorgerufen, die jährlich ca. 2,7 Millionen Todesopfer fordert [MURPHY & BREMAN, 2001]. *E. tenella* verursacht Kokzidiose in Hühnern, die weltweit für Verluste in einer geschätzten Höhe zwischen 500 Millionen £ [SHIRLEY *et al.*, 2007] und unfassbaren 3 Milliarden US\$ [DALLOUL & LILLEHOJ, 2006] verantwortlich zeichnet. Eine prophylaktische Vakzinierung gegen diese Krankheit ist meist ökonomisch ineffektiv [KINNAIRD *et al.*, 2004; MCDONALD & SHIRLEY, 2009], und eine Behandlung mit Kokzidiostatika wird durch häufige Resistenzbildung gegen bekannte Wirkstoffe erschwert [KINNAIRD *et al.*, 2004; MEHLHORN *et al.*, 1993; MELENDEZ, 2000; SHIRLEY, 1989]. Dies erfordert die Entwicklung neuer preiswerter Alternativen. Geeignete Zielproteine zur Entwicklung neuartiger Arzneistoffe sind die Zyklin-abhängigen Kinasen (CDK), zu denen auch die CDK-related Kinase 2 (EtCRK2) aus *E. tenella* gehört. Diese Proteine sind maßgeblich an der Regulation des Zellzyklus beteiligt [NELSON & COX, 2001]. Durch chemische Validierung mit dem CDK-Inhibitor Flavopiridol konnte nachgewiesen werden, dass ein Funktionsverlust von CDKs in *E. tenella* die Vermehrung des Parasiten in Kultur inhibiert. CDKs in *E. tenella* sind daher als Zielproteine für die Entwicklung einer Chemotherapie der Kokzidiose geeignet.

Die in dieser Arbeit durchgeführten bioinformatischen Tiefenanalysen sollten das Vorkommen von CDK Proteinen im Parasiten *E. tenella* überprüfen. Das Genom von *E. tenella* liegt als ein erster Entwurf vor (*Draft Assembly*) [<ftp://ftp.sanger.ac.uk>]. Die Erstellung der Tiefenanalyse der homologen CDK Proteine erfolgte durch Sammlung und Bewertung von Sequenzinformationen von anderen Organismen des Stammes der *Apicomplexa*. Dazu wurden, neben der Abfrage biologischer Datenbanken sowie Literaturdatenbanken, auch eine Reihe weiterer bioinformatischer Methoden, wie paarweise und multiple Sequenzvergleiche (*Alignments*), angewandt [THOMPSON *et al.*, 1994]. Darüber hinaus wurde eine Analyse der entsprechenden Zykline - der Aktivatoren der CDKs - bezüglich Funktion und Struktur, sowie eine Datenbanksuche nach bisher nicht beschriebenen Zyklinen in *E. tenella* durchgeführt. Durch diese bioinformatischen Analysen konnten neben der bereits bekannten EtCRK2, drei weitere, bislang nicht annotierte CDKs in *E. tenella* identifiziert werden (EtCRK1, EtCRK3



sowie EtMRK). Weitere BLAST-Suchen ergaben vier neue potentielle Zyklone für *E. tenella*, wovon EtCYC3a als Aktivator der EtCRK2 von María L. Suárez Fernández (Intervet Innovation GmbH, Schwabenheim) bestätigt werden konnte. Bioinformatische Analysen lassen vermuten, dass auch EtCYC1 und EtCYC3b in der Lage sind EtCRK2 zu aktivieren. Außerdem ist anzunehmen, dass EtCYC4 als Aktivator der EtCRK1 fungiert.

Inhibitoren der EtCRK2, die in vorangegangenen Arbeiten gefunden wurden, konnten bereits in enzymatischen Tests bestätigt werden [BEYER, 2007]. Durch die Anwendung chemoinformatischer Methoden wie Substruktur- und Ähnlichkeitssuchen, konnten darüber hinaus weitere Inhibitoren der EtCRK2 identifiziert werden. In *in vitro* Experimenten konnte gezeigt werden, dass mittels *in silico* Methoden vier Substanzen als Leitstrukturen der EtCRK2 identifiziert werden konnten. Eine Leitstruktur aus der Strukturklasse der Benzimidazol-Carbonitrile ist bislang nicht als Inhibitor anderer Kinasen beschrieben. Diese neu identifizierte Leitstruktur konnte darüber hinaus *in silico* optimiert werden. Im Rahmen dieser Arbeiten wurden Bindungsenergien mittels der LIE-Methode für Substanzen der Benzimidazol-Carbonitrile berechnet, um dadurch einen wahrscheinlichen Bindemodus vorherzusagen. Dieser konnte jedoch nicht eindeutig durch die Freie-Energie Simulationen bestimmt werden. Durch *in silico* Optimierung dieser Substanzklasse wurde ein Bindemodus festgelegt. Für diese Optimierung wurde eine virtuelle kombinatorische Substanzbibliothek erstellt. Die Auswahl geeigneter Verbindungen erfolgte durch molekulares *Docking*. Es wurde eine finale Auswahl von 15 Verbindungen für eine chemische Synthese vorgeschlagen.

Durch ein virtuelles *Screening* mittels *Docking* nach potentiellen Inhibitoren der PfMRK und EtMRK, konnten interessante Verbindungen aus einer Substanzdatenbank ausgewählt werden. Als *Docking*-Software wurde PLANTS [KORB *et al.*, 2006] verwendet. Diese Software wurde zunächst validiert, indem PLANTS mit der etablierten *Docking*-Software GOLD [JONES *et al.*, 1995] verglichen wurde. In dieser erfolgreichen Validierung zeigte sich, dass das *Docking* mit PLANTS mit einem geringen Rechenaufwand durchzuführen ist und dieses in der Lage ist, zwischen CDK-Inhibitoren und *Decoys* zu diskriminieren. Durch das virtuelle *Screening* konnten jeweils sieben Verbindungen als virtuelle *Hits* der PfMRK sowie der EtMRK identifiziert werden. Die Häufung von Strukturklassen mit bekannter CDK-Aktivität deutet darauf hin, dass während des virtuellen *Screenings* eine Anreicherung von CDK-Inhibitoren stattgefunden hat. Die Ergebnisse des virtuellen *Screenings* stellen eine gute Ausgangslage für die Weiterentwicklung neuer Wirkstoffe gegen Kokzidiose und Malaria dar.

## 6 Literaturverzeichnis

Accelrys Inc., S. D. U.. Insight II, Modelling Environment. 2003.

Almlof, M., Brandsdal, B. O., and Aqvist, J., 2004, Binding affinity prediction with different force fields: examination of the linear interaction energy method: *J Comput. Chem*, **25**, 1242-1254.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J., 1990, Basic local alignment search tool: *J Mol Biol*, **215**, 403-410.

Andersen, G., Busso, D., Poterszman, A., Hwang, J. R., Wurtz, J. M., Ripp, R., Thierry, J. C., Egly, J. M., and Moras, D., 1997, The structure of cyclin H: common mode of kinase activation and specific features: *EMBO J*, **16**, 958-967.

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., and Jewsbury, P. J., 2003, Imidazo[1,2-a]pyridines: a potent and selective class of cyclin-dependent kinase inhibitors identified through structure-based hybridisation: *Bioorg. Med Chem Lett*, **13**, 3021-3026.

Aqvist, J., Luzhkov, V. B., and Brandsdal, B. O., 2002, Ligand binding affinities from MD simulations: *Acc. Chem Res*, **35**, 358-365.

Aqvist, J. and Marelius, J., 2001, The linear interaction energy method for predicting ligand binding free energies: *Comb Chem High Throughput Screen.*, **4**, 613-626.

Arbiser, J. L. and Amblard, F., 2006, Honokiol Derivatives for the Treatment of Proliferative Disorders: PTC/US2006/006494.

Arvieux, C. and Tribut, O., 2005, Amprenavir or fosamprenavir plus ritonavir in HIV infection: pharmacology, efficacy and tolerability profile: *Drugs*, **65**, 633-659.

Ayoub, W. S., Geller, S. A., Tran, T., Martin, P., Vierling, J. M., and Poordad, F. F., 2005, Imatinib (Gleevec)-induced hepatotoxicity: *J Clin Gastroenterol*, **39**, 75-77.

Badawey, E. A. and Kappe, T., 1999, Benzimidazole condensed ring systems. XI. Synthesis of some substituted cycloalkyl pyrido[1,2-a]benzimidazoles with anticipated antineoplastic activity: *Eur J Med Chem*, **34**, 663-667.

Bareschino, M. A., Schettino, C., Troiani, T., Martinelli, E., Morgillo, F., and Ciardiello, F., 2007, Erlotinib in cancer treatment: *Ann Oncol.*, **18**, 35-41.

Barril, X. and Soliva, R., 2006, Molecular modelling: *Mol Biosyst.*, **2**, 660-681.

Bastarrachea, R. A., Montero, J. C., Saavedra-Gajardo, I., Cerda-Flores, R., Machado-Dominguez, A., and Comuzzie, A. G., 2008, [Molecular targets for new drug discovery to treat type 2 diabetes and obesity]: *Rev Med Chil.*, **136**, 107-117.

- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R., 2004, The Pfam protein families database: *Nucleic Acids Res*, **32**, D138-D141.
- Bays, H. E., 2004, Current and investigational antiobesity agents and obesity therapeutic treatment targets: *Obes. Res*, **12**, 1197-1211.
- Bemis, G. W. and Murcko, M. A., 1996, The properties of known drugs. 1. Molecular frameworks: *J Med Chem*, **39**, 2887-2893.
- Benson, D., Lipman, D. J., and Ostell, J., 1993, GenBank: *Nucleic Acids Res*, **21**, 2963-2965.
- Berman, H., Henrick, K., and Nakamura, H., 2003, Announcing the worldwide Protein Data Bank: *Nat Struct. Biol*, **10**, 980.
- Berthet, C. and Kaldis, P., 2007, Cell-specific responses to loss of cyclin-dependent kinases: *Oncogene*, **26**, 4469-4477.
- Beyer, C., 2007, Etablierung einer Grid-Computing Umgebung und virtuelles Screening an Cyclin-abhängigen Kinasen von veterinär- und human-relevanten Parasiten: Dissertation im Fachbereich Chemie und Pharmazie der Eberhard-Karls-Universität Tübingen.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M., 2003, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003: *Nucleic Acids Res*, **31**, 365-370.
- Böhm, H. J., Klebe, G., and Kubinyi, H., 2002, *Wirkstoffdesign*: Spektrum Akad. Verlag, Heidelberg, Berlin, Oxford., **1**, 427-458.
- Böhm, H. J. and Schneider, G., 2000, *Virtual Screening for Bioactive Molecules*: Wiley-VCH Verlag, Weinheim.
- Bould, J. G., Elsheikha, H. M., and Morsy, T. A., 2009, Avian coccidiosis: the basic pathology to control: *J Egypt Soc Parasitol*, **39**, 85-98.
- Bracchi-Ricard, V., Barik, S., Delvecchio, C., Doerig, C., Chakrabarti, R., and Chakrabarti, D., 2000, PfPK6, a novel cyclin-dependent kinase/mitogen-activated protein kinase-related protein kinase from *Plasmodium falciparum*: *Biochem J*, **347**, 255-263.
- 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., and Kuyper, L. F., 2001, Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis: *J Med Chem*, **44**, 4339-4358.
- Branden, C. and Tooze, J., 1999, *Introduction to Protein Structure - Folding and Flexibility*: Garland Publishing Inc. , New York, USA, **2**, 89-120.

- Brave, M., Goodman, V., Kaminskas, E., Farrell, A., Timmer, W., Pope, S., Harapanhalli, R., Saber, H., Morse, D., Bullock, J., Men, A., Noory, C., Ramchandani, R., Kenna, L., Booth, B., Gobburu, J., Jiang, X., Sridhara, R., Justice, R., and Pazdur, R., 2008, Sprycel for chronic myeloid leukemia and Philadelphia chromosome-positive acute lymphoblastic leukemia resistant to or intolerant of imatinib mesylate: *Clin Cancer Res*, **14**, 352-359.
- Breitenlechner, C. B., Bossemeyer, D., and Engh, R. A., 2005, Crystallography for protein kinase drug design: PKA and SRC case studies: *Biochim. Biophys. Acta*, **1754**, 38-49.
- Brickmann, J., 1995, Interactive visualization of molecular scenarios with MOL-CAD/SYBYL. In: *Data visualization in molecular science: Tools for insight and innovation.*: Addison-Wesley Publ. Comp, Reading, Mass, 83-97.
- Brickmann, J., Keil, M., Exner, T. E., and Marhöfer, R. J., 2000, *Molecular Graphics: Trends and Perspectives.*: *J. Mol. Model.*, **6**, 328-340.
- Brophy, D. F., Israel, D. S., Pastor, A., Gillotin, C., Chittick, G. E., Symonds, W. T., Lou, Y., Sadler, B. M., and Polk, R. E., 2000, Pharmacokinetic interaction between amprenavir and clarithromycin in healthy male volunteers: *Antimicrob Agents Chemother*, **44**, 978-984.
- Buijsman, R., 2004, *Chemogenomics in Drug Discovery: Structural Aspects of Kinases and Their Inhibitors*: Wiley-VCH Verlag GmbH & Co. KGaA, **22**, 191-216.
- Burge, C. and Karlin, S., 1997, Prediction of complete gene structures in human genomic DNA: *J Mol Biol*, **268**, 78-94.
- Cai, X., Fuller, A. L., McDougald, L. R., and Zhu, G., 2003, Apicoplast genome of the coccidian *Eimeria tenella*: *Gene*, **321**, 39-46.
- Canduri, F., Perez, P. C., Caceres, R. A., and de, A. W., Jr., 2007, Protein kinases as targets for antiparasitic chemotherapy drugs: *Curr Drug Targets*, **8**, 389-398.
- Carlson, W., Karplus, M., and Haber, E., 1985, Construction of a model for the three-dimensional structure of human renal renin: *Hypertension*, **7**, 13-26.
- Cavasotto, C. N., Kovacs, J. A., and Abagyan, R. A., 2005, Representing receptor flexibility in ligand docking through relevant normal modes: *J Am Chem Soc*, **127**, 9632-9640.
- Cavasotto, C. N. and Orry, A. J., 2007, Ligand docking and structure-based virtual screening in drug discovery: *Curr Top Med Chem*, **7**, 1006-1014.
- Cavasotto, C. N. and Phatak, S. S., 2009, Homology modeling in drug discovery: current trends and applications: *Drug Discov Today*, **14**, 676-683.
- Chalmers, R. M. and Davies, A. P., 2009, Minireview: Clinical cryptosporidiosis: *Exp Parasitol*.
- Chen, X., Liu, M., and Gilson, M. K., 2001, BindingDB: a web-accessible molecular recognition database: *Comb Chem High Throughput Screen.*, **4**, 719-725.

