

**Virtuelles Screening und *De Novo* Design von PPAR α
Agonisten mit Oberflächen-Deskriptoren**

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Ewgenij Proschak
aus Sankt Petersburg

Frankfurt 2008
(D30)

vom Fachbereich Biowissenschaften der Johann Wolfgang Goethe-Universität als
Dissertation angenommen.

Dekan: Prof. Dr. Volker Müller

Erster Gutachter Prof. Dr. Gisbert Schneider

Zweiter Gutachter Prof. Dr. Manfred Schubert Zsilavec

Datum der Disputation:

Eidesstattliche Erklärung

Hiermit bestätige ich an Eides statt, dass ich die vorliegende Dissertation selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegeben nicht bedient habe.

Frankfurt am Main, den.....

.....

(Ewgenij Proschak)

Zusammenfassung

Die Komplementarität der molekularen Oberflächen und der Pharmakophorpunkte ist ein verbreiteter Konzept im rechnergestützten Moleküldesign. Diesem Konzept folgend wurde die Software SQUIRREL neu entwickelt und in der Programmiersprache Java implementiert. Die Software generiert die Vorschläge für den bioisosteren Ersatz von Molekülen und Molekülfragmenten. SQUIRREL kombiniert Oberflächen- und Pharmakophoreigenschaften bioaktiver Substanzen und kann im virtuellen Screening und fragment-basierten *de novo* Design eingesetzt werden.

In einer prospektiven Studie wurde SQUIRREL verwendet, um neue selektive PPAR α -Agonisten aus einer kommerziellen Moleküldatenbank zu identifizieren. Die Software lieferte eine potente Substanz ($EC_{50} = 44$ nM) mit über 100facher Selektivität gegenüber PPAR γ .

In einer zweiten Studie wurde eine Leitstruktur *de novo* generiert und synthetisiert. Als Ausgangstruktur diente der bekannte PPAR α -Agonist GW590735. Während des Designvorgangs wurden zwei Teilstrukturen, die für die Aktivität von GW590735 verantwortlich sind, durch bioisostere Gruppen ersetzt, die von SQUIRREL_{nov} vorgeschlagen wurden. Die neue Leitstruktur aktiviert PPAR α in einem zellbasierten Reporter-Gen-Testsystem bei einem EC_{50} von 0.51 μ M.

Summary

Shape and pharmacophoric point complementarity is a widely accepted concept in computer-aided molecular design. Following this principle, the software SQUIRREL was developed to provide the possibility to generate proposals for bioisosteric replacement. SQUIRREL combines steric and pharmacophoric properties of bioactive compounds and can be applied to virtual screening and fragment-based *de novo* design.

In a prospective virtual screening study, SQUIRREL was employed to identify novel selective PPAR α agonists from a large collection of screening compounds. The software retrieved a potent hit ($EC_{50} = 44$ nM), with more than 100fold selectivity against PPAR γ .

In a second application, a lead structure was designed *de novo* and synthesized. The known PPAR α agonist GW590735 was chosen as a starting point. In the design process, two moieties were replaced by bioisosteric groups suggested by SQUIRREL_{nov}, which are responsible for compound potency. The new lead structure yielded submicromolar activity on PPAR α ($EC_{50} = 0.51$ μ M) in a cell-based reporter-gene assay.

Inhaltsverzeichnis

EINLEITUNG	7
PPARS – PEROXISOM PROLIFERATOR AKTIVIERTE REZEPTOREN	9
RECHNERGESTÜTZTE WIRKSTOFFDESIGN-METHODEN BEI DER SUCHE NACH NEUEN PPAR-AGONISTEN	14
THEORETISCHE METHODEN	16
SHAPE DESKRIPTOREN UND SHAPE MATCHING.....	16
ENTWICKELTE METHODEN	17
BENUTZTE METHODEN UND PROGRAMME	35
PRAKTISCHE METHODEN.....	42
PPAR-ASSAY.....	42
SYNTHESE DER SUBSTANZEN EP1 UND EP5.....	46
ERGEBNISSE UND DISKUSSION	53
ÜBERLAGERUNG VON MOLEKÜLEN BASIEREND AUF SHAPELETS	53
OBERFLÄCHEN-BASIERTES VIRTUAL SCREENING.....	60
VIRTUAL SCREENING MIT SQUIRREL.....	62
RETROSPEKTIVE VIRTUAL SCREENING STUDIE AUF PPARs.....	63
PROSPEKTIVE VIRTUAL SCREENING STUDIE AUF PPAR α	65
DE NOVO DESIGN UND SYNTHESE EINES PPAR α -AGONISTEN	73
SCHLUSSWORT UND AUSBLICK.....	79
LITERATURVERZEICHNIS	81
DANKSAGUNG	91
LEBENS LAUF	93
EIGENE PUBLIKATIONS LISTE	94
ANHANG	97

Einleitung

„Wirkstoffdesign ist Wissenschaft, Technologie und Kunst.“

Hans-Joachim Böhm [Böhm1996]

Die Arzneimittelforschung benötigt neue Wirkstoffe, die wirksam und sicher sind. Das rationale Design von Wirkstoffen gewinnt dabei immer mehr an Bedeutung. Computerbasierte Techniken können auf vielen Etappen der Entwicklung eines Medikaments zum Einsatz kommen. In dieser Arbeit werden insbesondere die Entwicklung und Anwendung der Methoden beleuchtet, die am Anfang der Entstehung eines Medikaments stehen und die Leitstrukturfindung erleichtern sollen.

Eine Leitstruktur ist eine Verbindung, die über die gesuchte biologische Aktivität verfügt. Sie zeichnet sich durch spezifische Bindung an das Zielprotein und synthetische Zugänglichkeit aus [Böhm1996].

Die Findung der Leitstruktur kann auf vielerlei Arten erfolgen, