

**VIRTUELLES SCREENING NACH
RNA-LIGANDEN:
ZUM UMGANG MIT EINER FLEXIBLEN
ZIELSTRUKTUR**

**Dissertation
zur Erlangung des Doktorgrades der
Naturwissenschaften**

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Manuel Manfred Nietert
aus Lahn-Gießen

Frankfurt am Main
2008

(D30)

vom Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Harald Schwalbe
erster Gutachter: Prof. Dr. Gisbert Schneider
zweiter Gutachter: Prof. Dr. Joachim Engels
Datum der Disputation: 23. Juli 2008

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	VIII
1. EINLEITUNG	1
1.1 Kontext der Arbeit.....	1
1.2 Der chemie-informatische Ansatz	1
1.3 Projektphasen	2
2. ALLGEMEINE METHODEN.....	8
2.1 Molekulare Deskriptoren	8
2.1.1 MOE-2D-Deskriptoren	9
2.1.2 CATS-2D	9
2.1.3 TripleCharge3D (TC3D).....	9
2.2 Bewertungsfunktionen.....	10
2.2.1 Mathews Korrelationskoeffizient (MCC)	10
2.2.2 „Prozent korrekte Vorhersage“	10
2.2.3 Mittlerer Quadratischer Fehler (MSE).....	10
2.3 Kolmogorov-Smirnov (KS) Statistik	11
2.3.1 Allgemeines zur KS-Statistik.....	11
2.3.2 Deskriptoren-Auswahl mit KS-Statistik	12
2.4 Künstliche Neuronale Netze (ANNs).....	13
2.4.1 Allgemeine ANN Definition:	13
2.4.2 Preprozessierung der Deskriptordaten:	14
2.4.3 Datensatz-Generation:.....	14
2.4.4 Generation der ANNs:	14
2.4.5 Training der ANNs:	14
2.4.6 MCC als Gütekriterium der Vorhersage der einzelnen ANNs.....	15
2.4.7 Selektion der besten ANNs für jeden Kreuz-Validationsdatensatz:	15
2.4.8 ANN-Jury-Zusammenstellung.....	15
2.4.9 ANN-Jury-Vorhersagen	15
2.5 Selbstorganisierende Karten (SOM)	15
2.6 <i>Hierarchical-k-means</i> – hk-means-Clusterung	16
2.7 Der Tat-TAR-FRET-Assay	17
2.7.1 Experimentelle Grundlagen	17
2.7.2 Durchführung der Messungen	18

3.	LIGANDEN-BASIERTE ANSÄTZE	28
3.1	RNA-Liganden-Datenbank	28
3.2	Allgemeiner RNA-Binder-Filter	29
3.2.1	Allgemeiner RNA-Liganden Datensatz – Datensatzzusammenstellung	29
3.2.2	Resultierender Arbeitsdatensatz: RNA-Binder gegen Protein-Binder	31
3.2.3	Untersuchungen zur Trennbarkeit der beiden Datensätze	32
3.2.4	KS-Statistik zur Analyse der Eigenschaftsverteilung	32
3.2.5	Abschätzung der strukturellen Diversität im RNA-Liganden-Datensatz	35
3.2.6	Allgemeiner Filter für RNA-Binder	37
3.2.7	Verteilung der Scorewerte für RNA-Binder gegen Protein-Binder	38
3.2.8	Suche mit den RNA-Binder-Filtern in den Chemikalien-Datenbanken der Versandhändler	40
3.2.9	Dockingexperimente mit den ausgewählten Substanzen in einen Konformer des PDB-Eintrags 1LVJ	41
3.2.10	„Cherry-Picking“ von Testsubstanzen aus den sortierten Trefferlisten zur experimentellen Bestimmung der Aktivität im Tat-TAR-FRET-Assay	53
3.2.11	Auswertung der RNA-Binder-Testsubstanzsammlung von Specs im Tat-TAR-FRET-Assay	54
3.2.12	Nachmessen der Substanzen	57
3.2.13	Auswertung der RNA-Binder-Testsubstanzsammlung von Asinex im Tat-TAR-FRET-Assay	58
3.2.14	Auswertung der RNA-Binder-Testsubstanzsammlung im CFivTT-Assay	60
3.2.15	Auswertung der Überschneidungen zwischen Tat-TAR-FRET-Assay und CFivTT-Assay	61
3.2.16	Struktur-Wirkungsbeziehung von M32 im CFivTT	63
3.2.17	Ableitung von Pharmakophormodellen aus den gemessenen Aktivitätswerten	66
3.2.18	Rückführung der Aktiven der Assays auf die verwendeten ANN-Filter	72
3.2.19	Bekannte Fehler bei der Zusammenstellung des RNA-Liganden-Datensatzes	75
3.3	Tat-TAR-FRET-Assay-spezifischer ANN-basierter Struktur-Filter	96
3.3.1	Erweiterter Datensatz und Ableitung eines spezifischeren Filters	96
3.3.2	Erstellung eines neuen Tat-TAR-Assay spezifischen Datensatzes	96
3.3.3	Vorbereitung der Strukturdatensätze / Konformergeneration	97
3.3.4	Deskriptorberechnungen für den Tat-TAR-FRET-Assay-spezifischen Datensatz	97
3.3.5	Deskriptorberechnungen für die COBRA und Specs	98
3.3.6	Abschätzung der Trennbarkeit des FRET-Assay-spezifischen Datensatzes mittels SOM-Projektionen	98
3.3.7	Vergleich der Molekülgerüste von COBRA, TAR-Liganden und TAR-Nichtliganden	100

3.3.8	Filterung der Deskriptoren zur Reduktionen der Anzahl der Deskriptordimensionen für das spätere Netztraining	101
3.3.9	KS-Statistiken für die Tat-TAR-Assay-spezifischen Datensätze.....	103
3.3.10	Netztraining der ANNs des Tat-TAR-Assay-spezifischen Filters	106
3.3.11	Screening der Specs mit den erhaltenen Jury-Filtern.....	108
3.3.12	Betrachtung der Auswahllisten	109
3.3.13	„Cherry-Picking“ der potentiell TAR-RNA-spezifischen Substanzen zum Testen	109
3.3.14	Test der Substanzen im Tat-TAR-FRET-Assay.....	110
3.3.15	Löslichkeitsprobleme	110
3.3.16	Ergebnisse der Tat-TAR-FRET-Assay-Messungen für den Assay-spezifischen Filter	113
3.3.17	Messungen zur Überprüfung der potenziell aktiven Verbindungen	117
3.3.18	Verbleibende aktive Verbindungen der LR-Reihe aus dem Tat-TAR-FRET-Assay	118
3.3.19	Methyleneaminoguanidin – LR9, AS20 und M23 – Vergleich der Substruktur motive	119
4.	ZIELSTRUKTURBASIERTE ANSÄTZE	120
4.1	Große Konformationsänderungen möglich	120
4.1.1	Klassische Docking-Probleme oder welche Struktur nimmt man	121
4.2	MD-Datensätze als Konformergeneratoren.....	126
4.3	Die MD-Trajektorien	126
4.4	Die RNA-Konformer-Datenbank.....	131
4.4.1	Hintergrund	131
4.4.2	Koordinierungsprofil.....	131
4.4.3	Berechnung der Taschen nach <i>PocketPicker</i>	133
4.4.4	Berechnen der Moleküloberfläche	133
4.4.5	Ableiten der Taschenoberfläche	133
4.4.6	Laden von Deskriptordateien	133
4.4.7	Ablegen und Extrahieren der Informationen in der Datenbank.....	133
4.4.8	Anwendung von <i>Pandora</i>	134
4.5	Ableiten eines Referenzdatensatzes aus der Datenbank	135
4.5.1	Überlegungen zur Auswahl der repräsentativen Taschen aus dem verfügbaren Konformerraum.....	135
4.5.2	Erstes Kriterium Taschendurchmesser.....	135
4.5.3	Zweites Kriterium „mittlere Vergrabenheit“ der Tasche	138
4.5.4	Vergrabenheit gegen Taschengröße als Fingerprint für Konformere	140
4.5.5	Pareto-Ranking zur Selektion der undominierten Lösungen der Multi-Objektiven-Optimierung	144

4.5.6	Die Diagonale zur Selektion der Taschen, die die mittlere Vergrabenheit und die Taschengröße gleichermaßen beinhalten ohne ein Kriterium zu bevorzugen.....	144
4.5.7	Analyse der Verteilung der Datenpunkte für die Auftragung mittlerer Vergrabenheit und Taschengröße	146
4.5.8	Auswahl radial um den optimalen Punkt P3 des Datensatzes	150
4.5.9	Das Taschenoberflächenbildungspotential PSGP als Gütekriterium für die Taschenauswahl.....	157
4.5.10	Erste Tests des PSGP zur Taschenselektion.....	164
4.6	Cluster-Methoden und ihre Schwächen	169
4.6.1	Ansatzpunkte und Probleme beim RMSD-basierten Clustern.....	169
4.6.2	Auswertung der RMSD-Berechnungen für Beispielkonformere der TAR.....	171
4.6.3	SIMPrints eine 3D-Form beschreibende Darstellung mit Substrukturauflösung.	176
4.7	Oberflächenanalyse für 1ANR aufgelöst nach Elementen	183
4.8	Abschließende Untersuchungen der Interaktion der bekannten Peptid-Liganden mit TAR.....	185
5.	SCHLUSSFOLGERUNG UND AUSBLICK	189
6.	ZUSAMMENFASSUNG	192
7.	ANHANG	194
7.1	KS-Statistik zur Trennbarkeit potentieller RNA-Liganden von den potentiellen Protein-Liganden	194
7.2	Die Strukturen der Substanzen der M-Reihe	198
7.3	Die Strukturen der Substanzen der M-Reihe	201
7.4	Die Strukturen der Substanzen der LR-Reihe.....	204
7.5	Absorptionspektren für ausgewählte Substanzen der Tat-TAR-FRET-Assay-Messreihen	208
7.6	CFivTT-Inhibitionsassay.....	212
7.6.1	Ergebnisse des CFivTT-Assay für die M-Reihe.....	213
7.7	Ganzzellassay – Wachstumsinhibition durch Substanzen der M-Reihe	213
7.8	DIGE-Proteomics-Ansatz	214
7.9	Liste der Publikationen die für den „Allgemeinen RNA-Liganden-Filter“ herangezogen wurden	215
7.10	PSGP-Projektion auf Atome – Basensequenz von TAR.....	219
8.	ABBILDUNGSVERZEICHNIS.....	224
9.	LITERATURVERZEICHNIS	232

Abkürzungsverzeichnis

2D	zwei-dimensional
3D	drei-dimensional
ADMET	A bsorption, D istribution, M etabolism, E xcretion, T oxicity
aRNALD	a llgemeiner RNA-Liganden-Datensatz
ARM	a rgininreiches M otiv
ANN	a rtificial n eural n etwork = künstliches neuronales Netz
bRNALD	(experimentell) b estätigter RNA-Liganden-Datensatz
CATS	C hemically A dvanced T emplate S earch
CFivTT-Assay	C ell F ree <i>in vitro</i> T ranscriptions T ranslations Assay
COBRA	C ollection of B ioactive R eference A nalogues
DEPC	Diethylpyrocarbonat
DMSO	Dimethylsulfoxid
Eppi	Safe-Lock Tubes
FRET	F luoreszenz- R esonanz E nergie T ransfer
GFP	g reen f luorescent p rotein = g rün- f luoreszierendes P rotein
GOLD	G enetic O ptimised L igand D ocking (Docking Programm)
GUI	G raphical U ser I nterface = Grafische Benutzeroberfläche
HIV	H uman I mmunodeficiency V irus
HL	H idden L ayer
HPLC	Hochleistungsflüssigkeitschromatographie
IC_{50}	Konzentration einer Substanz bei einer Inhibition von 50 % auftritt
IL	I nter- L ayer
IR	Infrarot
K_d	Dissoziationskonstante
LIQUID	L igand-based Q uantification of I nteraction D istributions
LTR	long terminal repeats
MCC	M atthews c orrelation c oefficient = Mathews Korrelationskoeffizient
MD	M olekül- D ynamik-(Simulation)
MDDB	M olekül- D ynamik-(Simulations-Trajektorien)- D aten B ank
MOE	M olecular O perating E nvironment
mRNA	m essenger- RNA
MSE	M ean s quare e rror
m-Skript	Matlab-Skript/Matlab-Programm
NIPALS	N onlinear i terative p artial l east s quares
NMR	N uklearmagnetische R esonanz
OL	O utput- L ayer

P3	der maximal erreichbare, als Optimum angesehene Datenpunkt der Auftragung von Taschengröße gegen die mittlere Vergrabenheit der Tasche für einen Datensatz - Synonym für die bestmöglich Tasche nach den beiden genannten Kriterien.
Pandora	<i>Pocket Analysis for Docking on Representative Aptamers</i>
PDB	P rotein D ata B ank
PPP	p otential p harmacophore p oint = potentieller P harmakophorpunkt
PSA	p olar s urface a rea
PSGP	P ocket S urface G eneration P otential/ P robability = Taschenoberflächenbildungswahrscheinlichkeit
QSAR	q uantitative s tructure a ctivity r elationship
RFA	relative F luoreszenzaktivität
RFI	relative F luoreszenz I nhibierung in Prozent
RFU	R elative F luorescence U nits
RMSD	root m ean s quare d eviation
RNA	Ribonukleinsäure
SA	S imulated A nnealing
SAR	s tructure a ctivity r elationship
SDF	s tructure d ata f ormat
SIMPrint	S kalierungs- I ndex- M ethod- I mprint
SMILES	S implified M olecular I nput L ine E ntry S pecification
SOM	s elf o rganizing m ap = selbst-organisierende Karte
SOMMER	<i>Self-Organizing Map Maker for Education and Research</i>
SVL	s cientific v ector l anguage
TAR	T rans- A ctivation R esponsive
TC3D	T riple C harge 3D (Deskriptor)
Triton-X	Alkylphenylpolyethylenglycol
tRNA	t ransfer- R NA
uRNALD	(experimentell) unbestätigter R NA- L iganden- D atensatz
UV	U ltraviolett
Well	Vertiefung der Mikrotiterplatte

1. Einleitung

1.1 Kontext der Arbeit

Diese Arbeit wurde im Rahmen der Mitarbeit im Sonderforschungsbereich 579 (SFB579) – RNA-Liganden-Wechselwirkungen – im Teilprojekt A11.2 als Mitglied im Arbeitskreis von Prof. Dr. Gisbert Schneider angefertigt. Das Teilprojekt A11.2 beschäftigt sich innerhalb des SFB primär mit den chemie-informatischen Fragestellungen, die für das *de Novo* Design von RNA-Liganden bedeutsam sind.

Die allgemeine Natur der Fragestellung, die sich auf das Design von potentiellen RNA-Liganden richtet, wurde hierbei anhand eines Modellsystems untersucht, welches auch im Rahmen des SFB579 bereits untersucht wird: dem Tat-TAR-Wechselwirkungsmodell.

Dieses Modell umfasst einen *in vitro* Verdrängungs-Assay, welcher mittels FRET-Technik eine Beobachtung des Tat-TAR-Komplexes ermöglicht (Matsumoto 2000).

1.2 Der chemie-informatische Ansatz

Die Chemie-Informatik beschäftigt sich allgemein mit der numerischen Repräsentation von chemischen Molekülen in einer *in silico* Umgebung und den daraus ableitbaren Lösungen für experimentelle Fragestellungen. Die möglichst genaue Modellierung der experimentellen Wirklichkeit ist das Ziel der Chemie-Informatik und als solches nur durch das Einbeziehen möglichst vieler experimentell bestätigter Informationen möglich. Letztlich bestimmt die Güte der experimentellen Daten sowohl die Güte des davon abgeleiteten Modells als auch die aus diesem Modell wiederum abgeleiteten Vorhersagen für experimentelle Ergebnisse. Somit muss sich ein chemie-informatisches Modell immer zur Prüfung seiner Güte dem experimentellen Vergleich stellen. Hierbei zeigt sich dann auch, ob das Modell die experimentelle Wirklichkeit ausreichend repräsentiert. Erkenntnisse über das in den experimentellen Eingabedaten vorhandene, zum Teil verborgene Wissen, sind leider von dieser Methode nicht zu erwarten, jedoch gilt auch hier das Zitat von George Box „all models are wrong, some models are useful“ (Box 1979). Alle Modelle sind somit erst einmal als falsch anzusehen und stellen immer eine Vereinfachung der Realität dar, jedoch ermöglichen einige dieser Vereinfachungen nützliches Wissen für weitere Experimente aus Datenbeständen abzuleiten, die für den Menschen manuell nur schwer erschließbar wären.

Ein Schwerpunkt dieser Arbeit beschäftigt sich mit dem Einsatz von adaptiven Maschinenlernverfahren, welche automatisiert einen funktionalen Zusammenhang zwischen Eingabewerten und definierten Zielwerten herstellen können und somit, falls das hierfür benutzte Wissen repräsentativ für die erlernte Fragestellung war, in der Lage sind, Vorhersagen über bisher unklassifizierte Eingabewert/Zielwert-Beziehungen eine relativ gute Näherung der Realität abzugeben.

Bezogen auf die Fragestellung der Medikamentenentwicklung ergeben sich hier allgemein ableitbare Projektphasen, die für jedes Zielstruktur-Liganden-System – im Hinblick auf eine ausreichend klare Modellierung der Wirklichkeit – durchlaufen werden müssen, um durch diese Modellierung eine Reduktion des experimentellen Aufwands herbeizuführen.

1.3 Projektphasen

Die Arbeit lässt sich zusammenfassend in vier Phasen einteilen:

Phase 1: Bestandsaufnahme bekannter Informationen über die Zielstruktur

- experimentell bestimmte Zielstrukturen
- experimentell bestimmte Liganden/Nichtliganden der Zielstruktur

Phase 2: Ableiten eines ligandenbasierten Ansatzes zur Vorhersage von potentiellen Bindern der Zielstruktur aus Substanzbibliotheken, der nicht auf Strukturdaten der Zielstruktur beruht.

Phase 3: Analyse der bekannten Konformere der Zielstruktur auf konstante Angriffspunkte für ein spezielles Ligandendesign.

Phase 4: Einbinden der bekannten Strukturinformationen der Zielstruktur zur weiteren Verfeinerung des Auswahlverfahrens neuer Kandidaten für die weitere experimentelle Bestimmung des Bindeverhaltens.

„Divide and Conquer“ – *In silico* Identifikation von neuen RNA-Liganden.

Das Ziel der *in silico* Identifikation von neuen Wirkstoffen muss in einzelne bewältigbare Schritte zerlegt werden, um sich diesem Ideal anzunähern. Diese Zerlegung erfolgt frei nach dem Philosophen Sun Tzu (Giles 1910) wie folgt:

Allgemeine Regeln zur Bekämpfung eines Gegners

1. Kenne deinen Gegner
2. Kenne die bekannten Gegner deines Ziels
3. Kenne die Freunde deines Ziels
4. Lerne aus den Fehlschlägen
5. Zeige keine generelle Angst vor Kollateralschäden wenn du das Ziel bekämpfst, der Gegner zeigt auch keine Gnade
6. Je genauer das Wissen über die spezifische Persönlichkeit des Ziels ist, desto präziser ist die Waffe der Wahl
7. Die Dosis macht das Gift, aber manchmal ist das Gift die einzige Heilung.

Dieser Sachverhalt wird im Rahmen des Wirkstoffdesigns wie folgt umgesetzt:

1. Kenne deinen Gegner
 - *Identifiziere die besonderen Eigenheiten deines Ziels, halte die Kollateralschäden so klein wie möglich*

Durch Kenntnis der spezifischen individuellen Eigenschaften (bspw. räumlicher Aufbau) der Zielstruktur, die diese von anderen Wirkstoffinteraktionszielen unterscheidet, ist erst die selektive Entwicklung von Liganden möglich die, nur an dieses spezielle Ziel binden. Bei fehlender Information darüber, was die Zielstruktur einzigartig macht, wird es unmöglich sein, selektive Vorhersagen zu erstellen. Gibt es außer der Zielstruktur keine Informationen über andere Wirkstoffinteraktionsziele, führt dies mit großer Wahrscheinlichkeit bei aktiven Liganden zu unvorhersehbaren Nebenwirkungen. Wenn jedoch weitere Informationen zu Nichtzielen vorhanden sind, können diese gleich zur Reduktion der Nebenwirkungswahrscheinlichkeit verwendet werden.

2. Kenne die bekannten Gegner deines Ziels
 - *Sie können dir beibringen, wie der Gegner zu treffen ist*

Durch Kenntnis der Struktur von bereits experimentell bestätigten Liganden kann ein Modell derjenigen kritischen Interaktionen abgeleitet werden, die durch diese Liganden gestört werden. In der vergleichenden Analyse des Aufbaus dieser Verbindungen und den Substanzen, die keine Wirkung auf das Ziel zeigen, ergeben sich Eigenschaften, die die aktiven von den inaktiven Substanzen unterscheiden. Diese können dazu genutzt werden, um in Substanzbibliotheken nach weiteren Substanzen mit dieser vorteilhaften Eigenschaftsverteilung zu suchen und so den Pool aktiver Verbindungen zu vergrößern. Somit kann der Substanzpool und damit der Informationsgehalt gesteigert werden, um kritische Eigenschaften immer mehr einzugrenzen.

3. Kenne die Freunde deines Ziels
 - *Beobachte sie, um mehr über kritische Interaktionen des Ziels zu erfahren*

Durch Kenntnis der natürlichen Wechselwirkungspartner der Zielstruktur kann einerseits abgeleitet werden, wo sich kritische Interaktionspunkte auf der Oberfläche der Zielstruktur befinden, andererseits kann anhand der natürlichen Liganden festgestellt werden wie diese mit der Zielstruktur interagieren. Aus diesen Informationen können dann Mimetika abgeleitet werden, die denselben Bindemechanismus wie die natürlichen Liganden besitzen und mit diesem um die Bindung an der Zielstruktur konkurrieren.

4. Lerne aus den Fehlschlägen
 - *Nicht funktionierende Ansätze beherbergen das Wissen, wie man es nicht probieren sollte.*

Indem die Substanzen, welche keine gewünschte Wirkung gezeigt haben, in die Modelle als Negativbeispiele eingebaut werden, ermöglicht man die Rate an weiteren Fehlschlägen zu senken.

5. Zeige keine generelle Angst vor Kollateralschäden, wenn du das Ziel bekämpfst, der Gegner zeigt auch keine Gnade.

Sollte es keine andere Wahl geben, ist die Inkaufnahme von Nebenwirkungen erst einmal wichtiger, als gar keine Versuche zur Inhibition zu unternehmen. Aus einer affinen Verbindung kann später vielleicht immer noch ein nebenwirkungsärmeres Strukturmotiv abgeleitet werden.

6. Je genauer das Wissen über die spezifische Persönlichkeit des Ziels ist, desto präziser ist die Waffe der Wahl.

Je genauer die Zielstruktur bekannt ist, desto präziser lässt sich eine Wirkstoffstruktur speziell für diese Zielstruktur entwickeln.

7. Die Dosis macht das Gift, aber manchmal ist das Gift die einzige Heilung.

Manchmal, wie im Fall der einfachen Chemotherapeutika, lassen sich Nebenwirkungen nicht vermeiden, jedoch rechtfertigt sich dieses Risiko, wenn der Wirkmechanismus des Giftes der einzige bekannte Ansatz in der Bekämpfung der Krankheit ist.

Zusammenfassend lässt sich somit sagen:

Das potenzielle Wirkstoffziel muss folglich eine zeitlich konstante Raumstruktur aufweisen, um als Ziel für strukturelle Gegenmaßnahmen angreifbar zu sein. Hinweise hierfür sind eine aufgeklärte Raumstruktur oder das Wissen über Effektmoleküle, welche dieses potentielle Wirkstoffziel beeinflussen können.

Ziel: die Störung der Tat-TAR-Interaktion

Einleitung:

Die Störung der Wechselwirkung des viralen Tat-Proteins (transactivation of transcription) mit der TAR-RNA-Zielstruktur (TAR = trans-activation response) des Humanen Immunodefizienz Virus (HIV) war das Primärziel dieser Arbeit. Die folgenden Absätze sollen den Kontext der Arbeit verdeutlichen, der zur Anwendung der in dieser Arbeit verwendeten Methoden und Überlegungen für den Umgang mit der Fragestellung des RNA-Liganden-Designs führten und noch in weiteren Abschnitten dieser Arbeit dargelegt werden. Ferner soll getreu der Philosophie aus dem obigen Abschnitt „Divide and Conquer“ der „(Haupt-)Gegner“ dieser Studie definiert werden sowie eine erste Abschätzung über die allgemeine Durchführbarkeit des Liganden-Designs für die gewählte Zielstruktur erfolgen, die wieder zur Definition der ersten anzuwendenden Strategie führte.

Kontext der Tat-TAR-Interaktion:

Wie von Wöhrl in einer Übersicht (Wöhrl 2003, Seite 4) beschrieben, „löst Tat einen positiven Rückkopplungsmechanismus aus, der ein hohes Transkriptionsniveau der viralen RNAs bewirkt“; siehe Abbildung 1. In Abbildung 2 ist die schematische Darstellung dieser durch die Bindung des Tat-Proteins an das TAR-Element bewirkten Aktivierung der RNA-Polymerase II, mittels der durch den Transkriptions-Elongationsfaktor pTEFb erfolgenden Phosphorylierung der RNA-Polymerase II nach Wöhrl (2003, Seite 4) dargestellt.

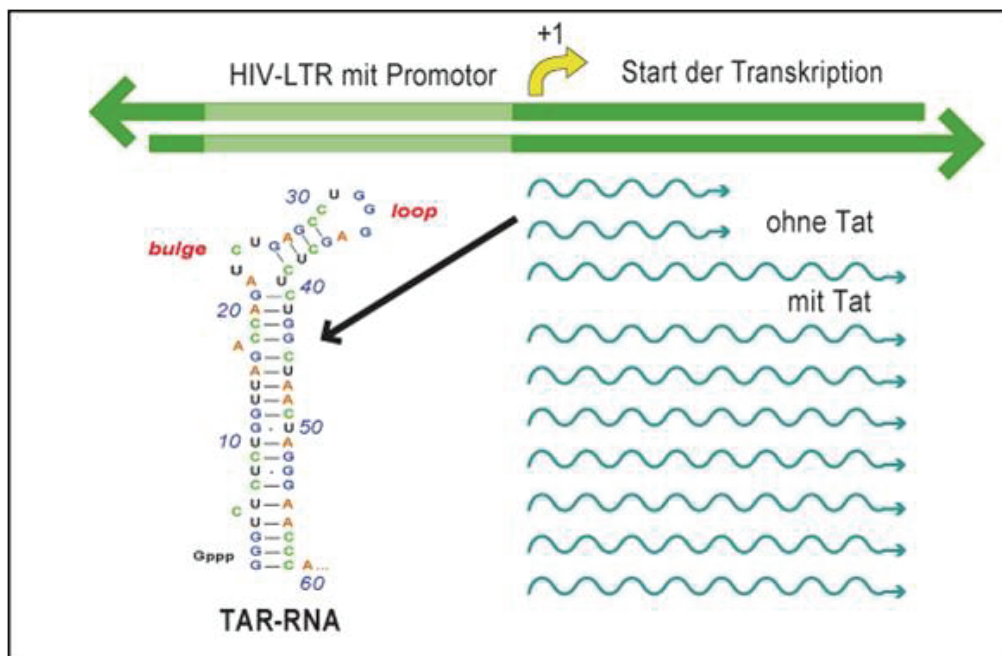


Abbildung 1: Aktivierung der Transkriptionselongation durch Tat am LTR von HIV, nach Wöhrl (2003, Seite 4); Verursacht durch die Bindung des Tat-Proteins an die TAR-RNA-Struktur kommt es zu einer deutlichen Steigerung der Transkriptionsrate längerer Viren-Transkripte und somit letztendlich zur gesteigerten Viren-Replikation. Dementsprechend ist diese Interaktion ein theoretisch guter Ansatzpunkt für die Wirkstoffentwicklung.

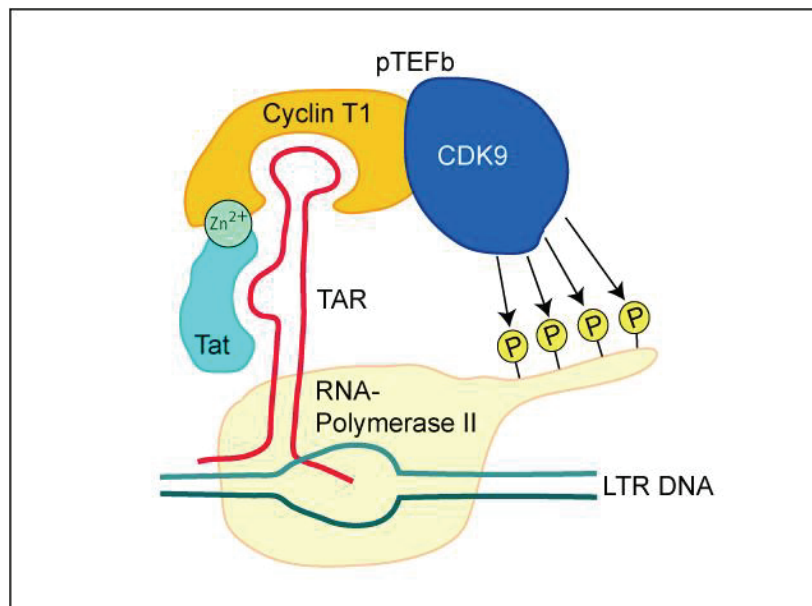


Abbildung 2: Vereinfachte schematische Darstellung des Antiterminationskomplexes; nach Wöhrl (2003, Seite 4)

Reduzierung der Tat-TAR-Interaktion auf ein isoliertes Modellsystem

Wie bei Wöhrl beschrieben, ist für die Bindung des Tat-Proteins an die TAR-Zielstruktur die basische Region des Tat-Proteins notwendig: „Die basische Region enthält ein argininreiches Motiv (ARM), welches wesentlich an der Bindung von Tat an die UCU-Ausbauchung von TAR beteiligt ist.“. Für die TAR-RNA konnten laut Wöhrl „Durch verschiedene Bindungs- und Deletionsstudien [konnte] gezeigt werden, dass die Region 19 – 42 der TAR-RNA notwendig und hinreichend für die Tat-Bindung ist. Die UCU-Ausbauchung ist für die Affinität und spezifische Bindung von Tat erforderlich; während die Schleifenregion von TAR für die in vivo Transaktivierung notwendig ist, jedoch für die direkte Tat/TAR Bindung nicht benötigt wird (Übersicht siehe: Karn. J., Mol. Biol. 1999).“.

Aus den vorliegenden Fakten kann somit ein Modellsystem aus diesen beiden reduzierten Strukturen abgeleitet werden, dass zur Untersuchung möglicher potentieller Inhibitoren der Tat-TAR-Interaktion geeignet ist. Dieses reduzierte Modell wurde bereits, im dem SFB579 zur Verfügung stehenden Tat-TAR-FRET-Assay (Matsumoto 2000) im Arbeitskreis Göbel angewandt.

Kurzbeschreibung des verbleibenden Interaktionssystems

Im Rahmen dieser Arbeit wird das Tat-Protein somit auf eine 9-Aminosäuren lange Sequenz reduziert: dem Tat-Peptid (RKKRRQRRR). Eine Reduktion der gesamten Zielstruktur der HIV-Transkripte erfolgt hierbei durch Fokussierung auf einen Teilausschnitt der Transkripte mit einer Länge von 31-Basen, der den Interaktionsbereich um die so genannte Bulge-Struktur enthält; siehe hierzu auch die Abbildung 3, welche eine schematische Darstellung der TAR-Sekundärstruktur beinhaltet. In dieser werden auch die einzelnen Substrukturbereiche der Zielstruktur einführend erklärt.

Definition der Zielstruktur für diese Arbeit

Da im Rahmen der Mitarbeit im SFB579 die RNA-Liganden-Interaktion im Vordergrund steht, wird somit die verbleibende TAR-RNA-Sequenz mit Bulge als Zielstruktur definiert.

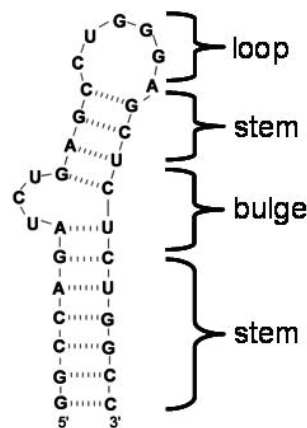


Abbildung 3: TAR-RNA, schematische Darstellung der Sekundärstruktur

Die Struktur besteht aus vier Substrukturmustern:

1. einem durch Basenpaarung von zwölf Basen hervorgerufenen Stammbereich,
2. einer darauf folgenden Bulge-Struktur, die durch drei standardmäßig ungepaarte Basen definiert wird und durch zwei Basenpaare begrenzt wird,
3. einem weiteren Stammbereich bestehen aus drei Basenpaaren,
4. einem Schleifenbereich (Loop), bestehend aus sechs ungepaarten Basen.

Besonderheiten der Zielstruktur im Bezug auf das Liganden-Design/Suche

Die jeweilige Raumstruktur mit gebundenem Liganden wird nach bisherigen Studien (Faber *et al.* 2000, Aboul-ela *et al.* 1996) stark durch den jeweiligen Liganden bedingt. Hinzukommt die offensichtliche generelle Flexibilität der TAR-Struktur, wie dies aus dem PDB-Eintrag 1ANR (Aboul-ela *et al.* 1996) ersichtlich ist. Dieser PDB-Eintrag zeigt die mittels NMR ermittelten 20 Strukturvorschläge einer freien TAR ohne Liganden, siehe hierzu auch Abbildung 4, welches die drei unterschiedlichsten Konformere des PDB-Eintrags 1ANR darstellt.

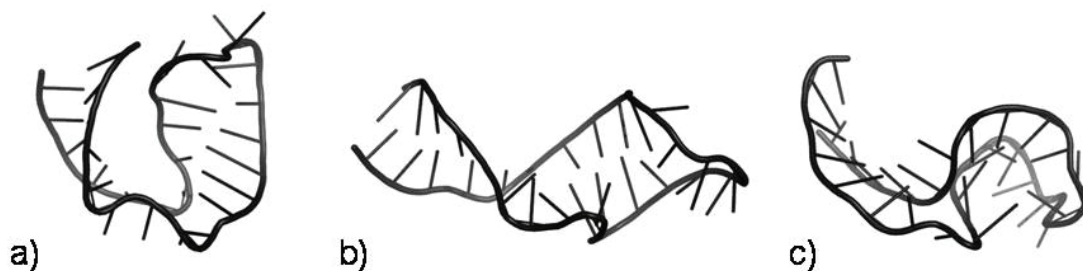


Abbildung 4: 1ANR – TAR-Konformere; a) Konformer Nr. 18 b) Konformer Nr. 7 c) Konformer Nr. 6; die Beipielkonformere wurden anhand des RMSD ausgewählt. Konformer 7 und 18 sind am weitesten im Ensemble von 1ANR entfernt, Konformer 6 ist wiederum am weitesten entfernt von den beiden erstgenannten Konformeren.

Ist ein Ligandendesign prinzipiell möglich

Das potenzielle Wirkstoffziel müsste laut Abschnitt 1.2 idealerweise eine zeitlich konstante Raumstruktur aufweisen, um als Ziel für strukturelle Gegenmaßnahmen angreifbar zu sein. Hinweise hierfür sind eine aufgeklärte Raumstruktur oder das Wissen über Effektormoleküle, welche dieses potentielle Wirkstoffziel beeinflussen können.

Beides trifft für die TAR-RNA des Humanen Immunodefizienz Virus (HIV) zu, da dessen als TAR-Element (trans-activation response) bezeichnete RNA-Sequenz mit ihrer charakteristischen Sekundärstruktur in sämtlichen Transkripten des Virus vorkommt und hierbei eine essentielle Wechselwirkung mit dem für die Translation essentiellen Tat-Protein nachgewiesen wurde (Karn 1999).

Abschlussbewertung der Durchführbarkeit

Vor Durchführung der weiteren Arbeiten im Rahmen dieser Doktorarbeit stand somit bereits fest, dass die Zielstruktur prinzipiell für die folgenden Untersuchungen geeignet erscheint, jedoch aufgrund der Variabilität der Konformere der TAR ein Umgang mit der zur Zeit nicht eindeutig festlegbaren „besten“ Zielstrukturkonformation Schwierigkeiten erwarten lässt. Das Wissen um das Vorhandensein des natürlichen Liganden (Tat-Protein und des daraus abgeleiteten Tat-Peptids) und weniger weiterer bekannter TAR-Liganden (z. B. Acetylpromazin, Neomycin, Argininamid), lässt aber die Hoffnung zu weitere mit TAR interagierende Moleküle identifizieren zu können.

Erste zu wählende Strategien

Es existierten jedoch zum Zeitpunkt des Beginns dieser Arbeit nur wenige frei verfügbare Informationen über die Wechselwirkung kleiner wirkstoffartiger Moleküle mit TAR, im speziellen gab es nur wenige Liganden die mit der TAR-RNA zusammen strukturaufgeklärt vorlagen. Von diesen ist auch noch bekannt, dass sie unterschiedliche Konformationen der TAR induzieren. Da somit nicht eine einzige Zielstruktur für das gezielte Liganden-Design vorlag, wurde hieraus geschlossen, dass eine einzige Konformation nicht als Referenzstruktur gewertet werden kann und dementsprechend primär die Liganden nach Gemeinsamkeiten begutachtet werden sollten, um eine ersten Ansatz abzuleiten. Hierzu können letztendlich auch TAR-RNA-Liganden eingesetzt werden, für die zwar keine TAR-Bindungs-Strukturdaten vorliegen, für die jedoch anderweitig experimentell bestätigt wurde, dass sie die Tat-TAR-Interaktion durch Bindung an die RNA inhibieren. Sollte sich zeigen, dass die zu erhaltende Strukturdatensatzgröße eine Erstellung eines – auf den Gemeinsamkeiten bekannter TAR-Liganden oder TAR-Nichtbindern basierenden – Filtersystems nicht nahe legt, ist eine Abstraktion auf die nächst-niedrigere Präzisions-Ebene anzustreben. Dieser Ansatz würde die Gesamtheit der Informationen früherer Ansätze zu kleinen wirkstoffartigen RNA-Liganden – beliebiger RNA-Zielstrukturen – beinhalten. Anstelle eines „Tat-TAR-Interaktions-spezifischen RNA-Liganden“-Filtersystems, würde man dann zumindest möglicherweise ein „unspezifisches RNA-Binder“-Filtersystem ableiten können.

Des Weiteren sollte die Möglichkeit einer begründbaren Definition der Auswahl von Referenzkonformeren – zur Ergänzung der Liganden-basierten Ansätze – erforscht werden. Dies erfolgt um später eine Zusammenführung der gewonnenen Liganden-basierten Informationen mit denen der vorhandenen Zielstruktur-basierten Informationen zu ermöglichen, welche möglicherweise neue Einblicke auf die Interaktionsmöglichkeiten von kleinen Liganden mit RNA-Zielstrukturen ermöglicht.

Erfolgte Umsetzung im Rahmen der Arbeit

Im Rahmen dieser Arbeit kam es zum Umsetzen der folgenden Ansätze:

- „Unspezifisches RNA-Binder“-Filtersystem
- „Tat-TAR-Interaktions-spezifisches RNA-Binder“-Filtersystem
- Untersuchungen zur Eingrenzung eines möglichen Referenzkonformer

2. Allgemeine Methoden

Die folgenden Abschnitte enthalten Methoden die im Rahmen dieser Arbeit vermehrt eingesetzt wurden. Diese Abschnitte dienen der ergänzenden Information zu den in den einzelnen Projektphasen angesprochenen Vorgehensweisen.

2.1 Molekulare Deskriptoren

Molekulare Deskriptoren sind Zahlenwerte für bestimmte Eigenschaften eines Moleküls oder die Kombination mehrerer dieser Eigenschaften zu einer kombinierten Größe.

Als molekularer Deskriptor kann jede bestimmbar – direkt messbare oder berechenbare – Größe dienen, welche mit dem zu beschreibenden Molekül zusammenhängt. Molekulare Deskriptoren werden anhand des zugrunde liegenden Molekül-Modells in zweidimensionale Deskriptoren (2D-Deskriptoren) und dreidimensionale Deskriptoren (3D-Deskriptoren) unterschieden. Beispielsweise ist die Summenformel $C_{10}H_{16}N_5O_{13}P_3$, welche aus Abbildung 5 abgeleitet werden kann, als eine Korrelationsvektorschreibweise für Adenosintriphosphat (ATP) zu sehen. Deskriptoren, die aus zweidimensionalen Abbildungen von Molekülen abgeleitet sind, werden dementsprechend 2D-Deskriptoren genannt. Wohingegen zur Berechnung der zugänglichen Oberfläche des ATP-Moleküls (Abbildung 6) eine Berechnung der dreidimensionalen Struktur (Konformation) nach einem Kräfte-Raum-Modell erforderlich ist, was diesen Wert somit zu einem 3D-Deskriptor macht. In dieser Arbeit wurden sowohl 2D- als auch 3D-Deskriptoren verwendet.

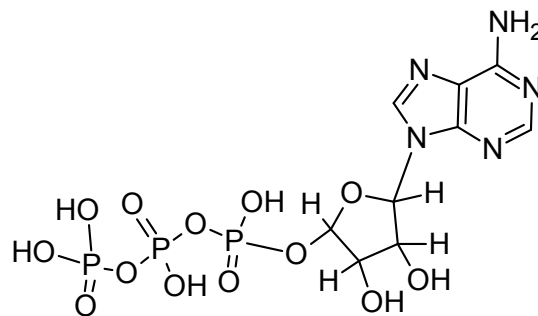


Abbildung 5: ATP als 2D-Darstellung

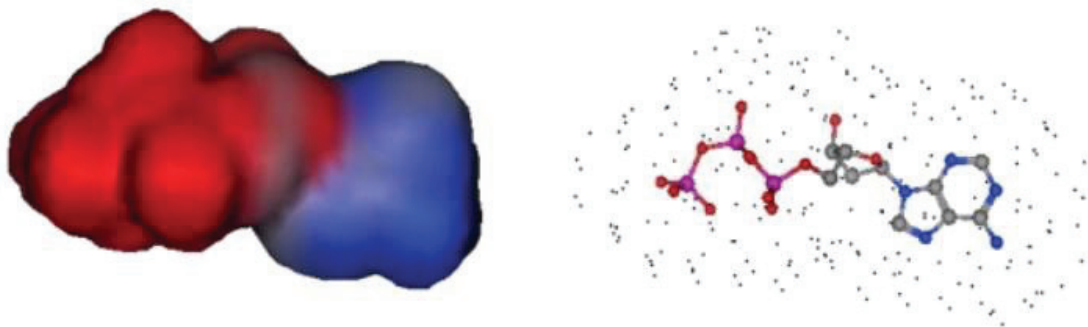


Abbildung 6: ATP 3D-Modell; links befindet sich die Darstellung der zugänglichen Oberfläche des Moleküls, rechts das zugrunde liegende energieminierte Molekülmodell

2.1.1 MOE-2D-Deskriptoren

Die Software MOE 2003.02 (Chemical Computing Group, Montreal) bietet eine Sammlung von 146 2D-Deskriptoren, welche in dieser Arbeit verwendet wurden. Die in dieser Sammlung aufgeführten Deskriptoren beziehen sich beispielsweise auf Eigenschaften wie «Durchmesser» und «Molekulares Gewicht», welche jeweils pro Molekül einen skalaren Wert ergeben. Für die jeweils im Rahmen dieser Arbeit verwendeten MOE-2D-Deskriptoren, sei hier auf die jeweiligen Anwendungs-Abschnitte verwiesen.

Insgesamt sind sieben verschiedene Klassen von MOE-2D-Deskriptoren vom Hersteller definiert worden:

- Physical Properties
- Subdivided Surface Areas
- Atom Counts and Bond Counts
- Kier&Hall Connectivity and Kappa Shape Indices
- Adjacency and Distance Matrix Descriptors
- Pharmacophore Feature Descriptors
- Partial Charge Descriptors

Die in MOE aufgrund von Partiaalladungsverteilungen erstellten Deskriptoren, basieren auf den zuvor in MOE berechneten Partiaalladungen. Die Partiaalladungen wurden für die Moleküle basierend auf dem Kraftfeld MMFF94x (Halgren 1996) erstellt, dies ist auch das verwendete Standard-Kraftfeld von MOE.

2.1.2 CATS-2D

Der CATS-2D-Deskriptor (Schneider et al. 1999) basiert auf der zweidimensionalen Struktur eines Moleküls und bildet diese in einem Korrelationsvektor ab. Hierzu werden die Distanzen – entsprechend der kürzesten Bindungsanzahl – zwischen Atompaaaren als Werte verwendet. Die Atome der Atompaaare, werden als **P**otentielle-**P**harmakophor-**P**unkte (PPP) betrachtet, deren fünf Grundeigenschaften wie folgt definiert wurden:

- Wasserstoffbrückendonator (D),
- Wasserstoffbrückenakzeptor (A),
- positiv geladen oder ionisierbar (P),
- negativ geladen oder ionisierbar (N),
- lipophil (L).

Diese Grundeigenschaften können 15-fach miteinander gepaart werden

(DD, DA, DP, DN, DL, AA, AP, AN, AL, PP, PN, PL, NN, NL, LL).

Die Anzahl solcher Paare, welche eine definierte Bindungsentfernung gemeinsam haben, werden in ein skaliertes Histogramm überführt (genormt durch Division mit der Anzahl der Nicht-Wasserstoffatome im Molekül), welches unter Verwendung des Intervalls von 0 bis 9 Bindungen zu einem 150-dimensionalen Korrelationsvektor führt.

2.1.3 TripleCharge3D (TC3D)

Der TripleCharge-Deskriptor (Fechner et al. 2003) basiert auf einer mit der Software CORINA 3.20 (Molecular Networks GmbH, Erlangen) erzeugten 3D-Struktur und bildet diese in einen Korrelationsvektor ab. Hierzu werden mit der Software PETRA 3.11 (Molecular Networks GmbH, Erlangen) die partiellen Atomladungen des Moleküls für jedes

Atomberechnet. Diese Ladungen werden für jedes Atompaar multipliziert, um einen Ladungswert für jedes Paar zu erhalten. Diese Ladungswerte werden entsprechend der Distanzen der Atompaare im Molekül in Gruppen zusammengefasst. Bei einer Gruppeneinteilung alle 0,16 Å und einer berücksichtigten maximalen Distanz von 16 Å ergibt sich ein 101-dimensionaler Korrelationsvektor. Es wird weiter zwischen den Ladungspaaren (+/+, +/- und -/-) bei der Gruppeneinteilung für jede gegebene Distanzgruppe unterschieden. Dies resultiert in einer Verdreifachung der Dimensionalität des Korrelationsvektors auf 303 Dimensionen.

2.2 Bewertungsfunktionen

2.2.1 Mathews Korrelationskoeffizient (MCC)

Der MCC (Gleichung 2.1) bietet eine Abschätzung der Genauigkeit einer binären Klassifizierung auch dann, wenn der Datensatz eine asymmetrische Verteilung der Klassen aufweist (Matthews 1975). Sein Wertebereich erstreckt sich von -1 bis +1. Ein Wert von 0 entspricht einer zufälligen Klassifizierung. Die Grundlage des MCC ist die Klassifizierung von Ausgabewerten in folgende vier Klassen:

- korrekt-Positive (P)
- korrekt-Negative (N)
- falsch-Positive (O)
- falsch-Negative (U)

Als korrekt klassifizierte Strukturen wurden Substanzen betrachtet, deren durch die Umwandlung des Score-Wertes erhaltene Klassenwerte auch dem *a priori* zugeordneten Klassenwert der Substanz entsprachen. Als falsch klassifizierte Substanzen erhielten somit einen von ihrer Klasse abweichenden Klassenwert.

$$MCC = \frac{PN - UO}{(N + U)(N + O)(P + O)(P + U)} \quad (2.1)$$

2.2.2 „Prozent korrekte Vorhersage“

Eine Angabe „Prozent korrekte Vorhersage“ besagt, wie viele der Strukturen im Datensatz korrekt klassifiziert wurden, egal zu welcher Klasse sie zugehörig waren. Dieses Kriterium ist weniger aussagekräftig als der MCC, wenn asymmetrische Datensätze vorliegen, da hier die *a priori* Wahrscheinlichkeiten für das Vorliegen der Klassen nicht gleich sind.

$$\text{Prozent korrekte Vorhersage} = \frac{P + N}{S} \times 100, \quad (2.2)$$

wobei S die Gesamtmenge der untersuchten Moleküle darstellt ($S = P + N + O + U$).

2.2.3 Mittlerer Quadratischer Fehler (MSE)

Beim MSE (Balaban 1982) handelt es sich - im Rahmen dieser Arbeit - um den mittleren quadratischen Fehler der Ausgabewerte der ANNs relativ zu den gewünschten Zielwerten, der zum Netztraining verwendet wurde. Hierzu muss man folgende Parameter definieren:

- \vec{x} ist der die Eingabewerte (Deskriptoren) enthaltende Vektor,
- o_{erwartet} ist der gewünschte Ausgabewert für einen Eingabevektor \vec{x} ,
- o_{realer} der berechnete Ausgabewert für einen Eingabevektor \vec{x} .

Abhängig vom Eingabevektor \vec{x} erhält man einen Ausgabewert. Gepaart mit den Erwartungswerten (Zielwerte) lässt sich somit für jede Konstellation aus Eingabewerten der dazugehörige Fehler in der Ausgabe als Differenz des tatsächlichen Ausgabewertes o_{realer} und des erwarteten Ausgabewertes o_{erwartet} bilden.

$$\text{mittlerer quadratischer Fehler} = \text{MSE} = \frac{1}{S} \sum_{i=1}^S (o_{\text{real}} - o_{\text{erwartet}})^2, \quad (2.3)$$

2.3 Kolmogorov-Smirnov (KS) Statistik

Die Kolmogorov-Smirnov-Statistik (KS-Statistik) (Byvatov und Schneider 2004) wurde angewendet, um unter den verwendeten Deskriptoren diejenigen auszuwählen, welche einen möglichst hohen Informationsgehalt für die Trennung der beiden Klassen liefern (RNABinder/ProteinBinder; Tat-TAR-FRET-Assay-Aktive/Tat-TAR-FRET-Assay-Inaktive). Dies geschah, um die Rechengeschwindigkeit zu erhöhen, dabei jedoch möglichst wenig des gesamten Informationsgehalts der verwendeten Deskriptorensätze für die Klassifikation zu verlieren.

Zusätzlich ermöglicht die KS-Statistik die Bestimmung der als Deskriptor kodierten Eigenschaften, die für ein gegebenes Trennproblem am augenscheinlichsten die größten Unterschiede aufweisen. Diese Unterschiede können einen Hinweis darauf enthalten, welche Eigenschaften wahrscheinlich die Aktivitätsunterschiede in den Assay-Ergebnissen herbeigeführt haben. Diese Informationen können dann dazu eingesetzt werden, um die Pharmakophormodelle zu erstellen oder bestehende weiter zu verfeinern.

2.3.1 Allgemeines zur KS-Statistik

Die KS-Statistik bestimmt das Maß der Unähnlichkeit für die Verteilung zweier Eigenschaften, ohne auf Modellvorstellungen des Verlaufs angewiesen zu sein. Bei der KS-Statistik werden folgende Bedingungen erwartet:

- ein Datensatz mit zwei definierten Klassen,
- Zahlenwerte, welche zu den Individuen der Klassen gehören (Deskriptor).

Die individuellen Zahlenwerte lassen sich auf einen Zahlenstrang projizieren. Ob ein solcher Deskriptorwert generell in der Lage ist, die Klassen zu trennen, zeigt sich durch folgende Berechnungsschritte:

1. Bestimmung von Minimum und Maximum des aus den Einzelwerten der beiden Klassen kombinierten Zahlenstrangs.
2. Projektion der Einzelwerte beider Klassen auf den kombinierten Zahlenstrang.

3. Beginnend vom Minimum wird schrittweise für jeden gefundenen Wert auf dem Strang - jeweils getrennt nach der Klasse - bestimmt, wie viel Prozent der Werte der Klasse bereits gefunden wurden. Dieser Wert wird auf der Ordinate aufgetragen.
4. Bestimmung des maximalen Abstands der Kurven der beiden Klassen.

Die Formel zur Berechnung der KS-Statistik für die RNABinder-Datensätze lautet:

$$K^* = \max_{-\infty < x < \infty} |RNABinder(x) - ProteinBinder(x)|, \quad (2.4)$$

wobei

$RNABinder(x)$ der Funktionsverlauf der RNABinder-Klasse

und

$ProteinBinder(x)$ der Funktionsverlauf der ProteinBinder-Klasse ist.

Zu Punkt 3 und 4 siehe den Graph in Abbildung 7.

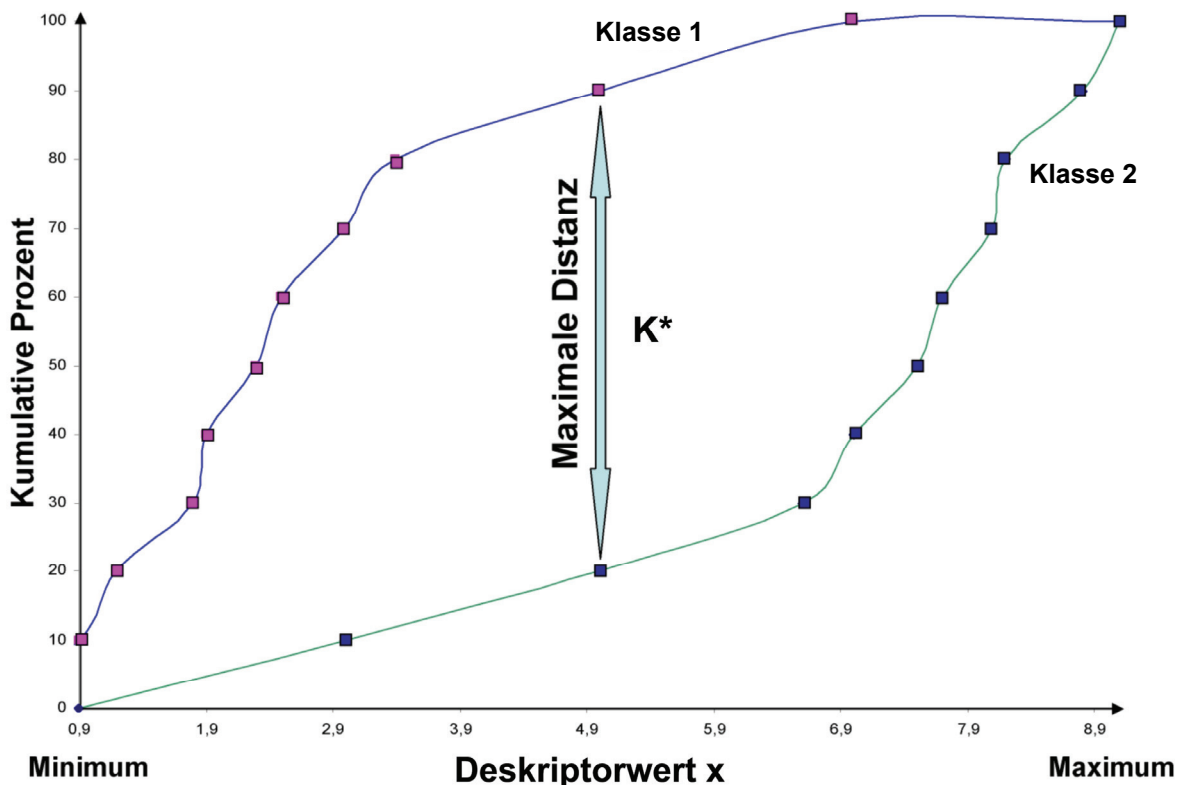


Abbildung 7: Beispiel eines Graphen einer KS-Statistik

2.3.2 Deskriptoren-Auswahl mit KS-Statistik

Zur Durchführung der KS-Statistik wurde ein Skript in Matlab erstellt. Mit diesem können Trainingsdatensätze darauf untersucht werden, ob die berechneten Deskriptoren für die Trennung der Klassen geeignet sind und welchen relativen Anteil an der gesamten Trennleistung jeder einzelne Deskriptor potentiell beitragen könnte.

Die Ergebnisse der KS-Statistik werden zuerst graphisch dargestellt, um einen Überblick über die Trennbarkeit der Klassen zu erhalten. Hiernach werden die Ergebnisse nach ihrer relativen Distanz sortiert und wieder graphisch dargestellt, um anhand des Kurvenverlaufs eine Einteilung nach der relativen Wichtigkeit der Deskriptoren vorzunehmen.

In der späteren Phase der Arbeit wurde für diese Arbeit KSandra verwendet, wobei es sich hier um ein Java-Programm zur Durchführung einer KS-Statistik handelt. Dieses Programm wurde speziell für die Textdatei-basierte Deskriptordatensatz-Verarbeitung geschrieben und vereinfacht die Auswahl geeigneter Deskriptoren für die weitere Verwendung in anderen Programmen.

2.4 Künstliche Neuronale Netze (ANNs)

Die Erstellung und das Training der in dieser Arbeit verwendeten künstlichen neuronalen Netze (ANNs – „Artificial Neuronal Network“) wurde durch die Benutzung folgender Software-Pakete umgesetzt:

Neural Network Toolbox 4.0.6 (The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098, U.S.A.) aus dem Software-Paket Matlab (*Matlab* 7.1.0.183 (R14) Service Pack 3 August 02, 2005).

2.4.1 Allgemeine ANN Definition:

Bei den verwendeten ANNs handelt es sich um zweilagige hetero-assoziative Feed-Forward Multi-Layer Perzeptrons (FFMLP); die von jedem ANN modellierte Gesamtfunktion ist in Gleichung 2.6 angegeben.

$$f(\text{net}_{out}) = o = f_{\log sig} \left(b_i + \sum_j w_{ij} f_{\log sig} \left(b_j + \sum_k w_{jk} x_k \right) \right), \text{ wobei} \quad (2.6)$$

x = Deskriptor-Vektor (k-dimensional)

w = Gewichtsvektor, der HL (k-dimensional) oder OL (HL-dimensional)

b = Bias-Vektor, der HL oder OL

o = Ausgabe des FFMLP

$f_{\log sig} = \text{logsig}(x) = (1 + \exp(-x))^{-1}$, sigmoide Übertragungsfunktion begrenzt die HL-Neuron Ausgabe auf den Intervall]0,1[

Als alternative Übertragungsfunktion kann auch f_{pureline} in der Ausgabeschicht verwendet werden, diese wird jedoch nur bei der direkten Vorhersage von Aktivitätswerten benutzt (Der Umfang dieser Arbeit beschränkt sich jedoch auf binäre Vorhersagen).

$f_{\text{pureline}} = \text{pureline}(x) = x$

Die ANNs bestehen allgemein aus einer Eingabe-Schicht (Input-Layer = IL), einer verborgenen Schicht (Hidden-Layer = HL) und einer Ausgabe-Schicht (Output-Layer = OL), die in einer geradlinigen/gerichteten Weise miteinander verknüpft sind.

Die Eingabe-Schicht dient dazu die einzelnen Eingabewerte des Deskriptorvektors an die einzelnen Neuronen in der verborgenen Schicht zu verteilen.

Jedes einzelne Neuron in der HL multipliziert darauf den eintreffenden Deskriptorwert x_i mit seinem für diese Deskriptorvektorposition i spezifischen Gewichtswert w_i . Anschließend werden diese Produkte durch das Neuron aufsummiert und der Bias (ein spezifischer Wert jeweils für jedes Neuron) zu dieser Summe hinzugezählt. Diese kombinierte Summe wird über eine Übertragungsfunktion (eine sigmoide oder lineare Funktion) in den Ausgabewert dieses Neurons umgewandelt.

Alle Ausgabewerte der Neuronen der HL werden als neue Eingabevektoren für das OL verwendet. Das OL verarbeitet die eingehenden Werte entsprechend demselben Algorithmus wie das HL. Eine Abweichung stellt hier jedoch das Vorhandensein nur eines Neurons in der OL dar. Dessen Ausgabewert stellt somit den Ausgabewert der in dieser Arbeit verwendeten Multi-Layer-Perzeptrons dar.

2.4.2 Preprozessierung der Deskriptordaten:

Die Daten wurden deskriptorweise auf einen Wertebereich von 0 bis 1 normiert. Die hierbei verwendeten Minima und Maxima wurden zur späteren Verwendung bei der Verarbeitung neuer Daten abgespeichert. Die Durchführung erfolgte am Anfang der Arbeit mit den Bordmitteln von Matlab, später erfolgte dies in Pandora.

2.4.3 Datensatz-Generation:

Die benutzten Deskriptordatensätze, wurden zu Kreuzvalidationszwecken zehnfach per Zufallsgenerator in gleichgroße Teilmengen zerlegt (50/50-Zerlegung). Diese Teilmengen wurden dann jeweils als Trainings- und Validationsdatensatz eingesetzt.

2.4.4 Generation der ANNs:

Ensembles von jeweils 450 ANNs wurden für jedes der Kreuzvalidierungssets erstellt. Diese 450 ANNs setzen sich wiederum aus jeweils 50 ANNs einer bestimmten Anzahl an Neuronen in der HL zusammen (2-20 Neuronen in der HL).

2.4.5 Training der ANNs:

Im Rahmen dieser Arbeit wurde der gdx-Trainingsalgorithmus von Matlab verwendet (gradient descent with momentum = Gradientenabstiegsverfahren mit Impulsterm). Dieser adaptiert die Gewichte und die Biase der erstellten ANNs zur Verringerung des mittleren quadratischen Fehlers (MSE) der Ausgabewerte der FFMLPs zu den Zielwerten.

Die Gewichte und Biase der ANNs wurden zufällig mit Werten aus dem Intervall $[-1, 1]$ initialisiert. Basierend auf Voruntersuchungen wurden folgende weitere Parameter verwendet:

- Als maximale Trainingsepochenanzahl wurde 10.000 festgesetzt.
- Die Initial-Lernrate wurde auf 0,05 gesetzt.
- Der MSE-Zielwert wurde auf 10^{-5} gesetzt.
- Die maximal zulässige Anzahl für aufeinander folgende Verschlechterungen des MSE wurde auf 500 gesetzt.

2.4.6 MCC als Gütekriterium der Vorhersage der einzelnen ANNs

Da innerhalb der Arbeit mit Aktivitätsklassen und nicht mit Bindekonstanten gearbeitet wurde, musste die Klassen-Kodierungen (Zahlwerte 0 oder 1) aus den Score-Werten der ANNs abgeleitet werden. Dies erfolgte mittels Runden des Ausgabewertes der ANNs auf die Klassenwerte 0 oder 1. Mit den so gewonnenen Klasseninformationen konnte der MCC für die Vorhersagegenauigkeit der ANNs berechnet werden. Die berechneten MCC-Werte für die Teilmengendatensätze wurden zur Bewertung der Vorhersagegüte der ANNs eingesetzt.

2.4.7 Selektion der besten ANNs für jeden Kreuz-Validationsdatensatz:

Basierend auf dem kombinierten Abschneiden bei der Bewertung der erreichten MCC-Werte für Trainings- und Validationsdatensatz wurde das beste ANN ausgewählt.

2.4.8 ANN-Jury-Zusammenstellung

Die jeweils besten Netze der einzelnen Kreuz-Validierungs-Trainingsrunden wurden zu einer Jury kombiniert. Die einzelnen Vorhersageergebnisse der ANN-Jury-Mitglieder werden gemittelt um einen einzigen Wert für einen bestimmten Deskriptorvektor (der jeweils ein Molekül kodiert) zu erhalten.

2.4.9 ANN-Jury-Vorhersagen

Die oben erwähnten ANN-Jurys wurden hiernach eingesetzt um die jeweilige Aktivitäts-Klasse für die Einträge der Chemikalien-Datensätze der Lieferanten prospektive vorherzusagen.

2.5 Selbstorganisierende Karten (SOM)

Hintergrund:

Selbstorganisierende Karten (**Self-Organizing Map** = SOM), sind eine Unterart der ANNs, die die Aufgabe übernehmen können höher-dimensionale Daten auf eine niedriger-dimensionale Repräsentation zu projizieren. Erstmals vorgestellt wurde dieses Verfahren von Kohonen (Kohonen 1982) beschrieben.

Bei den SOMs erfolgt die Adaptation ohne Vorgabe von spezifischen Zielwerten, so dass diese ANNs zur Typologie der ANNs mit unüberwachtem Lernverhalten gehören. In der Regel werden den erstellenden Programmen nur die hochdimensionalen Daten übergeben, sowie die Parameter festgelegt, wie die Topologie der resultierenden Karte aussehen soll. Die Erstellung einer SOM erfordert zwei Schritte: das Training und die Projektion der Daten auf die Karte.

Im ersten Schritt werden durch die Trainingsalgorithmen die Neuronen der Karte (Vektoren des höherdimensionalen Raums) angepasst.

Im zweiten Schritt werden mittels Nachbarschaftsanalyse (Euklidische-Distanz, Manhattan-Distanz, ...) die diesem Neuron zugehörigen Eingabedatenpunkte – repräsentiert durch ihre hochdimensionalen Vektoren – zugeordnet. Hieran anschließend kann die Häufigkeitsverteilung der Eingabedaten innerhalb der Neuronen in der beispielsweise 2-dimensionalen Kartenrepräsentation als Farbkodierung angezeigt werden. In den Karten ist jedem Neuron ein Feld zugewiesen, dass für die Visualisierung von Eigenschaften (Populationsverteilungen, die mittlere Abweichung der verwendeten Datenpunkte zum Neuronvektor, etc.) verwendet werden kann. In einigen Fällen kann die erstellte Karte auch

als Interface benutzt werden, um an die einem Neuron/Cluster zugehörigen Datenpunkt-Bezeichnungen zu gelangen.

Anwendung im Rahmen der Arbeit:

Im Rahmen dieser Arbeit wurden die SOM-Programme MOLMAP® von Prof. Dr. Gisbert Schneider (Schneider und Wrede 1998) und SOMMER von Michael Schmuker (Schmuker et al. 2007) eingesetzt.

Der Haupteinsatzbereich galt der vergleichenden Darstellung von Datensätzen mit hochdimensionalen Deskriptoren.

Mittels der Ausgabedateien von MOLMAP® wurde ebenfalls ein Modul für *Pandora* implementiert, das diese Eingabedateien verarbeiten kann, um die darin enthaltenden Clusterinformationen für die Annotation der Datenbank zu verwenden. Mittels dieser Kombination von SOM und Datenbanksystem ist es möglich, die nach den SOMs repräsentativen Konformere/Taschen aus der Datenbank zu extrahieren. Die Extraktion der Repräsentativen wurde in einem weiteren Modul von Pandora umgesetzt. Das Genannte gilt ebenso für die Extraktion ganzer Knoten-Populationen.

Ausblick:

Da in der Datenbank nicht nur die Informationen einer einzelnen Karte abgelegt werden können, sondern Pandora die Verwaltung verschiedener Karten erlaubt, wird es in Zukunft möglich sein, mittels SQL-Abfragen relativ einfach verschiedene Karten für die Selektion zu kombinieren. Denkbar ist hier die Anwendung als kombinierte Suche nach ähnlichen Taschen, die beispielsweise in ihrer Form (PocketPicker-Deskriptor) und Pharmakophorbeschreibungen (SIMPrint, LIQUID, etc.) bestimmte Eigenschaftskombinationen aufweisen sollen.

Da Pandora so ausgelegt wurde, dass generell Struktur-Informationen beliebiger Art abgelegt werden können – sobald ein Einlesemodul für diese Daten existiert – ist dieses System auch auf Liganden-Datensätze anwendbar. In Zukunft, soll die umgesetzte Docking-Vorbereitung auch um die Erstellung der Liganden-Datensätze erweitert werden, welche mittels SOM-Analyse beispielsweise Toxizitätsinseln im chemischen Raum der potentiellen Liganden automatisch aussortieren könnte. Dies gilt ebenso für weitere Auswahlkriterien wie beispielsweise die Löslichkeit, solange sich diese Merkmale als relativ zuverlässiger Deskriptoren umsetzen lassen.

2.6 Hierarchical-k-means – hk-means-Clusterung

Bei Hierarchical-k-means handelt es sich um ein Programm von Alex Böcker (Böcker et al. 2005). Es ist in der Lage Eingabe-Daten anhand des hk-mean-Algorithmus automatisiert zu Clustern und ermöglicht die Darstellung der Ergebnisse dieser Clusterung als graphische Baumansicht. Die Blätter des Baums enthalten dabei die Bezeichner der jeweiligen Datenpunkte im verwendeten Datensatz.

Anwendung im Rahmen der Arbeit:

Im Rahmen dieser Arbeit wurde Hierarchical-k-means dazu eingesetzt, um eine einfache Visualisierung der Ähnlichkeitsuntersuchung verschiedener Deskriptoren umzusetzen. Hierbei wurde die Software bei der Analyse der Taschen- und Konformervergleiche verwendet, um die anderen eingesetzten Verfahren (RMSD-Distanzmatrizen, SOMs) durch die Visualisierungsmöglichkeit der Baumansicht zu ergänzen; siehe Abschnitt 4.6.3.6.

Ausblick:

Es stehen zur Topologie der jeweils erstellten Bäume auch die erstellten Cluster-Mitgliederdaten in Form einer Textdatei zur Verfügung. Diese Datenaustausch-Option könnte in weiteren Projekten eingesetzt werden, um die darin erhaltenen Informationen für das Annotieren der Datenbank zu verwenden.

2.7 Der Tat-TAR-FRET-Assay

2.7.1 Experimentelle Grundlagen

Fluorescence Resonance Energy Transfer (FRET)

Der im Rahmen dieser Studie eingesetzte Assay basiert auf der Technik der Fluoreszenz-Donor-Akzeptor-Komplexe (Lottspeich und Zorbas 1998). Er beinhaltet ein Fluoreszenz-Akzeptor-Donor-Paar (Fluoreszein und Rhodamin), das zur Detektion verwendet wird. Dieses Paar ist in diesem Fall an eine dem Tat-Protein entnommene Peptid-Sequenz (Tat-Peptid, siehe Abbildung 8) endständig gebunden. Die Modifikationen entsprechen denen von Matsumoto (Matsumoto 2000). Das markierte Peptid kann aufgrund seiner dem Tat-Protein entnommenen Sequenz an die TAR-RNA binden (Karn 1999). Hierbei wird von der Annahme ist ausgegangen, dass die RNA das Peptid in einer gestreckten Form bindet. Dadurch wird das Fluoreszenzpaar räumlich getrennt, und eine beobachtbare Zunahme der Fluoreszenz des Fluoreszeins ist die Folge. Aus dem Verhältnis der Fluoreszenzintensität in Abwesenheit und bei Anwesenheit der TAR-RNA kann auf den Anteil an gebundenem zu ungebundenem Tat-Peptid geschlossen werden. Sollten die Messungen mit zugesetzten potentiellen Liganden eine Veränderung der beobachteten Fluoreszenzaktivität relativ zu den vorher bestimmten Referenzwerten ergeben, so kann davon ausgegangen werden, dass diese Veränderung im Gleichgewicht der freien und gebundenen Form durch den Liganden verursacht wurde. Somit ist eine Abschätzung der möglichen Störung der Tat-TAR-Interaktion anhand dieses Assay möglich. Direkte Aussagen zu den Bindekonstanten sind jedoch aus diesem Assay nicht ableitbar, da es sich nur um einen funktionalen Verdrängungsassay handelt, bei dem die gemessene Wirkung sich auf das Verhältnis des freien und ungebundenen Peptids bezieht und nicht um die Bindung der potentiellen RNA-Liganden an TAR. Sollte die Bindung der potentiellen RNA-Liganden an TAR keine Änderung der Affinität von TAR zum Peptid bewirken, kann diese Liganden-Bindung folglich nicht innerhalb dieses Assay nachgewiesen werden.

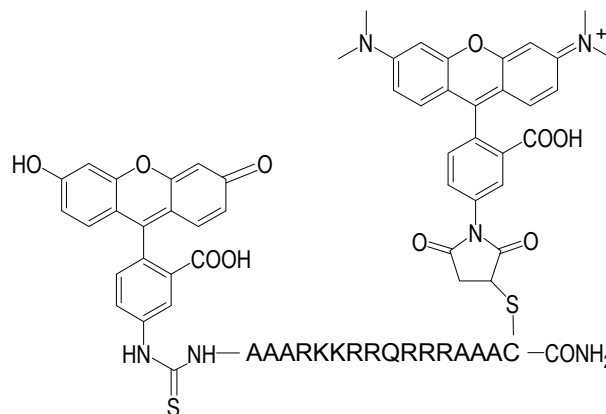


Abbildung 8: Fluoreszenzpaar-markiertes Tat-Peptid des verwendeten FRET-Assay

2.7.2 Durchführung der Messungen

Substanzlieferung

Die Substanzen werden von dem Chemikalienversandhändler SPECS in einem Paket mit 20 kg Trockeneis tief gefroren in ihrer DMSO-Lösung auf einer Platte in durch Gummipfropfen verschlossenen Vials geliefert.

Da die Platte keine Substanznamen sondern nur Reihen- und Spaltenangaben und Plattennummern beinhaltet, muss vor der Weiterverwendung die beiliegende Beschreibungsdatei ausgewertet werden um ein geeignetes Zuordnungssystem für das Labor zu schaffen, diese Indexierung dient dann der weiteren Etikettierung während der Versuche und späteren Auswertung.

Auspacken

Die Platte mit den Substanzen wird aus den oben genannten Gründen vom Paket in den „-80°C“-Gefrierschrank zur Lagerung verbracht.

Umfüllen und mit LaborIDs etikettieren

Nach dem Erstellen der Labels werden die etikettierten Tubes nach der Ordnung der Substanzen vorsortiert. Die Substanzen werden in ihren Platten bei Raumtemperatur langsam aufgetaut, da ansonsten beim schnellen zu starken Auftauen die Gummideckel bereits herausgedrückt werden können. Hiernach werden die Lösungen der Reihe nach in die vorbereiteten Tubes überführt.

Vorbereitung für das erste Screening

Erste Verdünnungsreihe – 1 mM Stammlösungen

Aus den erstellten Vorrattubes werden die 10-prozentigen DMSO-Lösungen abgeleitet, indem 100 µl der puren DMSO-Lösungen in neuen Tubes in 900 µl DEPC-behandeltem Wasser gelöst werden. Hierbei wird auch das Lösungsverhalten der Substanzen dokumentiert. Sollte es – wie es bei den bisherigen Bestellungen häufig der Fall gewesen ist – zu vermehrten Problemen mit ausfallenden Substanzen kommen, so wird eine Verdünnung mit Methanol vorgenommen, die dieses Löslichkeitsproblem deutlich reduziert.

Ansetzen der RNA-Stammlösung

Das Ansetzen der RNA-Lösungen, welche für den Assay benutzt werden, erfolgt aus Stammlösungen der RNA, die von Dr. Ute Scheffer aus dem Arbeitskreis-Göbel freundlicherweise bereitgestellt werden. Die RNA für die Lösungen wurde als Feststoff mit HPLC-Reinigung bei BioSpring (BioSpring GmbH, Frankfurt am Main, Deutschland; www.biospring.de) bestellt und enthält somit keine zusätzlichen Salze und wurde mit DEPC-behandeltem Wasser verdünnt.

Diese Stammlösungen enthalten je nach Charge eine unterschiedliche Konzentration der RNA, so dass aus diesen in einem ersten Verdünnungsschritt 1 µM-Stammlösungen zur einfacheren Handhabung erstellt werden.

Alle Verdünnungen erfolgen in Tubes vom Typ, „DNA LoBind Tube“ der Firma Eppendorf.

Die Gebrauchslösungen werden durch eine 1:10 Verdünnung aus den 1 µM-Stammlösungen zum Einsatz im Assay nach Bedarf vorbereitet.

Ansetzen der Peptid-Stammlösung

Das Ansetzen der Peptid-Lösungen welche für den Assay benutzt werden, erfolgt aus Stammlösungen des Peptids, die von Dr. Ute Scheffer aus dem Arbeitskreis-Göbel freundlicherweise bereitgestellt werden. Das Peptid für die Lösungen wurde als Feststoff mit HPLC-Reinigung bei Thermo Scientific (Thermo Fisher Scientific Inc.; www.thermo.com) bestellt und enthält somit keine zusätzlichen Salze und wurde mit DEPC-behandeltem Wasser verdünnt.

Diese Stammlösungen enthalten je nach Charge eine unterschiedliche Konzentration des Peptids, so dass aus diesen in einem ersten Verdünnungsschritt 1 µM-Stammlösungen zur einfacheren Handhabung erstellt werden.

Alle Verdünnungen erfolgen in speziellen Tubes vom Typ, „Protein LoBind Tubes“ der Firma Eppendorf.

Die Gebrauchslösungen werden durch eine 1:10 Verdünnung aus den 1 µM-Stammlösungen zum Einsatz im Assay nach Bedarf vorbereitet.

Ansetzen der TK-Puffer-Wasserlösung

Da im Assay der TK-Puffer in einfacher Konzentration vorliegen soll, wird zur Beschleunigung des Pipettierprozesses an der offenen Platte, bereits vor dem Beginn der Messungen eine zusammengesetzte Lösung des TK-Puffers mit der im Assay vorgesehenen DEPC-behandelten Wasser Menge erstellt, die dann im späteren Versuch direkt eingesetzt werden kann und so auch die Vermischung der Substanzen im Well beschleunigt. Hierzu werden ein Teil TK-10X-Puffer zu sechs Teilen DEPC-behandeltem Wasser hinzugegeben. Aufgrund der relativ großen benötigten Menge dieser Lösung im späteren Assay erfolgt dieser Ansatz im 14 ml Zentrifugen-Röhrchen. Die Lösung wird zum gründlichen Mischen gevortext und anschließend eine Weile stehen gelassen um das Austreiben der kleinen Luftblasen in der Lösung zu ermöglichen, die ansonsten den Assay durch Lichtstreuung verfälschen könnten. Dieser Vorgang kann auch durch ein Ultraschallbad beschleunigt werden.

Zusammensetzung der 10x TK-Puffer-Stammlösungen:

- 500 mM Tris-HCl,
- 200 mM KCl,
- 0.1% Triton X-100;

Eingestellt wurde ein pH von 7,4.

Vorbereitung des Arbeitsplatzes am Spektrometer

Da die Versuche bei 37°C im Spektrometer durchgeführt werden sollen, ist eine Vorbereitung des Arbeitsplatzes (siehe Abbildung 9) für das schnelle und saubere Pipettieren notwendige Bedingung für konstante Messergebnisse. Aufgrund der Zeit-/Temperaturabhängigkeit der Messergebnisse wird eine Umgebung gebraucht, die neben dem zügigen Arbeiten eine relative Temperaturkontrolle der eingesetzten Substanzen sowie der Lochplatte ermöglicht.

Hierzu wird ein Wärmeblock bereitgestellt in dem die zu verwendenden Substanz-, Puffer-, RNA- und Peptidlösungen auf die gewünschte Temperatur von 37°C vorgeheizt werden. Zur Temperaturisolation der Messplatte vom Labortisch wird eine Lage aus Papiertüchern eingesetzt. Des Weiteren werden die benötigten Pipettenspitzen sowie die Pipetten mit den benötigten Messvolumina bereitgelegt.



Abbildung 9: Fotos des Arbeitsplatzes für die Durchführung des Tat-TAR-FRET-Assay; Links der Platz an dem die Pipettierung der Lösungen auf die Mikrotiterplatte erfolgte. Rechts das Spektrometer mit dem Steuerungscomputer. Die beiden Arbeitsplätze befinden sich direkt gegenüberliegend im Labor.

Messungen am Gerät

Durchführung der ersten groben Screens

Zur Bestimmung der Substanzkandidaten, welche einen Effekt in dem Assay verursachen, wird ein erstes Screening bei einer Substanzkonzentration von 100 μM im Assay durchgeführt.

Hierbei werden die Substanzen jeweils in drei Vertiefungen der Mikrotiterplatte vermessen. Bei den verwendeten Mikrotiterplatten handelt es sich um: Corning 96 Well Half Area Black Flat Bottom Polystyrene NBS Microplate, without Lid, Nonsterile (cat-nr 3650).

Vorher werden die Fluoreszenz-Referenzwerte für das freie Peptid, als auch die RNA-Peptid-Mischung mittels Messung mehrerer Vertiefungen der Mikrotiterplatte bestimmt; siehe Abbildung 10. Der so bestimmte durchschnittliche Fluoreszenz-Referenzwert der freien Peptidlösung dient später als Basislinie der Versuchsauswertung und der Fluoreszenz-Referenzwert der RNA-Peptid-Lösung dient hierbei als maximale Aktivitätsgrenze, siehe hierzu auch den Abschnitt **Prozessierung der Rohdaten zu relativen Aktivitätswerten**. Aktivität im Assay ist somit definiert, als eine Verminderung der Fluoreszenzaktivität durch Zugabe der potentiellen Liganden.

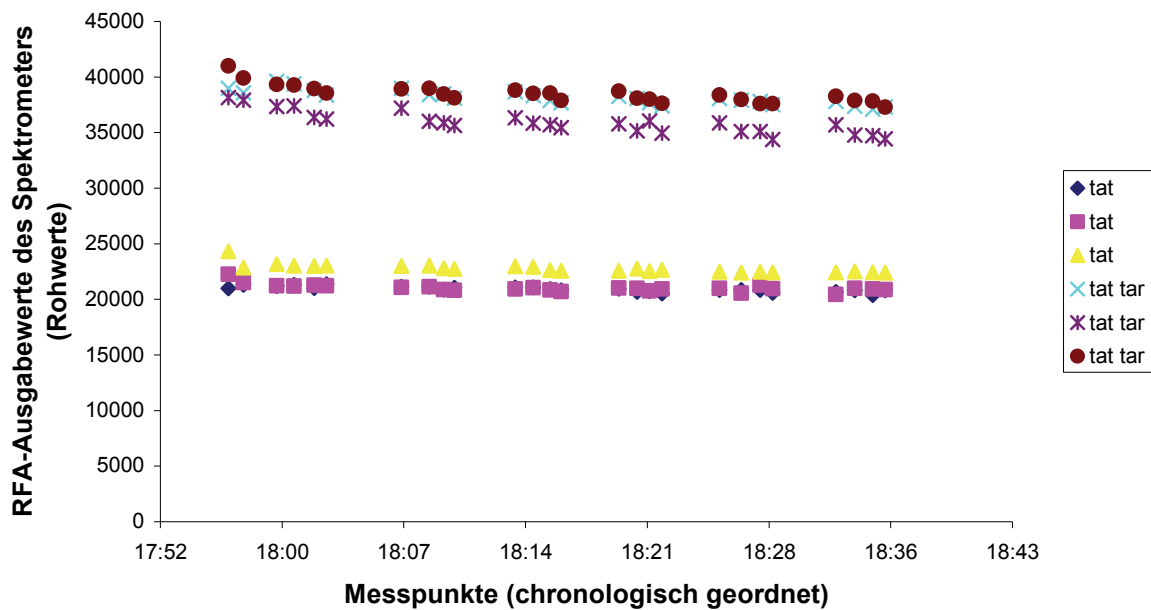


Abbildung 10: Tat-TAR-FRET-Assay-Beispielmessungen der Referenzen Tat und Tat TAR. Alle Messungen wurden parallel in sechs unterschiedlichen Vertiefungen der Mikrotiterplatte durchgeführt und zeitlich versetzt immer wieder gemessen. Die Sägezacken-Struktur die besonders bei den drei Tat-TAR-Messreihen zu sehen ist, wird hier auf das Ausbleichen des Fluoreszeins durch die Messung, oder das jeweilige Schütteln der Platte vor der Messung zurückgeführt. Beachtenswert ist hierbei jedoch die relativ größere Abweichung der Messwerte zwischen unterschiedlichen Vertiefungen (Pipettiergenauigkeit).

Um Effekte der verwendeten Lösungsmittel (DMSO und Methanol) aus den Messergebnissen der Substanzmessungen auszuschließen, werden die Referenzen mit freiem Peptid und RNA-Peptid-Mischung ebenfalls mit denselben Lösungsmittelkonzentrationen durchgeführt.

Pipettierschema für die freie Peptidreferenzlösung:

1. 70 μ l der Pufferlösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.
2. 10 μ l der Lösungsmittelreferenzlösung – Konzentration der Lösungsmittel wie bei den Ligandenlösungen – werden der Vertiefung der Mikrotiterplatte hinzugefügt.
3. 10 μ l DEPC-behandeltes Wasser als Volumenausgleich für die hierbei nicht verwendete RNA-Lösung werden der Vertiefung der Mikrotiterplatte hinzugefügt
4. 10 μ l der Peptidlösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.

Pipettierschema für die gebundene Peptidreferenzlösung/RNA-Peptid-Lösung:

1. 70 μ l der Pufferlösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.
2. 10 μ l der Lösungsmittelreferenzlösung – Konzentration der Lösungsmittel wie bei den Ligandenlösungen – werden der Vertiefung der Mikrotiterplatte hinzugefügt.
3. 10 μ l der Peptidlösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.
4. 10 μ l der RNA-Lösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.

Pipettierschema für die Aktivitätsbestimmung der Testsubstanzen in der RNA-Peptid-Lösung:

1. 70 µl der Pufferlösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.
2. 10 µl der Peptidlösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.
3. 10 µl der RNA-Lösung werden der Vertiefung der Mikrotiterplatte hinzugefügt.
4. 10 µl der Testsubstanzlösung (1 mM) werden der Vertiefung der Mikrotiterplatte hinzugefügt.

Durchführung der Bestimmung der IC_{50} -Werte für die erste Bestellrunde

Für die erste Runde wurden die IC_{50} -Werte der im Grobscreen als aktiv ermittelten Substanzen mittels Verdünnungsreihe bestimmt. Hierbei wurden sukzessive weiter Verdünnungen der Testsubstanzen erstellt und dann bei diesen verminderten Konzentrationen erneut Messungen in mehreren Vertiefungen der Mikrotiterplatte durchgeführt. Die gemittelten Aktivitätswerte der einzelnen Konzentrationen wurden als Messpunkte in ein gemeinsames Diagramm eingetragen und mittels Interpolation die Konzentration ermittelt, welche die relative Fluoreszenzaktivität auf einen Wert von 0,5 in diesem Assay reduziert wurden. Anschließend wurde diese so ermittelte Konzentration hergestellt und einem letzten Test unterzogen, ob die berechnete Konzentration experimentell auch den gewünschten Effekt zeigt.

Durchführung der verfeinerten Aktivitätsbestimmung für die zweite Bestellrunde

Anmerkungen:

Aufgrund starker Schwankungen der Aktivitätswerte zwischen einzelnen Vertiefungen der Mikrotiterplatte – selbst bei der Verwendung derselben Chargen der Lösungen – wurde hier auf eine intrinsische Referenzbestimmung in derselben Vertiefung der Mikrotiterplatte umgestiegen, in dem auch die Testsubstanz später gemessen wurde. Dieses Verfahren wird auch vom Arbeitskreis Göbel zur Bestimmung der Aktivität von Substanzen angewendet. Besonderer Vorteil hierbei ist die einfache Ausreißerbestimmung der Messung in einer Vertiefung der Mikrotiterplatte, indem das Abklingverhalten der Fluoreszenzaktivität hier direkt für die einzelne Vertiefung der Mikrotiterplatte zeitlich aufgelöst wird. Wie geschehen, können so einfach Vertiefungen der Mikrotiterplatte aus den Messreihen ausgeschlossen werden, die offensichtlich einen zu hohen RNA-Umsatz aufweisen, wie dies beispielsweise durch mögliche Verunreinigungen mit RNase geschehen kann, die sonst zu einem falsch positiven Befund geführt hätten; siehe hierzu Abbildung 11.

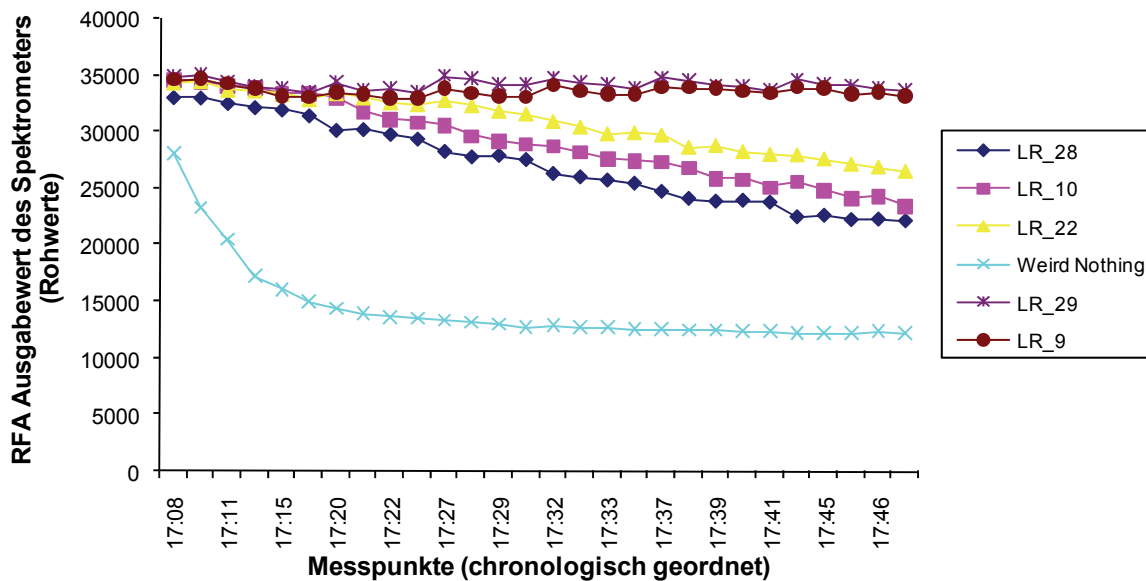


Abbildung 11: FRET-Assay – Titrationslauf mit Ausreißer-Messung: Wahrscheinlich beruht die türkisarbene Messkurve auf einer starken Verunreinigung der Vertiefung der Mikrotiterplatte mit RNAse und/oder „Fussel“. Ohne Überwachung der Equilibrierungsphase (erste sechs Schritte) wäre diese Stelle der Mikrotiterplatte ebenfalls für die Aktivitätsbestimmung herangezogen worden. Dies hätte dann zu einem falsch positiven Befund geführt. „Weird Nothing“ steht in diesem Fall für die Nicht-Hinzugabe von Testsubstanzlösung.

Durchführung:

Für die zweite Runde wurde eine Überprüfung der Aktivität der im Grobscreen als aktiv ermittelten Substanzen mittels Titration in einer Vertiefung der Mikrotiterplatte bestimmt. Hierbei wurde zunächst die relative Fluoreszenzaktivität in der Vertiefung der Mikrotiterplatte unter Ausschluss des Testsubstanzvolumens ermittelt und auf seine relative zeitliche Konstanz geachtet, welche durch mehrere zeitlich versetzte Messungen der Fluoreszenzaktivität in der Vertiefung der Mikrotiterplatte bestimmt wurde. Einem stabilem RNA-Peptid-Testsystem wurden dann sukzessive in 1 μ l- oder 2 μ l-Schritten Testsubstanzlösungen hinzugefügt und diese mehrfach zeitlich versetzt vermessen. Als Referenzkurve diente hierbei eine Kurve, welche durch Zugabe von purer Lösungsmittelösung derselben Lösungsmittelkonzentration analog mehrfach ermittelt wurde. Eine aktive Substanz kennzeichnet sich hier durch eine stufenweise Absenkung der Fluoreszenzaktivität jeweils nach Zugabe eines Aliquot an Testsubstanzlösung. Ein weiterer leichter Abfall der Messkurve ist bedingt durch die zeitliche Umsetzung der im Well enthaltenen RNA und somit Verminderung der Konzentration der RNA-Peptid-Komplexe. Als weiterer gewichtiger Faktor kommt das Ausbleichen des Fluoreszenzfarbstoffs durch schnell aufeinander folgende Messungen hinzu.

Die Referenzkurve steigt hingegen bei jeder Zugabe weiterer Volumina leicht an und zeigt dann ebenso wie bei einer aktiven Substanz eine zeitliche Umsetzung der im Well enthaltenen RNA durch Verminderung der Fluoreszenzaktivität. Der Anstieg ist sowohl bedingt durch die Annäherung der Oberfläche an den Messkopf des Spektrometers, als auch durch die größere Verdünnung der Komplexe, welche eine schwächere intermolekulare Quenchung der Komplexe bewirkt.

Aus der Differenz der Referenzkurve mit der Messkurve einer Testsubstanz kann somit die Aktivität, welche diese Testsubstanz verursacht, bestimmt werden.

Durch Mehrfachbestimmung dieser Kurven wird die statistische Sicherheit gewährleistet.

Durchführung der Bestimmung der IC_{50} -Werte für die zweite Bestellrunde

Für die zweite Runde wurden die IC_{50} -Werte, der im Grobscreen als aktiv ermittelten Substanzen, mittels Verdünnungsreihe genauer bestimmt. Hierbei wurden sukzessive weiter Verdünnungen der Testsubstanzen erstellt und dann bei diesen verminderten Konzentrationen erneut Messungen in mehreren Vertiefungen der Mikrotiterplatte durchgeführt. Die gemittelten Aktivitätswerte der einzelnen Konzentrationen wurden als Messpunkte in ein gemeinsames Diagramm eingetragen und mittels Interpolation die Konzentration ermittelt, welche die relative Fluoreszenzaktivität auf einen Wert von 0,5 in diesem Assay reduzieren würde. Anschließend wurde diese so ermittelte Konzentration hergestellt und einem letzten Test unterzogen, ob die berechnete Konzentration experimentell auch den gewünschten Effekt zeigt.

Rohdatenerfassung

Zur Erfassung der Rohdaten wurde die vorbereitete Mikrotiterplatte auf den dafür vorgesehenen Schlitten des Spektrometers gestellt und mittels Softwaresteuerung die Messung durchgeführt. Hierbei werden folgende Schritte automatisch nach Drücken des „Start Measurement“-Knopfes der Software durchlaufen.

1. Der Schlitten mit der Platte wird automatisch in das Spektrometer eingezogen.
2. Die Platte wird wie in den Voreinstellungen definiert geschüttelt.
3. Die Platte ruht wie in den Voreinstellungen definiert, um die Mischbewegung zur Ruhe kommen zu lassen.
4. Der Messkopf des Spektrometers fährt an die in der Software eingestellten Plattenbereiche (Vertiefungen der Mikrotiterplatte).
5. Die Messung wird wie in der Software eingestellt x-mal automatisch durchgeführt, und der Mittelwert der Messung einer Vertiefung wird im von der Software für die jeweilige Messung neu erstellten Excel-Tabellenblatt abgelegt.
6. Auswurf der Mikrotiterplatte per Ausfahren des Schlittens
7. Rückkehr zur Steuerungssoftware

Für die Steuerungssoftware des Safire2-Spektrometers wurden die in Tabelle 1 angegebenen Parameter verwendet:

Tabelle 1: Parametereinstellungen der Steuerungssoftware des Safire²-Spektrometers für den Tat-TAR-FRET-Assay; A1-A12 gibt den in diesem Beispiel angegebenen Plattenbereich für die durchzuführenden Messungen an (die gesamte erste Zeile der Platte)

SAFIRE II; Serial number: 12904200051; Firmware: V 1.35 08/2005 Safire2; XFLUOR4SAFIREII Version: V 4.62n	
Measurement mode:	Fluorescence
Excitation wavelength:	489
Emission wavelength:	590
Excitation bandwidth:	20
Emission bandwidth:	20
Gain (Manual):	122
Number of reads:	4
FlashMode:	High sensitivity

Integration time:	40
Lag time:	0
Plate definition file:	COS96fb.pdf
Part of the plate:	A1 - A12
Z-Position (Manual):	6748
Time between move and flash:	3
Shake duration(Orbital Medium):	5
Shake settle time:	3
Target Temperature:	37
Current Temperature:	37

Auswertung der Messdaten

Prozessierung der Rohdaten zu relativen Aktivitätswerten:

Um relative Aktivitätswerte aus den Assay-Rohdaten zu erhalten, um die Ergebnisse mehrerer Messtage oder auch um unterschiedliche Chargen von RNA- und Peptidlösungen miteinander direkt vergleichen zu können, bedurfte es einer Transformation der Rohdaten aus dem Spektrometer auf eine relative Werteskala. Dies wurde mittels der „Range Scaling“-Methode erreicht, bei der die Messergebnisse wie folgt weiterverarbeitet wurden:

1. Man bestimmt den kleinsten Referenzwert aus den Messungen und definiert ihn als Basislinie.
2. Man bestimmt den größten Referenzwert aus den Messungen und definiert ihn als maximalen Aktivitätswert.
3. Hiernach wird von jedem Messwert, der aus derselben Messreihe stammt, die Basislinie abgezogen, um einen Zahlenstrang zu generieren, der Null als Minimum aufweist.
4. Anschließend wird jeder so erhaltene Wert noch durch die Distanz zwischen Minimalwert und Maximalwert der Rohdaten geteilt, um eine unverzerrte Projektion des originalen Datenraums auf einen Wertebereich zwischen 0 und 1 zu erreichen.

Die so gewonnenen relativen Aktivitätswerte werden dann benutzt um die relativen IC_{50} -Werte der Testsubstanzen zu bestimmen, soweit diese im Messbereich zu bestimmen sind.

Die Werte des „Safire²“-Spektrometer wurden nach folgender Formel der „Range-Scaling“-Methode transformiert um relative Fluoreszenzaktivitätswerte zu ermitteln:

$$x_{ij}^* = \frac{x_{ik} - x_k(\min)}{x_k(\max) - x_k(\min)} \quad (2.7)$$

wobei

x_{ij}^* = Skalierte relative Fluoreszenzaktivität ; $0 \leq x_{ij}^* \leq 1$

x_{ik} = Rohdatenwert

$x_k(\min)$ = Minimum der gemessenen Fluoreszenzaktivität (maximales Quenschen; Maximum an freiem Peptid; Referenzmessung)

$x_k(\max)$ = Maximum der gemessenen Fluoreszenzaktivität (maximaler Fluoreszenzwert; Maximum an gebundenem Peptid; Referenzmessung)

Zusätzliche Messungen zur Bestimmung des Wechselwirkungscharakters vermeintlich aktiver Substanzen

Da es aufgrund der Wechselwirkung mit den Fluorophoren des Messsystems zu einer falschen Klassifizierung von Testsubstanzen kommen kann, werden für die genauere Analyse der Aktivität einer Messsubstanz Absorptionsspektren der Testsubstanzen aufgenommen, um mittels der so ermittelten Absorptionswellenlängenbereiche eine Interaktion mit den Fluorophoren Rhodamin und Fluoreszein des Messsystems auszuschließen. Hierzu wird überprüft ob die Substanz eine Absorption bei den Wellenlängen des Lichtes aufweist, die für das FRET-Messverfahren eingesetzt werden.

Durchführung:

Die Testsubstanz-Lösungen werden zusammen mit dem TK-Puffer durch Zugabe von Wasser auf dieselbe Konzentration wie im FRET-Assay gebracht. Einziger Unterschied ist hierbei die Abwesenheit von Tat-Peptid und TAR, deren Lösungen durch die gleiche Menge Wasser ersetzt wurden. Als Referenz wird eine TK-Puffer/Wasser-Lösung eingesetzt bei der das Volumen der Testsubstanzlösung durch eine Referenzlösung gleicher Lösungsmittelzusammensetzung (Methanol-/DMSO-Anteil) ergänzt wird.

Die Referenz- und die zu testende Substanzlösung wurden in den Küvetten für das (im Arbeitskreis Göbel) vorhandene „Cary 100“-Spektrometer überführt und in das Gerät eingesetzt. Das Spektrum der Referenzzelle wurde zur automatischen Basislinienkorrektur verwendet. Die Aufzeichnung der Spektren erfolgte per Steuerungssoftware. Die weitere Auswertung erfolgte mit Excel.

Für die Steuerungssoftware wurden folgende Parameter verwendet:

Scan Software Version: 3.00(182)

Parameter List:

Instrument	Cary 100
Instrument Version	9,00
Start (nm)	800,00
Stop (nm)	355,00
X Mode	Nanometers
Y Mode	Abs
UV-Vis Scan Rate (nm/min)	600,000
UV-Vis Data Interval (nm)	1,000
UV-Vis Ave. Time (sec)	0,100
UV-Vis SBW (nm)	2,0
Beam Mode	Double
Signal-to-noise Mode	Off
UV Source	On
Vis Source	On
Source Changeover (nm)	350,00
Baseline Correction	On
Baseline Type	Baseline correction
Baseline File Name	

Baseline Std Ref File Name
Cycle Mode Off
Comments
Method Modifications:
Autoconvert Changed: 09.08.2007 16:33:31. Old:None. New:Ascii with Log
Baseline Correction Changed: 09.08.2007 16:33:31. Old:0. New:1
End Method Modifications
<SBW (nm)> 2,00
<Energy> 164,00
<Current Wavelength> 355,00

Abschlussbetrachtung – Aussagekraft der Messergebnisse:

Der gewählte Assay ist mit Vorsicht zu interpretieren, da die hieraus abgeleiteten relativen Aktivitätswerte in erster Linie die Aussage zulassen, ob ein gewählter Ligand in diesem speziellen Assay eine Aktivität zeigt. Sollten direkte unerwünschte Wechselwirkungen mit dem Assaysystem (z.B. Autofluoreszenz, Quenchung, Lichtstreuung) ausgeschlossen worden sein, kann hieraus immer noch keine Wirkung in anderen Assays direkt abgeleitet werden. Schwierigkeiten wie die Zellwandpermeation sind in diesem Assaysystem explizit vermieden worden.

Die Natur des zellfreien Verfahrens ermöglicht einzig einen schnellen Test, ob eine Substanz in den Bindemodus des Peptids an die RNA eingreift und so die beiden FRET-Partner in räumliche Nähe zueinander bringt, was dann wiederum zu einer Verschiebung der relativen Fluoreszenzaktivität in Richtung des Referenzwertes des ungebundenen freien Tat-Peptids ohne RNA als Bindepartner führt.

Die Interpretation der Ergebnisse basiert auf folgenden Annahmen, die für das Testsystem formuliert wurden:

1. Das Peptid wird stark von der RNA gebunden und es bilden sich zeitlich stabile binäre Komplexe aus RNA und Peptid aus mit spezifischen Interaktionsflächen.¹
2. Der Hauptbindemodus des Peptids an die RNA ist lang gestreckt um die beiden Fluorophore voneinander räumlich zu trennen.
3. Eine Zugabe von Inhibitoren hat eine Verschiebung des Gleichgewichts vom gebundenen Peptid hin zu seiner freien Form zur Folge.

¹ Nach neuesten Untersuchungen mittels des LILBID-Massenspektrometrieverfahrens (Arbeitskreis Brutschy, (Morgner et al. 2006)), ist eher davon auszugehen, dass zwei Tat-Peptide an die TAR-RNA binden (Morgner et al. 2008)

3. Liganden-basierte Ansätze

Die folgenden Abschnitte behandeln die Untersuchungen, der Erstellung der Liganden-basierten Filter-Systeme, sowie der Validation dieser Ansätze durch Messungen in verschiedenen zur Verfügung stehenden Labor-Assays. Hierbei erfolgte eine Unterteilung in die zwei durchgeführten Hauptvorhersagemodelle: Abschnitt 3.2 „Allgemeiner RNA-Binder-Filter“ und 3.3 „Tat-TAR-FRET-Assay-spezifischer ANN-basierter Struktur-Filter“. Jeder der Abschnitte enthält Angaben zur Datensatzzusammenstellung, Erstellung der ANN-Filter, Vorhersagelisten der ANN-Filter, Ergebnisse der Labortest und Bewertungen der gefundenen potentiellen RNA-Liganden. Abschnitt 3.1 „RNA-Liganden-Datenbank“ gibt vorab ein paar Informationen über die verwendete Software.

Vorabinformationen zu den verwendeten Substanz-Kurzbezeichnungen der folgenden Abschnitte:

- **M-Reihe**, oder auch **M1-36** bezieht sich auf Substanzen, die mit den „Allgemeiner RNA-Binder-Filtern“ in den Substanz-Datensätzen von Specs (Delft, Niederlande. <http://www.specs.net>) vorhergesagt wurden und von dort für die Tests im Labor geliefert werden konnten.
- **AS-Reihe**, oder auch **AS1-23** bezieht sich auf Substanzen, die mit den „Allgemeiner RNA-Binder-Filtern“ in den Substanz-Datensätzen von ASINEX Ltd., Moscow, Russland. (<http://www.asinex.com>) vorhergesagt wurden und von dort für die Tests im Labor geliefert werden konnten.
- **LR-Reihe**, oder auch **LR1-42** bezieht sich auf Substanzen, die mit den „Tat-TAR-FRET-Assay-spezifischer ANN-basierter Struktur-Filtern“ in den Substanz-Datensätzen von Specs (Delft, Niederlande. <http://www.specs.net>) vorhergesagt wurden und von dort für die Tests im Labor geliefert werden konnten.

Die Darstellungen aller Strukturen der drei Messreihen erfolgten im Anhang.

Die Strukturen der Substanzen (mit Ausnahme der COBRA [Schneider P. und Schneider G.]), die als Trainingsdatensätze verwendet wurden können auf Anfrage vom Autor erhalten werden. Dies gilt ebenso für die in dieser Arbeit verwendeten Skripte, soweit das Urheberrecht beim Autor liegt.

3.1 RNA-Liganden-Datenbank

Hintergrund:

Im Rahmen dieser Arbeit mussten die bestehenden Informationen über RNA-Liganden aus der verfügbaren Literatur zusammengetragen werden. Da diese Informationen später der Erstellung verschiedener Modellierungsansätze für die RNA-Liganden-Vorhersage unterzogen werden sollten, wurde zu diesem Zweck ein Projektansatz gewählt der die Verwaltung der gesammelten Informationen übernehmen konnte. Da diese letztendlich mit den aus der Zielstruktur (RNA) abgeleiteten Informationen vereint werden sollten, wurde hier ebenfalls auf eine gute Verknüpfbarkeit mit dem im Rahmen dieser Arbeit ebenfalls zu entwickelnden Zielstruktur-Informationssystemen geachtet. Deshalb wurde hierbei ebenfalls auf die Option einer MySQL-Datenbank (MySQL AB, Uppsala, Schweden) zurückgegriffen, die mittels Java-Programm-Schnittstelle angesteuert werden sollte. Die fortschreitende Integration dieses Liganden-basierten Ansatzes führte letztlich zur Integration der einzelnen, für diesen Teil der Arbeit erstellten Programm-Module in Pandora – **P**ocket **A**nalysis for **D**Ocking on **R**epresentative **A**ptamers, dem für die Konformeranalyse entstandenem Programm aus dem Zielstruktur-basierten Ansatz auf. Abbildung 12 zeigt den Informationsfluss innerhalb des Liganden-basierten Teils dieser Arbeit. Der Status dieses Teils von Pandora befindet sich noch in der Prototypphase, umfasst jedoch bereits die Klassifikation mittels ANN-Filter und verschiedener Deskriptor-Verarbeitungsroutinen. Im Rahmen dieser Arbeit wurden jedoch parallel hierzu die in den noch folgenden Abschnitten genannten Umsetzungen bis zur vollständigen Portierung der fehlenden Module

angewendet. Insbesondere wurde die Liganden-Datensatz-Verwaltung noch mit dem ChemOffice-Plugin Version 7.0.1 (CambridgeSoft, Cambridge, USA) für Excel 2002 (Microsoft Corporation, USA) umgesetzt. Als Datei-Format für den Austausch von Strukturdaten unter den verwendeten Programmen diente für die Berechnung der Deskriptoren ausnahmslos das SDF-Format (Dalby et al. 1992).