- Chen, Y., Jirage, D., Caridha, D., Kathcart, A. K., Cortes, E. A., Denuff, R. A., Geyer, J. A., Prigge, S. T., and Waters, N. C., 2006, Identification of an effector protein and gain-of-function mutants that activate Pfmrk, a malarial cyclin-dependent protein kinase: *Mol Biochem Parasitol*, **149**, 48-57.
- Chong Wesley, K. M., Chu, S. S., Li, L., Duvadie, R. K., Yang, Y., and Xiao, W., 2003, Substituted 4-amino-thiazol-2-yl compounds as cyclin-dependent kinase inhibitors: US Patent 6569878.
- Chouhan, J. D., Zamarripa, D. E., Lai, P. H., Oramasionwu, C. U., and Grabinski, J. L., 2007, Sunitinib (Sutent): a novel agent for the treatment of metastatic renal cell carcinoma: *J Oncol. Pharm Pract*, **13**, 5-15.
- Clark, R. D., 2009, Prospective ligand- and target-based 3D QSAR: state of the art 2008: *Curr Top Med Chem*, **9**, 791-810.
- Cogolludo, A., Perez-Vizcaino, F., and Tamargo, J., 2005, New insights in the pharmacological therapy of arterial hypertension: *Curr Opin Nephrol Hypertens.*, **14**, 423-427.
- Cohen, F. E. and Sternberg, M. J., 1980, On the prediction of protein structure: The significance of the root-mean-square deviation: *J Mol Biol*, **138**, 321-333.
- Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y., and Humblet, C., 2009, Comparison of several molecular docking programs: pose prediction and virtual screening accuracy: *J Chem Inf Model*, **49**, 1455-1474.
- Curry, H. B., 1944, The method of steepest descent for nonlinear minimization problems.: *Quart. Appl. Math.*, **2**, 258-263.
- Dalloul, R. A. and Lillehoj, H. S., 2006, Poultry coccidiosis: recent advancements in control measures and vaccine development: *Expert Rev Vaccines.*, **5**, 143-163.
- Davies, T. G., Tunnah, P., Meijer, L., Marko, D., Eisenbrand, G., Endicott, J. A., and Noble, M. E., 2001, Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/indirubin-5-sulphonate: *Structure*, **9**, 389-397.
- de Amorim, H. L., Caceres, R. A., and Netz, P. A., 2008, Linear interaction energy (LIE) method in lead discovery and optimization: *Curr Drug Targets*, **9**, 1100-1105.
- Dean, P. M., 1994, Recent advances in drug design methods: where will they lead?: *Bioessays*, **16**, 683-687.
- Del Sal, G., 1997, The Mammalian cell cycle and its aberrations in cancer cells: *Adv. Clin Path.*, **1**, 123-136.
- DeRocher, A., Hagen, C. B., Froehlich, J. E., Feagin, J. E., and Parsons, M., 2000, Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system: *J Cell Sci*, **113**, 3969-3977.
- Dhaliwal, B. and Chen, Y. W., 2009, Computational Resources for Protein Modelling and Drug Discovery Applications: *Infect Disord. Drug Targets*.

- Doerig, C., 2004, Protein kinases as targets for anti-parasitic chemotherapy: *Biochim. Biophys. Acta*, **1697**, 155-168.
- Doerig, C., Chakrabarti, D., Kappes, B., and Matthews, K., 2000, The cell cycle in protozoan parasites: *Prog. Cell Cycle Res*, **4**, 163-183.
- Doerig, C., Doerig, C., Horrocks, P., Coyle, J., Carlton, J., Sultan, A., Arnot, D., and Carter, R., 1995, Pfcrk-1, a developmentally regulated cdc2-related protein kinase of *Plasmodium falciparum*: *Mol Biochem Parasitol*, **70**, 167-174.
- Doerig, C., Endicott, J., and Chakrabarti, D., 2002, Cyclin-dependent kinase homologues of *Plasmodium falciparum*: *Int J Parasitol*, **32**, 1575-1585.
- Dorigo, M., Maniezzo, V., and Colorni, A., 1996, Ant system: optimization by a colony of cooperating agents: *IEEE Trans Syst Man. Cybern. B Cybern.*, **26**, 29-41.
- Druker, B. J., 2002, STI571 (Gleevec) as a paradigm for cancer therapy: *Trends Mol Med*, **8**, S14-S18.
- Dubinina, G. G., Chupryna, O. O., Platonov, M. O., Borisko, P. O., Ostrovska, G. V., Tolmachov, A. O., and Shtil, A. A., 2007, In silico design of protein kinase inhibitors: successes and failures: *Anticancer Agents Med Chem*, **7**, 171-188.
- Duda, R. O., Hart, P. E., and Stork, D. G., 2001, *Pattern Classification.*: John Wiley and Sons., **2**.
- Eddy, S. R., 2004, What is a hidden Markov model?: *Nat Biotechnol.*, **22**, 1315-1316.
- Eglen, R. M. and Reisine, T., 2009, The current status of drug discovery against the human kinome: *Assay Drug Dev. Technol*, **7**, 22-43.
- el Hawash, S. A., Badawey, e., and Kappe, T., 1999, Benzimidazole condensed ring systems. XII. Synthesis and anticancer evaluation of certain pyrido[1,2-a]benzimidazole derivatives: *Pharmazie*, **54**, 341-346.
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P., 1997, Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes: *J Comput. Aided Mol Des*, **11**, 425-445.
- Elledge, S. J., 1996, Cell cycle checkpoints: preventing an identity crisis: *Science*, **274**, 1664-1672.
- EMA, 2009, <http://www.emea.europa.eu/>.
- Endicott, J. A., Noble, M. E., and Tucker, J. A., 1999, Cyclin-dependent kinases: inhibition and substrate recognition: *Curr Opin Struct. Biol*, **9**, 738-744.
- Engels, K., 2005, Trehalose-6-Phosphat synthase als Zielprotein zur Entwicklung eines Antiparasitikums.: Diplomarbeit im Fachbereich Verfahrenstechnik, Fachhochschule Bingen..

