

Adaptive Optimierung fokussierter Substanzbibliotheken

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

Vorgelegt beim Fachbereich 14 Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Andreas Schüller
aus Düsseldorf

Frankfurt 2008
(D 30)

Vom Fachbereich 14 Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Harald Schwalbe

Gutachter: Prof. Dr. Gisbert Schneider, Prof. Dr. Michael Göbel

Datum der Disputation: 15. April 2009

Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass aus Schriften Entlehnungen, soweit sie in der Dissertation nicht ausdrücklich als solche mit Angabe der betreffenden Schrift bezeichnet sind, nicht stattgefunden haben.

Andreas Schüller

Für Leyla

„It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change.“

– Charles Darwin

Danksagung

Ich bedanke mich bei meinem Doktorvater, Prof. Dr. Gisbert Schneider, für die Betreuung meiner Doktorarbeit und für die Ermöglichung eines überaus interdisziplinären Projektes.

Ich bedanke mich bei Prof. Dr. Michael Göbel für die Betreuung des experimentellen Teils meiner Arbeit.

Ich bedanke mich beim Fonds der Chemischen Industrie für die freundliche Gewährung eines zweijährigen Stipendiums.

Ich bedanke mich beim Sonderforschungsbereich 579 „RNA-Liganden-Wechselwirkungen“ in dessen Rahmen der experimentelle Teil meiner Arbeit eingebunden war.

Mein besonderer Dank gilt Dr. Ute Scheffer, in deren Labor die Messungen des experimentellen Teils stattfanden und die mir stets mit Rat und Tat zur Seite stand.

Ich bedanke mich bei Dr. Lutz Weber und Mitarbeitern (Roche, Basel) für die Bereitstellung einer kombinatorischen Substanzbibliothek Ugi-artiger Verbindungen.

Ich bedanke mich bei Dr. Ursula Dietrich und Michaela Stoll für die Durchführung von Zellkulturexperimenten am Georg-Speyer-Haus in Frankfurt am Main.

Ich bedanke mich herzlich bei meiner Arbeitsgruppe, dem Modlab-Team. Mein besonderer Dank gilt hier Michael Meissner für die Bereitstellung einer Java-Implementierung von Partikelschwärmen. Ich danke Matthias Rupp für die zahlreichen interessanten Gespräche. Ich bedanke mich bei Dr. Ulrich Fechner, der mir die konsequente objektorientierte Programmierung beigebracht hat und bei meinem Biochemiekollegen Dr. Manuel Nietert, der mich in den experimentellen Teil eingeführt hat. Ich bedanke mich bei meinen Kollegen Dr. Michael Schmuker, Kristina Grabowski, Dr. Lutz Franke, Ewgenij Proschak, bei unserer Sekretärin Brigitte Scheidemantel-Geiß und unserem fleißigen Systemadministrator Norbert Dichter für ihre Unterstützung.

Ich bedanke mich bei der Arbeitsgruppe von Prof. Dr. Michael Göbel für Hilfe und Unterstützung.

Mein ganz lieber Dank gilt Kirsten Büsgen, die in zahlreichen freiwilligen Stunden das professionelle Setzen dieser Arbeit übernommen hat.

Ich bedanke mich bei meinen lieben Eltern für ihre fortwährende Unterstützung in allen Belangen des Lebens und bei meinem Bruder und seiner Freundin für ihre Hilfe.

Ganz besonders bedanke ich mich bei meiner lieben Leyla, die mit ihrer Engelsgeduld und pausenlosen Unterstützung ganz maßgeblich zum Gelingen dieser Arbeit beigetragen hat.

Zusammenfassung

Im Laufe der Evolution haben sich Mechanismen entwickelt, die optimal an ihre jeweilige Problemstellung angepasst sind. Einige dieser Mechanismen waren Vorlage für biologisch inspirierte Optimierungsalgorithmen. In dieser Arbeit wurden drei solcher Optimierungsalgorithmen für die adaptive Optimierung von Substanzbibliotheken im Wirkstoffdesign eingesetzt.

Ziel dieser Arbeit war ein vertieftes Verständnis der Anwendung der adaptiven Optimierung fokussierter Substanzbibliotheken im Niedrigdurchsatzscreening. Dazu wurden im ersten Teil die Optimierungsalgorithmen „Zufallssuche“, „Simulated Annealing“, „Evolutionsstrategie“ und „Partikelschwarmoptimierung“ in der Programmiersprache Java implementiert. Die korrekte Funktionsweise der implementierten Algorithmen wurde anhand der Multidimensionalen Skalierung (MDS) überprüft, einer Projektionsmethode hochdimensionaler Räume, für die sowohl deterministisch-analytische als auch stochastisch-heuristische Lösungswege existieren. Durch den Vergleich der heuristischen Lösungen der Optimierungsalgorithmen mit der analytischen Lösung der Hauptkoordinatenanalyse konnten die Algorithmen erfolgreich validiert werden. Es zeigte sich, dass die Heuristiken qualitativ bessere Lösungen finden konnten als das deterministische Verfahren.

Obwohl in der Vergangenheit verschiedene Optimierungsalgorithmen mit Erfolg für den Entwurf fokussierter Substanzbibliotheken eingesetzt wurden, wurden diese noch nicht systematisch miteinander verglichen. Im zweiten Teil dieser Arbeit wurden die implementierten Algorithmen deshalb in Bezug auf den Entwurf von Substanzbibliotheken einem systematischen Vergleich unterzogen. Dazu wurden retrospektive Optimierungsexperimente mit Serinproteaseinhibitoren in verschiedenen Deskriptorräumen durchgeführt. Um Laborbedingungen zu simulieren, wurde die Gesamtzahl verfügbarer Substanztestungen auf 300 beschränkt. Die Ergebnisse zeigten, dass Evolutionsstrategien effizient mit großen Populationen und wenigen Iterationen arbeiteten. Sie sind damit ausgezeichnet für einen Einsatz in einer Niedrigdurchsatzscreening-Kampagne geeignet. Partikelschwarmoptimierungen arbeiteten mit mittleren Populationsgrößen effizient und sind ebenfalls für die Optimierung fokussierter Substanzbibliotheken geeignet. Das Simulated Annealing war für den Entwurf fokussierter Bibliotheken im Niedrigdurchsatzscreening weniger geeignet, da es eine große Anzahl an Iterationen mit einer Populationsgröße von eins bevorzugte. Es konnte gezeigt werden, dass die Optimierung fokussierter

Substanzbibliotheken ohne Vorwissen über das biologische Zielmolekül möglich ist. Analysen des Suchraumes durch Selbstorganisierende Karten (SOM) ergaben, dass die Voraussetzungen dafür eine ausreichende Häufigkeit aktiver Verbindungen in der Screeningbibliothek (mind. 2%) und eine moderate Neigung zur Clusterbildung aktiver Verbindungen waren.

Der dritte Teil dieser Arbeit beschreibt den Entwurf einer fokussierten Bibliothek von RNA-Liganden mit Hilfe der adaptiven Optimierung. Alle bislang veröffentlichten Studien konzentrierten sich auf den Entwurf von Proteinliganden. Hier wird gezeigt, dass sich die adaptive Optimierung fokussierter Substanzbibliotheken auch auf Ribonukleinsäuren als biologisches Zielmolekül anwenden lässt. In einer hybriden, prospektiven Optimierungsstudie wurden nach dem Vorbild einer iterativen Niedrigdurchsatzscreening-Kampagne vom Computer vorgeschlagene Moleküle im Labor auf Aktivität getestet. Die Substanzen wurden auf Inhibition einer spezifischen molekularen Wechselwirkung im Replikationszyklus von HIV getestet (Tat-TAR-Interaktion), die für die Replikation des Virus essentiell ist. Der Suchraum umfasste etwa 200.000 Screeningverbindungen des Anbieters Specs (Delft, Niederlande). In insgesamt vier Generationen wurden 170 Screeningverbindungen auf Aktivität überprüft. Darunter wurden neun Substanzen positiv auf Inhibition der Tat-TAR-Interaktion getestet ($IC_{50} = 506 \pm 497 \mu\text{M}$). Dies entspricht einer Trefferquote von 5,3%, wobei lediglich 0,089% der Verbindungen der Specs-Bibliothek untersucht wurden. Die zwei potentesten Kandidaten **AS_S10/06_33** ([2-(4-Chloro-phenoxy)-ethyl]-(2-nitro-5-piperazin-1-yl-phenyl)-amin) und **AS_S10/06_16** (2-(5-Amino-4-cyano-1*H*-pyrazol-3-yl)-*N*-hydroxy-acetamidin) wiesen einen IC_{50} von 51 μM bzw. 116 μM auf. Eine Literatursuche (durchgeführt mit SciFinder Scholar 2007) ergab, dass es sich um zwei neuartige RNA-Liganden handelt.

Eine Analyse der Struktur-Wirkungs-Beziehungen (SAR) der aktiven Analoga von **AS_S10/06_33** deckte eine endständige primäre oder sekundäre Aminogruppe auf, die als essentiell für die Bioaktivität dieser Substanzen ausgemacht werden konnten. Eine durchgeführte Dockingstudie lieferte eine mögliche Erklärung für diese Beobachtung (Interaktion mit dem RNA-Rückgrat). In Zellkulturexperimenten konnte mittels einer Reportergenuntersuchung für **AS_S10/06_16** auch *in vivo* ein inhibitorischer Effekt auf die Tat-TAR-Interaktion gemessen werden. Das Nitroaren **AS_S10/06_33** erwies sich in Zellkulturexperimenten schon bei geringer Konzentration als cytotoxisch.

English Summary

In the course of Evolution mechanisms have evolved that are optimally adapted to their respective tasks. Some of those mechanisms served as prototypes for biologically-inspired optimization algorithms. In this study we employed three such optimization algorithms in the adaptive optimization of compound libraries as an application of drug design.

The aim of this thesis was to deepen the understanding of the application of adaptive optimization of focussed compound libraries in low-throughput screening. In the first part of this study, the optimization algorithms “Random Search”, “Evolution Strategy”, and “Particle Swarm Optimization” were implemented in the programming language Java. The correct operation of those algorithms was validated by means of Multidimensional Scaling (MDS) – a projection method of high-dimensional spaces. Both, deterministic-analytical as well as stochastic-heuristic approaches exist to solve the MDS problem. By comparing the heuristic solutions of the optimization algorithms with the analytical solution of the Principal Coordinate Analysis the algorithms could successfully be validated. We show that the heuristics were able to find better solutions than the deterministic approach.

Although different optimization algorithms have been employed with success in focussed library design in the past, they have not yet been systematically compared with each other. Therefore, in the second part of this thesis a systematic comparison of the implemented algorithms with respect to focussed library design was performed. Retrospective optimization experiments were conducted with serine protease inhibitors in different descriptor spaces. To simulate laboratory conditions the number of compound tests was restricted to 300. Results show that Evolution Strategies work efficiently with large populations and few cycles. This makes them well suited for an application in a low-throughput screening campaign. Particle Swarm Optimizations operated efficiently with moderate population sizes and are also suited for focussed library design. Simulated Annealing was not as suited since it preferred a large number of iterations with a population size of one. Results indicate that optimization of focussed compound libraries without prior knowledge of the biological target is possible. The requirements were a sufficient frequency of active compounds in the screening library (at least 2%) and moderate clustering of active compounds as determined by a Self-Organizing Map analysis of the search space.

In the third part of this thesis we describe the design of a focussed compound library of RNA ligands by adaptive optimization. Studies published so far concentrated on the design of protein ligands. Here we show the successful application of the adaptive optimization of focussed compound libraries with ribonucleic acids as the biological target. Compounds suggested by a computer program were tested for activity in a hybrid, prospective optimization study. The substances were tested for inhibition of a specific molecular interaction in the replication cycle of HIV (Tat-TAR interaction) which is essential for virus replication. The search space was comprised of ca. 200,000 screening compounds of the vendor Specs (Delft, The Netherlands). In a total of four generations 170 screening compounds were tested for activity. Nine compounds were found to be active Tat-TAR interaction inhibitors ($IC_{50} = 506 \pm 497 \mu\text{M}$). This corresponds to a hit rate of 5.3% while only 0.089% of the compounds of the Specs library were screened. The two most potent candidates **AS_S10/06_33** ([2-(4-Chlorophenoxy)-ethyl]-(2-nitro-5-piperazine-1-yl-phenyl)-amine) and **AS_S10/06_16** (2-(5-Amino-4-cyano-1*H*-pyrazole-3-yl)-*N*-hydroxy-acetamidine) exhibited an IC_{50} of 51 μM , respectively, 116 μM . A literature search (performed with SciFinder Scholar 2007) revealed that novel RNA ligands were found.

An analysis of the structure-activity relationships (SAR) of the active analogues of **AS_S10/06_33** revealed a terminal primary or secondary amino group which was found to be essential for bioactivity of those substances. An analysis by automated docking resulted in a possible explanation of this observation (interaction with the RNA backbone). An inhibitory effect on the Tat-TAR interaction by **AS_S10/06_16** could also be detected *in vivo* by a reporter gene assay in cell culture. The nitroarene **AS_S10/06_33** proved to be cytotoxic in the cell culture experiments.

Inhaltsverzeichnis

Danksagung	XI
Zusammenfassung	XIII
English Summary	XV
Abkürzungsverzeichnis	XXI
1 Einleitung	1
1.1 Adaptive Optimierung von Substanzbibliotheken	1
1.2 Algorithmen der adaptiven Optimierung	4
1.3 Zielsetzung	6
2 Biologische Zielmoleküle	9
2.1 Serinproteasen	9
2.1.1 Chymotrypsin	11
2.1.2 Faktor X _a	11
2.1.3 Trypsin	11
2.1.4 Tryptase	12
2.1.5 Urokinase-Typ-Plasminogenaktivator	12
2.2 Das menschliche Immunschwächevirus und AIDS	13
2.3 Der Lebenszyklus des HI-Virus	15
2.4 Antiretrovirale Therapie	16
2.5 Die Antitermination bei HIV	17
3 Material und Methoden	19
3.1 Fluoreszenzbasierte Bioaktivitätsbestimmung	19
3.2 Experimentelle Durchführung der Aktivitätsbestimmung	21
3.3 Datensätze	23
3.3.1 Ugi-Bibliothek	23
3.3.2 COBRA - Sammlung bioaktiver Referenzanaloga	24
3.3.3 RNA-Liganden-Bibliothek	25
3.3.4 Specs-Substanzkatalog	25
3.3.5 Präprozessierung der Substanzbibliotheken	26
3.4 Molekulare Repräsentationen	27
3.4.1 Deskriptorberechnung für retrospektive Optimierungsstudien	30
3.4.2 Deskriptorberechnung für eine prospektive Optimierungsstudie	32
3.4.3 Deskriptorberechnung für eine SOM-Analyse von RNA-Liganden	32
3.5 Optimierungsalgorithmen	33
3.5.1 Zufallssuche	37
3.5.2 Simulierte Abkühlung	38

3.5.3	Evolutionsstrategie	39
3.5.4	Partikelschwarmoptimierung	42
3.6	Selbstorganisierende Karten (SOM)	44
3.7	Hauptkomponentenanalyse (PCA)	46
3.8	Pharmakophorsuche	47
3.9	Automatisiertes Docking	48
4	Multidimensionale Skalierung zur Validation von Optimierungsalgorithmen	49
4.1	Zielsetzung	49
4.2	Die Multidimensionale Skalierung	49
4.3	Ergebnisse und Diskussion	50
4.4	Schlussfolgerungen	55
5	Retrospektive Vergleichsstudie von Optimierungsalgorithmen	57
5.1	Zielsetzung	57
5.2	Methodik	58
5.2.1	Strategieparameter	60
5.3	Komparative Parametersелеktion	61
5.5	Optimierung mit binären Fitnessfunktionen	80
5.6	Modifikation der Optimierungsalgorithmen	86
5.7	Schlussfolgerungen	88
6	Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion	91
6.1	Zielsetzung	91
6.2	Planung der prospektiven Studie	91
6.3	Die Initialpopulation	94
6.4	Durchführung	98
6.5	Übersicht der experimentellen Ergebnisse	99
6.6	Optimierungsverhalten und Konvergenz	111
6.7	Grundgerüstspringen bei RNA-Liganden	116
6.8	Löslichkeitsprobleme	119
6.9	Fehlerhafte Repräsentation der Nitro-Gruppe	123
6.10	Suche nach Nicht-Nitroverbindungen	126
6.11	Struktur-Wirkungs-Beziehungen der Nachkommen von OB108	131
6.11.1	Pharmakophormodell der aktiven Nachkommen von OB108	136
6.11.2	Evaluierung des automatisierten Dockings in Nukleinsäuren	139
6.11.3	Automatisiertes Docking der Nachkommen von OB108	142
6.11.4	SAR des OB108-Nachkommens AS_S06/07_40	143

6.12	Struktur-Wirkungs-Beziehungen der Nachkommen von OB080d	145
6.13	Zellkultorexperimente	151
6.14	Schlussfolgerungen	155
7	Schlussfolgerungen und Ausblick	159
8	Anhang	167
8.1	Liste Molekularer Deskriptoren	167
8.2	Multidimensionale Skalierung (mathematische Einführung)	176
8.3	Nachkommen der prospektiven Optimierungsstudie	178
8.4	Dosis-Wirkungs-Kurven der aktiven Verbindungen	184
8.4.1	AS_S10/06_16	184
8.4.2	AS_S10/06_33	185
8.4.3	AS_S01/07_19	186
8.4.4	AS_S01/07_35	187
8.4.5	AS_S01/07_62	188
8.4.6	AS_S06/07_27	189
8.4.7	AS_S06/07_28	190
8.4.8	AS_S06/07_29	191
8.4.9	AS_S06/07_40	192
8.5	Über den Metabolismus von Nitrobenzol	193
8.6	Versuchsbedingungen der Zellkultorexperimente	194
9	Literaturverzeichnis	197
10	Lebenslauf	213
11	Publikationen	217

Abkürzungsverzeichnis

2D	zweidimensional
3D	dreidimensional
A	Adenin
ADME	Absorption, Distribution, Metabolismus und Exkretion
AIDS	Erworbenes Immunschwächesyndrom (für engl. „Acquired Immune Deficiency Syndrome“)
AK	Arbeitskreis
Arg	Arginin
ART	Antiretrovirale Therapie
Asp	Asparaginsäure
C	Cytosin
CAT	Chloramphenicol-Acetyltransferase
COBRA	Sammlung bioaktiver Referenzanaloga (engl. „Collection Of Bioactive Reference Analogues“)
CV	Korrelationsvektor (für engl. „Correlation Vector“)
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure (für engl. „Deoxyribonucleic Acid“)
EA	Evolutionärer Algorithmus
ES	Evolutionsstrategie
FDA	Arzneimittelzulassungsbehörde der USA (engl. „Food and Drug Administration“)
FRET	Fluoreszenzresonanzenergietransfer
FTatRh _d	Fluorescein-AAARKKRRRQRRRAAAC-Rhodamin
FX _a	Faktor X _a
G	Guanin
GPCR	G-Protein-gekoppelter Rezeptor (engl. „G-Protein Coupled Receptor“)
HAART	Hochaktive Antiretrovirale Therapie
His	Histidin
HIV	Menschliches Immunschwächevirus (auch „Humanes Immundefizienzvirus“)
HTS	Hochdurchsatzscreening (engl. „High-Throughput Screening“)

IC_{50}	Mittlere inhibitorische Konzentration (Konzentration, bei der 50% Inhibition vorliegt)
k -NN	k -Nächste-Nachbarn-Algorithmus
Leu	Leucin
LTR	Lange terminale Wiederholungseinheiten (für engl. „Long Terminal Repeats“)
Lys	Lysin
MeOH	Methanol
MG	Molekulargewicht
MDS	Multidimensionale Skalierung
MOE	Molecular Operating Environment
mRNA	Boten-Ribonukleinsäure (für engl. „Messenger Ribonucleic Acid“)
NNRTI	Nicht-nukleosidischer Reverse-Transkriptaseinhibitor
NRTI	Nukleosidischer oder nukleotidischer Reverse-Transkriptaseinhibitor
NP	nichtdeterministisch, polynominelle Zeit
P	Phosphat
PCA	Hauptkomponentenanalyse (engl. „Principle Component Analysis“)
PCO	Hauptkoordinatenanalyse (engl. „Principal Coordinate Analysis“)
Phe	Phenylalanin
PI	Proteaseinhibitor
pK_a	Negativer dekadischer Logarithmus der Säuredissoziationskonstante (Säurestärke, Azidität)
PPP	potentieller Pharmakophorpunkt
PSO	Partikelschwarmoptimierung
QSAR	Quantitative Struktur-Wirkungsbeziehung (für engl. „Quantitative Structure-Activity Relationship“)
Rhd	Rhodamin
RMSD	Quadratwurzel der mittleren quadratischen Abweichung (für engl. „Root Mean Square Deviation“)
RFU	Relative Fluoreszenzeinheit (für engl. „Relative Fluorescence Unit“)
RNA	Ribonukleinsäure (für engl. „Ribonucleic Acid“)

RRE	Rev-Antwortelem (für engl. „Rev Response Element“)
SA	Simulated Annealing (engl. „Simulierte Abkühlung“)
SAR	Struktur-Wirkungs-Beziehung (für engl. „Structure-Activity Relationship“)
Ser	Serin
SFB	Sonderforschungsbereich
SOM	Selbstorganisierende Karte (für engl. „Self-Organizing Map“)
RT	Reverse Transkriptase
Std	Standardabweichung (für engl. „Standard Deviation“)
SVL	Scientific Vector Language
TAR	Transaktivierungs-Antwortelem (für engl. „Transactivation Response Element“)
Tat	Transaktivator der Transkription
Trp	Tryptophan
Tyr	Tyrosin
U	Uracil
UNAIDS	Gemeinsames Programm der Vereinten Nationen für HIV/AIDS
uPA	Urokinase-Typ-Plasminogenaktivator
Val	Valin
WHO	Weltgesundheitsorganisation (für engl. „World Health Organisation“)

1 Einleitung

Das Streben nach Optimalität ist ein Grundprinzip in der Entwicklung der bekannten Lebewesen. So haben sich im Laufe einer über 3 Milliarden Jahre währenden Evolution Lebensformen entwickelt, die optimal an ihre jeweilige Umwelt angepasst sind.¹ Man könnte sie als das Ergebnis eines langen Optimierungsprozesses im Versuchsraum der Erde betrachten.¹ Das grundlegende Prinzip der Evolution sind Zyklen von Variation und Selektion. Diese bewirken eine schrittweise Annäherung an einen optimalen Zustand, ausgehend von einer suboptimalen Anfangslösung. Wird dieses Prinzip abstrahiert und in Form eines Algorithmus aufgeschrieben, können damit mathematische Optimierungsprobleme gelöst werden.² Bei geeigneter Formulierung der Fragestellung kann fast jedes Optimierungsproblem mathematisch beschrieben werden. So kann das Prinzip der schrittweisen Annäherung auf eine Vielzahl von Problemen angewendet werden, wie z.B. dem Auffinden der schnellsten Bahnverbindung,³ der Verringerung des Strömungswiderstandes im Windkanal,¹ dem Auffinden der kürzesten Route eines Handlungsreisenden⁴ oder dem Design eines potenten, neuen Wirkstoffes.^{5,6}

1.1 Adaptive Optimierung von Substanzbibliotheken

Die Identifizierung neuer Leitstrukturen ist das entscheidende Ziel der frühen Phase der Wirkstoffentwicklung.⁷ Leitstrukturen sind wichtige Vorstufen auf dem Weg zum Wirkstoff. Es handelt sich dabei um chemische Substanzen, die bereits einige gewünschte Eigenschaften des finalen Produktes aufweisen und die das Potential haben, zu einem Wirkstoff weiterentwickelt zu werden.⁸ Gewünschte Eigenschaften könnten die Bioaktivität gegen ein bestimmtes biologisches Zielmolekül, eine einfache synthetische Zugänglichkeit oder ein ausgewogenes Verhältnis zwischen Lipophilie und Wasserlöslichkeit sein.

Das grundsätzliche Ziel der Leitstrukturfindung ist die Auslese einiger weniger chemischer Verbindungen mit gewünschten Eigenschaften aus der Vielzahl an real oder hypothetisch verfügbaren Molekülen.⁹ Der chemische Raum ist dabei extrem groß. Die Anzahl synthetisch zugänglicher organischer Moleküle wird auf $10^{60} - 10^{100}$ geschätzt.¹⁰⁻¹² Es liegt auf der Hand, dass eine systematische Testung dieser riesigen Anzahl von Substanzen nicht machbar ist. Durch die Fortschritte im Hochdurchsatzscreening (*HTS*, für engl. „High-Throughput Screening“) und in der parallelen Synthese stehen seit den frühen 1990er Jahren wertvolle Werkzeuge für die pharmazeutische Forschung zur Verfügung.¹³ Mit Hilfe der Robotik können so

große kombinatorische Substanzbibliotheken synthetisiert und systematisch getestet werden.¹⁴

Dennoch, HTS-Kampagnen erfordern einen nicht unerheblichen finanziellen Aufwand und ergeben nicht immer viele validierte Treffer.¹⁵⁻¹⁷ Alternative Ansätze setzen auf Niedrigdurchsatzmethoden und konzentrieren sich auf das „Rosinenpicken“ ausgewählter Verbindungen – meist nur weniger dutzend oder hundert – mit gewünschter Aktivität.^{9,18} Unter diesen Methoden befinden sich auch die rechnergestützten, chemieinformatischen Verfahren. Mit ihrer Hilfe ist es möglich, die Verbindungen auch sehr großer molekularer Bibliotheken mit verhältnismäßig wenig Aufwand zu priorisieren.¹⁹ Chemieinformatische Methoden bauen auf vereinfachende Modelle und Annahmen, um Aussagen über die Wirkstoffartigkeit und voraussichtliche Potenz einer chemischen Verbindung zu treffen. Typische Vertreter dieser Methoden sind das Virtuelle Screening und automatisierte molekulare Docking,²⁰ die Pharmakophorsuchen,²¹ QSAR-Methoden (Quantitative Struktur-Wirkungs-Beziehung, für engl. „Quantitative Structure-Activity Relationship“)²² und das *De-novo*-Design.²³ Einige dieser Verfahren wurden auch in dieser Arbeit angewendet und werden in Kapitel 3 vertiefend behandelt. Trotz der zweifellosen Attraktivität dieser rechnergestützten Verfahren kann bis zum heutigen Tage kein Computerprogramm das abschließende Experiment im Labor ersetzen. Die rechnergestützten Verfahren sind dennoch erfolgreich. Bis zum Jahr 2005 erhielten zehn anhand der dreidimensionalen Rezeptorstruktur rational entworfene Wirkstoffe die Zulassung der amerikanischen Arzneimittelzulassungsbehörde FDA (für engl. „Food and Drug Administration“).^{23a}

In dieser Arbeit wurde der Ansatz des adaptiven Entwurfs fokussierter Substanzbibliotheken gewählt.^{16,24,25} Das Ziel des adaptiven Entwurfs von Substanzbibliotheken ist es, die vollständige biologische Testung einer molekularen Screeningbibliothek zu vermeiden. Stattdessen soll eine „intelligente“ Navigation durch den chemischen Raum erfolgen, um so nur eine geringe Anzahl an Substanzen mit gewünschten Eigenschaften auszuwählen.^{9,18} Die Navigation im chemischen Raum erfolgt mit Hilfe stochastischer Optimierungsalgorithmen (Abbildung 1-1):²⁶ ① Eine Initialpopulation von Molekülen wird aus einer Substanzbibliothek ausgewählt. ② Die ausgewählten Verbindungen werden synthetisiert oder eingekauft. ③ Die Bioaktivität der Moleküle wird im Labor experimentell bestimmt. ④ Die gemessenen Aktivitätsdaten werden wieder in den Optimierungsalgorithmus eingespeist, welcher eine neue und hoffentlich aktivere Generation von Molekülen durch „*In silico*“-Evolution erzeugt. Die Generierung neuer Moleküle erfolgt je nach eingesetztem Optimierungsalgorithmus

unterschiedlich und greift auf den Molekülvorrat der verwendeten Substanzbibliothek zurück (gestrichelte Linie). ⑤ Diese Schleife aus Synthese und Testung wird so lange wiederholt bis ein Abbruchkriterium erfüllt ist. Das Verfahren ist ein hybrider Ansatz, das den Computer als Ideengenerator für neue Molekülvorschläge einsetzt und tatsächlich gemessene Aktivitätsdaten zur Bewertung der Moleküle verwendet. Neue aktive Verbindungen können potentiell in jeder Iteration des Verfahrens durch die experimentelle Testung der synthetisierten Substanzen im Labor gefunden werden. Dieser der Bionik entlehnte Ansatz verwendet die „Natur“ nicht nur als Inspirationsquelle für einen Optimierungsalgorithmus, sondern setzt sie in Form biologischer Aktivitätsmessungen in jedem Durchgang des Optimierungszyklus zur Validierung der vorgeschlagenen Zwischenlösungen ein. Anders als bei den meisten chemieinformatischen Verfahren erfolgt hier die validierende oder falsifizierende Rückmeldung durch ein Laborexperiment unmittelbar nach jedem Designschritt.

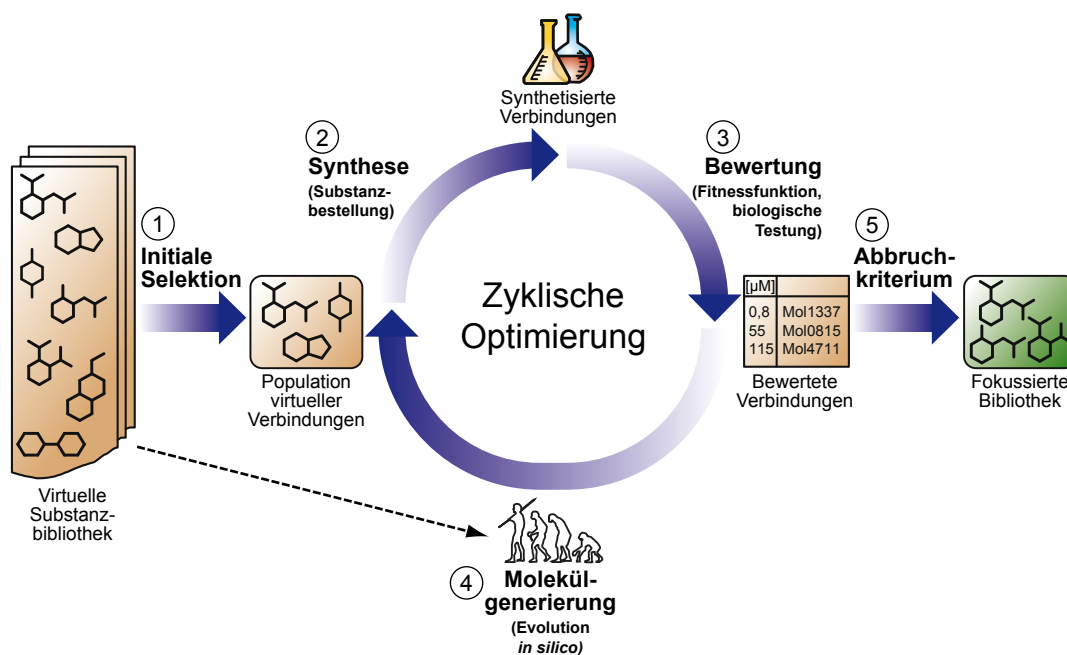


Abbildung 1-1: Schema der adaptiven Optimierung von fokussierten Substanzbibliotheken. ① Eine Initialpopulation von Molekülen wird aus einer Substanzbibliothek ausgewählt. ② Die ausgewählten Verbindungen werden synthetisiert oder eingekauft. ③ Die Bioaktivität der Moleküle wird im Labor experimentell bestimmt. ④ Die gemessenen Aktivitätsdaten werden wieder in den Optimierungsalgorithmus eingespeist, welcher eine neue und hoffentlich aktivere Generation von Molekülen durch „In silico“-Evolution erzeugt. Die Generierung neuer Moleküle greift auf den Molekülvorrat der verwendeten Substanzbibliothek zurück (gestrichelte Linie). ⑤ Diese Schleife aus Synthese und Testung wird so lange wiederholt bis ein Abbruchkriterium erfüllt ist. Wird in Schritt ③ anstatt eines Laborexperiments eine rechnerbasierte Fitnessfunktion zur Bestimmung der Qualität ausgewertet, kann auf eine Synthese der Moleküle vollständig verzichtet werden. In diesem Fall wird Schritt ② des Schemas ausgelassen.

Wird in Schritt ③ anstatt eines Laborexperiments eine rechnerbasierte *Fitnessfunktion* zur Bestimmung der Qualität der Moleküle ausgewertet, kann auf eine Synthese der Verbindungen vollständig verzichtet werden. Eine Fitnessfunktion ist die zu optimierende Zielfunktion eines Optimierungsalgorithmus. In diesem Fall wird Schritt ② des Schemas in Abbildung 1-1 ausgelassen. Diese Variante wurde in dieser Arbeit für retrospektive Optimierungsexperimente angewendet. Bei retrospektiven Experimenten werden Substanzbibliotheken eingesetzt, deren Moleküle bereits zuvor systematisch im Labor auf Bioaktivität getestet wurden. Die Bewertung der Moleküle erfolgt hier durch Nachschlagen der Aktivitätswerte aus einer Liste aller verfügbaren, retrospektiven Aktivitätsdaten. Die Aktivitätsdaten bleiben dabei dem Algorithmus so lange verborgen, bis dieser die Daten explizit anfragt. Dies entspricht einem pseudoprospektiven Vorgehen.

Das Ergebnis der adaptiven Optimierung ist schließlich eine kleine Substanzbibliothek möglichst potenter Verbindungen, die auf einen spezifischen biologischen Zielrezeptor fokussiert sind.²⁷

1.2 Algorithmen der adaptiven Optimierung

Evolutionäre Algorithmen (EA) und unter diesen besonders die Genetischen Algorithmen (GA)²⁸ werden seit Jahren für den Entwurf von Substanzbibliotheken eingesetzt.^{24,25,27,29-37} EA basieren auf einem stark vereinfachten Konzept der Darwinschen Evolution. Durch Anwendung des Prinzips der „natürlichen Selektion“ werden Eltern aus Populationen von Individuen ausgewählt, aus denen anschließend mit Hilfe der genetischen Operatoren „Mutation“ und „Cross-Over“ Nachkommen erzeugen werden. Bei den GA werden Lösungskandidaten gewöhnlicher Weise durch „Chromosomen“ repräsentiert. Dies sind binäre Vektoren, welche ganzzahlige Werte kodieren. Diese Form der Repräsentation ist ideal für die Beschreibung kombinatorischer Reaktionsprodukte in Eduktschreibweise geeignet. Im adaptiven Design von Substanzbibliotheken wurde dem entsprechend auch vorwiegend von der Optimierung virtueller kombinatorischer Bibliotheken berichtet.⁵ Seltener fungierten Screeningbibliotheken als Molekülvorrat.³⁷ Letztere werden klassischerweise durch Vektoren molekularer Deskriptoren repräsentiert. Molekulare Deskriptoren beschreiben Moleküle durch berechenbare physikochemische Eigenschaften und sind ganzzahliger oder reellwertiger Natur. Für diese Form der Repräsentation sind Evolutionsstrategien (ES)^{1,2,38} intuitiv geeignet. Die ES gehört ebenfalls zur Gruppe der evolutionären Algorithmen, arbeitet mit Mutations- und Cross-Over-Operatoren

und beschreibt Lösungskandidaten durch reellwertige Vektoren (vgl. Abschnitt 3.5.3). ES wurden schon vielfach mit Erfolg für das engverwandte ligandenbasierte *De-novo*-Design eingesetzt.^{6,23,39} Beim *De-novo*-Design werden Moleküle aus molekularen Fragmenten von Grund auf neu erschaffen. Ein Optimierungsalgorithmus hilft bei der iterativen Verknüpfung molekularer Bausteine zu Verbindungen mit gewünschten Eigenschaften.

Ein weiterer Optimierungsalgorithmus, der erfolgreich für das Design von Substanzbibliotheken eingesetzt wurde, ist das Simulated Annealing (*SA*, engl. „Simulierte Abkühlung“).^{33,40,41} SA ist ein globaler Optimierungsalgorithmus, der die Abkühlung eines physikalischen Systems simuliert.⁴²⁻⁴⁴ Abhängig von einem Temperaturfaktor werden beim SA nicht nur Fortschritte, sondern auch Rückschritte in der Qualität einer Lösung akzeptiert (vgl. Abschnitt 3.5.2). Von proprietären iterativen Suchstrategien, die nicht auf einem etablierten Optimierungsalgorithmus basieren, wurde ebenfalls berichtet.^{37,45,46}

In dieser Arbeit wird die Partikelschwarmoptimierung (*PSO*) als alternativer Optimierungsalgorithmus für den Entwurf fokussierter Substanzbibliotheken eingeführt. PSO ist eine biologisch inspirierte, populationsbasierte Optimierungstechnik, die auf einer Abstraktion biologischer Schwärme basiert.^{47,48} Ein Partikelschwarm ist eine Population von Partikeln, die sich simultan durch den Suchraum bewegen und dabei gegenseitig Information austauschen (vgl. Abschnitt 3.5.4). PSO wurden bereits erfolgreich auf eine Reihe chemieinformatischer Probleme angewendet.⁴⁹⁻⁵⁴

Die meisten Optimierungsstudien der wissenschaftlichen Literatur verwendeten ein rechnergestütztes Verfahren zur Auswertung der Fitnessfunktion.²³ Die Notwendigkeit der gleichzeitigen Optimierung mehrerer Ziele wurde erkannt und mit Hilfe gewichteter Summen,^{30,41} Attraktivitätswerten (engl. „Desirability scores“)³⁴ und Pareto-Rangfolgen (engl. „Pareto ranking“)^{31,32} realisiert. Nur wenige Studien operierten direkt auf experimentell bestimmten Aktivitätsdaten. Berichtet wurde hier u.a. das Design von Thrombin-Inhibitoren,^{25,27} peptidischen Stromelysin-Substraten,³⁵ peptidischen Trypsin-Inhibitoren,³⁶ Verbindungen gegen verschiedene biologische Zielmoleküle (antibakteriell, antifungal, antialzheimer und Antagonisten G-Protein-gekoppelter Rezeptoren (*GPCR*, für engl. „G-Protein Coupled Receptor“),³⁷ antibakteriellen Verbindungen³⁴ und das *De-novo*-Design von Cannabinoidrezeptor-Liganden.⁵⁵

Unabhängig vom gewählten Optimierungsalgorithmus besteht die generelle Arbeitsweise der adaptiven Optimierung fokussierter Substanzbibliotheken aus sich wiederholenden Zyklen von Molekülgenerierung und -bewertung. Diese entsprechen iterativen Runden von Synthese und Testung im Laborexperiment. Die drei integralen Bestandteile der adaptiven Optimierung sind 1.) ein Vorrat an Screeningverbindungen, 2.) ein Optimierungsalgorithmus und 3.) eine Fitnessfunktion. Eine Kombination dieser drei Bestandteile nach Abbildung 1-1 stellt ein *adaptive System* dar. Ein adaptives System ist ein System, das über eine Vorrichtung verfügt, die eigene Leistung in Bezug auf ein gegebenes Qualitätsmaß oder eine Optimalitätsbedingung beständig selbst zu überwachen und das ferner über eine Vorrichtung verfügt, die eigenen Parameter durch eine Rückkopplungsschleife so anzupassen, dass eine Annäherung an dieses Optimum erreicht wird.⁵⁶ Ein entscheidendes Merkmal von adaptiven Systemen ist Kontinuität (z.B. durch Vererbung oder ein Gedächtnis).^{57,58} So vererben erfolgreiche Elternindividuen bei den EA ihre variierten (mutierten) Eigenschaften an ihre Kinder weiter. Beim SA werden die Lösungskandidaten zweier aufeinander folgender Generationen gespeichert und mit einander verglichen (implizites Gedächtnis). Die PSO sieht ein explizites Gedächtnis vor. Jeder Partikel speichert die individuell beste und die global beste Position des Suchraumes. ES und PSO verfügen zudem über autoadaptive Verfahren zur Regelung der Schrittweite im Suchraum (vgl. Abschnitte 3.5.3 und 3.5.4).

1.3 Zielsetzung

Ziel dieser Arbeit war die Durchführung und Evaluierung der adaptiven Optimierung fokussierter Substanzbibliotheken im Niedrigdurchsatzscreening. Die Kombination von adaptiver Optimierung mit Niedrigdurchsatzexperimenten ist vollständig automatisierbar und kommt im Vergleich zum HTS mit geringeren Ressourcen aus.⁹ Dennoch sind nur wenige Beispiele dieses Ansatzes beschrieben.^{25,27,34-37,55}

Das erste Ziel dieser Arbeit war die Implementierung der Optimierungsalgorithmen „Zufallssuche“, „Simulated Annealing“,⁴²⁻⁴⁴ „Evolutionsstrategie“^{41,2,38} und „Partikelschwarmoptimierung“^{47,48} in der Programmiersprache Java (Version 1.5, Sun Microsystems, Inc., Santa Clara, USA).⁵⁹ Die korrekte Funktionsweise der implementierten Algorithmen wurde anhand der Multidimensionalen Skalierung (MDS) überprüft, einer Projektionsmethode hochdimensionaler Räume, für die sowohl deterministisch-analytische als auch stochastisch-heuristische Lösungswege existieren. Das Ergebnis dieser Validierung ist in Kapitel 4 beschrieben.

Obwohl in der Vergangenheit verschiedene Optimierungsalgorithmen mit Erfolg für den Entwurf fokussierter Substanzbibliotheken eingesetzt wurden, wurden diese noch nicht systematisch miteinander verglichen. Das zweite Ziel dieser Arbeit war deshalb der systematische Vergleich der implementierten Algorithmen in Bezug auf den Entwurf von Substanzbibliotheken. Dazu wurden retrospektive Optimierungsexperimente mit Serinproteaseinhibitoren in verschiedenen Deskriptorräumen durchgeführt. Es wurde untersucht, ob bei begrenzter Gesamtzahl von Substanztestungen eine optimale Einstellung von Populationsgröße und Anzahl der Iterationen existiert. Ferner wurde analysiert, ob sich von der Beschaffenheit des Suchraumes Rückschlüsse auf das Suchverhalten ziehen lassen. Dazu wurden Projektionen der Suchräume mit Hilfe selbstorganisierender Karten (SOM, vgl. Abschnitt 3.6) durchgeführt. Schließlich wurde verglichen, wie sich die Optimierungsalgorithmen bei Verwendung einer binären Fitnessfunktion verhalten. Die Ergebnisse dieser Vergleichsstudie sind in Kapitel 5 präsentiert.

Das dritte Ziel dieser Arbeit war der Entwurf einer fokussierten Bibliothek von RNA-Liganden mit Hilfe der adaptiven Optimierung. Alle bislang veröffentlichten Studien konzentrierten sich auf den Entwurf von Proteinliganden. Hier wird die Hypothese aufgestellt, dass sich die adaptive Optimierung fokussierter Substanzbibliotheken auch auf Ribonukleinsäuren als biologisches Zielmolekül anwenden lässt. Es wurde deshalb eine hybride, prospektive Optimierungsstudie durchgeführt. Nach dem Vorbild einer iterativen Niedrigdurchsatzscreening-Kampagne wurden vom Computer vorgeschlagene Moleküle im Labor auf Aktivität getestet. Als Testsystem wurde eine spezifische molekulare Wechselwirkung im Replikationszyklus des menschlichen Immunschwächevirus (*HIV*) ausgewählt (Tat-TAR-Interaktion), die für die Replikation des Virus essentiell ist.^{60,61} Für die prospektive Optimierung standen etwa 200.000 Screeningverbindungen des Anbieters Specs (Delft, Niederlande)⁶² als Suchraum zur Verfügung. Es wurden das Konvergenzverhalten, die Trajektorie der Suche und die Neigung zum Grundgerüstspringen mit Hilfe selbstorganisierender Karten analysiert. Des Weiteren wurden die Struktur-Wirkungs-Beziehungen (*SAR*, für engl. „Structure-Activity Relationship“) der gefundenen RNA-Liganden mit verschiedenen chemieinformatischen Methoden untersucht. Zwei ausgewählte TAR-RNA-Liganden wurden ferner in Zellkulturexperimenten einer alternativen, funktionalen Testung unterzogen. Die Ergebnisse dieser hybriden Studie sind ausführlich in Kapitel 6 beschrieben.

2 Biologische Zielmoleküle

2.1 Serinproteasen

Serinproteasen gehören zur Gruppe der Endopeptidasen und können mit ihrer enzymatischen Aktivität Proteine hydrolytisch spalten. Sie sind die am besten untersuchte Klasse der Proteasen. Ihren Namen verdanken sie einem essentiellen Serinrest im aktiven Zentrum der Enzyme. Proteasen übernehmen sehr vielfältige Aufgaben im Stoffwechsel. Als Verdauungsenzyme sind sie für die Proteolyse von Nahrungsproteinen verantwortlich. In der Blutgerinnung vermittelt eine Kaskade proteolytischer Aktivierungen eine schnelle Reaktion auf Verletzungen. Auch einige Protein hormone (z.B. Insulin), Proteine des Bindegewebes (z.B. Kollagen) und an Entwicklungsprozessen beteiligte Proteine (z.B. die Kollagenase) werden aus inaktiven Vorstufen (Zymogene, Proenzyme) durch Hydrolyse bereitgestellt. In Form von Caspasen sind Proteasen am programmierten Zelltod (Apoptose) beteiligt. Dieser Abschnitt stellt im Wesentlichen eine Zusammenfassung aus zwei Lehrbüchern der Biochemie und des Wirkstoffdesigns dar.^{8,63}

Charakteristisch für alle Serinproteasen ist eine katalytische Triade aus Aspartat (Asp) 102, Histidin (His) 57 und Serin (Ser) 195. Der Histidinrest positioniert die Serinseitenketten und polarisiert dessen Hydroxylgruppe durch Ausbildung einer Wasserstoffbrücke. Histidin wirkt auf diese Weise als Basenkatalysator. Das entstehende Alkoxidion ($-O^-$) der Serinseitenkette wirkt nun stärker nukleophil. Die Seitenkette von Aspartat 102 hilft bei der Orientierung des Histidinrests und macht dessen Imidazolgruppe durch eine elektrostatische Wechselwirkung zu einem besseren Protonenakzeptor. Die Spaltung einer Peptidbindung beginnt mit dem nukleophilen Angriff des Alkoxidions auf das Carbonylkohlenstoffatom des Substrates. Es folgt eine Reihe von Reaktionsschritten, an denen neben Ser 195 und His 57 auch ein Wassermolekül beteiligt ist. Die Nettoreaktion entspricht einer Hydrolyse. Für weitergehende Details zum Reaktionsmechanismus sei hier auf Lehrbücher der Biochemie verwiesen.⁶³

Viele Serinproteasen weisen große Sequenz- und Strukturähnlichkeit zueinander auf. So stimmen die Sequenzen der homologen Verdauungsenzyme Trypsin und Chymotrypsin zu 40% überein. Die Wurzel der mittleren quadratischen Abweichung (*RMSD*, für engl. „Root Mean Square Deviation“) der α -Kohlenstoffatome der überlagerten 3D-Strukturen beträgt nur 1,7 Å. Trotzdem besitzen beide Enzyme eine unterschiedliche Substratspezifität. So spaltet Trypsin nach langen, positiv geladenen Seitenketten, also

Arginin (Arg) und Lysin (Lys). Chymotrypsin hingegen spaltet bevorzugt nach großen, aromatischen Aminosäureresten wie Phenylalanin (Phe), Tyrosin (Tyr) und Tryptophan (Trp). Der Grund dafür sind Unterschiede in der Struktur einer bestimmten Bindetasche an der Oberfläche der Enzyme. Bei Chymotrypsin wird diese Substratspezifitätstasche S_1 am Grund durch die Aminosäure Serin 189 ausgekleidet. Bei Trypsin ist Ser 189 am Grund der Tasche durch die negativ geladene Aminosäure Asp 189 ausgetauscht. Daher werden hier positiv geladene Seitenketten bevorzugt. Für den rationalen Entwurf eines spezifischen Serinproteaseinhibitors sollte also in erster Linie auf eine möglichst hohe Komplementarität zur S_1 -Tasche geachtet werden. Daneben gibt es weitere Taschen, die von einem spezifischen Inhibitor ausgefüllt werden könnten (Abbildung 2-1). Im Anschluss werden die für diese Arbeit relevante Serinproteasen vorgestellt (Tabelle 2-1).

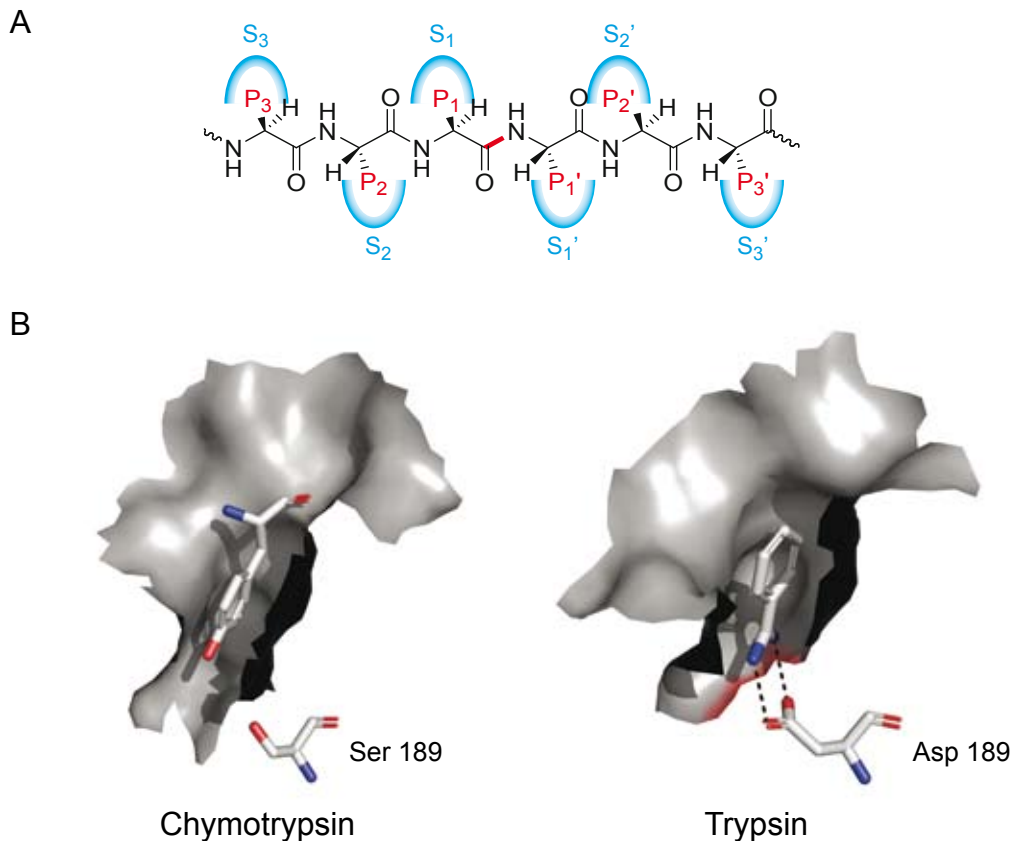


Abbildung 2-1: Nomenklatur der Protease-Substrat-Wechselwirkungen (A) und Vergleich der Bindetasche von Chymotrypsin und Trypsin (B). A: Stellen, an denen das Substrat in Wechselwirkung tritt sind mit „P“ (rot) bezeichnet. Die entsprechenden Bindetaschen der Protease sind mit „S“ (blau) bezeichnet. Die Bezeichner auf der carboxyterminalen Seite, rechts der Spaltstelle (rote Bindung), sind mit einem Apostroph versehen. B: Die S_1 -Tasche von Chymotrypsin wird am Grund von Ser 189 ausgekleidet und bevorzugt große aromatische Seitenketten (PDB-Struktur 1ab9 im Komplex mit Tyrosin, Auflösung: 1,6 Å)⁶⁴. Bei Trypsin befindet sich am Grund der S_1 -Tasche ein Aspartat (Asp 189) weshalb lange, positiv geladene Seitenketten bevorzugt gebunden werden (PDB-Struktur 2oxs im Komplex mit Benzamidin, Auflösung: 1,32 Å)⁶⁵.

2.1.1 Chymotrypsin

Chymotrypsin ist ein Verdauungsenzym, das im Bereich des Dünndarms Proteine hydrolysiert. Es wird in den Acinuszellen des Pankreas als inaktives Zymogen (Chymotrypsinogen) gebildet und durch Trypsin proteolytisch aktiviert. Als erste Serinprotease wurde die etwa kugelförmige 3D-Struktur bereits 1967 von David Brown in Cambridge, England aufgeklärt. Das etwa 25 kDa schwere Enzym besteht aus drei Ketten, die über zwei Disulfidbrücken miteinander verbunden sind. Es dient als diagnostischer Nachweis im Stuhl bei Pankreaserkrankungen. Erniedrigte Werte sind z.B. bei chronischer Pankreatitis mit exokriner Pankreasinsuffizienz zu beobachten. In der Medizin wird Chymotrypsin als Antiphlogistikum angewendet.^{8,63,66}

2.1.2 Faktor X_a

Faktor X_a (Stuart-Prower-Faktor, FX_a) ist ein Enzym der Gerinnungskaskade. Diese Serinprotease wird als Zymogen (Faktor X) in der Leber synthetisiert. Die proteolytische Aktivierung kann sowohl über den intravaskulären Pfad als auch über den extravaskulären Pfad der Gerinnungskaskade erfolgen. Die Aktivierung des intravaskulären Systems erfolgt, wenn bei einer Verletzung des Endothels in den Blutgefäßen anionische Oberflächen präsentiert werden. Faktor X wird in der Folge von Faktor IX_a (im Komplex mit Faktor VIII_a) proteolytisch aktiviert. Das extravaskuläre System wird in Gang gesetzt, wenn Gewebeverletzungen vorliegen. Die Aktivierung von Faktor X erfolgt hier durch einen Komplex aus dem Gewebefaktor und Faktor VII_a. Faktor X_a aktiviert durch proteolytische Spaltung Thrombin aus dessen Proenzym Prothrombin. Thrombin wiederum aktiviert Fibrin, was letztendlich zur Bildung von Gerinnseln führt. Faktor X_a ist ein globuläres Glykoprotein, das aus einer leichten und einer schweren Kette besteht und ein Molekulargewicht von ca. 58 kDa besitzt. Eine Faktor X_a-Defizienz führt zu einer Beeinträchtigung der Blutgerinnung, was schweren Blutungen zur Folge haben kann (vgl. hämorrhagische Diathese und Stuart-Prower-Krankheit). Eine unkontrollierte Aktivierung kann Thrombosen und Infarkte auslösen. Pharmakologisch ist Faktor X_a ein Angriffsziel von Antikoagulationen.^{63,66,67}

2.1.3 Trypsin

Das Verdauungsenzym Trypsin wird als inaktive Vorstufe (Trypsinogen) im Pankreas gebildet. Die proteolytische Aktivierung erfolgt im Duodenum durch das Enzym Enteropeptidase. Dieser Schritt initiiert eine sich selbst verstärkende Kaskade, in deren Verlauf alle Zymogene des Pankreas aktiviert werden. Trypsin wird daher spezifisch mit Hilfe des Pankreas-Trypsininhibitors kontrolliert, welcher mit einer Dissoziationskonstante von 0,1 pM sehr fest an Trypsin bindet. Trypsin ist das

Hauptenzym der enteralen Eiweißverdauung und spaltet bei pH 7–8 Polypeptide im Dünndarm. Das annähernd kugelförmige Enzym besteht aus einer einzigen Polypeptidkette mit einer Masse von 24 kDa. Bei zu hohen Konzentrationen von Trypsin kann das körpereigene Gewebe angegriffen und so Nekrosen und Pankreatitis auslösen. Bei Neugeborenen mit Mukoviszidose kann ein Trypsinmangel zu *Meconium Ileus* führen. In der Medizin wird Trypsin therapeutisch bei Entzündungs- und Schwellungszuständen angewendet.^{63,66,68}

2.1.4 Tryptase

Tryptase ist das häufigste Protein, das in den Granuli der Mastzellen gespeichert wird. Mastzellen spielen eine Rolle bei Hypersensibilität, einschließlich allergischer Reaktionen. Mastzellen werden auch im Zusammenhang mit Autoimmunerkrankungen wie Arthritis und Multipler Sklerose diskutiert. Mastzellen speichern in ihren Granuli eine Reihe verschiedener Proteine, die durch einen Auslöser sezerniert werden. Tryptasen werden durch Heparin stabilisiert und liegen in ihrer aktivierten Form in den Granuli vor, sind jedoch aufgrund des sauren pH-Wertes von ca. 5,5 reaktionsträge (Tryptase weist ein pH-Optimum im neutralen Bereich auf). Die Protease besitzt eine tetramere Struktur mit einem Molekulargewicht von 30–36 kDa je Untereinheit. Die vier Monomere sind auf den Ecken eines quadratischen Rahmens angeordnet, wobei die aktiven Zentren alle in Richtung einer zentralen, abgeschirmten Pore zeigen. Man nimmt an, dass Tryptasen eine Rolle bei der Regulierung von Entzündungsreaktionen spielen. Insbesondere wird ein Einfluss auf die allergische Antwort von Atemwegsentzündungen wie Asthma diskutiert. Es konnte außerdem gezeigt werden, dass Tryptasen einen mitogenen Effekt auf Fibroblasten haben, was Fibrosen auslösen könnte.⁶⁹

2.1.5 Urokinase-Typ-Plasminogenaktivator

Der Urokinase-Typ-Plasminogenaktivator (*uPA*, auch: Urokinase) wird als inaktives Proenzym (Prourokinase) in den Nieren gebildet. Nach seiner Aktivierung durch Plasmin (und anderen Proteasen) konvertiert uPA das Zymogen Plasminogen proteolytisch in Plasmin. Plasmin ist eine weitere Serinprotease, welche Fibrin hydrolysiert und damit Blutgerinnsel abbauen kann. Außerdem spaltet Plasmin Proteine der extrazellulären Matrix, aktiviert Proteasen und setzt Wachstumsfaktoren frei. uPA besteht aus zwei mit einer Disulfidbrücke verbundenen Ketten und hat ein Molekulargewicht von 53 kDa. Die Protease besitzt drei Domänen. Eine trypsinähnliche Domäne trägt die enzymatische Aktivität und eine weitere fungiert als Bindedomäne für den uPA-Rezeptor (uPAR). Die Bildung eines uPA-uPAR-Komplexes an der Zelloberfläche hat zudem einen positiven Effekt auf Zellmigration, -adhäsion und Chemotaxis.

Die proteolytischen, zellmigratorischen und proliferatorischen Effekte von uPA werden von Tumoren ausgenutzt. Vor allem ein Effekt auf Tumormigration und Metastase wird diskutiert. Tumorzellen sezernieren zudem weit größere Mengen an Urokinase als gesunde Zellen. In der Medizin wird uPA bei der Fibrinolysetherapie als thrombolytische Substanz bei kardiovaskulären Erkrankungen und Gehirnschlag eingesetzt. Ein potentes neues Einsatzgebiet ist zudem die Krebstherapie. Allgemein scheint ein erhöhter uPA-Spiegel einen gravierenderen Effekt auf den Organismus zu haben, als ein erniedrigter.^{70,71}

Tabelle 2-1: Zusammenstellung fünf in dieser Studie betrachteten Serinproteasen.

Enzym	Syntheseort	Spaltstelle ^{72,68}	Funktion	Medizinische Bedeutung
Chymotrypsin	Pankreas	Tyr-X, Trp-X, Phe-X, Leu-X	Verdauung	Pankreatitis, Anw. als Antiphlogistikum
Faktor X _a	Leber	Arg-Thr, Arg-Ile	Blutgerinnung	Blutungen (z.B. hämorrhagische Diathese, Stuart-Prower-Krankheit), Thrombosen, Infarkt (Ziel von Antikoagulantien)
Trypsin	Pankreas	Arg-X, Lys-X	Verdauung	Gewebenekrose, Pankreatitis, Meconium Ileus bei Mukoviszidose. Anw. bei Entzündungs- u. Schwellungszuständen.
Tryptase	Mastzellen	Arg-X, Lys-X	Entzündungsreaktion	Allergische Reakt. (z.B. Asthma), Autoimmunerkrank. (z.B. Arthritis), Fibrose
Urokinase (uPA)	Niere	Arg-Val	Thrombolyse, Abbau der extrazellulären Matrix	Krebs (Metastase von Tumorzellen), Anw. in der Fibrinolysetherapie (z.B. bei kardiovaskulären Erkrank. u. Hirnschlag)

2.2 Das menschliche Immunschwächevirus und AIDS

Das menschliche Immunschwächevirus (*HIV*, auch „Humanes Immundefizienzvirus“) ist ein Retrovirus aus der Gattung der Lentiviren. Eine Infektion führt nach langer, meist mehrjähriger Inkubationsphase zur bislang unheilbaren Immunschwächekrankheit AIDS (Erworbenes Immunschwächesyndrom, für engl. „Acquired Immune Deficiency Syndrome“).⁷³ Nach Schätzungen von UNAIDS, dem gemeinsamen Programm der Vereinten Nationen für HIV/AIDS und der Weltgesundheitsorganisation (*WHO*, für engl. „World Health Organisation“) lebten im Jahr 2007 33,2 Millionen Menschen

mit einer HIV-Infektion. Im gleichen Jahr infizierten sich 2,5 Millionen Menschen neu mit HIV und 2,1 Millionen starben an AIDS.⁷⁴ In Deutschland lebten 2007 laut Schätzungen des Robert Koch Instituts rund 59.000 Menschen mit HIV; es kam zu rund 3000 Neuinfektionen und 650 Todesfällen in Folge von AIDS.⁷⁵ Insgesamt wird die Zahl der Todesfälle seit Beginn der Epidemie auf über 20 Millionen weltweit und 27.000 in Deutschland geschätzt. Seit der Entdeckung des HI-Virus im Jahr 1981 hat sich AIDS damit zu einer Pandemie entwickelt.⁷⁶ Einer Studie zufolge, die die weltweiten Mortalitätsraten für verschiedene Todesursachen bis ins Jahr 2030 projizierte, wird die Zahl der HIV/AIDS-bedingten Todesfälle bis 2030 zunehmen, wohingegen die Mortalitätsraten anderer Infektionskrankheiten sinken werden (Abbildung 2-2).⁷⁷ Von 2001 bis 2006 ist die Zahl von HIV-Neuinfektionen in Deutschland um 80% gestiegen.⁷⁸ Die Bekämpfung von HIV und AIDS ist eines der herausragenden Ziele unserer Zeit.^{74,79} Erste Erfolge dieser Anstrengungen lassen sich bereits verzeichnen. So stagniert die weltweite HIV-Prävalenz (Prozentsatz der infizierten Personen) seit 2001 auf einem Wert von 0,8%. Die Gründe für diese Stagnation liegen in einem verbesserten Zugang zu Medikamenten, besseren Behandlungsmöglichkeiten und der einsetzenden Wirkung von Präventivprogrammen.⁷⁴ Aufgabe der Wissenschaft wird es sein – gerade im Hinblick auf die hohe Mutationsrate von HIV und der einhergehenden schnellen Ausbildung von Resistenzen – neue Therapieansätze im Kampf gegen das Virus zu finden.

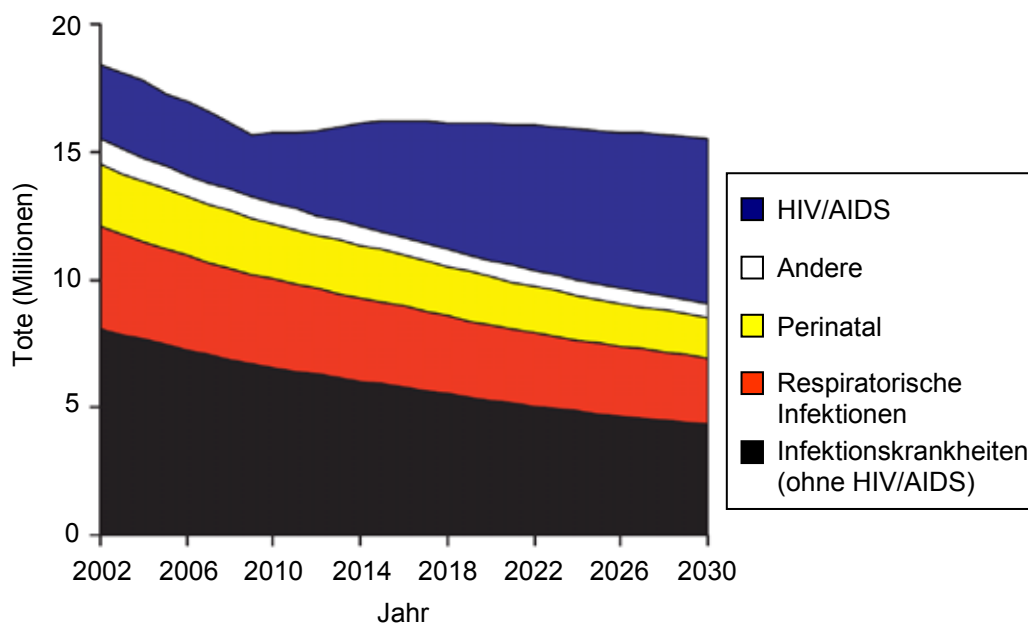


Abbildung 2-2: Projektion der weltweiten Todesfälle durch Infektionskrankheiten, perinatale und maternale Bedingungen, sowie Ernährungsdefizite. Der Projektion liegt ein Grundlinienszenario zu Grunde, demnach die globale Abdeckung mit der antiretroviralen Therapie bis 2012 80% erreicht und danach konstant bleibt. Abbildung nach Mathers und Loncar, 2006.⁷⁷

2.3 Der Lebenszyklus des HI-Virus

Das HI-Virion besitzt einen Durchmesser von ca. 100 nm, ist annähernd kugelförmig und von einer Lipidmembran umgeben (Abbildung 2-3).⁸⁰ In diese Membran integriert sind Komplexe des externen, membranständigen Glykoproteins gp120 und dem Transmembranprotein gp41. Im Inneren an die Lipidmembran assoziiert folgt eine Matrix aus Proteinen. Im Lumen des Virions befindet sich das Viruscapsid, welches das Virusgenom umschließt. Das Genom besteht aus zwei Kopien einzelsträngiger Plus-RNA. Neben dem viralen Genom finden sich im Capsid noch die Enzyme Reverse Transkriptase (RT) und Integrase sowie einige akzessorische Proteine. Die HIV-Protease befindet sich außerhalb des Capsids.⁸⁰

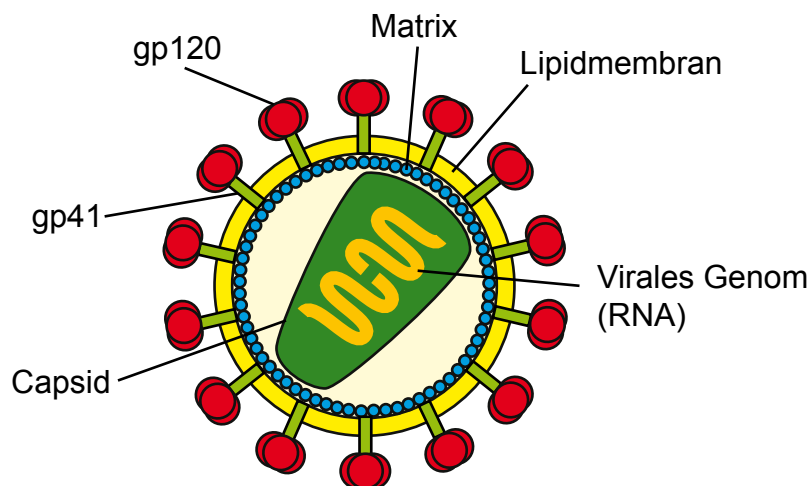


Abbildung 2-3: Schematische Darstellung des HI-Virions.

Zielzellen des Virus sind CD-4-Rezeptor-tragende Zellen der Immunabwehr, wie den T-Lymphozyten (T-Helferzellen), Monozyten, Makrophagen und dendritischen Zellen. Zu Beginn einer Infektion nimmt das Virus durch seine gp120-Rezeptoren Kontakt zu CD4-Rezeptoren einer Zielzelle auf.⁸¹⁻⁸⁵ Die Adhäsion wird durch die Chemokinrezeptoren CCR5 (monozytäre Zellen) und CXCR4 (T-Lymphozyten) unterstützt. Es kommt zur Fusion der Membranen und zum Eintritt des Post-Eintrittskomplexes. Nach teilweisem Verlust des Capsids wird das RNA-Genom durch RT in doppelsträngige DNA umgeschrieben. Es entsteht ein Präintegrationskomplex, der in den Zellkern geschleust wird. Dort wird das virale Genom mit Hilfe des Enzyms Integrase in das Wirtsgenom integriert. Nach Aktivierung erfolgt die Transkription viraler RNA vom Promotor im LTR-Bereich (für engl. „Long Terminal Repeats“). In der frühen Transkriptionsphase schreitet die RNA-Synthese nicht effektiv voran.

Zur effektiven Transkription bedarf es der Antitermination durch den Transaktivator der Transkription (Tat), einem viralen Protein. Durch die Bindung von Tat an TAR-RNA (Transaktivierungs-Antwortelement, für engl. „Transactivation Response Element“) wird ein prozessiver Elongationskomplex gebildet, durch den die virale RNA vollständig synthetisiert wird (vgl. Abschnitt 2.5). Die Bindung eines weiteren viralen Proteins, Rev, an das Rev-Antwortelement (RRE, für engl. „Rev Response Element“) der viralen RNA leitet die späte Transkriptionsphase ein. Rev bewahrt das Virusgenom vor dem Spleißen und geleitet es aus dem Zellkern ins Zytoplasma. Dort kommt es durch Interaktion viraler Proteine, Vorläuferproteine und viralen Genoms zur Selbstassemblierung viraler Partikel. Diese schnüren sich von ihrer Wirtszelle ab und bilden nach weiterer Reifung neue, infektiöse HI-Viren. An dieser Reifung ist unter anderem das Enzym Protease beteiligt.⁸¹⁻⁸⁵

2.4 Antiretrovirale Therapie

Hat man sich einmal mit HIV infiziert, ist die Krankheit bis heute unheilbar. Fortschritte in der antiretroviralen Therapie (ART) haben jedoch die Lebensqualität und Lebenserwartung von Infizierten erheblich gesteigert. Heute üblich ist eine Kombinationstherapie verschiedener Präparate im Rahmen der Hochaktiven Antiretroviralen Therapie (HAART).⁸⁶ Die heute zugelassenen Medikamente gliedern sich in vier Gruppen:

- nukleosidische und nukleotidische Reverse-Transkriptase-Inhibitoren (NRTI)
- nicht-nukleosidische Reverse-Transkriptase-Inhibitoren (NNRTI)
- Proteaseinhibitoren (PI)
- Fusionsinhibitoren

NRTI und NNRTI verhindern durch die Inhibition der RT ein Umschreiben des RNA-Genoms in seine DNA-Form. PI hemmen die Aktivität der viralen Protease, wodurch Vorläuferproteine nicht gespalten werden können und nicht-infektiöse Viruspartikel gebildet werden. Fusionsinhibitoren verhindern schließlich durch Interaktion mit Oberflächenrezeptoren den Eintritt des Virus in seine Zielzellen.⁸⁶

Trotz der Fortschritte in der Therapie stellt vor allem die Ausbildung von resistenten Viren durch die hohe genetische Variabilität von HIV und die natürliche Selektion während einer Therapie ein Problem dar. Es wird deshalb ständig nach neuen Substanzen und neuen Wirkmechanismen gesucht. Ein Ziel für neue Wirkstoffe ist die Tat-TAR-Interaktion im Verlauf der Antitermination der Transkription, welche für die Replikation von HIV essentiell ist (vgl. Abschnitt 2.5).^{60,61}

2.5 Die Antitermination bei HIV

Die Antitermination bei HIV erfolgt durch das Zusammenspiel verschiedener viraler und zellulärer Faktoren und ist zwingend für die Replikation von HIV notwendig. Ist die Antitermination z.B. durch die Abwesenheit des Tat-Proteins gestört, wird die Transkription viraler RNA zwar initiiert, schreitet jedoch nicht effizient voran, sondern terminiert frühzeitig. In diesem Fall werden nur kleine RNA-Fragmente synthetisiert, selten jedoch das vollständige Genom von HIV. Für eine vollständige RNA-Synthese ist die Bildung eines Elongationskomplexes erforderlich, der durch die kooperative Bindung von Tat und humanem CyclinT1 (hCycT1) an TAR-RNA beginnt (Abbildung 2-4).

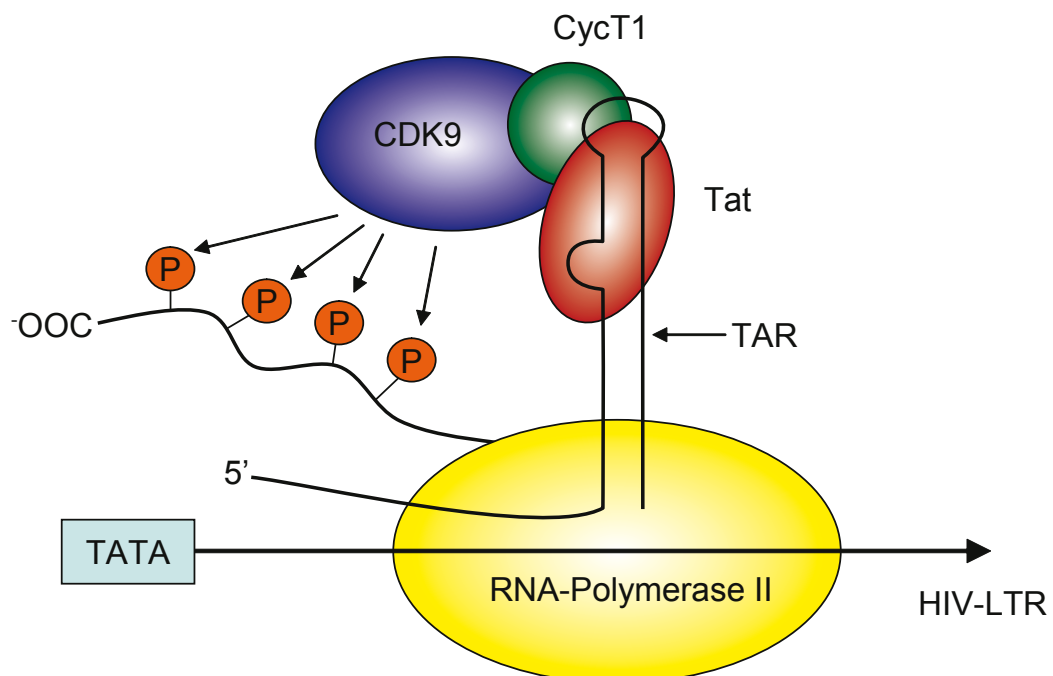


Abbildung 2-4: Schema der Antitermination bei HIV. Bei Abwesenheit von Tat ist die Transkription unvollständig und terminiert frühzeitig. Der Elongationskomplex bildet sich durch die kooperative Bindung von Tat und CycT1 an TAR-RNA. CycT1 und CDK9 bilden zusammen den Faktor pTEFb. Durch die Kinaseaktivität von CDK9 wird die C-terminale Domäne der RNA-Polymerase II hyperphosphoryliert und damit die Antitermination bewirkt. Abbildung nach Richter, 2006.⁶¹

hCycT1 ist ein regulatorischer Partner der Cyclin-abhängigen Kinase 9 (CDK9). Durch gegenseitige Wechselwirkung bilden hCycT1 und CDK9 den positiven Transkriptionselongationsfaktor (pTEFb). TAR ist eine RNA-Struktur, die zwischen den Nukleotiden +19 bis +42 am 5'-Ende naszierender, viraler RNA gelegen ist. Die Sekundärstruktur von TAR zeichnet sich durch eine Schleife aus sechs Nukleotiden und eine charakteristische Bulge (engl. „Ausbauchung“) aus drei Pyrimidinnukleotiden aus, welche über Stammstrukturen mit einander verbunden sind (Abbildung 2-5).

Das Tat-Protein interagiert über eine argininreiche Region mit TAR-RNA, und wechselwirkt mittels einer Aktivierungsdomäne mit pTEFb. Die Assemblierung des Tat-TAR-pTEFb-Komplexes am HIV-Promotor aktiviert die CDK9-Kinaseaktivität. Dadurch wird pTEFb autophosphoryliert und RNA-Polymerase II hyperphosphoryliert. Damit ist die Antitermination erreicht und vollständige RNA-Moleküle können durch den Elongationskomplex synthetisiert werden.^{60,61,82}

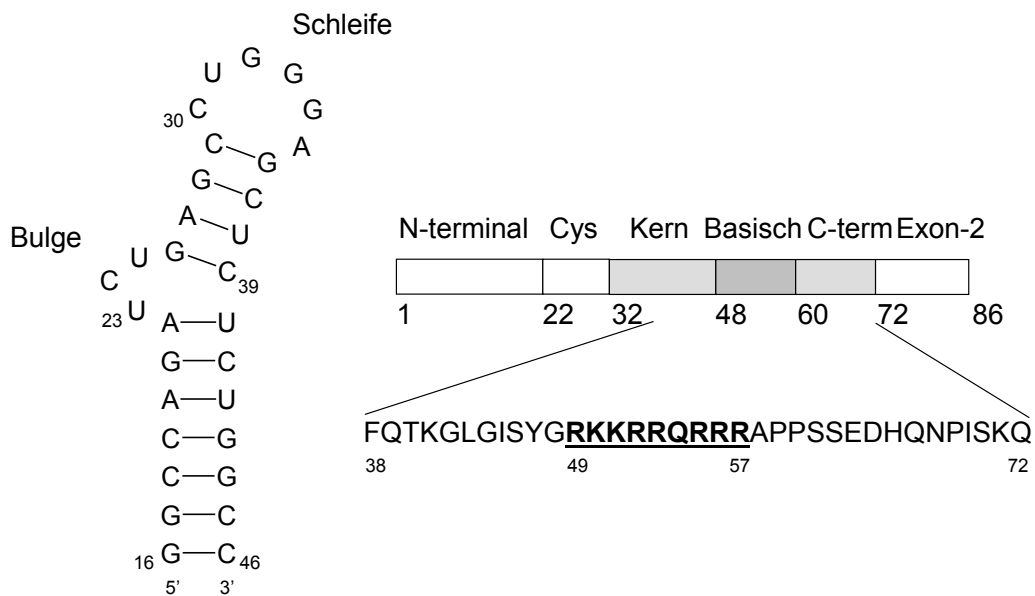


Abbildung 2-5: Sekundärstruktur von HIV-1 TAR-RNA (links) und Domänen des Tat-Proteins (rechts). Dargestellt ist der minimale Bereich von TAR, der für eine Interaktion mit Tat hinreichend ist. TAR bildet eine Haarnadelstruktur mit einer UCU-Bulge (Nuk. 23-25) und einer Schleife (Nuk. 30-35). Tat besteht aus einer N-terminalen Region, gefolgt von einer Cystein-reichen Region, der Kernregion, einer basischen Region, einer C-terminalen Region und schließlich der Region des Exons-2. Das minimale Motiv für TAR-Bindung ist im Sequenzausschnitt hervorgehoben. Die basische Region trägt ferner ein Kernlokalisierungssignal. Abbildung nach Richter, 2006.⁶¹

Durch ihre zentrale Rolle im Replikationszyklus von HIV ist die Interaktion von Tat mit TAR-RNA ein Angriffsziel für neue Wirkstoffe in der HIV-Therapie. Tat ist nicht nur für die Expression viraler Gene während der exponentiellen Wachstumsphase des Virus notwendig, sondern auch während der Aktivierung von integrierten proviralen Genomen, die potentiell resistente Stämme hervorbringen können.⁶⁰

3 Material und Methoden

3.1 Fluoreszenzbasierte Bioaktivitätsbestimmung

Die Fähigkeit potentieller Liganden, die Wechselwirkung zwischen TAR-RNA und Tat zu inhibieren, wurde mit Hilfe einer fluoreszenzbasierten Untersuchung bestimmt. Das Verfahren macht sich den Fluoreszenzresonanzenergietransfer (FRET) zu nutze, bei dem die Energie eines angeregten Donorfluorophors strahlungsfrei auf einen Akzeptorfluorophor übertragen wird. Der FRET-Effekt wurde 1946 von Theodor Förster entdeckt.⁸⁷ Die Effizienz des Energietransfers E ist stark vom Abstand der Fluorophore abhängig (Gl. 3-1):

$$E = \frac{R_0^6}{R_0^6 + r^6} \quad (3-1)$$

R_0 bezeichnet die Förster-Distanz, bei der 50% der Energie übertragen wird und r ist der Abstand von Donor und Akzeptor.

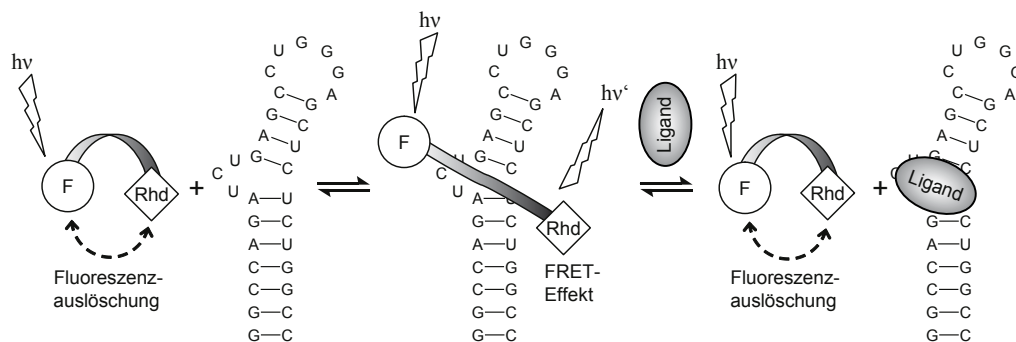


Abbildung 3-1: Schema der FRET-basierten Aktivitätsmessung. Im freien Zustand bildet ein Fluorescein (F)- und Rhodamin (Rhd)-markiertes Peptid ein intramolekulares Dimer, wodurch es zur Fluoreszenzauslöschung kommt. Bindet das markierte Peptid an TAR-RNA, so besitzen die beiden Fluorophore einen Abstand, der einen FRET-Effekt ermöglicht. Ein konkurrierender Ligand ist in der Lage das gebundene Peptid zu verdrängen und damit den Energietransfer zu unterbinden. Abbildung nach Matsumoto, 2000.⁸⁸

Die Bioaktivitätsbestimmung basiert auf einer modifizierten Version des von Matsumoto *et al.* (2000) beschriebenen Protokolls.⁸⁸ Hierbei wird ein Tat-Peptidfragment am N-Terminus mit Fluorescein (F, Fluoreszenzdonor) und am C-Terminus mit Rhodamin (Rhd, Fluoreszenzakzeptor) markiert. In Lösung kommen die Fluorophore dieses Konstrukts in räumliche Nähe (sie bilden ein „intramolekulares Dimer“), wodurch es zu Fluoreszenzauslöschung kommt (Abbildung 3-1).⁸⁹ Interagiert das markierte Peptid jedoch mit TAR-RNA, so bindet es in einer gestreckten Konformation im

Bereich der TAR-Bulge. Im gebundenen Zustand besitzen Donor und Akzeptor einen günstigen Abstand für FRET. Wird nun Licht einer passenden Wellenlänge auf den Donor eingestrahlt, so gibt dieser ein Teil der Energie strahlungsfrei an den Akzeptor weiter, der die Energie in Form von Licht einer anderen Wellenlänge wieder abstrahlt. Gibt man einen kompetitiven Liganden zum System hinzu, so konkurriert dieser mit Tat um die Bindung an TAR. Je stärker die Affinität des Liganden zu TAR ist, desto mehr Tat-Peptid wird aus der Bulge verdrängt. Die Fluorophore der verdrängten Tat-Peptide kommen in Lösung wiederum in räumliche Nähe und es kommt zur Fluoreszenzauslöschung. Die Abnahme des FRET-Effekts ist folglich ein Indikator für die Affinität eines potentiellen Liganden.