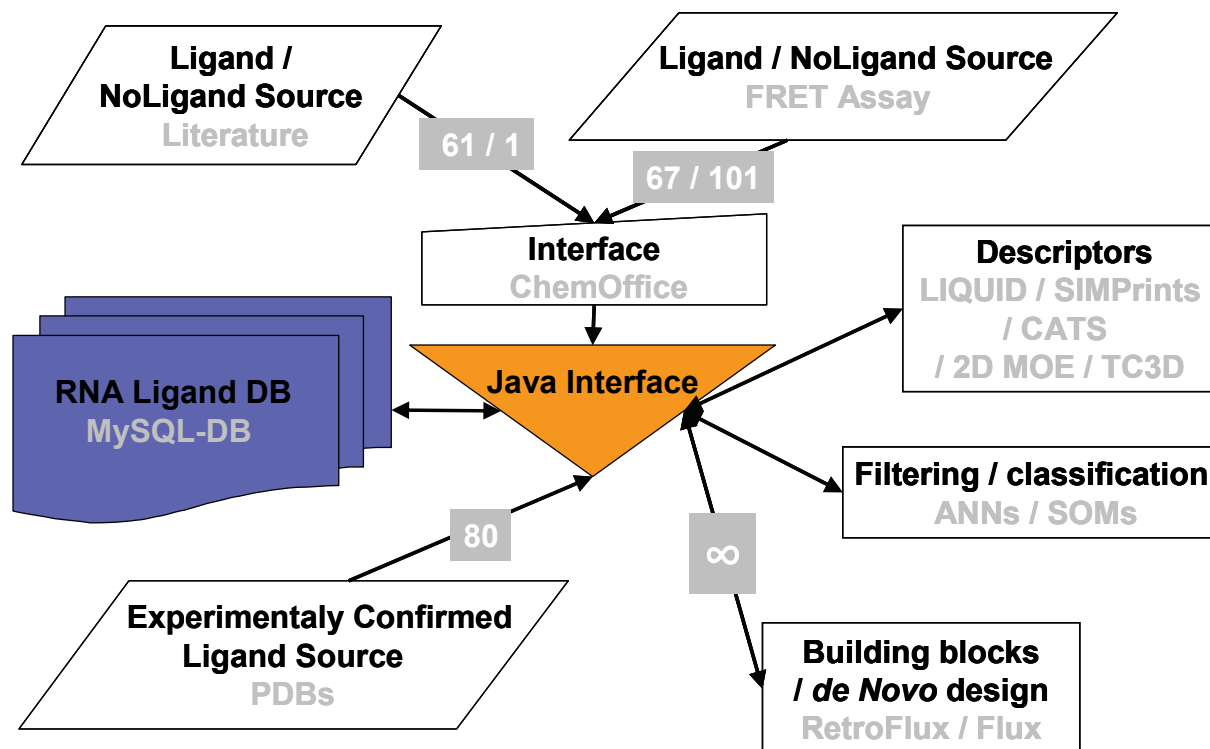


Abbildung 12: Schema des RNA-Liganden-Informationsflusses im Projekt

3.2 Allgemeiner RNA-Binder-Filter

3.2.1 Allgemeiner RNA-Liganden Datensatz – Datensatzzusammenstellung

Hintergrund:

Zu Beginn der Arbeiten wurde ein Datensatz zusammengestellt, der die bis dahin aus der Literatur frei verfügbaren Informationen zum RNA-Bindeverhalten von kleinen wirkstoffartigen Substanzen zusammenfassen sollte.

Zunächst war hierbei die Suche strikt auf TAR-Liganden beschränkt worden, jedoch sukzessive um weitere RNA-Strukturmotive als RNA-Bindestellen und den dazugehörigen Liganden erweitert worden, da die Liste bekannter TAR-Liganden mangels hochaffiner Ligandenvorbilder ($K_d < 1 \mu\text{M}$ und $IC_{50} < 1 \mu\text{M}$ im FRET-Assay) nicht zum Ableiten eines Pharmakophormodells geeignet war und die Menge der Substanzen mit schwächerer Aktivität ebenfalls nicht den Kriterien für die Erstellung eines spezifischen TAR-Liganden-ANN²-Filters entsprach. Ferner konnte für den in dieser Arbeit eingesetzten FRET-Assay eine Diskrepanz zwischen dort erreichten Inhibitionskonstanten (IC_{50} -Werten) und der in anderen Assays bestimmten Bindekonstantenmessungen (Lind et al. 2002) festgestellt werden [pers. Komm. Dr. Ute Scheffer]. Der im Rahmen dieser Arbeit einzusetzende

² ANN = artificial neural network -> künstliches Neuronales Netz

Verdrängungsassay – als einzusetzendes Gütekriterium – erforderte somit eine Neuausrichtung der Literatursuche nach Hinweisen auf eine erfolgreiche Inhibition der Tat-TAR-Interaktion gegenüber einer einfachen Betrachtung der Möglichkeit einer Bindung an TAR. Hierzu wurde folgende Arbeitshypothese aufgestellt:

Da die Bindung an die TAR-RNA, nicht zwangsläufig zu einer messbaren Inhibition der Tat-TAR-Interaktion im verwendeten FRET-Assay führen würde, ist eine Auswahl von Substanzen anzustreben, die einen solchen messbaren Effekt am wahrscheinlichsten hervorrufen könnte. Hierzu sind diejenigen Substanzen am geeignetsten anzusehen, die eine direkte Konkurrenz zum Bindeverhalten des Tat-Peptids bewirken können. Bekannte RNA-Liganden müssen, um die gemessene Bindung zu bewirken die Substrukturelemente tragen, die ihre Bindung an die jeweilige RNA begünstigen. Ein „allgemeiner RNA-Ligand“ sollte somit die bestmögliche Kombination aller RNA-bindenden Substrukturmuster enthalten. Allgemeine RNA-Liganden sollten somit nicht nur eine spezifische Stelle auf der TAR besetzen können, sondern die TAR-RNA großflächig dekorieren können. Hierzu müssen die auszuwählenden Substanzen ein allgemeines Interaktionsmuster mit RNA eingehen können. Um ein derartiges allgemeines RNA-Bindeverhalten ableiten zu können, müssen die gemeinsamen Strukturmerkmale der bekannten RNA-Liganden extrahiert werden. Sollte der zusammenstellbare Datensatz, eine ausreichende Anzahl an Substanzen mit unterschiedlichen RNA-Bindestellen aufweisen und strukturell divers sein, so sollten sich mittels der Kodierung der Substanzen als Deskriptordarstellung die allgemeinsten Substrukturmuster ableiten, die eine RNA-Bindung unterstützen ableiten lassen. Um jedoch auszuschließen, dass die so gefundenen Gemeinsamkeiten der RNA-Liganden nicht den Substrukturteil der Strukturen umfassen, die an einer RNA-Bindung nicht beteiligt sind, muss noch ein Vergleichsdatsatz erstellt werden, der im Idealfall aus Substanzen besteht, die nachweislich nicht mit RNA interagieren können. Da hauptsächlich Ergebnisse von erfolgreichen Bindungsmessungen veröffentlicht werden, sind Ergebnisse mit schlechtem Bindeverhalten von Substanzen für bestimmte Ziele ein rares Gut (Turner et al. 2008). Wie kann man diesem Problem begegnen? Da die bisherige Medikamentenentwicklung eher auf Proteine als auf RNA ausgelegt ist (Lind et al. 2002; Foloppe et al. 2006), müsste theoretisch ein Datensatz aus bekannten Medikamenten mit ihrer auf Proteinwechselwirkung ausgelegten Auslese bei der Entwicklung dazu führen, dass diese Substanz-Gruppe eine allgemeine Affinität zu proteinartigen Wirkstoffzielen in sich kodiert trägt. Die letztendlich mögliche Wechselwirkung mit RNA kann hierbei zwar nicht ausgeschlossen, aber in ihrer Wahrscheinlichkeit womöglich reduziert werden. Dieselbe Betrachtungsweise gilt natürlich ebenfalls für die RNA-Liganden-Datensätze: Der positive Datensatz mit bekannten(experimentell gesicherten) RNA-affinen Substanzen, wird zwar definitiv die Strukturmerkmale ausweisen, um jeweils an bestimmte RNA-Struktur motive zu binden, ob sich Substanzen jedoch selektiv nur an RNA und nicht an Proteine binden, kann wiederum nicht ausgeschlossen werden. Als ein direktes Gegenbeispiel für die Selektivität lässt sich hier bereits Acetylpromazin anführen: Bei Acetylpromazin handelt es sich um den als RNA-Ligand des (mittels NMR strukturaufgeklärten) PDB-Eintrags 1LVJ, jedoch ist diese Substanz auch für ihre Wirkung auf die D₂- und α₁-Rezeptoren bekannt (Gross 2001; Hall et al. 2001; Hall et. al 2001b; Ludders et al. 1983).

Durch Kombination beider Datensätze sollte es jedoch immer noch möglich sein, diejenigen Eigenschaften, welche die RNA-Binder von den Protein-Bindern unterscheiden, aus den Strukturen abzuleiten. Vorausgesetzt, dass es diesen Unterschied wirklich gibt, sollte dieser sich in Statistiken über die Eigenschaftsverteilung der Moleküle niederschlagen.

Durchführung:

Für das Erstellen der Datensätze wurde die zugängliche Literatur nach RNA-Liganden durchsucht. Zusätzlich wurde die PDB (Berman et al. 2000, <http://www.rcsb.org/pdb/home/home.do>) nach RNA-Strukturen mit Liganden durchsucht, um die Ligandenstrukturen aus den entsprechenden PDB-Dateien zu extrahieren.

3.2.1.1 COBRA Datensätze

Der COBRA-Datensatz (Version 4.6) (Schneider, P. und Schneider, G. 2003) ist eine Sammlung von Referenzmolekülen zum Entwurf Liganden-basierter Datenbanken und wurde aus neuerer wissenschaftlicher Literatur zusammengetragen. Dieser Datensatz enthält Informationen über ca. 2.600 Medikamente und weitere bioaktive pharmakologisch aktive Substanzen und steht dem Arbeitskreis Schneider zur Verfügung.

Im Rahmen dieser Arbeit wurde ein Teildatensatz aus dem COBRA-Datensatz erstellt. Dies geschah wie folgt:

1. Die Einträge zu den Zielstrukturen wurden auf RNA oder DNA-spezifische Einträge untersucht.
→ Unter den verwendeten Suchbegriffen wurden jedoch nur RNA- oder DNA-prozessierende Proteine gefunden.
2. Die 6.046 Strukturen wurden mittels MOE auf doppelte Struktureinträge zu unterschiedlichen Zielstrukturen überprüft („Nebenwirkungsprofile“).

→ 395 Einträge konnten so aus dem Datensatz entfernt werden, da sie bereits in der verbleibenden Strukturliste genannt wurden. Dies reduzierte den Protein-Liganden-Datensatz auf 5.651 Einträge.

3.2.2 Resultierender Arbeitsdatensatz: RNA-Binder gegen Protein-Binder

Die beiden Datensätze mit den aus der verfügbaren Literatur gefundenen bestätigten³ RNA-Bindern (Quellen: PDB-Einträge und Veröffentlichungen aus PubMed) wurde zusammengeführt und aus diesem gemeinsamen Datensatz mittels MOE ein gemeinsamer Datensatz mit den Strukturen erstellt; siehe Abbildung 13. Hierbei konnten 83 Strukturvorschläge aus dem Datensatzteil der PDB entnommen werden und wiederum 63 Strukturen aus dem Datensatz, welcher aus Publikationen der PubMed-Datenbank entnommen wurde. Dieser Datensatz sollte bei doppelter Nennung einer Substanz diese nur noch einmal enthalten. Hieraus resultierte ein gemeinsamer Datensatz mit 99 strukturell unterschiedlichen RNA-Liganden, von denen 98 zur 3D-Konformer-Generation verwendet werden konnten. Dieser Teil des Datensatzes wurde entsprechend als *COBRA46_RNALigandNrX@NameDesLigandenImSet* im späteren gemeinsamen Datensatz benannt. Die verbliebenen COBRA-Strukturen wurden analog mit ihrer Bezeichnung in der ursprünglichen Datenbank als *COBRA46_EintragsNummer@Zielstrukturbezeichnung* bezeichnet. Resultierend wurden hieraus zwei SDF-Dateien erstellt, welche für die weitere Deskriptorberechnung die Strukturinformationen beinhalten.

83 Strukturen aus der PDB-Datenbank
63 Strukturen aus den Literaturangaben der PubMed-Datenbank
<hr/>
99 strukturell einzigartige RNA-Liganden

Abbildung 13: Zusammenführung der Strukturen der beiden Recherche-Quellen zum verwendeten RNA-Liganden-Datensatz

³ Siehe hierzu Abschnitt 3.2.19.1 Seite 78 über experimentell unbestätigte Filikovsubstanzen

3.2.3 Untersuchungen zur Trennbarkeit der beiden Datensätze

Einleitung:

Die Überprüfung der in Abschnitt 3.2.1 genannten Arbeitshypothese, dass Protein-Liganden und RNA-Liganden unterschiedliche Substrukturmuster enthalten und somit trennbar sein sollten, musste überprüft werden. Die vorhergesagten Unterschiede sollten sich in Unterschieden bei der geeigneten Beschreibung (Deskriptoren) niederschlagen. Da die Deskriptoren eine Reduktion von Merkmalen der jeweiligen Strukturen auf konkrete Zahlwerte darstellen, sollten mathematische Verfahren zur Bestimmung der Überlappung zweier Zahlenverteilungen deren mögliche Güte für die Trennung in die beiden Ursprungsklassen vorherzusagen vermögen. Dieser Sachverhalt kann für jeden einzelnen Deskriptor (der kleinsten Einheit einer möglichen ganzheitlichen Beschreibung) mittels Kolmogorov-Smirnov-Statistik (KS-Statistik; siehe auch Abschnitt 2.3)(Byvatov und Schneider 2004) durchgeführt werden, welche ein einfach anzuwendendes Maß für die Überlappung zweier Zahlenverteilungen darstellt.

3.2.4 KS-Statistik zur Analyse der Eigenschaftsverteilung

Hintergrund:

Die Untersuchung der MOE-2D-Deskriptoren erlaubt – aufgrund ihrer meist recht gut verständlichen Beschreibungen und ihrer meist von anderen Deskriptoren unabhängigen Aufbau – eine Abschätzung darüber, welche Eigenschaften für ein gegebenes Struktur-Trennungsproblem am wichtigsten erscheinen.

Durchführung:

Die hier gezeigten Ergebnisse beruhen auf der Anwendung der im Rahmen dieser Arbeit erstellten m-Skripte zur KS-Statistik-Erstellung in Matlab nach dem Import der jeweiligen Deskriptordateien.

Ergebnisse:

Als Ergebnis der KS-Statistik konnte eine nach KS-Distanz sortierte Liste der verwendeten MOE-2D-Deskriptoren erhalten werden. Die Verteilung der KS-Werte wurde in Abbildung 14 dargestellt und zeigt die KS-Distanz Verteilung für den genannten Deskriptorensatz. Tabelle 2 enthält den Auszug der nach KS-Distanz die Top15 der Deskriptoren des gesamten MOE-2D-Deskriptorensatzes; die vollständige Liste ist im Anhang 7.1 auf Seite 194 zu finden.

Tabelle 2: „Top 15“ der MOE-2D Deskriptoren, welche laut KS-Statistik besonders gut zur Trennung des Protein-Liganden-Datensatzes vom RNA-Liganden-Daten geeignet sind.

Deskriptor (MOE-2D)	KS-Wert (Distanz)	Position im Ranking
SlogP_VSA0'	0,71	1
'Q_VSA_FPPOS'	0,66	2
'Q_RPC+'	0,65	3
'RPC+'	0,65	4
'PEOE_VSA_FNEG'	0,63	5
'PEOE_VSA_FPOS'	0,63	6
'Q_VSA_PPOS'	0,62	7
'SlogP'	0,61	8
'PEOE_RPC+'	0,59	9

Deskriptor (MOE-2D)	KS-Wert (Distanz)	Position im Ranking
'vsa_pol'	0,56	10
'Q_VSA_FHYD'	0,54	11
'Q_VSA_FPOL'	0,54	12
'PEOE_VSA-6'	0,52	13
'Q_VSA_FNEG'	0,51	14
'Q_VSA_FPOS'	0,51	15

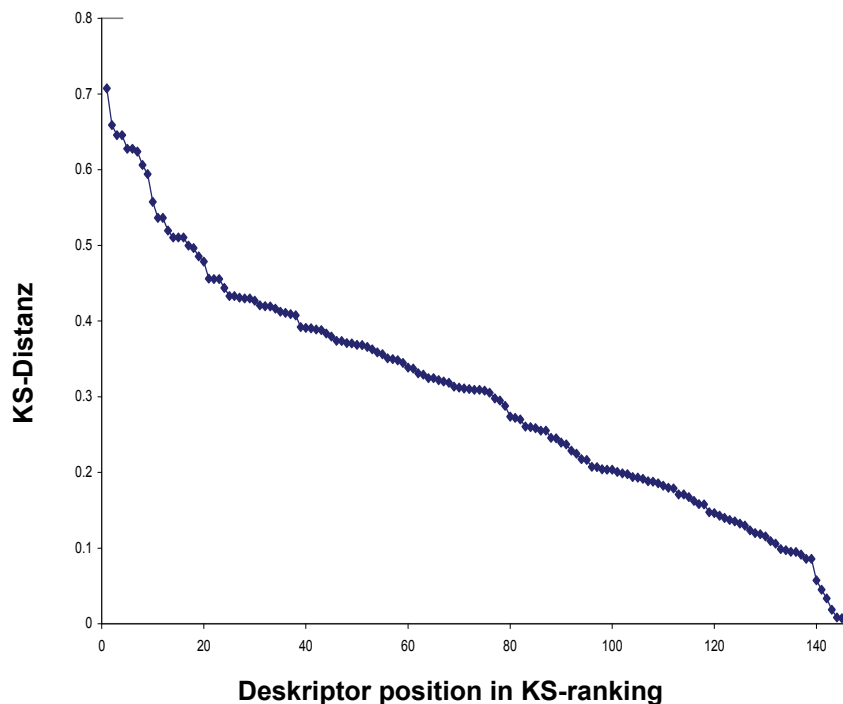


Abbildung 14: Plot der KS-Distanzen für die MOE-2D-Deskriptoren zur Trennung der Protein-Liganden von den RNA-Liganden-Datensätzen. Zur besseren Abschätzung des Trennverhaltens sortiert nach der KS-Distanz.

Diskussion:

Die Betrachtung der Abbildung 14, von der die Tabelle 2 nur einen Teilausschnitt („Top 15“) enthält, zeigt, dass es durchaus Eigenschaften (KS-Distanz > 0.1) innerhalb des hier untersuchten MOE-2D-Deskriptorraums gibt, welche zur Trennung der beiden postulierten Klassen (Protein-Binder und RNA-Binder) geeignet erscheinen.

Bei der Betrachtung der Deskriptornamen der nach KS-Distanz sortierten Liste fiel auf, dass innerhalb der „Top 15“ der untersuchten sortierten Deskriptorliste kein einziger Deskriptor vorkommt, der nicht auf einer Berechnung der Ladungsverteilungen innerhalb des Moleküls aufbaute. Dies führte zur genaueren Untersuchung der Grundlagen der in den „Top 15“ auftauchenden Deskriptoren; siehe Abschnitt 3.2.4.1 unten.

3.2.4.1 Analyse der Zusammensetzung der „Top15“ des MOE-2D-Deskriptor KS-Rankings

Hintergrund:

Bei der Analyse der Zusammensetzung der nach KS-Distanz gerankten Deskriptorenliste fiel eine Häufung der ladungsbezogenen Deskriptoren auf. Dies veranlasste eine genauere Betrachtung der Deskriptordefinitionen für die erkennbaren Deskriptorklassen.

Erklärung der Deskriptorklassen:

Die Klasse der Deskriptoren, die ein **VSA** im Namen enthalten ist auf der Approximation der zugänglichen „van der Waals“-Oberfläche aufgebaut, welche basierend auf der Verknüpfungstabelle der Atome errechnet wird. Hierbei werden basierend auf der Verknüpfung des jeweiligen Atoms anhand von Referenzlisten die zugängliche Oberfläche angenähert.

Die Klasse der **PEOE**-Deskriptoren beruht auf der „Partial Equalization of Orbital Electronegativities“-Methode von Gasteiger (Gasteiger 1980) und fasst somit Deskriptoren zusammen, die auf den Verteilungen der so berechneten Partialladungen beruhen.

Die Klasse der „**Q_***“-Deskriptoren beruht, wie die oben genannten PEOE-Deskriptoren, wieder auf der Verteilung von Partialladungen in den Molekülen. Im vorliegenden Fall wurden diese mit den Standardeinstellungen von MOE vor der eigentlichen Deskriptor-Berechnung berechnet und in den Molekülrepräsentationen abgelegt.

SlogP beruht auf einem Modell (Crippen 1999), welches für jedes Molekül – basierend auf seiner jeweiligen atomaren Zusammensetzung – einen Octanol/Wasser-Verteilungskoeffizienten berechnet. Den Atomtypen sind hier bestimmte Polaritätseigenschaften zugewiesen, die wiederum auf der Partialladungsverteilung beruhen.

RPC+ ist ein Deskriptor für die „relative positive Partialladung“ eines Moleküls und wird ebenfalls aus den mit MOE berechneten Partialladungen erstellt.

Diskussion:

Bereits in dieser Analyse ist somit der Grundtenor der späteren Arbeiten zu erkennen, welcher vereinfacht als „Positive Ladungen“ beschrieben werden kann; siehe hierzu auch den Abschnitt 4.7 unten über die Oberflächenanalyse der Zielstruktur.

Schlussfolgerung:

Aufgrund der Untersuchung der KS-Distanzen der verwendeten MOE-Deskriptoren für den RNA-Liganden/Protein-Liganden-Datensatz wurde offensichtlich, dass die Verwendung von ladungsbasierten Deskriptoren von Vorteil zu sein scheint. Hierbei sollten ferner diejenigen Deskriptoren geeignet sein, die die Ladungsverteilung innerhalb eines Moleküls mit größtmöglicher Genauigkeit wiedergeben können.

Aus der Untersuchung abgeleitete Deskriptorenwahl:

Der von Lutz Franke im Arbeitskreis Schneider entwickelte Charge3D-Deskriptor und seine Weiterentwicklung zum TC3D-Deskriptor (Fechner et al. 2003) wurde bereits in der Diplomarbeit „Vorhersage der Blut-Hirnschrankengängigkeit von kleinen Molekülen mittels künstlicher neuronaler Netze“ getestet und beschreibt genau diese gesuchten Eigenschaften: Die Abbildung der Partialladungsverteilung innerhalb eines Moleküls als Korrelationsvektordarstellung. Da im Rahmen dieser Arbeit eine häufige Benutzung dieses Deskriptorsystems zu erwarten war, wurde die zu langsame Perl-Skript-Variante im Rahmen dieser Arbeit in einer C-basierten Variante reimplementiert (Geschwindigkeitsgewinn \approx Faktor 3.000) und mit einem Webinterface versehen. Das Programm ChargeCyclonev2.2 steht dem Arbeitskreis auf dem internen Webserver zur Verfügung. Dieses erledigt auch die automatische Vorbereitung der Daten vor der eigentlichen Deskriptorberechnung:

- wahlweise Berechnung der benötigten 3D-Konformere mittels CORINA 3.20 (Molecular Networks GmbH, Erlangen)
- Berechnung der Partialladungen mittels PETRA 3.11 (Molecular Networks GmbH, Erlangen)

analog zur Umsetzung in den Perl-Skripten.

3.2.5 Abschätzung der strukturellen Diversität im RNA-Liganden-Datensatz**Einleitung:**

Da die Trennung dieses als komplex zu betrachtenden Systems sich nicht auf einen Zahlenwert herunter reduzieren lässt (entspreche einem KS-Wert = 1), wurde die Projektion des höherdimensionalen Deskriptorraums auf eine SOM durchgeführt, um abzuschätzen wie divers die Subdatensätze sich im chemischen Deskriptorraum verteilen.

Sollten sich bei einer solchen Projektion alle Substanzen des erwünschten Subsets in einem Bereich der SOM vereinen, so ist nicht von einem diversen Wechselwirkungsprofil des Datensatzes auszugehen. In diesem Fall würde eine einfache Ähnlichkeitssuche zu einer gemittelten Pharmakophor-Struktur die einfachste und beste Erfolgchance darstellen.

Die Erfolgsaussichten einer solchen Suche wiederum hängen stark von der Stabilität der Wechselwirkungspunkte der Zielstruktur ab; siehe hierzu für TAR die Abbildung 75 auf Seite 120. Sollte die Zielstruktur sich vielmehr mit dem jeweiligen Liganden in ihrer zeitlich aufgelösten Konformerzusammensetzung anpassen, so ist dieser Ansatz mit großer Wahrscheinlichkeit zum Scheitern verurteilt, da sich in der so ermittelten Durchschnittsstruktur die wichtigen Wechselwirkungsmöglichkeiten gegenseitig auslöschen könnten und sich das Vorhersagesystem somit auf die zwar verbleibenden, aber nicht mehr das Bindungsverhalten entscheidenden Merkmale der Struktur reduziert. Durch Zufall ist auch hier wieder ein Erfolg denkbar, wenn um dieses Grundgerüst, die richtigen Wechselwirkungsmöglichkeiten eingeführt werden. Der negative Fall ist aber mindestens genauso wahrscheinlich.

Um die Diversitätsuntersuchung mehr auf eine Beschreibung der Verteilung von Pharmakophorpunkten basierend umzusetzen, wurde hierfür im Rahmen dieser Arbeit auf den CATS-Deskriptorraum (Schneider et al. 1999) zurückgegriffen. Dieser bietet die Möglichkeit die Verteilung von Pharmakophorpunkten (Wasserstoffbrücken-Akzeptor, -Donor, usw.; siehe auch Abschnitt 2.1.2 oben) innerhalb eines Molekülmodells als Deskriptor zu verwenden. Hieraus sollte sich ableiten lassen, ob in den anzuwendenden Datensätzen, die Verteilung möglicher Interaktionspunkte mit der RNA divers ist oder die Strukturen alle denselben Aufbau bei der Verteilung der potentiellen Pharmakophorpunkte aufweisen.

Durchführung:

Der gemeinsame Datensatz mit RNA-Liganden und Protein-Liganden wurde für das SOM-Training eingesetzt.

Es wurden folgende Einstellungen für das Training verwendet:

- distance metric: -euklid
- -xNeurons 15
- -yNeurons 10
- -cycles 60.000
- -neuronDim 150
- -radius 10
- -tau 1.0

Anschließend erfolgte zur Abschätzung der Diversität innerhalb des RNA-Liganden-Datensatzes eine Projektion des CATS-Deskriptorraums des kombinierten RNA-Liganden-Datensatzes mit dem COBRA-Datensatz, ebenso wie die alleinige Projektion des RNA-Liganden-Datensatzes; siehe hierzu Abbildung 15. Diese Abbildung zeigt die Projektion des RNA-Liganden-Datensatzes relativ zur Verteilung der ebenfalls für das SOM-Training verwendeten COBRA-Substanzen.

Ergebnis:

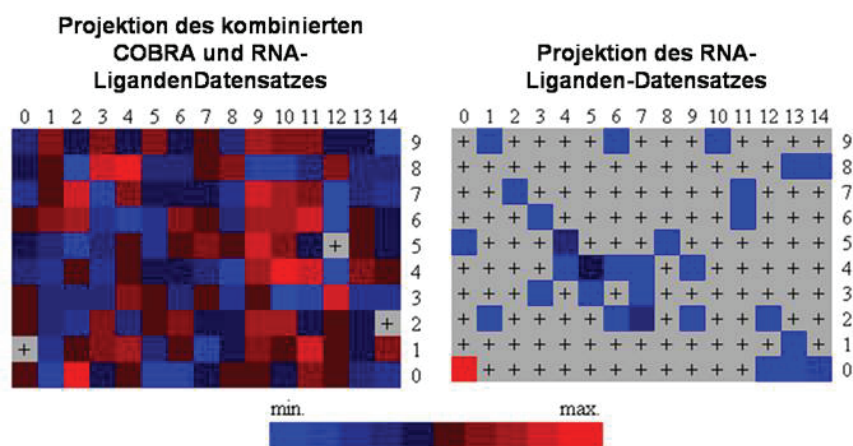


Abbildung 15: CATS-Deskriptor-SOM zur Bestimmung der Diversität innerhalb des RNA-Liganden-Datensatzes relativ zur Diversität des Protein-Liganden-Datensatzes. Die Einfärbung der Knoten gibt die relative Anzahl an Strukturen des jeweiligen Knotens in der Karte wieder.

Diskussion:

Wie in Abbildung 15 rechts zu sehen ist, verteilen sich die durch RNA-Liganden bevölkerten Knoten der SOM über die Karte, wobei es lokal durchaus einige stärker beieinander liegende Bereiche gibt. Jedoch sind die Knoten grob gesehen breitgestreut über die Karte verteilt. Die Verteilung des RNA-Liganden-Datensatzes auf der Karte suggeriert somit eine diverse Verteilung des RNA-Liganden-Datensatzes im benutzten chemischen Raum.

Schlussfolgerung:

Das Ergebnis der SOM-Analyse zur Verteilung der Teilmenge des RNA-Liganden-Datensatzes, hat den Entschluss weiter bestärkt, während dieser Studie Abstand von Methoden zu nehmen, die auf einem einzelnen Referenzliganden basieren.

Da ferner zum Beginn dieser Studie kein „hochaffiner“ wirkstoffartiger Ligand bekannt war, der in der Lage war die Tat-TAR-Interaktion ($IC_{50} < 100$ nM) im verwendeten FRET-Assay zu inhibieren, konnte zu diesem Zeitpunkt keine Auswahl für eine Leitstruktur-Optimierung aus dem möglichen Ligandenraum gerechtfertigt werden. Deshalb wurde auch aufgrund dieser SOM-Untersuchungen eine Strategie mit ANN-basierten Filtern angestrebt. Zusätzlich hat die offensichtliche Ähnlichkeit des hier beobachteten KS-Statistik-Rankings im Vergleich zum Ranking in der Diplomarbeit über „Vorhersage der Blut-Hirnschrankgängigkeit von Molekülen mit künstlichen neuronalen Netzen“ ebenfalls diesen, als relativ zuverlässiges Lernverfahren erprobten Weg der ANN-Filter als erste zu ergreifende Maßnahme nahe gelegt.

3.2.6 Allgemeiner Filter für RNA-Binder

Hintergrund:

Basierend auf den zwei folgenden Annahmen wurde ein Filter entwickelt, welcher der groben Trennung von Liganden dienen soll, die eher an RNA-Zielstrukturen als an Protein-Zielstrukturen binden sollen:

1. Der gesammelte RNA-Liganden-Datensatz enthält nur Moleküle, die an RNA-Zielstrukturen binden.
2. Die Substanzen des COBRA-Datensatzes besitzen aufgrund ihrer Zusammenstellung aus bekannten bioaktiven Substanzen einen eher proteinbindenden Charakter, da bisher eher Proteine als Arzneistoffzielstrukturen verwendet wurden.

Durchführung:

Abbildung 16 gibt den Ablauf der Filtererstellung wieder. Eine genauere Darstellung der Erstellung von ANN-basierten Filtersystemen im Rahmen dieser Arbeit ist in Abschnitt 2.4 wiedergegeben. Der in Abbildung 16 dargestellte Arbeitsablauf wurde – abgesehen vom Einlesen der Deskriptoren – mittels Matlab-Skripten voll automatisiert. Analog wurde so auch – im Rahmen dieser Arbeit – eine Untersuchung zur Dopamin-D2/D3-Liganden Vorhersage durchgeführt (Böcker & Sasse et al. 2007). Die jeweils resultierende Jury wird zur Klassifikation der Datensätze der Chemikalien von Lieferanten-Datenbanken eingesetzt.

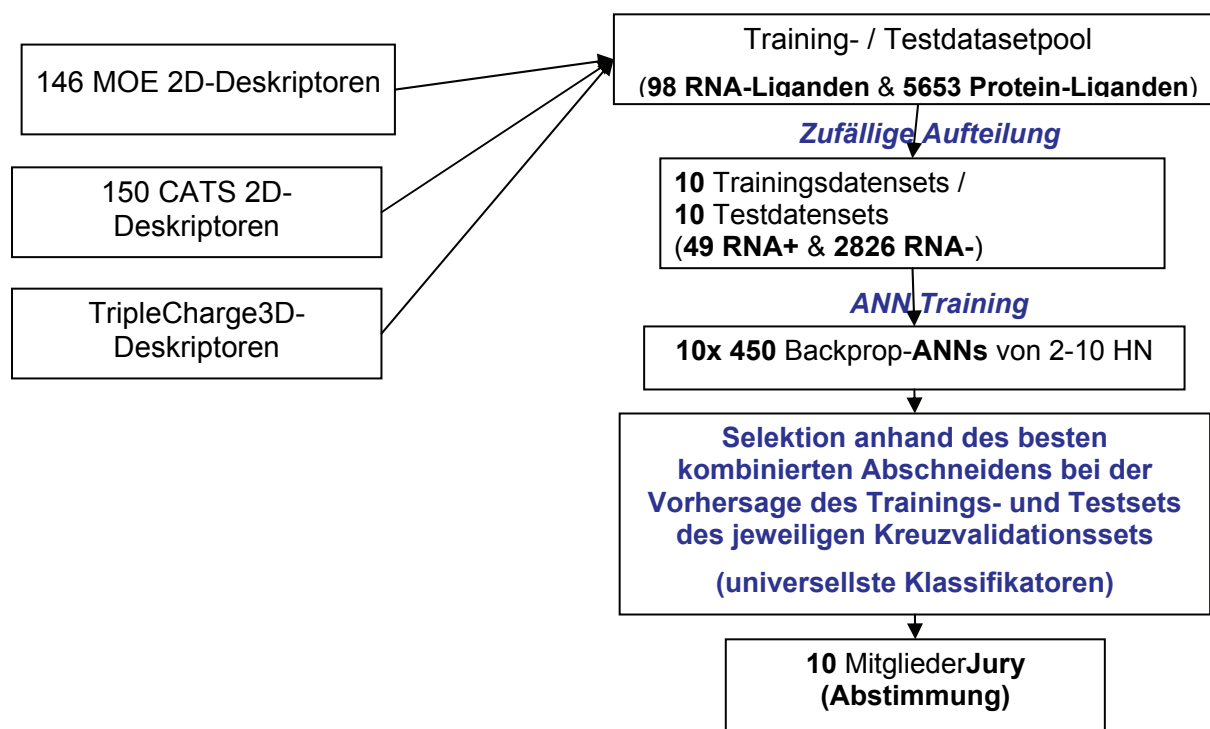


Abbildung 16: Trainings- und Auswahlschema der RNA-Liganden-GrobFilter.

Diskussion:

Die nach dem Schema in Abbildung 16 erstellten ANN-Filter können in Matlab zum Screening von weiteren Datensätzen verwendet werden, wenn zu diesen die gleichen Deskriptorinformationen vorliegen. Des Weiteren wurden die Parameter der Filter mittels erstellter Export-Skripte auch für den Einbau in ein im Rahmen dieser Arbeit erstelltes Java-Modul verfügbar gemacht. Durch deren Einbau in Entwicklungsumgebungen wie Pandora oder weitere Java-Programmen werden diese gewonnenen Erkenntnisse einfach verfügbar gemacht. Die Software Matlab muss nur zur Erstellung der Filter verwendet werden, nicht jedoch zur weiteren Benutzung der Filter in „Virtuellen Screening Kampagnen“ verfügbar sein.

3.2.7 Verteilung der Scorewerte für RNA-Binder gegen Protein-Binder

Überprüfung der Trenngüte:

Um die zu erwartende Trenngüte der erstellten Filter – innerhalb der geplanten Screening-Kampagne nach potentiellen Liganden in den Chemikalien-Datenbanken der Versandhändler und der letztendlichen Testung von ausgewählten Substanzen im Labor – zu überprüfen, wurde wie folgt vorgegangen:

Durchführung:

Zu diesem Zweck wurden die verwendeten RNA-Liganden- und Protein-Liganden-Datensätze erneut von den ANN-Filtern bewertet. Die erhaltenen Scorewerte-Verteilungen wurden in den Abbildungen 17 bis 19 für die drei erstellten RNA-Binder-Filter graphisch aufbereitet.

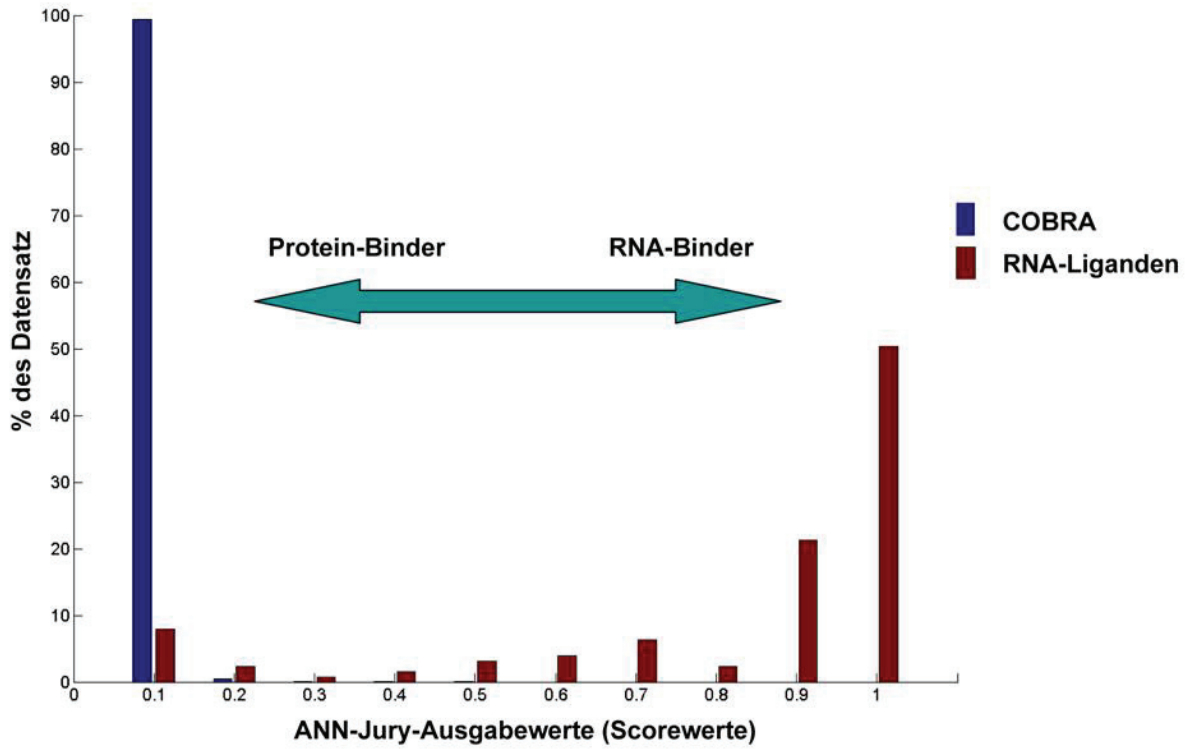


Abbildung 17: Histogramm der Scorewerteverteilung für den RNA-Binder-Filter basierend auf 146 MOE2D-Deskriptoren für den COBRA- und den RNA-Liganden-Datensatz

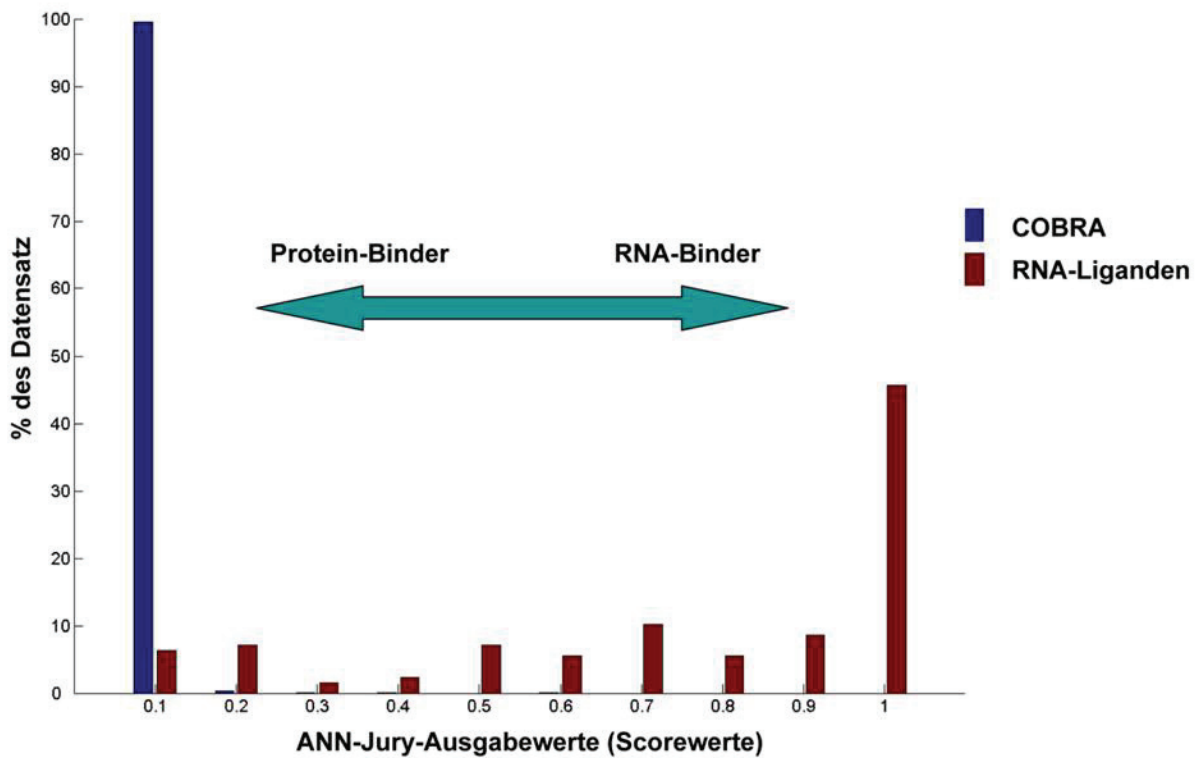


Abbildung 18: Histogramm der Scorewerteverteilung für den RNA-Binder-Filter basierend auf 150CATS-2D-Deskriptoren für den COBRA- und den RNA-Liganden-Datensatz

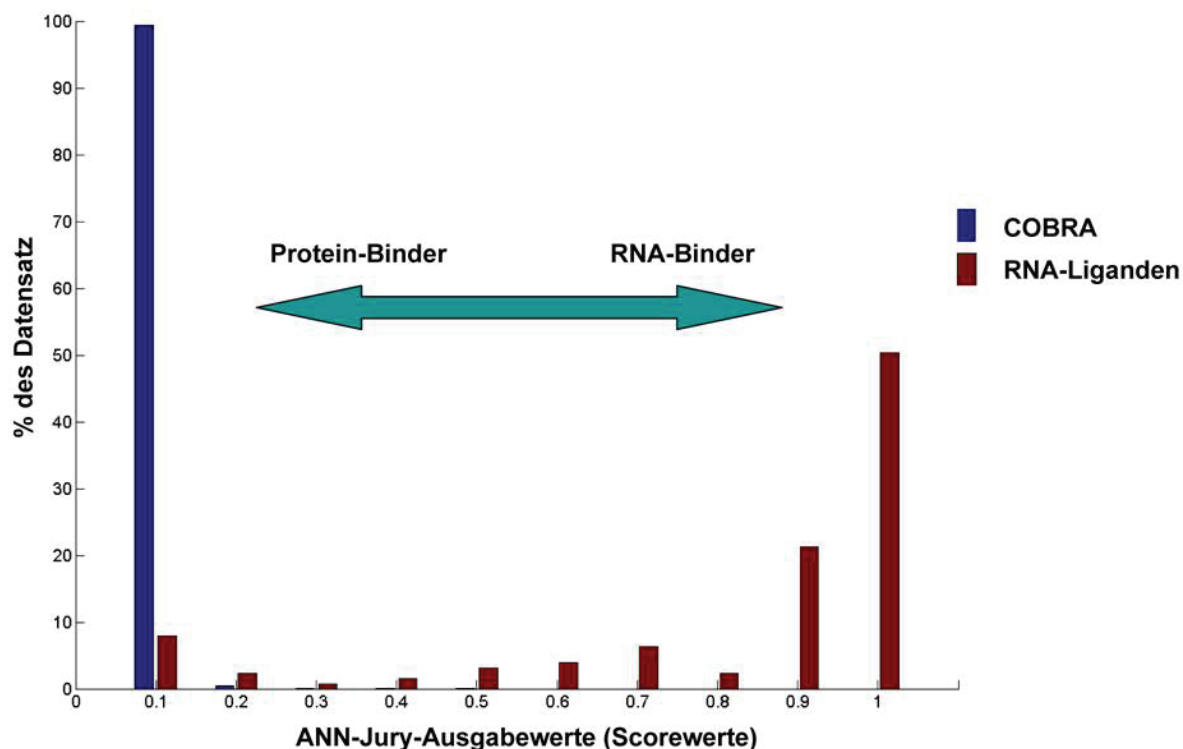


Abbildung 19: Histogramm der Scorewerteverteilung für den RNA-Binder-Filter basierend auf 300 TC3D-Deskriptoren für den COBRA- und den RNA-Liganden-Datensatz

Diskussion:

Die erhaltenen Scorewerte-Verteilungen für die drei erstellten Filter – in Abbildungen 17 bis 19 gezeigt – zeigen ausnahmslos eine deutliche Trennung des COBRA-Datensatzes von den RNA-Liganden. Ersichtlich ist dies durch den Anteil von beinahe 100% der COBRA-Datensatz-Einträge, welchem Score-Werte im Bereich von 0 bis 0,2 zugeteilt wurden. Dahingegen bekam der größte Teil der verwendeten RNA-Liganden einen Scorewert > 0,5 zugewiesen.

Schlussfolgerung:

Ausgehend von den erstellten RNA-Liganden-Grob-Filtern ist somit eine gute Trennung von offensichtlich Protein-Binder-artigen Substanzen aus den Chemikalien-Lieferantendatensätzen zu erwarten. Dieser sehr restriktive Charakter wurde auch beim Screening der besagten Lieferantendatensätze mittels der hier gewonnenen Filter weiter bestätigt; siehe hierzu den folgenden Abschnitt 3.2.8 „Suche mit den RNA-Binder-Filtern in den Chemikalien-Datenbanken der Versandhändler“.

3.2.8 Suche mit den RNA-Binder-Filtern in den Chemikalien-Datenbanken der Versandhändler

Mittels der im vorigen Abschnitt beschriebenen RNA-Liganden-Grob-Filter wurden drei dem Arbeitskreis zugängliche Chemikalien-Datenbanken von zwei Versandhändlern nach potentiellen RNA-Liganden durchsucht.

Ergebnisse:

Tabelle 3 zeigt die Trefferraten für potentielle RNA-Liganden innerhalb der Chemikalien-Datensätze verschiedener Lieferanten. Gezeigt werden die jeweilige Anzahl an vorgeschlagenen Substanzen und der prozentuale Anteil am jeweiligen gesamten Datensatz. Als potentieller RNA-Ligand wurde hier eine Substanz ab einem Score von 0,5 gewertet.

Tabelle 3: Virtuelle Hitrate der RNA-Liganden-Filter beim Screening der Chemikalien-Datenbanken der Versandhändler

	MOE2D-GF	CATS-GF	TripleCharge3D-GF
Specs	42	40	71
(229,658)	(0,02%)	(0,02%)	(0,03%)
AsinexGold	113	100	108
Collection	(0,06%)	(0,05%)	(0,06%)
(194,566)			
AsinexPlatinum	keine	198	59
Collection	(0,00%)	(0,17%)	(0,05%)
(111,211)			

Diskussion:

Bei Verwendung des auch zum Training der Netze eingesetzten Schwellenwerts von 0,5 zeigt sich eine deutliche Reduktion der erhaltenen Hitlisten. Dies war bereits durch die Auswertung der Score-Werteverteilung für die verwendeten Trainingsdatensätze erwartet worden. Die virtuelle Trefferquote lag hier je nach Datenbank und verwendetem Filter zwischen 0 und 0,17% der Gesamtmenge der jeweiligen Datenbank; siehe hierzu auch die Tabelle 3, die diese Ergebnisse in einer Übersicht darstellt.

Schlussfolgerung:

Die Suche mittels der RNA-Liganden-Grob-Filter nach potentiellen RNA-Liganden führte zu einer deutlichen Reduktion der potentiellen Testkandidaten aus den Chemikalien-Datensätzen der Lieferanten auf ein nun menschlich überschaubares Maß.

Weiterführende Arbeit:

Die Größen der erhaltenen Teilmengen der Chemikalien-Datensätze der Lieferanten konnten in einem weiteren Schritt dem „Cherry Picking“ unterworfen werden, um eine kombinierte Bestellliste von aussichtsreichen Kandidaten für die Tests im Labor zu erhalten.

3.2.9 Dockingexperimente mit den ausgewählten Substanzen in einen Konformer des PDB-Eintrags 1LVJ

Hintergrund:

Da im Rahmen dieser Arbeit eine messbare Störung der Tat-TAR-Interaktion (FRET-Assay) erreicht werden sollte und die bisher erarbeiteten Vorhersagemodelle einen voraussichtlich eher allgemeinen RNA-Binde-Mechanismus beachten sollten, galt es einen Weg zu finden, diese potentiellen allgemeinen RNA-Liganden auf eine zunehmende Selektivität für die TAR-RNA zu selektieren. Als ersten Ansatz kann hierbei auf eine Diskriminierung der Auswahllisten mittels Docking zurückgegriffen werden. Das zu erwartende Docking-Ergebnis hängt hierbei von der verwendeten Zielstruktur und der zu verwendenden Bewertungsfunktion ab.

Im Arbeitskreis stand die Docking-Software GOLD v2.2 (The Cambridge Crystallographic Data Centre, Cambridge) für die Dockingstudien zur Verfügung. Dieses Programm wurde jedoch laut Hersteller ("GOLD is a program for calculating the docking modes of small molecules into protein binding sites.") zur Vorhersage der Docking-Posen von kleinen Molekülen in die potentiellen Bindetaschen von Proteinen entwickelt. Das Programm enthält bereits zwei Bewertungsfunktionen (GoldScore und ChemScore), die laut Dokumentation von Protein-Liganden-Komplexen abgeleitet worden waren. Im Rahmen des SFB579 wurde zur Zeit dieser Arbeit im dem SFB579 angehörigen AK-Gohlke eine spezielle von RNA-Liganden-Komplexen abgeleitete DrugScore-Variante (DrugScore^{RNA}) (Pfeffer et al. 2007) entwickelt, konnte jedoch zu diesem Zeitpunkt noch nicht für diesen Versuchsteil eingesetzt werden. Dies führte dazu, dass die beiden in GOLD enthaltenen Bewertungsfunktionen anhand eines RNA-Konformer-Liganden-Komplexes (PDB-Eintrag 1LVJ entnommen) beispielhaft zur bestmöglichen Auswahl einer Bewertungsfunktion untersucht wurden. Dies erfolgte ebenso, um eine Vorstellung für die Güte der zu erwartenden Ergebnisse zu entwickeln.

Anschließend sollten die durch ANN-Filter erstellten Auswahllisten mittels der durch die Voruntersuchung ausgewählten Bewertungsfunktion ebenfalls einer Docking-Studie unterzogen werden.

Durchführung der Voruntersuchung:

Auswahl eines TAR-Konformer-Liganden-Komplexes

Als Beispiel-Komplex wurde der PDB-Eintrag 1LVJ benutzt, der die TAR-RNA und Acetylpromazin enthält. Diese Struktur wurde mittels NMR-Spektroskopie aufgeklärt und zeigte einen relativ geringen RMSD zwischen den einzelnen Konformervorschlägen. Als Alternative hätten die PDB-Einträge 1QD3 und 1ARJ zur Verfügung gestanden, welche aber aufgrund der größeren RMSD-Werte-Verteilung [siehe auch Abschnitt 4.6.2 auf Seite 171] innerhalb der jeweiligen PDB-Konformer-Ensembles zurückgestellt wurden. Da für diese Untersuchung aus Zeitgründen nur ein RNA-Liganden-Komplex untersucht werden sollte, bot sich die Verwendung des dem PDB-Eintrag 1LVJ entnommenen Komplexes an, weil dieser die wenigsten Unterschiede in den zu erhaltenden Werten erwarten ließ, egal welcher letztlich aus dem PDB-Ensemble (12 Konformere) entnommen würde.

Vorbereitung auf das Docking mit GOLD:

Mittels PyMOL Version 0.99rc6 (DeLano Scientific LLC, South San Francisco, California, U.S.A.) wurden die Koordinaten des Schwefelatoms des Acetylpromazin als zu verwendende Taschenmitte extrahiert.

Der PDB-Eintrag wurde so manipuliert, dass sich nur noch das erste Modell in der Datei befand. Zur Erzeugung der beiden benötigten Strukturdateien (RNA-Konformer und Ligand) wurden jeweils die Zeilen, die sich auf nicht benötigte Atome bezogen, ebenfalls gelöscht.

Mittels GOLD-Programmoberfläche wurden die benötigten Konfigurationsdateien erstellt. Folgend ist der Inhalt der Konfigurationsdatei – unter Benutzung der GoldScore-Bewertungsfunktion durchgeführten Dockingversuche – wiedergegeben.

```
GOLD CONFIGURATION FILE
generated by gold front end (GOLD v2.2)

POPULATION
popsiz = 100
select_pressure = 1.100000
n_islands = 5
maxops = 100000
```



```
niche_siz = 2
  GENETIC OPERATORS
pt_crosswt = 95
allele_mutatewt = 95
migratewt = 10

  FLOOD FILL
radius = 10
origin = -21.985 8.802 6.0991
do_cavity = 1
floodfill_atom_no = 0
cavity_file = cavity.atoms
floodfill_center = point

  DATA FILES
protein_datafile = 1LVJ_f_docken_gold.pdb
ligand_data_file = Inhib_1LVJ.pdb 25
param_file = DEFAULT
set_ligand_atom_types = 1
set_protein_atom_types = 1
directory = .
tordist_file = DEFAULT
make_subdirs = 0
save_lone_pairs = 1
fit_points_file = fit_pts.mol2
read_fitpts = 0

  FLAGS
display = 0
internal_ligand_h_bonds = 0
n_ligand_bumps = 0
flip_free_corners = 0
flip_amide_bonds = 0
flip_planar_n = 1
flip_pyramidal_n = 0
use_tordist = 1
start_vdw_linear_cutoff = 4
initial_virtual_pt_match_max = 2.5
```

```
TERMINATION
early_termination = 0
n_top_solutions = 3
rms_tolerance = 1.5

CONSTRAINTS
force_constraints = 0

COVALENT BONDING
covalent = 0

SAVE OPTIONS
save_score_in_file = 1
save_protein_torsions = 1
```

Bei Verwendung des ChemScore als Bewertungsfunktion enthält die verwendete Konfigurationsdatei noch folgende zusätzliche Parameterzeilen.

```
FITNESS FUNCTION SETTINGS
gold_fitfunc_path chemscore
score_param_file = DEFAULT
```

Bei Verwendung der potentielle RNA-Liganden enthaltenden Auswahllisten verweist der Eintrag „ligand_data_file“ auf die entsprechende SDF-Datei mit den Strukturen.

Durchführung des Dockinglaufs:

Alle Dockingläufe wurden mittels des entsprechenden Kommandozeilenaufrufs gestartet.

gold_auto xy_goldconf.conf; xy_goldconf.conf = relevante Konfigurationsdatei des Dockinglaufs.

Ergebnis und Diskussion der Voruntersuchung:

Durch die Untersuchung der Dockingposen sollte geklärt werden, inwieweit die Software in der Lage war, die ursprüngliche Position des Liganden aus der PDB zu reproduzieren. Hierbei viel auf, dass die erwartete Platzierung innerhalb des Bulgetunnels von den beiden Bewertungsfunktionen unterschiedlich umgesetzt wurde. Auf dem GoldScore als Bewertungsfunktion basierende Dockingläufe platzierten den Liganden zwar um 180° gedreht immerhin noch relativ nahe der ursprünglichen Position des Acetylpromazins im PDB-Eintrag, dahingegen vermieden Dockingläufe basierend auf der ChemScore-Bewertungsfunktion den Bulgetunnel komplett. Die Visualisierung der Dockingergebnisse wurde für diese Arbeit auf die Darstellung der besten Dockingpose laut Goldscore (Goldscore = 64,44) gegenüber der Originalpose in der PDB in Abbildung 20 beschränkt.

Der RMSD für die Paarung, der laut GoldScore besten erhaltenen Docking-Pose zur Originalpose, beträgt 6,9 Å.

Somit war eine sinnvolle Selektion der potentiellen Testkandidaten mit den bisher gewählten Programmeinstellungen mehr als fraglich, insbesondere mittels der ChemScore-Bewertungsfunktion. Da der Durchschnitt der erreichten GoldScores für das Docking von Acetylpromazin 57,89 betrug und als kleinster Wert 52,89 erreicht wurde, konnte jedoch zu diesem Zeitpunkt noch nicht ausgeschlossen werden, dass mittels *GOLD* eine Diskriminierung der potentiellen RNA-Liganden möglich wäre (GoldScore > 0 werden als potentiell gute Vorhersage der Bindung gewertet).

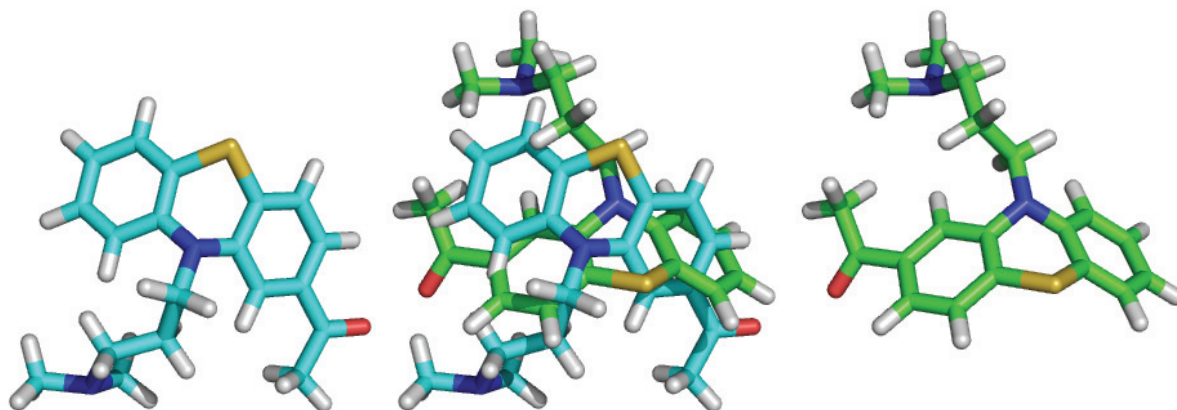


Abbildung 20: Acetylpromazin - Beste Dockingpose laut GoldScore gegen Originalpose für das erste Modell im PDB-Eintrag 1LVJ; links Originalpose, rechts Dockingpose, Mitte Überlappung im TAR-Konformer. Aus Übersichtsgründen wurde auf die Darstellung der RNA verzichtet. Goldscore = 64,44. Der RMSD zwischen den Posen beträgt 6,9 Å.

Docking der potentiellen RNA-Liganden aus den ANN-basierten Auswahllisten

Da die Voruntersuchungen zwar zeigten, dass die Originalbindepose von Acetylpromazin nicht reproduziert werden konnte, jedoch durchwegs positive GoldScores erhalten werden konnten, wurden – zur weiteren Überprüfung der Dockingvorhersagen von *GOLD* – die auf die ANN-Filter basierenden Auswahllisten der potentiellen RNA-Liganden ebenfalls dem Docking mittels GoldScore unterzogen. Zusätzlich zu den ANN-Filter basierenden Auswahllisten wurde zur besseren Abschätzung der Diskriminierung von RNA-Liganden von potentiellen Nicht-Bindern ein Datensatz aus dem verwendeten COBRA-Datensatz erstellt (willkürliche Auswahl jeder 50ten Struktur aus dieser Datenbank), der ebenfalls dem Docking unterzogen werden sollte. Dessen erreichte Scorewerte sollten als Referenzwerte bei der späteren Beurteilung dienen. Erwartet wurde hierbei, dass diese Werte im Durchschnitt niedriger ausfallen würden als die der potentiellen RNA-Liganden laut ANN-Filter. Als Alternative zum verwendeten Konformer aus dem PDB-Eintrag 1LVJ wurde hierbei jedoch zusätzlich auch auf das erste Modell im PDB-Eintrag 1ANR (ungebundene TAR) als TAR-Konformer zurückgegriffen. Dies erfolgte, um eine mögliche Stabilisierung der freien TAR durch einen potentiellen RNA-Liganden zu untersuchen, bzw. um die möglichen Unterschiede der Bindungsvorhersagen aufgrund unterschiedlicher verwendeter Konformere zu überprüfen.

Durchführung:

Die Strukturen der Auswahllisten der ANN-Filter wurden mit denselben Parametern wie bei den Voruntersuchungen mit Acetylpromazin gedockt. Ausnahme war hier nur das Setzen der Anzahl durchzuführender Dockingversuche pro Struktur auf 10. Dasselbe geschah für die in der COBRA-Referenzliste enthaltenen Strukturen. Für die Dockingläufe mit dem Konformer

3. Liganden-basierte Ansätze

aus dem PDB-Eintrag 1LVJ wurden somit dieselben Taschenkoordinate eingesetzt, wie dies auch bei den Voruntersuchungen mit Acetylpromazin der Fall war. Als Zielstruktur wurde zusätzlich auch auf das erste Modell im PDB-Eintrag 1ANR (freie TAR) als TAR-Konformer zurückgegriffen.

Die Koordinate zur Definition der potentiellen Bindetasche des 1ANR-Konformer wurde aus den Positionen folgender Atome der TAR berechnet.

Residue 24 (U40),Atom N3 x: -0,286
y: 3,714
z: 9,760

Residue 8 (C24), Atom N1 x: 4,528
y: -5,675
z: 8,088

Residue 10 (G26),Atom O6 x: -2,178
y: -2,071
z: 11,513

Mittelpunkt der Bulge abgeleitet aus den obigen Koordinaten:

x: 0,688
y: -1,344
z: 9,787

Ergebnisse:

Abbildung 21 zeigt die Auftragung der GoldScore-Werte des durchgeführten Dockinglaufs in den ersten RNA-Konformer des PDB-Eintrags 1LVJ. Um eine genauere Bewertung der potentiell sinnvollen Scorewertebereiche (Goldscore > -100) zu ermöglichen, wurde in Abbildung 22 die Auftragung auf Dockingergebnisse mit einem GoldScore > -100 beschränkt wiederholt. Offensichtliche Ausreißer wurden somit bei dieser Auswertung ignoriert.

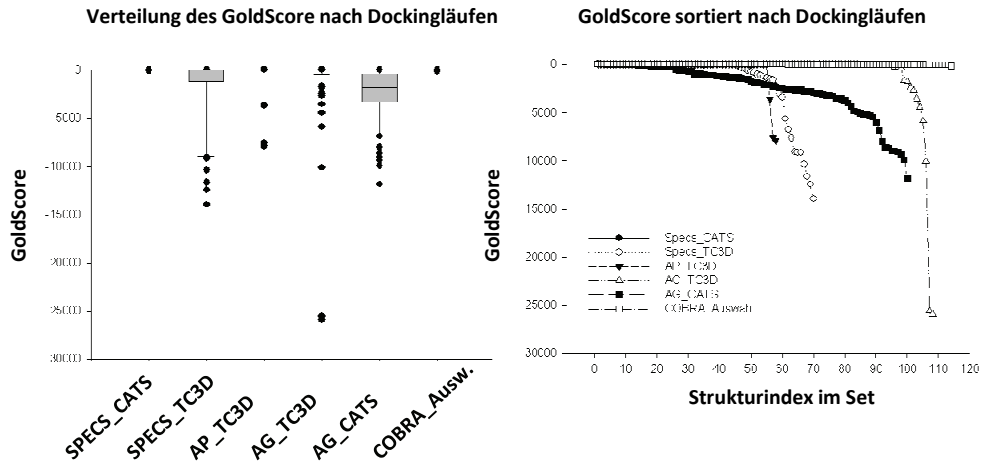


Abbildung 21: Scorewerte der Docking-Runde mit den potentiellen RNA-Liganden sowie einem COBRA-Subset (jeder 50te Eintrag); links Boxplot der Scoreverteilung; rechts sortierte Liste der GoldScore-Verteilung; Grundlage war der 1LVJ-Konformer

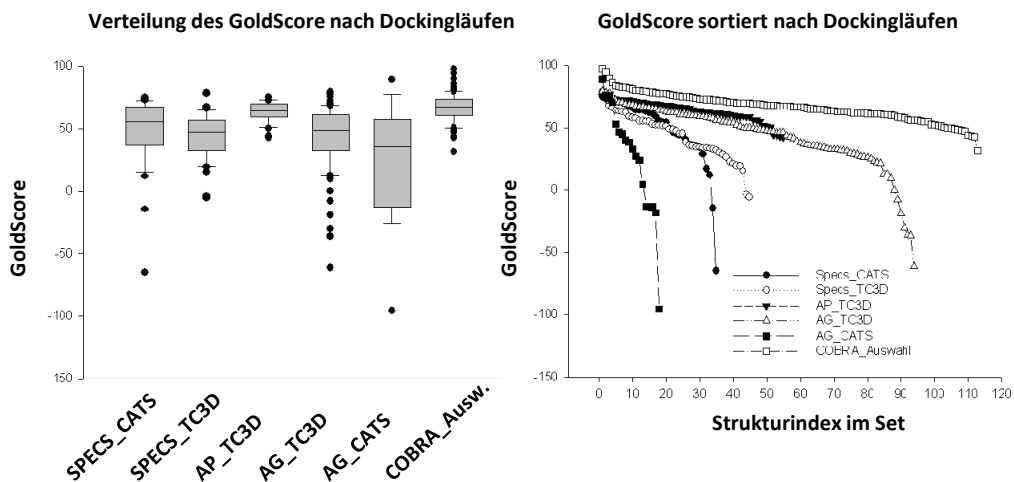


Abbildung 22: Scorewerte der Docking-Runde mit den potentiellen RNA-Liganden sowie einem COBRA-Subset (jeder 50te Eintrag); links Boxplot der Scoreverteilung; rechts sortierte Liste der GoldScore-Verteilung; die Darstellungen sind diesmal auf einen Wertebereich > -100 beschränkt, um einen genaueren Einblick in den Bereich der „TopScorer“ zu ermöglichen. Grundlage war der 1LVJ-Konformer

Um eine genauere Untersuchung der „Top10“-Score-Bereiche der einzelnen Läufe zu ermöglichen, wurden die Scorewerte der Läufe in Abbildung 23 nochmals einzeln gezeichnet.

3. Liganden-basierte Ansätze

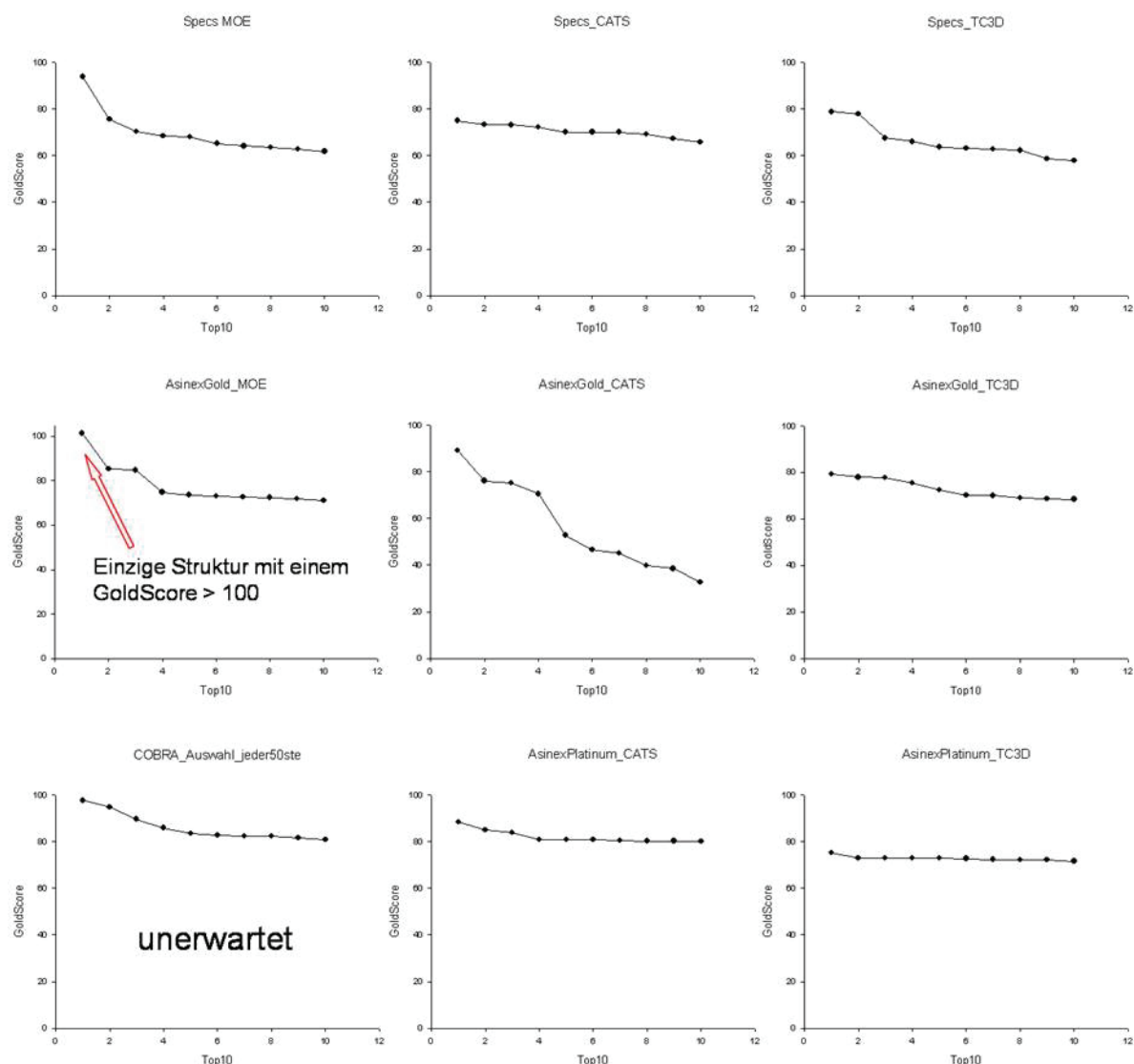


Abbildung 23: „Top 10“-Scores der Dockingrunden der einzelnen Screening-Hitlisten-Subsets als Subplots; Grundlage war der 1LVJ-Konformer

Diskussion:

Die Abbildung 21 bis Abbildung 23 zeigt jeweils die Auftragungen der erhaltenen Scorewerte der Dockingläufe für die unterschiedlichen Auswahllisten zu potentiellen RNA-Liganden und dem COBRA-Referenzdatensatz in unterschiedlichen Auflösungen zur Scorewerteverteilung. Prinzipiell positiv zu bewerten ist hierbei, dass die Scorewerte einen relativ großen Wertebereich innerhalb der einzelnen Datensätze abbilden und somit die Dockingsoftware eine Diskriminierung der einzelnen Substanzen mittels der gewählten Bewertungsfunktion (GoldScore) für den gewählten Konformer erreicht. Jedoch zeigte sich auch entgegen den vorher geäußerten Erwartungen (dass die Substanzen des COBRA-Referenz-Datensatzes im Durchschnitt schlechter als die potentiellen RNA-Liganden abschneiden sollten), dass das genaue Gegenteil der Fall war. So erhielt keine der verwendeten Substanzen des COBRA-Datensatzes einen GoldScore < 0, wohingegen bei den potentiellen RNA-Liganden zum Teil auch relativ extreme GoldScores von bis zu -25.000 auftraten. Der durchgängig positive GoldScore der Substanzen der COBRA-Datensatzauswahl legt zudem nahe, dass sich unter den bisher bekannten Medikamenten (welche sich zum Teil in der COBRA befinden) sich sehr viele RNA-Liganden befinden sollten. Diese Vermutung könnte zwar eventuelle Nebenwirkungen erklären, gehört aber zum jetzigen Zeitpunkt ins Reich der Spekulation und ist, da die bisherige Suche nach potenten RNA-Liganden (siehe eigene Literatursuche und den Übersichtsartikel „Towards the discovery of drug-like RNA ligands?“ (Foloppe et al. 2006)) eher spärlich ausgefallen ist, wohl eher unwahrscheinlich. Trotz allem

überlappen sich durchaus die Wirkbereiche, wenn man das Wirkungsprofil von Acetylpromazin betrachtet (Acetylpromazin ist auch ein bekannter Dopamin-Rezeptor-Antagonist (Hall et al. 2001, Starke und Freiburg 2001), als auch für seine α -adrenerge Wirkung bekannt (Ludders et al. 1983; Gross 2001; Hall et al. 2001)). Ebenso zeigt das breite Nebenwirkungsprofil dieser Substanz als Mitglied in der Wirkstoffgruppe Promazine, speziell der Phenothiazine mit Dimethylaminopropylgruppe (Müller-Oerlinghausen *et al.* 1999; Ammon 2001), dass ihr Bindemechanismus wohl nicht auf einzelne Zielstrukturen begrenzt zu sein scheint.

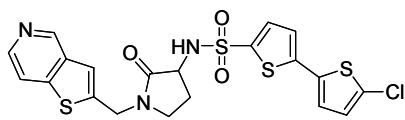
Da zur Auswahl von Substanzen basierend auf den Dockingergebnissen die TopScore-Bereiche noch verwendet werden würden, wurde eine genauere Untersuchung der TopScore-Bereiche der einzelnen Läufe ebenfalls umgesetzt. Hierzu wurden die Scorewerte der Läufe in Abbildung 23 nochmals einzeln zum besseren Vergleich gezeichnet. Hier zeigt sich, dass die Darstellung der „Top 10“ der COBRA-Auswahl (unten links) wiederum nicht den Erwartungen entspricht. Relativ zu den jeweiligen „Top 10“ der potentiellen RNA-Liganden-Vorschlagslisten schnitten sie wider Erwarten nicht schlechter ab, sondern im Durchschnitt deutlich besser. Nur in drei der acht untersuchten Fälle erreichte überhaupt eine Substanz mit ihrem Score den Bereich der COBRA-„Top 10“, was auf den gesamten untersuchten „Top 10“-Bereich von 80 Substanzen eine Quote von nur 3,75 Prozent ergibt. Dies ließ auch die Erwartungen an die Diskriminationsgüte der RNA-Ligandenvorhersage der GoldDocking-Runde weiter sinken. Da es jedoch nicht gänzlich ausgeschlossen werden konnte, dass es sich bei den COBRA-Einträgen vielleicht doch um RNA-Liganden handelt, wurden auch die „Top 10“ der 1LVJ-Dockingrunden ebenfalls zur manuellen Auswahl (Cherry-Picking) vorgelegt. Dies erfolgte trotz der widersprüchlichen Docking-Ergebnisse, da die so erstellten zusätzlichen Auswahllisten – welche auf den Dockingergebnissen basierten – sowieso durch die vorherige Auswahl mittels ANN-Filter einen RNA-bindenden Charakter aufweisen sollten und dieses durch die Dockingergebnisse auch nicht *per se* für die jeweiligen „Top 10“ abgelehnt wurde (durchgängig positive GoldScores). Ferner wurde im Übersichtsartikel „Towards the discovery of drug-like RNA ligands?“ von Foloppe, Matassova und Aboul-ela (2006) ebenfalls auf die Möglichkeit der Verwendung von Kinase-Liganden-Datensätzen als potentielle RNA-Liganden eingegangen. Die als „Top 10“ bewerteten Strukturen der COBRA-Auswahl werden in Abbildung 24 gezeigt; auf Rang 2 befindet sich ein Tyrosin-Kinase-Inhibitor; siehe Tabelle 4 mit den Docking-Rang und dem Bezeichner in der COBRA.

Tabelle 4: "Top 10" der Dockingrunde mit der COBRA-Auswahl; Rang = Platzierung in der Dockingrundenwertung, Name in COBRA-Datenbank enthält Informationen zu den bekannten Zielstrukturen

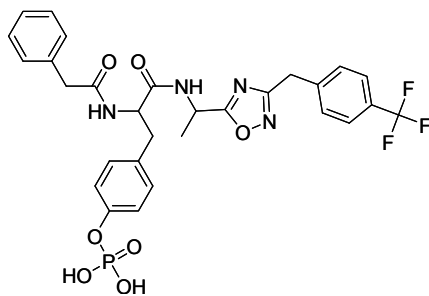
Rang	Name in COBRA-Datenbank
1	Protease_Serine_protease_(Trypsin-like)_Factor_Xa_fXa_2206
2	Kinase_Tyrosine_Kinase_Inhibitor_ZAP-70_SH-2_Domain_ZAP-70_SH-2_5789
3	GPCR_Tachykinin_Neurokinin1_Receptorantagonists_NK1_antagonist_3700
4	GPCR_adrenergic_receptor_adrenergic_receptor_beta_3_agonist_beta3_agonist_403 Protease_aggregation_Inhibitor_Ab_Aggregation_Inhibitors_...
5	Ab_Aggregation_inhibitor_IDOX_4860
6	GPCR_Somatostatin_sst2_selective_somatostatin_agonist_sst2_agonist_5246 Protease_Angiotensin_Converting_Enzyme_(metalloprotease)_...
7	ACE_Inhibitor_ACE_Inhibitor_SCH54470_605
8	GPCR_Serotonin_receptor_5-Hydroxytryptamin_1B/D_5-HT1B/D_agonist_5025
9	Protease_Serine_protease_(Trypsin-like)_Factor_Xa_fXa_2155
10	Nuclear_Receptor_liver_x_receptor_liver_x_receptor_alpha_LXR_alpha_agonist_3316

3. Liganden-basierte Ansätze

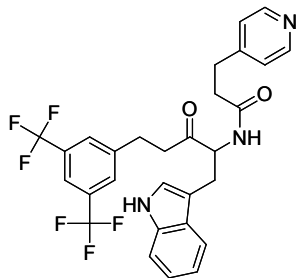
1



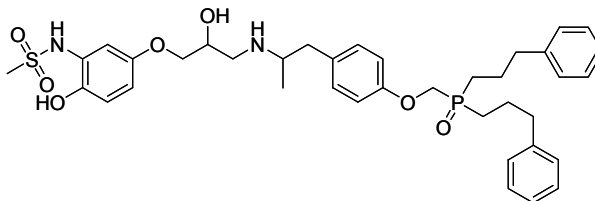
2



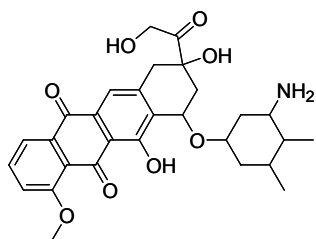
3



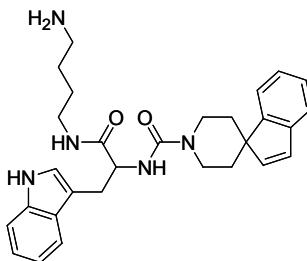
4



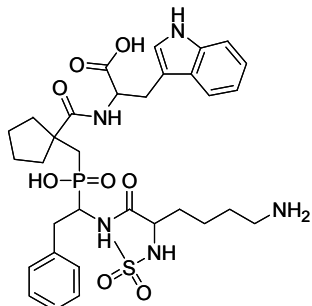
5



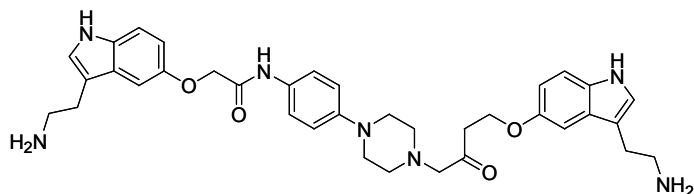
6



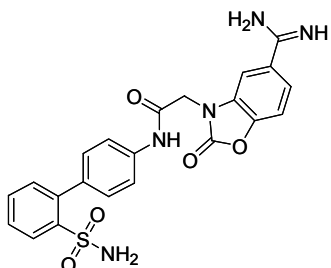
7



8



9



10

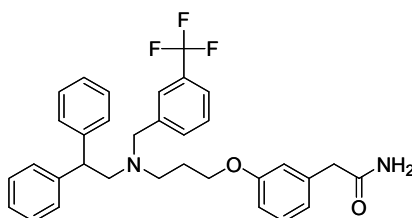


Abbildung 24: Strukturen der "Top 10" der Dockingrunde mit der COBRA-Auswahl; die Zahl gibt die Platzierung in der Dockingrundenwertung an

Abschließend lässt sich ausschließlich sagen, dass die in der COBRA-Datenbank enthaltenen Substanzen in sich die bevorzugten Interaktionsmöglichkeiten – welche durch die GoldScore-Bewertungsfunktion bewertet wurden – in verstärktem Maße gegenüber den per ANN-Filter aus der Specs-Datenbank ausgewählten Substanzen beinhalten.

Retrospektive Betrachtung der Ergebnisse nach den Tests im Labor

Die Auswertung der Verteilung der im Labor aktiven Verbindungen auf die „Cherry-Picking“-Hitlisten (siehe Abschnitt 3.2.18) zeigte ebenfalls kein verallgemeinerbares Verhalten, da dort zwar für einzelne Listen je nach verwendeten Konformer zum Teil Aktive (potentielle RNA-Binder) in den Top10-Hitlisten angereichert wurden, jedoch auch aktive Verbindungen aus diesen verdrängt wurden.

Für die drei im FRET-Assay aktiven Verbindungen wurde aufgrund der Tatsache, dass es sich bei der Zielstruktur der Dockingvorhersage und der RNA-Struktur im Assay jeweils um TAR handelt, eine genauere Untersuchung der Vorhersagen durchgeführt. Diese konnte jedoch kein eindeutiges Vorhersageverhalten bestätigen; siehe hierzu Tabelle 5 bis Tabelle 7.

Alle drei Verbindungen erhalten zwar immer einen positiven GoldScore, was auf eine potentielle Eignung als TAR-Ligand hindeutet. Jedoch unterscheiden sich die erreichten Werte nicht nur zwischen den TAR-Konformeren (1LVJ und 1ANR), sondern sind auch je nach durchgeführtem Docking-Lauf mit den gleichen Liganden-Konformer-Kombinationen unterschiedlich; siehe Ergebnisse der MOE-Liste (Tabelle 6) und TC3D-Liste (Tabelle 7), die beide M8 und M32 enthalten, aber unterschiedliche Ergebnisse (GoldScores) liefern. Dies zeigt den Nichtdeterminismus der Docking-Software, bei Verwendung der hier benutzten Parameter (Anzahl an Dockingversuchen pro Lauf). Da GOLD auf einem genetischen Algorithmus zur Findung der besten Dockingpose basiert, ist dies auch nicht weiter verwunderlich.

Tabelle 5: Dockingergebnisse für M23, Rang und Goldscore in den ausgewählten TAR-Konformeren 1LVJ und 1ANR als Teil des Docking-Laufs der CATS-ANN-Filter Auswahlliste

CATS-Liste	1LVJ	1ANR	1LVJ	1ANR
40 Strukturen	Rang	Rang	GoldScore	GoldScore
M23	13	8	63,79	53,14

Tabelle 6: Dockingergebnisse für M8 und M32, Rang und Goldscore in den ausgewählten TAR-Konformeren 1LVJ und 1ANR als Teil des Docking-Laufs der MOE-ANN-Filter Auswahlliste

MOE-Liste	1LVJ	1ANR	1LVJ	1ANR
42 Strukturen	Rang	Rang	GoldScore	GoldScore
M8	9	25	62,8	34,53
M32	10	11	61,82	43,11

Tabelle 7: Dockingergebnisse für M8 und M32, Rang und Goldscore in den ausgewählten TAR-Konformeren 1LVJ und 1ANR als Teil des Docking-Laufs der TC3D-ANN-Filter Auswahlliste

TC3D-Liste	1LVJ	1ANR	1LVJ	1ANR
71 Strukturen	Rang	Rang	GoldScore	GoldScore
M8	19	25	51,42	45,02
M32	23	27	47,24	42,9

Dies führt zu dem Schluss, dass dieses Verfahren zwar eine weitere Perspektive auf die ANN-Filter basierten Hitlisten erlaubt, jedoch keine eindeutig statistisch belegbare Anreicherung ableitbar ist, solange kein einzelner Referenzkonformer definiert werden kann. Aufgrund des noch immer geringen Anteils an aktiven Verbindungen in den Listen ist hier jedoch keine abschließende Bewertung möglich.

Schlussfolgerung:

Aus der Benutzung der Docking-Software GOLD im Rahmen dieser Arbeit konnte keine eindeutige Empfehlung für die weitere Verwendung in RNA-Liganden-Screening-Kampagnen abgeleitet werden.

Die Bedenken bei der Nutzung dieser Docking-Ergebnisse hatten folgende Gründe:

Bereits der ursprüngliche Ligand der verwendeten Zielstruktur, Acetylpromazin konnte mittels Software nicht mehr in die Zielstruktur in seiner ursprünglichen Ausrichtung eingepasst werden. Die beiden zum damaligen Zeitpunkt zur Verfügung stehenden Bewertungsfunktionen (GoldScore und ChemScore) waren laut Dokumentation von Protein-Liganden-Komplexen abgeleitet worden und waren - wie sich zeigte - beide nicht in der Lage, die korrekte Positionierung des Ursprungsliganden zu gewährleisten: GoldScore platzierte den Liganden um 180° gedreht immerhin noch relativ nahe der ursprünglichen Position, wohingegen ChemScore den Bulgetunnel komplett vermied. Somit war eine sinnvolle Selektion der potentiellen Testkandidaten mehr als fraglich, zumindest mittels der ChemScore-Bewertungsfunktion. Da zu diesem Zeitpunkt bereits im dem SFB579 angehörigen AK-Gohlke eine spezielle von RNA-Liganden-Komplexen abgeleitete DrugScore-Variante (DrugScore^{RNA}) (Pfeffer und Gohlke 2007) in Entwicklung war, wurde während dieser Studie von den Docking-Ergebnissen Abstand genommen, um auf die Fertigstellung besagter neuer Scoringfunktion zu warten.

Eine genauere Betrachtung der Ergebnisse der GoldScores-Docking-Runde ließ auch – wegen der dem ursprünglich postulierten Selektionsmechanismus widersprechenden Bewertung einer Auswahl der als Negativdatensatz verwendeten COBRA (willkürliche Auswahl jeder 50ten Struktur aus dieser Datenbank) – kein Vertrauen in die Diskriminierungsfähigkeit der Default-Scoringfunktion erwarten; siehe hierzu Abbildung 21 und Abbildung 22, welche die Scoreverteilungen der untersuchten Subsets darstellen. Hierbei ist deutlich zu sehen, dass dem COBRA-Subset konstant höhere Werte zugewiesen wurden, als dies für die eigentlich „speziellen“ RNA-Liganden-Subsets der Fall war.

Schlussfolgerungen aus der Analyse der GOLD-Dockinglisten:

Die retrospektive Analyse der Gold-Dockinglisten mit den erzielten Laborergebnissen ergaben zwar vergleichsweise gute Trefferquoten bei der Rückführung der Aktiven aus dem Labor – im Vergleich zur alleinigen Verwendung der „Top 10“ der ANN-Filter-Hitlisten. Da jedoch die zum Docking verwendeten Auswahllisten nur einen Anteil der gesamten Specs-Datenbank ausmachen, die zuvor mittels der besagten ANN-Filter erhalten wurden, ist der Vergleich der beiden Methoden durch diese Vorauswahl beeinflusst. Für einen fairen Vergleich der Verfahren müsste dementsprechend die gesamte Specs-Datenbank durch das Dockingverfahren bewertet, dann die Auswahl von Substanzen basierend auf den „Top 10“ durchgeführt und diese dann im Labor getestet werden. Ein solcher Vergleich wäre zwar interessant gewesen, war jedoch aus Zeitgründen nicht im Rahmen dieser Arbeit durchführbar:

Da ein Docking-Lauf für eine Substanz im Durchschnitt 20 Sekunden (3 GHz Opteron-Prozessor, 8GB RAM) dauerte, würde dies bei den damals 229.658 Strukturen des Specs-Katalogs zu einer benötigten Rechenzeit von ca. 53 Tagen führen und bei einer durchschnittlichen produzierten Datengröße von 0,34 MB pro „gedockter“ Struktur in einem Gesamtdatenvolumen von ca. 77 GB resultieren. Da ferner der Einfluss des jeweiligen gewählten Konformers – im Rahmen des „Schlüssel-Schloss“-Prinzips – eine nicht zu vernachlässigende Unsicherheit bei der Bewertung nur eines einzigen solchen Docking-Experiments beinhaltet, ist, um eine verallgemeinerbare umfassende Betrachtung des Sachverhaltes zu erlangen, dies nur durch die vergleichende Bewertung multipler Konformere erreichbar, die alle mit diesem standardisierten Verfahren bewertet werden müssten. Bezogen auf die alleinige Verwendung der im PDB-Eintrag 1ANR enthaltenen 20 Konformere hätte dies dementsprechend zu einer Rechenzeit von 1060 Tagen und 1,5 TB

an Daten geführt. Dies alleine hätte bereits den Rahmen dieser Doktorarbeit gesprengt (zumal für die Auswertung dieser Datenmenge neue Software angefertigt hätte werden müssen). Des Weiteren müsste der Anteil jedes verwendeten Konformers an der Gesamtkonformerpopulation bekannt sein, um eine begründete Aussage über die Gewichtung der Läufe untereinander zu ermöglichen. Die Auswahl nur eines einzelnen Konformers entspräche hier der Willkür oder einem nicht einschätzbaren Zufallsereignis. Dementsprechend sollte ohne eine differenzierte Bewertungsmöglichkeit jeder bekannte Konformer gleich in die endgültige Bewertung einfließen (Laplace Wahrscheinlichkeit / Gleichwahrscheinlichkeit (Papula 1999)).

Bei Einbeziehung der ebenfalls bekannten TAR-Konformere der PDB-Einträge 1LVJ (12 Konformere), 1QD3 (16 Konformere) und 1ARJ (20 Konformere) stellt sich ferner die Frage, wie die durch den jeweiligen Liganden bedingten Konformere gewertet werden sollten.

Der Auswahl möglicherweise geeigneter Konformere für das strukturbasierte Liganden Design ist der Abschnitt 4 der RNA-Konformer-Datenbank gewidmet.

3.2.10 „Cherry-Picking“ von Testsubstanzen aus den sortierten Trefferlisten zur experimentellen Bestimmung der Aktivität im Tat-TAR-FRET-Assay

Hintergrund:

Zur Überprüfung der Vorhersageergebnisse der eingesetzten RNA-Liganden-Grob-Filter wurden für die Tests im Labor einige Substanzen anhand der im Virtuellen-Screening gewonnenen Hitlisten ausgewählt. Die manuelle Auswahl der Substanzen erfolgte basierend den Erfahrungen der auswählenden Personen zur Auswahl bioaktiver Substanzen („Cherry-Picking“). Die ausgewählten Substanzen sollten hierbei eine subjektive Ähnlichkeit zu bekannten Wirkstoffmolekülen aufweisen.

Durchführung:

Das „Cherry-Picking“ erfolgte anhand der erstellten Einzellisten für die beschriebenen Auswahlverfahren (Abschnitt 3.2.8 RNA-Liganden-Grob-Filter-Hitlisten; Abschnitt 3.2.9 Gold-Docking-Hitlisten). Meist wurde nur innerhalb der „Top 10“ der jeweiligen Screening-Runden ausgewählt. Die Auswahl nahm hauptsächlich Prof. Dr. Gisbert Schneider basierend auf seinen Erfahrungen zur Auswahl bioaktiver Substanzen für die Testung im Labor vor, eine Ausnahme stellt hier die Liste des CATS-RNA-GF dar.

Für eine Übersicht der „Cherry-Pick“-Anteile der jeweiligen Einzellisten siehe Tabelle 8.

Ergebnisse:

Insgesamt wurden 42 Substanzen zur Testung in den Labor-Assays ausgewählt, von denen 36 Substanzen geliefert wurden. Tabelle 8 führt die Aufteilung der ausgewählten Substanzen auf die jeweiligen Auswahllisten zurück und enthält auch die Gesamtgröße der jeweiligen Vorschlagsliste. Da bei der Auswahl der auszuwählenden Substanzen die bereits durchgeführte Auswahl in einer anderen Liste mit betrachtet wurde, wurde zusätzlich untersucht, welche Anzahl Substanzen inklusive der Überlappung zu anderen Auswahllisten somit theoretisch maximal gewählt wurden.

Tabelle 8: Cherry-Picking Anteile der Einzelvorschlagslisten; GF = Grobfilter, GD = GOLD Docking; Overlap = Zusammengeführte Liste aus Überlappungen der erwähnten ANN-Filter-Vorschlagslisten

	"Cherry-Picked" in dieser Liste	Total "Cherry-Picked"	Größe der Vorschlagsliste
TC3D GF Top10	7	8	10
MOE GF Top10	4	5	10
CATS GF Score > 0,5	17	18	40
Overlap MOE Triple GF Score > 0,5	2	3	6
TC3D GD Top10	7	8	10
MOE GD Top10	9	9	10

3.2.11 Auswertung der RNA-Binder-Testsubstanzsammlung von Specs im Tat-TAR-FRET-Assay

Hintergrund:

Die 36 von Specs gelieferten Substanzen wurden mit dem Tat-TAR-FRET-Assay im AK Göbel vermessen. Eine ausführliche Beschreibung der Tests kann dem Abschnitt 2.7 „Der Tat-TAR-FRET-Assay“ entnommen werden. Dieser Test bewertet Substanzen nach ihrem Einfluss auf die relative Fluoreszenzaktivität einer Lösung eines mit einem FRET-Paar-markierten Tat-Peptids und der TAR. Die Bindung des Peptids an die RNA erhöht hierbei die Fluoreszenzausbeute auf ein zu beobachtendes Maximum, wohingegen freies Peptid eine verminderte Fluoreszenzaktivität zeigt und somit das Minimum des Messbereichs darstellt. Als „aktiv“ wurde eine Substanz betrachtet, wenn diese den relativen Fluoreszenzwert mindestens um 0,2 Einheiten senkte; zur Umwandlung der Rohdaten in die relativen Aktivitätswerte siehe Abschnitt 2.7.2 „Prozessierung der Rohdaten zu relativen Aktivitätswerten“.

Ergebnisse:

Tabelle 9 gibt die ermittelten relativen Fluoreszenz-Aktivitäten für die Messungen der M-Reihe an, als zusätzliche Referenz wurden ebenfalls noch die Werte für Tri-Arg ermittelt. Abbildung 25 zeigt die Auftragung der in Tabelle 9 angegebenen RFA-Werte für die M-Reihe und Abbildung 26 zeigt die in dieser Messreihe nach der Auswertung der relativen Messwerte identifizierten potentiellen RNA-Liganden.

Tabelle 9: RFA-Werte für die M-Reihe im Tat-TAR-FRET-Assay; Anzahl Messpunkte $n \geq 3$

	Mittelwert RFA	Stdabw. RFA
Ref: Tat	0,00	0,09
Ref: Tat & TAR	1,00	0,06
Tri-Arg	0,09	0,03
M1	0,97	0,14
M2	1,20	0,10
M3	1,05	0,02
M4	0,99	0,12
M5	1,06	0,03
M6	1,08	0,02
M7	0,90	0,11
M8	0,72	0,08
M9	1,02	0,04
M10	0,74	0,18
M11	1,02	0,01
M12	0,94	0,12
M13	1,09	0,04
M14	1,07	0,02
M15	0,98	0,04
M16	1,00	0,15
M17	1,08	0,02
M18	1,09	0,09
M19	0,99	0,01
M20	0,92	0,12
M21	1,00	0,07
M22	1,03	0,10
M23	0,40	0,03
M24	0,83	0,11
M25	0,94	0,13
M26	0,94	0,09
M27	-0,52	0,02
M28	0,94	0,09
M29	1,04	0,11
M30	0,90	0,06
M31	0,91	0,24
M32	0,68	0,01
M33	0,93	0,21
M34	0,92	0,10
M35	0,86	0,02
M36	0,96	0,04

“FRET II” (A3)

@ 100 µM, 1% DMSO

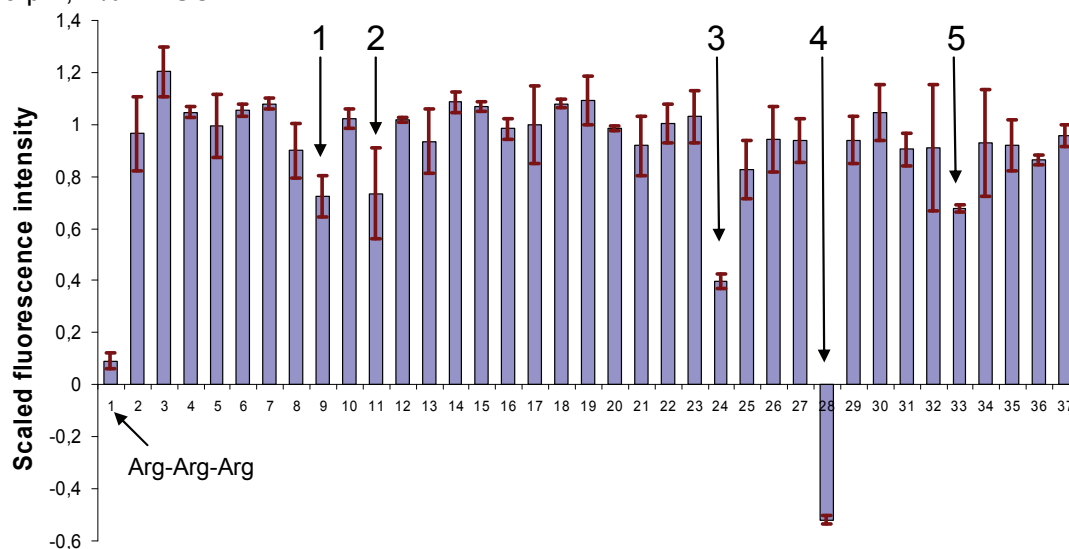


Abbildung 25: Relative Fluoreszenz-Aktivität der 36 von Specs gelieferten Substanzen; Ein Y-Wert von 1 = inaktiv im Assay; An Position 1 ist die Referenz Triple-Arginin dargestellt; Aktive Substanz Nr4 stellte sich als Peptid-Binder⁴ heraus. Für Nr.2 konnte gezeigt werden, dass die starke Absorption dieser Substanz, das Messergebnis verfälschte (schwarz).

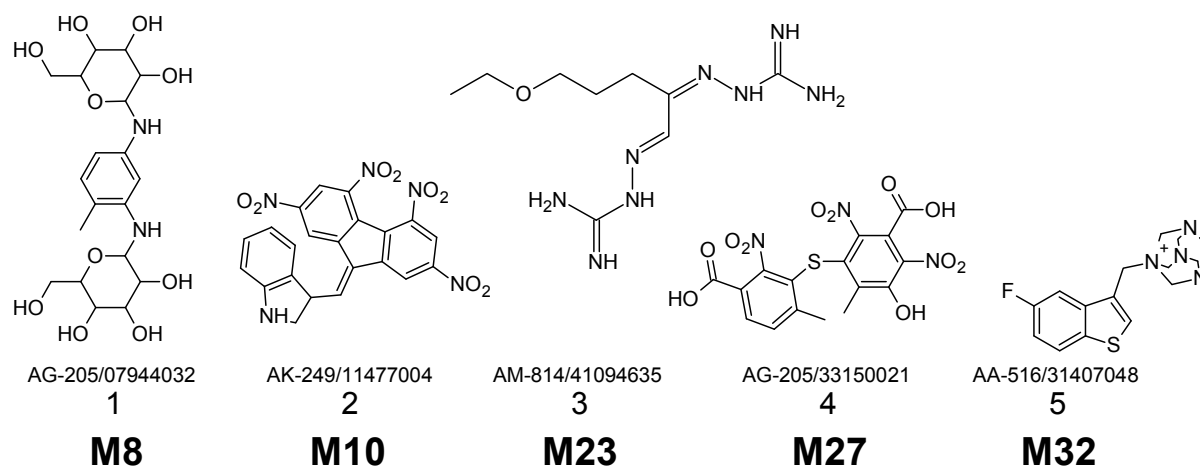


Abbildung 26: Aktive Substanzen im TAR-FRET-Assay. Angegeben sind die Specs-IDs und die Aktiven-Nummer im Assay; Substanz 2 ist schwarz und stört dementsprechend den Assay, Substanz 4 zeigt Aktivität mit dem Peptid; verbleibende RNA-Binder sind somit die Substanzen 1, 3 und 5.

⁴ **Definition Peptid-Binder:** Als Peptid-Binder wird im Rahmen dieser Arbeit eine Substanz bezeichnet, welche im Tat-TAR-FRET-Assay eine Reduktion der Fluoreszenzausbeute verursacht, die unabhängig vom Vorhandensein der TAR in der Lösung ist und ohne bei einer Fluoreszein-Vergleichslösung diese Reduktion zu verursachen [pers. Kommunikation Dr. Ute Scheffer].

3.2.12 Nachmessen der Substanzen

Hintergrund:

Nach zwei Jahren Lagerung bei -80°C wurden die Substanzen im Rahmen der Messungen der neuen Bestellrunde der FRET-Assay-Spezifischen ANN-Filter erneut vermessen. Des Weiteren wurden Absorptionsspektren der Substanzen unter Assay-Bedingungen angefertigt (AK-Göbel), als auch Massenspektren aufgenommen (AK-Karras: Markus Weidlich). Die Absorptionsspektren können im Anhang 7.5 ab Seite 208 gefunden werden.

Ergebnisse und Diskussion:

Die Ergebnisse dieser Tests zeigten, dass die Aktivitäten der Substanzen M23 und M32 für diesen Assay bestätigt werden konnten, ohne Auffälligkeiten bei den Spektren zu zeigen. M10 und M27 zeigten weiterhin Aktivität. Jedoch musste M10 aufgrund des Absorptionsspektrums als verlässlicher RNA-Binder ausgeschlossen werden. M27 wurde aufgrund seiner Aktivität mit dem Peptid ausgeschlossen, wies im Absorptionsspektrum jedoch keine Auffälligkeiten auf. Substanz M8 zeigte keine Aktivität mehr im Assay. Das daraufhin aufgenommene Massenspektrum offenbarte jedoch, dass diese Substanz auch nicht mehr im Substanzvorrat nachgewiesen werden konnte. Die Aktivität am Anfang der Studie wird somit unter Vorbehalt der Instabilität dieser Substanz weiter angenommen. Das Problem der potentiellen Instabilität von Testsubstanzen nach Lagerung in DMSO-Lösungen wurde bereits von anderen Gruppen beschrieben (Cheng et al. 2003; Kozikowski et al. 2003a; Kozikowski et al. 2003b).

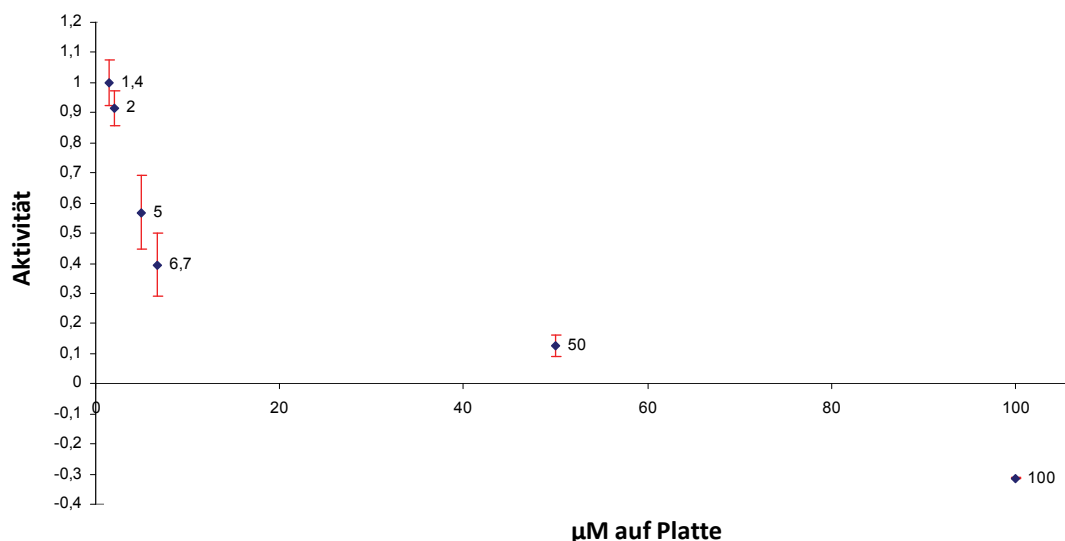


Abbildung 27: Messungen der Substanz M27 zur Bestimmung des IC_{50} im Tat-TAR-FRET-Assay; die X-Achse repräsentiert die Konzentration des jeweiligen Tests, die Y-Achse zeigt die ermittelte relative Fluoreszenzaktivität. Erkennbar ist das Unterschreiten der Basislinie bei einer Konzentration von 100 μM .

Schlussfolgerung:

Es wird davon ausgegangen, dass M27 aufgrund seiner Eigenschaft als Polyanion eher mit dem Peptid (Polykation) wechselwirkt, als mit der als Polyanion anzusehenden TAR. Dies würde auch die deutlich gesteigerte Inhibierung der Fluoreszenzaktivität (unter die Basislinie) erklären; siehe Abbildung 27.

3.2.13 Auswertung der RNA-Binder-Testsubstanzsammlung von Asinex im Tat-TAR-FRET-Assay

Hintergrund:

Da im Rahmen der Virtuellen-Screening-Kampagne mit den RNA-Liganden-Grob-Filtern die Datenbanken der Firma Asinex ebenfalls einem Screening unterzogen wurden, erfolgte auch für die „Top 10“ dieser virtuellen Hitlisten eine Auswahl mittels Cherry-Picking. Aus diesen Hitlisten wurden ebenfalls 23 Substanzen durch manuelle Auswahl für die Tests im Labor ausgewählt.

Ergebnisse:

Die Ergebnisse der Labortestreihe sind in Abbildung 28 und Tabelle 10 wiedergegeben. Einzig die Substanz AS20 - Abbildung 29 - konnte in der Testreihe als aktiv bestimmt werden. Diese Substanz lieferte bei den gegebenen Messbedingungen eine Reduktion der relativen Fluoreszenz-Aktivität auf $0,84 \pm 0,04$. Dies entspricht einer schwachen Inhibition der Tat-TAR-Wechselwirkung.