- English, N. J., 2007, Calculation of binding affinities of HIV-1 RT and beta-secretase inhibitors using the linear interaction energy method with explicit and continuum solvation approaches: *J Mol Model*, **13**, 1081-1097.
- 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., and Lockhart, D. J., 2005, A small molecule-kinase interaction map for clinical kinase inhibitors: *Nat Biotechnol.*, **23**, 329-336.
- Ferby, I., Blazquez, M., Palmer, A., Eritja, R., and Nebreda, A. R., 1999, A novel p34(cdc2)-binding and activating protein that is necessary and sufficient to trigger G(2)/M progression in *Xenopus* oocytes: *Genes Dev.*, **13**, 2177-2189.
- Fichera, M. E. and Roos, D. S., 1997, A plastid organelle as a drug target in apicomplexan parasites: *Nature*, **390**, 407-409.
- Fischer, E., 1894, Einfluss der Configuration auf die Wirkung der Enzyme.: *Ber. Dtsch. Chem. Ges.*, **27**, 2993.
- Fisher, R. P., 2005, Secrets of a double agent: CDK7 in cell-cycle control and transcription: *J Cell Sci*, **118**, 5171-5180.
- Foloppe, N. and Hubbard, R., 2006, Towards predictive ligand design with free-energy based computational methods?: *Curr Med Chem*, **13**, 3583-3608.
- Forster, M. J., 2002, Molecular modelling in structural biology: *Micron.*, **33**, 365-384.
- Foth, B. J. and McFadden, G. I., 2003, The apicoplast: a plastid in *Plasmodium falciparum* and other Apicomplexan parasites: *Int Rev Cytol*, **224**, 57-110.
- Frampton, J. E., 2009, Lapatinib: a review of its use in the treatment of HER2-overexpressing, trastuzumab-refractory, advanced or metastatic breast cancer: *Drugs*, **69**, 2125-2148.
- Gao, H. X. and Peng, L. M., 1999, Parameterization of the temperature dependence of the Debye-Waller factors: *Acta Crystallogr. A*, **55**, 926-932.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B., 2002, Genome sequence of the human malaria parasite *Plasmodium falciparum*: *Nature*, **419**, 498-511.
- Geyer, J. A., Prigge, S. T., and Waters, N. C., 2005, Targeting malaria with specific CDK inhibitors: *Biochim. Biophys. Acta*, **1754**, 160-170.

- Gibson, A. E., Arris, C. E., Bentley, J., Boyle, F. T., Curtin, N. J., Davies, T. G., Endicott, J. A., Golding, B. T., Grant, S., Griffin, R. J., Jewsbury, P., Johnson, L. N., Mesguiche, V., Newell, D. R., Noble, M. E., Tucker, J. A., and Whitfield, H. J., 2002, Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O(6)-substituted guanine derivatives: *J Med Chem*, **45**, 3381-3393.
- Gorre, M. E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P. N., and Sawyers, C. L., 2001, Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification: *Science*, **293**, 876-880.
- Grant, K. M., Dunion, M. H., Yardley, V., Skaltsounis, A. L., Marko, D., Eisenbrand, G., Croft, S. L., Meijer, L., and Mottram, J. C., 2004, Inhibitors of *Leishmania mexicana* CRK3 cyclin-dependent kinase: chemical library screen and antileishmanial activity: *Antimicrob Agents Chemother*, **48**, 3033-3042.
- Gray, N., Detivaud, L., Doerig, C., and Meijer, L., 1999, ATP-site directed inhibitors of cyclin-dependent kinases: *Curr Med Chem*, **6**, 859-875.
- Halgren, T. A., 1996, Merck Molecular Force Field. I-V.: *J. Comput. Chem.*, **17**, 490-641.
- Hall, T. A., 2009, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.: *Nucl. Acids. Symp. Ser.*, 95-98.
- Hammarton, T. C., Mottram, J. C., and Doerig, C., 2003, The cell cycle of parasitic protozoa: potential for chemotherapeutic exploitation: *Prog. Cell Cycle Res*, **5**, 91-101.
- Hanks, S. K. and Hunter, T., 1995, Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification: *FASEB J*, **9**, 576-596.
- Harmse, L., van Zyl, R., Gray, N., Schultz, P., Leclerc, S., Meijer, L., Doerig, C., and Havlik, I., 2001, Structure-activity relationships and inhibitory effects of various purine derivatives on the in vitro growth of *Plasmodium falciparum*: *Biochem Pharmacol*, **62**, 341-348.
- Hassan, P., Fergusson, D., Grant, K. M., and Mottram, J. C., 2001, The CRK3 protein kinase is essential for cell cycle progression of *Leishmania mexicana*: *Mol Biochem Parasitol*, **113**, 189-198.
- Hauschild, A., 2007, Clinical studies with sorafenib (Nexavar) in metastatic melanoma: *J Dtsch. Dermatol Ges*, **5**, 345.
- Henikoff, S. and Henikoff, J. G., 1993, Performance evaluation of amino acid substitution matrices: *Proteins*, **17**, 49-61.
- Hochegger, H., Takeda, S., and Hunt, T., 2008, Cyclin-dependent kinases and cell-cycle transitions: does one fit all?: *Nat Rev Mol Cell Biol*, **9**, 910-916.
- Holm, M., Lehmann, F., and Laufer, S., 2008, [Medical chemistry and molecular mechanisms of tyrosine kinase inhibitors. The viewpoint of pharmaceutical research]: *Pharm Unserer Zeit*, **37**, 382-392.



- Huang, N., Shoichet, B. K., and Irwin, J. J., 2006, Benchmarking sets for molecular docking: *J Med Chem*, **49**, 6789-6801.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Pagni, M., and Sigrist, C. J., 2006, The PROSITE database: *Nucleic Acids Res*, **34**, D227-D230.
- IHGSC, 2004, Finishing the euchromatic sequence of the human genome: *Nature*, **431**, 931-945.
- Ihlenfeldt, W. D., Voigt, J. H., Bienfait, B., Oellien, F., and Nicklaus, M. C., 2002, Enhanced CACTVS browser of the Open NCI Database: *J Chem Inf Comput. Sci*, **42**, 46-57.
- Iqbal, K., Alonso, A. C., El Akkad, E., Gong, C. X., Haque, N., Khatoon, S., Pei, J. J., Tanimukai, H., Tsujio, I., Wang, J. Z., and Grundke-Iqba, I., 2003, Alzheimer neurofibrillary degeneration: therapeutic targets and high-throughput assays: *J Mol Neurosci*, **20**, 425-429.
- Jaccard, P., 1912, The distribution of flora in the alpine zone: *New Phytologist*, **11**, 37-50.
- Jain, A. N., 1996, Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities: *J Comput. Aided Mol Des*, **10**, 427-440.
- Jautelat, R., Brumby, T., Schafer, M., Briem, H., Eisenbrand, G., Schwahn, S., Kruger, M., Lucking, U., Prien, O., and Siemeister, G., 2005, From the insoluble dye indirubin towards highly active, soluble CDK2-inhibitors: *Chembiochem*, **6**, 531-540.
- Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N. P., 1995, Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex: *Nature*, **376**, 313-320.
- Jiang, Z., 2008, Computational analysis of the interaction between ligand-receptor pairs: *Curr Pharm Des*, **14**, 588-592.
- Johansson, M. and Persson, J. L., 2008, Cancer therapy: targeting cell cycle regulators: *Anti-cancer Agents Med Chem*, **8**, 723-731.
- Johnson, D. G. and Walker, C. L., 1999, Cyclins and cell cycle checkpoints: *Annu. Rev Pharmacol Toxicol*, **39**, 295-312.
- Johnson, L. N., 2009, Protein kinase inhibitors: contributions from structure to clinical compounds: *Q. Rev Biophys.*, **42**, 1-40.
- Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D., and Beck, E., 1999, Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs: *Science*, **285**, 1573-1576.
- Jones, G., Willett, P., and Glen, R. C., 1995, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation: *J Mol Biol*, **245**, 43-53.

- Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R., 1997, Development and validation of a genetic algorithm for flexible docking: *J Mol Biol*, **267**, 727-748.
- Kaldis, P., 1999, The cdk-activating kinase (CAK): from yeast to mammals: *Cell Mol Life Sci*, **55**, 284-296.
- Kanter, P., 2005, Eine Evaluation der Genauigkeit verschiedener Verfahren der Genvorhersage in groen DNA-Sequenzen.: [http://theorie.informatik.uni-ulm.de/Lehre/SS5/GV/Ausarbeitung\\_Kanter.pdf](http://theorie.informatik.uni-ulm.de/Lehre/SS5/GV/Ausarbeitung_Kanter.pdf).
- Kellenberger, E., Rodrigo, J., Muller, P., and Rognan, D., 2004, Comparative evaluation of eight docking tools for docking and virtual screening accuracy: *Proteins*, **57**, 225-242.
- Khire, U. R., Bankston, D., Barbosa, J., Brittelli, D. R., Caringal, Y., Carlson, R., Dumas, J., Gane, T., Heald, S. L., Hibner, B., Johnson, J. S., Katz, M. E., Kennure, N., Kingery-Wood, J., Lee, W., Liu, X. G., Lowinger, T. B., McAlexander, I., Monahan, M. K., Natero, R., Renick, J., Riedl, B., Rong, H., Sibley, R. N., Smith, R. A., and Wolanin, D., 2004, Omega-carboxypyridyl substituted ureas as Raf kinase inhibitors: SAR of the amide substituent: *Bioorg. Med Chem Lett*, **14**, 783-786.
- Kim, H. Y. and Cho, Y., 1997, Structural similarity between the pocket region of retinoblastoma tumour suppressor and the cyclin-box: *Nat Struct. Biol*, **4**, 390-395.
- Kinnaird, J., Logan, M., Tait, A., and Langsley, G., 2001, TaCRK3 encodes a novel *Theileria annulata* protein kinase with motifs characteristic of the family of eukaryotic cyclin dependent kinases: a comparative analysis of its expression with TaCRK2 during the parasite life cycle: *Gene*, **279**, 127-135.
- Kinnaird, J. H., Bumstead, J. M., Mann, D. J., Ryan, R., Shirley, M. W., Shiels, B. R., and Tomley, F. M., 2004, EtCRK2, a cyclin-dependent kinase gene expressed during the sexual and asexual phases of the *Eimeria tenella* life cycle: *Int. J Parasitol.*, **34**, 683-692.
- Kitagawa, M., Okabe, T., Ogino, H., Matsumoto, H., Suzuki-Takahashi, I., Kokubo, T., Higashi, H., Saitoh, S., Taya, Y., Yasuda, H., and ., 1993, Butyrolactone I, a selective inhibitor of cdk2 and cdc2 kinase: *Oncogene*, **8**, 2425-2432.
- Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J., 2004, Docking and scoring in virtual screening for drug discovery: methods and applications: *Nat Rev Drug Discov*, **3**, 935-949.
- Klebe, G., 2006, Virtual ligand screening: strategies, perspectives and limitations: *Drug Discov Today*, **11**, 580-594.
- Klebe, G., 2009, Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen: Spektrum Akad. Verlag, Heidelberg, Berlin, Oxford., **2**, 427-458.
- Klinkert, M. Q. and Heussler, V., 2006, The use of anticancer drugs in antiparasitic chemotherapy: *Mini Rev Med Chem*, **6**, 131-143.
- Knighton, D. R., Xuong, N. H., Taylor, S. S., and Sowadski, J. M., 1991, Crystallization studies of cAMP-dependent protein kinase. Cocrystals of the catalytic subunit with a 20

- amino acid residue peptide inhibitor and MgATP diffract to 3.0 Å resolution: *J Mol Biol*, **220**, 217-220.
- Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., el Rifai, W., Hemmer, S., Wasenius, V. M., Vidgren, V., and Zhu, Y., 1998, DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies: *Am J Pathol*, **152**, 1107-1123.
- Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J. P., Gannon, J., and Hunt, T., 1992, Identification of the domains in cyclin A required for binding to, and activation of, p34cdc2 and p32cdk2 protein kinase subunits: *Mol Biol Cell*, **3**, 1279-1294.
- Kolb, P., Huang, D., Dey, F., and Caflisch, A., 2008, Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model: *J Med Chem*, **51**, 1179-1188.
- Korb, O., Stützle, T., and Exner, T. E., 2006, Swarm Intelligence. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design: Springer Verlag, Heidelberg, Berlin, New York, **1**, 115-134.
- Korb, O., Stutzle, T., and Exner, T. E., 2009, Empirical scoring functions for advanced protein-ligand docking with PLANTS: *J Chem Inf Model*, **49**, 84-96.
- Korf, I., 2003, Serial BLAST searching: *Bioinformatics*, **19**, 1492-1496.
- Koros, E. and Dorner-Ciossek, C., 2007, The role of glycogen synthase kinase-3beta in schizophrenia: *Drug News Perspect.*, **20**, 437-445.
- Koski, L. B. and Golding, G. B., 2001, The closest BLAST hit is often not the nearest neighbor: *J Mol Evol*, **52**, 540-542.
- Krebs, E. G. and Beavo, J. A., 1979, Phosphorylation-dephosphorylation of enzymes: *Annu. Rev Biochem*, **48**, 923-959.
- Kroemer, R. T., 2007, Structure-based drug design: docking and scoring: *Curr Protein Pept. Sci*, **8**, 312-328.
- Kryshtafovych, A., Prlic, A., Dmytriv, Z., Daniluk, P., Milostan, M., Eyrich, V., Hubbard, T., and Fidelis, K., 2007, New tools and expanded data analysis capabilities at the Protein Structure Prediction Center: *Proteins*, **69**, 19-26.
- Kubinyi, H. and Müller, G., 2004, Chemogenomics in Drug Discovery. A Medicinal Chemistry Perspective: Wiley-VCH Verlag, Weinheim, **1**, 191-219.
- Kummerfeld, N., 2005, Krankheiten der Heimtiere. Kapitel 18: Hühnervogel: Schlütersche Verlagsgesellschaft mbH & Co. KG, **6**, 613-651.
- Kvaal, C. A., Radke, J. R., Guerini, M. N., and White, M. W., 2002, Isolation of a *Toxoplasma gondii* cyclin by yeast two-hybrid interactive screen: *Mol Biochem Parasitol*, **120**, 187-194.

- Lange, A., McLane, L. M., Mills, R. E., Devine, S. E., and Corbett, A. H., 2009, Expanding the Definition of the Classical Bipartite Nuclear Localization Signal: Traffic..
- Le Roch, K., Sestier, C., Dorin, D., Waters, N., Kappes, B., Chakrabarti, D., Meijer, L., and Doerig, C., 2000, 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, **275**, 8952-8958.
- Leach, A. R., Shoichet, B. K., and Peishoff, C. E., 2006, Prediction of protein-ligand interactions. Docking and scoring: successes and gaps: J Med Chem, **49**, 5851-5855.
- Lees, E. M. and Harlow, E., 1993, Sequences within the conserved cyclin box of human cyclin A are sufficient for binding to and activation of cdc2 kinase: Mol Cell Biol, **13**, 1194-1201.
- Lewin, B., 2008, Genes IX: Jones and Bartlett Publishers, Inc. , London, UK.
- Li, J. L., Robson, K. J., Chen, J. L., Targett, G. A., and Baker, D. A., 1996, Pfmrk, a MO15-related protein kinase from Plasmodium falciparum. Gene cloning, sequence, stage-specific expression and chromosome localization: Eur J Biochem, **241**, 805-813.
- Lindberg, R. A., Quinn, A. M., and Hunter, T., 1992, Dual-specificity protein kinases: will any hydroxyl do?: Trends Biochem Sci, **17**, 114-119.
- Liu, J. J., Dermatakis, A., Lukacs, C., Konzelmann, F., Chen, Y., Kammlott, U., Depinto, W., Yang, H., Yin, X., Chen, Y., Schutt, A., Simcox, M. E., and Luk, K. C., 2003, 3,5,6-Trisubstituted naphthostyrils as CDK2 inhibitors: Bioorg. Med Chem Lett, **13**, 2465-2468.
- Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K., 2007, BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities: Nucleic Acids Res, **35**, D198-D201.
- Lucius, R. and Loos-Frank, B., 2008, Biologie von Parasiten.: Springer Verlag, Berlin, Heidelberg, New York, 107-256.
- Macdonald, L., 2000, New influenza drugs zanamivir (Relenza) and oseltamivir (Tamiflu): unexpected serious reactions: CMAJ., **163**, 879-5.
- Malizzia, L. J. and Hsu, A., 2008, Temsirolimus, an mTOR inhibitor for treatment of patients with advanced renal cell carcinoma: Clin J Oncol. Nurs., **12**, 639-646.
- Malumbres, M., 2005, Revisiting the "Cdk-centric" view of the mammalian cell cycle: Cell Cycle, **4**, 206-210.
- Malumbres, M. and Barbacid, M., 2005, Mammalian cyclin-dependent kinases: Trends Biochem Sci, **30**, 630-641.
- Malumbres, M. and Barbacid, M., 2009, Cell cycle, CDKs and cancer: a changing paradigm: Nat Rev Cancer, **9**, 153-166.

- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S., 2002, The protein kinase complement of the human genome: *Science*, **298**, 1912-1934.
- McAllister, M. M., 2005, A decade of discoveries in veterinary protozoology changes our concept of "subclinical" toxoplasmosis: *Vet Parasitol*, **132**, 241-247.
- McDonald, V. and Shirley, M. W., 2009, Past and future: vaccination against *Eimeria*: *Parasitology*, **136**, 1477-1489.
- McInnes, C., 2007, Virtual screening strategies in drug discovery: *Curr Opin Chem Biol*, **11**, 494-502.
- McInnes, C. and Fischer, P. M., 2005, Strategies for the design of potent and selective kinase inhibitors: *Curr Pharm Des*, **11**, 1845-1863.
- Mehlhorn, H., 2001, *Encyclopedic reference of parasitology*: pringer Verlag, Berlin, Heidelberg, New York.
- Mehlhorn, H., 2008, *Encyclopedia of parasitology*: Springer Verlag, Berlin, Heidelberg, New York.
- Mehlhorn, H., Düwel, D., and Raether, W., 1993, *Diagnose und Therapie der Parasitosen von Haus-, Nutz- und Heimtieren*: Gustav Fischer Verlag, Stuttgart, Jena, New York, 303-368.
- Meier, U. T. and Blobel, G., 1990, A nuclear localization signal binding protein in the nucleolus: *J Cell Biol*, **111**, 2235-2245.
- Meijer, L., Borgne, A., Mulner, O., Chong, J. P., Blow, J. J., Inagaki, N., Inagaki, M., Delcros, J. G., and Moulinoux, J. P., 1997, Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5: *Eur J Biochem*, **243**, 527-536.
- Melendez, R. D., 2000, Future perspectives on veterinary hemoparasite research in the tropics at the start of this century: *Ann N Y Acad Sci*, **916**, 253-258.
- 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., and Doerig, C., 2003, Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite *Plasmodium falciparum*: *J Biol Chem*, **278**, 39839-39850.
- Monaco, E. A. and Vallano, M. L., 2003, Cyclin-dependent kinase inhibitors: cancer killers to neuronal guardians: *Curr Med Chem*, **10**, 367-379.
- Montoya, J. G. and Liesenfeld, O., 2004, Toxoplasmosis: *Lancet*, **363**, 1965-1976.
- Morgan, D. O., 1997, Cyclin-dependent kinases: engines, clocks, and microprocessors: *Annu. Rev Cell Dev. Biol*, **13**, 261-291.

- Mottram, J. C., McCready, B. P., Brown, K. G., and Grant, K. M., 1996, Gene disruptions indicate an essential function for the LmmCRK1 cdc2-related kinase of *Leishmania mexicana*: *Mol Microbiol*, **22**, 573-583.
- Moult, J., Fidelis, K., Zemla, A., and Hubbard, T., 2001, Critical assessment of methods of protein structure prediction (CASP): round IV: *Proteins*, **5**, 2-7.
- Mount, S. M., 1982, A catalogue of splice junction sequences: *Nucleic Acids Res*, **10**, 459-472.
- 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., and Wu, C. H., 2005, InterPro, progress and status in 2005: *Nucleic Acids Res*, **33**, D201-D205.
- Murphy, S. C. and Breman, J. G., 2001, Gaps in the childhood malaria burden in Africa: cerebral malaria, neurological sequelae, anemia, respiratory distress, hypoglycemia, and complications of pregnancy: *Am J Trop Med Hyg*, **64**, 57-67.
- Muthas, D., Sabnis, Y. A., Lundborg, M., and Karlen, A., 2008, Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering: *J Mol Graph. Model*, **26**, 1237-1251.
- Naula, C., Parsons, M., and Mottram, J. C., 2005, Protein kinases as drug targets in trypanosomes and *Leishmania*: *Biochim. Biophys. Acta*, **1754**, 151-159.
- Needleman, S. B. and Wunsch, C. D., 1970, A general method applicable to the search for similarities in the amino acid sequence of two proteins: *J Mol Biol*, **48**, 443-453.
- Nelson, D. L. and Cox, M. M., 2001, *Lehninger - Biochemie.*: Springer Verlag, Berlin, Heidelberg, New York.
- Nervall, M., Hanspers, P., Carlsson, J., Boukharta, L., and Aqvist, J., 2008, Predicting binding modes from free energy calculations: *J Med Chem*, **51**, 2657-2667.
- Nicholas, K. B., Nicholas, H. B., and Deerfield, D. W., GeneDoc: Analysis and Visualization of Genetic Variation. Embnew. News. 1997.
- Niculescu-Duvaz, D., Whittaker, S., Springer, C., and Marais, R., 2007, The EGF receptor Hokey-Cokey: *Cancer Cell*, **11**, 209-211.
- Nissink, J. W., Murray, C., Hartshorn, M., Verdonk, M. L., Cole, J. C., and Taylor, R., 2002, A new test set for validating predictions of protein-ligand interaction: *Proteins*, **49**, 457-471.
- Noble, M. E., Endicott, J. A., and Johnson, L. N., 2004, Protein kinase inhibitors: insights into drug design from structure: *Science*, **303**, 1800-1805.

- Noro, R. and Gemma, A., 2008, EGFR and gefitinib (Iressa): *Gan To Kagaku Ryoho*, **35**, 1076-1079.
- Nwaka, S. and Ridley, R. G., 2003, Virtual drug discovery and development for neglected diseases through public-private partnerships: *Nat Rev Drug Discov*, **2**, 919-928.
- Oellien, F., Cramer, J., Beyer, C., Ihlenfeldt, W. D., and Selzer, P. M., 2006, The impact of tautomer forms on pharmacophore-based virtual screening: *J Chem Inf Model*, **46**, 2342-2354.
- Oellien, F., Engels, K., Cramer, J., Marhöfer, R. J., Kern, C., and Selzer, P. M., 2009, Searching new Antiparasitics in Virtual Space *in* From Molecular Targets to Drug Candidates: Selzer, P. M., Ed., WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
- Oellien, F., Ihlenfeldt, W. D., and Gasteiger, J., 2005, InfVis--platform-independent visual data mining of multidimensional chemical data sets: *J Chem Inf Model*, **45**, 1456-1467.
- Onodera, K., Satou, K., and Hirota, H., 2007, Evaluations of molecular docking programs for virtual screening: *J Chem Inf Model*, **47**, 1609-1618.
- Orzaez, M., Gortat, A., Mondragon, L., Bachs, O., and Perez-Paya, E., 2009, ATP-noncompetitive inhibitors of CDK-cyclin complexes: *ChemMedChem*, **4**, 19-24.
- Oslob, J. D., Heumann, S. A., Yu, C. H., Allen, D. A., Baskaran, S., Bui, M., Delarosa, E., Fung, A. D., Hashash, A., Hau, J., Ivy, S., Jacobs, J. W., Lew, W., Maung, J., McDowell, R. S., Ritchie, S., Romanowski, M. J., Silverman, J. A., Yang, W., Zhong, M., and Fuchs-Knotts, T., 2009, Water-soluble prodrugs of an Aurora kinase inhibitor: *Bioorg. Med Chem Lett*, **19**, 1409-1412.
- Pavletich, N. P., 1999, Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors: *J Mol Biol*, **287**, 821-828.
- Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., and Joannopoulos, J. D., 1992, Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients: *Reviews of Modern Physics*, **64**, 1045-1097.
- Perola, E., 2006, Minimizing false positives in kinase virtual screens: *Proteins*, **64**, 422-435.
- Pratt, D. J., Bentley, J., Jewsbury, P., Boyle, F. T., Endicott, J. A., and Noble, M. E., 2006, Dissecting the determinants of cyclin-dependent kinase 2 and cyclin-dependent kinase 4 inhibitor selectivity: *J Med Chem*, **49**, 5470-5477.
- Qu, X., Swanson, R., Day, R., and Tsai, J., 2009, A guide to template based structure prediction: *Curr Protein Pept. Sci*, **10**, 270-285.
- Quintas-Cardama, A. and Cortes, J., 2007, Nilotinib therapy in chronic myelogenous leukemia: *Drugs Today (Barc. )*, **43**, 691-702.
- Quintus, F., Sperandio, O., Grynberg, J., Petitjean, M., and Tuffery, P., 2009, Ligand scaffold hopping combining 3D maximal substructure search and molecular similarity: *BMC Bioinformatics*, **10**, 245.

- Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V., 1963, Stereochemistry of polypeptide chain configurations: *J Mol Biol*, **7**, 95-99.
- Raymond, J. W., Watson, I. A., and Mahoui, A., 2009, Rationalizing lead optimization by associating quantitative relevance with molecular structure modification: *J Chem Inf Model*, **49**, 1952-1962.
- Regazzi, M. B., Villani, P., Maserati, R., Cocchi, L., Giacchino, R., Burrioni, D., and Rettani, M., 1999, Pharmacokinetic variability and strategy for therapeutic drug monitoring of saquinavir (SQV) in HIV-1 infected individuals: *Br J Clin Pharmacol*, **47**, 379-382.
- Rogers, D., Brown, R. D., and Hahn, M., 2005, Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up: *J Biomol. Screen.*, **10**, 682-686.
- Roos, D. S., Crawford, M. J., Donald, R. G., Fraunholz, M., Harb, O. S., He, C. Y., Kissinger, J. C., Shaw, M. K., and Striepen, B., 2002, Mining the Plasmodium genome database to define organellar function: what does the apicoplast do?: *Philos. Trans R Soc Lond B Biol Sci*, **357**, 35-46.
- Ross-Macdonald, P. B., Graeser, R., Kappes, B., Franklin, R., and Williamson, D. H., 1994, Isolation and expression of a gene specifying a cdc2-like protein kinase from the human malaria parasite Plasmodium falciparum: *Eur J Biochem*, **220**, 693-701.
- Sadowski, J., 1997, A hybrid approach for addressing ring flexibility in 3D database searching: *J Comput. Aided Mol Des*, **11**, 53-60.
- Satyanarayana, A. and Kaldis, P., 2009, Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms: *Oncogene*, **28**, 2925-2939.
- Scheen, A. J. and Paquot, N., 2009, Vildagliptin (Galvus) and fixed combination vildagliptine-metformin (Eucreas) in the treatment of type 2 diabetes: *Rev Med Liege*, **64**, 161-167.
- Schlick, T. and Fogelson, A., 1992, TNPACK -- A truncated Newton minimization package for large-scale problems.: *ACM. Trans. Math. Software*, **18**, 46-111.
- Schnecke, V. and Bostrom, J., 2006, Computational chemistry-driven decision making in lead generation: *Drug Discov Today*, **11**, 43-50.
- Schneider, G. and Fechner, U., 2005, Computer-based de novo design of drug-like molecules: *Nat Rev Drug Discov*, **4**, 649-663.
- Schwardt, O., Kolb, H., and Ernst, B., 2003, Drug discovery today: *Curr Top Med Chem*, **3**, 1-9.
- Sclafani, R. A., 1996, Cyclin dependent kinase activating kinases: *Curr Opin Cell Biol*, **8**, 788-794.