In dieser Arbeit wurde ein farbstoffmarkiertes Tat-Peptid, das die minimale TAR-bindende Sequenz aus der basischen Domäne von Tat enthält (Tat₄₉₋₅₇-Sequenz, vgl. Abbildung 2-5) verwendet: Fluorescein-AAARKKRRQRRRAAC-Rhodamin (FTatRhd, Thermo Electron Corporation, Ulm). Die Fluorophore Fluorescein und Tetramethyl-Rhodamin sind jeweils über einen Abstandshalter aus drei Einheiten Alanin mit dem Peptid verknüpft (Abbildung 3-2).

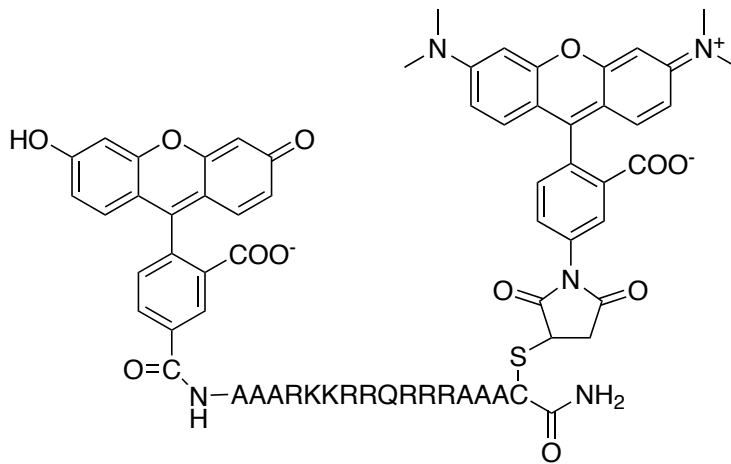


Abbildung 3-2: Struktur des fluorophormarkierten Tat-Peptids FTatRhd.

3.2 Experimentelle Durchführung der Aktivitätsbestimmung

Die fluoreszenzbasierte Bestimmung der Rezeptorbindung (engl. „Binding assay“) wurde in 96er-Mikrotiterplatten (Corning 6860, schwarz, antihafbeschichtet) bei physiologischen Bedingungen (37°C, pH 7,4) durchgeführt. Die Platten wurden mit einem Mikrotiterplattenleser (Safire², Tecan, Crailsheim) bei einer Anregungswellenlänge von 489 nm und einer Emissionswellenlänge von 590 nm gemessen. Nukleasfreie Materialien wurden während der Handhabung von RNA verwendet. TAR-RNA (HIV-1, BioSpring, Frankfurt) und farbstoffmarkiertes Tat₄₉₋₅₇-Peptid (FTatRh_d, Thermo Electron Corporation, Ulm) wurden bei einer Konzentration von 10 nM in einem finalen Volumen von 100 µL in TK-Puffer (50 mM Tris-HCl, 20 mM KCl, 0,01% Triton-X 100, pH 7,4) verwendet. Vor den Analysen wurde die RNA (100 nM in 0,25x TK-Puffer) für 5 min. auf 90°C erhitzt und dann sofort 5 min. lang auf Eis gestellt (Rehybridisierung). Die Emission des Hintergrundes (nur TK-Puffer), des reinen Peptids (minimale Emission) und des Tat-TAR-Komplexes (maximale Emission) wurde zuerst bestimmt. Zur Abschätzung der inhibitorischen Konzentration potentieller Liganden wurden dann Messungen bei einer Substanzkonzentration von 100 µM durchgeführt. Der relative Effekt, also die relative Fluoreszenzemission, wurde nach Gl. 3-2 berechnet:

$$\%Effekt = \frac{RFU_{Pot.Inhibitor} - RFU_{Tat-Peptid}}{RFU_{Tat-TAR-Komplex} - RFU_{Tat-Peptid}} \cdot 100\% \quad (3-2)$$

wobei die Fluoreszenzemission in relativen Fluoreszenzeinheiten (RFU, für engl. „Relative Fluorescence Unit“) gemessen wurde. Liganden mit einem %Effekt kleiner 100% wurden in Wiederholungsmessungen bei einer Konzentration von 100 µM oder 500 µM validiert oder falsifiziert. Die inhibitorische Konzentration, bei der 50% des Tat-Peptids vom Liganden verdrängt werden (IC_{50}), wurde schließlich durch eine Verdünnungsreihe von hohen zu niedrigen Konzentrationen bestimmt. An die entstehende Dosis-Wirkungs-Kurve wurde mit Hilfe des Programms Sigmaplot (Version 9.1, SyStat Software, Inc., San Jose, USA)⁹⁰ eine logistische Vierparameterfunktion angepasst Gl. 3-3:

$$\%Effekt = D + \frac{A - D}{1 + 10^{(x - \log C)B}} \quad (3-3)$$

mit dem Logarithmus der Ligandenkonzentration x , dem maximalen Plateauwert A (z.B. 100%), der Kurvensteigung B , dem IC_{50} C und dem minimalen Plateauwert D (z.B. 0%). In Abbildung 3-3 ist exemplarisch eine Dosis-Wirkungs-Kurve des unspezifischen TAR-RNA-Liganden Tripel-Arginin (RRR) dargestellt. Die Abszisse ist logarithmisch aufgetragen. Der quadratische Korrelationskoeffizient der Funktionsanpassung (engl. „Fitting“) ist $R^2 = 0,9987$. Der bestimmte IC_{50} ist $10 \mu\text{M}$, das 95%-Konfidenzintervall beträgt $8 - 14 \mu\text{M}$. Die Funktionsanpassung wurde mit Sigmaplot (Version 9.1), SyStat Software, Inc.)⁹⁰ durchgeführt.

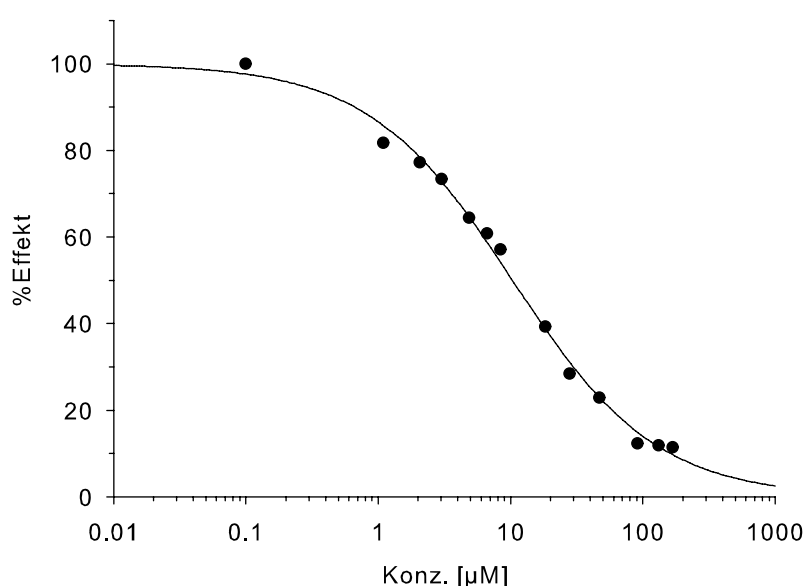


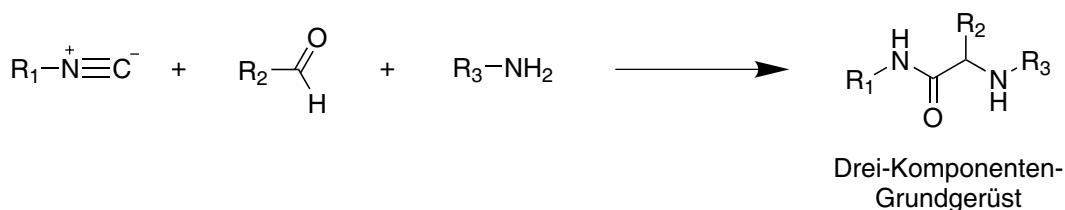
Abbildung 3-3: Dosis-Wirkungs-Kurve des unspezifischen TAR-RNA-Liganden Tripel-Arginin (RRR). Die Abszisse ist logarithmisch aufgetragen. Der quadratische Korrelationskoeffizient der Funktionsanpassung ist $R^2 = 0,9987$. Der bestimmte IC_{50} ist $10 \mu\text{M}$, das 95%-Konfidenzintervall beträgt $8 - 14 \mu\text{M}$. Die Funktionsanpassung wurde mit Sigmaplot (Version 9.1, SyStat Software, Inc.)⁹⁰ durchgeführt.

Alle Einzelpunktmessungen wurden in dreifacher Wiederholung durchgeführt. Substanzen, die Licht mit einer Wellenlänge von 489 nm absorbierten, wurden zusätzlich bei alternativen Bedingungen gemessen (Anregungswellenlänge: 550 nm , Emissionswellenlänge: 570 nm). Hier wird der Fluorophor Rhodamin direkt angeregt, der ein Extinktionsmaximum von 554 nm und ein Emissionsmaximum von 573 nm aufweist (Quelle: Fluorophore Reference Guide der Firma Biorad). Der gemessene Effekt beruht hier nicht auf einem FRET-Effekt, sondern auf dem Absorptions- und Emissionsverhalten des Fluorophors Rhodamin und der Fluoreszenzauslöschung des freien Tat-Peptids in Lösung.

3.3 Datensätze

3.3.1 Ugi-Bibliothek

Für die vorliegende Studie wurde mit vier verschiedenen molekularen Datensätzen gearbeitet. Der erste Datensatz – von nun an Ugi-Bibliothek genannt – ist eine kombinatorische Molekülbibliothek, die aus 15.840 Produkten einer Ugi-artigen Dreikomponentenreaktion^{91,92} besteht (Schema 3-1). Die Bibliothek wurde von Dr. Lutz Weber und Mitarbeitern (Roche, Basel) nach dem Prinzip einer „Eintopfreaktion“ (Mehrkomponentenreaktion) aus 15 Aminen, 44 Aldehyden und 24 Isonitrilen synthetisiert und uns freundlicherweise zur Verfügung gestellt.^{27,93} Alle Ugi-Produkte waren auf Inhibition der fünf Serinproteasen Chymotrypsin, Faktor X_a, Trypsin, Trypsinase und des Urokinase-Typ-Plasminogenaktivators (uPA) getestet worden. Diese Daten lagen vor.⁹³ Die inhibitorische Konzentration (IC_{50}), bei der eine Substratspaltungsaktivität von 50% im Vergleich zur ungehemmten Reaktion vorlag, wurde durch *In-vitro*-Experimente für jedes einzelne Enzym bestimmt.⁹³ Die Experimente besaßen eine obere Bestimmungsgrenze von 110 μM bei Chymotrypsin, Faktor X_a, Trypsinase und uPA, bzw. 100 μM im Falle von Trypsin. Moleküle mit einem $IC_{50} \geq 100 \mu\text{M}$ wurden als inaktiv deklariert. Eine Übersicht der Anzahl aktiver Verbindungen und ihrer IC_{50} -Wertebereiche für jede Serinprotease ist in Tabelle 3-1 zusammen gestellt. Aufgrund der Annotierungen mit Inhibitionskonzentrationen für fünf Serinproteasen und ihrer Größe von rund 16.000 Verbindungen ist die Ugi-Bibliothek ideal zur Anwendung für QSAR-Modellierung und verwandte Disziplinen geeignet. Der Datensatz wurde im Rahmen dieser Arbeit für retrospektive Experimente zum Design fokussierter Substanzbibliotheken verwendet (vgl. Kapitel 5).



Schema 3-1: Reaktionsschema einer Ugi-artigen Drei-Komponenten-Reaktion. Ein Isonitril, ein Aldehyd und ein Amin reagieren in einer Mehrkomponentenreaktion zu einem Drei-Komponenten-Produkt. Diese Reaktion findet in der kombinatorischen Chemie Anwendung.⁹⁴

Tabelle 3-1: Anzahl aktiver Verbindungen der Ugi-Bibliothek (erste Reihe) und Bereiche der IC_{50} -Werte der Verbindungen für jede Serinprotease (zweite bis vierte Reihe).^a

	Chymotrypsin	Faktor X _a	Trypsin	Tryptase	uPA
Aktive	34 (0,2%)	1703 (10,8%)	305 (1,9%)	4723 (29,8%)	1390 (8,8%)
Median- IC_{50} [μ M]	37,2 (4,1 – 72,0)	47,9 (33,9 – 67,6)	67,4 (50,0 – 84,1)	38,0 (27,5 – 52,5)	67,1 (52,4 – 85,4)
Min. – max. IC_{50} [μ M]	0,09 – 110	0,16 – 110	3,80 – 100	0,09 – 110	0,37 – 110
Mittlerer Top-20- IC_{50} [μ M]	13,7	2,3	17,7	0,6	10,4

^a Die zweite Reihe zeigt den Median aktiver Verbindungen. Zahlen in Klammern geben das untere und obere Quartil der IC_{50} -Verteilung an. Die dritte Reihe gibt den minimalen und maximalen IC_{50} an. Die vierte Reihe zeigt den mittleren IC_{50} der 20 aktivsten Verbindungen für jedes Enzym. Dieser Wert repräsentiert das globale Optimum einer fokussierten Bibliothek mit 20 Molekülen.

3.3.2 COBRA - Sammlung bioaktiver Referenzanaloga

Als zweiter Datensatz fand die Sammlung bioaktiver Referenzanaloga (COBRA, Version 6.1, für engl. „Collection of Bioactive Reference Analogues“) Verwendung.⁹⁵ COBRA ist eine Sammlung biologisch aktiver Referenzverbindungen, deren Einträge handverlesen aus ausgewählten wissenschaftlichen Zeitschriften und Büchern zusammengetragen wurde. Die Sammlung ist mit detaillierten Informationen zu Zielrezeptoren sowie mit verschiedenen Aktivitätsdaten annotiert und beinhaltet wirkstoff- und leitstrukturartige Verbindungen. Die hier verwendete Version 6.1 enthielt 8311 Verbindungen zu über 300 verschiedenen pharmakologischen Zielrezeptoren, darunter 228 (2,7%) Inhibitoren der für diese Arbeit relevanten Serinprotease Faktor X_a. COBRA wurde im Rahmen dieser Arbeit für das Design fokussierter Substanzbibliotheken (vgl. Abschnitt 5.5), zur Häufigkeitsanalyse funktioneller Gruppen (vgl. Abschnitt 6.10) und als Referenzbibliothek zur Kartierung des chemischen Raumes eingesetzt (vgl. Abschnitt 6.7). Die Originalversion der COBRA enthält einige Moleküldopplungen, die für einige Anwendungen nicht erwünscht waren. Im Einzelfall wurden doppelte und mehrfach vorhandene Moleküle aus der COBRA entfernt. Je nach eingesetztem Verfahren zur Erkennung von molekularer Identität enthielt die duplikatfreie COBRA 7450 Verbindungen (bearbeitet mit dem Molekülkonverter CLIFF, Version 1.14, Molecular Networks GmbH, Erlangen)⁹⁶ oder 7576 Verbindungen (bearbeitet mit der Software „Molecular Operating Environment“ (MOE), Version 2007.09, Chemical Computing Group, Montreal, Kanada).⁹⁷

3.3.3 RNA-Liganden-Bibliothek

Des Weiteren wurde eine Sammlung verschiedener RNA-Liganden verwendet. Die RNA-Liganden-Bibliothek entstand als begleitendes Projekt der RNA-Forschung der Arbeitsgruppen von Prof. Dr. G. Schneider und Prof. Dr. M. Göbel an der Goethe-Universität und wurde aus der wissenschaftlichen Literatur der letzten 15 Jahre und aus den verschiedenen RNA-Forschungsprojekten der beiden Arbeitsgruppen entnommen. Die Molekülsammlung beinhaltet zum größten Teil Liganden der HIV-1 TAR-RNA und zu einem kleineren Teil Liganden anderer RNA-Moleküle aus verschiedenen Organismen. Die Bibliothek wurde von Dr. Manuel Nietert initiiert, der die Liste getesteter Substanzen der Arbeitsgruppe Göbel mit den Ergebnissen einer Literatursuche vereinigte. Diese ursprüngliche Version wurde beständig um Liganden aus weiteren Literaturrecherchen und um Verbindungen aus den experimentellen Studien der Arbeitsgruppen Schneider und Göbel erweitert. In der letzten Fassung beinhaltete diese Sammlung 688 Verbindungen aus der Literatur und den Forschungsprojekten von (in alphabetischer Reihenfolge): Dr. Oliver Boden,⁹⁸ Dipl.-Chem. Sven Breitung, Dipl.-Ing. Alireza Givehchi, Dr. Andreas Krebs,⁹⁹ Dr. Verena Ludwig,¹⁰⁰ Dr. Manuel Nietert,¹⁰¹ Dr. Steffen Renner,¹⁰² Dipl.-Inf. Matthias Rupp, Dr. Ute Scheffer,¹⁰⁰ Dr. Gunther Seifert,¹⁰³ Dipl. Chem. Marcel Suhartono,¹⁰⁴ Dipl.-Bioinf. Yusuf Tanrikulu,¹⁰⁵ und dem Autor dieser Arbeit. Die Sammlung teilt sich in 267 aktive Substanzen und 421 inaktive Substanzen auf. Da die RNA-Liganden-Bibliothek im Laufe dieser Arbeit beständig erweitert wurde, kam jeweils die zum aktuellen Stand verfügbare Version zum Einsatz. Die Sammlung wurde im Rahmen dieser Arbeit als Referenzbibliothek zur Kartierung des chemischen Raumes (vgl. Abschnitt 6.7) und als Teil der Initialpopulation für eine prospektive, evolutionäre Optimierung einer fokussierten Substanzbibliothek eingesetzt (vgl. Abschnitt 6.3).

3.3.4 Specs-Substanzkatalog

Die Substanzbestellungen für die Validierungsexperimente des prospektiven Teils dieser Arbeit wurden beim Chemikalienanbieter Specs (Delft, Niederlande) getätigt.⁶² Specs bietet einen Chemikalienkatalog mit rund 200.000 wirkstoffartigen Screeningverbindungen für den Einsatz in der Medizinalchemie und der pharmazeutische Forschung an. Der Substanzkatalog wird monatlich aktualisiert, weshalb die Auswahl an Screeningverbindungen und deren Anzahl von Monat zu Monat leicht variiert. Für Screeningexperimente dieser Arbeit wurde der Specs-Katalog der Screeningverbindungen mit 10 mg Verfügbarkeit vom September 2006 herangezogen (191.431 Verbindungen). Substanzen wurden entweder in Pulverform oder bei einer Konzentration von 10 mM und einem Volumen von 500 μ L gelöst in DMSO auf Trockeneis geliefert. Für die

prospektive, adaptive Optimierungsstudie einer fokussierten Substanzbibliotheken wurde die Schnittmenge des Specs-Substanzkatalogs mit der RNA-Liganden-Bibliothek (Stand: Sept. 2006) aus dem Specs-Datensatz entfernt. Danach wurden zehn Verbindungen manuell aus der wissenschaftlichen Literatur und der RNA-Liganden-Bibliothek für eine Initialpopulation einer evolutionären Optimierungsstrategie ausgewählt und zum Specs-Datensatz hinzugefügt (vgl. Abschnitt 6.3). Der finale 191.405 Screeningverbindungen umfassende Datensatz stellte den Suchraum für die prospektive, adaptive Optimierungsstudie dar (vgl. Kapitel 6).

3.3.5 Präprozessierung der Substanzbibliotheken

Die Moleküle aller Bibliotheken wurden grundsätzlich vor ihrer eigentlichen Verarbeitung präprozessiert. Dieser Prozess diente der Standardisierung der Computerrepräsentation aller Verbindungen. Die Präprozessierung erfolgte mit Hilfe des Molekülkonverters CLIFF (Version 1.14, Molecular Networks GmbH, Erlangen)⁹⁶ und dem 3D-Konformationsgenerator CORINA (Version 3.20, Molecular Networks GmbH, Erlangen)¹⁰⁶ und folgte diesem Schema:

- Füge explizite Wasserstoffatome hinzu.
- Entferne Salze (entferne Gegenionen).
- Entferne Ladungen (protoniere Säuregruppen und deprotoniere basische Gruppen).
- Erstelle eine dreidimensionale Konformation mit CORINA. CORINA verfolgt dabei einen regelbasierten Ansatz mit Nachschlagelisten bevorzugter Bindungswinkel und -längen.

Für automatisierte Dockingstudien und Pharmakophorsuchen werden Moleküle benötigt, deren protonierbare Gruppen auf ca. pH 7 eingestellt sind. Daher wurde für diese Anwendungen eine alternative Präprozessierung mit MOE (Version 2007.09, Chemical Computing Group, Montreal, Kanada)⁹⁷ durchgeführt:

- Füge explizite Wasserstoffatome hinzu.
- Entferne Salze (behalte nur das größte Molekülfragment).
- Stelle die Protonierung auf ca. pH 7 ein (deprotoniere starke Säuren, protoniere starke Basen).
- Berechne Partiaalladungen (anhand des Kraftfeldes MMFF94x).
- Erstelle eine dreidimensionale, energieminierte Konformation im Vakuum (Kraftfeld: MMFF94x)

Das Kraftfeld MMFF94x ist eine in MOE enthaltene, modifizierte Fassung des publizierten Kraftfeldes MMFF94s.^{107,108}

3.4 Molekulare Repräsentationen

Eines der fundamentalen Probleme der Chemieinformatik ist die Repräsentation von Molekülen in einer für den Computer adäquaten Weise. Die exakteste heute bekannte mathematische Beschreibung chemischer Strukturen kann uns die Quantenmechanik in Form von Wellenfunktionen liefern.¹⁰⁹ Wellenfunktionen stellen Lösungen der Schrödinger-Gleichung dar – der zentralen Grundgleichung der nichtrelativistischen Quantenmechanik.¹⁰⁹ Mit Hilfe von Wellenfunktionen lassen sich viele Eigenschaften von Atomen und Molekülen sehr exakt beschreiben.⁷ Da eine exakte Lösung der Schrödinger-Gleichung für organische Moleküle nicht möglich ist, wird in der Theoretischen Chemie auf Näherungsverfahren zurückgegriffen.^{7,109} Für eine Anwendung im Wirkstoffdesign oder im virtuellen Screening sind diese Verfahren jedoch zu rechenaufwändig.⁷ In der Chemieinformatik wird zumeist mit wesentlichen Vereinfachungen bei der Beschreibung von Molekülen gearbeitet.⁷ Eine weit verbreitete Form der Repräsentation von Molekülen ist die Beschreibung durch molekulare Deskriptoren (Abbildung 3-4a).¹¹⁰ Deskriptoren beschreiben berechenbare, numerische Eigenschaften von Molekülen wie z.B. die Anzahl der Atome, das Molekulargewicht oder der Oktanol-Wasser-Verteilungskoeffizienten ($\log P$). Im Allgemeinen geben molekulare Deskriptoren die physikoschemischen und pharmakologischen Eigenschaften von Molekülen wieder. Viele solcher Eigenschaftswerte an einander gereiht ergeben einen mathematischen Vektor. Dieser Vektor ist Teil eines artifiziellen *chemischen Raumes*, dessen Ausdehnung und Dimension durch die Art und Anzahl der molekularen Deskriptoren bestimmt ist. Moleküle sind in diesem Raum durch jene Punkte gekennzeichnet, auf die ihre Deskriptorvektoren zeigen.¹¹

In der vorliegenden Arbeit wurden Moleküle zur Anwendung chemieinformatischer Methoden fast ausschließlich durch molekulare Deskriptoren repräsentiert. Diese Form der Beschreibung hat den Vorteil, dass viele in anderen Disziplinen bereits etablierte, vektorbasierte Verfahren und Konzepte direkt auf Moleküle anwendbar sind. So lässt sich z.B. die Ähnlichkeit zweier Moleküle durch einfache Messung ihres Abstandes im chemischen Raum bestimmen.¹¹¹ Dazu wird z.B. die euklidische Distanz oder die Manhattan-Distanz zwischen zwei Deskriptorvektoren berechnet (Tabelle 3-2).

Molekulare Deskriptoren haben jedoch den Nachteil, dass sie eine abstrakte Beschreibung chemischer Strukturen liefern. Eine ebenfalls weit verbreitete Art der Beschreibung, die für einen Chemiker intuitiver ist, ist der molekulare Graph (Abbildung 3-4b). Hierbei werden Atom durch Knoten und chemische Bindungen

durch Kanten eines mathematischen Graphen beschrieben.¹¹² Den Knoten (Atomen) und Kanten (Bindungen) können verschiedene Eigenschaften zugewiesen werden (z.B. Elementsymbol und Bindungstyp). Diese Form der Beschreibung hat den Vorteil, dass sich Konzepte der Graphentheorie ohne weiteres auf Molekülgraphen übertragen lassen. So kann z.B. die Ähnlichkeit zweier Moleküle durch Bestimmung ihrer maximalen Knoten- bzw. Kantenüberlagerung (maximaler Subgraphenisomorphismus) der zwei Molekülgraphen bestimmt werden. Nachteilig ist, dass die Suche nach Subgraphenisomorphismen rechenaufwendig sein kann, da dieses Problem NP-vollständig ist.¹¹³ Molekulare Graphen werden in chemieinformatischen Programmen oft in Form einer Verbindungstabelle (engl. „Connection table“) gespeichert.

Tabelle 3-2: Distanzmetriken zur Berechnung des Abstandes d zweier beliebiger Punkte x und y im n -dimensionalen metrischen Raum. Ist der Raum ein chemischer Raum, so kann der Abstand als Unähnlichkeit zweier Moleküle betrachtet werden.

Metrik	Gleichung
Euklidische Metrik	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Manhattan-Metrik	$d(x, y) = \sum_{i=1}^n x_i - y_i $

Eine weitere gängige Form der Beschreibung sind molekulare Fingerabdrücke (engl. „Fingerprints“, Abbildung 3-4c).⁷ Dabei handelt es sich um Folgen oder Vektoren numerischer Werte. Diese werden in die holografischen Fingerabdrücke (ganzzahlig oder Gleitkommazahlen) und die binäreren Fingerabdrücke (Binärzahlen „0“ und „1“) unterteilt. Bei den Substrukturfingerabdrücken beschreiben die Zahlenwerte die Anzahl (holografisch) oder die Präsenz (binär) einer chemischen Substruktur in einem Molekül. Komplexere Fingerabdrücke berücksichtigen nicht nur die Anzahl/Präsenz molekularer Eigenschaften, sondern auch ihre Konnektivität (z.B. CATS-Deskriptor, Abschnitt 3.4.1). Vor allem die binären molekularen Fingerabdrücke können effizient gespeichert werden und ermöglichen so eine sehr schnelle Verarbeitung.¹¹⁴

Als letzte Beschreibungsform chemischer Strukturen sei hier die dreidimensionale Darstellung erwähnt (Abbildung 3-4d). Molekulare Graphen sind reine 2D-Beschreibungen. Und auch die meisten Fingerabdrücke und molekularen Deskriptoren beschränken sich auf ein- und zweidimensionale Eigenschaften von

Molekülen. Bei der 3D-Darstellung von Molekülen werden Atomzentren durch ihre Koordinaten im euklidischen Raum beschrieben.¹¹² Die Atome selber werden durch Punkte, Kugeln oder approximierende Oberflächen beschrieben, denen verschiedene Eigenschaften zugeordnet werden können.

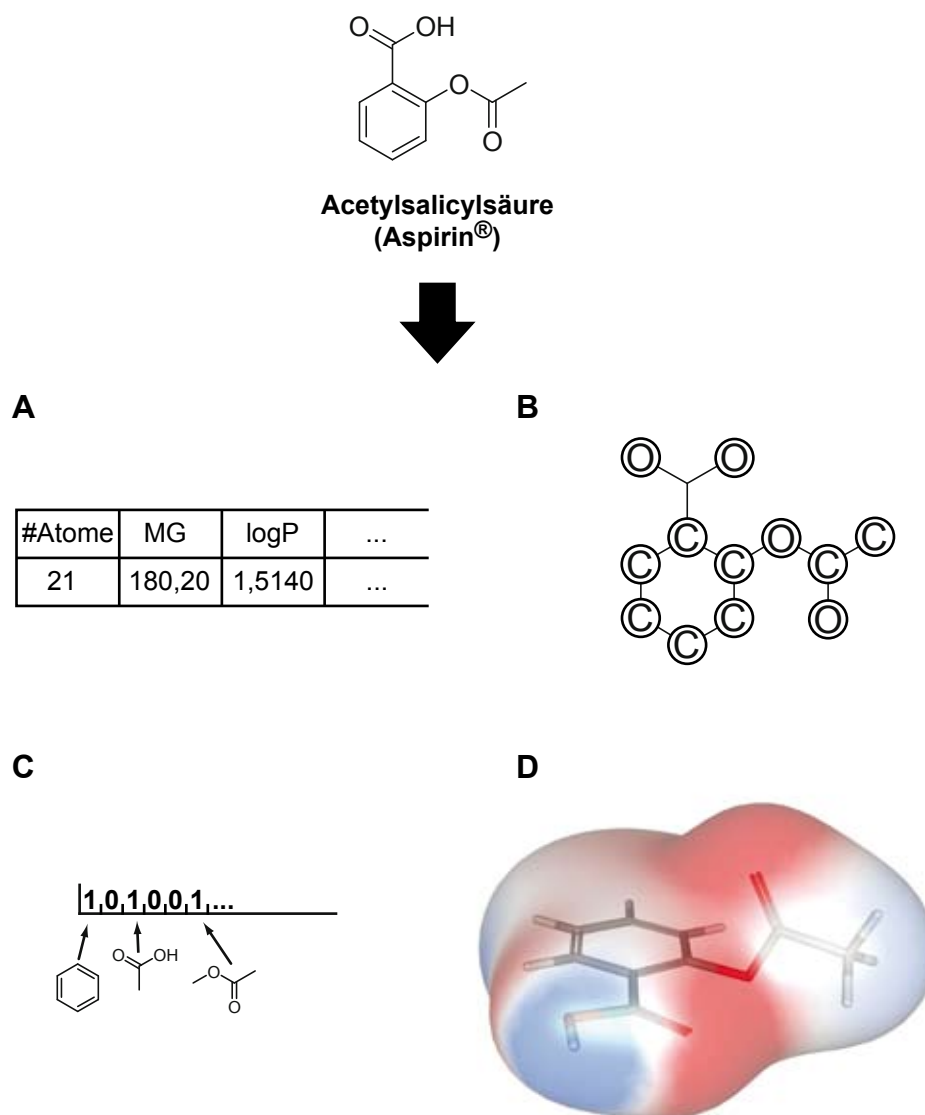


Abbildung 3-4: Verschiedene chemieinformatische Repräsentationen von Acetylsalicylsäure (Aspirin®). A: Molekulare Deskriptoren beschreiben Moleküle über ihre physikochemischen Eigenschaften, z.B. über die Anzahl der Atome (#Atome), das Molekulargewicht (MG), oder den Oktanol-Wasser-Verteilungskoeffizienten (logP). B: Molekulare Graphen beschreiben chemische Strukturen als Knoten (Atome), die über Kanten (Bindungen) verknüpft sind. C: Molekulare Fingerabdrücke beschreiben Moleküle zumeist über Bitstrings, deren verschiedene Bits z.B. die Anwesenheit einer Substruktur andeutet. D: Dreidimensionale Molekülmodelle weisen den Atomen Raumkoordinaten zu. Approximierende Oberflächen können berechnet werden, auf die verschiedene Eigenschaften, wie hier z.B. die Partialladungsverteilung, projiziert werden können (rot: negative Ladung, blau: positive Ladung, weiß: neutral).

So könnte man ein Molekül z.B. als die äußere Hülle aller Van der Waals-Oberflächen seiner Atome betrachten und auf diese Hülle Partialladungen projizieren.^{115,116}

Die dreidimensionale Beschreibung ist die Form der Repräsentation, die einem realen Molekül wahrscheinlich am nächsten kommt. 3D-Beschreibungen sind Voraussetzung für das automatisierte Docking und für Pharmakophormodelle und wurden in diesem Kontext auch im Rahmen dieser Arbeit eingesetzt. Ein Nachteil der 3D-Beschreibung ist das Konformerenproblem: Schon kleine organische Moleküle können so viele unterschiedliche Konformationen einnehmen, dass diese Form der Beschreibung auf heutigen Computersystemen rasch an ihre Grenzen kommt.

Als praktikable Lösung wird zumeist entweder eine einzige, energetisch günstige oder potentiell bioaktive Konformation berechnet, oder es werden systematisch oder stochastisch gleich einige hundert alternative Konformationen berechnet.¹¹⁷ Für Pharmakophorsuchen wurde in dieser Arbeit auf letztere Variante zurückgegriffen und die eingesetzte Dockingsoftware (GOLD, The Cambridge Crystallographic Data Centre, Cambridge, Vereinigtes Königreich)^{118,119} verwendet einen genetischen Algorithmus zur flexiblen Adaptation der Molekülkonformation an die Bindetasche.

3.4.1 Deskriptorberechnung für retrospektive Optimierungstudien

Die Verbindungen der Ugi-Bibliothek und der COBRA wurden für die retrospektiven Vergleichsstudien zur Optimierung fokussierter Substanzbibliotheken (vgl. Kapitel 5) durch molekulare Deskriptoren beschrieben. Für die Verbindungen der beiden Bibliotheken wurden nach ihrer Präprozessierung 146 2D-Deskriptoren mit Hilfe der Software „Molecular Operating Environment“ (MOE, Version 2004.03, Chemical Computing Group, Montreal, Kanada) berechnet.⁹⁷ Die Deskriptoren, die nur auf Grundlage der Verbindungstabelle (engl. „Connection table“) der Moleküle berechnet wurden und somit konformationsunabhängig sind, lassen sich in die Gruppen „physikalische Eigenschaften“ (Ladung, clogP, Molekulargewicht, ...), „Teiloberflächen“ (Oberflächeneigenschaften von clogP und der molaren Refraktivität), „Atom- und Bindungsanzahlen“, „Kier&Hall-Konnektivität und Kappa-Gestaltindizes“, „Adjazenz- und Distanzmatrixdeskriptoren“ (Petitjean, Wiener, ...), „Pharmakophoreigenschaften“ und „Partialladungsdeskriptoren“ einteilen.

Eine Auflistung aller Deskriptoren ist im Anhang in Abschnitt 8.1 angefügt. Für einige Deskriptoren wurden Partialladungen benötigt, die zuvor in MOE anhand des MMFF94x-Kraftfeldes berechnet wurden.^{107,108} Nach der Berechnung wurden

die Deskriptoren autoskaliert (mittelpunkzentriert und auf Einheitsvarianz skaliert, Gl. 3-4) um den endgültigen Deskriptorsatz zu liefern, der von nun an mit *MOE 2D* abgekürzt wird.

$$\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma} \quad (3-4)$$

mit dem autoskalierten Vektor \mathbf{Z} , dem Vektor der Originaldaten \mathbf{X} , dem Mittelwert μ und der Standardabweichung σ von \mathbf{X} .

Um die Dimensionalität und Redundanz des MOE-2D-Deskriptorsatzes zu reduzieren, wurde des Weiteren eine Hauptkomponentenanalyse (*PCA*, für engl. „Principle Component Analysis“)¹²⁰ auf den unskalierten MOE-2D-Deskriptoren mit Hilfe von MOE durchgeführt. Alle Hauptkomponenten mit einem Eigenwert größer oder gleich eins bildeten den zweiten Deskriptorsatz mit der Bezeichnung *MOE 2D PCA*.¹²¹ Für die Ugi-Bibliothek waren dies 16 Hauptkomponenten, die insgesamt 93,4% der Varianz des Datensatzes erklärten. Für COBRA wurden 22 Hauptkomponenten bestimmt, die 90,1% der Gesamtvarianz erklärten. Während der Hauptkomponentenanalyse wurden die Deskriptoren automatisch autoskaliert.

Als dritter Deskriptorsatz wurde der 150-dimensionale CATS-Deskriptor berechnet.¹²² CATS („Chemically Advanced Template Search“) ist ein alignierungsfreies Verfahren zur Beschreibung organischer Moleküle, das graphenbasierte Moleküleigenschaften auf einen Deskriptor abbildet. CATS definiert die fünf Pharmakophortypen „Wasserstoffbrückendonator“ (D), „Wasserstoffbrückenakzeptor“ (A), „lipophil“ (L), „positiv geladen oder ionisierbar“ (P) und „negativ geladen oder ionisierbar“ (N). Das Verfahren von CATS weist den Atomen und funktionellen Gruppen eines Moleküls bis zu zwei dieser Pharmakophoreigenschaften zu. Auf diese Weise annotierte Atome/Gruppen werden als potentielle Pharmakophorpunkte (PPP) bezeichnet. Es wird dann der topologisch kürzeste Pfad zwischen allen PPP-Paaren gesucht, der durch die Anzahl der chemischen Bindungen quantifizierbar ist.

Eine Abbildungsfunktion kombiniert daraufhin diese topologischen Distanzen und die PPP-Zuordnungen in einem Korrelationsvektor (*CV*, für engl. „Correlation Vector“). Die Abbildungsfunktion zählt die Häufigkeit aller auftretenden, paarweisen PPP-Distanzen und summiert sie in zehn Intervallen, die die Distanzen von 0 bis 9 Bindungen abdecken. Die 15 möglichen PPP-Permutationen (DD, DA, DP,

DN, DL, AA, AP, AN, AL, PP, PN, PL, NN, NL, LL) und die zehn verschiedenen Distanzintervalle führen schließlich zu einem 150-dimensionalen CV. Der letzte Schritt der CATS-Deskriptor-Berechnung ist eine Skalierung des Vektors.¹²³ Hier wurde die „sensitive Skalierung“ verwendet, bei der durch die summierten Häufigkeiten der zwei an einem PPP-Paar beteiligten Pharmakophortypen dividiert wird.¹²³ Wie die MOE-2D und MOE-2D-PCA-Deskriptorsätze, wurde auch der CATS-Deskriptor anschließend autoskaliert, um vergleichbare Wertebereiche sicherzustellen.

3.4.2 Deskriptorberechnung für eine prospektive Optimierungsstudie

Für eine prospektive, adaptive Optimierungsstudie einer großen Substanzbibliothek (vgl. Kapitel 6) wurden 184 2D-Deskriptoren mit MOE Version 2005.06 berechnet.⁹⁷ Die Deskriptorberechnung erfolgte wie im vorherigen Abschnitt beschrieben. Im Vergleich zu der dort eingesetzten MOE-Version 2004.03, wurden der Nachfolgeversion unter anderem „BCUT“- und „GCUT“-Deskriptoren¹²⁴ und Deskriptoren zur Wirkstoff- und Leitstrukturartigkeit (Lipinski & Oprea)^{125,126} hinzugefügt. Auf den 184 2D-Deskriptoren wurde eine Hauptkomponentenanalyse durchgeführt und 27 Hauptkomponenten mit einem Eigenwert größer oder gleich eins wurden als finaler Deskriptorsatz verwendet (erklärte Varianz: 89,8%).

Aufgrund eines Programmfehlers in CLIFF (Version 1.14, Molecular Networks GmbH, Erlangen)⁹⁶ während der Molekülvorbereitung musste die Deskriptorberechnung wiederholt werden (vgl. Abschnitt 6.9). Es wurden 184 2D-Deskriptoren mit MOE Version 2006.08 berechnet. Nach einer Hauptkomponentenanalyse wurden 28 Hauptkomponenten mit einem Eigenwert größer oder gleich eins für den finalen Deskriptorsatz ausgewählt (erklärte Varianz: 89,2%). Die Neuberechnung ist ausführlich im Abschnitt 6.9 beschrieben.

3.4.3 Deskriptorberechnung für eine SOM-Analyse von RNA-Liganden

Für das Training einer selbstorganisierenden Karte (SOM, für engl. „Self-Organizing Map“, vgl. Abschnitt 3.6)¹²⁷ auf einem gemischten Datensatz aus RNA-Liganden (RNA-Liganden-Bibliothek) und Protein-Bindern (COBRA) wurden 184 2D-Deskriptoren mit MOE Version 2007.09 berechnet. Die Deskriptoren wurden anschließend autoskaliert. Die SOM-Analyse ist in Abschnitt 6.7 beschrieben.

3.5 Optimierungsalgorithmen

Optimierungsalgorithmen sind Lösungsverfahren mathematisch definierter Minimierungs- oder Maximierungsprobleme.¹²⁸ Sie versuchen, einen Satz an Eingangsvariablen $\mathbf{x}_{\text{opt}} = (x_1, x_2, \dots, x_{n-1}, x_n)$ zu finden, für den eine gegebene Funktion $f(\mathbf{x})$ ein Optimum besitzt. \mathbf{x} kann als Positionsvektor eines Lösungskandidaten von $f(\mathbf{x})$ in einem n -dimensionalen Suchraum aufgefasst werden. Je nach Kontext wird $f(\mathbf{x})$ als Zielfunktion, Kostenfunktion, Bewertungsfunktion oder – vor allem im Zusammenhang mit evolutionären Algorithmen – als Fitnessfunktion bezeichnet.

Die generelle Arbeitsweise vieler numerischer Optimierungsalgorithmen ist iterativ. Nur im einfachsten Fall lassen sich die Extrempunkte durch Nullsetzen der ersten Ableitung bestimmen (analytische Optimierung). Dazu muss die Zielfunktion algebraisch gegeben und an jedem Punkt differenzierbar sein.² Ist sie dies wie z.B. im *chemischen Raum* (vgl. Abschnitt 3.4) nicht, bleibt zum Auffinden eines Optimums nur das Abtasten des Suchraumes, indem die Zielfunktion an verschiedenen Positionen im Raum berechnet wird. Dazu könnte man den Suchraum sehr engmaschig rastern, an jedem Rasterpunkt die Zielfunktion auswerten und auf diese Weise das Optimum bestimmen (Gittersuche). In der Praxis sind Zielfunktionen im Allgemeinen weder überall differenzierbar noch ist der Suchraum klein genug, um ihn systematisch zu durchsuchen.⁴⁴ Würde man z.B. im zehndimensionalen Raum jede Achse des Suchraumes an nur zehn Stellen rastern, so müsste die Zielfunktion 10.000.000.000 (10^{10})-mal ausgewertet werden. Bei rechenintensiven Zielfunktionen wäre dies nicht machbar. Stattdessen kann ein Optimierungsalgorithmus seine Optimumssuche z.B. an einer beliebigen Stelle im Suchraum beginnen. Die Zielfunktion an dieser Stelle wird ausgewertet und die Suche springt auf eine benachbarte Stelle, an der die Zielfunktion wiederum ausgewertet wird. Aus der Differenz der Funktionswerte können Rückschlüsse über die lokale Beschaffenheit des Suchraumes abgeleitet werden. Im einfachsten Fall wird die schlechtere der beiden Lösungen verworfen und von der besseren ausgehend wird abermals eine Nachbarlösung gesucht. Der Prozess wird iterativ wiederholt, bis ein Optimum gefunden wurde.

Dieses einfache Verfahren wird als „Bergsteigeralgorithmus“ (engl. „Hill climbing“) bezeichnet.¹²⁹ Der Definitions- und der Wertebereich einer Zielfunktion spannen einen $n+1$ -dimensionalen Suchraum auf. Optimierungswissenschaftler stellen sich diesen Suchraum oft als ein Gebirge vor, dessen Plankordinaten den Eingangsvariablen entsprechen und dessen Höhenlinien den Fitnesswerten der zu optimierenden

Funktion gleichkommen.² In Abbildung 3-5 ist ein solches Fitnessgebirge schematisch dargestellt. Soll die Zielfunktion minimiert werden, gilt es das tiefste Tal zu finden, und soll sie maximiert werden, gilt es den höchsten Berg zu finden. Auch der chemische Raum von Substanzbibliotheken lässt sich so visualisieren, wenn auf der x - und y -Achse molekulare Deskriptoren aufgetragen werden und die z -Achse z.B. die mittlere inhibitorische Konzentration der Bibliothekssubstanzen gegen ein ausgewähltes pharmakologisches Ziel wiedergibt. Der chemische Suchraum weist eine Besonderheit auf: Im Gegensatz zur kontinuierlichen Fitnesslandschaft in Abbildung 3-5 ist der chemische Raum diskret. Die Zielfunktion kann nur an jenen Orten des Suchraumes evaluiert werden, die realen Moleküle entsprechen.¹³⁰ Wird der Suchraum durch eine Substanzbibliothek aufgespannt, so gibt es genau so viele diskrete Punkte im Raum, wie es Substanzen in der Bibliothek gibt.

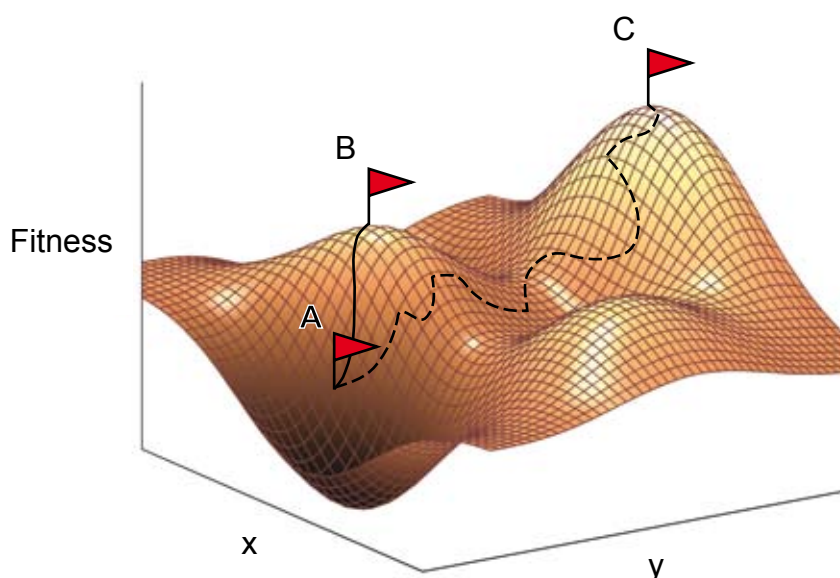


Abbildung 3-5: Optimierung im Fitnessgebirge. Der direkte Weg von A nach B (durchgezogene Linie) folgt dem steilsten Gradienten, erreicht aber nur ein lokales Optimum. Die alternative, gestrichelte Route (von A nach C) ist zwar weniger zielstrebig und länger, findet jedoch das globale Optimum.

Angenommen die in Abbildung 3-5 dargestellte Fitnessfunktion solle maximiert werden und eine Optimierung starte an Punkt A. Folgte man von diesem Punkt dem steilsten Gradienten (durchgezogenen Linie), so landete man schnell auf dem „Gipfel“ bei Punkt B. Diese Route würde z.B. von einem Gradientenverfahren oder einem Bergsteigeralgorithmus beschritten werden. Von Punkt B aus betrachtet führen alle Pfade augenscheinlich nur bergab und die Optimierung würde an dieser Stelle abbrechen. Trotzdem befänden wir uns hier nur in einem lokalen Optimum, da es

noch einen höher gelegenen Gipfel bei Punkt C gibt, welcher von B nur durch das Überwinden eines Tals zu erreichen ist. Die alternative, gestrichelte Route folgt einer anderen Strategie. Sie ist weniger zielstrebig, wesentlich länger und erlaubt neben Aufstiegen auch Abstiege. Jedoch ist diese Strategie *robuster*, weil sie nicht im lokalen Optimum bei Punkt B stecken bleibt, sondern schließlich das globale Optimum bei Punkt C findet. Eine Fitnessfunktion, die mehr als ein (lokales) Optimum aufweist, wird als „multimodal“ bezeichnet.

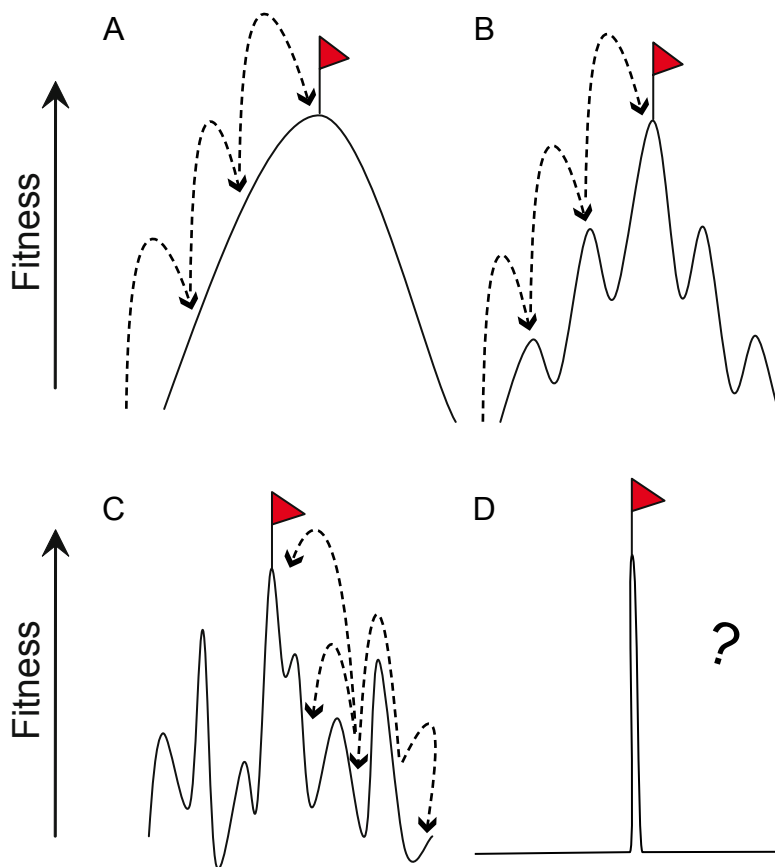


Abbildung 3-6: Verschiedenenartige Fitnesslandschaften. A) Eine konvexe Funktion ist der einfachste Fall einer Fitnesslandschaft. B) Eine multimodale Fitnesslandschaft stellt die Optimierung vor größere Schwierigkeiten, kann jedoch auch überwunden werden, wenn die Gipfel einer Systematik folgen („Gipfelspringen“). C) Eine stark multimodale Fitnesslandschaft ohne erkennbare Systematik in der Anordnung der Gipfel ist für einen Optimierungsalgorithmus schwer zu meistern. Populationsbasierte Ansätze können hier helfen. D) Besteht die Fitnesslandschaft aus einem (oder mehreren) sehr steilen, nadelförmigen Gipfeln ist sie für Optimierungsalgorithmen praktisch unlösbar. Hier hilft nur eine Zufallssuche. Abbildung modifiziert nach Schneider, 2000.^{130a}

Iterative Optimierungsstrategien basieren allgemein auf dem *Prinzip der starken Kausalität*: „Ähnliche Ursachen haben ähnliche Wirkungen“.² Dieses Prinzip besagt, dass eine kleine Änderung der Eingangsvariablen auch eine kleine Änderung der

Zielfunktion bewirkt. Starke Kausalität gilt, wenn in der nahen Umgebung eines jeden Ortes ähnliche Höhen im Fitnessgebirge anzutreffen sind. Somit sollten benachbarte Positionen im Suchraum vergleichbare Fitnesswerte besitzen. Der Idealfall ist eine konvexe Fitnessfunktion und lässt sich bildlich als Landschaft mit einem einzigen sanften Hügel beschreiben (Abbildung 3-6a).^{2,7,131} Eine multimodale Fitnesslandschaft stellt die Optimierung vor größere Schwierigkeiten, kann jedoch auch überwunden werden, wenn die Gipfel einer gewissen Systematik folgen („Gipfelspringen“, Abbildung 3-6c). Eine stark multimodale Fitnesslandschaft ohne erkennbare Systematik in der Anordnung der Gipfel ist für einen Optimierungsalgorithmus schwer zu meistern. Populationsbasierte Ansätze können hier helfen. Diese arbeiten nicht mit einem einzigen Lösungskandidaten pro Optimierungsrunde, sondern mit einer ganzen Schar, ähnlich einer biologischen Population. Dadurch können gleichzeitig verschiedene Alternativen verfolgt werden. Besteht die Fitnesslandschaft aus weiten Ebenen, die nur an sehr wenigen, punktförmigen Stellen von nadelförmigen Gipfeln überragt werden, ist sie für Optimierungsalgorithmen praktisch unlösbar (Abbildung 3-6d). In diesem ungünstigsten Fall bewirkt eine Änderung der Eingangsvariablen in weiten Teilen des Suchraumes gar keine Änderung der Zielfunktion. Es ist also nicht möglich, durch schrittweise Verbesserung das Optimum zu finden. In einem solchen Suchraum versagen kausalitätsbasierte Suchverfahren und als einzige Strategie verbleibt nur eine Zufallssuche (oder eine systematische Suche für diskrete Suchräume).⁴⁴

Bei einer Zufallssuche werden beliebige Punkte im Suchraum zufällig aufgesucht und evaluiert (vgl. Abschnitt 3.5.1). Das Prinzip der starken Kausalität wird in der Chemieinformatik mit dem Begriff des Nachbarschaftsverhaltens umschrieben.¹³²⁻¹³⁴ Eine molekulare Repräsentation besitzt ein gutes Nachbarschaftsverhalten, wenn ähnlich repräsentierte Moleküle ähnliche Eigenschaften besitzen. In diesem Fall können ausgehend von einem betrachteten Molekül Rückschlüsse auf vergleichbare Verbindungen in dessen Nachbarschaft gezogen werden, wodurch eine iterative, kausalitätsbasierte Optimierung möglich wird. In der Praxis reicht es, wenn das Prinzip der starken Kausalität nur approximativ Gültigkeit besitzt. Es kann daher auch in „verrauschten“ Fitnesslandschaften optimiert werden, so lange das Prinzip der starken Kausalität in der Mehrzahl der Variationsschritte gilt.

In dieser Arbeit wurden vier verschiedene Suchalgorithmen mit einander verglichen:

- a) Zufallssuche,
- b) Simulated Annealing,⁴²⁻⁴⁴
- c) Evolutionsstrategie,^{1,2,38}
- d) Partikelschwarmoptimierung.^{47,48}

Mit Ausnahme der Zufallssuche gehören diese Algorithmen zur Gruppe der stochastischen Optimierungsalgorithmen, deren gemeinsame Eigenschaft die Verwendung eines Zufallselementes (z.B. einer (Pseudo-)Zufallszahl) während der Optimierung ist.¹³⁵ Es handelt sich im Allgemeinen um Heuristiken. Dies sind Verfahren, die zulässige, meist gute Lösungen, aber nicht die beweisbar optimale Lösung finden. Wie nahe eine Lösung dem globalen Optimum ist, ist im Normalfall unbekannt oder kann nur geschätzt werden. Den stochastischen Verfahren gegenübergestellt ist die deterministische Optimierung. Hierbei werden z.B. die partiellen Ableitungen in einem lokalen Punkt gebildet und es wird dem Gradienten bis zum nächsten (lokalen) Optimum gefolgt.

Gradientenbasierte Verfahren wie z.B. der Gradientenabstieg bieten den Vorteil einer schnellen Arbeitsweise (kurze Laufzeit) und eines exakten Ergebnisses. Sie sind dafür nicht sehr robust, weil sie dazu neigen, das nächste lokale Optimum, anstatt das globale Optimum zu finden. Zudem muss die Zielfunktion algebraisch gegeben und zumindest lokal partiell differenzierbar sein. Dem gegenüber besitzen die stochastischen Verfahren eine universellere Anwendungsdomäne und verhalten sich in zerklüfteten Fitnessgebirgen robuster. Stochastische Optimierungsalgorithmen arbeiten mit beliebigen Zielfunktionen und besitzen Strategien zum Entkommen lokaler Optima. Eine solche Strategie ist z.B. das Verwenden von Populationen von Lösungskandidaten oder eine adaptive Schrittweitenregelung.^{2,128} Diese Universalität wird meist durch einen vielfach höheren Rechenaufwand erkauft.

Die Optimierungsalgorithmen wurden in der Programmiersprache Java (Version 1.5, Sun Microsystems, Inc., Santa Clara, USA)⁵⁹ in einer konsequent objektorientierten Weise unter Benutzung von Entwurfsmustern (engl. „Design patterns“)¹³⁶ implementiert. Im Rahmen dieser Arbeit wurde so eine flexible, modulare Programmbibliothek für die stochastische Optimierung entwickelt. Diese Bibliothek trägt den Namen *Cyclops* („Cyclische Optimierung“).

3.5.1 Zufallssuche

Die Zufallssuche wird in dieser Arbeit als Referenz herangezogen, um zu evaluieren wie gut ein Ergebnis hätte sein können, wenn es rein zufällig bestimmt worden wäre. Während einer Zufallssuche wird eine bestimmte Anzahl von Lösungen zufällig aus einer Lösungsmenge (hier: Moleküle einer Screeningbibliothek) gezogen, ihr Zielfunktionswert wird bestimmt und die Lösungen werden entsprechend sortiert. Eine effiziente Suchstrategie sollte bessere Ergebnisse als eine Zufallssuche liefern.

Pseudocode einer Zufallssuche:

```
solange (nicht beendet) {  
    Lösung = wähleZufallseintrag(Datensatz);  
    Qualität = evaluiere(Lösung);  
}
```

3.5.2 Simulierte Abkühlung

Das Simulated Annealing (SA) ist ein vielseitiger, globaler Optimierungsalgorithmus, der die Abkühlung eines physikalischen Systems simuliert, dessen mögliche Energiezustände der zu minimierenden Zielfunktion entsprechen.⁴²⁻⁴⁴ Da physikalische Systeme Niedrigenergiezustände annehmen, wenn die Temperatur auf den absoluten Nullpunkt gesenkt wird, sollte sich ein auf dieser Analogie basiertes Optimierungsverfahren ähnlich verhalten. Bei SA wird zu einem aktuellen Lösungskandidaten eine benachbarte Lösung gesucht. Diese wird als aktuelle Lösung übernommen, wenn sie qualitativ besser ist. Ist ihre Qualität schlechter, wird sie mit einer gewissen Wahrscheinlichkeit trotzdem als aktuelle Lösung akzeptiert. Die Wahrscheinlichkeit der Akzeptanz ist durch das Metropolisakzeptanzkriterium gegeben (Gl. 3-5):⁴³

$$R < e^{-\Delta E/T} \quad (3-5)$$

wobei R eine in $[0, 1]$ gleichverteilte Zufallszahl, ΔE die Energiedifferenz, oder allgemeiner formuliert, der Qualitätsunterschied zweier benachbarter Lösungen und T ein systemspezifischer Temperaturfaktor ist. Ist Gl. 3-5 erfüllt, so wird ein Lösungskandidat selbst dann akzeptiert, wenn seine Qualität schlechter als die der vorherigen Lösung ist. Der Temperaturfaktor bestimmt die Höhe der Akzeptanzwahrscheinlichkeit. Durch die simulierte Abkühlung ist diese Wahrscheinlichkeit zu Beginn der Optimierung groß und nimmt im Verlauf der Iterationen ab.

In der hier verwendeten Implementierung wurde ein exponentielles Abkühlungsschema verwendet (Gl. 3-6):

$$T(t) = T_0 \alpha^t \quad (3-6)$$

wobei $T(t)$ die Temperatur zum Zeitpunkt t ist, T_0 die Initialtemperatur bezeichnet und α eine Abkühlungsrate ist. Der Wert für T_0 wurde so gewählt, dass ein halbmaximaler Qualitätsunterschied zwischen zwei Lösungen einer

Metropolisakzeptanzwahrscheinlichkeit von ca. 50% zu Beginn der Optimierung entspricht. α wurde so gesetzt, dass mit der letzten Iteration eine Temperatur von 1 erreicht wird. Eine Nachbarlösung wurde durch Anwenden eines Evolutionsstrategie-ähnlichen Mutationsoperators mit fester Schrittweite erzeugt (vgl. Abschnitt 3.5.3). Die obere Bestimmungsgrenze von experimentellen Aktivitätsmessungen (hier: $IC_{50} = 100 \mu\text{M}$) wurde in besonderer Weise berücksichtigt (Plateauproblem): Das Metropolisakzeptanzkriterium wurde so modifiziert, dass inaktive Verbindungen selbst dann zurückgewiesen werden, wenn sie nach Gl. 3-5 akzeptiert worden wären. Diese Modifikation soll verhindern, dass von einem „aktiven“ Lösungskandidat auf einen „inaktiven“ gesprungen werden kann. Die hier verwendete SA-Implementierung lag als Multistart-Algorithmus vor, bei dem mehrere Optimierungsläufe parallel durchgeführt werden können und die besten Lösungen dieser Läufe archiviert werden.

Pseudocode des Simulated Annealings:

```

Iteration = 0;
Lösung = wähleInitiallösung(Datensatz);
Qualität = evaluiere(Lösung);
solange (nicht beendet) {
    NächsteLösung = wähleNachbar(Lösung);
    QualitätNächsteLösung = evaluiere(NächsteLösung);
    ΔQualität = QualitätNächsteLösung - Qualität;
    wenn (ΔQualität ≤ 0)
        Lösung = NächsteLösung;
    sonst {
        R = generiereZufallszahl();
        wenn (R < exp(-ΔQualität/T) &&
            QualitätNächsteLösung < Qualitätsgrenzwert)
            Lösung = NächsteLösung;
    }
    T = T0 αIteration;
    Iteration++;
}

```

3.5.3 Evolutionsstrategie

Evolutionsstrategien (ES) sind heuristische, populationsbasierte Optimierungsalgorithmen, die in den späten 1960er Jahren von Rechenberg entwickelt wurden.^{1,2,38} Sie gehören zur Klasse der evolutionären Algorithmen (EA) und ihre Arbeitsweise ist von Konzepten der Darwinschen Evolution inspiriert. Nachbarlösungen, hier als „Nachkommen“ bezeichnet, werden durch Anwendung von Mutations- und Cross-Over-Operatoren aus Eltern-Individuen erschaffen. Diese wiederum wurden zuvor

durch simulierte natürliche Selektion aus einer Population von Individuen ausgewählt. Nur die fittesten Individuen werden als Eltern ausgewählt; alle anderen Individuen „sterben“, ohne sich fortzupflanzen. Eine (μ, λ) -Evolutionstrategie selektiert μ Eltern, die in jeder Generation zusammen λ Nachkommen erzeugen. Evolutionstrategien existieren in den Varianten „Komma-Strategie“ und „Plus-Strategie“. Bei der Komma-Strategie überleben Eltern-Individuen nur jeweils eine Generation. Bei der Plus-Strategie können Eltern-Individuen auch in die Filialgeneration übernommen werden (Elitismus). Die in dieser Arbeit eingesetzte Implementierung verwendet ausschließlich einen Mutationsoperator. Dieser erzeugt Nachkommen, indem er $(0, \sigma)$ -normalverteilte Pseudozufallszahlen auf jede Komponente des elterlichen Positionsvektors addiert. σ ist hier die Standardabweichung einer Gaußschen Normalverteilung. Auf diese Weise werden Nachkommen gemäß einer $(\sigma\sqrt{n}, \sigma/\sqrt{2})$ -Normalverteilung um ihre Eltern herum gestreut. Dies bedeutet bildlich, dass neue Kindindividuen in der Schale einer Hyperkugel mit Radius $r = \sigma \cdot \sqrt{n}$ und der Stärke $w = \sigma/\sqrt{2}$ zu liegen kommen, wobei n die Dimension des Suchraumes bezeichnet.² Im Mittelpunkt der Hyperkugel befindet sich das Eltern-Individuum. σ stellt einen allgemeinen Schrittweitenparameter von Evolutionstrategien dar, der die Schrittlänge und Streuung neuer Nachkommen bestimmt. Um lokalen Minima effizient zu entkommen, wurde dieser Parameter in einer autoadaptiven Form implementiert. Nachkommen erben mit einer 50%igen Wahrscheinlichkeit die mit 1,3 multiplizierte oder die durch 1,3 dividierte Schrittweite ihres Elters.² ES arbeiten mit einem hohen Selektionsdruck, da nur die μ fittesten Individuen als Eltern ausgewählt werden. Der oberen Bestimmungsgrenze von Aktivitätsmessungen wurde auch bei der ES Rechnung getragen. Der Schritt der Elternselektion wurde so modifiziert, dass der tatsächliche Wert von μ an die Anzahl „aktiver“ Individuen (Moleküle) einer Population adaptiert wird (Gl. 3-7):

$$\mu' = \begin{cases} \mu, & \text{wenn } \textit{Aktive} = 0 \\ \textit{Aktive}, & \text{wenn } 0 < \textit{Aktive} < \mu \\ \mu & \text{wenn } \textit{Aktive} \geq \mu \end{cases} \quad (3-7)$$

mit der resultierenden Zahl von Eltern μ' , der maximalen Anzahl von Eltern μ , und der Anzahl aktiver Individuen in der Population *Aktive*. Diese Modifikation erhöht den Selektionsdruck und verhindert die Selektion inaktiver Individuen als Eltern.

Eine weitere essentielle Modifikation wurde im Schritt der Nachkommenerzeugung vorgenommen. Die Modifikation betrifft sowohl ES als auch SA, da SA hier ebenfalls mit einem ES-ähnlichen Mutationsoperator arbeitet. Evolutionstrategien operieren

in kontinuierlichen Suchräumen. Eine durch molekulare Deskriptoren repräsentierte Molekülbibliothek stellt jedoch einen dünn besiedelten, diskreten Suchraum dar. Der chemische Raum ist diskret, da nur solche Positionen besetzt sind, an denen „echte“ Moleküle realisiert sind. Der Mutationsschritt der Optimierungsalgorithmen erzeugt jedoch Vektoren, die auf beliebige Punkte im chemischen Raum zeigen. An diesen beliebigen Punkten sind mit ziemlicher Sicherheit keine „echten“ Moleküle realisiert. Um diese Diskrepanz zu kompensieren, wird für jeden Nachkommen eine Nächste-Nachbarn-Suche mittels eines k -Nächste-Nachbarn-Algorithmus (k -NN)¹³⁷ durchgeführt. Zu einem beliebigen Punkt im chemischen Raum wird also der nächstgelegene diskrete Punkt (ein „echtes“ Molekül) gesucht und als Nachkomme übergeben. Die nächsten Nachbarn werden bestimmt, indem zunächst die Manhattan-Distanz zwischen einem Positionsvektor und den Deskriptorvektoren aller Datenbankmoleküle berechnet wird. Die resultierende Liste wird nach aufsteigender Distanz sortiert und der erste Listeneintrag (mit minimaler Distanz) wird als nächster Nachbar zurückgegeben. Es bleibt anzumerken, dass, obwohl molekulare Screeningbibliotheken diskrete Suchräume darstellen, die Lösungskandidaten (hier: Moleküle) dennoch durch reellwertige Deskriptorvektoren repräsentiert werden. Als Distanzmaß wurde die Manhattan-Metrik ausgewählt, da sie gerade in hochdimensionalen Räumen Vorteile bietet.¹³⁸

Pseudocode einer Evolutionsstrategie:

```
Iteration = 0;
Population = wähleInitialpopulation(Datensatz);
Fitness = evaluiere(Population);
solange (nicht beendet) {
    Eltern = wähleEltern(Population, Fitness);
    Population = pflanzefort(Eltern);
    Fitness = evaluiere(Population);
    Iteration++;
}

Funktion pflanzefort(Eltern) {
    für (Elter in Eltern) {
        für (i in 1...λ/μ) {
            R = generiereZufallszahl();
            wenn (R < 0.5)
                Schrittweite = Schrittweite * 1,3;
            sonst
                Schrittweite = Schrittweite / 1,3;
            x = Elter.Positionsvektor;
            für (j = 1...Dimension)
                x[j] = x[j] +
                    generiereNormalverteilteZufallszahl(Schrittweite);
            Nachkomme = sucheNächstenNachbarn(x);
        }
    }
}
```

3.5.4 Partikelschwarmoptimierung

Die Partikelschwarmoptimierung (PSO) ist ein biologisch inspiriertes, populationsbasiertes Optimierungsverfahren, das 1995 von Kennedy und Eberhart eingeführt wurde.^{47,48} PSO basiert auf einer Abstraktion biologischer Schwärme, wie Vögelschwärmen oder Fischeschwärmen, und imitiert deren kollektive Bewegungen und sozialen Interaktionen. Während sich Partikel (also die Lösungskandidaten der Zielfunktion eines Schwarmes) durch den Suchraum bewegen, behalten sie zwei Arten von Gedächtnis: a) in ihrem lokalen, individuellen Gedächtnis speichern sie die persönlich beste Position, die sie bislang entdeckt haben und b) in einem globalen Gedächtnis speichern sie die gesamtbeste Position, die der Schwarm bislang entdeckt hat. Nachbarlösungen werden erzeugt, indem sich Partikel durch den Suchraum bewegen. Zusätzlich zu einem Positionsvektor besitzt jeder Partikel einen Geschwindigkeitsvektor, der die Richtung und Weite der Fortbewegung bestimmt. In

der vorliegenden Arbeit wurde ein Konstriktionspartikelschwarm verwendet.¹³⁹ Eine Bewegungsoperation besteht aus der Aktualisierung des Geschwindigkeitsvektors nach Gl. 3-8 und der Änderung des Positionsvektors nach Gl. 3-9:

$$v_i(t+1) = K \cdot (v_i(t) + c \cdot R_1 \cdot (m_i^c - x_i(t)) + s \cdot R_2 \cdot (m_i^s - x_i(t))) \quad (3-8)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (3-9)$$

wobei $v_i(t)$ die i -te Komponente des Geschwindigkeitsvektors zur Zeit t bezeichnet, x den Positionsvektor bezeichnet, R_1 und R_2 gleichverteilte Zufallszahlen in $[0, 1[$ sind, m^c das individuelle Gedächtnis und m^s das globale Gedächtnis darstellen und c („cognitiver“ Faktor) und s („sozialer“ Faktor) positive Gewichtungsfaktoren sind, die den Einfluss des individuellen, respektive, des globalen Gedächtnisses beeinflussen. Die Bezeichnungen „cognitiver“ Faktor und „sozialer“ Faktor leiten sich aus der Analogie zu biologischen Schwärmen ab. Der Konstriktionsfaktor K , der eine adaptive Konvergenz des Schwarmes vermittelt, ist definiert als (3.10):¹³⁹

$$K = \frac{2}{\left| 2 - \varphi - \sqrt{\varphi^2 - 4\varphi} \right|}, \quad \text{mit} \quad \varphi = c + s \quad (3-10)$$

Die Geschwindigkeitsvektoren wurden zufällig innerhalb eines Begrenzungsbereiches initialisiert, der durch den Minimal- und Maximalwert jeder Dimension des Suchraumes gegeben war. Um es Partikelschwärmen zu ermöglichen, effizient mit binären Zielfunktionswerten (z.B. den binären Aktivitätswerten „aktiv“ und „inaktiv“) zu arbeiten, musste die Routine zur Aktualisierung der Gedächtnisse leicht modifiziert werden: Gewöhnlicherweise wird das Gedächtnis der Partikel nur dann aktualisiert, wenn eine bessere Lösung gefunden wurde. Mit der Modifikation werden nun auch Lösungen akzeptiert, wenn sie gleich gut oder besser sind. Eine gleich gute Lösung wird jedoch abgewiesen, falls ihre Fitness (IC_{50}) oberhalb eines Schwellenwertes liegt (wie es für inaktive Moleküle der Fall ist). Dies berücksichtigt wieder die obere Bestimmungsgrenze von Aktivitätsmessungen. Wie schon zuvor bei SA und ES, wurde eine zusätzliche Nächste-Nachbarn-Suche angeschlossen, um die Partikel vom kontinuierlichen Suchraum auf den diskreten, chemischen Raum abzubilden.

Pseudocode einer Partikelschwarmoptimierung:

```
Iteration = 0;
Population = wähleInitialpopulation(Datensatz);
initialisiereGeschwindigkeitsvektorenZufällig(Population);
Fitness = evaluiere(Population);
Population.aktualisiereLokalesGedächtnis(Fitness);
Population.aktualisiereGlobalesGedächtnis(Fitness);
solange (nicht beendet) {
    für (Partikel in Population) {
        v = Partikel.Geschwindigkeitsvektor;
        x = Partikel.Positionsvektor;
        für (i in 1...Dimension) {
            R1 = generiereZufallszahl();
            R2 = generiereZufallszahl();
            v[i] = K*(v[i]+c*R1*(mc[i]-x[i])+s*R2*(ms[i]-x[i]));
            x[i] = x[i] + v[i];
        }
        Partikel = sucheNächstenNachbarn(x);
        Partikel.Geschwindigkeitsvektor = v;
    }
    Fitness = evaluiere(Population);
    Population.aktualisiereLokalesGedächtnis(Fitness);
    Population.aktualisiereGlobalesGedächtnis(Fitness);
    Iteration++;
}
```

3.6 Selbstorganisierende Karten (SOM)

Selbstorganisierende Karten (SOM, für engl. „Self-Organizing Map“) gehören zur Klasse der unüberwachten, künstlichen neuronalen Netze und wurden in den frühen 1980er Jahren von Kohonen eingeführt.¹²⁷ Man kann sie unter anderem zur Erzeugung niedrigdimensionaler Projektionen hochdimensionaler Daten einsetzen. So lassen sich z.B. hochdimensionale Deskriptorrepräsentationen chemischer Substanzbibliotheken auf zweidimensionale Karten abbilden. SOMs sind aus einer einzigen Schicht von Neuronen aufgebaut, welchen jeweils ein Gewichtsvektor mit der selben Dimension wie die der Eingabemuster (z.B. Deskriptorvektoren) zugeordnet ist. Während des Trainingsprozesses einer SOM wird der ursprünglich hochdimensionale Raum in Cluster zerlegt. Es werden genau so viele Cluster gebildet, wie die SOM Neuronen besitzt. Damit ähnliche Eingabemuster auf benachbarte Regionen des Ausgaberaumes abgebildet werden, wird für die Neuronenschicht der SOM eine Topologie eingeführt. Die Topologie kann verschiedene Formen annehmen (z.B. Rechteck, Torus, Kugel) und hat zur Aufgabe, die Neuronen der SOM untereinander zu verbinden. Zu Beginn des Trainings einer SOM sind die Neuronen zufällig im Datenraum verteilt.

Zum Training der SOM wird ein Eingabemuster mit allen Neuronen verglichen und das ähnlichste Neuron („Gewinnerneuron“) bestimmt. Dann passt eine Aktualisierungsroutine den Gewichtsvektor des Gewinnerneurons in Richtung des Eingabemusters an. Zusätzlich werden die topologischen Nachbarneuronen des Gewinnerneurons in abgeschwächter Form in Richtung des Eingabemusters angepasst. Als Ergebnis entsprechen topologisch benachbarte Neuronen ähnlichen Eingabemustern. Das Training erfolgt iterativ mit einer abnehmenden Lernrate. Formal erfolgt die Aktualisierung der Gewichtsvektoren nach Gl. 3-11:¹⁴⁰

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \varepsilon(t) h_{si} d(\mathbf{x}_j(t) - \mathbf{w}_i(t)) \quad (3-11)$$

mit dem Gewichtsvektor $\mathbf{w}_i(t)$ des i -ten Neurons zum Zeitpunkt t , der Lernrate $\varepsilon(t)$, der Nachbarschaftsfunktion h_{si} , der Distanzmetrik d und dem j -ten Eingabemuster $\mathbf{x}_j(t)$.

Die zeitabhängige Lernrate ist definiert als (Gl. 3-12):

$$\varepsilon(t) = \varepsilon_{Start} \left(\frac{\varepsilon_{Ende}}{\varepsilon_{Start}} \right)^{\frac{t}{t_{max}}} \quad (3-12)$$

mit der Startlernrate ε_{Start} , der Lernrate am Ende des Lernprozesses ε_{Ende} und der maximalen Lernzeit (Anzahl der Iterationen) t_{max} .

Häufig wird eine Gaußsche Nachbarschaftsfunktion verwendet (Gl. 3-13):

$$h_{si} = \exp\left(\frac{-d(s,i)^2}{2\sigma(t)^2}\right) \quad (3-13)$$

mit der topologischen Distanz d zwischen dem Gewinnerneuron s und dem i -ten Neuron. Die Standardabweichung σ beschreibt den zeitabhängigen Radius der Nachbarschaft (Gl. 3-14):

$$\sigma(t) = \sigma_{Start} \left(\frac{\sigma_{Ende}}{\sigma_{Start}} \right)^{\frac{t}{t_{max}}} \quad (3-14)$$

mit der Standardabweichung zu Beginn σ_{Start} und der Standardabweichung am Ende des Lernprozesses σ_{Ende} .

Als letzter Schritt werden die trainierten Neuronen entsprechend der Topologie in einem regelmäßigen Muster (z.B. Schachbrettmuster) als Karte dargestellt. Die Felder dieser Karte, welche dem Einflussgebiet eines Neurons entsprechen, werden „rezeptive Felder“ genannt.

Im Rahmen dieser Arbeit wurden mit der Software MOLMAP® zweidimensionale Karten der hochdimensionalen Deskriptorräume erstellt.¹⁴¹ MOLMAP arbeitet mit einer toroidalen Topologie. Der Torus (in der Form eines Doughnuts) repräsentiert einen unendlichen zweidimensionalen Raum. Deshalb sind das obere und untere Ende sowie die linke und die rechte Seite der Karten mit einander verknüpft. MOLMAP verwendet eine Gaußsche Glockenkurve, um das Nachbarschaftsverhalten der Neuronen zu definieren. Distanzmessungen erfolgten in der Regel mit der Manhattan-Metrik, in Ausnahmen mit der euklidischen Metrik. Die jeweils verwendete Metrik ist im Text angegeben. Weitere Einstellungen waren: Anzahl der Trainingszyklen (t_{max}): ca. $10 \times$ Anzahl der Trainingsmuster; Radius der Nachbarschaft (σ_{Start}): $max(x\text{-Neuronen}, y\text{-Neuronen}) / 2$; Lernrate tau (ϵ_{Start}): 1,0.

3.7 Hauptkomponentenanalyse (PCA)

Die Hauptkomponentenanalyse (PCA, für engl. „Principal Component Analysis“) ist ein statistisches Verfahren, das häufig z.B. für die Dimensionsreduzierung, die verlustbehaftete Datenkompression, die Eigenschaftenextraktion und die Datenvisualisierung eingesetzt wird.^{120,142} Eine PCA berechnet die Achsen eines Datenraumes neu, sodass diese in Richtung der maximalen Varianz der Datenpunkte zu liegen kommen. Die neuen Achsen werden Hauptkomponenten genannt. Die erste Hauptkomponente verläuft in Richtung der maximalen Varianz der Gesamtdaten, die zweite Hauptkomponente orthogonal dazu in Richtung der verbliebenen, noch nicht erklärten Varianz, usw. Dies kommt einer Rotation des Koordinatensystems gleich, wobei die Hauptkomponenten durch eine gewichtete Linearkombination der ursprünglichen Raumachsen ausgedrückt werden. Der neue Raum ist dekorreliert und besitzt genau so viele Achsen (Hauptkomponenten) wie der ursprüngliche Raum. Eine Dimensionsreduzierung wird erreicht, wenn die Datenpunkte nicht durch alle Hauptkomponenten ausgedrückt werden, sondern nur die ersten Hauptkomponenten, welche für gewöhnlich einen Großteil der Varianz erklären. Formal umfasst die Hauptkomponentenanalyse die Berechnung des Mittelwertes und der Kovarianzmatrix der Datenpunkte und das anschließende Auffinden jener Eigenvektoren der Kovarianzmatrix, die den größten Eigenwerten entsprechen.¹⁴²

Die Hauptkomponentenanalyse wurde in dieser Arbeit mit der Software „Molecular Operating Environment“ (Chemical Computing Group, Montreal, Kanada) durchgeführt.⁹⁷ Bei molekularen Deskriptordaten kann die Hauptkomponentenanalyse z.B. zur Dimensionsreduzierung und zur Dekorrelierung und Standardisierung der Daten eingesetzt werden (vgl. Abschnitt 3.4.1).

3.8 Pharmakophorsuche

Ein Pharmakophor ist das Ensemble der sterischen und elektronischen Eigenschaften, das notwendig ist, um die optimale molekulare Interaktion mit einer spezifischen biologischen Zielstruktur sicher zu stellen und um deren biologische Wirkung zu aktivieren oder zu inhibieren (IUPAC-Definition).¹⁴³ Ein Pharmakophor repräsentiert kein reales Molekül oder eine reale Assoziierung funktionaler Gruppen, sondern ist ein rein abstraktes Konzept, das die gemeinsamen molekularen Wechselwirkungsmöglichkeiten einer Gruppe von Verbindungen darstellt. Ein Pharmakophor kann als der größte gemeinsame Nenner einer Gruppe aktiver Moleküle angesehen werden.¹⁴³

Die Erstellung von Pharmakophormodellen erfolgte in dieser Arbeit mit der Software „Molecular Operating Environment“ (MOE, Version 2007.09, Chemical Computing Group, Montreal, Kanada) mittels der Funktion „Pharmacophore Elucidation“ (Pharmakophorerkennung).⁹⁷ Die Pharmakophorerkennung berechnet mit Hilfe einer stochastischen Suche bis zu 300 dreidimensionale Konformere der untersuchten Moleküle. Im nächsten Schritt werden den verschiedenen funktionellen Gruppen und Substrukturen der Konformere potentielle Pharmakophoreigenschaften zugewiesen. Hier wurden die Eigenschaften „aromatisch ODER π -Ringsystem“, „hydrophob“, „Protonendonator-Projektionspunkt“, „Protonenakzeptor-Projektionspunkt“, „kationisch“, „anionisch“, „Carboxy-bioisostere Gruppe“ und „Amidinium-bioisostere Gruppe“ verwendet. Die Pharmakophoreigenschaften werden als harte Kugeln mit definiertem Radius und definierten Koordinaten auf die 3D-Konformere projiziert. Im darauf folgenden Schritt überlagert ein Minimierungsalgorithmus gleichartige Pharmakophoreigenschaften der Moleküle, um so eine räumliche Alignierung der Konformere zu erreichen. Als Ausgabe erhält man 3-Punkt-, 4-Punkt und 5-Punktpharmakophorüberlagerungen der Eingabemoleküle. Ein fertiges Modell kann in einer Pharmakophorsuche zur Identifizierung potentiell aktiver Verbindungen eingesetzt werden. Dies ist eine Variante des virtuellen Screenings.²¹

3.9 Automatisiertes Docking

Beim automatisierten molekularen Docking werden kleine Moleküle in eine Rezeptorbindetasche eingepasst. Das Ziel des Dockings ist die Identifizierung einer bioaktiven Konformation eines potentiellen Rezeptorliganden. Grundsätzlich müssen dabei zwei Probleme gelöst werden: a) Eine passende Konformation des Liganden muss gefunden werden (das Dockingproblem) und b) die Qualität des Rezeptor-Ligandenkomplexes muss abgeschätzt werden (das Bewertungsproblem).⁷

Es wurden verschiedene Heuristiken entwickelt um das NP-vollständige Dockingproblem in der Praxis in akzeptabler Zeit zu lösen. In dieser Arbeit wurde die Software GOLD (Cambridge Crystallographic Data Centre, Cambridge, Vereinigtes Königreich) eingesetzt, welche einen genetischen Algorithmus zum Platzieren des Liganden und zum Durchsuchen seines Konformationsraumes einsetzt.^{118,119} Eine Konformation wird durch ein „Chromosom“ repräsentiert – ein binärer Vektor der die verschiedenen Parameter des Dockingproblems (Translation, Rotation, Rotationswinkel frei drehbarer Bindungen) kodiert. Populationen von Chromosomen repräsentieren parallel verschiedene Lösungsvorschläge und die genetischen Operatoren „Mutation“ und „Cross-Over“ werden zur Variation erfolgreicher Kandidaten eingesetzt.