Tabelle 10: RFA-Werte für die AS-Reihe im Tat-TAR-FRET-Assay; Anzahl Messpunkte $n \geq 3$

	Mittelwert	
	RFA	Stdabw. RFA
Ref: Tat	0,00	0,07
Ref: Tat &TAR	1,00	0,09
triArg	0,12	0,02
AS1	1,06	0,01
AS2	1,05	0,02
AS3	1,14	0,06
AS4	1,04	0,11
AS5	1,08	0,02
AS6	1,08	0,08
AS7	1,04	0,15
AS8	0,96	0,08
AS9	0,91	0,06
AS10	1,15	0,02
AS11	1,07	0,12
AS12	1,01	0,05
AS13	1,07	0,05
AS14	1,00	0,09
AS15	1,03	0,07
AS16	0,88	0,10
AS17	1,01	0,08
AS18	1,08	0,16
AS19	1,09	0,06
AS20	0,84	0,04
AS21	1,03	0,04
AS22	1,13	0,06
AS23	0,93	0,05

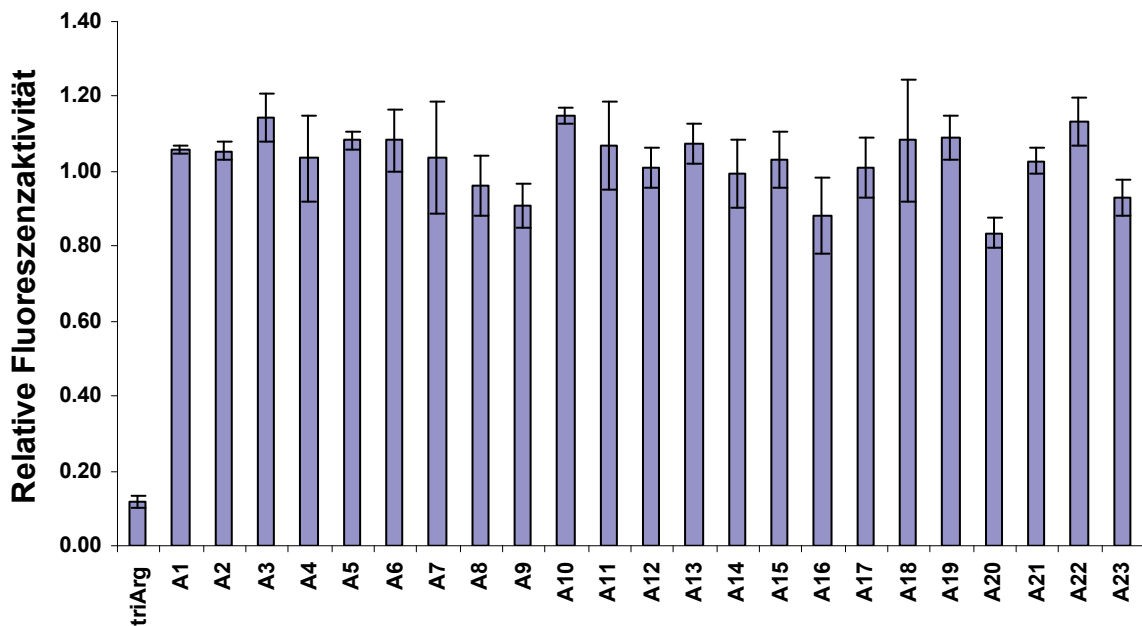
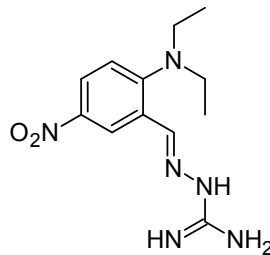


Abbildung 28: Relative Fluoreszenz-Aktivität der 23 von Asinex gelieferten Substanzen; Ein Y-Wert von 1 = inaktiv im Assay; An Position 1 ist die Referenz Triple-Arginin dargestellt; Substanz Nr6 = A20 konnte als einzige Substanz in ihrer Aktivität bestätigt werden.



BAS 01367404
AS20

Abbildung 29: Asinex AS20 – einzige aktive Substanz im TAR-FRET-Assay. Angegeben sind sowohl die Asinex-ID, als auch die LabID; AS20 war die einzige Substanz die in dieser Bestellrunde als aktiv bestätigt werden konnte.

Diskussion:

Die Struktur von AS20 ähnelt der – in der Tat-TAR-FRET-Assay-spezifischen Filterrunde gefundenen und im Labor gemessenen – aktiven Substanz LR9 stark in ihren Substrukturelementen; siehe hierzu 3.3.19. Es wird davon ausgegangen, dass diese Substanz das Vorbild für die später gefundene Substanz LR9 darstellt, da die Substanzen der ersten Testrunde in den Trainingsdatensatz der Tat-TAR-FRET-Assay-spezifischen Filter eingingen.

3.2.14 Auswertung der RNA-Binder-Testsubstanzsammlung im CFivTT-Assay

Hintergrund:

Im Rahmen der Zusammenarbeit mit Markus Weidlich (AK-Karas) wurden die Substanzen, die zuvor im Tat-TAR-FRET-Assay getestet wurden, von ihm im CFivTT-Assay getestet. Dieser Assay dient der Bestimmung einer potentiell antibakteriellen Wirkung von Testsubstanzen in einem *in vitro* Assay; siehe für eine genauere Beschreibung des Assays den Abschnitt 7.6 „CFivTT-Assay“. Hierbei wird davon ausgegangen, dass durch die in diesem Assay-modellierten Schritte der Proteinexpression (Referenzprotein = GFP) – welche Transkription und Translation einschließt – eine Vielzahl an möglichen Interaktionspunkten der Testsubstanzen mit RNA-Zielstrukturen besteht (mRNA, tRNA, Ribosom). Somit sollten im Rahmen der Vorhersage von allgemeinen RNA-Bindern ebenfalls Substanzen gefunden werden können, welche auch Aktivität in diesem Assay zeigen würden. Zusätzlich lässt sich anhand dieses weiteren Assays die Spezifität einer gefundenen Wirksubstanz im Tat-TAR-FRET-Assay weiter eingrenzen.

Ergebnisse:

Durch Mehrfachbestimmung der Aktivitätswerte im CFivTT-Assay konnten letztendlich fünf Substanzen identifiziert werden, die eine reproduzierbare Aktivität innerhalb des genannten Assays aufweisen; siehe hierzu die Abbildung 30, welche die relativen Inhibitionsraten in Prozent relativ zur Kontrollreaktion ohne den Einsatz von Antibiotika zeigt. Die Strukturen der fünf Aktiven sind in Abbildung 31 wiedergegeben. Die gesamten Ergebnisse der Testreihe befinden sich in Tabelle 31 im Anhang, die Auswahl der Ergebnisse für die fünf aktiven Verbindungen wurde in Tabelle 11 ausschnittsweise dargestellt. Die Werte der gesamten Messreihe können dem Anhang in Abschnitt 7.6.1 Seite 213 entnommen werden.

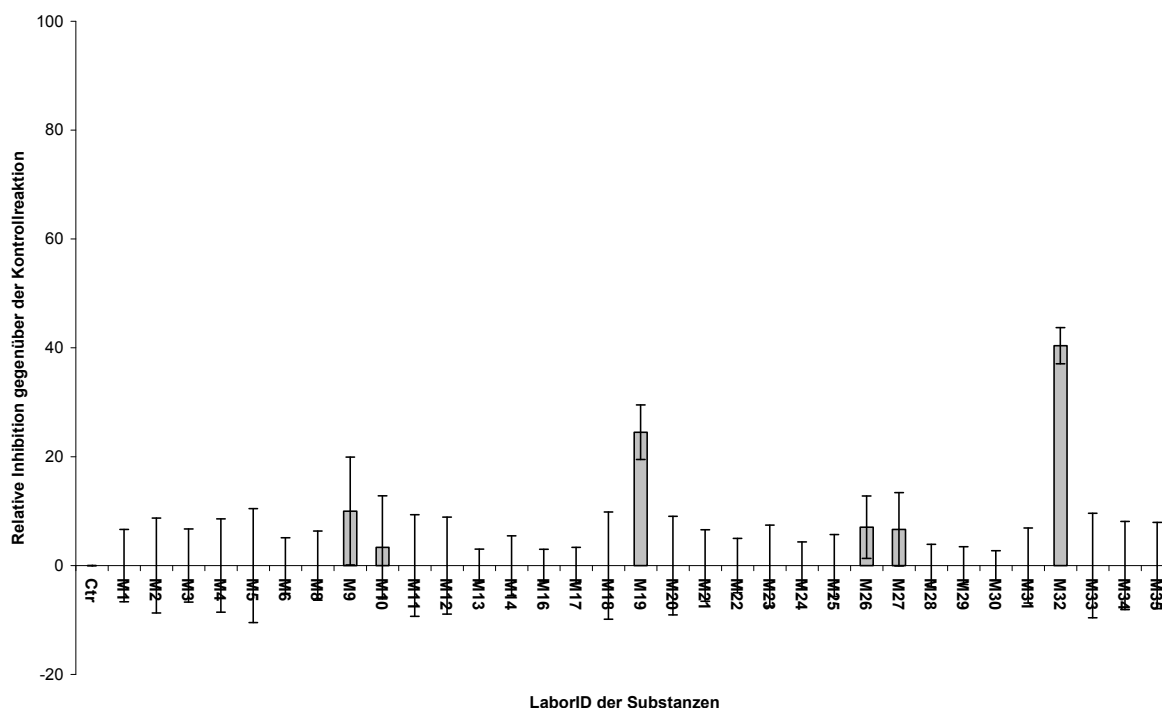


Abbildung 30: Ergebnisse des CFivTT-Assay für die Substanzen der M-Reihe (RNA-Binder-Vorhersagen). Dargestellt ist die relative Inhibition bezogen auf die Kontrollreaktion.

Tabelle 11: Inhibition der gekoppelten Transkription/Translation des CFivTT-Assay für die fünf aktiven Verbindungen der M-Reihe. M19 und M32 zeigen Inhibitionswirkung > 20%, bei gleichzeitiger relativ geringer Standardabweichung während der Testreihen. Testanzahl = 3.

LaborID	mittlere relative Inhibition [%]	Standardabweichung [%]
M9	10,01	9,91
M19	24,5	5,02
M26	7,04	5,74
M27	6,65	6,73
M32	40,4	3,32

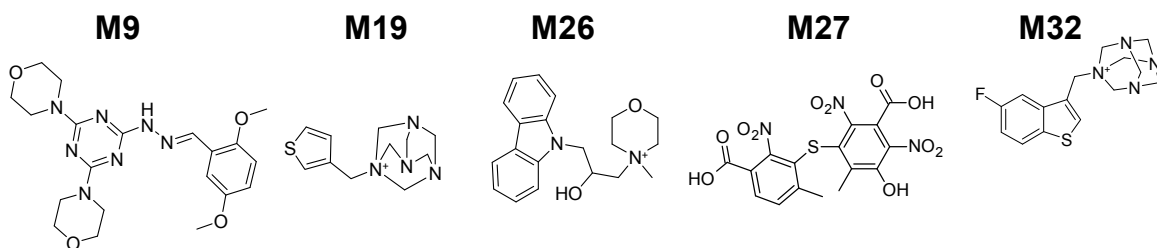


Abbildung 31: Strukturen der Verbindungen, welche im CFivTT-Assay Inhibition gegenüber der Kontrollreaktion zeigten. Die höchste beobachtete Inhibition bei gleichzeitiger geringster Standardabweichung wurde bei M19 und M32 beobachtet.

3.2.15 Auswertung der Überschneidungen zwischen Tat-TAR-FRET-Assay und CFivTT-Assay

Hintergrund:

Das Vorliegen der Aktivitätswerte für zwei unterschiedliche Assays, welche zwei unterschiedliche biologische Zielwirkungen modellieren (antiviral / antibiotisch), lässt die Überprüfung möglicher Gemeinsamkeiten zwischen den Hitlisten der einzelnen Assays zu. Diese Betrachtung der Überschneidungen der Hitlisten ermöglicht eine weitergehende Aussage über die Spezifität der beobachteten Wechselwirkungen (Aktivität) mit den jeweils verwendeten Assay-Systemen (Tat-TAR-FRET-Assay/CFivTT-Assay). Als spezifisch können nur Substanzen angesehen werden, die ausschließlich in einem der beiden Assays Aktivität zeigen. In beiden Assays aktive Substanzen können aufgrund der hier anzunehmenden allgemeinen Affinität zu verschiedenen Assay-Komponenten als potentielle „frequent hitter“ betrachtet werden. Im Rahmen dieser Arbeit könnte aber gerade diese Eigenschaft der unspezifischen Bindung an mögliche RNA-Zielmoleküle der weiteren Aufklärung der RNA-Liganden-Interaktionsmöglichkeiten dienen.

Durchführung:

Die Hitlisten der beiden Assays wurden manuell miteinander verglichen und basierend auf den Überschneidungen das Schnittbild in Abbildung 32 erstellt. Diesem wurden die Darstellungen der Moleküle hinzugefügt.

Dies erfolgte bisher für die Substanzen der M-Reihe und ist in Abbildung 32 dargestellt. Für die AS-Reihe wurde dies nicht mehr durchgeführt, da es hierbei nur einen Aktiven insgesamt

im Tat-TAR-FRET-Assay gab und im CFivTT-Assay aufgrund fehlender Aktivität in den Vortests der AS-Reihe dort auch keine Dreifachmessungen mehr durchgeführt wurden.

Ergebnisse:

Für die Substanzen der M-Reihe ist die Überlappungsanalyse der Hitlisten in Abbildung 32 dargestellt. Hierbei ist zu erkennen, dass die Substanz M32 in beiden Assays Aktivität zeigt. Anzumerken ist hierbei, dass für Substanz M27 weiterhin dasselbe gelten würde, wäre sie nicht im Rahmen der Überprüfung der Art ihrer Wirkung im FRET-Assay als Peptid-Binder aus der TAR-Hitliste ausgeschlossen worden; siehe hierzu Abschnitt 3.2.11. Für den CFivTT-Assay lässt sich diese Aussage nicht genauer spezifizieren, so dass diese Substanz dort weiter als aktiv geführt wird.

Für die Asinex-Bestellrunde (AS-Reihe) erübrigte sich eine derartige Auswertung, da keine der Verbindungen eine Aktivität im CFivTT-Assay zeigte und im Tat-TAR-FRET-Assay auch nur AS20 eine Aktivität zeigte.

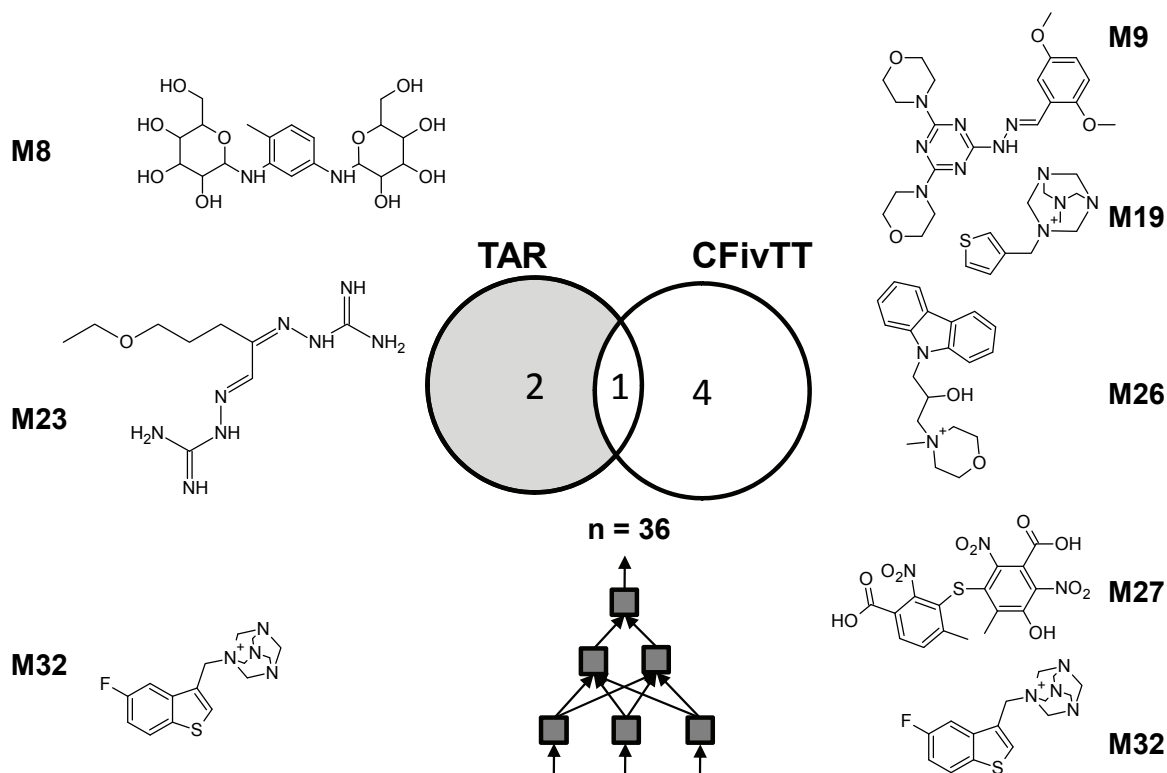


Abbildung 32: Schnittbild der Überlappungen der Aktiven aus Tat-TAR-FRET-Assay für die M-Reihe.

Diskussion:

Basierend auf der Überlappungsanalyse der Hitlisten der Assays verbleibt nur eine Substanz (M32), welche die Eigenschaften aufzuweisen scheint, die für eine messbare Bindung an multiple RNA-Strukturen erforderlich ist.

Struktur M19 zeigt eine strukturelle Ähnlichkeit zu M32 auf Substrukturebene. Die Tatsache, dass die Substanz M19 jedoch im benutzten Tat-TAR-FRET-Assay keine Aktivität aufweist, führte zu der Annahme, dass die Thienyl-Gruppe, die wiederum eine Substruktur von M32s Fluoro-Benzothiophenyl-Gruppe darstellt, bei TAR nicht zur erwünschten starken Bindung führt und somit die in beiden Strukturen vorkommende fixierte positive Ladung des Tetraaza-adamantyl-Gerüsts alleine nicht zu einer messbaren Inhibition der Tat-TAR-

Wechselwirkung ausreicht. Dementsprechend wäre es interessant zu klären, wie genau das Vorhandensein des Fluoro-Benzol-Rings bei M32 dessen Bindung an TAR gegenüber M19 begünstigt. Eventuell könnte dieser Substrukturbereich eine verbesserte Interaktion mit dem Bulge-Bereich der TAR bewirken, der über „Base-Stacking“ die Bindung stabilisiert.

Substanz M23 zeigt hier nur im Tat-TAR-FRET-Assay eine messbare Wirkung, daraus abgeleitet ergibt sich somit eine relative Selektivität dieser Verbindung zu TAR, da diese Verbindung im Gegentest ohne Auswirkungen auf den Proteinexpressionsweg im CFivTT-Assay zu sein scheint. Da der natürliche Ligand – Tat-Protein – mit einer sechs Arginine enthaltenden Sequenz an TAR bindet, ist die strukturelle Ähnlichkeit einer gefundenen aktiven Verbindung gegenüber Arginin nicht weiter verwunderlich. In diesem Fall ist die Ähnlichkeit anhand der Substrukturgruppen klar ersichtlich, da in tri-Arginin (dem benutzten Referenzliganden) das Guanidin-Motiv dreimal vorkommt, und in M23 diese Gruppe zweimal vorhanden ist; siehe hierzu Abbildung 33, Gegenüberstellung der Verbindung M23 und tri-Arginin.

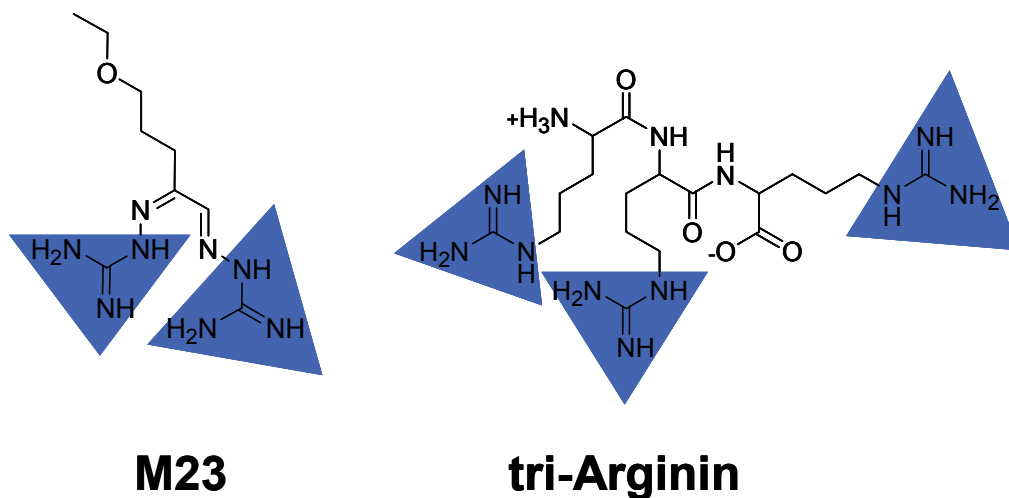


Abbildung 33: Gegenüberstellung M23 und tri-Arg; blau hinterlegt sind die Guanidin-Gruppen in beiden Molekülen; unter physiologischen Bedingungen sind diese Gruppen protoniert, so dass M23 zwei und tri-Arg hierdurch drei positive Ladungen zusätzlich trägt.

Schlussfolgerung:

Die Gegenüberstellung der beiden Assays (Tat-TAR-FRET-Assay und CFivTT-Assay), erlaubt eine differenziertere Betrachtung der beobachteten Aktivitäten. Insbesondere die relative Selektivität der Substanzen M23 und AS20 legt den Schluss nahe, dass aufgrund dieser Strukturen ein für TAR relativ spezifisches Interaktionsmuster bestätigt werden konnte: die Guanidin-Gruppe, hier speziell als „Guanylhydrazon“ (Methylenaminoguanidin).

3.2.16 Struktur-Wirkungsbeziehung von M32 im CFivTT

Hintergrund:

Der hier gezeigte Abschnitt über die SAR-Studie zu M32 mittels CFivTT-Assay, wurde von Markus Weidlich im Rahmen seiner Doktorarbeit alleinig initiiert und durchgeführt. Da die Ergebnisse aber zur Erstellung des in Abschnitt 3.2.17.1 gezeigten Pharmakophormodells herangezogen wurden, werden die Versuche an dieser Stelle kurz vorgestellt. Eine gemeinsame Veröffentlichung über M32 ist ebenfalls in Vorbereitung.

Durchführung:

Zur Überprüfung der im CFivTT-Assay gezeigten Aktivität wurde von Markus Weidlich die Substanz MW1 (1-Prop-2-ynyl-3,5,7-triaza-1-azonia-adamantanbromid) bei Sigma Aldrich (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany, <http://www.sigmaaldrich.com/>) zusammen mit Benzo[b]thiophen geordert. (5-Chloro-1-benzothiophen-3-yl)methylamin, wurde von ihm bei ACROS (Acros Organics, Geel, Belgien, <http://www.acros.com/>) bestellt. Diese drei Verbindungen – in Abbildung 34 – wurden im Rahmen der SAR-Studie im CFivTT-Assay auf Aktivität getestet, siehe Abbildungen 35 und 36. Abbildung 34 zeigt die Strukturen oben beschriebener Verbindungen.

Ergänzende Anmerkungen zur Durchführung:

Die Assay-Bedingungen für diese Testreihen unterscheiden sich zum Teil von den vorher besprochenen Triplikate-CFivTT-Screening-Runden, so dass hier nur ein qualitativer Vergleich der Inhibition möglich ist; siehe hierzu die folgenden Erklärungen.

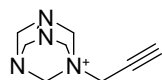
Größter Unterschied war hier einerseits die zwischenzeitlich erfolgte Umstellung des Assay-Volumens vom 100 µl- auf einen 25 µl-Maßstab, sowie der folgenden versuchsspezifischen Änderungen.

Bei Versuch 1 (M32 gegen Thianaphthen und Methylamin) wurde zusätzlich zum veränderten Volumen, auch noch die Kontrollreaktion mit einer DMSO-Konzentration von 2,5% durchgeführt, so dass die angegebenen Inhibitionswerte dem alleinigen Effekt der eingesetzten Testsubstanzen entsprechen, da die DMSO-Konzentration nunmehr in allen Messungen dieses Versuchs gleich war.

Bei Versuch 2 (M32 gegen MW1) wurde die Kontrollreaktion ohne Zugabe von DMSO durchgeführt, analog zur Durchführung bei den Triplikate-Screening Runden. So muss bei diesen Inhibitionsangabe berücksichtigt werden, dass ein Teil der beobachteten Inhibition durch das eingesetzte DMSO in den Testsubstanzlösungen verursacht wurde. Da hier aber nur die relative Inhibition von M32 gegen MW1 überprüft werden sollte und die DMSO-Konzentration in beiden Testsubstanzlösungen gleich war, ist dieser qualitative Vergleich erlaubt.

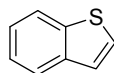
Ergebnisse:

Bei der Untersuchung zeigte sich, dass die beiden getesteten heteroaromatischen Strukturen MW2 und MW3 nur eine geringfügige Inhibition aufweisen (Abbildung 35), wohingegen das Triaza-1-azonia-adamantyl enthaltende MW1 beinahe gleichwertige Aktivität zu M32 im CFivTT-Assay aufwies (Abbildung 36).



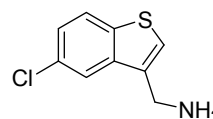
1-Prop-2-ynyl-3,5,7-triaza-1-azonia-tricyclo[3.3.1.1^{3,7}]decane

MW1



Benzo[b]thiophene

MW2



C-(5-Chloro-benzo[b]thiophen-3-yl)-methylamine

MW3

Abbildung 34: Substanzen aus dem Struktur-Wirkungsbeziehungstest für M32 (durchgeführt von Markus Weidlich im CFivTT-Assay).

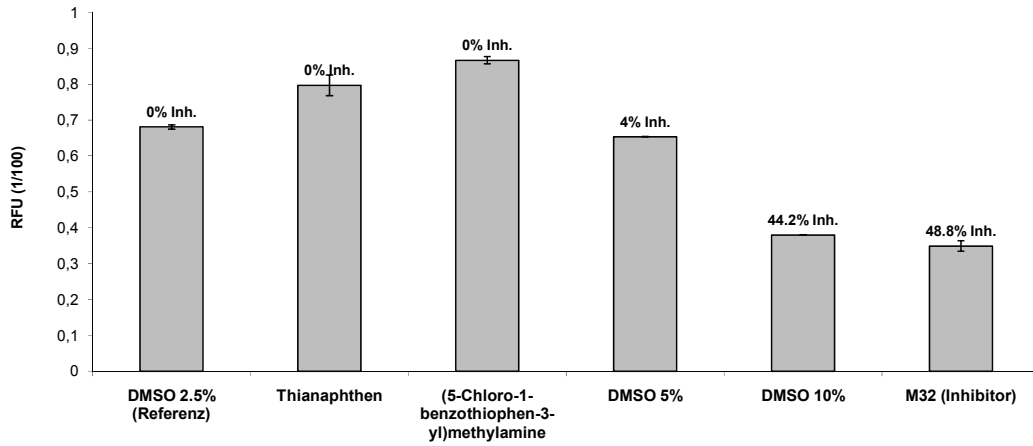


Abbildung 35: SAR M32 gegen Thianaphthen und 5-CI-BTMA; Konzentration der Substanzen im Test = 100 µM (Versuch 1)

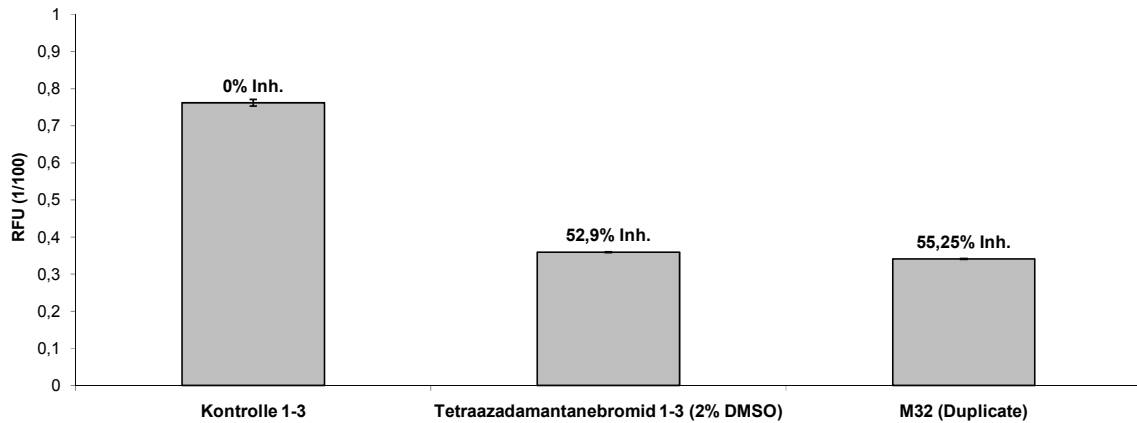


Abbildung 36: SAR M32 gegen 1-prop-2-ynyl-3,5,7-triaza-1-azonia-adamantan, Bromid; Konzentration der Substanzen im Test = 100 µM (Versuch 2)

3.2.17 Ableitung von Pharmakophormodellen aus den gemessenen Aktivitätswerten

Hintergrund:

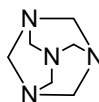
In einer retrospektiven Betrachtung der Substanzen konnten beim Heraussuchen der als aktiv klassifizierten Verbindungen aus der Datenbank einige sich augenscheinlich wiederholende Substruktur-Muster erkannt werden, die daraufhin dazu benutzt wurden, um Pharmakophor-Modelle für die Aktivität und Inaktivität innerhalb der Assays abzuleiten. Diese Modelle können dazu dienen, für zukünftige Versuche gezieltere Interaktionsmodelle zu entwickeln, die die Vorhersagequalität steigern können und somit letztendlich auch das Verständnis für die Raumstruktur der Zielmoleküle verbessern. Dies ist möglich, da die Struktur-Informationen in den Ligandensammlungen wiederum komplementär zu den Zielstrukturen sein müssen, um mit diesen auf die gemessene Art und Weise zu interagieren (den beobachteten Aktivitäten).

Die Vorgehensweise zur Ableitung von Pharmakophor-Modellen entwickelte sich aus der Suche nach den aktiven Verbindungen in der Substanzdatenbank aller getesteter Verbindungen, welche zur Zurückführung der Aktiven auf die verwendeten Auswahllisten erfolgte; siehe Abschnitt 3.2.18. Die einzelnen Auswahllisten stehen ihrerseits bereits für bestimmte Modellierungsversuche. Hinzu kam die Rückführung der Aktiven auf die Substanzen im Trainingsdatensatz, der in Abschnitt 3.2.19 beschrieben wurde.

Hierbei fiel auf, dass sich einige Struktur motive innerhalb der Listen immer wieder zu wiederholen schienen, aus denen sich folgende Pharmakophormodelle ergaben.

3.2.17.1 Tetra-aza-adamantyl

Durch die Verwendung der Substanzen in zwei unterschiedlichen Laborassays ergab sich - basierend auf den jeweiligen Messergebnissen - die Ableitung zweier unterschiedlich gut belegbarer Pharmakophormodelle, welche jeweils auf der enthaltenen Substruktur Urotropin basierten; siehe Abbildung 37. Hierbei konnte jeweils ein Modell für die potentielle Pharmakophorzusammensetzung für den FRET- als auch für den CFivTT-Assay abgeleitet werden. Die Verbindung M32 zeigte in beiden verwendeten Assays eine messbare Aktivität, die basierend auf den Ergebnissen der Untersuchungen mittels des Tat-TAR-FRET-Assays auf die Bindung an (in den jeweiligen Assays) vorhandene RNA-Strukturen zurückgeführt wird; siehe hierzu Abschnitt 3.2.15 – „Auswertung der Überschneidungen zwischen Tat-TAR-FRET-Assay und CFivTT-Assay“.



1,3,5,7-Tetraaza-tricyclo[3.3.1.1^{3,7}]decane

Abbildung 37: Tetra-aza-adamantan (Urotropin), das Grundgerüst der aktiven Substanzen M19 und M32

Tetra-aza-adamantyl-Modell für den CFivTT-Assay

Die Suche nach den Verbindungen M19 und M32 - beide zeigen ein Tetra-aza-adamantyl-Motiv (Abbildung 37) mit Bindung der Seitenkette an einen Stickstoff dieses Grundgerüst, was zu einer positiven Ladung am resultierenden Azonia-Stickstoff führte - lieferte die Erkenntnis, dass innerhalb des Datensatzes weitere Adamantyl-Grundgerüst enthaltende Strukturen ebenfalls zu finden waren, die jedoch ausnahmslos inaktiv waren. Diese unterschieden sich in ihrem Anteil an Stickstoffen und der Art der Verknüpfung zur

Seitenkette. Bei den beiden aktiven Substanzen handelte es sich um eine Tetra-aza-adamantyl mit Anschluss der Seitenkette an einem der Stickstoffe, was zum Fixieren einer positiven Ladung im Molekül führte. Dahingegen handelte es sich bei den inaktiven Verbindungen um ein Tri-aza-adamantyl, bei dem der Anschluss der Seitenkette über ein in diesem Fall den Stickstoff ersetzendes Kohlenstoffatom erfolgte; siehe hierzu Abbildung 38.

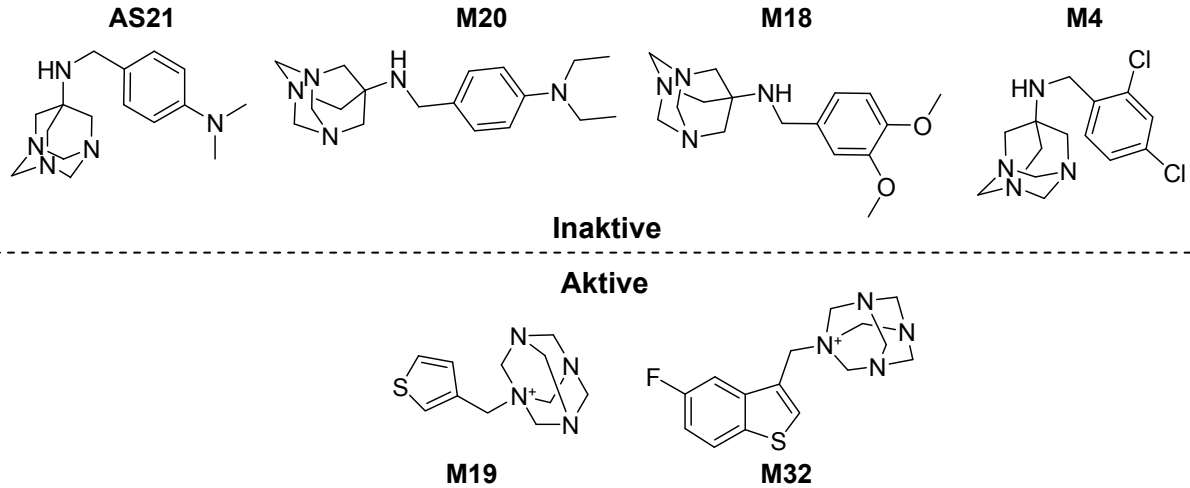


Abbildung 38: Adamantyl-Grundgerüste in den Testreihen für das Labor – Gegenüberstellung Aktive gegen Inaktive (CFivTT-Assay); M32 zeigt auch Aktivität im Tat-TAR-FRET-Assay

Aus der Abbildung 38 lassen sich somit zwei generelle Strukturmuster für Aktive und Inaktive ableiten, wobei eine davon für die RNA-Bindung von Vorteil zu sein scheint. Diese Strukturmuster – welche sich die jeweiligen Gruppen Teilen – wurden in Abbildung 39 als reduzierte Molekülgrundgerüste dargestellt.

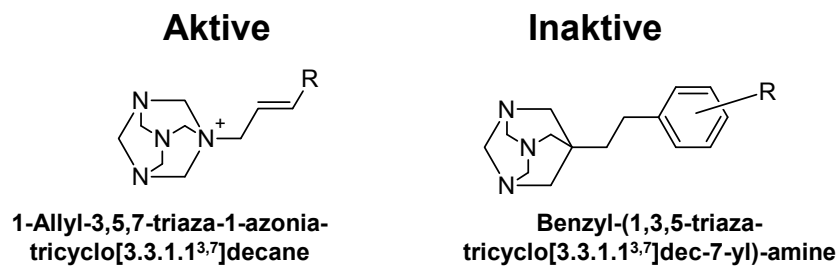


Abbildung 39: Scaffoldvergleich der aktiven/inaktiven Verbindungen im CFivTT-Assay: Betrachtet wurden ausschließlich die ein Adamantyl-Grundgerüst enthaltenden Strukturen.

Größter Unterschied der Aktiven von den Inaktiven ist hierbei der Übergang vom Adamantyl-Gerüst zur Seitenkette über ein positiv geladenes Azonia-Stickstoffatom (Abbildung 39).

Zur Klärung der Frage, ob die beobachtete Aktivität allein auf dem Einführen einer positiven Ladung in eine Testsubstanz beruhen könnte, wurde Abbildung 40 erstellt. Diese zeigt die in den drei Messreihen getesteten Moleküle mit enthaltener positiver Ladung. Hierbei zeigte sich, dass innerhalb der Aktiven zwar eine weitere Verbindung (M26) mit einer positiven Ladung enthalten ist, jedoch auch fünf weitere (positiv geladene) inaktive Verbindungen vorhanden sind. Da die Aktivität der aktiven Verbindung M26 ($RFI = 7,04 \pm 5,74$) jedoch deutlich unter der beobachteten Aktivitäten von M19 und M32 liegt, wurde aus den Strukturen der beiden Letzteren geschlossen, dass das hierin enthaltene Tetra-aza-adamantyl-Strukturmotiv für die beobachtete Aktivität von M19 und M32 hauptverantwortlich ist. So kann davon ausgegangen werden, dass das beliebige Einbinden einer positiven Ladung alleine noch nicht das Optimum für einen allgemein affinen RNA-Binder darstellt,

3. Liganden-basierte Ansätze

sondern zusätzlich durch die drei weiteren Stickstoffe - welche durch den Atomkäfig der Adamantyl-Gruppe räumlich fixiert werden – hierbei eine weiter unterstützende Wirkung zur Verbesserung der Bindungsstärke erreicht werden kann.

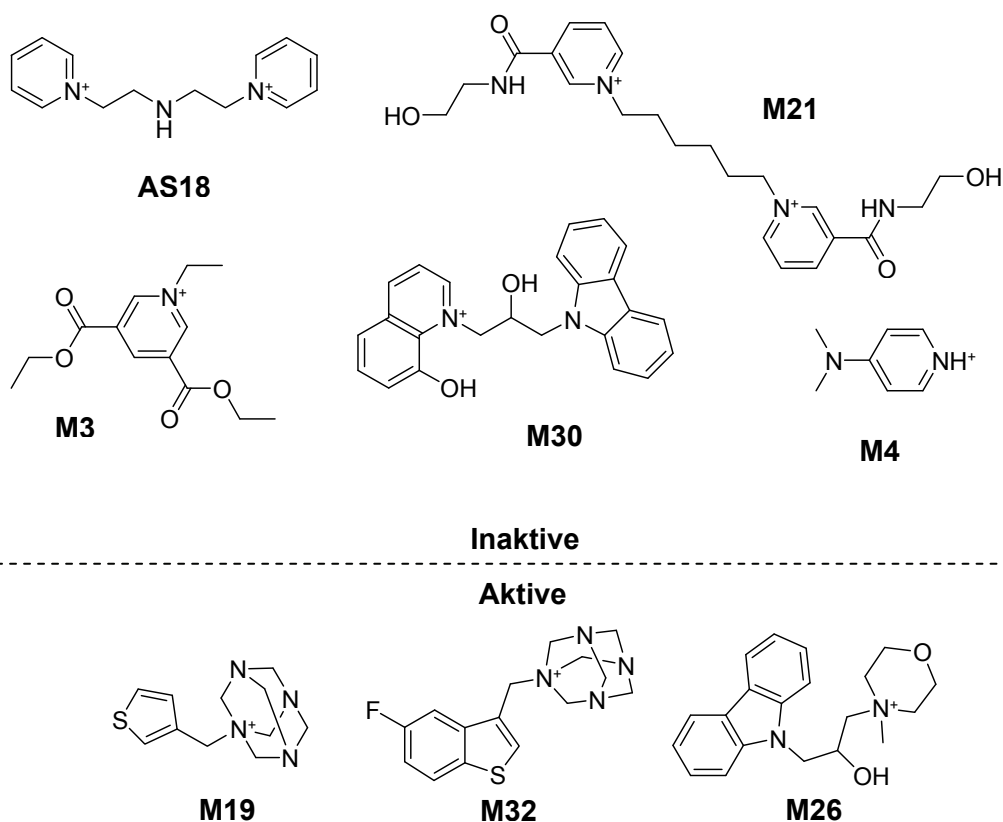


Abbildung 40: Positiv geladene Testsubstanzen. Die Auftragung erfolgte getrennt nach Aktivität im CFivTT-Assay.

Die Stichprobe mit nur zwei aktiven und vier inaktiven Verbindungen (Abbildung 38) konnte durch die beschriebenen Versuche (Abschnitt 3.2.16) von Markus Weidlich auf drei aktive Verbindungen gegenüber vier inaktiven Verbindungen erhöht werden. Somit standen für die weitere Modellerstellung für die Aktiven die Verbindungen M19, M32 und MW1 zur Verfügung, wohingegen sich die Verbindungen M4, M18, M20 und AS21 in der Gruppe der inaktiven Verwendung befanden.

Aus den nun zur Verfügung stehenden Verbindungen wurden die offensichtlichen Gemeinsamkeiten der drei Aktiven Verbindungen als einfaches Pharmakophormodell abgeleitet; siehe Abbildung 41. Dieses reduziert sich auf eine Tri-aza-1-azonia-adamantyl-Struktur mit einer positiven Ladung, auf die in einer Entfernung von zwei gesättigten Kohlenstoffbindungen mindestens eine ungesättigte Bindung folgt. Ob das Vorhandensein dieser ungesättigten Bindung notwendig für eine messbare Aktivität ist, kann hier mangels fehlender Gegenbeispiele (ohne ungesättigter Bindung an dieser Stelle) noch nicht endgültig geklärt werden. Relativ sicher kann hier jedoch aufgrund der vorhandenen Negativbeispiele (M4, M18, M20 und AS21) mit Tri-aza-adamantyl-Gerüst davon ausgegangen werden, dass die räumliche Anordnung der Stickstoffe innerhalb des Adamantylgerüsts allein genommen nicht für eine messbare Aktivität ausreicht.

Die Auslagerung des Azonia-Stickstoff aus dem Adamantyl-Käfig auf die Seitenkette bewirkt zwar immerhin noch eine positive Partialladung in relativer Nähe zur Originalposition der positiven Ladung, bei der ferner auch die weiteren spezifizierten Abstände eingehalten werden. Jedoch konnten für diese Verbindungen keine Aktivitäten in den Assays

nachgewiesen werden. Daher scheint die Fixierung einer „permanenten“ positiven Ladung gegenüber einer Partialladung hier einen Vorteil bei der Bindung der untersuchten Verbindungen an RNA-Strukturen zu bieten, obwohl durch die Auslagerung des Stickstoffs eine weitere Möglichkeit für den Aufbau einer Wasserstoffbrücke theoretisch gegeben ist.

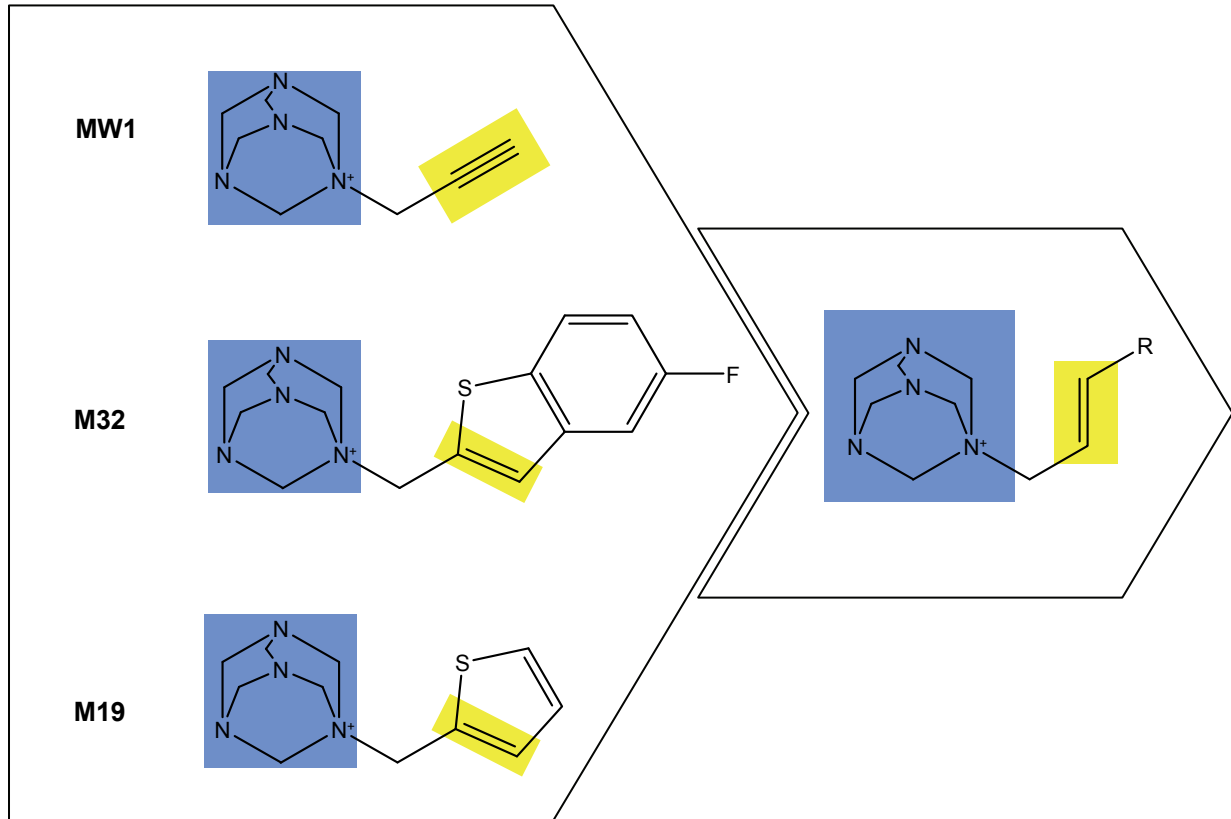


Abbildung 41: Ableitung des Tri-aza-1-azonia-adamantyl-Pharmakophormodells aus den im CFivTT-Assay getesteten Verbindungen: MW1, M19 und M32

Aus der Kombination der Ergebnisse der Screening-Runden (M-Reihe und AS-Reihe), sowie aus der Studie zur SAR von M32 im CFivTT konnte somit das Modell in Abbildung 41 als neue potentielle Antibiotika-Klasse abgeleitet werden.

Anmerkung:

Die Struktur M32 wurde bereits in weiteren Test auf ihre Aktivität als potentielles Antibiotika untersucht (Ausführung erfolgte durch Markus Weidlich im Rahmen seiner Doktorarbeit), siehe hierzu auch den ergänzenden Abschnitt 7.7 im Anhang auf Seite 213 „Ganzzellassay – Wachstumsinhibition durch Substanzen der M-Reihe“, sowie den Abschnitt 7.8 im Anhang auf Seite 214 „DIGE-Proteomics-Ansatz“.

Tetra-aza-adamantyl-Modell für den FRET-Assay

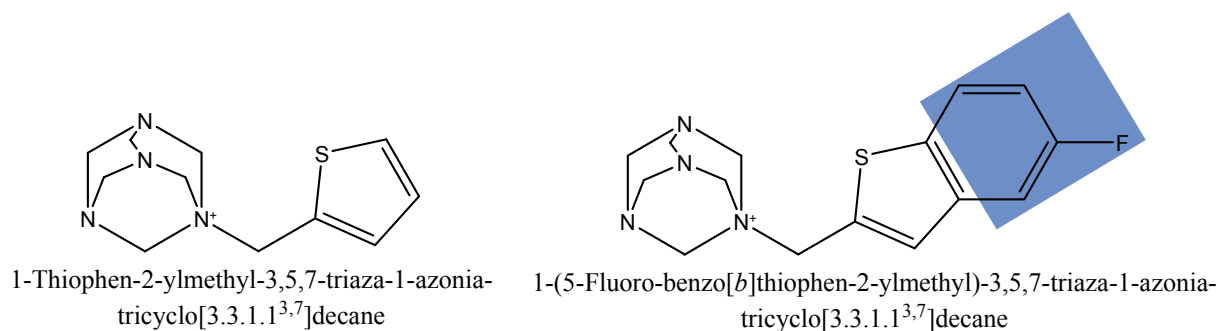


Abbildung 42: Gegenüberstellung von M19 und M32. M19 zeigte keine messbare Aktivität im FRET-Assay, M32 zeigte eine Aktivität (RFA = 0,68 ± 0,01), blau hinterlegt ist der Teil von M32, der für diesen Unterschied wahrscheinlich verantwortlich zu machen ist.

Im Rahmen der Untersuchungen zur Rückführung der Strukturen der M-Reihe auf die verwendeten ANN-Filter, fiel auf, dass sich die Strukturen M19 und M32 in ihrer Substrukturzusammensetzung stark ähneln; siehe hierzu auch Abbildung 42. M19 zeigt im Assay eine RFA von $0,99 \pm 0,01$, wohingegen M32 eine RFA von $0,68 \pm 0,01$ aufweist. Ferner weisen weitere als inaktiv betrachtete Strukturen der Messreihen (M4, M18, M20, AS21) ebenfalls ein Adamantyl-Gerüst (Tri-aza-adamantyl) auf, das sich jedoch in der Art der Verknüpfung unterscheidet. Die Verknüpfung der Seitenkette erfolgt hier nicht wie bei M19 und M32 über ein Stickstoffatom unter Bildung einer permanenten, fixierten Ladung, sondern über ein Kohlenstoffatom; siehe hierzu auch Abbildung 38 im vorherigen Abschnitt über das Tetra-aza-adamantyl-Modell für den CFivTT-Assay, sowie die dortige Argumentation. Im Falle des FRET-Assays, kann somit nur abgeleitet werden, dass das pure Vorhandensein einer Tri-aza-1-azonia-adamantyl-Gruppe für eine messbare Wirkung im FRET-Assay allein nicht ausreicht.

Der zusätzliche eingeführte aromatische Ring mit einem zusätzlichen Fluorsubstituenten in M32 – als einziger Unterschied zu M19 – scheint hierbei den Unterschied in den gemessenen RFA-Werten zwischen M19 und M32 zu bewirken. Als Grund für diesen beobachteten Unterschied wird eine Verbesserung der möglichen Base-Stacking-Interaktion in der Bulge von TAR angenommen, zusätzlich könnte es durch die Einführung des Fluoratoms zu einer weiteren stabilisierenden Wasserstoffbrücke mit TAR kommen, die mit M19 nicht umgesetzt werden kann.

Die genannten Punkte könnten eventuell auch die beobachteten unterschiedlichen Aktivitätswerte der Verbindungen M19 und M32 im CFivTT-Assay (Abschnitt 3.2.14) erklären, da M19 eine RFI von $24,5 \pm 5,02\%$ und M32 dort eine RFI von $40,4 \pm 3,32\%$ zeigt, aufgrund der Unbestimmtheit, mit welchen Assay-Komponenten des CFivTT-Assay jeweils genau die Testsubstanzen interagieren, ist hier jedoch von einer frühzeitigen Interpretation abzusehen.

Ausblick:

Bei der Suche innerhalb der Gruppe der durch manuelle Auswahl aus den ANN-Filter-Hitlisten ausgewählten, jedoch nicht gelieferten Substanzen, konnte eine weitere Verbindung ausgemacht werden, welche dem genannten Grundtypus (Abbildung 43) folgt; siehe Abbildung 44. Die Ähnlichkeit zu den beiden Aktiven M19 und M32 legt hier ebenfalls eine potentielle Wirkung im CFivTT-Assay nahe, und die Größe der Seitenkette lässt hier ebenfalls auf eine mögliche Wirkung im Tat-TAR-FRET-Assay hoffen. Diese Substanz ist bis heute (1.12.2007) nicht lieferbar. Nur die Substanzen M19 und M32 sind weiterhin lieferbar.

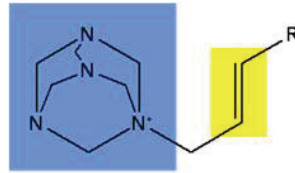


Abbildung 43: Reduziertes Pharmakophormodell für potentiell allgemeine RNA-Binder-Eigenschaften; Aktivitäten in CFivTT und Tat-TAR-FRET-Assay hängen noch vom verwendeten Rest ab.

AG-690/40719331

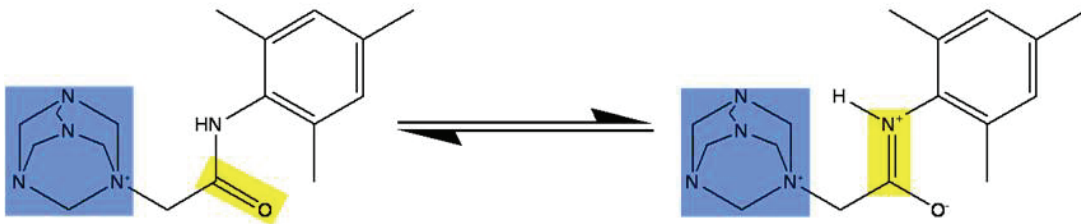


Abbildung 44: AG-690/40719331, Substanz aus dem Cherry-Picking (bis heute nicht lieferbar); links ist die Struktur abgebildet, wie sie bei Specs in der Datenbank zu finden ist, rechts ist eine mögliche Resonanzstruktur der Peptidbindung angegeben, beide erfüllen die Doppelbindungshypothese des Pharmakophormodells aus Abbildung 43

In einer weiterführenden Studie sollte dieses Tri-aza-1-azonia-Strukturmotiv aus Abbildung 44 weiter in ihrem Potential erforscht werden. Des Weiteren bietet die durch die Resonanzstruktur der Peptidbindung mögliche Einführung eines weiteren Ladungspaars eine interessante Perspektive für die weiterführende Untersuchung des Einflusses von Ladungen auf das Bindeverhalten und dies ohne die durch eine fixierte Ladung hervorgerufenen Membran-Permeationsprobleme weiter zu verstärken. Ferner könnte die Struktur AG-690/40719331 durch ihre Peptidbindung für Synthesegruppen als Startstruktur interessant sein, um - ausgehend von diesem postulierten allgemeinen RNA-Binder-Grundmotiv – Derivate-Kataloge mit unterschiedlichen Seitengruppen mittels Peptidchemie zu erstellen, die eine unterschiedliche Selektivität zu bestimmten RNA-Zielstrukturen aufweisen. Diese könnten zumindest in zellfreien/membranfreien Assays, wie dem CFivTT-Assay dazu dienen, mittels Etablierung einer durch dieses Substrukturelement in die Testsubstanzen eingeführten potentiell erworbenen Grundaffinität für RNA, den Einfluss bestimmter Seitenketten auf das Bindeverhalten mittels QSAR-Studien weiter zu erforschen.

3.2.17.2 SAR von M9 für den CFivTT-Assay

Die Untersuchung einer weiteren aktiven Verbindung (M9) des CFivTT-Assay und aller negativen Verbindungen der Testreihen (M-Reihe und AS-Reihe) ergab wieder die Übereinstimmung von Substrukturelementen von Strukturen aus beiden Listen; siehe hierzu Abbildung 45. Hierbei war die Überschneidungs-Liste jedoch kleiner als beim vorigen Modell, da für dieses Modell insgesamt nur fünf Beispiele aus der Liste der im CFivTT gemessenen Strukturen herangezogen werden konnten: M9, M11, M35, AS16 und AS17, siehe Abbildung 45.

Am auffälligsten ist hier die Ähnlichkeit zwischen der aktiven Verbindung M9 und der inaktiven Verbindung AS16, die sich von M9 nur durch den Austausch einer Methyl-Gruppe gegen ein Wasserstoffatom unterscheidet.

Ferner ist der geringe Unterschied von M9 zur inaktiven M11 erkennbar, welche sich nur in einer zusätzlichen Methoxy-Gruppe am Benzolring der Verbindung M11 unterscheidet.

Bei M11 handelt es sich um N-(2,5-Dimethoxy-benzyliden)-N-(4,6-di-morpholin-4-yl-[1,3,5]triazin-2-yl)-hydrazin, bei M11 jedoch um N-(4,6-Di-morpholin-4-yl-[1,3,5]triazin-2-yl)-N'-(2,4,5-trimethoxy-benzyliden)-hydrazin.

Die ebenfalls abgebildeten inaktiven Substanzen M35 und AS17 mit dem Substrukturelement 2,4-Di-morpholin-4-yl-[1,3,5]triazine, welches sowohl in M9, M11 und AS16 vorhanden ist, verdeutlicht ferner, dass dieses Substrukturelement alleine nicht hinreichend für eine Aktivität ist. Aufgrund der geringen Datenlage, lässt sich hier aber nicht mehr aus den Daten ableiten.

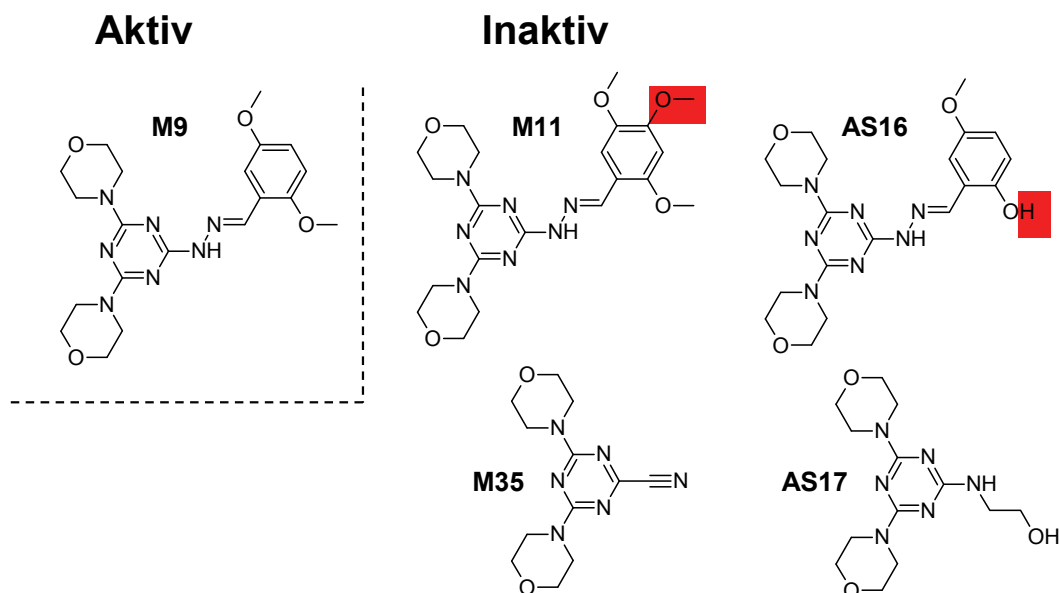


Abbildung 45: Gegenüberstellung der Verbindungen M9, M11 und AS16, welche im CFivTT-Assay, trotz kleinster struktureller Unterschiede unterschiedliche Aktivitäten zeigten; rot unterlegt ist die bei M9 fehlende Methoxy-Gruppe, die für dessen Aktivität gegenüber M11 verantwortlich zu sein scheint. Durch Hinzunahme der Messdaten aus der AS-Reihe, konnte zusätzlich AS16 als weiteres strukturell sehr ähnliches Molekül hinzugenommen werden, der Austausch der Methyl-Gruppe durch ein Wasserstoffatom in AS16 (rot hinterlegt), führte ebenfalls zum Verlust der Aktivität. Das Substrukturmotiv 2,4-Di-morpholin-4-yl-[1,3,5]triazin ist, da es auch in M35 und AS17 vorkommt eindeutig nicht hinreichend für eine Aktivität im Assay.

3.2.18 Rückführung der Aktiven der Assays auf die verwendeten ANN-Filter

Hintergrund:

Zur Bewertung der verwendeten Auswahlverfahren (ANNs) wurde eine retrospektive Betrachtung der erzielten Treffer in den Assays durchgeführt. Diese Bewertung erfolgte anhand der Rückführung der gefundenen Treffer auf die einzelnen Auswahl-Listen, die basierend auf einzelnen Auswahlverfahren (ANNs) oder Weiterverarbeitungen der Listen mit anderen Methoden (Kombination zweier ANN-Filter und GOLD-Docking-Scorelisten) erstellt wurden und zur manuellen Auswahl im Rahmen dieser Arbeit vorgelegt wurden.

Durchführung:

Um die Herkunft der in den Labortests als aktiv befundenen Substanzen zu bestimmen, wurde eine Auswertung der Verteilungen der Aktivenliste auf die Hitlisten der verwendeten

Filter und den daraus abgeleiteten GOLD-Docking-TopScorer-Listen durchgeführt. Hierzu wurden die aktiven Substanzen innerhalb der zur manuellen Auswahl vorgelegten Listen gesucht. Basierend auf dem Anteil der gefundenen Substanzen wurde einerseits eine prozentuale Trefferquote für das Auffinden aller möglicher Treffer in den beiden Assays (Gesamtaktivenanzahl = 7) berechnet, die sich aus der Gesamtanzahl der Aktiven und der in der jeweiligen Vorschlagsliste zu findenden Trefferanzahl ergab; andererseits wurde die Trefferquote innerhalb der jeweiligen Vorschlagsliste bestimmt indem die Gesamtlänge der Liste ins Verhältnis zu den hierin identifizierten Aktiven gesetzt wurde.

Die Ergebnisse der Berechnungen wurden in den Tabellen 12 und 13 zusammengefasst. Tabelle 13 geht hierbei auf die einzelnen Assays ein (Tat-TAR-FRET-Assay und CFivTT-Assay). Eine graphische Aufarbeitung der überlappenden Vorhersage-Ergebnisse für potentielle RNA-Liganden wurde in Abbildung 46 umgesetzt.

Ergebnisse:

Tabelle 12 zeigt die Verteilung der Aktiven auf die Auswahllisten der jeweiligen Filter. Hierbei zeigen sich Unterschiede bei der Anreicherung der gefundenen Hits in den Listen. Die meisten der Hits wurden von den TC3D- und MOE-Filtern identifiziert. Für eine Darstellung der Überlappungen zwischen den Hitlisten siehe das Schnittbild Abbildung 46.

Tabelle 12: Verteilung der aktiven Verbindungen auf die gesamten Filter-basierten Hitlisten; Gesamtanzahl an potentiellen RNA-Liganden = 7; % Hitrate für Aktive = $\text{Aktive}/7 \cdot 100$; % der Liste = $\text{Aktive}/(\text{Größe der Liste}) \cdot 100$.

Filter	Aktive	% Hitrate für Aktive	% der Liste	Größe der Liste
TC3D > 0,5	4	57,1	5,6	71
CATS > 0,5	2	28,6	5,0	40
TC3D & MOE > 0,5	3	42,9	50,0	6
MOE > 0,5	5	71,1	12,5	40

Tabelle 13: Verteilung der aktiven Verbindungen auf die Auswahllisten des Cherry-Picking; gewertet wurde ob die jeweilige Verbindung in der Liste zu finden war, nicht ob sie dort ausgewählt wurde; GF = ANN-Filter Ranking, GD = GOLDDockingRanking der ANN-Filter-Hitlisten.

Cherry-Pick-Listen	auswählbar in Liste	Aktivität in		Peptid-Binder	Substanz-Bezeichnungen
		FRET / CFivTT			
TC3D GF Top10	0	0	0	---	---
MOE GF Top10	2	1	1	---	M8, M19
CATS GF Score > 0,5	2	1	1	1	M23, M27
Overlap MOE Triple GF Score > 0,5	3	2	2	---	M8, M19, M32
CATS GD Top10 (1ANR)	1	1	0	---	M23
TC3D GD Top10 (1LVJ)	1	0	1	---	M9
MOE GD Top10 (1LVJ)	4	2	3	1	M8, M26, M27, M32

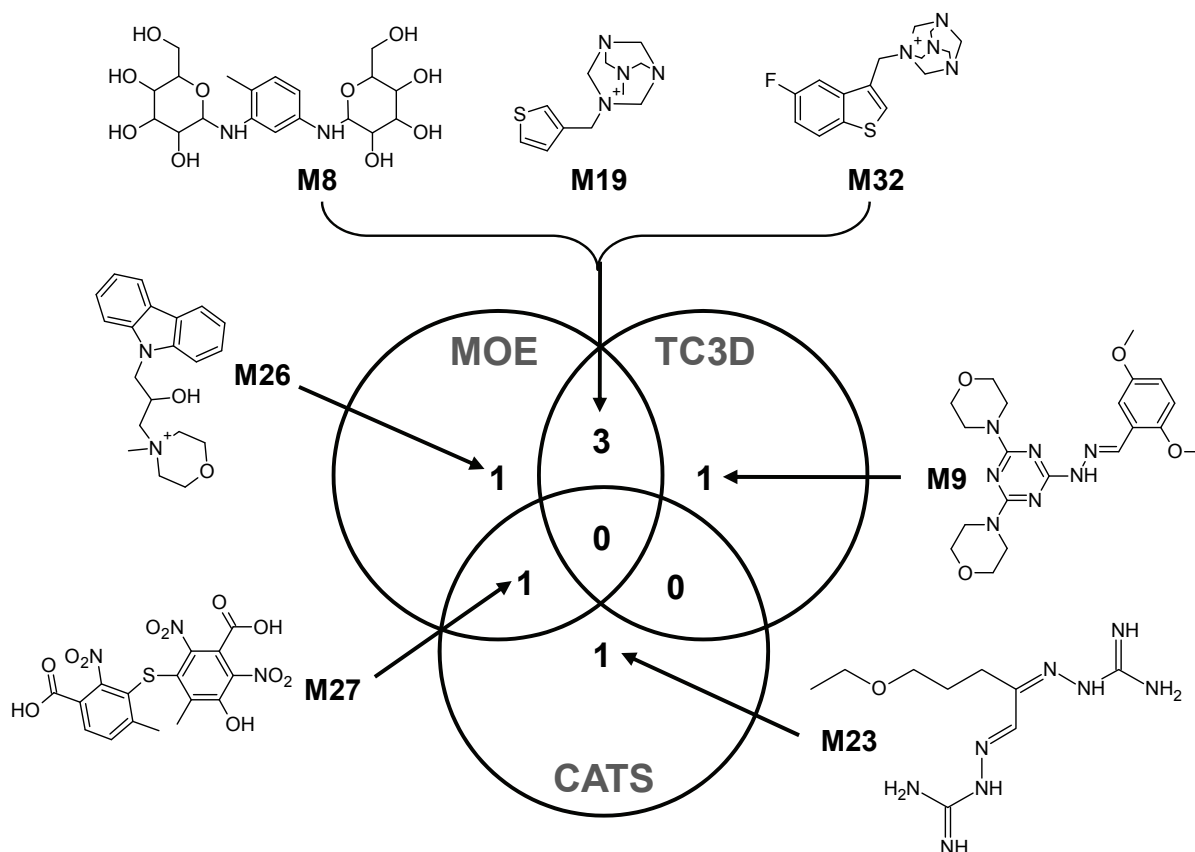


Abbildung 46: Verteilung der Aktiven der M-Reihe auf die Hitlisten der drei Filter

Diskussion:

Abbildung 46 zeigt, dass es keine Überlappungen für alle drei in dieser Arbeit verwendeten ANN-Filter bei der Vorhersage der Aktiven der M-Reihe gibt. Ebenfalls zeigt sich keine Überlappung im Bereich der vom CATS-RNA-GF-Filter und TC3D-RNA-GF-Filter. Der Überlappungsbereich zwischen TC3D-RNA-GF-Filter und MOE-RNA-GF-Filter stellt im Rahmen dieser Arbeit die größte Anzahl an gefundenen aktiven Verbindungen in den Tests. Insbesondere ist die Größe der Anreicherung durch Kombination der Auswahllisten der beiden Filter als beste Möglichkeit zu werten, um fokussierte Listen potentieller Liganden für RNA-Zielstrukturen zu erstellen. Dies beruht auf der zusätzlichen Untersuchung der Größe dieser kombinierten Hitlisten, wie sie in Tabelle 12 angegeben wurde. Hierbei erreicht diese Kombination der Auswahllisten, zwar nur eine relative Trefferquote von 42,9% für die Vorhersage aller möglichen gefundenen Liganden der beiden Assays (Tat-TAR-FRET-Assay und CFivTT-Assay), jedoch ist die Quote der gefundenen potentiellen RNA-Liganden für diese Kombination der Auswahllisten mit 50% der Einträge dieser Liste die höchste Trefferquote die im Rahmen dieser Arbeit erzielt werden konnte. Der MOE-RNA-GF-Filter erzielte die insgesamt höchste Trefferquote (74,1%) bei der Identifizierung der potentiellen RNA-Liganden dieser Arbeit, fällt jedoch mit seiner Anreicherungsquote innerhalb der Hitliste von 12,5% deutlich hinter der TC3D-MOE-Kombination zurück (ca. 4x). Der CATS-RNA-GF-Filter lieferte mit M23 den besten potentiellen RNA-Liganden für den Tat-TAR-FRET-Assay, der mit den anderen Filtern nicht identifiziert worden wäre. Somit lässt sich zusammenfassend sagen, dass die eingesetzten Filter jeweils einen Teilausschnitt möglicher potentieller RNA-Liganden abdecken. Die größere Überlappung der MOE- und TC3D-Filter ist angesichts der in Abschnitt 3.2.4 durchgeführten KS-Statistik, welche zur Reimplementierung und dem Einsatz im Rahmen der Arbeit des TC3D-Deskriptors führte, nicht weiter erstaunlich: Da die Verwendung des TC3D-Deskriptors basierend auf der MOE-2D-Deskriptor-KS-Analyse durchgeführt wurde, die eine Präferenz für ladungsbasierte Deskriptoren innerhalb der MOE-2D-Deskriptoren-Menge nahe legte, sollte ein nur auf den

Ladungsverteilungen innerhalb der Moleküle abgeleiteter mehrdimensionaler Deskriptor zu ähnlichen Ergebnissen führen, wie dies das Ensemble der verwendeten MOE-2D-Deskriptoren in anderer Form bereits beinhaltet.

3.2.19 Bekannte Fehler bei der Zusammenstellung des RNA-Liganden-Datensatzes

Hintergrund:

Im Rahmen der Literatursuche kam es leider zu einer Reihe von Fehlern bei der Zusammenstellung des Datensatzes, die jedoch nicht sofort bei der Überprüfung des Datensatzes gefunden wurden, so dass hier zum Teil „falsch positive“ als auch unbestätigte Strukturen mit in die Modellerstellung eingebracht wurden. Ferner wurden so insgesamt vier Strukturen übersehen, die für die Modellerstellung als „echt positiv“ hätten verwendet werden können.

Die Fehler beruhten auf der im Rahmen der Betreuung eines OCII/B Arbeitskreispraktikums erfolgten Auslagerung eines Teils der Literatursuche an die hierbei betreute Person. Diese nachvollziehbaren Fehler wurden jedoch alleinig durch die unpräzise Aufgabenformulierung im Zusammenhang mit der unzureichend erfolgten Hintergrundvermittlung seitens des Betreuers (dem Autor dieser Arbeit) verursacht. Letztendlich wurden diese Fehler jedoch zum Teil erst bei der Rückführung der aktiven Verbindungen aus den verwendeten Assays auf die zugrunde liegenden Strukturen des Trainingsdatensatzes aufgedeckt, welche eine Überprüfung der Quell-Veröffentlichungen umfasste.

Ergebnisse:

Es handelt sich hierbei um folgende drei Fehler:

1.) Ein Fluoreszeinderivat wurde aufgrund folgender fehlerhafter Anweisung in den Datensatz übernommen:

„Nimm die PDB-Strukturen und extrahiere das kleinste Molekül; überprüfe, ob dieses Molekül sich in Kontakt mit der RNA befindet. Wenn ja, trage dieses Molekül als Ligand ein.“

Dies führte dazu, dass ein Konstrukt aus RNA und einem Fluoreszeinderivat zerlegt und das Fluoreszeinderivat in den Datensatz übernommen wurde. Aufgrund der relativen Ähnlichkeit des Fluoreszeinderivats mit dem bekannten TAR-RNA-Binder Acetylpromazin und der damaligen Unkenntnis des Aussehens von Fluoreszein wurde dieser bedauerlicherweise „falsch positive“ Eintrag akzeptiert; siehe Abbildung 47.

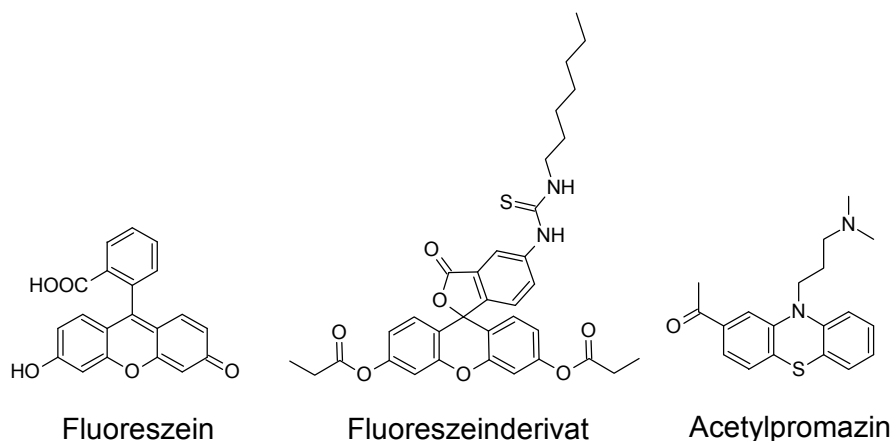


Abbildung 47: Fluoreszein, Fluoreszeinderivat und Acetylpromazin

2.) Der nächste entdeckte Fehler entstand beim Abzeichnen von zwei Ethidium-Arginin-Konjugaten, wo aus der Veröffentlichung (Patino et al. 2002) die Abbildung des Arginin durch ein Arg abgekürzt wurde; siehe Abbildung 48.

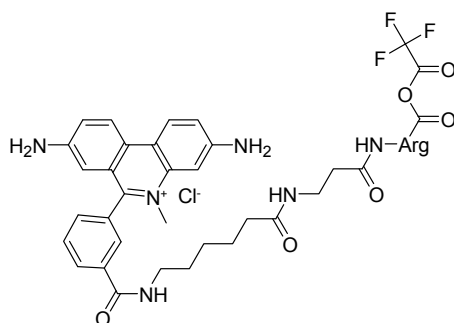


Abbildung 48: Ethidium-Arginin-Konjugat; Arginin ist als Arg abgekürzt

Dieser Fehler führte in der anschließenden Überprüfung, ob CORINA 3.20 (Molecular Networks GmbH, Erlangen) 3D-Konformere erzeugen kann, zum automatischen Entfernen der beiden Verbindungen aus dem Datensatz, da dieses Programm mit der Arg-Schreibweise nicht umgehen konnte. Jedoch wären diese Strukturen auch spätestens bei der Partialladungsberechnung mit PETRA 3.11 (Molecular Networks GmbH, Erlangen) entfernt worden.

Dies führte zum Verlust zweier RNA-Binder aus dem Datensatz, die sonst für das ANN-Training eingesetzt hätten werden können. Dieser Fehler basierte erneut auf einer falsch gestellten Anweisung, die besagte: „Zeichne die Strukturen aus den Veröffentlichungen ab, welche als RNA-Liganden klassifiziert wurden“. Es wurde nicht explizit darauf hingewiesen, dass Abkürzungen für Substrukturen in eine alle Atome umfassenden Form umgewandelt werden sollten.

3.) Der letzte bekannte Fehler bezieht sich auf die Strukturen aus der Veröffentlichung (Filikov et al. 2000), welche hauptsächlich Ergebnisse zweier „Virtueller Screening“-Runden enthält, von denen nur 1 aus 18 übernommenen Strukturen auch experimentell überprüft und bestätigt wurde. Zwei weitere experimentell bestätigte Strukturen für TAR wurden leider nicht aus der Veröffentlichung (Filikov et al. 2000) übernommen. Sie waren nur in einer kleinen Extratablelle für die experimentell bestätigten Strukturen abgebildet, zusammen mit der einzigen außerdem bestätigten Struktur, die jedoch zusätzlich in der anderen Tabelle abgebildet war.

Diskussion:

Die angeführten Fehler bei der Erstellung des RNA-Liganden-Datensatzes führten hauptsächlich zur Einführung von „Virtuellen Screening“-Ergebnissen in einen eigentlich auf rein experimentellen Daten basierenden Datensatz und dem Verlust von vier als „echt positiv“ zu betrachtenden Strukturen. Da es sich bei den genannten Einträgen zwar um auf weiteren experimentellen Daten beruhende Ableitung von Aktivität handelt, werden diese aber wiederum – aufgrund der reinen Vorhersagenatur – einen zum jetzigen Zeitpunkt unüberprüfbar Fehler beinhalten. Somit ist die Fehlerabschätzung der Güte des resultierenden Gesamtdatensatzes, solange keine experimentelle Bestätigung der genannten nur vorhergesagten Strukturen vorliegt, nicht genau präzisierbar. Jedoch wurden diese Strukturen basierend auf den Ergebnissen von Docking-Algorithmen ausgewählt, dessen Bewertungsfunktionen nicht auf DNA oder RNA speziell angepasst wurden, so dass hierbei eine Unterrepräsentation der „van der Waals“-Interaktion zu Gunsten der elektrostatischen Wechselwirkung anzunehmen ist. Da dies für alle aus dieser Veröffentlichung genannten Strukturen gilt, handelt es sich hierbei möglicherweise um einen

systematisch auftretenden Fehler des Subdatensatzes, der – falls dies ein Unterschied zu den anderen Strukturen des RNA-Liganden-Datensatzes darstellt – sich durch die Repräsentation eines deutlich unterschiedlichen Bereichs des chemischen Raums zeigen sollte; siehe hierzu Abschnitt 3.2.19.1.

Die Tatsache, dass die elektrostatischen Wechselwirkungspartner (Substrukturelemente der Zielstruktur als auch der potentiellen Liganden) sich bei der Bewertung in relativer räumlicher Nähe befinden sollten (Docking-Studie), impliziert zwar das Vorhandensein von „Spacer“-Gruppen, die diese Interaktionspunkte in ihren räumlichen Positionen halten können, jedoch müssen diese nicht zwangsläufig einen hydrophoben Charakter aufweisen, geschweige denn zu „Base-Stacking“-Interaktionen in der Lage sein. Die ausgewählten Substanzen tragen diese Bevorzugung bestimmter elektrostatischer Wechselwirkung implizit in sich, wie in folgender Aussage aus der oben genannten Veröffentlichung (Filikov et al. 2000) ersichtlich ist: „...shows that many compounds predicted to be good ligands for TAR RNA are likely to be false positives as drug candidates because of their chemical nature: too many net charges or too polar to be orally active.“ Diesen Sachverhalt kann auch noch bei der Auswertung des in Rahmen dieser Arbeit durchgeführten Tat-TAR-FRET-Assays, sowie des von Markus Weidlich im Rahmen seiner Doktorarbeit durchgeführten zellbasierten Assays beobachtet werden; siehe Anhang 7.7 auf Seite 213.

Um die Ergebnisse dieser Tests zusammenfassend vorwegzunehmen, sei hier gesagt, dass die mittels der durch die Tat-TAR-Assay-spezifischen Filter ausgewählten Strukturen ebenfalls diese Eigenschaftsverteilung aufweisen. Als wichtigstes Argument für eine Relativierung dieses Fehlers sei erwähnt, dass auch mit diesem fehlerhaften Datensatz Substanzen gefunden wurden, die eine Aktivität in den RNA-Assays zeigten und deren Struktur sich auf Teile dieses Subdatensatzes zurückführen lässt. Möglich ist jedoch, dass hierdurch die allgemeine Erfolgsrate reduziert wurde. Dies ist aber zum jetzigen Zeitpunkt reine Spekulation und müsste durch eine erneute Versuchsrunde gezeigt werden. Als ein Versuch der Abschätzung der Größe des möglicherweise eingeführten Fehlers wurden die in Abschnitt 3.2.19.1 und 3.2.19.2 durchgeführten Analysen vorgenommen.

3.2.19.1 Versuch der Abschätzung des Fehlers mittels KS-Statistiken für die Teilmengen

Hintergrund:

Die in Abschnitt 3.2.19 dargelegten Fehler bei der Erstellung des RNA-Liganden-Datensatzes führen dazu, dass die Frage offen bleibt, inwieweit die Verwendung der Vorhersagedaten der Filikov-Veröffentlichung (Filikov et al. 2000) dazu geführt hat, die Vorhersageergebnisse der ANN-Filter zu beeinflussen, bzw. ob die Strukturen aus dieser Veröffentlichung die Grundlage für die Auswahl der als aktiv befundenen Substanzen der M-Reihe gestellt haben können?

Vorwegnehmend sei gesagt, dass die in den angewendeten Labor-Assays als aktiv befundene Substanz M32 das Adamantyl-Substrukturmotiv mit der Filikov-Struktur 00151463 teilt, welche zwar nicht für die TAR, so aber für den L-Arginin RNA Aptamer vorgeschlagen wurde; siehe hierzu den Abschnitt 3.2.17.1 und Abbildung 61 auf Seite 93.

Eine kleine Abschätzung des durch diesen Datensatz möglicherweise eingeführten Fehlers ist aus der Anwendung der Kolmogorov-Smirnov-Statistik (KS-Statistik, siehe Abschnitt 2.3) auf die Teilmengen überprüfter Substanzen gegen die unüberprüften Strukturen aus der Veröffentlichung (Filikov et al. 2000) möglich. Die KS-Statistik sollte hierbei als Maß, für die Ähnlichkeit der Datenwerte der jeweiligen Deskriptoren der jeweils verwendeten Datensätze zueinander, eingesetzt werden können, welches als Kriterium für die Bewertung der

Ähnlichkeit der den Datensätzen zugrunde liegenden Substanzen als jeweils einzelner numerischen Wert anzunähern und somit zu bewerten ist.

Der Anteil der unüberprüften Substanzen am gesamten „RNA-Liganden-Datensatz“ (alle potentiellen RNA-Liganden laut Literatursuche, siehe Abschnitt 3.2.1) beträgt ca. 17%. Somit sollte sich durch Anwendung der KS-Statistik überprüfen lassen, wie divers unterschiedlich dieser Teil des Datensatzes vom Rest des RNA-Datensatzes ist. Ebenso gilt dies dann bei Anwendung der KS-Statistik für die Unterschiede der Teilmengen der RNA-Liganden-Datensätze zum Protein-Liganden-Datensatz.

Arbeitshypothese:

Sollten signifikante Unterschiede zwischen den Teilmengen des RNA-Liganden-Datensatzes auftreten und es ferner für die Teilmenge der unüberprüften Substanzen eine höhere KS-Distanz zum Protein-Liganden-Datensatz geben als dies bei der Verwendung des kompletten RNA-Liganden-Datensatzes der Fall wäre, so kann davon ausgegangen werden, dass die ANN-Filter sich primär auf diesen Unterschied konzentrieren würden. Sollte dies der Fall sein, so ist die Vorhersagegüte der Filter kompromittiert, d.h. man weiß nicht genau, was das System vorherzusagen gelernt hat. Vielmehr wird den Filtern nur die Auswahl von Substanzen zugestanden, die nicht wie bekannte Protein-Liganden aussehen, ohne zu wissen, was für Aktivitäten in den verbleibenden Substanzen hauptsächlich kodiert sind; siehe hierzu die Ergebnisse der Auftrenngüte der RNA-Filter für die verwendeten Trainingsdatensätze in Abschnitt 3.2.7.

Durchführung:

Eine Betrachtung der maximal erreichbaren KS-Distanzen in der Kombination der folgenden Teilmengen des Trainingsdatensatzes wurde durchgeführt:

- COBRA (5651 Strukturen)
- Gesamter verwendeter RNA-Liganden-Datensatz (aRNALD; 98 Strukturen)
- Anteil experimentell bestätigter Strukturen des RNA-Liganden-Datensatzes (bRNALD; 81 Strukturen)
- Anteil der unbestätigten Strukturen des RNA-Liganden-Datensatzes (uRNALD; 17 Strukturen)

Verwendet wurde hierzu der MOE-2D-Deskriptordatensatz.

Ergebnisse:

Aus den beschriebenen Berechnungen ergaben sich folgende erreichte maximale KS-Distanzen innerhalb der Datensatzkombination für jeweils alle MOE-2D-Deskriptoren:

- a) aRNALD gegen COBRA → 0,67
- b) bRNALD gegen COBRA → 0,75
- c) uRNALD gegen COBRA → 0,68
- d) bRNALD gegen uRNALD → 0,79

Diskussion:

Eine Vermischung der RNA-Datensätze (bRNALD und uRNALD) führte zu einer Verschlechterung der Auftrennbarkeit des resultierenden gesamten RNA-Liganden-Datensatzes (aRNALD) von der COBRA. Dies wird aus folgenden Betrachtungen ersichtlich: Die maximal erreichbare KS-Distanz zwischen aRNALD und COBRA mit 0,67 ist kleiner als die erreichte maximale KS-Distanz von bRNALD gegen COBRA mit 0,75. Selbst die maximale KS-Distanz zwischen uRNALD und COBRA von 0,68 ist, wenn auch nur geringfügig, größer als die maximale KS-Distanz beim Vergleich von aRNALD und COBRA. Ferner ist die beobachtete maximale KS-Distanz von bRNALD gegen uRNALD mit 0,79 die größte, in den hier durchgeführten Vergleichen gefundene KS-Distanz.

Dies legt nahe, dass die Einführung der Teilmenge der unbestätigten Strukturen des RNA-Liganden-Datensatzes (uRNALD) durch die bewirkte veränderte Verteilung im chemischen Raum eine schlechtere Trennung der Klassen erlaubt. Dies wird weiter durch den KS-Distanz-Wert der Kombination „d“ deutlich, deren KS-Distanz das hier gefundene Maximum darstellt und die größte Differenz in den als Deskriptoren kodierten chemischen Strukturen der Teilmengen der Datensatzkombinationen offen legt. Ferner zeigt sich, dass aufgrund des KS-Distanz-Wertes der Kombination „b“ eine stärkere Trennung der potentiellen RNA-Binder von den Protein-Bindern möglich gewesen wäre⁵, als dies nach Einführung der uRNALD-Teilmenge noch der Fall war (Kombination „a“).

Abschlussüberlegungen:

Da die Betrachtung mittels KS-Statistik eine deutliche Reduktion des chemischen Raums auf einzelne Merkmale bedeutet und die verwendeten ANNs auf eine Kombination der verschiedenen Deskriptoren hin trainiert wurden, ist diese Auswertung nur bedingt auf das Lernverhalten der ANNs übertragbar. Deshalb wurde im nächsten Schritt eine Analyse des chemischen Raums mittels SOM-Analyse, welche die Nachbarschaftsbeziehungen innerhalb des untersuchten Raums auf eine 2D-Darstellung reduziert durchgeführt. Die relativen Verteilungen der Datensätze auf den Karten sollten ein besseres Verständnis des Einflusses des uRNALD auf die Bewertungsfunktionen der ANNs ermöglichen.

3.2.19.2 Untersuchung der Verteilung der Filikov-Daten im chemischen Raum mittels SOMs und Analyse der Auswirkungen auf die Selektion der als im Labor als aktiv befundenen Substanzen**Hintergrund:**

Das Problem bei der aus der Filikov-Veröffentlichung (Filikov *et al.* 2000) entnommenen Teilmenge des verwendeten Gesamtdatensatzes für die potentiellen RNA-Liganden liegt in ihrem nicht definierten Anteil an falsch positiven Trainingsbeispielen für das ANN-Training. Die Einführung der als RNA-Liganden vorhergesagten Strukturen nach Filikov (Filikov *et al.* 2000) in die verwendeten Trainingsdaten der ANNs könnte zu einer Beeinflussung der Vorhersagen der ANNs geführt haben, siehe hierzu auch Abschnitt 3.2.19.1. Abgeleitet aus

⁵ Die Auswirkungen der Einführung der Filikov-Strukturen aus der Veröffentlichung (Filikov *et al.* 2000) sind laut KS-Statistik-Analyse zur Trennbarkeit der RNA-Liganden-Datensätze vom COBRA-Datensatz bedauerlich, da dies aber erst beim erneuten Überprüfen und Zusammenstellen für diese Arbeit entdeckt wurde, konnte diese Erkenntnis nicht mehr in eine neue Filtergeneration und letztlich eine neue Bestellrunde für die Überprüfung im Labor umgesetzt werden.

den Ergebnissen der KS-Statistiken aus Abschnitt 3.2.19.1 blieb die Frage offen, inwieweit diese Substanzen jedoch an der Auswahl der Testsubstanzen beteiligt sein könnten.

Das Hauptaugenmerk der Untersuchung richtete sich abschließend jedoch vor allem auf den CATS-Deskriptorraum, da mit diesem Filter die besagten – anhand ihrer markanten Substruktur (Tetra-aza-adamantyl-) offensichtlich aus dem uRNALD abstammenden – aktiven Substanzen M19 und M32 nicht als potentielle RNA-Liganden vorhergesagt wurden. Dahingegen konnte anhand der Hitlisten der TC3D- und MOE-Filter überlappend/bestätigend diese beiden Liganden M19 und M32 bereits ihren Hitlisten zugewiesen werden; vgl. Abschnitt 3.2.18. Somit ist davon auszugehen, dass nur der CATS-Filter den Einfluss der uRNALD-Substanzen beim Training ignoriert hat. Ob dieser Sachverhalt ein Zufall des Trainingsvorganges war oder am durch den CATS-Deskriptor kodierten chemischen Raum lag, sollte hier noch geklärt werden.

Die Untersuchung wurde mittels der Anwendung von SOMs (Erklärung erfolgte in Abschnitt „Selbstorganisierende Karten (SOM)“ auf Seite 15) durchgeführt, die in diesem Fall eine Projektion des chemischen Raums auf eine toroidale zweidimensionale Oberflächenrepräsentation zwecks Visualisierung erlaubt. Diese Karte bewahrt die räumliche Nähe im chemischen Raum zum Teil auch in der Projektion und ist somit tendenziell geeignet um Nachbarschaftsverhältnisse des chemischen Raums zu untersuchen (Schneider und Wrede 1998; Schmuker et al. 2007).

Anmerkungen zur Verwendung von SOMs für diese Analyse:

Da die jeweilige resultierende Projektion des hochdimensionalen chemischen Raums auf eine Karte aber nicht zwangsläufig benachbarte Strukturen des chemischen Raum als benachbart auf der Karte ergeben muss, müsste für die eingehende Untersuchung die folgende Analyse für mehrere Trainingsvarianten der SOMs durchgeführt werden, um eine statistische Signifikanz abzuleiten. Dies wurde jedoch – da hierzu im Rahmen der Arbeit keine automatisierten Methoden vorlagen – nicht mehr durchgeführt. Für zukünftige Arbeiten wäre es somit besser die Analyse anhand der relativen Nähe direkt aus dem hochdimensionalen chemischen Raum abzuleiten ohne die möglichen Fehler bei der Reduktion der Dimensionen zu erhalten. Hierbei könnten die sortierten Nachbarschaftslisten relevanter Moleküle im hochdimensionalen Deskriptorraum als Kriterium für die mögliche Einflussnahme auf die Vorhersage der ANNs dienen.

Angewendete Arbeitshypothese:

Als Arbeitshypothese für die Überprüfung des möglichen Einflusses des uRNALD wird hierbei angenommen, dass die fraglichen Filikov-Substanzen, sollten sie einen Einfluss bei der Auswahl der Labor-Substanzen durch die ANNs gehabt haben, sie sich in unmittelbarer räumlicher Nähe zu den Knoten der Karte befinden sollen, welche die Labor-Substanzen auf eben dieser Karte beinhalten. Sollten jedoch bestätigte Literatur-Substanzen näher an den aktiven Laborsubstanzen liegen, so wird angenommen, dass die vorher verwendeten ANNs ebendiese als Selektionskriterium bei der Bewertung verwendet haben werden.

Des Weiteren wurde mit dieser Methode der Projektionsanalyse ebenfalls überprüft, inwieweit sich die bestätigten RNA-Liganden aus der Literatur von den unbestätigten Strukturen der Filikov-Veröffentlichung (Filikov *et al.* 2000) überhaupt in der Kartenprojektion unterscheiden, wie dies durch die KS-Statistik aus Abschnitt 3.2.19.1 nahe gelegt wurde.

Durchführung:

Sämtliche Untersuchungen wurden an allen drei verwendeten Deskriptorräumen durchgeführt (MOE-2D-Deskriptoren, CATS, TC3D).

Hierbei wurde wie folgt vorgegangen:

1. **Strukturdatensatzerstellung**
2. **Strukturdatensatzvorbereitung / Konformergeneration**
3. **Deskriptorberechnung**
4. **Training der SOMs**
5. **Projektion der Einzeldatensätze auf die SOMs**
6. **Analyse der Überlappungen der Datensätze innerhalb der SOMs**
7. **Analyse der Überlappungspunkte der Laborsubstanzen in den SOMs um die Strukturen des Filikov-Datensatzes zu extrahieren, die wahrscheinlich für ihre Auswahl verantwortlich waren.**

1. Datensatzerstellung

Zur Untersuchung des Sachverhalts wurden neue Datensatzzusammenstellungen aus den ursprünglich verwendeten Dateien erneut abgeleitet. Dies erfolgte um diese Datensätze mittels der Merge-Funktion in MOE auf ihrer Übereinstimmung mit den Datensätzen aus der KS-Statistik-Analyse des Abschnitts 3.2.19.1 zu überprüfen und diente der Eliminierung vermeintlich erneut aufgetretener Fehler bei der Extraktion. Ferner mussten hierfür noch die zusätzlichen Datensätze aus der M-Reihe abgeleitet werden.

Es erfolgte die Überprüfung des COBRA-Datensatzes auf korrekte Extraktion des Protein-Liganden-Datensatzes. So wurde auch ein weiteres Mal überprüft, ob sich RNA-Binder-Klassifizierungen unter den vermeintlichen Proteinliganden befinden.

Der RNA-Liganden-Datensatz (aRNALD) wurde erneut aus der Datenbank extrahiert und ebenfalls überprüft.

Aus dem so gewonnenen Datensatz wurde der Filikov-Teildatensatz extrahiert, welcher nur ungeprüfte Vorhersagen enthielt (uRNALD) und der Rest des Literaturdatensatzes, der noch eine bestätigte Filikov-Substanz (bRNALD) enthielt.

Es wurde ferner der Datensatz der von SPECS gelieferten Substanzen zusammengestellt.

Aus diesem Datensatz wurde jeweils ein Teildatensatz für die als aktiv befundenen Substanzen des FRET- und CFivTT Assays extrahiert.

2. Strukturdatensatzvorbereitung / Konformergeneration

Die im Absatz 1 („Datensatzerstellung“) erstellten Datensätze wurden für die Deskriptorberechnung alle mit dem „db_prepare“-Shellskript (Version: 1.6) von Andreas Schüller, das die Datensatz-Bearbeitung für Virtual-Screenings standardisiert, durchgeführt.

Dieses Skript standardisiert die Verwendung verschiedener Tools aus dem Arbeitskreis zum automatischen Vorbereiten der SDFs für die spätere Nutzung. Es wird dadurch sichergestellt, dass etwaige Gegen-Ionen – welche die Deskriptorberechnung verfälschen könnten – entfernt wurden, als auch das Anfügen von Wasserstoffen an alle Strukturen nach demselben Programmstandard gewährleistet ist. Weiterhin werden durch dieses Skript ebenfalls die 3D-Konformere erzeugt, wie dies wahlweise im ChargeCyclone3D-Webinterface zur TC3D-Deskriptorberechnung durchgeführt werden kann. Die Verwendung dieses Skripts ermöglicht somit wiederum eine potentielle Fehlerreduktion gegenüber dem vorher angewendeten manuellen Aufruf dieser Programme.

Verwendet wurden in diesem Teil der Studie ausschließlich die so generierten 3D-Konformer-SDF-Dateien.

3. Deskriptorberechnung

CATS-Deskriptor-Berechnung

Für die Berechnung der CATS Deskriptoren wurde auf Ulrich Fechners jSpeedcats (Version 0.32) (Fechner et al. 2003) zurückgegriffen, welches mit den Standardparametern aufgerufen wurde, wie sie auch für die ANN-Filter von 2005 verwendet wurden. Hierbei reduziert sich die erforderliche Parameterübergabe auf:

- -i Eingabedateiname (SDF-File).
- -o Ausgabedateiname (Deskriptortextdatei).
- -s „suppress header“ = keine Kopfzeile.

Eine genauere Definition des CATS-Deskriptors erfolgte in Abschnitt 2.1.2.

Exemplarisch ist hier der Originalaufruf von 2005 aufgeführt:

```
jspeedcats -i RNA_COBRA_MOE_gefiltert_unikate_01maerz05.sdf ...  
-o RNA_COBRA_MOE_gefiltert_unikate_01maerz05.txt -s
```

Die Berechnung wurde analog für alle oben genannten Datensätze durchgeführt.

TC3D-Deskriptor-Berechnung

Für die Berechnung der TC3D-Deskriptoren wurde wieder auf das im Rahmen dieser Arbeit entstandene ChargeCyclone3D v2.2 zurückgegriffen. Insbesondere wurde hierbei auf das Webinterface auf dem AK-internen Webserver zugegriffen, da auch dieses wieder die mögliche Fehlbedienung deutlich reduziert (aufgrund der automatisierten Teilschritte und Eingabewerteüberprüfung). Die Berechnung erfolgte analog zu 2005 mit jeweils 100 Bins für jede der Klassen (++/+-/--) und einem maximalen zu beachtenden Abstand von 15 Ångström. Auf die Generierung der benötigten 3D-Konformere wurde diesmal jedoch verzichtet, da diese im Rahmen der Strukturdatensatzvorbereitung bereits erfolgte.

MOE-Deskriptor-Berechnung

Für die Berechnung der 2D-MOE-Deskriptoren wurde MOE (Version 2006.08) eingesetzt.

Hierzu wurden folgende Schritte durchgeführt:

1. Die vorher erstellten 3D-SDF-Dateien wurden in MOE-Datenbanken importiert,
2. Mit den Standard-Einstellungen von MOE wurden Partialladungen berechnet und den Struktureinträgen hinzugefügt (verwendetes Kraftfeld = MMFF94x (Halgren 1996)).
3. Basierend auf den Struktureinträgen wurden alle 2D-Deskriptoren ausgewählt und berechnet.
4. Die so gewonnenen Deskriptoren wurden, zusammen mit den SMILES der Strukturen, als Textdateien exportiert.

Anmerkung:

Alle Operationen mit dem Programm wurden hier mittels Programmoberfläche durchgeführt. Hierzu gibt es jedoch auch ein SQL-Skript, welches diese Aufgabe ebenfalls erledigen kann, verwendet wird dieses jedoch nur durch die MOE-basierten Filter auf dem internen Webserver. Bei Benutzung der Programmoberfläche sind die 2D-Deskriptoren zwar deterministisch jedoch nicht alphabetisch sortiert; in der SQL-Skript-Variante werden sie jedoch erst einmal alphabetisch sortiert erstellt und dann entsprechend der Referenzliste umsortiert. Da sich die mögliche Referenzliste durch Weiterentwicklung von MOE jedoch erweitert hat und von Hand geändert hätte werden müssen, wurde aus Zeitgründen für die hier beschriebenen Versuche die Version der konsequenten Oberflächenbenutzung verfolgt.

4. Training der SOMs

Die verwendeten SOMs wurden mittels des im Rahmen dieser Arbeit entstandenen Webinterface „Gisos SOMs“ auf dem AK-internen Webserver erstellt. Hinter der Oberfläche verbirgt sich das Programm MOLMAP® von Prof. Dr. Gisbert Schneider.

Alle drei hier weiter behandelten SOMs wurden mit dem jeweiligen kombinierten Datensatz aus Protein-Liganden und allen Literatur-RNA-Liganden trainiert. Die Strukturen der SPECS-Bestellungen wurden bewusst noch nicht verwendet, um zu erzwingen, dass sie später bei der Projektion der Daten auch einem Knoten des obigen kombinierten Datensatzes zugewiesen würden.

Es wurden folgende Einstellungen für das Training verwendet:

TC3D

- distance metric: -euklid
- -xNeurons 20
- -yNeurons 20
- -cycles 120000
- -neuronDim 300
- -radius 10
- -tau 1.0

MOE-2D

- distance metric: -euklid
- -xNeurons 20
- -yNeurons 20
- -cycles 120000
- -neuronDim 184
- -radius 10
- -tau 1.0

CATS

- distance metric: -euklid
- -xNeurons 20
- -yNeurons 20
- -cycles 600000
- -neuronDim 150
- -radius 10
- -tau 1.0

Jedem Trainingslauf wurden seine jeweilige Deskriptordatei (-patternFile), sowie ein Name für die zu erstellende SOM-Datei (-molmapFile) zugewiesen.

Anmerkung zu CATS

Die deutlich erhöhte Anzahl an Trainingszyklen wurde deshalb gewählt, weil die Untersuchung des „Quantization Error“ der resultierenden Karte auf ein unzureichendes Training bei kleiner gewählter Anzahl hindeutete.

5. Projektion der Einzeldatensätze auf die SOMs

Die Projektion der Datensätze auf die SOM wurde ebenfalls mittels besagten Webinterfaces durchgeführt. Hierbei wurden die Standardeinstellungen verwendet und die Option „overall frequency“ ausgewählt. Diese resultiert in Karten, in denen die relative Populationsmenge eines Knotens als Temperatur-Skalierung angegeben ist; siehe hierzu den relativen Werteverlauf des SOM-Farbbalkens in Abbildung.



Abbildung 49: Farbbalken der SOMs; die Werte für den jeweiligen Farbverlauf innerhalb einer SOM, beziehen sich immer auf den relativen Min- und Max-Wert, der in der jeweiligen SOM auftritt.

Anmerkung zur Kartenbetrachtung:

Eine direkte Ableitung von Zahlwerten aus den Farben zweier SOMs ist nicht möglich, da diese jeweils eine relativ zur Häufigkeit innerhalb des jeweiligen projizierten Datensatzes berechnete Farbe darstellt und somit keinen direkten Bezug zu anderen Projektionen aufweist und keine Eigenschaft der Karte an sich ist. In der hier durchgeführten Studie wird somit nur eine qualitative Verteilungsanalyse der Daten durchgeführt. Sollte der prozentuale Anteil der Strukturen in einem Knoten von Interesse sein, wird über den zu diesem Knoten gehörenden Hyperlink die Anzahl der Strukturen dieser Projektion in dem Knoten ermittelt und dann mit der Gesamtzahl der in dieser jeweiligen Projektion verwendeten Strukturen in Relation gesetzt.

6. Analyse der Überlappungen der Datensätze innerhalb der SOMs

Zur vergleichenden Analyse der einzelnen Kartenprojektionen der jeweiligen Deskriptorräume wurden von den Kartenbereichen der Ausgabedateien

Bildschirmaufnahmen erstellt und diese in Powerpoint zurechtgeschnitten und nebeneinander platziert.

Die erste Maßnahme war die Überprüfung der Projektionen der Datensätze aller Literatur-RNA-Liganden (aRNALD) und des Protein-Liganden (COBRA) mit anschließender Aufteilung des Literatur-RNA-Liganden-Datensatzes in die beiden Datensätze „RNA-Liganden-Datensatz ohne unbestätigte Filikov“ (bRNALD) und „Filikov Subset mit unbestätigter Aktivität“ (uRNALD). Diese wird in den Abbildungen 50 - 52 gezeigt.

Ergebnisse:

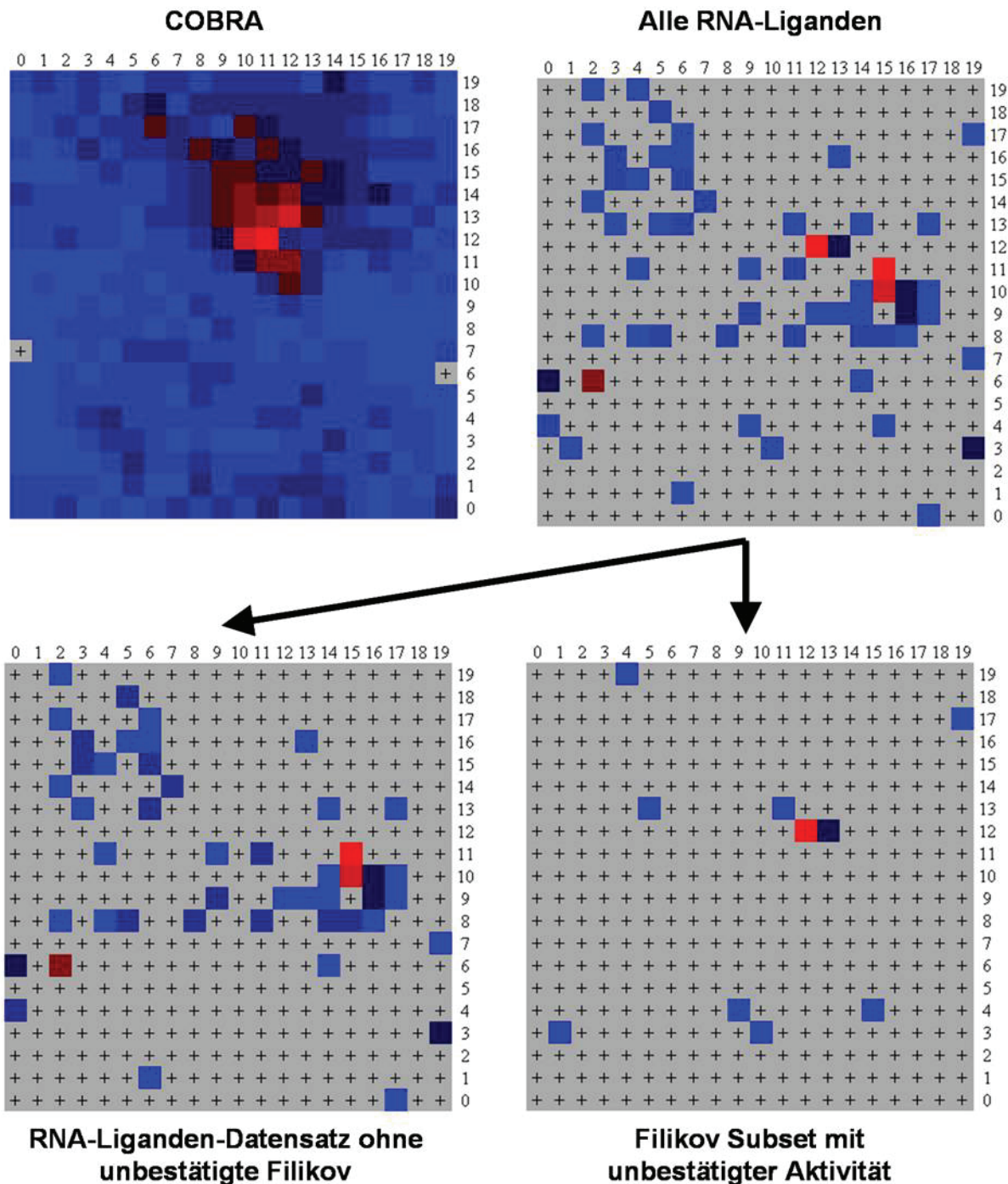


Abbildung 50: MOE-Deskriptor SOM: Projektionen des COBRA-Datensatzes und des gesamten Literatur-RNA-Liganden Datensatzes; Aufteilung der Projektion des RNA-Liganden-Datensatz in die Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen

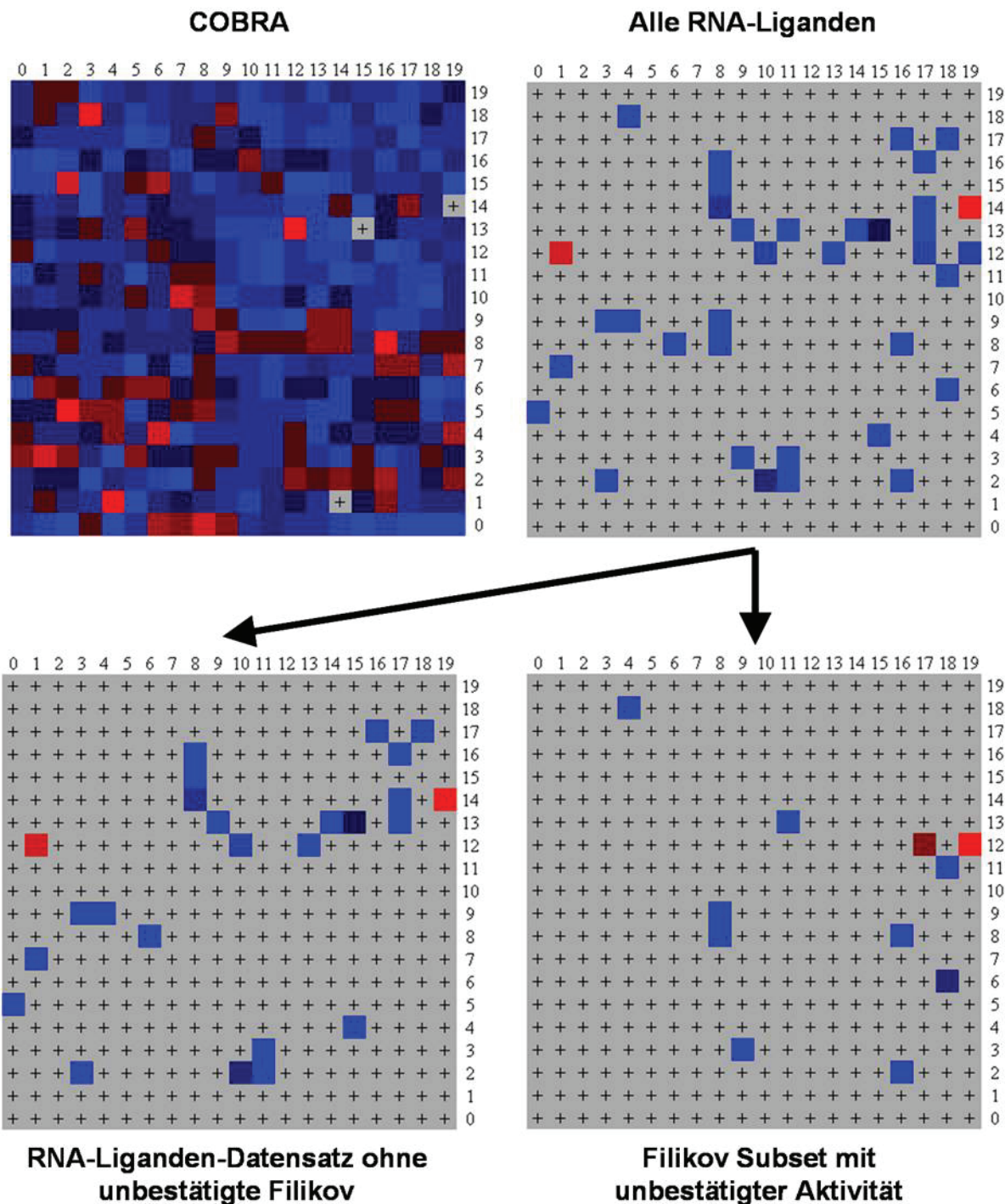


Abbildung 51: CATS-Deskriptor SOM: Projektionen des COBRA-Datensatzes und des gesamten Literatur-RNA-Liganden Datensatzes; Aufteilung der Projektion des RNA-Liganden-Datensatz in die Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen.