- Selzer, P. M., 2005, Structure-based rational drug design: Neue *Wege* der modernen Wirkstoff-entwicklung. In: Lucius, R.; Hiepe, T.; Gottstein, B. (Hrsg) Grundzüge der allgemeinen Parasitologie: Parey, Berlin.
- Selzer, P. M., Marhofer, R. J., and Rohwer, A., 2008, Applied Bioinformatics: An Introduction.: Springer Verlag, Berlin, Heidelberg, New York.
- Sheinerman, F. B., Giraud, E., and Laoui, A., 2005, High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding: *J Mol Biol*, **352**, 1134-1156.
- Shirley, M. W., 1989, Development of a live attenuated vaccine against coccidiosis of poultry: *Parasite Immunol*, **11**, 117-124.
- Shirley, M. W., Smith, A. L., and Blake, D. P., 2007, Challenges in the successful control of the avian coccidia: *Vaccine*, **25**, 5540-5547.
- Soni, R., O'Reilly, T., Furet, P., Muller, L., Stephan, C., Zumstein-Mecker, S., Fretz, H., Fabbro, D., and Chaudhuri, B., 2001, Selective in vivo and in vitro effects of a small molecule inhibitor of cyclin-dependent kinase 4: *J Natl Cancer Inst*, **93**, 436-446.
- Sotriffer, C. A. and Drumburg, I., 2005, "In situ cross-docking" to simultaneously address multiple targets: *J Med Chem*, **48**, 3122-3125.
- Sousa, S. F., Fernandes, P. A., and Ramos, M. J., 2006, Protein-ligand docking: current status and future challenges: *Proteins*, **65**, 15-26.
- Stahl, M. A., 2000, Structure-Based Library Design, in *Virtual Screening for Bioactive Molecules*: Wiley-VCH Verlag, Weinheim, 229-264.
- Stephen, B., Rommel, M., Dauschies, A., and Haberkorn, A., 1997, Studies of resistance to anticoccidials in Eimeria field isolates and pure Eimeria strains: *Vet Parasitol*, **69**, 19-29.
- Stjernschantz, E., Marelius, J., Medina, C., Jacobsson, M., Vermeulen, N. P., and Oostenbrink, C., 2006, Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method: *J Chem Inf Model*, **46**, 1972-1983.
- Taft, C. A., Da Silva, V. B., and Da Silva, C. H., 2008, Current topics in computer-aided drug design: *J Pharm Sci*, **97**, 1089-1098.
- Tetko, I. V., 2005, Computing chemistry on the web: *Drug Discov Today*, **10**, 1497-1500.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J., 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice: *Nucleic Acids Res*, **22**, 4673-4680.
- Tominaga, Y. and Jorgensen, W. L., 2004, General model for estimation of the inhibition of protein kinases using Monte Carlo simulations: *J Med Chem*, **47**, 2534-2549.

- Traxler, P., Green, J., Mett, H., Sequin, U., and Furet, P., 1999, Use of a pharmacophore model for the design of EGFR tyrosine kinase inhibitors: isoflavones and 3-phenyl-4(1H)-quinolones: *J Med Chem*, **42**, 1018-1026.
- Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., and McFadden, G. I., 1998, Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*: *Proc Natl Acad Sci U S A*, **95**, 12352-12357.
- Waller, R. F., Reed, M. B., Cowman, A. F., and McFadden, G. I., 2000, Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway: *EMBO J*, **19**, 1794-1802.
- Walters, W. P., Stahl, M. A., and Murcko, M. A., 1998, Virtual Screening - An Overview: *Drug Discov. Today*, **3**, 160-178.
- Ward, P., Equinet, L., Packer, J., and Doerig, C., 2004, Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote: *BMC Genomics*, **5**, 79.
- Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E., and Head, M. S., 2006, A critical assessment of docking programs and scoring functions: *J Med Chem*, **49**, 5912-5931.
- Waters, N. C., Woodard, C. L., and Prigge, S. T., 2000, Cyclin H activation and drug susceptibility of the Pfmrk cyclin dependent protein kinase from *Plasmodium falciparum*: *Mol Biochem Parasitol*, **107**, 45-55.
- Wheeler, D. L., Church, D. M., Edgar, R., Federhen, S., Helmberg, W., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Suzek, T. O., Tatusova, T. A., and Wagner, L., 2004, Database resources of the National Center for Biotechnology Information: update: *Nucleic Acids Res*, **32**, D35-D40.
- Willett, P., 2003, Similarity-based approaches to virtual screening: *Biochem Soc Trans*, **31**, 603-606.
- Willett, P., 2006, Similarity-based virtual screening using 2D fingerprints: *Drug Discov Today*, **11**, 1046-1053.
- Williams, R. B., Carlyle, W. W., Bond, D. R., and Brown, I. A., 1999, The efficacy and economic benefits of Paracox, a live attenuated anticoccidial vaccine, in commercial trials with standard broiler chickens in the United Kingdom: *Int J Parasitol*, **29**, 341-355.
- Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., and Williamson, D. H., 1996, Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*: *J Mol Biol*, **261**, 155-172.
- Wissler, C., 1905, The Spearman Correlation Formula: *Science*, **22**, 309-311.

- 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., and Waters, N. C., 2003, Oxindole-based compounds are selective inhibitors of *Plasmodium falciparum* cyclin dependent protein kinases: *J Med Chem*, **46**, 3877-3882.
- Wu, S. Y., McNae, I., Kontopidis, G., McClue, S. J., McInnes, C., Stewart, K. J., Wang, S., Zheleva, D. I., Marriage, H., Lane, D. P., Taylor, P., Fischer, P. M., and Walkinshaw, M. D., 2003, Discovery of a novel family of CDK inhibitors with the program LI-DAEUS: structural basis for ligand-induced disordering of the activation loop: *Structure*, **11**, 399-410.
- Xiang, Z., 2006, Advances in homology protein structure modeling: *Curr Protein Pept. Sci*, **7**, 217-227.
- Yoon, S. and Welsh, W. J., 2004, Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring: *J Chem Inf Comput. Sci*, **44**, 88-96.
- Yung, S., Unnasch, T. R., and Lang-Unnasch, N., 2001, Analysis of apicoplast targeting and transit peptide processing in *Toxoplasma gondii* by deletional and insertional mutagenesis: *Mol Biochem Parasitol*, **118**, 11-21.
- Zhai, S., Senderowicz, A. M., Sausville, E. A., and Figg, W. D., 2002, Flavopiridol, a novel cyclin-dependent kinase inhibitor, in clinical development: *Ann Pharmacother.*, **36**, 905-911.
- Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I., and Schneider, G., 2001, Deciphering apicoplast targeting signals--feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins: *Gene*, **280**, 19-26.