Die Qualität der Bindung an einen Rezeptor wird mittels einer Bewertungsfunktion berechnet und durch eine Punktzahl (engl. „Score“) angegeben. Neben der sterischen Passgenauigkeit geht auch eine Pseudoenergie der nicht kovalenten Wechselwirkungen zwischen Rezeptor und potentiellm Ligand in die Bewertungsfunktion ein. Je komplementärer Bindetasche und potentieller Ligand in Passform und Ausrichtung funktioneller Gruppen sind, desto höher die Punktzahl. Alle Dockingexperimente wurden mit GOLD (Versionen 3.2 und 4.0, The Cambridge Crystallographic Data Centre, Cambridge, Vereinigtes Königreich) durchgeführt. Als Bewertungsfunktion wurde *GoldScore* verwendet. *GoldScore* ist eine empirische Bewertungsfunktion und stellt eine gewichtete Summe aus den vier Termen „Wasserstoffbrückenenergie des Protein-Liganden-Komplexes“, „Van-der-Waals-Energie des Protein-Liganden-Komplexes“, „interne Van-der-Waals-Energie des Liganden“ und „interne Torsionsenergie des Liganden“ dar.^{118,119}

4 Multidimensionale Skalierung zur Validation von Optimierungsalgorithmen

4.1 Zielsetzung

Mit der Programmbibliothek *Cyclops* wurde ein modularer Rahmen für die iterative stochastische Optimierung in der Programmiersprache Java (Version 1.5, Sun Microsystems, Inc., Santa Clara, USA)⁵⁹ entwickelt. Innerhalb dieses Rahmens wurden die Optimierungsalgorithmen „Zufallssuche“, „Simulated Annealing“,⁴²⁻⁴⁴ „Evolutionstrategie“^{41,2,38} und „Partikelschwarmoptimierung“^{47,48} implementiert. Da es sich bei diesen Algorithmen um stochastische Verfahren handelt, ist eine analytische Überprüfung der korrekten Funktionsweise nicht möglich. Um die Optimierungsalgorithmen dennoch validieren zu können, wurde ein Optimierungsproblem gewählt, für das sowohl analytisch-deterministische als auch stochastisch-heuristische Lösungswege existieren.¹⁴⁴ Durch den Vergleich der analytischen Lösung mit der heuristischen Lösung der Optimierungsalgorithmen sollte so eine Validierung durchgeführt werden.

4.2 Die Multidimensionale Skalierung

Die korrekte Funktionsweise der in der Programmbibliothek *Cyclops* implementierten Optimierungsalgorithmen wurde anhand der Multidimensionalen Skalierung (MDS) überprüft.^{120,144} Ziel der klassischen MDS ist die Rekonstruktion der Raumkoordinaten von einander in Beziehung stehender Objekte, von denen nur deren Distanzmatrix bekannt ist. Die Distanzmatrix enthält alle paarweisen Abstände dieser Objekte. Zumeist ist die MDS gleichzeitig mit der Reduktion der Dimensionalität verbunden. In diesem Fall soll eine niederdimensionale Projektion der Objekte gefunden werden. Ist eine Koordinatenrekonstruktion oder -projektion z.B. auf Grund einer ungenauen Distanzmatrix oder auf Grund der Dimensionsreduzierung nicht exakt möglich, so verbleibt ein Fehlerterm zwischen der Eingabedistanzmatrix und der Distanzmatrix der rekonstruierten Koordinaten. Diesen Fehler gilt es zu minimieren. Eine analytische Lösung des MDS-Problems erfolgt mittels der Hauptkoordinatenanalyse (PCO, für engl. „Principal Coordinate Analysis“).¹⁴⁵ Bei der PCO wird die Distanzmatrix zunächst transformiert und dann einer Eigenzerlegung unterzogen. Aus Eigenvektoren und Eigenwerten können schließlich die rekonstruierten Koordinaten berechnet werden. Um das MDS-Problem mit Hilfe von Optimierungsalgorithmen zu lösen, wird eine Fehlerfunktion (STRESS-Funktion, Gl. 8-4) der Koordinatenrekonstruktion berechnet

und diese minimiert. Eine Minimierung läuft z.B. nach diesem Schema ab: Eine initiale Lösung der MDS wird generiert, indem allen Objekten zufällige Koordinaten zugewiesen werden. Die initiale Lösung wird durch die STRESS-Funktion bewertet. Dann wird eine Nachbarlösung durch Perturbation der Koordinaten der Initiallösung erzeugt und diese wird ebenfalls bewertet. Die bessere von beiden Lösungen wird beibehalten und dient als Ausgangspunkt für eine weitere Optimierungsrunde. Dieser Zyklus schreitet fort, bis ein geeignetes Abbruchkriterium erreicht ist. Eine mathematische Einführung in die MDS ist im Anhang im Abschnitt 8.2 zu finden.

4.3 Ergebnisse und Diskussion

Im folgenden Experiment wurde die Qualität der MDS der Hauptkoordinatenanalyse mit der von vier verschiedenen Optimierungsverfahren verglichen: Zufallssuche, Simulated Annealing, Evolutionsstrategie und Partikelschwarmoptimierung. Die PCO wurde mit Hilfe der Funktion *‘cmdscale’* des Statistikwerkzeugkastens von Matlab (Version 7.04.365 R14 SP2, The MathWorks, Natick, USA) berechnet.¹⁴⁶ Ziel der MDS war die Rekonstruktion der relativen Koordinaten der zehn größten deutschen Städte: Berlin, Hamburg, München, Köln, Frankfurt am Main, Stuttgart, Dortmund, Essen, Düsseldorf und Bremen (Stand: Februar 2008). Die Distanzmatrix enthielt die gegenseitigen Entfernungen entlang der Orthodromen („Luftlinie“). Diese wurden mit Hilfe des Programms Google Earth (Version 4.2.0205.5730, Google Inc., Mountain View, USA, zugegriffen am 14.02.2008)¹⁴⁷ auf 10 Meter genau bestimmt (Tabelle 4-1). Ein Lösungskandidat der vier Optimierungsalgorithmen repräsentierte die Koordinaten aller zehn Objekte (Städte). Ein Lösungskandidat wurde durch einen reellwertigen Vektor der Länge $n \cdot k$ repräsentiert, wobei k die Dimension des Projektionsraumes (hier: $k = 2$) und n die Anzahl der Objekte darstellt (hier: $n = 10$). Für die Initiallösungen wurden zufällige Koordinaten festgesetzt. Die Optimierungsalgorithmen wurden mit Strategieparametern initialisiert, die sich zuvor in einigen Probeläufen als vorteilhaft herausstellten (Tabelle 4-2). Das Simulated Annealing lieferte mit abgeschaltetem Metropolisakzeptanzkriterium die besten Ergebnisse. Der Temperaturfaktor mit exponentieller Abkühlungsrate floss in die Schrittweite bei der Erzeugung von Nachbarlösungen ein.

Tabelle 4-1: Distanzmatrix der zehn größten deutschen Städte. Distanzen entlang der Orthodromen in Kilometern, bestimmt mit dem Programm Google Earth (Version 4.2.0205.5730, zugegriffen am 14.02.2008). D'mund = Dortmund, D'dorf = Düsseldorf.

	Hamburg	München	Köln	Frankfurt	Stuttgart	D'mund	Essen	D'dorf	Bremen
Berlin	256,08	504,22	478,50	424,58	512,28	422,94	454,43	478,56	316,15
Hamburg		611,89	356,48	393,03	533,79	284,33	308,21	338,79	95,15
München			456,40	304,42	190,90	477,30	493,53	487,14	582,34
Köln				152,77	288,45	72,57	57,63	34,13	268,91
Frankfurt					152,61	177,65	190,14	182,76	329,47
Stuttgart						327,65	335,63	321,88	478,31
D'mund							31,71	57,51	196,37
Essen								30,79	217,33
D'dorf									248,11

Tabelle 4-2: Parameter der Optimierungsalgorithmen

Parameter	Wert
Allgemein	
Projektionsdimension	2
Zulässiger Koordinatenbereich	[-500,0; 500,0]
Wiederholungen (für Statistik)	100
Zufallssuche	
Iterationen	30000
Simulated Annealing	
Iterationen	30000
Metropolisakzeptanzkriterium	desaktiviert
Initialschrittweite σ_0	50,0
Finale Schrittweite σ_{Ende}	10^{-9}
Reduktionsrate der Schrittweite	$(\sigma_{\text{Ende}} / \sigma_0)^{1/\text{Iterationen}}$
Evolutionsstrategie	
Eltern	2
Populationsgröße	30
Iterationen	1000
Strategie	(μ, λ)
Schrittweite (mit Autoadaptation)	50,0
Maximale Schrittweite	100,0
Partikelschwarmoptimierung	
Schwarmgröße	30
Iterationen	1000
Maximalgeschwindigkeit	50,0
Kognitiver Faktor	2,05
Sozialer Faktor	2,05
Typ	Konstriktionsschwarm

Die Ergebnisse des Experiments bestätigten die korrekte Funktionsweise der implementierten Algorithmen (Tabelle 4-3). SA, ES und PSO waren gleichermaßen gut in der Lage, eine MDS durchzuführen. Die Endwerte ihrer Fitnessfunktionen unterscheiden sich erst ab der 14. Nachkommastelle. Interessanterweise lieferten die drei Optimierungsalgorithmen doppelt so gute Ergebnisse wie eine PCO. Die Standardabweichung der Distanzen lag mit rund 20 Metern im Bereich der Messungengenauigkeit der Ausgangsdistanzen. Eine Zufallssuche verlief erwartungsgemäß ineffizient.

Abbildung 4-1 zeigt ein Diagramm des Optimierungsverlaufes der MDS. Auf der Abszisse sind die Generationen aufgetragen und auf der Ordinate ist die Fitness (hier: der STRESS-Wert) der bis zur jeweiligen Generation besten gefundenen Lösung aufgetragen. Der Verlauf der Optimierungen in Abbildung 4-1 zeigt ein typisches Optimierungsverhalten. Zu Beginn einer Optimierung sind sehr große Schwankungen zwischen den Lösungskandidaten zu beobachten und die Qualität der Lösungen wird in wenigen Generationen dramatisch besser. Nach einer gewissen Anzahl von Iterationen konvergiert die Optimierung, die Schrittweite wird sehr klein und es wird nur noch eine Feinabstimmung der bis dahin besten Lösung vorgenommen (hier etwa ab Generation 200 – 300 bei ES und PSO und ab Generation 12.000 bei SA). Die populationsbasierten Strategien ES und PSO konvergieren schnell, müssen dafür je Generation eine Schar von Lösungen evaluieren. Das Simulated Annealing bevorzugt lange Optimierungszeiten mit vielen Iterationen.

Die mit Hilfe einer Evolutionsstrategie rekonstruierten Koordinaten der deutschen Großstädte sind in Abbildung 4-2 in zeitlicher Abfolge dargestellt (5., 11., 51. und 1000. Generation). Auffällig ist, dass schon nach nur 51 Generationen (Abbildung 4-2c) die ungefähren Positionen aller zehn Städte gefunden wurden. Die finale Lösung nach 1000 Generationen ist in Abbildung 4-2e mit einer Landkarte des Bundesgebietes überlagert. Da die Koordinaten nur anhand der Objektabstände untereinander optimiert wurden, sind die rekonstruierten Lösungen translations- und rotationsinvariant, sowie invariant bezüglich der Richtung der Achsen. Die Überlagerung in Abbildung 4-2e ist nicht exakt, da sie manuell mit einem Grafikprogramm vorgenommen wurde. Jedoch ist eine exakte Reproduktion der Koordinaten schon theoretisch nicht möglich, da eine zweidimensionale Projektion (wie hier geschehen) die Erdkrümmung außer Acht lässt.

Tabelle 4-3: Ergebnisse der Multidimensionalen Skalierung. Angegeben ist der Mittelwert ‚STRESS‘ und die Standardabweichung ‚Std (STRESS)‘ der STRESS-Funktion (Gl. 8-4) über 100 Optimierungsläufe. Für PCO wurde der Wert der STRESS-Funktion nachträglich berechnet. Die Wurzel des über alle Distanzen gemittelten STRESS-Wertes ‚Std (Distanzen) [m]‘ ist ein Maß für die mittlere Abweichung der Interstädtedistanzen.

Verfahren	STRESS	Std. (STRESS)	Std. (Distanzen) [m]
PCO	0,03863		29,3
Zufallssuche	1.070.688	151.734	154.250
SA	0,015578156952358	1,87E-10	18,6
ES	0,015578156952347	0,0	18,6
PSO	0,015578156952354	4,59E-10	18,6

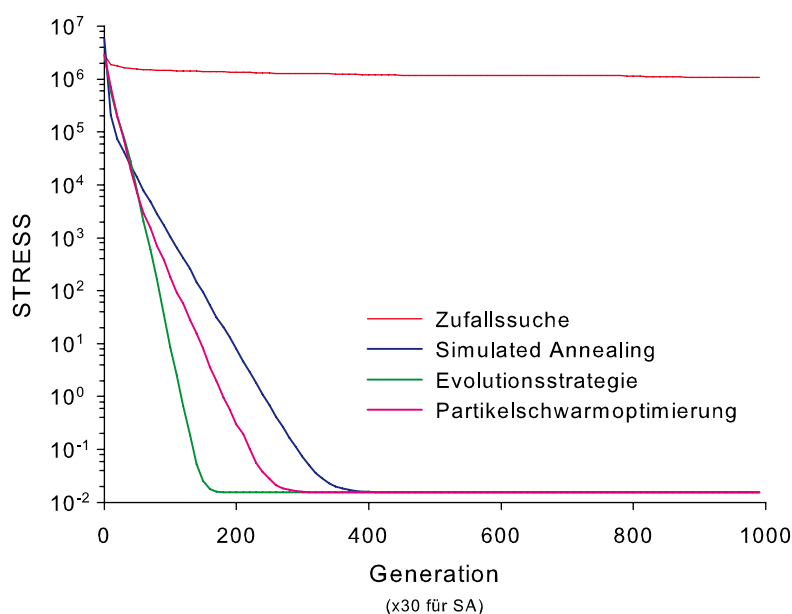


Abbildung 4-1: Die MDS mittels stochastischer Optimierungsverfahren. Optimierungsziel war die Minimierung der STRESS-Funktion (Ordinate, logarithmische Auftragung). Es wurde eine Zufallssuche (rot), das Simulated Annealing (blau), die Evolutionsstrategie (grün) und die Partikelschwarmoptimierung (pink) verglichen. Dargestellt ist der Median (100 Wiederholungen) der besten Lösung, die bis zur jeweiligen Generation gefunden wurde. Das Abbruchkriterium war für die Zufallssuche, ES und PSO das Erreichen der 1000sten Generation. Für die Zufallssuche und SA war dies die 30000ste Generation; die Abszisse (Generation) wurde hier entsprechend skaliert.

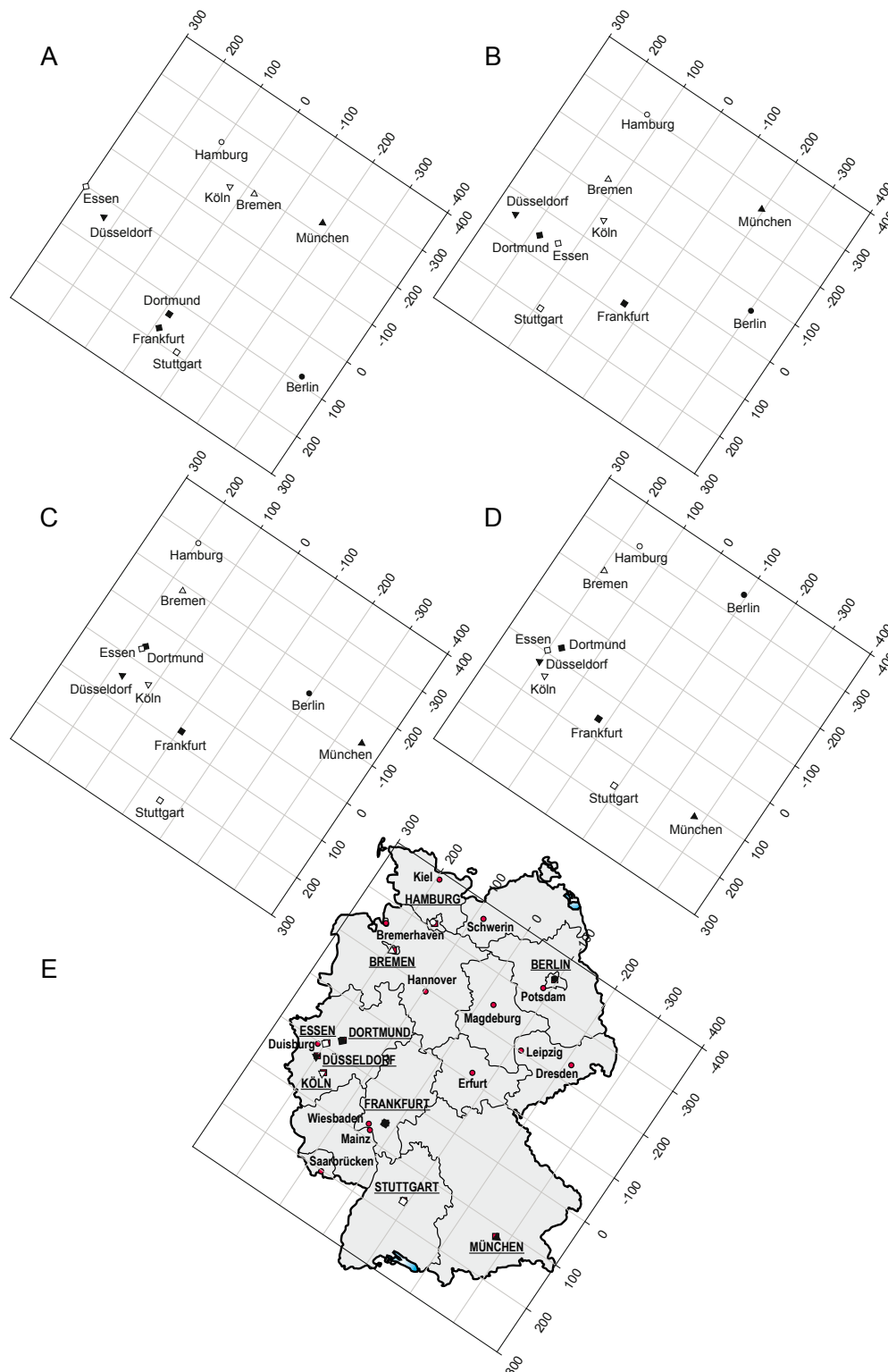


Abbildung 4-2: Evolution der Koordinaten der zehn größten deutschen Städte. Die MDS wurde mit Hilfe einer (2,30)-Evolutionstrategie (1000 Generationen) durchgeführt. Skalenbereich der Achsen in Kilometern. A: 5. Generation, STRESS = 1.378.012; B: 11. Generation, STRESS = 695.014; C: 51. Generation, STRESS = 321,082; D: 1000. Generation, STRESS = 0,0156; E: Die Koordinaten der 1000. Generation wurden hier mit einer Karte des Bundesgebietes unterlegt.

4.4 Schlussfolgerungen

Vier Optimierungsalgorithmen (Zufallssuche, Simulated Annealing (SA), Evolutionsstrategie (ES) und Partikelschwarmoptimierung (PSO)) wurden auf das Problem der Multidimensionalen Skalierung angewendet und mit dem analytischen Verfahren der Hauptkoordinatenanalyse (PCO) verglichen. Die stochastischen Optimierungsalgorithmen SA, ES und PSO lieferten doppelt so gute Ergebnisse wie die PCO. Eine Zufallssuche verlief erwartungsgemäß sehr ineffizient. In einem MDS-Experiment zur Rekonstruktion der relativen Position der zehn größten deutschen Städte konnten die ursprünglichen Koordinaten auf ± 20 Meter genau bestimmt werden (Standardabweichung). Die PCO erreichte eine Genauigkeit von ± 30 Metern. Die Anwendung der Optimierungsalgorithmen auf das Problem der multidimensionalen Skalierung zeigt die korrekte Funktionsweise der implementierten Algorithmen.

5 Retrospektive Vergleichsstudie von Optimierungsalgorithmen

5.1 Zielsetzung

Obwohl in der Vergangenheit verschiedene Optimierungsalgorithmen mit Erfolg für den Entwurf fokussierter Substanzbibliotheken eingesetzt wurden, wurden diese noch nicht systematisch miteinander verglichen. Die implementierten Algorithmen wurden deshalb in diesem Teil der Arbeit in Bezug auf den Entwurf von Substanzbibliotheken einem systematischen Vergleich unterzogen. Dazu wurden retrospektive Optimierungsexperimente mit Serinproteaseinhibitoren in verschiedenen Deskriptorräumen durchgeführt. Die zwei wesentlichen Fragestellungen waren:

- Welcher Optimierungsalgorithmus ist für das Design fokussierter Substanzbibliotheken am besten geeignet?
- Wie sollte die Populationsgröße gewählt werden und wie viele Iterationen sollten durchgeführt werden?

Die Aufklärung dieser Fragen war zusätzlich Voraussetzung für die Planung einer prospektiven Studie. In den folgenden Experimenten wurden die drei Optimierungsalgorithmen „Simulated Annealing“,⁴²⁻⁴⁴ „Evolutionstrategie“^{1,2,38} und „Partikelschwarmoptimierung“^{47,48} mit einer Zufallssuche verglichen.

Diese Studie wurde im Hinblick auf das Niedrigdurchsatzscreening von TAR-RNA-Liganden mit limitierten finanziellen und personellen Ressourcen durchgeführt (vgl. Kapitel 6). Es sollten daher zwei zusätzliche Einschränkungen gelten, um Laborbedingungen zu simulieren:

- Die maximale Anzahl an Substanztestungen ist beschränkt.
- Eine kleine Anzahl an Iterationen ist bevorzugt.

Für die Anzahl der Substanztestungen wurde ein Maximalbudget festgelegt. Dieses Budget gibt in einer prospektiven Studie an, wie viele Moleküle für eine Synthese oder den Erwerb mit anschließender biologischer Testung zur Verfügung stehen. Hält man diese Zahl klein, so kann der finanzielle und zeitliche Aufwand einer Niedrigdurchsatzscreening-Kampagne entscheidend verringert werden.

Die zweite Einschränkung hat hauptsächlich einen Einfluss auf den zeitlichen Aspekt einer solchen Screeningkampagne. Wegen der langwierigen Arbeitsschritte der Substanzsynthese (oder des Einkaufs/der Lieferung) und der biologischen Testung, sollte die Anzahl der Iterationen klein gehalten werden, damit neue Verbindungen gebündelt, *en gros* beschaffen und getestet werden können. Wesentliche Teile der hier beschriebenen Vergleichsstudie wurden erst kürzlich publiziert.¹³⁰

5.2 Methodik

Für die Optimierungsexperimente standen zwei molekulare Bibliotheken zur Verfügung, die bereits mit biologischen Aktivitätsdaten annotiert waren. Die Ugi-Bibliothek ist eine kombinatorische Bibliothek aus 15.840 Reaktionsprodukten einer Ugi-artigen Drei-Komponenten-Reaktion, die uns freundlicher Weise von Dr. L. Weber (Roche, Basel) zur Verfügung gestellt wurde (vgl. Abschnitt 3.3.1).⁹³

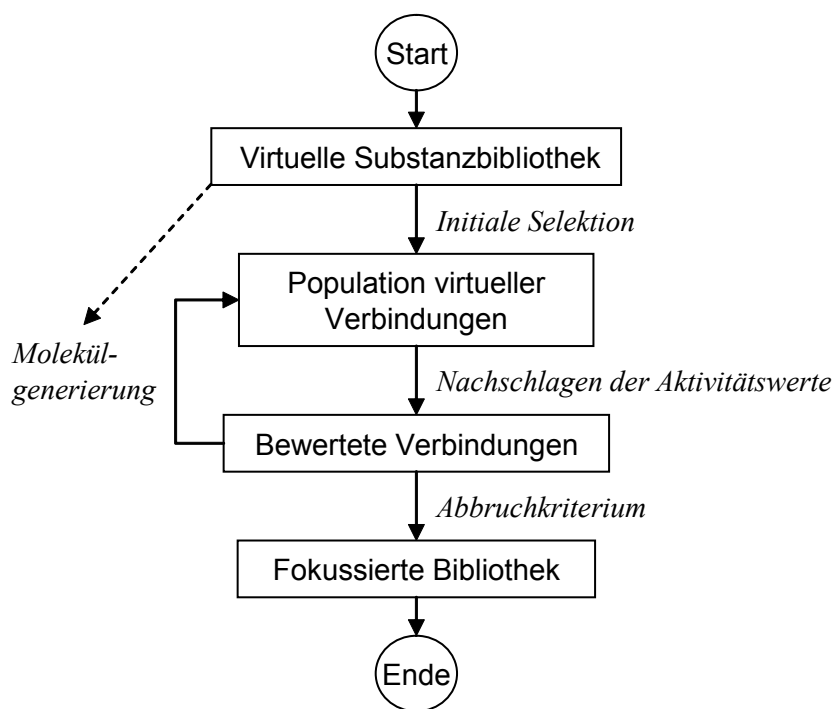


Abbildung 5-1: Schema der retrospektiven Optimierungsexperimente. Eine Initialpopulation von Molekülen wird aus einer Substanzbibliothek ausgewählt. Die Aktivitätswerte der ausgewählten Verbindungen werden in der Liste aller Aktivitätsdaten nachgeschlagen. Mit Hilfe der bewerteten Verbindungen wird eine neue und hoffentlich aktivere Generation von Molekülen erzeugt. Die Generierung neuer Moleküle greift auf den Molekülvorrat der verwendeten Substanzbibliothek zurück (gestrichelte Linie). Diese Schleife aus Molekülgenerierung und Bewertung wird so lange wiederholt bis ein Abbruchkriterium erfüllt ist.

Für die Moleküle der Ugi-Bibliothek wurden von Dr. L. Weber und Mitarbeitern die mittleren inhibitorischen Konzentrationen (IC_{50}) gegen fünf verschiedene Serinproteasen bestimmt. Als zweite Bibliothek stand COBRA (Version 6.1) zur Verfügung.⁹⁵ COBRA ist eine Sammlung von 8311 bioaktiven, wirkstoffartigen Referenzanaloga. Diese waren mit Bioaktivitätsdaten ihres jeweiligen Zielrezeptors versehen (vgl. Abschnitt 3.3.2). Für die zwei Bibliotheken wurden drei verschiedenen computerchemische Repräsentationen berechnet (vgl. Abschnitt 3.4.1): a) 146 MOE-2D-Deskriptoren, b) 16 MOE-2D-Hauptkomponenten (MOE 2D PCA) und c) der 150-dimensionale CATS-Deskriptor. Die Optimierungen wurden mit den Optimierungsalgorithmen „Zufallssuche“, „Simulated Annealing“,⁴²⁻⁴⁴ „Evolutionstrategie“^{41,2,38} und „Partikelschwarmoptimierung“^{47,48} durchgeführt und folgten dem Schema in Abbildung 5-1.

Die maximale Anzahl der Substanztestungen wurde für die Ugi-Bibliothek auf 300 (1,9% der Bibliotheksmoleküle), für COBRA auf 150 (1,8%) beschränkt. Die Anzahl von 300, respektive 150, bezeichnet das oben erwähnte Maximalbudget an Verbindungen, die in einer prospektiven Studie für die Testung zur Verfügung stehen.

Das beschränkte Gesamtbudget der zu testenden Verbindungen kann in unterschiedlichen Verhältnissen auf die Populationsgröße und die Anzahl der Generationen aufgeteilt werden. Das Produkt aus Populationsgröße und Anzahl der Generationen darf dabei nicht größer als das Maximalbudget werden. Um die bestmögliche Kombination der zwei Parameter zu finden, wurden retrospektive Optimierungsläufe durchgeführt und die Ergebnisse verglichen. 14 verschiedenen Kombinationen von „Populationsgröße“ \times „Anzahl der Generationen“ wurden verglichen. Für die Ugi-Bibliothek waren diese: 1×300 , 2×150 , 3×100 , 5×60 , 10×30 , 15×20 , 20×15 , 30×10 , 37×8 , 50×6 , 60×5 , 75×4 , 100×3 und 150×2 . Die verglichenen Kombinationen für COBRA lauteten: 1×150 , 2×75 , 3×50 , 4×37 , 5×30 , 7×21 , 10×15 , 15×10 , 21×7 , 25×6 , 30×5 , 37×4 , 50×3 und 75×2 . Retrospektive Optimierungsexperimente wurden mit Hilfe der entwickelten Programmbibliothek *Cyclops* durchgeführt.

Das Optimierungsziel war die Minimierung des IC_{50} -Wertes ausgewählter Bibliotheksmoleküle. Als Qualität eines Optimierungslaufes wurde der gemittelte IC_{50} -Wert der Top-20 aktivsten Verbindungen aller Generationen eines Laufes gewertet. Das Optimierungsziel war also die Erstellung einer fokussierten Substanzbibliothek aus 20 Verbindungen mit minimalen IC_{50} -Werten. Die Optimierungsexperimente der Vergleichsstudie wurden jeweils 100-fach wiederholt. In den Tabellen und Abbildungen

dieses Kapitels ist der Median dieser Wiederholungen aufgetragen. Der aufgetragene Wert berechnete sich damit nach Gl. 5-1:

$$IC_{50} = Median \left(\frac{1}{20} \sum_{i=1}^{20} (IC_{50}^{Top-20})_{i,j} \right) \quad , \text{ mit } \quad j = 1 \dots 100 \quad (5-1)$$

5.2.1 Strategieparameter

Ein Optimierungsalgorithmus benötigt gewisse Parameter, um an die jeweiligen Problemstellungen angepasst zu werden („Strategieparameter“). In der vorliegenden Arbeit lag das Augenmerk auf der Entwicklung einer robusten, allgemeingültigen Methode für das Design fokussierter Substanzbibliotheken.

Tabelle 5-1: Parameter der verwendeten Optimierungsalgorithmen

Parameter	Wert
Simulated Annealing	
Initialtemp. T_0	75,0
Initialtemp. T_0 (mit Initialaktiven) ^a	20,0
Endtemp. T_{Ende}	1,0
Abkühlrate	$(T_{Ende} / T_0)^{1/Iterationen}$
Evolutionsstrategie	
Strategie ^b	(μ, λ) (Elitismus: 1)
Anzahl der Eltern μ^c	Populationsgröße / 6
Maximale Schrittweite σ_{max}	5,0
ES-Mutationsoperator (verwendet von ES und SA) ^d	
Schrittweite σ_0 (CATS)	2,0
Schrittweite σ_0 (MOE 2D)	1,5
Schrittweite σ_0 (MOE 2D PCA)	0,75
Partikelschwarmoptimierung	
Kognitiver Faktor c	2,05
Sozialer Faktor s	2,05
Maximale Geschwindigkeit	unbeschränkt
Konstruktionsfaktor	vgl. Gl. 3-10

^a Für einige Experimente wurde die Initialpopulation mit bekannten Aktiven angereichert.

^b Die Strategie entspricht einer (μ, λ) -ES, wobei jedoch ein Elternindividuum pro Generation in die Nachkommengeneration übernommen werden kann (Elitismus: 1).

^c Die Anzahl der Eltern wurde in Abhängigkeit von der jeweiligen Populationsgröße gewählt, um für verschiedene Populationsgrößen einen konstanten Selektionsdruck zu gewährleisten.

^d ES verwendet eine autoadaptive Schrittweitenregelung, SA eine fixe Schrittweite.

Es wurde deshalb keine erschöpfende Optimierung aller Suchparameter für die zahlreichen Kombinationen aus Suchräumen, Optimierungsalgorithmen und Optimierungszielen durchgeführt. Vielmehr wurde ein allgemeiner Satz an Parametern mit Hilfe der Ugi-Bibliothek bestimmt (Tabelle 5-1). Zur Bedeutung der einzelnen Parameter sei auf Abschnitt 3.5 verwiesen. Das Optimierungsverhalten von SA, ES und PSO wurde so eingestellt, dass ein Lösungskandidat niemals zweimal ausgewählt werden konnte. Dies kommt einer globalen *Tabu*-Regel gleich. Die Positionierung der Lösungskandidaten war auf einen Bereich beschränkt, der durch den Minimal- und Maximalwert jeder Dimension des Suchraumes gegeben war. Durch die Anwesenheit von vielen inaktiven Molekülen in den Datensätzen, die alle den gleichen Fitnesswert besaßen („inaktiv“), konnte eine effiziente Optimierung erst dann erfolgen, wenn eine erste aktive Verbindungen vom Suchalgorithmus gefunden wurde. Die zufällige Auswahl der Initialpopulation wurde deshalb so lange wiederholt, bis zumindest ein aktives Molekül gefunden wurde, oder die Abbruchbedingung erreicht wurde. Die gemeinsamen Eigenschaften aller vier implementierten Suchalgorithmen waren: a) ein Lösungskandidat repräsentierte genau ein Molekül, b) Moleküle wurden durch molekulare Deskriptorvektoren kodiert und c) das Optimierungsziel war die Minimierung der Zielfunktion (IC_{50}).

5.3 Komparative Parameterselektion

Die Ergebnisse der besten Kombinationen von Populationsgröße und Anzahl der Generationen sind in Tabelle 5-2 (CATS), Tabelle 5-3 (MOE 2D) und Tabelle 5-4 (MOE 2D PCA) zusammengestellt. Die Tabellen zeigen die Ergebnisse für zwei verschiedene Arten der Selektion der Initialpopulation: a) zufällige Selektion und b) Einbeziehung von vier aktiven Verbindungen in die Initialpopulation. Zunächst werden die Ergebnisse des ersten Selektionstyps diskutiert.

Mit der Ausnahme von Chymotrypsin übertrafen alle drei Optimierungsalgorithmen klar die Zufallssuche beim Auffinden von Serinproteaseinhibitoren. Das schlechte Abschneiden beim Auffinden von Chymotrypsin-Inhibitoren ist durch deren geringe Häufigkeit in der Ugi-Bibliothek erklärt: Nur 34 der insgesamt 15.840 Moleküle (0,2%) weisen Aktivität gegen Chymotrypsin auf. Das zugrunde liegende Fitnessgebirge ist durch die überwiegende Anzahl inaktiver Moleküle informationsarm. Das fundamentale Prinzip der iterativen Optimierungsalgorithmen ist das der starken Kausalität: ähnliche Änderungen im Suchraum sollten ähnliche Änderungen in der Qualität eines Lösungskandidaten bewirken.^{2,132-134} Wenn die lokale Nachbarschaft

eines Optimums jedoch keinerlei Hinweise auf das selbst Optimum bietet (z.B. in der Form eines Gradienten), kann ein Optimierungsalgorithmus im besten Fall so gut wie eine Zufallssuche arbeiten. Da Optimierungen von Trypsin-Inhibitoren (1,9% Aktive in der Ugi-Bibliothek) erfolgreich durchgeführt werden konnten, wird die Häufigkeitsuntergrenze von Molekülen für die adaptive Optimierung von Substanzbibliotheken auf den Bereich zwischen 0,2% und 1,9% geschätzt.

Es gibt keinen klaren Gewinner unter den analysierten Suchalgorithmen. Alle Optimierungsalgorithmen (mit Ausnahme der Zufallssuche) waren in der Lage, gute Ergebnisse zu produzieren. Mit geeigneten Parametern schnitten alle drei Algorithmen vergleichbar gut ab. Die mittlere Qualitätsdifferenz (ΔIC_{50}) aller optimierter, fokussierter Substanzbibliotheken (Tabellen 5-2, 5-3 und 5-4) war $3,3 \pm 3,9 \mu\text{M}$ und überschritt nie einen Wert von $20 \mu\text{M}$. Mit einer Ausnahme (Trypsin-Inhibitoren) lieferten Optimierungen im CATS-Deskriptorraum die besten Ergebnisse. Interessanterweise schnitten die 146 MOE-2D-Deskriptoren und die 16 MOE-2D-Hauptkomponenten ähnlich ab. Mit nur 16 Hauptkomponenten (~11% von 146) konnten vergleichbar gute Ergebnisse wie mit dem vollständigen Deskriptorsatz erzielt werden. Die Dimensionsreduzierung beschleunigte die Optimierungsläufe zudem merklich.

In vorläufigen Optimierungsexperimenten wurde realisiert, dass die Häufigkeit der Aktiven in den Datensätzen nicht nur den minimalen IC_{50} der fokussierten Bibliotheken beeinflusste, sondern auch die Größe der statistischen Schwankung der Lösungen. Ein Optimierungsverfahren, das mit geringer statistischer Schwankung gute Ergebnisse liefert, sei hier als ein *robustes* Verfahren definiert. In einer prospektiven Studie würde man nur einen oder einige wenige Optimierungsläufe durchführen, jedoch sicherlich keine 100 Wiederholungen, wie es hier zur statistischen Datenerhebung geschehen ist. Große Fehlerspannen (ausgedrückt durch die Differenz des oberen und unteren Quartils, Tabellen 5-2, 5-3 und 5-4) sind deshalb ein Indikator für schwache Robustheit. Eine nicht robuste Optimierungsstrategie birgt in einer prospektiven Studie die Gefahr, mit einer hohen Wahrscheinlichkeit schlechte Ergebnisse zu liefern. Optimierungen der Trypsin-Inhibitoren (1,9% Aktive) wiesen die größten Fehlerspannen auf (ungefähr $30 \mu\text{M}$). Im Vergleich dazu waren Fehlerspannen von Optimierungen der mindesten vier Mal so häufigen Faktor X_a -, Trypsin-, und uPA-Inhibitoren deutlich kleiner.

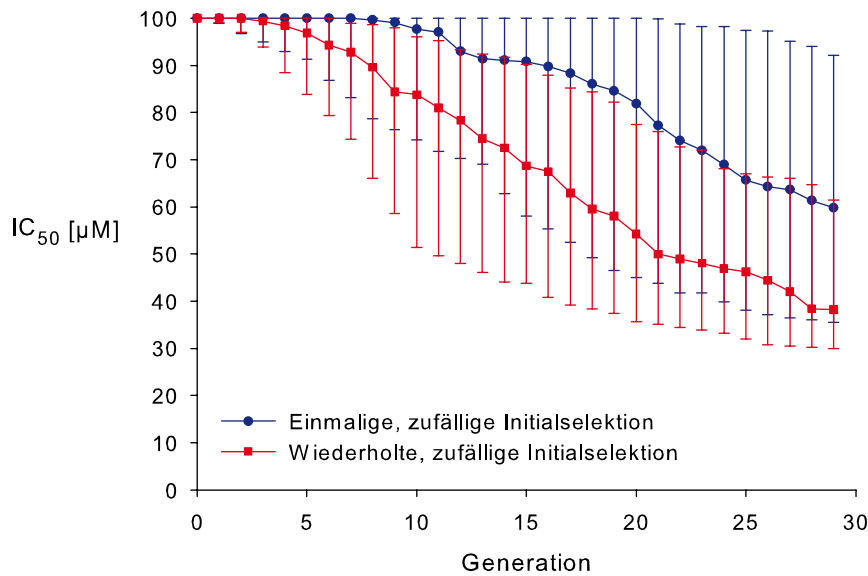


Abbildung 5-2: Auswirkung der Selektion der Initialpopulation auf den Optimierungsverlauf. Aufgetragen ist der gemittelte IC_{50} der Top-20 aktivsten Verbindungen als Median von 100 Wiederholungen (Gl. 5-1). Fehlerbalken geben das untere und obere Quartil an. Bei Optimierungsläufen mit wiederholter, zufälliger Selektion der Initialpopulation setzt die Minimierung der Fitness (IC_{50}) früher ein, die Fitness erreicht einen besseren Endwert und die Fehlerbalken sind merklich kleiner als bei einmaliger, zufälliger Initialselektion. Es wurde eine Optimierung von Trypsin-Inhibitoren im CATS-Deskriptorraum mittels einer (2, 10)-ES (30 Generationen) durchgeführt.

Ausgehend von den vorläufigen Experimenten konnten die Optimierungsläufe stabilisiert werden, indem die zufällige Selektion der Initialpopulation so oft wiederholt wurde, bis mindestens eine aktive Verbindung gefunden wurde. Der Erfolg dieser Vorgehensweise ist in Abbildung 5-2 exemplarisch für eine Optimierung von Trypsin-Inhibitoren im CATS-Deskriptorraum mittels einer (2, 10)-ES (30 Generationen) dargestellt. Bei Optimierungsläufen mit wiederholter zufälliger Selektion der Initialpopulation setzt die Minimierung der Fitness (IC_{50}) früher ein, die Fitness erreicht einen besseren Endwert und die Fehlerbalken sind merklich kleiner als bei einmaliger, zufälliger Initialselektion. Alle in Tabellen 5-2, 5-3 und 5-4 gezeigten Ergebnisse wurden deshalb mit wiederholter, zufälliger Initialselektion berechnet. Trotz dieser deutlichen Verbesserung sind die Fehlerspannen bei hundertfacher Wiederholung noch groß. Würde ein solches Verfahren in einer prospektiven Studie eingesetzt, bestünde die Möglichkeit ein schlechtes Ergebnis zu erhalten.

Die Qualität eines Optimierungslaufs war stark von der Qualität der Initialpopulation abhängig. Wurden schon früh im Optimierungszyklus hochaktive Verbindungen

gefunden, so waren die Optimierungen im Allgemeinen erfolgreicher. Die Wahrscheinlichkeit P , mindestens einen aktiven Inhibitor in einer Initialpopulation aus z.B. 30 Molekülen zu finden, ist (Gl. 5-2):

$$P = 1 - \prod_{i=0}^{29} \frac{\text{gesamt} - i - \text{Aktive}}{\text{gesamt} - i} \quad (5-2)$$

wobei *gesamt* die Anzahl aller Bibliothekssubstanzen und *Aktive* die Anzahl aktiver Verbindungen ist. Gemäß Gl. 5-2 ist die Wahrscheinlichkeit, wenigstens einen aktiven Faktor X_a -Inhibitor in einer Initialpopulation aus 30 Verbindungen zu finden, 0,97. Für Trypsin ist sie 0,44 und für Chymotrypsin 0,06. Analog waren die Optimierungsergebnisse für Faktor X_a gut, für Trypsin waren sie zufrieden stellend, aber mit großen Fehlerspannen und für Chymotrypsin nicht besser als der Zufall. Um die statistischen Schwankungen zu reduzieren und damit die Robustheit zu erhöhen, wurde deshalb die Entscheidung getroffen, vier aktive Verbindungen in die Initialpopulation mit einzubeziehen. Dies sollte die Optimierungsläufe weiter stabilisieren. Die vier Aktiven wurden zufällig aus den 50 aktivsten Faktor X_a -Inhibitoren, den 20 aktivsten Trypsin-Inhibitoren und den 10 aktivsten Chymotrypsin-Inhibitoren ausgewählt. Die Enzyme Faktor X_a , Trypsin und Chymotrypsin wurden wegen ihres ansteigenden Schwierigkeitsfaktors für die Optimierung – der sich wiederum aus den abnehmenden Anzahlen aktiver Verbindungen ergibt – als Testsystem ausgewählt. Die Wahrscheinlichkeiten, zufällig wenigstens vier Aktive in der Initialpopulation zu finden, sind nun wesentlich geringen: 0,41 (Faktor X_a), 0,0025 (Trypsin) und $4,7 \times 10^{-7}$ (Chymotrypsin, berechnet mit Matlab, Version 7.04.365 R14 SP2 mit installiertem Statistikwerkzeugkasten, The MathWorks, Natick, USA).¹⁴⁶ Die Einbeziehung aktiver Verbindungen sollte den Optimierungen folglich einen besseren Start ermöglichen, da nun die Zeit bis zum Auffinden einer ersten aktiven Verbindung auf Null reduziert ist und somit die volle Optimierungszeit für die Minimierung des IC_{50} -Wertes zur Verfügung steht.

Tatsächlich führte die Einbeziehung vier initialaktiver Verbindungen zu einer merklichen Verbesserung des erreichten minimalen IC_{50} -Wertes für Faktor X_a und Trypsin. Die beste Strategie für Faktor X_a erreichte einen minimalen IC_{50} von 5,8 μM (Tabelle 5-3, ES), was dem globalen Optimum von 2,3 μM (Tabelle 3-1) sehr nahe kommt. Gleichzeitig waren die Ergebnisse robuster: Die Fehlerspannen für fokussierte Trypsininhibitor-Bibliotheken reduzierten sich von 30 μM auf rund 8 μM .

Für Chymotrypsin-Inhibitoren, dem schwierigsten Optimierungsziel, ergab sich nur eine marginale Verbesserung. Es sei angemerkt, dass die Initialtemperatur bei SA-Experimenten mit bekannten aktiven Verbindungen in der Initialpopulation von 75 auf 20 reduziert werden musste. Andernfalls würde SA die initialaktiven Verbindungen wegen einer zu hohen Metropolisakzeptanzwahrscheinlichkeit (Gl. 3-5) schon in der ersten Iteration wieder verwerfen.

Alle drei verglichenen Optimierungsalgorithmen bevorzugten unterschiedliche Kombinationen von Populationsgröße und Anzahl der Generationen (Abbildungen 5-3 und 5-4). SA bevorzugte eindeutig eine große Anzahl an Generationen und folglich eine kleine Populationsgröße. In diesem Fall bezieht sich die Populationsgröße auf die Anzahl der simultanen Initialisierungen der Multi-Start-Implementierung. SA zeigte eine näherungsweise lineare Abhängigkeit von der Anzahl der Iterationen, was inhärent durch die Funktionsweise von SA bedingt ist: SA basiert auf einer thermischen Äquilibrierung eines physikalischen Systems, das allmählich abgekühlt wird. Je langsamer die Temperatur gesenkt wird, desto mehr Zeit bleibt für den Temperatureausgleich. Dies bedeutet für das Simulated Annealing, dass es umso effizienter arbeitet, je mehr Iterationen durchlaufen werden, also je länger die Abkühlungsphase ist. Jamois et al. machten 2003 in ihrem Vergleich von SA und GA bereits eine ähnliche Beobachtung.³³ Diese Charakteristik des SA steht im Widerspruch zur zweiten Nebenbedingung der Optimierungsexperimente: „Eine kleine Anzahl an Iterationen ist bevorzugt“. Angenommen ein Zyklus aus Synthetisierung und Testung dauerte in einem Niedrigdurchsatzscreening-Experiment ca. zwei Wochen, so würde eine Studie mit 60 Generationen 2,3 Jahre in Anspruch nehmen – für eine Studie mit einem Budget von nur 300 Substanzen ist dies sicherlich inakzeptabel.

PSO lieferte mit mittleren bis hohen Iterationszahlen (also mit mittleren bis kleinen Populationsgrößen) die besten Ergebnisse. PSO scheint deshalb generell für das Design fokussierter Substanzbibliotheken geeignet zu sein. PSO und ES schnitten ähnlich gut ab, obgleich PSO-Ergebnisse bei Optimierungen mit einbezogenen Aktiven in der Initialpopulation eine größere Fehlerspanne aufwiesen. Jedoch produzierte die PSO im Falle des MOE-2D-Dekriptorraumes auch ohne die Einbeziehung initialaktiver Substanzen hervorragende Ergebnisse.

ES hingegen bevorzugte kleine bis mittlere Iterationszahlen und somit große Populationen. Dies macht sie besonders für Niedrigdurchsatzscreening-Kampagnen attraktiv. In fast allen Fällen lieferten schon Evolutionsstrategien mit nur vier

Generationen (Populationsgröße: 75) überzeugende Ergebnisse. Um das obige Beispiel eines Niedrigdurchsatzscreening-Experimentes noch einmal anzuführen: Vier Zyklen aus Synthese und Testung könnten innerhalb eines akzeptablen Zwei-Monats-Zeitraums abgeschlossen werden. Wurde die Initialpopulation mit Aktiven angereichert, so reichten die bevorzugten Kombinationen von „Populationsgröße“ \times „Anzahl der Generationen“ von 20×15 bis 75×4 und bieten somit ausreichend Spielraum für die experimentelle Planung einer Niedrigdurchsatzstudie.

Ein ungewöhnlicher Einbruch in der Qualität der Optimierungsergebnisse für ES ist für Populationsgrößen von 1 bis 5 in Abbildungen 5-3 und 5-4 sichtbar. Dieses Verhalten war eine Konsequenz der Regel, nach der die Anzahl der Eltern berechnet wurde: *Populationsgröße / 6* (vgl. Tabelle 5-1). Diese Regel sorgte in der Vergleichsstudie über alle Populationsgrößen hinweg für einen gleichmäßigen Selektionsdruck. Für Populationen von einem bis fünf Individuen wurde deshalb ein einziges Elternindividuum pro Generation ausgewählt. Von vier Initialaktiven wurden somit drei in der ersten Generation verworfen. Gleichzeitig sank die Anzahl der Generationen von 300 auf 60, was die Optimierungsbedingungen weiter verschärfte.

Tabelle 5-2: Ergebnisse einer vergleichenden Optimierungsstudie mit einem Budget von 300 Verbindungen auf der durch den CATS-Deskriptor repräsentierten Ugi-Bibliothek.^a

	Zufallssuche	Simulated Annealing	Evolutionsstrategie	Partikelschwarm
Chymotrypsin				
Min. IC ₅₀ [μM]	109,4	108,6	108,3	109,0
Quartile	105,0 – 110,0	104,7 – 110,0	104,7 – 110,0	104,6 – 110,0
Aktive	0,0	1,0	1,0	1,0
Quartile	0,0 – 1,0	0,0 – 1,0	0,0 – 2,0	0,0 – 1,3
Pop.-Gr. x Gen.		10x30	2x150	5x60
Chymotrypsin mit vier Initialaktiven				
Min. IC ₅₀ [μM]	109,5	96,0	101,0	94,2
Quartile	106,3 – 110,0	92,2 – 104,7	95,9 – 105,0	87,5 – 105,0
Aktive	0,0	3,0	2,0	3,0
Quartile	0,0 – 1,0	1,0 – 4,0	1,0 – 3,0	1,0 – 5,0
Pop.-Gr. x Gen.		3x100	100x3	2x150
Faktor X_a				
Min. IC ₅₀ [μM]	36,1	8,4	8,3	7,5
Quartile	32,6 – 40,0	6,9 – 11,0	6,6 – 10,2	6,6 – 11,2
Aktive	32,0	158,0	164,0	101,5
Quartile	29,0 – 36,0	129,8 – 174,3	133,0 – 177,3	86,8 – 137,0
Pop.-Gr. x Gen.		1x300	20x15	50x6
Faktor X_a mit vier Initialaktiven				
Min. IC ₅₀ [μM]	36,4	6,6	6,4	6,0
Quartile	32,7 – 40,5	5,6 – 8,3	5,2 – 7,4	5,2 – 6,6
Aktive	32,0	151,5	123,0	127,0
Quartile	28,0 – 35,0	127,8 – 168,0	113,8 – 140,5	118,8 – 169,3
Pop.-Gr. x Gen.		3x100	37x8	5x60
Trypsin				
Min. IC ₅₀ [μM]	90,4	35,1	30,8	31,5
Quartile	86,8 – 93,6	27,7 – 48,5	26,8 – 41,3	26,2 – 47,6
Aktive	5,0	46,0	52,5	56,0
Quartile	4,0 – 7,0	30,0 – 58,3	33,0 – 66,0	31,0 – 64,0
Pop.-Gr. x Gen.		1x300	15x20	2x150
Trypsin mit vier Initialaktiven				
Min. IC ₅₀ [μM]	90,8	26,7	26,8	24,0
Quartile	87,3 – 93,8	24,1 – 30,3	24,9 – 30,0	22,7 – 27,2
Aktive	6,0	67,0	65,5	70,0
Quartile	4,0 – 7,0	57,8 – 74,0	62,0 – 70,3	65,0 – 75,3
Pop.-Gr. x Gen.		3x100	50x6	5x60
Tryptase				
Min. IC ₅₀ [μM]	18,6	3,9	2,7	2,8
Quartile	17,1 – 20,2	2,9 – 5,3	2,2 – 4,3	2,4 – 5,1
Aktive	89,0	244,0	235,0	262,0
Quartile	84,0 – 94,0	234,0 – 250,0	224,8 – 243,0	254,8 – 269,0
Pop.-Gr. x Gen.		1x300	30x10	3x100
uPA				
Min. IC ₅₀ [μM]	59,4	21,3	17,4	15,5
Quartile	54,9 – 64,4	16,7 – 31,5	16,0 – 22,5	14,2 – 19,7
Aktive	26,0	118,0	125,0	144,0
Quartile	23,0 – 30,0	97,8 – 135,0	115,8 – 137,0	131,8 – 160,3
Pop.-Gr. x Gen.		1x300	37x8	20x15

^a Pop.-Gr. x Gen. bezeichnet die beste Kombination von Populationsgröße und Anzahl der Generationen. Ausschlaggebend war der gemittelte IC₅₀-Wert der Top-20 aktivsten Verbindungen eines Optimierungsexperiments (Min. IC₅₀ [μM], Gl. 5-1). Minimale IC₅₀-Werte sind als Mediane von 100 Wiederholungen zusammen mit ihren unteren und oberen Quartilen angegeben. Aktive repräsentiert die mediane Zahl der gefundenen aktiven Moleküle.

Tabelle 5-3: Ergebnisse einer vergleichenden Optimierungsstudie mit einem Budget von 300 Verbindungen auf der durch die MOE-2D-Deskriptoren repräsentierten Ugi-Bibliothek.^a

	Zufallssuche	Simulated Annealing	Evolutionsstrategie	Partikelschwarm
Chymotrypsin				
Min. IC ₅₀ [μM]	109,5	108,5	108,2	108,0
Quartile	105,0 – 110,0	104,6 – 110,0	104,4 – 110,0	104,0 – 110,0
Aktive	0,0	1,0	1,0	1,0
Quartile	0,0 – 1,0	0,0 – 1,0	0,0 – 2,0	0,0 – 3,0
Pop.-Gr. x Gen.		30x10	3x100	20x15
Chymotrypsin mit vier Initialaktiven				
Min. IC ₅₀ [μM]	109,5	93,9	93,7	90,3
Quartile	106,3 – 110,0	89,1 – 99,1	88,7 – 99,1	83,6 – 95,4
Aktive	0,0	3,0	3,0	4,0
Quartile	0,0 – 1,0	2,0 – 4,0	2,0 – 4,0	3,0 – 5,0
Pop.-Gr. x Gen.		3x100	100x3	3x100
Faktor X_a				
Min. IC ₅₀ [μM]	36,1	16,7	13,5	7,9
Quartile	32,2 – 40,0	11,6 – 21,8	6,4 – 21,4	6,2 – 13,9
Aktive	32,0	103,5	112,0	112,5
Quartile	29,0 – 36,0	90,8 – 113,0	99,0 – 120,0	99,0 – 127,3
Pop.-Gr. x Gen.		3x100	37x8	20x15
Faktor X_a mit vier Initialaktiven				
Min. IC ₅₀ [μM]	36,3	6,3	5,8	6,7
Quartile	33,1 – 40,2	5,8 – 8,0	5,5 – 6,5	6,2 – 12,1
Aktive	32,0	113,0	118,0	98,5
Quartile	28,0 – 35,0	95,5 – 122,5	108,0 – 124,0	69,0 – 108,0
Pop.-Gr. x Gen.		2x150	30x10	2x150
Trypsin				
Min. IC ₅₀ [μM]	90,4	53,3	45,7	33,8
Quartile	86,7 – 93,4	39,8 – 74,7	27,8 – 69,1	27,9 – 58,7
Aktive	6,0	24,0	30,5	40,0
Quartile	4,0 – 7,0	16,0 – 33,3	19,0 – 55,3	23,8 – 56,0
Pop.-Gr. x Gen.		2x150	20x15	20x15
Trypsin mit vier Initialaktiven				
Min. IC ₅₀ [μM]	90,6	30,7	29,4	30,9
Quartile	87,5 – 93,8	28,8 – 37,1	27,9 – 32,2	27,4 – 40,0
Aktive	5,5	51,0	58,0	49,5
Quartile	4,0 – 7,0	37,8 – 57,0	53,5 – 62,0	32,8 – 65,0
Pop.-Gr. x Gen.		2x150	30x10	2x150
Tryptase				
Min. IC ₅₀ [μM]	18,7	5,4	2,4	2,2
Quartile	16,9 – 20,2	3,2 – 9,0	1,7 – 8,0	1,7 – 3,2
Aktive	90,0	202,0	206,5	195,0
Quartile	84,0 – 95,0	182,8 – 211,3	196,5 – 216,0	180,0 – 206,0
Pop.-Gr. x Gen.		3x100	20x15	20x15
uPA				
Min. IC ₅₀ [μM]	59,5	32,8	31,2	27,7
Quartile	54,9 – 64,5	27,9 – 38,6	24,6 – 36,3	20,4 – 31,3
Aktive	26,0	85,0	92,5	107,0
Quartile	23,0 – 29,0	77,0 – 94,0	84,8 – 100,0	93,0 – 115,0
Pop.-Gr. x Gen.		2x150	30x10	3x100

^a Pop.-Gr. x Gen. bezeichnet die beste Kombination von Populationsgröße und Anzahl der Generationen. Ausschlaggebend war der gemittelte IC₅₀-Wert der Top-20 aktivsten Verbindungen eines Optimierungsexperiments (Min. IC₅₀ [μM], Gl. 5-1). Minimale IC₅₀-Werte sind als Mediane von 100 Wiederholungen zusammen mit ihren unteren und oberen Quartilen angegeben. Aktive repräsentiert die mediane Zahl der gefundenen aktiven Moleküle.

Tabelle 5-4: Ergebnisse einer vergleichenden Optimierungsstudie mit einem Budget von 300 Verbindungen auf der durch den MOE-2D-PCA-Deskriptor repräsentierten Ugi-Bibliothek.^a

	Zufallssuche	Simulated Annealing	Evolutionsstrategie	Partikelschwarm
Chymotrypsin				
Min. IC ₅₀ [μM]	109,5	108,3	108,2	108,5
Quartile	105,0 – 110,0	104,7 – 110,0	104,5 – 110,0	103,4 – 110,0
Aktive	0,0	1,0	1,0	1,0
Quartile	0,0 – 1,0	0,0 – 1,0	0,0 – 2,0	0,0 – 2,0
Pop.-Gr. x Gen.		15x20	3x100	75x4
Chymotrypsin mit vier Initialaktiven				
Min. IC ₅₀ [μM]	109,5	89,2	94,4	93,5
Quartile	106,4 – 110,0	84,1 – 94,7	89,3 – 99,7	88,6 – 105,8
Aktive	0,0	4,0	3,0	3,0
Quartile	0,0 – 1,0	3,0 – 5,0	2,0 – 4,0	1,0 – 4,0
Pop.-Gr. x Gen.		3x100	100x3	1x300
Faktor X_a				
Min. IC ₅₀ [μM]	35,8	13,6	11,8	12,3
Quartile	32,3 – 39,8	8,9 – 19,3	6,2 – 19,0	8,7 – 19,7
Aktive	32,0	106,0	123,0	95,0
Quartile	29,0 – 36,0	96,0 – 116,0	108,8 – 134,0	81,3 – 110,0
Pop.-Gr. x Gen.		2x150	20x15	20x15
Faktor X_a mit vier Initialaktiven				
Min. IC ₅₀ [μM]	36,5	6,6	6,2	8,9
Quartile	33,2 – 40,3	5,9 – 8,0	5,7 – 6,6	6,9 – 10,5
Aktive	32,0	105,5	113,5	78,0
Quartile	28,0 – 35,3	96,8 – 116,0	105,8 – 119,3	69,0 – 92,3
Pop.-Gr. x Gen.		3x100	30x10	5x60
Trypsin				
Min IC ₅₀ [μM]	90,3	48,0	37,8	42,7
Quartile	86,8 – 93,3	29,5 – 71,1	27,6 – 59,9	29,8 – 62,2
Aktive	6,0	29,0	38,0	31,0
Quartile	4,0 – 7,0	17,8 – 49,0	21,8 – 51,0	20,0 – 47,0
Pop.-Gr. x Gen.		1x300	60x5	10x30
Trypsin mit vier Initialaktiven				
Min. IC ₅₀ [μM]	90,6	28,3	27,0	29,6
Quartile	87,0 – 93,6	26,0 – 29,9	25,2 – 28,6	28,2 – 48,2
Aktive	6,0	54,0	62,0	51,5
Quartile	4,0 – 7,0	49,8 – 59,0	58,0 – 65,0	25,8 – 55,0
Pop.-Gr. x Gen.		3x100	30x10	1x300
Tryptase				
Min. IC ₅₀ [μM]	18,7	5,2	2,3	2,5
Quartile	17,2 – 20,2	3,0 – 9,7	1,7 – 6,9	1,9 – 7,3
Aktive	89,0	190,0	197,5	177,0
Quartile	84,0 – 95,0	182,8 – 198,0	189,8 – 205,3	162,5 – 188,3
Pop.-Gr. x Gen.		2x150	30x10	5x60
uPA				
Min. IC ₅₀ [μM]	59,4	28,4	26,4	19,5
Quartile	54,6 – 64,6	25,2 – 33,1	23,9 – 30,2	16,2 – 28,5
Aktive	26,0	97,0	103,0	110,0
Quartile	23,0 – 30,0	85,0 – 106,0	87,0 – 115,3	97,8 – 120,3
Pop.-Gr. x Gen.		2x150	10x30	2x150

^a Pop.-Gr. x Gen. bezeichnet die beste Kombination von Populationsgröße und Anzahl der Generationen. Ausschlaggebend war der gemittelte IC₅₀-Wert der Top-20 aktivsten Verbindungen eines Optimierungsexperiments (Min. IC₅₀ [μM], Gl. 5-1). Minimale IC₅₀-Werte sind als Mediane von 100 Wiederholungen zusammen mit ihren unteren und oberen Quartilen angegeben. Aktive repräsentiert die mediane Zahl der gefundenen aktiven Moleküle.

5 Retrospektive Vergleichsstudien von Optimierungsalgorithmen

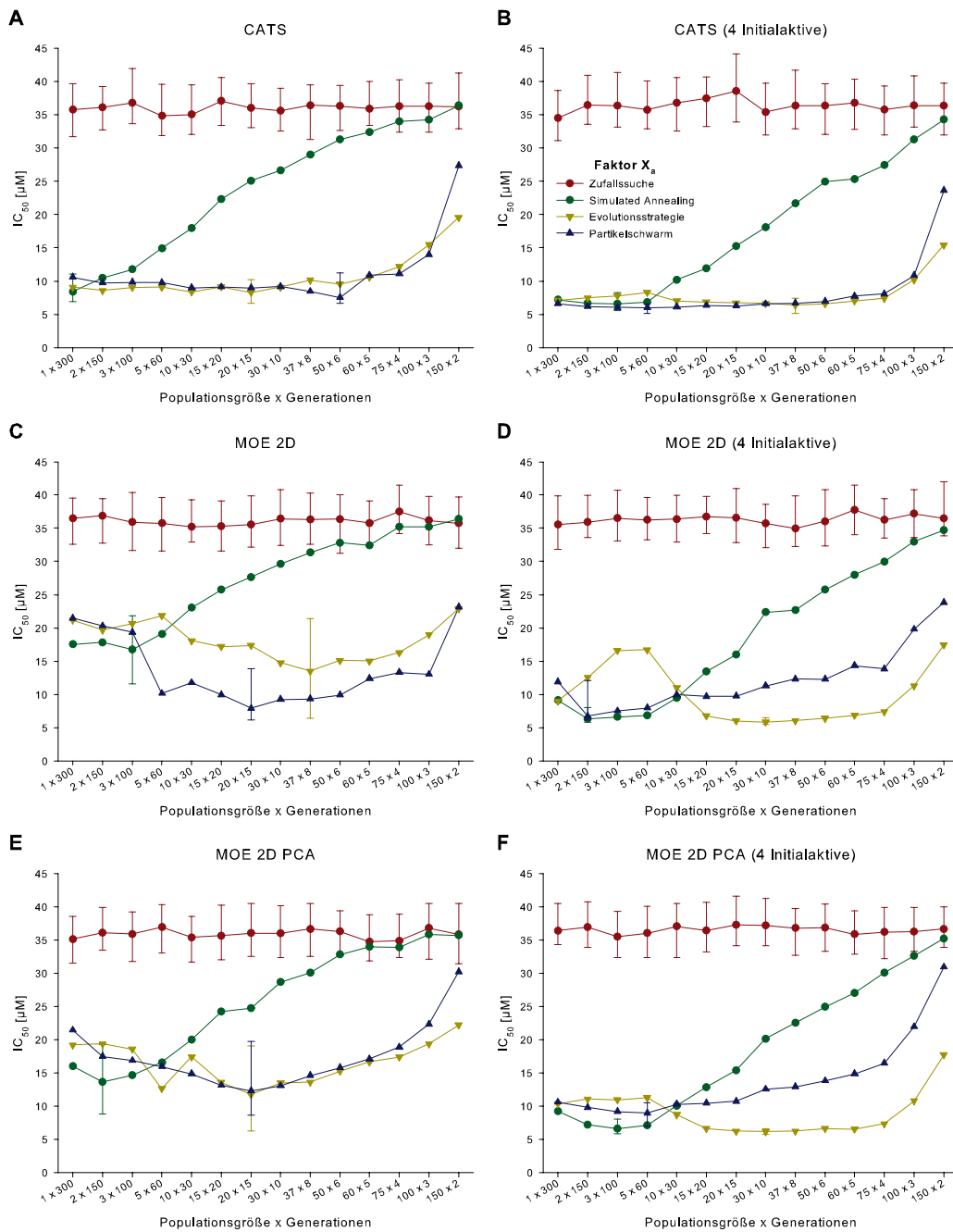


Abbildung 5-3: Vergleich verschiedener Kombinationen von Populationsgröße und Anzahl der Generationen für Faktor- X_a -Inhibitoren der Ugi-Bibliothek. Aufgetragen ist der gemittelte IC_{50} der Top-20 aktivsten Verbindungen als Median von 100 Wiederholungen (Gl. 5-1). Fehlerbalken zeigen das untere und obere Quartil an und sind für die Zufallssuche und die optimalen Kombinationen der anderen Algorithmen gegeben.

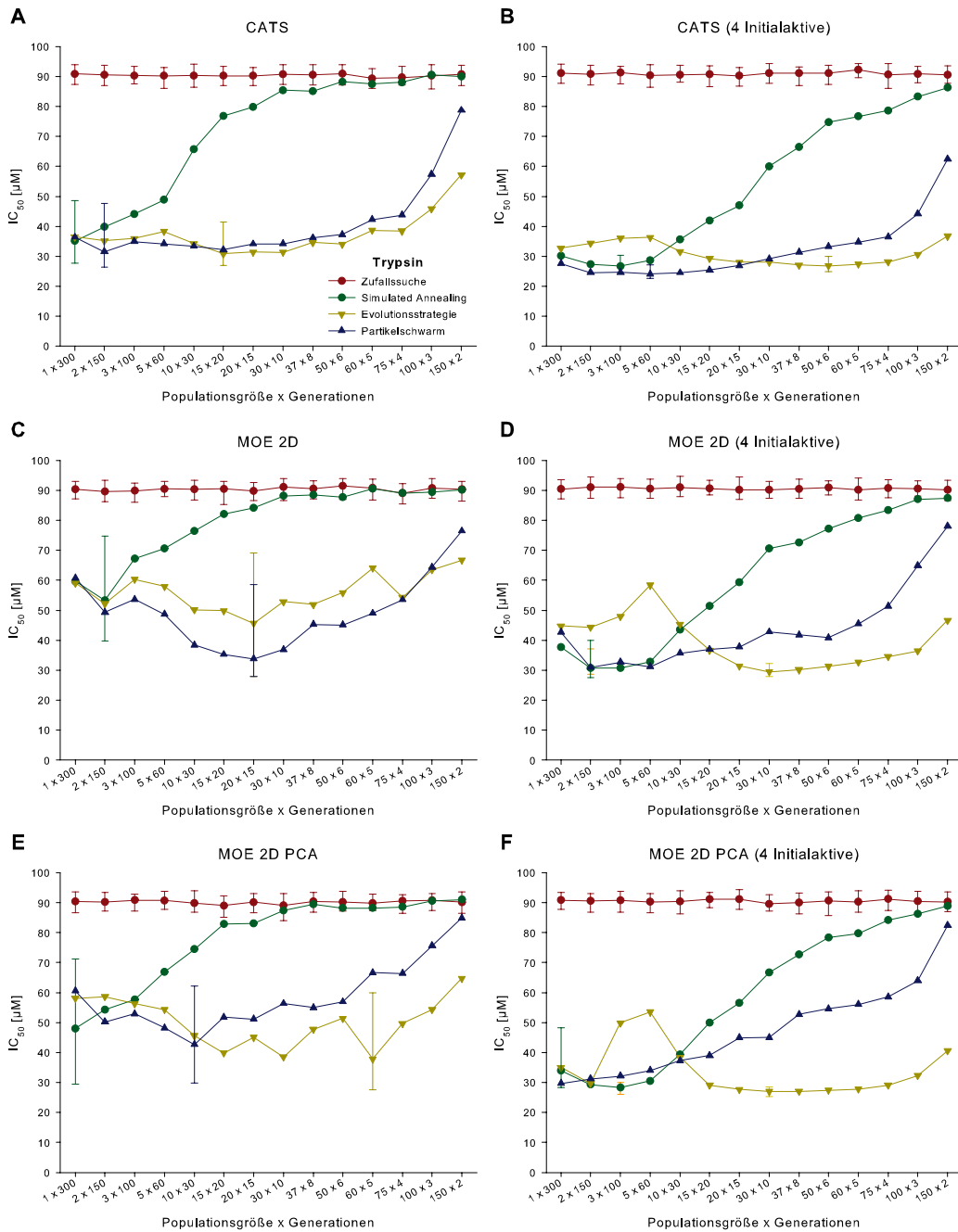


Abbildung 5-4: Vergleich verschiedener Kombinationen von Populationsgröße und Anzahl der Generationen für Trypsin-Inhibitoren der Ugi-Bibliothek. Aufgetragen ist der gemittelte IC_{50} der Top-20 aktivsten Verbindungen als Median von 100 Wiederholungen (Gl. 5-1). Fehlerbalken zeigen das untere und obere Quartil an und sind für die Zufallssuche und die optimalen Kombinationen der anderen Algorithmen gegeben.

5.4 Detailliertes Verständnis des adaptiven Designs von Substanzbibliotheken

Für eine Analyse der verschiedenen Deskriptorräume wurden selbstorganisierende Karten (SOM) erstellt. 20×20 -SOMs mit toroidaler Topologie wurden über 1.000.000 Iterationen auf allen 15.840 Ugi-Verbindungen mit Hilfe der Manhattan-Metrik mit der Software MOLMAP®¹⁴¹ trainiert (Radius = 10,0; $\tau = 5,0$).

In Abbildung 5-5 sind gemittelte IC_{50} -Werte für Trypsin- und Faktor- X_a -Inhibitoren auf die generierten SOMs projiziert. Die Mittelwerte wurden durch arithmetische Mittelung der IC_{50} -Werte der Verbindungen je eines rezeptiven Feldes der Karten berechnet. Helle Bereiche deuten eine Vorherrschaft inaktiver Verbindungen an, dunkle Bereiche deuten eine Vorherrschaft aktiver Verbindungen an und Pluszeichen „+“ stehen für leere Neuronen. Diese Form der Präsentation wurde ausgewählt, weil sie der Repräsentation des Suchraumes in den Optimierungen entspricht. Schattierte Felder der Karten deuten nicht nur die Anwesenheit aktiver Substanzen an, sondern geben durch die Mittelung der IC_{50} -Werte auch ihre relative Häufigkeit und Potenz im Vergleich zu den inaktiven Substanzen wieder. Inaktive Substanzen besaßen einen einheitlichen IC_{50} von 100 μM (Trypsin), bzw. 110 μM (Faktor X_a , vgl. Abschnitt 3.3.1). Eine klare Trennung der aktiven von den inaktiven Verbindungen war in den SOM-Projektionen nicht erkennbar. Inaktive Verbindungen deckten den Projektionsraum der SOMs fast vollständig ab. Aktive Verbindungen bildeten jedoch Cluster. Trypsin- und Faktor- X_a -Inhibitoren besetzten ähnliche Bereiche der Karten. Dieses Nachbarschaftsverhalten wird auch durch ihre Aktivitätswerte wieder gespiegelt: 92% aller Trypsin-Inhibitoren sind auch auf Faktor X_a aktiv. Trypsin und Faktor X_a sind dafür bekannt, sehr ähnliche Bindetasche zu besitzen.¹⁴⁸ Sowohl Trypsin- als auch Faktor- X_a -Inhibitoren bilden im vom CATS-Deskriptor aufgespannten Suchraum deutlich abgrenzbare Cluster. Diese Beobachtung liefert einen Hinweis darauf, warum Optimierungsexperimente auf dem CATS-Deskriptor meist bessere Ergebnisse lieferten, als auf den MOE-2D- und MOE-2D-PCA-Deskriptoren. Aktive Verbindungen bildeten auf den MOE-2D- und MOE-2D-PCA-Karten zwar auch Cluster, diese waren aber weniger stark ausgeprägt. Die Verteilungen der Aktiven ähneln sich auf diesen beiden Karten und reflektieren so den gemeinsamen Deskriptorursprung der beiden Datensätze. Konsequenterweise war die Leistung der beiden Deskriptorsätze in den Optimierungsexperimenten ebenfalls vergleichbar.

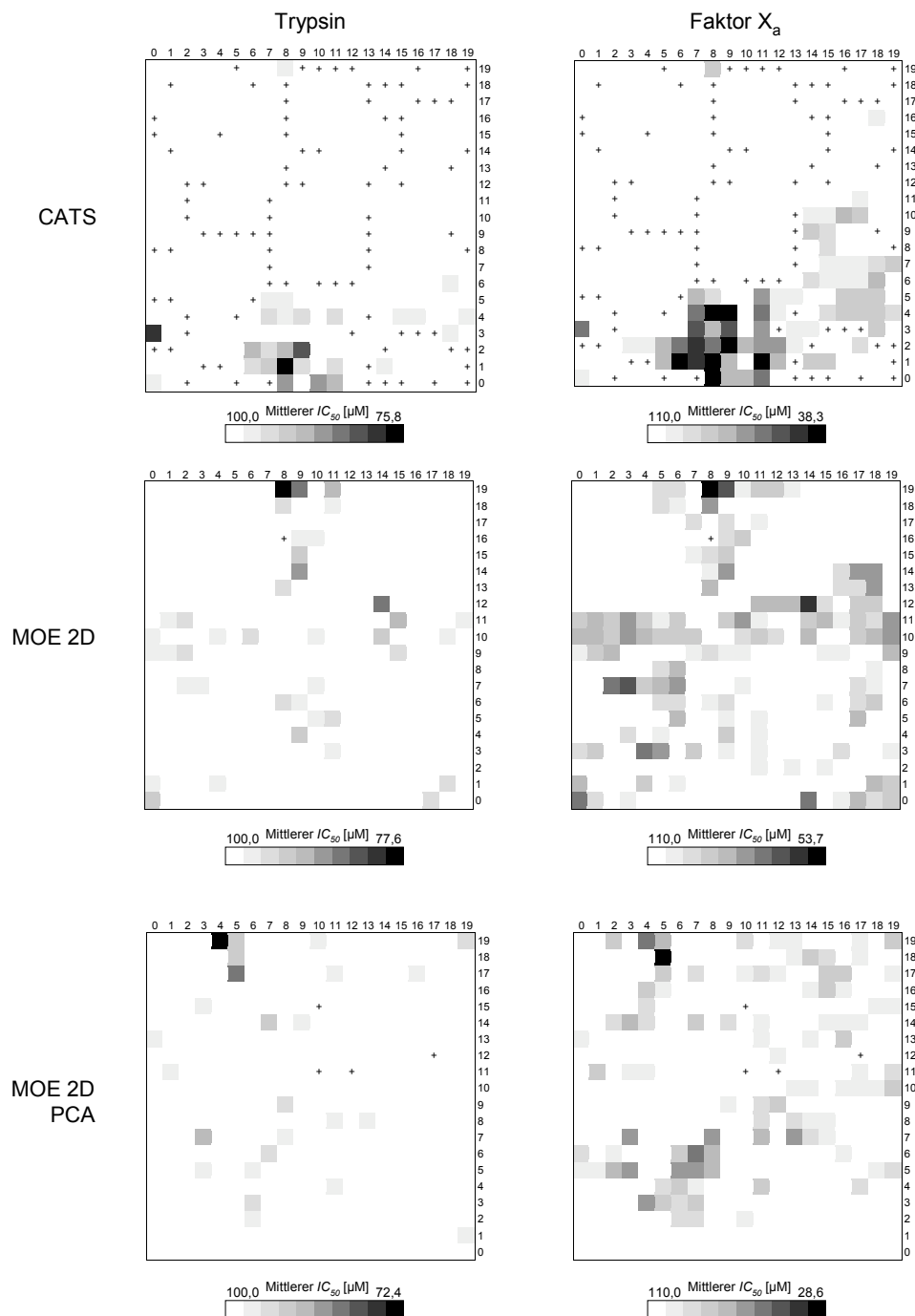


Abbildung 5-5: Toroidale selbstorganisierende Karten (SOM) der hochdimensionalen Deskriptorräume der Ugi-Bibliothek. Auf die Karten sind die gemittelten IC_{50} -Werte der Trypsin-Inhibitoren (linke Seite) und der Faktor- X_a -Inhibitoren (rechte Seite) projiziert. Die Mittelwerte wurden durch Mittelung der IC_{50} -Werte der Verbindungen je eines rezeptiven Feldes berechnet. Helle Bereiche deuten eine Vorherrschaft inaktiver Verbindungen an, dunkle Bereiche deuten eine Vorherrschaft aktiver Verbindungen an. Pluszeichen „+“ stehen für leere Neuronen. 20×20 -SOMs mit toroidaler Topologie wurden über 1.000.000 Iterationen auf allen 15.840 Ugi-Verbindungen mit Hilfe der Manhattan-Metrik mit MOLMAP®¹⁴¹ trainiert (Radius = 10,0; $\tau = 5,0$).

Der verschiedene Grad der Clusterbildung könnte auch die unterschiedlichen Effekte der Einbeziehung aktiver Verbindungen in die Initialpopulation erklären. Im CATS-Deskriptorraum bewirkte die Einbeziehung Initialaktiver nur eine geringe Verbesserung der Optimierungsergebnisse. Wegen des klar abgrenzbaren Clusters aktiver Verbindungen konnten die Optimierungsalgorithmen hier viel Zeit auf die *Ausnutzung* der lokalen Nachbarschaft aufwenden und mussten weniger Zeit für die *Erforschung* des Suchraumes aufbringen. In den MOE-2D- und MOE-2D-PCA-Deskriptorräumen hatte die Einbeziehung bekannter Inhibitoren eine größere Auswirkung auf Optimierungen. Hier mussten die Optimierungsalgorithmen wegen der diffuseren Clusterbildung erst ein größeres Hypervolumen des Suchraumes erforschen, um schließlich potente Inhibitoren zu finden. Die Einbeziehung Initialaktiver kommt hier einer Abkürzung der Erforschungsphase gleich, da die Suchen direkt in Gebieten hoher Aktivität („Aktivitätsinseln“) beginnen.¹⁴⁹ Durch diese „Abkürzung“ war somit mehr Zeit für die Ausnutzung der lokalen Aktivitätsinseln verfügbar und die Optimierungsergebnisse verbesserten sich.

Um das Optimierungsverhalten ES- und PSO-basierter Optimierung von Substanzbibliotheken besser verstehen zu lernen, wurden zusätzliche Daten von der besten Evolutionsstrategie (50×6) und der besten Partikelschwarmoptimierung (5×60) für eine Optimierung von Trypsin-Inhibitoren im CATS-Deskriptorraum mit Einbeziehung vier Aktiver in der Initialpopulation erhoben. Das Simulated Annealing wurde in dieser Analyse ausgegrenzt, weil dessen Optimierungsverhalten weniger für eine prospektive Niedrigdurchsatzscreening-Kampagne geeignet ist.

Für die diskreten Suchräume von durch Deskriptoren beschriebenen Substanzbibliotheken gilt eine Besonderheit: ES und PSO operieren in kontinuierlichen Suchräumen. Eine durch molekulare Deskriptoren repräsentierte Molekülbibliothek stellt jedoch einen dünn besiedelten, diskreten Suchraum dar. Der chemische Raum ist diskret, da nur solche Positionen besetzt sind, an denen „echte“ Moleküle realisiert sind. Zur Erzeugung eines Nachkommens wurde der Positionsvektor eines Elters mit Hilfe eines Mutationsoperators variiert wodurch ein „virtueller Nachkomme“ an einer neuen Position im kontinuierlichen Raum entstand. Da der virtuelle Nachkomme im kontinuierlichen Raum erzeugt wurde, ist es so gut wie sicher, an dieser Position kein „echtes“ Molekül anzutreffen. Es wurde deshalb ein Diskretisierungsschritt in Form einer Nächste-Nachbar-Suche¹³⁷ eingeführt um den virtuellen Nachkommen auf die Position eines benachbarten Moleküls im diskreten, chemischen Raum abzubilden (Abbildung 5-6).

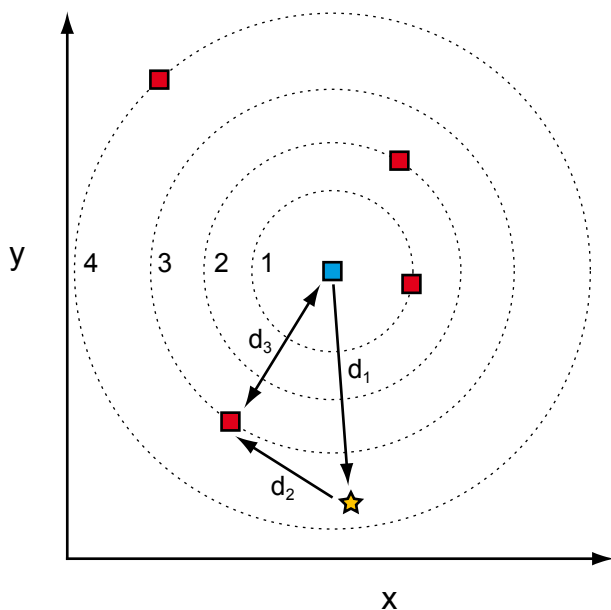


Abbildung 5-6: Der Diskretisierungsschritt für chemische Suchräume. Das Elternindividuum (blaues Quadrat) erzeugt durch Variation (z.B. Mutation) seines Genotyps (Positionsvektor) einen *virtuellen* Nachkommen im kontinuierlichen Suchraum (gelber Stern). Eine Nächste-Nachbar-Suche findet das nächst gelegene Molekül im diskreten, chemischen Raum (rotes Quadrat unten links). Dieser nächste Nachbar ist der resultierende *echte* Nachkomme des Variationsschrittes. d_1 bezeichnet den Abstand „Elter \leftrightarrow virtueller Nachkomme“, d_2 den Abstand „virtueller Nachkomme \leftrightarrow echter Nachkomme“ und d_3 den Abstand „Elter \leftrightarrow echter Nachkomme“. Der Variationsprozess verläuft über d_1 und d_2 in Pfeilrichtung. Der echte Nachkomme ist hier der 3. Nachbar seines Elters (Ordinalzahl: 3).