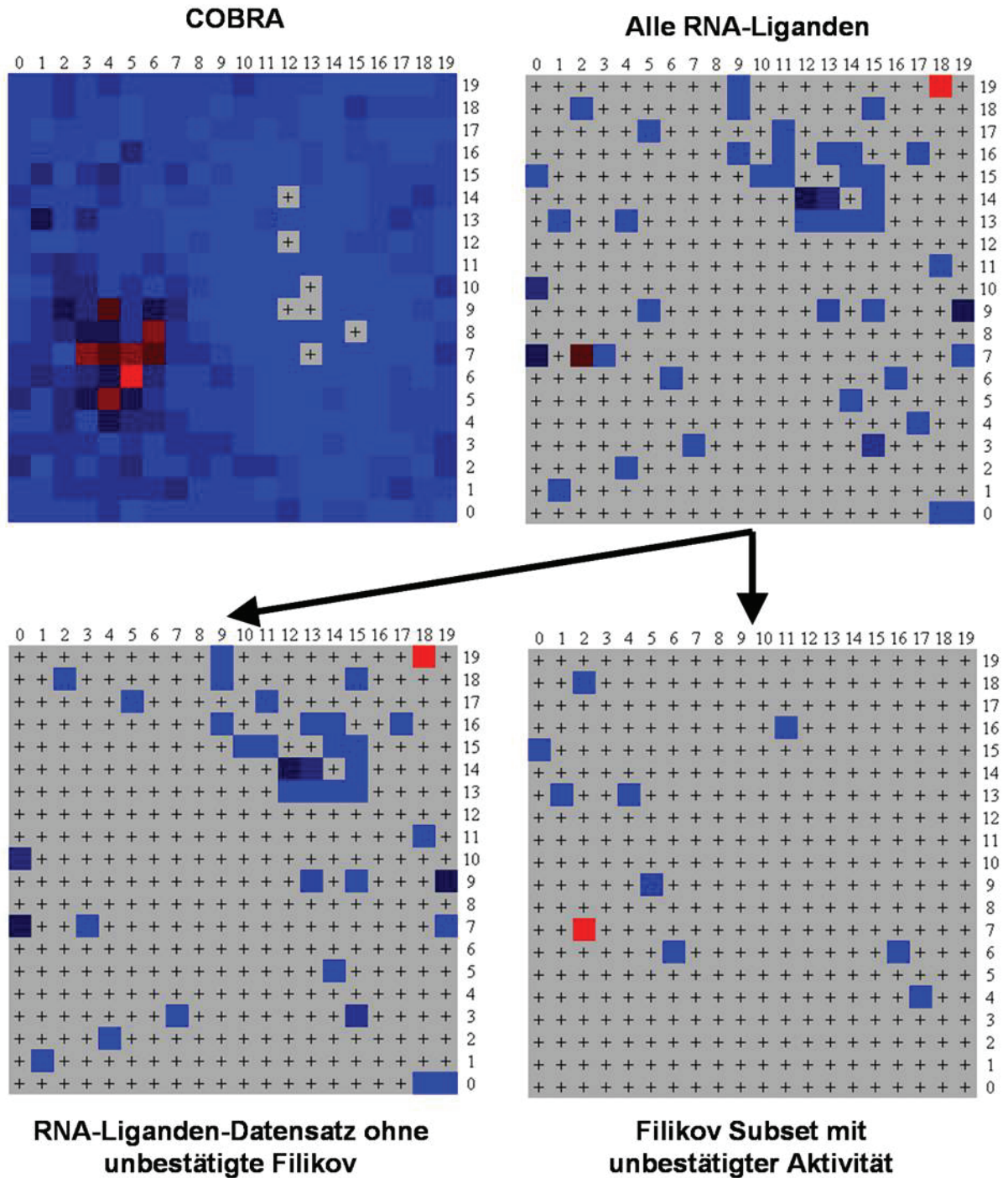


Abbildung 52: TC3D-Deskriptor SOM: Projektionen des COBRA-Datensatzes und des gesamten Literatur-RNA-Liganden Datensatzes; Aufteilung der Projektion des RNA-Liganden-Datensatz in die Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen.

Diskussion:

Zusammenfassend legt die Streuung der jeweiligen Literatur-RNA-Liganden-Datensatz Projektionen auf den Karten eine Diversität des zugrunde liegenden chemischen Raums nahe. Wenn man jedoch die Projektion der RNA-Liganden-Teilmengen darunter betrachtet, ist die geringere Diversität innerhalb des uRNALD erkenntlich, und die Positionen der belegten Knoten für diese beiden Teilmengen (uRNALD und bRNALD) scheinen sich voneinander stark zu unterscheiden. Um diesen Sachverhalt genauer zu untersuchen

wurden die Abbildungen 53 bis 55 erstellt, welche die beiden Subsets wieder nebeneinander zeigen, in denen aber dieses Mal zusätzlich die Überlappungspunkte mittels weißer Quadrate markiert wurden. Dies zeigte, dass für den CATS- wie den MOE-2D-Deskriptorraum keine direkten Überlappungen vorhanden sind und es bei dem TC3D-Deskriptorraum lediglich einen Knoten gemeinsam zwischen diesen Teildatensätzen gibt. Somit ist davon auszugehen, dass die unbestätigten Filikovstrukturen deutlich andere Bereiche im chemischen Raum bevölkern, als dies die restlichen bestätigten Literatur-RNA-Liganden tun. Dies bestätigt auch die Annahmen aus der KS-Statistik-Analyse aus dem Vorfeld dieses Teils der Studie in Abschnitt 3.2.19.1.

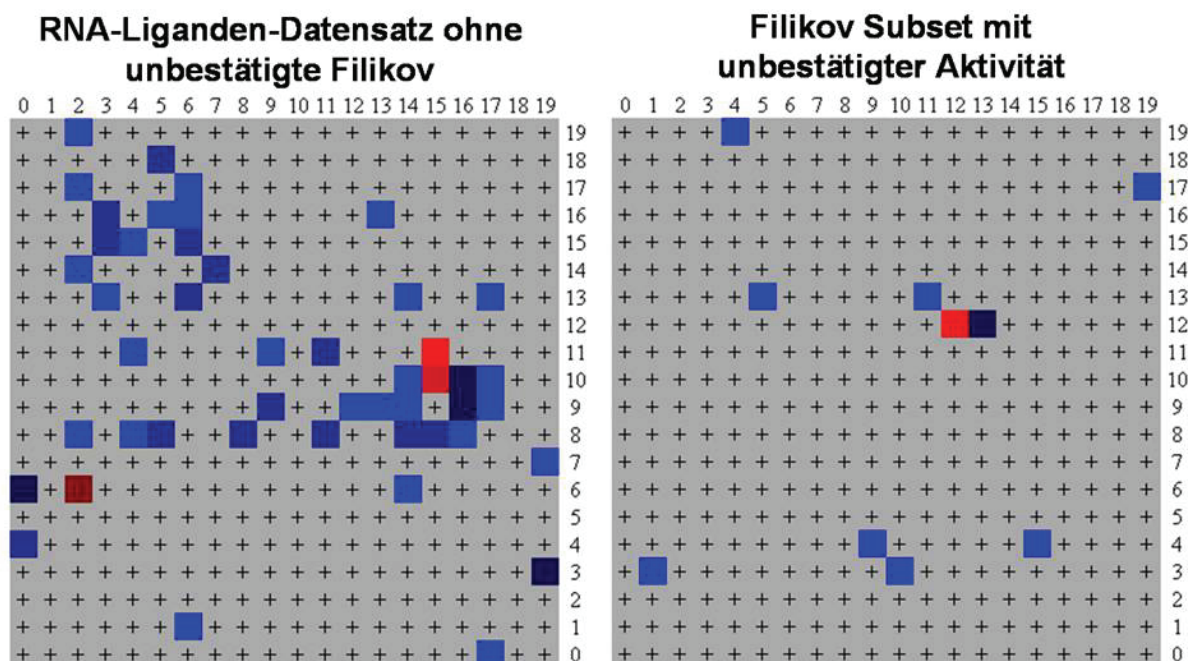


Abbildung 53: MOE-2D-Deskriptor SOM: Projektion der RNA-Liganden-Datensätze der Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt keine Übereinstimmungen zwischen den belegten Knoten der SOM-Projektionen.

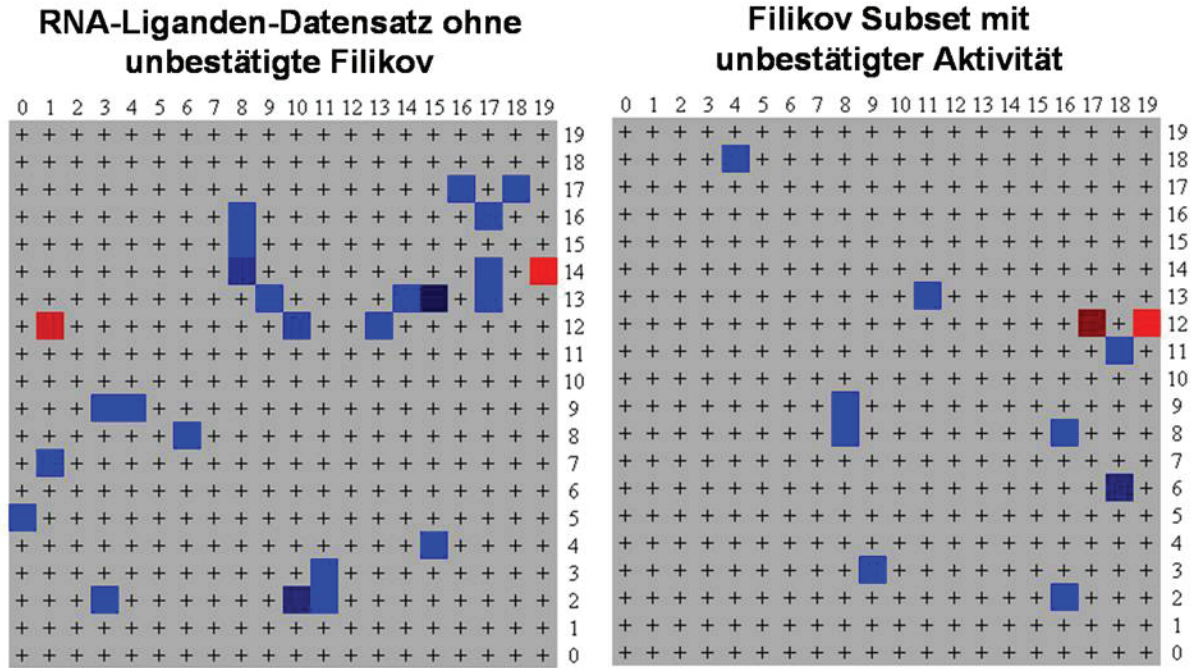


Abbildung 54: CATS-Deskriptor SOM: Projektion der RNA-Liganden-Datensätze der Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt keine Übereinstimmungen zwischen den belegten Knoten den belegten Knoten der SOM-Projektionen.

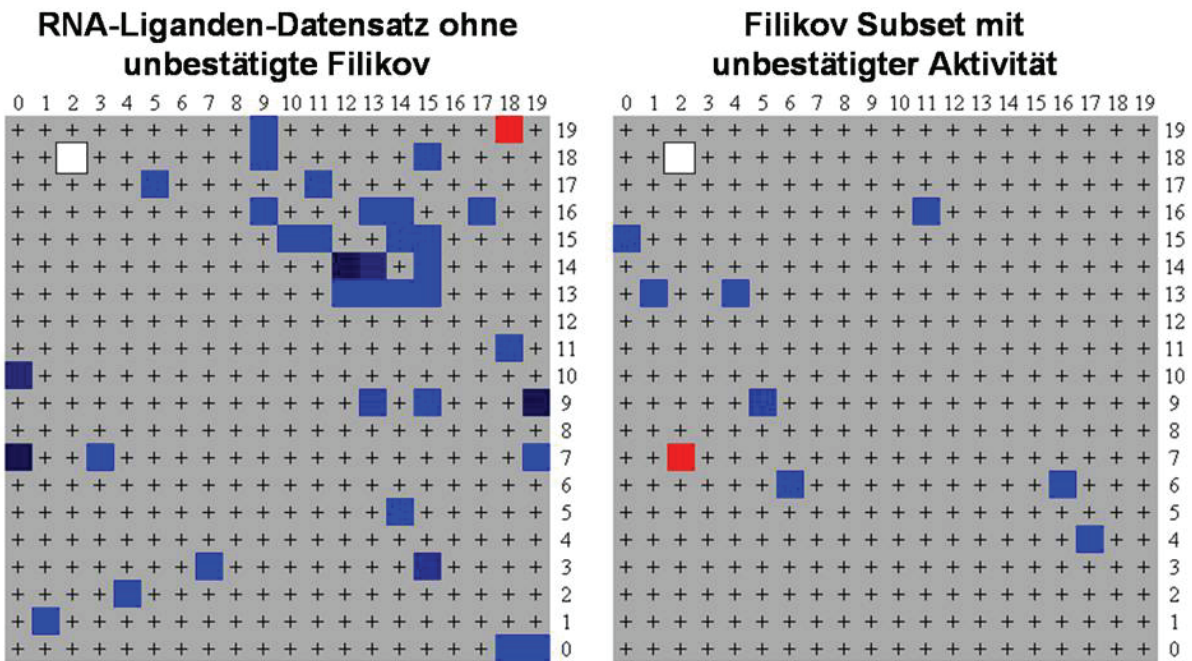


Abbildung 55: TC3D-Deskriptor SOM: Projektion der RNA-Liganden-Datensätze der Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt eine Übereinstimmung zwischen den belegten Knoten der SOM-Projektionen.

Diskussion:

Eine Schlussfolgerung aus dem Ergebnis der Überlappungsanalysen der bRNALD- und uRNALD-Projektionen lautet, dass die Filikovsubstanzen – da sie einen anderen Bereich des chemischen Raums abdecken – durch ihre Anzahl von immerhin 17 % im Gesamtdatensatz durchaus in der Lage gewesen sein können, die Auswahl der Testsubstanzen in diesen speziellen Bereich des chemischen Raums zu schieben.

Dieser Sachverhalt wurde weitergehend untersucht, indem als nächster Schritt überprüft wurde, inwieweit es zu Überlappungen auf den Karten zwischen dem „Filikov Subset mit unbestätigter Aktivität“ und den jeweiligen Projektionen der Specs-Lieferung vom 21. Juli 2005 kommt. Dies ist in Abbildungen 56 bis 58 wieder für die drei Deskriptorräume dargestellt. Direkte Überlappungen wurden – wieder wie bereits oben beschrieben – mit weißen Quadraten markiert. Ersichtlich ist hierbei, dass es nur an wenigen Stellen zu einer direkten Überlappung gekommen ist und somit der Einfluss der Filikovdaten auf wenige Strukturen begrenzt zu sein scheint.

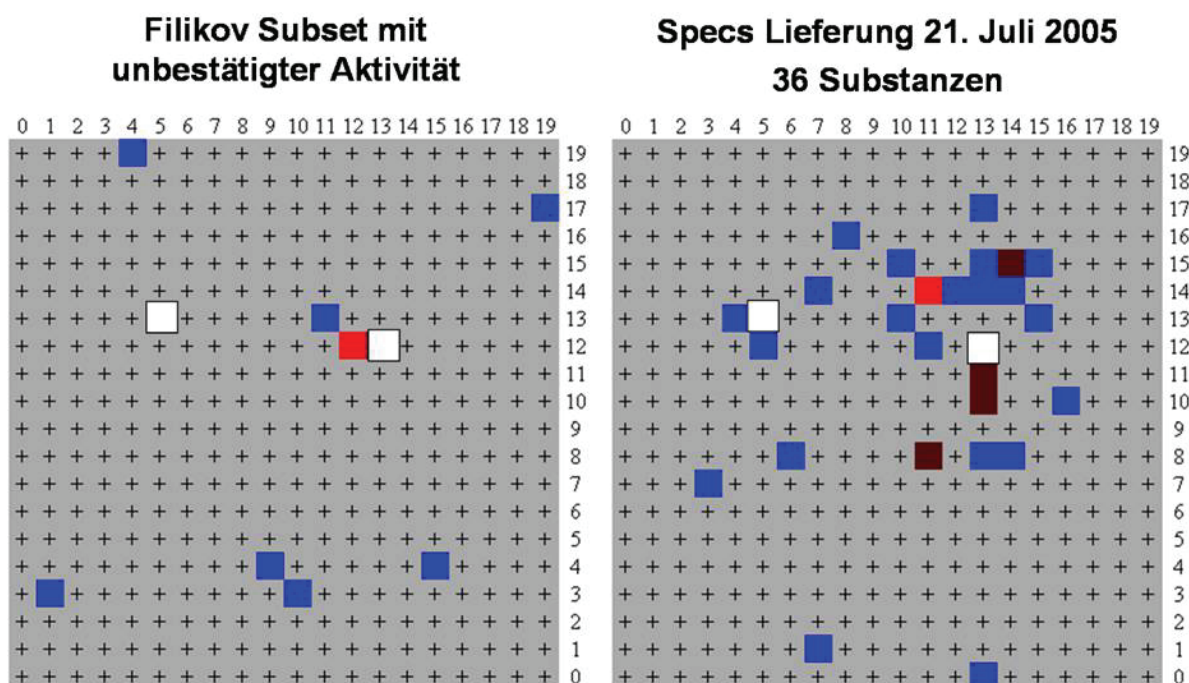


Abbildung 56: MOE-2D-Deskriptor SOM: Projektion Filikov-Subsets mit den experimentell unbestätigten Substanzen, sowie die Projektion aller bestellten Substanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt zweidirekte Übereinstimmungen zwischen den belegten Knoten und ferner deren Nähe zu weiteren belegten Knoten in der Projektion der bestellten Substanzen. Weiße Quadrate = direkte Überlappung.

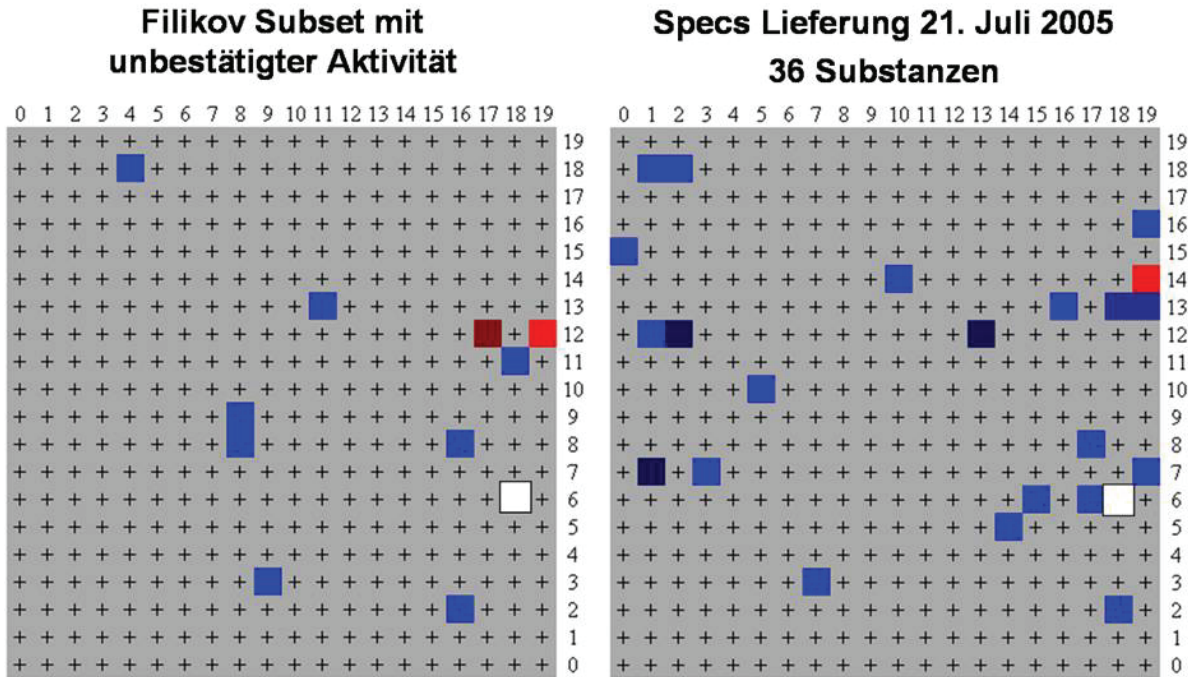


Abbildung 57: CATS-Deskriptor SOM: Projektion Filikov-Subsets mit den experimentell unbestätigten Substanzen, sowie die Projektion aller bestellten Substanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt eine direkte Übereinstimmungen zwischen den belegten Knoten und ferner deren Nähe zu weiteren belegten Knoten in der Projektion der bestellten Substanzen. Weißes Quadrat = direkte Überlappung.

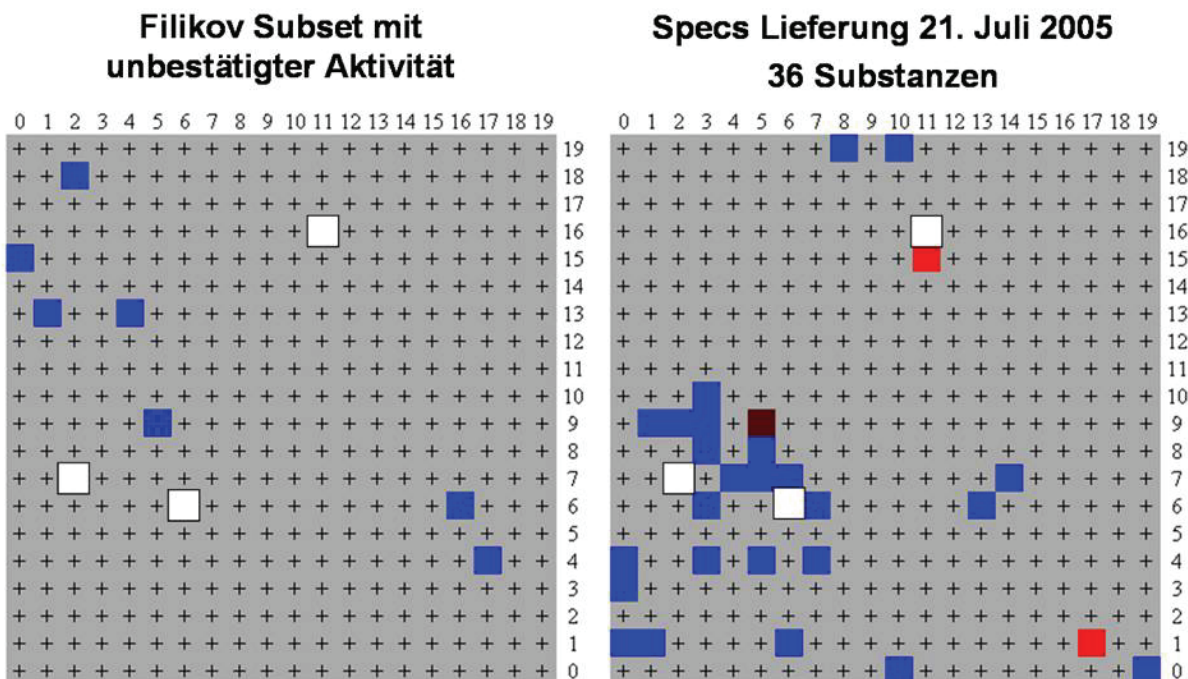


Abbildung 58: TC3D-Deskriptor SOM: Projektion Filikov-Subsets mit den experimentell unbestätigten Substanzen, sowie die Projektion aller bestellten Substanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt eine direkte Übereinstimmungen zwischen den belegten Knoten und ferner deren Nähe zu weiteren belegten Knoten in der Projektion der bestellten Substanzen. Weißes Quadrate = direkte Überlappung.

Die Tabelle 14 zeigt die zusammenfassende Aufschlüsselung der Auswertung der SOM-Projektionsvergleichsanalyse zwischen dem vermeintlich fehlerhaften Filikov-Subdatensatz und den bestellten Substanzen zur Testung im Labor. Des Weiteren ist der Anteil der bevölkerten Knoten der Specs-Bestellung angegeben, der wahrscheinlich beeinflusst wurde.

Tabelle 14: Aufschlüsselung der belegten Knoten innerhalb der SOMs für uRNALD und die Bestellung bei Specs sowie deren Überlappung.

	Anzahl im Subset (Filikovs unbestätigt)	Anzahl Specs Lieferung	Overlap	% kompromittiert in Specs Lieferung
MOE	10	28	2	7,1
CATS	11	23	1	4,3
TC3D	10	31	3	9,7

Diskussion:

Eine Überlappung würde, falls es sich bei den Strukturen in uRNALD tatsächlich nur um falsch positive RNA-Liganden handelt, in einer Senkung der Hitrate im Labor enden. Die bisherige Analyse sagt jedoch nur etwas über die mögliche Beeinflussung der Auswahl vor der Bestellung aus und hat noch keinen Bezug auf die Erfolgsrate im Labor, welches hier in der Studie als Erfolgskriterium verwendet wurde. Dieser Sachverhalt wurde im nächsten Abschnitt untersucht, indem wieder mittels Überlappungsanalyse der Karten geprüft wurde, welche Strukturen aus dem Trainingsdatensatz bei diesen Projektionen den als aktiv bewerteten Substanzen aus den Labor-Assays FRET und CFivTT am nächsten lagen. Zu sehen ist diese Überprüfung für die CATS-SOM in Abbildungen 59 und 60.

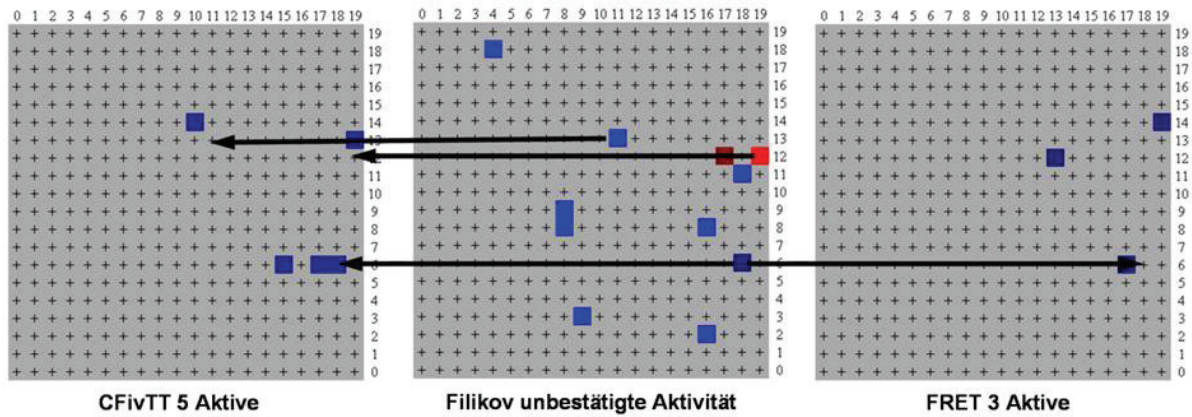


Abbildung 59: CATS SOM - Überlappungsanalyse des RNA-Liganden-Datensatzes (unbestätigte Filikov) mit den aktiven Laborsubstanzen

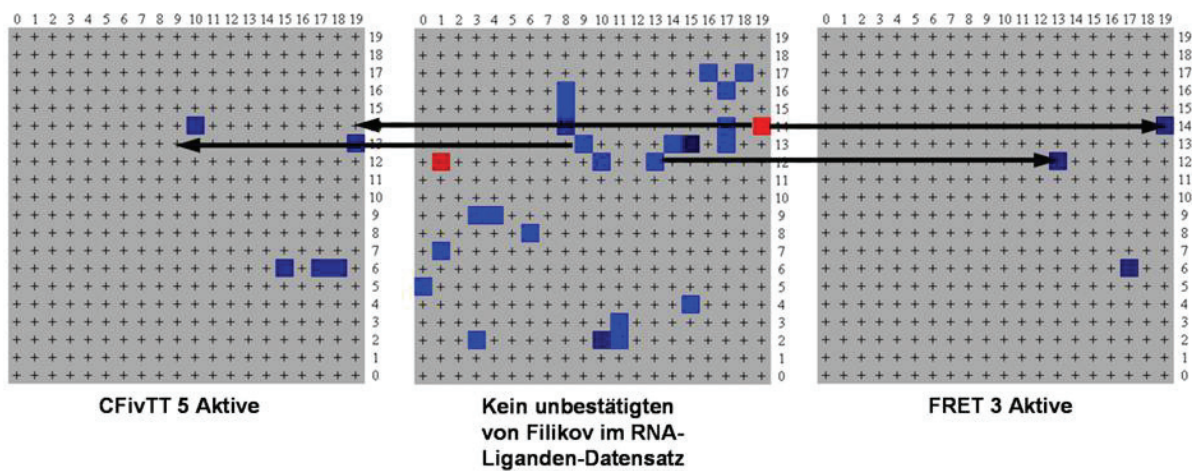


Abbildung 60: CATS SOM - Überlappungsanalyse des RNA-Liganden-Datensatzes (ohne unbestätigte Filikov) mit den aktiven Laborsubstanzen

Die Pfeile in Abbildung 59 und Abbildung 60 deuten an, an welchen Stellen es zu direkten Überlappungen kam und, falls keine direkten Überlappungen in den Trainingsdaten zu finden waren, wo sich die nächste mögliche Struktur im Trainingsdatensatz befand.

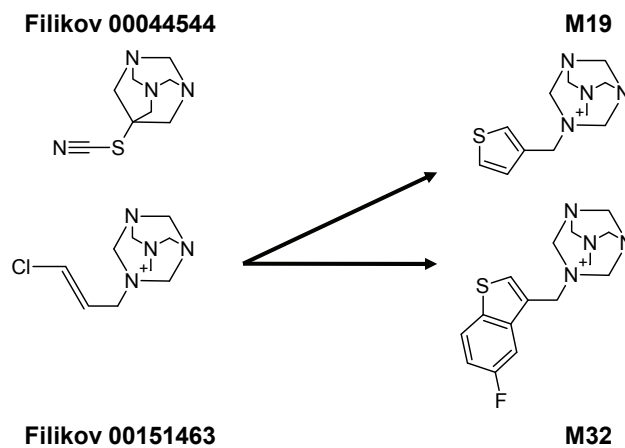


Abbildung 61: Wahrscheinlicher Ableitungsweg der aktiven Laborsubstanzen aus dem verwendeten Trainingsdatensatz; links CATS-SOM Mitglieder des Knotens [18; 6] (uRNALD), rechts die Mitglieder der CATS-Knoten [18; 6](CFivTT) und [17; 6](CFivTT).

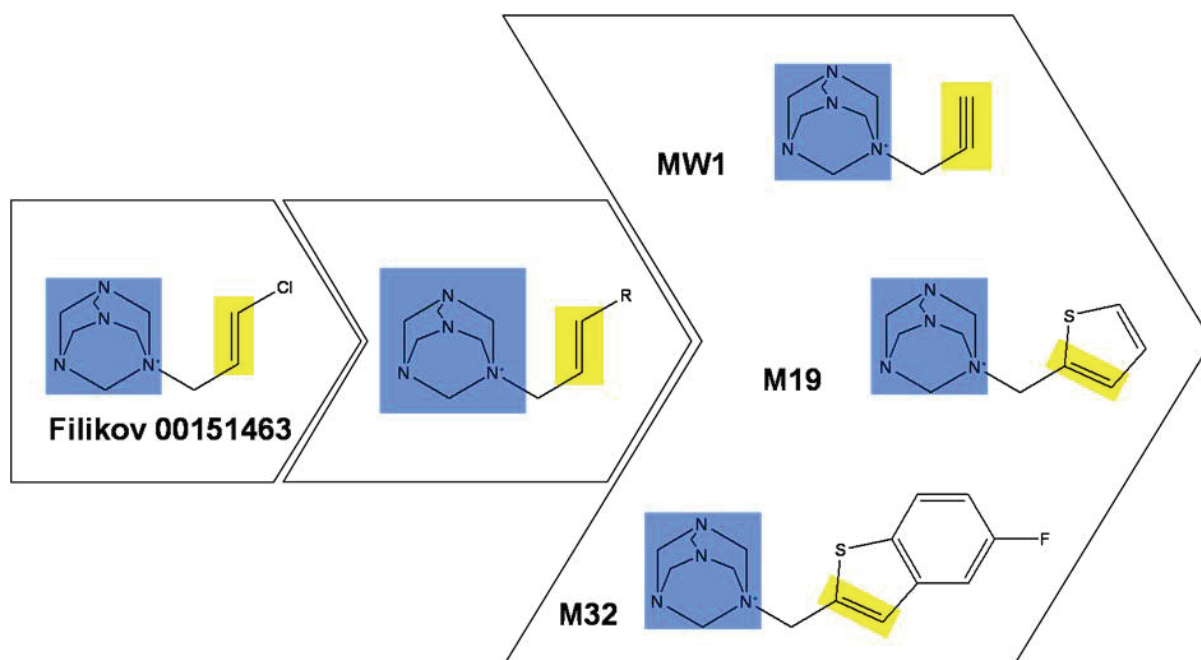


Abbildung 62: Vermuteter Informationsfluss innerhalb des ANN-Filter-Trainings.

Links in Abbildung 62 ist die Startstruktur aus dem Trainingsdatensatz, in der Mitte die wahrscheinlich abgeleitete Substruktur und rechts sind die bekannten Aktiven aus den Labor-Assays FRET und CFivTT zu sehen. MW1 wurde im Rahmen einer weiterführenden Struktur-Wirkungsbeziehungsstudie für M32 von Markus Weidlich mittels Substruktursuche gefunden. Blau hinterlegt ist 3,5,7-triaza-1azonia-adamantyl; gelb hinterlegt ist die ungesättigte Bindung, welche immer nach zwei gesättigten Bindungen in der Seitenkette folgt.

Diskussion:

Da wie bereits erwähnt der Anteil der mit CATS-Filtern gefundenen aktiven Substanzen 1/7 an den gesamten Aktiven beträgt, wird hier nur auf die Analyse des CATS-Deskriptorraums eingegangen. Dies geschah, um zu klären, wieso ausgerechnet bei diesem eine relative Unempfindlichkeit für den Filikov-Datensatzanteil festzustellen war.

Für den Knoten CATS [18; 6] ist bei Betrachtung der Projektion der Aktiven aus FRET und CFivTT eine direkte Überlappung für CFivTT festzustellen und eine direkte Nachbarschaft für FRET ersichtlich; siehe Abbildung 59. Für den benachbarten Knoten CATS [17; 6] ist jeweils dieselbe Substanz verantwortlich (M32). Da in diesem Bereich nur der Knoten CATS [18; 6] in der Projektion des „Filikov Subset mit unbestätigter Aktivität“ zuzuordnen ist, der auch einer direkten Überlappung in der Projektion des CFivTT-Datensatzes entspricht, und diesem Bereich der Karte keine benachbarten bevölkerten Knoten der Projektion des „RNA-Liganden-Datensatz ohne unbestätigte Filikov“ zuzuordnen ist, kann hier davon ausgegangen werden, dass für diese gefundenen aktiven Substanzen in den Labor-Assays die Filikov-Daten die Grundlage für die Selektion gebildet haben. Dies ließ sich durch den Vergleich der in diesen Knoten gefundenen Substanzen und ihrer Gegenüberstellung mit den aktiven Laborsubstanzen in Abbildung 61 weiter erhärten. Jedoch wurden besagte Verbindungen M19 und M32 nicht vom CATS-Filter vorhergesagt, obwohl dies nach der SOM-Projektion möglich gewesen wäre. Somit lässt sich hier ableiten, dass dieser Filter sich auf einen anderen Bereich des chemischen Raums konzentriert hat. Dies erklärt wiederum, wieso die einzige exklusiv mit diesem Filter erfolgreich als aktiv vorhergesagte Verbindung M23 nur von diesem Filter und nicht auch von den anderen Filtern identifiziert wurde.

Zusammenfassend lässt sich somit ein Grundmuster erkennen, das vom Ursprung im Trainingsdatensatz über die Abstraktion der ANN-Filter basierend auf den MOE- und TC3D-Deskriptoren die Auswahl der bestellten Substanzen durchzieht und zum Erfolg in beiden Labor-Assays geführt hat. Dieses Grundmuster ist in Abbildung 62 angegeben. Da der Ursprung dieses Strukturmusters in der Vorhersage für einen l-Arginin-Aptamer zu finden ist, ist die Promiskuität des Bindevhaltens der gewonnenen Liganden als eine Bestätigung der allgemeinen Güte der RNA-Bindervorhersage durch die verwendeten ANN-Filter zu werten. Interessant ist hierbei anzumerken, dass die Filikov-Studie, welcher der Prototyp entspringt, ebenfalls Vorhersagen zu TAR-Liganden enthält, jedoch wurde der Prototyp (Filikov 00151463) nicht als solcher vorhergesagt. Dies legt – im Angesicht der in dieser Studie durchgeführten Analyse der Diversität der TAR-Zielstruktur – den Schluss nahe, dass bei Filikov (Filikov et al. 2000) möglicherweise nicht der passende Konformer für ihre Dockingstudien verwendet wurde und sie deshalb den Liganden nicht in ihrer „Top10“-Liste der TAR-Liganden führen. Jedoch zeigte auch die zu „Filikov 00151463“ ähnelnde Verbindung M19 (siehe Abbildung 62) keine Aktivität im hier verwendeten Tat-TAR-FRET-Assay sondern wies diese nur im CFivTT-Assay auf. Hieraus kann geschlossen werden, dass der Ansatz von Filikov (Filikov et al. 2000) eine relativ gute Diskriminierungsgüte aufweisen könnte: Es wäre somit interessant zu wissen, wie die Substanz M32 durch den Filikov-Ansatz bewertet werden würde. Im Rückblick passt jedoch die Übereinstimmung der Grundstruktur, welche für den l-Arginin-Aptamer entdeckt wurde, ebenfalls gut zum Wissen, dass die TAR-RNA ebenfalls an ein Triple-Arginin-Motiv des Tat-Proteins bindet. Somit sollten auch beide Zielstrukturen in der Lage sein Arginin-artige Struktur motive zu binden.

3.3 Tat-TAR-FRET-Assay-spezifischer ANN-basierter Struktur-Filter

3.3.1 Erweiterter Datensatz und Ableitung eines spezifischeren Filters Hintergrund:

Die Motivation für das erneute Erstellen eines KNN-basierten Filters ist ein TAR-spezifischer Filter für das fokussierte *de novo* Design. Die fehlende Selektivität des RNA-Binder-Filters wurde aus den Aktivitäts-Überlappungsuntersuchungen für die beiden Assays Tat-TAR und CFivTT ersichtlich, siehe hierzu Abschnitt 3.2.15. Dies war jedoch keine Überraschung, da der RNA-Binder-Filter von vorneherein auf eine allgemeine RNA-Binder-Vorhersage ausgelegt war. Jedoch zeigte die erzielte Trefferquote beim Tat-TAR-Assay auf einen noch weiter zu optimierenden Stand, dessen Lösung ein für diesen Assay spezifisch trainierter Filter sein könnte.

3.3.2 Erstellung eines neuen Tat-TAR-Assay spezifischen Datensatzes

Da im Zeitraum dieser Arbeit Tests mit weiteren Substanzen verschiedener Quellen des SFB579 durchgeführt wurden, konnte ein neuer Datensatz ausreichender Größe für ein Netztraining zusammengetragen werden, der einzig auf Strukturen basierte, welche im Tat-TAR-Assay des AK-Göbel vermessen wurden. Die Hoffnung hierbei war die Steigerung der Trefferquote in einer neuen Bestellrunde zur Testung in eben jenem Assay und der späteren Anwendung als Filter von *de novo* Strukturvorschlägen zur Synthese potentieller neuartiger RNA-Binder für TAR.

Der neue Datensatz setzte sich aus Strukturen zusammen, welche folgenden Quellen entnommen wurden:

- Verbindungen, welche von Andreas Schüller im Rahmen seiner Doktorarbeit zur „Adaptive Optimierung fokussierter Substanzbibliotheken für die Leitstrukturfindung“ von ihm getestet wurden.
- Verbindungen von Steffen Renner, welche im Rahmen seiner Doktorarbeit von Verena Ludwig getestet wurden (Renner 2005; Renner et al. 2005).
- Weitere Verbindungen, welche im AK-Göbel von Dr. Ute Scheffer getestet wurden und aus verschiedenen Doktorarbeiten stammten (O. Boden, V. Ludwig, A. Krebs, G. Seifert) (Boden 2005; Ludwig 2005; Ludwig et al. 2007; Seifert 2008; pers. Komm. Dr. Ute Scheffer).
- Verbindungen, welche im Rahmen dieser Arbeit als potentielle RNA-Binder vorhergesagt und dort im Rahmen dieser Arbeit getestet wurden; siehe Abschnitt 3.2.11 und 3.2.13.
- Verbindungen, welche im Rahmen der Zusammenarbeit mit Yusuf Tanrikulu als Vorhersagen seiner LIQUID-Modelle dort getestet wurden (Tanrikulu et al. 2007).

Das Zusammenführen dieser Datensätze führt bei einer festgelegten Aktivitätsgrenze der Substanzen von 1 mM im Assay zu einem kombinierten Datensatz, der aus 64 im Assay aktiven und 242 im Assay inaktiven Verbindungen bestand. Anmerkung: Dieser kann auf Anfrage vom Autor erhalten werden.

3.3.3 Vorbereitung der Datensätze / Konformergeneration

Zur Vorbereitung der Datensätze auf die Deskriptorberechnung wurde jeweils für die beiden Subdatensätze der aktiven und inaktiven Verbindungen eine Datensatzvorbereitung, wie bereits im Abschnitt 3.3.3 „Vorbereitung der Strukturdatensätze / Konformergeneration“ beschrieben, durchgeführt. Dies wurde wieder mit dem „db_prepare“-Shellskript (Version: 1.6) von Andreas Schüller, durchgeführt. Dies resultierte in zwei 3D-Konformere enthaltende SDF-Dateien, welche zur weiteren Deskriptorberechnung eingesetzt werden konnten.

3.3.4 Deskriptorberechnungen für den Tat-TAR-FRET-Assay-spezifischen Datensatz

Für die zu trainierenden Filter wurden folgende vier Deskriptoren-Sets berechnet:

3.3.4.1 MOE2D-Deskriptoren

Für die Berechnung der 2D-MOE-Deskriptoren wurde MOE (Version 2006.08) eingesetzt.

Hierzu wurden folgende Schritte durchgeführt:

5. Die in Abschnitt 3.3.3 erstellten 3D-SDF-Dateien wurden in MOE-Datenbanken importiert.
6. Mit den Standard-Einstellungen von MOE wurden Partialladungen berechnet und den Struktureinträgen hinzugefügt (verwendetes Kraftfeld = MMFF94x (Halgren 1996)).
7. Basierend auf den Struktureinträgen wurden alle 2D-Deskriptoren ausgewählt und berechnet.
8. Die so gewonnenen Deskriptoren wurden, zusammen mit den SMILES der Strukturen, als Textdateien exportiert.

Anmerkung:

Alle Operationen mit dem Programm wurden hier mittels Programmoberfläche durchgeführt; siehe hierzu auch Abschnitt 2.1.1 „MOE-2D-Deskriptoren“ Seite 9.

3.3.4.2 CATS-Deskriptoren

Die Berechnung wurde analog für alle oben genannten Datensätze durchgeführt.

Für die Berechnung der CATS Deskriptoren wurde auf Uli Fechners jSpeedcats (Version 0.32) (Fechner et al. 2003) zurückgegriffen, welches mit den Standardparametern aufgerufen wurde, wie sie auch für die ANN-Filter von 2005 verwendet wurden. Hierbei reduziert sich die erforderliche Parameterübergabe auf:

- -i Eingabedateiname (SDF-File).
- -o Ausgabedateiname (Deskriptortextdatei).
- -s „suppress header“ = keine Kopfzeile.

Eine genauere Definition des CATS-Deskriptors erfolgte in Abschnitt 2.1.2.

Die Berechnung wurde analog für alle oben genannten Datensätze durchgeführt.

TC3D-Deskriptoren

Für die Berechnung der TC3D-Deskriptoren wurde wieder auf das im Rahmen dieser Arbeit entstandene ChargeCyclone3D v2.2 zurückgegriffen. Die Berechnung erfolgte analog zu 2005 mit jeweils 100 Bins für jede der Klassen (++)/+/-- und einem maximalen zu beachtenden Abstand von 10 Anströme. Verwendet wurden hierzu die 3D-Konformere enthaltenden Datensätze.

3.3.4.3 LIQUID-Deskriptoren

Die Berechnung der LIQUID-Deskriptoren erfolgte mittels des Programms LIQUID-VS (Version 1.2) von Yusuf Tanrikulu unter Angabe folgender Parameter für alle angewendeten Strukturdateien.

Clusterradii:

L = 2.0 A

D = 2.0 A

A = 2.0 A

Correlation vector scaling mode:

Surface to one

3.3.5 Deskriptorberechnungen für die COBRA und Specs

Die oben genannten Berechnungen wurden analog zu Abschnitt 3.3.4 ebenfalls für den COBRA-Datensatz (Schneider, P. und Schneider, G.) und die Chemikalien-Datensätze von Specs (Delft, The Netherlands. <http://www.specs.net>) durchgeführt, um diese später im Virtuellen Screening einsetzen zu können.

3.3.6 Abschätzung der Trennbarkeit des FRET-Assay-spezifischen Datensatzes mittels SOM-Projektionen

Hintergrund:

Die aus den FRET-Assay-spezifischen Datensätzen ableitbaren Verfahren (ANN-Filter), sollten aufgrund der in diesen Datensätzen möglicherweise vorhandenen Unterschiede lernen zwischen den aktiven und inaktiven Verbindungen zu unterscheiden und so letztlich im Einsatz mit fremden Datensätzen eine Einstufung der darin enthaltenen Verbindungen vornehmen zu können. Der erste gewählte Ansatz für die Überprüfung der möglichen „Erlernbarkeit“ der Fragestellung wurde mittels der unüberwachten Lernmethode der SOM umgesetzt. Dieses Mal erfolgte die Umsetzung mittels SOMMER- *Self-Organizing Map Maker for Education and Research* (Schmucker et al. 2007).

Durchführung:

Zur ersten Untersuchung der Trennbarkeit der beiden Teildatensätze wurden SOMs eingesetzt, welche mit den MOE-2D-Deskriptoren des Gesamtdatensatzes trainiert wurden.

Es wurden folgende Parameter eingesetzt:

Topologie: Toroidal Topology

Höhe: 10

Breite: 15

Distanz-Funktion: Euclidian

Max time (maximale Anzahl an Trainingszyklen): 10.000

Initial Neighborhood: 5,0
 Final Neighborhood: 0,5
 Initial learning rate: 1,0
 Final learning rate: 0,01
 Training type: "Train by Vector"

Ergebnisse:

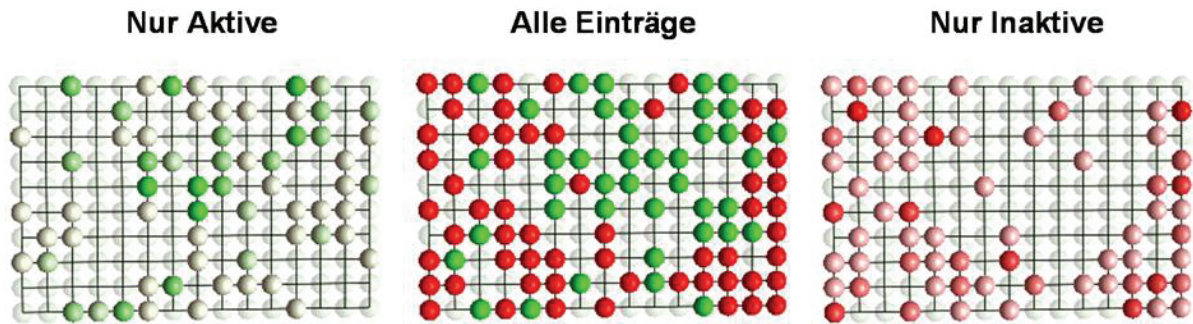


Abbildung 63: SOMMER SOM des Tat-TAR-spezifischen Datensatzes angewendet auf die MOE-2D-Deskriptoren; grün = Aktive, rot = Inaktive; in der gemeinsamen Projektion sind die Knoten nach der häufigsten Klasse eingefärbt.

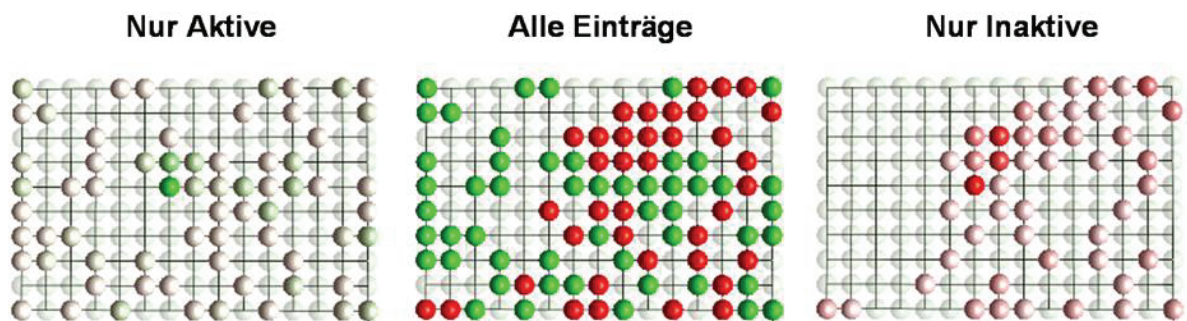


Abbildung 64: SOMMER SOM des Tat-TAR-spezifischen Datensatzes angewendet auf die TC3D-Deskriptoren; grün = Aktive, rot = Inaktive; in der gemeinsamen Projektion sind die Knoten nach der häufigsten Klasse eingefärbt.

Die Trennung erfolgt aufgrund physikochemischer und topologischer Deskriptoren, welche bereits beim ersten Ansatz erfolgreich getestet worden waren.

Diskussion:

Die anschließende Projektion der Subdatensätze zeigte eine relative Trennbarkeit der Datensätze anhand der eingesetzten Merkmale; siehe hierzu Abbildungen 63 und 64, welche diese Trennbarkeit anhand der unterschiedlichen Verteilung innerhalb der SOM aufzeigen. Die Trennung ist hier nicht durch eine deutliche Trennung der Bereiche beider Klassen ersichtlich, sondern vielmehr durch das Vorhandensein einzelner „Inselbereiche“ innerhalb der Karten, welche durch einzelne Klassen belegt werden. In Abbildung 64 ist beispielsweise der linke Kartenbereich nur durch Einträge der Aktiven belegt.

3.3.7 Vergleich der Molekülgerüste von COBRA, TAR-Liganden und TAR-Nichtliganden

Hintergrund:

Um eine Abschätzung der möglichen Beeinflussung der Aktivität durch bestimmte verwendete Grundgerüste im Tat-TAR-FRET-Assay zu untersuchen, wurde eine Analyse der einzelnen Datensätze (COBRA, TAR-Liganden und TAR-Nichtliganden) durchgeführt. Diese Analyse sollte klar stellen, ob bestimmte Grundstruktur motive innerhalb der RNA-Datensätze Unterschiede zu den bisher in der COBRA dokumentierten Grundgerüsten bioaktiver Verbindungen zeigen.

Durchführung:

Die dargestellte Untersuchung der Molekülgerüst-Extraktion erfolgte mittels zweier MOE-Skripte, die freundlicherweise von Kristina Grabowski (Grabowski und Schneider 2007) zur Verfügung gestellt wurden.

Ergebnisse:

Eine Untersuchung der Molekülgerüste des erstellten Tat-TAR-Assay-Datensatzes zeigte sowohl Unterschiede im Aufbau der TAR-Aktiven zu den TAR-Inaktiven, als auch zu den biologisch aktiven Substanzen des COBRA-Datensatzes, siehe hierzu Abbildung 65, welche die 11 häufigsten Molekülgerüste der COBRA darstellt und die Anzahl des jeweiligen Vorkommens dieser Molekülgerüste innerhalb der beiden Teildatensätze des Tat-TAR-spezifischen Datensatzes zeigt.

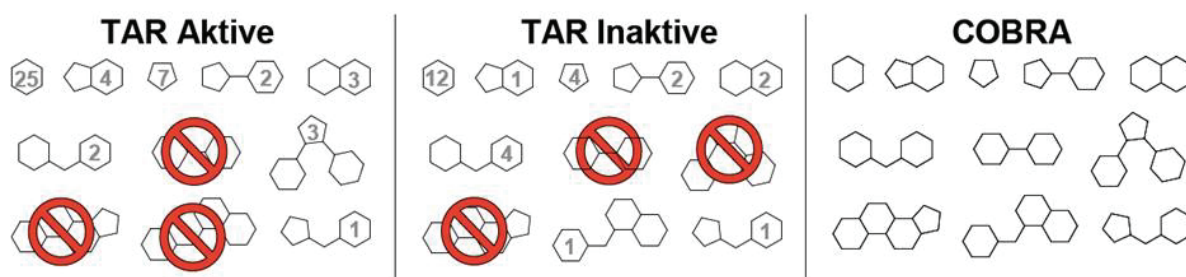


Abbildung 65: Darstellung der Molekülgerüste des Tat-TAR-spezifischen Datensatzes aufgeteilt nach Aktivitätssubklassen; als Referenzauswahl wurden die 11 häufigsten Molekülgerüste des COBRA-Datensatzes verwendet, welche rechts zu sehen sind; Innerhalb des Tat-TAR-spezifischen Datensatzes wurden die Vorkommenshäufigkeiten der Molekülgerüste als Zahlen eingefügt. Kommen diese überhaupt nicht vor, wurden sie entsprechend als durchgestrichen markiert.

Diskussion:

Die Größe der untersuchten RNA-Datensätze erlaubt hier nur eine vorsichtige Interpretation über mögliche sterisch vorteilhafte Grundgerüste für die Störung der Tat-TAR-Wechselwirkung. Jedoch werden folgende Annahmen von der Ergebnissen unterstützt.

Aufbauend auf diesen Ergebnissen könnten in weiteren Studien diese bevorzugten Molekülgrundgerüste für die Auswahl möglicher Testsubstanzen herangezogen werden. In den Grundgerüsten der Liganden sollte die theoretische Passform an ihre Zielstruktur kodiert vorliegen. Da die Anzahl der bisher getesteten Substanzen und die weitere Reduktion auf verschiedene Molekülgerüste bedingt durch die bisher geringe Anzahl verschiedener Ansätze für die RNA-Liganden-Vorhersage jedoch nur einen Teil des chemisch zugänglichen Strukturraums ausmacht, sollte aus diesen Vorschlägen nicht verfrüht ein Schema abgeleitet werden. Da ein solches Schema nur den Bereich der bisher erforschten Bereiche des chemischen Raums abdecken würde, wäre somit eine starke Einschränkung des Testraums die Folge. In letzter Konsequenz würden möglicherweise eventuelle hochaffine Bereiche ($IC_{50} < 1 \mu\text{M}$) des chemischen Raums für TAR zu früh in der weiteren

Entwicklungsphase ausgeschlossen werden. Trotzdem zeigt sich auch hier wieder eine grundsätzliche Unterscheidbarkeit der untersuchten Substanzen anhand dieses Merkmals, auch wenn die Ausprägung aufgrund der Datensatzgröße zum jetzigen Zeitpunkt einen direkten Einsatz als Filtersystem noch nicht nahe legt, so könnte dieser Ansatz in einer späteren Phase der Erforschung der Tat-TAR-Wechselwirkung wieder eingeführt werden.

3.3.8 Filterung der Deskriptoren zur Reduktionen der Anzahl der Deskriptordimensionen für das spätere Netztraining

Hintergrund:

Die Anwendung von Deskriptoren für das Netztraining bedarf eines ausgewogenen Verhältnisses der Anzahl der verwendeten Deskriptoren in Relation zur Anzahl der verwendeten Trainingsbeispiele. Da die Anzahl der Trainingsdaten in dieser Studie von den bereits gemessenen Verbindungen im Tat-TAR-FRET-Assay abhängt, ist der zugängliche Datensatz (64 Beispiele der Aktiven-Klasse/242 Beispiele der Inaktiven-Klasse) als relativ klein zu werten. Zusätzlich hierzu ist die Trainingsdauer (Rechenzeit pro Anpassungsschritt) eines ANN abhängig von der Dimension der Eingabedaten, so dass durch eine Verminderung der Deskriptor-Dimensionen eine Beschleunigung der Trainingsdauer unkompliziert bewirkt werden kann. Solange diese Reduktion der Deskriptoren hauptsächlich auf der Entfernung der für das gegebene Trennungsproblem irrelevanten (keine Informationen tragenden) Deskriptoren beruht, ist auch davon auszugehen, dass die Trennbarkeit durch diese Maßnahme nicht weiter beeinträchtigt werden sollte, da die ANNs lernen sollten, diese Informationen durch Einführen eine Gewichts von nahe 0,0 zu ignorieren.

Als eine weitere Maßnahme zur Beschleunigung der Trainingszyklen können identische Deskriptoren entfernt werden, die keine Zusatzinformationen zur Unterscheidung der beiden Klassen beinhalten. Deskriptoren mit einer starken Korrelation zu einem bereits verwendeten Deskriptor bieten demnach keine Erhöhung des Informationsgehaltes und könnten dementsprechend aussortiert werden.

Zur automatischen Reduktion der verwendeten Anzahl der Deskriptoren sowie der automatischen Preprozessierung der Daten für das Netztraining wurde ein Modul („Analyse Descriptors“) für das Java-Projekt geschrieben, welches erlaubt Deskriptordaten einzulesen und innerhalb dieser Sets Deskriptoren spaltenweise auszuwerten. Dieses Modul entfernt automatisch Deskriptoren ohne Informationsgehalt (konstante Werte über alle Einträge). Ferner kann es mit weiteren Deskriptorverarbeitungsmodulen des Java-Projekts kombiniert werden, um die Deskriptoren anhand ihrer geringsten Korrelation zu sortieren („Correlation Matrix Analysis looking for smallest“) und, um bei einer reduzierten Auswahl der Deskriptordimensionalität, ein Maximum an unkorrelierten Informationen innerhalb des Datensatzes zu gewährleisten. Hierzu bedarf es der vorhergehenden Korrelationsmatrixberechnung als Eingabeparameter für die anschließende Selektion.

Später ist eine einfache Anwendung der Selektions- und Transformationsdaten auf die gesamten Deskriptordimensionen weiterer zu testender Datensätze möglich, was die weitere Anwendung der erstellten ANN-Filter erleichtert.

Durchführung:

Mittels des genannten Werkzeugs wurden die oben erstellten Datensätze bearbeitet und die resultierenden Datensätze konnten anschließend zum Netztraining eingesetzt werden. Hierbei wurden folgende Schritte durchgeführt:

- Import der Deskriptor-Dateien in Pandora
- Normalisierung der einzelnen Spalten mit Pandora

- Extraktion der Deskriptoren ohne konstante Werte mit Pandora
- Export der Deskriptor-Dateien aus Pandora

Ergebnisse:

Aus der Anwendung des Normalisierungs-Moduls von Pandora – mit Entfernung der Deskriptoren mit konstanten Werten über alle Einträge – ergaben sich folgende verbleibende Deskriptordimensionen:

- CATS-2D 132 von ursprünglich 150 = 88%
- TC3D 270 von ursprünglich 300 = 90%
- MOE-2D 178 von ursprünglich 184 = 97%
- LIQUID 120 von ursprünglich 120 = 100%

Diskussion:

Mittels der Anwendung des Normalisierungs-Moduls von Pandora, konnten je nach Deskriptortyp eine unterschiedliche Anzahl an Deskriptoren entfernt werden, die für alle untersuchten Einträge jeweils denselben Wert trugen. Nur der LIQUID-Deskriptor bildete hier die Ausnahme und enthielt als einziger keine Spalten mit konstanten Werten. Somit können mit diesem Verfahren unkompliziert einzelne Deskriptor-Spalten von den Gesamt-Ensembles der jeweiligen Deskriptorklassen entfernt werden, die nur eine unnötige Verlängerung der Trainingszyklen zur Folge gehabt hätten.

Anmerkung zur Korrelationsanalyse:

Das in *Pandora* erstellte Dekorrelationsmodul wurde an diesem Punkt der Arbeit zwar zum Teil auch eingesetzt, jedoch wurden die erhaltenen Listen nicht zum Netztraining verwendet. Als Eingabeparameter wurden die mittels *ChemSAR* für Excel (Version 7.0.0.212 November 16, 2001) erstellten Korrelationsmatrizen eingesetzt. Da Excel nur 255 Spalten verarbeiten kann, konnte dieser Ansatz in der Form jedoch nicht für den gesamten TC3D-Deskriptor umgesetzt werden.

Das erwähnte Modul in *Pandora* kann aufgrund einer Korrelationsmatrix (oder anderer symmetrischer Matrizen) eine nach dem MaxMin-Algorithmus (Kennard und Stone 1996) definierte Rangfolge der Einträge vornehmen. Im Beispiel mit den hier verwendeten Korrelationsmatrizen werden folgende Schritte umgesetzt:

1. Das Paar mit der geringsten Korrelation wird aus der Matrix abgeleitet und deren Deskriptor-Bezeichnungen einer Sammelliste hinzugefügt.
2. Es wird der nächste Deskriptor bestimmt, der den bereits in der Sammelliste enthaltenen Deskriptoren am unkorreliertesten ist (Mittelwert der paarweisen Korrelationswerte) und die Deskriptor-Bezeichnungen der Sammelliste angehängt.
3. Wiederholung von Schritt 2, bis sich alle Deskriptor-Bezeichnungen in der Sammelliste befinden

Von der Untersuchung der Korrelation der einzelnen Deskriptoren wurde im Rahmen dieser Untersuchung wieder Abstand genommen, da dieses Verfahren, nicht für alle Deskriptor-Klassen eingesetzt werden konnte. Ferner zeigte sich, wie zu erwarten war, dass dieses Verfahren erst nach dem Einsatz der KS-Statistik oder eines ähnlichen Verfahrens eingesetzt werden sollte, da durch die Art der Ableitung des Datensatzes so zuerst die Deskriptoren mit einem hohen KS-Wert und niedrigen KS-Wert abwechselnd kombiniert wurden. Zur Verdeutlichung dieses Sachverhalts ist Abbildung 66 dieser Arbeit hinzugefügt

worden. Die Graphik zeigt die Auftragung der KS-Distanz gegen die Position nach dem MinMax-Ranking für die verbliebenen 178 Deskriptoren (nach Abtrennung der Deskriptoren mit konstanten Werten). Insbesondere der Bereich der „Top 30“-Deskriptoren nach dem erfolgten MinMax-Ranking, zeigt eine deutlich verschobene Verteilung in den KS-Distanz-Bereich unter 0,3. Nur ein Deskriptor mit einem KS-Distanz-Wert von $> 0,4$ wurde hier innerhalb der „Top 30“ ausgewählt. Deutlich erkennbar ist die „Pendelbewegung“ der Kurve zwischen Datenpunkten mit hohem und niedrigem KS-Distanzwert. Da die KS-Distanz im Rahmen dieser Arbeit immer wieder als Gütekriterium eingesetzt wurde um die Trainierbarkeit der ANNs im Voraus abzuschätzen, lässt das gezeigte Verhalten in Abbildung 66 darauf schließen, dass der Ansatz mittels MaxMin-Algorithmus zur Auswahl der am „besten“ geeigneten Deskriptoren noch der weiteren Optimierung bedarf. Diese Optimierungstätigkeit wurde im Rahmen dieser Arbeit wieder an die ANN-Trainingsalgorithmen übertragen, da zu diesem Zeitpunkt keine begründbare Reduktion der Deskriptorklassen-Dimensionen ableitbar war.

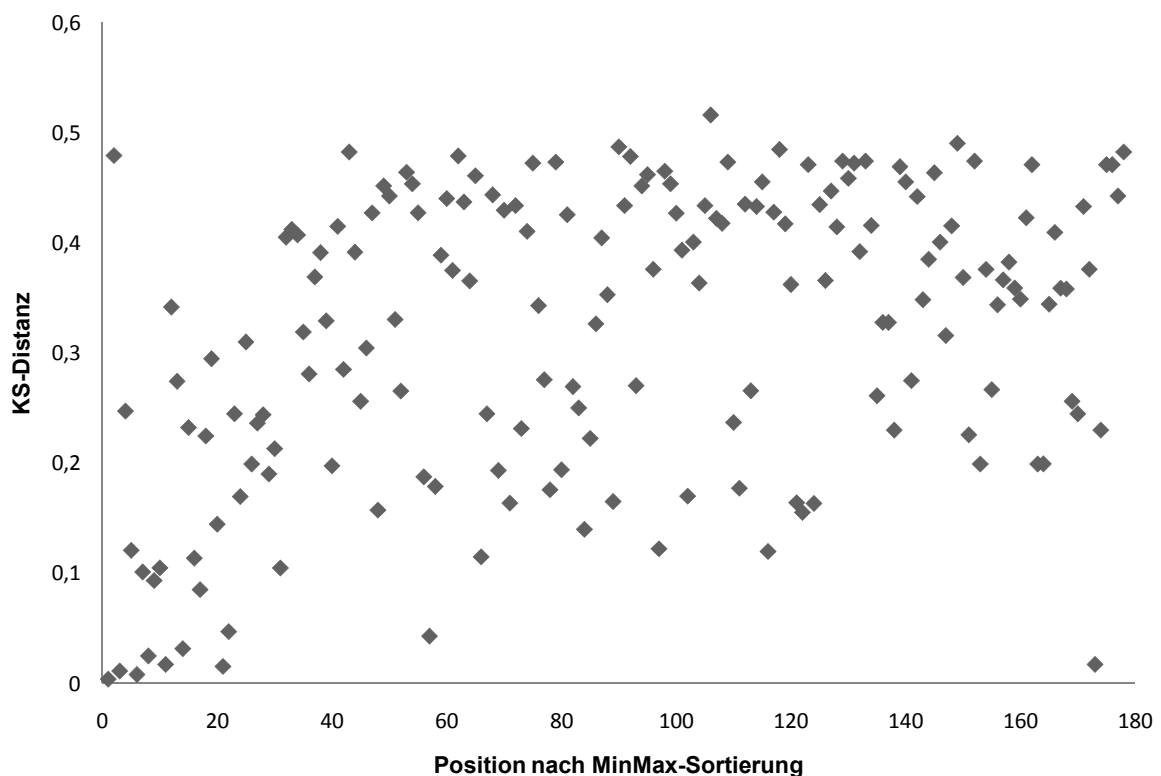


Abbildung 66: KS-Distanz gegen Position in MinMax-Sortierung für den MOE-Deskriptorraum nach Abtrennung der Deskriptoren mit konstanten Werten für alle Einträge.

3.3.9 KS-Statistiken für die Tat-TAR-Assay-spezifischen Datensätze

Hintergrund:

Im Rahmen dieser Arbeit wird die KS-Statistik eingesetzt, um nach Unterschieden innerhalb der Werteverteilungen zweier Klassen zu suchen. Diese Unterschiede bei den Verteilungen der Werte zweier Klassen, lassen direkte Schlüsse über die Trainierbarkeit von ANNs oder vergleichbaren Systemen zu. Als erster Test ist somit vor der Anwendung eines Deskriptordatensatzes für das ANN-Training zur Unterscheidung zweier Klassen eine KS-Statistik durchgeführt worden. So erfolgte dieser Schritt auch bei den Tat-TAR-Assay-spezifischen Datensätzen. Die KS-Statistik kann darüber hinaus verwendet werden, um die „markantesten“ Unterschiede zweier Klassen aus den Werten ihrer Deskriptorensembles

abzuleiten. In diesem Fall interessierte aber nur die relative Trennbarkeit der beiden Klassen, wie sie durch die KS-Distanz-Werte ausgedrückt wird.

Durchführung:

Die jeweiligen Deskriptordateien wurden mittels KSandra verarbeitet und anschließend diese Ergebnisse in Excel aufbereitet. Hierzu wurden die erhaltenen Ergebnislisten nach den KS-Distanzwerten sortiert und als Graphik gegen ihre Position im Ranking aufgetragen. Hieraus resultierten die Abbildungen 67 bis 70.

Ergebnisse:

Die Abbildungen 67 bis 70 zeigen die berechneten KS-Distanz-Werte für die einzelnen Deskriptorklassen (TC3D, LIQUID, MOE-2D und CATS-2D). Die Auftragung erfolgte sortiert nach der jeweils erreichten KS-Distanz.

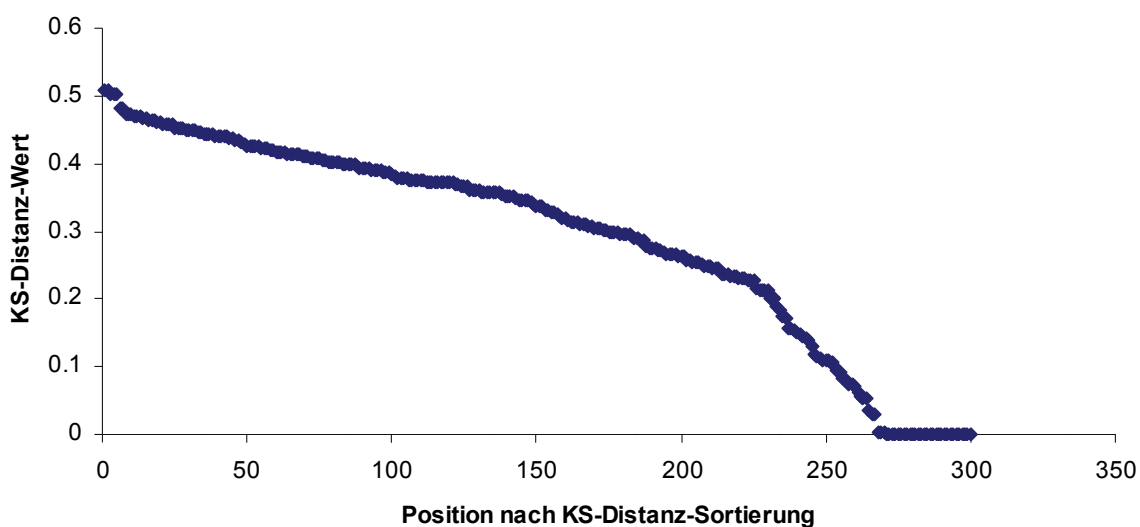


Abbildung 67: KS-Statistik der TC3D-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze

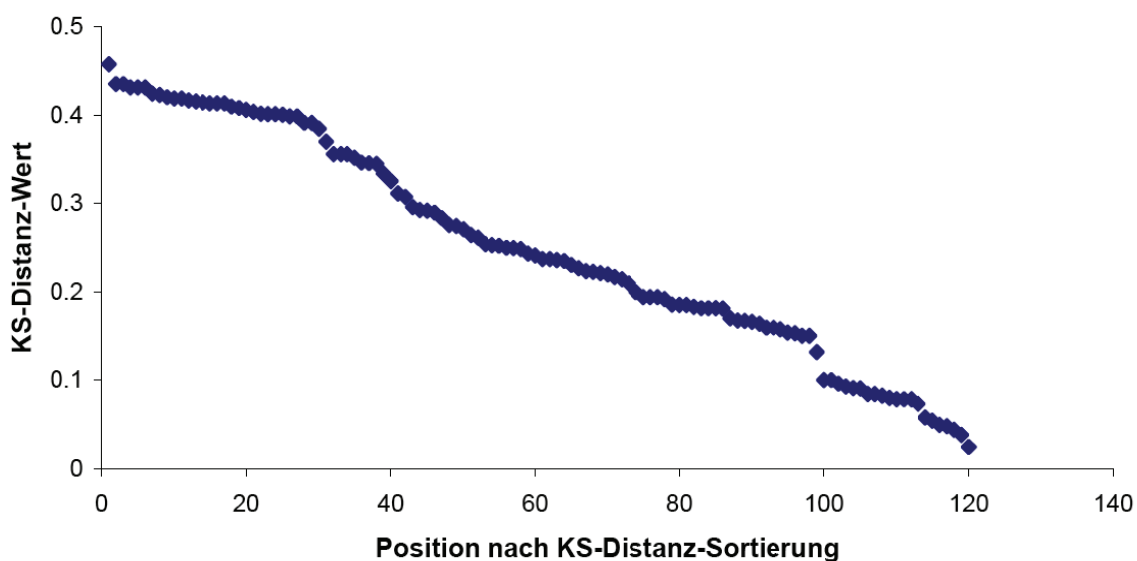


Abbildung 68: KS-Statistik der LIQUID-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze

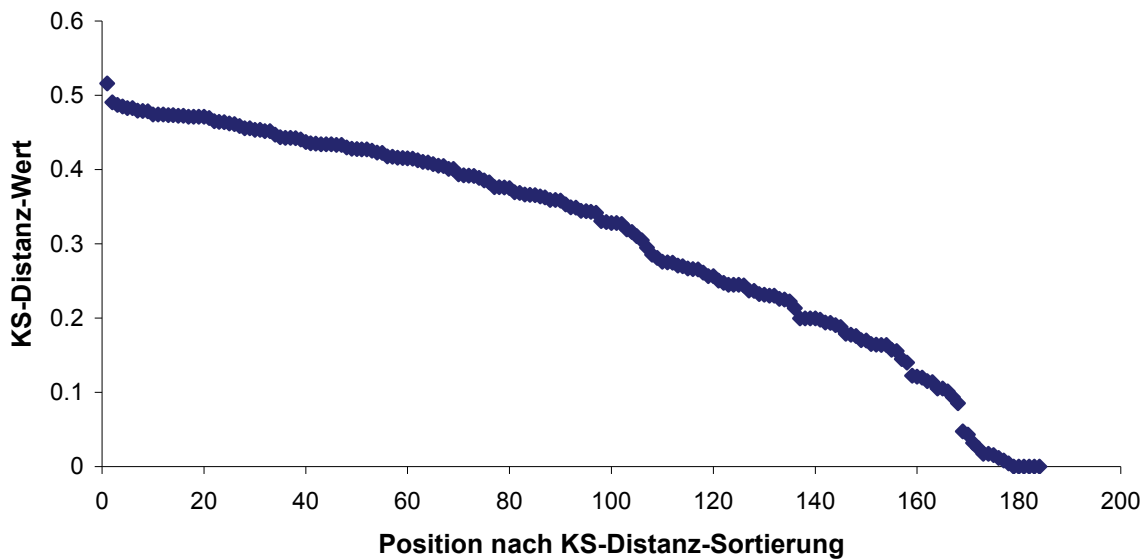


Abbildung 69: KS-Statistik der MOE-2D-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze

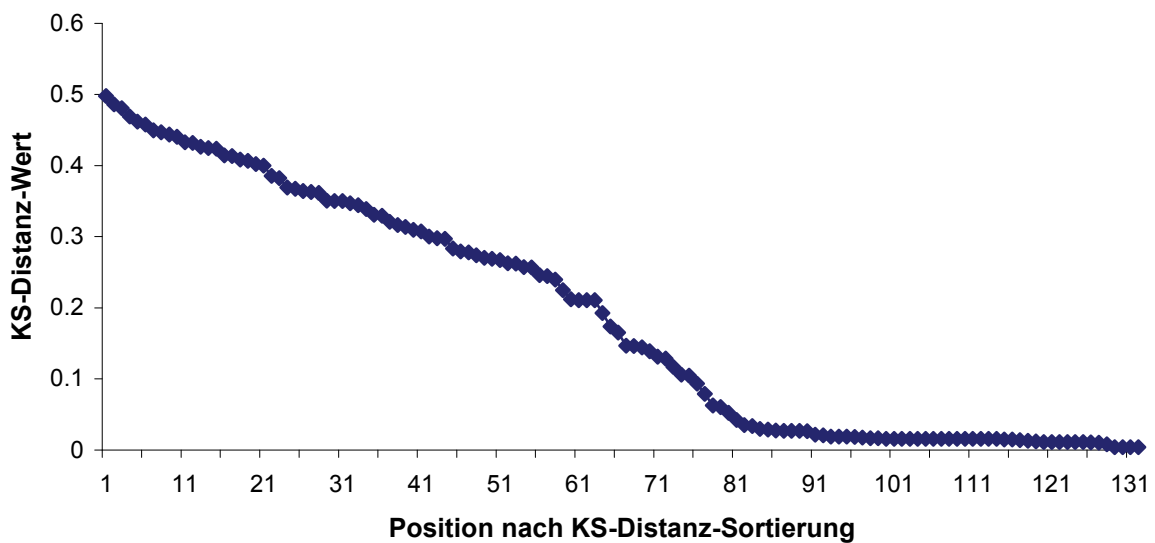


Abbildung 70: KS-Statistik der CATS-2D-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze; die 18 Deskriptoren mit einer KS-Distanz von 0,0 wurden bereits vor der Auftragung entfernt.

Diskussion:

Die Abbildungen 67 bis 70 zeigen die erreichten KS-Distanz-Werte für die Tat-TAR-Assay-spezifischen Datensätze, diese zeigen durchgehend eine relative Trennbarkeit der beiden Datensätze anhand der gewählten Deskriptorklassen. Innerhalb dieser Klassen gibt es jedoch zum Teil deutliche Unterschiede im Anteil der Deskriptoren, die einen KS-Distanzwert $> 0,1$ erreichen. Besonders beim CATS-2D-Deskriptor ist dies sehr augenfällig, da 50% der einzelnen Korrelationsvektor-Deskriptoreinheiten einen Wert unter 0,1 zugewiesen bekamen. Dahingegen wurden für mindestens 84% der einzelnen Deskriptoreinheiten der anderen Deskriptorklassen (MOE-2D, TC3D und LIQUID ein KS-Distanz-Wert über 0,1 bestimmt.

Schlussfolgerung:

Die gezeigten KS-Distanzen für die Tat-TAR-Assay-spezifischen Datensätze legen eine Trennbarkeit der Aktiven von der Inaktiven-Klasse nahe. Eine Kombination der laut KS-Statistik als geeignet erscheinenden Deskriptoren erscheint hier als eine weitere Option um die Vorhersageeigenschaften der ANNs möglicherweise zu verbessern. Hierbei könnte die Auswahl mittels Korrelationsanalyse erfolgen, die die benötigten Dimensionen des resultierenden – aus mehreren Deskriptorklassen gemischten – Deskriptorvektors auf ein für das ANN-Training anwendbare Anzahl reduzieren. Aufgrund der erwarteten Anzahl der laut KS-Distanz „sinnvollen“ Dimensionen, müsste für diesen Ansatz eine Korrelationsmatrixberechnung in einem dazu geeigneten Programmumfeld erfolgen. Dies könnte einerseits die Java-Umgebung von Pandora sein, als auch in Matlab mittels m-Skript erfolgen.

Mit dem Erweitern der Messdaten für den Tat-TAR-FRET-Assay könnte in Zukunft bei geeigneter Größe der Datensätze ein Übergang zur direkten Aktivitätsvorhersage potentieller Verbindungen übergegangen werden.

3.3.10 Netztraining der ANNs des Tat-TAR-Assay-spezifischen Filters

Für das Netztraining der Filter zum Selektieren von spezifischen Tat-TAR-Assay-spezifischen Ligandenvorschlägen wurde auf die im Rahmen dieser Arbeit erstellten Matlab-Skripte zurückgegriffen, die nach dem Import der Deskriptordatensätze in Matlab eine automatische Generation der trainierten Netze ermöglichen. Dabei werden folgende Punkte automatisch abgearbeitet:

- Verteilung der Datensätze auf zehn Kreuzvalidierungsläufe
- Automatische zufallsgesteuerte Zerlegung der Datensätze in Trainings- und Testdatensätze (wahlweise auch noch Abspalten eines Validationsdatensatzes)
- Generierung von ANNs mit unterschiedlicher Anzahl von Neuronen in der verborgenen Schicht (Neuronenanzahl = 2-10) (Hidden Layer)
- Training der erstellten Netze mittels der erstellten Datensätze
- Protokollierung der Trainingsergebnisse und Abspeicherung der ANNs

Eine Auswertung der so erstellten Trainingsläufe wurde ebenfalls mittels weiterer Matlab-Skripte ermöglicht. Diese automatisieren wieder folgende Abläufe:

- Extraktion der Statistiken für die einzelnen ANNs zu Gesamtstatistiken
- Extraktion der besten Netze der einzelnen Kreuzvalidierungsläufe zu den im Screening verwendeten Filter-Jurys, basierend auf dem kombinierten Abschneiden der Vorhersagen für die Trainings- und Testdatensätze
- Abspeichern der einzelnen Parameter der Jury-Netze für die Verwendung in den erstellten Java-Modulen

Ergebnisse:

Tabelle 15: MCCs der auf MOE-2D-Deskriptor basierten Tat-TAR-FRET-Assay-spezifischen ANN-Filter-Jurymitglieder für Trainings- und Testdatensatz

ANN	MCC	
	Training	MCC Test
1	0,8594	0,5294
2	0,8594	0,5852
3	0,773	0,5986
4	0,7053	0,5595
5	0,8244	0,642
6	0,728	0,5595
7	0,6587	0,6411
8	0,8387	0,6381
9	0,773	0,5852
10	0,773	0,6364

Tabelle 16: MCCs der auf CATS-2D-Deskriptor basierten Tat-TAR-FRET-Assay-spezifischen ANN-Filter-Jurymitglieder für Trainings- und Testdatensatz

ANN	MCC	
	Training	MCC Test
1	0,7062	0,6103
2	0,8159	0,6595
3	0,71	0,6178
4	0,7969	0,5341
5	0,7511	0,5551
6	0,611	0,6602
7	0,7062	0,6411
8	0,6835	0,6103
9	0,8387	0,5679
10	0,8594	0,5373

Tabelle 17: MCCs der auf TC3D-Deskriptor basierten Tat-TAR-FRET-Assay-spezifischen ANN-Filter-Jurymitglieder für Trainings- und Testdatensatz

ANN	MCC	
	Training	MCC Test
1	0,8994	0,6381
2	0,9203	0,5888
3	0,8387	0,6587
4	0,9001	0,5373
5	0,8594	0,611
6	0,8994	0,5595
7	0,9605	0,6835
8	0,9404	0,6364
9	0,8594	0,6171
10	0,7511	0,6348

Tabelle 18: MCCs der auf LIQUID-Deskriptor basierten Tat-TAR-FRET-Assay-spezifischen ANN-Filter-Jurymitglieder für Trainings- und Testdatensatz

ANN	MCC	
	Training	MCC Test
1	0,7969	0,5986
2	0,6587	0,6381
3	0,6875	0,586
4	0,6875	0,5852
5	0,5595	0,479
6	0,8594	0,5871
7	0,7969	0,5871
8	0,8419	0,483
9	0,7969	0,5736
10	0,7541	0,6411

Tabelle 19: ANN-Jury MCC Performance der Tat-TAR-FRET-Assay-spezifischen Filter, Vergleichsstatistik zu den unterschiedlichen Kreuzvalidationsläufen (10 Jurymitgliedern)

Jury-Performance	MCC	MCC
	Mittelwert	Standardabweichung
MOE Training	0,78	0,07
MOE Test	0,60	0,04
CATS Training	0,75	0,08
CATS Test	0,60	0,05
TC3D Training	0,88	0,06
TC3D Test	0,62	0,04
LIQUID Training	0,74	0,09
LIQUID Test	0,58	0,05

Diskussion:

Im Durchschnitt konnte ein MCC > 0,58 für die Vorhersagegüte der ANN-Filter-Jury für die verwendeten Trainings- und Testdatensätze erreicht werden; siehe hierzu Tabelle 19. Es ist somit anzunehmen, dass mithilfe der Filter eine relativ gute Abtrennung der aktiven von den inaktiven Substanzen für diesen FRET-Assay möglich sein sollte. Da aufgrund der immer noch relativ geringen Datensatzgröße (306 Substanzen) jedoch das Verhältnis zur Anzahl der verwendeten Deskriptoren immer noch als sehr klein für ein ANN-Training anzusehen ist (Anzahl Strukturen / Anzahl Deskriptoren $\approx 0,3 - 1$), kann erst die Überprüfung mit einem unabhängigen Datensatz eine abschließende Bewertung der Vorhersagegüte erlauben. Aufgrund der kleinen Datensatzgröße wurde jedoch auf die zusätzliche Abtrennung eines Validationsdatensatzes im Rahmen dieser Studie für die zur Vorhersage bestimmten ANN-Jury-Filter verzichtet.

3.3.11 Screening der Specs mit den erhaltenen Jury-Filtern

Die vier erhaltenen Tat-TAR-FRET-Assay-spezifischen Jury-Filter für CATS-2D-, TC3D-, MOE-2D- und LIQUID-Deskriptoren wurden dazu genutzt, um die Chemikalien-Datenbanken des Vertriebshändlers Specs nach potentiellen TAR-Liganden zu durchsuchen. Hierzu wurden die bereits in Abschnitt 3.3.5 erstellten Datensätze mit den Filtern bewertet und die Substanzen nach ihren Scores sortiert und mittels MOE 2006.08 (Chemical Computing Group, Montreal) Ausdrücke mit 2D-Ansichten der Strukturen für die manuelle Auswahl vorbereitet.

3.3.12 Betrachtung der Auswahllisten

Für die einzelnen Filtertypen wurden jeweils die Top20 der sortierten Einzellisten für die manuelle Auswahl ausgewählt.

Ausschluss des MOE-2D-Deskriptor-basierten Jury-Filter

Bei der Betrachtung der „Top 20“ des MOE-2D-Deskriptor-basierten Filters zeigte sich, dass dieser offensichtlich eine Präferenz für Fluor-Atome entwickelt hatte und dies als Hauptdiskriminierungsmerkmal zu verwenden schien. Keine der in diesen „Top 20“ enthaltenen Strukturen beinhaltet weniger als sechs Fluoratome, wobei die Mehrzahl von 19 Strukturen, vereinfacht gesagt, fluoridierte Kohlenstoffketten zwischen 4 und 8 Atomen Länge als Strukturmotiv aufweist. Zusätzlich weisen die Strukturen zwar normale Ringstrukturen ohne Fluoratome auf, deren Aufbau und Heteroatomanteil (Sauerstoff, Stickstoff und Schwefel) denen der Trainingsdaten entsprechen, jedoch wurde die besagte Präferenz für besagtes Artefakt – der genannten Polyfluorierung – als Ausschlusskriterium gewertet. Abschließend erklären lässt sich bis heute dieser „Fehler“ nicht, da in den Trainingsdaten zwei fluoridierte aktive Verbindungen neun fluoridierten inaktiven Verbindungen gegenüber standen, die auch vom Filtersystem jeweils korrekt klassifiziert wurden. Ein einfaches Vertauschen der Datensätze kann aufgrund der Scorewerte der ANN-Filter ausgeschlossen werden.

Kombination der CATS-2D-, TC3D- und LIQUID-Hitlisten zu einem gemeinsamen Triple-Filter

Hintergrund:

Da in der ersten Bestellrunde mit der Kombination der Hitlisten – einzelner auf verschiedene Deskriptor-Klassen basierender Jury-Filter – zu einer gemeinsamen Überlappungsliste die besten Ergebnisse erzielt wurden (siehe Abschnitt 3.2.18), sollte dieser Ansatz in der Tat-TAR-FRET-Assay-spezifischen Auswahlrunde ebenfalls umgesetzt werden. Nach Ausschluss der MOE-2D-Deskriptor basierten Jury verblieben noch drei Jury-Filter, deren als plausibel erscheinende verbleibende Filter-Hitlisten, wurden zu einem kombinierten Filterergebnis zusammengeführt.

Durchführung:

Die Erstellung der gemeinsamen Ranking-Liste wurde wie folgt umgesetzt.

- Mittels MOE wurden die drei Hitlisten zusammengeführt, hierbei wurde die „Merge“-Funktion von MOE eingesetzt um die Überlappungen der besagten drei Filter-Hitlisten zu extrahieren.
- Um ein Ranking der überlappenden Hitliste durchführen zu können, wurden die Einzelscores der betreffenden drei Filter gemittelt.
- Die „Top 100“ dieser gemeinsamen Hitliste wurden dann ebenfalls als Grundlage für die manuelle Auswahl verwendet.

3.3.13 „Cherry-Picking“ der potentiell TAR-RNA-spezifischen Substanzen zum Testen

Aus den Auswahllisten wurden mittels „Cherry-Picking“ Strukturen ausgewählt, die bei Specs zur weiteren Testung in den Laborassays bestellt wurden.

Hierbei wurde folgende Auswahl getroffen:

- TC3D Top20 → 8 Strukturen (2 mögliche Peptidbinder; auf Rang 8 und 13)
- CATS Top20 → 12 Strukturen (1 Aktiver; auf Rang 17)
- LIQUID Top20 → 8 Strukturen (kein Aktiver)

- Gemittelter TC3D-CATS-LIQUID-Score Top100 → 15 Strukturen (kein Aktiver)

Anmerkung:

Hinter den Auswahlanteilen wurden auch gleich vorwegnehmend die Ergebnisse der Untersuchungen mittels Tat-TAR-FRET-Assay, angegeben. Dies erfolgte analog zur Rückführung der Aktiven auf die ANN-Filter, wie in Abschnitt 3.2.18 für den allgemeineren RNA-Binder-Fall. Der einzige „echte“ Treffer wurde somit mittels des CATS-ANN-Filters erhalten. Aufgrund des hohen Anteils (ca. 73%) an unlöslichen/„nicht messbaren“ Verbindungen konnte hier jedoch die Rückführungsanalyse nicht in derselben Qualität wie bei der ersten Vorhersagerunde durchgeführt werden.

3.3.14 Test der Substanzen im Tat-TAR-FRET-Assay

Einleitung:

Insgesamt konnten 42 Substanzen von Specs geliefert werden, 41 davon wurden im Rahmen dieser Arbeit vermessen. Eine Substanz stellte eine Überlappung zu den Bestellungen von Andreas Schüller dar, so dass diese Substanz im Rahmen der Tests zu seiner Doktorarbeit von ihm vermessen wurde.

Durchführung:

Die Substanzen wurden, wie im Abschnitt 2.7 „Der Tat-TAR-FRET-Assay“ behandelt, im Labor getestet. Jedoch zeigten sich Löslichkeitsprobleme für die Mehrzahl der ausgewählten Substanzen, so dass eine sinnvolle Vermessung aller ausgewählten Substanzen nicht möglich war; siehe hierzu den Abschnitt 3.3.15.

3.3.15 Löslichkeitsprobleme

Hintergrund:

Im Rahmen der Arbeit im Labor bei den Messungen zu Yusuf Tanrikulus Vorhersagen (Tanrikulu et al. 2007) waren Löslichkeitsprobleme aufgetreten, daher wurde dieses Mal ein besonderer Wert auf die Beobachtungen der Verdünnungsreihen gelegt.

Die Ergebnisse dieses Abschnitts wurden neben der eigentlichen Durchführung des Tat-TAR-FRET-Assay erhoben. Die Löslichkeitsbeobachtungen beziehen sich auf die bei den Verdünnungsschritten beobachteten Löslichkeitsproblemen der Testlösungen in den Vorratsgefäßen.

Die hier beschriebenen Verdünnungsschritte beinhalten die Verdünnung von der 10 mM Stammlösung in DMSO, zur 1:10 verdünnten FRET-Assay-Stammlösung mit 1 mM Testsubstanzkonzentrationen in einer 90-volumenprozentigen wässrigen Testlösung, wie sie im FRET-Assay eingesetzt wird, um auf der Mikrotiterplatte letztlich eine Konzentration von 100 µM im Assay zu erreichen. Als Alternative wurde die erste Verdünnung ebenfalls mit Methanol erprobt, da der FRET-Assay auch mit einer Methanol-Konzentration < 10% noch verwendet werden kann (pers. Komm. Dr. Ute Scheffer).

Wie sich zeigen sollte, enthielten die Laboraufzeichnungen zu den Löslichkeitsbeobachtungen für die Testsubstanzen bereits die Informationen (Ausschlusskriterien), die zur begründeten Reduktion der Anzahl der testbaren Substanzen nötig waren. Hierbei zeigte sich, dass Substanzen mit einer offensichtlichen Auffälligkeit (augenscheinliche Ausfällungen) beim ersten 1:10 Verdünnungsschritt mit Wasser, ebenfalls keine zuverlässigen Messergebnisse in den Messreihen mit Methanol mehr liefern konnten oder nach der Überprüfung der Absorptionsspektren als unlöslich (Trübung der Lösung) aussortiert werden mussten. Die temporäre Löslichkeit in den Methanol-Lösungen konnte somit das generelle Löslichkeitsproblem der LR-Reihe zwar für den Anfang der Messungen beheben, jedoch fielen die Substanzen bei weiterer Verdünnung in den Vertiefungen der

Mikrotiterplatte wieder aus, so das dieses Vorgehen nicht für die weitere Umsetzung empfohlen werden kann.

Durchführung:

Die Durchführung der Verdünnungsreihen erfolgte wie unter Abschnitt 2.7.2 beschrieben, jedoch wurden zusätzlich Tests mit Methanol-verdünnten Stammlösungen durchgeführt.

Auf Anregung von Andreas Schüller und Dr. Ute Scheffer wurde versucht die Substanzen in Methanol zur Lösung zu bringen, was wie erwartet zu einer deutlichen Steigerung der beobachteten Löslichkeit führte.

Hierzu wurden die Stammlösungen 1:10 mit Methanol statt mit Wasser verdünnt.

Für alle Verdünnungsschritte wurden die Verdünnungsbeobachtungen und Stabilitätsbetrachtungen der erstellten Lösungen protokolliert. Diese umfassten folgende Beobachtungspunkte:

- Die gelieferten DMSO-Stammlösungen nach dem Auftauen in den Original-Gefäßen
- Die gelieferten DMSO-Stammlösungen nach Überführung in Tubes und kurzer Zentrifugation
- Direkte Mischungsbeobachtungen 1:10 Verdünnungen der DMSO-Stammlösungen mit Wasser oder Methanol
- 1:10 Verdünnungen der DMSO-Stammlösungen (Wasser oder Methanol) nach kurzer Zentrifugation
- Beobachtungen der 1:10 verdünnten Lösungen nach Lagerung bei 4°C.
- Beobachtungen der Lösungen in den Küvetten für die Absorptionsspektren (Puffer- und Testsubstanzkonzentration wie im FRET-Assay)

Anmerkung zu weitergehenden Untersuchungen:

Nach dem Verdünnen der DMSO-Stammlösungen der Substanzen mit Methanol konnten so immerhin 38 der 41 Testsubstanzen in eine „stabile Lösung“⁶ gebracht werden. Mit diesen Verdünnungslösungen wurde ebenfalls eine Dreifachmessungs-Reihe der zur erzielenden Reduktion der Fluoreszenzaktivität durchgeführt, siehe Abschnitt 3.3.16 Abbildung 72.

Ergebnisse:

In Tabelle 20 angegeben, sind die Löslichkeitsbeobachtungen nach Lagerung der Substanzen bei 4°C für 3 Tage. Der Großteil der beobachteten Ausfällungserscheinungen (Flocken, Pellets, Kristallnadeln, Trübung) konnte auch bereits bei den entsprechenden Verdünnungsschritten beobachtet werden.

⁶ Stabile Lösung = eine Lösung der Testverbindungen ohne erkennbare Löslichkeitsprobleme in einer 1 mM Lösung der Testsubstanzen nach einer 1:10 Wasser- oder Methanol-Verdünnung der 10 mM DMSO-Stammlösungen. Beobachtung erfolgte bei Raumtemperatur, Lagerung bei 4°C

3. Liganden-basierte Ansätze

Tabelle 20: Löslichkeitsbeobachtungen bei den Verdünnungen der LR-Reihe. Alle Lösungen enthalten 10% DMSO und 90% Wasser oder Methanol. Beobachtet wurden die Lösungen nach Lagerung bei 4°C. Die jeweilige Verdünnungsstabilität wurde umschrieben, wie sie in den Vorratsgefäßen (Tube) zu sehen war. klar oder ok = deuten stabile Lösungen an, anderslautende Beschreibungen verdeutlichen die augenscheinlichen Löslichkeitsprobleme.

LabID = LR + X	Wasser Verdünnungsstabilität	Methanol Verdünnungsstabilität
1	klar	ok
2	klar	ok
3	Flocken und Beschichtung der Oberfläche des Tube	ok
4	klar	ok
5	kleine Flocken	ok
6	kleine Kristallnadeln beginnend bei Tubeoberfläche plus Pellet	ok
7	Stäbchen, Flusen	ok
8	Flocken	ok
9	klar	ok
10	Plättchen Pellet	ok
11	klar	ok
12	Flocken groß	ok
13	klare gelbe Lösung. mit rotem Niederschlag	ok
14	Flocken Matrix	ok
15	schwarze kleine Punkte Pellet	ok
16	Flockiges Pellet	ok
17	flocken Pellet	ok
18	ölige Phase am Tubeboden	ok
19	klar	ok
20	klar	ok
21	kompaktes Pellet	ok
22	goldgelbes Pellet, Flocken, schmieriger Film auf Oberfläche klare Lösung	leichtes gelborange Pellet am Boden
23	große Flocken lose gepackt	ok
24	klar	ok
25	Pellet Flocken	ok
26	Flockiges Pellet	Weißer Niederschlag 5 mm hoch
27	Pellet gelb klein flockig	ok
28	klar	ok
29	kompaktes Pellet	ok
30	klar	ok
31	gelbe Flocken Pellet klare farblose Lösung großflächige Kristallbeschichtung der Tubeoberfläche	gelbe Flocken Matrix unübersehbar
32	Kokonflocken	ok
33	kleine Kristallnadeln beginnend bei Tubeoberfläche	ok
34	klar	ok
35	kleine Kristallnadeln beginnend bei Tubeoberfläche	ok
36	klar, aber roter Faden drin	ok
37	Schleimfaden	ok
38	drei kleine Kristalle „knubbelig“ fest an Tubewand	ok
39	kleiner Niederschlag	ok
40	starker gelber Niederschlag	ok
41		

Diskussion:

Die Stabilität der Lösungen, d.h. sie zeigen keine Entmischungerscheinungen, konnte durch die Anwendung von Methanol als Lösungsmittel deutlich erhöht werden (von 11 auf 38 gelöste Substanzen). Jedoch ergab die Analyse der Absorptionsspektren, dass sich unter den – aus Methanol-verdünnten Lösungen stammenden – ermittelten, potentiell aktiven Verbindungen für diesen Assay sich wahrscheinlich auch Substanzen befanden, die in den Vertiefungen der Mikrotiterplatte ausgefallen sind. Die erneute 1:10 Verdünnung – diesmal mit wässriger Lösung – verursachte dementsprechend die beobachtete Ausfällungsreaktion aus den Wasserverdünnungsschritten, jedoch dieses Mal zunächst verdunkelt in den Vertiefungen der Mikrotiterplatte (schwarze Mikrotiterplatte).

Schlussfolgerung:

In zukünftigen Bestellrunden sollte ein deutlicher Beitrag der Vorhersagen auf die genaue Vorhersage der Löslichkeit als weiteres Auswahlkriterium gelegt werden. Da es hierzu jedoch zurzeit noch keine automatisierten Verfahren gibt, die eine ausreichende Vorhersagegüte erreichen, sollte eventuell das Wissen von praktisch arbeitenden Chemikern für die Überprüfung der Auswahllisten in die manuelle Auswahl mit einbezogen werden.

Ferner kann die Analyse der Absorptionsspektren vor Durchführung des Tat-TAR-FRET-Assays helfen, Zweifelsfälle ungenau bestimmter Löslichkeit der Testsubstanzen, die noch nicht in den Wasserverdünnungen der DMSO-Stammlösungen auffällig wurden, vorzeitig aus den weiteren Tests als negativ (d. h. nicht genauer bestimmbar) auszuschließen.

3.3.16 Ergebnisse der Tat-TAR-FRET-Assay-Messungen für den Assay-spezifischen Filter

Hintergrund:

Die hier gezeigten Ergebnisse dienen im Rahmen dieser Arbeit zum ersten Screening der Testsubstanzen auf Aktivität im Assay.

Die hier vorgestellten Ergebnisse beruhen auf dem Versuchsablauf, wie er in Abschnitt 2.7 detaillierter geschildert wurde. Die angegebenen Aktivitätswerte beziehen sich auf mittels „range scaling“ transformierten Fluoreszenz-Aktivitätsdaten. Ein Wert von 1 steht für dieselbe Fluoreszenzintensität wie sie in den Kontrollvertiefungen auf der Mikrotiterplatte für die Tat-TAR-Lösungen ohne Zusatz einer Testverbindung ermittelt wurde. Ein Wert von 0 entspräche der Fluoreszenzintensität bei freiem Tat-Protein ohne TAR in der Lösung (vollständige Dissoziation des Tat-TAR-Komplexes), ebenfalls durch Kontrollmessungen in den Vertiefungen der Mikrotiterplatte. Die Pufferkonzentrationen und Tat- oder TAR-Konzentrationen entsprechen denen der Tests mit den Testsubstanzen, ebenso die DMSO oder Methanolkonzentrationen.

Durchführung:

Die Ergebnisse der Messungen am „Safire²“-Spektrometer wurden für die einzelnen Messungen nach den enthaltenen Testsubstanzen zusammengeführt. Anhand der ebenfalls bestimmten Referenzpunkte für die Tat-Lösungen oder Tat-TAR-Lösungen wurden die relativen Aktivitätswerte für die Testsubstanzen mittels „Range-Scaling“ transformiert, um einzelne Messreihen mit unterschiedlichen Lösungs-Chargen vergleichen zu können. Die Versuche wurden dreifach parallel in unterschiedlichen Vertiefungen auf der Mikrotiterplatte durchgeführt. Sollten die erzielten Messergebnisse dieser zeitgleich durchgeführten Messungen stark schwanken, so wurden für diese Substanz weitere Testreihen angesetzt, ebenso wenn die Messwerte eine Aktivität implizierten (relative Fluoreszenzaktivität < 0,9). Die Messungsanzahl für jede Substanz beträgt somit mindestens drei Messungen.

Ergebnisse:

Die Abbildung 71 und Abbildung 72 zeigen die Auftragungen zu den Ergebnissen aus Tabelle 21 und Tabelle 22.

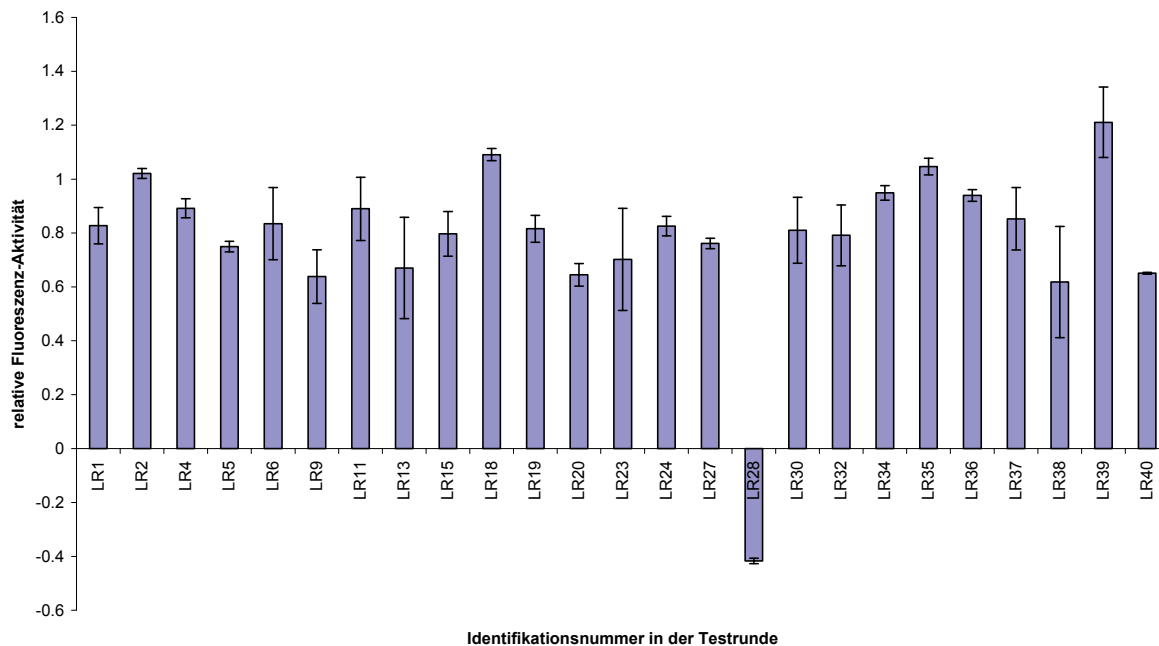


Abbildung 71: Testreihe der potentiellen Tat-TAR-FRET-spezifischer Liganden (in H₂O / DMSO)

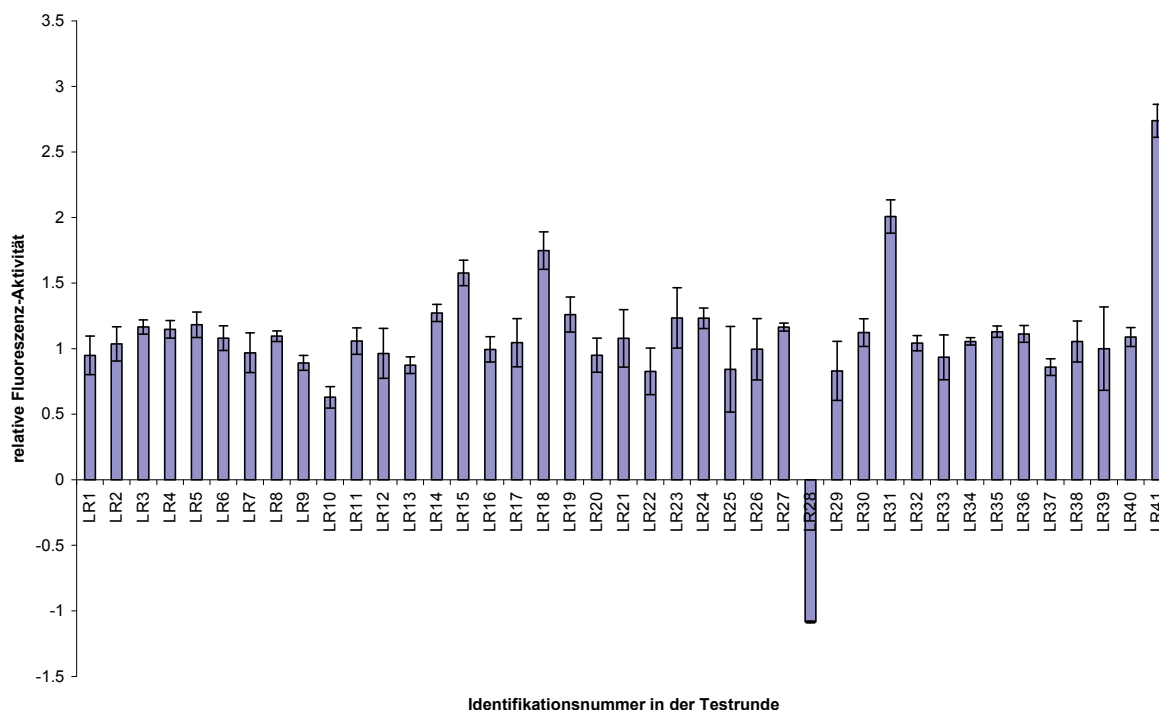


Abbildung 72: Testreihe der potentiellen Tat-TAR-FRET-spezifischer Liganden (in H₂O / Methanol)

Tabelle 21: Ergebnisse der Messungen für die potentiellen Tat-TAR-FRET-Assay-spezifischen RNA-Liganden-Filter als wasserverdünnte Testreihe

	Mittelwert RFA	Stdabw. RFA
Ref: Tat TAR	1,00	0,07
Ref: TAT	0,00	0,02
LR1	0,83	0,07
LR2	1,02	0,02
LR4	0,89	0,04
LR5	0,75	0,02
LR6	0,83	0,13
LR9	0,64	0,10
LR11	0,89	0,12
LR13	0,67	0,19
LR15	0,80	0,08
LR18	1,09	0,02
LR19	0,82	0,05
LR20	0,65	0,04
LR23	0,70	0,19
LR24	0,83	0,04
LR27	0,76	0,02
LR28	-0,42	0,01
LR30	0,81	0,12
LR32	0,79	0,11
LR34	0,95	0,03
LR35	1,05	0,03
LR36	0,94	0,02
LR37	0,85	0,12
LR38	0,62	0,21
LR39	1,21	0,13
LR40	0,65	0,00

Tabelle 22: Ergebnisse der Messungen für die potentiellen Tat-TAR-FRET-Assay-spezifischen RNA-Liganden-Filter als Testreihe mit Methanol-verdünnten Stammlösungen

	Mittelwert RFA	Stdabw. RFA
Ref: Tat TAR	1,00	0,17
Ref: TAT	0,00	0,11
LR1	0,95	0,15
LR2	1,04	0,13
LR3	1,16	0,05
LR4	1,15	0,07
LR5	1,18	0,10
LR6	1,08	0,09
LR7	0,97	0,15
LR8	1,10	0,04
LR9	0,89	0,06
LR10	0,63	0,08
LR11	1,06	0,10
LR12	0,96	0,19
LR13	0,87	0,06
LR14	1,27	0,07
LR15	1,58	0,10
LR16	0,99	0,10
Ref: Tat TAR	1,00	0,25
Ref: TAT	0,00	0,14
LR17	1,05	0,18
LR18	1,75	0,14
LR19	1,26	0,13
LR20	0,95	0,13
LR21	1,08	0,22
LR22	0,83	0,18
Ref: Tat TAR	1,00	0,21
Ref: TAT	0,00	0,18
LR23	1,23	0,23
LR24	1,23	0,08
LR25	0,84	0,33
LR26	1,00	0,23
LR27	1,16	0,03
LR28	-1,08	0,01
LR29	0,83	0,23
LR30	1,12	0,11
LR31	2,01	0,13
LR32	1,04	0,06
LR33	0,93	0,17
LR34	1,05	0,03
LR35	1,13	0,04
LR36	1,11	0,06
LR37	0,86	0,06
LR38	1,05	0,16
LR39	1,00	0,32
LR40	1,09	0,07
LR41	2,74	0,13

Diskussion:

Deutlich zu erkennen, ist die unterschiedliche Anzahl an getesteten Verbindungen der Methanol- und Wasser-verdünnten Testreihe (38 zu 25 Substanzen). Die 25 Substanzen in der Wasser-verdünnten Testreihe kamen jedoch nur zustande indem alle vorhandenen klaren Lösungen – inklusive klarer Überstände der Verdünnungsreihen – hierzu verwendet wurden.

Besonders markant stechen die Messungen zu LR28 aus den beiden Messreihen heraus, da diese Verbindung den relativen Fluoreszenzwert ins Negative drückt und somit mit dem Peptid zu interagieren scheint. Diese Beobachtung konnte in der Interaktion mit dem Tat-Peptid alleine reproduziert werden, so dass keine TAR für diese Wirkung vorhanden sein muss. Basierend auf den Messergebnissen und Löslichkeitsbeobachtungen konnten nur noch sechs Verbindungen für die weitere Überprüfung der potentiellen Aktivität abgeleitet werden: LR9, LR10, LR13, LR16, LR20 und LR28.

3.3.17 Messungen zur Überprüfung der potenziell aktiven Verbindungen

Aus den ersten Screening-Ergebnissen ließen sich zwar eine Reihe potenzieller RNA-Liganden ableiten, diese wurden aber in anschließenden Untersuchungen zum größten Teil wieder verworfen, weil eine eingehende Betrachtung des Löslichkeitsverhaltens der Kandidaten eine weiterführende Überprüfung der Aktivität ausschloss.

Die gewonnene Auswahl an Substanzen (LR9, LR10, LR13, LR16, LR20, LR28), welche zuvor aus der Kombination der Screening-Ergebnisse der beiden Screeningrunden (wasser- und methanolverdünnt) anhand der Reduktion der relativen Fluoreszenzaktivität $> 0,2$ entstand, wurde ferner durch die Überprüfung der Absorptionsspektren weiter reduziert. Für eine Erklärung der Erstellung der Absorptionsspektren siehe Abschnitt 2.7 Seite 18.

Hierbei zeigte sich, dass die in den mit Methanol-Verdünnungen durchgeführten Messungen gewonnenen Aktivitäten der Substanzen LR10, LR13 und LR16 aufgrund der beobachteten Absorptionsspektren wohl durch ihr Absorptionsverhalten die Aktivität in diesem Assay verursacht haben. Kombiniert mit den bereits erwähnten Beobachtungen der Löslichkeit während der ersten 1:10 Wasserverdünnung ergibt sich, dass diese Substanzen in den Vertiefungen der Mikrotiterplatte wieder ausfallen und durch die verteilten Schwebstoffe eine Trübung der Lösung verursachen, die wiederum die Verminderung der Fluoreszenzaktivität verursacht; siehe Tabelle 23, für eine Beschreibung aller Löslichkeitsbeschreibungen siehe Tabelle 20.

Tabelle 23: Ausschnitt der Löslichkeitsbeobachtungen zur LR-Reihe beim Erstellen der Verdünnungen mit Wasser

LabID	Löslichkeitsbeobachtungen
LR9	klar
LR10	Plättchen/Pellet
LR13	klare gelbe Lösung mit rotem Niederschlag
LR16	flockiger Niederschlag/Pellet
LR20	klar
LR28	klar

3.3.18 Verbleibende aktive Verbindungen der LR-Reihe aus dem Tat-TAR-FRET-Assay

Hintergrund:

Da durch die vorhergehenden Untersuchungen sukzessive die Anzahl an potentiellen RNA-Liganden immer weiter reduziert werden konnte, musste für die noch verbleibenden drei Verbindungen geklärt werden, ob die beobachtete Aktivität im Tat-TAR-FRET-Assay auf eine Wechselwirkung mit dem Peptid beruhen könnte. Als abschließende Untersuchung wurde getestet, ob die Substanzen mit dem Peptid direkt wechselwirken können.

Durchführung:

Hierzu wurden die relativen Fluoreszenzwerte der Verbindung in einer Lösung mit dem Peptid – ohne zugesetzte RNA – gegen eine solche Lösung ohne Zugabe der Substanzen getestet.

Ergebnis:

- LR9 zeigt hierbei keine Veränderung der gemessenen Fluoreszenzintensität,
- LR28 zeigt eine deutliche Verringerung der Fluoreszenzintensität,
- LR20 zeigt eine leichte Verringerung der Fluoreszenzintensität.

Diskussion:

Aus dem Ergebnis der Interaktion mit den Tat-Peptid-Lösungen ohne TAR wird abgeleitet, dass die beobachtete Wirkung der Substanzen LR20 und LR28 primär durch eine direkte Wechselwirkung mit dem Peptid verursacht wird. Dahingegen wird von LR9 angenommen, dass die erzielte Wirkung auf einer Interaktion mit der RNA beruht, da hier keine Auffälligkeiten zu beobachten waren.

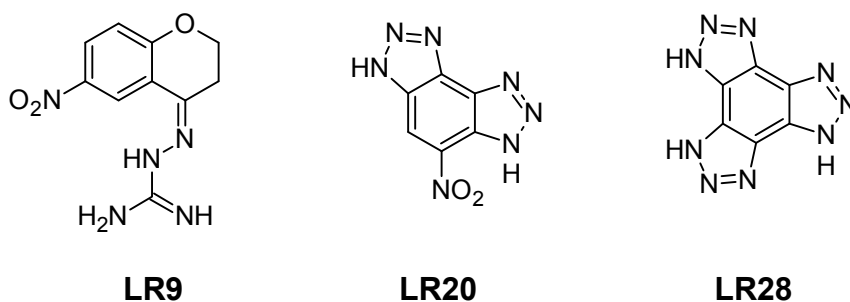


Abbildung 73: Aktive im Tat-TAR-FRET-Assay; LR9 bindet möglicherweise an der RNA, LR20 und LR28 binden möglicherweise am Peptid

Tabelle 24: relative Fluoreszenzaktivitäten im Tat-TAR-FRET-Assay für die Messungen der Aktiven aus der LR-Reihe

LabID	relative Fluoreszenzaktivität
LR9	0,64 ±0,10
LR20	0,65 ±0,04
LR28	-0,42 ±0,01

3.3.19 Methylenaminoguanidin – LR9, AS20 und M23 – Vergleich der Substruktur motive

Hintergrund:

Da im Rahmen dieser Arbeit bereits bei der Auswertung der RNA-Liganden-Binder-Filter (siehe Abschnitt 3.2.17), basierend auf Vergleichen der potentiell aktiven RNA-Liganden dieser alten Testrunden, Interaktions-Muster/Pharmakophormodelle abgeleitet werden konnten, wurde diese Betrachtungsweise auch für diese Testreihe angestrebt. Da hierbei jedoch nur LR9 als potentieller RNA-Ligand in Frage kam, wurde hier ebenso ein Vergleich mit den Aktiven der vorherigen Testreihen durchgeführt. Auffällig war hierbei die Ähnlichkeit der einzigen in der AS-Reihe als aktiv gefundenen AS20 mit dem neuen LR9. Bei genauerer Betrachtung konnte später die Verbindung M23 ebenfalls der Klasse mit diesem Substruktur-Muster hinzugefügt werden.

Durchführung:

Die Strukturen wurden gezeichnet und die Darstellungen manuell paarweise miteinander verglichen.

Ergebnis:

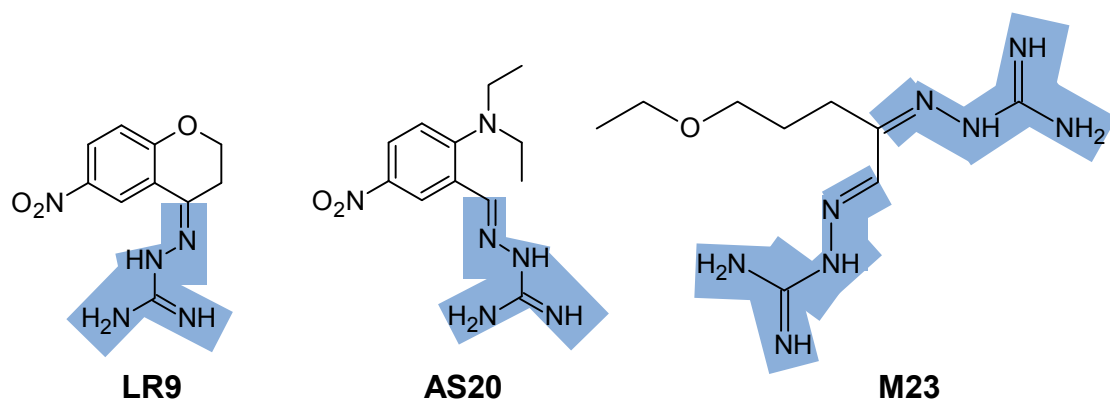


Abbildung 74: Gegenüberstellung der drei aktiven LR9, AS20 und M23 des Tat-TAR-FRET-Assay; blau hinterlegt sind die erweiterten Guanidin-Gruppen (Methylenaminoguanidin) in den Molekülen; unter physiologischen Bedingungen sind diese Gruppen protoniert. M23 trägt somit zusätzlich zwei positive Ladungen und LR9 und AS20 jeweils eine positive Ladung.

Diskussion:

Die in der zweiten Screening-Runde gefundene Substanz LR9 und die in der ersten Screening-Runde gefundenen Substanzen AS20 und M23 weisen alle ein Guanidin-Substrukturmotiv (Methylenaminoguanidin) auf, welches unter physiologischen Bedingungen protoniert vorliegen würde; siehe Abbildung 74. Guanidin kommt als Substrukturmotiv auch im Aufbau des Tat-Peptids sechsmal als Arginin vor. Das führt zu der Annahme, dass elektrostatische Wechselwirkungen bei der Bindung dieser Substanzen eine Hauptrolle spielen, welche die Interaktionen der Arginine des Peptids imitieren und somit die beobachtete Wirkung (Verdrängung des Peptids) im Assay verursachen. Eine Ähnlichkeit zum entwickelten Tetra-aza-adamantyl-Pharmakophormodell ist auch hier zu erkennen: Die Hauptgemeinsamkeit besteht in einer positiven Ladung und ungesättigten Bindungen.

4. Zielstrukturbasierte Ansätze

Die folgenden Abschnitte beziehen sich auf Zielstrukturbasierte Ansätze, die im Rahmen dieser Arbeit umgesetzt wurden. Prinzipiell gehört auch der Abschnitt 3.2.9 – „Dockingexperimente mit den ausgewählten Substanzen in einen Konformer des PDB-Eintrags 1LVJ“ auf Seite 41 in diesen Kontext, da die Dockingversuche aber auch Teil der Auswahl der Screening-Runden des ersten ligandenbasierten Ansatzes waren, wurden sie bereits dort geschildert. Die Erfahrungen aus den Docking-Experimenten führten zur weiteren Beschäftigung mit der Fragestellung repräsentative Konformere für Dockingverfahren auszuwählen. In den folgenden Abschnitten dreht es sich somit hauptsächlich um die Frage, wie man im Falle einer als flexibel anzusehenden Zielstruktur, deren Konformerraum bestmöglich für das Liganden-Design zugänglich machen kann, um begründbar die verwendete Menge an Referenzkonformeren senken zu können.

4.1 Große Konformationsänderungen möglich

Als Verdeutlichung der Flexibilität der TAR-RNA wurde der PDB-Eintrag 1ANR mit 20 Frames auf die Relativbewegungen seiner Atome von Frame zu Frame überprüft. Hierbei ergab sich für die Atome H21 des Residue 33 und HO3' des Residue 45 ein Maximum zwischen den Entfernungen der Frames 2 und 15. Dabei bewegte sich dieses Atoppaar von einer maximalen Distanz von 59,21 Ångström insgesamt 25,67 Ångström auf ein Minimum von nur noch 33,54 Ångström zu; siehe Abbildung 75.

Dies bedeutet eine relative Reduktion der Distanz um ca. 43 Prozent. Eine derartige Umorganisation der Konformation verbietet somit die numerische Erfassung einer bestimmten Konformation als alleinige Referenz für das Liganden-Design.

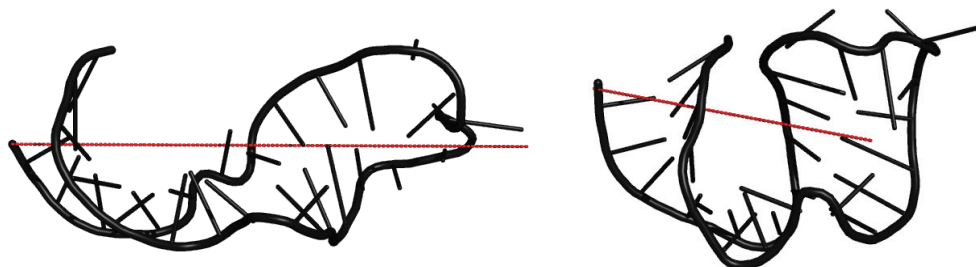


Abbildung 75: TAR-Konformer: starke Umorganisation der Raumstruktur zwischen einzelnen Modellen des PDB-Eintrags 1ANR, Entfernung zwischen den Atomen H21 des Nukleotids 33 und HO3' des Nukleotids 45 als rote Linie wiedergegeben. Die relative Verkürzung des Abstands der beiden Atome zwischen Modell 7 und 18 beträgt 25,67 Ångström.

Auswahl der „Referenzstruktur“ – Welche Struktur soll man also nehmen?

Sollte keine Populationsgewichtung möglich sein, so muss jede der Konformere als gleich wahrscheinlich und somit gleich legitim für die Wirkstoffentwicklung betrachtet werden. Problematisch ist hierbei jedoch die Tatsache, dass die Fitness eines Liganden von seiner Positionierung vor der Bewertung mit den Scoringfunktionen abhängt und die Güte mit der ein Ligand platziert werden kann, wiederum von der Kombination der Form des Liganden mit der Zielstruktur abhängt. Anders ausgedrückt kann es dazu kommen, dass die vermeintlich besten Liganden nur in diesen einen Konformer passen, der aber nur selten anzutreffen ist und nicht durch den Liganden induziert werden kann, während experimentell besser abschneidende Liganden in diesen einen Konformer gar nicht hinein passen und somit als falsch negative Kandidaten verworfen werden. Diese Art des „Glücksspiels“ könnte verglichen werden mit der „Suche nach der Nadel im Heuhaufen, während dieser

durchgeschüttelt wird, ohne Wissen über die konkrete Form der Nadel“. Auch hier gibt es keine explizite Lösung, sondern nur eine Wahrscheinlichkeit die Nadel an einem bestimmten Punkt zu finden.

Nutzung von Wahrscheinlichkeiten

Als geeignete Wahrscheinlichkeitsverteilung ergibt sich aus den Überlegungen zur Populationsverteilung der Konformere, der Einsatz dieser Verteilungen zur Gewichtung der als geeignet erscheinender Konformere, anhand der Auftrittswahrscheinlichkeit innerhalb des möglichen Konformerraums. Hierzu sollte eine populationsbasierte Erfassung der Auftrittswahrscheinlichkeiten der Konformere durch Clusterung ähnlicher Strukturen ein quantifizierbares Kriterium bei der Auswahl der wahrscheinlichsten Konformere ermöglichen. Für den Erfolg dieses Ansatzes bedarf es der automatischen Bewertung der einzelnen Konformere der Datenbank aufgrund ihrer relativen Ähnlichkeit zu allen anderen in der Datenbank vorhandenen Konformeren. Dieser Ansatz wurde noch nicht implementiert, da zurzeit noch kein geeignetes Kriterium für diese Bewertung verfügbar ist. Ein erster Schritt könnte hier die Umsetzung der RMSD-basierten Clusterung der Datenbank sein. Jedoch könnte – da es im Rahmen dieser Arbeit hauptsächlich um die Analyse der potentiellen Bindestellen geht – die Clusterung basierend auf der Verteilung ähnlicher Taschen auf verschiedenen Konformeren umgesetzt werden. Die Beschreibung der Taschen befindet sich aber momentan noch in der Prototyp-Phase, so dass eine „sinnvolle“ Umsetzung erst die Bewältigung der erfolgreichen Beschreibung der Ähnlichkeit der Taschen voraussetzt. Deshalb beschränkt sich im Rahmen der Arbeit der Fokus auf die Aufklärung der möglichen Beschreibungen der Taschen.

4.1.1 Klassische Docking-Probleme oder welche Struktur nimmt man

Die im Abschnitt 4.1 geschilderte Unbestimmtheit einer klar definierten Zielstruktur erstreckt sich nicht nur auf das Problem, eine Referenzstruktur innerhalb des möglichen Konformerraums der freien TAR auszuwählen, sondern auch auf die Auswahl in bereits eingeschränkten Konformerensembles, wie der durch Ligandenbindung in seiner konformellen Freiheit eingeschränkteren TAR-Struktur 1LVJ. Diese strukturelle Unbestimmtheit ist in den Abbildung 76 dargestellt und zeigt die Überlagerung des Konformerensembles, Letzteres lässt erahnen, dass eine möglichst exakte numerische Lösung basierend auf einem Konformer unweigerlich zu abweichenden Bewertungen einzelner Liganden führen würde. Weiter verdeutlicht wird dieser Sachverhalt durch die in Abbildung 77 bis Abbildung 79 gezeigten Überlagerungen der Ensemble-Mitglieder der PDB-Einträge 1ANR und 1ARJ, welche im Fall von 1ARJ, ebenso bedingt durch einen Liganden relativ geringe Abweichungen der Ensemblemitglieder voneinander aufzeigen, jedoch ebenso wie 1LVJ „verrauschte“ Koordinaten enthalten. Besonders groß ist die zu beobachtende Vielfalt an deutlich unterschiedlichen Konformeren bei 1ANR; deren Modell ohne Ligand strukturaufgelöst wurde TAR-RNA („freie TAR“).

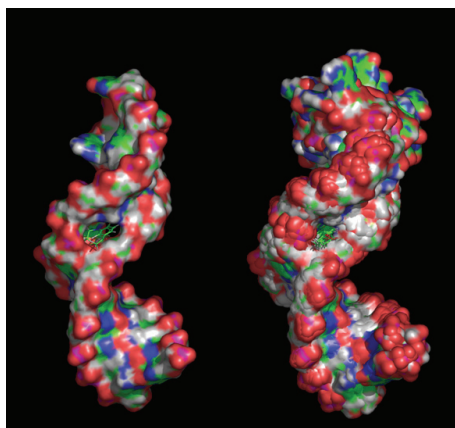


Abbildung 76: PDB-Eintrag 1LVJ – Vergleich eines Mitglieds des Ensembles gegen die Überlagerung aller Mitglieder

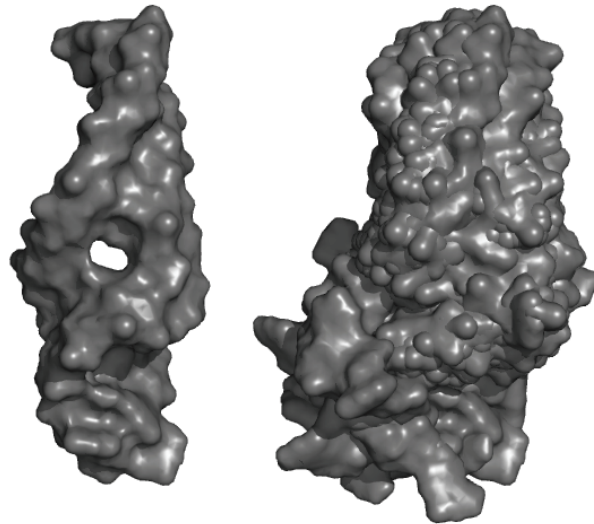


Abbildung 77: PDB-Eintrag 1ANR– Vergleich des siebten Mitglieds des Ensembles (aufgrund relativer Ähnlichkeit zu 1LVJ) gegen die Überlagerung aller Mitglieder

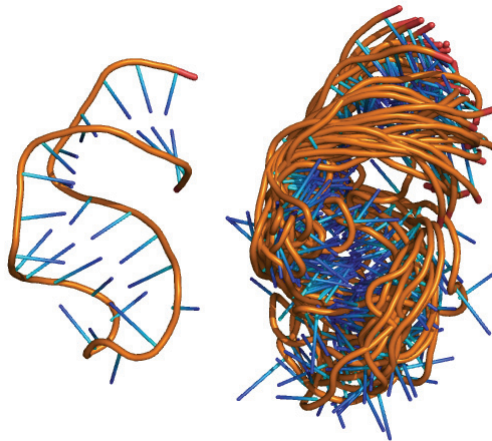


Abbildung 78: PDB-Eintrag 1ANR– Vergleich des ersten Mitglieds des Ensembles gegen die Überlagerung aller Mitglieder

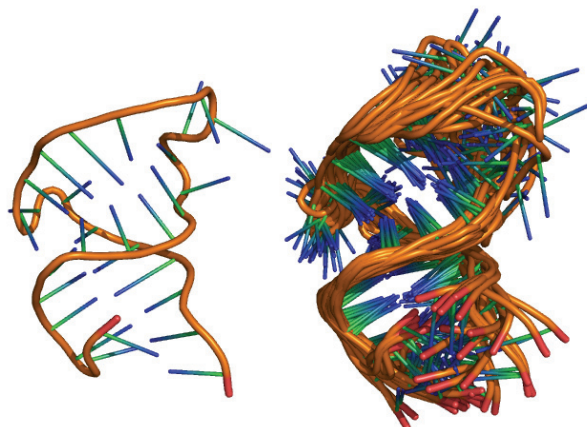


Abbildung 79: PDB-Eintrag 1ARJ – Vergleich des ersten Mitglieds des Ensembles gegen die Überlagerung aller Mitglieder

Wie sollte also die Gesamtheit der Lösungen zur Auswahl von Testverbindungen bewertet werden?

Durch willkürliche Entnahme eines Konformers aus dem konformellen Raum und seinem Einsatz als Dockingstruktur können zwar bezogen auf diese einzelne Konformation gute Vorhersagen ermöglicht werden, jedoch steht dieses Ergebnis in einem unbestimmten Verhältnis zu den anderen möglichen Konformer-Liganden-Lösungen, so dass eine repräsentative Aussage über das Bindeverhalten fragwürdig erscheint (Liebeschuetz 2008).

Insbesondere fließt hierbei kein Wissen über die Populationshäufigkeiten der jeweils relevanten Konformere in die Bewertung ein, was einen deutlichen Mangel dieser Methode darstellt. Da in der Natur die beiden Wechselwirkungspartner RNA und Ligand durch Diffusion aufeinander treffen müssen, und damit es bei dieser Begegnung (Stoß) auch zu einer zeitlich stabilen Komplexbildung kommen kann, müssen die beiden Moleküle in einer Weise aufeinander treffen, so dass es zu relativ stabilen Bindungen zwischen den beiden Partnern kommt. Ist somit eine spezifische RNA-Liganden-Konformation identifiziert worden, muss auch eine Bewertung erfolgen, wie wahrscheinlich die Bildung dieses Komplexes überhaupt ist.

Eine diskrete Bewertung aller zum Docking verwendeten Konformere benötigt demnach zusätzliches Wissen, um die einzelnen Konformere im Kontext des Gesamtensembles bewerten zu können (Liebeschuetz 2008). Ansonsten gleicht das willkürliche Beschränken der Auswahl auf einen Beispielkonformer einem Blick auf ein Zufallsereignis und lässt demzufolge keine verallgemeinerbaren Aussagen zu.

Die kombinatorische Explosion möglicher RNA-Liganden-Komplexe, welche durch die Kombinationsvielfalt aus dem Konformerraum der TAR und dem chemisch zugänglichen Raum möglicher Ligandenverbindungen entsteht, lässt sich mit heutigen Mitteln der Docking-Programme nicht mehr systematisch bewältigen. In dem kürzlich erschienenen Review „Overcoming the Inadequacies or Limitations of Experimental Structures as Drug Targets by Using Computational Modeling Tools and Molecular Dynamics Simulations“ von Marco und Gago in ChemMedChem (Marco und Gago 2007) wird auf diesen Sachverhalt eingegangen. Sie beleuchten in ihrem Überblicksartikel den Wechsel von der Vorstellung eines starren Zielmoleküls hin zu einer Repräsentation als „Macrostate“, bei dem eine bestimmte Struktur als Teil eines Ensembles – aller möglichen Zustände des Moleküls – bei den gegebenen Bedingungen aufgebaut ist. Es wird ferner beschrieben, wie man den experimentell erschlossenen Konformerraum für eine vollständigere Betrachtung erweitern kann, indem man mittels verschiedener Konformererzeugungsalgorithmen (Normal mode analysis, NMA; Gaussian network model, GNM; Floppy inclusion and rigid substructure topography; Molecular dynamics simulations MD-Simulation) zu mehr oder weniger repräsentativen Ensembles gelangen kann.

Die Betrachtungsweise, ein Molekül nicht mehr als statisches Individuum zu sehen, sondern sich analog zu seinen möglichen Faltungswegen ein Ensemble von gleichzeitig nebeneinander bestehenden, ineinander übergehenden Einzelkonformeren vorzustellen, ermöglicht gerade im Hinblick auf die hier verwendete TAR-Struktur mit ihrer bekannten konformellen Vielfalt eine Analyse der Zielstruktur unter allgemeineren Gesichtspunkten als dies die Einzelkonformeranalyse erlauben würde. Dies ist besonders gut zu motivieren, indem man sich immer wieder klar macht, dass die meisten der Einträge in der PDB mittels eines Computer-Algorithmus aus experimentellen paarweisen Abstandsdatensätzen generiert wurden, wobei die Generation der Konformere mittels einer vereinfachten MD-Simulation (Simulated Annealing: SA) erfolgte. Somit liegt es nahe, die relativ restriktiven Einstellungen der SA, welche im Normalfall möglichst zu einer einzigen Lösung führen sollen, mit dem Wissen aufzuweichen, so dass man mit den „Experimentellen Methoden“, für die jeweilige Konstellation aus Zielstruktur und Effektmolekülen bestenfalls die häufigste Lösung beschreiben wird, allerdings nicht die Gesamtheit der möglichen Lösungen.

Dieser Disput wird in ähnlicher Form schon seit Jahren bereits zwischen den „Anhängern“ der NMR und Röntgenkristallographie ausgefochten, der sich vereinfacht als „Welche

Methode beschreibt die Wirklichkeit genauer?“ ausdrücken lässt. Kristallographen erreichen unverkennbar die höheren Auflösungen bei der Strukturaufklärung, vorausgesetzt ihnen gelingt es Kristalle zu züchten. „NMRler“ hingegen sind bereits in der Lage die Struktur von Molekülen in einer flüssigen Umgebung abzuleiten und kommen dadurch dem „wirklichen“ dynamischen Treiben in einer Zelle näher als es der in relativer Bewegungsunfähigkeit verharrende Kristall jemals sein kann. Jedoch sind auch hier aufgrund des nach unten begrenzten zeitlichen Auflösungsvermögens der Messverfahren, der Güte der Strukturaufklärung Grenzen gesetzt. Überlagerungen einzelner Mikrozustände zum Makrozustand erlauben durch die ungenügende Schärfe des Messsignals keine vollständige Trennung des Makrozustands in seine Subpopulationen der zugrunde liegenden Mikrozustände. Zum jetzigen Zeitpunkt wird dieses Problem häufig durch „relatives Einfrieren“ der Proben gelöst, wodurch reduzierte Teilchenbewegung eine Schärfung des zu erhaltenden Messsignals erfolgt. Unweigerlich bedeutet dies aber auch eine Einschränkung der konformellen Freiheit der zu untersuchenden Zielstruktur, bedingt durch eine Reduktion der verfügbaren Bewegungsenergie. Dies resultiert somit in einer Verschiebung des Populationsgleichgewichts von den höherenergetischen Zuständen hin zu den niederenergetischen Zuständen, dessen Auflösungsvermögen durch Erhöhung des Populationsanteils relativ zu den höherenergetischen Zuständen verbessert wird. Jedoch führt diese Umlagerung innerhalb des Ensembles an möglichen Konformeren zu einem Informationsverlust für das Ligandendesign, da die abgeleiteten Strukturinformationen und ihre teils integral ableitbaren Populationsgrößen nicht der Realität unter nativen Bedingungen entsprechen können, sondern einen eingeschränkten Teil dieser Wirklichkeit wiedergeben. Wobei die oben genannten Methoden gemeinhaben, dass wenig bevölkerte Subzustände nicht erfasst werden können. Die so experimentell bestimmten Abstandswerte und ihre daraus abgeleiteten Strukturvorschläge geben somit die bei den jeweiligen experimentellen Bedingungen vorherrschenden Subzustände wieder, welche durchaus für ein Ligandendesign als primäre Strukturvorlage dienen können. Auf diesem Wege können also strukturelle Gegenstücke respektive Liganden gefunden werden, die exakt zu diesen Strukturen passen würden. Der Erfolg dieses Ansatzes hängt aber damit zusammen, ob der so entwickelte Ligand auch wirklich auf diesen Konformer unter nativen Bedingungen einwirken und das Populationsgleichgewicht der Konformere zu dessen Gunsten dauerhaft verschieben kann, so dass eine Bindung des Liganden dadurch auch einen Effekt auf die Wechselwirkungsrate mit den Interaktionspartnern der Zielstruktur hat.

Wie sehen die Konformationen der natürlichen Wechselwirkung aus?

Diese Frage ist weiterhin wichtig – falls experimentell erreichbar – da die durch den Liganden induzierten/stabilisierten Konformationen im Falle einer gewünschten Inaktivierung der natürlichen Wechselwirkung nicht die Konformationen enthalten sollte, wie sie auch eine natürliche Interaktion verursachen würde. Dieser potentiell synergistische Effekt sollte vermieden werden, um die Chance auf einen funktionalen Inhibitor zu optimieren. Im Fall der HIV-1 TAR RNA sind leider bis zur Durchführung dieser Arbeit keine Strukturen aufgeklärt worden, die sowohl die RNA als auch das Tat-Protein interagierend zeigen. Bis heute wurden somit nur Teilausschnitte dieser Wechselwirkung aufgeklärt, die für sich genommen Informationen über potentielle Wechselwirkungsmechanismen enthalten, aber diese nicht explizit festlegen. Die beste Annäherung an diese Wechselwirkung wurde bis jetzt durch die Verwendung von Peptiden erbracht, welche dem Interaktionsbereich des Tat-Proteins mit der TAR-RNA entnommen wurden und meist nur einige Aminosäuren lang sind. Die Entwicklung wurde in den letzten 10 Jahren von den ersten Versuchen mit der BIV-TAR und dem dazugehörigen Tat-Peptid-Fragment (PDB-Einträge 1BIV und 1MNB) bis zur Entwicklung eines Beta-Hairpin Peptidomimetischen Inhibitors der BIV Tat-TAR Interaktion (PDB-Eintrag 2NS4) vorangetrieben.

Motivation für „Molekül Dynamik Simulationen“ (MD)

Die beschriebenen Probleme bei der Auswahl eines geeigneten Konformers zur strukturbasierten Entwicklung von RNA-Liganden, hier im Speziellen von TAR-Liganden, führten zur Beschäftigung mit der Frage, welche Konformere für das Ziel möglich sind und wie wahrscheinlich die einzelnen Konformere mit ihren charakteristischen Pharmakophor-Raumstrukturen in der Gesamtheit des Konformerraums auftreten.

Hierbei galt es zwei Hauptprobleme zu lösen:

1. Erzeugung eines Ensembles von Konformeren, welches den möglichen Konformerraum möglichst gut in Relation zur Wirklichkeit wiedergibt.
2. Die Verwaltung des so abgebildeten Konformerraums, zur einfachen (schnellen) Analyse umzusetzen.

Da die verfügbaren Konformer-Quellen, wie die Einträge der PDB (1ANR, 1ARJ, 1LVJ, 1QD3 und 1AJU), einerseits durch ihre Anzahl einen eher kleinen Konformerraum abdecken und diese Bereiche im Fall der Strukturen mit gebundenem Liganden auch noch durch diese induziert wurden, handelt es sich bei dieser Basis um einen Blick auf Eckpunkte des Konformerraums. Zur Erstellung weiterer, physiologisch möglicher Konformere wurde in dieser Arbeit auf die „Molekül Dynamik Simulation“ (MD) (Karplus und McCammon 2002) zurückgegriffen, da sie unter geeigneter Parametrisierung eine gute Näherung der Wirklichkeit ermöglicht. Die Automatisierbarkeit der MD ermöglicht hierbei das Erstellen großer Mengen an Konformeren, welche durch die Art der Simulation bedingt außerdem eine zeitliche Entwicklungskomponente beinhaltet, die für spätere Auswertungen und Erweiterungen des Auswahlverfahrens von Konformeren für das Docking verwendet werden können. Aus diesen Informationen ist die weitere Entwicklung eines relativ schnellen, pseudo-flexiblen Dockingansatzes aus den bestehenden schnellen rigiden Verfahren möglich und dies einfach durch Erweiterung der bisherigen Auswertungsverfahren mittels Ähnlichkeitssuche und Bewertung der konformellen benachbarten Strukturen.

Bei der Generierung einer TAR-RNA-Konformeren-Datenbank wird durch ausreichendes Sampling der Läufe (unterschiedliche Startkonformere und Starttemperaturvektoren) im Idealfall die Oberfläche des Konformerraums komplett abgebildet. Die so generierten Trajektorien verknüpfen dann eventuell die bereits experimentell bekannten Strukturen (1ANR, 1ARJ, etc.) in diesem Konformerraum. Da mit einer begrenzten Anzahl von MD-Läufen (einzelnen MD-Simulationen mit derselben Zielstruktur) aus der Vielzahl von Möglichkeiten immer nur ein Teilausschnitt des ganzen Bildes dargestellt wird, ist auf eine einfache Erweiterung der Verwaltungs- und Analysetools zu achten, um sich sukzessive der Gesamtheit des Konformerraums langsam annähern zu können. Eine einzige Simulation ausreichender Länge sollte ebenfalls zu einem vollständigen Bild der Konformerraumoberfläche führen, ist jedoch aus Gründen der Rechengeschwindigkeit unter heutigen Bedingungen nicht möglich. Dementsprechend ist das heuristische Sampling der Konformerraumoberfläche unter den gegebenen Umständen der beste Weg, um einen wenn – auch lückenhaften Eindruck – der Beschaffenheit des Konformerraums zu erhalten.

Basierend auf einer derartigen Datensammlung lassen sich theoretisch folgende Punkte ableiten:

- Identifizierung von Bindestellen, welche wahrscheinlich durch einen Liganden besetzt werden können
- Definierung des wahrscheinlichsten Konformers und Stabilisierung desselben mit Hilfe eines Liganden
- Entwicklung einer auf MD-Trajektorien basierenden flexiblen Docking-Lösung für das virtuelle Screening.

Die Umsetzung dieser Punkte ist das endgültige Ziel des im Rahmen dieser Arbeit begonnenen Software-Pakets: Pandora – **P**ocket **A**nalysis for **D**Ocking on **R**epresentative **A**ptamers. Der erste Punkt wurde im Rahmen dieser Arbeit bereits umgesetzt.

4.2 MD-Datensätze als Konformergeneratoren

Hintergrund:

Die Voruntersuchungen an den verfügbaren PDB-Einträgen mit TAR-RNA, insbesondere die freie TAR im Eintrag 1ANR, zeigen eine starke Diversität innerhalb des Konformerraums. Abgeleitet hieraus ergab sich die Schwierigkeit einen Konformer als Referenz aus dem vorhandenen Konformerraum für das weitere Liganden-Design zu spezifizieren. Da die TAR-RNA nach der vorhandenen Datenlage große Bereiche des Konformerraums zu bevölkern scheint, stellte sich die Frage, ob dieser Konformerraum mittels einer geeigneten Methode genauer beschrieben werden könnte. Als einfachster Ansatz wäre der Einsatz eines Konformergenerators wie das für die Liganden-Konformationserzeugung einsetzbare CORINA 3.20 (*Molecular Networks GmbH, Erlangen*) für die RNA denkbar. Jedoch ist bei einem solchen Sampling-Verfahren die Abdeckung des Konformerraums nach physiologischen Bedingungen nicht zwingend gegeben. Hieraus ergab sich die Suche nach einer Alternative, die die physiologischen Bedingungen bei der Konformererstellung möglichst exakt nachbildet. Optimalerweise würde dieses Verfahren ebenfalls Informationen über die Populationshäufigkeit unter den gegebenen Bedingungen liefern müssen, um die einzelnen Konformervorschläge des resultierenden Gesamtensembles bewerten zu können. Diese Bewertung der Antreffwahrscheinlichkeit innerhalb des Ensembles ermöglicht dann wiederum die differenzierte Bewertung/Gewichtung der aus den jeweiligen Konformeren abgeleiteten potentiellen Bindemöglichkeiten im Kontext des Ensembles. Dieses Abbild der „Wirklichkeit“ bietet die MD-Simulation.

Im Rahmen der Arbeit konnten nach eigenen anfänglichen Tests mit der Software GROMACS 3.0 (Hess et al. 2008), von einer Kooperationspartnerin (Eva Kestner, Arbeitskreis Gohlke) umfangreichere MD-Datensätze für die TAR erhalten werden. Diese Simulationen wurden jedoch mit dem „AMBER8“-Software-Paket durchgeführt. Hierbei handelt es sich um die Konformerdaten von vier MD-Simulationsläufen, die von unterschiedlichen Startstrukturen aus unter denselben Simulationsbedingungen durchgeführt wurden. Die in dieser Arbeit verwendeten MD-Daten beinhalten nur die eigentlichen MD-Simulationsdaten nach erfolgten Energieminimierungs- und Pre-Äquilibrierung-Schritten.

4.3 Die MD-Trajektorien

Hintergrund:

Im Rahmen der Arbeit konnte eine Kooperation mit Eva Kestner (AK Gohlke) eingegangen werden. Dies brachte vier unterschiedliche Trajektorien von MD-Simulationen eines TAR-Konformer in das Projekt ein. Insgesamt handelt es sich bis heute um 153.000 Konformationen, welche den Simulationen entnommen werden konnten. Es existieren mittlerweile jedoch 47.000 weitere Konformationen, die in weiteren Studien noch in das Projekt eingebracht werden sollen. Aufgrund der Laufzeit der Simulationen war eine frühere Integration im Rahmen dieser Arbeit jedoch nicht mehr möglich.

Bei den MD-Läufen handelt es sich um die Simulation der solvatisierten, mit Gegenionen versehenen Modell-TAR, wie sie im PDB-Eintrag 1ANR zu finden ist. Alle Systeme (TAR, Ionen, Wasser) wurden vor dem Start der eigentlichen Simulation energie-minimiert und pre-equilibriert. Sämtliche Arbeiten bezüglich der Durchführung der MD-Simulationen wurden von Eva Kestner durchgeführt, die kurze Angabe zur Durchführung dieser Simulation dient hier nur der Vollständigkeit. Im Rahmen dieser Arbeit wurden die Ergebnisse der MD-Simulationen als physiko-chemische Konformer-Generatoren gewertet, die einen Einblick in die Bewegungsmöglichkeiten der TAR liefern sollten.

Vorbereitung:*Vorbereitung der TAR-Strukturen:*

Als Startstrukturen wurden auch Konformere der PDB-Einträgen 1LVJ, 1QD3 und 1ARJ entnommen, die sich zum Teil von der Sequenz des PDB-Eintrags 1ANR unterscheiden (Länge und Endbasenzusammensetzung). Deshalb wurden – um eine bessere Vergleichbarkeit zu erzielen – diese Konformere angepasst, um der Sequenz des 1ANR-Eintrags der PDB zu entsprechen. Die verwendete Sequenz lautet wie folgt:

GGCAGAUUCUGAGCCUGGGAGCUCUCUGCC

Vorberechnungen:

Das solvatisierte System wurde unter Festhalten der Atomkoordinaten der TAR-RNA 20.000 Schritte energie-minimiert. Anschließend erfolgte die Temperatureinstellung auf 300 K mittels Berendsen Thermostat (Berendsen et al. 1984) unter konstantem Volumen (NVT) für 50 ps. Die Dichte des Solvents wurde durch eine Simulation von 50 ps Länge bei konstantem Druck (NPT, $p = 1\text{ atm}$) angepasst. Hieran schloss sich das langsame Entfernen der Bewegungsbeschränkungen der TAR an, die über 50 ps durchgeführt wurde. Innerhalb der letzten 50 ps der Pre-Equilibrierungsphase wurde dem System gestattet sich frei bei 300 K zu entwickeln.

MD-Simulation:

Für die MD-Simulationen wurden folgende Parameter verwendet:

Software-Paket = AMBER 8 (University of California, San Francisco)

Wassermodell = TIP3P

Größe der Box = mindestens 11 Å um die RNA

Gegenionen = Natriumionen

Kraftfeld = Cornell *et al.* 94 all atom force field

Alle Bindungen, die Wasserstoffatome einschlossen, wurden hierbei mit dem SHAKE Algorithmus (Cheatham et al. 1999) beschränkt.

Periodische Rand-Bedingungen unter Verwendung der Particle Mesh Ewald (PME) Methode (Essmann et al. 1995)

Schrittgröße = 2 fs

Vorgesehene Gesamtdauer = 50 ns

Samplingrate = jede 500te Konformation ($500 * 2\text{ fs} = 1\text{ ps}$)

Temperatur = 300 K.

Direct-space non-bonded cut-off of 8 Å

Datenübergabe:

Die Daten der MD-Simulationen (MD-Läufe) wurden im AMBER-Format als mdcrd-Dateien während der Berechnung abgespeichert und enthalten die Koordinaten der Atome in der Reihenfolge, wie sie in der zum MD-Lauf zugehörigen im PDB-Format vorliegenden Datei abgelegt wurden. Mithilfe dieser Dateikombination aus mdcrd-Datei und PDB-Datei können aus diesen Quelldateien die Konformere der Trajektorie abgeleitet werden. Dies erfolgte mittels des im Rahmen dieser Arbeit entstandenen Pandora, welches dazu benutzt wurde die Konformere der MD-Läufe in eine Datenbank zu übertragen.

Ergebnisse:

Die folgenden Abbildungen 80 bis 83 zeigen die Startstrukturen und vorläufigen Endstrukturen der verwendeten MD-Simulationen. Angegeben wurde ferner der RMSD zwischen Start- und Endstruktur sowie die Entwicklungszeit als auch die Anzahl der abgespeicherten Einzelbilder, die zur Auswertung zur Verfügung standen. Als eine weitere Teil-Untersuchung wurde der für jeden MD-Lauf erzielte RMSD-Wert (alle Atome oder nur Phosphoratome) gegen die Entwicklungszeit der MD-Simulation aufgetragen; dies ist in Abbildung 84 zu sehen.

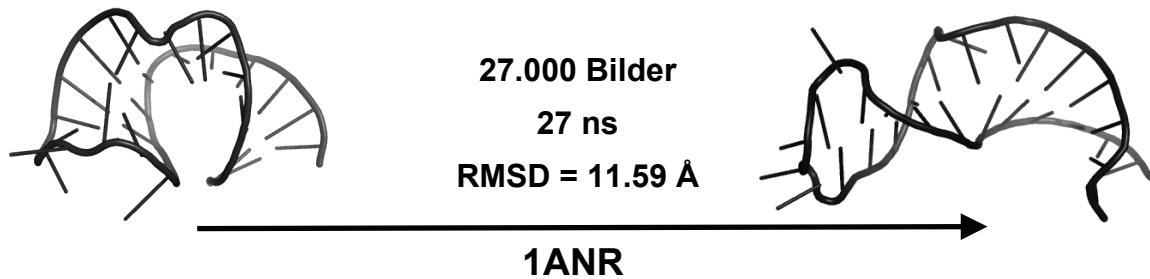


Abbildung 80: Start- und Endbild des 1ANR-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.

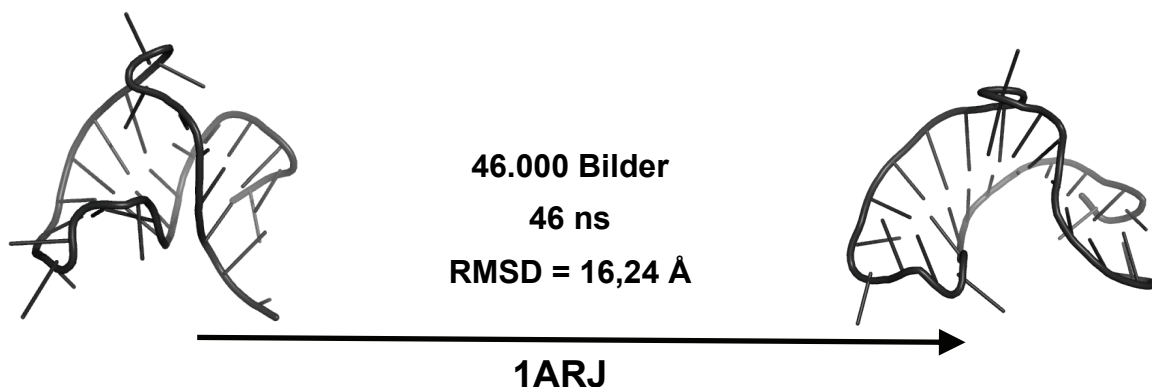


Abbildung 81: Start- und Endbild des 1ARJ-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.

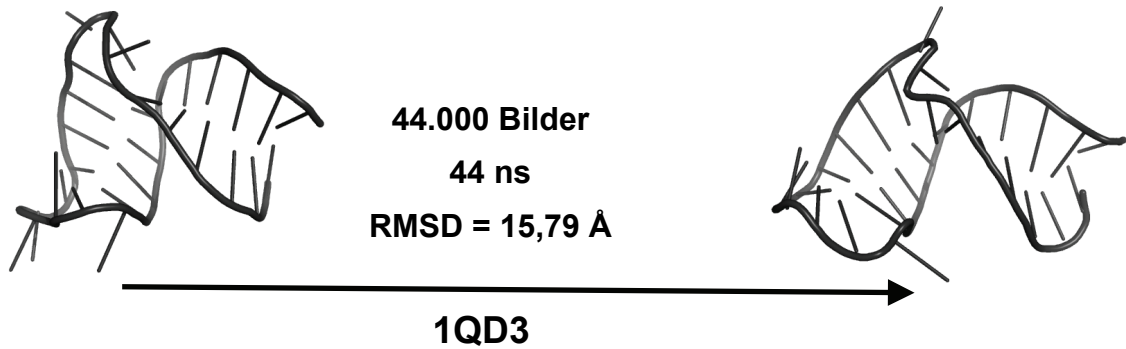


Abbildung 82: Start- und Endbild des 1QD3-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.

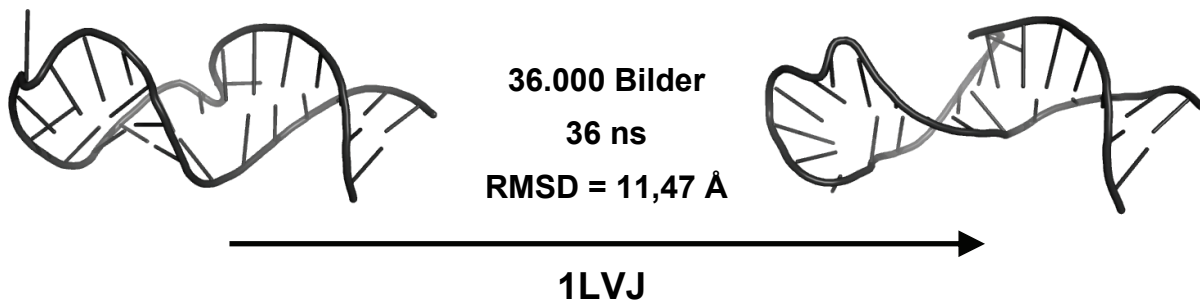


Abbildung 83: Start- und Endbild des 1LVJ-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.

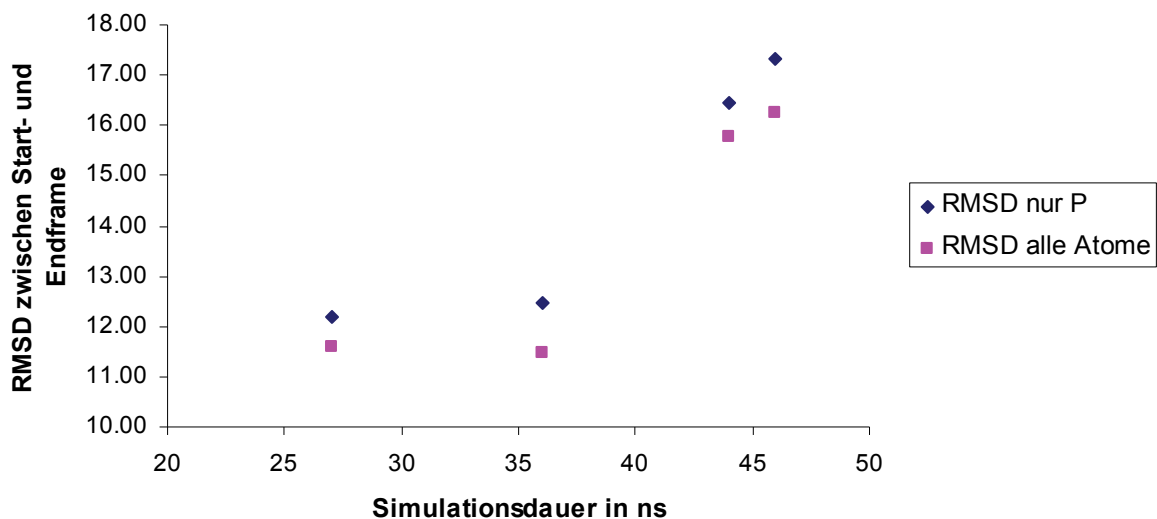


Abbildung 84: Auftragung des RMSD zwischen Start- und Endstruktur der vier MD-Läufe gegen die MD-Entwicklungszeit. Für die RMSD-Werte basierend auf allen Atomen ergibt sich ein $r = 0,90$, basierend nur auf den Phosphoraten ein $r = 0,92$.

Tabelle 25: RMSD-Distanzmatrix der MD-Lauf Start- und Endbilder

1ANR: 1	1ANR: 27000	1ARJ: 1	1ARJ: 46000	1QD3: 1	1QD3: 44000	1LVJ: 1	1LVJ: 36000	
0,00	11,59	16,88	21,59	5,85	17,59	18,65	15,99	1ANR: 1
	0,00	17,34	19,93	10,78	16,69	15,33	14,42	1ANR: 27000
		0,00	16,24	17,99	14,17	11,58	18,18	1ARJ: 1
			0,00	21,27	12,53	14,91	19,54	1ARJ: 46000
				0,00	15,79	18,31	14,98	1QD3: 1
					0,00	17,97	21,94	1QD3: 44000
						0,00	11,47	1LVJ: 1
							0,00	1LVJ: 36000

Diskussion:

Die MD-Läufe waren aufgrund der andauernden Rechenzeit nicht in ihrer vollen, geplanten Länge von 50 ns in diese Arbeit zu integrieren. Der Aufbau von Pandora ermöglicht jedoch die Integration der restlichen Daten zu einem späteren Zeitpunkt. Aufgrund des Teilcharakters der bisher erhaltenen MD-Laufdaten muss bei einem Vergleich der Läufe innerhalb dieser Arbeit auf die Anzahl der momentan – in den einzelnen MD-Läufen – enthaltenen Bilder eingegangen werden. Die Schrittweiten innerhalb der einzelnen MD-Simulationen sind bei allen Läufen jedoch immer 1 ps von Bild zu Bild und beinhalten jeweils 500 Einzelschritte der MD-Simulation.

Beachtenswert ist die relative Korrelation der RMSD-Werte mit der Entwicklungszeit der MD-Simulationen; siehe Abbildung 84. Da hier aber ausschließlich die jeweiligen vier Start- und Endpunkte der Simulationen verglichen wurden, ist eine vorsichtige Interpretation anzustreben. Da die RMSD-Werte bis jetzt aber mit steigender Simulationszeit zunehmen, muss in weiteren Studien darauf geachtet werden, dass die einzelnen MD-Simulationen auch wirklich konvergieren werden. Bisher entfernen sich die RMSD-Werte der MD-Läufe im Durchschnitt stärker voneinander, als dies im Vergleich der jeweiligen Start-Endbildvergleiche mittels RMSD-Berechnung der Fall ist; siehe Tabelle 25. Diese Distanzmatrix zeigt in vier der möglichen sechs Kombinationen aus Endbildern der MD-Läufe einen RMSD-Wert > 16,24 Å, der als größte Entfernung innerhalb eines MD-Laufs klassifiziert wurde (1ARJ-Lauf). In einem Fall sind sich (laut RMSD-Wert) die Startstrukturen unterschiedlicher MD-Läufe ähnlicher als die kleinste bisher aufgetretene RMSD-Werte-Abweichung (11,47 Å zwischen Start- und Endbild bei 1LVJ) innerhalb der MD-Läufe (Startbilder von 1ANR und 1QD3 RMSD = 5,85 Å).

4.4 Die RNA-Konformer-Datenbank

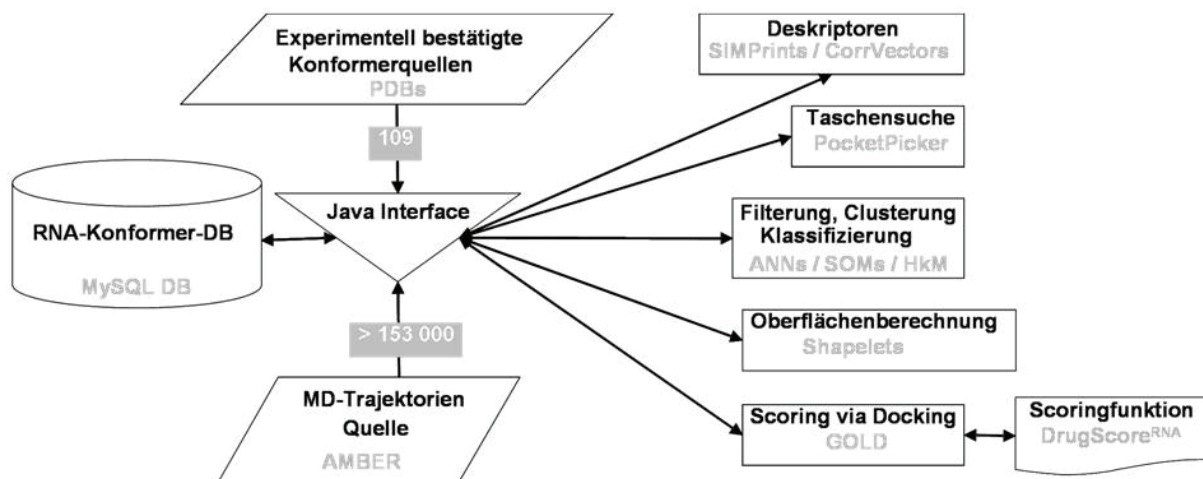


Abbildung 85: Schema des Informationsflusses innerhalb des RNA-Konformer-Datenbank-Projekts. Mit Ausnahme der Verwendung der DrugScore^{RNA}-Scoringfunktion wurden alle Teile im Rahmen dieser Arbeit integriert.

4.4.1 Hintergrund

Die Verwendung von MD-Läufen als Konformerquelle lässt die einfache Generierung hunderttausender Konformere zu. Jedoch ist die Software, welche klassischerweise zur Auswertung dieser Daten bereit steht, auf andere MD-spezifischere Fragestellungen als die Bindestellenidentifikation angepasst. Ein MD-Lauf wird hier typischerweise immer als Ganzes betrachtet, so dass ein gezielter, schneller Zugriff auf einzelne Bilder sowie die Annotation dieser Bilder noch nicht umgesetzt sind. Mit Programmen wie der Visualisierungssoftware VMD (Humphrey et al.1996) lassen sich zwar einfach einzelne Trajektorien betrachten und in vielen verschiedenen Arten rendern, jedoch ist dieses Programm nicht in der Lage eine Datenbank mit den einzelnen Bildern anzulegen, die dann als erweiterte Speicherform für die Auswertung der Strukturdaten und ihrer abgeleiteten Metadaten dauerhaft und schnell zugänglich ist.

Des Weiteren bietet eine Integration der Daten in eine, durch die hier hauptsächlich verwendete Programmiersprache, anzusteuernde Entwicklungsumgebung den flexiblen schnellen Zugriff auf die Daten während des Entwicklungsprozesses.

4.4.2 Koordinierungsprofil

Das Projekt erfordert die Verwaltung von Strukturdaten und den daraus abzuleitenden weiteren Eigenschaften. Hierzu werden folgende Anforderungen an das Koordinierungsprogramm gestellt, das die erforderlichen Subprogramme verwaltet:

- Verwaltung von Strukturdaten aus mindestens zwei Quellen (PDB-Dateien und AMBER-MD-Koordinaten-Dateien) sowie die Erstellung von Deskriptoren für diese Strukturdaten.
- Die Interaktion mit deskriptor-verarbeitenden Programmen (ANNs, SOMs, hkMeans (Böcker et Al. 2005)).
- Die Möglichkeit Strukturdaten an andere Programme zu übergeben (PDB-Format), um diese beispielsweise zum Docken (GOLD) oder Rescoren (DrugScore^{RNA}) zu verwenden.
- Die Berechnung und Speicherung der Moleküloberfläche.

- Die Entgegennahme der Ergebnisse externer Taschensuchprogramme.
- Die eigenständige Berechnung der Taschenpositionen und Ableitung der molekularen Taschenoberfläche.

Für eine schematische Übersicht siehe Abbildung 85 auf Seite 131.

4.4.2.1 Datenverwaltung mit SQL

Das Profil der Datenverwaltung beinhaltet das einfache, automatisierbare Verwalten der Daten als auch die ständige Erweiterung der Datensammlung mit neuen Daten, so dass eine Abschätzung der benötigten Speicherkapazität nicht möglich ist. Die generelle Lösung für dieses Problem bietet das Konzept der Datenbank mit ihrem nichtlinearen Zugriff auf die in ihr abgelegten Daten mit dazugehöriger Steuerungssoftware.

Als Auswahlkriterium für die zu verwendende Datenbank wurden folgende Punkte mit einbezogen:

- Zur Integration in die laufende Entwicklung der Analysesoftware sollte die Datenbank eine Schnittstelle zur automatischen Steuerung der Datenbank in der verwendeten Programmiersprache besitzen.
- Da die verwendete Sprache (Java) systemunabhängig ist, sollte dies auch für den verwendeten Datenbanktyp gelten, um möglichst unabhängig von der Hardware und dem jeweiligen Betriebssystem zu bleiben.
- Um unnötige Kosten bei der Nutzung in Kooperation mit anderen Arbeitskreisen zu ermöglichen, sollte die Software wenn möglich kostenfreie Open-Source-Software sein.

Alle diese Bedingungen wurden vom MySQL-Software-Paket erfüllt, weswegen es die Basis der Datenverwaltung in diesem Projekt übernimmt. Da die Datenbank aber prinzipiell unabhängig von Java mit SQL⁷-Befehlen gesteuert wird, wurde ebenfalls eine Trennung der Datenverwaltung von der Datenverarbeitung erreicht. Daher kann die Analysesoftware durch den Austausch entsprechender Treiber im Java-Programm auch auf andere Datenbank-Softwarepakete angepasst werden, ohne dass eine Änderung im Hauptprogramm nötig wird.

4.4.2.2 Serialisierung des Variablenspeichers

Zum Abspeichern der Ergebnisse einzelner Arbeitsschritte während der Analyse wurde die Möglichkeit integriert, den jeweiligen aktuellen Variablenspeicher des Programms dauerhaft auf die Festplatte zu schreiben. Somit kann auch unabhängig von der Datenbank mit dem Programm gearbeitet und Ergebnisse für späteres Weiterverarbeiten gesichert werden.

4.4.2.3 Laden von Strukturdaten

Das entwickelte Programm ist in der Lage Strukturinformationen aus PDB-Files zu extrahieren und beherrscht das Einlesen von zu einem PDB gehörenden AMBER-MD-Koordinaten-Dateien (run00X.mdcrd).

4.4.2.4 Einlesen von PDB-Dateien

Da im Rahmen dieser Arbeit erste Tests mit den Strukturdaten der PDB durchgeführt wurden, war ein Modul nötig, welches in der Lage ist, die Information der PDB-Strukturdateien auszuwerten und um die darin enthaltenen Atomkoordinaten und dazugehörige Atomlabel zu extrahieren. Hierzu wurde ein Modul für das Programm Pandora geschrieben, welches diese Aufgabe übernimmt und die enthaltenen Daten in die interne Speicherrepräsentation des Programms überführt.

⁷ SQL = Structured Query Language

4.4.2.5 Einlesen der *AMBER*-Koordinaten

Durch die Zusammenarbeit mit Eva Kestner waren diesem Projekt die Atom-Koordinaten der MD-Studien von Eva Kestner zugänglich (pers. Komm. Eva Kestner). Um diese Daten in die Datenbank laden zu können, wurde ein Programm-Modul für das Einlesen der im AMBER-Format abgespeicherten MD-Simulations-Ergebnisse entwickelt. Dieses basiert auf einer kombinierten Nutzung der für die MD-Simulation – die Startstruktur enthaltenden – erstellten PDB-Dateien, als auch den AMBER-MD-Trajektorien-Textdateien.

Mittels der – aus der jeweiligen zu einem MD-Lauf gehörenden PDB-Datei – extrahierten Informationen (Anzahl Atome der Strukturen und deren Reihenfolge in den Trajektorien-Dateien) ist es nun möglich, Bild für Bild aus den Trajektorien-Dateien in die Datenbank zu überführen.

4.4.3 Berechnung der Taschen nach *PocketPicker*

Die Berechnung der Taschen erfolgt mit einem aus der Software *PocketPicker* (Weisel et al. 2007) abgeleiteten Java-Modul von Martin Weisel, welches in dieses Projekt integriert wurde. Mittels der erstellten Interfaces ist es möglich für Strukturen, welche in Pandora geladen werden, sowohl die Taschenvolumina abzuleiten, als auch für diese die Deskriptoren abzuleiten (identisch zum *PocketPicker*).

Diese Deskriptoren beschreiben die Größe und Vergrabenheit der Taschen und können verwendet werden, um anhand dieser Kriterien Taschen zu clustern.

4.4.4 Berechnen der Moleküloberfläche

Die Berechnung der Moleküloberfläche wurde anhand des Java-Moduls „Shapelets“ (Proschak et al. 2008) umgesetzt. Mittels der erstellten Interfaces kann hierüber die Oberfläche des Moleküls als Sammlung diskreter Oberflächenpunkte extrahiert werden. Auf diese Oberflächenpunkte-Sammlungen können wiederum mittels eigener Module von Pandora diesen Punkten Eigenschaften zugewiesen werden. Im Standardfall ist dies das Atomlabel des zugrunde liegenden Atoms, das aufgrund der kleinsten Abstände zwischen Atomrepräsentation und Oberflächenrepräsentation ermittelt wurde.

4.4.5 Ableiten der Taschenoberfläche

Im Rahmen dieser Arbeit wurde ein Modul geschrieben, das aus der Oberflächenrepräsentation des Moleküls und den Taschengridpunktkoordinaten einen Teilausschnitt erstellt, bei dem die Oberflächenpunkte des Moleküls extrahiert werden, welche sich in unmittelbarer Nachbarschaft (Cut-Off < 1,5 Ångström) zu den Taschengridpunktkoordinaten befinden. Diese können dann zusammen mit den Taschengridpunkt-Ensembles in der Datenbank abgelegt werden. Die so abgeleiteten Koordinaten-Ensembles stellen somit die Repräsentation der Taschenoberfläche dar.

4.4.6 Laden von Deskriptordateien

Das Programm besitzt die Routinen, um eine Tabulator-separierte Deskriptordatei in den Speicher des Programms zu laden.

4.4.7 Ablegen und Extrahieren der Informationen in der Datenbank

Im Rahmen dieser Arbeit wurden folgende Informationen in der Datenbank gespeichert:

- Die Atomkoordinaten von verschiedenen PDB-Einträgen (PDB-Code: 1ANR, 1ARJ, 1LVJ, 1QD3, usw.).
- Atomkoordinaten zu MD-Simulationsläufen (Startstrukturen entstammen 1ANR, 1ARJ, 1QD3, 1LVJ).
- Taschenkoordinaten zu sämtlichen Atomkoordinaten-Sets (plus Vergrabenheit der Gridpunkte).

- Taschendescriptoren zu sämtlichen Taschenkoordinaten (skaliert und unskaliert).
- Taschenoberflächenkoordinaten (plus Atomlabel).
- SOM-Ergebnisse für verschiedene Clusterversuche mittels der abgeleiteten Deskriptoren.

4.4.8 Anwendung von *Pandora*

Da es sich bei Pandora um den Prototyp einer Analysesoftware für die Bewertung von Konformerdatensätzen handelt, bei dessen Entwicklungsbeginn noch nicht abzusehen war, was für Funktionen später genau benötigt werden würden und auch in welcher Reihenfolge Einzelfunktionen eingesetzt werden mussten, wurde auf eine starke Modularisierung dieses Projektes geachtet, für die jeweils eine GUI⁸ erstellt wurde. Daher gibt es innerhalb der Arbeit nicht nur einen möglichen Weg ans Ziel zu gelangen. Ferner wurde verstärkt an der Optimierung der Datenbankindizierung gearbeitet, um diese möglichst schnell auf Suchanfragen reagieren lassen zu können.

Der nachfolgende Abschnitt erläutert das Standardvorgehen beim Einfügen neuer Strukturdaten in die Konformer-Datenbank (PDB-Einträge oder MD-Läufe). Generell kann aber fast jeder Schritt ohne die Verwendung der Datenbank durchgeführt werden.

4.4.8.1 Standardvorgehen bei der Einbindung neuer Strukturdaten

Liegen neue Strukturdaten in Form einer einzelnen PDB-Datei oder als Kombination aus PDB-Datei und MD-Trajektorie vor, so können die enthaltenen Strukturen in die Datenbank überführt werden. Hierbei können bereits eine Reihe von Berechnungen durchgeführt werden und diese Ergebnisse ebenfalls als Annotation zur jeweiligen Struktur in der Datenbank abgelegt werden. Hierbei handelt es sich um folgende – im Rahmen dieser Arbeit als Standardvorgehen – erarbeiteten Punkte:

1. Laden der Strukturinformationen in Pandora.
2. Berechnung der Taschenpunkte und Deskriptoren für jede Struktur.
3. Berechnung der Taschenoberflächen für jede Struktur.
4. Ablegen der Strukturinformationen mit den jeweiligen Taschenpunkten, Taschendescriptoren und Taschenoberfläche in der Datenbank.

In der Datenbank sind nach dem Hochladen der Atomkoordinaten und dem anschließenden Berechnen der Taschen folgende zusätzliche Informationen abrufbar:

- Taschenkoordinaten (Gridpunkte mit Vergrabenheit).
- Taschenoberflächenkoordinaten (Oberflächenpunkte mit Atomlabel).
- Anzahl der Gridpunkte und Oberflächenpunkte zu jeder Tasche.
- Taschendescriptor nach PocketPicker (skaliert und unskaliert).

Daraus ableitbar sind im Folgenden weiterhin:

- die „Mittlere Vergrabenheit“ jeder Tasche,
- die Partition der Taschenoberfläche nach Atomtypen,
- der Korrelationsvektor der Taschenoberflächen basierend auf Atomtypen bis zu einer Entfernung von 20 Ångström.

⁸ GUI – graphical user interface – „Graphische Benutzer Oberfläche“

4.5 Ableiten eines Referenzdatensatzes aus der Datenbank

Basierend auf den in die Datenbank eingetragenen Konformeren sollen im Rahmen dieses Projekts möglichst repräsentative Referenztaschen bestimmt werden.

Die Fähigkeit Taschen – in den jeweiligen Konformeren – anzusehen wurde mittels eines kleinen MDDB-Browsers umgesetzt, der auch die Extraktion aus der Datenbank für eine mögliche Verwendung, beispielsweise in Dockingprogrammen, bereits umsetzen würde. Jedoch gilt es noch zu klären, welche Taschen „genau“ als repräsentativ zu betrachten sind.

4.5.1 Überlegungen zur Auswahl der repräsentativen Taschen aus dem verfügbaren Konformerraum

Zur Durchführung des Liganden-Designs anhand von Zielstrukturen müssen diese Zielstrukturen möglichst repräsentativ für den verfügbaren Konformerraum ausgesucht werden. Zugrunde liegend ist hierbei die Annahme, dass die repräsentativsten Strukturen diejenigen sind, welche bezogen auf die Gesamtmenge der möglichen Konformere für die Zielstruktur die höchste Antreffwahrscheinlichkeit innerhalb einer Stichprobe beinhalten. Genauer gesagt, sind für das erfolgreiche Design eines Liganden jene Strukturen vorzuziehen, auf die der Ligand in der Lösung auch treffen wird, so dass sich zwischen der Zielstruktur und dem Liganden eine Bindung aufbauen kann. Die relativ zu beobachtende Stärke der Bindung sollte hierbei auch direkt mit der zeitlichen Dauer des Auftretens des Zielkonformers und der Menge ähnlicher Zielkonformere zusammenhängen, da die Bindung eines Liganden nicht einen binären Prozess darstellt, sondern durch viele verschiedene Bindeposen ein gradueller Übergang zwischen Nichtbindern und Bindern möglich ist.

Die Aktivität im beschriebenen FRET-Assay hängt hierbei von der Gleichgewichtsverteilung zwischen freier und gebundener Form des Liganden ab, wobei die gebundene Form sich wiederum in wirkungstragende Bindung und wirkungslose Bindung aufteilen lässt.

Zwischen den beiden Formen der wirkungstragenden Bindung und wirkungsloser Bindung verläuft analog der Gradient zwischen positionsspezifischer Wechselwirkung und dem einfachen unspezifischen Dekorieren der Zielstruktur, solange bei der Bindung des natürlichen Liganden von einer positionsspezifischen Bindung an die Zielstruktur ausgegangen werden kann. Hierbei ist das erstrebenswerte Ziel die Identifikation der maximal möglichen spezifischen Interaktion, welche trotzdem eine ausreichende Grundaffinität zum Ziel aufweist, insbesondere, wenn noch keine hochaffinen Liganden für diese Zielstruktur vorliegen.

Für die hier bearbeitete Fragestellung der Identifikation neuer TAR-Liganden wurde ein Ansatz gewählt in dem sich diesem optimierten Referenzzielstrukturendatensatz schrittweise durch sukzessives Filtern des Gesamtdatensatzes genähert wird.

4.5.2 Erstes Kriterium Taschendurchmesser

Ableitung des Intervalls der optimalen Taschengröße

Hintergrund:

Die jeweilige Taschengröße des Datensatzes kann als ein Ausschlusskriterium verwendet werden, um den Datensatz weiter zu reduzieren: Indem die Taschen aus diesem entfernt werden, welche den Anforderungen der „optimalen“ Taschengröße entsprechen. Die „optimale“ Taschengröße kann hierbei aus Vergleichsanalysen eines Medikamenten-Datensatzes abgeleitet werden. Dazu wird ein Datensatz verwendet, welcher den konformellen Raum der bekannten Arzneistoffe abdeckt.

Durchführung:

In diesem Fall handelt es sich um die COBRA 6.1, welcher 8311 bekannte bioaktive Verbindungen enthält. Aus diesem Datensatz wurde ein 3D-Konformer Datensatz abgeleitet,

dem mittels Software jeweils der größte Abstand zweier Atome innerhalb der Strukturen für jeden Eintrag extrahiert wurde.

Ergebnis:

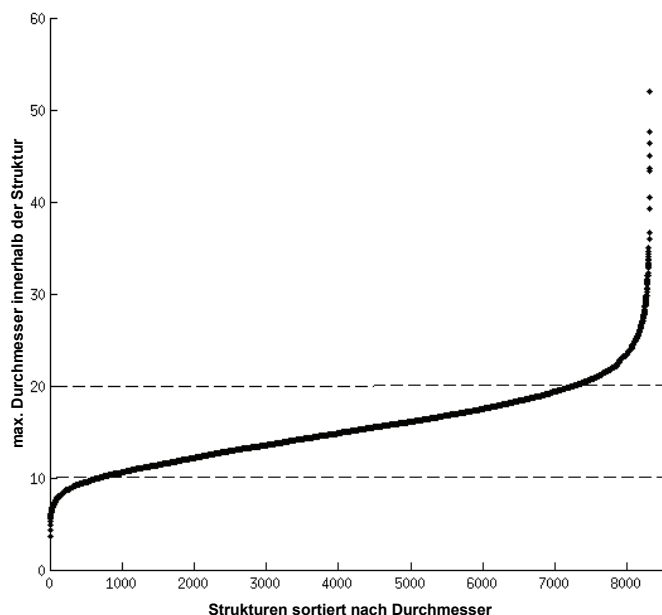


Abbildung 86: Maximale Durchmesser der Strukturen des COBRA-Referenzdatensatzes, Einträge sortiert nach diesem Kriterium zum Abschätzen der benötigten oberen und unteren Taschengröße

Abbildung 86 zeigt die Verteilung des maximalen Durchmessers der Strukturen der COBRA, sortiert nach der dem berechneten maximalen Durchmesser der 3D-Konformere. In Abbildung 86 sind die Grenzen von 10 und 20 Ångström als gestrichelte Linie angegeben. Ca. 80% der untersuchten Strukturen zeigen einen maximalen Durchmesser innerhalb dieses Intervalls. 7,8% entfallen auf Strukturen kleiner als 10 Ångström, 12,7% entfallen auf Strukturen größer als 20 Ångström. Als kleinstes Molekül wurde hierbei ein Sulfoxid mit einem Durchmesser von 3,655 Ångström und als größtes Molekül ein Peptid mit 51,997 Ångström gefunden.

Diskussion:

In der durchgeführten Untersuchung der COBRA zeigt sich, dass ca. 80 Prozent der Strukturen einen maximalen Durchmesser zwischen 10 und 20 Ångström aufweisen. Der aus dem COBRA-Datensatz abgeleitete Referenzgrößenbereich (maximaler Durchmesser = 10-20 Ångström) für bioaktive Substanzen deckt deren maximal auftretende Durchmesser bei ca. 80% der untersuchten Verbindungen ab. Abgeleitet aus diesen Informationen lässt sich als weiteres Auswahlkriterium somit ein „optimaler“ Größenbereich für Taschen bezogen auf bioaktive Referenz-Liganden festlegen. Bei Anwendung dieses Größenkriteriums auf die Taschen der Datenbank sollten somit nur Taschen übrig bleiben, die im Durchschnitt weder viel zu klein noch zu groß sind, um „klassische“ bioaktive Verbindungen vollständig aufzunehmen. Erwartet wird somit eine Optimierung der Oberflächengrößenüberlappungsmöglichkeiten, bei der die potentiellen Liganden weder stark aus der Tasche herausragen oder nur Teile dieser Tasche ausfüllen können.

Schlussfolgerung:

Abgeleitet aus den maximalen Durchmessern des – bekannte bioaktive Referenzsubstanzen enthaltenden – COBRA-Datensatzes, lässt sich als weiteres Auswahlkriterium somit ein „optimaler“ Durchmesserbereich von 10-20 Ångström für Taschen bezogen auf bioaktive Referenz-Liganden festlegen.

Anwendung des optimalen Intervalls der Taschengrößen zur Selektion der Taschen in der DB

Hintergrund:

Basierend auf den Befunden in Abschnitt 4.5.2 kann der Bereich der optimalen Größe für eine Tasche somit unter diesem sterischen Gesichtspunkt bei einer Abdeckung von 80% des gesamten bioaktiven Referenzmoleküldatensatzes zwischen 10 und 20 Ångström in diesem Intervall eingegrenzt werden. Dies erlaubt eine Eingrenzung des Taschenraums auf diejenigen Taschen, die durch die potentiellen Liganden optimal ausgefüllt werden, ohne größere ungenutzte Bereiche innerhalb der Tasche oder Herausragen der Liganden aus den Taschen zu verursachen.

Durchführung:

Für die Umsetzung des Filters wurde ein Modul geschrieben, das die Taschenvolumina der Datenbank extrahiert und für jeden Taschenkoordinatensatz den maximalen Durchmesser bestimmt. Die berechneten Werte werden mit den ID-Listen der Taschen, welche die Taschen eindeutig in der Datenbank identifizieren, an den Benutzer zurückgegeben.

Die weitere Aufbereitung der Datenliste erfolgte mit Matlab. Hierbei wurden die Einträge nach dem maximalen Taschendurchmesser mittels erstellter m-Skripte gefiltert. Basierend auf den gefilterten Listen wurden die Auftragungen in Abbildung 87 mittels weiterer m-Skripte erstellt. Die Anzahl Taschenmitglieder der Teilmengen ergab sich direkt aus der Länge der erstellten Listen.

Ergebnisse:

Die Anwendung des Größenfilters ($10 \leq \text{Taschengröße} \leq 20$) ergibt eine Reduktion des Taschenraums von 733.663 auf 148.286 Taschen, wobei eine Reduktion auf 150.237 bereits durch die Verwendung des unteren Auswahlkriteriums durch Abtrennung der zu kleinen Taschen erfolgt. Somit reduziert dieser Ansatz den ursprünglichen Datenraum auf 20 Prozent.

Abbildung 87 zeigt die Auftragung der Taschenhäufigkeit einer bestimmten Taschenpunktzahl für den ungefilterten Datensatz (links), für eine gemeinsame Auftragung des ungefilterten und gefilterten Datensatz (Mitte) und des gefilterten Datensatzes (rechts). Für die gemeinsame Abbildung wurden die Punkte des ungefilterten Datensatzes als ungefüllte Kreise, die des gefilterten Datensatzes als gefüllte Kreise wiedergegeben.

Eine weitere Darstellung kann in der Abbildung 92 und 93 auf Seite 148 im Abschnitt 4.5.7 gefunden werden, die die Darstellung der beiden Datensätze als Auftragung der „mittleren Vergrabenheit“ und Größe beinhalten.

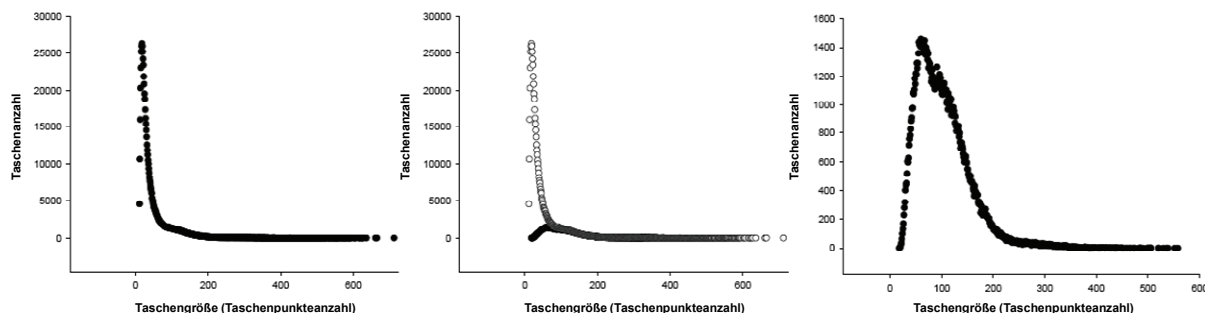


Abbildung 87: Taschengröße gegen Taschenanzahl; links nur der ganze Datensatz; mitte ganzer Datensatz (gefüllte Kreise) und gefiltert nach Durchmesser (ungefüllte Kreise); rechts gefiltert nach Durchmesser; auf der X-Achse ist jeweils die Größe der Tasche (Taschenpunktzahl), auf der Y-Achse ist jeweils die Anzahl der Taschen aufgetragen.

Diskussion:

Die durch die Filterung des Datensatzes mit dem maximal vorkommenden Durchmesser hervorgerufene Reduktion des Taschendatensatzes auf ca. 20 Prozent der Ursprungsgröße erlaubt das einfache Aussortieren von Taschen, die in ihrer maximalen Ausdehnung nicht dem Kriterium der Komplementarität zum Hauptintervall von 80% der bioaktiven Referenzliganden nicht vereinbar sind. Im Fall der TAR-RNA wird hierbei hauptsächlich die Anzahl der kleinsten Taschen reduziert, jedoch verbleibt auch nach dieser Filterung noch ca. eine Tasche pro Konformer in der Datenbank. 153.000 Konformere → 148.286 Taschen.

Vorausgreifend auf die Ergebnisse des Abschnitts 4.5.7 sei hier angemerkt, dass die entfernten Taschen hauptsächlich die größten, jedoch relativ unvergrabenen Taschen und die kleinsten, stark vergrabenen Taschen beinhalten. Hierzu kamen die kleinsten Taschen mit kleiner mittlerer Vergrabenheit.

Ausblick:

In weiteren Experimenten sollte genauer geklärt werden, wo genau auf der TAR sich diese verbleibenden Taschen befinden.

4.5.3 Zweites Kriterium „mittlere Vergrabenheit“ der Tasche

Hintergrund:

Die „mittlere Vergrabenheit“ als Kriterium beruht auf den Einzelvergrabenheiten der Taschenpunkte. Sie ergibt als gemittelter Wert ein Maß zur „globalen Vergrabenheit“ einer Tasche. Hierbei ist zu beachten, dass im Rahmen des Liganden-Designs die Stellen in der Zielstruktur von besonderem Interesse sind, welche einen Liganden aufnehmen und diesen dort durch sterische Umgrenzung halten können, selbst wenn Teile der Wechselwirkungen mit der Zielstruktur zeitweilig nicht ausgebildet werden. Somit ist die „mittlere Vergrabenheit“ ein weiteres Kriterium mit denen sich strukturell eher flache Taschen von tieferen Taschen unterscheiden lassen.

Durchführung:

Die Vergrabenheit jedes Taschenpunktes wurde im Rahmen der Taschenberechnung in der Datenbank abgelegt und kann mittels SQL-Abfrage (beispielsweise durch MySQL-Query-Browser) als relativ einfache Aggregat-Abfrage ermittelt werden.

Beispielabfrage für runID 26:

```
SELECT coordTable4pockets_Grid_MPPS.PIT_ID,  
AVG(coordTable4pockets_Grid_MPPS.burriedFactor),  
COUNT(coordTable4pockets_Grid_MPPS.burriedFactor), pocket_info_table_MPPS.runID  
FROM coordTable4pockets_Grid_MPPS, pocket_info_table_MPPS WHERE  
coordTable4pockets_Grid_MPPS.PIT_ID = pocket_info_table_MPPS.PIT_ID AND  
pocket_info_table_MPPS.runID = 26 GROUP BY coordTable4pockets_Grid_MPPS.PIT_ID  
ORDER BY AVG(coordTable4pockets_Grid_MPPS.burriedFactor)
```


Beispielergebnis:

In Tabelle 26 ist das Ergebnis der Abfrage für den PDB-Eintrag 1UUI mit nur einem Konformer angegeben. Wiedergegeben sind hier die Identifikationsnummer der Tasche in der Datenbank (PIT_ID), die Identifikationsnummer des Konformer-Datensatzes (runID), die „mittlere Vergrabenheit“ der Tasche und die Größe der Tasche als Anzahl an Volumenpunkten der Tasche.

Tabelle 26: Beispiel-Abfrageergebnis zur Taschengröße und „mittlere Vergrabenheit“; Angegeben sind die Ergebnisse für runID26 = PDB-Eintrag 1UUI. PIT_ID = Identifikationsnummer in der Datenbank, AVG(burriedFactor) = „mittlere Vergrabenheit“ der Tasche, COUNT(burriedFactor) = Größe der Tasche (Anzahl der Taschenpunkte), runID = Identifikationsnummer des Konformer-Datensatzes in der Datenbank.

PIT_ID	AVG(burriedFactor)	COUNT(burriedFactor)	runID
172340	15,8478	46	26
172339	16,1429	70	26
172349	16,1786	28	26
172355	16,2800	25	26
172347	16,3793	29	26
172341	16,4000	40	26
172344	16,4324	37	26
172338	17,5890	146	26

Diskussion:

Die „mittlere Vergrabenheit“ einer Tasche ist aus der Datenbank leicht abzuleiten und kann im Rahmen der Taschenauswahl als weiteres Kriterium der Taschenselektion eingesetzt werden. Die Verknüpfung der Taschenvergrabenheit mit der eindeutigen Identifikationsnummer PIT_ID ermöglicht weitergehende Berechnungen auf denselben Taschenkoordinaten oder weiteren hiermit über die PIT_ID verknüpften abgeleiteten Datensätzen wie der Taschenoberfläche.

Schlussfolgerungen:

Die Kombination der– mittels SQL-Abfragen– gewonnenen „mittleren Vergrabenheit“, mit der ebenfalls mit dieser Abfrage ermittelten Taschenpunktanzahl als Maß der Größe der Tasche, erlaubt eine weitere Selektion des Taschenraums nach diesem Kriterium. Des Weiteren ist die Kombination aus gefundenen Taschengrößen und ihren mittleren Vergrabenheitswerten ein Charakteristikum für den jeweiligen Konformer und kann als solches als reduziertes Unterscheidungskriterium zwischen verschiedenen Konformeren benutzt werden (Siehe Abschnitt 4.5.4).

4.5.4 Vergrabenheit gegen Taschengröße als Fingerprint für Konformere

Hintergrund:

Allgemein besteht das Ziel, den möglichen Konformerraum der Zielstruktur möglichst vollständig abzubilden. Die jeweilige Kombination der Taschengröße zur mittleren Vergrabenheit jeder Tasche ist als Resultat der Raumstruktur des jeweiligen Konformers anzusehen und beschreibt somit indirekt die Form des Konformers. Die Diversität der Beschreibungen innerhalb eines Ensembles von potentiellen Bindetaschengibt somit Auskunft darüber, wie divers der Konformerraum durch dieses Ensemble abgebildet wird.

Fragestellung:

In diesem Teil der Arbeit sollte geklärt werden, ob die Verteilung der „mittleren Vergrabenheit“ gegen die Taschengröße der MD-Läufe denen der PDB-Einträge entsprechen würde. Insbesondere sollte auch geklärt werden ob die, durch die geringere Größe der Konformer-Ensemble verursachte, starke Diskretisierung des Konformerraums durch die MD-Läufe weiter aufgelöst werden können, um zusätzliche Bereiche des Konformerraums zu erreichen.

Durchführung:

Zur Untersuchung dieses Abbildungsvermögens für die Güte-Analyse von MD-Läufen wurden vier exemplarische MD-Läufe unterschiedlicher Startkonformere mit ihren Referenzdatensätzen aus der PDB verglichen.

Hierbei wurde sich auf Strukturen beschränkt, die der TAR-Form des HIV-1 entsprechen. Dies sind im Folgenden die Einträge 1ANR, 1ARJ, 1QD3, 1LVJ, 1UUI, 1UUD, 1UTS.

In Anlehnung an die ersten vier genannten PDB-Einträge wurden auch die MD-Simulationen durchgeführt (pers. Komm. Eva Kestner).

Wichtige Anmerkung:

Bei den MD-Läufen wurden jedoch zwecks besserer Vergleichbarkeit der einzelnen Läufe die Strukturen aneinander angepasst, so dass ausschließlich mit Konformationsisomeren gearbeitet wurde. Als Beispiel sei hier die Modifikation des 1LVJ-Eintrags genannt: Dabei wurden die beiden endständigen Basen entfernt, um diese dem 1ANR-Eintrag anzupassen, bevor die Simulationen durchgeführt wurden. Die so angepassten Startstrukturen und ihre daraus resultierenden Läufe sollten somit gleichberechtigt behandelt werden dürfen und sich nur durch ihre Startkonformation unterscheiden. So wurde gewährleistet, dass die Läufe bei ausreichender Laufzeit ineinander konvergieren können, um demzufolge gemeinsam von verschiedenen Startpunkten aus letztlich den gesamten gemeinsamen Konformerraum abzubilden. Durch die genannte Anpassung der Startkonformere sind die MD-Läufe jedoch nicht alle vorbehaltlos mit den dazugehörigen PDB-Einträgen gleichen Namens vergleichbar, da es sich hierbei nicht mehr um dieselbe Molekularstruktur handelt.

Arbeitsablauf:

Für jede der Konformer-Sammlungen (MD-Läufe und PDB-Einträge) wurden folgende Schritte durchgeführt:

1. Extraktion der mittleren Vergrabenheit und Anzahl der Taschenpunkte der Taschen.
2. Auftragung der mittleren Vergrabenheit gegen die Taschenpunktzahl.
3. Gegenüberstellung der MD-Lauf-Projektionen und PDB-Einträge.

Ergebnisse:*Einzelbetrachtungen:*

Resultierend aus der Extraktion und Analyse der Daten konnten die Abbildungen 88 und 89 erstellt werden. Abbildung 88 zeigt die Gegenüberstellung der MD-Läufe gegen die PDB-Einträge aufgrund der mittleren Vergrabenheit und der Taschengröße. Abbildung 89 zeigt die Auftragung der mittleren Vergrabenheit gegen die Taschengröße für die Taschen der PDB-Einträge 1UUI, 1UUD und 1UTS. Für die letztgenannten drei stehen keine MD-Läufe zur Verfügung. Die PDB-Einträge unterscheiden sich wie die MD-Läufe in der Anzahl der zur Verfügung stehenden Konformere, so dass dies in die Bewertung der Graphiken einfließen muss. Insbesondere 1UUI und 1UUD bestehen jeweils ausschließlich aus einem Konformer.

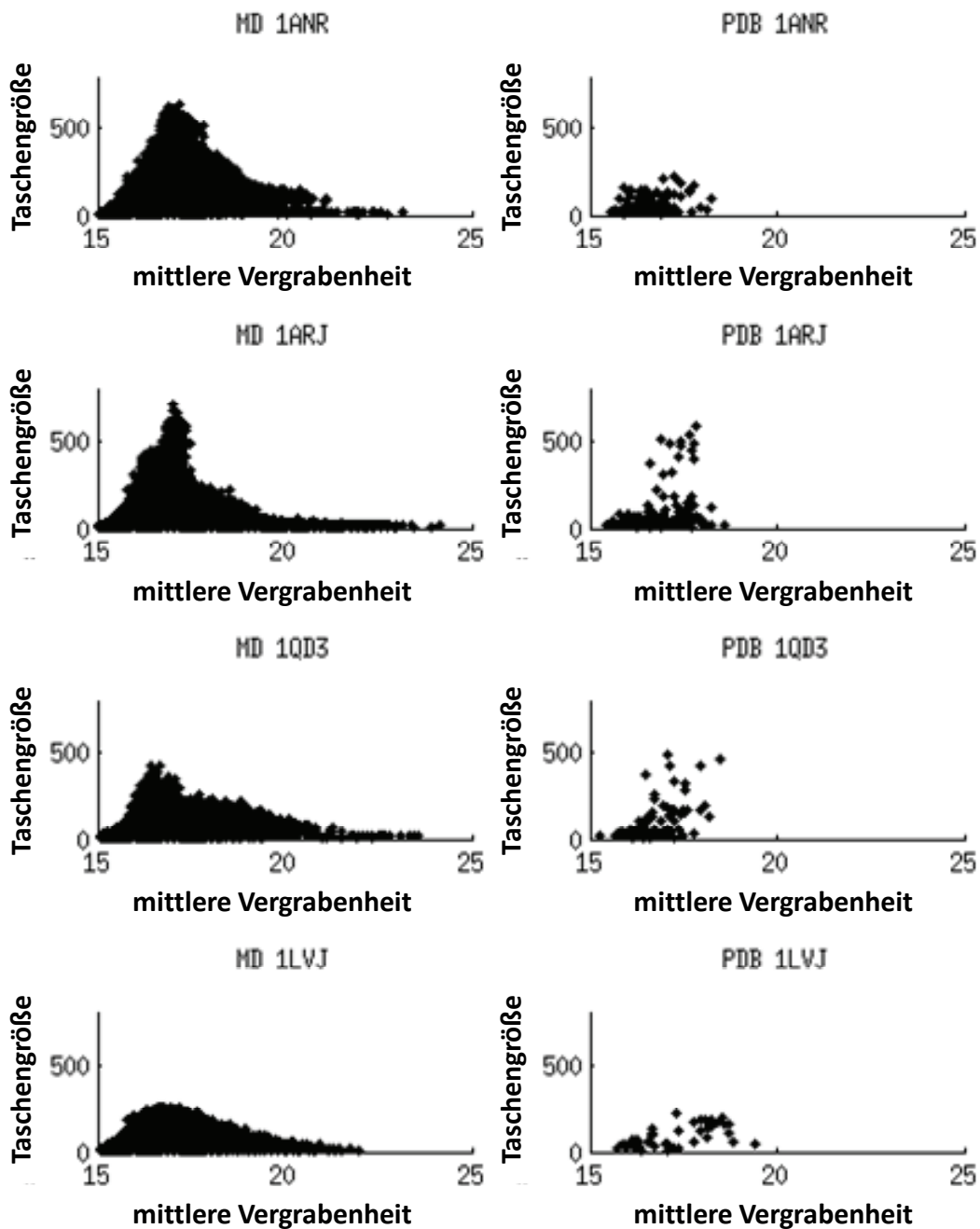


Abbildung 88: Gegenüberstellung MD-Läufe und der PDB-Einträge für die Darstellungen der Taschengröße gegen die mittlere Vergrabenheit

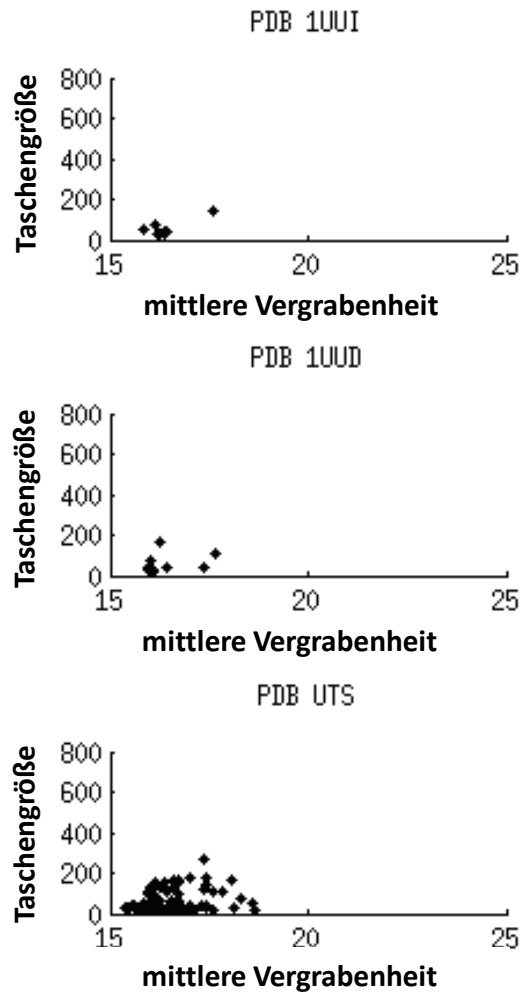


Abbildung 89: Taschengröße gegen mittlere Vergrabenheit für die PDB-Einträge 1UUI, 1UUD und UTS

Gesamtbetrachtung der MD-Läufe:

Abbildung 90 enthält die gemeinsame Projektion der vier MD-Läufe, welche in vier verschiedenen Farben erfolgte. Aufgrund der unterschiedlichen Flächengröße der Projektionen der einzelnen Läufe wurden diese Projektionen sequenziell durch Überlagerung der jeweils größere Flächen beanspruchenden MD-Läufe durch die kleinere Flächen beanspruchenden MD-Läufe aufgebaut. Die durch die kleineren Flächen überlagerten Bereiche beinhalten auch bei den größeren Flächen beanspruchenden MD-Läufen Datenpunkte in diesen Bereichen, wie aus Abbildung 88 ersichtlich ist.

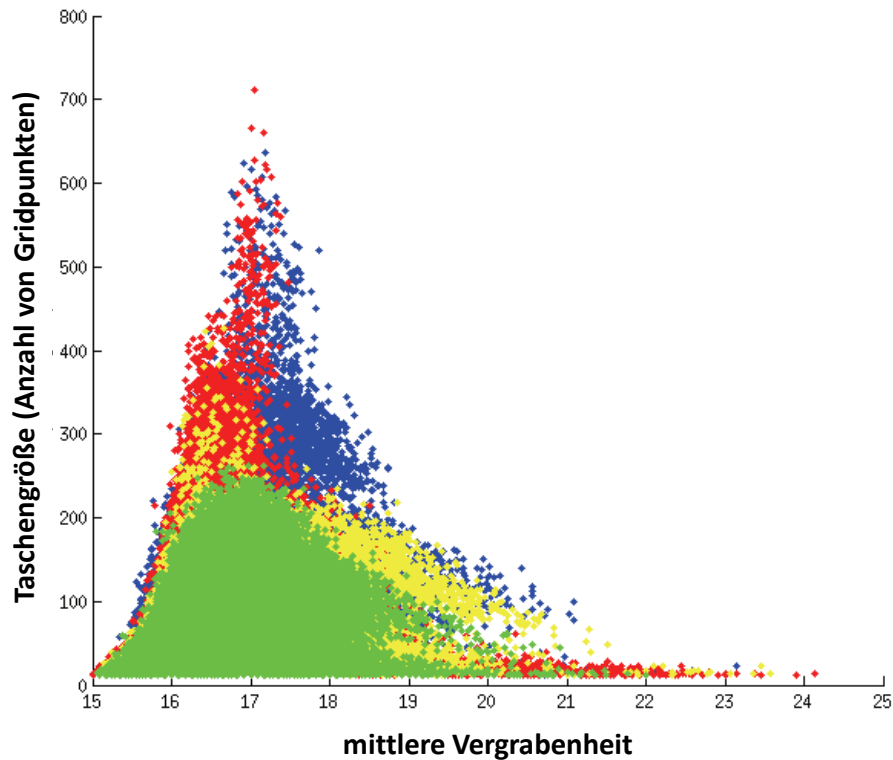


Abbildung 90: Taschengrößen gegen „mittlere Vergrabenheit“ für die MD-Läufe; blau = 1ANR, rot = 1ARJ, gelb = 1QD3; grün = 1LVJ

Diskussion:

Einzelbetrachtungen:

Die Abbildung 90 mit ihrer Gegenüberstellung der MD-Läufe gegen die PDB-Einträge gleichen Namens zeigt, dass die MD-Läufe einen weiteren strukturellen Bereich des möglichen Konformerraums für die Taschen abdecken, als dies die PDB-Einträge tun. Insbesondere ist hier nicht zuletzt durch die höhere Sampling-Rate eine stärkere Abdeckung des Konformerraums zu erwarten gewesen, der sich auch in der Verteilung möglicher mittlerer Vergrabenheit und Größe der Taschen zeigt.

Für den 1ANR-Fall zeigt sich, dass die Bereiche des MD-Laufs deutlich über den der ursprünglich im Ensemble des 1ANR-PDB-Eintrags beinhalteten Variationen hinausgehen. Für den 1ARJ-Fall sind die Variationen nicht so prägnant, jedoch deckt hier das Ensemble des PDB-Eintrags auch einen weiteren Bereich, insbesondere in der Region der größeren Taschen mit größerer mittlerer Vergrabenheit ab. Dieser Bereich wird auch vom Ensemble des MD-Laufs erfasst.

Für den 1LVJ-Fall deckt das Ensemble des PDB-Eintrags Bereiche ab, die nicht vom dazugehörigen MD-Lauf erreicht wurden. Hier konnte die MD-Simulation den erfassten Taschenraum zwar ebenfalls glätten, aber nicht alle Bereiche abbilden, die laut experimenteller Vorgabe möglich gewesen wären.

Für den 1QD3-Fall wurden wiederum – wie bei 1ANR und ARJ oben erwähnt – die Bereiche des PDB-Ensembles abgebildet. Auffällig ist hierbei aber die insgesamt für beide Ensembles (PDB und MD-Lauf) geltende Vermeidung größerer und gleichzeitig stärker vergrabener Taschen, wie dies in den anderen drei PDB-Ensembles erkennbar war.

Die Projektionen für die PDB-Einträge 1UUI, 1UUD und 1UTS für die keine MD-Läufe vorliegen, decken sich mit den Bereichen der bereits diskutierten Projektionen. Insbesondere zeigen sich hier keine Datenpunkte in den Bereichen, die durch die MD-Läufe nicht bereits

abgedeckt wurden. Die Betrachtungen der Einzelkonformere der PDB-Einträge von 1UUI und 1UUD zeigen in Relation zu den anderen Auswertungen nur einen kleinen Bereich des Konformerraums.

Gesamtbetrachtung der MD-Läufe:

Der Vergleich der Verteilungen der Datenpunkte für die MD-Läufe zeigt zum Teil deutliche Unterschiede in der Verteilung zwischen den Läufen. Insbesondere die Läufe der 1QD3- und 1LVJ-Startstruktur wiesen eine geringere Abdeckung des so erfassten Konformerraums für die Taschenbildung auf, als dies bei 1ANR und 1ARJ zu erkennen ist. Hierbei wurde während der MD-Läufe besonders der Bereich der großen Taschen mit starker mittlerer Vergrabenheit von den Läufen 1QD3 und 1LVJ vermieden.

Schlussfolgerungen:

Die MD-Läufe scheinen den möglichen Konformerbereich und somit den verfügbaren Taschenraum relativ gleichmäßig zu repräsentieren. Jedoch werden einige Bereiche, die durch die PDB-Einträge vorgeschlagen werden, im Rahmen dieser MD-Simulationen nicht erschlossen. Dies könnte einerseits daran liegen, dass die MD-Simulationen aufgrund der bisherigen Laufzeit (Simulationszeit) noch nicht den gesamten möglichen Konformerraum erreichen. Andererseits könnte dies dadurch bedingt sein, dass die Strukturaufklärungen (NMR-Spektren) nicht bei Raumtemperatur durchgeführt wurden, wie dies bei den MD-Läufen simuliert wurde. Hieraus ergäbe sich, dass die Konformere in niederenergetische Bereiche „kristallisiert“ wurden, die so in den MD-Lauf-Ensembles nur äußerst unwahrscheinlich anzutreffen gewesen wären. Dies ist jedoch zu diesem Zeitpunkt rein spekulativ.

Hingegen trat die größte Abweichung zwischen dem PDB-Eintrag 1QD3 und den MD-Läufen auf. Da es sich bei 1QD3 um eine gleichzeitige Strukturaufklärung eines Liganden mit der TAR handelt, ist hier die wahrscheinlichste Erklärung, dass das Vorhandensein des Liganden diese „extremen“ Strukturen induziert hat, die ohne diesen Ligand nicht vorkommen würden. Interessant wäre zu klären, ob diese Konformere bei einer veränderten (erhöhten) Simulationstemperatur jedoch ebenfalls – wenn auch wahrscheinlich sehr kurzlebig – auftreten würden.

4.5.5 Pareto-Ranking zur Selektion der undominierten Lösungen der Multi-Objektiven-Optimierung

Die „optimale“ Tasche zeichnet sich durch den bestmöglichen Kompromiss aus Taschengröße (maximale Anzahl an möglichen Wechselwirkungen) und der „mittleren Vergrabenheit“ der Tasche (maximale „mittlere Vergrabenheit“ ergibt die maximal begrenzte Tasche) aus.

Diese Taschen befinden sich auf der Pareto-Front der unter diesen Kriterien undominierten Lösungen des Optimierungsproblems, welche für diesen Datensatz aus 44 undominierten Punkten besteht und dessen Punkte nach dem Pareto-Ranking (Gillet et al. 2002a; Gillet et al. 2002b) einer gleich optimalen Kombination der zwei verwendeten Merkmale entsprechen. Das hier durchgeführte Pareto-Ranking wurde mittels einer in *Pandora* integrierten Java-Klasse umgesetzt, die hierzu von Andreas Schüller erhalten werden konnte.

4.5.6 Die Diagonale zur Selektion der Taschen, die die mittlere Vergrabenheit und die Taschengröße gleichermaßen beinhalten ohne ein Kriterium zu bevorzugen.

Auf der Diagonale der Auftragung der Taschengröße gegen die mittlere Vergrabenheit liegen Taschen, welche einen ausgewogenen Kompromiss aus den genannten Kriterien erfüllen.

Jedoch sind auf dieser Achse ebenfalls sehr kleine flache Taschen zu finden, so dass dieses Auswahlkriterium noch weiter eingeschränkt werden muss; siehe hierzu den folgenden Abschnitt 0.

Gleichgewichtete Maximierung der Taschengröße und mittleren Vergrabenheit

Die denkbar beste Tasche wäre demzufolge die Tasche, die beide Werte relativ gleichmäßig maximiert und somit wäre es die größte Tasche bei gleichzeitiger Beachtung der Vergrabenheit. Diese Tasche kann am Schnittpunkt der Pareto-Front mit der Diagonalen, welche zwischen den beiden Kriterien liegt, ermittelt werden.

Diese Taschenposition ist in Abbildung 91 als P3 markiert worden und stellt das erreichbare Optimum für die genannten Kriterien innerhalb der Taschenpopulation der MD-Läufe dar. Nachfolgend sind Beschreibungen für die sonstigen als am charakteristischsten geltenden Merkmalskombinationen von Abbildung 91 erläutert:

- P1 stellt den ungünstigsten Fall dar – kleine flache Taschen.
- P2 stellt die größte Tasche des MD-Laufs dar, unter relativer Missachtung des Vergrabenheitskriteriums.
- P3 stellt das erreichbare Optimum der Taschensuche basierend auf den bisherigen Überlegungen dar.
- P4 stellt die vergrabendste Tasche dar, unter relativer Missachtung des Größenkriteriums.
- P5 stellt den Median der Verteilung auf der Diagonalen zwischen den beiden Kriterien dar.

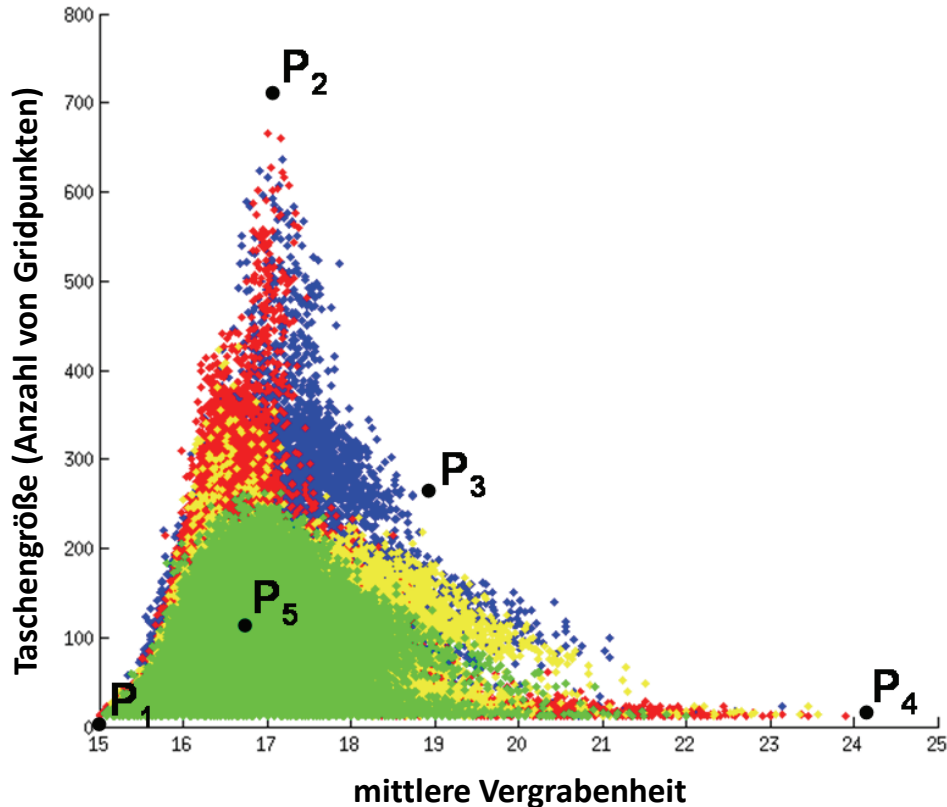


Abbildung 91: Taschengröße gegen die mittlere Vergrabenheit für alle Taschen des gesamten MD-Datensatzes; Als charakteristischste Positionen für die Betrachtung wurden P1 bis P5 markiert; siehe hierzu die Ausführungen im Text.

4.5.7 Analyse der Verteilung der Datenpunkte für die Auftragung mittlerer Vergrabenheit und Taschengröße

Hintergrund:

Im Rahmen der Arbeit wurde eine Analyse durchgeführt, die die Auswirkungen der Filterung – basierend auf den maximalen Taschendurchmessern – auf die Populationshäufigkeit bestimmter Merkmalskombinationen der Taschenteilmengen untersuchte. Hierzu wurde die Populationshäufigkeit der einzelnen Datenpunkte bestimmter „mittlerer Vergrabenheit“ und Größe der Taschen bestimmt. Untersucht wurde ferner der Effekt auf diese Populationsverteilung nach Anwendung des „maximalen Durchmesser“-Selektionskriteriums.

Durchführung:

Für die Untersuchung wurde ein m-Skript entwickelt, welches die Populationshäufigkeit der einzelnen Datenpunkte in einer Farbkodierung den erstellten Abbildungen zur mittleren Vergrabenheit und Taschengröße umsetzt. Hierzu wird jeweils die Taschenanzahl mit derselben Merkmalskombination zusammengefasst und dieser Anzahl in Relation zur maximal in dieser Auftragung gefundenen Populationsmenge eine Farbe aus dem Farbbalken zugewiesen. Die Auswertung umfasst an dieser Stelle zwei Datensätze:

- Alle gefundenen Taschen für die Konformere der MD-Laufe 1ANR, 1ARJ, 1LVJ und 1QD3.
- Teilmenge der Taschen aus dem Datensatz aller gefundenen Taschen für die Konformere der MD-Laufe 1ANR, 1ARJ, 1LVJ und 1QD3, ausgewählt mit dem „maximalen Durchmesser“-Filter.

Ergebnisse:

In Abbildung 92 ist die Auswertung der Populationsverteilung der Kombinationen der „mittleren Vergrabenheit“ und Größe der Taschen für den gesamten MD-Datensatz dargestellt.

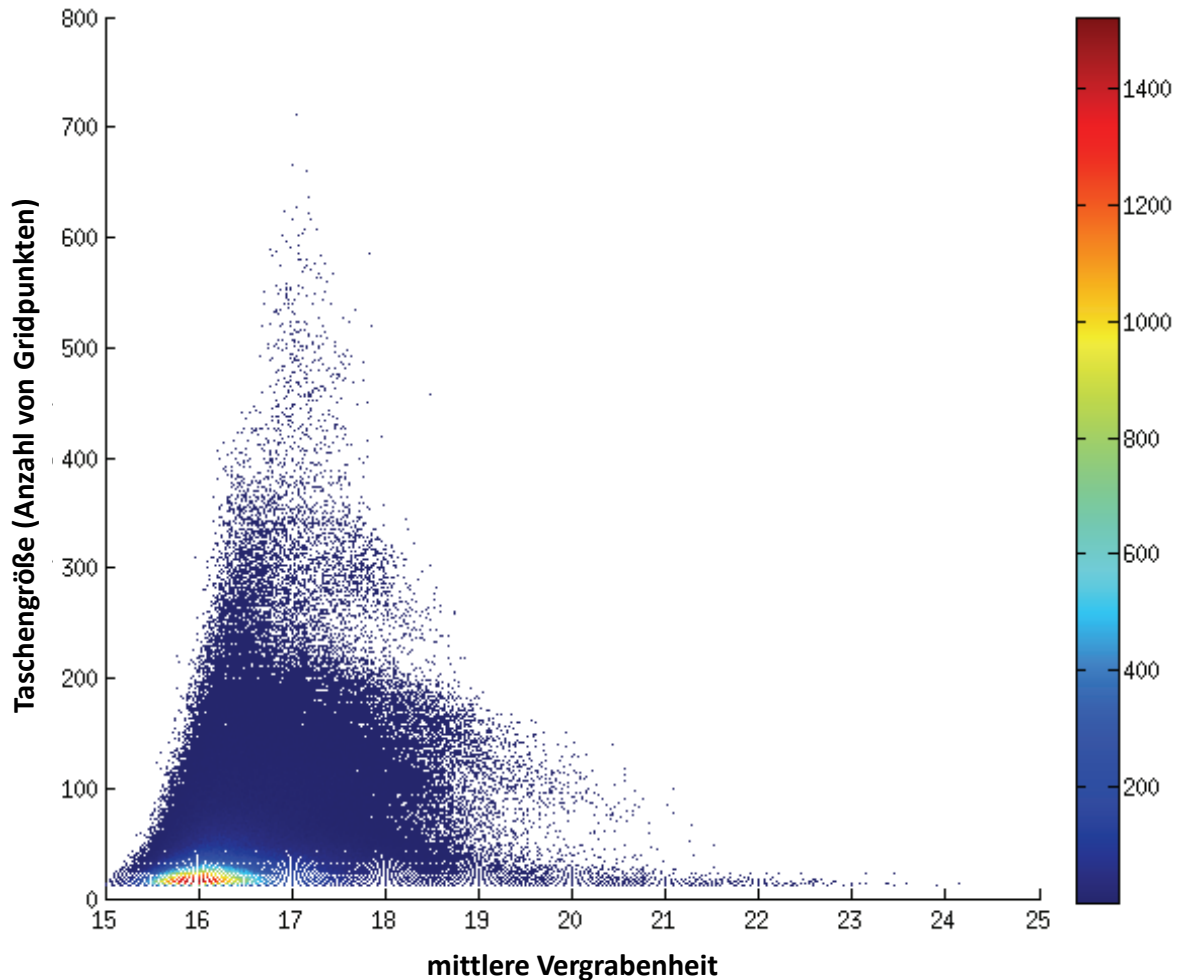


Abbildung 92: Gesamter MD-Datensatz – Taschengröße gegen mittlere Vergrabenheit; die Farbkodierung gibt die Häufigkeit der Kombination im Set an; die meisten Taschen sind somit klein und wenig vergraben.

In Abbildung 93 ist die Auswertung der Populationsverteilung der Kombinationen der „mittleren Vergrabenheit“ und Größe der Taschen für die Teilmenge des gesamten MD-Datensatz dargestellt, welcher durch Anwendung des „maximalen Durchmesser“-Filters aus Abschnitt 4.5.2 erhalten wurde.

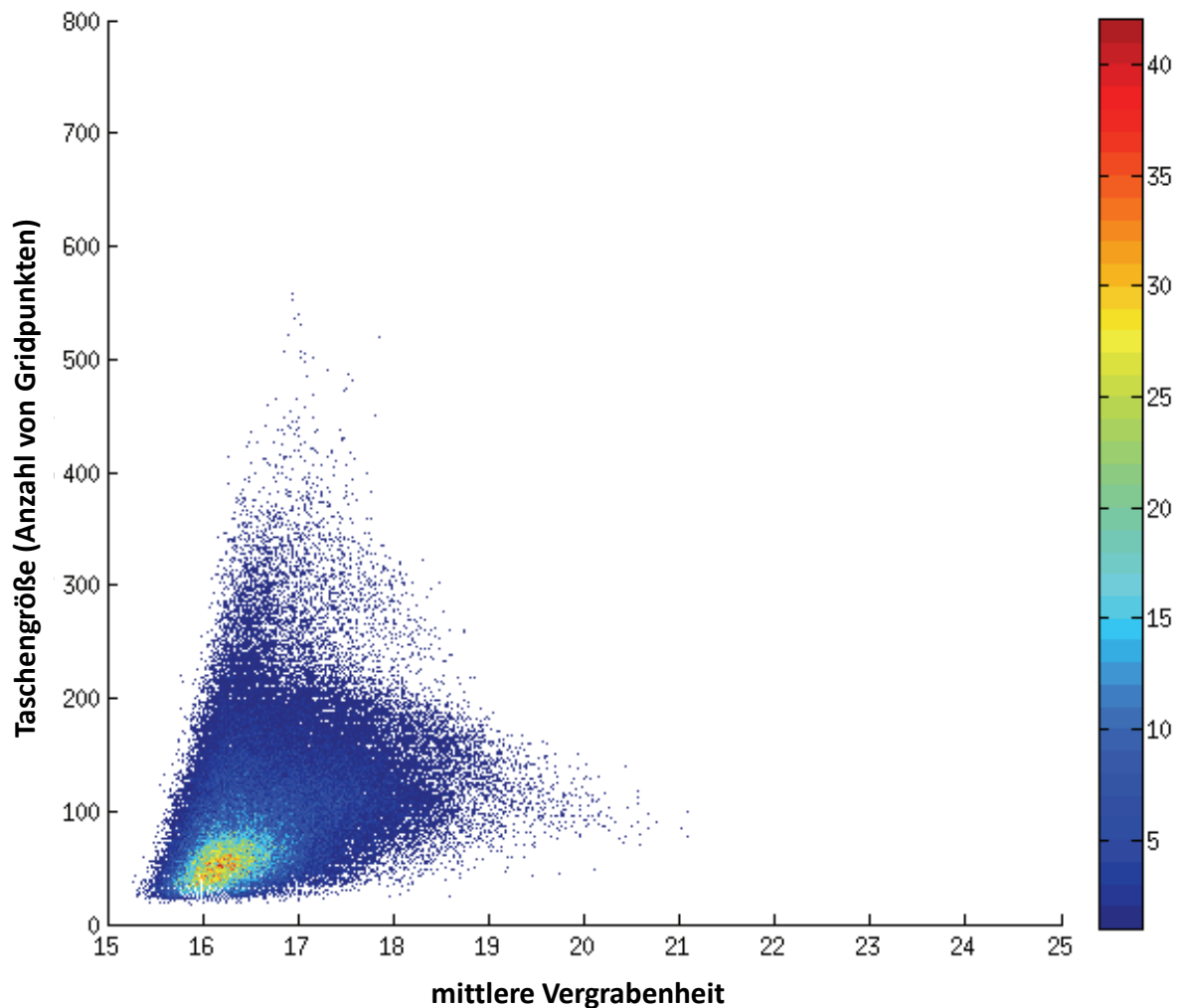


Abbildung 93: Selektion der Taschen des MD-Datensatzes basierend auf dem optimalen Taschendurchmesserbereich– Taschengröße gegen „mittlere Vergrabenheit“; die Farbkodierung gibt die Häufigkeit der Kombination im Set an; die meisten Taschen sind zwar immer noch klein und wenig vergaben, jedoch wurde ein Großteil der als unnütz definierten Extreme entfernt.

Diskussion:

Abbildung 92 und 93 zeigen beide eine Hauptverteilung der Taschen auf relativ kleine Taschen. Größere Taschen und größere Vergrabenheit kommen in den Konformeren der MD-Läufe seltener vor. Die Anwendung des „maximalen Durchmesser“-Filters bewirkt eine Reduktion des Taschendatensatzes, durch die die beiden extremen Kombinationen aus großer Taschengröße und relativ kleiner „mittlerer Vergrabenheit“ sowie kleiner Taschengröße mit schwacher bis starker Vergrabenheit aus dem Datensatz abgetrennt werden können. Die extrem kleinen flachen Taschen, welche einen Großteil des ursprünglichen Datensatzes in Abbildung 92 ausmachen, konnten durch diesen Filter zuverlässig beseitigt werden; siehe hierzu Abbildung 93 im Vergleich. Der Filter bewirkt somit eine Fokussierung des Datensatzes um die als optimal anzusehende Diagonale zwischen „mittlerer Vergrabenheit“ und Taschengröße.

Schlussfolgerung:

Da der „maximalen Durchmesser“-Filter die Taschenpopulation auf ca. 20 Prozent des Ursprungsdatensatzes (siehe Abschnitt 4.5.2) reduziert, kann er zur Vorbereitung der Daten für weitere Analysen eingesetzt werden. Die Reduktion erfolgt hier anscheinend ohne einen Verlust des als optimal angesehenen Zielbereichs, um P3 in Kauf nehmen zu müssen.

Auswahl radial um den optimalen Punkt P3 des Datensatzes

Da der Punkt P3 als angestrebtes Optimum einer Tasche unter gleichzeitiger Beachtung der Taschengröße und Vergrabenheit gilt, sollten Taschen in seiner relativen Umgebung ebenfalls interessanter für die Auswahl als potentielle Dockingtaschen sein, als Taschen aus anderen Bereichen der Taschenpopulation. Deshalb wurde nach einem Verfahren gesucht, um diese aus der Menge der Taschen zu extrahieren. Hierbei bot sich – da keine Richtungsbevorzugung innerhalb dieser Optimierungslandschaft stattfinden sollte – die radiale Auswahl der Taschen um P3 an.

Durchführung:

Zur Auswahl um den optimalen Punkt P3 wurden die Wertebereiche der einzelnen Achsen normiert und anschließend mittels euklidischer Distanz die Entfernung jeder Tasche zu P3 ermittelt. Als Basis für die Normierung wurden die jeweils auf dieser Achse befindlichen Minima und Maxima der Werteverteilungen des Gesamtdatensatzes herangezogen. Basierend auf dieser Rechnung ist es möglich, Teilmengen des Datensatzes verschiedener Größe um P3 für die weitere Betrachtung zu extrahieren.

Zur ersten Analyse der Populationsverteilung wurde eine Auftragung der sortierten Radien zu P3 für die gesamte MD-Lauf-Taschenpopulation durchgeführt; siehe Abbildung 94.

Die Darstellung der auf unterschiedliche Radien bezogenen Auswahl der Taschen ist in Abbildung 95 wiedergegeben. Die Erstellung erfolgte mittels m-Skripten. Die Populationsgrößen dieser Teilmengen des Taschendatensatzes befindet sich in Tabelle 27.

Zusätzlich zur einfachen Darstellung der ausgewählten Taschenbereiche wurde noch eine populationsbasierte Projektion dieser Teilmengen durchgeführt, wie dies im Abschnitt 4.5.7 beschrieben wurde. Diese dient der Verdeutlichung der Häufigkeitsverteilungen der Merkmalskombinationen innerhalb der ausgewählten Radien. Als Beispiele wurden hier die Radien 0,3; 0,2 und 0,1 gewählt.

Ergebnisse:

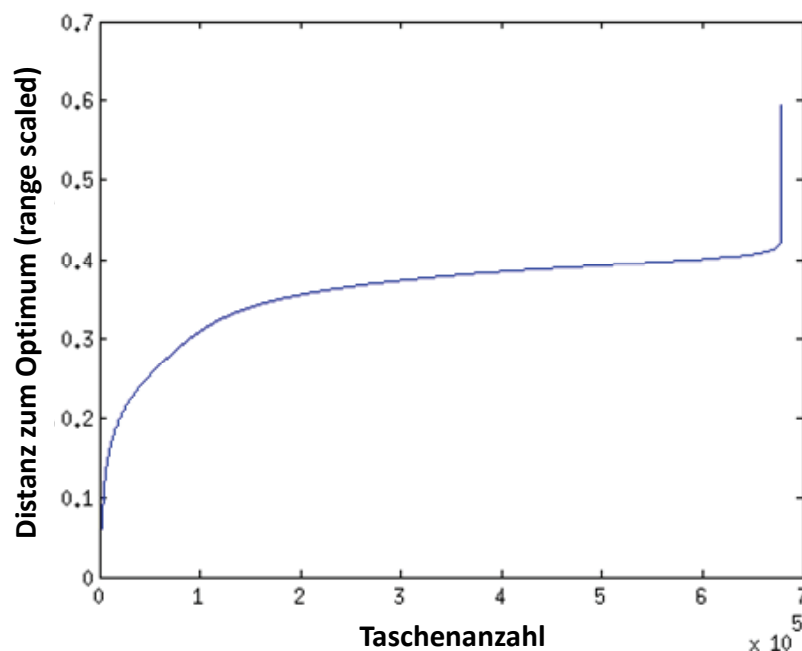


Abbildung 94: Verteilung der Radien um P3 für den von den gesamten MD-Läufen abgeleiteten Taschendatensatz, Auftragung erfolgte sortiert nach der Größe des Radius.

Tabelle 27: Populationsgrößen der Teilmengen des Taschendatensatzes nach Einsatz des Radiusfilters um Punkt P3 der Auftragung von „mittlerer Vergrabenheit“ und Taschengröße.

Radius	verbleibende Population	Prozent der Gesamtpopulation; n = 678259
0,3	70.071	10,331
0,2	13.512	1,992
0,1	2.329	0,343
0,05	608	0,090
0,04	364	0,053
0,03	181	0,027
0,02	63	0,009
0,01	14	0,002

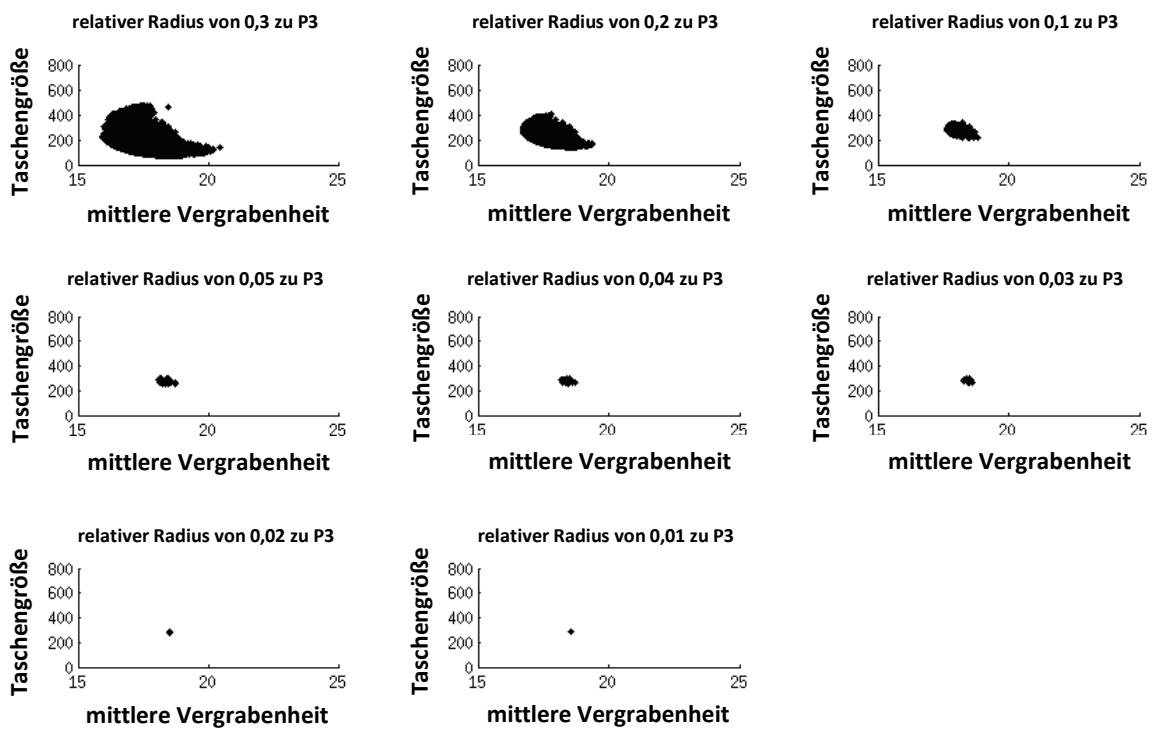


Abbildung 95: Fokussierung der Taschenauswahl auf den als optimal angesehenen Bereich um P3; Auswahl erfolgte aufgrund verschieden großer Selektionsradien von 0,3 bis 0,01

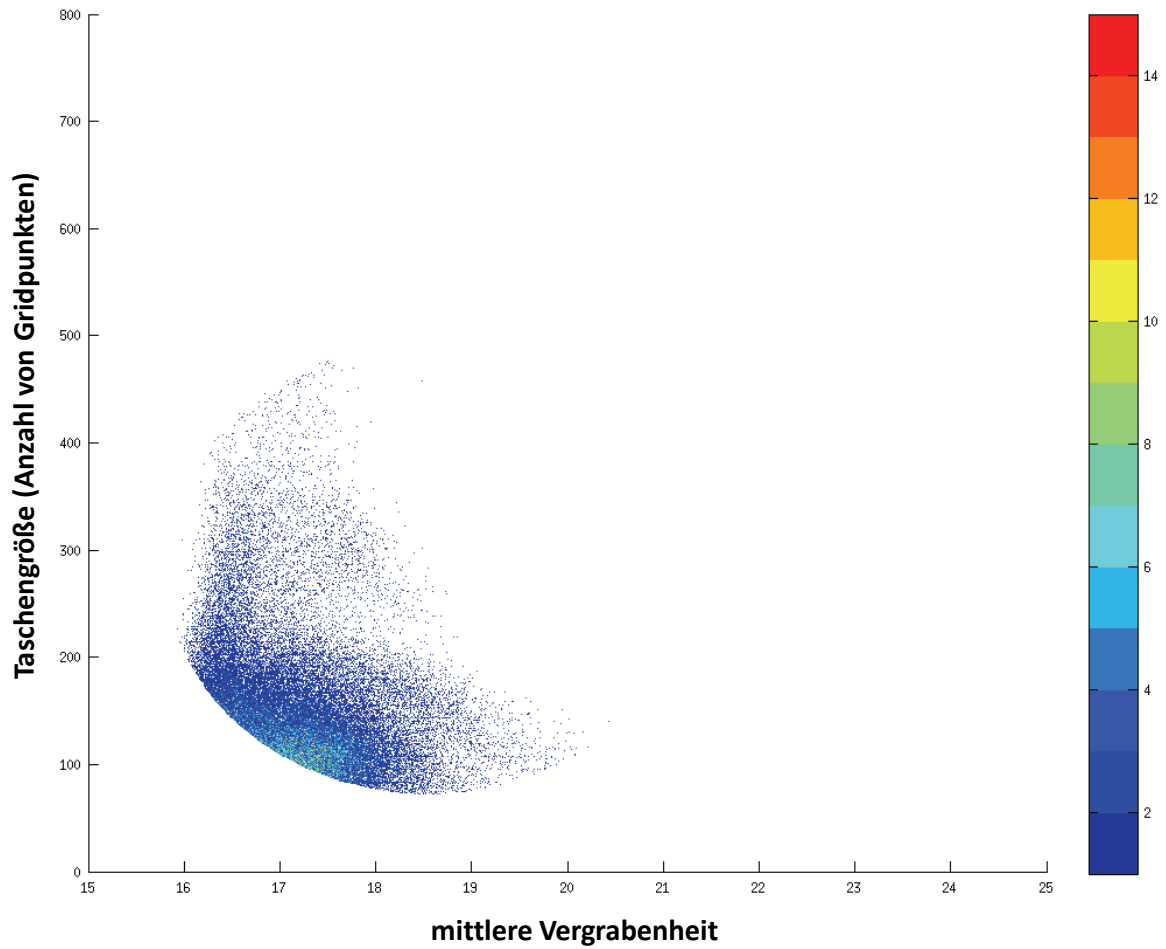


Abbildung 96: Populationsverteilung innerhalb der Auftragung der Taschengröße (Y-Achse) und der „mittleren Vergrabenheit“ (X-Achse). Auswahlradius = 0,3; 70.071 Taschen.

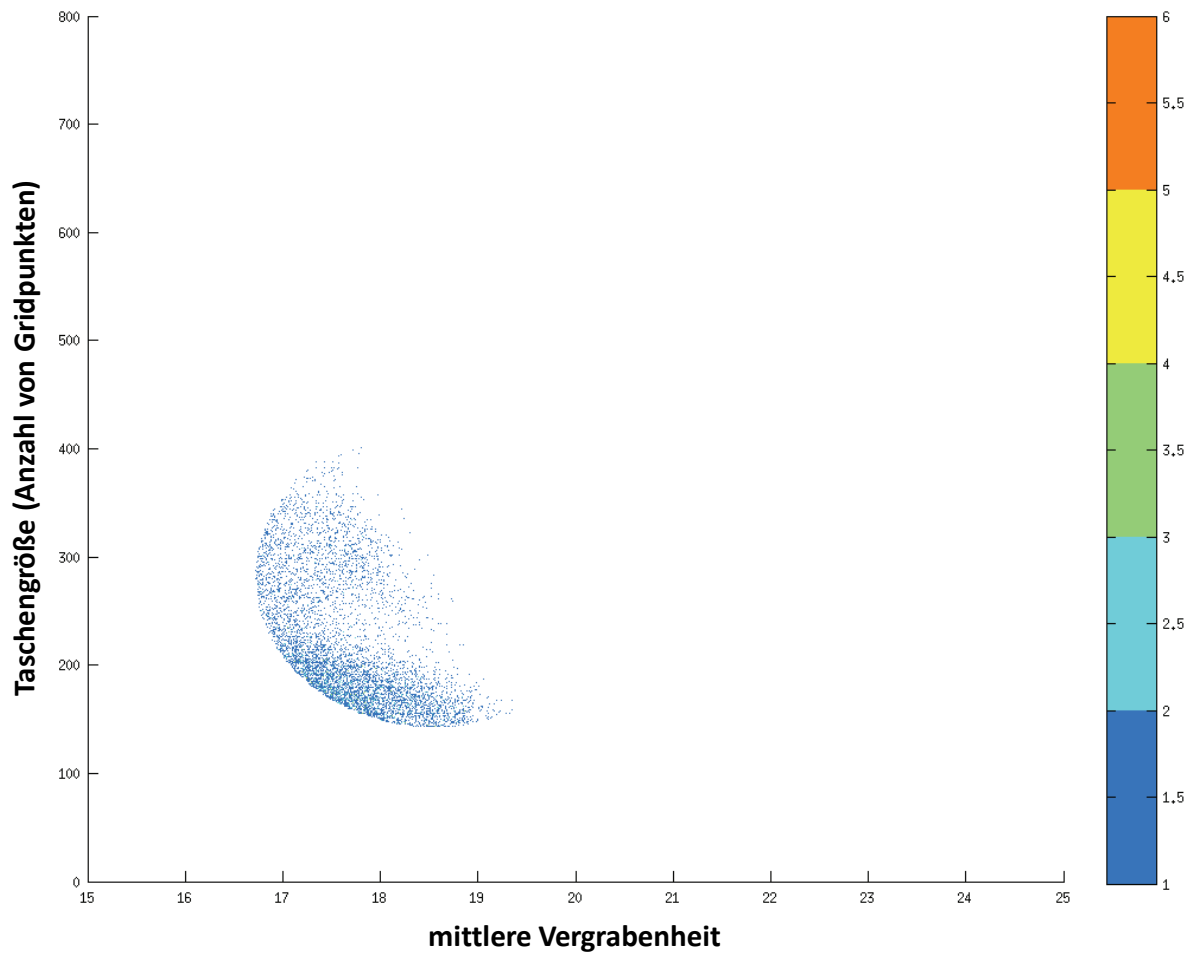


Abbildung 97: Populationsverteilung innerhalb der Auftragung der Taschengröße (Y-Achse) und der „mittleren Vergrabenheit“ (X-Achse). Auswahlradius = 0,2; 13.512 Taschen.

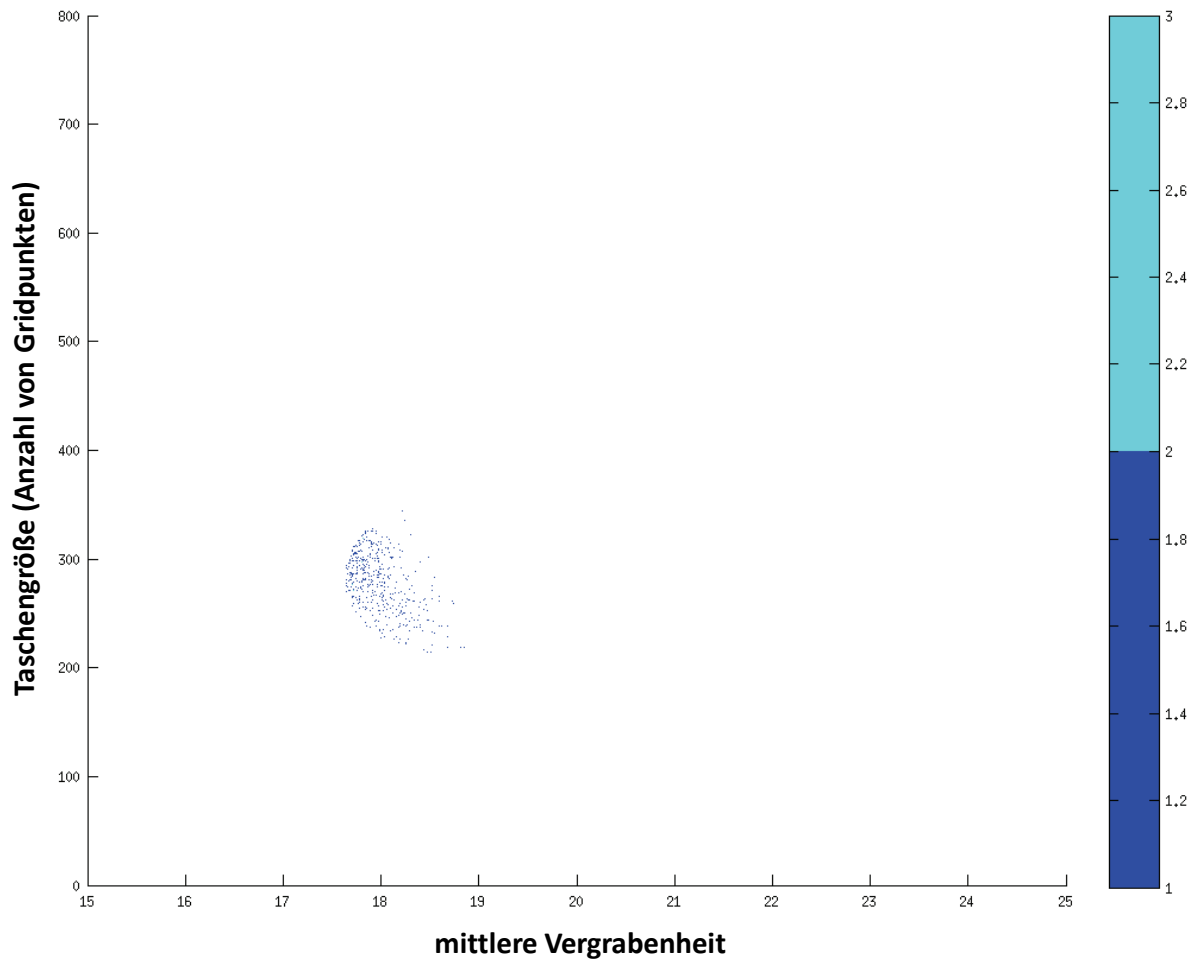


Abbildung 98: Populationsverteilung innerhalb der Auftragung der Taschengröße (Y-Achse) und der „mittleren Vergrabenheit“ (X-Achse). Auswahlradius = 0,1; 2.329 Taschen.

Diskussion:

Die Populationsverteilung um den Punkt P3 zeigt laut Abbildung 94 für den hier untersuchten von den gesamten MD-Läufen abgeleiteten Taschendatensatz keine gleichförmige Verteilung an. Direkt um P3 herum nimmt die Anzahl der gefundenen Punkte mit steigendem Radius erst einmal nahezu quadratisch bis zu einem Radius von ca. 0,32 zu. Im darauf folgenden Bereich bis etwa 0,4 ist eine relativ konstante Steigerung der Populationszahlen feststellbar. Ab einem Radius von etwa 0,4 ist eine starke Verminderung des weiteren Zuwachses der Populationszahlen zu verzeichnen, wie sie ähnlich im Anfangsbereich beobachtet wurde. Dieser letzte Sprung in der Kurve rührt daher, dass der Punkt P3 sich in der relativen Mitte des Graphen befindet, und sich somit von diesem relativen Zentralpunkt aus die Extrempunkte der dreieckigen Datenpunktverteilung erst durch die letzten Radienerweiterungsschritte erschlossen werden. Diese Merkmalskombinationen waren innerhalb des MD-Datensatzes jedoch nur äußerst selten zu finden. Hierbei handelt es sich insbesondere um die besonders kleinen, aber mit starker mittlerer Vergrabenheit bestimmten Taschen.

Die jeweiligen Populationsgrößen der Teilmengen können als Kriterien gewertet werden, wie weit der Radius gezogen werden muss, um eine ausreichende Teilmengengröße für die weitere Verarbeitung zu erreichen. Diese Radiengröße kann auch aus

Tabelle 27 erschlossen werden, die diskrete Beispielgrenzen enthält. Die Gesamtpopulationen der Teilmenge bei einem gewählten Selektionsradius von 0,3 = 70.071 Taschen; 0,2 = 13.512 Taschen; 0,1 = 2.329 zeigen eine unübersehbare Reduktion der Taschenpopulationsgrößen bei kleiner werdendem Radius.

Abbildung 95 zeigt hier die Umsetzung des radialen Auswahlverfahrens anhand unterschiedlicher Radien zur Erzeugung kleinerer Teilmengen für verschiedene Radien; verwendet wurden die Beispielradien aus.

Um eine genauere Analyse der Populationsverteilung innerhalb der erhaltenen Teilmengen zu ermöglichen wurden Abbildungen 96, 97 und 98 angefertigt. Diese zeigen die relative Populationsgröße einer Merkmalskombination innerhalb der Graphik als Farbkodierung des jeweiligen Datenpunktes. Hierbei ist erkennbar, dass um den Punkt P3 keine relativen Häufungen von Merkmalskombinationen auftreten.

Schlussfolgerung:

Da sich laut den Voruntersuchungen mittels des „maximaler Durchmesser“-Filters in Abschnitt 4.5.7 zeigen ließ, dass die Taschen mit einer zu geringen Größe und unzureichender Vergrabenheit keine guten potentiellen Bindestellen für bioaktive Liganden ergeben würden, ist die Auswahl radial um den Punkt P3 eine weitere geeignete Maßnahme, um den zu betrachtenden Taschenraum automatisch zu verkleinern.

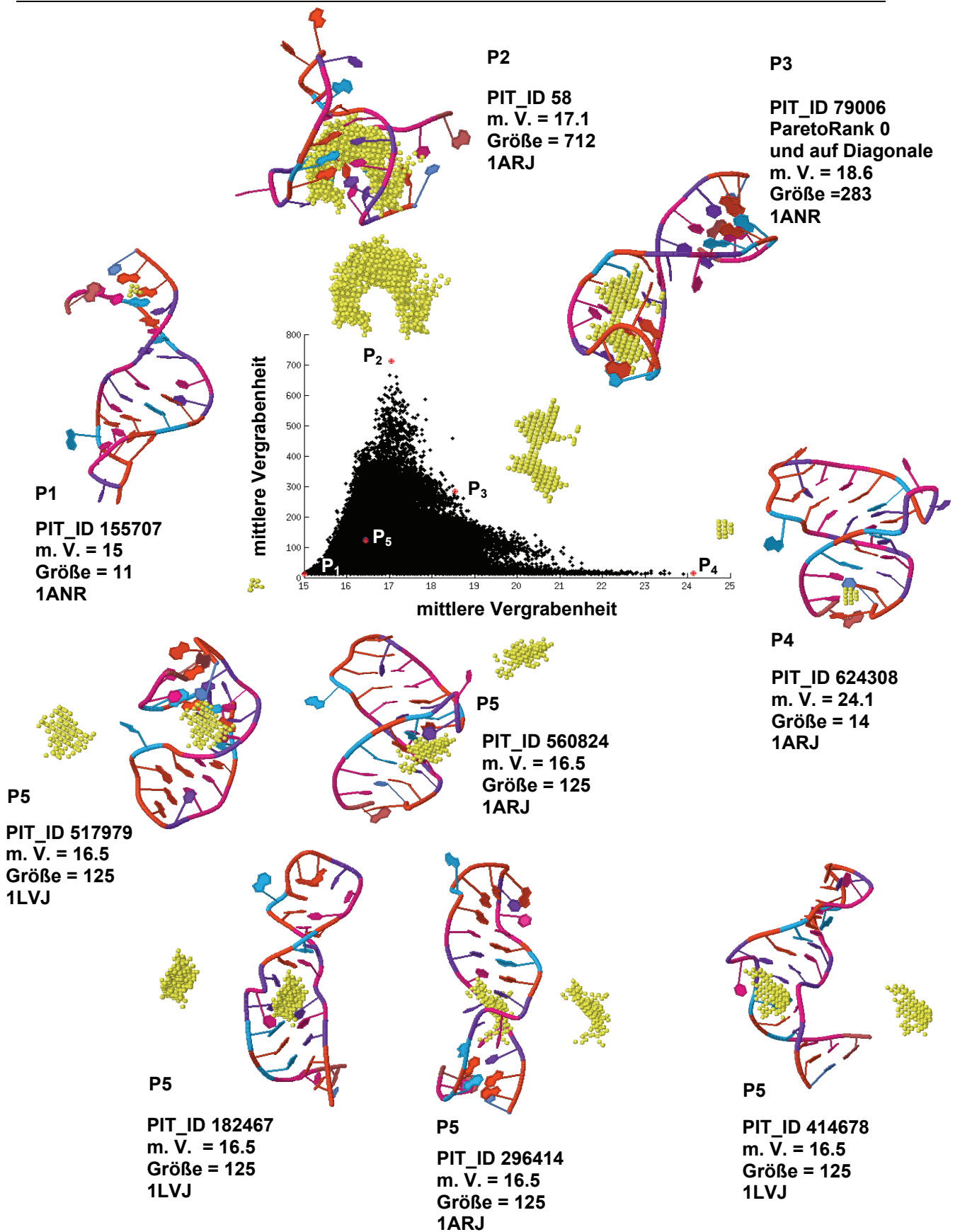


Abbildung 99: Visualisierung der Referenztaschenauswahl nach Vergrabenheit und Größe der Tasche; mittlere Vergrabenheit = „m. V.“, Größe = Anzahl Taschenpunkte; 1ANR, 1ARJ, 1LVJ stehen jeweils für die MD-Läufe aus denen die Konformere entstammen. PIT_ID = eindeutiger Bezeichner der jeweiligen Tasche in der Datenbank.

4.5.8 Das Taschenoberflächenbildungspotential PSGP als Gütekriterium für die Taschenauswahl

Hintergrund:

Die optimale Tasche für das Docking wäre diejenige innerhalb einer Population, die für die Interaktion mit einem Liganden die zeitlich stabilsten Interaktionen bietet.

Da durch die brownische Molekularbewegung die Taschen jedoch ständig entstehen und vergehen sowie ihre Form dabei variieren, sind die Taschen am interessantesten, die innerhalb der Population aus den zeitlich stabilsten Taschenoberflächenbereichen bestehen.

Diese Taschenoberflächenbereiche können als diejenigen Taschenoberflächen definiert werden, die über alle einzelnen Konformere der Datenbank am häufigsten in Taschen zu finden sind.

Da ferner nicht nur das pure Vorhandensein eines Oberflächenstückes eines Atoms innerhalb einer Tasche ausschlaggebend ist, sondern noch deren räumliche Größe einen Einfluss auf die Interaktionsmöglichkeit/-wahrscheinlichkeit eines Liganden innerhalb der Tasche hat, ist somit zur Betrachtung der Wechselwirkungswahrscheinlichkeit die Größe der Oberfläche des Atoms mitzubewerten.

Taschen, die aus einem hohen Anteil solch hoch-bewerteter Atome bestehen, sollten demnach die optimalen Taschen für die festesten Bindungen innerhalb des denkbaren Taschenraums für dieses Konformerensemble darstellen.

Der Wert, der sich bei der Zählung aller jemals von einem Atom erzeugten Taschenoberflächenpunkte innerhalb eines Konformerensembles ergibt, kann umgedreht als Potential dieses Atoms gesehen werden, erfolgreich eine Taschenoberfläche auszukleiden. Im Rahmen der Arbeit an *Pandora* wird diese Wahrscheinlichkeit für das Auftauchen in einer Tasche als Pocket Surface Generation Potential (PSGP) bezeichnet. Dies entspricht im Deutschen dem Taschenoberflächenbildungspotential (TOBP).

Je höherwertig dieses PSGP, desto interessanter wird es für das strukturbasierte Design von Liganden, für diese Wechselwirkungsmöglichkeiten mit relativ konstanten Taschenoberfläche Liganden zu entwickeln. Durch ein hohes PSGP kann somit die Wahrscheinlichkeit einer tatsächlich möglichen Interaktion auf ein Maximum gesteigert werden (allerdings sollte hier noch nachträglich differenziert werden, was für Wechselwirkungen das betreffende Atom überhaupt eingehen kann). Umgekehrt sind Atome, welche so gut wie nie in Taschen vorkommen oder in diesen nur einen kleinen Teil der Oberfläche ausmachen, eher uninteressant als Wechselwirkungspartner.

Somit wäre es interessant herauszufinden, welche Atome immer wieder am Aufbau von Taschen beteiligt sind, um gezielt zwischen diesen relativ konstant vorkommenden Interaktionspartnern und den zu entwerfenden Liganden Wechselwirkungen zu etablieren und um die Wahrscheinlichkeit einer dauerhaften Bindung zu optimieren.

Fragestellung:

Angenommen es existiert eine Datenbank mit einer möglichst guten Abdeckung des Konformerraums mit einer Populationsverteilung, wie sie auch *in vivo* Auftreten würde, so könnte man folgende Fragestellungen ableiten:

Welche Atome sind für die Taschenoberflächen innerhalb der MD-Läufe in der Datenbank verantwortlich?

Decken sich diese Atome mit den bereits bekannten Bindestellen für TAR?

Durchführung:

Da innerhalb der Datenbank diese Information möglicherweise verborgen war, wurde ein Modul geschrieben, welches für jeden MD-Lauf eine Rückprojektion der Taschenoberflächenpunkte der Taschen auf das jeweilige Atom ermöglicht. Die Summe aller Oberflächenanteilshistogramme konnte wiederum zu einem Gesamtoberflächenanteil dieser Atome verbunden werden. Hierbei ist davon auszugehen, dass die Größe eine Aussage darüber beinhaltet, ob dieses Atom innerhalb des Konformer-Ensembles häufig in einer Tasche anzutreffen ist und somit einen optimalen Interaktionspartner innerhalb der möglichen Taschen darstellt.

Extraktion der PSGPs:

Nach Auswahl des MD-Laufs wird für diese Trajektorie ein PSGP-Atom-Vektor zum Zählen der Oberflächenpunkte erstellt. Hiernach werden für jeden Konformer des MD-Laufs, das Atomkoordinatenset und die Taschenoberflächensets aus der Datenbank extrahiert. Anschließend wird für jede Taschenoberfläche für jeweils jeden Taschenoberflächenpunkt bestimmt, welches Atom innerhalb des jeweilig zugrunde liegenden Konformers, diesem Taschenoberflächenpunkt am nächsten ist; der Zähler dieses Atoms wird im PSGP-Atom-Vektor eins hochgesetzt. Nach Abarbeiten der gesamten Konformerliste des MD-Laufs wird der PSGP-Atom-Vektor ausgegeben; siehe hierzu Abbildung 100 auf Seite 159.

Pseudocode:

1. Auswahl des MD-Laufs → Erstellen des PSGP-Atom-Vektor zum Zählen der Oberflächenpunkte.
2. Extraktion der Konformer-ID-Liste für den MD-Lauf.
3. Extraktion des ersten Konformer aus der Liste.
4. Extraktion der Taschen-ID-Liste für den gerade verwendeten Konformer
5. Extraktion der ersten Taschenoberflächen für den Konformer.
6. Für jeden Taschenoberflächenpunkt wird das nächste Atom im aktuellen Konformer bestimmt und dessen Zähler im PSGP-Atom-Vektor eins hochgesetzt.
7. Falls weitere Taschen für den Konformer vorliegen → Wiederholung des Schritts 6 für jede weitere Tasche.
8. Falls weitere Konformere für diesen MD-Lauf vorliegen → Wiederholung der Schritte 4-7.
9. Ausgabe des PSGP-Atom-Vektors für diesen MD-Lauf.

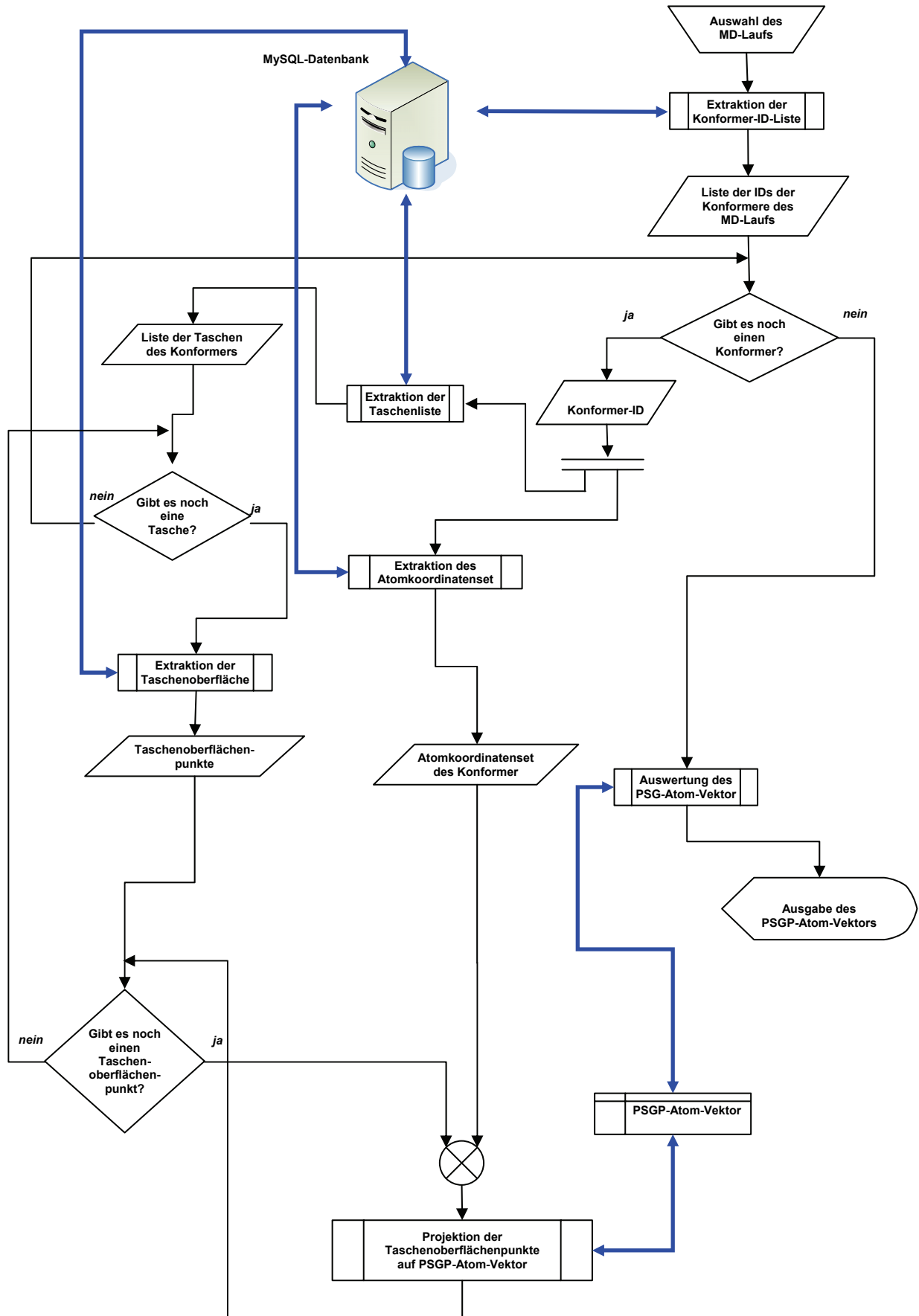


Abbildung 100 : Ablaufdiagramm der PSGP-Extraktion mit Pandora aus der MySQL-Datenbank.

Projektion der PSGP auf die Strukturen:

Die so gewonnenen PSGP-Scorelisten können wieder auf die Konformere zurück projiziert werden.

Im Rahmen der Arbeit wurde diese Rückprojektion der Taschenoberflächenpunkte auf die sie erzeugenden Atome mittels eines weiteren Moduls umgesetzt. Als Ergebnis können die einzelnen Nukleotid-Bausteine als 2D-Betrachtung (nach ihrem PSGP eingefärbt) einzeln betrachtet werden; siehe Abbildung 101. Bildschirmaufnahmen der Darstellungen aller 29 Nukleotide können im Anhang 7.10ab Seite 219 gefunden werden. Diese Darstellung soll eine vorzeitige Interpretation der PSGP basierend auf der jeweilig verwendeten Konformer-Struktur unterbinden, da diese Interpretation eine Vermengung der Statistik über viele Konformere mit einem Beispielkonformer bedeuten würde und so prinzipiell nicht zulässig ist.

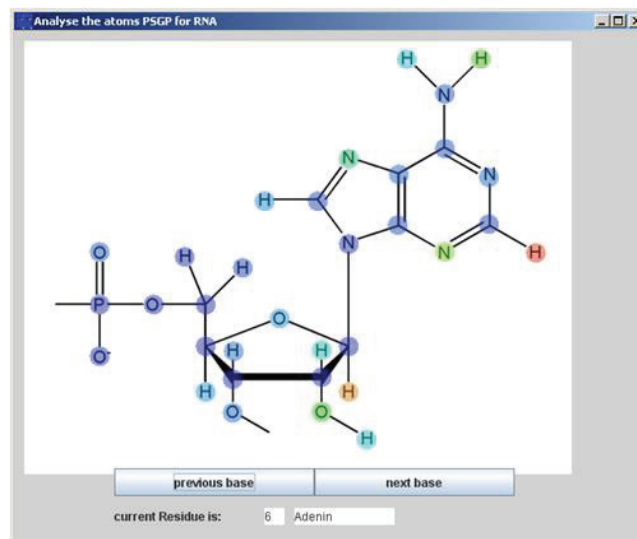


Abbildung 101: PSGP-Projektor für die 2D-Darstellungen der Nukleotide. Als Beispiel wird hier die Bildschirmaufnahme der Projektion des sechsten Nukleotids in der verwendeten TAR-Sequenz im Pandora-Modul gezeigt. Die Einfärbung der Atome, erfolgt nach dem jeweiligen PSGP; blau = niedriges PSGP und rot = hohes PSGP.

Basierend auf dem nukleotidweise gemittelten PSGP wurden Abbildungen 106 und 107 erstellt, bei denen die Nukleotide mit einem PSGP < 1,5 rot und die PSGP > 1,5 blaugefärbt wurden. Als Beispiel wurden hierbei die drei laut RMSD unähnlichsten Strukturen aus dem 1ANR-PDB-Ensemble verwendet.

Ergebnisse:

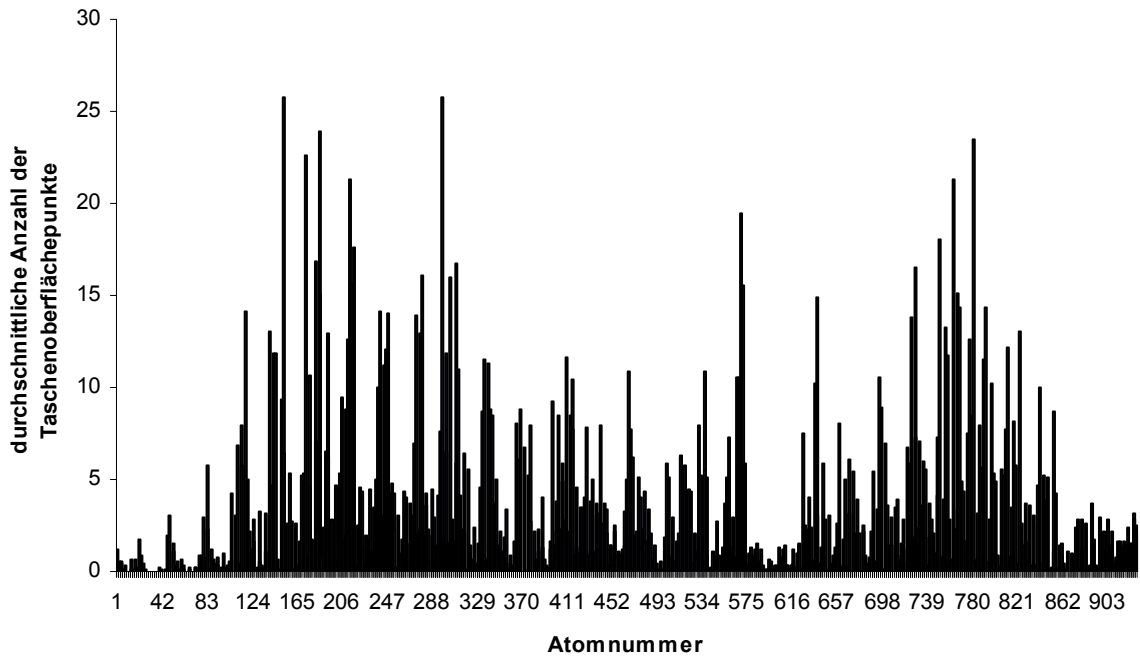


Abbildung 102: Histogramm der durchschnittlichen Taschenoberflächenpunkteanzahl für jedes Atom über die Bilder aller verfügbaren MD-Läufe in der MDDB für TAR

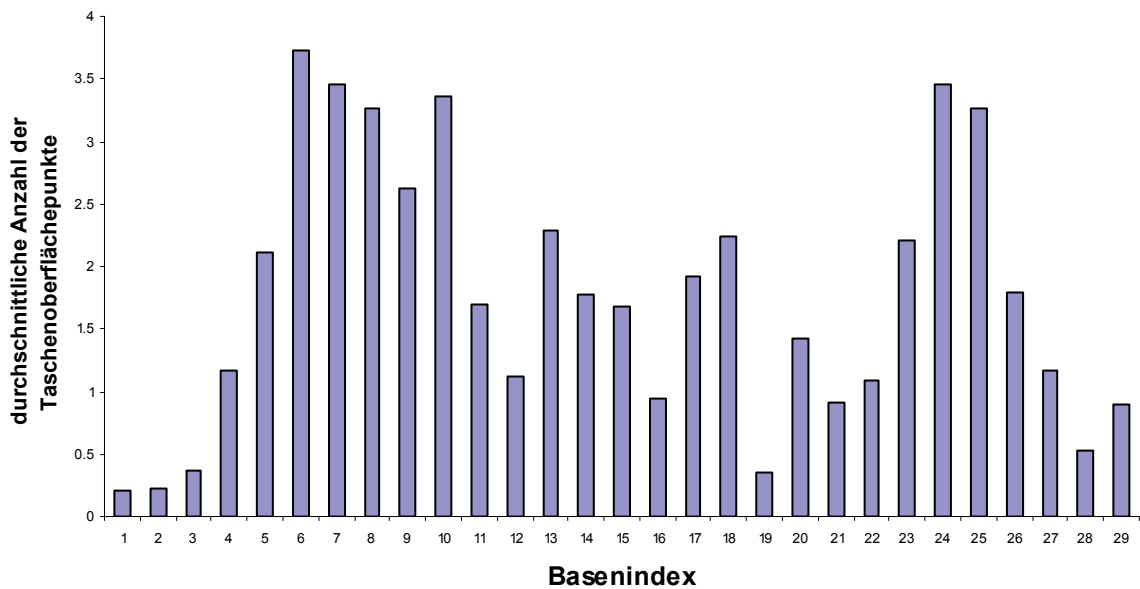


Abbildung 103: Histogramm der durchschnittlichen Taschenoberflächenpunkteanzahl für jede Base über alle Bilder aller verfügbaren MD-Läufe in der MDDB für TAR

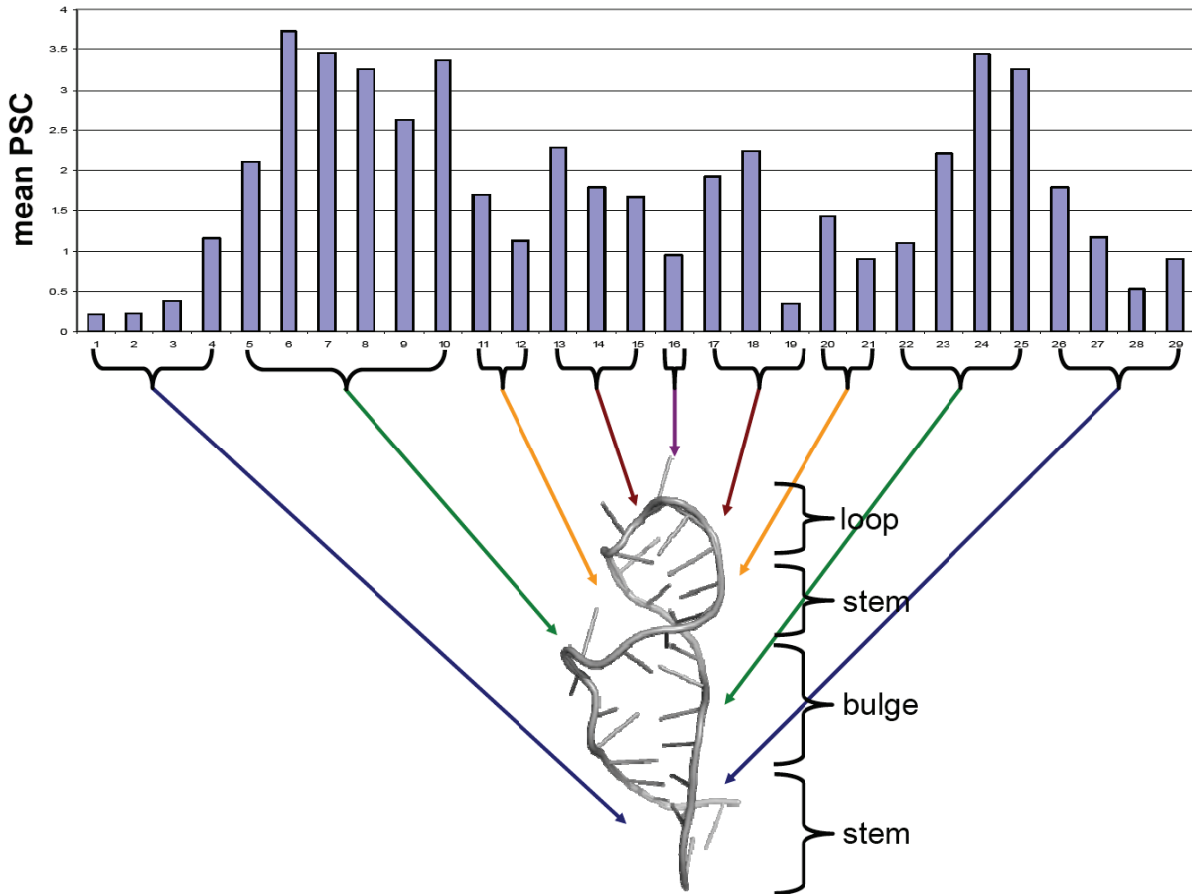


Abbildung 104: Durchschnittliche Taschenoberflächenpunkteanzahl für die einzelnen Basen und Zuordnung zu einer Beispiel TAR-Struktur, gerendert als Cartoon-Basenmodell

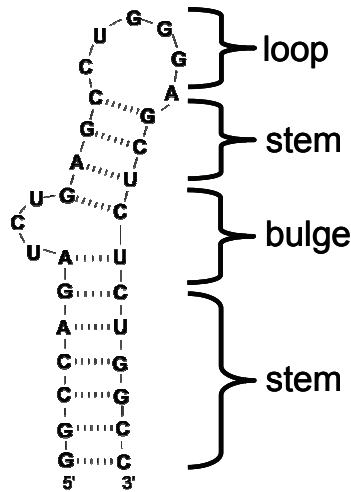


Abbildung 105: TAR-RNA, schematische Darstellung der Sekundärstruktur

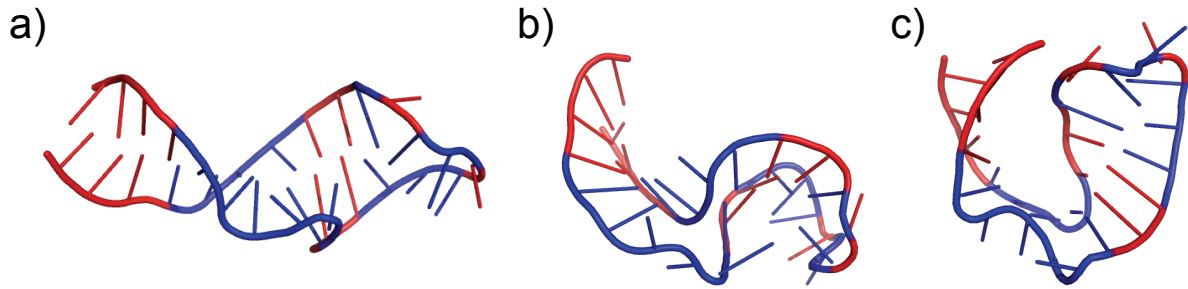


Abbildung 106: Drei TAR-Konformere des PDB-Eintrags 1ANR eingefärbt nach ihrem durchschnittlichen PSGP; Darstellung als Cartoon-Repräsentation; PSGP < 1,5 rot und PSGP > 1,5 blau

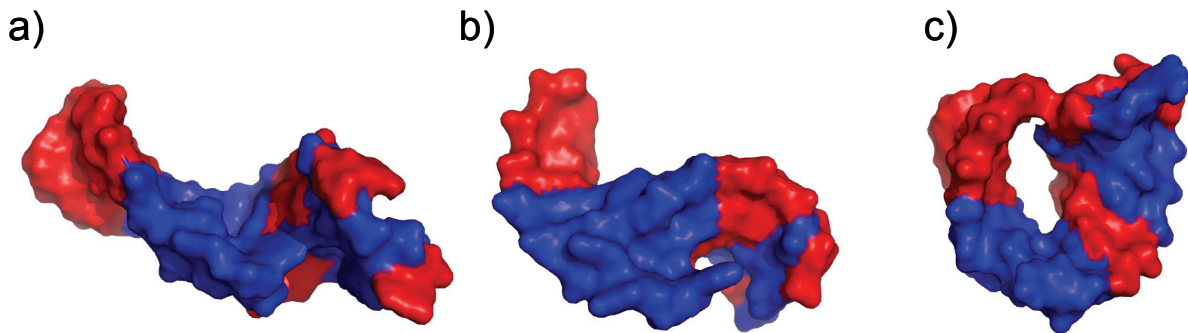


Abbildung 107: Drei TAR-Konformere des PDB-Eintrags 1ANR eingefärbt nach ihrem durchschnittlichen PSGP; Darstellung als Oberflächenrepräsentation; PSGP < 1,5 rot und PSGP > 1,5 blau

Diskussion:

Atombasiertes Bestimmen des PSGP

Abbildung 102 zeigt das Histogramm der durchschnittlichen Taschenvergrabenheit für jedes Atom einzeln aufgelöst. Die dargestellte Reihenfolge der Atome entspricht abschnittsweise der Reihenfolge in einzelnen Nukleotiden, welche die TAR darstellen. Hierbei zeigt sich, dass einzelne Atome in bestimmten Bereichen der TAR-RNA hierbei einen höheren PSGP erhalten als in anderen Bereichen. Besonders der endständige Stammbereich zeigt hier ein verringertes PSGP seiner Atome.

Nukleotid-basiertes Bestimmen des PSGP

Abbildung 103 zeigt den durchschnittlichen PSGP für die einzelnen Nukleotide der TAR-RNA. In der Abbildung zeigt sich – wie auch in der atomaufgelösten Darstellung in Abbildung 102 – die Präferenz der Taschen, für bestimmte Bereiche der TAR-RNA.

Um die Bereiche der TAR-RNA zu identifizieren, die an am meisten von Taschen bevölkert werden, wurden die Projektionen 104, 106 und 107 durchgeführt.

Hierbei zeigen Abbildungen 106 und 107 die drei – laut RMSD – unterschiedlichsten RNA-Konformere des PDB-Eintrags 1ANR, eingefärbt nach einem PSGB < 1,5 in rot und bei einem PSGB > 1,5 blau. Aus diesen Darstellungen ist ersichtlich, dass die jeweiligen Bereiche komplementär aufgebaut sind. Genauer betrachtet, sind jeweils die Stamm-Bereiche rot eingefärbt und die „Bulge“- und „Loop“-Region erscheinen blau. Innerhalb des „Loop“-Bereichs sind einige Nukleotide ebenfalls rot eingefärbt.

Nukleotid-basierte Projektion der PSGP auf Beispielkonformere

Zur besseren Analyse wurden die gemittelten PSGB der Nukleotide auf einen Beispielkonformer zurückgeführt; siehe hierzu Abbildung 104 die diese Rückprojektion des Histogramms auf einen Konformer darstellt. Für diese wurde der Konformer unter das Histogramm platziert und mittels Pfeilen angedeutet, welche Bereiche der TAR-RNA jeweils das PSGP verursacht haben. Zusätzlich enthält diese Darstellung auch die Bezeichnungen „Stamm“, „Bulge“ und „Loop“, die in den meisten Veröffentlichungen zu TAR als Struktur motive verwendet wurden (Karn 1999). Hierbei zeigt sich, dass die auch in der Literatur (Karn 1999) als Bindestellen vorgeschlagenen Bereiche der TAR ebenfalls die größten PSGPs tragen.

Schlussfolgerungen:

Das hier abgeleitete PSGP deckt sich mit den aus der Literatur (Karn 1999) bekannten Bindestellen. Die PSGP könnten somit dazu eingesetzt werden, die Taschen anhand dieses Kriteriums zu bewerten und um geeignete Docking-Referenztaschen zu identifizieren. Dieser Ansatz wurde in einer ersten Bewertung in Abschnitt 4.5.9 durchgeführt.

Ferner zeigt sich, dass einige Nukleotide in der „Loop“-Region keine so großen PSGP tragen wie die benachbarten Loop-Nukleotide. Dies könnte daran liegen, dass einige Basen, um den „Loop“ bilden zu können in das Lösungsmittel ragen müssen und somit im Durchschnitt von den Taschenbereichen innerhalb des „Loops“ wegzeigen, so dass deren Atome nicht am Taschenaufbau beteiligt sein können. Hierzu müssen aber noch weitere Untersuchungen durchgeführt werden. Erste Betrachtungen der Konformere des PDB-Eintrags 1ANR scheinen dies zu bestätigen.

Um die gewonnenen Informationen nicht wieder durch die Mittelwertbildung zu verwischen, wie dies bei der Nukleotid-Betrachtung der Fall war, sollte mit der Atomauflösung weiter gearbeitet werden. In zukünftigen Arbeiten sollte ferner bei einer Differenzierung der Statistiken auch die Größe der Oberflächenbereiche relativ zu ihrer Häufigkeit innerhalb des Konformerraums ermöglicht werden. Momentan können einzelne kleine, konstante Taschenoberflächenbereiche noch nicht von großen, unregelmäßig auftretenden Taschenoberflächenbereichen unterschieden werden. Dies ist jedoch für eine bessere Diskriminierung der Güte der möglichen Wechselwirkungen erforderlich. Allerdings zeigt bereits dieser Ansatz eine gute Lokalisierbarkeit von scheinbar wichtigen Atomen für potentiell feste/zeitlich stabile Bindungen.

4.5.9 Erste Tests des PSGP zur Taschenselektion

Hintergrund:

Basierend auf der Ableitung des PSGP, sollten erste Tests zur Bewertung der Taschen nach diesem neuen Kriterium durchgeführt werden. Hierzu wurden zwei erste Bewertungsfunktionen abgeleitet: der PSGP-Taschenscore und das PSGP-Taschengewicht. Die hiernach gezeigten Ansätze stellen den ersten Schritt in einer noch weiter zu entwickelnden Kette von Optimierungsschritten dar und sollten nur zeigen, dass anhand des PSGP bereits Taschen ausgewählt werden können.

Der PSGP-Taschenscore

Der PSGP-Taschenscore beruht auf der Summe aller PSGP-Werte für jeden Taschenoberflächenpunkt der jeweiligen zu bewertenden Taschen.

$$pocketscore_{PSGP} = \sum_i PSGP_i \quad (4.1)$$

Diese Funktion diskriminiert nicht nach der relativen Lage der „interessanten“ Taschenoberflächenpunkte zueinander, sondern bewirkt einfach eine relative Höhergewichtung von Taschen gleicher Taschenpunktzahl anhand der globalen PSGP-

Summe über all diese Taschenoberflächenpunkte. Prinzipiell werden somit auch Taschen mit größerer Taschenoberfläche, denen mit deutlich kleinerer Taschenoberfläche vorgezogen. Ein Relativierung dieses Effekts hätte durch die Einführung eines weiteren Faktors reduziert werden können: Dem Teilen des Pocketscores durch die Anzahl der verwendeten Taschenoberflächenpunkte. Jedoch könnte dies wiederum zu einer Bevorzugung extrem kleiner Taschen mit hohen PSGP-Werten der beteiligten Oberflächenpunkte führen. Mit der in Gleichung 4.1 gezeigten Formel wurde eine Bewertung der verfügbaren Taschenpopulation durchgeführt und die Taschen in Abbildung 108 dementsprechend als „Top 10“ vorhergesagt.

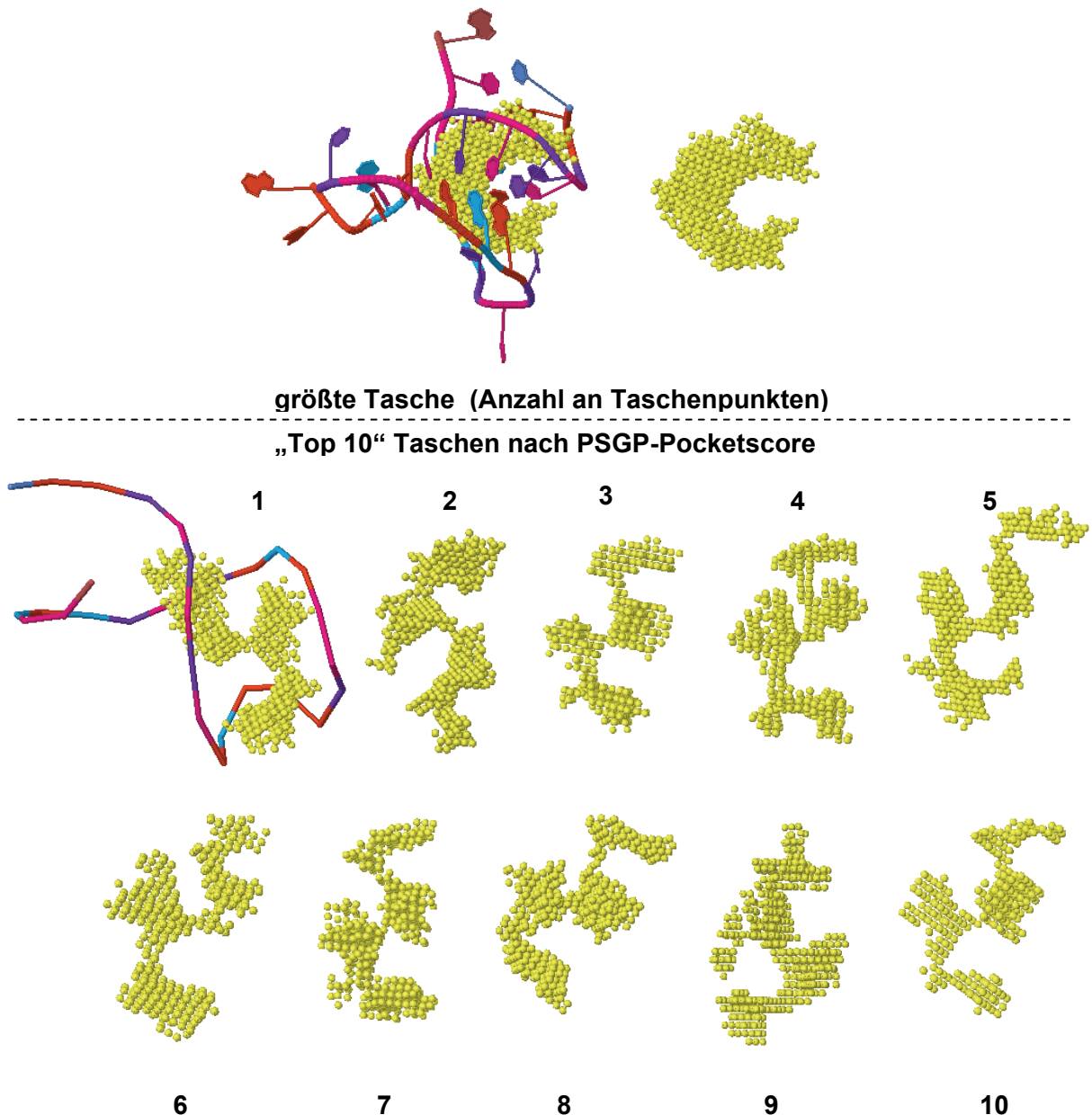


Abbildung 108: "Top 10" Taschen nach PSGP-PocketScore; oben ist zusätzlich die größte Tasche im Datensatz zu sehen. Die nach PSGP-Taschenscore ausgewählten Taschen sind zwar relativ groß, aber augenscheinlich differenzierter aufgebaut als die Referenztasche (größte Tasche im Datensatz). Aus Übersichtsgründen wurde nur bei der ersten Tasche der Verlauf des RNA-Rückgrats gezeigt.

Das PSGP-Taschengewicht

Das Taschengewicht wurde eingeführt, um diejenigen Taschen auswählen zu können, deren Oberflächenpunkten, bzw. deren Atome mit hohem PSGP möglichst dicht beieinander liegen. Dies erfolgte, da die Taschen, welche bisher vom PSGP-Taschenscore ausgewählt wurden, bevorzugt diejenigen Taschen waren, die sowohl die bevorzugten Bereiche der Bulge als auch Loop der TAR-RNA beinhalten. Um diesen Effekt zu relativieren, wurde eine Funktion in *Pandora* geschrieben, die für jede Tasche, prüft, welche Atome für die Taschenoberfläche verantwortlich sind, um dann abschließend für jede Paarung der identifizierten relevanten Atome für diese Tasche, den jeweiligen Paarungswert relativ zur Entfernung der beiden Atome zu bestimmen. Durch die Verwendung der Atomdistanzen als Nenner wird somit eine Skalierung der Gewichtung (Taschengewichte) anhand der Dichteverteilung der Atome mit hohen PSGP-Werten ermöglicht.

$$pocketweight_{PSGP} = \sum \frac{PSGP_{Atom_i} * PSGP_{Atom_j}}{d(Atom_i, Atom_j)}, \text{ wobei} \quad (4.2)$$

$PSGP_{Atom_i}$ dem PSGP-Wert des $Atom_i$ entspricht

$PSGP_{Atom_j}$ dem PSGP-Wert des $Atom_j$ entspricht

$d(Atom_i, Atom_j)$ der Distanz zwischen $Atom_i$ und $Atom_j$ entspricht

Verteilung der PSGP-Taschenscore und PSGP-Taschengewicht

Als erste Anwendung wurde der Konformerraum der Datenbank mit diesen Taschenscores bewertet. Gezeigt werden hier beispielhaft die Auftragungen der PSGP-Taschenscores und PSGP-Taschengewichte für den mit dem „Taschendurchmesserfilter“ eingeschränkten Taschenraum. Die Abbildung 109 und Abbildung 110 zeigen den gleichen Ausschnitt des Taschenraums, jedoch jeweils eingefärbt nach dem maximalen PSGP-Taschenscore und PSGP-Taschengewicht. Die Erstellung der Abbildungen erfolgte mittels hierfür geschriebener m-Skripte. Die unterschiedliche Farbverteilung innerhalb der beiden Abbildungen zeigt die Möglichkeiten der Selektion des Taschenraums anhand dieser beiden Kriterien. In weiterführenden Studien gilt es diese Kriterien nach ihrer Anwendbarkeit weiter zu untersuchen. In Abbildung 108 wurden die als „Top 10“ bewerteten Taschen nach dem PSGP-Taschenscore dargestellt, diese zeigen, dass dieses Kriterium bereits eine unterschiedliche Differenzierung als die alleinige Auswahl nach der Taschengröße ermöglicht. Wie bereits aus den anderen Untersuchungen zur automatischen Einschränkung des zu verwendenden Taschenraums nahe gelegt wurde, ist letztendlich eine Kombination aus den verschiedenen Kriterien anzustreben.

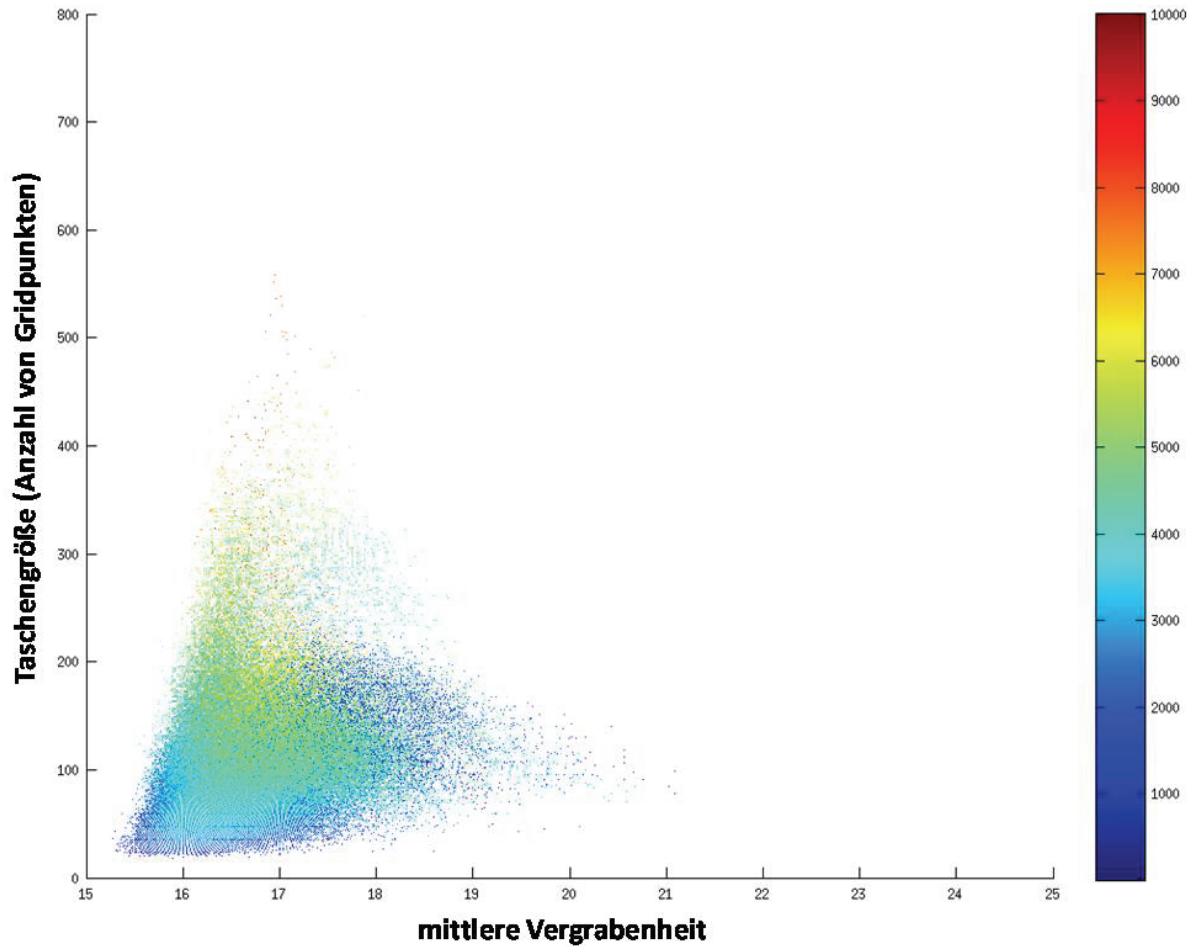


Abbildung 109: Verteilung der maximalen PSGP-Taschenscore für den Taschenraum nach Anwendung des Taschendurchmesserfilters. Auftragung erfolgte nach Taschengröße und "mittlerer Vergrabenheit", Farbcodierung erfolgte nach dem maximalen gefundenen PSGP-Taschenscore für den jeweiligen Datenpunkt (Taschengröße und "mittlerer Vergrabenheit")

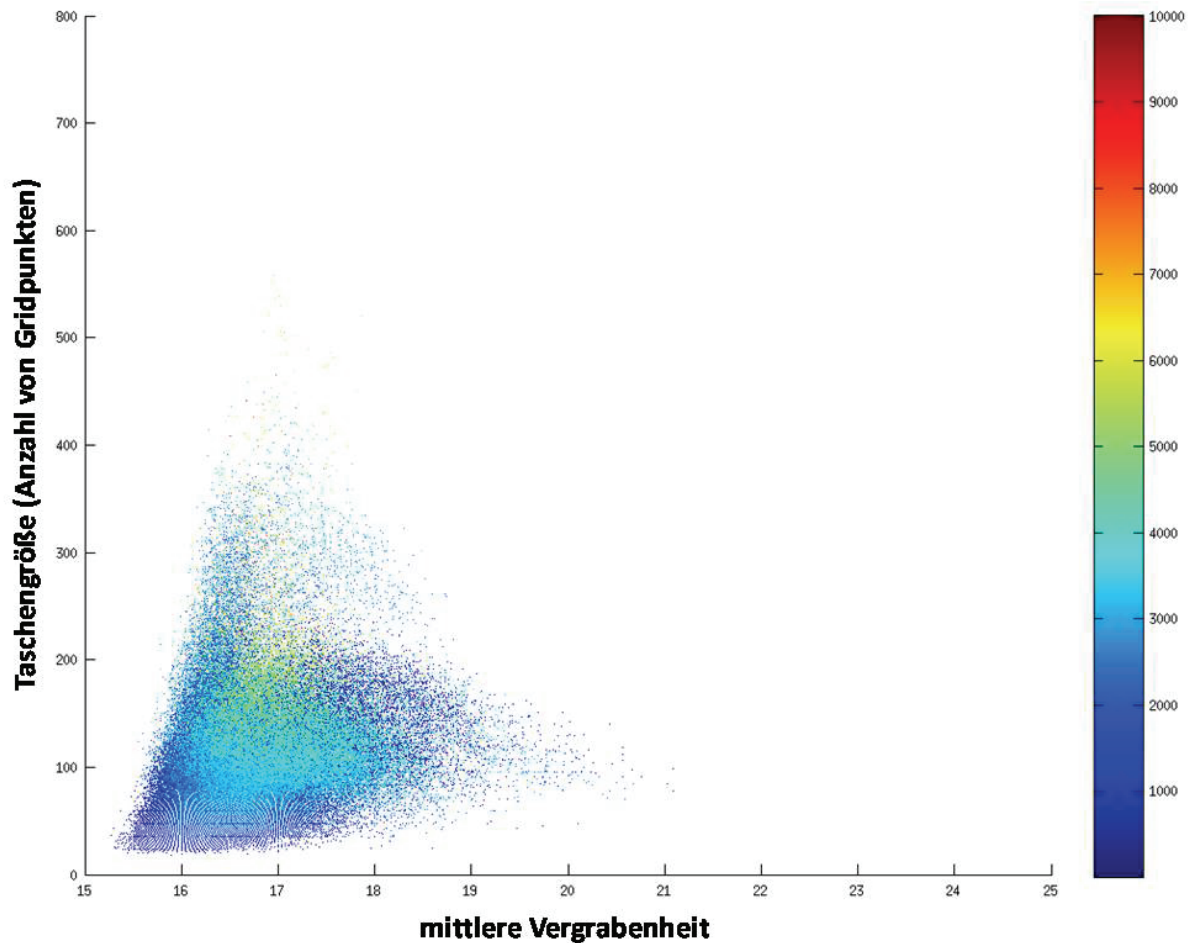


Abbildung 110:Verteilung der maximalen PSGP-Taschengewichte für den Taschenraum nach Anwendung des Taschendurchmesserfilters. Auftragung erfolgte nach Taschengröße und "mittlerer Vergrabenheit", Farbcodierung erfolgte nach dem maximalen gefundenen PSGP-Taschenscore für den jeweiligen Datenpunkt (Taschengröße und "mittlerer Vergrabenheit")

Unterschiedliche Verteilung der PSGP-Taschenscores für einzelne MD-Läufe

Die Verteilungen der PSGP-Taschenscore-Werte für die einzelnen Taschenräume der MD-Läufe (Trajektorien) unterscheiden sich von Lauf zu Lauf, so dass auch hier wieder eine Differenzierung des PSGP-Taschenscores in Abhängigkeit zu den zugrunde liegenden Konformeren ersichtlich wird; siehe hierzu Abbildung 111. Dementsprechend sollten zukünftige Studien die repräsentative Abbildungsgüte der einzelnen Konformerraumteile (Trajektorien) in die Bewertung weiter mit einbeziehen, solange noch keine umfassende repräsentative Konformerraumabbildung definiert werden kann.

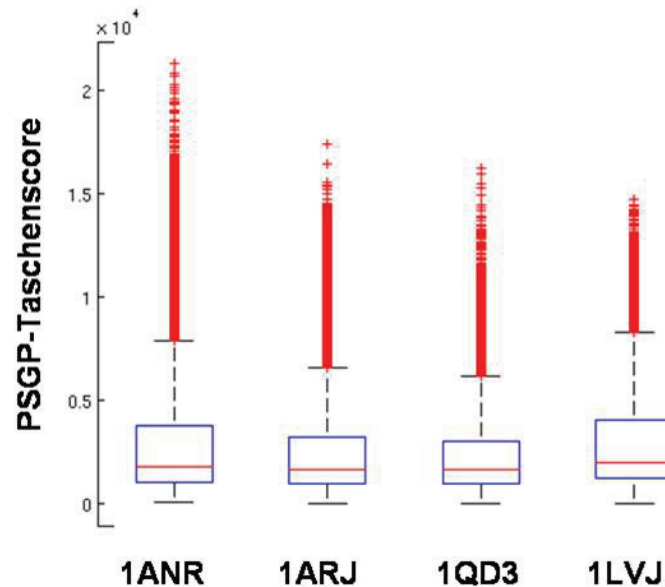


Abbildung 111: Boxplots der PSGP-Taschenscores für die vier MD-Trajektorien

4.6 Cluster-Methoden und ihre Schwächen

Die nachfolgenden Überlegungen beziehen sich auf die Auswertung des Konformerraums der genannten Datenbank, die zurzeit ca. 153.000 MD-generierte Konformere umfasst. Die Ausführungen beschäftigen sich hierbei mit den Problemen und Lösungen, die sich bei der notwendigen Aufgabe ergaben, die Datenbank mit Ähnlichkeitsmaßen für die Suche zu indizieren. Insbesondere gilt dies für die Indizierung der gewonnenen Taschenableitungen für besagte Konformere, da diese nicht eine vorgeschriebene (definierte) Anzahl an paarweise zu interpretierenden Punkten enthalten, wie dies bei den Konformeren der Fall ist. Auch die resultierende Zahl an potentiellen Taschen übersteigt um ein Vielfaches die der Konformeren-Population (ca. 5x bei TAR).

4.6.1 Ansatzpunkte und Probleme beim RMSD-basierten Clustern

Der RMSD-Wert (Mittlere Standardabweichung) als Maß der Ähnlichkeit beruht auf einer paarweisen Berechnung der Abstände von bestimmten Merkmalen einer Struktur. Daher erfordert die Berechnung des RMSD-Wertes für ein Strukturpaar folgende Bedingungen:

- Es müssen genau definierte Paare für die Struktur vorliegen; eine automatische Berechnung für unterschiedliche Strukturen ist nicht möglich.
- Die Strukturen müssen, falls der RMSD ihre strukturelle Ähnlichkeit ausdrücken soll, vor der Berechnung bestmöglich überlagert (aligniert) werden.

Dies ist nur für Strukturen gleichen Aufbaus anwendbar, so dass Taschenvergleiche (unterschiedliche Anzahl an Punkten mit nicht definierter Paarung) hiermit nicht durchgeführt werden können. Dieser Ansatz ist somit einzig und allein für die Konformere umsetzbar.

4.6.1.1 Gedankenexperiment: Konformerraum-RMSD-Matrix

Dies bedeutet, dass bei der Auswertung der Ähnlichkeit zweier RNA-Konformere diese vorher mittels geeigneter Algorithmen im Raum so gedreht werden müssen, dass sie eine optimale Überlagerung zeigen. Letzteres ist meist mit der Berechnung des besagten RMSD

verbunden. Steht ein Vergleich einer Vielzahl von Konformeren an, bedarf es einer $(n^2/2)$ -n Anzahl an Berechnungen dieser Art, wobei n für die Anzahl der Strukturen steht. Bei kleinen Systemen ist das ohne weiteres möglich, bei großen Systemen vervielfacht sich jedoch die benötigte Rechenzeit nahezu quadratisch.

Ferner ist der RMSD-Wert eine Beschreibung genau eines Paares, und ein Wertevergleich ist somit immer nur relativ zu einer Referenzstruktur sinnvoll. Zum Suchen ähnlicher Strukturen in der Datenbank ist daher erneut ein Abgleich aller Strukturen gegen die Suchanfrage nötig. Hieraus ergeben sich folgende Überlegungen: Die derartig berechnete Distanzmatrix könnte zwar als Endergebnis festgehalten werden, dennoch wäre ihr Nutzen für zukünftige Abfragen gering.

Da die gespeicherten Abstandsinformationen bereits Näherungen für weitere relativ ähnliche Strukturen enthalten, könnte man dieses Wissen direkt nutzen, wenn man *a priori* wüsste, welche der in der Matrix beschriebenen Strukturen der neuen Struktur am ähnlichsten sieht. Allerdings wäre – um die relevanteste Position innerhalb der bestehenden Matrix zu bestimmen – besagter Abgleich mit allen bisherigen Strukturen durchzuführen, so dass durch die RMSD-Berechnung und der Ablage der Daten in einer Distanzmatrix kein Geschwindigkeitsgewinn für die Suche erzielt werden könnte.

Die bei dieser Suche somit erweiterten Relationsbeziehungen könnten zwar wieder genutzt werden, um die Matrix zu erweitern, gleichwohl ist hier der zu erwartende quadratisch ansteigende Speicherbedarf ebenfalls als praktisches Problem zu werten. Dies resultiert letztendlich in einer immer langsameren Suche und Eingliederung neuer Daten.

Des Weiteren gilt, dass die Art wie der Mittelwert gebildet wird, dazu führt, dass Strukturen mit einem gleichen Zahlwert beim RMSD zu einer Referenzstruktur nicht zwangsläufig zueinander ähnlich sein müssen, weil die Information der Richtung einer Abweichung bei diesem Verfahren vernichtet wird.

Schlussfolgerung:

Somit kann eine Distanzmatrix als solche nicht direkt zur Ähnlichkeitssuche für die Integration neuer Daten eingesetzt werden.

Nutzung zur Suche ähnlicher Strukturen im Bestand:

Ferner kann mittels der erstellten Distanzmatrix nach den ähnlichsten oder unähnlichsten Strukturen zu einer ausgewählten Struktur innerhalb der Datenbank gesucht werden. Dies wäre auch relativ „berechnungsarm“ umsetzbar, da hier nur einzelne Spalten der Distanzmatrix nach der Größe der RMSD-Werte ausgewertet (sortiert) werden müssten, um anhand der Indizes einen direkten Verweis auf die abgelegten Strukturen zu erhalten.

Clustern von Einträgen:

Wie bereits oben geschildert, eignen sich die Einzelpositionen der Matrix (respektive RMSD-Werte) nicht dazu, die Strukturen direkt in Gruppen ähnlicher Strukturen zusammenzufassen.

Jedoch könnte die so erstellte Distanzmatrix die Möglichkeit bieten, die Datenbank nach diesem Kriterium zu Clustern; siehe Abschnitt 4.6.1.2.

4.6.1.2 Distanzmatrix als Deskriptor für die Clusterung

Die resultierende Distanzmatrix aus den Berechnungen des Abschnitts 4.6.1.1 könnte dazu benutzt werden, diese spaltenweise als Deskriptor (Korrelationsvektor) zu benutzen. Hieraus könnten wiederum Distanzberechnungen erstellt werden, weil in einer solchen Spalte die einzelnen Positionen innerhalb der Spalte als Beschreibung für die Ähnlichkeit zu anderen Spalten (Strukturen) verwendet werden können. Die daraus resultierende Klassifizierung ist statisch und müsste für jede Erweiterung der Matrix durch neue Einträge wiederholt werden,

da die Spaltenlänge mit jedem neuen Eintrag zunimmt, wobei das Auflösungsvermögen jedoch stetig zunimmt.

Eine Alternative zur ständigen Erweiterung der Distanz-Matrix könnte bei Erreichen einer gewissen Grundgröße die Nutzung eines zusätzlichen Cluster-Verfahrens sein, welches nach Extraktion eines repräsentativen diversen Ensembles (Abdeckung des Konformerraums) anhand der Ähnlichkeit (RMDS-Korrelationsvektor) zu diesen Ensembles eine Clusterung weiterer Einträge vornehmen kann, um so Populationsschwerpunkte innerhalb des Konformerraums aufzuspüren. Die Umsetzung könnte mittels SOM als beschriebenes Cluster-Verfahren erfolgen; siehe Abschnitt 2.5.

Motivation für zukünftige Arbeiten:

Der Ansatz mittels SOM ein repräsentatives Ensemble zu extrahieren, wurde bereits im Rahmen dieser Arbeit für die Taschenkorrelationsvektoren umgesetzt und kann im vorliegenden Software-Paket leicht auf die oben beschriebene RMSD-Korrelationsvektor-Variante zur Konformer-Clusterung erweitert werden (Kabsch-Alignment (Kabsch 1976) und RMSD-Berechnungen sind als Module bereits integriert; ferner sind Distanzmatrixverwaltung und Rechenoperationen im unten beschriebenen SIMPrints-Ansatz implementiert worden; siehe Abschnitt 4.6.3).

Momentan ist dies nur für die Konformer-Clusterung basierend auf verschiedenen Taschendescriptoren umgesetzt. Ferner kann der bereits in die Software eingebaute Element-basierte Korrelationsvektor für die Erstellung dieser SOMs auf Basis der Atomkoordinaten analog zum RMSD verwendet werden. Weiterhin wäre hier ein Vergleich von RMSD- und Element-basiertem Korrelationsvektor interessant, da letzterer noch eine Auflösung der Verteilung der Atome im Raum beinhaltet und diese Information nicht auf einen relativen Zahlwert reduziert wurde, wie im Fall des RMSD-Deskriptors.

Eine starke Korrelation zwischen den, bestimmte Taschentypen enthaltenden, Konformeren und den nach RMSD-Korrelationsvektor ähnlichen Konformeren ist zwar anzunehmen, dennoch würde dieser Vergleich der Aufteilung der Konformere auf die Cluster der verschiedenen SOMs interessante Einblicke auf das Vorkommen gleicher Taschentypen in unterschiedlichen Konformeren ermöglichen. Dieser Abgleich wäre zurzeit noch manuell vorzunehmen. Dies würde die Frage leichter klären lassen, ob gleichartige Taschen in unterschiedlich geformten Konformeren – eventuell an unterschiedlichen Positionen – der TAR vorkommen.

Hiermit sei auf die diesbezüglich relevante Fragestellung in Abschnitt 4.8 auf Seite 185 („Abschließende Untersuchungen der Interaktion der bekannten Peptid-Liganden mit TAR“) verwiesen.

4.6.2 Auswertung der RMSD-Berechnungen für Beispielkonformere der TAR

Hintergrund:

Zum strukturbasierten Liganden Design bedarf es einer definierten Zielstruktur, die in ihrer räumlichen Struktur als Referenz herangezogen werden kann. Die TAR liegt in der PDB jedoch in mehreren Strukturaufklärungslösungsvorschlägen vor. Diese Vorschläge beinhalten ferner nicht nur einen einzigen Konformer, sondern stellen jeweils ein Konformer-Ensemble dar, welches mit den experimentellen bestimmten Entfernungsparametern (NMR-Studien) am besten in Einklang gebracht werden konnte. Die Grundlagen für diese Ensembles unterscheiden sich sowohl in der Anwesenheit strukturell co-aufgeklärter Liganden, als auch in der unterschiedlichen Größe der TAR-Aptamere sowie in ihrer endständigen Basenzusammensetzung. Des Weiteren wurde eine unterschiedliche Anzahl an Konformeren in den einzelnen PDB-Einträgen hinterlegt.

Fragestellungen:

1. Als Teil dieser Arbeit sollte untersucht werden, inwieweit einzelne PDB-Einträge besonders zur Charakterisierung der Flexibilität der TAR geeignet sind.
2. Eine weitere Fragestellung bezog sich auf den relativen Einfluss von Liganden auf die beobachtete Flexibilität innerhalb der verschiedenen PDB-Ensembles.

Durchführung:

- Die TAR-Konformere von 1ANR, 1ARJ, 1QD3 und 1LVJ wurden jeweils als Ensemble in Pandora geladen; eventuell vorhandene Liganden wurden entfernt.
- Mittels des RMSD-Moduls „Kabsch Alignment for RMSD“, wurde ein Alignment – jedes Konformers gegen jeden anderen Konformer – durchgeführt.
- Mit dem Modul wurde ferner die RMSD-Distanzmatrix aller Konformer-Kombinationen erstellt.
- Für jede Distanzmatrix wurden – mittels Excel – Statistiken erstellt, die den durchschnittlichen RMSD, seine Standardabweichung und das jeweilige Minimum und Maximum (jeder unterschiedlichen Konformer-Kombination) festhalten.

Ergebnisse:

Basierend auf den RMSD-Distanzmatrizen wurde untersucht, inwieweit die einzelnen Ensembles strukturelle Diversität innerhalb ihrer jeweiligen Zusammensetzung zeigen. Die Ergebnisse für die Untersuchung wurden in den Tabellen 28 und Tabelle 2929 zusammengefasst. Die Referenzstrukturen der Tabelle 28 unterscheiden sich von denen insofern, als dass die RMSD-Berechnungsgrundlage der zweiten durch Reduktion der benutzten Atomkoordinaten auf die Phosphoratom beschränkt wurde. Dies erfolgte, um ein Maß für die strukturelle Ähnlichkeit der einzelnen Konformere anhand der Rückradstruktur der TAR zur ermöglichen.

Die genannten Tabellen enthalten die Statistiken der RMSD-Distanzmatrizen, die für Durchschnitts-RMSD, Standardabweichung und auftretenden minimal oder maximal RMSD berechnet wurden.

Tabelle 28: Statistik der RMSD-Werte-Verteilung der RMSD-Distanzmatrizen, basierend auf den alle Atome einschließenden Konformer-Alignments für die PDB-Einträge 1ANR, 1ARJ, 1QD3, 1LVJ; die Zahl (n) gibt die Anzahl der Konformere in der PDB wieder.

alle Atome	Mittelwert	Standardabweichung	Minimum	Maximum
1ANR (20)	7,00	2,12	2,90	12,28
1ARJ (20)	4,80	1,11	2,18	8,00
1QD3 (17)	4,16	1,22	1,87	7,17
1LVJ (12)	3,35	0,81	1,09	5,12

Tabelle 29: Statistik der RMSD-Werte-Verteilung der RMSD-Distanzmatrizen, basierend auf den Phosphor-Atomen einschließenden Konformer-Alignments für die PDB-Einträge 1ANR, 1ARJ, 1QD3, 1LVJ; die Gewichtung der Alignments ist somit auf das TAR-Rückgrat gelegt; die Zahl (n) gibt die Anzahl der Konformere in der PDB wieder.

nur P Atome	Mittelwert	Standardabweichung	Minimum	Maximum
1ANR (20)	7,07	2,61	2,65	14,29
1ARJ (20)	4,58	1,19	1,93	7,73
1QD3 (17)	4,13	1,35	1,60	7,47
1LVJ (12)	2,82	0,71	1,35	4,65

Diskussion:

Vorbemerkung:

Ein qualitativer Vergleich der beiden Tabellen 28 und Tabelle 29 ergab, dass diese sich nur in ihren jeweiligen relativen Werten zueinander unterscheiden, jedoch die Reihenfolgen der einzelnen, sortierten Kategorien nicht variiert. Daher wird hier die Betrachtung der Einzelversuchsreihen stellvertretend nur an einer der Tabellen durchgeführt, der Tabelle 29.

Welches Ensemble ist am flexibelsten/strukturell diversesten?

Anhand der ermittelten RMSD-Werte-Verteilungen in Tabelle 29 geht 1ANR als das Ensemble hervor, welches in allen vier Kategorien die größten Werte aufweist und somit relativ zu den anderen drei Ensembles als flexibelstes betrachtet werden kann.

Insbesondere zeigt der maximal gefundene RMSD-Wert von 14,29 Ångström innerhalb dieses Ensembles einen fast doppelt so hohen Wert, wie dies beispielsweise 1ARJ und 1QD3 zeigen; gegenüber 1LVJ handelt sich in dieser Kategorie um den Faktor drei.

Diese qualitative Rangliste zieht sich durch alle Kategorien, sei es nun der durchschnittliche RMSD-Wert, dessen Standardabweichung, und selbst innerhalb des geringsten gefunden Abstands innerhalb der Ensembles (Minimum der RMSD-Werte).

Hat das Vorhandensein von Liganden einen Einfluss auf die Diversität der gefundenen Konformere?

1ANR ist das einzige PDB-Konformer-Ensemble, das ohne Ligand strukturell bestimmt wurde. 1ARJ (Argininamid), 1QD3 (Neomycin) und 1LVJ (Acetylpromazin) beinhalten jeweils einen Liganden bei ihren Strukturvorschlägen.

1ANR ist von den vier untersuchten Konformer Ensembles anhand der RMSD-Distanzmatrizen-Statistiken als diversestes Ensemble identifiziert worden. Wie bereits im obigen Abschnitt „*Welches Ensemble ist am flexibelsten/strukturell diversesten?*“ beschrieben, beträgt das Maß der Einschränkung der RMSD-Werte den Faktor 2-3. Besonders stark ist hier das Ensemble 1LVJ eingeschränkt, welches selbst zum ebenfalls mit einem Liganden besetzten 1QD3 eine weitere Reduktion des RMSD-Mittelwertes um weitere ca. 32% aufweist.

Schlussfolgerung:

Zur Bestimmung des besten Referenzensembles wurde die RMSD-Distanzmatrizen-Analyse erfolgreich eingesetzt, um aus einem kleinen Datensatz von unterschiedlichen Konformerensembles, das zu identifizieren, welches die größte strukturelle Vielfalt aufweist.

Ferner konnte bestätigt werden, dass die Ensembles mit Liganden eine eingeschränkte Flexibilität gegenüber dem so definierten Referenzensemble (1ANR) aufweisen. Dies konnte schon aus der Anzahl der abgelegten Strukturen in den PDB-Einträgen vermutet werden. Jedoch konnte hierdurch weiter zwischen den beiden gleich groß bevölkerten Ensembles 1ANR und 1ARJ unterschieden werden.

1ANR mit ihren 20 Konformeren – für eine TAR ohne Ligand – wurde im Rahmen dieser Arbeit als Referenz festgelegt. Dies erfolgte basierend auf den Auswertungen der Statistiken für die einzelnen RMSD-Distanzmatrizen von 1ANR, 1ARJ, 1QD3 und 1LVJ, die auf Basis der Konformere in den gleichnamigen PDB-Einträgen berechnet wurden. Die Festlegung auf das 1ANR-Ensemble wurde basierend auf dessen relativ größter struktureller Diversität – im Vergleich zu der der anderen PDB-Ensembles – vorgenommen; siehe hierzu Tabellen 28 und Tabelle 29. 1ANR liefert demzufolge den bestmöglichen Referenzdatensatz an Konformeren, der aus den vier PDB-Einträgen ausgewählt werden kann. Wird das strukturell eingeschränkste Ensemble gesucht, bietet sich dementsprechend 1LVJ an.

Interpretationsmöglichkeit:

Wieso ist 1LVJ so starr?

Acetylpromazin liegt als einziger der Liganden in einer Tasche, die quer durch den TAR-Konformer greift. Dagegen liegen Neomycin und Argininamid in einer Furche der RNA, auf der Oberfläche des Konformers. Laut dem Vortrag von Al-Hashimi anlässlich des Internationalen Symposiums „RNA-Ligand-Interaction“ vom 27.-29. September 2007 des SFB579 in Frankfurt am Main, kann bei der „Bulge“ von einer Art „Scharnier“ ausgegangen werden. Dies würde erklären, wieso ein Ligand, der quer in diesem Scharnier steckt die Bewegungsfreiheit der TAR massiv einschränkt. Eine Betrachtung der postulierten Interaktionsmöglichkeiten des Acetylpromazins mittels Stacking-Interaktion seiner Ringe durch eine Verlängerung der beiden Stamm-Bereiche zu einem einzigen Superstamm, würde ebenfalls ein Knicken dieses Superstamms energetisch ungünstig erscheinen lassen (Du 2002). Da Acetylpromazin aber auch keine besonders hydrophile Verbindung (Du 2002) ist, könnte dieser Bereich aber auch der einzige sein, an dem Acetylpromazin strukturell aufgeklärt werden kann. Zur Klärung dieser Frage müssten die Signalstärkeverhältnisse zwischen dieser und den anderen drei Strukturen betrachtet werden.

Möglicher Zusatznutzen der berechneten Statistiken:

Im Rahmen der Untersuchungen zu den Gesamtstatistiken der RMSD-Distanzmatrizen, wurden auch Einzelstatistiken für jeden einzelnen Konformer innerhalb der Ensembles angelegt. Mittels dieser kann die Struktur innerhalb des Ensembles ermittelt werden, welche allen anderen Konformeren im Durchschnitt am ähnlichsten sieht. Und somit, falls man postulieren kann, dass die Einzelstrukturen der Ensembles Momentaufnahmen einer Pendelbewegung des Konformers im Konformerraum entsprechen, dass diese Struktur dem wahren Mittelpunkt der Pendelbewegung am ähnlichsten sind. Diese Struktur sollte auf dem gemittelten kürzesten Pfad zwischen den einzelnen Strukturen liegen. Somit wäre diese Struktur diejenige, welche in einer Population am häufigsten vorkommen sollte. Dies würde sie als Hauptziel der Ligandeninteraktion interessant machen, um weitere Liganden nach dieser Hauptstruktur zu entwickeln. Beispielhaft sind die so bestimmten „Hauptstrukturen“ für die PDB-Einträge 1QD3, 1ARJ und 1LVJ in Abbildung 112 dargestellt. Es handelt sich um folgende Modelle

- 1QD3 (Modell 3 mit einem durchschnittlichen RMSD von 3,37)
- 1ARJ (Modell 5 mit einem durchschnittlichen RMSD von 4,21)
- 1LVJ (Modell 4 mit einem durchschnittlichen RMSD von 3,07)

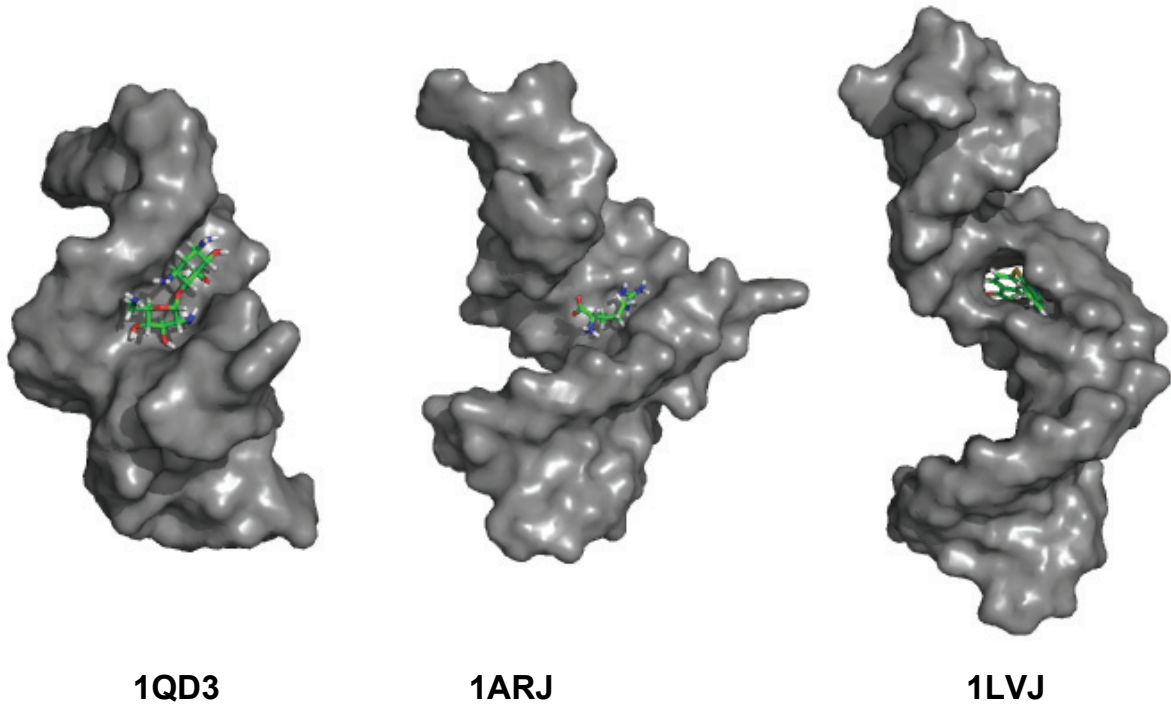


Abbildung 112: Durchschnittlichste Strukturen der TAR bezogen auf die RMSD-Distanzmatrixberechnungen der Einzelensembles 1QD3, 1ARJ, 1LVJ; die gezeigten Strukturen haben den kleinsten durchschnittlichen RMSD zu allen anderen Konformeren ihres Ensembles.

Ausblick:

Sollte eine weiterführende Untersuchung bezüglich der durch die Liganden erfolgten strukturellen Diversitätsverminderung der Datensätze angestrebt werden, würde sich das Rigiditätsbeurteilungs-Werkzeug von Simone Fulle eignen, um die durch den jeweiligen Liganden verursachte Versteifung zu untersuchen (Fulle und Gohlke 2007).

4.6.3 SIMPrints eine 3D-Form beschreibende Darstellung mit Substrukturauflösung

Die Skalierungs-Index-Methode (SIM) ist eine vom Max-Planck-Institut für extraterrestrische Physik entwickelte Methode zur Extrahierung der strukturellen Information aus 3D-Daten, welche aus der Theorie komplexer Systeme abgeleitet wurde.

Diese Methode wurde ursprünglich entwickelt, um eine automatische Charakterisierung von Galaxien anhand ihrer charakteristischen räumlichen Ausdehnung vornehmen zu können und wurde bereits auf weitere nicht-astronomische Fragestellungen angewendet (Bunk *et al.* 2001). Da die Charakterisierung räumlicher Verteilungsmuster analog zu der hier gestellten Fragestellung der Konformeranalyse bei Makromolekülen verläuft, erschien ein Test der Übertragbarkeit dieser Klassifizierungsmethode auf die hier geschilderten Probleme nahe liegend und wurde im Rahmen dieser Doktorarbeit ergänzend getestet.

Eine Beschreibung der Skalierung-Index-Methode kann derzeit der Webseite des Max-Planck-Instituts entnommen werden.

http://www.mpe.mpg.de/theory/new_technology/begin/index.html

4.6.3.1 Die Grundidee

Die Skalierungs-Index-Methode basiert auf dem Vermessen der lokalen Dichteverteilung als quantitative Charakterisierung der Punkteverteilung in n-dimensionalen Zustandsräumen. Wie auf der Webseite angegeben, ist dieses Verfahren wegen seiner allgemeinen Formulierung in ganz ähnlicher Weise auf die Untersuchung von zeitlichen Mustern in Messreihen, als auch für die Analyse von Strukturen in statischen Bildern anwendbar.

Da es sich bei den einzelnen Konformeren einer MD-Simulation um eine zeitliche Abfolge dreidimensionaler Bilder handelt, lag der Schluss nahe, dass dieses Verfahren auch zur automatischen Klassifizierung großer MD-Datensammlungen geeignet sein muss.

Betrachtet wird bei diesem Verfahren – ausgehend vom jeweiligen Referenzpunkt – die Verteilung weiterer Punkte bestimmter Eigenschaften im relativen Abstand zum Referenzpunkt, ausgehend von diesem Referenzpunkt (die Funktion der Eigenschaftendichteverteilung um diesen Punkt im 3D-Raum).

Im angewandten Verfahren, wird die Menge der gefundenen Eigenschaftspunkte zwischen den beiden verwendeten Radien ins Verhältnis zur Distanz der beiden Radien gesetzt. Der so erhaltene Wert lässt eine Aussage zu, ob der zentrale Referenzpunkt sich in einer punkttartigen, linearen oder flächigen Raumstruktur befindet.

Hierzu wird folgende Formel verwendet:

$$\alpha = \frac{\log N(r_2) - \log N(r_1)}{\log r_2 - \log r_1}, \text{ wobei} \quad (4.3)$$

$N(r_1)$ = der Anzahl der beachteten Punkte bis zum Radius r_1 entspricht

$N(r_2)$ = der Anzahl der beachteten Punkte bis zum Radius r_2 entspricht

r_1 und r_2 = stehen für die beiden eingesetzten Radien, wobei $r_1 < r_2$ gilt

Für den zweidimensionalen Fall gilt hier dann folgendes:

$\alpha \sim 0$; punkttartig

$\alpha \sim 1$; linear

$\alpha \sim 2$; flächig

4.6.3.2 Anwendung auf chemische Strukturen oder Taschenoberflächenpunkte

Nach dem im Abschnitt 4.6.3.1 beschriebenen Ansatz sollte sich die Punkteverteilung der einzelnen Koordinaten Punkte für eine beliebige dreidimensionale Struktur (Konformer oder Taschenoberflächenpunktesammlung) in eine diese Raumstruktur beschreibende Sammlung von α -Werten umsetzbar sein. Diese Sammlung kann dann mittels zusammenfassender Umwandlung in eine geeignete Vektorschreibweise (gruppierender Korrelationsvektor) für die Berechnung der Ähnlichkeit anhand einfacher Distanzmaße – wie Euklidischer- oder Manhattan-Distanz – verwendet werden.

Diese so abgeleiteten Korrelationsvektoren beschreiben die Zusammensetzung der Struktur aus unterschiedlichen Dichteverteilungen (Substrukturformen).

Für diese Korrelationsvektoren liefert somit die Bestimmung der Ähnlichkeit mittels Distanzmaß eine Aussage über die Substrukturähnlichkeit dieses jeweiligen Paares.

Da die Berechnung des Korrelationsvektors einzig auf der Verteilung innerhalb der zu kodierenden Struktur beruht, ermöglicht dies die Berechnung und Speicherung der so kodierten Substrukturzusammensetzungsinformationen im Voraus für spätere Vergleichsanalysen.

Die Natur dieses Korrelationsvektors lässt ferner die Suche nach Substrukturelementen spezifischer α -Werte-Zusammensetzung zu, welche beispielsweise „interessante“ Taschenformen kodieren können und somit die Suche innerhalb der Taschenkoordinaten-Datensammlung ohne die Probleme einer „Alignment“-Analyse (wie zu definierende Wechselwirkungspaare) zu beinhalten.

Hierbei gilt: sollte eine bestimmte α -Werteverteilung für eine Substruktur in einem solchen Korrelationsvektor auftreten, besteht die Möglichkeit, dass diese Substruktur in der betreffenden größeren kodierten Struktur vorhanden ist. Sollte jedoch diese α -Werteverteilung nicht in dem betreffenden Korrelationsvektor vorhanden sein, kann bestätigt werden, dass auch die so kodierte Substruktur nicht in der größeren kodierten Struktur vorhanden ist.

4.6.3.3 Welche Koordinaten sind zum Einsatz geeignet?

Generell lassen sich alle molekularen Koordinatensammlungen verwenden, deren Punkteverteilung eine ausreichend hohe Auflösung für die Bildung eines differenzierenden Dichtegradienten in der Untersuchung verwendeten Radiuspaarung aufweisen können – oder anders gesagt: es müssen genügend Punkte zwischen den beiden Radien vorhanden sein können, um eine Differenzierung zu ermöglichen.

Für Atomkoordinaten ist dies nur zur globalen Betrachtung sehr großer Komplexe sinnvoll anzuwenden. Bei der Verwendung kleinerer Strukturen muss das Auflösungsvermögen eventuell – wie im Rahmen dieser Arbeit mittels Oberflächenberechnung – durch die Berechnung weiterer 3D-Punktendarstellungen der gewählten Strukturen erfolgen.

4.6.3.4 Molekulare SIMPrints

Im Rahmen dieser Arbeit wurde der besagte Ansatz der Skalierungsindexmethode in die Strukturauswertungssoftware-Software *Pandora* implementiert.

Hierbei wurde besagte Berechnungsweise wie folgt umgesetzt:

Für eine in das Programm geladene Struktur können die α -Werte für jeden Koordinatenpunkt berechnet werden. Dies erfolgt basierend auf einer Auswahl hierzu zu verwendender Atomklassen. Im Prototyp wurden die einzelnen Elemente zur Unterscheidung verwendet, oder der gesamte Koordinatensatz (alle Elemente einbeziehend) angewendet.

Abgeleitet aus den obigen Überlegungen ergeben sich somit Korrelationsvektoren, die die Form der Verteilung der ausgewählten Koordinatenpunkte innerhalb der betrachteten Strukturen kodieren. Diese werden weiterhin in Anlehnung an den „molekularen Fingerprint“ als SIMPrint bezeichnet. Ein SIMPrint, ist somit das Histogramm der lokalen Eigenschaftendichteverteilung innerhalb eines Moleküls.

4.6.3.5 Darstellung der Berechnung eines SIMPrints anhand von zwei Beispiel-Referenzpositionen in einem TAR-Konformer

Zur Visualisierung der Erstellung eines SIMPrint wurde die Bild-Sequenz in Abbildung 113 erstellt. Sie zeigt die sequenzielle erfolgende Auswahl von Strukturausschnitten aus der Molekül-Superstruktur, welche auch im Rahmen der SIMPrint-Berechnung erfolgen könnte und einem besseren Verständnis des erläuterten Sachverhalts dienen soll. Hierfür wurde sich auf zwei Bereiche des Moleküls beschränkt, die durch ihren unterschiedlichen räumlichen Aufbau zur Visualisierung geeignet erschienen.

Hierbei werden für jedes Atom, welches den Kriterien eines Referenzpunktes entspricht (in diesem Fall: ein Kohlenstoff), die Populationen an Atomen desselben Kriteriums (in diesem Fall: alle Kohlenstoffatome) ausgewählt, welche sich innerhalb eines zu bestimmenden Radius um den Referenzpunkt befinden. Es wird für die Berechnung des α -Werts eines Referenzpunktes jeweils eine Kombination aus den Populationsmengen zweier unterschiedlich großer radialer Ausschnitte benötigt.

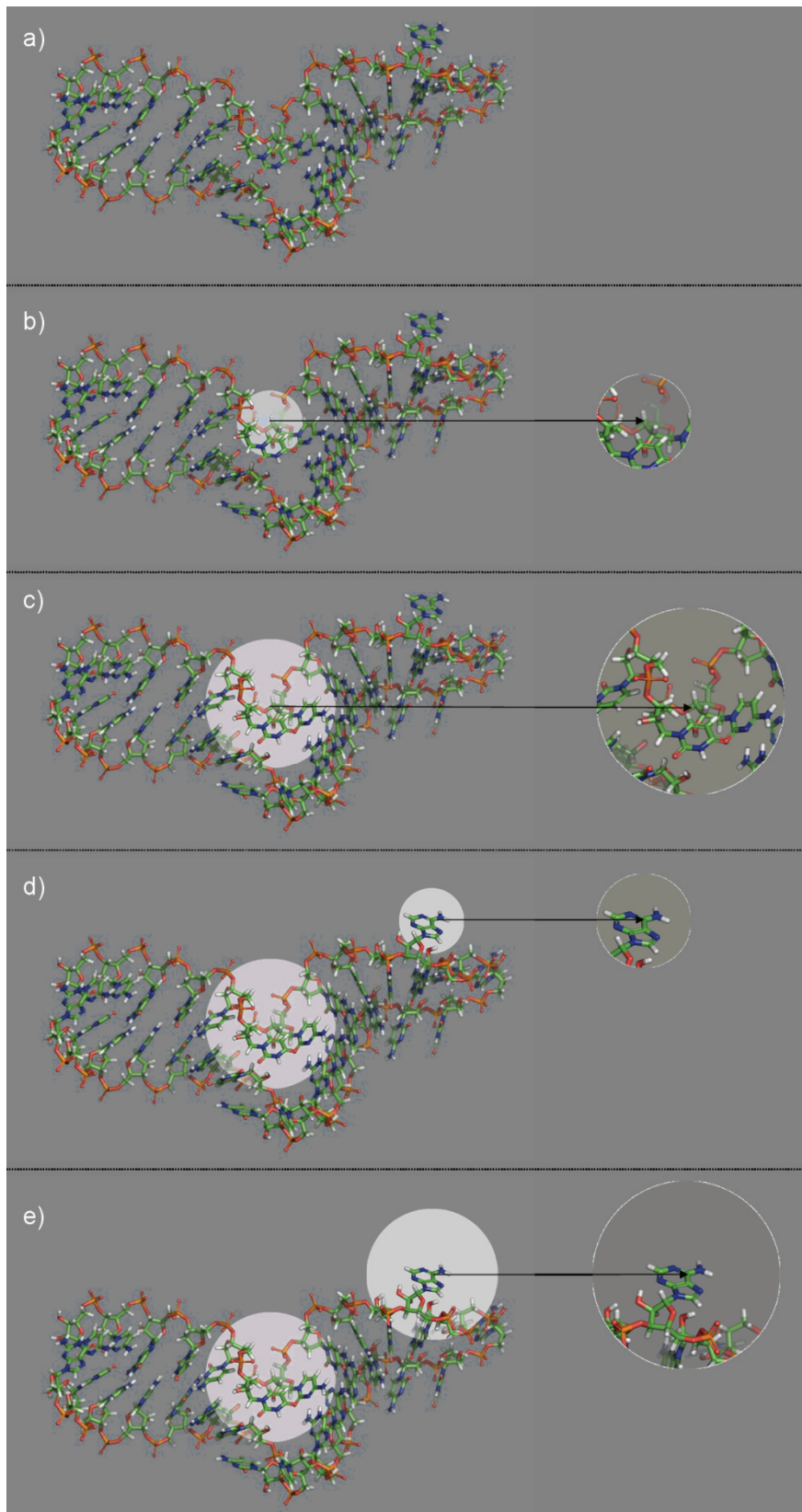


Abbildung 113: Visualisierung zweier Berechnungsschritte für einen SIMPrint einer TAR-Struktur; a) zeigt den TAR-Konformer als ganzes; b) und c) zeigen die Erstellung der Radialausschnitte um den ersten Referenzpunkt; d) und e) zeigen dies für den zweiten Referenzpunkt

4.6.3.6 Erste Anwendungen der SIMPrints

Im Rahmen dieser Arbeit wurden erste Tests mit dem SIMPrint-Deskriptorsystem anhand zweier Beispielsets für den molekularen Einsatz durchgeführt: atombasiert und oberflächenbasiert.

Hierbei wurde geprüft, ob die bisherigen Implementierungen bereits in dieser Phase des Projekts in der Lage sind, Abbildungen der 3-dimensionalen Raumstruktur auf die abgeleiteten Korrelationsvektoren zu ermöglichen.

Durchführung:

Die Berechnung der SIMPrints erfordert das Festlegen zweier Radien für die Berechnung. Im hier angewendeten Fall wurden diese Radien willkürlich festgesetzt, wobei folgende Überlegungen angewendet wurden.

- Radius 1 muss größer als eine CC-Bindungslänge sein, da bei kleinerem Radius keine Differenzierung der Nachbarschaft möglich ist. Wenn nur das Referenzatom selber innerhalb des Radius gefunden werden kann, ergäbe sich sonst nur ein konstanter Wert.
- Radius 2 sollte kleiner als der maximale halbierte Durchmesser der zu betrachtenden Struktur sein, um eine Differenzierung der Form zu ermöglichen. Bei zu großen Radien, geht der Wert wieder gegen einen konstanten Wert und das Diskriminierungsvermögen nimmt ab. Somit wurde ein Wert von 8,5 Ångström als sinnvoll im Rahmen der Aufklärung der lokalen Dichteverteilung angenommen.

Test 1:

Atomkoordinaten von PDB-Eintrag 1ANR (20 Konformere)

Es wurden die SIMPrints für die Atomkoordinaten mit folgenden Einstellungen erstellt:

- Radius 1 = 3,5 Ångström
- Radius 2 = 8,5 Ångström
- Binanzahl = 50
- Maximaler betrachteter α -Wert = 5,0
- Atomtypenauswahl = alle Atome (spezifische Auswahl jeweils für C, N, O, P und H)

Es wurden die so erstellten fünf atomspezifischen Deskriptor-Vektoren hintereinander zu einem kombinierten 250 dimensional Deskriptor zusammengefügt.

Mit diesem kombinierten Deskriptor wurden verschiedene Clusterverfahren getestet, von denen hier der HK-means-Baum wiedergegeben wurde; siehe Abbildung 114.

Test 2:

Taschenoberflächenkoordinaten der jeweils größten Tasche eines jeden Konformers des PDB-Eintrags 1ANR (20 Konformere).

Die SIMPrints für die Taschenoberflächenkoordinaten wurden mit folgenden Parametern erstellt:

- Radius 1 = 3,5 Ångström
- Radius 2 = 8,5 Ångström
- Binanzahl = 100
- Maximaler betrachteter α -Wert = 5,0

- Atomtypenauswahl = alle Oberflächenkoordinaten (keine Auswahl nach Atomtypen, jede Taschenoberflächenkoordinate wurde verwendet)

Mit diesem kombinierten Deskriptor wurden verschiedene Clusterverfahren getestet, von denen hier der HK-means-Baum wiedergegeben wurde; siehe Abbildung 115.

Ergebnisse:

Test 1:

Abbildung 114 zeigt die Einteilung der Strukturen in Subklassen nach dem 250-dimensionalen Test-Deskriptor erstellt durch Hk-mean-Clustering. Für den erstellten Beispielbaum wurden die einem Blatt zugrunde liegenden Strukturen, mittels Kabsch-Alignment (Kabsch 1976) ausgerichtet und je nach Anzahl der Strukturen in unterschiedlicher Farbe gerendert. Die Visualisierung erfolgte in PyMol, als Cartoon.

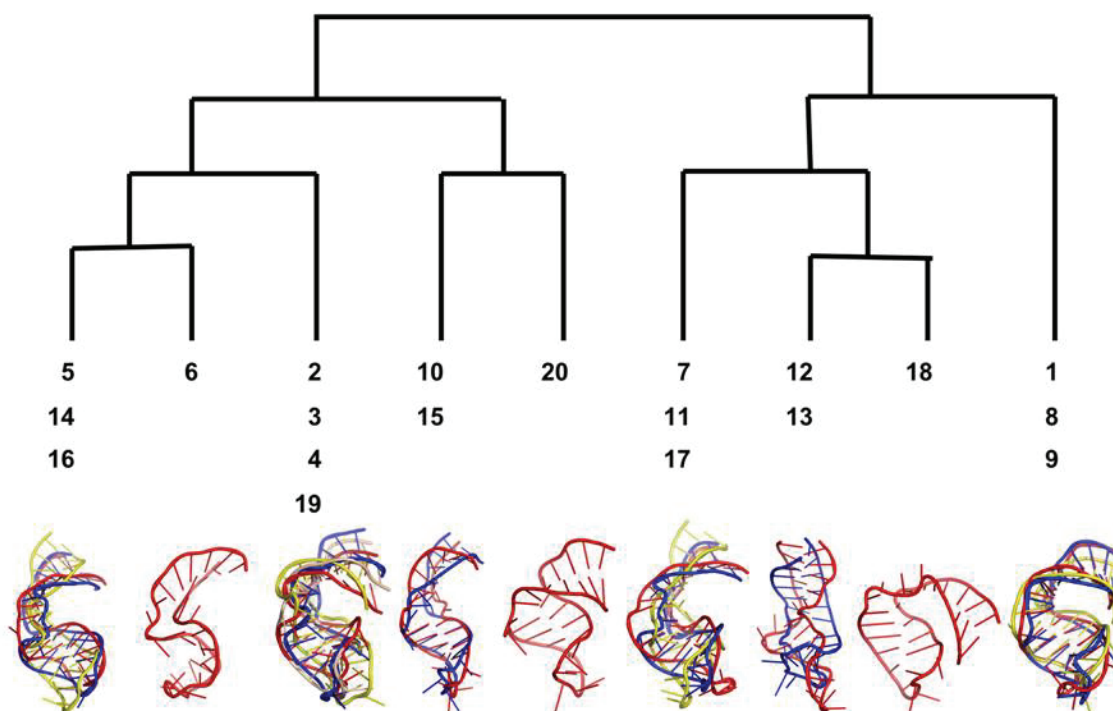


Abbildung 114: HK-means-Baum der Konformere der 20 Konformere des PDB-Eintrags 1ANR basierend auf einer alle Atomkoordinaten einschließenden SIMPrint-Berechnung; Als Blätter des Baums sind die Visualisierungen der alignierten Strukturen, die den SIMPrints zugrundeliegenden Konformere dargestellt. Die Zahlen unter dem Baum geben die Modellnummern im PDB-Eintrag wieder.

Test 2:

Abbildung 115 zeigt die Einteilung der Taschenoberflächen in Subklassen nach Hk-mean-Clustering – durchgeführt mit dem 100-dimensionalen Test-Deskriptor für die Taschenoberfläche ohne Diskriminierung zugrunde liegenden Atomtypen.

Beispielhaft wurden hier die Taschenpaare, die nach dem Baum in den vier Hauptästen in den jeweils benachbarten Blättern vorkommen als Taschenoberflächenrepräsentation gerendert in PyMol. Das Alignment erfolgte aufgrund fehlender geeigneter Software per Hand und sollte die Ähnlichkeit der jeweiligen Paare hervorheben.

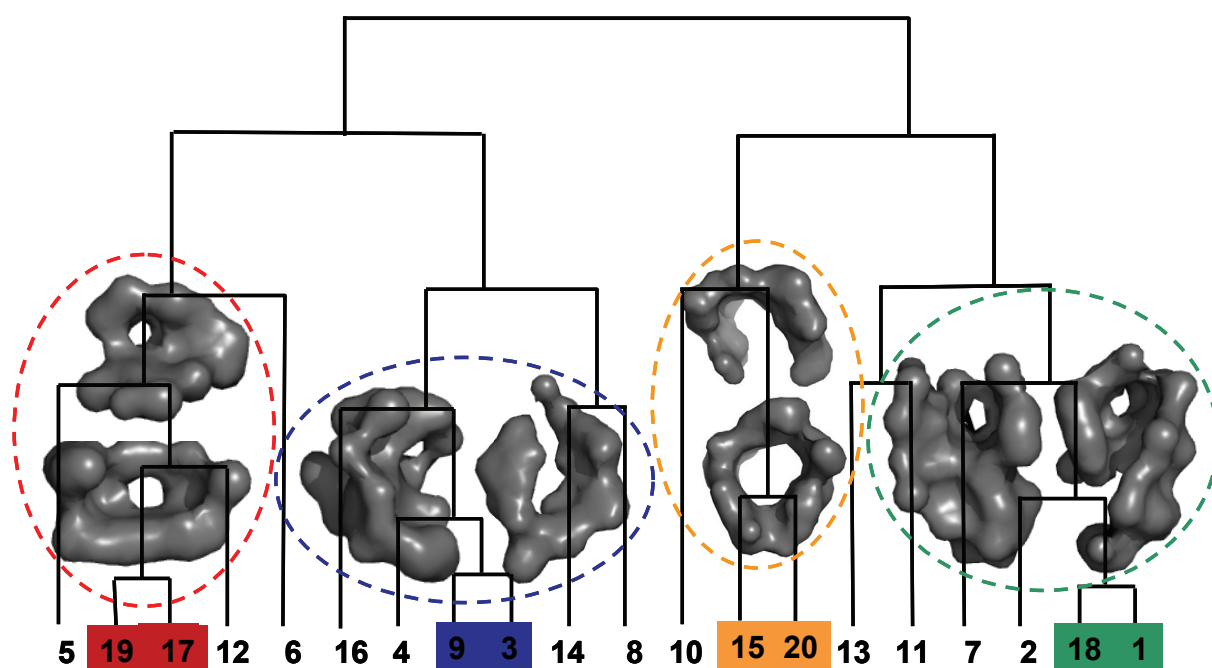


Abbildung 115: Baum aus dem Hk-Means-Clustering der SIMPrints der Taschenoberflächen der jeweils größten Tasche jeden Konformers der 1ANR; exemplarisch wurde die Oberfläche von vier, der nach diesem Kriterium ähnlichsten Taschenpaare, visualisiert.

Diskussion:

Die in Abbildungen 114 und 115 gezeigten Bäume und die Visualisierungen der zugrunde liegenden Strukturen, zeigen eine Unterscheidungsfähigkeit der hier erprobten Deskriptoren für die Form der Strukturen. Aufgrund des hierbei jedoch noch enthaltenden Prototypcharakters der verwendeten Parameter der bisherigen Tests, ist zwar die generelle Einsetzbarkeit dieses Ansatzes erkennbar, jedoch müssen noch durch chemieinformatische Anpassungen, wie beispielsweise Pharmakophorrepräsentationen, der Nutzen dieser Methode erhöht werden. Da es sich ferner um ein neues Ähnlichkeitsverfahren handelt, das aufgrund seiner auf lokalen Dichteunterschieden basierenden Ableitungen auf eine relativ große Anzahl an Punkten angewiesen ist, fällt der Vergleich zum als Standard zu sehenden RMSD schwer. Dieser Vergleich kann auch nur auf den Atomkoordinaten-basierten Anwendungen beruhen, so dass für die Güte der Unterscheidungsfähigkeit für die Taschenoberflächenrepräsentationen noch kein anderes Standard-Verfahren zur Verfügung steht. Dementsprechend ist eine genaue Bewertung hier nicht vorzunehmen. Jedoch zeigt sich nach optischer Betrachtung der vier Paare eine größere Ähnlichkeit innerhalb der benachbarten Taschen, als dies zu anderen Taschenpaaren der Fall ist.

Schlussfolgerung:

Die hier vorgestellten SIMPrints, befinden sich zwar noch in der Prototyp-Phase, was Bestimmung der optimalen Radien und Auswahl, sowie Kombination unterschiedlicher Merkmale der Koordinaten angeht, jedoch zeigen die ersten Clusterungsversuche mit dieser speziellen Form von Deskriptoren und den resultierenden Visualisierungen der Ergebnisse eine grundsätzliche Anwendbarkeit auf die Molekül- oder Taschenbeschreibung zur eventuellen späteren Anwendung bei der Indizierung der Datenbank.

Weiterführende Versuche:

Im Rahmen der Arbeit wurden ebenfalls Tests zur Optimierung der ausgewählten Radien durchgeführt. Dies umfasste den Einsatz des hierzu angepassten Cyclops-Java-Moduls (Schüller und Schneider 2008) in einem Javamodul von Pandora bei dem heuristische Verfahren angewendet wurden, um eine auf zwei zu optimierende Testradien resultierende Testdeskriptor-basierende Distanzmatrix einer RMSD-basierten Distanzmatrix anzupassen. Dies konnte allerdings nur für den atomkoordinatenbasierten Ansatz durchgeführt werden, deren geringe Punkteanzahl jedoch nicht für den SIMPrint-Ansatz geeignet scheint. Jedoch konnte hier bereits gezeigt werden, dass dieses Modul in der Lage ist, eine Optimierung vorzunehmen.

Ausblick:

Zur genaueren Beschreibung der aus der Datenbank abgeleiteten Taschenoberflächen, sollte innerhalb der Anwendung die Verwendung von Pharmakophor-Beschreibungen anstelle der Element-basierten Einteilung der Oberflächenpunkte umgesetzt werden. Dies sollte eine genauere Beschreibung der Taschen für ihre Eignung als potentielle Bindestelle ermöglichen, so dass in Zukunft damit begonnen werden kann, Taschen anhand des Vorkommens von in ihrer Oberflächenbeschaffenheit ähnlich aufgebauter Taschen in der Datenbank zu gewichten.

4.7 Oberflächenanalyse für 1ANR aufgelöst nach Elementen

Hintergrund:

Die Interaktion der TAR mit potentiellen Interaktionspartnern wie dem Tat-Peptid, erfolgt nach Diffusion durch Ausbildung von gemeinsamen komplementären Interaktionsflächen. Diese Oberflächenstücke sollten, falls es sich um eine spezifisch Interaktion der beiden Bindungspartner handelt, in dieser Kombination nur bei diesem Paar vorkommen. Dementsprechend scheint es interessant für die Untersuchung der TAR zu sein, ob diese sich in der Oberflächenzusammensetzung zu anderen RNA-Strukturen unterscheidet. Ferner sollte die Diversität der TAR-Konformere, wie in PDB-Eintrag 1ANR zu sehen, möglicherweise in einer unterschiedlichen Zusammensetzung der Atome, die die jeweilige Oberfläche ausmachen resultieren. Sollte dies nicht der Fall sein, kann nicht davon ausgegangen werden, dass spezifische Wechselwirkung zu nur einem Konformer möglich sind. Besonders interessant wäre hierbei die Überprüfung der Taschenoberflächenzusammensetzungen, da die Taschen als Bereiche der Konformere definiert werden, an denen es zu den dauerhaftesten Bindungen kommen sollte.

Erste Tests:

Im Rahmen der Arbeit sollte mit einem schnellen Test ermittelt werden, ob sich die Oberflächenzusammensetzung des TAR-Konformers durch die unterschiedliche Raumstruktur stark verändert. Hierzu wurde die Zusammensetzung der Oberfläche aus den jeweils durch bestimmte Elemente erzeugten Punkten analysiert.

Durchführung:

Für diese erste Prototypuntersuchung wurde dieses Mal die gesamte Oberfläche des jeweiligen TAR-Konformers, auf die Anteile der diesen Oberflächenpunkten zugrunde liegenden Elementtypen (P, O, C, H und N) untersucht. Hierzu wurde die Summe aller gefundenen Oberflächenpunkte eines bestimmten Elementtyps verwendet, anschließend wurden diese Anteilssummen jeweils mit der Gesamtsumme von Oberflächenpunkten für diesen Konformer in Relation gesetzt um einen Prozentanteil an der Oberfläche anzuleiten; siehe hierzu Abbildung 116.

Ergebnisse:

Die Abbildung 116 zeigt die Auftragung der berechneten Oberflächenprozentangaben für die Elemente C, N, H, O für den PDB-Eintrag 1ANR; für P wurde keine Auftragung durchgeführt, da deren Anteil durchgängig < 0,01 Prozent war.

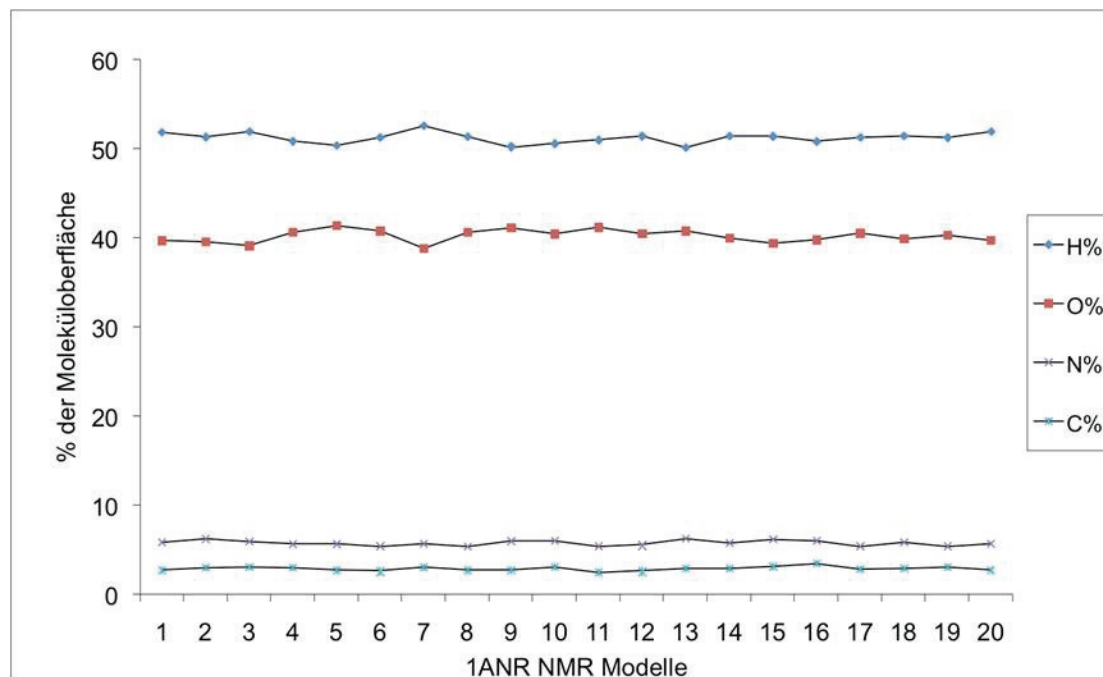


Abbildung 116: Oberflächenaufteilung des PDB-Eintrags 1ANR nach Elementen C, N, O und H für jeden der zwanzig Konformere; P wurde nicht aufgetragen, da deren Anteil durchgängig < 0,01 % war.

Diskussion:

Als Arbeitshypothese wurde angenommen, dass die sich relativ stark in ihrer räumlichen Struktur unterscheidenden Konformeren aufgrund dieser relativen Bewegungen unterschiedliche Atomtypen an die Oberfläche stellen würden. Dies konnte jedoch nicht bestätigt werden; siehe hierzu Abbildung 116. Zumindest unter dieser globalen Betrachtung lässt sich kein unterschiedlicher Aufbau der resultierenden Oberflächen der Einzelkonformere ausmachen.

Schlussfolgerung:

Als – wenn auch noch sehr gewagtes – Modell lässt sich somit anhand der ca. 40 Prozent der Oberfläche, welche durch Sauerstoff ausgemacht wird, davon ausgehen, dass sobald ein Ligand eine gewisse Polarisierung mit partial positiv geladenen Bereichen oder besser noch positiven Ladungen besitzt, er zwangsläufig über elektrostatische Wechselwirkungen an die Oberfläche gebunden werden kann. Dies erfolgt ohne besondere Diskriminierung bestimmter Oberflächenbereiche, wobei dem Rückgrat der TAR ein besonderes Gewicht zufallen wird, was aber aus dieser Analyse nicht ersichtlich ist; siehe Abbildung 116.

Die Analyse der Oberflächen erlaubt anscheinend eine schnelle und einfache Ableitung globaler Oberflächeneigenschaften für den generell wahrscheinlichsten, jedoch ortsunspezifischen Bindemechanismus. Allerdings ist die räumliche Auflösung dieses Ansatzes, mit seiner Mittelung über alle Bereiche der Zielstruktur noch wenig informativ. Später kann

dies jedoch auch für die Taschenoberflächen erfolgen, wobei hier eher auf eine Pharmakophorrepräsentation übergegangen werden sollte, die zu diesem Zeitpunkt noch nicht implementiert wurde. Wie in Abschnitt 4.8 noch vertieft wird könnten hierbei bevorzugt „elektronenreiche Taschen“ (Hermann und Westhof 1999), respektive Taschenoberflächen als potentiell primär zu identifizierende Bindestellen dienen.

4.8 Abschließende Untersuchungen der Interaktion der bekannten Peptid-Liganden mit TAR

Hintergrund:

Im Rahmen dieser Arbeit wäre es interessant gewesen, die Wechselwirkung der TAR mit ihrem „natürlichen“ Liganden, dem Tat-Peptid, zu untersuchen. Da das Tat-Peptid, im Rahmen dieser Arbeit auch in den Versuchen im Labor (FRET-Assay) als Referenzligand innerhalb des Verdrängungsassays verwendet wurde, wäre somit die Raumstruktur dieser Bindung eine interessante Zusatzinformation bei der Betrachtung der Fragestellung der Störung dieser Wechselwirkung gewesen. Für die hier untersuchte TAR-Struktur von HIV-1 liegt jedoch bisher keine aufgeklärte Struktur mit dem gebundenen Tat-Peptid vor. Als beste Alternativen liegen hierzu bis jetzt leider nur die Raumstrukturen der BIV-TAR mit dem jeweiligen Tat-Referenz-Peptid vor (PDB-Eintrag 1BIV (Ye et al. 1995)). Zusätzlich existiert die Raumstruktur eines weiteren zyklischen Peptid-Analogons mit Bindung an BIV-TAR (PDB-Eintrag 2A9X (Leeper et al. 2005)).

Für das im FRET-Assay verwendete Tat-Peptid liegen solche Informationen nicht vor. Jedoch zeigen neue Versuche mittels Massenspektroskopie, dass das Peptid, wohl nicht nur einen einzigen Bindemodus an die TAR aufweist. Bei der Untersuchung mittels des LILBID (Laser Induced Liquid Bead Ion Desorption)-Verfahrens (Morgner *et al.* 2006) konnte gezeigt werden, dass an die untersuchte TAR zwei gelabelte Tat-Peptide binden [pers. Komm. Dr. Ute Scheffer; Manuskript in Vorbereitung]. Dementsprechend scheint es nicht nur einen Bindemodus für das Tat-Peptid an TAR zu geben.

Fragestellung:

Welche Bindemöglichkeiten weist das Tat-Peptid an der TAR-RNA auf?

Welche Strukturelemente sind für die Bindung in den bereits aufgeklärten Tat-TAR-Komplexen zu erkennen?

Mit welchem Bindungstyp konkurrieren die potentiellen RNA-Liganden?

Kommen diese Struktur motive auch bei den in dieser Arbeit gefundenen kleinen Liganden vor?

Tat-Sequenzen:

1BIV: SGPRPRGTRGKGRRIIR

2A9X: RVRTRGKRRIRVPP

FRET-Assay-Tat: „Fluoreszein-Tag“-AAARKKRRQRRRAAA-„Rhodamin-Tag“

Analyse 1: Wie viele Ladungen sind in den Peptiden zu finden?

Analyse 2: Wie viele polare Aminosäuren sind in den Peptiden zu finden?

Analyse 3: Wie viele aromatische Aminosäuren sind in den Peptiden zu finden?

Ergebnisse:

1BIV: SGPRPRGTRGKGRIR
7x R, 1x K, 0x H : 8 positive Ladungen
polare Aminosäuren: 1x S, 1x T (zwei OH-Gruppen)
unpolare Aminosäuren: 1x I, 4x G, 2x P
aromatische Aminosäuren: keine

2A9X: RVRTRGKRRIRVPP
6x R, 1x K, 0x H : 7 positive Ladungen
polare Aminosäuren: 1x T (eine OH-Gruppe)
unpolare Aminosäuren: 2x V, 1x I, 1x G, 2x P
aromatische Aminosäuren: keine

FRET-Assay-Tat: „Fluoreszein-Tag“-AAARKRRQRRRAAA-„Rhodamin-Tag“
6x R, 2x K, 0x H : 8 positive Ladungen
Zusätzlich trägt das Rhodamin eine weitere positive Ladung: 9 positive Ladungen
polare Aminosäuren: 1x Q (eine Säureamid-Gruppe)
unpolare Aminosäuren: 6x A
aromatische Aminosäuren: keine

Diskussion:

Die Bindungsmöglichkeiten des Tat-Peptids

Die Untersuchung der Zusammensetzung der verwendeten an TAR-RNA-Strukturen bindenden Tat-Peptide ergab eine Häufung von Arginin und Lysin als verwendete Bausteine. Ferner konnte in keiner der untersuchten Peptide eine aromatische Aminosäure gefunden werden. Dies erlaubt eine Abschätzung der möglichen Wechselwirkungen, die das jeweilige Peptid eingehen kann.

In der hauptsächlich hier untersuchten, für den Tat-TAR-FRET-Assay relevanten Peptid-Sequenz (RKKRRQRRR, die tri-A-Sequenz am Anfang und Ende sind Modifikationen für das Anbringen der FRET-Label und werden für diese Betrachtung genauso ignoriert wie die FRET-Label selber) kommen acht unter physiologischen Bedingungen als positiv geladen anzusehende Seitenketten vor sowie eine weitere polare Aminosäure.

Alle Aminosäuren des betrachteten 9-mers zeigen einen Aufbau aus einer längeren gesättigten und damit als flexibel zu betrachtenden Kohlenstoffkette, an deren Ende das Substrukturelement mit positiver Ladung gebunden ist. Das Peptid-Rückgrat ist ebenfalls als Quelle von Wasserstoffbrücken-Interaktionen zu sehen und durch die Abwesenheit von Prolin ebenfalls als flexibel zu betrachten.

Das Fehlen von aromatischen Seitenketten schränkt somit die Interaktionsmöglichkeiten dieses Peptids auf die Ausbildung von Wasserstoffbrücken, sowie Salzbrücken und eventueller Kation- π -Interaktionen ein. Aromatische Stacking-Interaktionen des Peptids, konnten somit ausgeschlossen werden (nur in den FRET-Labels befinden sich aromatische Systeme). Diese Ableitung an möglichen Wechselwirkungen ist auch mit den in der Veröffentlichung von Davis *et al.* – Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots" (Davis et al. 2004) – geschilderten Wirkungsweise der dort ermittelten Liganden konform.

Da die RNA aufgrund des Aufbaus des Molekülrückgrats aus Phosphatgruppen einen Polyanion-Charakter trägt, kann hierbei die Bindung des Tat-Peptid-Polykations nicht verwundern. Ferner konnten die durch die Analyse der Atom-Oberflächenverteilung der TAR in Abschnitt 4.7 gewonnenen Werte einer zu ca. 40% aus Sauerstoffatomen aufgebauten Oberfläche der TAR (1ANR – 20 Modelle), die Möglichkeit zur Ausbildung multipler Wasserstoffbrücken erkennen lassen.

Kann das Tat-Peptid spezifisch an die RNA-Binden?

Geht man davon aus, dass es sich bei dem Tat-Peptid tatsächlich um ein aus strukturell ähnlichen, flexiblen Bausteinen aufgebautes Polykation handelt, so ist die Festlegung auf nur einen Hauptbindemodus als unwahrscheinlich zu betrachten. Hinzu kommt noch die Flexibilität der Zielstruktur. Jedoch liegen hierzu keine genaueren Strukturdaten vor.

Die in einer massenspektroskopischen Untersuchung festgestellte Bindung zweier Tat-Peptide an eine TAR-RNA legt jedoch zumindest das Vorhandensein zweier gleichzeitig vorkommender unterschiedlicher Bindemoden nahe (Morgner et al. 2008). Da das Tat-Peptid und die TAR-RNA jedoch jeweils nur Teile größerer Strukturen sind, lässt sich über die Interaktion des HIV-1-Transkripts mit dem Tat-Protein zurzeit nur spekulieren (Hermann und Westhof 1999).

Kommen diese Interaktionsmöglichkeiten auch bei den in dieser Arbeit gefundenen kleinen Liganden vor?

Lässt man die Möglichkeit für die Ausbildung von Wasserstoffbrücken einmal in der folgenden Betrachtung außen vor (alle getesteten Verbindungen, wiesen prinzipiell die Möglichkeit auf Wasserstoffbrücken einzugehen), so bleiben die Interaktionsmöglichkeiten mittels Salzbrücken und eventueller Kation- π -Interaktionen übrig. Für dieses Interaktionsmuster sind folglich fixierte positive Ladungen oder hauptsächlich – unter physiologischen Bedingungen – protoniert vorliegende Substrukturmuster erforderlich. Die aus den Liganden-basierten Screeningrunden gewonnenen potentiellen TAR-Liganden, zeigen genau diese Interaktionsmöglichkeiten:

- M32 trägt eine fixierte positive Ladung; siehe Abbildung 42 Seite 70
- M23, AS20 und LR9 tragen jeweils mindestens eine erweiterte Guanidin-Gruppen (Methylenaminoguanidin); siehe Abbildung 74 Seite 119.

Wie könnten die gefundenen RNA-Liganden mit der TAR interagieren?

Laut der Studie von Hermann und Westhof (1999) könnten elektronegative Atomen besetzte Taschen in der TAR für die Bindung von kationischen RNA-Liganden geeignet sein. Die vier in den Messreihen gefundenen, positive Ladungen tragenden RNA-Liganden (M23, M32, AS20 und LR9) könnten mit diesen Zielbereichen stärker interagieren und somit die Bindung der Guanidin-Gruppen des Tat-Peptids ersetzen. Ferner sollte in einer Studie des Arbeitskreises Göbel der Einfluss des Protonierungsgrads auf die Aktivitätswerte von potentiellen RNA-Liganden für eine Testreihe von Pyrazol-Derivaten bei unterschiedlichen pH-Werten im Tat-TAR-Assay überprüft werden. Hierbei konnte bei erniedrigtem pH eine Steigerung der Aktivitätswerte für die untersuchten Pyrazol-Derivate gezeigt werden (pers. Komm. Dr. Ute Scheffer). Wie in der Gegenüberstellung in Abbildung auf Seite 63 für M23 und das tri-Arg-Peptid gezeigt, ähneln sich die gezeigten Strukturen im Vorhandensein von Guanidin-Gruppen den Arginin-Seitenketten des Tat-Peptids. Die Einführung der

Methylenamino-Linkergruppe kann durch die Veränderung des pK_A der Guanidin-Gruppe als Hauptursache gewertet werden, weshalb die Verbindungen M23, AS20 und LR9 im Gegensatz zu den weiteren vorhandenen Substanzen mit Guanidin-Gruppen der Messreihen im Tat-TAR-FRET-Assay als aktiv klassifiziert wurden. Eine einfache Reduktion des Bindeverhaltens auf das bloße Vorhandensein von positiven Ladungen als einziges Argument für die Aktivität im Tat-TAR-Assay kann jedoch analog zur bereits im Abschnitt 3.2.17.1 erfolgten Gegenüberstellung der im CFivTT-Assay aktiven und inaktiven kationischen Testsubstanzen ausgeschlossen werden. Wie in Abbildung 40 auf Seite 68 zu sehen ist, enthalten die Messreihen weitere kationische Verbindungen, von denen jedoch nur M32 als potentieller TAR-Ligand im Tat-TAR-FRET-Assay bestätigt werden konnte.

Da ferner M32 auch eine Aktivität in den Assay-Systemen von Markus Weidlich zeigt (Abschnitt 3.2.15), kann davon ausgegangen werden, dass die Aktivität von M32 im Tat-TAR-FRET-Assay nicht auf eine spezifische Interaktion mit der TAR-RNA beruht. Dahingegen zeigten die Testverbindungen M23, AS20 keine Wirkung im CFivTT-Assay (Die Aktivitätswerte der Substanzen der LR-Reihe wurde noch nicht vollständig bestimmt), woraus sich schließen lässt, dass diese Verbindungen ein eher spezifisches Bindeverhalten für die TAR-RNA zeigen könnten.

5. Schlussfolgerung und Ausblick

Zusammenfassend scheint es für zukünftige Studien besonders wichtig/lohnenswert, die Lage der potentiellen Bindestellen der RNA-Zielstruktur mit elektronenreichen Oberflächen (Hermann und Westhof 1999) zu identifizieren. Der Einfluss der Konformation auf die Bindevorhersage und der Erweiterung der experimentellen Datengrundlage mittels MD-Simulation wurde bereits von Marco und Gago (2007) beschrieben. So dass als Konsequenz hieraus und den eigenen Arbeiten die Zusammenführung der Identifikation potentieller Bindestellen und ihre Bewertung innerhalb des möglichen Konformerraums wünschenswert erscheint. Dies wird auch durch den Übersichtartikel von Foloppe *et al.* (2006) nahegelegt, die in ihrem Artikel folgendes erwähnen:

„For example, we are not aware of a program dedicated to detect and analyze the pockets and clefts in large folded RNA molecules, such as ribosomes. The sheer size of these structures precludes the equivalent manual analysis still frequently performed with proteins. Ideally, such software would rank cavities according to their druggability, possibly for virtual screening. Of course, the targeted sites need to be functionally relevant“(ebd. 1022).

Das Gesagte gilt in gleichem Maße für die Bindetaschenvorhersage in großen Konformerdatensammlungen, wie sie bei der MD-Simulation anfallen.

Das im Rahmen dieser Arbeit entstandene Programm Pandora ermöglicht die automatische Identifikation und Verwaltung von potentiellen Bindestellen für MD-Datensätze oder PDB-Einträge, wie dies als wünschenswert beschrieben wurde. Jedoch bedarf die Charakterisierung der Bindestellen nach den gewünschten, relevanten Kriterien noch weiterer zukünftiger Arbeit. Erste Schritte wurden hier bereits im Rahmen der Arbeit durchgeführt – jedoch hauptsächlich als Klassifizierung nach der Form der Taschen – daher bedarf es noch der Implementierung einer Bewertungsmöglichkeit der Taschenoberflächen nach potentiellen Bindestellen für kationische Substruktur motive der potentiellen RNA-Liganden. Basierend auf den aus dieser Arbeit gewonnenen „relevanten“ Eigenschaften der identifizierten Liganden sowie der aus weiteren Veröffentlichungen zu entnehmenden Fokussierung auf elektrostatische Wechselwirkungen (Hermann und Westhof 1999, Chin *et al.* 1999, Davis *et al.* 2004, Cai *et al.* 2008) wird diese Vorgehensweise ebenfalls nahegelegt.

Basierend auf den Ergebnissen der Liganden-basierten Screeningrunden, sollten für ein relativ TAR-RNA-spezifisches Liganden-Design Strukturen mit Methylenaminoguanidyl-Substrukturen (siehe auch Abschnitt 3.3.19) in ihrer Anwendung weiter erforscht werden, da diese Verbindungen im Vergleich mit dem CFivTT-Assay eine spezifische Aktivität im Tat-TAR-FRET-Assay zeigten.

Die im Rahmen dieser Arbeit als „allgemeine RNA-Binder-Filter“ vorgestellten Screeningverfahren (Abschnitt 3.2) können für zukünftige Arbeiten eine Reduzierung der potentiell zu testenden Substanzen aus Chemikalien-Datenbanken von Lieferanten ermöglichen. Sie sind als stark verallgemeinernder Ansatz jedoch nur als Vorfilter für weitere Tests einzusetzen, bzw. für eine parallele Anwendung der Vorhersagen für mehrere potentielle RNA-Zielstrukturen gleichzeitig empfehlenswert. Dies gilt, da so erneut das bei der Erstellung der Filter verwendete Wissen über vielfältige RNA-Bindungsstellen optimal genutzt werden kann, indem die Wahrscheinlichkeit eine erfolgreichen RNA-Liganden-Wechselwirkung vorherzusagen, mit jeder überprüften Zielstruktur für einen potentiellen Liganden steigt. Hierbei ist auch positiv zu bewerten, dass basierend auf den Erfahrungen im Rahmen dieser Arbeit die Kombination der Ergebnisse mehrerer Assay-Systeme (hier Tat-TAR-FRET-Assay mit dem CFivTT-Assay) weitergehende Einblicke in das mögliche Bindevverhalten der potentiellen RNA-Liganden ermöglichten, als dies nur mit dem Tat-TAR-FRET-Assay alleine möglich gewesen wäre (siehe exemplarisch Abschnitt 3.2.17.1). Desweiteren konnten so im Rahmen der Arbeit – zusätzlich zum angestrebten Fall der Tat-TAR-Inhibitoren – ebenfalls Substanzen entdeckt werden, die einen Effekt im zu

Testzwecken verwendete CFivTT-Assay-System hatten, so dass mit diesem Filter-Ansatz sowohl potentiell antivirale wie antibakterielle Verbindungen identifiziert werden können. Das Potential dieses Ansatzes sollte somit durch weitere Studien mit den mittlerweile vergrößerten Datensätzen durch Neuerstellung der ANN-basierten Filtersysteme einer weiteren Überprüfung unterzogen werden. Insbesondere könnten in einer weiterführenden Untersuchung noch eine Einbindung von SOMs zur Identifikation potentieller „RNA-Binder-Inseln“ im chemischen Raum herangezogen werden; siehe Abschnitt 3.2.19.2.

Da die ersten Versuche zur Clusterung der Taschen vielversprechend aussehen, aber sich noch in der Prototypphase befinden, sollten in weiteren Studien diese Ansätze weiter verfeinert werden. Desweiteren sollte in einer zukünftigen Projektphase zusätzlich zu den Ansätzen die Populationsverteilungen als Bewertungskriterium einzusetzen, auch die Bewertung der potentiellen Taschenstabilität eines Konformers anhand der aus den MD-Simulationen entnehmbaren Energiebetrachtungen (potentielle und kinetische Energie) einbezogen werden. Das Problem möglichst repräsentative Teilmengen des Konformerraums für das Docken in flexiblen RNA-Zielstrukturen zu definieren, konnte im Rahmen dieser Arbeit nicht abschließend geklärt werden, jedoch liegen einige Filterroutinen für die Selektion von Taschen mit nach sterischen Gesichtspunkten als derzeit „optimal“ zu betrachtenden Eigenschaften vor, wie sie in Abschnitt 4.5 allgemein und insbesondere durch die Definition der Taschen um den „Punkt P3“ in Abschnitt 0 erläutert wurden.

Die Möglichkeit Taschenoberflächen anhand des als SIMPrint eingeführten Ansatzes (siehe Abschnitt 4.6.3) zu beschreiben, ermöglicht die Clusterung der Taschen ohne auf den in diesem Fall nicht anwendbaren RMSD zurückgreifen zu müssen. In Kombination mit den Taschenvolumenformen beschreibenden Deskriptoren von PocketPicker von Martin Weisel (Weisel et al.2007), sollte die Beschreibung der Taschenoberflächen-Eigenschaftsverteilungen (als Pharmakophorrepräsentationen der Taschenoberflächenpunkte noch zu implementieren) eine verbesserte Clusterung ähnlicher Taschen erlauben. Dies sollte dann in zukünftigen Arbeiten für eine Bewertung der „Repräsentativität“ von Taschenteilmengen eine genauere Abbildung des Konformerraums ermöglichen, um diese dann für gezielte Docking-Studien verwenden zu können. Für diese Dockingstudien sollte dann auch die nun verfügbare DrugScore^{RNA}-Bewertungsfunktion (Pfeffer und Gohlke 2007) verwendet werden, um die bekannten Probleme mit den für das Protein-Liganden-Docking abgeleiteten Bewertungsfunktionen zu umgehen und so eine bessere Diskriminierung der einzelnen potentiellen Liganden zu ermöglichen; siehe Abschnitt 3.2.9.

Das im Rahmen dieser Arbeit definierte Taschenoberflächenbildungspotential (PSGP, siehe Abschnitt 4.5.8) – für die Atome einer Zielstruktur aus den vorhergesagten Taschenpositionen von Konformerensembeln abgeleitet – ermöglicht eine neuartige automatisierte Betrachtung von potentiellen Bindestellen in größeren Konformer-Datensätzen. Hierbei wird anhand des Auftretens möglichst konstanter Kontaktmöglichkeiten, bzw. dem durchschnittlichen Vorhandensein von Kontaktflächen für potentielle Liganden, welche durch bestimmte Atome der Zielstruktur innerhalb der auftretenden Taschenbereiche zur Verfügung gestellt werden, eine Einschätzung über die Zugänglichkeit dieser Interaktionsmöglichkeiten getroffen. Eine möglichst konstante Zugänglichkeit wiederum sollte einen Einfluss auf das durchschnittliche Bindegleichgewicht von Liganden haben, die mit den potentiellen Kontaktstellen interagieren können. Getreu dem Motto „Kontaktmöglichkeit ermöglicht Bindung/Bindung bedarf Kontaktmöglichkeiten“ können somit mit dieser Methode diejenigen Bereiche einer Zielstruktur identifiziert werden, die für ein Liganden-Design aufgrund der Antreffwahrscheinlichkeit in der Lösung (bei „Kontaktstoß“ von Ligand und Zielstruktur) am interessantesten erscheinen. Erste Tests hiermit für die Taschenselektion wurden bereits im Rahmen dieser Arbeit durchgeführt; siehe Abschnitt 4.5.9. Jedoch befindet sich auch dieser Ansatz noch in Prototypphase und bedarf der weiteren Untersuchung. In zukünftigen Arbeiten, könnte somit das Ausbilden einer Interaktion mit einem nach dem PSGP-Kriterium bevorzugten Atom zu einer Höherbewertung bei der Auswahl von potentiellen Liganden in zukünftigen Docking-Studien führen. Hierfür müsste aber wiederum erst die Fragestellung geklärt werden, welche

Konformere als repräsentative Teilauswahl für diese Docking-Studien herangezogen werden können.

Das im Rahmen dieser Arbeit erstellte Softwarepaket *Pandora* ist in seiner Anwendbarkeit nicht auf die hier untersuchte TAR-RNA beschränkt. In zukünftigen Studien können somit dieselben Analysen, die in dieser Arbeit beschrieben wurden mittels Pandora auch auf andere Zielstrukturen angewandt werden. Einige Module von Pandora sind zwar momentan auf die Analyse von RNA-Strukturen ausgelegt, jedoch sollten sich diese RNA-Sequenz-spezifischen Anwendungen (Projektion der PSGP-Werte auf die 2D-Darstellungen der Nukleotide) noch mittels einfacher Zusatzimplementierungen auf die Sequenz-spezifische Analyse von Proteinen erweitern lassen. Das PSGP ist jedoch auch ohne diese Visualisierungsmöglichkeit bereits anwendbar und mit einem anderen bereits erstellten Modul können beispielsweise die PSGP-Werte als B-Werte in einer im PDB-Format vorliegenden Strukturdatei abgelegt werden. Prinzipiell können somit mit Pandora beliebige molekulare Zielstrukturen untersucht werden. Als ein nächster Schritt könnte beispielsweise damit begonnen werden die verfügbare Strukturinformation für das Ribosom zu untersuchen. Ferner könnten der Ansatz des PSGP dazu benutzt werden, um bei verschiedenen Makromolekülen mittels der durch MD-Simulationen vorhandenen Bewegungsinformationen die Teile der Oberfläche zu identifizieren, welche relativ konstante jedoch leicht vergrabene Kontaktmöglichkeiten innerhalb der auftretenden potentiellen Bindestellen aufweisen. Diese Bereiche könnten möglicherweise für die Komplexbildung zwischen zu untersuchenden Makromolekülen besonders relevant sein. Denkbar ist hier auch die Anwendung zur Identifikation möglicher Bindestellen von kleinen Liganden auf den Oberflächen von Makromolekülen, um diese Bereiche dann gezielt einem Liganden-Design zu unterziehen. Die so zu erhaltenden potentiellen Liganden könnten dann möglicherweise die Bindung von anderen Makromolekülen in diesen Bereichen reduzieren.

Abschließende Anmerkung zur Namensgebung:

Da es sich bei Pandora– **P**ocket **A**nalysis for **D**Ocking on **R**epresentative **A**ptamers um einen vorläufigen Projektnamen handelt, der den Kern der Zielsetzung für die Implementierung dieser Software beschreibt, die jedoch immer noch nicht erreicht wurde, da die automatisierte Beantwortung der Frage „was ein repräsentativer Konformer ist“ noch nicht im Rahmen dieser Arbeit abgeschlossen wurde, lässt sich – mit einer gewissen Ironie – eine Ähnlichkeit zu einem von Johann Wolfgang von Goethes Werken (**Pandora**) nicht abstreiten: beide liegen bis jetzt nur als Fragment vor. In Anspielung auf die klassische Pandora mit ihrer „Büchse“ enthält das Software-Paket – auch als Pandora-Box bezeichnet – jedoch die „Hoffnung“ diesem Ziel mit seinen bisher implementierten Anwendungsmöglichkeiten und zum Erreichen dieses Ziels noch zu implementierenden Modulen, diesem letztendlich noch näher zu kommen.

6. Zusammenfassung

Das Ziel dieser Arbeit war es, RNA-Strukturen als potentielle Zielstrukturen für die Medikamentenentwicklung zu untersuchen. Hierbei ging es im Speziellen um die Anwendung Virtueller Screening Verfahren für die RNA-Liganden-Vorhersage. Hierzu wurde die als TAR-Motiv (transactivating response element) bekannte RNA-Struktur der mRNAs des HI-Virus ausgewählt. Diese Struktur wurde gewählt, da mit den vier PDB-Einträgen 1ANR, 1ARJ, 1LVJ und 1QD3 bereits experimentell motivierte Strukturmodelle zum Beginn der Untersuchung vorlagen. Ausschlaggebend war hierbei auch das Vorhandensein eines Tat-TAR-FRET-Assays im Rahmen des SFB 579, in welchem diese Arbeit angefertigt wurde. Die Aufmerksamkeit, welche dem HI-Virus im Rahmen der Bekämpfung der Immunschwächekrankheit bereits zukam, führte bei dem gewählten Testmodell ebenfalls zu einem, wenn auch immer noch überschaubaren Datensatz bereits getesteter Substanzen, der als Grundlage für einen Liganden-basierten Ansatz als erste Basis dienen konnte. Basierend auf diesen Voruntersuchungen ergaben sich die weiteren Schritte dieser Arbeit.

Die Arbeit lässt sich zusammenfassend in vier zum Teil parallel verlaufende Phasen einteilen:

Phase 1: Bestandsaufnahme bekannter Informationen über die Zielstruktur

- experimentell bestimmte Zielstrukturen
- experimentell bestimmte Liganden/Nichtliganden der Zielstruktur

Phase 2: Ableiten eines ligandenbasierten Ansatzes zur Vorhersage von potentiellen Bindern der Zielstruktur aus Substanzbibliotheken, der nicht auf Strukturdaten der Zielstruktur beruht.

Phase 3: Analyse der bekannten Konformere der Zielstruktur auf konstante Angriffspunkte für ein spezielles Liganden-Design.

Phase 4: Einbinden der bekannten Strukturinformationen der Zielstruktur zur weiteren Verfeinerung der Auswahlverfahren neuer Kandidaten für die weitere experimentelle Bestimmung des Bindeverhaltens.

Im Rahmen dieser Arbeit konnten mittels der Anwendung von künstlichen neuronalen Netzen in einem ligandenbasierten Ansatz durch virtuelles Screening der Chemikalien-Datenbanken verschiedener Lieferanten fünf neue potentielle TAR-RNA-Liganden identifiziert werden (drei davon mit einem Methylenaminoguanidyl-Substrukturmotiv), sowie als „Spin-Off“ durch die Anwendung der ursprünglich nur für den Tat-TAR-FRET-Assay vorgesehenen Testsubstanzen in einem Kooperationsprojekt (mittels CFivTT-Assay) zwei neue potentiell antibakterielle Verbindungen identifiziert werden.

Die Beschäftigung mit der offensichtlichen Flexibilität der TAR-RNA und damit einer nicht eindeutig zu definierenden Referenz-Zielstruktur für das Liganden-Docking führte zur Erstellung eines Software-Pakets, mit dem flexible Zielstrukturen – basierend auf den Konformer-Datensätzen von MD-Simulationen – auf konstante Angriffspunkte untersucht werden können. Hierbei wurde ausgehend von der Integration eines Taschenvorhersage-Programms (PocketPicker) eine Reihe von Filtern implementiert, die auf den hierzu in einer MySQL-Datenbank abgelegten Strukturinformationen eine Einschränkung des möglichen Taschenraums für das zukünftige Liganden-Design automatisiert vornehmen können. Des Weiteren ermöglicht dieser Ansatz einen einfachen Zugriff auf die einzelnen Konformere und die Möglichkeit Annotationen zu den Konformeren und den daraus abgeleiteten Tascheninformationen hinzuzufügen, so dass diese Informationen für die Erstellung von Liganden-Docking-Versuchen verwendet werden können. Ferner wurden im Rahmen dieser Arbeit ein neuer Deskriptor für die Beschreibung von Taschenoberflächen eingeführt: der auf der „Skalierungs-Index-Methode“ basierende molekulare SIMPrint. Die Beschäftigung mit der Verteilung der potentiellen Bindetaschen auf der Oberfläche der Konformerensemble führte ferner zur Definition der Taschenoberflächenbildungswahrscheinlichkeit (Pocket

Surface Generation Probability – PSGP) für einzelne Atome einer Zielstruktur, die tendenziell für die Einschätzung der Ausbildung einer potentiell langlebigen Interaktion eines Liganden mit der Zielstruktur herangezogen werden kann, um beispielsweise Docking-Posen zu bewerten.

7. Anhang

7.1 KS-Statistik zur Trennbarkeit potentieller RNA-Liganden von den potentiellen Protein-Liganden

Die folgende Tabelle enthält die vollständige KS-Statistik zum Abschnitt 3.2.4 „KS-Statistik zur Analyse der Eigenschaftsverteilung“ Seite 32

Tabelle 30: KS-Statistik zur Trennbarkeit von Protein-Liganden-Datensätzen von RNA-Liganden-Datensätzen

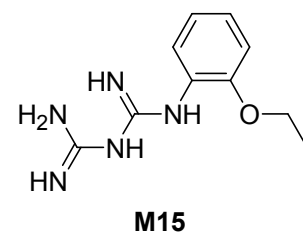
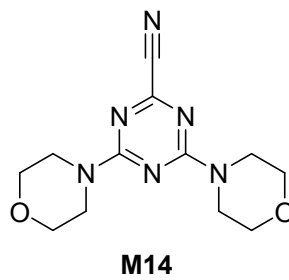
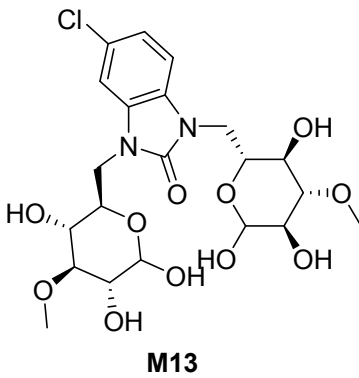
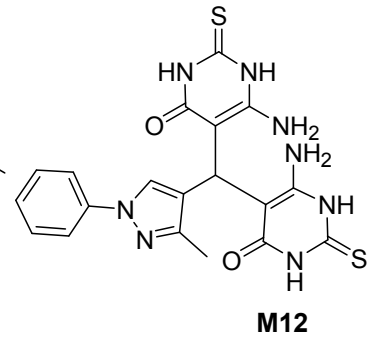
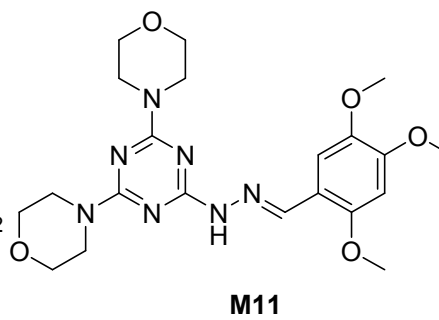
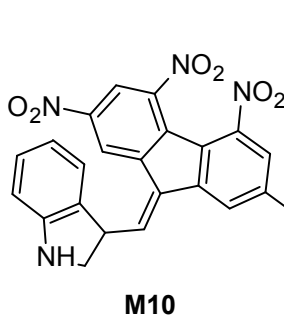
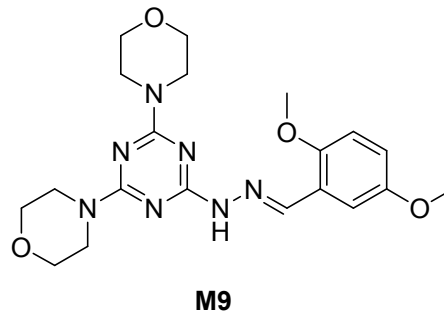
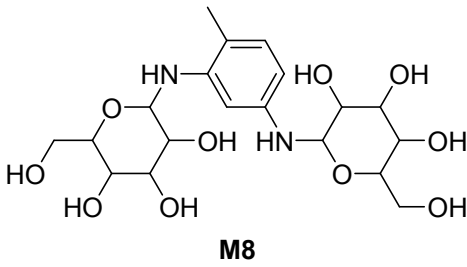
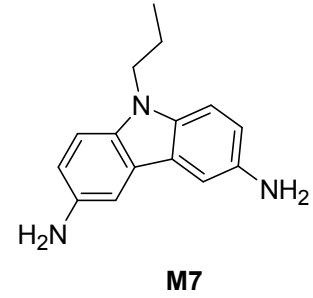
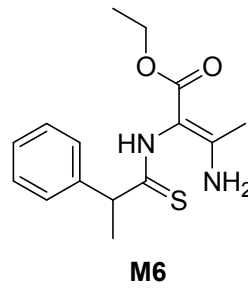
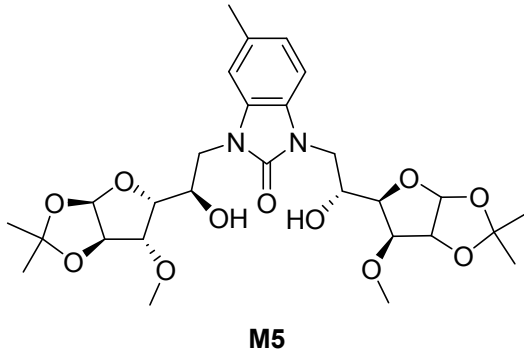
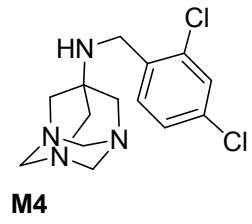
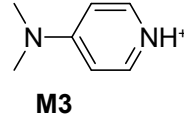
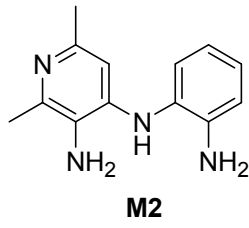
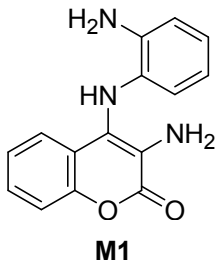
MOE-Deskriptor	KS-Wert
SlogP_VSA0	0,70729
Q_VSA_FPPOS	0,65902
Q_RPC+	0,64559
RPC+	0,64559
PEOE_VSA_FNEG	0,62753
PEOE_VSA_FPOS	0,62753
Q_VSA_PPOS	0,62376
SlogP	0,60625
PEOE_RPC+	0,59418
vsa_pol	0,55739
Q_VSA_FHYD	0,53624
Q_VSA_FPOL	0,53624
PEOE_VSA-6	0,51938
Q_VSA_FNEG	0,51018
Q_VSA_FPOS	0,51018
logP(o/w)	0,51018
Q_VSA_POL	0,49951
a_don	0,49649
a_base	0,48548
FCharge	0,47843
TPSA	0,456
PC+	0,45531
Q_PC+	0,45531
SlogP_VSA1	0,44368
a_aro	0,43284
b_ar	0,43284
PEOE_VSA_POS	0,4307
PC-	0,42981
Q_PC-	0,42981
KierA3	0,4268
PEOE_VSA_PPOS	0,42067
PEOE_PC+	0,41948

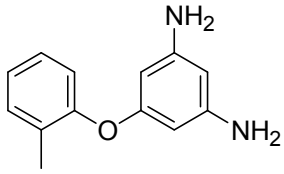
MOE-Deskriptor	KS-Wert
b_rotN	0,41927
b_1rotN	0,41626
KierFlex	0,41244
Q_VSA_POS	0,41082
PEOE_VSA-1	0,40905
SMR_VSA2	0,40751
KierA2	0,392
PEOE_VSA-4	0,39079
PEOE_VSA_POL	0,39047
b_1rotR	0,38876
PEOE_VSA+4	0,38795
SlogP_VSA4	0,38348
SMR_VSA3	0,37949
PEOE_PC-	0,37388
Q_VSA_HYD	0,37348
a_nO	0,37085
SlogP_VSA7	0,37041
SMR_VSA5	0,36845
a_acc	0,36813
PEOE_RPC-	0,36564
b_rotR	0,36263
b_single	0,35852
PEOE_VSA+2	0,35614
Kier3	0,3508
SMR_VSA4	0,3496
vsa_don	0,348
Q_VSA_NEG	0,34463
a_nH	0,33848
PEOE_VSA+1	0,33702
PEOE_VSA_FPPOS	0,33088
PEOE_VSA+3	0,32907
PEOE_VSA_FHYD	0,32456
PEOE_VSA_FPOL	0,32456
a_nN	0,32174
vsa_acc	0,31994
SlogP_VSA9	0,31825
SMR_VSA1	0,31316
PEOE_VSA_NEG	0,31186

MOE-Deskriptor	KS-Wert
weinerPol	0,31074
SlogP_VSA3	0,31011
Q_RPC-	0,30887
RPC-	0,30887
Kier2	0,30801
vsa_base	0,3051
SlogP_VSA2	0,29776
SMR_VSA6	0,29485
a_count	0,28785
PEOE_VSA_PNEG	0,27366
a_IC	0,27212
PEOE_VSA+6	0,26986
vdw_area	0,2604
Q_VSA_PNEG	0,25977
KierA1	0,25845
b_count	0,25507
bpol	0,25506
balabanJ	0,24545
Kier1	0,24509
SMR_VSA7	0,23934
vsa_other	0,23702
chi1_C	0,22862
VDistEq	0,22475
a_ICM	0,21728
PEOE_VSA-0	0,21627
SlogP_VSA8	0,20735
apol	0,20697
VDistMa	0,20391
chi0	0,20345
chi1	0,20339
VAdjEq	0,20058
SlogP_VSA5	0,19861
zagreb	0,19771
chi0v	0,19383
diameter	0,19302
Weight	0,1913
PEOE_VSA-3	0,18815
vdw_vol	0,18768
a_heavy	0,18555

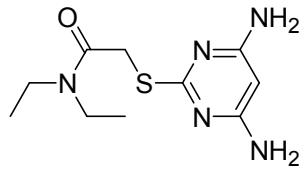
MOE-Deskriptor	KS-Wert
weinerPath	0,18226
PEOE_VSA+0	0,17973
PEOE_VSA_FPNEG	0,17885
b_heavy	0,17087
VAdjMa	0,17087
vsa_hyd	0,16739
chi1v	0,16231
PEOE_VSA_HYD	0,15797
radius	0,15773
a_nF	0,1472
PEOE_VSA-5	0,14619
SMR_VSA0	0,14259
mr	0,13988
a_nCl	0,13705
SMR	0,13522
b_double	0,13233
SlogP_VSA6	0,12957
a_nS	0,12353
chi0_C	0,1197
PEOE_VSA+5	0,11827
chi1v_C	0,11541
density	0,10932
chi0v_C	0,10564
a_hyd	0,098454
a_nCl	0,097305
petitjean	0,095007
petitjeanSC	0,094841
Q_VSA_FPNEG	0,091299
a_acid	0,085787
vsa_acid	0,085787
PEOE_VSA-2	0,057283
b_triple	0,045054
a_nBr	0,033576
a_nP	0,018465
a_nI	0,0082699
a_nB	0,0071121

7.2 Die Strukturen der Substanzen der M-Reihe

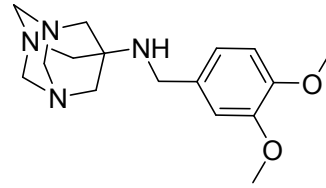




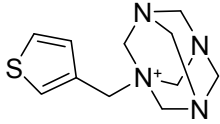
M16



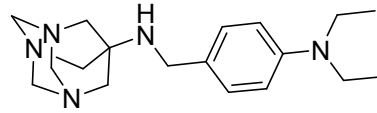
M17



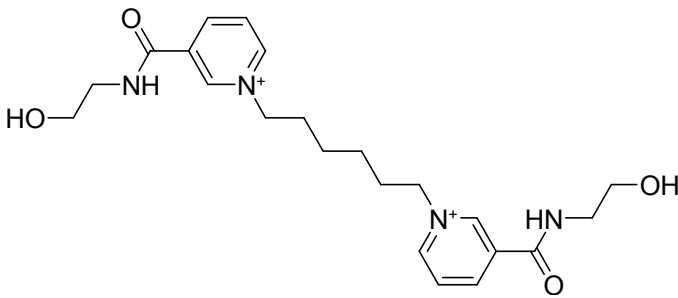
M18



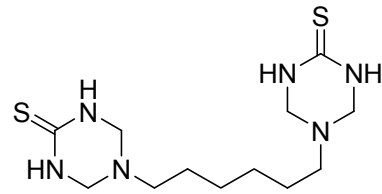
M19



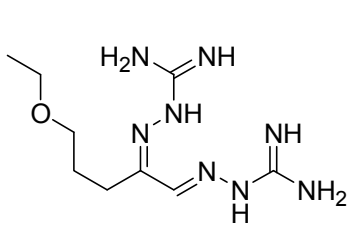
M20



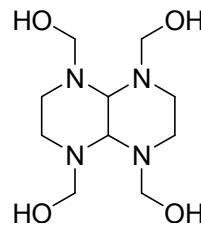
M21



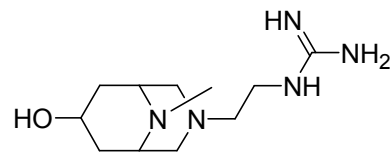
M22



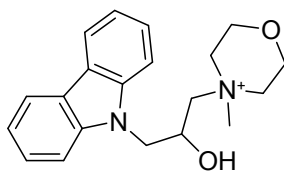
M23



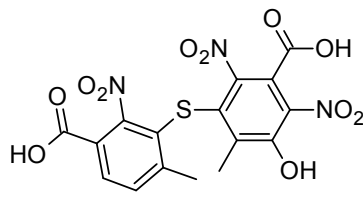
M24



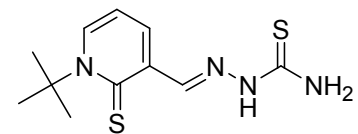
M25



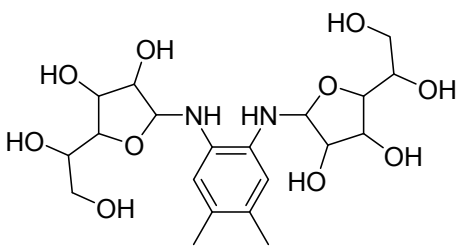
M26



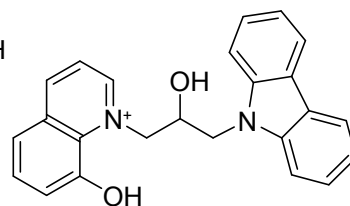
M27



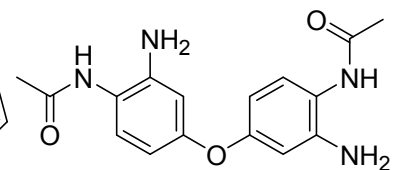
M28



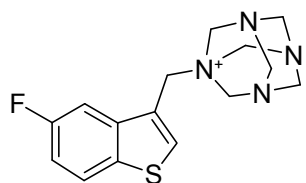
M29



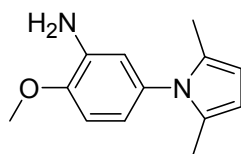
M30



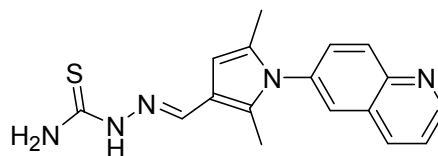
M31



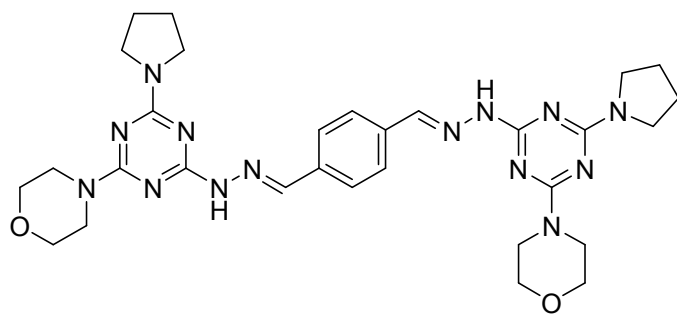
M32



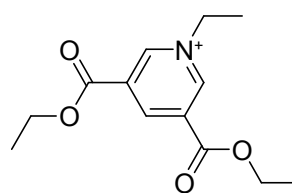
M33



M34

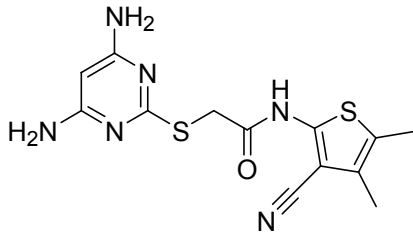


M35

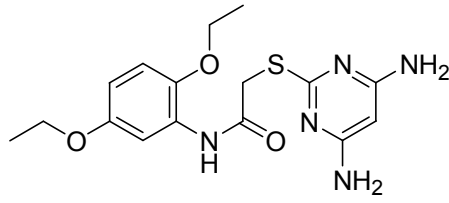


M36

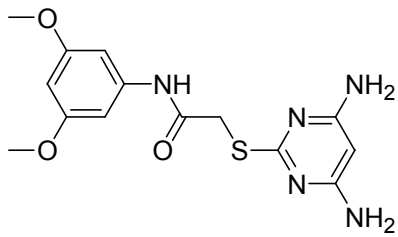
7.3 Die Strukturen der Substanzen der AS-Reihe



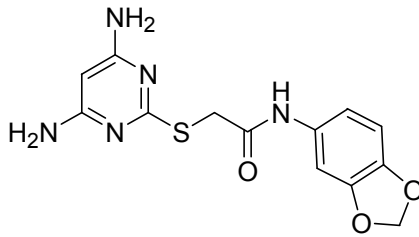
AS1



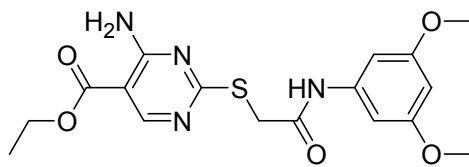
AS2



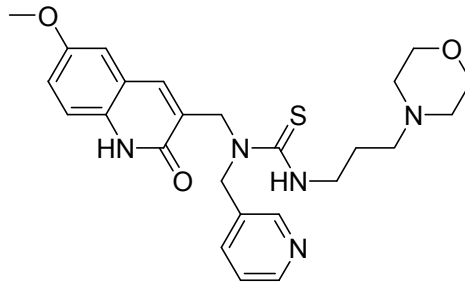
AS3



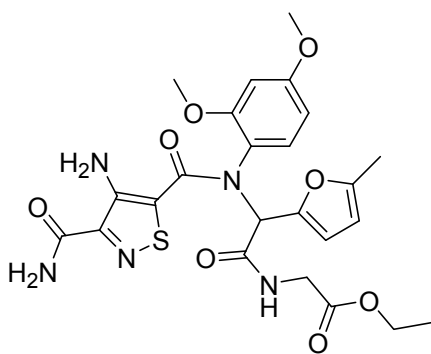
AS4



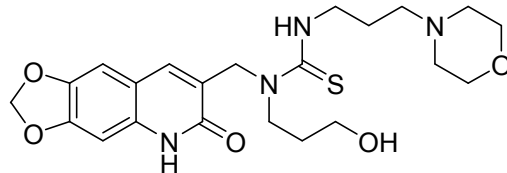
AS5



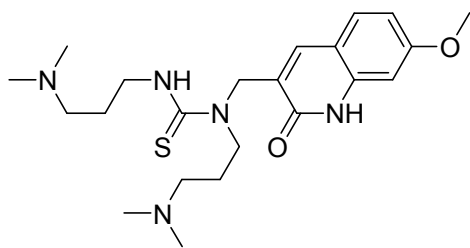
AS6



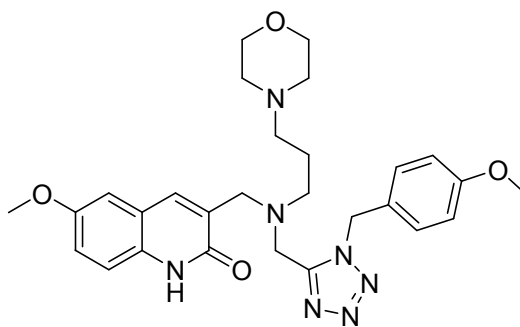
AS7



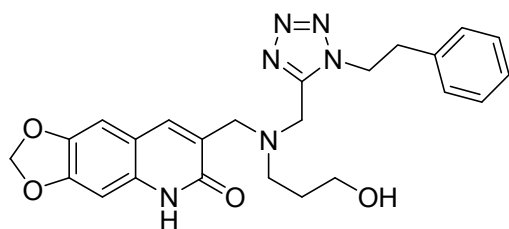
AS8



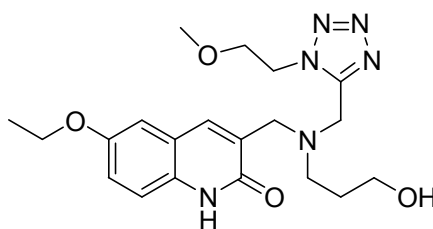
AS9



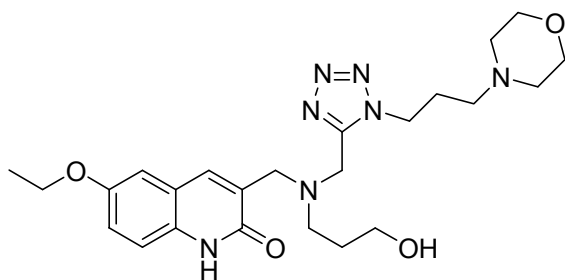
AS10



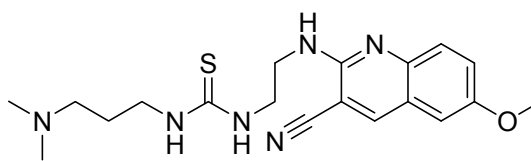
AS11



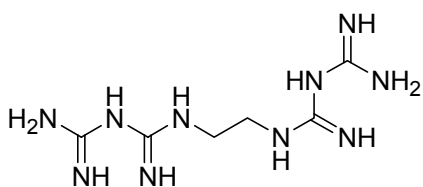
AS12



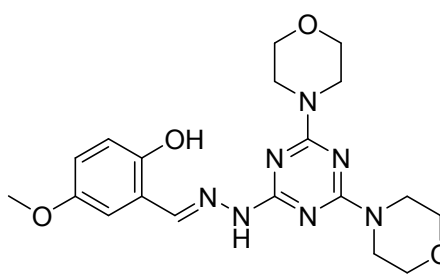
AS13



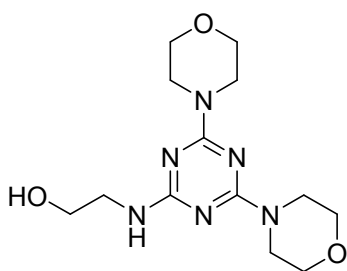
AS14



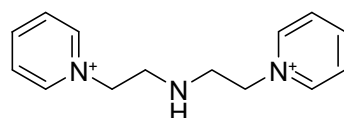
AS15



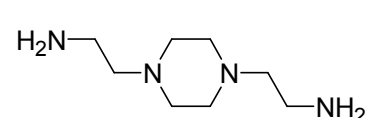
AS16



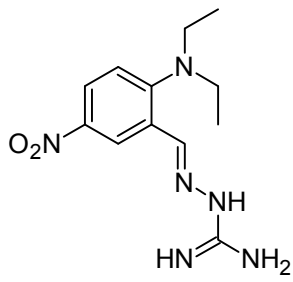
AS17



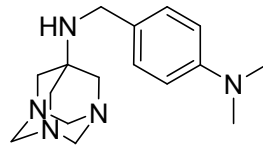
AS18



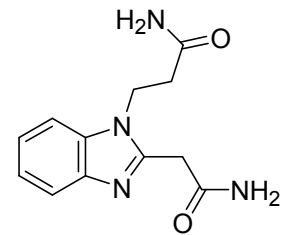
AS19



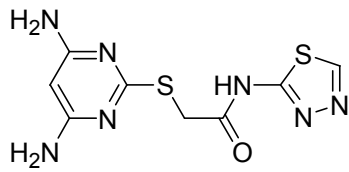
AS20



AS21

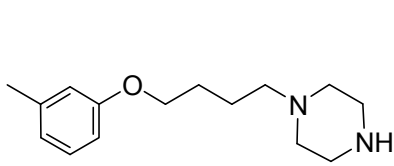


AS22

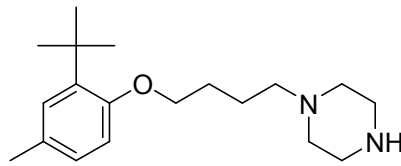


AS23

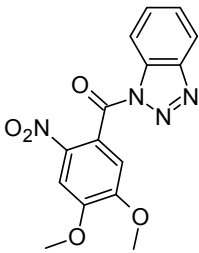
7.4 Die Strukturen der Substanzen der LR-Reihe



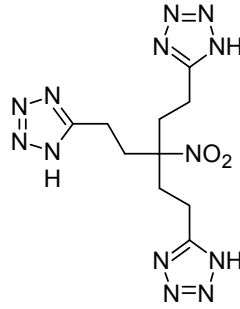
L1



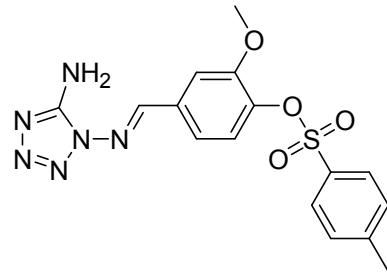
L2



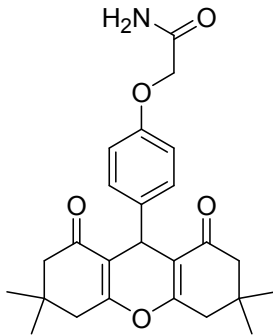
L3



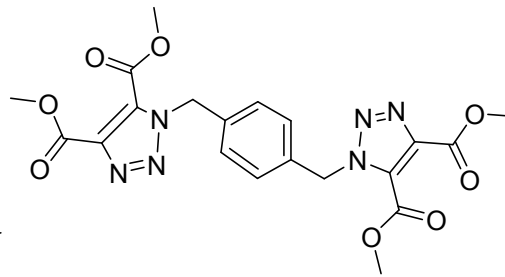
L4



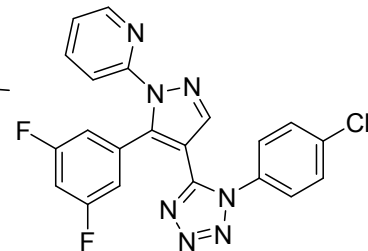
L5



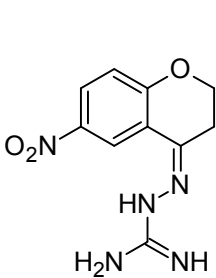
L6



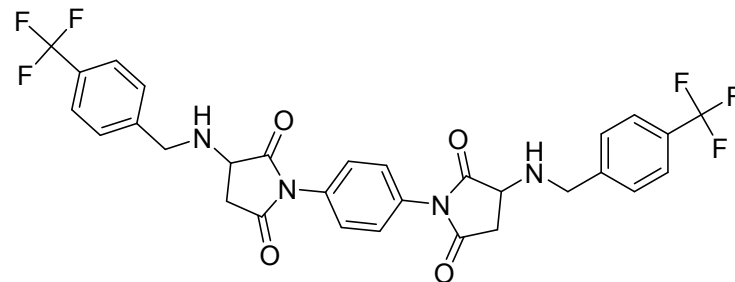
L7



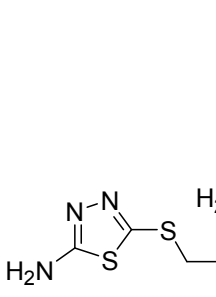
L8



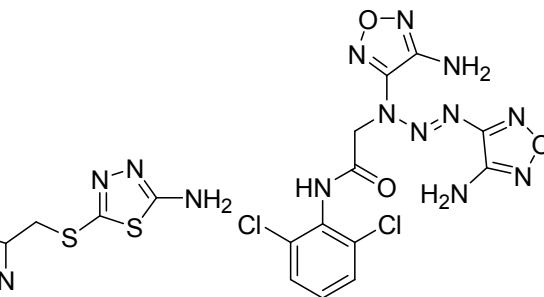
L9



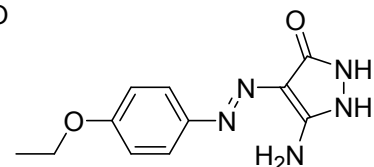
L10



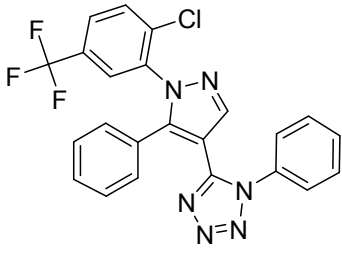
L11



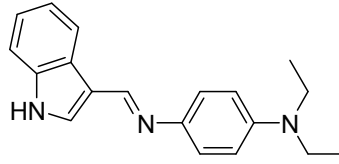
L12



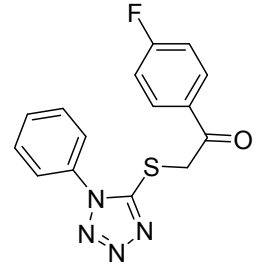
L13



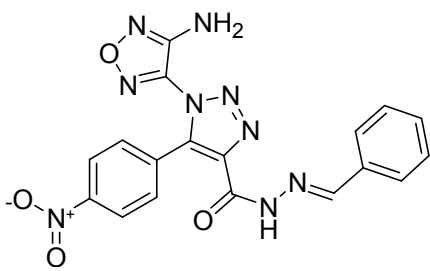
L14



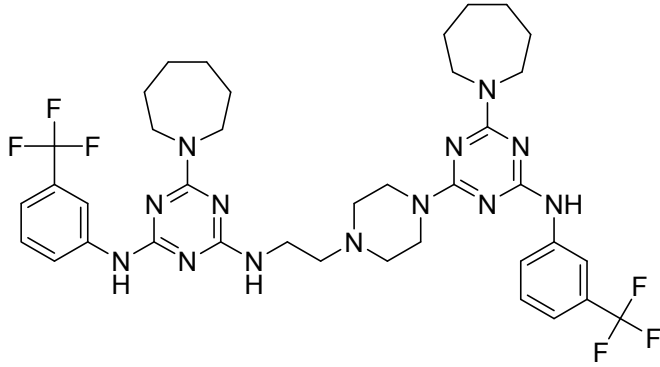
L15



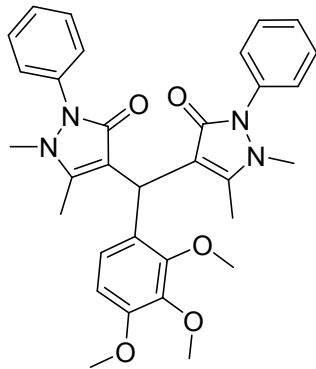
L16



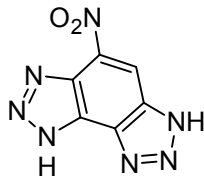
L17



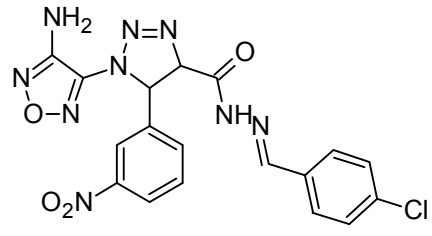
L18



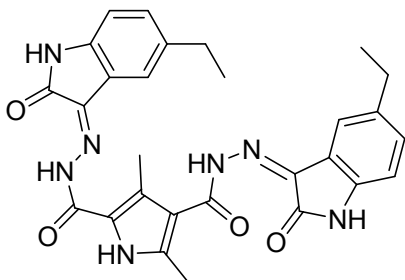
L19



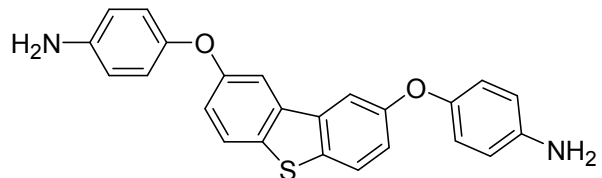
L20



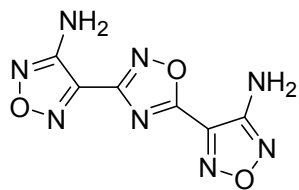
L21



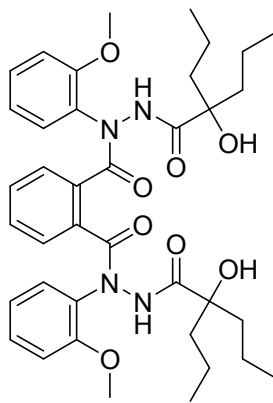
L22



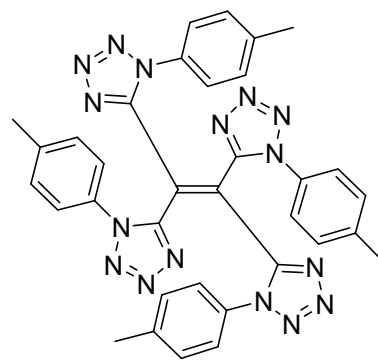
L23



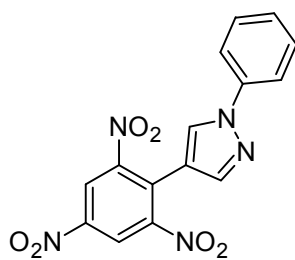
L24



L25



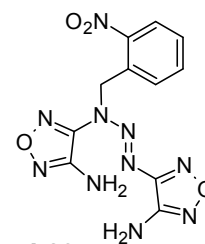
L26



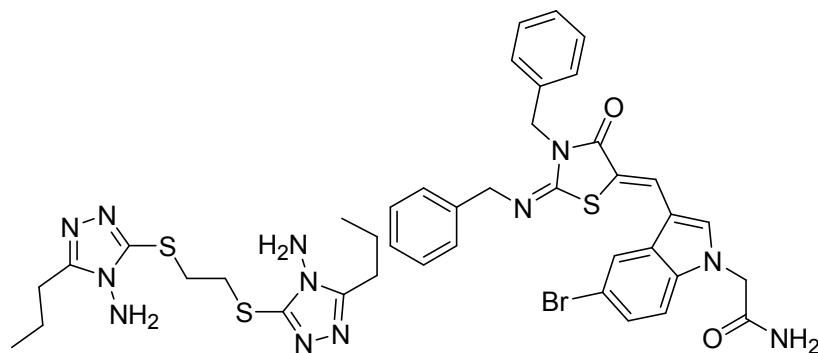
L27



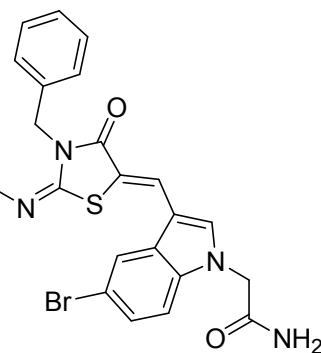
L28



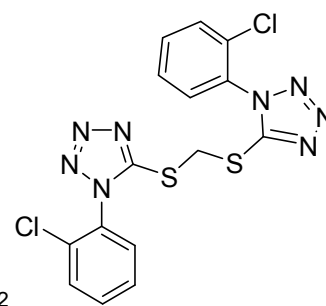
L29



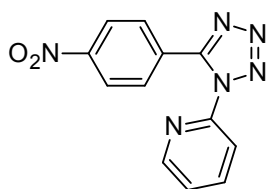
L30



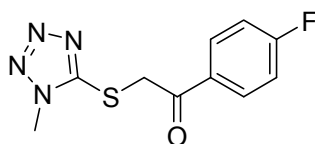
L31



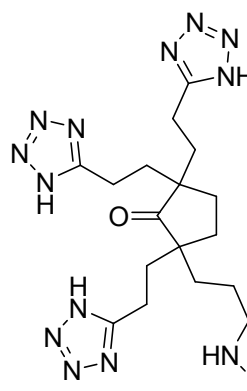
L32



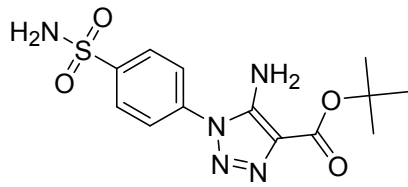
L33



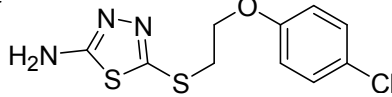
L34



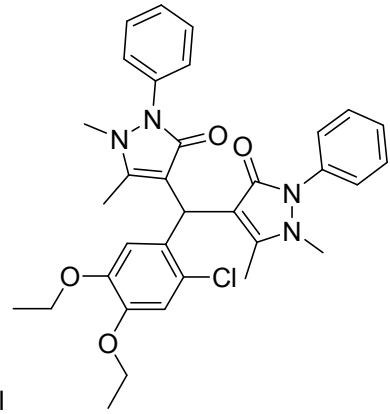
L35



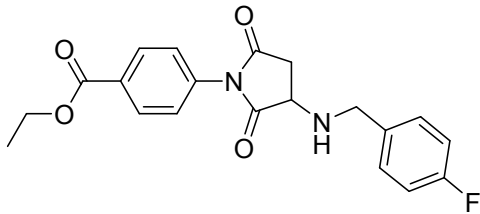
L36



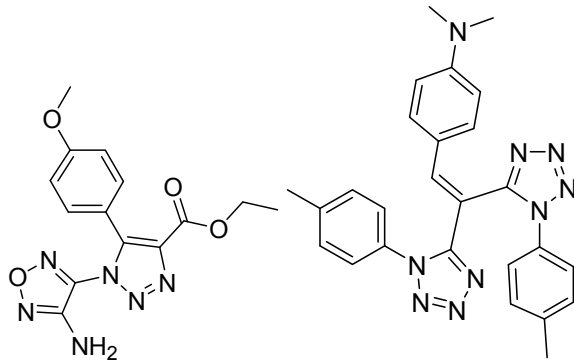
L37



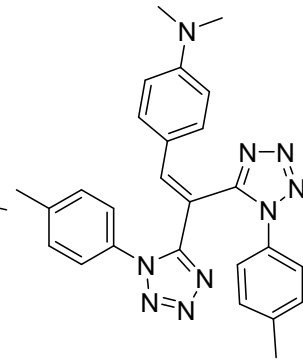
L38



L39

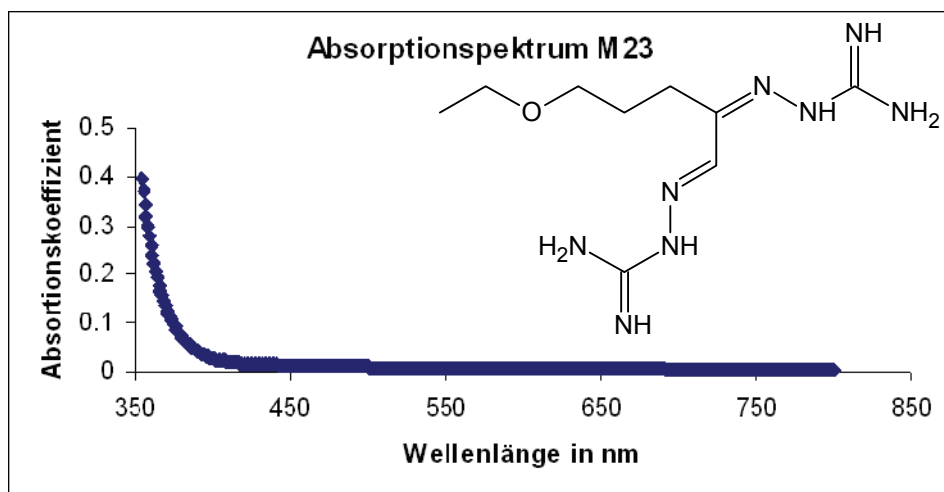
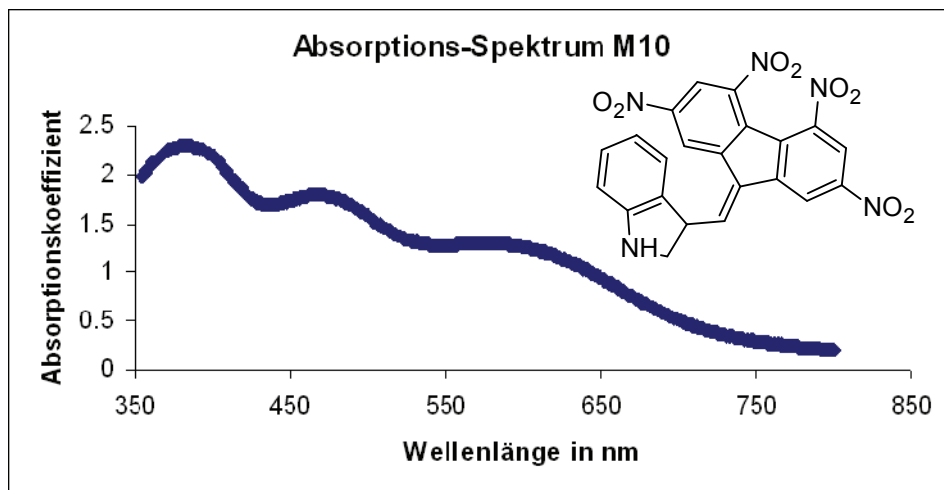
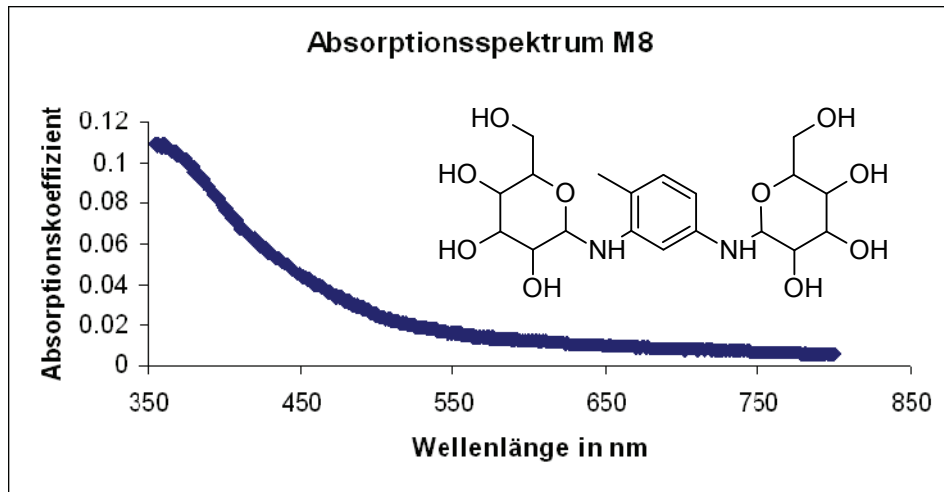