Bei Optimierungen mit der ES war ein Nachkomme im Schnitt der 33igste Nachbar seines Elters (Tabelle 5-5). Der Median der Distanz zwischen einem Elter und seinem virtuellen Nachkommen war 94 (gemessen in willkürlichen Einheiten mit der Manhattan-Metrik). Die tatsächliche Distanz eines Elters zu seinem echten Nachkommen betrug jedoch nur 28 Einheiten. Diese überraschend große Abweichung könnte durch das Phänomen der „stillen Mutationen“ erklärt werden.¹⁵⁰ Einige Punktmutationen des Positionsvektors („Genotyp“) resultieren nicht gleichzeitig in einer Variation des Phänotyps (Molekül), da weite Bereiche des chemischen Raumes nicht mit Molekülen besetzt sind. Da der zusätzliche Diskretisierungsschritt offensichtlich einen weitreichenden Eingriff in das Suchverhalten darstellte, wurden seine Auswirkungen durch Variation des Schrittweitenparameters der ES genauer analysiert. Schrittweiten von 0 bis 3 waren für eine effiziente Optimierung akzeptabel. Diese Schrittweiten entsprechen „Elter \leftrightarrow virtueller Nachkomme“-Abständen von 0 bis 126 Einheiten. Diese wiederum entsprechen „Elter \leftrightarrow echter Nachkomme“-Abständen von 19 bis 38

Einheiten. Dies deutet an, dass Evolutionsstrategien effektiv in lokal beschränkten Nachbarschaften operieren. Des Weiteren hatte der Diskretisierungsschritt einen beachtlichen Dämpfungseffekt auf zu klein oder zu groß gewählte Schrittweiten. Es sei angemerkt, dass eine Schrittweite von 0 nur durch eine globale *Tabu*-Regel ermöglicht wurde, die besagte, dass während eines Optimierungslaufes jede Substanz nur maximal einmal ausgewählt werden durfte. Es wurde daher nicht immer der räumlich nächste Nachbar als echter Nachkomme ausgewählt, sondern der nächste „erlaubte“ Nachbar, der nicht durch ein *Tabu* gesperrt war. Die Nächste-Nachbar-Suche verhinderte zusammen mit der globalen *Tabu*-Regel effizient eine frühzeitige Konvergenz.

Tabelle 5-5: Daten zum Optimierungsverhalten von ES und PSO. Angegeben sind Eltern-Nachkommen-Distanzen (Manhattan-Metrik) und die Ordinalzahl (der wievielte Nachbar war der echte Nachkomme ausgehend vom Elter?).^a

	Evolutionsstrategie	Partikelschwarmoptimierung
„Elter↔echter Nachkomme“-Distanz	28	47
„Elter↔virtueller Nachkomme“-Distanz	94	92
Ordinalzahl	33	356

^a Eingesetzt wurde eine (9,50)-ES (6 Gen.) und eine PSO (6 Partikel, 50 Gen.) zur Optimierung von Trypsin-Inhibitoren im CATS-Deskriptorraum mit Einbeziehung vier Aktiver in der Initialpopulation.

Die statistischen Daten, die von PSO-Läufen gewonnen wurden, weisen auf ein weniger lokales Optimierungsverhalten für diese Verfahren hin. Im analysierten Fall betrug die „Elter↔echter Nachkomme“-Distanz 47 Einheiten, im Vergleich zu 92 Einheiten bei der „Elter↔virtueller Nachkomme“-Distanz (Tabelle 5-5). Ein bewegter Partikel war im Schnitt der 356igste Nachbar seiner ursprünglichen Position. Diese Daten können jedoch die Eigendynamik der PSO nicht richtig wiedergeben. Es wurden deshalb individuelle PSO-Läufe separat analysiert:

Zu Beginn eines PSO-Laufes wurden sehr weite Sprünge im Suchraum vollzogen. Der Schwarm konvergierte jedoch innerhalb weniger Iterationen auf einen der drei Referenzpunkte der PSO: der aktuellen Partikelposition, der lokal besten Position eines Partikels und der global besten Position des gesamten Schwarms. In dieser Anfangsphase reduzierten sich die „Elter↔echter Nachkomme“-Abstände von 143 Einheiten (dies entspricht der Ausdehnung des gesamten Suchraumes) auf 8.6 Einheiten (einem direkten Nachbarn). Die anfängliche Explorationsphase wurde zur Kalibrierung der

Geschwindigkeitsvektoren benötigt, da diese zufällig initialisiert worden sind. Das weitere Verhalten kann am besten als „pulsierend“ beschrieben werden (Abbildung 5-7). Der Schwarm expandiert und kontrahiert periodisch und erkundet dabei Positionen, die aus der aktuellen Position eines Partikels, seiner lokal besten Position und der global besten Position des Schwarms extrapoliert werden. Die Schrittweiten der „fliegenden“ Partikel variierten dabei von Sprüngen über die halbe Ausdehnung des Suchraumes bis hin zu nur marginaler Bewegung. Die Gesamtrajektorie orientierte sich am globalen Gedächtnis des Schwarms. Sobald eine neue global beste Position entdeckt wurde, migrierte der Schwarm schrittweise darauf hin. Der Effekt des zusätzlichen Diskretisierungsschrittes wird auch bei PSO deutlich: „Elter \leftrightarrow virtueller Nachkomme“-Abstände waren im CATS-Deskriptorraum rund doppelt so groß wie „Elter \leftrightarrow echter Nachkomme“-Abstände. PSOs profitierten stark vom *Tabu*, dass Substanzen während eines Optimierungslaufes nur genau einmal ausgewählt werden durften. Besonders in Optimierungsexperimenten mit einer großen Anzahl an Iterationen verhinderte die *Tabu*-Regel effektiv eine frühzeitige Konvergenz.

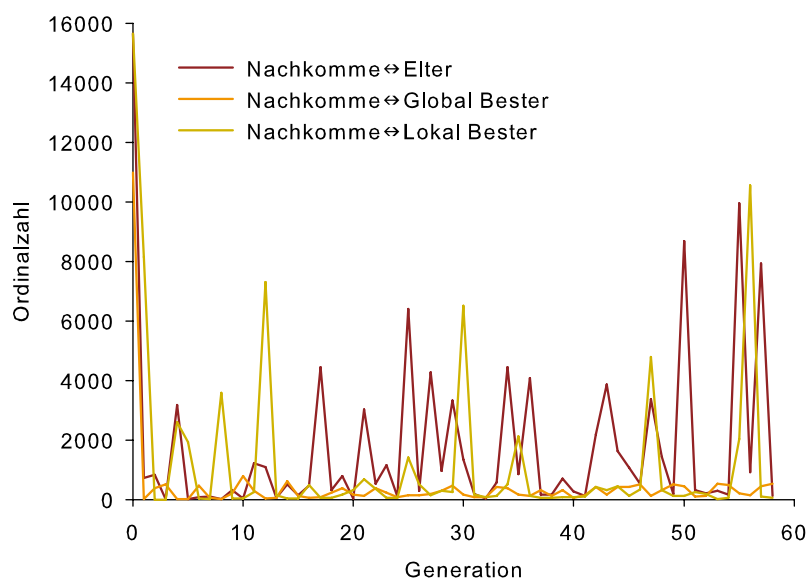
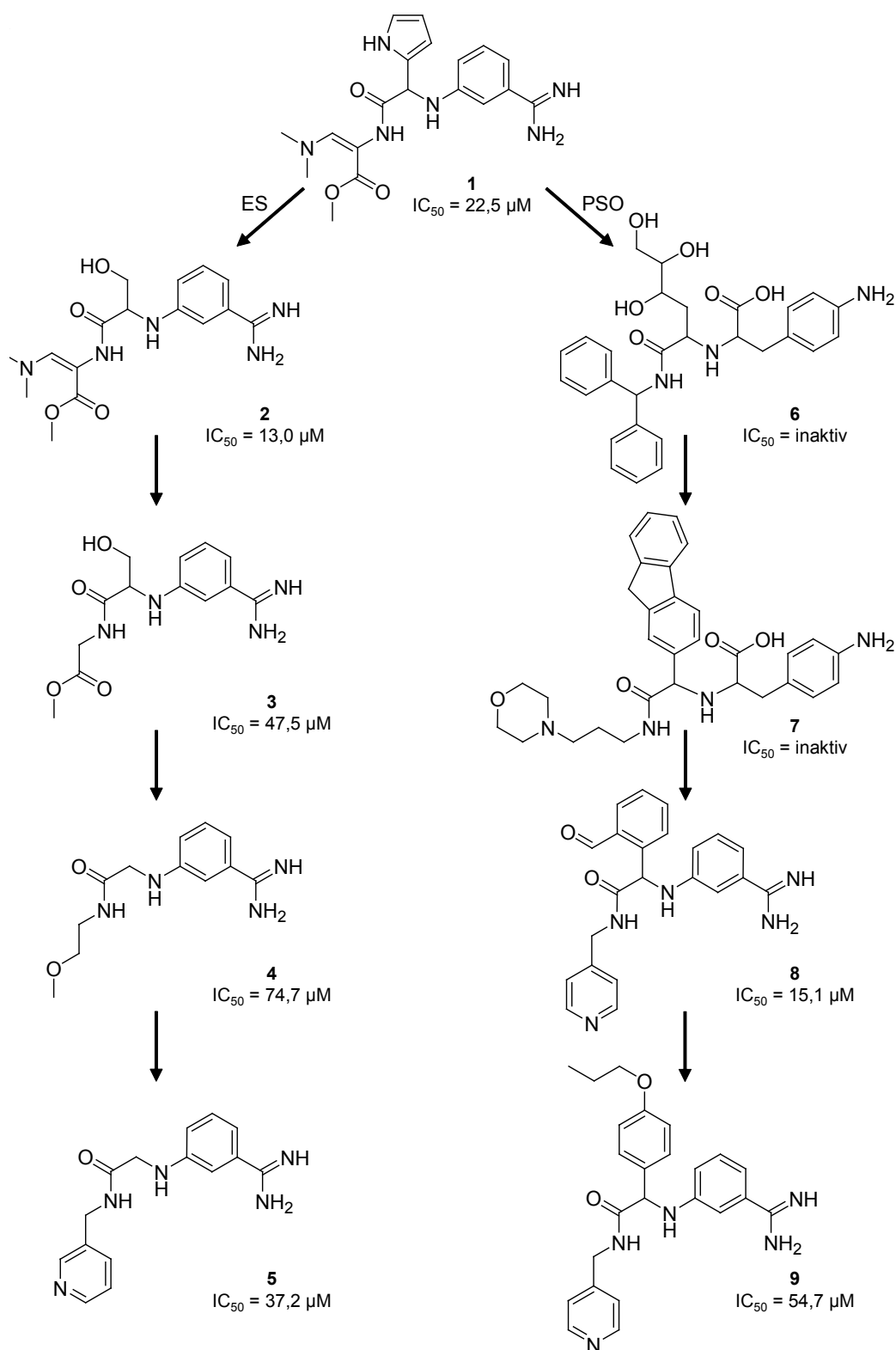


Abbildung 5-7: Das pulsierende Suchverhalten der Partikelschwarmoptimierung. Dargestellt ist die Ordinalzahl (der wievielte Nachbar?) der Nachkommen zum Elternpartikel (rot), zum global besten Partikel (orange) und zum lokal besten Partikel (gelb) eines einzelnen PSO-Laufes (5 Partikel, 60 Generationen).

Die Analysen des Suchverhaltens ergaben zusammenfassend folgendes Bild: Evolutionsstrategien bewegen sich in kleinen Schritten von Eltern zu Nachkommen



Schema 5-1: Vergleich der ersten fünf Generationen eines typischen ES- und PSO-Laufes. IC_{50} -Werte geben Trypsin-Inhibierung an. Beide Algorithmen sind mit Referenzstruktur 1 gestartet. Für ES sind erfolgreiche Eltern-Individuen abgebildet und für PSO die Stationen der Bewegung eines Partikels. Eingesetzt wurde eine (9,50)-ES (6 Generationen) und eine PSO mit 5 Partikeln (60 Generationen).

und Eltern gekoppelt mit der simulierten Darwinschen „natürlichen Selektion“. Der Erfolg der Partikelschwarmoptimierung basiert auf der Bewegung von Partikeln, die Positionen einnehmen, welche aus der aktuellen Position und den Positionen der beiden Gedächtnisse extrapoliert werden. Folglich weist das Molekül an der neuen Partikelposition weniger Ähnlichkeit mit dem Molekül an der alten Position auf. Dieses Verhalten ist in Schema 5-1 exemplarisch dargestellt. Die Strukturen auf der linken Seite (2-5) stellen erfolgreiche Eltern einer Evolutionsstrategie dar, die von Referenzverbindung 1 in der Initialpopulation ausging. Auf der rechten Seite sind die Stationen (Verbindungen 6-9) eines Partikels der Partikelschwarmoptimierung abgebildet. Die ES arbeitete mit wiederholten, kleinen strukturellen Variationen der Referenzverbindung. Größere strukturelle Abwandlungen waren in der Bewegung eines Partikels zu beobachten. In dem hier vorgestellten Beispiel waren die zwei dem Referenzmolekül folgenden Partikel komplett inaktiv. Jedoch kehrte der Partikel geleitet durch sein lokales und globales Gedächtnis mit Struktur 8 wieder auf eine Aktivitätsinsel zurück – vermutlich in die Nähe seiner Ausgangsposition. Die ES trifft während der Optimierung ebenfalls auf inaktive Substanzen. Diese „sterben“ allerdings ab und werden nicht als Eltern einer neuen Generation in Betracht gezogen. Eine Besonderheit des Trypsin-Inhibitor-Designs soll hier nicht unerwähnt bleiben: Die Benzamidingruppe der Verbindungen 1-5, 8 und 9 ist ein privilegiertes Bindefragment der S_1 -Substraterkennungstasche des Enzyms (vgl. Abschnitt 2.1). 97% aller Ugi-artigen Trypsininhibitoren des Datensatzes besaßen dieses Fragment als molekularen Baustein. Aus diesem Grund wiesen auch alle erfolgreichen Designs des Beispiels in Schema 5-1 diese Substruktur auf. Eine ähnliche Beobachtung wurde bereits von Weber *et al.* beim evolutionären Design Ugi-artiger Thrombin-Inhibitoren gemacht.^{5,25}

5.5 Optimierung mit binären Fitnessfunktionen

Im Anschluss wurden vergleichende Optimierungsstudien für das Design fokussierter Substanzbibliotheken auf der COBRA-Sammlung wirkstoff- und leitstrukturartiger Verbindungen (Version 6.1, 8311 Referenzverbindungen) durchgeführt.⁹⁵ Das Ziel war hier, das Verhalten der Optimierung fokussierter Substanzbibliotheken mit einer binären Fitnessfunktion zu analysieren. Des Weiteren wurde überprüft, ob die auf dem Ugi-Datensatz optimierten Parameter auf einem fremden Datensatz Gültigkeit besitzen würden.

Alle Verbindungen der COBRA waren mit Informationen zum Zielrezeptor und mit biologischen Aktivitäten annotiert. Einzig Faktor- X_a -Serinproteaseinhibitoren waren ausreichend in COBRA vorhanden (228 Verbindungen, 2,7%; Chymotrypsin, Trypsin, uPA, Tryptase: < 1%). Im Gegensatz zur Ugi-Bibliothek wurden in diesem Fall nur binäre Aktivitätsbezeichnungen („aktiv“ vs. „inaktiv“) benutzt. Folglich wurde die Erfolgsrate eines Optimierungsalgorithmus als seine Fähigkeit gemessen, möglichst viele aktive Verbindungen zu identifizieren. Da COBRA ungefähr halb so viele Moleküleinträge enthält wie die Ugi-Bibliothek, wurde das Budget der Substanztestungen von 300 auf 150 gesenkt. Bei der Auswahl der Initialpopulation wurden wieder die Verfahren „(wiederholte) zufällige Auswahl“ und „Einbeziehung vier aktiver Substanzen in die Initialpopulation“ verglichen. Die vier Aktiven wurden zufällig aus allen 228 Faktor- X_a -Inhibitoren ausgewählt.

Auf dem ersten Blick ähneln die Versuchsergebnisse denen, die auf der Ugi-Bibliothek gewonnen wurden (Tabelle 5-6, Abbildung 5-8). Die besten Ergebnisse konnten hier mit einer PSO (1 Partikel, 150 Iterationen) auf MOE-2D-Deskriptoren, mit einer Erfolgsquote von 56%, gewonnen werden. Alle drei Algorithmen konnten binäre Aktivitätsdaten effizient verarbeiten und übertrafen die Zufallssuche deutlich. Bei PSO musste die Kondition zur Aktualisierung der Partikelgedächtnisse jedoch zuvor modifiziert werden, um diese Suchstrategie für binäre Aktivitätsräume anzupassen (vgl. Abschnitt 5.6). Diese Modifikation hatte keine Auswirkungen auf Optimierungen der Ugi-Bibliothek.

Tabelle 5-6: Ergebnisse einer vergleichenden Optimierungsstudie mit einem Budget von 150 Verbindungen auf COBRA.^a

	Zufallssuche	Simulated Annealing	Evolutionsstrategie	Partikelschwarm
Faktor X_a CATS				
Max. Aktive	4,0	18,0	47,5	52,5
Quartile	3,0 – 5,0	10,8 – 28,3	28,8 – 63,0	29,8 – 70,3
Max. %Aktive	2,7%	12,0%	32,3%	35,5%
Quartile	2,0% – 3,3%	7,2% – 18,8%	19,6% – 42,9%	20,1% – 47,5%
Pop.-Gr. x Gen.		3x50	7x21	4x37
Faktor X_a CATS mit vier Initialaktiven				
Max. Aktive	4,0	57,5	61,5	71,5
Quartile	3,0 – 5,0	44,0 – 63,0	45,0 – 69,0	49,3 – 80,0
Max. %Aktive	2,7%	38,3%	41,0%	47,7%
Quartile	2,0% – 3,3%	29,3% – 42,0%	30,0% – 46,0%	32,8% – 53,3%
Pop.-Gr. x Gen.		3x50	15x10	5x30
Faktor X_a MOE 2D				
Max. Aktive	4,0	29,5	53,5	72,0
Quartile	3,0 – 5,0	7,8 – 46,3	39,3 – 77,3	53,5 – 89,0
Max. %Aktive	2,7%	19,7%	35,7%	48,0%
Quartile	2,0% – 3,3%	5,2% – 30,8%	26,2% – 51,5%	35,7% – 59,3%
Pop.-Gr. x Gen.		1x150	2x75	1x150
Faktor X_a MOE 2D mit vier Initialaktiven				
Max. Aktive	4,0	60,0	66,5	84,5
Quartile	2,0 – 5,0	52,8 – 66,0	46,0 – 81,0	69,0 – 94,0
Max. %Aktive	2,7%	40,0%	44,3%	56,3%
Quartile	1,3% – 3,3%	35,2% – 44,0%	30,7% – 54,0%	46,0% – 62,7%
Pop.-Gr. x Gen.		2x75	2x75	1x150
Faktor X_a MOE 2D PCA				
Max. Aktive	4,0	24,5	43,0	50,0
Quartile	3,0 – 5,0	8,0 – 37,3	27,8 – 56,3	35,8 – 64,3
Max. %Aktive	2,7%	16,3%	28,7%	33,3%
Quartile	2,0% – 3,3%	5,3% – 24,8%	18,5% – 37,5%	23,8% – 42,8%
Pop.-Gr. x Gen.		1x150	3x50	2x75
Faktor X_a MOE 2D PCA mit vier Initialaktiven				
Max. Aktive	4,0	47,0	54,0	64,5
Quartile	3,0 – 5,0	38,0 – 54,0	50,0 – 58,0	52,0 – 79,0
Max. %Aktive	2,7%	31,3%	36,0%	43,0%
Quartile	2,0% – 3,3%	25,3% – 36,0%	33,3% – 38,7%	34,7% – 52,7%
Pop.-Gr. x Gen.		1x150	15x10	1x150

^a Pop.-Gr. x Gen. bezeichnet die beste Kombination von Populationsgröße und Anzahl der Generationen. Ausschlaggebend war die maximale Anzahl gefundener aktiver Verbindungen eines Optimierungsexperiments (Max. Aktive). Die Anzahl der Aktiven ist als Median von 100 Wiederholungen zusammen mit dem unteren und oberen Quartil angegeben. ‚Max. %Aktive‘ bezeichnet den maximalen Prozentsatz an gefundenen Aktiven im Verhältnis zur Anzahl der Evaluationen. Ein Wert von 100% wird erreicht, wenn genau 150 aktive Verbindungen in einem Optimierungslauf gefunden wurden.

5 Retrospektive Vergleichsstudien von Optimierungsalgorithmen

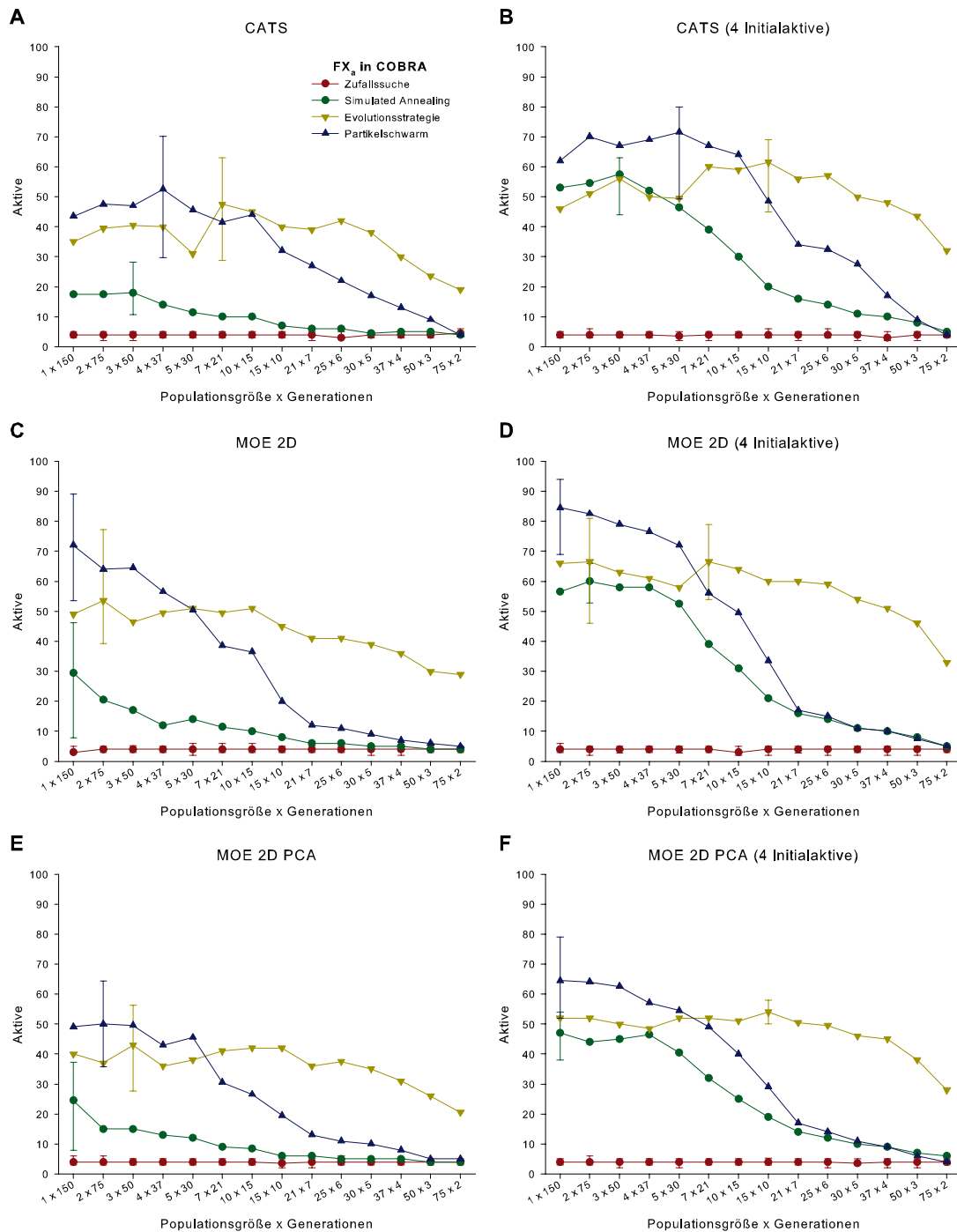


Abbildung 5-8: Vergleich verschiedener Kombinationen von Populationsgröße und Anzahl der Generationen für Faktor- X_a -Inhibitoren der COBRA. Aufgetragen ist die Anzahl gefundener Aktiver als Median von 100 Wiederholungen. Fehlerbalken zeigen das untere und obere Quartil an und sind für die Zufallssuche und die optimalen Kombinationen der anderen Algorithmen gegeben.

Der größte Unterschied zu den vorherigen Optimierungsergebnissen der Ugi-Bibliothek liegt im schlechteren Abschneiden des Simulated Annealings, besonders in Ansätzen ohne die Einbeziehung aktiver Verbindungen in die Initialpopulation. SA reagierte sensible auf Änderungen der Strategieparameter. Mit einer separat optimierten Initialtemperatur und Schrittweite konnten auch mit SA ähnliche Erfolgsquoten wie mit ES und PSO erreicht werden (Abbildung 5-9). Für das Design der vergleichenden Optimierungsstudie wurde zu Demonstrationszwecken jedoch auf eine separate Optimierung der Strategieparameter verzichtet.

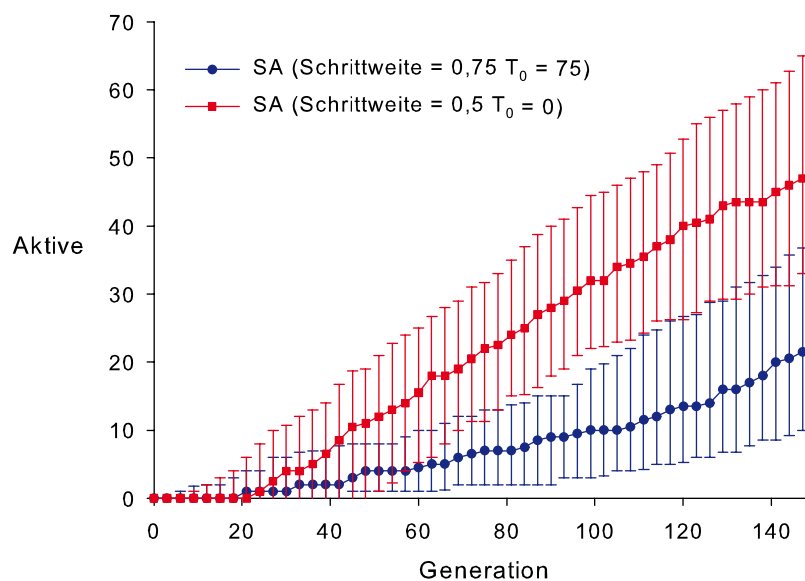


Abbildung 5-9: Vergleich zweier SA-Ansätze mit verschiedenen Strategieparametern. Der Ansatz mit zuvor auf der Ugi-Bibliothek optimierten Parametern (blaue Kurve) funktioniert ineffizient. Ein Ansatz mit speziell optimierten Parametern liefert bessere Ergebnisse. Mit der Initialtemperatur $T_0 = 0$ ist das Metropolisakzeptanzkriterium außer Kraft gesetzt und das Verfahren gleicht einem Bergsteigeralgorithmus. Aufgetragen ist die Anzahl gefundener Aktiver als Median von 100 Wiederholungen. Es wurden Faktor- X_a -Inhibitoren der COBRA optimiert (Populationsgröße: 1, Generationen: 150, Deskriptorraum: MOE 2D PCA).

In einer prospektiven Screeningstudie ist eine separate Optimierung in den allermeisten Fällen nicht machbar. SA erwies sich hier – in der vorliegenden Implementierung – als nicht robust, da Suchparameter auf verschiedenen Datensätzen keine allgemeine Gültigkeit besaßen. Der Effekt der suboptimalen Strategieparameter ist bei Optimierungen mit Initialaktiven weniger stark ausgeprägt. Interessant ist die Beobachtung, dass in Abbildung 5-9 eine Initialtemperatur von Null beste Ergebnisse liefert. Ein Wert von Null deaktiviert *de facto* das Metropolisakzeptanzkriterium (Gl. 3-5). Der Algorithmus entspricht mit diesen Parametern einem Bergsteigeralgorithmus, der das nächste lokale

Optimum sucht und keine Maßnahmen gegen frühzeitige Konvergenz ergreift. Dies lässt auf ein unimodales oder nur oligomodales Fitnessgebirge schließen. Weitere Hinweise auf ein weniger komplexes Fitnessgebirge als im Fall der Ugi-Bibliothek werden die SOM-Analysen weiter unten liefern.

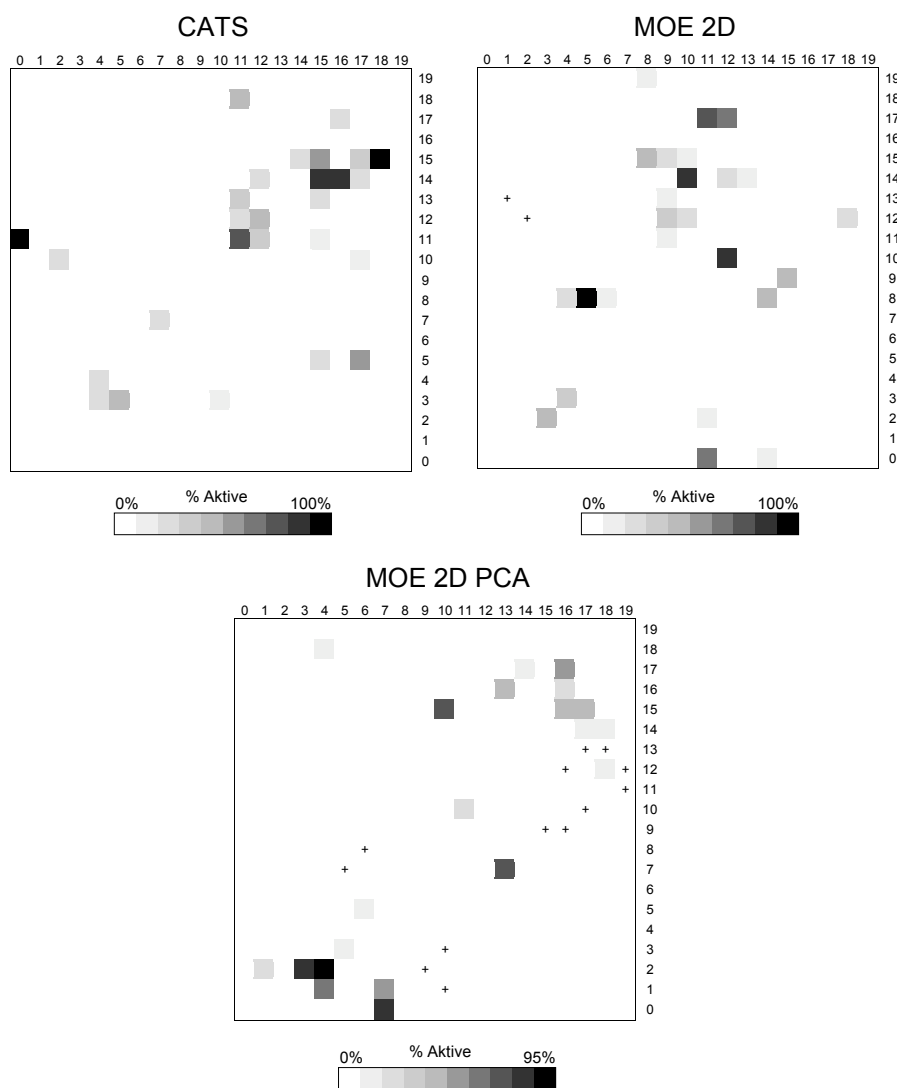


Abbildung 5-10: Toroidale selbstorganisierende Karten (SOM) der hochdimensionalen Deskriptorräume der COBRA. Auf die Karten ist der prozentuale Anteil von Faktor- X_n -Inhibitoren projiziert. Die Prozentwerte wurden durch Bildung des Verhältnisses von aktiven und inaktiven Substanzen je eines rezeptiven Feldes berechnet. Helle Bereiche deuten eine Vorherrschaft inaktiver Verbindungen an, dunkle Bereiche deuten eine Vorherrschaft aktiver Verbindungen an. Pluszeichen „+“ stehen für leere Neuronen.

In Übereinstimmung mit den vorherigen Ergebnissen bevorzugte SA auch auf COBRA eine große Anzahl an Iterationen und ist deshalb weniger für eine Anwendung in einer Niedrigdurchsatzscreening-Kampagne geeignet. ES bevorzugte wieder große Populationen und nur wenige Generationen. Die Evolutionsstrategien schnitten etwas schlechter als PSO ab. Der Grund sind hier ebenfalls leicht suboptimale Strategieparameter (hier: die Schrittweite). Ihr populationsbasierter Charakter macht die ES jedoch robuster gegenüber Änderungen der Schrittweite. Die Partikelschwarmoptimierung – welche wieder mehr Iterationen und kleinere Populationen bevorzugte – schnitt auf COBRA in allen analysierten Fällen am besten ab. Der Konstriktionsschwarm zeigte hier aufgrund seiner autoadaptiven Schrittweitenkontrolle ein exzellentes Optimierungsverhalten.

Die SOM-Analyse der drei Deskriptorräume deckte eine andersartige Verteilung der Faktor- X_a -Inhibitoren auf (Abbildung 5-10). 20×20 -SOMs mit toroidaler Topologie wurden auf allen 8311 COBRA-Substanzen trainiert. Dieses Mal wurde die SOM des CATS-Deskriptorraumes mit der euklidischen anstatt mit der Manhattan-Distanz trainiert und der CATS-Deskriptor wurde nicht autoskaliert, um die Qualität der Projektion zu verbessern. Auf die Karten wurde der prozentuale Anteil der Faktor- X_a -Inhibitoren projiziert. Die Prozentwerte wurden durch Bildung des Verhältnisses von aktiven und inaktiven Substanzen je eines rezeptiven Feldes berechnet. Helle Bereiche deuten eine Vorherrschaft inaktiver Verbindungen an, dunkle Bereiche eine Vorherrschaft aktiver Verbindungen und Pluszeichen „+“ bezeichnen leere Neuronen. Schattierte Bereiche zeigen somit nicht nur die Anwesenheit aktiver Verbindungen an, sondern auch deren relative Häufigkeit im Vergleich zu den Inaktiven. Wie auch bei den SOM, die auf der Ugi-Bibliothek trainiert wurden, bildeten aktive Verbindungen Cluster auf den Karten. Im Gegensatz zu vorher sind jetzt dagegen mehrere sehr dichte Cluster mit bis zu 100% Aktive auszumachen. Die entsprechenden Gebiete des Suchraumes scheinen hoch mit Faktor- X_a -Inhibitoren angereichert zu sein. Sobald solch ein Cluster von einem Suchalgorithmus getroffen wird, schreitet die weitere Optimierung in der hoch angereicherten Nachbarschaft deutlich erleichtert voran. Dies könnte das gute Ergebnis von rund 50% Aktiven unter nur 150 ausgewerteten Bibliothekssubstanzen erklären (75 aktive Substanzen entsprechen 33% aller Faktor- X_a -Inhibitoren der COBRA). Im Gegensatz zu der auf Serinproteasen fokussierten kombinatorischen Ugi-Bibliothek stellt die Sammlung bioaktiver Referenzanaloga (COBRA) eine weit diversere Substanzbibliothek dar, die Verbindungen gegen mehr als 300 biologische Ziele enthält. Schneider *et al.* konnten bereits früher zeigen, dass Liganden verschiedener Zielrezeptoren der COBRA durch SOM-Analysen partiell separierbar sind.^{95,151}

5.6 Modifikation der Optimierungsalgorithmen

Alle drei Optimierungsalgorithmen mussten im Vergleich zur Originalimplementierung modifiziert werden, um an die speziellen Anforderungen chemischer Suchräume angepasst zu werden. Der größte Eingriff in die Algorithmen war die Einführung eines zusätzlichen Diskretisierungsschrittes mittels einer Nächste-Nachbarn-Suche. Eine weitere Modifikation aller drei Optimierungsalgorithmen berücksichtigt die obere Bestimmungsgrenze von biologischen Aktivitätsmessungen (hier: $IC_{50} = 100 \mu\text{M}$). Bei der Evolutionsstrategie wurde die Elternselektion so angepasst, dass nur tatsächlich aktive Eltern zulässig sind (Gl. 3-7). Die resultierende Anzahl der Eltern μ' wird damit fitnessadaptiv und wird je nach Anzahl aktiver Individuen auf einen Wert zwischen 1 und μ gesetzt. In Abbildung 5-11a wurden zwei Evolutionsstrategien mit normaler und fitnessadaptiver Elternselektion verglichen. Auf Datensätzen mit einer IC_{50} -Obergrenze bietet das modifizierte Verfahren Vorteile. In ähnlicher Weise wurde auch eine Anpassung des Simulated Annealings vorgenommen: Hier wurde das Metropolisakzeptanzkriterium (Gl. 3-5) so modifiziert, dass inaktive Verbindungen in jedem Fall zurückgewiesen werden. Diese Modifikation verhindert, dass von einem „aktiven“ Lösungskandidat auf einen „inaktiven“ gesprungen werden kann. Die Partikelschwarmoptimierung wurde so modifiziert, dass inaktive Partikel für das lokale und globale Gedächtnis unzulässig sind.

Eine weitere Modifikation betrifft den Elitismus bei den Evolutionsstrategien. Damit diese auch bei kleinen Populationsgrößen von 1 bis 10 Individuen noch effizient funktionieren, musste ein Elitismus von einem Individuum eingeführt werden. Normalerweise arbeitet die ES entweder gänzlich ohne Elitismus (Komma-Strategie, die Elterngeneration stirbt vollständig aus) oder mit vollständigem Elitismus (Plus-Strategie, die Elterngeneration geht vollständig in den Selektionspool über). Im Rahmen dieser Arbeit lieferte ein „partieller Elitismus“ beste Ergebnisse (Abbildung 5-11b). Hierbei wird eine definierte Anzahl von Elternindividuen in den Selektionspool übernommen.

Bei diskreten Optimierungsproblemen kann es sinnvoll sein, mit *Tabus* zu arbeiten. Ein *Tabu* verhindert für eine einstellbare Dauer die Selektion eines schon ausgewerteten Lösungskandidaten. Die hier verwendeten Optimierungsalgorithmen arbeiten für gewöhnlich ohne eine *Tabu*-Regel. Im Rahmen dieser Arbeit lieferte eine globale *Tabu*-Regel die besten Ergebnisse. Bereits ausgewertete Individuen werden für den weiteren Verlauf der Optimierung mit einem *Tabu* belegt (gesperrt). Besonders Partikelschwarmoptimierungen profitierten von dieser Regel (Abbildung 5-11c).

Um es den Partikelschwarmoptimierungen zu ermöglichen, effizient mit binären Fitnessfunktionen zu arbeiten, musste die Routine zur Aktualisierung der Gedächtnisse modifiziert werden.

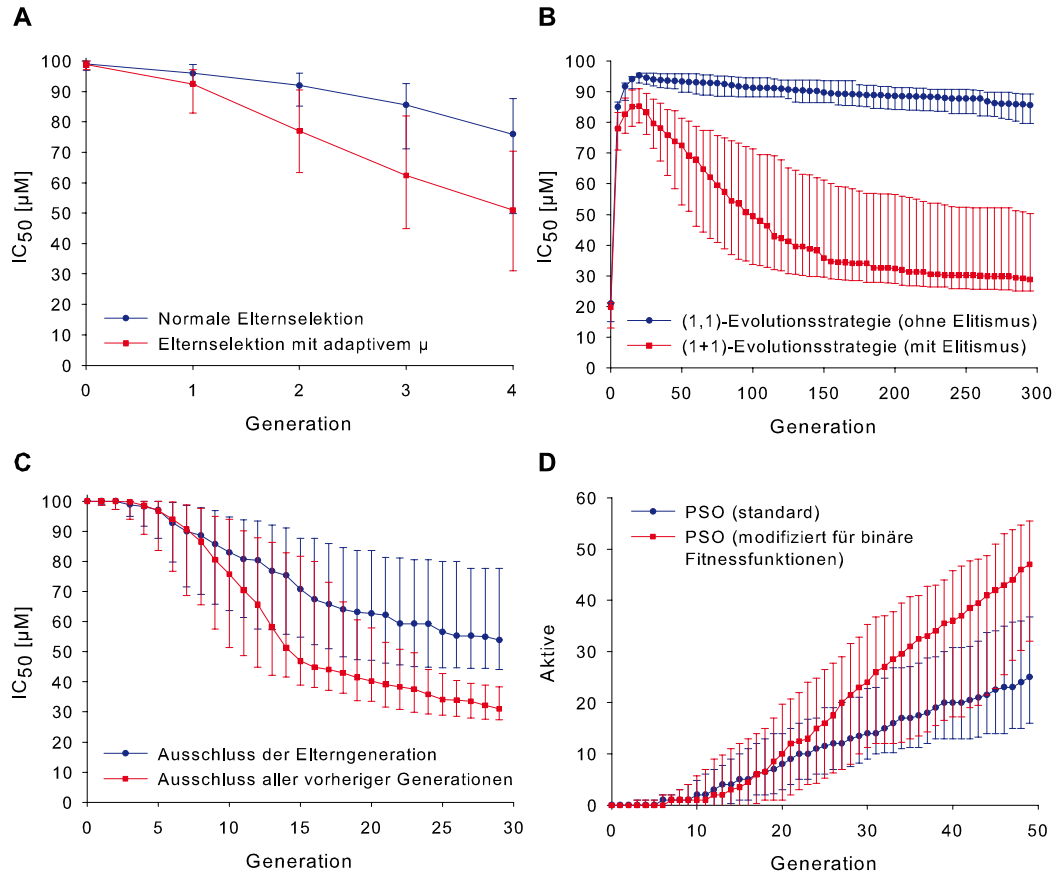


Abbildung 5-11: Modifizierungen der Optimierungsalgorithmen. Bei A)-C) ist der gemittelte IC_{50} der Top-20 aktivsten Verbindungen als Median von 100 Wiederholungen aufgetragen (vgl. Gl. 5-1). Bei D) ist der Median der Anzahl der Aktiven von 100 Wiederholungen aufgetragen. A) (10,60)-ES (5 Generationen, Trypsin-Inhibitoren (Ugi-Bib.), MOE-2D-PCA-Deskriptoren) mit normaler und fitnessadaptiver Elternselektion. Wird der Wert von μ an die Anzahl aktiver Verbindungen angepasst, verbessert sich die Optimierungsleistung, da keine inaktiven Eltern mehr ausgewählt werden können. B) (1,1)-ES und (1+1)-ES (300 Generationen, Trypsin-Inhibitoren (Ugi-Bib.), MOE-2D-PCA-Deskriptoren). Der Elitismus macht die ES-Optimierung mit kleinen Populationsgrößen erst möglich. C) PSO (10 Partikel, 30 Generationen, Trypsin-Inhibitoren (Ugi-Bib.), CATS-Deskriptor) mit verschiedenen *Tabu*-Regeln. Die Optimierungsergebnisse verbessern sich, wenn alle bislang evaluierten Individuen vom Selektionspool ausgeschlossen werden. Wird nur jeweils die Elterngeneration ausgeschlossen, verschlechtern sich die Ergebnisse. D) PSO (3 Partikel, 50 Generationen, Faktor X_a -Inhibitoren (COBRA), MOE-2D-PCA-Deskriptoren) ohne und mit Modifikation für binäre Fitnessfunktionen. Die Routine zur Aktualisierung der Partikelgedächtnisse musste modifiziert werden, um bessere Ergebnisse zu erhalten.

In der Originalimplementierung werden die Gedächtnisse nur angepasst, wenn eine bessere Lösung gefunden wurde. Bei binären Fitnessfunktionen gibt es jedoch unter den aktiven Lösungen keine „besseren“ oder „schlechteren“. Da sie alle den gleichen Wert „aktiv“ besitzen, wird das Gedächtnis niemals aktualisiert und es kommt zur Stagnation der Optimierung. Abhilfe schafft hier die Modifizierung der Aktualisierungsroutine, nicht nur bessere, sondern auch gleich guten Lösungen zuzulassen (Abbildung 5-11d).

5.7 Schlussfolgerungen

Mit Hilfe verschiedener retrospektiver Vergleichsstudien wurde die Nutzbarkeit dreier Optimierungsalgorithmen für das adaptive Design fokussierter Substanzbibliotheken unter der Einschränkung einer limitierten Anzahl an Substanztestungen gezeigt. Für zwei der Algorithmen wurde eine detaillierte Analyse der internen Arbeitsweise vorgestellt. Schon vor über einer Dekade berichteten Weber *et al.* von der erfolgreichen Optimierung einer Bibliothek aus 160.000 Molekülen mit einem Budget von nur 400 Substanztestungen.²⁵ Damals wurde ein genetischer Algorithmus verwendet, um bislang unbekannte submikromolare Thrombin-Inhibitoren zu entdecken.

In der voran gegangenen Studie wurden die drei konzeptionell unterschiedlichen Optimierungsalgorithmen „Simulated Annealing“, „Evolutionsstrategie“ und „Partikelschwarmoptimierung“ mit einander verglichen. Mit sinnvollen Parametern waren alle drei Algorithmen in der Lage, gute Ergebnisse zu liefern. Die Evolutionsstrategie erfüllte jedoch als einziges Verfahren das sekundäre Nebenziel der Experimente vollständig: Die Optimierungsalgorithmen sollten bevorzugt mit nur wenigen Zyklen und dafür großen Populationen arbeiten, da die Studie im Hinblick auf das Niedrigdurchsatzscreening durchgeführt wurde. Evolutionsstrategien sind demnach gut für eine Anwendung im Niedrigdurchsatz-screening geeignet. Auch die Partikelschwarmoptimierung arbeitete mit mittelgroßen Anzahlen von Iterationen zufrieden stellend und ist deshalb ebenfalls für Niedrigdurchsatzexperimente geeignet. Die Partikelschwarmoptimierungen waren zudem einfach zu parametrisieren. Benötigte Strategieparameter waren entweder autoadaptiv (Schrittweitenkontrolle) oder lieferten mit bewährten Grundeinstellungen gute Ergebnisse (Gewichtungsfaktoren des lokalen und globalen Gedächtnisses)⁵⁴. In der Zukunft sollten weitere Anstrengungen unternommen werden, um es der Partikelschwarmoptimierung durch gezielte algorithmische Veränderungen zu ermöglichen, auch mit weniger Iterationen effizient zu arbeiten. Das Simulated Annealing bevorzugte als einziges nicht

populationsbasiertes Verfahren sehr große Iterationszahlen und ist daher in der hier vorgestellten Implementierung weniger für das Niedrigdurchsatzscreening geeignet.

An allen drei Optimierungsalgorithmen wurden bestimmte Modifikationen vorgenommen, um beste Ergebnisse liefern zu können. Es konnte gezeigt werden, dass eine Nächste-Nachbarn-Suche effektiv einen kontinuierlichen Suchraum auf einen diskreten chemischen Raum abbilden kann. Bei den algorithmischen Modifikationen hatten die gesonderte Behandlung inaktiver Verbindungen und eine globale *Tabu*-Regel den größten Einfluss auf die Versuchsergebnisse.

Es konnte weiter gezeigt werden, dass die Optimierung fokussierter Substanzbibliotheken ohne Vorwissen über das biologische Zielmolekül möglich ist. Die Voraussetzungen dafür waren eine ausreichende Häufigkeit aktiver Verbindungen in der Screeningbibliothek (mind. 2%) und eine moderate Neigung zur Clusterbildung im verwendeten Suchraum. Sind aktive Verbindungen im Voraus bekannt, ist es sinnvoll, dieses Vorwissen in den Optimierungsexperimenten zu nutzen. Mit nur vier bekannt aktiven Verbindungen in der Initialpopulation konnten bessere und deutlich robustere Ergebnisse erzeugt werden, als mit zufälliger Selektion der Initialpopulation.

Für das adaptive Design einer fokussierten Ligandenbibliothek im Rahmen einer prospektiven Niedrigdurchsatzscreening-Kampagne lassen sich folgende Schlussfolgerungen ziehen: Als Optimierungsalgorithmus sollte eine Evolutionsstrategie oder eine Partikelschwarmoptimierung eingesetzt werden. Evolutionsstrategien operieren chemisch intuitiver, da Eltern- und Nachkommen-Moleküle eine große Ähnlichkeit besitzen. Die mit Hilfe der Ugi-Bibliothek erarbeiteten Strategieparameter sollten übernommen werden und die Initialpopulation sollte mit bekannten aktiven Liganden als Referenzmoleküle angereichert werden.

6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion

6.1 Zielsetzung

Alle bislang veröffentlichten Studien zum adaptiven Design von Substanzbibliotheken konzentrierten sich auf den Entwurf von Proteinliganden. Hier wird die Hypothese aufgestellt, dass sich die adaptive Optimierung fokussierter Substanzbibliotheken auch auf Ribonukleinsäuren als biologisches Zielmolekül anwenden lässt. Das dritte Ziel dieser Arbeit war daher der Entwurf einer fokussierten Bibliothek von RNA-Liganden mit Hilfe der adaptiven Optimierung.

Es wurde eine hybride, prospektive Optimierungsstudie durchgeführt. Nach dem Vorbild einer iterativen Niedrigdurchsatzscreening-Kampagne wurden vom Computer vorgeschlagene Moleküle im Labor auf Aktivität getestet. Die Ergebnisse der theoretischen Vorarbeiten (vgl. Kapitel 5) dienten als Fundament dieser Studie. Das Ziel der Optimierung war die Findung neuer Liganden der HIV-1 TAR-RNA, die zur Inhibition deren Interaktion mit Tat (Transaktivator der Transkription) befähigt sind. Als Suchraum wurde eine Bibliothek mit Screeningsubstanzen der Firma Specs (Delft, Niederlande) vom September 2006 gewählt.⁶² Die Originalbibliothek umfasste 191.461 Verbindungen. Nach Vorbereitung der Bibliothek für das Screening (vgl. Abschnitt 3.3.4) standen 191.405 Substanzen für die Suche zur Verfügung. Die prospektive Optimierungsstudie besaß ein Budget von maximal 300 Substanztestungen. Die Studie wurde im Rahmen des Sonderforschungsbereiches (SFB) 579 „RNA-Liganden-Wechselwirkungen“, Projekt A11.2, durchgeführt. Die Aktivität potentieller Inhibitoren wurde mittels einer fluoreszenzbasierten Methode bestimmt. In Kooperation mit Dr. Ute Scheffer aus der Arbeitsgruppe von Prof. Dr. Göbel (SFB 579, Projekt A3) stand eine FRET-basierte, kompetitive Verdrängungsmethode für die Bioaktivitätsbestimmung *in vitro* zur Verfügung.

6.2 Planung der prospektiven Studie

Ausgehend von den Ergebnissen der theoretischen Studie boten sich die Partikelschwarmoptimierung und die Evolutionsstrategie als Optimierungsalgorithmen an. Nach sorgfältiger Abwägung fiel die Wahl auf die Evolutionsstrategie. Neben der günstigeren Bevorzugung großer Populationen stellen die evolutionären Algorithmen, zu denen auch die Evolutionsstrategie gehört, ein etabliertes Verfahren für das Design

fokussierter Substanzbibliotheken dar. Das Auffinden von RNA-Liganden mit chemieinformatischen Methoden ist eine Herausforderung. Ribonukleinsäuren sind flexible Makromoleküle, für die das Konzept der Rezeptorrigidität des rezeptorbasierten Designs nur eingeschränkt Gültigkeit besitzt.^{101,152-154} Auch ligandenbasierte Ansätze müssen mit einer großen strukturellen Variabilität und geringer Wirkstoffartigkeit bekannter RNA-Liganden umgehen.¹⁵² Auch auf Grund dieser potentiellen Schwierigkeiten wurde die Evolutionsstrategie als etabliertes Verfahren der unerprobten Partikelschwarmoptimierung vorgezogen.

Es wurde eine Populationsgröße von 60 Molekülen und 5 Generationen festgelegt (Tabelle 6-1). In den retrospektiven Experimenten hatte sich die Initiierung der Optimierung mit bekannten aktiven Verbindungen bewährt. Für die Initialpopulation wurden deshalb zehn aktive TAR-RNA-Liganden aus der Literatur und aktuellen Forschungsprojekten der Universität ausgewählt (RNA-Liganden-Bibliothek; zur Initialpopulation vgl. Abschnitt 6.3). Auf eine zufällige Auswahl der Initialpopulation wurde deshalb komplett verzichtet. Es verblieben daher vier Synthese-und-Test-Zyklen, die bei einer veranschlagten Dauer von zwei Monaten pro Zyklus in rund acht Monaten zu bewältigen sein sollten. Die Anzahl der Eltern wurde nach der Regel (*Populationsgröße / 6*) auf 10 eingestellt, wobei dies den Maximalwert darstellte und sich die tatsächliche Elternzahl adaptiv an die Anzahl aktiver Verbindungen anpasste. Das Maximalbudget betrug 300 Testungen. Durch die Wahl der Initialpopulation konnte das Budget auf 240 Testungen gekürzt werden. Die *Tabu*-Regel wurde auf „Elternausschluss“ (vgl. Abbildung 5-11) eingestellt. Beim „Elternausschluss“ werden die jeweilige Elternindividuen für einen Zyklus mit einem *Tabu* belegt. Im Lauf der Optimierung können daher Substanzen mehr als einmal ausgewählt werden. Dies kann Ressourcen sparen. Der Suchraum bestand aus 27 Hauptkomponenten von 184 MOE-2D-Deskriptoren. Zwar lassen die Ergebnisse der retrospektiven Experimente eine leichte Überlegenheit des CATS-Deskriptors erkennen, jedoch lagen diese Daten zum Planungszeitpunkt noch nicht vollständig vor.

Es wurde auf jegliche Vorfilterung der Screeningmoleküle (z.B. Filterung nach Wirkstoffartigkeit, Filterung nach Wasserlöslichkeit, etc.) verzichtet, um keine potentiell falschen Annahmen über RNA-Liganden zu machen. Vielmehr sollten sich Moleküle mit ungewünschten Eigenschaften im Laufe der Optimierung selbstständig mit Hilfe simulierter Evolution eliminieren.

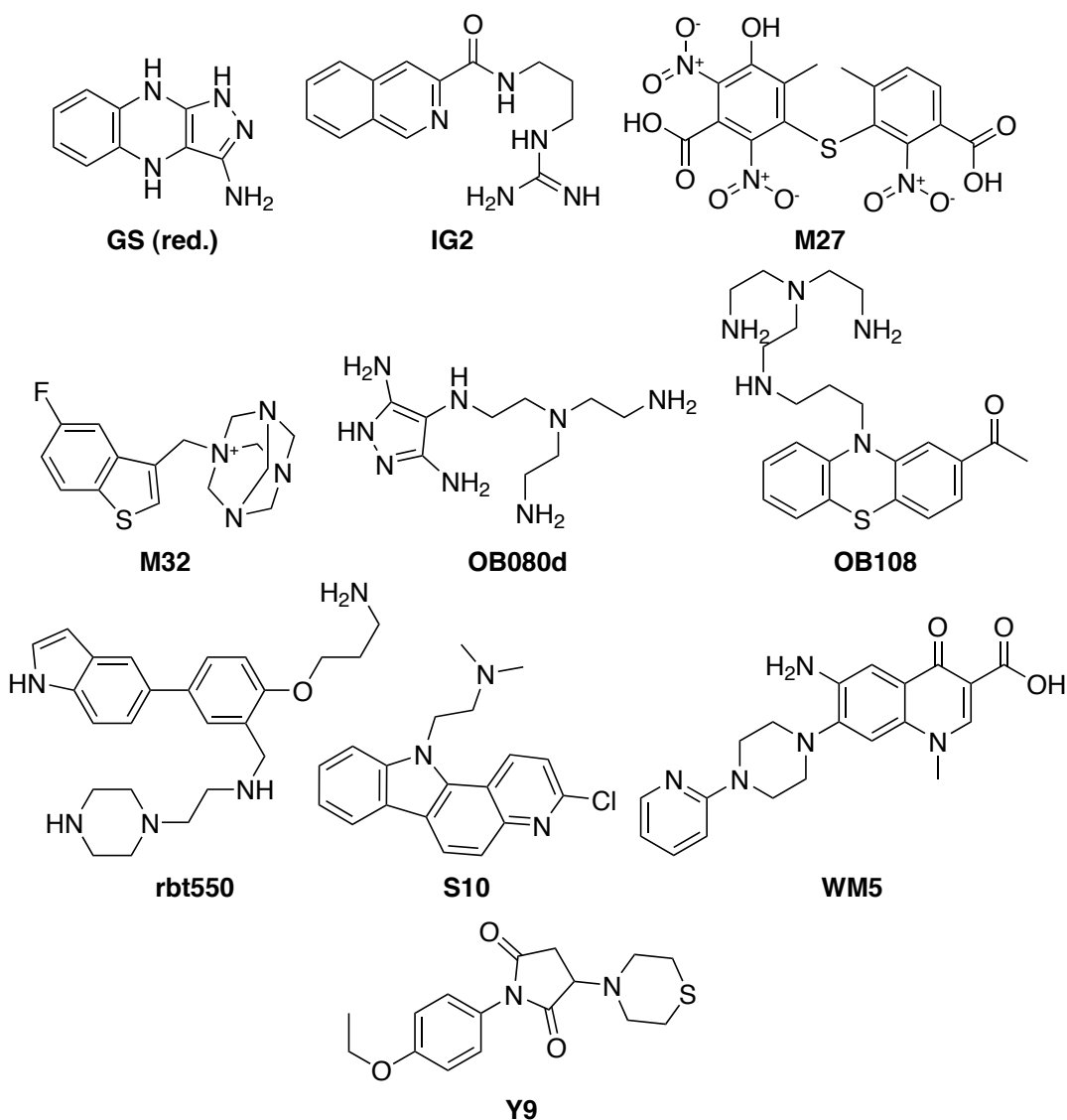
Ausgehend von der Initialpopulation wurde die Optimierung drei Mal gestartet, um drei Vorschläge für die erste Generation zu erhalten. Von den drei Vorschlägen wurde einer nach persönlichem Ermessen ausgewählt. Dabei wurde auf die Diversität der Strukturklassen, Wirkstoffartigkeit und Ähnlichkeit zu privilegierten RNA-Bindemotiven geachtet. Danach wurde die Optimierung sich selbst überlassen und ohne weiteres geplantes Eingreifen durchgeführt.

Tabelle 6-1: Optimierungsparameter der prospektiven TAR-RNA-Liganden-Designstudie

Parameter	Wert
Optimierungsalgorithmus	Evolutionsstrategie
Eltern	1 – 10 (fitnessadaptiv)
Populationsgröße	60
Generationen	5
Schrittweite	0,5 (autoadaptiv)
Strategie	(μ, λ)
Taburegel	Elternausschluss
Substanzbibliothek	Specs (Sept. 2006, 191.405 Screeningverbindungen)
Chemischer Raum	27 Hauptkomponenten von 184 MOE-2D-Deskriptoren
Initialpopulation	10 bekannte Aktive
Budget	300 Testungen

6.3 Die Initialpopulation

Sieben der zehn Verbindungen der Initialpopulation wurden der zum damaligen Stand (Okt. 2006) aktuellen Forschung der Universität Frankfurt entnommen (RNA-Liganden-Bibliothek). Drei weitere sind der Literatur zur TAR-RNA-Forschung entnommen. Es folgt eine kurze Beschreibung der Moleküle der Initialpopulation (Schema 6-1):



Schema 6-1: Moleküle der Initialpopulation

GS (red.)

Reduzierte Spezies der Verbindung **GS** aus der Dissertation von Dr. Gunther Seifert (AK Prof. Dr. Göbel).¹⁰³ Zum Planungszeitpunkt wurde der IC_{50} mit 5 μ M angegeben. Die Substanz stellte damit die zweitaktivste kleine organische Verbindung der laufenden TAR-RNA-Forschung an der Universität dar. Diese Aussage wurde jedoch nach weitergehenden Untersuchungen revidiert. Die Reduktion konnte nur durch eine große Menge des Reduktionsmittels Dithionit erreicht werden. Dithionit verfälscht das Ergebnis der Messmethode jedoch sehr erheblich. Mangels einer Negativkontrolle ist das Versuchsergebnis deshalb auf den Effekt des Reduktionsmittels zurück zu führen. Die oxidierten Spezies sind im Vergleich gar nicht oder nur marginal aktiv. Verbindung **GS (red.)** ist *falsch positiv*.

IG2

Bei dieser Verbindung handelt es sich um ein Isoquinolin-Derivat aus der Publikation von He *et al.* 2005.¹⁵⁵ Die Verbindung bewirkte in einem colorimetrischen Enzymtest 61% Inhibition der CAT (Chloramphenicol-Acetyltransferase)-Aktivität und in einem antiviralen zellbasierten Test wurde ein EC_{50} von 3,96 μ M bestimmt. Besonderes Interesse kam **IG2** zu, da diese Verbindung als Isoquinolin-Derivat Ähnlichkeit zu Tetrahydroquinolinen aufweist, welche in einer parallelen Studie untersucht wurden.

M27

Nitroverbindung aus der Dissertation von Dr. Manuel Nietert (AK Prof. Dr. Schneider) mit einer Inhibitionsrate von $152\% \pm 2\%$.¹⁰¹ Diese Verbindung wurde zum Planungszeitpunkt der prospektiven Studie als aktiver RNA-Ligand betrachtet. Diese Aussage wurde jedoch später revidiert und die Substanz in ihrer Eigenschaft als Polyanion als Tat-Peptid-Binder betrachtet. Verbindung **M27** ist deshalb *falsch positiv*. Specs-Nr.: AG-205/33150021

M32

Eine von Dr. Manuel Nietert (AK Prof. Dr. Schneider) im Rahmen seiner Doktorarbeit gefundene Struktur mit charakteristischem Urotropin-Motiv (1,3,5,7-Tetraazaadamantyl-Gruppe).¹⁰¹ Die Verbindung wies $32\% \pm 1\%$ Inhibition der Tat-TAR-Interaktion in FRET-Messungen auf. In einem Transkriptions-/Translationsexperiment zur Bestimmung einer antibakteriellen Wirkung erreichte die Verbindung eine Inhibitionsrate von $40\% \pm 3\%$. Specs-Nr.: AA-516/31407048

OB080d

Triaminopyrazol-Derivat aus der Dissertation von Dr. Oliver Boden (AK Prof. Dr. Göbel).⁹⁸ Zum Planungszeitpunkt wurde die Aktivität in den FRET-Messungen als $IC_{50} = 80 \mu\text{M}$ angegeben. Die Verbindung war damit das fünftpotenteste kleine organische Molekül aus den TAR-RNA-Projekten der Universität. Weitergehende Studien haben jedoch gezeigt, dass die Verbindung nicht rein vorlag. Es wurde festgestellt, dass Triaminopyrazole allgemein zu einer schnellen Polymerisierung innerhalb weniger Stunden neigen (z.B. über Nacht). Sie liegen damit bei Versuchsbedingungen als Polykationen vor und besitzen somit *per se* eine hohe Affinität zu Nukleinsäuren. Z.B. verbesserte sich der IC_{50} von Triaminopyrazol über Nacht von 2-3 mM durch Polymerisierung auf 200-300 μM (persönliche Korrespondenz mit Prof. Dr. Michel Göbel und Dr. Ute Scheffer). Das Grundprinzip der Triaminopyrazole war die Schaffung eines „Superguanidiniums“, wobei Guanidinium die kleinste kationische Einheit der Arginin-Seitenkette ist und mehrere Einheiten Arginin wesentlicher Bestandteil der TAR-Bindedomäne von Tat sind. Interessanterweise verbessert sich die Aktivität der Triaminopyrazole um den Faktor 10, wenn der pH von 7,4 (Versuchsbedingung) auf 6,0 eingestellt wird. Die Verbindungen liegen dann als Dikationen vor und entsprechen somit besser dem Konzept eines Superguanidiniums. Die pH-Wert-Änderung verbessert den IC_{50} des Triaminopyrazols von 2-3 mM auf 700-800 μM . Argininamid weist als Kontrolle sowohl bei pH 7,4 also auch bei pH 6,0 einen IC_{50} von rund 1,5 mM auf. Wegen der ungewissen Reinheit der Verbindung und deren Neigung zur Polymerisation ist **OB080d** kein guter Kandidat zum Ableiten von Struktur-Wirkungs-Beziehungen. Sie wird deshalb als *potentiell falsch positiv* betrachtet.

OB108

Von Acetylpromazin (Schema 6-2) abgeleitete Verbindung aus der Dissertation von Dr. Oliver Boden (AK Prof. Dr. Göbel).⁹⁸ Zum Planungszeitpunkt war diese Substanz mit einem IC_{50} von 50 nM die potenteste kleine organische Verbindung aus den laufenden TAR-RNA-Projekten der Universität. Jedoch haben weitergehende Studien gezeigt, dass die Substanz instabil ist. Gelöste Proben verfärbten sich mit der Zeit von gelblich nach schwarz, was einen Zersetzungsprozess vermuten lässt. Die Reinheit der Proben konnte deshalb nicht gewährleistet werden. Des Weiteren könnte die Verbindung unter Imin-Bildung polymerisieren. Als Polykation wäre deshalb *per se* eine hohe Affinität zu Nukleinsäuren zu erwarten. Wenn die Substanz in Reinform vorliegt, ist ihr IC_{50} mit einiger Sicherheit niedriger als der von Acetylpromazin ($IC_{50} = 500 \mu\text{M}$). Wegen der Instabilität ist sie jedoch kein guter Kandidat zum Ableiten von Struktur-Wirkungs-Beziehungen. **OB108** wird als *potentiell falsch positiv* betrachtet.

rbt550

Diese Verbindung ist ein Vertreter aus einer ganzen Reihe ähnlicher Derivate, die systematisch von Murchie *et al.* 2004 untersucht wurden.¹⁵⁶ Bei **rbt550** handelt es sich mit einem K_i von 51 nM um den potentesten Vertreter der untersuchten kationischen Substanzen mit Benzolgrundgerüst. Die Aktivität der Verbindung wurde zusätzlich durch die hochauflösende NMR-Aufklärung eines Komplexes von **rbt550** mit TAR-RNA bestätigt.

S10

Mit Hilfe einer alignierungsfreien Fuzzy-Pharmakophorsuche wurde diese Verbindung abgeleitet von Acetylpromazin und dem Acridinderivat CGP40336A (Schema 6-2) im Rahmen der Doktorarbeit von Dr. Steffen Renner (AK Prof. Dr. Schneider) gefunden.¹⁰² Die Aktivität wurde mit $IC_{50} = 46 \mu\text{M}$ in den FRET-Messungen bestimmt. **S10** repräsentierte zudem zum Planungszeitpunkt das viertpotenteste Molekül aus der TAR-RNA-Forschung an der Universität.

WM5

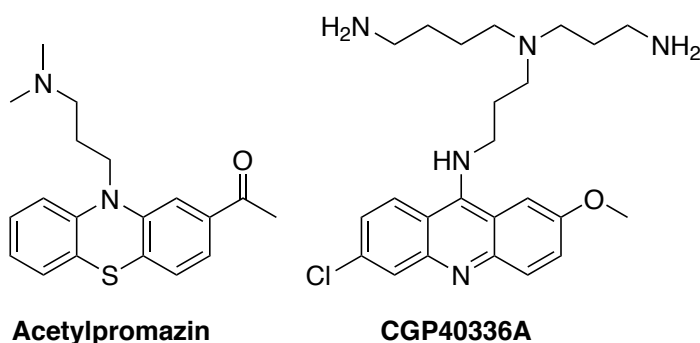
Dieses 6-Aminoquinolon-Derivat wurde von Cecchetti *et al.* 2000 und Parolin *et al.* 2003 publiziert.^{157,158} Neben der gesicherten Aktivität von $K_i = 3,5 \mu\text{M}$ weist diese Verbindung Ähnlichkeit zu den Tetrahydroquinolinen auf, die in einer parallelen Studie untersucht wurden. Es konnte gezeigt werden, dass die Anwesenheit von Mg^{2+} -Ionen im Reaktionspuffer essentiell für die Bindung von Quinolon-Derivaten an Nukleinsäuren sind. Die Versuchsbedingungen für die FRET-Messungen geben jedoch einen magnesiumfreien Puffer vor. Die Aktivität von WM5 könnte somit bei den gewählten Bedingungen nicht reproduziert werden. WM5 ist damit – zumindest bei FRET-Versuchsbedingungen – *potentiell falsch positiv*.

Y9

Verbindung aus der Dissertation von Yusuf Tanrikulu (AK Prof. Dr. Schneider) mit einem neuartigen Succinimidyl-Grundgerüst.¹⁰⁵ Die Inhibitionsrate der Tat-TAR-Interaktion betrug $19\% \pm 4\%$. Dabei handelte es sich jedoch um eine optimistische Abschätzung durch zu Hilfenahme eines Korrekturfaktors gegen den Effekt von 1% DMSO im Lösungsmittel. Eine pessimistische Betrachtung unter Auslassung eines Positiv-Ausreißers der Dreifachmessung liefert eine Inhibitionsrate von $8\% \pm 0,4\%$. Aus der persönlichen Erfahrung mit dieser Messmethode muss jedoch mit einer Schwankung der Messwerte um 10 – 15% gerechnet werden.

Die Aktivität dieser Verbindung ist deshalb nicht gesichert und **Y9** wird als *potentiell falsch positiv* betrachtet. Specs-Nr.: AN-919/14384247

Trotz sorgfältiger Auswahl der Moleküle der Initialpopulation muss retrospektiv festgestellt werden, dass nur 4 von 10 Substanzen einwandfreie Tat-TAR-Interaktionsinhibitoren waren. 4 Verbindungen waren *potentiell falsch positiv* und 2 Verbindungen gesichert *falsch positiv*.



Schema 6-2: Strukturformeln der Tat-TAR-Interaktionsinhibitoren Acetylpromazin ($IC_{50}^{FRET} = 500 \mu\text{M}$)¹⁵⁹ und CGP40336A ($IC_{50} = 22 \text{ nM}$ *in vitro* und $IC_{50} = 1,2 \mu\text{M}$ in einem Zellkultur-Experiment).¹⁶⁰

6.4 Durchführung

Für die prospektive Optimierung einer fokussierten TAR-RNA-Ligandenbibliothek wurde ein auf die *Cyclops*-Programm-Bibliothek aufsetzendes Java-Programm geschrieben. Mit dessen Hilfe wurden ausgehend von der Initialpopulation drei Vorschläge für die erste Generation von Substanzen generiert. Von den drei Vorschlägen wurde einer manuell nach den subjektiven Kriterien „Diversität“, „Wirkstoffartigkeit“ und „RNA-Liganden-Ähnlichkeit“ ausgewählt. Die Substanzen wurden als 10 mM Stammlösungen (Lösungsmittel: DMSO) in Volumina von 500 μL bei Specs (Delft, Niederlande)⁶² bestellt und gefroren auf Trockeneis geliefert. Die Lieferzeit betrug in der Regel 3 Woche. Von den Stammlösungen wurden zunächst wässrige Verdünnungen (1 mL, 10% DMSO) mit einer Substanzkonzentration von 1 mM angesetzt. Bei diesen Bedingungen unlösliche Substanzen wurden bei einer Konzentration von 1 mM in 10% DMSO und 90% Methanol gelöst. Weiterhin unlösliche Proben wurden mit einer Tischzentrifuge zentrifugiert (20800 rfc) und der Überstand wurde weiterverwendet. Die FRET-Messungen wurden daraufhin wie beschrieben durchgeführt (vgl. Abschnitt 3.2). Von aktiven Substanzen wurden Titrationskurven (Dosis-Wirkungs-

Kurven) zum Abschätzen des IC_{50} bestimmt. Die Analyse der Substanzen nahm je Generation ungefähr 1,5 Monate in Anspruch. Die gemessenen Aktivitäten wurden dem Optimierungsprogramm zu geführt, mit dem im Anschluss eine neue Generation von Molekülvorschlägen generiert wurde und der Zyklus wiederholte sich (vgl. Abbildung 1-1).

6.5 Übersicht der experimentellen Ergebnisse

In der Zeit von Herbst 2006 bis Sommer 2007 wurden vier der fünf geplanten Generationen realisiert. Da die Initialpopulation aus Liganden bekannter Aktivität bestand, verblieben drei Zyklen aus Synthese und Testung. Es wurden insgesamt 180 Specs-Moleküle als potentielle Tat-TAR-Interaktionsinhibitoren vorgeschlagen. 8 Substanzen wurden in mehr als einer Generation vorgeschlagen und 2 Substanzen waren nicht lieferbar, sodass insgesamt 170 Verbindungen von Specs bezogen und in FRET-Messungen auf Aktivität überprüft wurden. Der vollständige evolutionäre Stammbaum der 180 Molekülkandidaten ist in Schema 6-3 abgebildet. Weitere Daten zu den Verbindungen (z.B. Specs-Bestellnummer) sind im Anhang in Abschnitt 8.3 zusammengestellt.

Unter den 170 getesteten Verbindungen wurden insgesamt 9 Substanzen positiv auf Inhibition der Tat-TAR-Interaktion getestet (Tabelle 6-2). Dies entspricht einer Trefferquote von 5,3%, wobei lediglich 0,089% der Verbindungen der Specs-Bibliothek untersucht wurden. Die mittleren inhibitorischen Konzentrationen (IC_{50}) lagen im mittleren mikromolaren Bereich ($IC_{50} = 506 \pm 497 \mu\text{M}$), wobei der potenteste Kandidat **AS_S10/06_33** ([2-(4-Chloro-phenoxy)-ethyl]-(2-nitro-5-piperazin-1-yl-phenyl)-amin) einen IC_{50} von $51 \mu\text{M}$ aufwies (Abbildung 6-1). **AS_S10/06_33** ist ein Nachkomme des Elters **OB108**. Zahlreiche zu **AS_S10/06_33** analoge Nitroarene wurden vom evolutionären Algorithmus vorgeschlagen, von denen sechs ebenfalls als aktiv identifiziert werden konnten (Abbildung 6-1). In einem anderen Zweig des evolutionären Stammbaumes konnte Verbindung **AS_S10/06_16** (2-(5-Amino-4-cyano-1H-pyrazol-3-yl)-N-hydroxy-acetamidin) mit $IC_{50} = 116 \mu\text{M}$ als TAR-RNA-Ligand identifiziert werden (Abbildung 6-1). Diese von **OB080d** abstammende Verbindung weist ein 1H-Pyrazol-Grundgerüst auf. In Abbildung 6-1 ist der evolutionäre Stammbaum der gefundenen aktiven Verbindungen dargestellt. Die zur Abschätzung des IC_{50} -Wertes gemessenen Dosis-Wirkungs-Kurven sind im Anhang in Abschnitt 8.4 abgebildet.

Bei der verwendeten fluoreszenzbasierten Messmethode mussten einige Bedingungen beachtet werden: Die Methode setzt für eine korrekte Funktionsweise voraus, dass sowohl das farbstoffmarkierte Tat-Peptid als auch ein potentieller Ligand in die Bulge von TAR-RNA binden und dass ein potentieller Ligand kompetitiv mit dem Tat-Peptid um die Bindestelle konkurriert. Des Weiteren wird angenommen, dass binäre Komplexe aus Tat-Peptid und TAR-RNA vorliegen. Es konnte gezeigt werden, dass letztere Bedingung nicht immer erfüllt ist.¹⁶¹ Für die Diskussion der Ergebnisse wird in dieser Arbeit von binären Komplexen ausgegangen.

Verbindung **AS_S10/06_73** wurde in Einzelpunktmessungen zunächst als aktiv gemessen, jedoch wurde diese Aussage durch spätere Untersuchungen wieder falsifiziert. Es zeigte sich, dass **AS_S10/06_73** potentiell an das Tat-Peptid bindet, wie ein Titrationsexperiment in Abwesenheit von TAR-RNA ergab. Damit war eine Bedingung der Messmethode nicht erfüllt und die Verbindung wurde als *falsch positiv* eingestuft.

Verbindung **AS_S01/07_37** wurde ebenfalls zunächst als aktiv gemessen und wurde später als *falsch positiv* eingestuft. Diese Verbindung war schlecht löslich und präzipitierte im Laufe der Messung in den Vertiefungen der Mikrotiterplatte. Es besteht die Möglichkeit, dass der weißliche Niederschlag das Anregungslicht streute und somit eine artifizielle Absenkung des Fluoreszenzwertes verursachte. Die Verbindung wird daher als inaktiv betrachtet.

Eine weitere Möglichkeit der artifiziellen Absenkung des Fluoreszenzwertes im Messsystem besteht durch Absorption der Anregungswellenlänge von 489 nm. Einige der hier getesteten Nitroarene zeigten diesen Effekt. Für die schwachaktiven Verbindungen **AS_S01/07_62**, **AS_S06/07_27** und **AS_S06/07_28** wurden deshalb zusätzlich Messungen bei alternativen Bedingungen durchgeführt (Anregungswellenlänge: 550 nm, Emissionswellenlänge: 570 nm). Es zeigte sich, dass die ursprünglich bestimmten IC_{50} -Werte durch Absorptionseffekte eine zu optimistische Abschätzung darstellten. Es wurden daher die bei alternativen Bedingungen bestimmten Aktivitätswerte verwendet. Verbindungen **AS_S06/07_27** und **AS_S06/07_28** binden TAR-RNA nur schwach im niedrigen millimolaren Bereich. Die Angabe eines IC_{50} -Wertes ist hier nur als Richtwert zu verstehen. Wegen ihres reproduzierbaren inhibitorischen Effektes, werden sie jedoch für die weitere Diskussion in dieser Arbeit als aktiv betrachtet.

Bei Substanz **AS_S10/06_16** wurden reproduzierbar negative %Effekt-Werte bestimmt (z.B. -36% bei 500 μM Substanzkonzentration). Die angepasste Dosis-Wirkungs-Kurve steht jedoch in guter Übereinstimmung zu den Messwerten ($R^2 = 0,9966$; vgl. Abschnitt 8.4.1). Es kann spekuliert werden, dass Verbindung **AS_S10/06_16** bei Konzentrationen $> 200 \mu\text{M}$ Aggregationseffekten unterliegt. Dies konnte jedoch nicht gezeigt werden.

Für die Verbindungen **AS_S01/07_19**, **AS_S01/07_35**, **AS_S01/07_62**, **AS_S06/07_27**, **AS_S06/07_28** und **AS_S06/07_29** war aufgrund schwacher Aktivität und schlechter Löslichkeit bei hohen Konzentrationen kein minimaler Plateauwert der Dosis-Wirkungs-Kurve bestimmbar. Diese Werte wurden während der Kurvenanpassung mit Sigmaplot (Version 9.1, SyStat Software, Inc., San Jose, USA)⁹⁰ extrapoliert.

Die Titrationsmessung von Verbindung **AS_S06/07_40** war aufgrund schlechter Löslichkeit nur bis zu einer Konzentration von 200 μM bei einem hohen Anteil organischen Lösungsmittels (10% DMSO) durchführbar. Eine Messung des Absorptionsspektrums bei verschiedenen Konzentrationen (bis zu einer Konzentration von 475 μM) zeigte jedoch eine annähernd lineare Abhängigkeit der Extinktion von der Konzentration (gemessen in Abwesenheit von Tat-Peptid und TAR-RNA), was nicht auf einen Präzipitationseffekt hindeutet. Es besteht die Möglichkeit, dass Verbindung **AS_S06/07_40** im Komplex mit TAR-RNA und/oder dem Tat-Peptid präzipitiert. Dies konnte nicht eindeutig geklärt werden. Für die weitere Diskussion in dieser Arbeit wird die Verbindung wegen ihres reproduzierbaren inhibitorischen Effektes als aktiv betrachtet.

Ein Hinweis zur Nomenklatur der anschließenden Betrachtungen: Die Initialgeneration wird als Generation Null bezeichnet. Sie ist die *nullte* Generation. Darauf folgt Generation eins, die erste Generation, und darauf Generation zwei, die zweite Generation, usw.

Tabelle 6-2: Liste der aktiven Verbindungen. Der IC_{50} und das 95%-Konfidenzintervall beziehen sich auf die Funktionsanpassung der Messwerte an eine logistische Vier-Parameter-Funktion (Dosis-Wirkungs-Kurve). Die Funktionsanpassung wurde mit Sigmaplot (Version 9.1, SyStat Software, Inc.) durchgeführt.⁹⁰ Gen. bezeichnet die Generation, MG das Molekulargewicht.

Gen.	Nachkomme	Specs-Nr.	Elter	IC_{50}	95%-Konfidenzintervall	MG
1	AS_S10/06_16	AN-989/40886562	OB080d	116 μ M	102 – 131 μ M	180 Da
		2-(5-Amino-4-cyano-1H-pyrazol-3-yl)-N-hydroxy-acetamidin				
1	AS_S10/06_33	AG-690/15428920	OB108	51 μ M	46 – 56 μ M	378 Da
		([2-(4-Chloro-phenoxy)-ethyl]-(2-nitro-5-piperazin-1-yl-phenyl)-amin				
2	AS_S01/07_19	AG-670/40809134	AS_S10/06_33	269 μ M	180 – 403 μ M	398 Da
		1-(4-Methoxy-phenyl)-4-(2-nitro-5-piperazin-1-yl-phenyl)-piperazin				
2	AS_S01/07_35	AG-670/40741579	AS_S10/06_16	823 μ M	551 – 1230 μ M	214 Da
		2-(4,6-Diamino-pyrimidin-2-ylsulfanyl)-propionsäure				
2	AS_S01/07_62	AG-670/12382002	AS_S10/06_33	493 μ M ^a	218 – 1114 μ M	313 Da
		Benzyl-(2-nitro-5-piperazin-1-yl-phenyl)-amin				
3	AS_S06/07_27	AG-670/12382010	AS_S01/07_62	(1114 μ M) ^{a,b}	118 - 104 μ M	267 Da
		2-(2-Nitro-5-piperazin-1-yl-phenylamino)-ethanol				
3	AS_S06/07_28	AG-670/12382031	AS_S01/07_62	(1409 μ M) ^{a,b}	keine Angabe ^c	281 Da
		2-[5-(3-Methyl-piperazin-1-yl)-2-nitro-phenylamino]-ethanol				
3	AS_S06/07_29	AG-670/12382017	AS_S01/07_62	162 μ M	125 – 210 μ M	265 Da
		N1-(4-Nitro-3-piperidin-1-yl-phenyl)-ethane-1,2-diamin				
3	AS_S06/07_40	AK-968/40319755	AS_S01/07_19	115 μ M	102 – 129 μ M	421 Da
		2-(4-Nitro-phenyl)-5-phenyl-4-[(2-piperazin-1-yl-ethylimino)-methyl]-1,2-dihydro-pyrazol-3-on				

^a Die Substanz absorbierte im Anregungsbereich von 489 nm und verfälschte die Messungen hin zu einer optimistischeren Abschätzung der mittleren inhibitorischen Konzentration. Die Angaben beziehen sich daher auf eine Analyse bei alternativen Bedingungen (Anregungswellenlänge: 550 nm, Emissionswellenlänge: 570 nm).

^b Der IC_{50} -Wert ist wegen schwacher Aktivität nur als Richtwert zu verstehen.

^c Für die Kurvenanpassung wurde ein artifizieller Messpunkt bei Konz. = 10^4 μ M mit %Effekt = 0% hinzugefügt. Das Konfidenzintervall kann deshalb nicht angegeben werden.

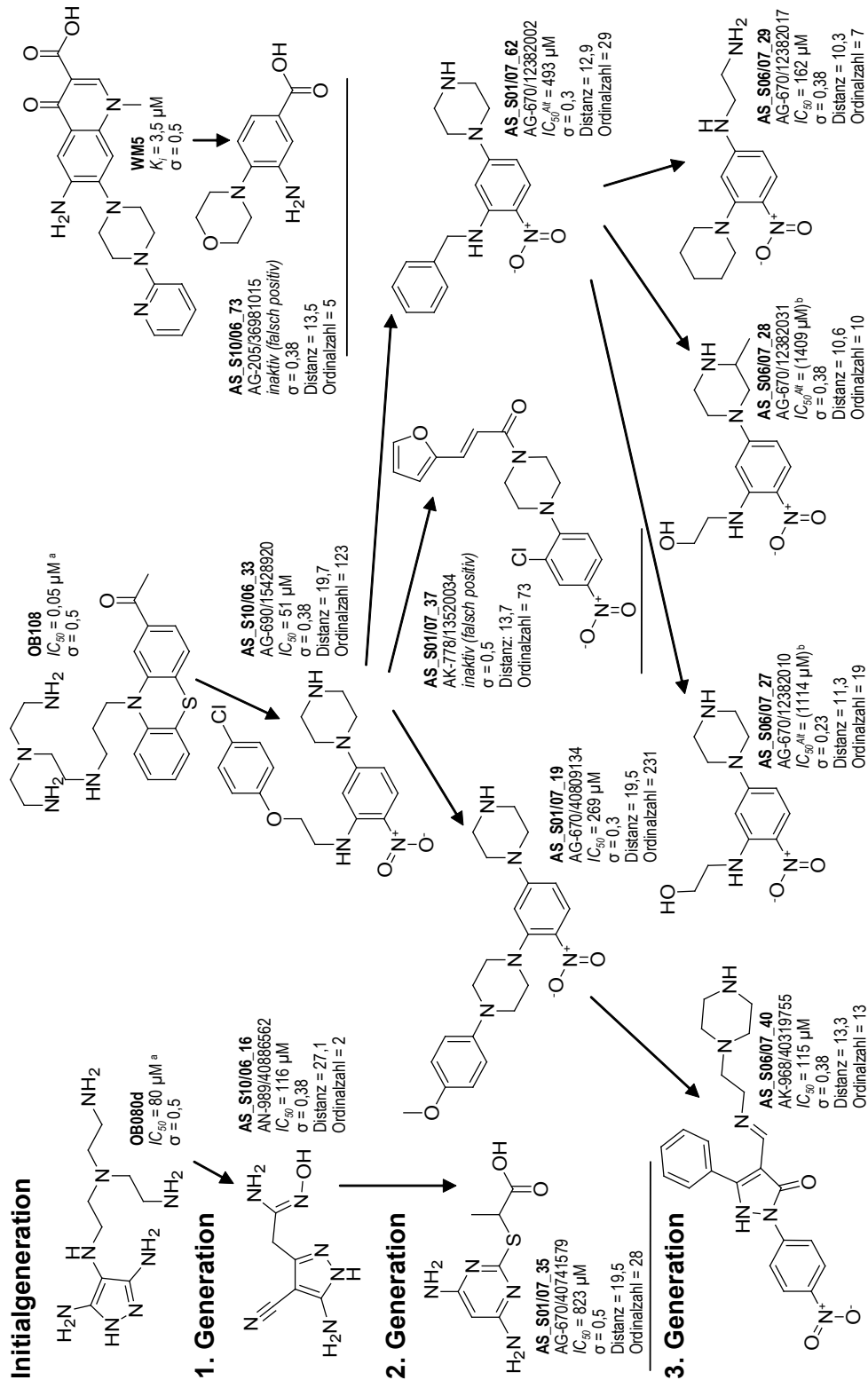
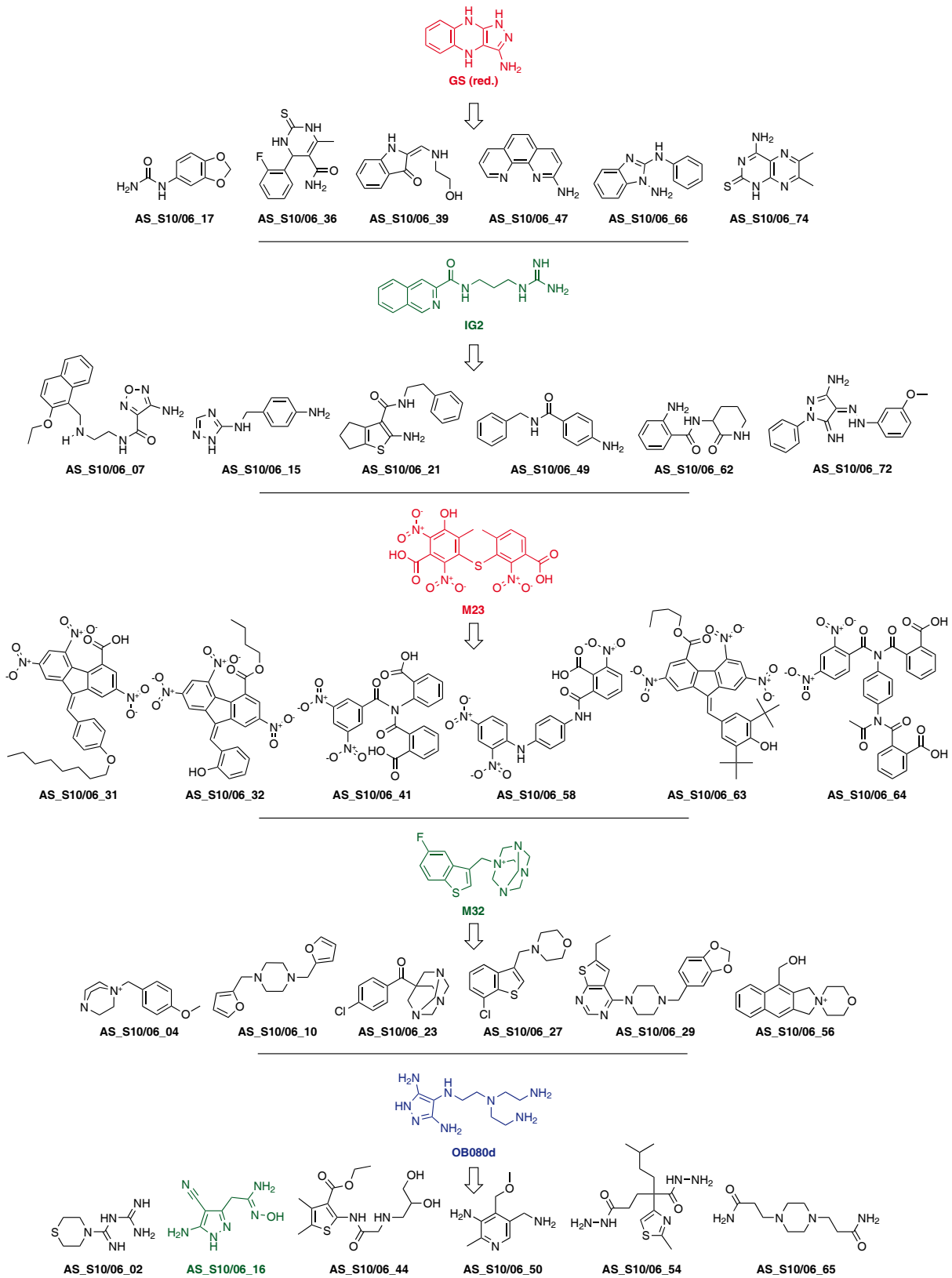


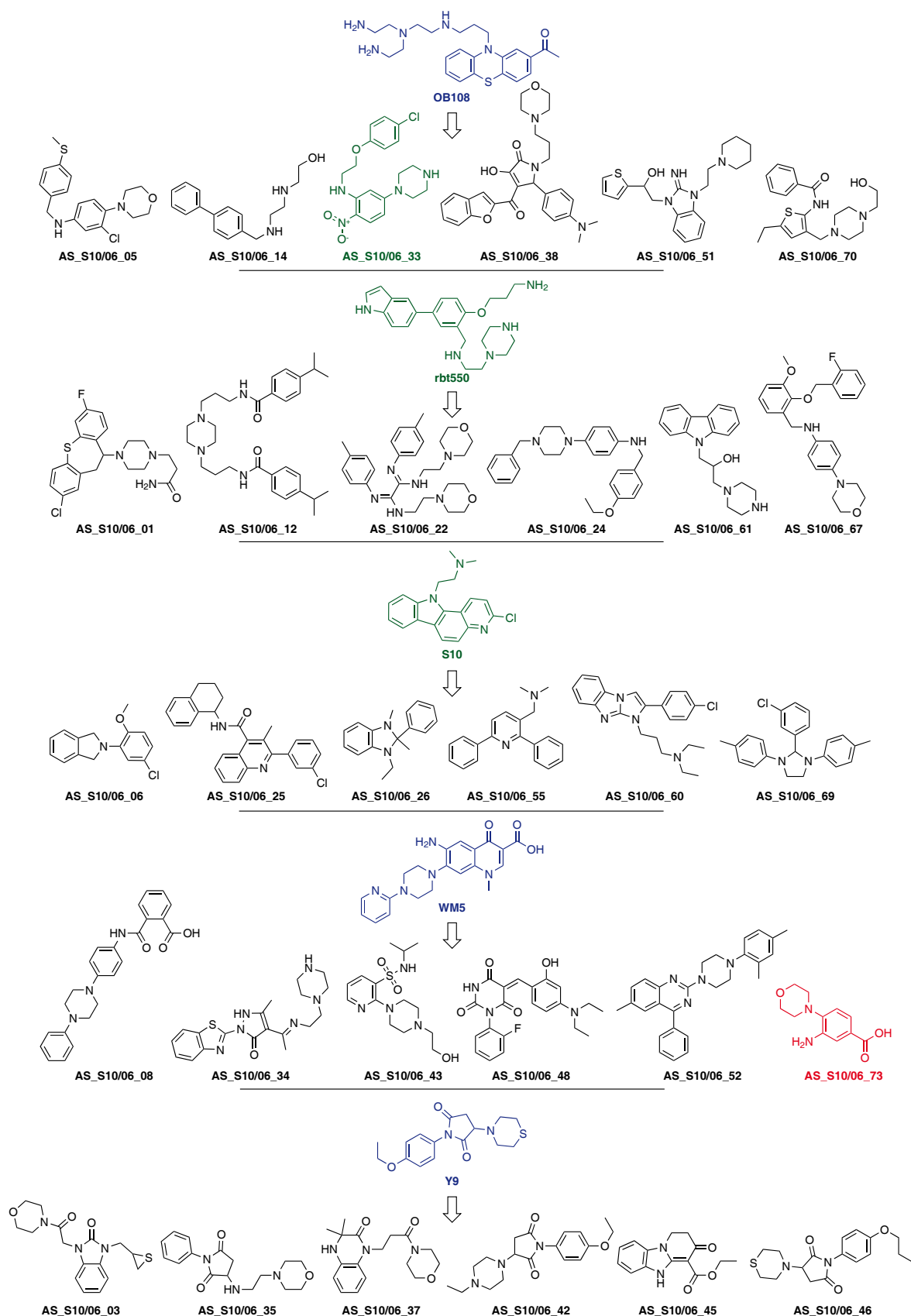
Abbildung 6-1: Stammbaum der adaptiven Optimierung von Tat-TAR-Interaktionsinhibitoren. σ bezeichnet die Schrittweite der ES; Distanz (Manhattan) und Ordinalzahl („der wievielte Nachbar?“) sind jeweils relativ zum Elter angegeben. IC_{50}^{Alt} -Werte wurden bei alternativen Bedingungen bestimmt (vgl. Abschnitt 3.2). ^a Unbestätigte Werte (siehe Text). ^b Richtwerte (siehe Text).

6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



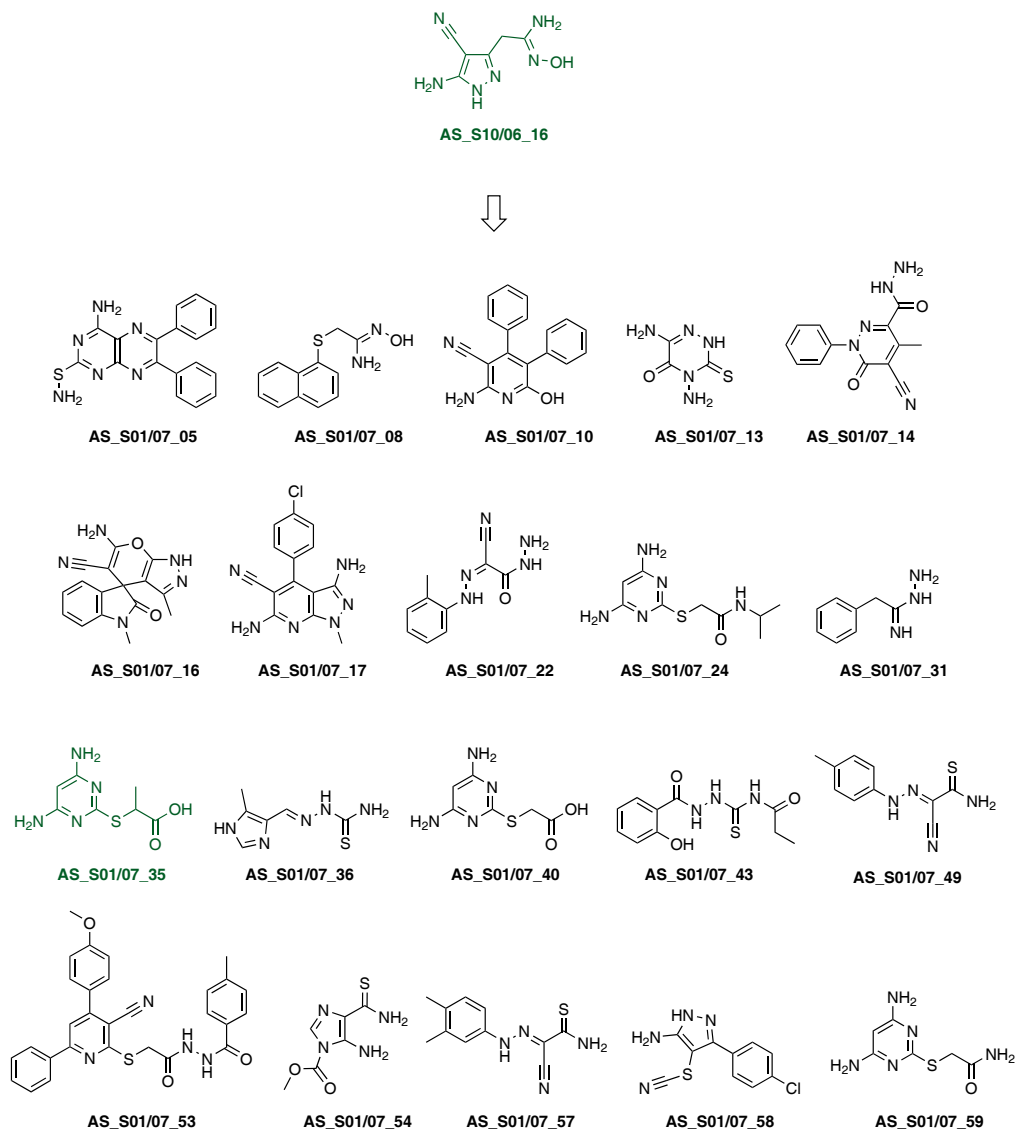
Schema 6-3a: Eltern der Initialpopulation (0. Generation) mit ihren Nachkommen der 1. Generation (Teil 1). Farblegende: grün: aktiv, schwarz: inaktiv, blau: potentiell falsch positiv, rot: falsch positiv.

6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



Schema 6-3b: Eltern der Initialpopulation (0. Generation) mit ihren Nachkommen der 1. Generation (Teil 2). Farblgende: grün: aktiv, schwarz: inaktiv, blau: potentiell falsch positiv, rot: falsch positiv.

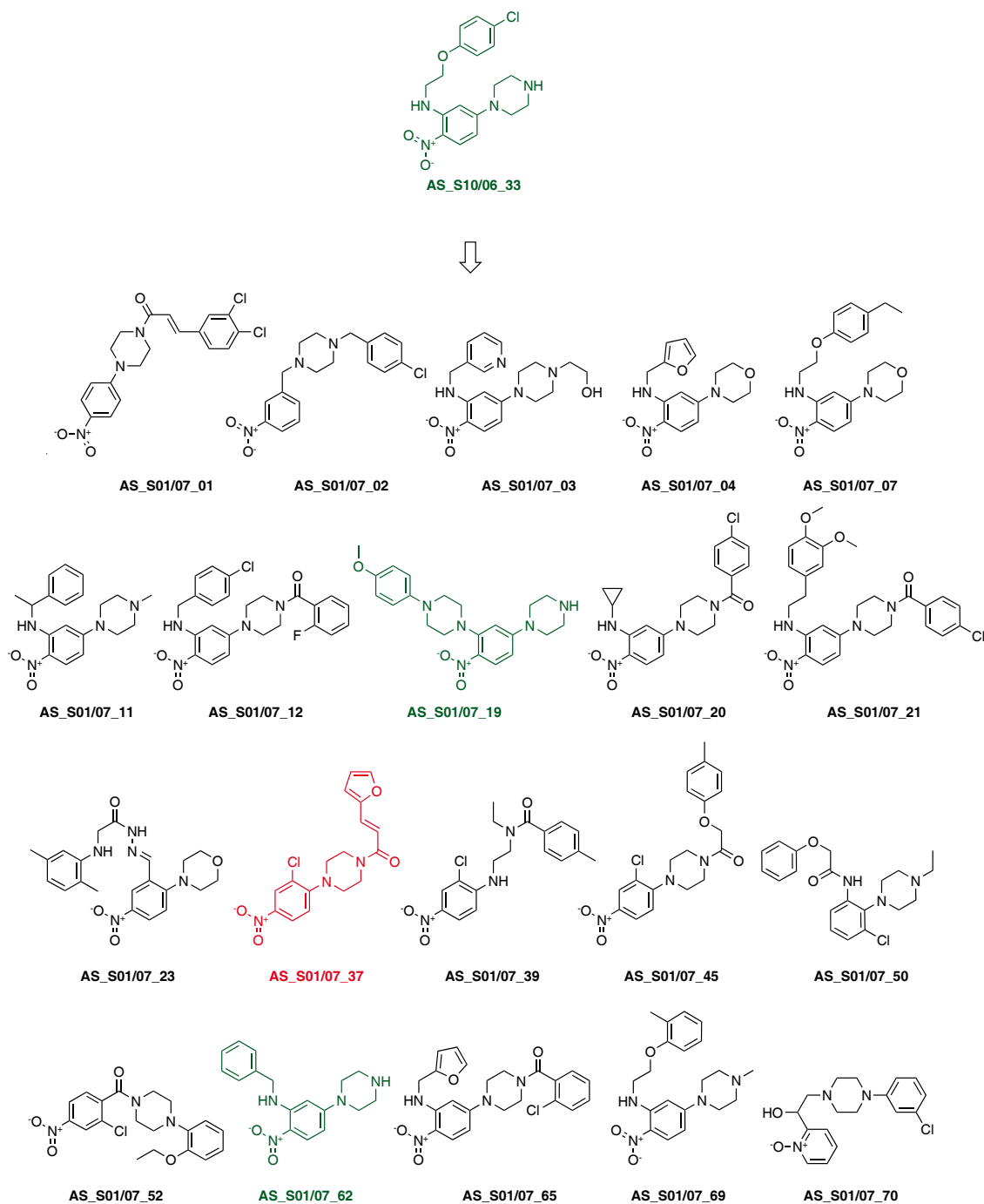
6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



Schema 6-3c: Elter der 1. Generation mit seinen Nachkommen der 2. Generation (Teil 1).

Farblgende: grün: aktiv, schwarz: inaktiv.

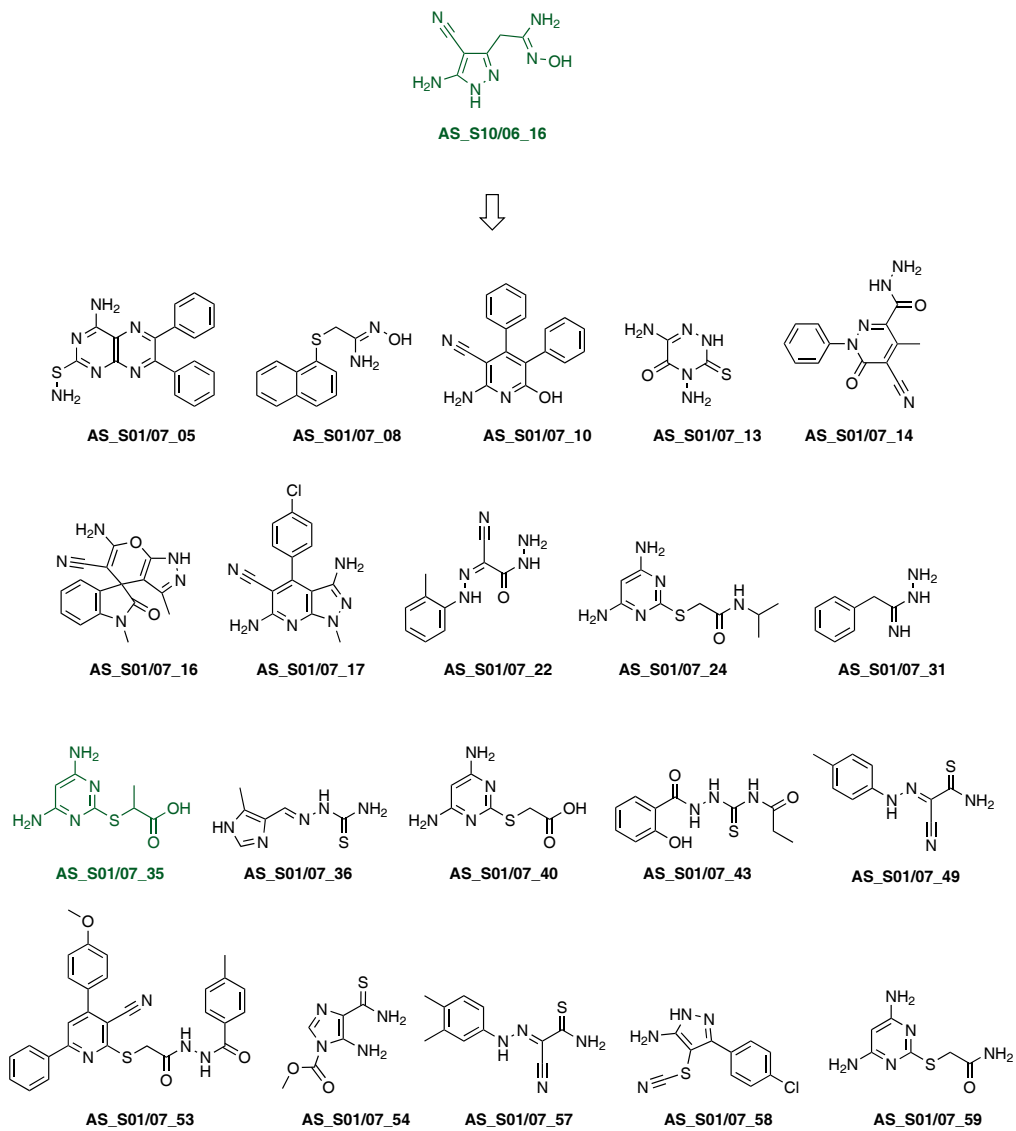
6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



Schema 6-3d: Elter der 1. Generation mit seinen Nachkommen der 2. Generation (Teil 2).

Farblegende: grün: aktiv, schwarz: inaktiv, rot: falsch positiv.

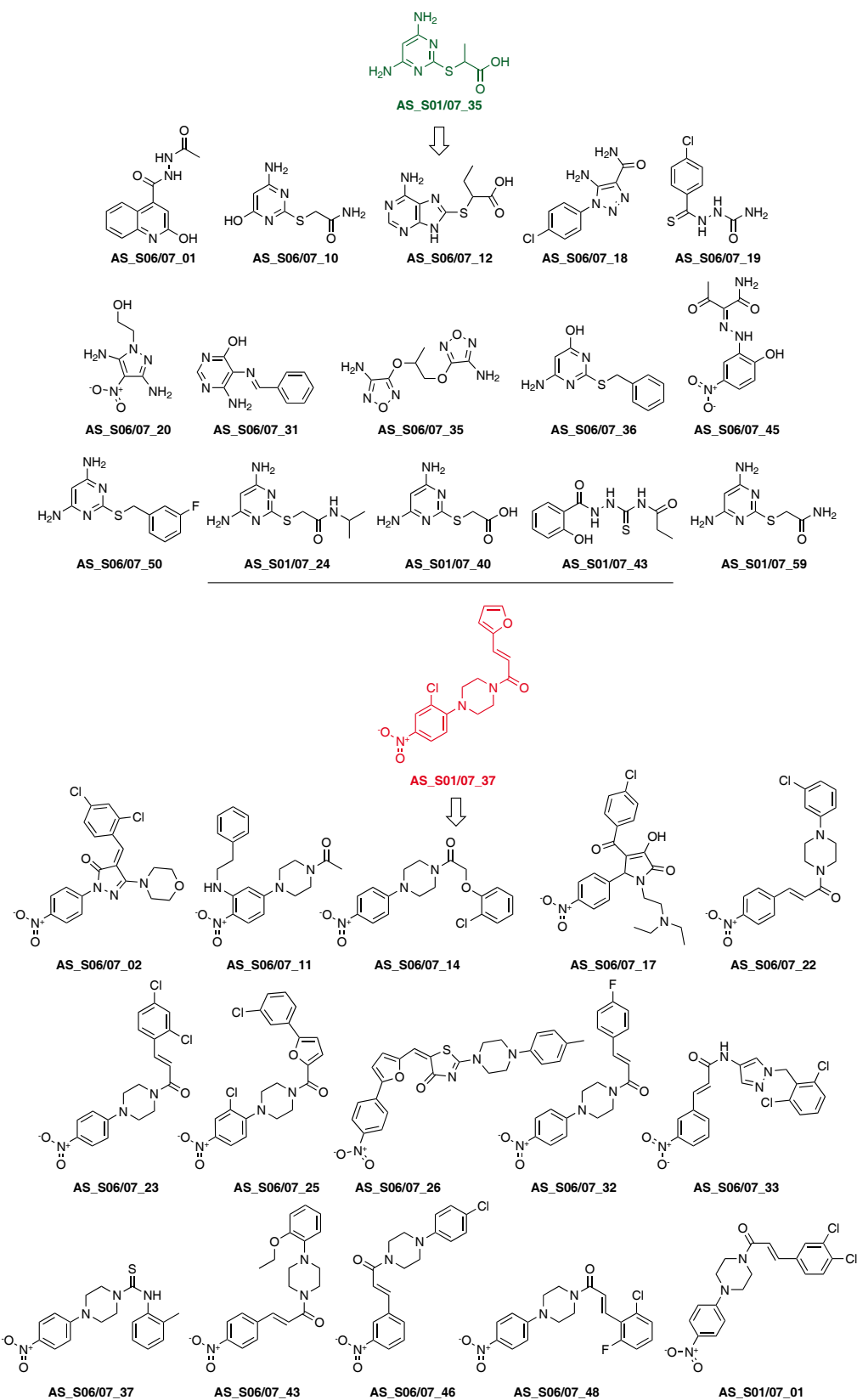
6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



Schema 6-3c: Elter der 1. Generation mit seinen Nachkommen der 2. Generation (Teil 3).

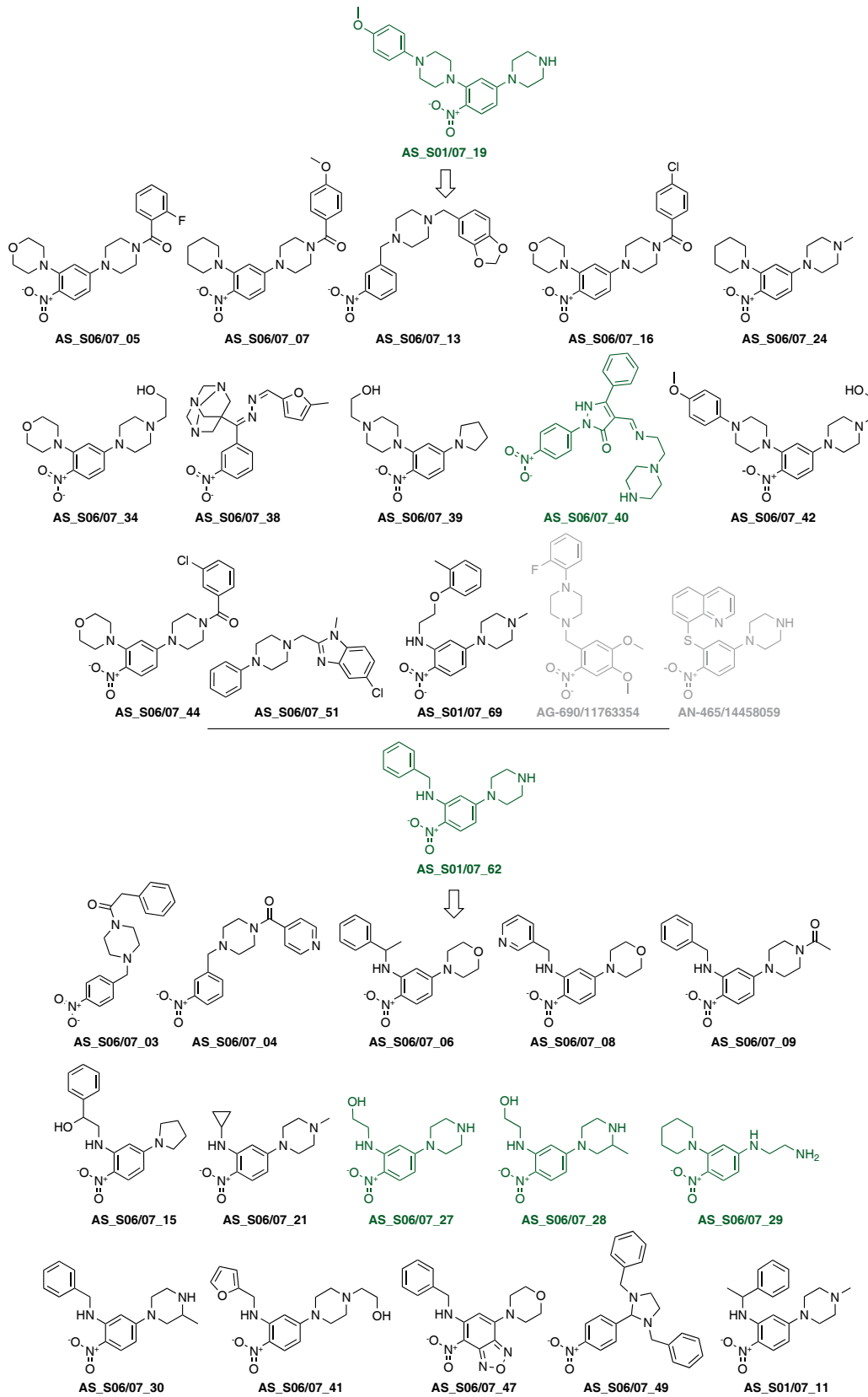
Farblgende: grün: aktiv, schwarz: inaktiv.

6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



Schema 6-3f: Eltern der 2. Generation mit ihren Nachkommen der 3. Generation (Teil 1).
 Farblegende: grün: aktiv, schwarz: inaktiv, rot: falsch positiv.

6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion



Schema 6-3g: Eltern der 2. Generation mit ihren Nachkommen der 3. Generation (Teil 2). Farblegende: grün: aktiv, schwarz: inaktiv, grau: nicht lieferbar.

6.6 Optimierungsverhalten und Konvergenz

Die Nachkommenmoleküle wiesen eindeutige strukturelle Ähnlichkeit zu ihren Elternmolekülen auf (Schema 6-3). Erfolgreiche strukturelle Motive wurden erkannt und variiert. So wiesen z.B. 13 von 15 Nachkommen von **AS_S01/06_19** gleichzeitig ein Nitrobenzol- und Piperazinmotiv auf (Schema 6-3g). Der mediane Eltern-Kind-Abstand betrug rund 14 Einheiten (Manhattanmetrik, willkürliche Einheiten) und ein Nachkomme war im Median der 15. Nachbar seines Elters (Abbildung 6-2). Der mediane Abstand zufälliger Moleküle im gleichen Suchraum betrug zum Vergleich 27 Einheiten und ein Nachkomme war im Median der 83.144ste Nachbar seines Elters (bestimmt durch eine Zufallssuche mit Populationsgröße 60, 5 Generationen, 100 Wiederholungen). Dies zeigt, dass die Evolutionsstrategie in einem lokal begrenzten Nachbarschaftsraum operierte. Das Optimierungsverhalten basierte auf der Ähnlichkeit von Eltern zu ihren Nachkommen.

Nach nur vier Generationen ist teilweise Konvergenz eingetreten. Von anfänglich drei Zweigen des Stammbaumes hat sich ein einziger (**OB108**-Nachkommen, Nitroarene) mit der dritten Generation vollständig durchgesetzt (Abbildung 6-1). Der Median der Schrittweite der ES verringerte sich im Verlauf der Optimierung autoadaptiv vom Startwert 0,5 auf 0,38 (Abbildung 6-2a). Die Eltern-Kind-Distanzen sanken von anfangs 16,5 auf 12,0 und die Ordinalzahl („Der wie viele Nachbar ist ein Nachkomme im Vergleich zu seinem Elter?“) reduzierte sich leicht von 13 auf 11 (Abbildung 6-2b und c). Auffällig ist, dass die mediane Ordinalzahl der 2. Generation mit 34 sehr viel größer ist, als die der 1. Generation mit 13, obwohl in beiden Generationen die mediane Eltern-Kind-Distanz mit 16,53 und 16,35 ungefähr konstant blieb. Dies erklärt sich durch die Zusammenstellung der Initialpopulation aus größtenteils Nicht-Specs-Verbindungen. Zwar ist die Specs-Substanzbibliothek eine diverse Screeningbibliothek, jedoch besteht sie bedingt durch ihre vielen ähnlichen Derivate verschiedener Substanzklassen aus vielen Clustern. Eine Specs-Verbindung aus einem solchen Cluster kann daher im chemischen Raum in einem gegebenen Radius von vielen Molekülen umgeben sein. Eine externe Substanz jedoch, die in diesen Raum projiziert wurde, hat eine größere Wahrscheinlichkeit in einem dünnbesiedelten Bereich zu liegen zu kommen. Sie kann deshalb im gleichen Radius weniger Nachbarn besitzen.

Der mediane Eltern-Kind-Abstand der aktiven Verbindungen betrug 13 Einheiten und ein Nachkomme war im Median der 19. Nachbar seines Elters. Die Schrittweite der Aktiven betrug im Median 0,38 (abgeleitet aus Abbildung 6-1, ohne Falsch-Positive).

Konvergenz war auch bei ausschließlicher Betrachtung der aktiven Verbindungen zu beobachten. Die Eltern-Kind-Distanz sank von 23 über 16 auf 11 Einheiten (von 1. über 2. zur 3. Generation). Die Ordinalzahl in ähnlicher Weise von 63 über 29 auf 12. Die Schrittweite reduzierte sich schließlich geringfügig von 0.5 über 0.38 über 0.30 auf 0.38. Die Anzahl aktiver Verbindungen pro Generation erhöhte sich von 2 (1. Generation) über 3 (2. Generation) bis auf 4 (3. Generation).

Zur besseren Veranschaulichung des Optimierungsverlaufes wurde der hochdimensionale, chemische Raum der Specs-Bibliothek mit Hilfe von selbstorganisierenden Karten (SOM) auf zwei Dimensionen abgebildet.¹²⁷ Der projizierte chemische Raum bestand aus 191.394 Specs-Substanzen und den 10 Verbindungen der Initialpopulation. Es wurden alle 184 MOE-2D-Deskriptoren verwendet. Die Karte besitzt eine Auflösung von 20×20 Neuronen und wurde für 2.000.000 Iterationen mit dem Programm MOLMAP®¹⁴¹ trainiert (Manhattan-Metrik, Radius = 5,0, $\tau = 1,0$). Auf die SOM projiziert wurden jeweils die Moleküle einer Generation (Abbildung 6-3).

Demnach liegen die Verbindungen der Initialpopulation (Abbildung 6-3a) in verschiedenen Bereichen des chemischen Raumes, wobei auf der SOM chemisch ähnliche Moleküle in lokalen Clustern vorliegen. So bilden **OB108**, **rbt550**, **WM5**, **M32** und **S10** – allesamt große Moleküle mit rigiden Strukturen – einen Cluster. **OB080d** und **Y9** bilden einen zweiten sowie **M27**, **IG2** und **GS (red.)** einen weit gefassten dritten Cluster. Mit der ersten Generation (Abbildung 6-3b) ist eine Explosion in der Kartenabdeckung der Molekülkandidaten zu beobachten. Der Optimierungsalgorithmus scheint mit der ersten Generation relativ diverse Molekülvorschläge hervorgebracht zu haben. Dennoch ist deutlich eine Lokalisierung der Nachkommen in Nachbarschaft zu ihren Eltern erkennbar (unter Berücksichtigung der toroidalen Topologie der Karte). Die relative Diversität der Nachkommen bei gleichzeitiger Eltern-Ähnlichkeit ist auch in Schema 6-3 subjektiv nachvollziehbar. Die Verteilung der Moleküle der 2. Generation (Abbildung 6-3c) lässt deutlich die Selektion von nur drei Eltern erkennen: Die Nachkommen sind in drei Clustern, jeweils um ihre Eltern verteilt, wobei die Cluster der **OB080d**-Nachkommen und der **WM5**-Nachkommen teilweise überlappen. Mit der letzten Generation (Abbildung 6-3d) ist eine weitere Konvergenz der Optimierung eingetreten. Zwar sind immer noch versprengte Felder auf der Karte eingezeichnet, jedoch befindet sich in diesen jeweils nur ein Molekül (rote Felder, 11. bis 14. Spalte). Interessant ist ein Sprung von acht Feldern vom Elter **AS_S01/07_62** zu seinen aktiven Nachkommen **AS_S06/07_27**, **AS_S06/07_28**

und **AS_S06/07_29**. Im hochdimensionalen Raum waren diese Verbindungen sehr direkte Nachbarn (Ordinalzahlen: 19, 10 und 7). Es könnte sich hierbei um einen Projektionsfehler der SOM vom hochdimensionalen zum zweidimensionalen Raum handeln. Diese Art von Fehler muss bei der Dimensionsreduzierung grundsätzlich beachtet werden und lässt sich nie völlig ausschließen.

Abbildung 6-2 und Abbildung 6-3 belegen deutlich das den Evolutionsstrategien zugrunde liegende Prinzip der Eltern-Kind-Ähnlichkeit. Der Vergleich der medianen Ordinalzahl von 15 (ES) mit 83.144 (Zufallssuche) zeigt dies auf eindrucksvolle Weise. Das Prinzip der ähnlichkeitsbasierten Optimierung mit iterativer Variation erfolgreicher Entwürfe lässt sich also auch mit Erfolg auf die Suche von RNA-Liganden anwenden.

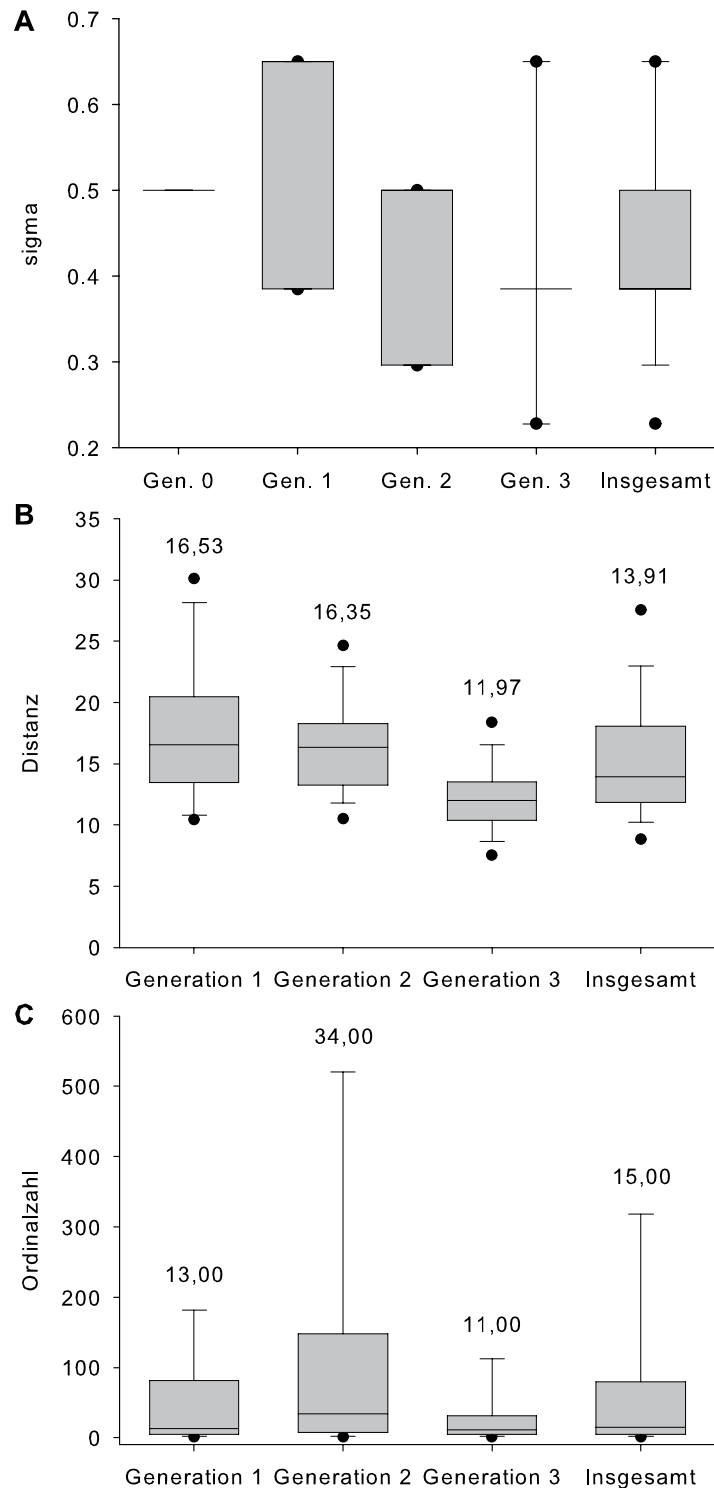


Abbildung 6-2: Boxplots zu den Verteilungen der Schrittweite σ (A), der Eltern-Kind-Distanz (B) und der Eltern-Kind-Ordinalzahl (C). Punkte geben das 5. und 95. Perzentil an, Fehlerbalken das 10. und 90. Perzentil und Boxen das 1. und 3. Quartil und den Median. Zahlenwerte repräsentieren den Median. Distanzen wurden mit der Manhattanmetrik in einer beliebigen Einheit gemessen. Die Ordinalzahl gibt an, der wievielte Nachbar ein Nachkomme im Vergleich zu seinem Elter war. Bei C liegt das 95. Perzentil zum Zwecke einer besseren Übersichtlichkeit außerhalb der Ordinatenkalierung.

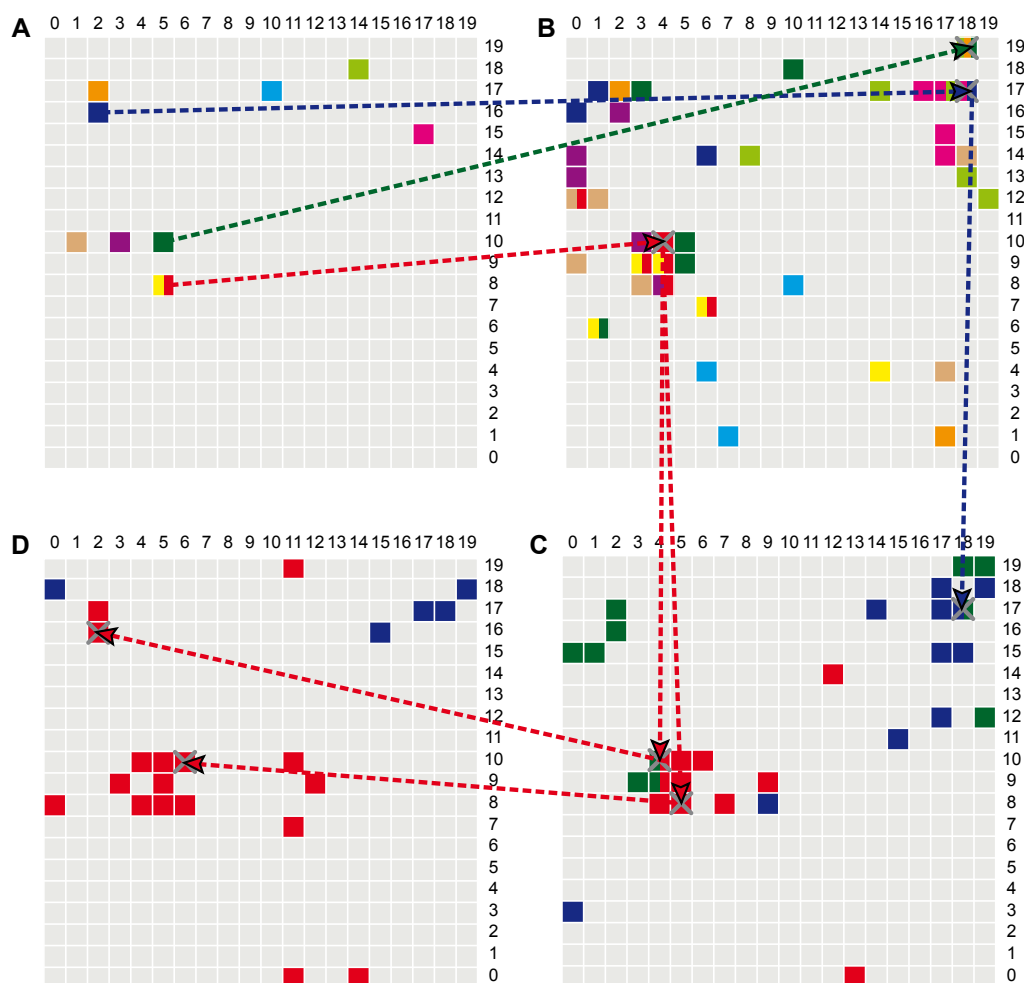


Abbildung 6-3: SOM der Specs-Bibliothek. Neuronen, die den Moleküle der vier Generationen zugeordnet sind, sind farblich markiert. Gleiche Farbe bedeutet gleichen evolutionären Ursprung (gleicher Elter). Graue Kreuze markieren Neuronen mit aktiven Molekülen (inkl falsch positiv). Pfeile verbinden Eltern und ihre Nachkommen (Trajektorien der Optimierung). A: Initialgeneration, B: 1. Generation, C: 2. Generation, D: 3. Generation. Farbkodierung (jeweils der Urelter): rot: **OB108**, gelb: **rbt550**, dunkelgrün: **WM5**, lila: **M32**, hellbraun: **S10**, dunkelblau: **OB080d**, orange: **Y9**, hellblau: **M27**, hellgrün: **IG2** und pink: **GS (red.)**. Aktive Substanzen: rote Linie: **AS_S10/06_33**: B (4/9), **AS_S01/07_19**: C (5/8), **AS_S01/07_37**: C (4/10) falsch positiv, **AS_S01/07_62**: C (4/10), **AS_S06/07_27**: D (2/16), **AS_S06/07_28**: D (2/16), **AS_S06/07_29**: D (2/16), **AS_S06/07_40**: D (5/10); dunkelblaue Linie: **AS_S10/06_16**: B (18/17), **AS_S01/07_35**: C (18/17); dunkelgrüne Linie: **AS_S10/06_73**: B (18/19) falsch positiv. SOM-Training mit MOLMAP^{®141}: 191.404 Screeningsubstanzen, 184 MOE-2D-Deskriptoren, 20×20 Neuronen, 2.000.000 Iterationen, Manhattan-Metrik, Radius = 5,0, $\tau = 1,0$.

6.7 Grundgerüstspringen bei RNA-Liganden

Die meisten heute bekannten RNA-Liganden besitzen einige eindeutige Merkmale, die sie von Proteinbindern unterscheiden.¹⁵² So tragen viele RNA-Liganden eine oder mehrere positive Ladung (z.B. Guanidinium-Derivate), besitzen große interkalierende Ringsysteme (Farbstoffe, Acridin-Derivate, Ethidium-Derivate, Phenothiazine), weisen ein Kohlenhydratgrundgerüst auf (z.B. Aminoglykoside), bestehen aus großen Zyklen (z.B. Makrolide) oder sind peptidischer Natur. Um die gefundenen neun aktiven TAR-RNA-Liganden in die Reihe der RNA-Liganden einordnen zu können, wurde eine selbstorganisierende Karte auf einem Datensatz trainiert, der aus 267 bekannten RNA-Liganden (zusammen getragen aus Projekten der Universität und wissenschaftlichen Publikationen, RNA-Liganden-Bibliothek) und 7576 (nicht doppelten) Verbindungen der COBRA-Sammlung wirkstoffartiger Moleküle bestand. Verbindungen der COBRA wurden pauschal als Protein-Binder eingeordnet. Da die Molekülsammlung im Hinblick auf Proteine erstellt wurde, sollte diese Annahme in den meisten Fällen zutreffend sein. Die SOM wurde mit dem Computerprogramm MOLMAP® erstellt, besitzt eine Ausdehnung von 15×10 Neuronen und wurde 200.000 Zyklen lang trainiert (Manhattan-Metrik, Radius = 5,0, $\tau = 1,0$, 184 autoskalierte MOE-2D-Deskriptoren).¹⁴¹ Die Visualisierung der SOM wurde mit SOMVIS 1.2 von Michael Reutlinger durchgeführt, der das Programm im Rahmen eines Praktikums für die Arbeitsgruppe entwickelt hat.

In Abbildung 6-4a ist farbkodiert der prozentuale Anteil an RNA-Liganden im kartographierten Raum dargestellt. In dieser Art der Darstellung werden nicht jene Bereiche hervorgehoben, die besonders viele RNA-Liganden beinhalten, sondern solche, in denen der relative Anteil an RNA-Liganden im Vergleich zu Proteinbindern besonders hoch ist. In einem deutlich abgegrenzten Cluster (Bereich um Neuron (11/6)) finden sich relativ große Moleküle. Darunter Oligopeptide und Peptidanaloga, Farbstoffkonjugate (Acridin- und Ethidium-Konjugate) und viele Antibiotika (Aminoglykoside und Makrolide). In versprengten Feldern finden sich daneben kationische Liganden (z.B. Guanidinium-Derivate) und Interkalatoren. Einen weiteren eindeutigen Cluster bilden die kationischen Liganden der rbt-Reihe in Neuron (0/8) (vgl. Murchie 2004)¹⁵⁶. In einem großen Bereich in der Mitte der Karte ist der Anteil an RNA-Liganden nur minimal. Dieser Bereich entspricht der Domäne der Proteinbinder. Ziel einer RNA-Wirkstoffentwicklung sollte es sein, sich von der Domäne der heutigen bekannten RNA-Liganden hin zur Domäne der wirkstoffartigen Protein-Binder zu bewegen. Große, peptidische und kationische Verbindungen besitzen schlechte ADME

(Absorption, Distribution, Metabolismus und Exkretion)-Eigenschaften und taugen deshalb nur bedingt als Wirkstoffe. Bei antibiotischen Substanzen kommt zusätzlich das Resistenzrisiko als Problematik hinzu. Die TAR-RNA-Leitstruktur Acetylpromazin **1** ist so ein wirkstoffähnlicher Vertreter, der auf die SOM in den Bereich der Proteinbinder projiziert wurde (Neuron (4/7)). Wie die Projektion der absoluten Zahlen an Proteinbindern (Abbildung 6-4b) und RNA-Liganden (Abbildung 6-4c) verdeutlicht, sind RNA-Liganden durchaus auch in der Domäne der Proteinbinder anzutreffen. Jedoch in verhältnismäßig geringer Anzahl. Die Projektion der neu gefundenen, neun TAR-RNA-Liganden in Abbildung 6-4d gibt einen Hinweis darauf, dass diese Liganden wirkstoffartigen Charakter besitzen. Sie wurden in den linken Teil der SOM projiziert, in einen Bereich, der vermehrt Proteinbinder beinhaltet. Dies ist einerseits eine Konsequenz der Eltern-Moleküle, welche ebenfalls nicht im Hauptcluster der RNA-Liganden lokalisiert sind (Abbildung 6-4d, kleine Karte). Andererseits besitzen die Screening-Substanzen der Specs-Bibliothek bereits potentiell wirkstoffartigen Charakter, oder können als Vorstufen dazu angesehen werden.⁶² Interessant ist, dass sich besonders einige Nachkommen des Elters **OB108** relativ weit von diesem entfernt haben. Befindet sich **OB108** noch in einem Cluster interkalierender RNA-Liganden (Neuron (6/0)), so haben sich dessen Nachkommen alle von diesem Cluster weg, in Richtung der wirkstoffartigen (proteinbindenden) Domäne bewegt.

Die SOM-Analyse von RNA-Liganden *versus* Proteinbinder hat gezeigt, dass eine adaptive Optimierung von Substanzbibliotheken auf Basis einer Evolutionsstrategie in der Lage ist, ein Grundgerüstspringen zu betreiben. Es konnte in neue Bereiche des chemischen Raumes vorgedrungen werden, der durch eine reine Neudekoration der vorhandenen Grundgerüste der Elternmoleküle verborgen geblieben wäre. Vor allem der Sprung vom Phenothiazin **OB108** auf das Nitrophenylpiperazin-Derivat **AS_S10/06_33** kann als klassisches Grundgerüstspringen betrachtet werden. Die wichtige Fähigkeit des Grundgerüstspringens wird maßgeblich vom zu Grunde liegenden Deskriptorraum beeinflusst. Der verwendete MOE-2D-Deskriptorraum abstrahiert von einer rein strukturellen Beschreibung von Molekülen auf ein weiter gefasstes Verständnis derselben. Unter anderem wurden mit den MOE-Deskriptoren auch Pharmakophoreigenschaften berücksichtigt, von denen bekannt ist, dass sie ein Grundgerüstspringen vermitteln können.^{105,122}

6 Evolutionäre Suche nach Inhibitoren der Tat-TAR-Interaktion

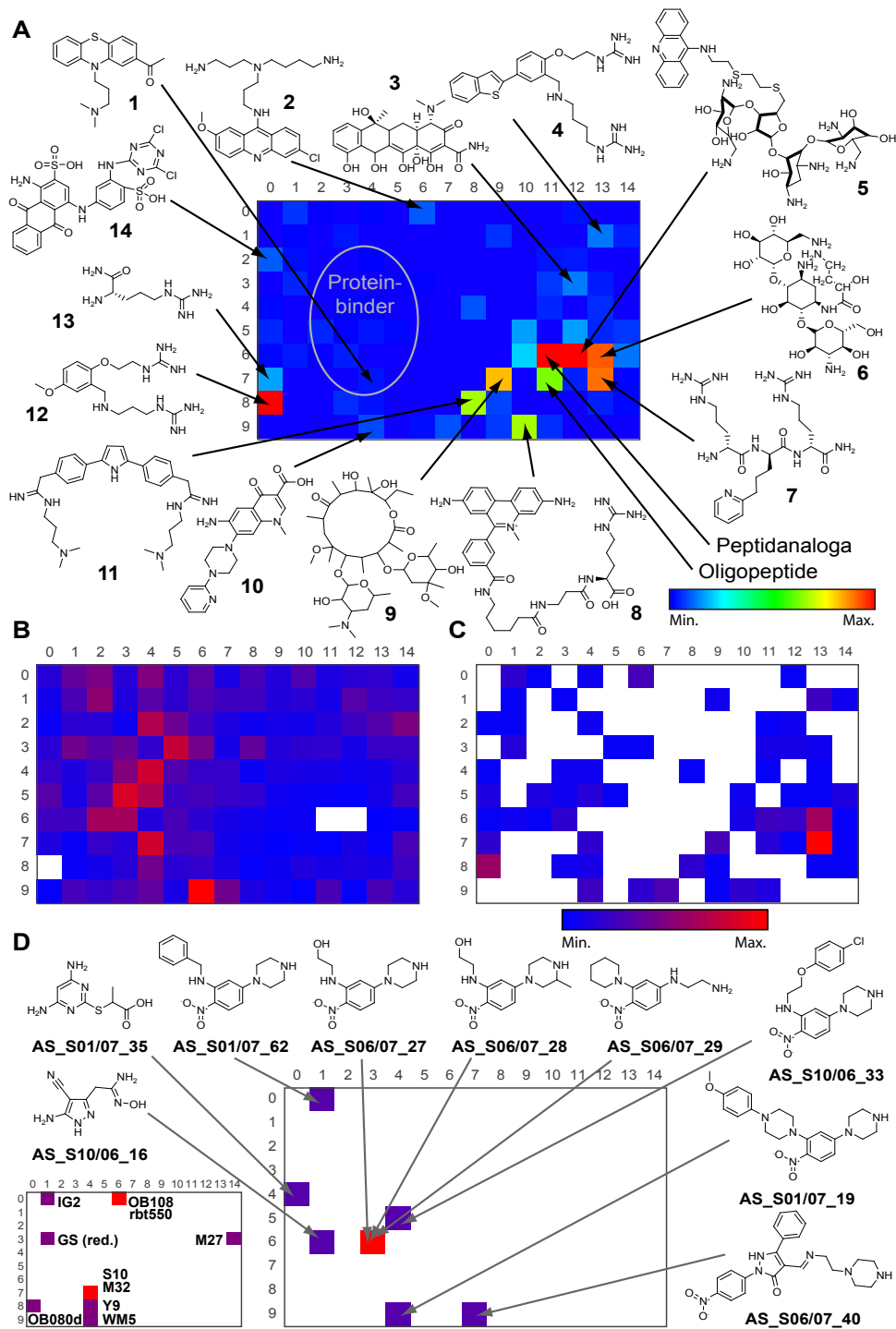


Abbildung 6-4: SOM von RNA-Liganden *vs.* Proteinbindern. Die SOM wurde mit Hilfe bekannter RNA-Liganden und der Moleküle der COBRA trainiert. A) Prozentualer Anteil an RNA-Liganden (dunkelblau: 0%, grün: 50%, rot: 100%). 1: Acetylpromazin,¹⁵⁹ 2: CGP 40336A,¹⁶⁰ 3: Tetracyclin,^{161a} 4: rbt417,¹⁵⁶ 5: Neo-C-Acridin-Aminoglykosid-Konjugat,^{161b} 6: Amikacin,^{161c} 7: D-Arg-D-2Pyridyl-D-Arg,¹⁰⁰ 8: Ethidium-Arginin-Konjugat,^{161d} 9: Clarithromycin,^{161e} 10: WM5,^{157,158} 11: A132,^{161f} 12: rbt202,¹⁵⁶ 13: Argininamid,¹⁰² 14: Reaktives Blau 4.^{161g} B) Projektion von 7576 COBRA-Verbindungen (weiß: 0, blau: 1, rot: 235). C) Projektion von 267 RNA-Liganden (weiß: 0, blau: 1, rot: 33). D) Große Karte: Projektion von 9 neu gefundenen RNA-Liganden (weiß: 0, blau: 1, rot: 3). Kleine Karte: Projektion der Initialpopulation (weiß: 0, lila: 1, rot: 2).

6.8 Löslichkeitsprobleme

Während der experimentellen Aktivitätsbestimmung der Verbindungen sind Löslichkeitsprobleme aufgetreten. Alle Substanzen sind bei Specs als 10 millimolare Lösungen in DMSO bestellt worden. So gut wie alle der 170 bestellten Substanzen waren in dieser Aufbereitung löslich (5 Moleküle waren unlöslich). Im ersten Verdünnungsschritt auf 1 millimolare wässrige Lösungen mit 10% DMSO-Anteil sind jedoch insgesamt über ein Drittel (39%) der Substanzen ausgefallen. Ähnliche Löslichkeitsprobleme wurden von Dr. Manuel Nietert für die gleiche Messmethode berichtet.¹⁰¹ Das Problem der Löslichkeit ist zuerst bei Messungen der Verbindungen der ersten Generation aufgefallen. 22 der 60 Verbindungen (37%) waren nach dem ersten Verdünnungsschritt schlecht löslich (Tabelle 6-3). Die Proben wurden zwar zentrifugiert und der Überstand wurde dennoch vermessen, jedoch war dies ob der Konzentrationsverluste durch Präzipitation keine zufrieden stellende Herangehensweise an das Problem. Eine Ausfallrate von 37% bei einer Populationsgröße von 60 Individuen stellt eine echte Herausforderung für einen Optimierungsalgorithmus dar. Die tatsächliche Populationsgröße betrug deshalb nur 63% des angestrebten Wertes.

Tabelle 6-3: Anzahl gut und schlecht löslicher Moleküle in den Populationen der Generationen. Die Löslichkeit wurde in 10%igen wässrigen Verdünnungen von DMSO bei einer Substanzkonzentration von 1 mM per Augenmaß bestimmt. „Wieder aufgetaut“ bezeichnet die Löslichkeit der Substanzen nachdem sie eine Nacht bei -25°C gelagert wurden. Die Bestimmung der Löslichkeit erfolgte hier mittels Zentrifugation. „Mit MeOH“ bezeichnet die Löslichkeit in Verdünnungen von 10% DMSO und 90% Methanol bei einer Substanzkonzentration von 1 mM. Die Löslichkeit wurde hier per Augenmaß bestimmt.

	gut löslich	schlecht löslich	gesamt
1. Generation	38 (63%)	22 (37%)	60 (100%)
2. Generation	40 (68%)	19 (32%)	59 (100%)
- wieder aufgetaut	35 (59%)	24 (41%)	
- mit MeOH	58 (98%)	1 (2%)	
3. Generation	26 (51%)	25 (49%)	51 (100%)
- wieder aufgetaut	22 (43%)	29 (57%)	
- mit MeOH	48 (94%)	3 (6%)	
Insgesamt	104 (61%)	66 (39%)	170 (100%)

Das Problem wurde von zwei Seiten angegangen. Zum einen wurden, einem Hinweis von Dr. Ute Scheffer folgend, einige der schlecht löslichen Proben mit Erfolg in einer Verdünnung von DMSO und Methanol (10% DMSO in 90% Methanol) bei 1 mM Substanzkonzentration in Lösung gebracht und dann mit der FRET-Methode letztlich bei 1% DMSO und 9% MeOH in einer wässrigen Lösung gemessen. Zum anderen wurde ein einfacher Löslichkeitsfilter auf Grundlage berechneter (vorhergesagter) wässriger Löslichkeitskoeffizienten und Oktanol-Wasser-Verteilungskoeffizienten entwickelt.

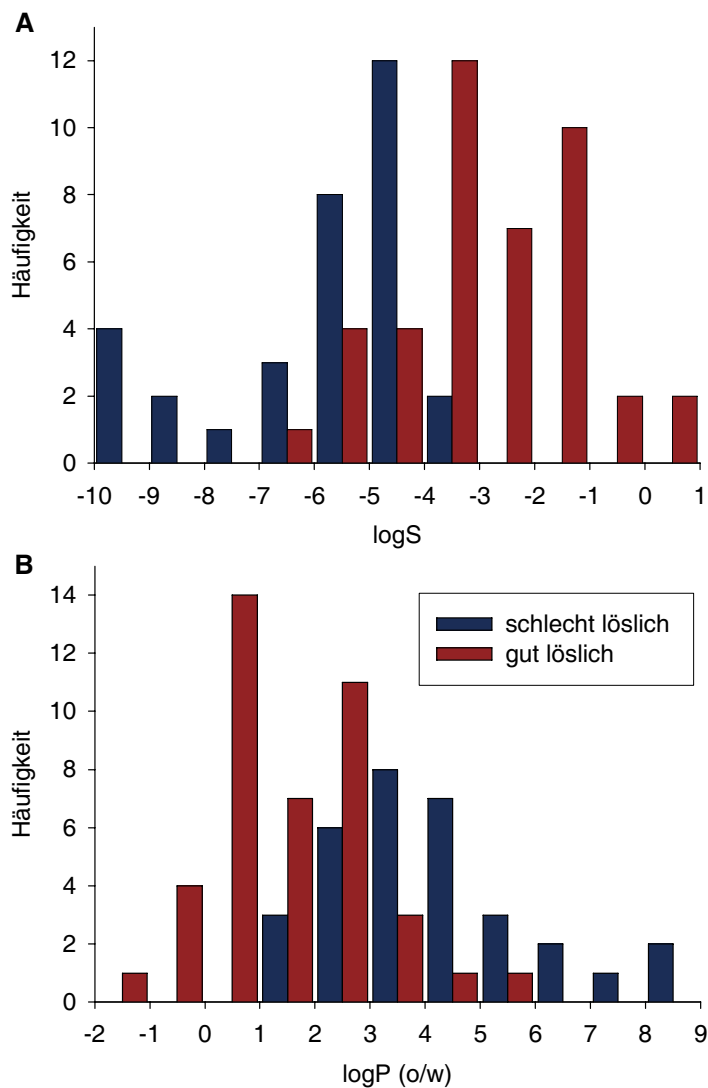


Abbildung 6-5: Histogramme der MOE 2006.08-Deskriptoren „logS“ (A) und „logP(o/w)“ (B). „logS“ ist der dekadische Logarithmus eines berechneten, wässrigen Löslichkeitsquotienten (mol/L). Höhere Werte stehen für bessere wässrige Löslichkeit. „logP(o/w)“ ist der dekadische Logarithmus eines berechneten Oktanol-Wasser-Verteilungskoeffizienten. Niedrigere Werte stehen für bessere wässrige Löslichkeit.

Für den Filter wurden zunächst die Deskriptoren „logS“ (berechneter dekadischer Logarithmus des wässrigen Löslichkeitskoeffizienten) und „logP(o/w)“ (berechneter dekadischer Logarithmus des Oktanol-Wasser-Verteilungskoeffizienten) mit dem Programm MOE (Version 2006.08, Chemical Computing Group, Montreal, Kanada)⁹⁷ für alle 60 Verbindungen der ersten Generation und zusätzlich für 14 weitere Verbindungen aus dem Projekt eines Kollegen berechnet. Die logS-Vorhersage beruht auf einem linearen, atombasierten Modell ($r^2 = 0,90$, ca. 1.200 Moleküle).¹⁶² Die logP(o/w)-Vorhersage beruht ebenfalls auf einem linearen, atombasierten Modell ($r^2 = 0,931$; Std = 0,393; 1827 Moleküle).¹⁶³ Um die beobachtete Löslichkeit im ersten Verdünnungsschritt mit den berechneten Werten ins Verhältnis zu setzen, wurden Histogramme für beide Deskriptoren erstellt (Abbildung 6-5). Die Vorhersagegenauigkeit der berechneten Werte war jedoch schlecht. In den Histogrammen beider Deskriptoren gibt es einen Bereich, indem die Häufigkeitsverteilungen von gut und schlecht löslichen Substanzen überlappen. Im Idealfall läge eine klare Trennung der beiden Klassen vor. Anhand von Abbildung 6-5 wurde beschlossen, alle Verbindungen mit einem berechneten $\log S \leq -7,0$ und (logisches UND) einem berechneten $\log P(o/w) > 6,0$ aus der Specs-Bibliothek heraus zu filtern und zu verwerfen. Die Grenzwerte sind konservativ gewählt, um wenige falsch negative Vorhersagen zu riskieren (Moleküle, die fälschlicherweise als unlöslich gefiltert wurden). Auch die Verknüpfung der Bedingungen über ein logisches UND ist konservativ gewählt. Von den 191.404 Verbindungen der Specs-Bibliothek besaßen 54.086 Moleküle einen $\log S \leq -7,0$ und 29.040 einen $\log P(o/w) > 6,0$. Die Schnittmenge (25.516 Moleküle) wurde herausgefiltert. Die verbliebene löslichkeitsgefilterte Specs-Bibliothek umfasste 165.888 Verbindungen.

Wie in Abbildung 6-6 dargestellt, hat der Löslichkeitsfilter seine angestrebte Wirkung verfehlt. Im Gegenteil: der Anteil schlecht löslicher Verbindungen hat in der zweiten und dritten Generation nochmals zugenommen. Für die Zukunft empfehle ich deshalb einen stringenteren Löslichkeitsfilter für Niedrigdurchsatzscreenings. Es besteht die Gefahr, dass durch Anwendung eines stringenteren Filters potentiell aktive Verbindungen außen vor bleiben. Jedoch verbrauchen unlösliche Verbindungen in der Praxis wertvolle Ressourcen (Material, Zeit, Geld). Es scheint deshalb vorteilhaft, die vorhandenen Ressourcen in die experimentelle Analyse viel versprechender Kandidaten zu stecken, als sie an schlecht löslichen Molekülen zu verschwenden. Die neun gefundenen Aktiven besitzen einen mittleren berechneten logS von $-2,77 \pm 1,36$ (Min. = -4,41) und einen mittleren berechneten logP (o/w) von $1,25 \pm 1,24$ (Max. = 2,86). Der niedrige logS-Wert ist ein Indikator für eine nur mäßige Löslichkeit der meisten gefundenen Aktiven. Dies hat sich auch experimentell bestätigt. Bei einem

zu rigorosen Filter wären diese Verbindungen eventuell ausselektiert und dann auch niemals entdeckt worden. Der Grat des passenden Grenzwertes ist schmal. Eine elegantere Lösung des Problems wäre eine Verbesserung der Vorhersagegenauigkeit der Löslichkeitsdeskriptoren. Die Filterproblematik belegt vor allem einen Mangel an zuverlässigen Vorhersagemodellen von Löslichkeitskenngrößen.

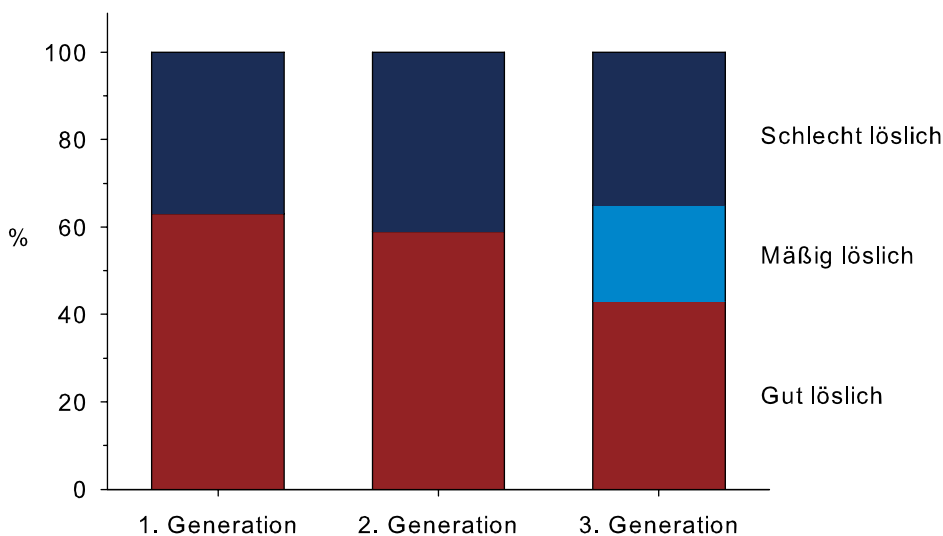


Abbildung 6-6: Prozentuale Verteilung gut, mäßig und schlecht löslicher Verbindungen. Für die ersten beiden Generationen wurde nur eine binäre Klassifizierung in gut und schlecht löslich vorgenommen. Mäßig lösliche Moleküle der 3. Generation unterscheiden sich von schlecht löslichen darin, dass sie nur einen geringen Niederschlag bildeten.

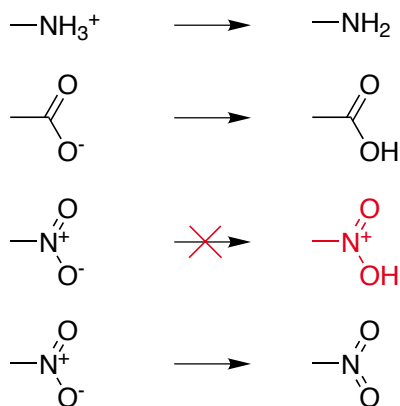
Bewährt hat sich der Zusatz von Methanol zu den Verdünnungen. Mittels der methanolischen Verdünnungen konnten in der zweiten und dritten Generation jeweils 98% und 94% der Verbindungen auch nach dem ersten Verdünnungsschritt in Lösung gebracht werden (Tabelle 6-3). Einige dieser Substanzen sind dann jedoch bei Messbedingungen (1% DMSO, 9% MeOH, pH 7,4) wieder ausgefallen. Die Präzipitation erfolgte hier direkt in den Vertiefungen der Mikrotiterplatten, was die akute Gefahr falsch positiver Detektion mit sich brachte. Besonders der pH-Wert von 7,4 förderte bei Molekülen mit basischen Aminogruppen die Präzipitation. Potentiell aktive Verbindungen mussten deshalb gesondert z.B. durch UV/VIS-Absorptionsspektroskopie auf Löslichkeit überprüft werden. Verbindung **AS_S01/07_37** ist ein Falsch-Positiver, dessen vermeintliche Aktivität ein Artefakt seiner schlechten Löslichkeit war. Die Substanz bildete in den Vertiefungen der Mikrotiterplatten mit der Zeit einen Niederschlag. Bei geeigneten Vorsichtsmaßnahmen haben sich die methanolischen Verdünnungen jedoch bewährt. Die aktiven

Nachkommen von **OB108** mit Nitrophenyl-Grundgerüst hätten sonst nicht oder nur teilweise entdeckt werden können.

In zukünftigen Niedrigdurchsatz-Screeningprojekten sollte die Problematik schlechter wässriger Löslichkeit schon im Vorfeld berücksichtigt werden. Hier bietet sich die Entwicklung, Implementierung oder der Erwerb genauerer Vorhersagemethoden an. Ein Übersichtsartikel von Balakin *et al.* (2006) diskutiert den aktuellen Stand der Forschung auf diesem Gebiet.¹⁶⁴ Llinàs *et al.* (2008) präsentieren in einem aktuellen Artikel eine List von 100 Molekülen, deren Löslichkeit mit einer sehr exakten potentiometrischen Methode bestimmt wurde.¹⁶⁵ Diese Daten könnten eine Basis für genauere Vorhersagemodelle sein. Eine Alternative ist die von Dr. Manuel Nietert vorgeschlagene Einbeziehung experimentell arbeitender Chemiker, die gestützt durch Erfahrung und Intuition bei der Vorhersage wässriger Löslichkeit von Hilfe sein könnten.¹⁰¹

6.9 Fehlerhafte Repräsentation der Nitro-Gruppe

Zum verwendeten Standardprotokoll der Aufbereitung molekularer Bibliotheken gehört die Ladungsneutralisierung der Verbindungen durch Addition oder Subtraktion von Protonen (vgl. Abschnitt 3.3.5). Da die pK_a -Werte saurer und basischer Gruppen in den aller meisten Fällen nicht bekannt sind und zudem durch verschiedene Effekte lokal in der Bindetasche modifiziert werden können, wurde versucht, die Gruppen in der Form ihrer freien Säure oder Base darzustellen (Schema 6-4). Für diese Aufgabe wurde das Computerprogramm CLIFF (Version 1.14; Molecular Networks GmbH, Erlangen)⁹⁶ mit der Option „-uncharge 1“ verwendet.



Schema 6-4: Ladungsneutralisierung verschiedener funktioneller Gruppen. Die rot markierte Gruppe ist chemisch inkorrekt.

Nach abgeschlossener experimenteller Analyse der ersten Generation, jedoch rechtzeitig vor Bestellung der zweiten Generation, wurde ein Programmfehler in dieser Option entdeckt. Nitrogruppen wurden fälschlicherweise wie in Schema 6-4 rot markiert „neutralisiert“ - eine chemisch unsinnige Modifikation. Eine korrekte Darstellung von Nitrogruppen ohne explizite Ladungen ist über einen pentavalenten Stickstoff zu erreichen. Diese Art der Darstellung wird von CLIFF unterstützt und kann über einen separaten Schalter aktiviert werden, schützte jedoch nicht vor dem Softwarefehler.

Zur Bereinigung des Fehlers wurden folgende Schritte unternommen:

1. Das Standardprotokoll zur Molekülaufbereitung wurde geändert. Die Ladungsneutralisierung wurde ab der zweiten Generation vom Computerprogramm CORINA (Version 3.20, Molecular Networks GmbH, Erlangen)¹⁰⁶ übernommen.
2. Die Specs-Bibliothek wurde mit korrekten Nitrogruppen neu aufbereitet und in den Optimierungsprozess eingeschleust.
3. Die Auswirkungen des Fehlers wurden analysiert.

Die erneute (fehlerkorrigierte) Aufbereitung der Specs-Bibliothek machte eine Wiederholung der Deskriptorberechnung nötig. Durch die Neuberechnung der molekularen Deskriptoren wurden die Positionen der Moleküle im chemischen Raum verändert. In der ersten Generation befanden sich sieben Nitroverbindungen: **AS_S10/06_31**, **AS_S10/06_32**, **AS_S10/06_33**, **AS_S10/06_41**, **AS_S10/06_58**, **AS_S10/06_63** und **AS_S10/06_64**. Davon stammen bis auf **AS_S10/06_33** alle vom Elter **M27**, welcher ebenfalls eine Nitroverbindung ist. Für die sieben Nitroverbindungen wurden die Eltern-Kind-Distanzen im alten chemischen Raum (inkorrekte Nitrogruppen, erste Generation) und neuen chemischen Raum (korrekte Nitrogruppen, ab der zweiten Generation) mit einander verglichen. Dazu wurde die Quadratwurzel der mittleren quadratischen Abweichung (*RMSD*, für engl. „Root Mean Square Deviation“) der Eltern-Kind-Distanzen berechnet (Gl. 6-1):

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - y_i)^2} \quad (6-1)$$

mit der Gesamtzahl der betrachteten Verbindungen N , der Eltern-Kind-Distanz der i -ten Substanz im fehlerhaften Datenraum x_i und der Eltern-Kind-Distanz der

i -ten Substanz im Nitrogruppen-korrigierten Datenraum y_i . Der RMSD der Eltern-Kind-Distanzen der sieben Verbindungen betrug 4,5 Einheiten und der RMSD der Ordinalzahl betrug 389,6. Diese Änderungen wurden für gering befunden (vgl. mediane Distanz zwischen beliebigen Datenpunkten: 27 Einheiten; mediane Ordinalzahl: 83.144; Abschnitt 6.6). Die absolute Abweichung der Eltern-Kind-Distanz der aktiven Nitroverbindung **AS_S10/06_33** betrug 1,0 und die der Ordinalzahl 149. Auch diese Abweichung war gering. Die größten Abweichungen wurden in Van-der-Waals-Oberflächendeskriptoren (Partiellladung, saure/basische/polarisierbare/hydrophobe Oberfläche, SlogP-Oberfläche) und Molekülgrößendeskriptoren (Van-der-Waals-Oberfläche, -volumen, Anzahl rotierbarer Bindungen) festgestellt. Auffällig waren große Änderungen des Deskriptorwertes für die polarisierbare Van-der-Waals-Oberfläche (Deskriptor „*rsa_pol*“) für alle Moleküle, auch solchen ohne Nitrogruppe. Die Analysen ergaben, dass dies eine Folge einer undokumentierten Änderung einer MOE-Funktion war. Während die Deskriptoren der ursprünglichen Bibliothek mit MOE 2005.06 berechnet wurden, wurden die Deskriptoren der Nitrogruppen-korrigierten Bibliothek mit MOE 2006.08 berechnet. Die undokumentierte Änderung betraf die Definition von polaren Gruppen, welche bei Version 2005.06 als „*Wasserstoffbrückendonator UND Wasserstoffbrückenakzeptor*“ definiert war und in Version 2006.08 in „*Wasserstoffbrückendonator ODER Wasserstoffbrückenakzeptor*“ geändert wurde (jeweils mit logischem UND und logischem ODER). Da die MOE-Deskriptoren für die Nitrogruppen-korrigierte Bibliothek neu berechnet werden mussten, wurde mit der neuen Definition des „*rsa_pol*“-Deskriptors weiter gearbeitet.

Die erforderliche Neuberechnung der molekularen Deskriptoren machte auch eine erneute Hauptkomponentenanalyse der Deskriptordaten notwendig. Nach dem „Eigenwert ≥ 1 “-Kriterium¹²¹ wurden diesmal 28 Hauptkomponenten ausgewählt. Dies ist eine Hauptkomponente mehr als zuvor. Die Dimensionalität des Suchraumes hat sich damit von 27 auf 28 geändert.

Die Tatsache, dass die Optimierung nach diesem massiven Eingriff dennoch weiter erfolgreich verlaufen ist (vgl. Abschnitt 6.5) spricht für die Robustheit des eingesetzten Optimierungsverfahrens. Selbst der Austausch der kompletten Suchlandschaft durch einen höherdimensionalen Raum wurde von der Evolutionsstrategie toleriert. Dies muss als herausragende Leistung der ES festgehalten werden.

6.10 Suche nach Nicht-Nitroverbindungen

Mit den Molekülvorschlägen für die dritte Generation wurde offensichtlich, dass die evolutionäre Suche auf Nitroverbindungen konvergieren würde. Als Wirkstoffe können Nitroverbindungen problematische Eigenschaften besitzen. Nitroarene sind zwar nicht *per se* toxisch, sehr wohl jedoch ihre zahlreichen Stoffwechselprodukte.¹⁶⁶

In Tierversuchen mit Mäusen und Ratten und teilweise anhand von Unfällen mit Menschen wurden schädigende Auswirkungen von Nitrobenzol auf Milz, Niere, Hoden, Leber und zentrales Nervensystem beobachtet.¹⁶⁷ Der überwiegende, systemische, toxische Effekt von Nitrobenzol im Menschen ist die Methämoglobinämie (Symptom: Zyanose). Hierbei wird Fe²⁺-Hämoglobin in Erythrozyten zu Fe³⁺-Methämoglobin oxidiert. Daneben wurde in Experimenten mit Tieren eine karzinogene Wirkung von Nitrobenzol festgestellt.¹⁶⁷ Die Substanz induzierte ein breites Spektrum benignen und malignen Tumore in verschiedenen Organen (Lunge, Thymus, Milchdrüse, Leber, Niere). In einem Versuch wurden 10 von 16 Nitroarene (62,5%) positiv auf Karzinogenität geprüft (in Mäusen/Ratten).¹⁶⁸ Beim Metabolismus von Nitrobenzol entstehen im Körper freie reaktive Sauerstoff- und Nitroxidradikale, die für die karzinogenen Eigenschaften verantwortlich sein könnten.^{167,169,170} Eine Abhandlung zum Metabolismus von Nitrobenzol ist im Anhang zu finden (Abschnitt 8.5).

Trotz dieser Befunde zu Nitroarenen müssen Nitroverbindungen nicht zwingend schädliche Eigenschaften besitzen. Eine Struktur-Wirkungs-Beziehungsanalyse ergab, dass mehrere starke elektronenziehende Gruppen die karzinogene Wirkung von Nitroarenen in Tierversuchen unterdrücken konnten.¹⁶⁸ Die bloße Anwesenheit einer Nitrogruppe muss deshalb nicht zwingend eine karzinogene Wirkung bedeuten. Kommerziell eingesetzte Nitro-Wirkstoffe sind z.B. Nitrazepam (ein Hypnotikum), Nifedipin (Calciumkanalblocker, zeigt in hohen Dosen teratogene Effekte), Metronidazol (Antibiotikum aus der Gruppe der Nitroimidazole, bewirkt in Bakterien DNA-Defekte) und Chloramphenicol (Antibiotikum, bindet an das Peptidyltransferasezentrum der Ribosomen, Naturstoff).¹⁷¹ Zudem können Nitroverbindungen in der präklinischen Phase der Wirkstoffentwicklung ein wertvolles pharmakologisches Testsystem darstellen.¹⁷² Eine Analyse der COBRA-Sammlung wirkstoffartiger Substanzen (Version 6.1, 7576 nicht doppelte Verbindungen)⁹⁵ ergab, dass 202 Verbindungen (2,7%) eine Nitrogruppe aufweisen. Diese Nitroverbindungen waren gegen verschiedene biologische Ziele gerichtet (Ionenkanäle: 24%, GPCRs: 19%, nukleäre Rezeptoren: 5%, Kinasen: 6%, Enzyme: 40% und andere: 6%).

Wegen der potentiell schädlichen Eigenschaften von Nitroverbindungen wurde nach der Generierung der Molekülvorschläge für die dritte Generation entschlossen,

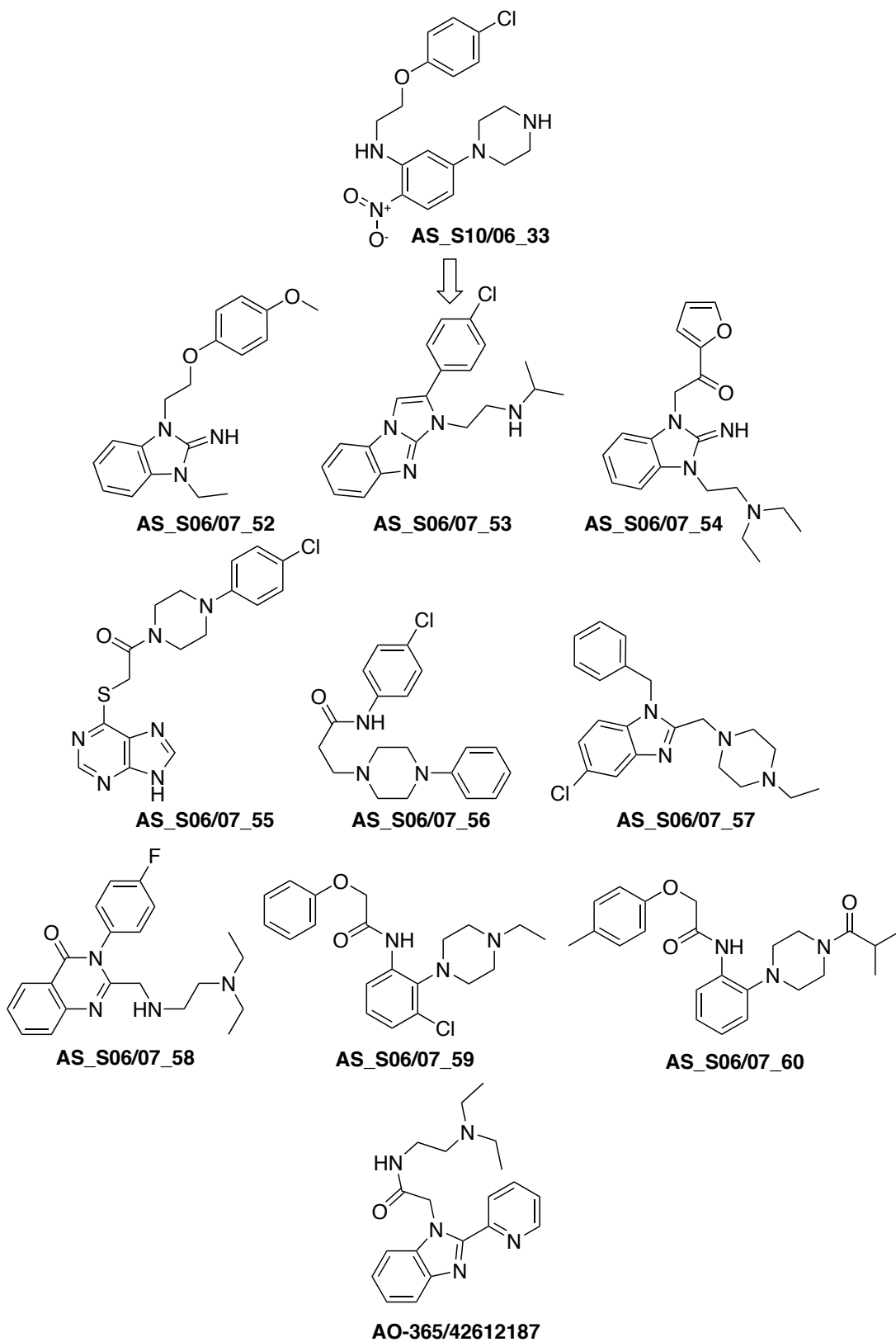
auf die Durchführung der vierten und letzten geplanten Generation zu verzichten. Stattdessen wurden zur aktivsten Nitroverbindung **AS_S10/06_33** zehn ähnliche Nicht-Nitroverbindungen aus dem Specs-Katalog ausgewählt. Dazu wurde eine chemische Ähnlichkeitssuche zur Verbindung **AS_S10/06_33** durchgeführt, indem der Deskriptorvektor von **AS_S10/06_33** mittels Manhattan-Metrik mit allen Deskriptorvektoren der Specs-Bibliothek verglichen wurde. Es wurde der gleiche Deskriptorraum wie in der Optimierungsstudie verwendet (28 Hauptkomponenten der 184 MOE 2D-Deskriptoren). Die Ähnlichkeitssuche wurde mit dem Programm *rankIt* (Version 0.4.7 vom 13. Juli 2004) durchgeführt.¹⁷³ Der Specs-Katalog wurde nach zunehmender Distanz zu **AS_S10/06_33** sortiert, alle Nitroverbindungen wurden entfernt und aus den 120 ähnlichsten Verbindungen wurden zehn Substanzen zur Bestellung und biologischen Testung ausgewählt (Tabelle 6-4 und Schema 6-5). Neun der zehn bestellten Substanzen waren lieferbar und wurden mit der FRET-Methode auf Aktivität gegen TAR-RNA geprüft. Eine Einzelpunktmessung (dreifache Wiederholung) bei 100 μM Substanzkonzentration ergab zunächst, dass sechs der neun Verbindungen eine Inhibition von mehr als 8% bewirkten (Tabelle 6-5). Zur weiteren Abschätzung der Substanzaktivitäten wurden Titrationsmessungen als Einzelmessungen, akkumulativ in je einer Mikrotiterplattenvertiefung vorgenommen. Diese Methode ist ungenau und erlaubt lediglich die Bestimmung eines Aktivitätstrends. Das Ergebnis dieser Abschätzung zeigte, dass nun vier der sechs titrierten Substanzen eine Fluoreszenzerhöhung bewirkten, eine Substanz keine Änderung des Fluoreszenzwertes bewirkte und nur eine Substanz den Fluoreszenzwert nach wie vor absenkte (**AS_S06/07_53**). Verbindung **AS_S06/07_53** bewirkte eine Absenkung des Wertes auf 72% bei 500 μM Substanzkonzentration. Diese Antwort war jedoch schwächer als in der Einzelpunktmessung, bei der eine Absenkung auf 74% \pm 6% bei 100 μM Substanzkonzentration erreicht wurde. Die Messergebnisse deuten auf eine potentiell schwache Aktivität von **AS_S06/07_53** im niedrigmillimolaren Bereich hin. Ein IC_{50} -Wert konnte aufgrund von ungenügender Löslichkeit nicht bestimmt werden. Das abweichende Ergebnis der Titrationsmessungen im Vergleich zu den Einzelpunktmessungen könnte auf eine Degenerierung der Substanzen hindeuten. Möglich ist, dass diese Substanzen bei den gewählten Lagerungsbedingungen (-25 °C in DMSO) und nach wiederholten Gefrier-/Auftau-Zyklen nicht stabil waren. Zusammenfassend führte die chemische Ähnlichkeitssuche von Nicht-Nitroverbindungen, die sich von **AS_S10/06_33** ableiten, in eine Sackgasse. Als einziger potentiell schwachaktiver Kandidat konnte Verbindung **AS_S06/07_53** identifiziert werden. Es zeigte sich, dass chemische Ähnlichkeit allein keine hinreichende Bedingung für biologische Aktivität ist.¹⁷⁴

Tabelle 6-4: Zehn zu **AS_S10/06_33** ähnliche Nicht-Nitroverbindungen. Die ‚Ordinalzahl‘ gibt an, der wievielte Nachbar ein Molekül zu **AS_S10/06_33** ist. Bei ‚Ord.zahl (ohne Nitroverb.)‘ wurden alle Nitroverbindungen aus der Liste der ähnlichen Substanzen entfernt. Die ‚Distanz‘ wurde mit der Manhattan-Metrik zu **AS_S10/06_33** in einer beliebigen Einheit gemessen.

Name	Specs-Nr.	Ord.zahl	Ord.zahl ohne Nitroverb.	Distanz	Aktivität
AS_S06/07_52	AJ-292/42034521	124	7	14,51	inaktiv
AS_S06/07_53	AE-848/12530561	172	14	15,09	potentiell schwachaktiv
AS_S06/07_54	AG-690/13706286	348	78	16,28	inaktiv
AS_S06/07_55	AQ-088/42013598	278	50	15,88	inaktiv
AS_S06/07_56	AG-205/07906042	188	21	15,24	inaktiv
AS_S06/07_57	AQ-360/43172889	219	28	15,45	inaktiv
AS_S06/07_58	AP-064/42049163	138	9	14,66	inaktiv
AS_S06/07_59	AN-465/42889773	45	1	13,27	inaktiv
AS_S06/07_60	AN-465/42769193	312	62	16,07	inaktiv
nicht lieferbar	AO-365/42612187	179	18	15,19	

Tabelle 6-5: Ergebnisse der Aktivitätsmessungen der Nicht-Nitroverbindungen. Einzelpunktmessungen erfolgten in dreifacher Wiederholung bei 100 μM Substanzkonzentration in wässriger Lösung mit 1% DMSO (gut lösliche Substanzen) oder 1% DMSO und 9% MeOH (schlecht lösliche Substanzen). Titrations erfolgten als Einzelmessungen in je einer Mikrotiterplattenvertiefung akkumulativ. Bei hohen DMSO-Konzentrationen ($\geq 2\%$) wurde für die Messwerte ein Korrekturfaktor mit einbezogen.

Name	Effekt (100 μM in 1% DMSO)	Effekt (100 μM in 1% DMSO u. 9% MeOH)	Titration (Effekt bei verschiedenen Endpunkten)
AS_S06/07_52	88% \pm 2%		97% (500 μM in 5% DMSO)
AS_S06/07_53	74% \pm 6%		72% (500 μM in 0,5% DMSO u. 2% MeOH)
AS_S06/07_54	76% \pm 5%		107% (500 μM in 5% DMSO)
AS_S06/07_55		97% \pm 4%	
AS_S06/07_56		97% \pm 2%	
AS_S06/07_57	91% \pm 5%		238% (200 μM in 2% DMSO)
AS_S06/07_58	86% \pm 2%		138% (300 μM in 3% DMSO)
AS_S06/07_59	77% \pm 2%		113% (300 μM in 3% DMSO)
AS_S06/07_60		94% \pm 7%	



Schema 6-5: Strukturformeln der gemessenen Nicht-Nitroverbindungen. Die Verbindungen wurden mittels einer Ähnlichkeitssuche zu **AS_S10/06_33** und anschließendem „Rosinenpicken“ ausgewählt.

Bedingt durch die Tatsache, dass die evolutionäre Optimierung so eindeutig auf Nitroverbindungen konvergierte, bleibt zu klären, warum in der ersten Generation überhaupt eine Nitroverbindungen als Nachkomme einer Nicht-Nitroverbindung vorgeschlagen wurde. Unter den insgesamt 65 Nachkommen (verschiedenen Grades) der Nitroverbindung **AS_S10/06_33** sind gerade einmal 3 (4,6%) Nicht-Nitroverbindungen. Eine SOM-Analyse mit MOLMAP®¹⁴¹ ergab, dass sich Verbindung **AS_S10/06_33** in einem Bereich mit überdurchschnittlich vielen Nitroverbindungen befindet (Abbildung 6-7, Neuron (4/10), 136 Nitroverbindungen). Sein Elternmolekül **OB108** befindet sich jedoch in einem Bereich geringer Dichte an Nitroverbindungen (Neuron (5/8), 8 Nitroverbindungen). Des Weiteren weisen von den 165.888 Verbindungen des löslichkeitsgefilterten Suchraumes lediglich 12.225 (7%) das Nitrobenzol-Substrukturmotiv auf und nur 14.611 (9%) besitzen eine Nitrogruppe.

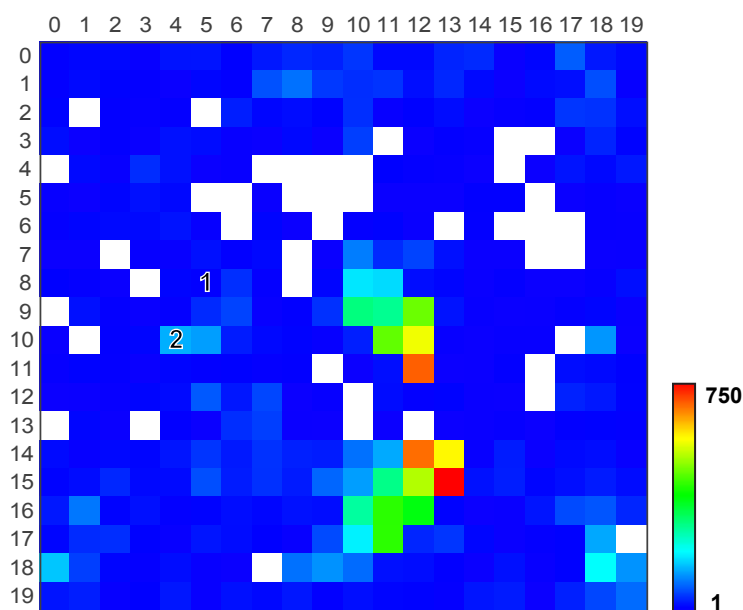


Abbildung 6-7: SOM-Projektion der Nitroverbindungen des Specs-Kataloges. Es wurden die selbe SOM wie in Abbildung 6-3 verwendet. Die Projektion erfolgte mit SOMVIS 1.2. Die Farbskala spiegelt die Anzahl der Nitroverbindungen wieder. Weiße Felder beinhalten keine Nitroverbindungen. **1: OB108** (5/8), 8 Nitroverbindungen im rezeptiven Feld; **2: AS_S10/06_33** (4/10), 136 Nitroverbindungen im rezeptiven Feld.

Dennoch befand sich in der ersten Generation unter den Nachkommen von **OB108** eine Nitroverbindung, die zudem noch als aktiv gemessen wurde. Dies scheint ein zufälliges und glückliches Ereignis gewesen zu sein. Mit einer Distanz von 19,7 Einheiten (123ste Nachbar) ist **AS_S10/06_33** weiter von seinem Elter entfernt, als

die meisten anderen aktiven Substanzen von ihren jeweiligen Eltern. Diese längere Distanz vermittelte einen zufälligen Sprung in einen Bereich mit höherer Dichte an Nitroverbindungen. Unter den 500 nächsten Nachbarn von **OB108** befanden sich gerade einmal 17 (3%) Nitroverbindungen. Ein weiteres Indiz für eine Erklärung durch den „Zufall“ liefert die Statistik. **AS_S10/06_33** war einer von sechs Nachkommen von **OB108**. Die Wahrscheinlichkeit bei sechs zufälligen Zügen (ohne Zurücklegen) mindestens eine Nitroverbindung auszuwählen ist mit rund 43% relativ hoch (berechnet mit Hilfe einer hypergeometrischen Wahrscheinlichkeitsdichtefunktion in Matlab, The MathWorks, Version 7.04.365 R14 SP2, mit installiertem Statistikwerkzeugkoffer, Befehl: „1 - hygecdf(0, 165888, 14611, 6)“).¹⁴⁶

Es war also zunächst Zufall, dass eine Nitroverbindungen wie **AS_S10/06_33** ausgewählt wurde. Jedoch hat in den folgenden Schritten die Logik der Evolutionsstrategie bewirkt, dass dieser Zufall zu einem evolutionären Vorteil ausgenutzt wurde. Es ist diese Charakteristik des gerichteten Zufalls, die die stochastischen Optimierungsverfahren so robust, und deshalb so attraktiv machen.

6.11 Struktur-Wirkungs-Beziehungen der Nachkommen von OB108

In den folgenden Abschnitten wurden Struktur-Wirkungs-Beziehungen (*SAR*, für engl. „Structure-Activity Relationship“) mit chemieinformatischen Methoden für die gefundenen aktiven Verbindungen abgeleitet.

Die Gegenüberstellung von sechs der sieben aktiven Nachkommen von **OB108** legt eine große Übereinstimmung der chemischen Struktur offen (Abbildung 6-8). Die größte gemeinsame Substruktur umfasst 16 Atome (Nicht-Wasserstoffatome). Sie besteht aus einem zentralen Nitrophenylmotiv, von dem in Ortho- und Parastellung vier Atome lange aliphatische Ketten ausgehen, welche mit dem zentralen Benzolring über eine Aminogruppe verknüpft sind. Die eine Kette der Substruktur weist dabei terminal als entscheidendes Merkmal eine primäre oder sekundäre Aminogruppe auf, welche unter Versuchsbedingungen wahrscheinlich protoniert, also kationisch vorlag. Die Übereinstimmung dieser klar definierten Struktur legt den Schluss nahe, dass hier ein Aktivitätsgrundmuster für Liganden der TAR-RNA gefunden wurde. Bedingt durch die Konvergenz des Suchalgorithmus und die Auswahl vorhandener Verbindungen im Specs-Katalog wurden zahlreiche analoge Verbindungen gefunden, die jedoch als inaktiv gemessen wurden. Diese Analoga weisen als entscheidenden Unterschied eine

Abweichung von der größten gemeinsamen Substruktur im Bereich der terminalen, kationischen Aminogruppe auf. Sobald diese terminale, primäre oder sekundäre Aminogruppe durch eine andere Gruppe (z.B. durch eine tertiäre Aminogruppe, eine Ethergruppe oder eine aliphatische Gruppe) ersetzt wird, verlieren die Verbindungen ihre Aktivität (Abbildung 6-9). Es wird deshalb postuliert, dass diese Gruppe eine für die erfolgreiche Bindung essentielle Interaktion mit TAR-RNA eingeht.

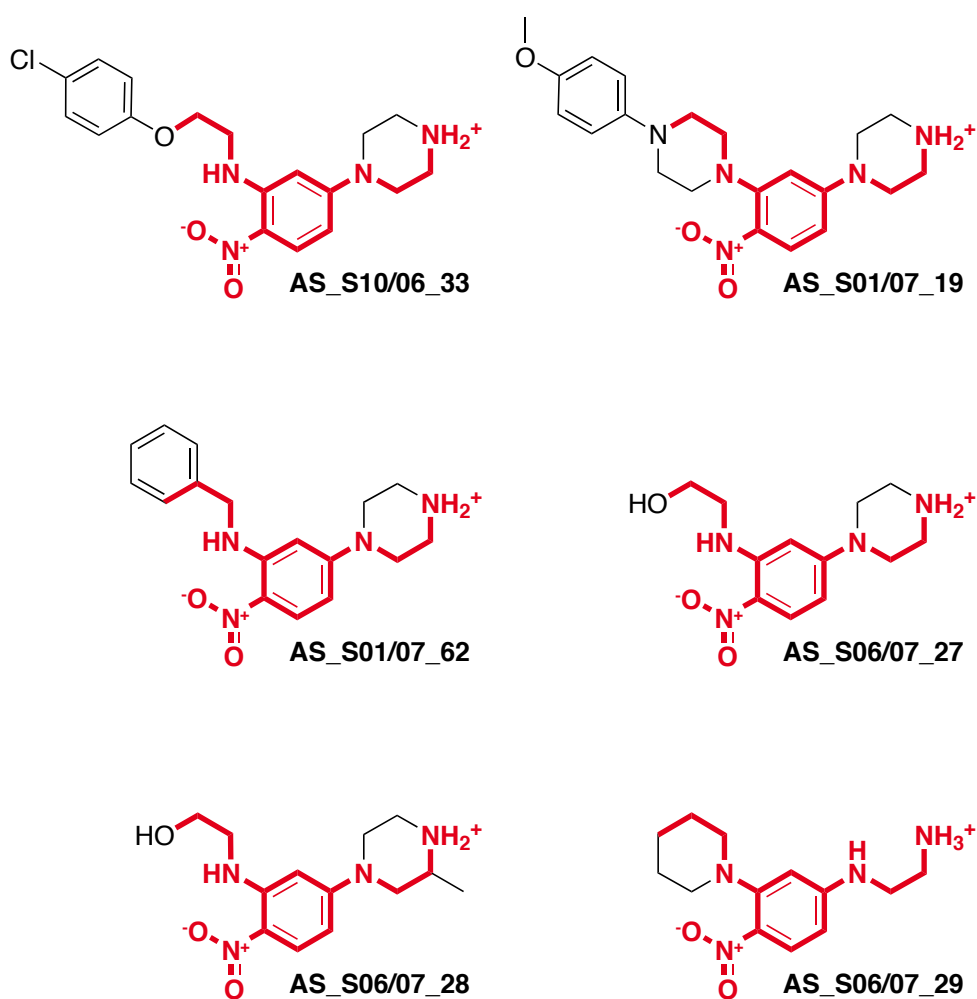


Abbildung 6-8: Gegenüberstellung sechs aktiver Nachkommen von **OB108**. Die größte gemeinsame Substruktur der Verbindungen ist rot einfärbt.

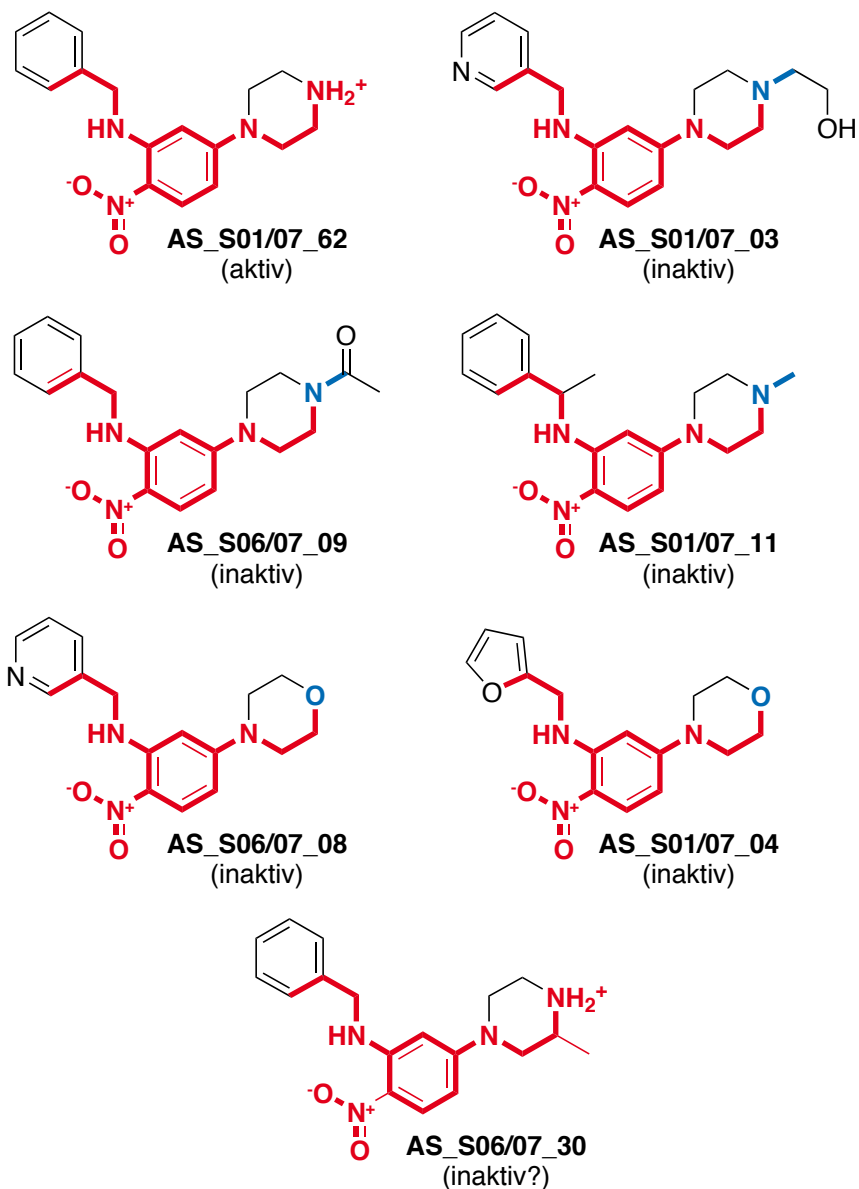


Abbildung 6-9a: Gegenüberstellung der aktiven Verbindung **AS_S01/07_62** und sechs inaktiver Kandidaten. Die größte gemeinsame Substruktur der sechs Aktiven aus Abbildung 6-8 ist rot eingefärbt. Abweichungen davon sind blau markiert. Verbindung **AS_S06/07_30** wurde als inaktiv gemessen, obwohl sie die aktive Substruktur voll erfüllt. Es könnte sich hierbei um eine falsch negative Verbindung handeln.

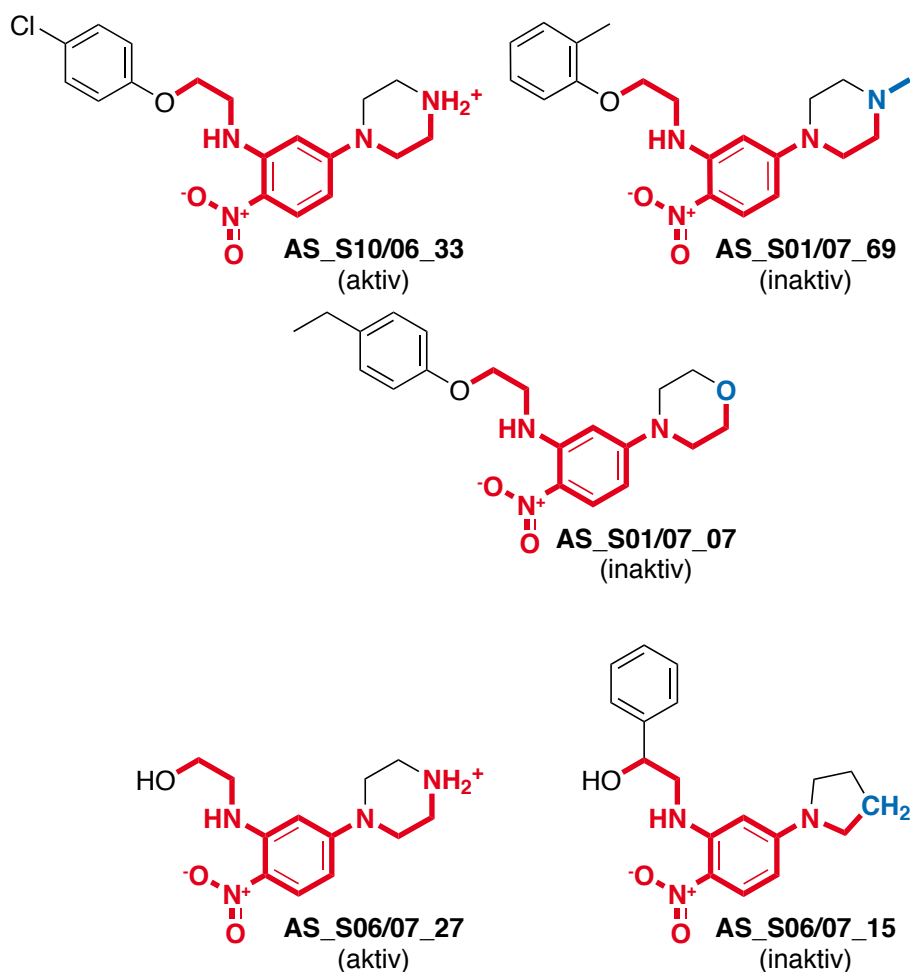


Abbildung 6-9b: Gegenüberstellung der aktiven Verbindungen **AS_S10/06_33** (oben) und **AS_S06/07_27** (unten) mit drei inaktiven Kandidaten. Die größte gemeinsame Substruktur der sechs Aktiven aus Abbildung 6-8 ist rot eingefärbt. Abweichungen davon sind blau markiert.

Aminogruppen sind zum einen als Protonendonoren dazu in der Lage, Wasserstoffbrücken auszubilden. Zum anderen liegen Aminogruppen bei physiologischem pH zumeist protoniert, also kationisch vor und können so elektrostatisch wechselwirken. Weder Ether- noch aliphatische Gruppen besitzen diese pharmakophoren Eigenschaften. Der Verlust der Aktivität bei Derivaten mit Ethergruppe oder aliphatischer Gruppe an Stelle einer Aminogruppe lässt sich leicht mit dem Verlust dieser Eigenschaften erklären. Ein sterischer Effekt kann ausgeschlossen werden, da eine Ethergruppe nicht mehr Raum beansprucht, als eine Aminogruppe. Bei tertiären Aminogruppen hängt die Ausbildung beider pharmakophoren Eigenschaften vom pH-Wert ab. Bedingt durch sterische Effekte reagieren tertiäre Amine in wässrigen Lösungsmitteln weniger stark basisch als primäre oder sekundäre Amine, obwohl deren

höherer Substituierungsgrad durch elektronenspendende Alkylgruppen die Basizität eigentlich erhöhen sollte (Abbildung 6-10).¹⁷⁵ Der berechnete pK_a für die aktive Verbindung **AS_S10/06_33** ist 8,8, wohingegen der pK_a für die nahverwandte, aber inaktive Verbindung **AS_S01/07_69** 7,6 beträgt (berechnet mit der Onlineversion von ALOGPS 2.1; <http://www.vcclab.org/>, zugegriffen am 29.05.2008).^{176,177} Bei einem pH von 7,4, wie unter Versuchsbedingungen, liegen nach der Henderson-Hasselbalch-Gleichung für Säure-Base-Gleichgewichte rund 39% der Verbindung **AS_S01/07_69** als unprotonierte Spezies vor. Lokale Einflüsse in der Bindetasche könnten den pK_a der tertiären Aminogruppe noch weiter in den sauren Bereich verschieben und damit eine Protonierung noch unwahrscheinlicher machen. Eine nicht protonierte, tertiäre Aminogruppe ist weder kationisch, noch kann sie Wasserstoffbrücken ausbilden. Diese fehlenden, pH-abhängigen Pharmakophoreigenschaften könnten den Verlust der Aktivität bei Derivaten mit tertiärer Aminogruppe erklären. Bedingt durch ihre Pharmakophoreigenschaften wird die essentielle primäre oder sekundäre Aminogruppe wahrscheinlich als Donor eine Wasserstoffbrücke ausbilden oder als Kation eine elektrostatische Wechselwirkung eingehen. Eine Überlagerung beider Wechselwirkungstypen wäre auch denkbar.

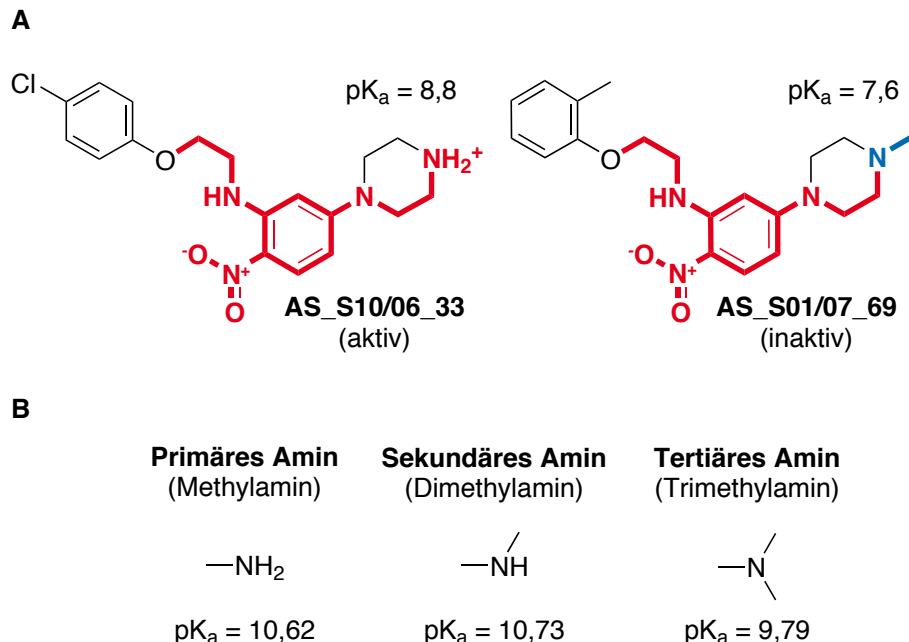


Abbildung 6-10: Vergleich der Basizität verschieden stark substituierter Amine. A) Berechnete pK_a -Werte für zwei engverwandte Moleküle unterschiedlicher Aktivität. Der pK_a wurde mit der Onlineversion von ALOGPS 2.1 berechnet.^{176,177} B) Experimentell bestimmte pK_a -Werte für Methylamin, Dimethylamin und Trimethylamin. Trotz des höheren Substituierungsgrades besitzen tertiäre Amine in wässrigen Lösungsmitteln aufgrund sterischer Effekte einen kleineren pK_a als primäre und sekundäre Amine.¹⁷⁵

Eine einzige Substanz, **AS_S06/07_30** (Abbildung 6-9a), stellt eine Ausnahme dieser klar definierten Struktur-Wirkungs-Beziehung dar. **AS_S06/07_30** erfüllt die gemeinsame Substruktur der sechs aktiven **OB108**-Nachkommen vollständig und müsste demnach ein potentieller Ligand der TAR-RNA sein, obwohl die Verbindung als inaktiv gemessen wurde. Tatsächlich wies diese Substanz bei ihrer ersten Messung bei 100 μM eine Inhibitionsrate von $56\% \pm 1\%$ auf (Dreifachmessung). In anschließenden Experimenten konnte diese Aktivität jedoch nicht reproduziert werden. Die Dosis-Wirkungs-Kurve ist „parabelförmig“, mit einer minimalen Inhibitionsrate von 30% bei 200-300 μM . Bei höheren Konzentrationen wurde die Inhibitionsrate jedoch wieder bis auf null reduziert. Aufgrund dieser Beobachtungen könnte eine Degenerierung der Substanz durch z.B. Lagerung oder Oxidation vorgelegen haben. Endgültige Sicherheit brächte nur eine erneute Bestellung und Messung der Substanz, evtl. mit einer anderen Messmethode unter anderen Bedingungen. Infolge dieser unsicheren Erkenntnisse wurde **AS_S06/07_30** als inaktiv betrachtet. Da die Verbindung jedoch die gemeinsame Substruktur der sechs aktiven **OB108**-Nachkommen vollständig erfüllt und anfangs auch als moderater Inhibitor gemessen wurde, bleibt die Möglichkeit offen, dass es sich hierbei um eine falsch negative Zuordnung handelt.

6.11.1 Pharmakophormodell der aktiven Nachkommen von **OB108**

Nach der Identifizierung struktureller Gemeinsamkeiten der sechs aktiven **OB108**-Nachkommen aus Abbildung 6-8 sollen nun auch gemeinsame Pharmakophoreigenschaften aufgedeckt werden. Es wurde deshalb in dem Programm MOE (Version 2007.09) mit der Funktion „Pharmacophore Elucidation“ eine Pharmakophoranalyse durchgeführt.⁹⁷ Um korrekte Ergebnisse liefern zu können benötigt diese Funktion Eingabemoleküle, die zu einander ähnlich sind. Es wurden deshalb nur Nachkommen von **OB108** (52 nicht doppelte Verbindungen, ohne Nachkommen der falsch positiven Verbindungen **AS_S01/07_37**) analysiert. Die Moleküle wurden im ersten Schritt mit MOE „gewaschen“, um ionisierbare Säure- und Basengruppen durch Hinzufügen oder Entfernen eines Protons auf ca. pH 7 einzustellen. Die Pharmakophorererkennung berechnet dann mit Hilfe einer stochastischen Suche bis zu 300 dreidimensionale Konformere aller Moleküle. Im nächsten Schritt werden den verschiedenen funktionellen Gruppen und Substrukturen der Konformere potentielle Pharmakophoreigenschaften zugewiesen. Hier wurden die Eigenschaften „aromatisch ODER π -Ringsystem“, „hydrophob“, „Protonendonor-Projektionspunkt“, „Protonenakzeptor-Projektionspunkt“, „kationisch“, „anionisch“, „Carboxy-bioisostere Gruppe“ und „Amidinium-bioisostere Gruppe“ verwendet.

Im darauf folgenden Schritt überlagert ein Minimierungsalgorithmus gleichartige Pharmakophoreigenschaften der Moleküle, um so eine räumliche Alignierung zu erreichen. Als Ausgabe erhält man 3-Punkt-, 4-Punkt und 5-Punktpharmakophor-überlagerungen der Eingabemoleküle. Anhand der Klassifizierung in aktive und inaktive Moleküle kann ein Gütewert für die verschiedenen Modelle berechnet werden. Es war mit dieser Vorgehensweise jedoch nicht möglich, ein diskriminierendes Pharmakophormodell für die sechs aktiven Nachkommen von **OB108** zu erstellen.

Das Problem kann auf eine unzureichend genaue Repräsentation der terminalen Aminogruppe eingegrenzt werden. Die „Wasch“-Funktion von MOE 2007.09 protoniert auch tertiäre Amine. Den entstehenden tertiären Ammoniumgruppen werden die Pharmakophoreigenschaften „Protonendonor“ und „kationisch“ zugeordnet. Der Vergleich in Abbildung 6-9 zeigt jedoch, dass dieser Aminogruppe eine besondere Bedeutung zukommt und dass die tertiäre Spezies offensichtlich nicht protoniert vorlag. Es wurde deshalb in der MOE-eigenen Skriptsprache SVL eine Funktion geschrieben, die speziell tertiäre Aminogruppen so „wäscht“, dass diese deprotoniert vorliegen. Eine anschließende Pharmakophorerkennung lieferte nun auch akkurate Pharmakophormodelle, von denen eines in einer manuell modifizierten Fassung in Abbildung 6-11 dargestellt ist. Die Genauigkeit eines Pharmakophormodells berechnet MOE 2007.09 nach Gl. 6-2:

$$\%Acc = \frac{TP + TN}{TP + FP + TN + FN} \cdot 100\% \quad (6-2)$$

mit den richtig positiven Treffern *TP*, den richtig negativen Treffern *TN*, den falsch positiven Treffern *FP* und den falsch negativen Treffern *FN*. Ein Wert von 100% wird erreicht, wenn alle positiven und negativen Beispiele richtig klassifiziert werden. Das Pharmakophormodell in Abbildung 6-11 erreicht eine Genauigkeit von $\%Acc = 96\%$ unter **OB108**-Nachkommen. Es findet alle sechs in Abbildung 6-8 dargestellten aktiven Liganden ($TP = 6$) und zusätzlich die inaktive Verbindung **AS_S06/07_30** ($FP = 1$), welche jedoch schon oben als potentiell aktiv diskutiert wurde. Die aktive Verbindung **AS_S06/07_40** wird nicht als positiver Treffer identifiziert und damit als falsch negativ gewertet ($FN = 1$). Da **AS_S06/07_40** nicht der gemeinsamen Substruktur aus Abbildung 6-8 entspricht, arbeitet das Pharmakophormodell hier dennoch korrekt. Das Bindeverhalten der sechs aktiven Liganden kann demnach mit einem 4-Punktpharmakophormodell bestehend aus zwei hydrophoben Eigenschaften,

einer aromatischen und einer kombinierten kationischen Protonendonoreigenschaft hinreichend beschrieben werden. Über den kationischen Protonendonor könnte eine essentielle Wasserstoffbrückenbindung vermittelt werden. Der aromatische Benzolring könnte π - π - oder π -Kation-Wechselwirkungen mit benachbarten Basen eingehen. Hydrophobe Eigenschaften werden ungerichtete Wechselwirkungen mit der Bindetasche eingehen.

Bemerkenswerterweise konnte dieses Modell nicht mit den Standardfunktionen von MOE erstellt werden. Es musste zunächst eine spezielle „Wasch“-Funktion programmiert werden, um der Pharmakophorerkennung die richtige Beschreibung der Eingabemoleküle zu ermöglichen. Eine Standardpharmokophorerkennung mit anschließender Pharmakophorsuche nach TAR-RNA-Liganden mit MOE würde deshalb auch retrospektiv andere Ergebnisse liefern, als das gewählte Verfahren der adaptiven Optimierung von Substanzbibliotheken. Pharmakophorsuche und adaptive Optimierung sind deshalb keinesfalls austauschbare, sondern allenfalls komplementäre Methoden.

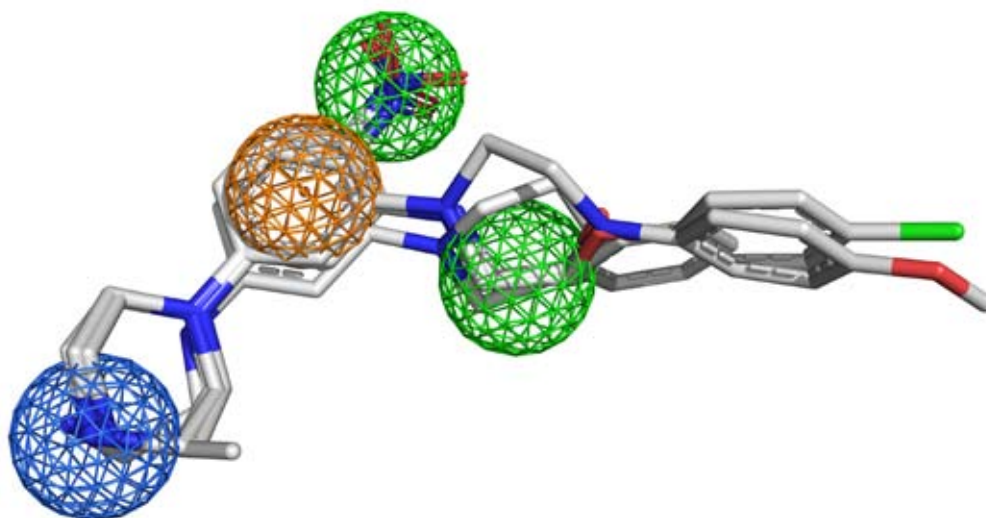


Abbildung 6-11: Pharmakophormodell von sechs aktiven Nachkommen von **OB108**. Das Modell wurde mit Hilfe der MOE-Funktion „Pharmacophore Elucidation“ und anschließender manueller Bearbeitung erstellt. Liganden: **AS_S10/06_33**, **AS_S01/07_19**, **AS_S01/07_62**, **AS_S06/07_27**, **AS_S06/07_28**, **AS_S06/07_29**. Pharmakophoreigenschaften: blau: Kation UND Protonendonor, orange: aromatisch, grün: hydrophob. Genauigkeit %Acc = 96%. Abb. erstellt mit PyMOL (DeLano Scientific, Palo Alto, CA, USA).^{177a}

6.11.2 Evaluierung des automatisierten Dockings in Nukleinsäuren

Die Hypothesen zu Struktur-Wirkungs-Beziehungen wurden durch automatisierte Dockingexperimente überprüft. Beim automatisierten Docking werden Moleküle durch Konformationsänderungen an eine vorgegebene Bindetasche angepasst. Die Qualität der Bindung an einen Rezeptor wird mittels einer Bewertungsfunktion berechnet und durch eine Punktzahl (engl. „Score“) angegeben. Neben der sterischen Passgenauigkeit geht vor allem auch eine Pseudoenergie der nicht kovalenten Wechselwirkungen zwischen Rezeptor und potentiellm Ligand in die Bewertungsfunktion ein. Je komplementärer Bindetasche und potentieller Ligand in Passform und Ausrichtung funktioneller Gruppen sind, desto höher die Punktzahl. Alle Dockingexperimente wurden mit dem Programm GOLD (Version 3.2, Cambridge Crystallographic Data Centre, Cambridge, Vereinigtes Königreich) durchgeführt.^{118,119} Als Bewertungsfunktion wurde *GoldScore* verwendet (vgl. Abschnitt 3.9).^{118,119} Das automatisierte Docking wurde ursprünglich für Protein-Liganden-Interaktionen entwickelt und anhand bekannter Protein-Liganden-Komplexe kalibriert. Das Docking in Nukleinsäuren als Rezeptoren stellt eine bekannte Herausforderung dar.^{152,178} Nukleinsäuren besitzen ausgeprägte negative Ladungen und eine höhere strukturelle Flexibilität als Proteine.^{152-154,178} Die Schwierigkeit des RNA-Dockings zeigte sich in einem ersten Validierungsexperiment. Der TAR-RNA-Inhibitor Acetylpromazin sollte in seine Bindetasche der TAR-RNA (TAR-Bulge) zurück gedockt werden. Dazu wurde das erste Modell der NMR-Struktur 1LVJ, einem Komplex aus Acetylpromazin und HIV-1 TAR-RNA, herangezogen (Abbildung 6-12, grüne Struktur).¹⁵⁹ Das Rückdocking von Acetylpromazin in seine eigene Bindetasche mit GOLD misslang. In keiner der zehn Durchläufe des Dockingverfahrens konnte die native Bindepose rekonstruiert werden. Die besten Posen des Liganden sind zum Lösungsmittel hin orientiert (Abbildung 6-12, blaue Struktur). Trotz der fehlgeschlagenen Positionierung lässt der positive *GoldScore*-Wert von 36 ± 11 jedoch auf eine begünstigte Bindung schließen. Es wurde also weder Bindepose noch Bindeaffinität richtig vorher gesagt.

In einem zweiten Rückdockingexperiment wurde eine essentielle Wasserstoffbrücke von Acetylpromazin zu Cytosin 24 der TAR-RNA vorgegeben (gestrichelte Linie rechts unten, Abbildung 6-12). Mit dieser Hilfestellung gelang es GOLD die native Pose mit einer mittleren Abweichung (RMSD) von 3,46 Å korrekt zu rekonstruieren (Abbildung 6-12, orangefarbene Struktur). Der *GoldScore*-Wert von 26 ± 2 ist für die native Pose im Mittel jedoch geringer als für die zuvor falsch vorhergesagte

Pose. GOLD war zwar durch Vorgabe einer essentiellen Wasserstoffbrücke in der Lage die native Pose zu rekonstruieren, jedoch versagte auch hier die Berechnung der Dockingpunktzahl. Aufgrund dieser ernüchternden Ergebnisse sollten alle hier gezeigten Dockingexperimente kritisch betrachtet werden. Auch das automatische Docking von Proteinliganden wurde in der letzten Zeit mit einiger Skepsis betrachtet.¹⁷⁹ Die Dockingergebnisse dieser Arbeit sollten deshalb generell nur als Vorschlag, und nicht als molekulare Tatsachen verstanden werden.

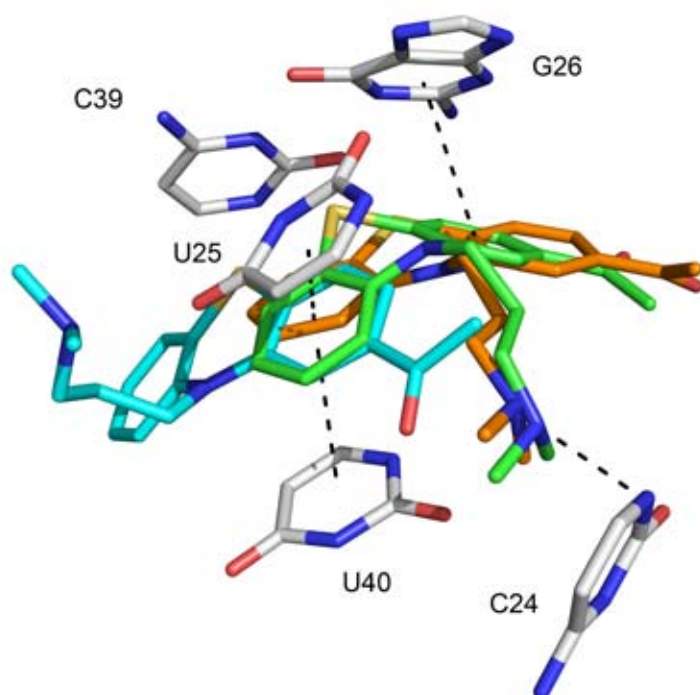


Abbildung 6-12: Vergleich von zwei Rückdockingexperimenten mit der NMR-Struktur von TAR-RNA (1LVJ, 1. Modell)¹⁵⁹ mit gebundenem Acetylpromazin (grün). Ein unbeschränktes Rückdockingexperiment konnte die native Bindepose nicht reproduzieren (blau, *GoldScore*-Wert: 36 ± 1). Wird die Wasserstoffbrückenbindung zur Base C24 im Dockingexperiment festgehalten, so wird eine der nativen Pose ähnliche Lage vorgeschlagen (orange, *GoldScore*-Wert: 26 ± 2 , RMSD zur NMR-Struktur: 3,46 Å). Abb. erstellt mit PyMOL (DeLano Scientific, Palo Alto, CA, USA).^{177a}

Die Probleme beim automatisierten Docking in Ribonukleinsäuren sind bekannt und wurden von anderen Gruppen auf verschiedene Weisen angegangen. Von Filikov *et al.* (2000) wurde das erfolgreiche Docking in RNA-Zielstrukturen mit der Dockingsoftware DOCK¹⁸⁰ und anschließender Verfeinerung der Bindeposen durch Monte-Carlo-Minimierung (mit der Software ICM¹⁸¹) beschrieben.¹⁸² Die Autoren beschreiben außerdem die Entwicklung einer speziellen Bewertungsfunktion für RNA-Liganden-Interaktionen. Die Arbeiten von Lind *et al.* (2002)¹⁸³ und Du *et al.* (2002)¹⁵⁹ bauen auf

die Arbeiten von Filikov auf. Du *et al.* konnten mit dieser Methode Acetylpromazin als TAR-RNA-Liganden identifizieren.¹⁵⁹ Detering & Varani (2004) berichten von guten Dockingergebnissen mit den Programmen DOCK und AutoDock¹⁸⁴, wobei sie nur geringfügige Anpassungen der Programme an Ribonukleinsäuren vornahmen.¹⁷⁸ Z.B. reduzierten sie die negative Ladung der Phosphatgruppen des RNA-Rückgrates. Barbault *et al.* (2006) beschreiben die Parametrisierung einer empirischen Bewertungsfunktion für das Docking in RNA-Strukturen mittels neuronaler Netze.¹⁸⁵ Im Rahmen des SFB 579 „Protein-Liganden-Wechselwirkungen“ der Universität Frankfurt wurde von Patrick Pfeffer und Dr. Holger Gohlke (2007) die wissensbasierte Bewertungsfunktion DrugScore^{RNA} entwickelt.¹⁸⁶ Diese stand jedoch bis zum Abschluss der vorliegenden Arbeit nicht zur Verfügung.

Für Dockingexperimente stand nur eine geringe Auswahl an hochaufgelösten TAR-RNA-Liganden-Komplexen zur Verfügung. Für aktive Nachkommen von **OB108** wurde die NMR-Struktur 1LVJ (1. Modell) mit gebundenem Acetylpromazin als Dockingrezeptor ausgewählt.¹⁵⁹ **OB108** wurde als Abkömmling von Acetylpromazin synthetisiert und beiden Molekülen ist das interkalierende Acetylphenothiazin-Motiv gemein. Für die aktiven Nachkommen von **OB080d** wurde die TAR-RNA-Struktur 1ARJ, im Komplex mit Arginin, ausgewählt (ebenfalls eine NMR-Struktur (1. Modell)).¹⁸⁷ Der Diaminopyrazol-Ring von **OB080d** wurde in Anlehnung an die Guanidiniumfunktion von Arginin synthetisiert und soll ein ähnliches Bindeverhalten aufweisen.

Tabelle 6-6: Mittlere GoldScore-Werte des Dockings der Verbindungen aller Generationen in die Bulge des ersten TAR-RNA-Modells der NMR-Strukturen 1LVJ und 1ARJ. Für jede Verbindung wurde der beste GoldScore-Wert von 10 Läufen berücksichtigt. Angegeben ist der Mittelwert mit Standardabweichung. Aktive Substanzen bezeichnen beim 1LVJ-Docking die aktiven Nachkommen von **OB108** und beim 1ARJ-Docking die aktiven Nachkommen von **OB080d** und **OB080d** selbst.

	GoldScore 1LVJ	GoldScore 1ARJ
Aktive Substanzen	61,64 ± 6,62	52,71 ± 6,59
Inaktive Substanzen	59,00 ± 10,03	54,12 ± 8,40
Alle Substanzen	59,28 ± 9,89	54,49 ± 8,87

Tabelle 6-6 zeigt das Ergebnis von automatisierten Dockingexperimenten mit den Verbindungen aller Generationen und den TAR-RNA-Strukturen 1LVJ und 1ARJ.

Jede der 180 Verbindungen wurde zehnmals gedockt. Bezeichnenderweise wurde nicht ein einziger negativer *GoldScore*-Wert berechnet. Die mittleren Punktzahlen für aktive, inaktive und alle Substanzen sind annähernd identisch. Für 1ARJ ist der *GoldScore*-Wert für die aktiven sogar geringer als für die inaktiven Verbindungen. Die Zahlen belegen mit Nachdruck, dass das hier durchgeführte automatisierte Docking mit GOLD (mit der Bewertungsfunktion *GoldScore*) als Methode für das virtuelle Screening von TAR-RNA-Liganden unbrauchbar ist. Die Dockingergebnisse sind mit Nichten diskriminierend – es wurden schlichtweg alle Verbindungen als potentielle Liganden klassifiziert.

6.11.3 Automatisiertes Docking der Nachkommen von OB108

Mit dem automatisierten Docking wurden im Anschluss mögliche Bindeposen der aktiven Substanzen in der Bulge von TAR-RNA generiert. Für sechs der sieben **OB108**-Nachkommen konnten übereinstimmende Bindeposen in der Bindetasche von 1LVJ gefunden werden (Abbildung 6-13). Da die *GoldScore*-Werte zu unzuverlässig waren, wurden sie nicht als Auswahlkriterium herangezogen, sondern die gezeigten Bindeposen wurden manuell aus den zehn gedockten Alternativen pro Ligand ausgewählt. Wie schon zuvor im Struktur- und Pharmakophorvergleich lässt sich eine wesentliche Übereinstimmung der Dockingposen der sechs Aktiven beobachten. Die Liganden liegen senkrecht zur Längsachse von TAR-RNA, wobei die kationische terminale Aminogruppe zur großen Furche orientiert ist. Wie auch bei Acetylpromazin scheinen π - π -Wechselwirkungen mit den Basen U25 und U40 möglich (gestrichelte Linien, Abbildung 6-13).

Des Weiteren deutet sich eine Wechselwirkung zwischen der terminalen Aminogruppe der Liganden und einem Sauerstoff des Phosphatrückgrates von C39 von TAR-RNA an. Diese Wasserstoffbrückenbindung und/oder elektrostatische Wechselwirkung könnte das entscheidende Aktivitätskriterium der **AS_S10/06_33**-analogen Verbindungen sein. Wird an dieser Stelle die Fähigkeit zur Ausbildung einer Wasserstoffbrücke zum Phosphatrückgrat durch das Fehlen einer primären oder sekundären Aminogruppe unterbunden, so könnte die gesamte Bindung destabilisiert werden und die Verbindungen wären somit inaktiv. Aufgrund der großen Ungenauigkeit des RNA-Dockings liegen in Wirklichkeit wahrscheinlich andere Bindeverhältnisse vor.

Der Hohlraum in Richtung C24 der Bindetasche von Acetylpromazin wird durch die sechs aktiven **OB108**-Nachkommen nur ansatzweise durch eine Nitrogruppe

ausgefüllt und bleibt größtenteils ungenutzt. Aufgrund der großen Flexibilität von TAR-RNA ist deshalb zu vermuten, dass die tatsächlichen Bindeverhältnisse durch „Induced fit“-Effekte (engl. „induzierte Passform“) anders gelagert sind. Trotz dieser Einschränkung stellt Abbildung 6-13 eine Hypothese für das Bindeverhalten der **OB108**-Nachkommen dar.

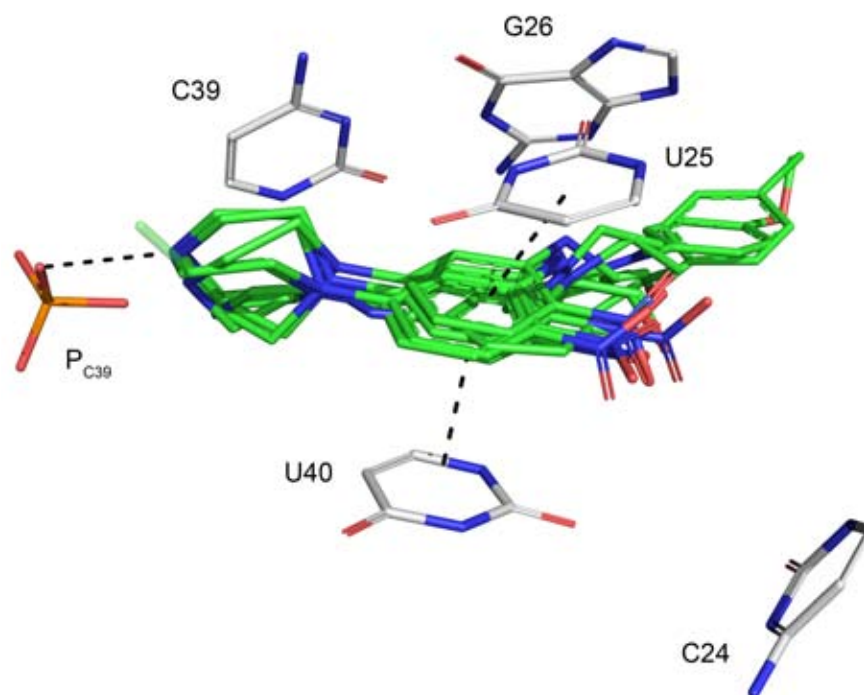


Abbildung 6-13: Dockingposen von sechs aktiven Nachkommen von **OB108** in der Bindetasche von TAR-RNA (1LVJ, 1. Modell,¹⁵⁹ Docking-Programm: GOLD 3.2). π - π -Wechselwirkungen mit den Basen U25 und U40 und eine Wasserstoffbrückenbindung/elektrostatische Wechselwirkung der terminalen Aminogruppe zum Phosphatrückgrat bei P_{C39} scheinen möglich (gestrichelte Linien). Die Tasche in Richtung der Base C24 bleibt weitestgehend unausgefüllt. Abgebildete Liganden: **AS_S10/06_33** Pose 8, **AS_S01/07_19** Pose 7, **AS_S01/07_62** Pose 5, **AS_S06/07_27** Pose 8, **AS_S06/07_28** Pose 6, **AS_S06/07_29** Pose 1. Abb. erstellt mit PyMOL (DeLano Scientific, Palo Alto, CA, USA).^{177a}

6.11.4 SAR des **OB108**-Nachkommens **AS_S06/07_40**

In der bisherigen Diskussion wurde der aktive **OB108**-Nachkomme **AS_S06/07_40** bewusst ausgeklammert. **AS_S06/07_40** besitzt als einziger der sieben aktiven Nachkommen von **OB108** trotz struktureller Ähnlichkeit nicht das gleiche Grundgerüst (Abbildung 6-14a). Eine Pharmakophorerkennung mit MOE lieferte keine überzeugenden Alignierungen mit **AS_S10/06_33**. In Dockingexperimenten in die Bindetasche von 1LVJ (1. Modell) nehmen beide Liganden eine um 180° gedrehte Konformation ein (Abbildung 6-14b). Bei beiden Liganden scheinen π - π -Wechselwirkungen mit den Basen U25 und U40 möglich (gestrichelte Linien, Abbildung

6-14b). Im Gegensatz zu den Analoga von **AS_S10/06_33** ist für **AS_S06/07_40** eine Wasserstoffbrückenbindung und/oder elektrostatische Wechselwirkung des Piperazin-Stickstoffs zum Phosphatrückgrat von U25 in der kleinen Furche denkbar. Aufgrund fehlender Analoga besitzen diese Aussagen allerdings einen sehr spekulativen Charakter. In Laborexperimenten erwies sich **AS_S06/07_40** zudem als schlecht löslich. Es kann nicht ausgeschlossen werden, dass die Aktivität der Substanz ein Artefakt der Messmethode ist (vgl. Abschnitt 6.5).

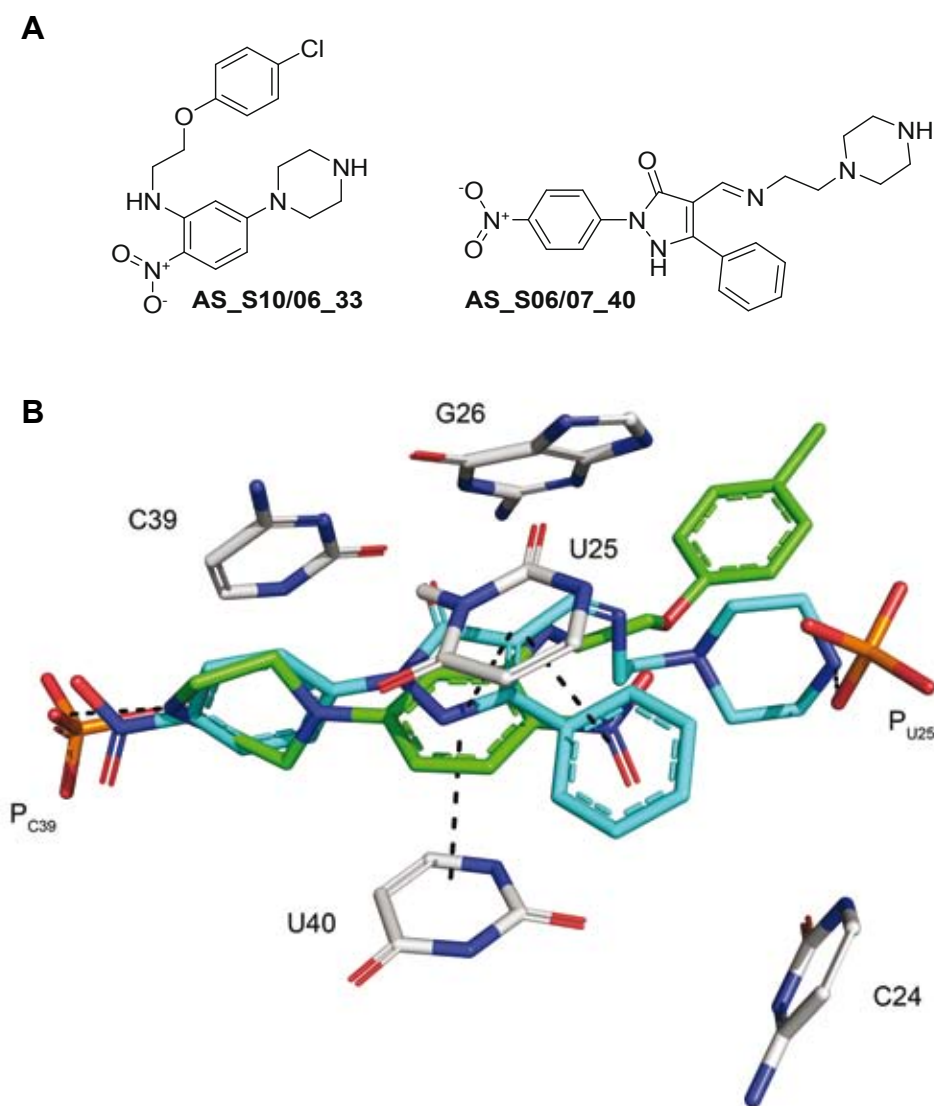


Abbildung 6-14: Vergleich der aktiven **OB108**-Nachkommen **AS_S10/07_33** und **AS_S06/07_40**. A) Der strukturelle Vergleich offenbart trotz starker Verwandtschaft ein unterschiedliches Grundgerüst. B) Dockingposen der aktiven Verbindungen **AS_S10/06_33** (grün) und **AS_S06/07_40** (blau) in der Bindetasche von TAR-RNA (1LVJ, 1. Modell,¹⁵⁹ Docking-Programm: GOLD 3.2). Für **AS_S06/07_40** scheinen π - π -Wechselwirkungen mit den Basen U25 und U40 und eine Wasserstoffbrückenbindung/elektrostatische Wechselwirkung des Piperazin-Stickstoffs zum Phosphatrückgrat P_{U25} möglich (gestrichelte Linien). Beide Liganden nehmen eine in Bezug auf ihre Nitro- und Piperazingruppen um 180° gedreht Lage ein. Abb. erstellt mit PyMOL.^{177a}

6.12 Struktur-Wirkungs-Beziehungen der Nachkommen von OB080d

Die Diskussion der Struktur-Wirkungs-Beziehungen der Nachkommen von **OB080d** erfolgt anhand einer Bindehypothese, die in der Arbeitsgruppe von Prof. Dr. Göbel entwickelt wurde.^{98,103} Demnach binden 2,4-Diaminopyrazole wie **OB080d** über bis zu fünf Wasserstoffbrücken mit der Base Guanin 26 und den Sauerstoffen der Phosphate P_{A22} und P_{U23} (Abbildung 6-15a). Der Pyrazolring würde demnach protoniert, also positiv geladen vorliegen. Des Weiteren sind in dieser Anordnung π - π - und π -Kation-Wechselwirkungen mit der Base Adenin 22 denkbar, welche sich unterhalb der Pyrazolringebene befindet. Anhand des ersten Modells der NMR-Struktur 1ARJ von TAR-RNA¹⁸⁷ wurde das Bindemodell wie in Abbildung 6-15b dargestellt modifiziert. In Modellierungsanalysen erschien eine Bindung von G26 über zwei Wasserstoffbrücken und die gleichzeitige Bindung von P_{A22} wegen eines zu großen Abstandes zwischen den interagierenden Gruppen unwahrscheinlich.

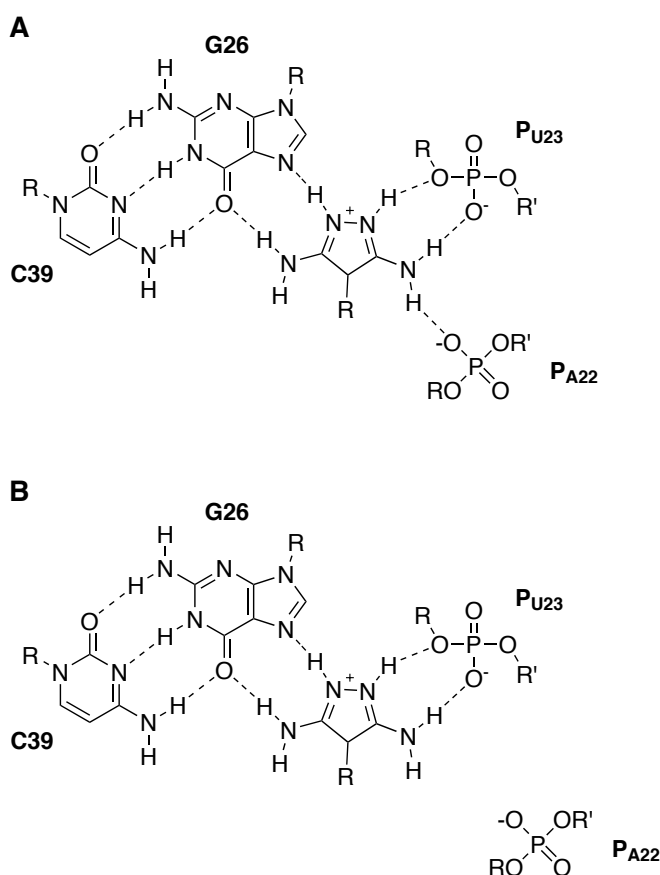


Abbildung 6-15: Hypothetische Bindemodelle von 2,4-Diaminopyrazolen in der Bindetasche von TAR-RNA. A) Arbeitshypothese der Arbeitsgruppe von Prof. Dr. Göbel. B) Anhand der NMR-Struktur 1ARJ (1. Modell)¹⁸⁷ modifiziertes Bindemodell. Eine Wechselwirkung mit dem Rückgrad bei P_{A22} scheint unwahrscheinlich.

Anhand der allgemeinen Bindehypothese wurden spezifische Bindemodelle für das Elternmolekül **OB080d** und dessen zwei aktiven Nachkommen **AS_S10/06_16** und **AS_S01/07_35** erstellt (Abbildung 6-16a-c). **OB080d** und **AS_S10/06_16** weisen beide ein 1*H*-Pyrazol-Grundgerüst auf; **AS_S01/07_35** besitzt ein analoges Pyrimidingrundgerüst. Aufgrund ihrer verschiedenen Aminosubstituenten sind alle drei Verbindungen in der Lage, ein ähnliches Wasserstoffbrückenmuster zu G26 und P_{U23} auszubilden, durch das sie stabile Bindungen zu TAR-RNA eingehen können. Die Bindung von **AS_S10/06_16** könnte zusätzlich durch eine Wasserstoffbrücke zu C39 stabilisiert werden (Abbildung 6-16b). Zur weiteren Validierung wurden die hypothetischen Bindemodelle dreidimensional mit MOE in der Bindetasche von TAR-RNA (1ARJ, 1. NMR-Modell)¹⁸⁷ modelliert (Abbildung 6-16d). Die Vorgehensweise der Modellierung entspricht einem manuellen Docking. Auch hier konnte das vorgeschlagene Wasserstoffbrückenmuster rekonstruiert werden. Zusätzlich wäre hier eine π - π - und/oder π -Kation-Wechselwirkung mit A22 denkbar.

Ein automatisiertes Dockingexperiment mit GOLD (hier: Version 4.0) schlug vergleichbare Bindeposen für die Verbindungen **OB080d** und **AS_S10/06_16** vor (Abbildung 6-17). Die Moleküle wurden für die Vorbereitung auf das Docking einem speziellen „Wasch“-Schritt unterzogen, der den Stickstoff an Position 2 der 1*H*-Pyrazole protonierte. Die Protonierung erfolgte in MOE mit Hilfe eines eigens dafür geschriebenen SVL-Skriptes. Die von GOLD vorgeschlagenen Posen unterschieden sich vom manuellen Docking durch einen fehlenden Kontakt zu P_{U23} (Abbildung 6-17). Die im automatisierten Docking für **AS_S01/07_35** vorgeschlagenen Bindeposen entsprachen nicht der vorgestellten Bindehypothese. **AS_S01/07_35** wurde von GOLD 4.0 in allen zehn Dockingläufen an den äußeren Rand der Bindetasche, in Richtung der zum Lösungsmittel offenen Seite orientiert (Abbildung 6-17). Diese Posen erschienen wenig plausibel und wurden deshalb verworfen.

Zusammenfassend konnten die entwickelten Bindehypothesen durch automatisierte Dockingexperimente nur eingeschränkt validiert werden. Durch die sehr begrenzte Anzahl von nur drei Liganden dieses Typs und die eingeschränkte Nutzbarkeit von automatisierten Dockingverfahren für das Docking in Nukleinsäuren ist das Ableiten von Struktur-Wirkungs-Beziehungen auf diesem Wege als problematisch zu betrachten.

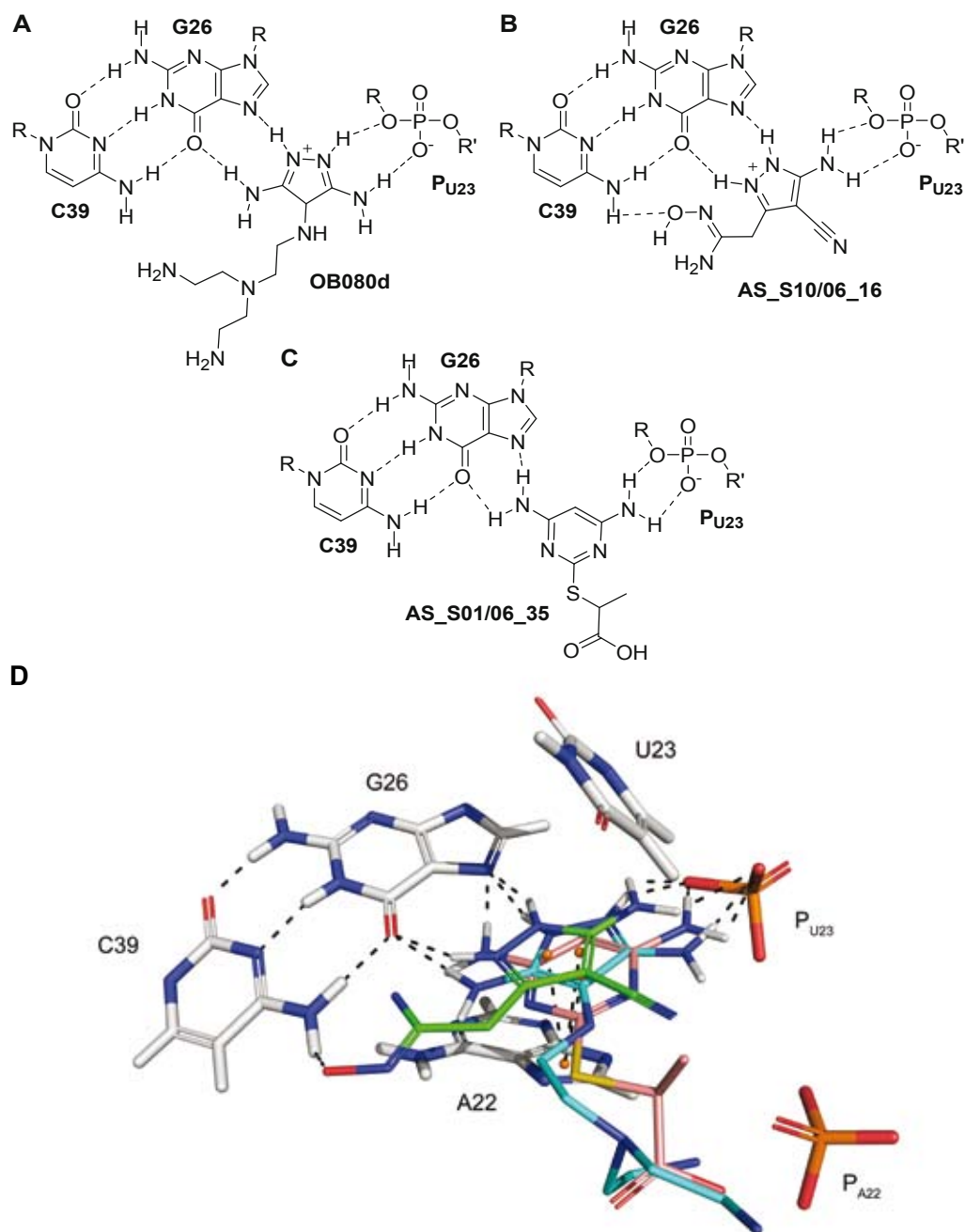


Abbildung 6-16: Hypothetische Bindemodelle der aktiven Verbindungen **OB080d** (A), **AS_S10/06_16** (B) und **AS_S01/07_35** (C) für HIV-1 TAR-RNA. Die Bindemodelle wurden auf Grundlage der Arbeitshypothese zu 2,4-Diaminopyrazolen des Arbeitskreises von Prof. Dr. Göbel erstellt. Alle drei Liganden zeichnen sich durch ein identisches Wasserstoffbrückenmuster aus. D) 3D-Bindeposen von **OB080d** (blau), **AS_S10/06_16** (grün) und **AS_S01/07_35** (rosa) in der Bindetasche von TAR-RNA (1ARJ, 1. Modell)¹⁸⁷. Die Posen wurden durch manuelles Docking mit dem Programm MOE modelliert. Wasserstoffbrückenbindungen zu den Basen C39 und G26, und zum Phosphatrückgrad P_{U23} sind denkbar (gestrichelte Linien). Des Weiteren erscheinen π - π -Wechselwirkungen und π -Kation-Wechselwirkungen zur Base A22 möglich (durch gestrichelte Linien verbundene orangefarbene Kugeln). Abb. D erstellt mit PyMOL (DeLano Scientific, Palo Alto, CA, USA).^{177a}

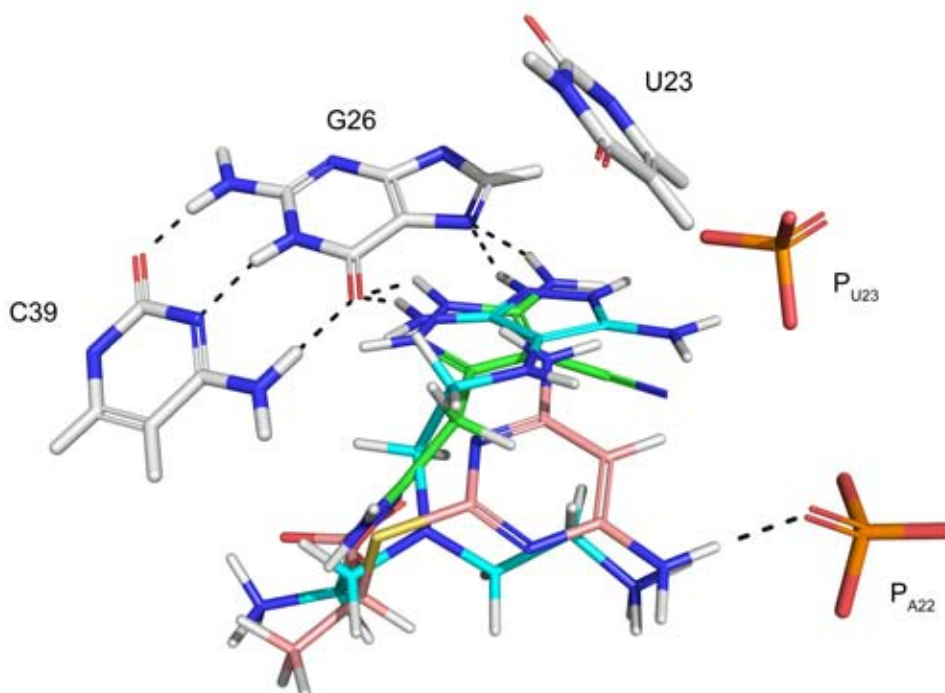


Abbildung 6-17: Docking-Bindeposen der aktiven Liganden von **OB080d**. Von GOLD (Version 4.0) vorgeschlagene Bindepose von **OB080d** (blau), **AS_S10/06_16** (grün) und **AS_S01/07_35** (rosa) in der Bindetasche von TAR-RNA (1ARJ, 1. Modell)¹⁸⁷. Die Posen für **OB080d** (Pose 1) und **AS_S10/06_16** (Pose 6) ähneln jenen des manuellen Dockings, unterscheiden sich aber durch einen fehlenden Kontakt zu P_{U23} (vgl. Abbildung 6-16). Die für **AS_S01/07_35** (Pose 9) vorgeschlagene Bindepose entspricht nicht der vorgestellten Bindehypothese, da die Verbindungen an den äußeren Rand der Bindetasche, in Richtung der zum Lösungsmittel offenen Seite orientiert wurde. Diese Pose erscheint wenig plausibel. Abb. erstellt mit PyMOL (DeLano Scientific, Palo Alto, CA, USA).^{177a}

Als alternatives Verfahren zur dreidimensionalen Analyse von Struktur-Wirkungs-Beziehungen wurde ein Pharmakophormodell anhand der drei Liganden **OB080d**, **AS_S10/06_16** und **AS_S01/07_35** erstellt. Das automatisierte Verfahren der Pharmakophorererkennung in MOE („Pharmacophore Elucidation“) lieferte keine zufrieden stellenden Ergebnisse. Die Pharmakophorererkennung wurde mit gewaschenen und nicht gewaschenen Molekülen durchgeführt. Bei den gewaschenen Molekülen wurden zusätzlich die Varianten „protonierter Pyrazolring“ und „nicht protonierter Pyrazolring“ getestet; jeweils ohne Erfolg. Das Pharmakophormodell in Abbildung 6-18 wurde manuell mit MOE erstellt. Die Grundlage des Modells waren die bereits vorgestellten Bindehypothesen aus Abbildung 6-16. Das Modell wurde des Weiteren im Hinblick darauf erstellt, möglichst nur **OB080d** und seinen zwei aktiven Nachkommen unter allen **OB080d**-Nachkommen zu entsprechen. In einer Pharmakophorsuche unter allen **OB080d**-Nachkommen und **OB080d** selbst wurden mit diesem Modell zehn Treffer erzielt; darunter die drei aktiven Liganden (Schema 6-6). Gleichzeitig wurden diesmal auch sieben eng verwandte Verbindungen gefunden, die jedoch als inaktiv gemessen wurden. Aktive und inaktive Analoga unterscheiden sich in Schema 6-6 teilweise nur durch die Präsenz oder Abwesenheit einer einzigen Methylgruppe (vgl. **AS_S01/07_35** (aktiv) und **AS_S01/07_40** (inaktiv)). Die Genauigkeit des Pharmakophormodells beträgt $\%Acc = 73\%$ (vgl. Gl. 6-2).

Die Diskussion der Struktur-Wirkungs-Beziehungen von **OB080d** und seinen Nachkommen bleibt auf einem spekulativen Level. Die geringe Anzahl von nur drei aktiven Liganden dieses Typs und die Ungenauigkeit der eingesetzten Computerverfahren erschweren die Diskussion. Das Bindeverhalten der drei Liganden wurde anhand ihrer relativen Ähnlichkeit zu Guanidinium diskutiert, der funktionellen Gruppe der Seitenkette der Aminosäure Arginin, welche gehäuft in der TAR-Bindedomäne des Tat-Proteins zu finden ist. Für TAR-RNA im Komplex mit Arginin existiert wiederum eine NMR-Struktur (1ARJ)¹⁸⁷, dessen erstes Minimierungsmodell die Grundlage der vorgestellten Bindehypothesen darstellte. Die vorgestellten Bindemodelle basieren also auf einer ganzen Reihe ungeprüfter Annahmen. Aufgrund der Flexibilität von TAR-RNA wäre es deshalb gut denkbar, dass **OB080d** und seine zwei aktiven Nachkommen komplett andere Bindeposen in bislang unbekanntem TAR-Konformationen eingehen. Die Computerchemie kann an diesem Punkt zunächst keine weitere Aussage treffen; eine weitere Analyse der Struktur-Wirkungs-Beziehungen muss experimentell im Labor erfolgen.

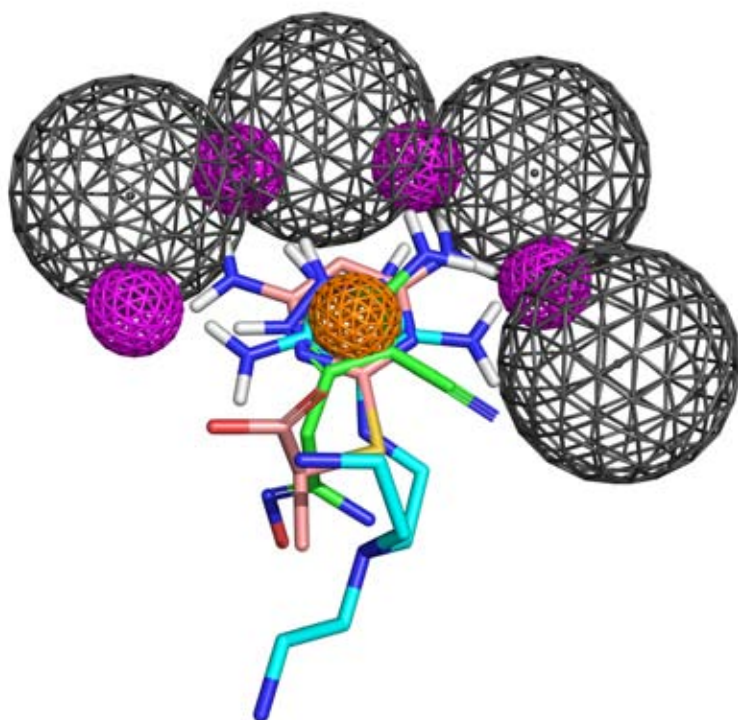
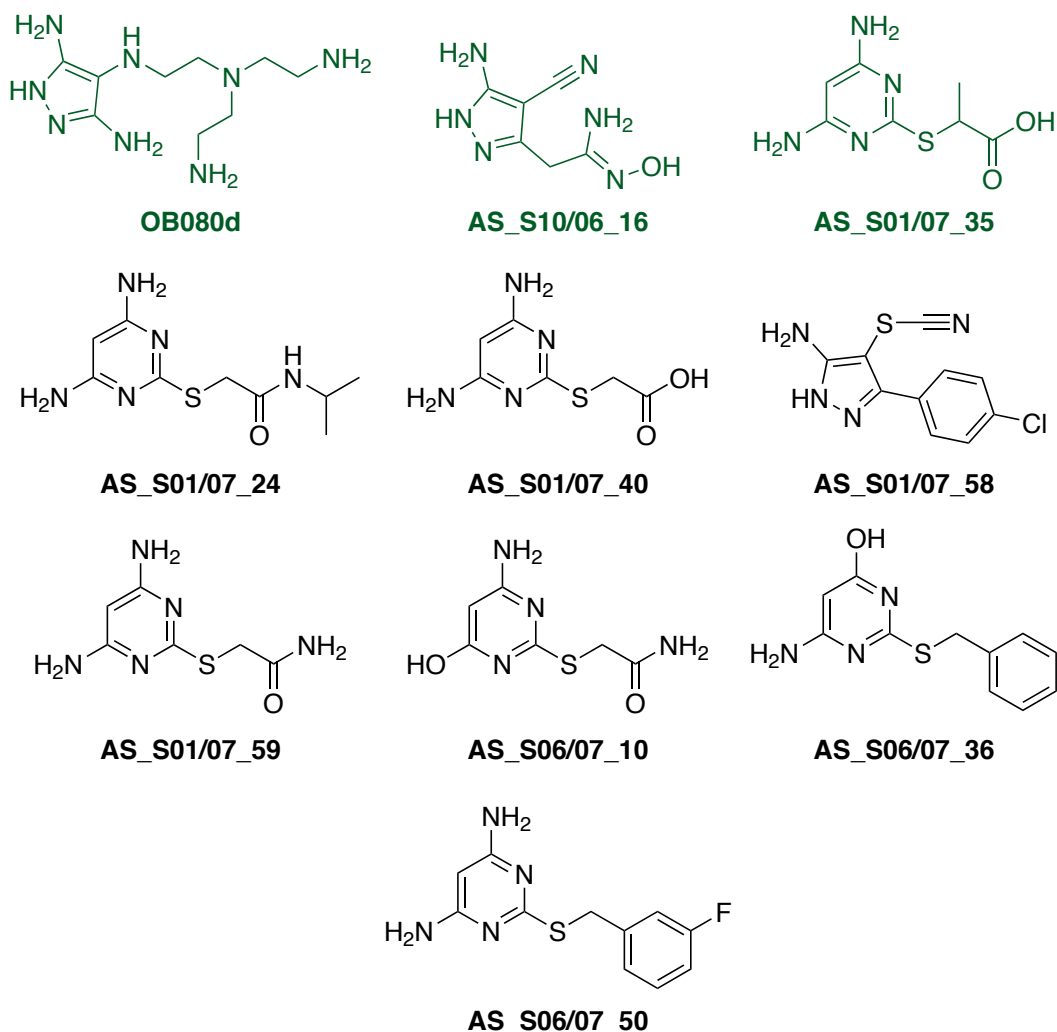


Abbildung 6-18: Pharmacophormodell der Verbindungen **OB080d** (blau), **AS_S10/06_16** (grün) und **AS_S01/06_35** (rosa). Die Alignierung der Liganden und die Erstellung des Pharmacophormodells erfolgten manuell auf Grundlage der Arbeitshypothese zu 2,4-Diaminopyrazolen (Abbildung 6-16). Pharmacophoreigenschaften: pink: Wasserstoffbrückendonor-Projektionspunkt, orange: aromatisch, grau: verbotenes Ausschlussvolumen. Genauigkeit $\%Acc = 73\%$; 3 von 10 Treffern unter allen **OB080d**-Nachkommen (inkl. **OB080d**) sind aktiv, d.h. 7 Verbindungen wurden fälschlicherweise durch die Pharmacophorsuche als positiv bestimmt. Abb. erstellt mit PyMOL (DeLano Scientific, Palo Alto, CA, USA).^{177a}



Schema 6-6: Treffer einer Pharmakophorsuche mit dem Pharmakophormodell aus Abbildung 6-18 in einer Molekülsammlung bestehend aus allen Nachkommen von **OB080d** (und einschließlich **OB080d**). Aktive Verbindungen sind grün gezeichnet, inaktive schwarz.

6.13 Zellkulturexperimente

Die beiden viel versprechenden Liganden **AS_S10/06_16** ($IC_{50} = 116 \mu\text{M}$) und **AS_S10/06_33** ($IC_{50} = 51 \mu\text{M}$) wurden zur weiteren Validierung in einem Zellkulturexperiment auf Aktivität überprüft. Die Untersuchungen erfolgten in Kooperation mit Dr. Ursula Dietrich am Georg-Speyer-Haus in Frankfurt am Main und wurden freundlicherweise von Michaela Stoll durchgeführt. Die Zellkulturexperimente ermöglichen die Testung von potentiellen HIV-Inhibitoren *in vivo* (Abbildung 6-19). Dafür wurden HeLa P4-Zellen, welche an ihrer Oberfläche CD4-Rezeptoren und CCR5/CXCR4-Corezeptoren tragen, mit einem *lacZ*-Gen transfiziert, das unter Kontrolle des HIV-Promotors im LTR-Bereich steht.¹⁰⁰ Das Reportergen wurde nach

der Sequenz von TAR in die DNA eingefügt. Kommt es nach der Infektion durch HIV-1 (Stamm: Lai) zur Antitermination durch Tat wird das Genprodukt von *lacZ*, die β -Galactosidase (β -Gal), exprimiert. Dem Nährmedium der Zellkulturen wurde ein Farbstoffreagenz (Galacton) zugesetzt, welches der β -Galactosidase als Substrat dient. Das enzymatisch gespaltene Galacton geht eine Farbreaktion ein, die colorimetrisch nachgewiesen werden kann (Blaufärbung). Inhibitoren der Antitermination verhindern die Expression der β -Galactosidase und unterbinden damit auch die Farbreaktion. Durch die Anzahl und Intensität blau gefärbter Zellen kann so die Stärke eines potentiellen Inhibitors quantifiziert werden.

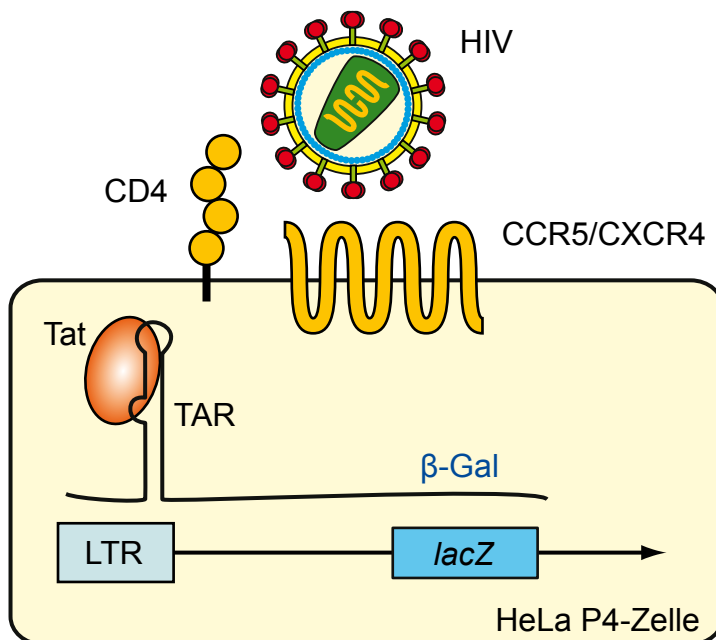


Abbildung 6-19: Schema der Reporter-Gen-Untersuchung in Zellkulturen. HeLa P4-Zellen, welche an ihrer Oberfläche CD4-Rezeptoren und CCR5/CXCR4-Corezeptoren tragen, wurden mit einem *lacZ*-Gen transfiziert, das unter Kontrolle des HIV-Promotors im LTR-Bereich steht. Kommt es nach der Infektion durch HIV-1_{Lai} zur Antitermination durch Tat wird das Genprodukt von *lacZ*, die β -Galactosidase (β -Gal), exprimiert. Dem Nährmedium der Zellkulturen wurde ein Farbstoffreagenz (Galacton) zugesetzt, welches β -Galactosidase als Substrat dient. Das enzymatisch gespaltene Farbstoffreagenz reagiert zu einem blauen Farbstoff, der colorimetrisch nachgewiesen werden kann. Inhibitoren der Antitermination verhindern die Expression der β -Galactosidase und unterbinden damit auch die Farbreaktion.

Nach der Infektion durch HIV-1 werden nicht inhibierte Zellen das Genom von HIV transkribieren und die Farbreaktion auslösen. Je stärker die Potenz eines Inhibitors, desto weniger Transkription wird erfolgen und umso geringer wird die Blaufärbung ausfallen. Die Reporter-Gen-Untersuchung liefert einen Hinweis darauf, ob eine Substanz die Antitermination der Transkription des viralen Genoms inhibieren kann.

Dies kann zum einen – wie auch hier gewünscht – über eine direkte Inhibition der Tat-TAR-Interaktion erfolgen. Zum anderen könnte ein Inhibitor aber auch Prozesse im Lebenszyklus von HIV stören, die zu einem früheren Zeitpunkt ablaufen. Ein prominentes Beispiel dafür wäre die Inhibition des Zelleintritts von HIV. Zur Bestimmung, welcher Wirkmechanismus im Einzelfall vorliegt, kann die Untersuchung zusätzlich in einer modifizierten Fassung durchgeführt werden. Darauf wurde hier jedoch verzichtet, da die Ergebnisse der Zellkulturuntersuchungen den zusätzlichen Aufwand nicht rechtfertigten. Eine Beschreibung der Versuchsbedingungen ist im Anhang in Abschnitt 8.6 zu finden.

In den Zellkulturexperimenten konnte für Substanz **AS_S10/06_16** erfolgreich eine inhibitorische Aktivität nachgewiesen werden (Abbildung 6-20a). Bei einer Substanzkonzentration von 100 μM sank die Farbintensität der Farbreaktion auf 51% im Vergleich zur nicht inhibierten Probe. Die absolute, mittlere inhibitorische Konzentration (IC_{50}^{Abs}) liegt demnach bei rund 100 μM . Dieser Wert steht in guter Übereinstimmung mit dem *in vitro* bestimmten IC_{50} -Wert von 116 μM . Bei Konzentration größer 100 μM wurden jedoch erhöhte Farbintensitätswerte gemessen. Bei 250 μM wurde z.B. ein Intensitätswert von rund 200% gemessen, was einer Verdopplung der Farbintensität im Vergleich zur Positivkontrolle darstellt. *In vitro* wurde bei den FRET-Experimenten im Konzentrationsbereich größer 100 μM eine Absenkung der Fluoreszenzintensität auf -30% im Vergleich zur Hintergrundintensität festgestellt (vgl. Abbildung 8-1 im Anhang). Beide Phänomene weisen auf ein irreguläres Verhalten dieser Verbindung hin. Ihre unmittelbare Elternstruktur **OB080d** wurde bereits bei der Diskussion der Initialpopulation (vgl. Abschnitt 6.3) als potentiell problematisch eingestuft. Proben von **OB080d** neigten zur Polymerisierung/Aggregation, wodurch komplexe, polykationische Strukturen entstanden. Die Kindverbindung **AS_S10/06_16** besitzt wie **OB080d** ein 1H-Pyrazol-Grundgerüst und könnte ebenfalls Polymerisierungs- oder Aggregationseffekten unterliegen. Eine weitere Aufklärung des Inhibitionsverhaltens von **AS_S10/06_16** ist nur durch weiterführende Experimente zu erreichen.

Für Verbindung **AS_S10/06_33** konnte in den Zellkulturexperimenten keine weitere Aussage zur Aktivität erhalten werden. Die Untersuchung konnte nicht erfolgreich durchgeführt werden, da **AS_S10/06_33** schon bei der niedrigsten Konzentration von 25 μM cytotoxisch wirkte (Abbildung 6-20b). Die toxische Wirkung auf HeLa-Zellen war nach Aussage von Frau Michaela Stoll im Mikroskop sichtbar. Wie im Abschnitt 6.10 diskutiert könnte die starke Toxizität eine Folge der Stoffwechselprodukte dieser

Nitroarenverbindung sein. Es wäre interessant analoge Verbindungen zu untersuchen, bei denen man die Nitrogruppe durch weniger problematische isostere Gruppen ersetzen würde.

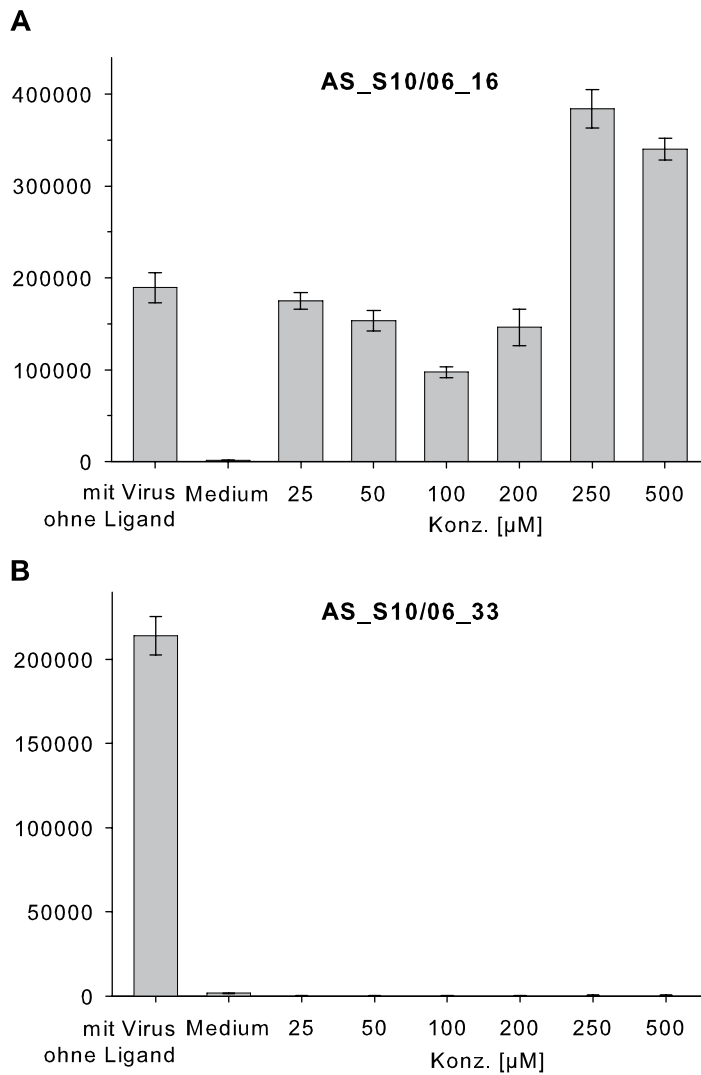


Abbildung 6-20: Ergebnisse der Zellkulturexperimente von **AS_S10/06_16** (A) und **AS_S10/06_33** (B). Der Balken „mit Virus ohne Ligand“ entspricht der Positivkontrolle ohne Zugabe der Substanzen (volle Farbreaktion). Der Balken „Medium“ entspricht der Negativkontrolle ohne Zugabe von HIV und ohne Zugabe der Substanzen (keine Farbreaktion). Die Konzentrationsreihe von 25 bis 500 μM zeigt den Intensitätswert der Farbreaktion nach Zugabe unterschiedlicher Konzentrationen der Liganden. Balken zeigen den Mittelwert einer dreifachen Messung; Fehlerbalken geben die Standardabweichung der Messwerte an.

Zusammenfassend konnten die Ergebnisse der Zellkulturexperimente nicht wesentlich für die weitere Charakterisierung der zwei TAR-RNA-Liganden **AS_S10/06_16** und **AS_S10/06_33** beitragen. Für **AS_S10/06_16** konnte ein absoluter IC_{50} bestimmt werden, der in guter Übereinstimmung mit den Daten der FRET-Messungen steht. Bei höheren Konzentrationen wurden im Zellkulturexperiment jedoch Messwerte mit einem Effekt $> 100\%$ bestimmt, die eine Diskussion erschweren. Ein relativer IC_{50} kann deshalb nicht angegeben werden. Substanz **AS_S10/06_33** wirkte schon in geringen Konzentrationen cytotoxisch und entzog sich somit der experimentellen Zugänglichkeit.

6.14 Schlussfolgerungen

Es wurde der erfolgreiche prospektive Entwurf einer fokussierten Substanzbibliothek mit Hilfe der adaptiven Optimierung unter eingeschränkten Ressourcen gezeigt. Ziel der Optimierung war das Auffinden von Liganden der HIV-1 TAR-RNA, welche zur Störung deren Interaktion mit dem viralen Transaktivator der Transkription (Tat) in der Lage waren. In einem der Bionik entlehnten Ansatz wurde eine hybride Niedrigdurchsatzscreening-Kampagne durchgeführt, bei der ein Optimierungsalgorithmus (hier: eine Evolutionsstrategie) Molekülvorschläge entwarf und diese anschließend in einem biochemischen Labor auf Aktivität überprüft wurden. Das Ergebnis der Labormessungen wurde wieder in den Algorithmus eingespeist, welcher daraufhin eine neue Generation an Molekülvorschlägen generierte. Die Bioaktivität wurde *in vitro* mittels einer FRET-basierten, kompetitiven Verdrängungsmethode bestimmt.⁸⁸ Moleküle wurden beim Chemikalienanbieter Specs (Delft, Niederlande) bezogen.⁶²

In insgesamt vier Generationen wurden 170 Screeningverbindungen auf Aktivität überprüft. Darunter wurden neun Substanzen positiv auf Inhibition der Tat-TAR-RNA-Interaktion getestet. Dies entspricht einer Trefferquote von 5,3%, wobei lediglich 0,089% der Verbindungen der Specs-Bibliothek (191.405 Substanzen) untersucht wurden. Die mittleren inhibitorischen Konzentrationen (IC_{50}) lagen im mittleren mikromolaren Bereich ($IC_{50} = 506 \pm 497 \mu\text{M}$), wobei der potenteste Kandidat **AS_S10/06_33** ([2-(4-Chloro-phenoxy)-ethyl]-(2-nitro-5-piperazin-1-yl-phenyl)-amin) einen IC_{50} von 51 μM aufwies. In einem anderen Zweig des evolutionären Stammbaumes konnte Verbindung **AS_S10/06_16** (2-(5-Amino-4-cyano-1H-pyrazol-3-yl)-N-hydroxy-acetamidin) mit $IC_{50} = 116 \mu\text{M}$ als TAR-RNA-Ligand identifiziert werden.

Eine Analyse des Optimierungsverhaltens der eingesetzten (10, 60)-Evolutionstrategie zeigte wie zuvor in den retrospektiven Studien eine große Eltern-Kind-Ähnlichkeit. Die Nachkommen eines Elters wurden in dessen lokalen Nachbarschaft gesucht. Lagen zwischen zwei zufällig ausgewählten Verbindungen der Specs-Bibliothek im Mittel 83.143 weitere Verbindungen, so waren es bei der Evolutionstrategie zwischen Eltern und Kindern nur 14. Trotz der Beschränkung auf die lokale Nachbarschaft konnte „Grundgerüstspringen“ beobachtet werden.¹²² Während das Elternmolekül **OB108** ein Phenothiazin-Grundgerüst aufweist, ist sein aktiver Nachkomme **AS_S10/06_33** ein Nitrophenylpiperazin-Derivat. In einem anderen Zweig des evolutionären Stammbaums wurde das 1*H*-Pyrazol-Motiv des Elters **AS_S10/06_16** bei seinem Nachkommen **AS_S01/07_35** auf ein Pyrimidin-Motiv abgewandelt.

Bei der experimentellen Handhabung der Molekülkandidaten sind Probleme mit der wässrigen Löslichkeit aufgetreten. Es wurde zunächst entschieden, auf jegliche Vorfilterung der Screeningverbindungen zu verzichten, dabei dem Optimierungsalgorithmus „freien Lauf“ zu lassen und Moleküle mit unerwünschten Eigenschaften vielmehr durch „natürliche Selektion“ zu entfernen. Dies hatte jedoch zur Folge, dass 39% aller getesteten Verbindungen nicht, oder nur schlecht löslich waren. Ein nachträglich installierter Löslichkeitsfilter führte auf Grund zu ungenauer Vorhersagemodelle der wässrigen Löslichkeit und des Wasser-Oktanol-Verteilungskoeffizienten nicht zum gewünschten Erfolg. Für die Zukunft wird deshalb empfohlen, gleich von Beginn an mit einem Löslichkeitsfilter zu arbeiten und weitere Anstrengungen für die Verbesserung der Qualität dieses Filter zu unternehmen.

Da die Optimierung auf Nitroarene konvergierte und diese im humanen Metabolismus problematisch sein können,¹⁶⁹ wurde auf die Durchführung der letzten Generation der Optimierung verzichtet, um stattdessen nach Nicht-Nitroverbindungen zu suchen. Mittels chemischer Ähnlichkeitssuche wurden neun Verbindungen ausgewählt, die eine relative Ähnlichkeit zur Nitroverbindung **AS_S10/06_33** aufwiesen, selber jedoch keine Nitrogruppe trugen. Unter diesen neun Verbindungen konnte jedoch nur eine einzige als potentiell schwach aktiv bestimmt werden. Dies zeigt, dass chemische Ähnlichkeit allein nicht hinreichend für eine biologische Wirkung ist.¹⁷⁴ Es war zunächst Zufall, dass eine Nitroverbindung wie **AS_S10/06_33** vom Optimierungsalgorithmus ausgewählt wurde. Jedoch hat in den folgenden Schritten die Logik der Evolutionstrategie bewirkt, dass dieser Zufall zu einem evolutionären Vorteil ausgenutzt wurde. Es ist diese Charakteristik des gerichteten Zufalls, die die stochastischen Optimierungsverfahren so robust, und deshalb so attraktiv machen.

Für Nachkommen des Elters **OB108** (darunter auch **AS_S10/06_33**) wurde mit chemieinformatischen Methoden eine Analyse der Struktur-Wirkungs-Beziehungen (SAR) durchgeführt. Sechs der sieben aktiven **OB108**-Nachkommen weisen ein konserviertes Substrukturmotiv aus 16 Atomen auf. Eine endständige primäre oder sekundäre Aminogruppe konnte als besonders wichtig für die Bioaktivität ausgemacht werden. Wird der Stickstoff durch ein Sauerstoff- oder ein Kohlenstoffatom ersetzt, geht die Aktivität verloren. Auch tertiäre Aminogruppen sind an dieser Stelle unzulässig.

Es wurde daher die Schlussfolgerung gezogen, dass eine terminale tertiäre Aminogruppe im lokalen Milieu der Bindetasche unprotoniert vorliegt und damit ihre essentielle Eigenschaft als Protonendonator/Kation verloren geht. Ein Pharmakophormodell der sechs aktiven Substanzen konnte erfolgreich erstellt werden. Dies war jedoch erst dann möglich, als tertiäre, terminale Aminogruppen manuell deprotoniert wurden. Die Standardeinstellungen des verwendeten Modellierungsprogramms (MOE) sahen dies nicht vor.

Es konnte weiter gezeigt werden, dass das automatisierte Docking kleiner organischer Moleküle in Nukleinsäuren mittels der Software GOLD¹¹⁸ nahezu unbrauchbar ist. GOLD bewertete die Bindung aller Substanzen unspezifisch als positiv (Bewertungsfunktion: *GoldScore*). Eine Priorisierung der Substanzen war damit unmöglich. Bei einem Docking in die Bindetasche der TAR-RNA-Struktur 1LVJ¹⁵⁹ konnten jedoch für sechs aktive **OB108**-Nachkommen sehr ähnliche Bindemodi gefunden werden. Demnach könnte die terminale Aminogruppe eine essentielle Wechselwirkung zum Phosphatrückgrat von C39 vermitteln. Der Benzolring könnte an π - π -Wechselwirkungen mit den Basen U25 und U40 der TAR-Bulge beteiligt sein.

Eine SAR-Analyse der **OB080d**-Nachkommen (darunter auch **AS_S10/06_16**) verlief weniger aufschlussreich. Da automatisierte Verfahren Probleme bereiteten, wurden Bindemodelle anhand einer vorhandenen Bindehypothese zu 1H-Pyrazolen per Hand erstellt.^{98,103} Demnach sind Wasserstoffbrückenbindungen zu den Basen C39 und G26, und zum Phosphatrückgrat P_{U23} denkbar. Auch eine π - π - und/oder π -Kation-Wechselwirkungen zur Base A22 scheint möglich. Ein darauf aufbauendes Pharmakophormodell fand inaktive Analoga, deren fehlende Bioaktivität mit den vorhandenen Daten nicht erklärt werden konnte.

Da nur zwei aktive Nachkommen von **OB080d** gefunden wurden, war die Datengrundlage für weitere Aussagen nicht ausreichend. Ein automatisiertes

Docking in die Bindetasche der TAR-RNA-Struktur 1ARJ¹⁸⁷ konnte für zwei der drei Verbindungen einen der Bindehypothese entsprechenden Bindemodus finden.

In Zellkulturexperimenten wurde mittels einer Reportergenuntersuchung für **AS_S10/06_16** auch *in vivo* eine inhibitorische Aktivität der Tat-TAR-Interaktion gemessen. Bei hohen Konzentrationen der Substanz wurde die Inhibition jedoch wieder aufgehoben. Für diese vielversprechende Substanz sollten weiterführende Experimente durchgeführt werden. Das Nitroaren **AS_S10/06_33** erwies sich in Zellkulturexperimenten schon bei geringer Konzentration als cytotoxisch. Von Nitroarenen ist bekannt, dass sie problematische Stoffwechselprodukte erzeugen.¹⁶⁹ Es sollte daher in weiterführenden Experimenten ein geeigneter isosterer Ersatz für die Nitrogruppe gefunden werden.

7 Schlussfolgerungen und Ausblick

Ziel dieser Arbeit war die Durchführung und Evaluierung der adaptiven Optimierung fokussierter Substanzbibliotheken im Niedrigdurchsatzscreening. Als Besonderheit wurde eine Vorgehensweise gewählt, die in verschiedenen Punkten von der Natur inspiriert war. So wurden Prinzipien der Darwinschen Evolutionstheorie zum Vorbild eines mathematischen Optimierungsalgorithmus (der Evolutionsstrategie).^{1,2,38} Die Bewertung potentieller Molekülkandidaten wurde nicht wie sonst üblich einem mathematischen Modell überlassen, sondern es wurden mittleren inhibitorischen Konzentrationen (IC_{50}) experimentell in einem biochemischen Labor bestimmt.⁸⁸ Der hier gewählte Ansatz ist der Bionik entlehnt – er ist eine Hybridisierung eines technischen Verfahrens (hier: ein Computerprogramm) mit Lösungsstrategien, die sich auf natürliche Weise entwickelt haben.

Zur Vorbereitung der Optimierung einer fokussierten Bibliothek wurden im ersten Schritt verschiedene Optimierungsalgorithmen in einer Programmibibliothek implementiert. Die korrekte Funktionsweise der implementierten Algorithmen wurde mittels der Multidimensionalen Skalierung validiert. Die Multidimensionale Skalierung ist ein Projektionsverfahren hochdimensionaler Räume, für das sowohl analytisch-deterministische als auch stochastisch-heuristische Lösungswege existieren.¹⁴⁴ Die heuristische Lösung der Optimierungsalgorithmen wurde mit der analytischen Lösung der Hauptkoordinatenanalyse verglichen.¹⁴⁵ Die Algorithmen konnten erfolgreich validiert werden. Es zeigte sich, dass die Heuristiken qualitativ bessere Lösungen finden konnten als das deterministische Verfahren.

Nach erfolgter Validierung der Optimierungsalgorithmen wurden zur weiteren Vorbereitung vergleichende retrospektive Optimierungsexperimente durchgeführt. Die Optimierungsalgorithmen „Simulated Annealing“⁴²⁻⁴⁴, „Evolutionsstrategie“^{1,2,38} und „Partikelschwarmoptimierung“^{47,48} wurden mit einer Zufallssuche verglichen. Um Laborbedingungen zu simulieren, wurde die Gesamtzahl verfügbarer Substanztestungen auf 300 beschränkt. In den Optimierungsexperimenten wurden kleine fokussierte Bibliotheken erzeugt, welche mit Serinproteaseinhibitoren angereichert waren. Die Optimierungen wurden mit variierenden Populationsgrößen und Iterationszahlen durchgeführt. Es stellte sich heraus, dass Evolutionsstrategien effizient mit großen Populationen und wenigen Iterationen arbeiteten. Sie sind damit ausgezeichnet für einen Einsatz in einer Niedrigdurchsatzscreening-Kampagne geeignet. Jedes Individuum einer Population entsprach einem Molekül. Große Populationen sind also von Vorteil, da so

die Substanzen einer Generation zeitsparend gebündelt bezogen und getestet werden können. Partikelschwarmoptimierungen arbeiteten mit mittleren Populationsgrößen noch effizient und sind ebenfalls für die Optimierung fokussierter Substanzbibliotheken geeignet. Das Simulated Annealing war für den Entwurf fokussierter Bibliotheken im Niedrigdurchsatzscreening weniger geeignet, da es eine große Anzahl an Iterationen mit einer Populationsgröße von eins bevorzugte. Die Partikelschwärme wurden mit dieser Arbeit zum ersten Mal als Optimierungsalgorithmus für den Entwurf fokussierter Substanzbibliotheken eingeführt. Sie besitzen einige interessante Eigenschaften, von denen die herausragendste ihre große Anpassungsfähigkeit war. Je nach Problemstellung passte sich die Schrittweite der Partikelschwarmoptimierung autoadaptiv an. Der Einsatz von Partikelschwärmen in der adaptiven Optimierung fokussierter Bibliotheken sollte in der Zukunft als Standardwerkzeug in Betracht gezogen werden. Ihrer Präferenz zu größeren Iterationszahlen sollte mit geeigneten algorithmischen Veränderungen begegnet werden.

Für die prospektive Optimierung einer Substanzbibliothek wurde schließlich eine Evolutionsstrategie ausgewählt, bei der maximal zehn Eltern je Generation 60 Nachkommen erzeugen. Die Zahl der Eltern wurde in einer Eigenentwicklung adaptiv eingestellt, um die Tatsache zu berücksichtigen, dass Aktivitätsmessungen eine nach oben begrenzte Detektionsschwelle besitzen. Das Ziel der prospektiven Optimierung war der Entwurf einer fokussierten Substanzbibliothek von Liganden der TAR-RNA, welche zur Inhibierung deren Interaktion mit dem viralen Transaktivator der Transkription (*Tat*) befähigt sein sollten.⁶¹ In insgesamt vier Generationen wurden so 170 Screeningverbindungen auf Aktivität überprüft. Neun Substanzen konnten als aktive Inhibitoren der Tat-TAR-Interaktion identifiziert werden. Ihre mittleren inhibitorischen Konzentrationen (IC_{50}) lagen im mittleren mikromolaren Bereich ($IC_{50} = 506 \pm 497 \mu\text{M}$), wobei die zwei potentesten Kandidaten einen IC_{50} von 51 μM (**AS_S10/06_33**, [2-(4-Chloro-phenoxy)-ethyl]-(2-nitro-5-piperazin-1-yl-phenyl)-amin) und 116 μM (**AS_S10/06_16**, 2-(5-Amino-4-cyano-1*H*-pyrazol-3-yl)-*N*-hydroxy-acetamidin) aufwiesen.

In dieser Arbeit konnte zum ersten Mal die Anwendung der adaptiven Optimierung auf den Entwurf fokussierter Substanzbibliotheken von Ribonukleinsäure-Liganden gezeigt werden. Frühere Anwendungen beschränkten sich auf das Design von Proteinliganden.⁹³ Ribonukleinsäuren besitzen als pharmakologisches Angriffsziel besondere Eigenschaften. Ihre dreidimensionale Struktur ist ausgesprochen flexibel. Die Bildung einer Bindetasche wird teilweise erst durch den Liganden

selbst induziert.¹⁵²⁻¹⁵⁴ Zudem besitzt RNA andere Wechselwirkungseigenschaften als Proteine. π - π - und π -Kation-Wechselwirkungen spielen eine große Rolle und RNA-Moleküle sind Polyanionen.^{152,178} Rezeptorbasierte Verfahren sind hier deshalb nicht unproblematisch. In dieser Arbeit wurde unter anderem gezeigt, dass das automatisierte Docking mit der Software GOLD (The Cambridge Crystallographic Data Centre, Cambridge, Vereinigtes Königreich)¹¹⁸ nur sehr eingeschränkt auf RNA anwendbar ist. Ein ligandenbasierter Ansatz umgeht die Rezeptorproblematik. Hier werden bekannte Referenzverbindungen als Startpunkt einer Wirkstoffentwicklung verwendet. Es konnte jedoch in dieser Arbeit anhand einer Pharmakophorsuche gezeigt werden, dass auch bei ligandenbasierten Verfahren Probleme auftreten können. So wiesen die aktiven Analoga von **AS_S10/06_33** eine essentielle terminale Aminogruppe auf. Ein Austausch des Stickstoffatoms durch ein Sauerstoff- oder Kohlenstoffatom führte zum Verlust der Aktivität. Auch musste eine primäre oder sekundäre Aminogruppe vorliegen. Lag eine tertiäre Aminogruppe vor, so führte dies zum Verlust der Aktivität. Es wurde daher die Schlussfolgerung gezogen, dass eine tertiäre terminale Aminogruppe im lokalen Milieu der Bindetasche (TAR-Bulge) unprotoniert vorlag und sie damit ihre essentielle Eigenschaft als Protonendonator/Kation verlor. Das Programm „Molecular Operating Environment“ (MOE, Chemical Computing Group, Montreal, Kanada)⁹⁷ protoniert solche Aminogruppen standardmäßig. Die Erstellung eines Pharmakophormodells mit MOE gelang erst, als tertiäre Aminogruppen manuell deprotoniert wurden. Das in dieser Arbeit eingesetzte Verfahren kommt ohne solche Annahmen aus. Weder wird eine dreidimensionale Struktur des Zielmoleküls vorausgesetzt, noch müssen Annahmen über den Protonierungszustand einzelner Gruppen gemacht werden. Die Verwendung eines Laborexperiments als „Fitnessfunktion“ macht solche Annahmen überflüssig. Anstatt ein molekulares System zu simulieren, wurde direkt mit dem real existierenden System gearbeitet. Die adaptive Optimierung fokussierter Substanzbibliotheken mit laborexperimenteller Aktivitätsbestimmung ist deshalb besonders bei schlecht definierten pharmakologischen Zielen attraktiv, bei denen detailliertes Vorwissen rar ist. So ist bei vielen G-Protein-gekoppelten Rezeptoren die genaue dreidimensionale Struktur unbekannt.⁷ Andererseits kennt man aus Genomikprojekten Rezeptoren, für die noch kein bekannter Ligand existiert (verwaiste Rezeptoren, engl. „Orphan receptors“).⁷ So könnte die adaptive Optimierung auch bei der „Deorphanisierung“ helfen. Es sein angemerkt, dass auch bei Verwendung eines Laborexperiments als „Fitnessfunktion“ die Qualität der Ergebnisse von der Qualität der Messmethode und der Durchführung der Experimente abhängt.

Die adaptive Optimierung von TAR-RNA-Liganden erzielte eine Trefferquote von 5,3%, wobei lediglich 0,089% aller 191.405 Screeningverbindungen (Anbieter: Specs, Delft, Niederlande)⁶² getestet wurden. Hochdurchsatzscreening-Kampagnen erreichen typischerweise Trefferquoten von 0,1% bis 5%.^{15-17,128} Im Vergleich zu einer HTS-Kampagne konnte so eine 1,1 – 53-fache Anreicherung erreicht werden. Bei einem ungefähren Preis von 10 € pro Verbindungen könnten so gleichzeitig rund zwei Millionen Euro im Vergleich zu einer HTS-Kampagne eingespart werden. Der IC_{50} der gefundenen Aktiven liegt im mittleren mikromolaren Bereich. Für HTS-Kampagnen gelten oft wesentlich niedrigere Aktivitätsschwellen (z.B. $\leq 10 \mu\text{M}$).^{23a} Ist das Zielmolekül ein Protein sind selbst subnanomolare Inhibitoren keine Seltenheit. Die Situation ist bei RNA-Liganden eine andere: Es gibt viele Beispiele von RNA-Liganden mit Bindungskonstanten im niedrigen mikromolaren Bereich.¹⁵² Darunter sind auch zugelassene Wirkstoffe, die klinisch eingesetzt werden (z.B. das Aminoglycosid-Antibiotikum Paromomycin, $K_d = 12,3 \mu\text{M}$).¹⁵² Wirkstoffartige, kleine organische Moleküle mit submikromolaren Bindeaffinitäten sind dagegen in der Unterzahl. Die in dieser Arbeit gefundenen TAR-RNA-Liganden sind keine Leitstrukturen. Es handelt sich jedoch um validierte Treffer – potentiellen Kandidaten für Leitstrukturen. HTS-Kampagnen liefern selten validierte Treffer und sind anfälliger für falsch positive Aussagen.¹⁵

Einen Hinweis auf die Eignung eines Treffers als Leitstruktur kann die Ligandeneffizienz (engl. „Ligand efficiency“) geben (Gl. 7-1):¹⁸⁸

$$\Delta g = \frac{\Delta G}{N} \quad (7-1)$$

mit der freien Bindungsenthalpie $\Delta G = -RT \ln K_d$ und der Anzahl der Nicht-Wasserstoffatome N . In der Praxis kann für einen relativen Vergleich von Liganden auch mit der mittleren inhibitorischen Konzentration IC_{50} anstatt mit der Dissoziationskonstanten K_d gerechnet werden.¹⁸⁸ Die Ligandeneffizienz gibt die Potenz eines Liganden pro Nicht-Wasserstoffatom, unabhängig von seiner Molekülgröße an ($\Delta g \approx \text{Potenz} / \text{Größe}$). Da das Molekulargewicht (MG) einer Leitstruktur in der Entwicklungsphase der Leitstrukturoptimierung durch Addition zusätzlicher Gruppen für gewöhnlich zunimmt, sollte nicht allein die Potenz eines Leitstrukturkandidaten, sondern auch seine Größe betrachtet werden.¹⁸⁸ Verbindung **AS_S10/06_16** ($N = 13$, MG = 180 Da) besitzt eine Ligandeneffizienz von $\Delta g = 0,42 \text{ kcal/mol}$. Für Verbindung **AS_S10/06_33** ($N = 26$, MG = 387 Da) ist

die Ligandeneffizienz hingegen nur $\Delta g = 0,23$ kcal/mol. Unter der Annahme, dass die Potenz während der Leitstrukturoptimierung linear mit dem Molekulargewicht zunimmt,¹⁸⁸ benötigt Verbindung **AS_S10/06_16** zum Erreichen eines IC_{50} von 10 nM 27 Nicht-Wasserstoffatome. Hopkins *et al.* (2004) geben das durchschnittliche Molekulargewicht eines Nicht-Wasserstoffatoms einer wirkstoffartigen Verbindung mit 13,286 Da an.¹⁸⁸ 27 Nicht-Wasserstoffatome entsprechen demnach ca. 359 Da. Verbindung **AS_S10/06_33** benötigt zum Erreichen eines IC_{50} von 10 nM 49 Nicht-Wasserstoffatome. Dies entspräche einem Molekulargewicht von ca. 651 Da, was die Lipinski-Regel wirkstoffartiger Moleküle überschreitet ($MG \leq 500$ Da).¹²⁵ Trotz der geringeren Potenz ist Verbindung **AS_S10/06_16** demnach wegen ihrer höheren Ligandeneffizienz dennoch potentiell der bessere Leitstrukturkandidat.

Da die Suche viele aktive und inaktive Analoga der Verbindung **AS_S10/06_33** hervorbrachte, konnte für diese Gruppe von Substanzen (Nitrophenylamine) eine schlüssige SAR-Analyse durchgeführt werden. Wahrscheinlich ist eine π - π -Wechselwirkung des Benzolrings mit den Basen U25 und U40 der TAR-Bulge und eine Wasserstoffbrückenbindung und/oder elektrostatische Wechselwirkung einer terminalen Aminogruppe mit dem Phosphatrückgrat P_{C39} . **AS_S10/06_33** wirkte jedoch in einem Zellkulturexperiment schon bei geringer Konzentration toxisch. Für die Toxizität der Substanz könnte ihre Nitrogruppe verantwortlich sein.¹⁶⁷ In weiterführenden Experimenten sollte die Nitrogruppe deshalb durch eine weniger problematische Gruppe ersetzt werden. Dieser bioisostere Ersatz könnte unter den drei Gesichtspunkten eines ähnlichen Volumens, ähnlicher Pharmakophoreigenschaften und ähnlicher elektronischer Eigenschaften durchgeführt werden. Eine Nitrogruppe ($-\text{NO}_2$) besitzt ein Van-der-Waals-Volumen von rund 32 \AA^3 (berechnet mit MOE 2007.09)⁹⁷, ist hydrophob oder besitzt eine schwache Protonenakzeptoreigenschaft¹⁸⁹ und besitzt einen stark elektronenziehenden Charakter ($-I$ - und $-M$ -Effekt)¹⁷⁵. Für den isosteren Ersatz käme z.B. eine Carboxylatgruppe ($-\text{COO}^-$, 41 \AA^3 , Protonenakzeptor, stark e^- -ziehend), eine Trifluormethangruppe ($-\text{CF}_3$, 45 \AA^3 , hydrophob/schwacher Protonenakzeptor, stark e^- -ziehend), ein Halogenid ($-\text{X}$, $10\text{-}42 \text{ \AA}^3$, hydrophob/schwacher Protonenakzeptor, schwach e^- -ziehend), ein Methylester ($-\text{COOCH}_3$, 72 \AA^3 , Protonenakzeptor, stark e^- -ziehend) oder eine Cyanogruppe ($-\text{CN}$, 36 \AA^3 , hydrophob/schwacher Protonenakzeptor, stark e^- -ziehend) in Frage.^{8,175} George *et al.* berichteten 2007 von einem erfolgreichen bioisosteren Ersatz der Nitrogruppen eines Dinitroannelins.¹⁹⁰ Dinitroanneline sind Tubulin bindende gegen Protozoenparasiten wirksame Herbizide. Unter den getesteten funktionellen Gruppen (Amino-, Chlorid-, Cyano-, Methylester-, Carboxyl-, Amid- und Acetyl-Gruppe) wiesen die Chloro- und

Cyanoderivate eine ähnliche oder leicht verbesserte Bindungskonstante auf. Es wurde diskutiert, dass Nitrogruppen bei der Bindung an Tubulin sowohl elektrostatische Wechselwirkungen eingehen als auch als schwacher Protonenakzeptor fungieren. Eine solche Funktion der Nitrogruppe wäre auch für die aktiven Analoga von **AS_S10/06_33** denkbar. Eine Literatursuche nach **AS_S10/06_33** (durchgeführt mit SciFinder Scholar 2007, zugegriffen am 14. Juli 2008)¹⁹¹ ergab zwei Treffer. In einer Publikation von Tasler *et al.* werden Nitrophenylpiperazine als Serotoninrezeptorliganden (5-HT₆-Rezeptor) mit submikromolaren Bindungskonstanten vorgestellt. Die zweite Publikation (eine Patentschrift) beschäftigt sich ebenfalls mit Nitrophenylpiperazinen als 5-HT₆-Rezeptorliganden.¹⁹² Eine Erwähnung im Zusammenhang mit HIV-1 TAR-RNA konnte nicht gefunden werden. Die Verbindung stellt somit einen neuartigen TAR-RNA-Liganden dar.

Im evolutionären Stammbaum von **AS_S10/06_16** konnte abgesehen von der bekannten Elternstruktur der Initialpopulation nur eine weitere aktive Verbindung gefunden werden. Auf Grund der geringen Anzahl aktiver Verbindungen in diesem Zweig der evolutionären Suche war eine SAR-Analyse mit automatisierten chemieinformatischen Methoden kaum möglich. Es wurden deshalb Bindemodelle anhand einer vorhandenen Bindehypothese zu 1*H*-Pyrazolen per Hand erstellt.^{98,103} Demnach sind Wasserstoffbrückenbindungen zu den Basen C39 und G26, und zum Phosphatrückgrad P_{U23} denkbar. Auch eine π - π - und/oder π -Kation-Wechselwirkungen zur Base A22 scheint möglich. In einem Zellkulturexperiment konnte die antivirale Wirkung von **AS_S10/06_16** für niedrige Konzentrationen bestätigt werden. Bei hohen Konzentrationen wurde die inhibitorische Wirkung jedoch wieder aufgehoben. Hier sollte die Aufklärung der Struktur-Wirkungs-Beziehungen durch weitere Experimente voran gebracht werden. Eine Literatursuche nach **AS_S10/06_16** (durchgeführt mit SciFinder Scholar 2007, zugegriffen am 14. Juli 2008)¹⁹¹ ergab keinen Treffer. Es handelt sich also um einen neuartigen RNA-Liganden.

Das Verfahren der adaptiven Optimierung fokussierter Substanzbibliotheken erwies sich als äußerst robust. Trotz sorgfältiger Auswahl der Moleküle der Initialpopulation musste retrospektiv festgestellt werden, dass nur vier von zehn Substanzen einwandfreie Tat-TAR-Interaktionsinhibitoren waren. Vier Verbindungen waren *potentiell falsch positiv* und zwei Verbindungen gesichert *falsch positiv*. Obgleich der eingeschränkten Qualität der Initialpopulation konnte die Optimierung dennoch mit Erfolg durchgeführt werden. Interessanterweise lieferten zwei als potentiell falsch positiv charakterisierte Elternverbindungen die ersten aktiven Nachkommen der ersten Generation. Dies

könnte zum einen zur weiteren Diskussion der Aktivität der Elternstrukturen beitragen. Zum anderen vermittelte in einem Fall ein deutliches Grundgerüstspringen (von einem Phenothiazin (**OB108**) auf ein Nitrophenylpiperazin-Derivat (**AS_S10/06_33**)) den Erfolg. In den experimentellen Substanztestungen zeichneten sich 39% aller Molekülkandidaten durch eine mangelhafte Löslichkeit aus. Diese Substanzen konnten entweder gar nicht oder nur mit Schwierigkeiten auf Aktivität getestet werden. Für die Evolutionsstrategie bedeutete dies pro Generation einen Ausfall von 39% der Individuen. Dass die Optimierung dennoch erfolgreich durchgeführt werden konnte, spricht für die Robustheit des Verfahrens. Des Weiteren waren in der ersten Generation Nitrogruppen auf Grund eines Programmfehlers falsch repräsentiert. Es wurde daher nötig den gesamten Suchraum auszutauschen und durch einen um eine Dimension erweiterten Raum zu ersetzen. Auch dieser Eingriff in das Verfahren wurde toleriert. IC_{50} -Werte waren teilweise schwer zu bestimmen oder konnten zunächst nur ungenau abgeschätzt werden. Durch den Einsatz einer adaptiven Elternanzahl konnte auch diese Unzulänglichkeit aufgefangen werden. Tatsächlich hatte sich in retrospektiven Experimenten gezeigt, dass auch mit binären Aktivitätswerten („aktiv“ vs. „inaktiv“) effizient optimiert werden kann.

Perspektivisch könnten die Ergebnisse dieser Arbeit bei der Diskussion und Weiterverfolgung von Treffern aus HTS-Kampagnen helfen.¹³⁰ Auf dem Gebiet der Chemogenomik könnte das Verfahren bei der Findung von Liganden verwaister Rezeptoren und Rezeptoren ohne bekannte dreidimensionale Struktur angewendet werden.⁷ Im Bereich der RNA-Forschung konnten schon jetzt neue, wirkstoffartige RNA-Liganden entdeckt werden.

8 Anhang

8.1 Liste Molekularer Deskriptoren

Die Software „Molecular Operation Environment“ (*MOE*, Chemical Computing Group, Montreal, Kanada) hält 184 zweidimensionale Deskriptoren bereit.⁹⁷ Diese unterteilen sich in die Klassen „physikalische Eigenschaften“, „Teiloberflächen“, „Anzahl der Atome und Bindungen“, „Kier&Hall Konnektivität und Kappa-Gestaltindizes“, „Adjazent- und Distanzmatrixdeskriptoren“, „Pharmakophoreigenschaften“ und „Partiellladungsdeskriptoren“.

Deskriptoren physikalischer Eigenschaften

Die folgenden physikalischen Eigenschaften werden aus der Verbindungstabelle (engl. „Connection table“) unabhängig von der Konformation eines Moleküls berechnet.

Deskriptor	Beschreibung
apol	Summe der atomaren Polarisierbarkeit nach ¹⁹⁴ .
bpol	Summe des Absolutwertes der Differenz der atomaren Polarisierbarkeit aller gebundenen Atome untereinander nach ¹⁹⁴ .
density	Molekulare Massendichte: Molekulargewicht dividiert durch das Van-der-Waals-Volumen [amu/Å ³].
FCharge	Gesamtladung eines Moleküls (Summe aller formaler Ladungen).
mr	Molekulare Refraktivität nach ¹⁹⁵ .
SMR	Molekulare Refraktivität nach ¹⁹⁶ .
Weight	Molekulargewicht [Da] nach ¹⁹⁴ .
logP(o/w)	Logarithmus des Oktanol-Wasser-Verteilungskoeffizienten nach ¹⁶³ .
logS	Logarithmus der wässrigen Löslichkeit [mol/L] nach ¹⁶² .
reactive	Indikator für die Anwesenheit reaktiver Gruppen nach ¹²⁶ .
SlogP	Logarithmus des Oktanol-Wasser-Verteilungskoeffizienten nach ¹⁹⁶ .
TPSA	Polare Oberfläche nach ¹⁹⁷ .
vdw_vol	Van-der-Waals-Volumen [Å ³].
vdw_area	Van-der-Waals-Oberfläche [Å ²].

Teiloberflächendesriptoren

Die Teiloberflächendesriptoren basieren auf einer Approximierung der Wasserzugänglichen Van-der-Waals-Oberfläche [\AA^2] für jedes Atom v_i und den Eigenschaften L_i (SlogP)¹⁹⁶ und R_i (Molekulare Refraktivität)¹⁹⁶.

Deskriptor	Beschreibung
SlogP_VSA0	Summe von v_i für die gilt: $L_i \leq -0,4$.
SlogP_VSA1	Summe von v_i für die gilt: L_i in $(-0,4; -0,2]$.
SlogP_VSA2	Summe von v_i für die gilt: L_i in $(-0,2; 0]$.
SlogP_VSA3	Summe von v_i für die gilt: L_i in $(0; 0,1]$.
SlogP_VSA4	Summe von v_i für die gilt: L_i in $(0,1; 0,15]$.
SlogP_VSA5	Summe von v_i für die gilt: L_i in $(0,15; 0,20]$.
SlogP_VSA6	Summe von v_i für die gilt: L_i in $(0,20; 0,25]$.
SlogP_VSA7	Summe von v_i für die gilt: L_i in $(0,25; 0,30]$.
SlogP_VSA8	Summe von v_i für die gilt: L_i in $(0,30; 0,40]$.
SlogP_VSA9	Summe von v_i für die gilt: $L_i > 0,40$.
SMR_VSA0	Summe von v_i für die gilt: R_i in $[0; 0,11]$.
SMR_VSA1	Summe von v_i für die gilt: R_i in $(0,11; 0,26]$.
SMR_VSA2	Summe von v_i für die gilt: R_i in $(0,26; 0,35]$.
SMR_VSA3	Summe von v_i für die gilt: R_i in $(0,35; 0,39]$.
SMR_VSA4	Summe von v_i für die gilt: R_i in $(0,39; 0,44]$.
SMR_VSA5	Summe von v_i für die gilt: R_i in $(0,44; 0,485]$.
SMR_VSA6	Summe von v_i für die gilt: R_i in $(0,485; 0,56]$.
SMR_VSA7	Summe von v_i für die gilt: $R_i > 0,56$.

Deskriptoren der Atom- und Bindungszahlen

Deskriptor	Beschreibung
a_aro	Anzahl aromatischer Atome.
a_count	Anzahl der Atome (einschließlich impliziter Wasserstoffe).
a_heavy	Anzahl Nicht-Wasserstoffatome.
a_ICM	Mittlerer Informationsinhalt (Entropie) eines Atoms.
a_IC	Informationsinhalt (Entropie) eines Atoms: a_ICM multipliziert mit n.
a_nH	Anzahl der Wasserstoffatome (einschließlich impliziter).
a_nB	Anzahl der Boratome.
a_nC	Anzahl der Kohlenstoffatome.
a_nN	Anzahl der Stickstoffatome.
a_nO	Anzahl der Sauerstoffatome.
a_nF	Anzahl der Fluoratome.
a_nP	Anzahl der Phosphoratome.
a_nS	Anzahl der Schwefelatome.
a_nCl	Anzahl der Chloratome.
a_nBr	Anzahl der Bromatome.
a_nI	Anzahl der Iodatome.
b_1rotN	Anzahl der rotierbarer Einfachbindungen (ohne konjugierte Bindungen).
b_1rotR	Anteil rotierbarer Einfachbindungen: b_1rotN dividiert durch b_heavy.
b_ar	Anzahl der aromatischen Bindungen.
b_count	Anzahl der Bindungen.
b_double	Anzahl der Doppelbindungen (ohne aromatische Bindungen).
b_heavy	Anzahl der Bindungen zwischen Nicht-Wasserstoffatomen.
b_rotN	Anzahl der rotierbaren Bindungen
b_rotR	Anzahl der rotierbaren Bindungen: b_rotN dividiert durch b_heavy.
b_single	Anzahl der Einfachbindungen (ohne aromatische Bindungen).
b_triple	Anzahl der Dreifachbindungen.
chiral	Anzahl der Chiralitätszentren.
chiral_u	Anzahl der unbeschränkten Chiralitätszentren.
lip_acc	Anzahl der O- und N-Atome.
lip_don	Anzahl der OH- und NH-Atome.
lip_druglike	Eins wenn lip_violation < 2 sonst null.
lip_violation	Anzahl der Übertretungen der Lipinski-Regel „Rule of Five“. ¹²⁵
nmol	Anzahl der Moleküle
opr_brigid	Anzahl der rigiden Bindungen. ¹²⁶
opr_leadlike	Eins wenn opr_violation < 2 sonst null.
opr_nring	Anzahl der Ringbindungen nach ¹²⁶ .
opr_nrot	Anzahl der rotierbaren Bindungen nach ¹²⁶ .
opr_violation	Anzahl der Übertretungen des Oprea-Tests auf Wirkstoffartigkeit. ¹²⁶
rings	Anzahl der Ringe.
VAdjMa	Vertex-Adjazenzinformation (Stärke).
VAdjEq	Vertex-Adjazenzinformation (Gleichheit).

Deskriptoren der Kier&Hall-Konnektivität und der Kappa-Gestaltindizes

Der Chi-Konnektivitätsindex nach Kier und Hall kodiert Informationen über die Anzahl und Adjazenz von gebundenen Atomen, Bindungen und Zwei-Bindungs-Fragmenten. Die Kappa-Gestaltindizes nach Kier und Hall liefern auf Grundlage des molekularen Graphen eine Aussage über die Form und Gestalt eines Moleküls.¹⁹⁸

Deskriptor	Beschreibung
chi0	Atomarer Konnektivitätsindex nullter Ordnung nach ¹⁹⁸ und ¹⁹⁹ .
chi0_C	Kohlenstoff-Konnektivitätsindex nullter Ordnung.
chi1	Atomarer Konnektivitätsindex erster Ordnung nach ¹⁹⁸ und ¹⁹⁹ .
chi1_C	Kohlenstoff-Konnektivitätsindex erster Ordnung.
chi0v	Atomarer Valenz-Konnektivitätsindex nullter Ordnung nach ¹⁹⁸ und ¹⁹⁹ .
chi0v_C	Kohlenstoff-Valenz-Konnektivitätsindex nullter Ordnung.
chi1v	Atomarer Valenz-Konnektivitätsindex erster Ordnung nach ¹⁹⁸ und ¹⁹⁹ .
chi1v_C	Kohlenstoff-Valenz-Konnektivitätsindex erster Ordnung.
Kier1	Erster Kappa-Gestaltindex nach ¹⁹⁸ .
Kier2	Zweiter Kappa-Gestaltindex nach ¹⁹⁸ .
Kier3	Dritter Kappa-Gestaltindex nach ¹⁹⁸ .
KierA1	Erster alpha-modifizierter Kappa-Gestaltindex nach ¹⁹⁸ .
KierA2	Zweiter alpha-modifizierter Kappa-Gestaltindex nach ¹⁹⁸ .
KierA3	Dritter alpha-modifizierter Kappa-Gestaltindex nach ¹⁹⁸ .
KierFlex	Kier-Index der molekularen Flexibilität nach ¹⁹⁸ .
zagreb	Zagreb-Index (Summe der quadrierten Anzahl direkt gebundener Nicht-Wasserstoffatommachbarn über alle Atome eines Moleküls).

Adjazenz- und Distanzmatrix-Deskriptoren

Die Adjazenzmatrix M einer chemischen Struktur ist definiert als alle Elemente $[M_{ij}]$, wobei M_{ij} eins ist, wenn Atome i und j direkt miteinander verbunden sind, null anderenfalls.

Die Distanzmatrix D einer chemischen Struktur ist definiert als alle Elemente $[D_{ij}]$, wobei D_{ij} die Länge des kürzesten topologischen Pfades von Atom i zu j ist.

Deskriptor	Beschreibung
balabanJ	Balabanscher topologischer Konnektivitätsindex nach ²⁰⁰ .
BCUT_PEOE_0 BCUT_PEOE_1 BCUT_PEOE_2 BCUT_PEOE_3	Die BCUT-Deskriptoren werden aus den Eigenwerten einer modifizierten Adjazenzmatrix berechnet. ¹²⁴ Jeder Eintrag ij der Adjazenzmatrix nimmt den Wert $1/\text{Wurzel}(b_{ij})$ an, wobei b_{ij} die formale Bindungsordnung der Bindung von i zu j ist. Die Diagonale nimmt Werte der PEOE-Partialladungen an. Die resultierenden Eigenwerte werden sortiert und der kleinste, der 1/3-ste, der 2/3-ste und der größte werden zurückgegeben.
BCUT_SLOGP_0 BCUT_SLOGP_1 BCUT_SLOGP_2 BCUT_SLOGP_3	Die BCUT-Deskriptoren unter Verwendung der SlogP-Werte anstatt der Partialladung.
BCUT_SMR_0 BCUT_SMR_1 BCUT_SMR_2 BCUT_SMR_3	Die BCUT-Deskriptoren unter Verwendung der SMR-Werte (molare Refraktivität) anstatt der Partialladung.
diameter	Größter Wert der Distanzmatrix. ²⁰¹
petitjean	Wert (diameter - radius) / diameter.
GCUT_PEOE_0 GCUT_PEOE_1 GCUT_PEOE_2 GCUT_PEOE_3	Die GCUT-Deskriptoren werden aus den Eigenwerten einer modifizierten topologischen Distanz-Adjazenzmatrix berechnet. Jeder Eintrag ij der Adjazenzmatrix nimmt den Wert $1/\text{Wurzel}(d_{ij})$ an, wobei d_{ij} die (modifizierte) topologische Distanz zwischen den Atomen i und j ist. Die Diagonale nimmt Werte der PEOE-Partialladungen an. Die resultierenden Eigenwerte werden sortiert und der kleinste, der 1/3-ste, der 2/3-ste und der größte werden zurückgegeben.
GCUT_SLOGP_0 GCUT_SLOGP_1 GCUT_SLOGP_2 GCUT_SLOGP_3	Die GCUT-Deskriptoren unter Verwendung der SlogP-Werte anstatt der Partialladung.
GCUT_SMR_0 GCUT_SMR_1 GCUT_SMR_2 GCUT_SMR_3	Die GCUT-Deskriptoren unter Verwendung der SMR-Werte (molare Refraktivität) anstatt der Partialladung.
petitjeanSC	Petitjeanscher Graphengestaltkoeffizient nach ²⁰¹ : (diameter - radius) / radius.
radius	Wenn r_i der größte Matrixeintrag in Reihe i der Distanzmatrix D ist, dann ist der Radius als der kleinste Wert von r_i definiert. ²⁰¹
VDistEq	Wenn m die Summe der Distanzmatrixeinträge ist, dann ist VDistEq definiert als die Summe von $\log_2 m - p_i \log_2 p_i / m$, wobei p_i die Anzahl der Distanzmatrixeinträge mit dem Wert i ist.
VDistMa	Wenn m die Summe der Distanzmatrixeinträge ist, dann ist VDistMa definiert als die Summe von $\log_2 m - D_{ij} \log_2 D_{ij} / m$ über alle i und j .
weinerPath	Wiener-Pfadzahl: Die Hälfte der Summe über alle Distanzmatrixeinträge nach ^{202,203} .
weinerPol	Wiener-Polaritätszahl: Die Hälfte der Summe über alle Distanzmatrixeinträge mit einem Wert von 3. ²⁰²

Deskriptoren der Pharmakophoreigenschaften

In einem regelbasierten Ansatz werden Atomen die Pharmakophoreigenschaften Donor, Akzeptor, polar (Donor oder Akzeptor), positiv (basisch), negativ (azide), hydrophob und sonstige zugeordnet. Es wird von einem physiologischen pH-Wert ausgegangen.

Deskriptor	Beschreibung
a_acc	Anzahl der Wasserstoffbrückenakzeptoren (ausschließlich aziden Atomen aber einschließlich kombinierten Donoren und Akzeptoren wie –OH).
a_acid	Anzahl der aziden Atome.
a_base	Anzahl der basischen Atome.
a_don	Anzahl der Wasserstoffbrückendonoren (ausschließlich basischen Atomen aber einschließlich kombinierten Donoren und Akzeptoren wie –OH).
a_hyd	Anzahl der hydrophoben Atome.
vsa_acc	Approximierung der Summe der VDW-Oberflächen [\AA^2] von reinen Wasserstoffbrückenakzeptoren (ausschließlich aziden Atomen und kombinierten Donoren und Akzeptoren wie –OH).
vsa_acid	Approximierung der Summe der VDW-Oberflächen von aziden Atomen [\AA^2].
vsa_base	Approximierung der Summe der VDW-Oberflächen von basischen Atomen [\AA^2].
vsa_don	Approximierung der Summe der VDW-Oberflächen [\AA^2] von reinen Wasserstoffbrückendonoren (ausschließlich basischen Atomen und kombinierten Donoren und Akzeptoren wie –OH).
vsa_hyd	Approximierung der Summe der VDW-Oberflächen von hydrophoben Atomen [\AA^2].
vsa_other	Approximierung der Summe der VDW-Oberflächen von „sonstigen“ Atomen [\AA^2].
vsa_pol	Approximierung der Summe der VDW-Oberflächen [\AA^2] von polaren Atomen (kombinierten Donoren und Akzeptoren wie –OH).

Partialladungsdeskriptoren

Es werden zwei verschiedene Arten von Partialladungsdeskriptoren angeboten. Die PEOP-Deskriptoren basieren auf der Methode der partiellen Angleichung von Orbitalelektronegativitäten (*PEOE*, für engl. „Partial Equalization of Orbital Electronegativities“) von Gasteiger.²⁰⁴ Bei dieser Methode werden Ladungen zwischen gebundenen Atomen transferiert bis ein Gleichgewicht erreicht ist. Die Q-Deskriptoren berechnen keine Partialladungen sondern übernehmen die im Voraus berechneten Partialladungen. In dieser Arbeit wurde das MMFF94x-Kraftfeld (vgl. Abschnitt 3.3.5) zur Berechnung von Partialladungen verwendet. In den nachfolgenden Deskriptorbeschreibungen bezieht sich q_i auf die Partialladung von Atom i und v_i auf die Van-der-Waals-Oberfläche [\AA^2] von Atom i .

Deskriptor	Beschreibung
Q_PC+ PEOE_PC+	Positive Gesamtpartialladung: Summe aller positiver q_i . PC+ ist mit Q_PC+ identisch und wurde aus Kompatibilitätsgründen beibehalten.
Q_PC- PEOE_PC-	Negative Gesamtpartialladung: Summe aller negativer q_i . PC- ist mit Q_PC- identisch und wurde aus Kompatibilitätsgründen beibehalten.
Q_RPC+ PEOE_RPC+	Relative positive Partialladung: Das größte positive q_i dividiert durch die Summe positiver q_i . RPC+ ist mit Q_RPC+ identisch und wurde aus Kompatibilitätsgründen beibehalten.
Q_PRC- PEOE_PRC-	Relative negative Partialladung: Das größte negative q_i dividiert durch die Summe negativer q_i . PRC- ist mit Q_PRC- identisch und wurde aus Kompatibilitätsgründen beibehalten.
Q_VSA_POS PEOE_VSA_POS	Positive Gesamt-Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i nicht negativ ist.
Q_VSA_NEG PEOE_VSA_NEG	Negative Gesamt-Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i negativ ist.
Q_VSA_PPOS PEOE_VSA_PPOS	Positive polare Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i größer 0,2 ist.
Q_VSA_PNEG PEOE_VSA_PNEG	Negative polare Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i kleiner -0,2 ist.
Q_VSA_HYD PEOE_VSA_HYD	Hydrophobe Gesamt-Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass $ q_i $ kleiner oder gleich 0,2 ist.
Q_VSA_POL PEOE_VSA_POL	Polare Gesamt-Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass $ q_i $ größer 0,2 ist.
Q_VSA_FPOS PEOE_VSA_FPOS	Anteilige positive Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i nicht negativ ist, dividiert durch die gesamte Oberfläche.
Q_VSA_FNEG PEOE_VSA_FNEG	Anteilige negative Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i negativ ist, dividiert durch die gesamte Oberfläche.
Q_VSA_FPPOS PEOE_VSA_FPPOS	Anteilige positive polare Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i größer 0,2 ist, dividiert durch die gesamte Oberfläche.
Q_VSA_FPNEG PEOE_VSA_FPNEG	Anteilige positive polare Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass q_i kleiner -0,2 ist, dividiert durch die gesamte Oberfläche.
Q_VSA_FHYD PEOE_VSA_FHYD	Anteilige hydrophobe Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass $ q_i $ kleiner oder gleich 0,2 ist, dividiert durch die gesamte Oberfläche.
Q_VSA_FPOL PEOE_VSA_FPOL	Anteilige polare Van-der-Waals-Oberfläche. Dies ist die Summe aller v_i für die gilt, dass $ q_i $ größer 0,2 ist, dividiert durch die gesamte Oberfläche.
PEOE_VSA+6	Summe aller v_i für die gilt, dass q_i größer als 0,3 ist.
PEOE_VSA+5	Summe aller v_i für die gilt, dass q_i in [0,25; 0,30) ist.
PEOE_VSA+4	Summe aller v_i für die gilt, dass q_i in [0,20; 0,25) ist.
PEOE_VSA+3	Summe aller v_i für die gilt, dass q_i in [0,15; 0,20) ist.
PEOE_VSA+2	Summe aller v_i für die gilt, dass q_i in [0,10; 0,15) ist.
PEOE_VSA+1	Summe aller v_i für die gilt, dass q_i in [0,05; 0,10) ist.
PEOE_VSA+0	Summe aller v_i für die gilt, dass q_i in [0,00; 0,05) ist.
PEOE_VSA-0	Summe aller v_i für die gilt, dass q_i in [-0,05; 0,00) ist.
PEOE_VSA-1	Summe aller v_i für die gilt, dass q_i in [-0,10; -0,05) ist.
PEOE_VSA-2	Summe aller v_i für die gilt, dass q_i in [-0,15; -0,10) ist.
PEOE_VSA-3	Summe aller v_i für die gilt, dass q_i in [-0,20; -0,15) ist.
PEOE_VSA-4	Summe aller v_i für die gilt, dass q_i in [-0,25; -0,20) ist.
PEOE_VSA-5	Summe aller v_i für die gilt, dass q_i in [-0,30; -0,25) ist.
PEOE_VSA-6	Summe aller v_i für die gilt, dass q_i kleiner als -0,30 ist.

8.2 Multidimensionale Skalierung (mathematische Einführung)

Es folgt eine mathematische Beschreibung der Multidimensionalenskalierung nach Basalaj 2001.¹⁴⁴

Gegeben seien n Objekte und ein Verfahren zur Berechnung einer Unähnlichkeitsmatrix $\Delta = [\delta_{ij} : i, j = 1, \dots, n]$. Gesucht ist eine optimale Konfiguration $\mathbf{X}^* = [x_{ia}^* : a = 1, \dots, p]$ von n Punkten in einem p -dimensionalen Raum (gewöhnlich ein euklidischer Raum), sodass der Punkt $\mathbf{x}_i^* = (x_{i1}^*, \dots, x_{ip}^*)^T$ das i -te Objekt eindeutig repräsentiert und die euklidische Distanz (Gl. 8-1)

$$d_{ij}(\mathbf{X}^*) = \|\bar{\mathbf{x}}_i^* - \bar{\mathbf{x}}_j^*\| = \sqrt{\sum_{a=1}^p (x_{ia}^* - x_{ja}^*)^2} \quad (8-1)$$

zwischen den Punkten $\bar{\mathbf{x}}_i^*$ und $\bar{\mathbf{x}}_j^*$ die korrespondierende Unähnlichkeit δ_{ij} für alle Paare der Objekte (i, j) approximiert (Gl. 8-2):

$$\forall_{i < j} d_{ij}(\mathbf{X}^*) \approx \delta_{ij} \quad (8-2)$$

Es wird von einer symmetrischen Unähnlichkeitsmatrix ausgegangen, weshalb es genügt jedes Objektpaar genau einmal zu untersuchen: $i < j$.

Wir definieren den Fehler der Approximierung der Unähnlichkeit zwischen den Objekten i und j für eine gegebene Konfiguration \mathbf{X} als (Gl. 8-3)

$$e_{ij} := |d_{ij}(\mathbf{X}) - \delta_{ij}| \quad (8-3)$$

Optimierungsverfahren bedienen sich einer Verlustfunktion, die die Summe über die Fehler aller Objektpaare (i, j) bildet und somit den Gesamtapproximierungsfehler bestraft. Eine Optimierung findet dann im nächsten Schritt das Minimum dieser Funktion für \mathbf{X} und liefert die gewünschte optimale Konfiguration \mathbf{X}^* .

In der vorliegenden Anwendung wurde eine simple Stressfunktion als Verlustfunktion verwendet, die die Summe der Fehlerquadrate bildet (Gl. 8-4):

$$STRESS := \sum_{i < j} (d_{ij}(\mathbf{X}) - \delta_{ij})^2 \quad (8-4)$$

Eine analytische Lösung des MDS-Problems erfolgt mittels der Hauptkoordinatenanalyse (PCO, für engl. „Principal Coordinate Analysis“).¹⁴⁵ Bei der PCO wird die Distanzmatrix zunächst geeignet transformiert und dann einer Eigenzerlegung unterzogen. Aus Eigenvektoren und Eigenwerten können schließlich die rekonstruierten Koordinaten berechnet werden.

Sei $\mathbf{A} = [a_{ij} : i, j = 1, \dots, n] = [-\frac{1}{2}\delta_{ij}^2]$, dann berechnet sich die zentrierte Matrix \mathbf{B} mit Hilfe der Gowerschen Zentriermatrix $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ als (Gl. 8-5)

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H} \quad (8-5)$$

mit der Identitätsmatrix \mathbf{I} und dem Spaltenvektor $\mathbf{1}$ der Länge n , bestehend aus Einsen. Es erfolgt eine Eigenzerlegung nach (Gl. 8-6)

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \quad (8-6)$$

mit der diagonalen Matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ der Eigenwerte und den Eigenvektoren als Spalten der Matrix \mathbf{V} . Die Matrix \mathbf{X}^* der rekonstruierten Koordinaten erhält man als (Gl. 8-7)

$$\mathbf{X}^* = \mathbf{V}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \quad (8-7)$$

im k -dimensionalen Projektionsraum.

8.3 Nachkommen der prospektiven Optimierungsstudie

Tabelle 8-1: Liste aller 180 Nachkommen der prospektiven Optimierungsstudie. Die Specs-Nr. bezieht sich jeweils auf den Nachkommen. Die Aktivität wurde in einem FRET-Experiment zur Inhibierung der Tat-TAR-RNA-Interaktion bestimmt. σ gibt den Wert der adaptiven Schrittweiten an, die zur Erzeugung des Nachkommens verwendet wurde. Die Distanz ist von Elter zu Nachkommen mit der Manhattan-Metrik in einer beliebigen Einheit gemessen. Die Ordinalzahl gibt an, der wievielte Nachbar der Nachkomme aus der Sicht seines Elters ist.

Genera- tion	Elter	Nachkomme	Specs-Nr.	IC_{50} [μ M]	σ	Distanz	Ordinal- zahl
1	GS (red.)	AS_S10/06_17	AQ-360/42570321	inaktiv	0,65	16,6	84
1	GS (red.)	AS_S10/06_36	AO-081/15570003	inaktiv	0,65	13,4	6
1	GS (red.)	AS_S10/06_39	AH-262/34227046	inaktiv	0,65	13,8	8
1	GS (red.)	AS_S10/06_47	AM-814/41095109	inaktiv	0,65	13,6	7
1	GS (red.)	AS_S10/06_66	AE-848/30538013	inaktiv	0,65	13,8	10
1	GS (red.)	AS_S10/06_74	AB-323/13887384	inaktiv	0,65	16,5	80
1	IG2	AS_S10/06_07	AN-465/42834738	inaktiv	0,38	13,9	7
1	IG2	AS_S10/06_15	AP-970/42166925	inaktiv	0,65	16,0	82
1	IG2	AS_S10/06_21	AF-399/41900017	inaktiv	0,38	13,7	6
1	IG2	AS_S10/06_49	AN-652/40072255	inaktiv	0,65	14,5	12
1	IG2	AS_S10/06_62	AE-641/00357001	inaktiv	0,65	13,0	2
1	IG2	AS_S10/06_72	AE-641/09369053	inaktiv	0,38	15,0	24
1	M27	AS_S10/06_31	AG-690/11669205	inaktiv	0,65	29,7	5
1	M27	AS_S10/06_32	AG-690/13153170	inaktiv	0,38	28,2	4
1	M27	AS_S10/06_41	AG-205/07460034	inaktiv	0,65	32,8	8
1	M27	AS_S10/06_58	AK-249/10569004	inaktiv	0,38	27,9	3
1	M27	AS_S10/06_63	AE-848/32718027	inaktiv	0,38	25,3	1
1	M27	AS_S10/06_64	AG-205/07459022	inaktiv	0,65	27,5	2
1	M32	AS_S10/06_04	AQ-917/42754483	inaktiv	0,65	20,4	38
1	M32	AS_S10/06_10	AK-968/13150007	inaktiv	0,65	20,5	39
1	M32	AS_S10/06_23	AE-848/06135019	inaktiv	0,38	17,3	2
1	M32	AS_S10/06_27	AE-534/30057010	inaktiv	0,65	21,7	92
1	M32	AS_S10/06_29	AO-364/42192154	inaktiv	0,38	15,4	1
1	M32	AS_S10/06_56	AG-690/12765083	inaktiv	0,65	18,5	7
1	OB080d	AS_S10/06_02	AE-641/11703121	inaktiv	0,38	27,4	3

1	OB080d	AS_S10/06_16	AN-989/40886562	116 µM	0,38	27,1	2
1	OB080d	AS_S10/06_44	AH-262/36085032	inaktiv	0,38	30,1	30
1	OB080d	AS_S10/06_50	AH-034/11366558	inaktiv	0,38	30,3	35
1	OB080d	AS_S10/06_54	AK-968/40223609	inaktiv	0,38	28,1	8
1	OB080d	AS_S10/06_65	AE-641/00347016	inaktiv	0,38	27,9	6
1	OB108	AS_S10/06_05	AN-465/42885938	inaktiv	0,65	18,1	33
1	OB108	AS_S10/06_14	AN-465/42886203	inaktiv	0,65	17,7	24
1	OB108	AS_S10/06_33	AG-690/15428920	51 µM	0,38	19,7	123
1	OB108	AS_S10/06_38	AF-399/42016971	inaktiv	0,38	17,2	16
1	OB108	AS_S10/06_51	AG-690/12869331	inaktiv	0,65	18,4	46
1	OB108	AS_S10/06_70	AP-906/42853956	inaktiv	0,38	14,2	2
1	rbt550	AS_S10/06_01	AE-641/00601025	inaktiv	0,38	15,3	2
1	rbt550	AS_S10/06_12	AK-968/41925657	inaktiv	0,65	22,6	998
1	rbt550	AS_S10/06_22	AN-689/41741413	inaktiv	0,65	20,0	177
1	rbt550	AS_S10/06_24	AN-465/41988090	inaktiv	0,38	17,0	16
1	rbt550	AS_S10/06_61	AG-690/33352020	inaktiv	0,38	15,7	4
1	rbt550	AS_S10/06_67	AN-465/14964068	inaktiv	0,65	19,5	111
1	S10	AS_S10/06_06	AH-487/42305586	inaktiv	0,65	10,4	12
1	S10	AS_S10/06_25	AK-968/41021147	inaktiv	0,65	17,0	5547
1	S10	AS_S10/06_26	AE-848/12530194	inaktiv	0,38	10,7	27
1	S10	AS_S10/06_55	AG-690/34151063	inaktiv	0,65	10,4	14
1	S10	AS_S10/06_60	AE-848/36048058	inaktiv	0,38	12,9	305
1	S10	AS_S10/06_69	AG-205/32455053	inaktiv	0,65	11,4	60
1	WM5	AS_S10/06_08	AQ-360/42570608	inaktiv	0,65	13,5	4
1	WM5	AS_S10/06_34	AJ-292/14921084	inaktiv	0,65	13,4	3
1	WM5	AS_S10/06_43	AJ-333/36117060	inaktiv	0,65	19,8	2080
1	WM5	AS_S10/06_48	AG-670/11900035	inaktiv	0,65	17,2	269
1	WM5	AS_S10/06_52	AG-690/11971136	inaktiv	0,38	16,2	113
1	WM5	AS_S10/06_73	AG-205/36981015	inaktiv	0,38	13,5	5
1	Y9	AS_S10/06_03	AG-690/13704918	inaktiv	0,65	11,3	125
1	Y9	AS_S10/06_35	AN-919/14229134	inaktiv	0,38	10,8	78
1	Y9	AS_S10/06_37	AE-848/34077014	inaktiv	0,38	10,9	90
1	Y9	AS_S10/06_42	AN-919/14384249	inaktiv	0,38	8,3	7

8 Anhang

1	Y9	AS_S10/06_45	AH-262/34902003	inaktiv	0,65	11,7	182
1	Y9	AS_S10/06_46	AG-205/37224030	inaktiv	0,38	5,4	1
2	AS_S10/06_16	AS_S01/07_05	AB-323/13887392	inaktiv	0,50	26,1	1561
2	AS_S10/06_16	AS_S01/07_08	AE-641/30118009	inaktiv	0,50	24,7	754
2	AS_S10/06_16	AS_S01/07_10	AQ-405/42300525	inaktiv	0,50	23,0	285
2	AS_S10/06_16	AS_S01/07_13	AF-826/30391008	inaktiv	0,50	17,2	4
2	AS_S10/06_16	AS_S01/07_14	AF-886/30580032	inaktiv	0,50	23,6	416
2	AS_S10/06_16	AS_S01/07_16	AK-777/40971894	inaktiv	0,50	24,1	526
2	AS_S10/06_16	AS_S01/07_17	AP-835/42125444	inaktiv	0,30	20,9	71
2	AS_S10/06_16	AS_S01/07_22	AF-886/30565005	inaktiv	0,30	16,6	1
2	AS_S10/06_16	AS_S01/07_24	AM-807/12740417	inaktiv	0,30	18,2	9
2	AS_S10/06_16	AS_S01/07_31	AE-484/32882054	inaktiv	0,50	20,1	39
2	AS_S10/06_16	AS_S01/07_35	AG-670/40741579	823 µM	0,50	19,4	28
2	AS_S10/06_16	AS_S01/07_36	AN-988/14610001	inaktiv	0,50	18,1	7
2	AS_S10/06_16	AS_S01/07_40	AG-205/12367052	inaktiv	0,50	18,3	11
2	AS_S10/06_16	AS_S01/07_43	AG-690/15432239	inaktiv	0,30	21,7	127
2	AS_S10/06_16	AS_S01/07_49	AI-031/31971063	inaktiv	0,50	18,9	19
2	AS_S10/06_16	AS_S01/07_53	AG-690/13703170	inaktiv	0,30	24,7	785
2	AS_S10/06_16	AS_S01/07_54	AF-407/34872024	inaktiv	0,50	18,1	8
2	AS_S10/06_16	AS_S01/07_57	AN-329/40926269	inaktiv	0,50	17,4	5
2	AS_S10/06_16	AS_S01/07_58	AH-487/40652376	inaktiv	0,50	18,3	12
2	AS_S10/06_16	AS_S01/07_59	AN-023/15594003	inaktiv	0,50	16,9	2
2	AS_S10/06_33	AS_S01/07_01	AK-968/12688125	inaktiv	0,30	13,2	42
2	AS_S10/06_33	AS_S01/07_02	AK-968/13150067	inaktiv	0,30	11,8	8
2	AS_S10/06_33	AS_S01/07_03	AG-670/40809147	inaktiv	0,30	9,6	2
2	AS_S10/06_33	AS_S01/07_04	AG-670/15543018	inaktiv	0,50	13,3	46
2	AS_S10/06_33	AS_S01/07_07	AG-205/34689009	inaktiv	0,30	13,6	65
2	AS_S10/06_33	AS_S01/07_11	AG-670/40809138	inaktiv	0,50	12,3	15
2	AS_S10/06_33	AS_S01/07_12	AG-670/15543055	inaktiv	0,50	14,3	108
2	AS_S10/06_33	AS_S01/07_19	AG-670/40809134	269 µM	0,30	15,6	231
2	AS_S10/06_33	AS_S01/07_20	AG-690/15432141	inaktiv	0,30	10,8	4
2	AS_S10/06_33	AS_S01/07_21	AN-465/14013008	inaktiv	0,50	12,8	25
2	AS_S10/06_33	AS_S01/07_23	AG-205/07922019	inaktiv	0,30	13,3	49

2	AS_S10/06_33	AS_S01/07_37	AK-778/13520034	inaktiv	0,50	13,7	73
2	AS_S10/06_33	AS_S01/07_39	AG-690/11662059	inaktiv	0,50	14,4	115
2	AS_S10/06_33	AS_S01/07_45	AK-778/13520035	inaktiv	0,30	10,5	3
2	AS_S10/06_33	AS_S01/07_50	AN-465/42889773	inaktiv	0,30	13,3	45
2	AS_S10/06_33	AS_S01/07_52	AK-968/12096065	inaktiv	0,30	11,4	5
2	AS_S10/06_33	AS_S01/07_62	AG-670/12382002	493 µM	0,30	12,9	29
2	AS_S10/06_33	AS_S01/07_65	AG-670/40809108	inaktiv	0,30	12,0	12
2	AS_S10/06_33	AS_S01/07_69	AG-205/36565031	inaktiv	0,50	9,4	1
2	AS_S10/06_33	AS_S01/07_70	AJ-333/13050201	inaktiv	0,50	17,7	762
2	AS_S10/06_73	AS_S01/07_09	AQ-360/42570636	inaktiv	0,50	18,0	385
2	AS_S10/06_73	AS_S01/07_15	AG-690/37079192	inaktiv	0,30	13,4	6
2	AS_S10/06_73	AS_S01/07_18	AH-487/42143608	inaktiv	0,50	20,0	1617
2	AS_S10/06_73	AS_S01/07_25	AQ-360/41615677	inaktiv	0,30	14,7	27
2	AS_S10/06_73	AS_S01/07_26	AM-807/43303469	inaktiv	0,30	18,2	438
2	AS_S10/06_73	AS_S01/07_27	AF-399/40634498	inaktiv	0,50	16,6	133
2	AS_S10/06_73	AS_S01/07_28	AG-205/12143049	inaktiv	0,30	18,3	473
2	AS_S10/06_73	AS_S01/07_29	AG-690/37079063	inaktiv	0,50	15,3	45
2	AS_S10/06_73	AS_S01/07_33	AQ-360/42570609	inaktiv	0,50	16,1	81
2	AS_S10/06_73	AS_S01/07_34	AM-807/43303470	inaktiv	0,50	17,7	320
2	AS_S10/06_73	AS_S01/07_38	AK-778/12433206	inaktiv	0,30	14,0	14
2	AS_S10/06_73	AS_S01/07_41	AJ-292/41694950	inaktiv	0,30	12,9	2
2	AS_S10/06_73	AS_S01/07_42	AO-476/43362740	inaktiv	0,30	16,8	153
2	AS_S10/06_73	AS_S01/07_46	AE-848/12529051	inaktiv	0,30	14,7	26
2	AS_S10/06_73	AS_S01/07_48	AN-465/42886318	inaktiv	0,50	16,6	127
2	AS_S10/06_73	AS_S01/07_55	AD-310/37069025	inaktiv	0,50	13,8	12
2	AS_S10/06_73	AS_S01/07_61	AE-641/12753359	inaktiv	0,30	16,9	177
2	AS_S10/06_73	AS_S01/07_63	AR-434/42446816	inaktiv	0,30	13,0	3
2	AS_S10/06_73	AS_S01/07_67	AK-968/11368963	inaktiv	0,50	12,3	1
2	AS_S10/06_73	AS_S10/06_39	AH-262/34227046	inaktiv	0,30	13,7	11
3	AS_S01/07_19	AS_S01/07_69	AG-205/36565031	inaktiv	0,38	12,3	4
3	AS_S01/07_19	AS_S06/07_05	AG-690/15432854	inaktiv	0,23	13,5	15
3	AS_S01/07_19	AS_S06/07_07	AG-670/15543046	inaktiv	0,38	13,1	11
3	AS_S01/07_19	AS_S06/07_13	AK-968/13150207	inaktiv	0,23	13,1	12

8 Anhang

3	AS_S01/07_19	AS_S06/07_16	AG-690/15432853	inaktiv	0,23	12,2	3
3	AS_S01/07_19	AS_S06/07_24	AN-465/13571005	inaktiv	0,38	15,3	50
3	AS_S01/07_19	AS_S06/07_34	AG-205/36981049	inaktiv	0,38	15,1	41
3	AS_S01/07_19	AS_S06/07_38	AG-205/37047129	inaktiv	0,38	12,9	5
3	AS_S01/07_19	AS_S06/07_39	AG-205/33674043	inaktiv	0,23	11,0	2
3	AS_S01/07_19	AS_S06/07_40	AK-968/40319755	115 µM	0,38	13,3	13
3	AS_S01/07_19	AS_S06/07_42	AG-670/40809143	inaktiv	0,23	10,8	1
3	AS_S01/07_19	AS_S06/07_44	AG-670/40809135	inaktiv	0,38	13,7	20
3	AS_S01/07_19	AS_S06/07_51	AQ-360/43173505	inaktiv	0,38	16,8	166
3	AS_S01/07_19	nicht lieferbar	AG-690/11763354		0,38	13,0	9
3	AS_S01/07_19	nicht lieferbar	AN-465/14458059		0,38	14,5	29
3	AS_S01/07_35	AS_S01/07_24	AM-807/12740417	inaktiv	0,38	12,9	8
3	AS_S01/07_35	AS_S01/07_40	AG-205/12367052	inaktiv	0,65	8,3	1
3	AS_S01/07_35	AS_S01/07_43	AG-690/15432239	inaktiv	0,38	14,2	15
3	AS_S01/07_35	AS_S01/07_59	AN-023/15594003	inaktiv	0,65	12,3	6
3	AS_S01/07_35	AS_S06/07_01	AF-399/15284214	inaktiv	0,38	13,3	9
3	AS_S01/07_35	AS_S06/07_10	AP-048/15613037	inaktiv	0,38	11,2	2
3	AS_S01/07_35	AS_S06/07_12	AR-422/41547520	inaktiv	0,65	13,5	11
3	AS_S01/07_35	AS_S06/07_18	AI-204/33265012	inaktiv	0,38	15,4	44
3	AS_S01/07_35	AS_S06/07_19	AE-484/32882045	inaktiv	0,65	20,6	1856
3	AS_S01/07_35	AS_S06/07_20	AH-262/15226028	inaktiv	0,65	13,9	12
3	AS_S01/07_35	AS_S06/07_31	AC-907/34129010	inaktiv	0,38	11,8	3
3	AS_S01/07_35	AS_S06/07_35	AG-690/36720011	inaktiv	0,65	18,9	640
3	AS_S01/07_35	AS_S06/07_36	AG-690/37012083	inaktiv	0,38	11,9	4
3	AS_S01/07_35	AS_S06/07_45	AN-329/41189552	inaktiv	0,65	16,6	114
3	AS_S01/07_35	AS_S06/07_50	AM-807/41929662	inaktiv	0,65	15,8	59
3	AS_S01/07_37	AS_S01/07_01	AK-968/12688125	inaktiv	0,38	7,4	1
3	AS_S01/07_37	AS_S06/07_02	AG-690/40125091	inaktiv	0,38	11,5	28
3	AS_S01/07_37	AS_S06/07_11	AG-670/15543050	inaktiv	0,65	13,5	89
3	AS_S01/07_37	AS_S06/07_14	AK-968/12269141	inaktiv	0,38	9,3	9
3	AS_S01/07_37	AS_S06/07_17	AG-205/04795027	inaktiv	0,65	16,6	455
3	AS_S01/07_37	AS_S06/07_22	AK-968/12861021	inaktiv	0,38	8,6	3
3	AS_S01/07_37	AS_S06/07_23	AK-968/12096029	inaktiv	0,38	7,5	2

3	AS_S01/07_37	AS_S06/07_25	AK-778/13520039	inaktiv	0,65	12,1	39
3	AS_S01/07_37	AS_S06/07_26	AG-690/13701889	inaktiv	0,38	11,6	30
3	AS_S01/07_37	AS_S06/07_32	AK-968/11841012	inaktiv	0,38	8,6	4
3	AS_S01/07_37	AS_S06/07_33	AK-968/41926613	inaktiv	0,65	18,4	1326
3	AS_S01/07_37	AS_S06/07_37	AG-205/33165011	inaktiv	0,38	11,6	31
3	AS_S01/07_37	AS_S06/07_43	AK-968/12345827	inaktiv	0,38	9,6	10
3	AS_S01/07_37	AS_S06/07_46	AK-968/12346606	inaktiv	0,38	9,0	6
3	AS_S01/07_37	AS_S06/07_48	AK-968/41925817	inaktiv	0,38	9,0	8
3	AS_S01/07_62	AS_S01/07_11	AG-670/40809138	inaktiv	0,23	10,7	12
3	AS_S01/07_62	AS_S06/07_03	AG-690/12781967	inaktiv	0,38	10,6	11
3	AS_S01/07_62	AS_S06/07_04	AG-690/12781913	inaktiv	0,38	13,4	94
3	AS_S01/07_62	AS_S06/07_06	AG-670/15543011	inaktiv	0,38	9,7	5
3	AS_S01/07_62	AS_S06/07_08	AG-670/15543023	inaktiv	0,23	8,9	3
3	AS_S01/07_62	AS_S06/07_09	AG-670/15543049	inaktiv	0,23	8,8	2
3	AS_S01/07_62	AS_S06/07_15	AG-690/09249007	inaktiv	0,38	10,2	6
3	AS_S01/07_62	AS_S06/07_21	AN-023/15004003	inaktiv	0,23	11,2	15
3	AS_S01/07_62	AS_S06/07_27	AG-670/12382010	(1114 μ M)	0,23	11,3	19
3	AS_S01/07_62	AS_S06/07_28	AG-670/12382031	(1409 μ M)	0,38	10,6	10
3	AS_S01/07_62	AS_S06/07_29	AG-670/12382017	162 μ M	0,38	10,3	7
3	AS_S01/07_62	AS_S06/07_30	AG-670/12382030	inaktiv	0,38	6,7	1
3	AS_S01/07_62	AS_S06/07_41	AG-670/40809141	inaktiv	0,23	11,4	22
3	AS_S01/07_62	AS_S06/07_47	AG-690/12618161	inaktiv	0,23	11,4	21
3	AS_S01/07_62	AS_S06/07_49	AG-205/32540027	inaktiv	0,38	12,1	39

8.4 Dosis-Wirkungs-Kurven der aktiven Verbindungen

8.4.1 AS_S10/06_16

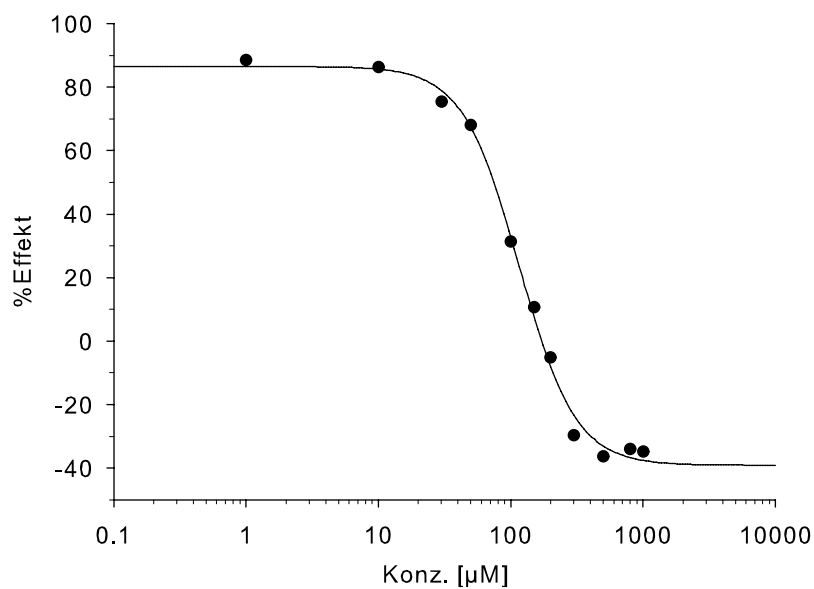


Abbildung 8-1: Dosis-Wirkungs-Kurve von AS_10/06_16.

Lösungsmittel:	H ₂ O
Ausreißer:	Messwert bei 0,1 μM ausgelassen
IC₅₀:	116 μM
95% Konfidenzintervall:	102 – 131 μM
R² der Kurvenanpassung:	0,9966

8.4.2 AS_S10/06_33

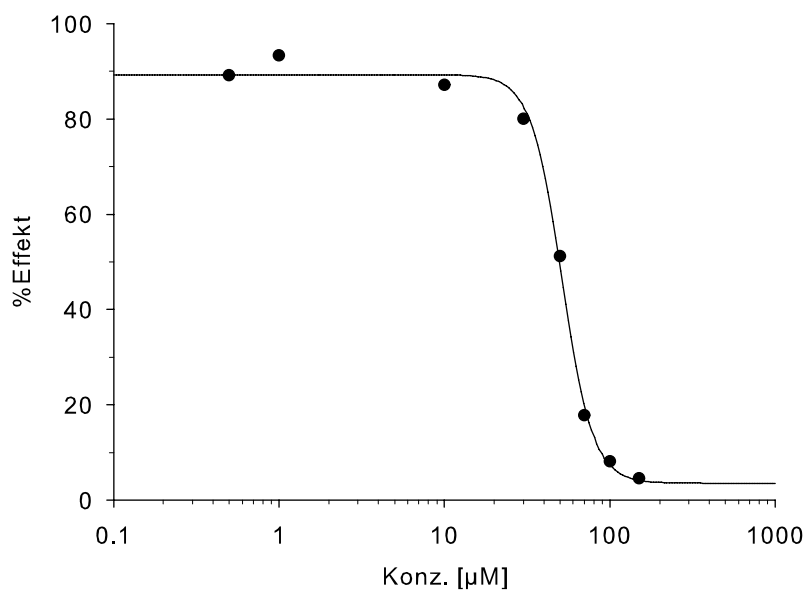


Abbildung 8-2: Dosis-Wirkungs-Kurve von AS_S10/06_33.

Lösungsmittel:	H ₂ O mit 1% DMSO und 9% MeOH
Ausreißer:	Messwerte bei 0,1; 200; 300 und 500 µM ausgelassen
Löslichkeit:	Unlöslich ab 200 µM
IC₅₀:	51 µM
95% Konfidenzintervall:	47 – 56 µM
R² der Kurvenanpassung:	0,9989

8.4.3 AS_S01/07_19

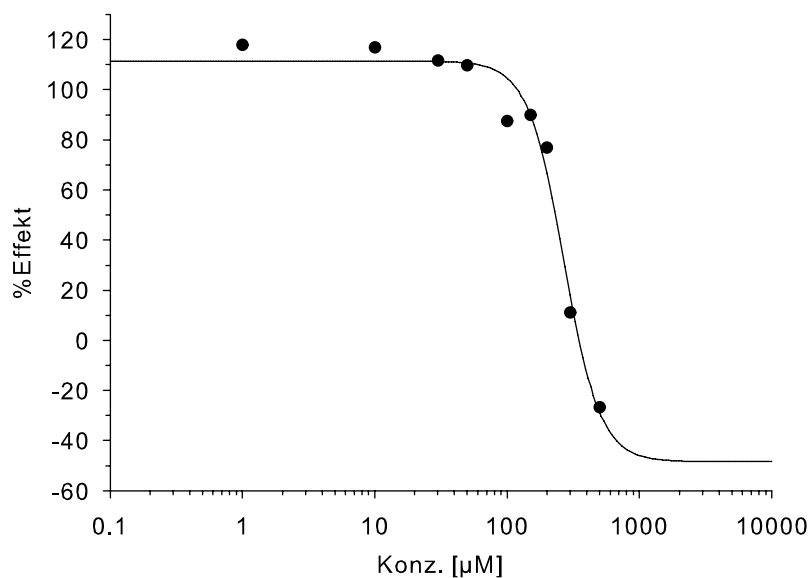


Abbildung 8-3: Dosis-Wirkungs-Kurve von AS_S01/07_19.

Lösungsmittel:	H ₂ O mit 1% DMSO und 10% MeOH
Ausreißer:	Messwerte bei 0,1; 800; und 1000 µM ausgelassen
Löslichkeit:	Unlöslich ab 800 µM
IC₅₀:	269 µM
95% Konfidenzintervall:	180 – 403 µM
R² der Kurvenanpassung:	0,9930
Anmerkung:	Minimaler Plateauwert nicht bestimmbar

8.4.4 AS_S01/07_35

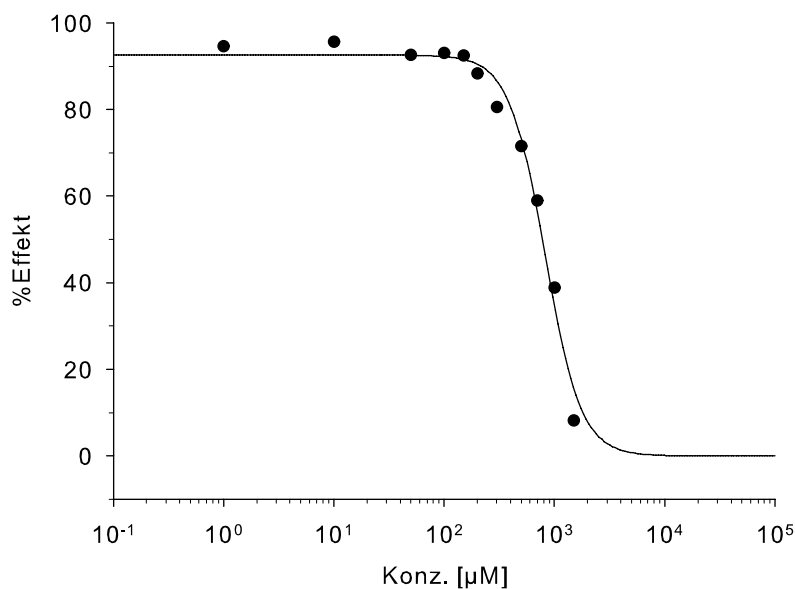


Abbildung 8-4: Dosis-Wirkungs-Kurve von AS_S01/07_35.

Lösungsmittel:	H ₂ O mit 5% DMSO und 10% MeOH
Ausreißer:	Messwert bei 0,1 µM ausgelassen
Löslichkeit:	Unlöslich ab 1500 µM
IC₅₀:	823 µM
95% Konfidenzintervall:	551 – 1230 µM
R² der Kurvenanpassung:	0,9979
Anmerkung:	Minimaler Plateaueffekt nicht bestimmbar. Kurvenanpassung mit D = 0% durchgeführt.

8.4.5 AS_S01/07_62

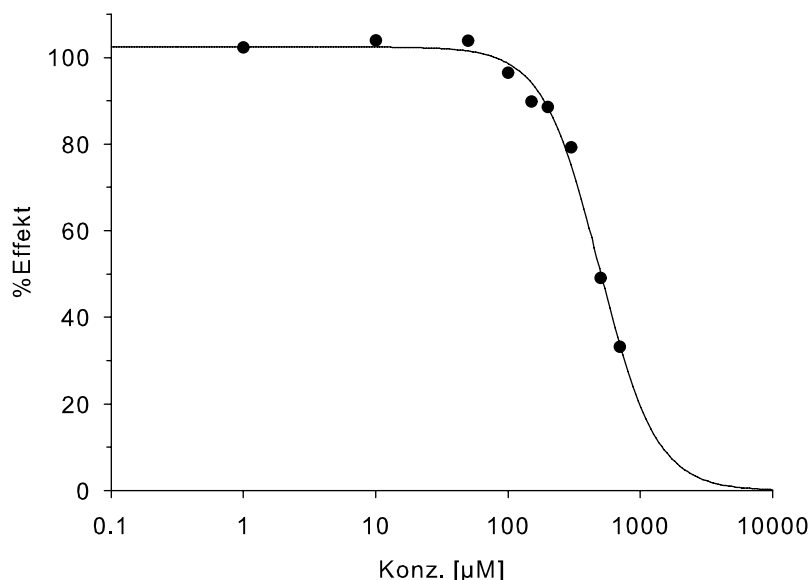


Abbildung 8-5: Dosis-Wirkungs-Kurve von AS_S01/07_62 (bei alternativen Bedingungen, s.h. Text).

Lösungsmittel:	H ₂ O mit 5% DMSO und 15% MeOH
Ausreißer:	Messwerte bei 0,1; 1000 und 1500 µM ausgelassen
Löslichkeit:	Unlöslich ab 1000 µM
IC₅₀:	329 µM
95% Konfidenzintervall:	182 – 593 µM
R² der Kurvenanpassung:	0,9969
Anmerkung:	Minimaler Plateaueffekt nicht bestimmbar. Die Substanz absorbiert im Anregungsbereich von 489 nm und verfälscht die Messungen daher hin zu einer optimistischeren Abschätzung der mittleren inhibitorischen Konzentration. Deshalb erfolgte auch eine Analyse bei alternativen Bedingungen (Anregungswellenlänge: 550 nm, Emissionswellenlänge: 570 nm):
IC₅₀^{Alt}:	493 µM
95%^{Alt} Konfidenzintervall:	218 – 1114 µM
R²_{Alt} der Kurvenanpassung:	0,9992.

8.4.6 AS_S06/07_27

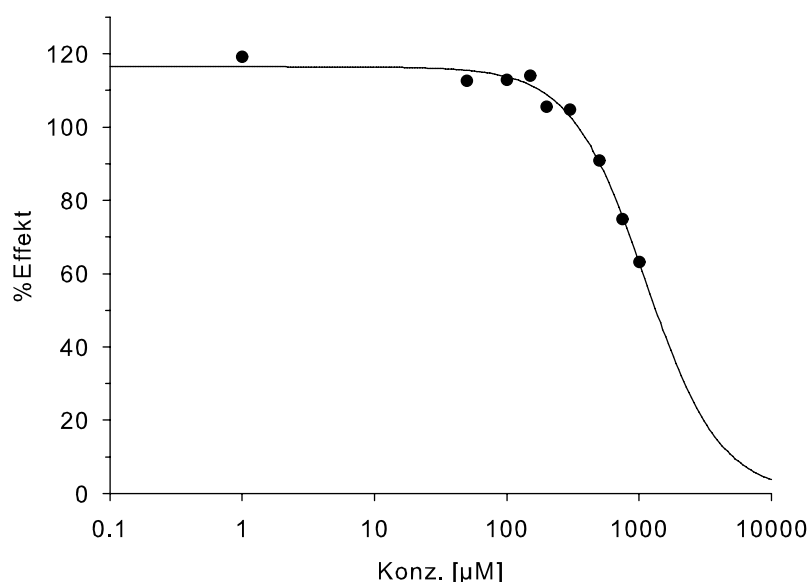


Abbildung 8-6: Dosis-Wirkungs-Kurve von AS_S06/07_27 (bei alternativen Bedingungen, s.h. Text).

Lösungsmittel:	H ₂ O mit 1,5% DMSO und 6% MeOH
Ausreißer:	Messwert bei 0,1 und 10 µM ausgelassen
IC₅₀:	484 µM
95% Konfidenzintervall:	294 – 795 µM
R² der Kurvenanpassung:	0,9988
Anmerkung:	Minimaler Plateauwert nicht bestimmbar. Die Substanz absorbiert im Anregungsbereich von 489 nm und verfälscht die Messungen daher hin zu einer optimistischeren Abschätzung der mittleren inhibitorischen Konzentration. Deshalb erfolgte auch eine Analyse bei alternativen Bedingungen (Anregungswellenlänge: 550 nm, Emissionswellenlänge: 570 nm):
IC₅₀^{Alt}:	1114 µM
95%^{Alt} Konfidenzintervall:	118 – 10 ⁴ µM
R²_{Alt} der Kurvenanpassung:	0,9996
Anmerkung:	Der IC ₅₀ -Wert ist wegen schwacher Aktivität nur als Richtwert zu verstehen.

8.4.7 AS_S06/07_28

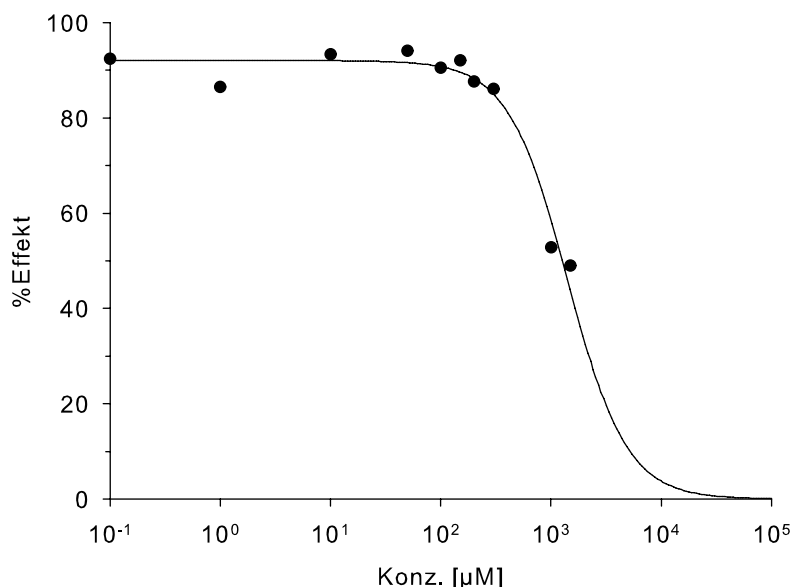


Abbildung 8-7: Dosis-Wirkungs-Kurve von AS_S06/07_28 (bei alternativen Bedingungen, s.h. Text).

Lösungsmittel:	H ₂ O mit 1,5% DMSO und 6% MeOH
Ausreißer:	Messwert bei 500 und 700 µM ausgelassen
IC₅₀:	529 µM
95%⁰ Konfidenzintervall:	389 – 719 µM
R² der Kurvenanpassung:	0,9964
Anmerkung:	Minimaler Plateaueffekt nicht bestimmbar. Die Substanz absorbiert im Anregungsbereich von 489 nm und verfälscht die Messungen daher hin zu einer optimistischeren Abschätzung der mittleren inhibitorischen Konzentration. Deshalb erfolgte auch eine Analyse bei alternativen Bedingungen (Anregungswellenlänge: 550 nm, Emissionswellenlänge: 570 nm):
IC₅₀^{Alt}:	1409 µM
95%^{Alt} Konfidenzintervall:	(1027 – 1933 µM)
R²_{Alt} der Kurvenanpassung:	(0,9983)
Anmerkung:	Die Kurvenanpassung war nur durch Hinzufügen eines artifiziellen Messpunktes bei Konz. = 10 ⁴ µM mit %Effekt = 0% möglich. Die Werte für das 95% ^{Alt} -Konfidenzintervall und R ² _{Alt} entsprechen daher nicht den tatsächlichen Daten, das Konfidenzintervall ist zu optimistisch. Der IC ₅₀ -Wert ist wegen schwacher Aktivität nur als Richtwert zu verstehen.

8.4.8 AS_S06/07_29

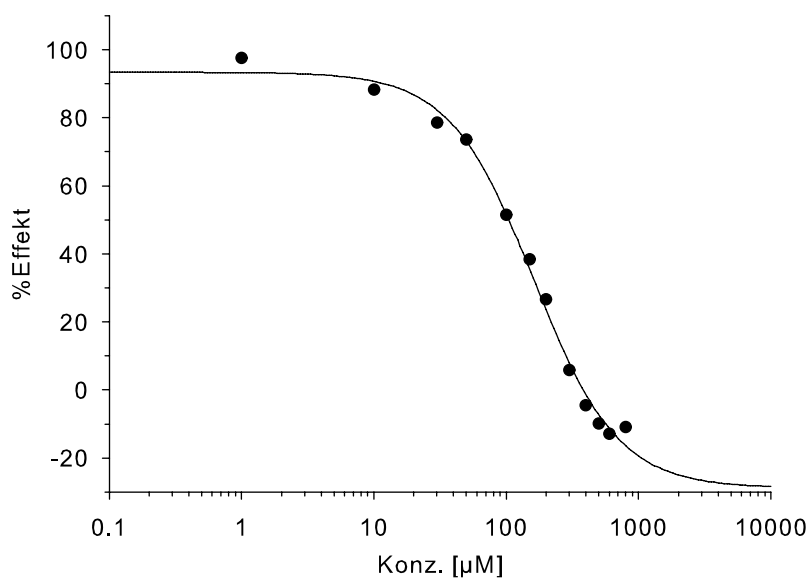


Abbildung 8-8: Dosis-Wirkungs-Kurve von AS_S06/07_29.

Lösungsmittel:	H ₂ O mit 0,8 % DMSO und 3,2 % MeOH
Löslichkeit:	Unlöslich ab 800 µM
IC₅₀:	162 µM
95% Konfidenzintervall:	125 – 210 µM
R² der Kurvenanpassung:	0,9967
Anmerkung:	Minimaler Plateau-effekt nicht bestimmbar.

8.4.9 AS_S06/07_40

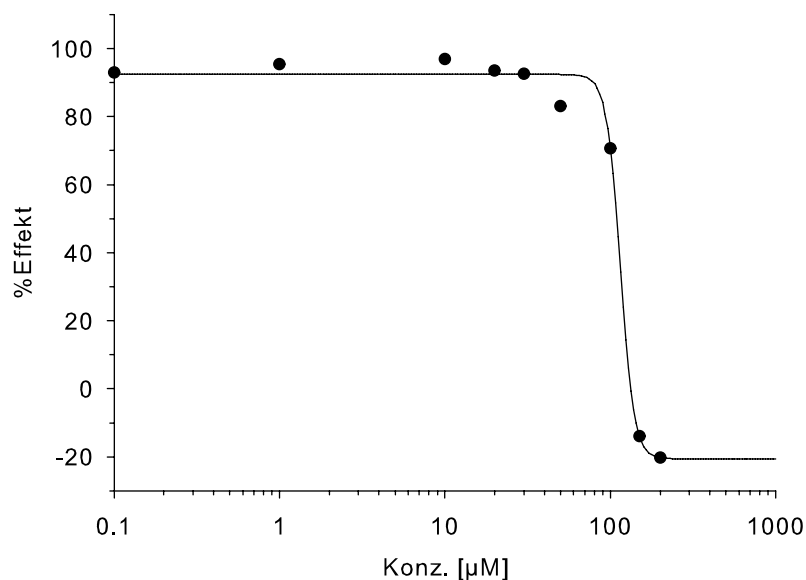


Abbildung 8-9: Dosis-Wirkungs-Kurve von AS_S06/07_40.

Lösungsmittel:	H ₂ O mit 10% DMSO
Ausreißer:	Messwerte bei 75, 300 und 400 µM ausgelassen. Nur zweifach Messung.
Löslichkeit:	Unlöslich ab 300 µM
IC₅₀:	115 µM
95% Konfidenzintervall:	102 – 129 µM
R² der Kurvenanpassung:	0,9980
Anmerkung:	Substanz war schlecht löslich und die Titration deshalb problematisch in der Durchführung.

tetrameres Hämoglobin durch Dimerbildung destabilisiert werden und Nitrosobenzol reagiert zudem mit Thiolgruppen von Cysteinen und kann so die Globinkette denaturieren.^{166,167,170} Diese hämopathischen Vorgänge bewirken eine Anreicherung von Methämoglobin im Blut was zu Methämoglobinämie und Zyanose führen kann.^{167,170}

In Körpergewebe wird Nitrobenzol durch Ein-Elektron-Reaktionen mikrosomal zu Anilin reduziert (Abbildung 8-10).^{167,169,170,205} Die dabei entstehenden freien Radikale Nitroanion, Hydnitroxid und ein Aminokation (postuliert) können bei ausreichend hohem Durchsatz der Nitrobenzolreduktion akkumulieren und oxidativen Stress auf angrenzende Gewebe ausüben. Zudem kann in Geweben mit ausreichender Sauerstoffversorgung das Nitroanion-Radikal durch Sauerstoff in einer „nutzlosen Reaktion“ oxidiert werden, wodurch gefährliche Menge an Superoxidradikalen freigesetzt werden. Bei dieser Reaktion wird außerdem Nitrobenzol regeneriert. Eine ganze Reihe der toxischen und karzinogenen Effekte von Nitrobenzol könnten durch diesen nutzlosen Zyklus und das dadurch permanent aufrechterhaltene Niveau an Nitroxidintermediaten und durch die bekannten karzinogenen Eigenschaften von Superoxidradikalen erklärt werden.^{167,169,170,205}

8.6 Versuchsbedingungen der Zellkulturexperimente

Die antivirale Wirkung einer Substanz gegen HIV-1 wird gemessen, indem Zellen bei verschiedenen Konzentrationen der zu bestimmenden Substanz mit HIV-1 infiziert werden und die Menge der Virusbildung relativ zur Kontrolle (in Abwesenheit der Substanz) bestimmt wird.¹⁰⁰ Es wurde die humane Zelllinie Hela P4 (transfiziert mit den HIV-1-Rezeptoren CD4, CCR5 und CXCR4) verwendet, die ein β -Galactosidasegen unter der Kontrolle eines HIV-1-LTR-Promotors aufweist. Ferner wurde der Laborstamm HIV-1_{Lai} für die Versuche benutzt.

$1,5 \times 10^4$ Zellen wurden in je eine Vertiefung einer 96er-Mikrotiterplatte in 100 μ L Medium gegeben und über Nacht bei 37°C in einem Zellinkubator inkubiert. Am nächsten Tag wurden die Zellen gewaschen und in Anwesenheit der Testsubstanzen mit 20 μ L der HIV-1_{Lai}-VirusstammLösung bei einem Gesamtvolumen von 100 μ L Medium infiziert. Die Testsubstanzen wurden bei verschiedenen Konzentrationen (25 μ M, 50 μ M, 100 μ M, 200 μ M, 250 μ M und 500 μ M) zugesetzt und bei 37°C inkubiert. Die Messungen erfolgten in dreifacher Ausführung. Nach zwei Tagen wurden die Zellen vorsichtig mit PBS (phosphatgepufferte Saline, für engl. „Phosphate-Buffered Saline“) gewaschen und anschließend lysiert. Ein Reaktionspuffer (enthielt Galacton, s.u.)

wurde hinzugefügt und für eine Stunde inkubiert. Zur Erzeugung des lumineszenten Signals wurden danach eine Verstärkerlösung (enthält Emerald, s.u.) dazu gegeben. Das Signal wurde 2 min. später mit einem Luminometer (Lumistar Galaxy, BMG Labtechnologies, Offenburg) quantifiziert.

Galacton®:

Applied Biosystems, Applera Deutschland GmbH, Frankfurter Straße 129 B, 64293 Darmstadt

Produkt-Nr. T2094

Chemilumineszentes Substrat für E. Coli- β -Galactosidase-Versuche. Zur Detektion von Femtogramm- bis Nanogramm-Spiegel des β -Galactosidase-Versuchs.

Chemischer Name: 3-(4-Methoxy Spiro {1,2-dioxetan-3,2'-(5'-chloro)-tricyclo-[3.3.1.1^{3,7}]decan}-4-yl)phenyl- β -D-galactopyranosid

Emerald™:

Applied Biosystems, Applera Deutschland GmbH, Frankfurter Straße 129 B, 64293 Darmstadt

Produkt-Nr. T2002

Lumineszenzverstärker für Versuche in Lösung, Emerald™ Enhancer. 10-fach-Konzentrat zur Verwendung mit chemilumineszenten Substraten (AMPPD®, CSPD®, Galacton®, Glucon™, oder Glucuron®).

9 Literaturverzeichnis

- (1) Rechenberg, I. *Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*; Frommann-Holzboog: Stuttgart, 1973.
- (2) Rechenberg, I. *Evolutionsstrategie '94*; Frommann-Holzboog: Stuttgart, 1994.
- (3) Chang, C.; Kwan, C. In *AI 2004: Advances in Artificial Intelligence*, 2005; S. 803-815.
- (4) Chatterjee, S.; Carrera, C.; Lynch, L. A. Genetic Algorithms and Traveling Salesman Problems. *Eur. J. Op. Res.* **1996**, *93*, 490-510.
- (5) Weber, L. Current Status of Virtual Combinatorial Library Design. *QSAR Comb. Sci.* **2005**, *24*, 809-823.
- (6) Schneider, G.; Lee, M. L.; Stahl, M.; Schneider, P. De Novo Design of Molecular Architectures by Evolutionary Assembly of Drug-Derived Building Blocks. *J. Comput.-Aided. Mol. Des.* **2000**, *14*, 487-494.
- (7) Schneider, G.; Baringhaus, K. *Molecular Design: Concepts and Applications for Beginners*; 1. Ed.; Wiley-Vch: Weinheim, 2007.
- (8) Böhm, H.; Klebe, G.; Kubinyi, H. *Wirkstoffdesign*; 1. Ed.; Spektrum Akademischer Verlag: Heidelberg, Berlin, Oxford, 2002.
- (9) Bleicher, K. H.; Böhm, H.; Müller, K.; Alanine, A. I. Hit and Lead Generation: Beyond High-Throughput Screening. *Nat. Rev. Drug. Discov.* **2003**, *2*, 369-378.
- (10) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16*, 3-50.
- (11) Lipinski, C.; Hopkins, A. Navigating Chemical Space for Biology and Medicine. *Nature* **2004**, *432*, 855-861.
- (12) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening - an Overview. *Drug Discov. Today* **1998**, *3*, 160-178.
- (13) Eglen, R. M.; Schneider, G.; Böhm, H. In *Virtual Screening for Bioactive Molecules*; Böhm, H.; Schneider, G., Ed.; Wiley-VCH: Weinheim, New York, 2000; Band 10, S. 1-14.
- (14) Drewry, D. H.; Young, S. S. Approaches to the Design of Combinatorial Libraries. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 1-20.
- (15) Bajorath, J. Integration of Virtual and High-Throughput Screening. *Nat. Rev. Drug. Discov.* **2002**, *1*, 882-894.
- (16) Miller, J. L. Recent Developments in Focused Library Design: Targeting Gene-Families. *Curr. Top. Med. Chem.* **2006**, *6*, 19-29.
- (17) Lahana, R. How Many Leads from HTS? *Drug Discov. Today* **1999**, *4*, 447-448.

- (18) Schneider, G. Trends in Virtual Combinatorial Library Design. *Curr. Med. Chem.* **2002**, *9*, 2095-2101.
- (19) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862-865.
- (20) Schneider, G.; Böhm, H. Virtual Screening and Fast Automated Docking Methods. *Drug Discov. Today* **2002**, *7*, 64-70.
- (21) Mason, J. S.; Good, A. C.; Martin, E. J. 3-D Pharmacophores in Drug Discovery. *Curr. Pharm. Des.* **2001**, *7*, 567-597.
- (22) Winkler, D. A. The Role of Quantitative Structure-Activity Relationships (QSAR) in Biomolecular Discovery. *Brief. Bioinf.* **2002**, *3*, 73-86.
- (23) Schneider, G.; Fechner, U. Computer-Based De Novo Design of Drug-Like Molecules. *Nat. Rev. Drug. Discov.* **2005**, *4*, 649-663.
- (23a) Congreve, M.; Murray C. W.; Blundell, T. L. Keynote review: Structural biology and drug discovery. *Drug Discov. Today* **2005**, *10*, 895-907.
- (24) Sheridan, R. P.; Kearsley, S. K. Using a Genetic Algorithm To Suggest Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 310-320.
- (25) Weber, L.; Wallbaum, S.; Gubernator, K.; Broger, C. Optimization of the Biological Activity of Combinatorial Compound Libraries by a Genetic Algorithm. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2280-2282.
- (26) Weber, L. Multi-Component Reactions and Evolutionary Chemistry. *Drug Discov. Today* **2002**, *7*, 143-147.
- (27) Illgen, K.; Enderle, T.; Broger, C.; Weber, L. Simulated Molecular Evolution in a Full Combinatorial Library. *Chem. Biol.* **2000**, *7*, 433-441.
- (28) Holland, J. H. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, 1975.
- (29) Brown, R. D.; Martin, Y. C. Designing Combinatorial Library Mixtures Using a Genetic Algorithm. *J. Med. Chem.* **1997**, *40*, 2304-2313.
- (30) Gillet, V. J.; Willett, P.; Bradshaw, J.; Green, D. V. S. Selecting Combinatorial Libraries to Optimize Diversity and Physical Properties. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 169-177.
- (31) Gillet, V. J.; Khatib, W.; Willett, P.; Fleming, P. J.; Green, D. V. S. Combinatorial Library Design Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 375-385.
- (32) Gillet, V. J.; Willett, P.; Fleming, P. J.; Green, D. V. S. Designing Focused Libraries Using MoSELECT. *J. Mol. Graph. Model.* **2002**, *20*, 491-498.
- (33) Jamois, E. A.; Lin, C. T.; Waldman, M. Design of Focused and Restrained Subsets from Extremely Large Virtual Libraries. *J. Mol. Graph. Model.* **2003**, *22*, 141-149.

- (34) Mandal, A.; Johnson, K.; Wu, C. F. J.; Bornemeier, D. Identifying Promising Compounds in Drug Discovery: Genetic Algorithms and Some New Statistical Techniques. *J. Chem. Inf. Model.* **2007**, *47*, 981-988.
- (35) Singh, J.; Ator, M. A.; Jaeger, E. P.; Allen, M. P.; Whipple, D. A.; Solowej, J. E.; Chowdhary, S.; Treasurywala, A. M. Application of Genetic Algorithms to Combinatorial Synthesis: A Computational Approach to Lead Identification and Lead Optimization. *J. Am. Chem. Soc.* **1996**, *118*, 1669-1676.
- (36) Yokobayashi, Y.; Ikebukuro, K.; McNiven, S.; Karube, I. Directed Evolution of Trypsin Inhibiting Peptides Using a Genetic Algorithm. *J. Chem. Soc., Perkin Trans. 1* **1996**, *20*, 2435-2437.
- (37) Shanmugasundaram, V.; Maggiora, G. M.; Lajiness, M. S. Hit-Directed Nearest-Neighbor Searching. *J. Med. Chem.* **2005**, *48*, 240-248.
- (38) Schwefel, H. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*; Interdisciplinary systems research; Birkhäuser Verlag: Basel, 1977; Band 26.
- (39) Schneider, G.; Wrede, P. The Rational Design of Amino Acid Sequences by Artificial Neural Networks and Simulated Molecular Evolution: De Novo Design of an Idealized Leader Peptidase Cleavage Site. *Biophys. J.* **1994**, *66*, 335-344.
- (40) Agrafiotis, D. K. Stochastic Algorithms for Maximizing Molecular Diversity. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 841-851.
- (41) Zheng, W.; Hung, S. T.; Saunders, J. T.; Seibel, G. L. PICCOLO: A Tool for Combinatorial Library Design Via Multicriterion Optimization. *Pac. Symp. Biocomput.* **2000**, 588-599.
- (42) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671-680.
- (43) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equations of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087-1092.
- (44) Salamon, P.; Sibani, P.; Frost, R. *Facts, Conjectures, and Improvements for Simulated Annealing (SLAM Monographs on Mathematical Modeling and Computation)*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 2002.
- (45) Agrafiotis, D. K.; Lobanov, V. S. Ultrafast Algorithm for Designing Focused Combinational Arrays. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1030-1038.
- (46) Le Bailly de Tillegem, C.; Beck, B.; Boulanger, B.; Govaerts, B. A Fast Exchange Algorithm for Designing Focused Libraries in Lead Optimization. *J. Chem. Inf. Model.* **2005**, *45*, 758-767.

- (47) Eberhart, R. C.; Kennedy, J. A New Optimizer Using Particle Swarm Theory. In *Proceedings of the Sixth International Symposium on Micromachine and Human Science*; Nagoya, Japan, 1995; S. 39-43.
- (48) Kennedy, J.; Eberhart, R. C. Particle Swarm Optimization. In *Proceedings of IEEE International Conference on Neural Networks*; Piscataway, NJ, 1995; S. 1942-1948.
- (49) Cedeño, W.; Agrafiotis, D. K. Using Particle Swarms for the Development of QSAR Models Based on K-Nearest Neighbor and Kernel Regression. *J. Comput.-Aided. Mol. Des.* **2003**, *17*, 255-263.
- (50) Shen, Q.; Jiang, J.; Jiao, C.; Huan, S.; Shen, G.; Yu, R. Optimized Partition of Minimum Spanning Tree for Piecewise Modeling by Particle Swarm Algorithm. QSAR Studies of Antagonism of Angiotensin II Antagonists. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2027-2031.
- (51) Shen, Q.; Jiang, J.; Jiao, C.; Lin, W.; Shen, G.; Yu, R. Hybridized Particle Swarm Algorithm for Adaptive Structure Training of Multilayer Feed-Forward Neural Network: QSAR Studies of Bioactivity of Organic Compounds. *J. Comput. Chem.* **2004**, *25*, 1726-1735.
- (52) Lin, W.; Jiang, J.; Shen, Q.; Shen, G.; Yu, R. Optimized Block-Wise Variable Combination by Particle Swarm Optimization for Partial Least Squares Modeling in Quantitative Structure-Activity Relationship Studies. *J. Chem. Inf. Model.* **2005**, *45*, 486-493.
- (53) Meissner, M.; Schneider, G. Protein Folding Simulation by Particle Swarm Optimization. *Open Struct. Biol. J.* **2007**, *1*, 1-6.
- (54) Meissner, M.; Schmuker, M.; Schneider, G. Optimized Particle Swarm Optimization (OPSO) and Its Application to Artificial Neural Network Training. *BMC Bioinformatics* **2006**, *7*, 125.
- (55) Rogers-Evans, M.; Alanine, A. I.; Bleicher, K. H.; Kube, D.; Schneider, G. Identification of Novel Cannabinoid Receptor Ligands Via Evolutionary De Novo Design and Rapid Parallel Synthesis. *QSAR Comb. Sci.* **2004**, *23*, 426-430.
- (56) Dracopoulos, D. C. *Evolutionary Learning Algorithms for Neural Adaptive Control*; Springer: London, 1997.
- (57) Koza, J. R. In *Dynamic, Genetic, and Chaotic Programming: The Sixth-Generation*; Soucek, B.; the IRIS Group, Ed.; John Wiley & Sons, Inc.: New York, 1992; S. 203-321.
- (58) Frank, S. A. In *Adaption*; Rose, M. R.; Lauder, G. V., Ed.; Academic Press: San Diego, 1996; S. 451-505.

- (59) *Java Platform, Standard Edition*, Version 1.5; Sun Microsystems, Inc.: 4150 Network Circle, Santa Clara, CA 95054, USA; <http://java.sun.com/>.
- (60) Karn, J. Tackling Tat. *J. Mol. Biol.* **1999**, *293*, 235-254.
- (61) Richter, S. N.; Palu, G. Inhibitors of HIV-1 Tat-Mediated Transactivation. *Curr. Med. Chem.* **2006**, *13*, 1305-1315.
- (62) *Specs-Screeningverbindungen*; Specs: Kluiverweg 6, 2629 HT Delft, Die Niederlande, 2006; <http://www.specs.net/>.
- (63) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemie*; 5. Ed.; Spektrum Akademischer Verlag: Heidelberg, Berlin, 2003.
- (64) Kashima, A.; Inoue, Y.; Sugio, S.; Maeda, I.; Nose, T.; Shimohigashi, Y. X-Ray Crystal Structure of a Dipeptide-Chymotrypsin Complex in an Inhibitory Interaction. *Eur. J. Biochem* **1998**, *255*, 12-23.
- (65) Alok, A.; Sinha, M.; Singh, N.; Sharma, S.; Kaur, P.; Singh, T. P. Crytal Structure of the Trypsin Complex with Benzamidine at High Temperature (35 C). *Noch zu veröffentlichen*.
- (66) Reiche, D. *Roche Lexikon Medizin*; 5. Ed.; Urban & Fischer: MÜNschen, Jena, 2003.
- (67) Krupiczojc, M. A.; Scotton, C. J.; Chambers, R. C. Coagulation Signalling Following Tissue Injury: Focus on the Role of Factor Xa. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 1228-1237.
- (68) Swiss Institute of Bioinformatics ExPASy Proteomics Server.
- (69) Hallgren, J.; Pejler, G. Biology of Mast Cell Tryptase. *FEBS J.* **2006**, *273*, 1871-1895.
- (70) Myöhänen, H.; Vaheri, A. Regulation and Interactions in the Activation of Cell-Associated Plasminogen. *Cell. Mol. Life Sci.* **2004**, *61*, 2840-2858.
- (71) Anderson, S.; Ryan, R. Inhibitors of Urokinase Plasminogen Activator.
- (72) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res* **2003**, *31*, 3784-8.
- (73) Weiss, R. A. How Does HIV Cause AIDS? *Science* **1993**, *260*, 1273-9.
- (74) AIDS Epidemic Update 2007. Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO), **2007**; <http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2007/default.asp>.
- (75) HIV/AIDS in Deutschland - Eckdaten. Robert Koch Institut, **2007**; www.rki.de/cn_048/nn_196014/DE/Content/InfAZ/H/HIVAIDS/hiv__aids__wat.html.

- (76) Simon, V.; Ho, D. D.; Abdool Karim, Q. HIV/AIDS Epidemiology, Pathogenesis, Prevention, and Treatment. *Lancet* **2006**, *368*, 489-504.
- (77) Mathers, C. D.; Loncar, D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Med.* **2006**, *3*, e442.
- (78) Epidemiologisches Bulletin: Halbjahresbericht I/2007. Robert Koch Institut, **2007**; http://www.rki.de/cln_048/nn_208878/DE/Content/InfAZ/H/HIVAIDS/Veroeffentlichungen/EpiBull/EpiBull__node.html?__nnn=true.
- (79) Hildebrand, V. In *Retrovirus Bulletin*; Nationales Referenzzentrum für Retroviren - Institut für Klinische und Molekulare Virologie: Erlangen, 2003; Band 02/2003.
- (80) Carter, J.; Saunders, V. *Virology: Principles and Applications*; 1. Ed.; WileyBlackwell: The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, 2007.
- (81) Lodish, H.; Berk, A.; Zipursky, S. L.; Darnell, J. *Molekulare Zellbiologie*; 4. Ed.; Spektrum Akademischer Verlag: Heidelberg, Berlin, 2001.
- (82) Wöhrl, B. In *Retrovirus Bulletin*; Nationales Referenzzentrum für Retroviren - Institut für Klinische und Molekulare Virologie: Erlangen, 2003; Band 02/2003.
- (83) Tor, Y. Targeting RNA with Small Molecules. *ChemBioChem* **2003**, *4*, 998-1007.
- (84) Fanning, G. C. RNA as A Target for Host Defense and Anti-HIV Drugs. *Curr. Drug Targ.* **2006**, *7*, 1607-1613.
- (85) Metzner, K. In *Retrovirus Bulletin*; Nationales Referenzzentrum für Retroviren - Institut für Klinische und Molekulare Virologie: Erlangen, 2003; Band 02/2003.
- (86) Preiser, W.; Rabenau, H. F.; Doerr, H. W. *Viren - Viruserkrankungen : Synopsis Der Epidemiologie, Klinik, Diagnostik Und Therapie Viraler Erkrankungen*; ZETT-Verlag: Steinen, 2002.
- (87) Förster, T. Zwischenmolekulare Energiewanderung Und Fluoreszenz. *Annalen der Physik* **1948**, *437*, 55-75.
- (88) Matsumoto, C.; Hamasaki, K.; Mihara, H.; Ueno, A. A High-Throughput Screening Utilizing Intramolecular Fluorescence Resonance Energy Transfer for the Discovery of the Molecules That Bind HIV-1 TAR RNA Specifically. *Bioorg. Med. Chem. Let.* **2000**, *10*, 1857-1861.
- (89) Wei, A. P.; Blumenthal, D. K.; Herron, J. N. Antibody-Mediated Fluorescence Enhancement Based on Shifting the Intramolecular Dimer. Dblarw. Monomer Equilibrium of Fluorescent Dyes. *Analytic. Chem.* **1994**, *66*, 1500-1506.

- (90) *Sigmaplot 2004*, Version 9.1; SyStat Software, Inc.: 1735, Technology Drive, Ste 430, San Jose, CA 95110, USA; <http://www.systat.com/>.
- (91) Ugi, I.; Steinbrückner, C. Über Ein Neues Kondensations-Prinzip. *Angew. Chem.* **1960**, *72*, 267-268.
- (92) Ugi, I.; Meyr, R.; Fetzer, U.; Steinbrückner, C. Versuche Mit Isonitrilen. *Angew. Chem.* **1959**, *71*, 386.
- (93) Schüller, A.; Fechner, U.; Renner, S.; Franke, L.; Weber, L.; Schneider, G. A Pseudo-Ligand Approach to Virtual Screening. *Comb. Chem. High-Throughput. Screen.* **2006**, *9*, 359-364.
- (94) Dömling, A.; Ugi, I. Multicomponent Reactions with Isocyanides. *Angew. Chem. Int. Ed.* **2000**, *39*, 3168-3210.
- (95) Schneider, P.; Schneider, G. Collection of Bioactive Reference Compounds for Focused Library Design. *QSAR Comb. Sci.* **2003**, *22*, 713-718.
- (96) *CLIFF (Chemical Library: Interconversion of File Formats)*, Version 1.14; Molecular Networks GmbH: Nägelsbachstraße 25, 91052 Erlangen; <http://www.molecular-networks.com/>.
- (97) *Molecular Operating Environment (MOE)*, Versionen 2004.03, 2005.06, 2006.08, 2007.09; Chemical Computing Group: 1010 Sherbrooke St. West, #910, Montreal, Canada, H3A 2R7; <http://www.chemcomp.com/>.
- (98) Boden, O. Dissertation: Design und Synthese heterocyclischer Liganden für die TAR-RNA aus HIV-1. *Johann Wolfgang Goethe-Universität* **2006**.
- (99) Krebs, A. *Chiral-Pool-Synthese Von aromatischen Aminosäuren als Bausteine für RNA-Liganden*; Der Andee Verlag: Osnabrück, 2004.
- (100) Ludwig, V.; Krebs, A.; Stoll, M.; Dietrich, U.; Ferner, J.; Schwalbe, H.; Scheffer, U.; Dürner, G.; Göbel, M. W. Tripeptides from Synthetic Amino Acids Block the Tat-TAR Association and Slow Down HIV Spread in Cell Cultures. *ChemBioChem* **2007**, *8*, 1850-1856.
- (101) Nietert, M. Dissertation: Virtuelles Screening nach RNA-Liganden: Zum Umgang mit einer flexiblen Zielstruktur. *Johann Wolfgang Goethe-Universität* **2008**.
- (102) Renner, S.; Ludwig, V.; Boden, O.; Scheffer, U.; Göbel, M.; Schneider, G. New Inhibitors of the Tat-TAR RNA Interaction Found with a „Fuzzy“ Pharmacophore Model. *ChemBioChem* **2005**, *6*, 1119-1125.
- (103) Seifert, G. Dissertation: Synthese und Charakterisierung substituierter Pyrazole und Chinoxalinderivate als potenzielle Inhibitoren des Tat-TAR-Komplexes von HIV-1. *Johann Wolfgang Goethe-Universität* **2008**.

- (104) Suhartono, M.; Weidlich, M.; Stein, T.; Karas, M.; Dürner, G.; Göbel, M. W. Synthesis of Non-Natural Aromatic Alpha-Amino Acids by a Heck Reaction. *Eur. J. Org. Chem.* **2008**, *2008*, 1608-1614.
- (105) 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**, *8*, 1932-1936.
- (106) *CORINA 3D Structure Generator*, Version 3.20; Molecular Networks GmbH: Nägelsbachstraße 25, 91052 Erlangen; <http://www.molecular-networks.com/>.
- (107) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720-729.
- (108) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *J. Comput. Chem.* **1999**, *20*, 730-748.
- (109) Atkins, P. W.; Höpfner, A. *Physikalische Chemie*; 3. Ed.; Wiley-VCH: Weinheim, New York, Chichester, Brisbane, Singapore, Toronto, 2002.
- (110) Todeschini, R.; Consonni, V.; Mannhold, R.; Kubinyi, H.; Timmerman, H. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
- (111) Willett, P.; Barnard, J.; Downs, G. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983-996.
- (112) Gasteiger, J.; Engel, T. *Cheminformatics: A Textbook*; 1. Ed.; Wiley-VCH: Weinheim, 2003.
- (113) Messmer, B. T. Efficient Graph Matching Algorithms. Institut für Informatik und Angewandte Mathematik, University of Bern, 1995.
- (114) Fechner, U.; Paetz, J.; Schneider, G. Comparison of Three Holographic Fingerprint Descriptors and Their Binary Counterparts. *QSAR Comb. Sci.* **2005**, *24*, 961-967.
- (115) Connolly, M. L. Analytical Molecular Surface Calculation. *J. Appl. Cryst.* **1983**, *16*, 548-558.
- (116) Zamora, I.; Oprea, T.; Cruciani, G.; Pastor, M.; Ungell, A. Surface Descriptors for Protein-Ligand Affinity Prediction. *J. Med. Chem.* **2003**, *46*, 25-33.
- (117) Perola, E.; Charifson, P. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization Upon Binding. *J. Med. Chem.* **2004**, *47*, 2499-2510.

- (118) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727-748.
- (119) GOLD (*Genetic Optimisation for Ligand Docking*), Versionen 3.2 & 4.0; CCDC Software Ltd.: The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom.
- (120) Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 11. Ed.; Springer: Berlin, 2006.
- (121) Lance, C. E.; Butts, M. M.; Michels, L. C. The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say? *Org. Res. Meth.* **2006**, *9*, 202-220.
- (122) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. „Scaffold-Hopping“ by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem. Int. Ed.* **1999**, *38*, 2894-2896.
- (123) Fechner, U.; Schneider, G. Optimization of a Pharmacophore-Based Correlation Vector Descriptor for Similarity Searching. *QSAR Comb. Sci.* **2004**, *23*, 19-22.
- (124) Pearlman, R. S.; Smith, K. Novel Software Tools for Chemical Diversity. *Persp. Drug. Discov. Des.* **1998**, *9-11*, 339-353.
- (125) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3-25.
- (126) Oprea, T. I. Property Distribution of Drug-Related Chemical Databases. *J. Comput.-Aided. Mol. Des.* **2000**, *14*, 251-264.
- (127) Kohonen, L. *Self-Organization and Associative Memory*; Springer-Verlag: Heidelberg, 1984.
- (128) Schneider, G.; So, S. *Adaptive Systems in Drug Design*; 1. Ed.; Landes Bioscience: Georgetown, 2002.
- (129) Gerdes, I.; Klawonn, F.; Kruse, R. *Evolutionäre Algorithmen*; 1. Ed.; Vieweg: Wiesbaden, 2004; S. 252.
- (130) Schüller, A.; Schneider, G. Identification of Hits and Lead Structure Candidates with Limited Resources by Adaptive Optimization. *J. Chem. Inf. Model.* **2008**, *48*, 1473-1491.
- (130a) Schneider, G. In *Virtual Screening for Bioactive Molecules*; Böhm, H.; Schneider, G., Ed.; Wiley-VCH: Weinheim, New York, 2000; Band 10, S. 161-186.
- (131) Maggiora, G. M.; Shanmugasundaram, V.; Lajiness, M. S.; Doman, T. N.; Schultz, M. W. In *Cheminformatics in Drug Discovery*; Oprea, T. I., Ed.; 2005; S. 315-332.

- (132) Horvath, D.; Jeandenans, C. Neighborhood Behavior of *in Silico* Structural Spaces with Respect to *in Vitro* Activity Spaces - a Novel Understanding of the Molecular Similarity Principle in the Context of Multiple Receptor Binding Profiles. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 680-690.
- (133) Maggiora, G. M. On Outliers and Activity Cliffs - Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535-1535.
- (134) Peltason, L.; Bajorath, J. SAR Index: Quantifying the Nature of Structure-Activity Relationships. *J. Med. Chem.* **2007**, *50*, 5571-5578.
- (135) Spall, J. C. *Introduction to Stochastic Search and Optimization*; John Wiley & Sons, Inc.: New York, NY, USA, 2003.
- (136) Freeman, E.; Freeman, E.; Bates, B.; Sierra, K. *Head First Design Patterns*; 1. Ed.; O'Reilly Media, Inc., 2004.
- (137) Aghbari, Z. A. Array-Index: A Plug&search K Nearest Neighbors Method for High-Dimensional Data. *Data Knowl. Engin.* **2005**, *52*, 333-352.
- (138) Aggarwal, C.; Hinneburg, A.; Keim, D. In *Database Theory — ICDT 2001*; 2001; S. 420-434.
- (139) Clerc, M.; Kennedy, J. The Particle Swarm - Explosion, Stability, and Convergence in a Multidimensional Complex Space. *IEEE Trans. Evol. Comput.* **2002**, *6*, 58-73.
- (140) Schneider, G. MOLMAP(R) User Guide. **2003**.
- (141) Schneider, G.; Wrede, P. Artificial Neural Networks for Computer-Based Molecular Design. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175-222.
- (142) Bishop, C. M. *Pattern Recognition and Machine Learning*; 1. Ed.; Springer: Singapore, 2007.
- (143) Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, *70*, 1129-1143.
- (144) Basalaj, W. Proximity Visualization of Abstract Data. **2001**; <http://www.pavis.org/essay/index.html>, zugegriffen am 04.07.2008.
- (145) Gower, J. C. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika* **1966**, *53*, 325-338.
- (146) *Matlab*, Version 7.04.365 R14 SP2; The MathWorks: 3 Apple Hill Drive, Natick, MA 01760-2098, USA; <http://www.mathworks.com/>.
- (147) *Google Earth*, Version 4.2.0205.5730; Google Inc.: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA; <http://earth.google.com/>, zugegriffen am 14.02.2008.

- (148) Whitlow, M.; Arnaiz, D. O.; Buckman, B. O.; Davey, D. D.; Griedel, B.; Guilford, W. J.; Koovakkat, S. K.; Liang, A.; Mohan, R.; Phillips, G. B.; Seto, M.; Shaw, K. J.; Xu, W.; Zhao, Z.; Light, D. R.; Morrissey, M. M. Crystallographic Analysis of Potent and Selective Factor Xa Inhibitors Complexed to Bovine Trypsin. *Acta Crystallogr. Sect. D* **1999**, *55*, 1395-1404.
- (149) Noeske, T.; Sasse, B. C.; Stark, H.; Parsons, C. G.; Weil, T.; Schneider, G. Predicting Compound Selectivity by Self-Organizing Maps: Cross-Activities of Metabotropic Glutamate Receptor Antagonists. *ChemMedChem* **2006**, *1*, 1066-1068.
- (150) Lewin, B. *Genes VIII*; International Ed.; Pearson Prentice Hall, Pearson Education, Inc.: Upper Saddle River, NJ 07458, 2004.
- (151) Schneider, G.; Schneider, P. In *Chemogenomics in Drug Discovery*; Kubinyi, H.; Müller, G., Ed.; Wiley-VCH: Weinheim, 2004; S. 341-376.
- (152) Foloppe, N.; Matassova, N.; Aboul-ela, F. Towards the Discovery of Drug-Like RNA Ligands? *Drug Discov. Today* **2006**, *11*, 1019-1027.
- (153) Faber, C.; Sticht, H.; Schweimer, K.; Rosch, P. Structural Rearrangements of HIV-1 Tat-Responsive RNA Upon Binding of Neomycin B. *J. Biol. Chem.* **2000**, *275*, 20660-20666.
- (154) Aboul-ela, F.; Karn, J.; Varani, G. Structure of HIV-1 TAR RNA in the Absence of Ligands Reveals a Novel Conformation of the Trinucleotide Bulge. *Nucl. Acids Res.* **1996**, *24*, 3974-3981.
- (155) He, M.; Yuan, D.; Lin, W.; Pang, R.; Yu, X.; Yang, M. Synthesis and Assay of Isoquinoline Derivatives as HIV-1 Tat-TAR Interaction Inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3978-3981.
- (156) Murchie, A. I. H.; Davis, B.; Isel, C.; Afshar, M.; Drysdale, M. J.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T. M.; Mirza, S.; Prescott, C. D.; Vaglio, P.; Aboul-ela, F.; Karn, J. Structure-Based Drug Design Targeting an Inactive RNA Conformation: Exploiting the Flexibility of HIV-1 TAR RNA. *J. Mol. Biol.* **2004**, *336*, 625-638.
- (157) Cecchetti, V.; Parolin, C.; Moro, S.; Pecere, T.; Filipponi, E.; Calistri, A.; Tabarrini, O.; Gatto, B.; Palumbo, M.; Fravolini, A.; Palu', G. 6-Aminoquinolones as New Potential Anti-HIV Agents. *J. Med. Chem.* **2000**, *43*, 3799-3802.
- (158) Parolin, C.; Gatto, B.; Del Vecchio, C.; Pecere, T.; Tramontano, E.; Cecchetti, V.; Fravolini, A.; Masiero, S.; Palumbo, M.; Palu, G. New Anti-Human Immunodeficiency Virus Type 1 6-Aminoquinolones: Mechanism of Action. *Antimicrob. Agents Chemother.* **2003**, *47*, 889-896.

- (159) Du, Z.; Lind, K. E.; James, T. L. Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor That Was Identified by Computational Screening. *Chem. Biol.* **2002**, *9*, 707-712.
- (160) Hamy, F.; Brondani, V.; Florsheimer, A.; Stark, W.; Blommers, M.; Klimkait, T. A New Class of HIV-1 Tat Antagonist Acting Through Tat-TAR Inhibition. *Biochemistry* **1998**, *37*, 5086-5095.
- (161) Morgner, N.; Scheffer, U.; Breitung, S.; Barth, H.; Göbel, M.; Brutschy, B. Binding Sites of the Viral RNA Element TAR and of TAR Mutants for Various Peptide Ligands, Probed with LILBID: A New Laser Mass Spectrometry. *J. Am. Soc. Mass. Spec.* **2008**, im Druck.
- (161a) Brodersen, D. E.; Clemons, W. M. Jr; Carter, A. P.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **2000**, *22*, 1143-1154.
- (161b) Luedtke, N. W.; Liu, Qi; Tor, Y. RNA-Ligand Interactions: Affinity and Specificity of Aminoglycoside Dimers and Acridine Conjugates to the HIV-1 Rev Response Element. *Biochemistry* **2003**, *42*, 11391-11403.
- (161c) Tor, Y. The ribosomal A-site as an inspiration for the design of RNA binders. *Biochimie* **2006**, *88*, 1045-1051.
- (161d) Peytou, V.; Condom, R.; Patino, N.; Guedj, R.; Aubertin, A. M.; Gelus, N.; Bailly, C.; Terreux, R.; Cabrol-Bass, D. J. Synthesis and Antiviral Activity of Ethidium-Arginine Conjugates Directed Against the TAR RNA of HIV-1. *J. Med. Chem.* **1999**, *42*, 4042-4053.
- (161e) Schlünzen, F.; Zarivach, R.; Harms, J.; Bashan, A; Tocilj, A; Albrecht, R; Yonath, A; Franceschi F. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **2001**, *413*, 814-821.
- (161f) Luedtke, N. W.; Tor, Y. A Novel Solid-Phase Assembly for Identifying Potent and Selective RNA Ligands. *Angew. Chem. Int* **2000**, *39*, 1788-1790.
- (161g) Chow, C. S.; Bogdan F. M. A Structural Basis for RNA-Ligand Interactions. *Chem. Rev.* **1997**, *97*, 1489-1513.
- (162) Hou, T.; Xia, K.; Zhang, W.; Xu, X. ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 266-275.
- (163) Labute, P. MOE LogP(Octanol/Water) Model. *Nicht publiziert, Quellcode in \$MOE/lib/svl/quasar.svl/q_logp.svl* **1998**.

- (164) Balakin, K. V.; Savchuk, N. P.; Tetko, I. V. In Silico Approaches to Prediction of Aqueous and DMSO Solubility of Drug-Like Compounds: Trends, Problems and Solutions. *Curr. Med. Chem.* **2006**, *13*, 223-241.
- (165) Llinàs, A.; Glen, R. C.; Goodman, J. M. Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a Database of 100 Reliable Measurements? *J. Chem. Inf. Model.* **2008**, *48*, 1289-1303.
- (166) Neumann, H. Health Risk of Combustion Products: Toxicological Considerations. *Chemosphere* **2001**, *42*, 473-479.
- (167) Jameson, C. W.; Lunn, R. M.; Jeter, S.; Sabella, A. *Report on Carcinogens Background Document for Nitrobenzene*; Technology Planning and Management Corporation: Canterbury Hall, Suite 310, 4815 Emperor Blvd, Durham, NC 27703, USA, 2002; <http://ntp.niehs.nih.gov/ntp/newhomeroc/roc11/NBpub.pdf>.
- (168) Rosenkranz, H. S.; Mermelstein, R. Mutagenicity and Genotoxicity of Nitroarenes. All Nitro-Containing Chemicals Were Not Created Equal. *Mutat. Res.* **1983**, *114*, 217-67.
- (169) Letelier, M. E.; Izquierdo, P.; Godoy, L.; Lepe, A. M.; Faúndez, M. Liver Microsomal Biotransformation of Nitro-Aryl Drugs: Mechanism for Potential Oxidative Stress Induction. *J. Appl. Toxicol.* **2004**, *24*, 519-525.
- (170) Davies, L. Nitrobenzene (Environmental Health Criteria 230). United Nations Environment Programme, International Labour Organization, World Health Organization **2003**; <http://whqlibdoc.who.int/publications/2003/9241572300.pdf>.
- (171) Mutschler, E.; Geisslinger, G.; Kroemer, H. K.; Ruth, P.; Schäfer-Korting, M. *Arzneimittelwirkungen: Lehrbuch der Pharmakologie und Toxikologie. Mit einführenden Kapiteln in die Anatomie, Physiologie und Pathophysiologie*; 9. Ed.; Wissenschaftliche Verlagsges.: Stuttgart, 2008; S. 1244.
- (172) Tasler, S.; Kraus, J.; Wuzik, A.; Müller, O.; Aschenbrenner, A.; Cubero, E.; Pascual, R.; Quintana-Ruiz, J.; Dordal, A.; Mercè, R.; Codony, X. Discovery of 5-HT₆ Receptor Ligands Based on Virtual HTS. *Bioorg. Med. Chem. Let.* **2007**, *17*, 6224-6229.
- (173) Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. Comparison of Correlation Vector Methods for Ligand-Based Similarity Searching. *J. Comput.-Aided. Mol. Des.* **2003**, *17*, 687-698.
- (174) Martin, Y.; Kofron, J.; Traphagen, L. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **2002**, *45*, 4350-4358.
- (175) Vollhardt, K. P. C.; Schore, N. E. *Organische Chemie*; 3. Ed.; Wiley-VCH: Weinheim, 2000.

- (176) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual Computational Chemistry Laboratory - Design and Description. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453-63.
- (177) *AILOGPS*, Version 2.1 (online); VCCLAB, Virtual Computational Chemistry Laboratory; <http://www.vcclab.org/>, zugegriffen am 29.05.2008.
- (177a) The PyMOL Molecular Graphics System, Version 1.0r2; DeLano, W. L. The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA, 2002; <http://www.pymol.org/>.
- (178) Detering, C.; Varani, G. Validation of Automated Docking Programs for Docking and Database Screening Against RNA Drug Targets. *J. Med. Chem.* **2004**, *47*, 4188-4201.
- (179) Cummings, M.; DesJarlais, R.; Gibbs, A.; Mohan, V.; Jaeger, E. Comparison of Automated Docking Programs as Virtual Screening Tools. *J. Med. Chem.* **2005**, *48*, 962-976.
- (180) Ewing, T. J. A.; Kuntz, I. D. Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. *J. Comput. Chem.* **1997**, *18*, 1175-1189.
- (181) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM - A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J. Comput. Chem.* **1994**, *15*, 488-506.
- (182) Filikov, A. V.; Mohan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. Identification of Ligands for RNA Targets Via Structure-Based Virtual Screening: HIV-1 TAR. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 593-610.
- (183) Lind, K. E.; Du, Z.; Fujinaga, K.; Peterlin, B. M.; James, T. L. Structure-Based Computational Database Screening, In Vitro Assay, and NMR Assessment of Compounds That Target TAR RNA. *Chem. Biol.* **2002**, *9*, 185-193.
- (184) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- (185) Barbault, F.; Zhang, L.; Zhang, L.; Fan, B. T. Parametrization of a Specific Free Energy Function for Automated Docking Against RNA Targets Using Neural Networks. *Chemom. Intell. Lab. Syst.* **2006**, *82*, 269-275.
- (186) Pfeffer, P.; Gohlke, H. DrugScoreRNA-Knowledge-Based Scoring Function To Predict RNA-Ligand Interactions. *J. Chem. Inf. Model.* **2007**, *47*, 1868-1876.

- (187) Aboul-ela, F.; Karn, J.; Varani, G. The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* **1995**, *253*, 313-32.
- (188) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discov. Today* **2004**, *9*, 430-431.
- (189) Robinson, J. M. A.; Philp, D.; Harris, K. D. M.; Kariuki, B. M. Weak Interactions in Crystal Engineering-Understanding the Recognition Properties of the Nitro Group. *New J. Chem.* **2000**, *24*, 799-806.
- (190) George, T. G.; Endeshaw, M. M.; Morgan, R. E.; Mahasenan, K. V.; Delfin, D. A.; Mukherjee, M. S.; Yakovich, A. J.; Fotie, J.; Li, C.; Werbovetz, K. A. Synthesis, Biological Evaluation, and Molecular Modeling of 3,5-Substituted-N1-Phenyl-N4,N4-Di-N-Butylsulfanilamides as Antikinetoplastid Antimicrotubule Agents. *Bioorg. Med. Chem.* **2007**, *15*, 6071-6079.
- (191) *SciFinder Scholar*, Version 2007; Chemical Abstracts Service: Columbus, Ohio, USA.
- (192) Quintana Ruiz, J. R. Nitro-Substituted Phenyl-Piperazine Compounds, Their Preparation and Use in Medicaments. Patent Nr. EP1676842, 2006.
- (193) Stewart, J. J. P. *Mopac*, 7. Edition; Stewart Computational Chemistry: Colorado Springs, CO, USA; <http://openmopac.net/>.
- (194) Lide, D. R. *CRC Handbook of Chemistry and Physics*; 74. Ed.; CRC Press: Boca Raton, 1994.
- (195) Labute, P. MOE Molar Refractivity Model. *Nicht publiziert, Quellcode in \$MOE/lib/svl/quasar.svl/q_logp.svl* **1998**.
- (196) Wildman, S.; Crippen, G. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868-873.
- (197) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714-3717.
- (198) Hall, L. H.; Kier, L. B. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Ed.; Reviews in Computational Chemistry; Wiley-VCH: New York, 1991; Vol. 2.
- (199) Kier, L. B.; Hall, L. H. The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity. *Eur. J. Med. Chem.* **1977**, *12*, 334.
- (200) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem Phys Lett* **1982**, *89*, 399-404.

- (201) Petitjean, M. Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331-337.
- (202) Balaban, A. T. Five New Topological Indices for the Branching of Tree-Like Graphs. *Theor. Chim. Acta* **1979**, *53*, 355-375.
- (203) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
- (204) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219-3228.
- (205) Holder, J. W. Nitrobenzene Carcinogenicity in Animals and Human Hazard Evaluation. *Toxicol. Indust. Health* **1999**, *15*, 445-57.

10 Lebenslauf

Curriculum Vitæ



Andreas Schüller

Geboren am 9. Januar 1979
in Düsseldorf

deutsch

ledig

Curriculum Vitæ

Studium

- Seit April 2005 **Doktorarbeit** „*Adaptive Optimierung fokussierter Substanzbibliotheken*“ in der Arbeitsgruppe von Prof. Dr. Gisbert Schneider, Beilstein-Stiftungsprofessur für Chemieinformatik, an der Johann Wolfgang Goethe-Universität Frankfurt.
- Juni 2004 Erhalt des akademischen Grades „**Diplom Biochemiker**“ an der Universität Frankfurt mit dem Gesamturteil „*Sehr gut, mit Auszeichnung*“
- 12/2003 bis 6/2004 **Diplomarbeit** „*Modellierung von Protein-Liganden-Interaktionen mit einer heuristischen Bewertungsfunktion*“ in der Arbeitsgruppe von Prof. Dr. Gisbert Schneider an der Universität Frankfurt
- 5/2003 bis 6/2003 **Praktikum** in der Arbeitsgruppe von Prof. Dr. Karl-Dieter Entian, Institut für molekulare Genetik and zelluläre Mikrobiologie an der Universität Frankfurt
- 3/2003 bis 5/2003 **Praktikum** in der Arbeitsgruppe von Prof. Dr. Gisbert Schneider, Universität Frankfurt
- 8/2000 bis 9/2000 **Praktikum** bei der Motor Presse Stuttgart, Geschäftsbereich Online (Datenbank- und Webprogrammierung)
- 10/1999 bis 6/2004 **Studium** an der Universität Frankfurt, *Studiengang: Biochemie*

Berufserfahrung

- 06/2008 bis 08/2008 Freier Mitarbeiter bei Merz Pharmaceuticals GmbH, Abteilung Research & Development, Frankfurt am Main

Erfahrungen in der Lehre

1/2007 bis 3/2007	Assistent im Praktikum <i>“Chemie für Mediziner”</i> :
1/2006 bis 3/2006	Theoretischer und laborpraktischer Blockkurs der allgemeinen und organischen Chemie (10 Wochen)
10/2006	Assistent im Kurs <i>“Prinzipien der Bioinformatik”</i> : Eine Einführung in die angewandte Bio- und Chemieinformatik.
7/2005	Assistent im Kurs <i>„Wirkstoffdesign“</i> : Ein zweiwöchiger Praktikumskurs über das rationale Wirkstoffdesign
2005 bis 2008	Betreuer mehrerer Arbeitskreispraktikanten
2005 bis 2008	Co-Betreuer von Diplomstudenten
2005 bis 2008	Beisitzer in mündlichen Abschlussprüfungen

Stipendien und Mitgliedschaften

5/2007 bis 12/2007	Mitglied im Sonderforschungsbereich 579 <i>„RNA-Liganden Interaktionen“</i> der Deutschen Forschungsgemeinschaft (DFG)
Seit 2006	Mitglied der Gesellschaft Deutscher Chemiker (GDCh)
Seit 2006	Rezensent für <i>QSAR & Combinatorial Science</i>
4/2005 bis 4/2007	Stipendiat des <i>„Chemiefonds-Stipendiums“</i> für Doktoranden des Fonds der Chemischen Industrie

Wissenschaftliche Interessen

- Biologisch inspirierte Algorithmen und k -NN-Methoden
- Design fokussierter Substanzbibliotheken und virtuelles Screening
- *De-novo*-Design
- Virtuelle kombinatorische Chemie
- Wissenschaftliche Programmierung

Fähigkeiten und Engagement

Programmierung	Java, C, Pascal, Python, Shell Skript und Webprogrammierung (PHP, SQL, HTML, JavaScript) auf Windows- und Unixsystemen
Wissenschaftliche Software	Molecular Operating Environment (inklusive SVL-Programmierung), Matlab, Pipeline Pilot (inklusive PilotScript-Programmierung)
Sprachen	Englisch, fließend Spanisch, gute Kenntnisse Französisch, Grundkenntnisse
10/2004 bis 3/2005	Sprach- und Kulturreise nach Südamerika (<i>Chile, Argentinien, Bolivien und Peru</i>)
10/2002 bis 10/2003	Mitglied der Fachschaft Biochemie, Universität Frankfurt
Seit 2000	Freiberuflicher Webprogrammierer, Frankfurt am Main
9/1996 bis 8/1999	Jugendgruppenleiter in der Deutschen Pfadfinderschaft Sankt Georg, Leonberg

11 Publikationen

G. Schneider and A. Schüller: Adaptive Combinatorial Design of Focused Compound Libraries. In: Ligand-Macromolecule Interactions in Drug Discovery (A. C. Roque, Ed.). Humana Press **2008**, Totowa, angenommen.

A. Schüller, G. Schneider: Identification of Hits and Lead Structure Candidates with Limited Resources by Adaptive Optimization, *J. Chem. Inf. Model.* **2008**, 48, 1473-1491.

M. Hartenfeller, E. Proschak, A. Schüller, G. Schneider: Concept of Combinatorial De Novo Design of Drug-like Molecules by Particle Swarm Optimization, *Chem. Biol. Drug Des.* **2008**, 72, 16-26.

A. Schüller, M. Suhartono, U. Fechner, Y. Tanrikulu, S. Breitung, U. Scheffer, M. W. Göbel, G. Schneider: The concept of template-based *de novo* design from drug-derived molecular fragments and its application to TAR RNA, *J. Comput.-Aided Mol. Design* **2008**, 22, 59-68.

E. Proschak, J. K. Wegner, A. Schüller, G. Schneider, U. Fechner: Molecular Query Language (MQL) - A context-free grammar for substructure matching, *J. Chem. Inf. Model.* **2007**, 47, 295-301.

A. Schüller, V. Hähnke, G. Schneider, SmiLib v2.0: A Java-based tool for rapid combinatorial library enumeration, *QSAR Comb. Sci.* **2006**, 26, 407-410.

A. Schüller, U. Fechner, S. Renner, L. Franke, L. Weber, G. Schneider: A Pseudo-Ligand Approach to Virtual Screening, *Comb. Chem. High Throughput Screen.* **2006**, 9, 359-364.

A. Schüller, G. Schneider, E. Byvatov: SMILIB: Rapid assembly of combinatorial libraries in SMILES notation, *QSAR Comb. Sci.* **2003**, 22, 719-721.

Konferenzbeiträge

A. Schüller, U. Scheffer, M. Göbel, G. Schneider: Evolutionary Computer-assisted Screening for HIV-1 TAR RNA Ligands. Posterpräsentation bei der 3rd German Conference on Chemoinformatics, Goslar, November **2007**.

K. Grabowski, M. Meissner, A. Schüller, M. Nietert, E. Proschak, Y. Tanrikulu, B. Stauch, C. Münk, U. Scheffer, K. Cichutek, M. Schubert-Zsilavec, M. Göbel, G. Schneider: Applications of MOE in an Academic Environment. Posterpräsentation beim 5th European Chemical Computing Group User Group Meeting, Berlin, Oktober **2007**.

A. Schüller, U. Scheffer, M. Göbel, G. Schneider: Evolutionary Computer-assisted Screening for HIV-1 TAR RNA Ligands. Posterpräsentation beim RNA-Symposium des Sonderforschungsbereiches „RNA-Liganden-Interaktionen“, Universität Frankfurt, September **2007**.

B. Krüger, A. Schüller, U. Scheffer, M. Göbel, G. Schneider: Seeking out novel lead candidates by *de novo* design and molecular library optimization. Posterpräsentation bei der 2nd German Conference on Cheminformatics, Goslar, November **2006**.

A. Schüller, G. Schneider, Ligand-based Multiobjective Focused Library Optimization. Posterpräsentation bei der 1st German Conference on Cheminformatics, Goslar, November **2005**.

Vorträge

A. Schüller, G. Schneider: Adaptive Optimization of Focused Compound Libraries for Lead Candidate Identification in Early Drug Discovery. Workshop on Bio-Inspired Computing in Computational Biology bei der 9th International Conference on: Parallel Problem Solving from Nature, Reykjavik, Island, September **2006**.

A. Schüller, G. Schneider: Adaptive Optimization of Focused Libraries for Lead Identification. Regionaltreffen der Stipendiaten des Fonds der Chemischen Industrie, Universität Würzburg, Januar **2006**.

Diplomarbeit

A. Schüller: Modellierung von Protein-Liganden-Interaktionen mit einer heuristischen Bewertungsfunktion, Juni **2004**.

Software

SmiLib v2.0 – Ein javabasiertes Open-Source-Programm zur schnellen Enumerierung kombinatorischer Bibliotheken im SMILES-Format, <http://gecco.org.chemie.uni-frankfurt.de/smilib/>.