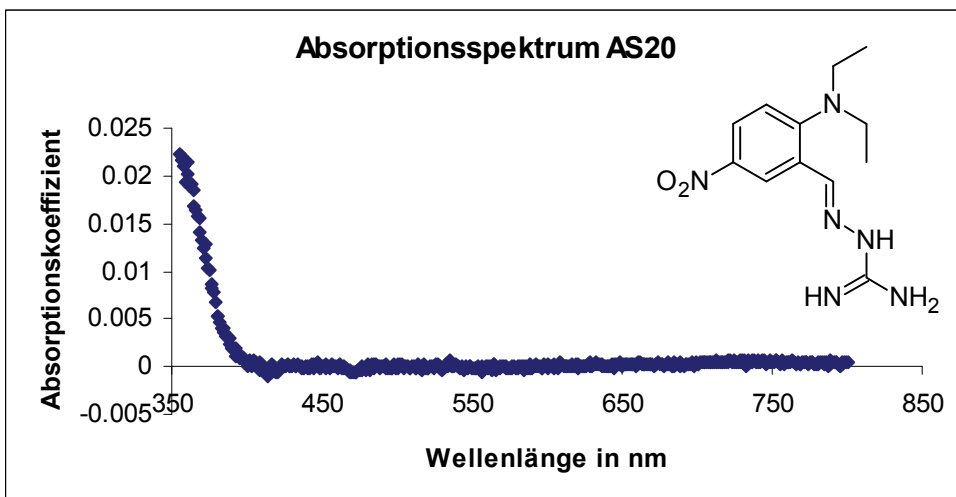
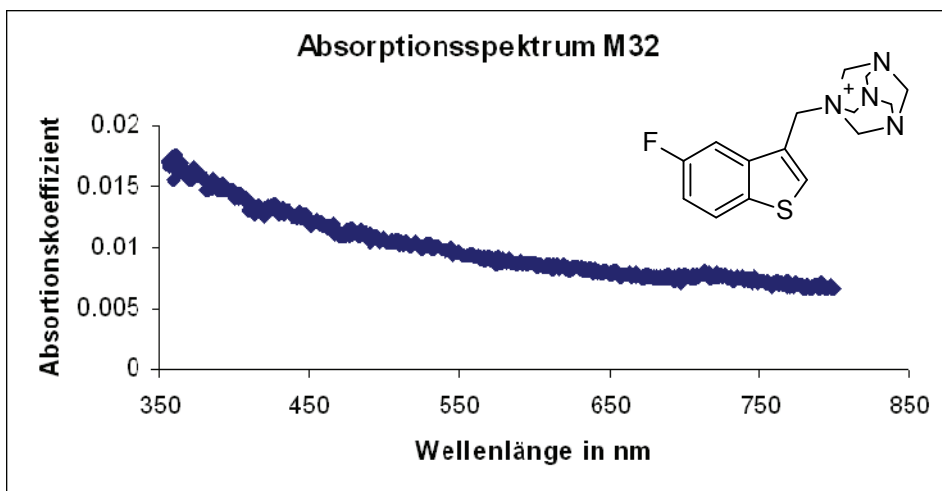
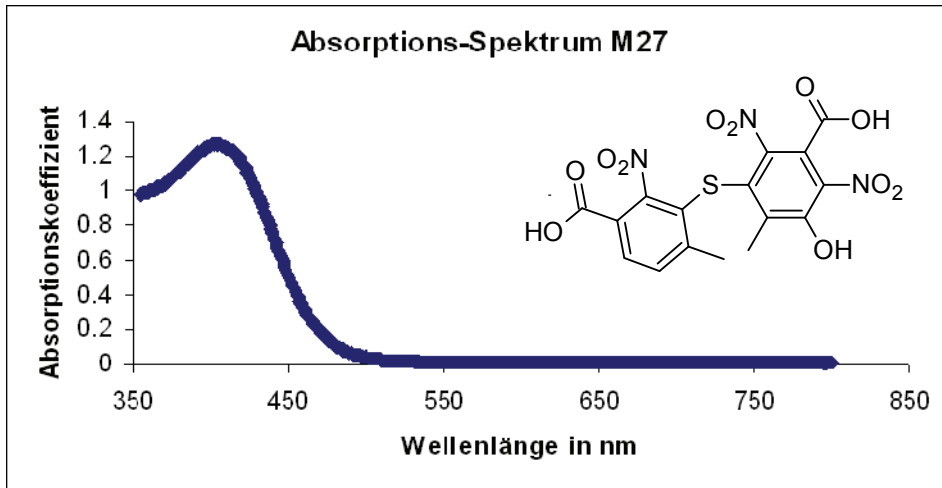
L40

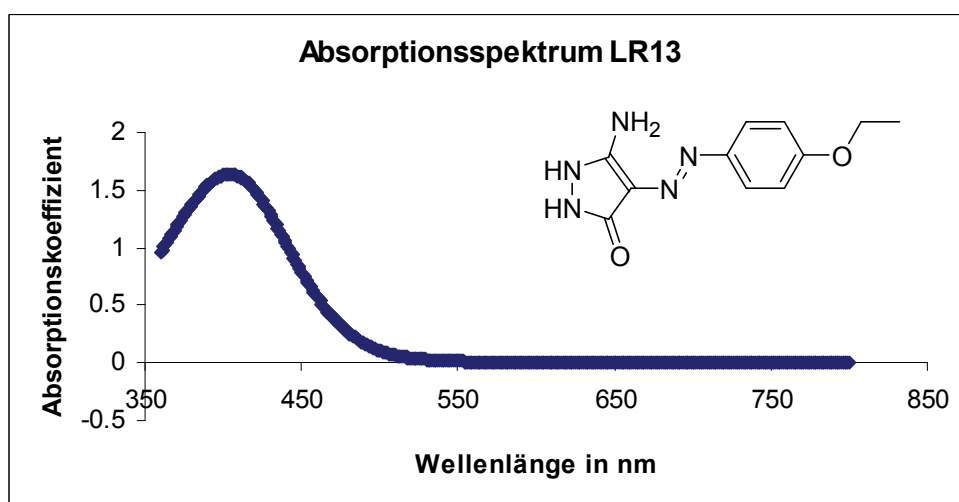
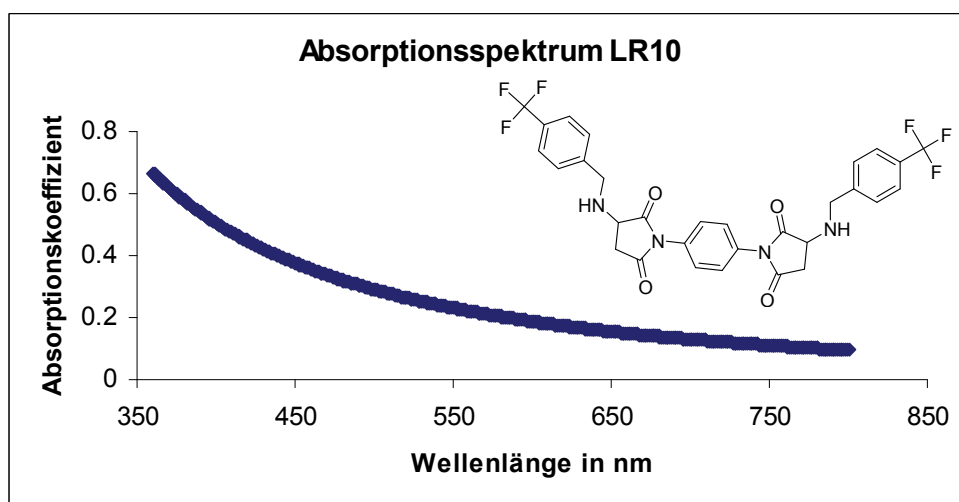
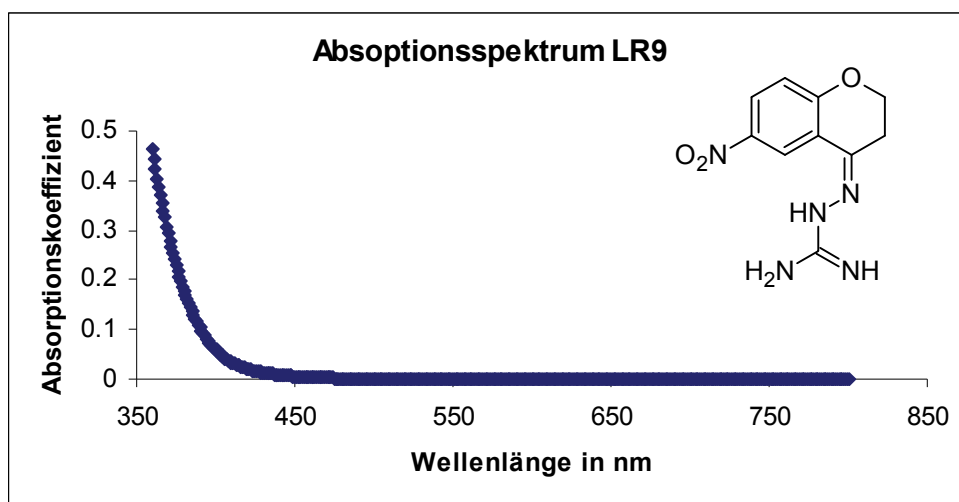


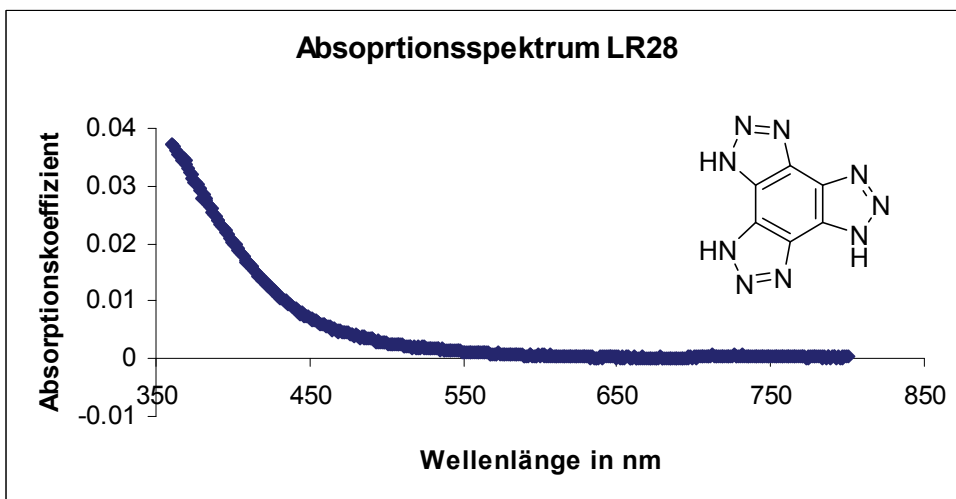
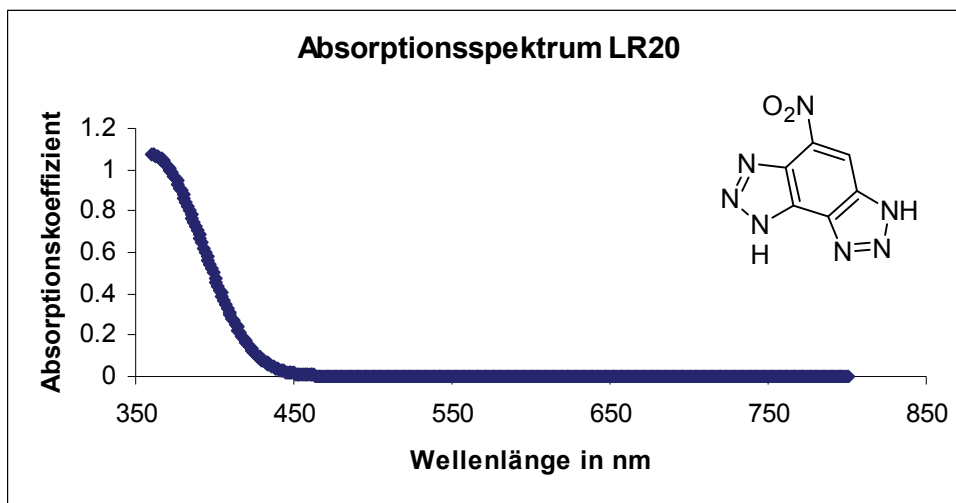
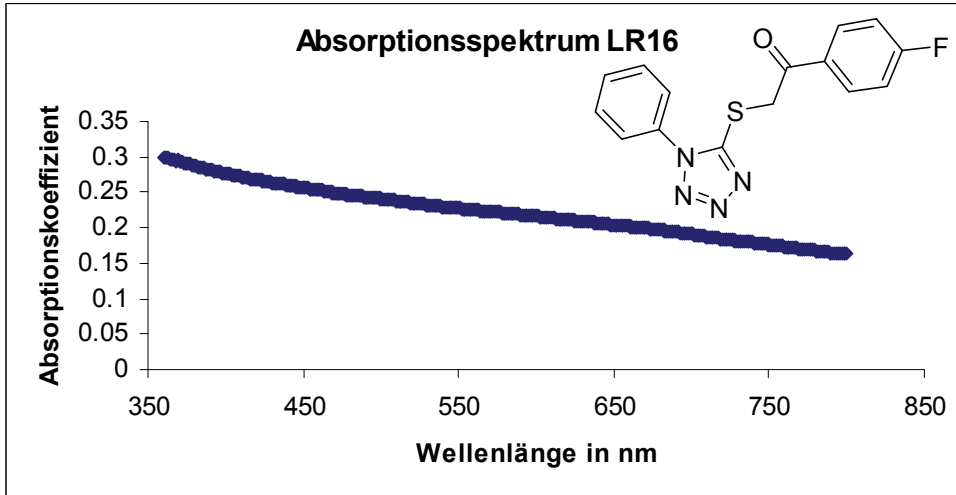
L41

7.5 Absorptionsspektren für ausgewählte Substanzen der TAT-TAR-FRET-Assay-Messreihen









7.6 CFivTT-Inhibitionsassay

Im Rahmen der Kooperation mit Markus Weidlich konnten von ihm die Testsubstanzen der Messreihen (M-Reihe und AS-Reihe, sowie zum jetzigen Zeitpunkt erste Substanzen der LR-Reihe) – die ursprünglich nur für den Tat-TAR-FRET-Assay bestimmt waren – auch im CFivTT-Inhibitionsassay auf potentielle Wirkung getestet werden.

Eine ausführlichere Darstellung und Analyse der Ergebnisse kann in Markus Weidlichs Doktorarbeit „Biochemische und Proteomanalytische Identifizierung und Charakterisierung neuer antimikrobieller RNA-Liganden“ gefunden werden. Die Methodenbeschreibung erfolgte durch Markus Weidlich bereits in „Differential activity profiles of translation inhibitors in whole-cell and cell-free approaches“ in (Weidlich 2007).

Die Angabe der folgenden Assay-Parameter dient nur der Vollständigkeit dieser Arbeit, um den Ursprung der im Rahmen dieser Arbeit verwendeten Messdaten (RFI-Werte) aufzuzeigen.

CFivTT-Inhibitionsassay: „M“-Proben:

Setup: gekoppelte zellfreie in-vitro transcription/translation im Batch-Verfahren

basierend auf S30 Extrakt aus *E. Coli* A19

Reporterprotein: GFP (Plasmid: pIVEX2.3-GFP; T7-promoter)

Assayvolumen: 25 µl; Probenkonzentration: 100 µM

Reaktionszeit: 5 hrs.@30°C; Wasserbad, 550 rpm (Rüttler)

Lösungsmittel: 2% (v/v) DMSO → Kontrollreaktion = 2% (v/v) DMSO [ohne Sample]

Messung der Fluoreszenz einer 1/100 Verdünnung des Reaktionsansatzes

Anregungswellenlänge: 395 nm; Emissionswellenlänge: 509 nm

(Hitachi F-4500 Fluoreszenzspektrophotometer)

Messwert als relativen Messwert in % auf TT der Kontrollreaktion

Jede Probe wurde mindestens 3mal unabhängig voneinander gemessen

7.6.1 Ergebnisse des CFivTT-Assay für die M-Reihe

Tabelle 31: Messergebnisse des CFivTT-Assay für die Substanzen der M-Reihe

LaborID	Mittlere relative Transkription/Translation [%]	mittlere relative Inhibition [%]	Standardabweichung [%]
Kontrolle	100	0	0
M1	115,27	0	6,64
M2	127	0	8,73
M3	111,75	0	6,71
M4	119,82	0	8,58
M5	115,24	0	10,47
M6	142,32	0	5,11
M8	131,97	0	6,35
M9	89,99	10,01	9,91
M10	96,66	3,34	9,46
M11	131,95	0	9,33
M12	104,38	0	8,91
M13	116,39	0	3,03
M14	103	0	5,49
M16	119,82	0	3
M17	125,8	0	3,35
M18	106,6	0	9,87
M19	75,5	24,5	5,02
M20	121,51	0	9,06
M21	110,9	0	6,56
M22	116,51	0	5,01
M23	106,81	0	7,43
M24	109,38	0	4,35
M25	119,13	0	5,69
M26	92,96	7,04	5,74
M27	93,35	6,65	6,73
M28	104,36	0	3,9
M29	106,03	0	3,46
M30	110,16	0	2,74
M31	105,26	0	6,9
M32	59,6	40,4	3,32
M33	116,62	0	9,61
M34	107,39	0	8,11
M35	117,36	0	7,93

7.7 Ganzzellassay – Wachstumsinhibition durch Substanzen der M-Reihe

Von Markus Weidlich wurden mit den Substanzen der M-Reihe - nach den Ergebnissen der CFivTT-Assay-Versuche - auch auf eine eventuelle Wirkung gegen eine Auswahl an pathogenen Keimen getestet. Dies erfolgte mittels Standard-Mikrodilutionsmethode. Eine ausführlichere Darstellung und Analyse der Ergebnisse kann in Markus Weidlichs Doktorarbeit „Biochemische und Proteomanalytische Identifizierung und Charakterisierung neuer antimikrobieller RNA-Liganden“ gefunden werden, die sich zurzeit in Vorbereitung befindet.

Eine Zusammenfassung der Ergebnisse der Versuche gegen die gewählten Keime für M32 kann in Tabelle 32 gefunden werden.

Ergebnisse/Schlussfolgerungen:

Tabelle 32: Aktivitäten von M32 gegen pathogene Keime; als Referenz dient Chloramphenicol

Testsubstanz	Molmasse	Lösungsmittel	<i>Staphylococcus aureus</i> (ATCC 29213)	<i>Escherichia coli</i> (ATCC 25922)	<i>Enterococcus faecalis</i> (ATCC 29212)
			Minimale Hemmkonzentration - MIC ($\mu\text{g/ml}$) / (μM)		
Chloramphenicol	323,13	DMSO	8 / 25	4 / 12,5	4 / 12,5
M32	385,29	DMSO	128 / 332	128 / 332	128 / 332

Es wurde Aktivität von M7 im Bereich 2 – 250 μM entdeckt (Substanz konnte aufgrund der Eigenfluoreszenz nicht im CFivTT-Assay gemessen werden). Dies stellt eine weitere Substanz dar, die aufgrund der angewendeten ANN-Filter gefunden wurde.

Ferner konnte Aktivität bei M32 und TAZA = MW1 beobachtet werden. Jedoch konnte M32 nur eine schwache Aktivität gegen Gram-negative als auch –positive Testkeime aufweisen; siehe Tabelle 32. Hier wird vermutet, dass aufgrund der fixierten positiven Ladung des Moleküls eine schlechte Zellmembranpermeation die Grundlage dieses schlechteren Abschneidens ist - als dies im Vergleich zum zellfreien und somit „Zellmembran“-freien CFivTT der Fall war.

7.8 DIGE-Proteomics-Ansatz

Von Markus Weidlich wurde - im Rahmen seiner Arbeiten zur Charakterisierung der Wirkungsweise von M32 auf pathogene Keime - ein DIGE-Proteomics-Ansatz durchgeführt. Eine ausführlichere Beschreibung und Analyse der Ergebnisse kann in Markus Weidlichs Doktorarbeit „Biochemische und Proteomanalytische Identifizierung und Charakterisierung neuer antimikrobieller RNA-Liganden“ gefunden werden. Hier beschrieben ist eine Kurzfassung des Ansatzes und seiner bisherigen Ergebnisse.

Methode:

DIGE-Proteomics-Ansatz

1. Bestimmung der IC_{50} /MHK in 5 ml Flüssigkulturen gegen *Bacillus subtilis* 168 WT
2. Inkubation der Kultur mit „M32“ bei IC_{50} (10 biologische Replikate);
Wachstumsbedingungen: 37°C, Inkubation ab $OD_{500} = 0,9 - 1,0$ (log-Phase), 30 Minuten
3. Zellaufschluss / 2 D-Gelelektrophorese nach Labeln mit DIGE-Farbstoffen
4. relative Quantifizierung gegen internen Standard nach Fluoreszenzscannen
5. Auswertung per Software inkl. Student t-Test / +/- 1.5 Up-/Downregulierung vom internen Standard als differentiell reguliert angesehen
6. Tryptischer Verdau und Identifikation per MALDI-MS bzw. MS/MS der relevanten Proteine.

Ergebnis:

Es wurden Markerproteine gefunden, die auf Interaktion sowohl mit der Zellwand als auch der Zellmembran hindeuten. Weiterhin sind Up-regulierte Markerproteine identifiziert worden,

die bereits bei anderen Proteomics-basierten antimikrobiellen Wirkstoffstudien aufgetreten sind (vor allem Stress-Proteine).

7.9 Liste der Publikationen die für den „Allgemeinen RNA-Liganden-Filter“ herangezogen wurden

Die folgende Liste enthält die Publikationen aus denen entweder direkt Strukturen entnommen wurden, oder zu denen PDB-Einträge existieren die für die Erstellung des „Allgemeinen RNA-Binder-Filters“ herangezogen wurden.

Aboul-ela F, Karn J, Varani G (1995): The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* 253(2): 313-332.

Aboul-ela F, Karn J, Varani G (1996): Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* 24(20): 3974-3981.

Bashan A, Agmon I, Zarivach R, Schluenzen F, Harms J, Berisio R, Bartels H, Franceschi F, Auerbach T, Hansen HA, Kossoy E, Kessler M, Yonath A (2003): Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol. Cell.* 11 (1): 91-102.

Berisio R, Harms J, Schluenzen F, Zarivach R, Hansen HA, Fucini P, Yonath A (2003): Structural insight into the antibiotic action of telithromycin against resistant mutants. *J. Bacteriol.* 185 (14): 4276-4279.

Berisio R, Schluenzen F, Harms J, Bashan A, Auerbach T, Baram D, Yonath A (2003): Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat. Struct. Biol.* 10 (5): 366-3670.

Brodersen DE, Clemons WM Jr, Carter AP, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000): The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell.* 103 (7): 1143-1154.

Brodsky AS, Williamson JR (1997) Solution Structure of the HIV-2 TAR Argininamide Complex. *J. Mol. Biol.* 267(3):624-4.

Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000): Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407(6802): 340-348.

Chang KY, Tinoco I Jr. (1997): The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* 269 (1): 52-66.

Chen X, Ramakrishnan B, Sundaralingam M (1995): Crystal structures of B-form DNA-RNA chimers complexed with distamycin. *Nat. Struct. Biol.* 2 (9): 733-735.

Conn GL, Draper DE, Lattman EE, Gittis AG (1999): Crystal structure of a conserved ribosomal protein-RNA complex. *Science* 284 (5417): 1171-1174.

Du Z, Lind KE, James TL (2002): Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.* 9 (6): 707-712.

Faber C, Sticht H, Schweimer K, Rosch P (2000): Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* 275 (27): 20660-20666.

Fourmy D, Recht MI, Blanchard SC, Puglisi JD (1996): Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* 274 (5291): 1367-1371.

Hansen JL, Ippolito JA, Ban N, Nissen P, Moore PB, Steitz TA (2002): The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Mol. Cell.* 10 (1): 117-128.

Hansen JL, Moore PB, Steitz TA (2003): Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit. *J. Mol. Biol.* 330 (5): 1061-1075.

Hansen JL, Schmeing TM, Moore PB, Steitz TA (2002): Structural insights into peptide bond formation. *Proc Natl. Acad. Sci. U.S.A.* 99 (18): 11670-11675.

Helgstrand C, Grahn E, Moss T, Stonehouse NJ, Tars K, Stockley PG, Liljas L (2002): Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. *Nucleic Acids Res.* 30 (12): 2678-2685.

Jaeger J, Restle T, Steitz TA (1998): The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.* 17 (15): 4535-4542.

Jiang F, Gorin A, Hu W, Majumdar A, Baskerville S, Xu W, Ellington A, Patel DJ (1999): Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples. *Structure Fold Des.* 7(12):1461-1472.

Jiang L, Majumdar A, Hu W, Jaishree TJ, Xu W, Patel DJ (1999): Saccharide-RNA recognition in a complex formed between neomycin B and an RNA aptamer. *Structure Fold Des.* 7 (7): 817-827.

Jiang L, Patel DJ (1998): Solution structure of the tobramycin-RNA aptamer complex. *Nat. Struct. Biol.* 5 (9): 769-774.

Jiang L, Suri AK, Fiala R, Patel DJ (1997): Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. *Chem. Biol.* 4 (1): 35-50.

Johansson C, Finger LD, Trantirek L, Mueller TD, Kim S, Laird-Offringa IA, Feigon J (2004): Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *J. Mol. Biol.* 337 (4): 799-816.

Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Kramer A, Sattler M (2001): Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 294(5544): 1098-1102.

Lynch SR, Puglisi JD (2001): Structural origins of aminoglycoside specificity for prokaryotic ribosomes. *J. Mol. Biol.* 306(5): 1037-1058.

Mikkelsen NE, Johansson K, Virtanen A, Kirsebom LA (2001): Aminoglycoside binding displaces a divalent metal ion in a tRNA-neomycin B complex. *Nat Struct Biol.* 8 (6): 510-514.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000a): The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* 289, 920.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000b): Large Ribosomal Subunit Complexed With A 13 Bp Minihelix- Puromycin Compound. *Science* 289, 920-930.

Ogle JM, Brodersen DE, Clemons WM Jr, Tarry MJ, Carter AP, Ramakrishnan V (2001): Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292 (5518): 897-902.

Ogle JM, Murphy FV, Tarry MJ, Ramakrishnan V (2002): Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* 111 (5): 721-732.

Park CJ, Bae SH, Lee MK, Varani G, Choi BS (2003): Solution structure of the influenza A virus cRNA promoter: implications for differential recognition of viral promoter structures by RNA-dependent RNA polymerase. *Nucleic Acids Res.* 31 (11): 2824-2832.

Pioletti M, Schlunzen F, Harms J, Zarivach R, Gluhmann M, Avila H, Bashan A, Bartels H, Auerbach T, Jacobi C, Hartsch T, Yonath A, Franceschi F (2001): Crystal structures of

complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.* 20 (8): 1829-1839.

Puglisi JD, Chen L, Blanchard S, Frankel AD (1995): Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science* 270 (5239): 1200-1203

Rosendal KR, Wild K, Montoya G, Sinning I (2003): Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc. Natl. Acad. Sci. U.S.A.* 100 (25):14701-14706.

Schluenzen F, Harms J, Franceschi F, Hansen HAS, Bartels H, Zarivach R, Yonath A (2003): Structural Basis for the Antibiotic Activity of Ketolides and Azalides. *Structure* 11: 329

Schluenzen F, Zarivach R, Harms J, Bashan A, Tocilj A, Albrecht R, Yonath A, Franceschi F (2001): Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* 413 (6858):814-21.

Schmeing TM, Seila AC, Hansen JL, Freeborn B, Soukup JK, Scaringe SA, Strobel SA, Moore PB, Steitz TA (2002): A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat. Struct. Biol.* 9 (3): 225-230.

Shi K, Pan B, Sundaralingam M (2003): Structure of a B-form DNA/RNA chimera (dC)(rG)d(ATCG) complexed with daunomycin at 1.5 Å resolution. *Acta Crystallogr D Biol Crystallogr.* 59 (Pt 8): 1377-1383.

Skordalakes E, Brogan A, Park BS, Kohn H, Berger JM (2005): Structural Mechanism of Inhibition of the Rho Transcription Termination Factor by the Antibiotic Bicyclomycin. *Structure* 13 (1): 99-109.

Steinbacher S, Kaiser J, Wungsintaweekul J, Hecht S, Eisenreich W, Gerhardt S, Bacher A, Rohdich F (2003): Structure of 2C-methyl-d-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. *J. Mol. Biol.* 316 (1): 79-88.

Tereshko V, Skripkin E, Patel DJ (2003): Encapsulating streptomycin within a small 40-mer RNA. *Chem. Biol.* 10 (2):175-187.

Valle M, Sengupta J, Swami NK, Grassucci RA, Burkhardt N, Nierhaus KH, Agrawal RK, Frank J (2002): Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J.* 21 (13):3557-3567.

Varani L, Spillantini MG, Goedert M, Varani G (2000): Structural basis for recognition of the RNA major groove in the tau exon 10 splicing regulatory element by aminoglycoside antibiotics. *Nucleic Acids Res.* 28(3):710-9.

Vicens Q, Westhof E (2001): Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure (Camb)* 8: 647-658.

Vicens Q, Westhof E (2002): Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding a site. *Chem. Biol.* 9 (6): 747-755.

Vicens Q, Westhof E (2003): Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide. *J. Mol. Biol.* 326 (4): 1175-1188.

Vila-Sanjurjo A, Ridgeway WK, Seymaner V, Zhang W, Santoso S, Yu K, Cate JH (2003): X-ray crystal structures of the WT and a hyper-accurate ribosome from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 100 (15): 8682-8687.

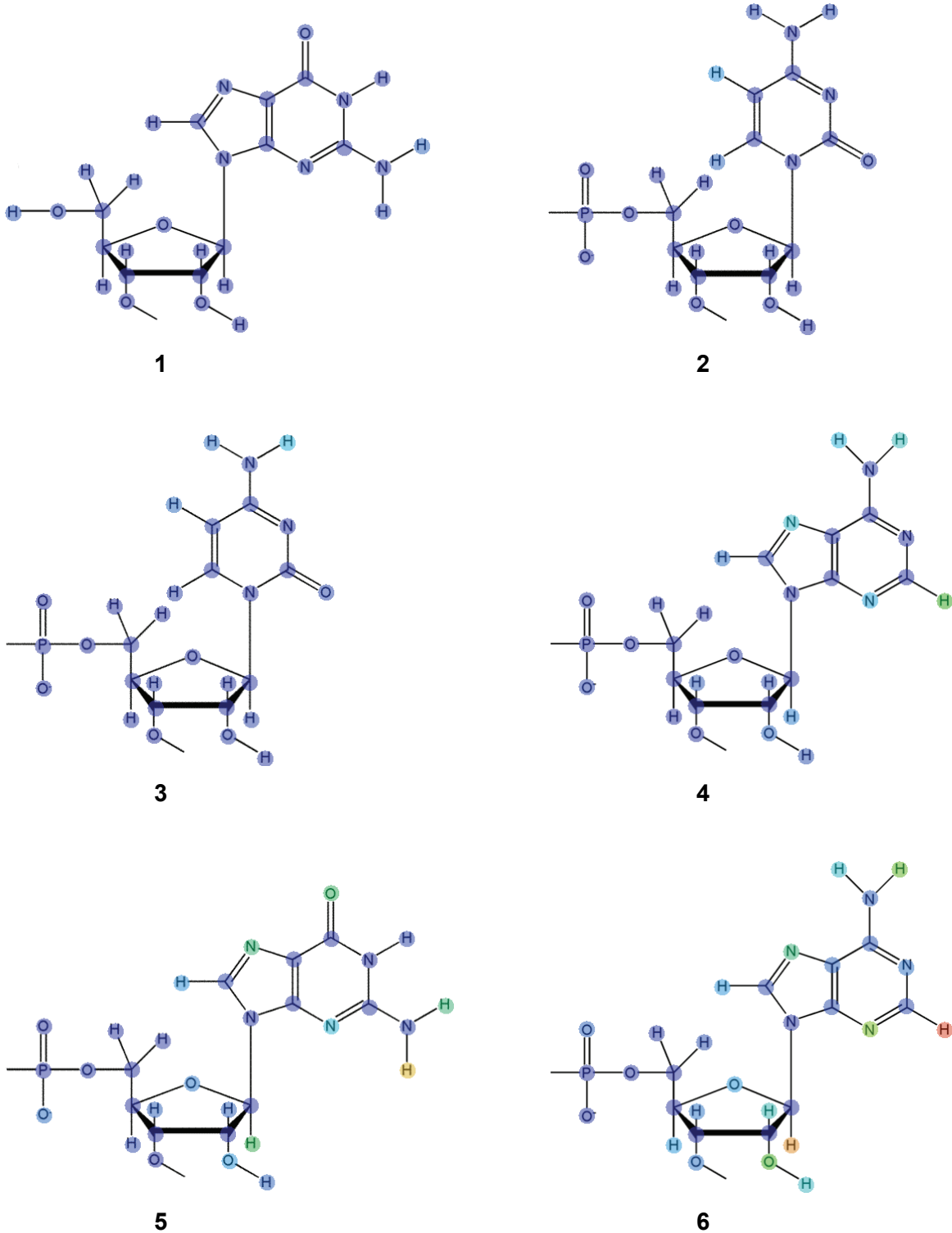
Yang Y, Declerck N, Manival X, Aymerich S, Kochoyan M (2002): Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT. *EMBO J.* 21 (8): 1987-1997.

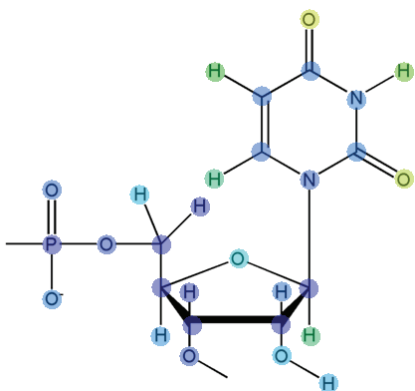
Ye X, Kumar RA, Patel DJ (1995): Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* 2(12): 827-840.

Yoshizawa S, Fourmy D, Puglisi JD (1998): Structural origins of gentamicin antibiotic action. *EMBO J.* 17(22):6437-6448.

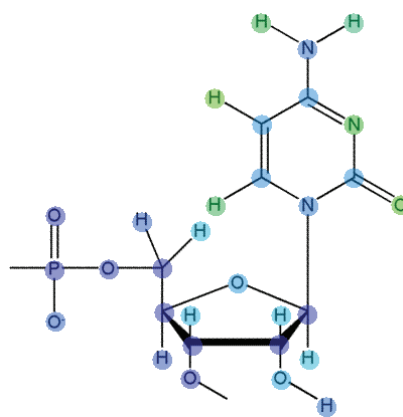
7.10 PSGP-Projektion auf Atome – Basensequenz von TAR

Die folgenden Projektionen der 29 Nukleotide der verwendeten TAR-RNA-Sequenz sind nach der Stärke des PSGP farblich kodiert worden. Blau = niedriges PSGP, Rot = hohes PSGP.

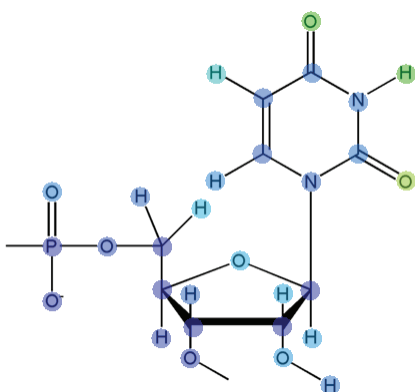




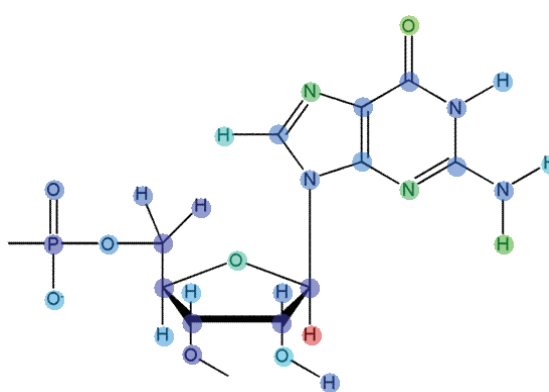
7



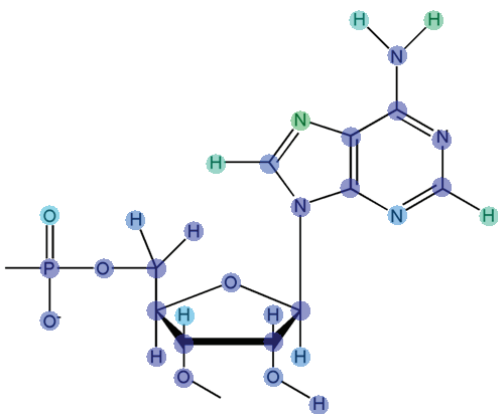
8



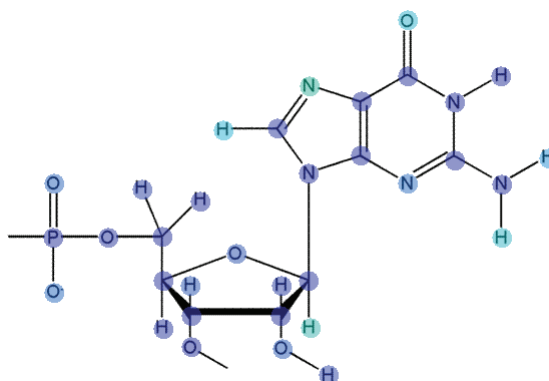
9



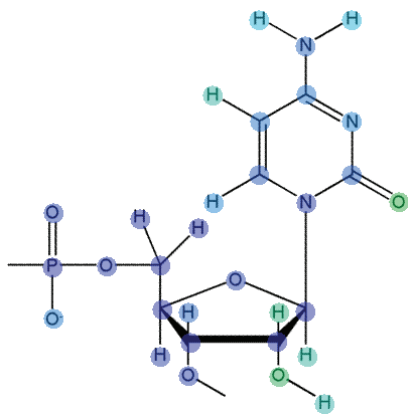
10



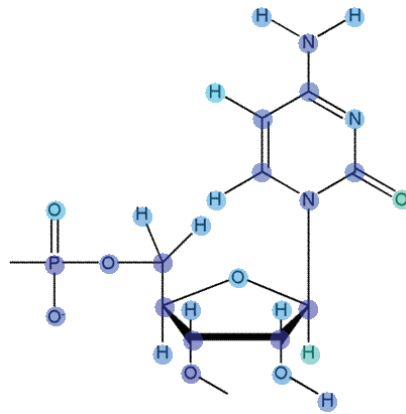
11



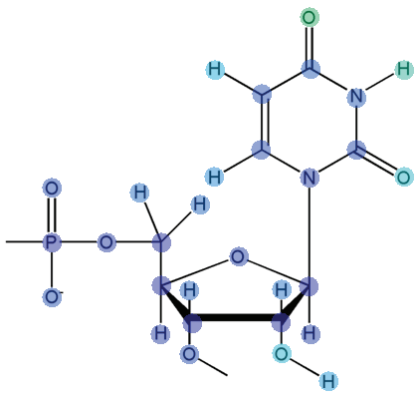
12



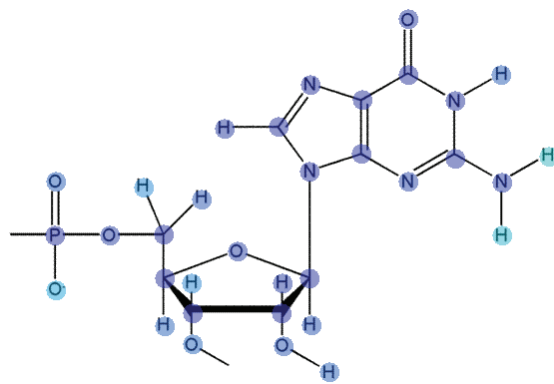
13



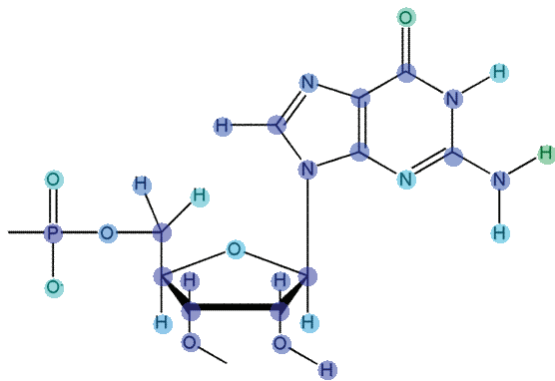
14



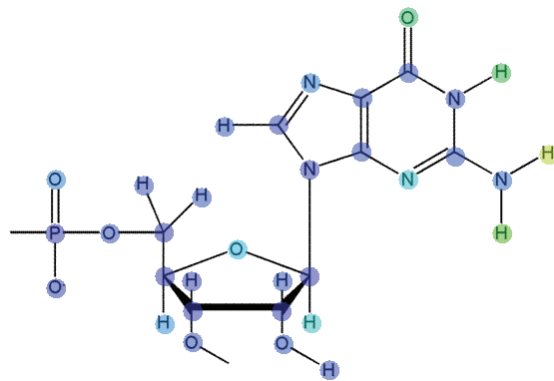
15



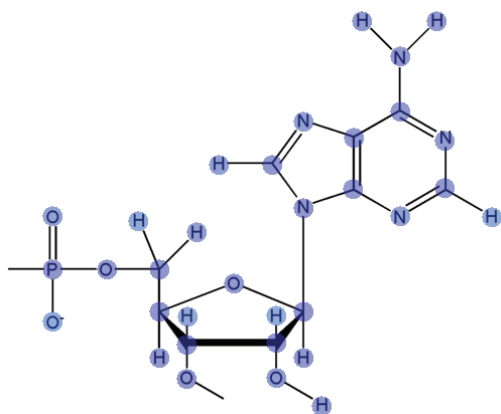
16



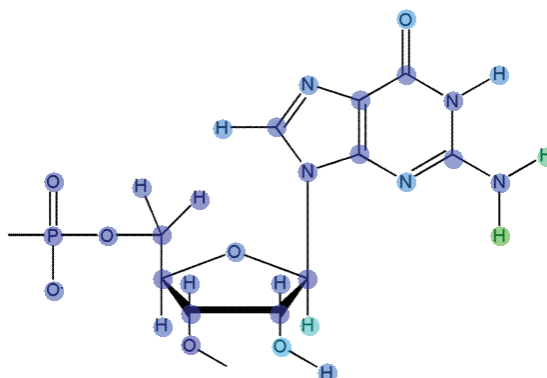
17



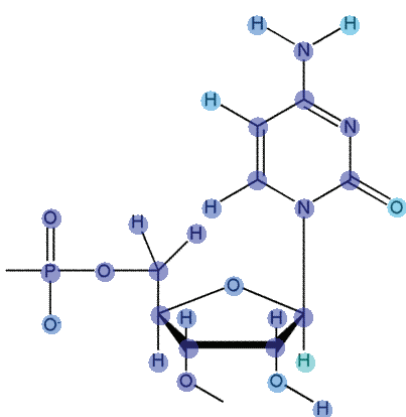
18



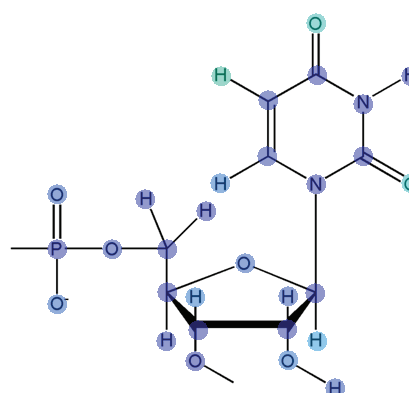
19



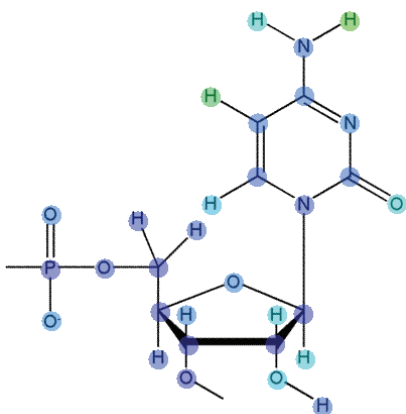
20



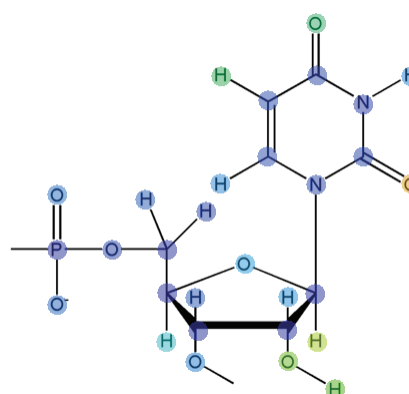
21



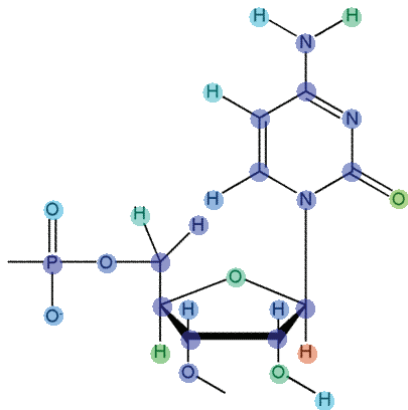
22



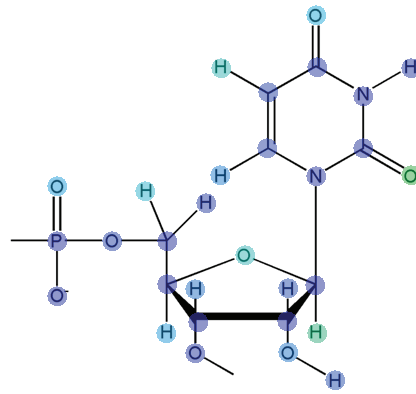
23



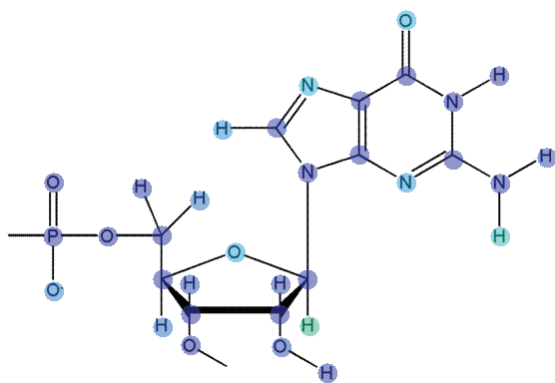
24



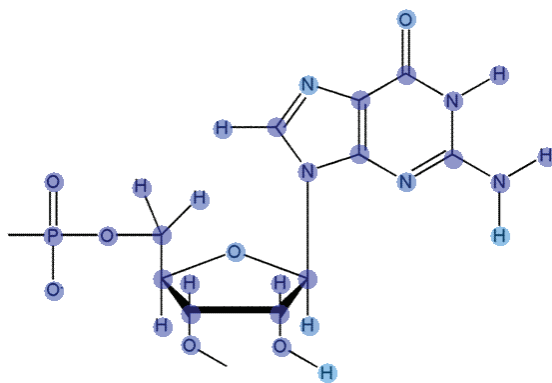
25



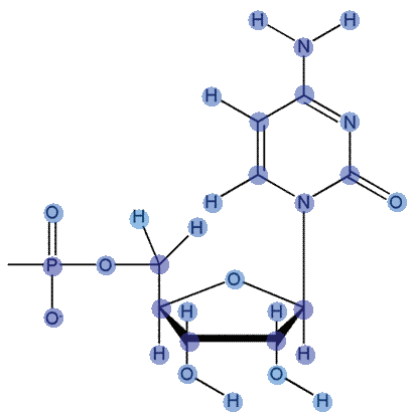
26



27



28



29

8. Abbildungsverzeichnis

sAbbildung 1:	Aktivierung der Transkriptionselongation durch Tat am LTR von HIV, nach Wöhl (2003, Seite 4); Verursacht durch die Bindung des Tat-Proteins an die TAR-RNA-Struktur kommt es zu einer deutlichen Steigerung der Transkriptionsrate längerer Viren-Transkripte und somit letztendlich zur gesteigerten Viren-Replikation. Dementsprechend ist diese Interaktion ein theoretisch guter Ansatzpunkt für die Wirkstoffentwicklung.	4
Abbildung 2:	Vereinfachte schematische Darstellung des Antiterminationskomplexes; nach Wöhl (2003, Seite 4)	5
Abbildung 3:	TAR-RNA, schematische Darstellung der Sekundärstruktur	6
Abbildung 4:	1ANR – TAR-Konformere; a) Konformer Nr. 18 b) Konformer Nr. 7 c) Konformer Nr. 6; die Beispielkonformere wurden anhand des RMSD ausgewählt. Konformer 7 und 18 sind am weitesten im Ensemble von 1ANR entfernt, Konformer 6 ist wiederum am weitesten entfernt von den beiden erstgenannten Konformeren.	6
Abbildung 5:	ATP als 2D-Darstellung	8
Abbildung 6:	ATP 3D-Modell; links befindet sich die Darstellung der zugänglichen Oberfläche des Moleküls, rechts das zugrunde liegende energieminierte Molekülmodell	8
Abbildung 7:	Beispiel eines Graphen einer KS-Statistik	12
Abbildung 8:	Fluoreszenzpaar-markiertes Tat-Peptid des verwendeten FRET-Assay	17
Abbildung 9:	Fotos des Arbeitsplatzes für die Durchführung des Tat-TAR-FRET-Assay; Links der Platz an dem die Pipettierung der Lösungen auf die Mikrotiterplatte erfolgte. Rechts das Spektrometer mit dem Steuerungscomputer. Die beiden Arbeitsplätze befinden sich direkt gegenüberliegend im Labor.	20
Abbildung 10:	Tat-TAR-FRET-Assay-Beispielmessungen der Referenzen Tat und Tat TAR. Alle Messungen wurden parallel in sechs unterschiedlichen Vertiefungen der Mikrotiterplatte durchgeführt und zeitlich versetzt immer wieder gemessen. Die Sägezacken-Struktur die besonders bei den drei Tat-TAR-Messreihen zu sehen ist, wird hier auf das Ausbleichen des Fluoreszeins durch die Messung, oder das jeweilige Schütteln der Platte vor der Messung zurückgeführt. Beachtenswert ist hierbei jedoch die relativ größere Abweichung der Messwerte zwischen unterschiedlichen Vertiefungen (Pipettiergenauigkeit).	21
Abbildung 11:	FRET-Assay – Titrationslauf mit Ausreißer-Messung: Wahrscheinlich beruht die türkis-farbene Messkurve auf einer starken Verunreinigung der Vertiefung der Mikrotiterplatte mit RNase und/oder „Fussel“. Ohne Überwachung der Equilibrierungsphase (erste sechs Schritte) wäre diese Stelle der Mikrotiterplatte ebenfalls für die Aktivitätsbestimmung herangezogen worden. Dies hätte dann zu einem falsch positiven Befund geführt. „Weird Nothing“ steht in diesem Fall für die Nicht-Hinzugabe von Testsubstanzlösung.	23
Abbildung 12:	Schema des RNA-Liganden-Informationsflusses im Projekt	29
Abbildung 13:	Zusammenführung der Strukturen der beiden Recherche-Quellen zum verwendeten RNA-Liganden-Datensatz	31
Abbildung 14:	Plot der KS-Distanzen für die MOE-2D-Deskriptoren zur Trennung der Protein-Liganden von den RNA-Liganden-Datensätzen. Zur besseren Abschätzung des Trennverhaltens sortiert nach der KS-Distanz.	33
Abbildung 15:	CATS-Deskriptor-SOM zur Bestimmung der Diversität innerhalb des RNA-Liganden-Datensatzes relativ zur Diversität des Protein-Liganden-Datensatzes. Die Einfärbung der Knoten gibt die relative Anzahl an Strukturen des jeweiligen Knotens in der Karte wieder.	36

Abbildung 16:	Trainings- und Auswahlschema der RNA-Liganden-GrobFilter.	38
Abbildung 17:	Histogramm der Scorewerteverteilung für den RNA-Binder-Filter basierend auf 146 MOE2D-Deskriptoren für den COBRA- und den RNA-Liganden-Datensatz	39
Abbildung 18:	Histogramm der Scorewerteverteilung für den RNA-Binder-Filter basierend auf 150 CATS-2D-Deskriptoren für den COBRA- und den RNA-Liganden-Datensatz	39
Abbildung 19:	Histogramm der Scorewerteverteilung für den RNA-Binder-Filter basierend auf 300 TC3D-Deskriptoren für den COBRA- und den RNA-Liganden-Datensatz	40
Abbildung 20:	Acetylpromazin - Beste Dockingpose laut GoldScore gegen Originalpose für das erste Modell im PDB-Eintrag 1LVJ; links Originalpose, rechts Dockingpose, Mitte Überlappung im TAR-Konformer. Aus Übersichtsgründen wurde auf die Darstellung der RNA verzichtet. Goldscore = 64,44. Der RMSD zwischen den Posen beträgt 6,9 Å.	45
Abbildung 21:	Scorewerte der Docking-Runde mit den potentiellen RNA-Liganden sowie einem COBRA-Subset (jeder 50te Eintrag); links Boxplot der Scoreverteilung; rechts sortierte Liste der GoldScore-Verteilung; Grundlage war der 1LVJ-Konformer	47
Abbildung 22:	Scorewerte der Docking-Runde mit den potentiellen RNA-Liganden sowie einem COBRA-Subset (jeder 50te Eintrag); links Boxplot der Scoreverteilung; rechts sortierte Liste der GoldScore-Verteilung; die Darstellungen sind diesmal auf einen Wertebereich > -100 beschränkt, um einen genaueren Einblick in den Bereich der „TopScorer“ zu ermöglichen. Grundlage war der 1LVJ-Konformer	47
Abbildung 23:	„Top 10“-Scores der Dockingrunden der einzelnen Screening-Hitlisten-Subsets als Subplots; Grundlage war der 1LVJ-Konformer	48
Abbildung 24:	Strukturen der "Top 10" der Dockingrunde mit der COBRA-Auswahl; die Zahl gibt die Platzierung in der Dockingrundenwertung an	50
Abbildung 25:	Relative Fluoreszenz-Aktivität der 36 von Specs gelieferten Substanzen; Ein Y-Wert von 1 = inaktiv im Assay; An Position 1 ist die Referenz Triple-Arginin dargestellt; Aktive Substanz Nr4 stellte sich als Peptid-Binder heraus. Für Nr.2 konnte gezeigt werden, dass die starke Absorbtion dieser Substanz, das Messergebniss verfälschte (schwarz).	56
Abbildung 26:	Aktive Substanzen im TAR-FRET-Assay. Angegeben sind die Specs-IDs und die Aktiven-Nummer im Assay; Substanz 2 ist schwarz und stört dementsprechend den Assay, Substanz 4 zeigt Aktivität mit dem Peptid; verbleibende RNA-Binder sind somit die Substanzen 1, 3 und 5.	56
Abbildung 27:	Messungen der Substanz M27 zur Bestimmung des IC_{50} im Tat-TAR-FRET-Assay; die X-Achse repräsentiert die Konzentration des jeweiligen Tests, die Y-Achse zeigt die ermittelte relative Fluoreszenzaktivität. Erkennbar ist das Unterschreiten der Basislinie bei einer Konzentration von 100 μ M.	57
Abbildung 28:	Relative Fluoreszenz-Aktivität der 23 von Asinex gelieferten Substanzen; Ein Y-Wert von 1 = inaktiv im Assay; An Position 1 ist die Referenz Triple-Arginin dargestellt; Substanz Nr6 = A20 konnte als einzige Substanz in ihrer Aktivität bestätigt werden.	59
Abbildung 29:	Asinex AS20 – einzige aktive Substanz im TAR-FRET-Assay. Angegeben sind sowohl die Asinex-ID, als auch die LabID; AS20 war die einzige Substanz die in dieser Bestellrunde als aktiv bestätigt werden konnte.	59
Abbildung 30:	Ergebnisse des CFivTT-Assay für die Substanzen der M-Reihe (RNA-Binder-Vorhersagen). Dargestellt ist die relative Inhibition bezogen auf die Kontrollreaktion.	60
Abbildung 31:	Strukturen der Verbindungen, welche im CFivTT-Assay Inhibition gegenüber der Kontrollreaktion zeigten. Die höchste beobachtete	

	Inhibition bei gleichzeitiger geringster Standardabweichung wurde bei M19 und M32 beobachtet.	61
Abbildung 32:	Schnittbild der Überlappungen der Aktiven aus Tat-TAR-FRET-Assay für die M-Reihe.	62
Abbildung 33:	Gegenüberstellung M23 und tri-Arg; blau hinterlegt sind die Guanindin-Gruppen in beiden Molekülen; unter physiologischen Bedingungen sind diese Gruppen protoniert, so dass M23 zwei und tri-Arg hierdurch drei positive Ladungen zusätzlich trägt.	63
Abbildung 34:	Substanzen aus dem Struktur-Wirkungsbeziehungstest für M32 (durchgeführt von Markus Weidlich im CFivTT-Assay).	64
Abbildung 35:	SAR M32 gegen Thianaphthen und 5-Cl-BTMA; Konzentration der Substanzen im Test = 100 μ M (Versuch 1)	65
Abbildung 36:	SAR M32 gegen 1-prop-2-ynyl-3,5,7-triaza-1-azonia-adamantan, Bromid; Konzentration der Substanzen im Test = 100 μ M (Versuch 2)	65
Abbildung 37:	Tetra-aza-adamantan (Urotropin), das Grundgerüst der aktiven Substanzen M19 und M32	66
Abbildung 38:	Adamantyl-Grundgerüste in den Testreihen für das Labor – Gegenüberstellung Aktive gegen Inaktive (CFivTT-Assay); M32 zeigt auch Aktivität im Tat-TAR-FRET-Assay	67
Abbildung 39:	Scaffoldvergleich der aktiven/inaktiven Verbindungen im CFivTT-Assay: Betrachtet wurden ausschließlich die ein Adamantyl-Grundgerüst enthaltenden Strukturen.	67
Abbildung 40:	Substanzen aus den Testreihen, die positive Ladungen enthalten sollen. Die Auftragung erfolgte getrennt nach Aktivität im CFivTT-Assay.	68
Abbildung 41:	Ableitung des Tri-aza-1-azonia-adamantyl-Pharmakophormodells aus den im CFivTT-Assay getesteten Verbindungen: MW1, M19 und M32	69
Abbildung 42:	Gegenüberstellung von M19 und M32. M19 zeigte keine messbare Aktivität im FRET-Assay, M32 zeigte eine Aktivität (RFA = $0,68 \pm 0,01$), blau hinterlegt ist der Teil von M32, der für diesen Unterschied wahrscheinlich verantwortlich zu machen ist.	70
Abbildung 43:	Reduziertes Pharmakophormodell für potentiell allgemeine RNA-Binder-Eigenschaften; Aktivitäten in CFivTT und Tat-TAR-FRET-Assay hängen noch vom verwendeten Rest ab.	71
Abbildung 44:	AG-690/40719331, Substanz aus dem Cherry-Picking (bis heute nicht lieferbar); links ist die Struktur abgebildet, wie sie bei Specs in der Datenbank zu finden ist, rechts ist eine mögliche Resonanzstruktur der Peptidbindung angegeben, beide erfüllen die Doppelbindungshypothese des Pharmakophormodells aus Abbildung 43	71
Abbildung 45:	Gegenüberstellung der Verbindungen M9, M11 und AS16, welche im CFivTT-Assay, trotz kleinster struktureller Unterschiede unterschiedliche Aktivitäten zeigten; rot unterlegt ist die bei M9 fehlende Methoxy-Gruppe, die für dessen Aktivität gegenüber M11 verantwortlich zu sein scheint. Durch Hinzunahme der Messdaten aus der AS-Reihe, konnte zusätzlich AS16 als weiteres strukturell sehr ähnliches Molekül hinzugenommen werden, der Austausch der Methyl-Gruppe durch ein Wasserstoffatom in AS16 (rot hinterlegt), führte ebenfalls zum Verlust der Aktivität. Das Substrukturmotiv 2,4-Di-morpholin-4-yl-[1,3,5]triazin ist, da es auch in M35 und AS17 vorkommt eindeutig nicht hinreichend für eine Aktivität im Assay.	72
Abbildung 46:	Verteilung der Aktiven der M-Reihe auf die Hitlisten der drei Filter	74
Abbildung 47:	Fluoreszein und Acetylpromazin	75
Abbildung 48:	Ethidium-Arginin-Konjugat; Arginin ist als Arg abgekürzt	76

Abbildung 49:	Farbbalken der SOMs; die Werte für den jeweiligen Farbverlauf innerhalb einer SOM, beziehen sich immer auf den relativen Min- und Max-Wert, der in der jeweiligen SOM auftritt.	84
Abbildung 50:	MOE-Deskriptor SOM: Projektionen des COBRA-Datensatzes und des gesamten Literatur-RNA-Liganden Datensatzes; Aufteilung der Projektion des RNA-Liganden-Datensatz in die Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen	85
Abbildung 51:	CATS-Deskriptor SOM: Projektionen des COBRA-Datensatzes und des gesamten Literatur-RNA-Liganden Datensatzes; Aufteilung der Projektion des RNA-Liganden-Datensatz in die Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen.	86
Abbildung 52:	TC3D-Deskriptor SOM: Projektionen des COBRA-Datensatzes und des gesamten Literatur-RNA-Liganden Datensatzes; Aufteilung der Projektion des RNA-Liganden-Datensatz in die Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen.	87
Abbildung 53:	MOE-2D-Deskriptor SOM: Projektion der RNA-Liganden-Datensätze der Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt keine Übereinstimmungen zwischen den belegten Knoten der SOM-Projektionen.	88
Abbildung 54:	CATS-Deskriptor SOM: Projektion der RNA-Liganden-Datensätze der Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt keine Übereinstimmungen zwischen den belegten Knoten den belegten Knoten der SOM-Projektionen.	89
Abbildung 55:	TC3D-Deskriptor SOM: Projektion der RNA-Liganden-Datensätze der Subsets mit und ohne die experimentell unbestätigten Filikovsubstanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt eine Übereinstimmung zwischen den belegten Knoten Knoten der SOM-Projektionen.	89
Abbildung 56:	MOE-2D-Deskriptor SOM: Projektion Filikov-Subsets mit den experimentell unbestätigten Substanzen, sowie die Projektion aller bestellten Substanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt zwei direkte Übereinstimmungen zwischen den belegten Knoten und ferner deren Nähe zu weiteren belegten Knoten in der Projektion der bestellten Substanzen. Weiße Quadrate = direkte Überlappung.	90
Abbildung 57:	CATS-Deskriptor SOM: Projektion Filikov-Subsets mit den experimentell unbestätigten Substanzen, sowie die Projektion aller bestellten Substanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt eine direkte Übereinstimmungen zwischen den belegten Knoten und ferner deren Nähe zu weiteren belegten Knoten in der Projektion der bestellten Substanzen. Weißes Quadrat = direkte Überlappung.	91
Abbildung 58:	TC3D-Deskriptor SOM: Projektion Filikov-Subsets mit den experimentell unbestätigten Substanzen, sowie die Projektion aller bestellten Substanzen. Die Überprüfung auf Überlappungen der Subdatensets zeigt eine direkte Übereinstimmungen zwischen den belegten Knoten und ferner deren Nähe zu weiteren belegten Knoten in der Projektion der bestellten Substanzen. Weißes Quadrate = direkte Überlappung.	91
Abbildung 59:	CATS SOM - Überlappungsanalyse des RNA-Liganden-Datensatzes (unbestätigte Filikov) mit den aktiven Laborsubstanzen	93
Abbildung 60:	CATS SOM - Überlappungsanalyse des RNA-Liganden-Datensatzes (ohne unbestätigte Filikov) mit den aktiven Laborsubstanzen	93
Abbildung 61:	Wahrscheinlicher Ableitungsweg der aktiven Laborsubstanzen aus dem verwendeten Trainingsdatensatz; links CATS-SOM Mitglieder des	

	Knotens [18; 6] (uRNALD), rechts die Mitglieder der CATS-Knoten [18; 6](CFivTT) und [17; 6] (CFivTT).	93
Abbildung 62:	Vermuteter Informationsfluss innerhalb des ANN-Filter-Trainings.	94
Abbildung 63:	SOMMER SOM des Tat-TAR-spezifischen Datensatzes angewendet auf die MOE-2D-Deskriptoren; grün = Aktive, rot = Inaktive; in der gemeinsamen Projektion sind die Knoten nach der häufigsten Klasse eingefärbt.	99
Abbildung 64:	SOMMER SOM des Tat-TAR-spezifischen Datensatzes angewendet auf die TC3D-Deskriptoren; grün = Aktive, rot = Inaktive; in der gemeinsamen Projektion sind die Knoten nach der häufigsten Klasse eingefärbt.	99
Abbildung 65:	Darstellung der Molekülgerüste des Tat-TAR-spezifischen Datensatzes aufgeteilt nach Aktivitätssubklassen; als Referenzauswahl wurden die 11 häufigsten Molekülgerüste des COBRA-Datensatzes verwendet, welche rechts zu sehen sind; Innerhalb des Tat-TAR-spezifischen Datensatzes wurden die Vorkommenshäufigkeiten der Molekülgerüste als Zahlen eingefügt. Kommen diese überhaupt nicht vor, wurden sie entsprechend als durchgestrichen markiert.	100
Abbildung 66:	KS-Distanz gegen Position in MinMax-Sortierung für den MOE-Deskriptorraum nach Abtrennung der Deskriptoren mit konstanten Werten für alle Einträge.	103
Abbildung 67:	KS-Statistik der TC3D-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze	104
Abbildung 68:	KS-Statistik der LIQUID-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze	104
Abbildung 69:	KS-Statistik der MOE-2D-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze	105
Abbildung 70:	KS-Statistik der CATS-2D-Deskriptoren für die Tat-TAR-Assay-spezifischen Datensätze; die 18 Deskriptoren mit einer KS-Distanz von 0,0 wurden bereits vor der Auftragung entfernt.	105
Abbildung 71:	Testreihe der potentiellen Tat-TAR-FRET-spezifischer Liganden; Ausgangsbasis = wasserverdünnte DMSO-Stammlösungen	114
Abbildung 72:	Testreihe der potentiellen Tat-TAR-FRET-spezifischer Liganden; Ausgangsbasis = wasserverdünnte Methanol-Lösungen	114
Abbildung 73:	Aktive im Tat-TAR-FRET-Assay; LR9 bindet möglicherweise an der RNA, LR20 und LR28 binden möglicherweise am Peptid	118
Abbildung 74:	Gegenüberstellung der drei aktiven LR9, AS20 und M23 des Tat-TAR-FRET-Assay; blau hinterlegt sind die erweiterten Guanidin-Gruppen (Methyleneaminoguanidin) in den Molekülen; unter physiologischen Bedingungen sind diese Gruppen protoniert. M23 trägt somit zusätzlich zwei positive Ladungen und LR9 und AS20 jeweils eine positive Ladung.	119
Abbildung 75:	TAR-Konformer: starke Umorganisation der Raumstruktur zwischen einzelnen Modellen des PDB-Eintrags 1ANR, Entfernung zwischen den Atomen H21 des Nukleotids 33 und HO3' des Nukleotids 45 als rote Linie wiedergegeben. Die relative Verkürzung des Abstands der beiden Atome zwischen Modell 7 und 18 beträgt 25,67 Ångström.	120
Abbildung 76:	PDB-Eintrag 1LVJ – Vergleich eines Mitglieds des Ensembles gegen die Überlagerung aller Mitglieder	121
Abbildung 77:	PDB-Eintrag 1ANR – Vergleich des siebten Mitglieds des Ensembles (aufgrund relativer Ähnlichkeit zu 1LVJ) gegen die Überlagerung aller Mitglieder	122
Abbildung 78:	PDB-Eintrag 1ANR – Vergleich des ersten Mitglieds des Ensembles gegen die Überlagerung aller Mitglieder	122

Abbildung 79:	PDB-Eintrag 1ARJ – Vergleich des ersten Mitglieds des Ensembles gegen die Überlagerung aller Mitglieder	122
Abbildung 80:	Start- und Endbild des 1ANR-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.	128
Abbildung 81:	Start- und Endbild des 1ARJ-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.	128
Abbildung 82:	Start- und Endbild des 1QD3-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.	129
Abbildung 83:	Start- und Endbild des 1LVJ-MD-Laufs. Angegeben ist die Anzahl an im Rahmen der Arbeit verfügbaren Einzelbildern, die bisher simulierte Zeitspanne des MD-Laufs, als auch der RMSD zwischen Start- und Endbild basierend auf allen 930 Atomen der TAR.	129
Abbildung 84:	Auftragung des RMSD zwischen Start- und Endstruktur der vier MD-Läufe gegen die MD-Entwicklungszeit. Für die RMSD-Werte basierend auf allen Atomen ergibt sich ein $r = 0,90$, basierend nur auf den Phosphoratomen ein $r = 0,92$.	129
Abbildung 85:	Schema des Informationsflusses innerhalb des RNA-Konformerend-Datenbank-Projekts. Mit Ausnahme der Verwendung der DrugScore ^{RNA} -Scoringfunktion wurden alle Teile im Rahmen dieser Arbeit integriert.	131
Abbildung 86:	Maximale Durchmesser der Strukturen des COBRA-Referendatensatzes, Einträge sortiert nach diesem Kriterium zum Abschätzen der benötigten oberen und unteren Taschengröße	136
Abbildung 87:	Taschengröße gegen Taschenanzahl; links nur der ganze Datensatz; mitte ganzer Datensatz (gefüllte Kreise) und gefiltert nach Durchmesser (ungefüllte Kreise); rechts gefiltert nach Durchmesser; auf der X-Achse ist jeweils die Größe der Tasche (Taschenpunkteanzahl), auf der Y-Achse ist jeweils die Anzahl der Taschen aufgetragen.	137
Abbildung 88:	Gegenüberstellung MD-Läufe und der PDB-Einträge für die Darstellungen der Taschengröße gegen die mittlere Vergrabenheit	141
Abbildung 89:	Taschengröße gegen mittlere Vergrabenheit für die PDB-Einträge 1UUI, 1UUD und UTS	142
Abbildung 90:	Taschengrößen gegen „mittlere Vergrabenheit“ für die MD-Läufe; blau = 1ANR, rot = 1ARJ, gelb = 1QD3; grün = 1LVJ	143
Abbildung 91:	Taschengröße gegen die mittlere Vergrabenheit für alle Taschen des gesamten MD-Datensatzes; Als charakteristischste Positionen für die Betrachtung wurden P1 bis P5 markiert; siehe hierzu die Ausführungen im Text.	145
Abbildung 92:	Gesamter MD-Datensatz – Taschengröße gegen mittlere Vergrabenheit; die Farbkodierung gibt die Häufigkeit der Kombination im Set an; die meisten Taschen sind somit klein und wenig vergraben.	147
Abbildung 93:	Selektion der Taschen des MD-Datensatzes basierend auf dem optimalen Taschendurchmesserbereich – Taschengröße gegen „mittlere Vergrabenheit“; die Farbkodierung gibt die Häufigkeit der Kombination im Set an; die meisten Taschen sind zwar immer noch klein und wenig vergaben, jedoch wurde ein Großteil der als unnütz definierten Extreme entfernt.	148
Abbildung 94:	Verteilung der Radien um P3 für den von den gesamten MD-Läufen abgeleiteten Taschendatensatz, Auftragung erfolgte sortiert nach der Größe des Radius.	150
		229

Abbildung 95:	Fokussierung der Taschenauswahl auf den als optimal angesehenen Bereich um P3; Auswahl erfolgte aufgrund verschieden großer Selektionsradien von 0,3 bis 0,01	151
Abbildung 96:	Populationsverteilung innerhalb der Auftragung der Taschengröße (Y-Achse) und der „mittleren Vergrabenheit“ (X-Achse). Auswahlradius = 0,3; 70.071 Taschen.	152
Abbildung 97:	Populationsverteilung innerhalb der Auftragung der Taschengröße (Y-Achse) und der „mittleren Vergrabenheit“ (X-Achse). Auswahlradius = 0,2; 13.512 Taschen.	153
Abbildung 98:	Populationsverteilung innerhalb der Auftragung der Taschengröße (Y-Achse) und der „mittleren Vergrabenheit“ (X-Achse). Auswahlradius = 0,1; 2.329 Taschen.	154
Abbildung 99:	Visualisierung der Referenztaschenauswahl nach Vergrabenheit und Größe der Tasche; mittlere Vergrabenheit = „m. V.“, Größe = Anzahl Taschenpunkte; 1ANR, 1ARJ, 1LVJ stehen jeweils für die MD-Läufe aus denen die Konformere entstammen. PIT_ID = eindeutiger Bezeichner der jeweiligen Tasche in der Datenbank.	156
Abbildung 100 :	Ablaufdiagramm der PSGP-Extraktion mit Pandora aus der MySQL-Datenbank.	159
Abbildung 101:	PSGP-Projektor für die 2D-Darstellungen der Nukleotide. Als Beispiel wird hier die Bildschirmaufnahme der Projektion des sechsten Nukleotids in der verwendeten TAR-Sequenz im Pandora-Modul gezeigt. Die Einfärbung der Atome, erfolgt nach dem jeweiligen PSGP; blau = niedriges PSGP und rot = hohes PSGP.	160
Abbildung 102:	Histogramm der durchschnittlichen Taschenoberflächenpunkteanzahl für jedes Atom über die Bilder aller verfügbaren MD-Läufe in der MDDB für TAR	161
Abbildung 103:	Histogramm der durchschnittlichen Taschenoberflächenpunkteanzahl für jede Base über alle Bilder aller verfügbaren MD-Läufe in der MDDB für TAR	161
Abbildung 104:	Durchschnittliche Taschenoberflächenpunkteanzahl für die einzelnen Basen und Zuordnung zu einer Beispiel TAR-Struktur, gerendert als Cartoon-Basenmodell	162
Abbildung 105:	TAR-RNA, schematische Darstellung der Sekundärstruktur	162
Abbildung 106:	Drei TAR-Konformere des PDB-Eintrags 1ANR eingefärbt nach ihrem durchschnittlichen PSGP; Darstellung als Cartoonrepräsentation; PSGP < 1,5 rot und PSGP > 1,5 blau	163
Abbildung 107:	Drei TAR-Konformere des PDB-Eintrags 1ANR eingefärbt nach ihrem durchschnittlichen PSGP; Darstellung als Oberflächenrepräsentation; PSGP < 1,5 rot und PSGP > 1,5 blau	163
Abbildung 108:	"Top 10" Taschen nach PSGP-PocketScore; oben ist zusätzlich die größte Tasche im Datensatz zu sehen. Die nach PSGP-Taschenscore ausgewählten Taschen sind zwar relativ groß, aber augenscheinlich differenzierter aufgebaut als die Referenztasche (größte Tasche im Datensatz). Aus Übersichtsgründen wurde nur bei der ersten Tasche der Verlauf des RNA-Rückgrats gezeigt.	165
Abbildung 109:	Verteilung der maximalen PSGP-Taschenscore für den Taschenraum nach Anwendung des Taschendurchmesserfilters. Auftragung erfolgte nach Taschengröße und "mittlerer Vergrabenheit", Farbcodierung erfolgte nach dem maximalen gefundenen PSGP-Taschenscore für den jeweiligen Datenpunkt (Taschengröße und "mittlerer Vergrabenheit")	167
Abbildung 110:	Verteilung der maximalen PSGP-Taschengewichte für den Taschenraum nach Anwendung des Taschendurchmesserfilters. Auftragung erfolgte nach Taschengröße und "mittlerer Vergrabenheit", Farbcodierung	

	erfolgte nach dem maximalen gefundenen PSGP-Taschenscore für den jeweiligen Datenpunkt (Taschengröße und "mittlerer Vergrabenheit")	168
Abbildung 111:	Boxplots der PSGP-Taschenscores für die vier MD-Trajektorien	169
Abbildung 112:	Durchschnittlichste Strukturen der TAR bezogen auf die RMSD-Distanzmatrixberechnungen der Einzelensembles 1QD3, 1ARJ, 1LVJ; die gezeigten Strukturen haben den kleinsten durchschnittlichen RMSD zu allen anderen Konformeren ihres Ensembles.	175
Abbildung 113:	Visualisierung zweier Berechnungsschritte für einen SIMPrint einer TAR-Struktur; a) zeigt den TAR-Konformer als ganzes; b) und c) zeigen die Erstellung der Radialausschnitte um den ersten Referenzpunkt; d) und e) zeigen dies für den zweiten Referenzpunkt	179
Abbildung 114:	HK-means-Baum der Konformere der 20 Konformere des PDB-Eintrags 1ANR basierend auf einer alle Atomkoordinaten einschließenden SIMPrint-Berechnung; Als Blätter des Baums sind die Visualisierungen der alignierten Strukturen, der den SIMPrints zugrundeliegenden Konformere dargestellt. Die Zahlen unter dem Baum geben die Modellnummern im PDB-Eintrag wieder.	181
Abbildung 115:	Baum aus dem Hk-Means-Clustering der SIMPrints der Taschenoberflächen der jeweils größten Tasche jeden Konformers der 1ANR; exemplarisch wurde die Oberfläche von vier, der nach diesem Kriterium ähnlichsten Taschenpaare, visualisiert.	182
Abbildung 116:	Oberflächenaufteilung des PDB-Eintrags 1ANR nach Elementen C, N, O und H für jeden der zwanzig Konformere; P wurde nicht aufgetragen, da deren Anteil durchgängig < 0,01 % war.	184

9. Literaturverzeichnis

Aboul-ela F, Karn J, Varani G (1995): The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* 253 (2): 313-332.

Aboul-ela F, Karn J, Varani G (1996): Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* 24 (20): 3974-3981.

Ammon HPT (2005): Arzneimittelneben- und -wechselwirkungen. Ein Handbuch und Tabellenwerk für Ärzte und Apotheker. Wissenschaftliche Verlagsgesellschaft, 4. Auflage.

Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984): Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81 (8): 3684-3690.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The Protein Data Bank *Nucleic Acids Research* 28: 235-242.

Böcker A, Derksen S, Schmidt E, Teckentrup A, Schneider G (2005): A hierarchical clustering approach for large compound libraries. *J. Chem. Inf. Model.* 45 (4): 807-815.

Böcker A, Sasse BC, Nietert M, Stark H, Schneider G (2007): GPCR Targeted Library Design: Novel Dopamine D(3) Receptor Ligands. *Chem. Med.Chem.* 2 (7): 1000-1005.

Boden O (2005): Design und Synthese heterocyclischer Liganden für die TAR-RNA aus HIV-1. Dissertation, Frankfurt.

Box GEP (1979): Robustness in the strategy of scientific model building. In: Launer, Wilkinson GN (Hrsg.): *Robustness in Statistics*. Academic Press, New York.

Bunk W, Thalhammer S, Rath C, Aschenbrenner T, Morfill GE, Heckl WM, Jamitzky F, Stark RW (2001): Scaling-index method as an image processing tool in scanning-probe microscopy. *Ultramicroscopy* 86: 241-246.

Cai Y, Xiao F, Guo P (2008): The effect of N- or C-terminal alterations of the connector of bacteriophage phi29 DNA packaging motor on procapsid assembly, pRNA binding, and DNA packaging. *Nanomedicine* [Jan. 15 2008 Epub ahead of print].

Cheatham TE, Cieplak P, Kollman PA (1999): A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* 16 (4): 845-862.

Cheng X, Hochlowski J, Tang H, Hepp D, Beckner C, Kantor S, Schmitt R (2003): Studies on repository compound stability in DMSO under various conditions. *J. Biomol. Screen.* 8 (3): 292-304.

Chin K, Sharp KA, Honig B, Pyle AM (1999): Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* 6: 1055-1061.

Davis B, Afshar M, Varani G, Murchie AI, Karn J, Lentzen G, Drysdale M, Bower J, Potter AJ, Starkey ID, Swarbrick T, Aboul-ela F (2004): Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.* 336 (2): 343-356.

Du Z, Lind KE, James TL. (2002): Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.* 9 (6): 707-712.

Essmann U, Perera L, Berkowitz ML, Darden T, Hsing Lee, Pedersen LG (1995):A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103 (19): 8577-8593.

- Faber C, Sticht H, Schweimer K, Rosch P (2000):** Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* 275: 20660-20666.
- Fechner U, Franke L, Renner S, Schneider P, Schneider G (2003):** Comparison of correlation vector methods for ligand-based similarity searching. *J. Comput. Aided Mol. Des.* 17: 687-698.
- Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, James TL (2000):** Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided Mol. Des.* 14 (6): 593-610.
- Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, James TL (2000):** Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided. Mol. Des.* 14 (6): 593-610.
- Foloppe N, Matassova N, Aboul-ela F (2006):** Towards the discovery of drug-like RNA ligands? *Drug Discovery Today* 11 (21/22): 1019-1027.
- Fulle S, Gohlke H (2008):** Analyzing the flexibility of RNA structures by constraint counting. *Biophys J.*; 94(11): 4202-19.
- Gillet J, Khatib W, Willett P, Fleming PJ, Green DVS (2002):** Combinatorial Library Design Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* 42: 375-385.
- Gillet VJ, Willett P, Fleming PJ, Green DVS (2002):** Designing focused libraries using MoSELECT. *J. Mol. Graph. Model* 20: 491-498.
- Grabowski K, Schneider G (2007):** Properties and Architecture of Drugs and Natural Products Revisited. *Curr. Chem. Biol.* 1: 115-27.
- Gross ME (2001):** Tranquilizers, α 2-Adrenergic Agonists and Related Agents, In: Adams HR, ed.: *Veterinary Pharmacology and Therapeutics*. Iowa State University Press: Ames, USA; 8. Auflage: 299-342.
- Halgren TA (1996):** Merck molecular force field.1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.*: 490-519.
- Hall LW, Clarke KW, Trim CM (2001):** Principles of sedation, analgesia and premedication. In: Hall LW, Clarke KW, Trim CM (Hrsg.): *Veterinary Anaesthesia*. WB Saunders: London, UK; 10. Auflage: 75-112.
- Hermann T, Westhof E (1999):** Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J. Med. Chem.* 42 (7): 1250-1261.
- Hess B., Kutzner C., van der Spoel D., Lindahl E. (2008):** GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, Web Release Date: February 2, 2008
- Humphrey W, Dalke A, Schulten K (1996):** VMD – Visual Molecular Dynamic. *J. Molec. Graphics* 14: 33-38.
- Kabsch WA (1976):** Solution of the best rotation to relate two sets of vectors. *Acta Crystallographica* 32: 922.
- Karn J (1999):** Tackling Tat. *Journal of Molecular Biology* 293 (2): 235-254.
- Karplus M, McCammon JA (2002):** Molecular dynamics simulations of biomolecules. *Nature Structural Biology* 9: 646-652.
- Kennard RW, Stone LA (1969):** Computer aided design of experiments. *Technometrics* 11: 137-148.
- Kohonen T (1982):** Self-organized formation of topologically correct feature maps. *Biological Cybernetics* V 43 (1): 59-69.

Kozikowski BA, Burt TM, Tirey DA, Williams LE, Kuzmak BR, Stanton DT, Morand KL, Nelson SL (2003): The effect of freeze/thaw cycles on the stability of compounds in DMSO. *J. Biomol. Screen.* 8 (2): 210-215.

Kozikowski BA, Burt TM, Tirey DA, Williams LE, Kuzmak BR, Stanton DT, Morand KL, Nelson SL (2003): The effect of room-temperature storage on the stability of compounds in DMSO. *J. Biomol. Screen.* 8 (2): 205-209.

Leeper TC, Athanassiou Z, Dias RL, Robinson JA, Varani G (2005): TAR RNA recognition by a cyclic peptidomimetic of Tat protein. *Biochemistry* 44: 12362-12372.

Lieberschuetz JW (2008): Evaluating docking programs: keeping the playing field level. *J. Comput. Aided Mol. Des.* [Epub ahead of print 13 Jan. 2008].

Lottspeich F, Zorbach H (1998): Bioanalytik. Spektrum, Akad. Verl.: Heidelberg, Berlin

Ludders JW, Reitan JA, Martucci R, Fung DL, Steffey EP (1983): Blood pressure response to phenylephrine infusion in halothane-anesthetized dogs given acetylpromazine maleate. *American Journal of Veterinary Research* 44 (6): 996-999.

Ludwig V (2005): Auxiliär-vermittelte Synthese von nicht-natürlichen Aminosäuren als Bausteine für RNA-Liganden. Dissertation, Frankfurt.

Ludwig V, Krebs A, Stoll M, Dietrich U, Ferner J, Schwalbe H, Scheffer U, Dürner G, Göbel MW (2007): Tripeptides from Synthetic Amino Acids Block the Tat-TAR Association and Slow Down HIV Spread in Cell Cultures. *Chem. Bio. Chem.* 8: 1850-1856.

Marco E, Gago F (2007): Overcoming the Inadequacies or Limitations of Experimental Structures as Drug Targets by Using Computational Modeling Tools and Molecular Dynamics Simulations. *Chem. Med. Chem.* 2: 1388-1401.

Matsumoto C, Hamasaki K, Mihara H, Ueno A (2000): Novel Screening Method Using Intramolecular FRET for the Discovery of the Molecules That Bind HIV-1 TAR RNA Specifically. *Bioorg. Med. Chem. Lett.* 10: 1857-1861.

Morgner N, Barth H-D, Brutschy B (2006): A New Way To Detect Noncovalently Bonded Complexes of Biomolecules from Liquid Micro-Droplets by Laser Mass Spectrometry. *Aust. J. Chem.* 59: 109-114.

Morgner N, Barth H-D, Brutschy B, Scheffer U, Breitung S., Göbel M. (2008): Binding sites of TAR and of TAR mutants for various ligands, probed with a new laser mass spectrometry. *Journal of The American Society for Mass Spectrometry*, eingereicht.

Müller-Oerlinghausen B, Düppenbecker H, Lasek R (1999): Handbuch der unerwünschten Arzneimittelwirkungen. Urban & Fischer:

Pfeffer P, Gohlke H (2007): DrugScoreRNA-Knowledge-Based Scoring Function to Predict RNA-Ligand Interactions. *J. Chem. Inf. Model* 47 (5): 1868-1876.

Proschak E, Rupp M, Derksen S, Schneider G (2008): Shapelets: Possibilities and limitations of shape-based virtual screening. *J. Comput. Chem.* 29 108-114.

Renner S, Ludwig V, Boden O, Scheffer U, Göbel M, Schneider G (2005): New inhibitors of the Tat-TAR RNA interaction found with a "fuzzy" pharmacophore model. *Chem. Bio. Chem.* 6 (6): 1119-1125.

Renner S (2005): Development and application of fast fuzzy pharmacophore-based virtual screening methods for scaffold hopping. Dissertation: Frankfurt.

Schmuker M, Schwarte F, Brück A, Proschak E, Tanrikulu Y, Givehchi A, Scheiffele K, Schneider G (2007): SOMMER: Self-Organizing Maps for Education and Research. *J. Mol. Model.* 13: 225-228.

- Schmuker M, Givhchi A, Schneider G (2004):** Impact of different software implementations on the performance of the Maxmin method for diverse subset collection. *Mol. Divers.* 8(4): 421-5.
- Schneider P, Schneider G (2003):** Collection of Bioactive Reference Compounds for Focused Library Design, *QSAR & Combinatorial Science* 22 (7): 713-718.
- Schneider G, Wrede P (1998):** Artificial neural networks for computer-based molecular design. *Prog. Biophys. Mol. Biol.* 70: 175-222.
- Schneider G, Neidhart W, Giller T, Schmid G (1999):** "Scaffold-Hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew. Chemie* 111: 3068-3070.
- Schneider G, Neidhart W, Giller T, Schmid G (1999):** "Scaffold-Hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew. Chemie Int.* 38: 2894-2896.
- Schüller A, Schneider G (2008):** Identification of Hits and Lead Structure Candidates with Limited Resources by Adaptive Optimization. *J. Chem. Inf. Model.* 48 (7): 1473–1491.
- Seifert G (2007):** Synthese und Charakterisierung substituierter Pyrazole und Chinoxalinderivate als potenzielle Inhibitoren des Tat-TAR-Komplexes von HIV-1. Dissertation: Frankfurt.
- Starke K, Freiburg I (2001):** Grundlagen der Pharmakologie des Nervensystems. In: Forth W, Henschler D, Rummel W, Förstermann U, Starke K (Hrsg.): *Allgemeine und Spezielle Pharmakologie und Toxikologie.* Urban & Fischer: München, Jena; 8. Auflage: 111-146.
- Sun Tzu** The Art of War, Übersetzer: Lionel Giles; www.gutenberg.org
- Tanrikulu Y, Nietert M, Scheffer U, Proschak E, Grabowski K, Schneider P, Weidlich M, Karas M, Göbel M, Schneider G (2007):** Scaffold hopping by "fuzzy" pharmacophores and its application to RNA targets. *Chem. Bio. Chem.* 8 (16): 1932-1936.
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. (2008):** Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* 17; 358 (3): 252-60.
- Weisel M, Proschak E, Schneider G (2007):** PocketPicker: Analysis of Ligand Binding-Sites with Shape Descriptors. *Chemistry Central J.* 01: 7.
- Wöhrl B. (2003):** Antiterminationskomplexe als Ziel neuer Therapeutika. *RetroVirus Bulletin* 2: 4-5; www.virologie.uni-erlangen.de
- Ye X, Kumar RA, Patel DJ (1995):** Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* 2: 827-840.

Danksagungen

Ich möchte mich an dieser Stelle herzlich beim Modlab-Team von Herrn Prof. Dr. Gisbert Schneider am Institut für Organische Chemie der Johann Wolfgang Goethe Universität in Frankfurt am Main bedanken. Insbesondere gilt mein Dank Prof. Dr. Gisbert Schneider für das Überlassen dieses interessanten Themas im Rahmen der Mitarbeit im SFB 579. Ferner gilt dies für die anregenden Diskussionen und die Freiheit neue Wege ausprobieren zu dürfen, sowie seine Hilfe beim „Cherry-Picken“ von Testsubstanzen und die Schaffung einer positiven Arbeitsatmosphäre.

Ich danke dem SFB 579 im Allgemeinen und stellvertretend für die anderen Mitglieder Prof. Dr. Joachim Engels für die freundliche Unterstützung und anregenden Seminare, die im Rahmen dieser Arbeit eine Inspirationsquelle waren. Ferner möchte ich ihm für seine freundlichen Rückmeldungen im Bezug auf meine Arbeit als Webadministrator des SFB 579 danken. Der Dank gilt auch Beate Conrady für die positive Zusammenarbeit im Rahmen der Arbeit an der SFB-Webseite.

Für die Möglichkeit diese Arbeit durchzuführen möchte ich hier auch der DFG für die finanzielle Unterstützung im Rahmen des Sonderforschungsbereiches RNA-Ligand-Wechselwirkungen (SFB 579) danken. Dasselbe gilt für die indirekte Unterstützung des Projekts durch das Beilstein-Institut zur Förderung der Chemischen Wissenschaften, das die „Beilstein Stiftungsprofessur“ erst ermöglicht hat, dessen Ausrüstung ich im Rahmen dieser Arbeit benutzen durfte.

Die Computer-Hardware würde natürlich nicht so reibungslos ohne den System-Administrator Norbert Dichter funktionieren, dem ich hiermit herzlichst danken möchte, für all die kleinen und großen Hilfen, die ich im Laufe der vergangenen Jahre erhalten habe. Vielen Dank für die Geduld und netten Erklärungen, die ich im Rahmen dieser Arbeit im Bezug auf Linux erfahren habe. Sowie die Hilfe beim Retten meiner Daten, wenn mal wieder eins meiner Script „Amok gelaufen“ ist, oder ich das eigentlich unmögliche geschafft habe und mit meiner Neugier mein Linux „unrettbar“ zerstört hatte. Ferner Danke ich ihm für die Unterstützung bei der Umsetzung extravaganterer Hardware- und Software-Wünsche, wie beispielsweise dem Aufbau eines experimentellen MySQL-Cluster-Systems und vielen weiteren Hilfen und Anregungen.

Ich danke Brigitte Scheidemantel-Geiß für die organisatorische Unterstützung im Rahmen der Arbeit, die vor allen Dingen bei der Klärung von „andersartigen“ Rechnungsfragen immer zur Hilfe war.

Ich möchte den vielen Mitgliedern des Modlab-Teams für die positive Arbeitsatmosphäre danke und auch für die vielen Hilfen die einem „Biochemiker auf Abwegen“ den weiteren Einstieg in die Chemieinformatik durch Vermittlung von weiterführenden Informatikkenntnissen ermöglicht haben. Weiterhin gilt dies für die Bereitstellung verschiedenster Code-Schnipsel und anregenden Diskussionen zu Fragestellungen über die Durchführbarkeit „eigenwilliger“ Ansätze. Mein besonderer Dank gilt hier den von mir am stärksten „strapazierten“ Helfern: Alexander Böcker, Andreas Schüller, Ewgenij Proschak, Jan Hiß, Kristina Grabowski, Martin Weisel, Matthias Rupp, Michael Meißner, Michael Schmuker, Svetlana Derksen, Yusuf Tanrikulu (als mein Büromitbewohner, wohl die am meisten befragte Person).

Katja Barthelmes möchte ich für die Unterstützung beim Sammeln der Liganden aus der verfügbaren Literatur danken.

Lutz Franke sei für seinen TC3D-Deskriptoransatz recht herzlichst gedankt, sowie für die Beantwortung von Fragen bei der Reimplementierung.

Andreas Schüller sei noch einmal besonders gedankt, da er als zweiter Biochemiker im AK eine wichtige Diskussionsquelle für alle biochemischen Fragestellungen darstellte.

Frau Dr. Ute Scheffer und ihren Mitarbeiterinnen Frau Elisabeth Kalden bin ich für die Bereitstellung von RNA-Proben und Puffern, sowie für die freundliche Anleitung und Unterstützung bei der Durchführung des Tat-TAR-FRET-Assay, sowie chemischer Fragestellungen sehr dankbar. Insbesondere danke ich Frau Dr. Ute Scheffer für die anregende Diskussion während meiner Zeit im Labor und auch darüber hinaus für die Beantwortung vieler chemischer Fragestellungen im Rahmen der Fertigstellung dieser Arbeit, sowie dem Bereitstellen der Liganden-Sammlung des AK-Göbels.

Ich danke hiermit recht herzlich Prof. Dr. Michael Göbel für die Möglichkeit die Versuche am Tat-TAR-FRET-Assay in seinem Arbeitskreis durchführen zu können, sowie die anregenden Diskussionen die sich daraus ergaben.

Ich danke hiermit Markus Weidlich aus dem Arbeitskreis Karas recht herzlich für die gute Zusammenarbeit, die es ermöglicht hat für die vorhergesagten potentiellen RNA-Binder auch Messwerte seines CFivTT-Assay und teils darüber hinaus auch in weiteren Assay-Systemen zu erhalten. Ferner danke ich ihm für die Durchführung der massenspektrometrischen Untersuchungen der Testsubstanzen. Durch seine Arbeit konnte dem „allgemeinen RNA-Binder“-Filter-Projekt ein zusätzlicher Blick auf die potentielle antibiotische Wirkweise der Testsubstanzen ermöglicht werden. Ferner waren die Diskussionen immer wieder motivierend für die weitere Arbeit. Dank sei hier auch Prof. Dr. Michael Karas der diese Zusammenarbeit unterstützt hat.

Hiermit möchte ich auch Eva Kestner meiner „RNA-Konformergeneratorin“ für das Verfügbarmachen ihrer MD-Trajektorien-Datensätze herzlichst danken. Diese Daten bildeten unbestreitbar eine starke, umfangreiche Grundlage für die in dieser Arbeit durchgeführten Analysen zur Taschenvorhersage in Konformerensembeln. Ferner möchte ich Dr. Holger Gohlke für seine anregenden Gespräche zu MD-Simulationen danken, in dessen Arbeitskreis auch die Arbeiten von Eva Kestner durchgeführt wurden.

Ich möchte mich ferner bei meinem neuen Arbeitskreis von Prof. Dr. H. Ulrich Göringer an der TU Darmstadt bedanken in deren Kreis ich diese Arbeit beenden konnte und natürlich auch für die finanzielle Unterstützung durch die TU Darmstadt in dieser Abschlussphase.

Der besondere Dank gilt meiner Familie, die mich während dieser Arbeit immer wieder angespornt und unterstützt hat. Insbesondere für das unermüdliche wiederholte Fehlersuchen/Korrekturlesen mehrerer Versionen dieser Arbeit möchte ich meinem Vater danken. Meinem Bruder möchte ich besonders für Rat und Tat beim Kampf mit den Formatierungen in Word danken. Und zum Schluss möchte ich besonders meiner Frau Michaela danken die mich immer wieder motiviert hat und mit mir diese Zeit durchgestanden hat in der es zum Schluss auch noch monströse kafkaeske Schachtelsätze zu bezwingen galt.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit von mir selbstständig durchgeführt und ausgearbeitet worden ist. Es wurde lediglich Bezug auf die angegebene Literatur genommen. Ich versichere ferner, dass diese Arbeit noch keinem anderen Prüfungsamt vorgelegen hat.

Offenbach am Main, 19. Februar 2008 Manuel Manfred Nietert

Lebenslauf

Zur Person: Manuel Manfred Nietert
geboren am 22.01.1979 in Lahn-Gießen
Staatsangehörigkeit: deutsch
verheiratet

Schulische Ausbildung:

1985 - 1989	Limesschule/Grundschule Pohlheim	<i>Pohlheim-Watzenborn-Steinberg</i>
1989 - 1995	Adolf-Reichwein-Schule/integrierte Gesamtschule	<i>Pohlheim-Watzenborn-Steinberg</i>
1995 - 1996	Coronado-Highschool	<i>Colorado-Springs/USA</i>
1996	Erwerb des High-School-Equivalency-Diploma	<i>Colorado/USA</i>
1996 - 1998	Liebigschule Gießen/gymnasiale Oberstufe	<i>Gießen</i>
1998	Erwerb der allgemeinen Hochschulreife an der Liebigschule Gießen	<i>Gießen</i>

Berufliche Ausbildung:

1998 Ausbildung zum Sanitäter bei der Bundeswehr

Hochschulausbildung:

1999 – 2004	Biochemiestudium an der J.W. Goethe-Universität	<i>Frankfurt am Main</i>
Sep. 2001	Erwerb des Vordiploms der Biochemie am 11. Sep. 2001 Prüfungsfächer: Biochemie, Biophysikalische Chemie, Pflanzenphysiologie	<i>Frankfurt am Main</i>
Juli 2004	Erwerb des Diploms der Biochemie am 16.7.2004 Prüfungsfächer: Biochemie, Biophysikalische Chemie, Bioinformatik Thema der Diplomarbeit: <i>„Vorhersage der Blut-Hirnschrankengängigkeit von Molekülen mit künstlichen neuronalen Netzen“</i>	<i>Frankfurt am Main</i>
2004-2007	Arbeit zur Promotion in der Arbeitsgruppe von Prof. Dr. Gisbert Schneider an der Johann Wolfgang Goethe–Universität Thema der Dissertation <i>“ Virtuelles Screening nach RNA-Liganden: Zum Umgang mit einer flexiblen Zielstruktur”</i>	<i>Frankfurt am Main.</i>

Publikationsverzeichnis

Publikationen

Tanrikulu Y, Nietert M, Scheffer U, Proschak E, Grabowski K, Schneider P, Weidlich M, Karas M, Göbel M, Schneider G., Scaffold hopping by "fuzzy" pharmacophores and its application to RNA targets. *Chembiochem*. 2007 Nov 5;8(16):1932-6.

Böcker A, Sasse BC, Nietert M, Stark H, Schneider G., GPCR Targeted Library Design: Novel Dopamine D(3) Receptor Ligands. *ChemMedChem*. 2007 Jul;2(7):1000-5

Poster

Weidlich M., Nietert M., Beckhaus T., Schneider G., Stein T. und Karas M. (2008) Proteomanalytische Charakterisierung neuer antimikrobieller Translationsinhibitoren mittels 2-D DIGE Technologie. 41. Jahrestagung der Deutschen Gesellschaft für Massenspektrometrie (DGMS), Gießen.

Weidlich M., Nietert M., Suhartono M., Göbel M., Schneider G., Stein T., and Karas K. (2007) Rational Approaches to Discover Novel Antimicrobial RNA-Ligands. *Joint Meeting on Medicinal Chemistry*, Portoroz/Slowenien

Tanrikulu Y., Proschak E., Nietert M., Grabowski K., Scheffer U., Göbel M., Schneider G. (2007) Virtual Screening for Novel TAR RNA Ligands with LIQUID: Scaffold-hopping Using a Combination of Automated Ligand Docking and Fuzzy Pharmacophore Modeling, *Fourth Joint Sheffield Conference on Cheminformatics*, Sheffield.

Grabowski K., Meissner M., Schüller A., Nietert M., Proschak E., Tanrikulu Y., Stauch B., Münk C., Scheffer U., Cichutek K., Schubert-Zsilavec M., Göbel M., Schneider G. (2007) Applications of MOE in an Academic Environment, *European User Group Meeting der Chemical Computing Group*, Berlin

Nietert M., Weisel, M., Proschak, E., Kestner E., Gohlke, H. Schneider, G. (2007) Automated Prediction of Putative Binding Sites in Flexible RNA-Targets, *International Symposium „RNA-Ligand-Interactions“*, Frankfurt am Main

Nietert M., Tanrikulu Y., Grabowski K., Scheffer U., Göbel M., Weidlich M., Karas M., Kestner E., Gohlke H., Schneider G. (2006) Virtual Screening for RNA Ligands – Coping with the Flexibility of Drug Targets, *2nd German Conference on Cheminformatics*, Goslar.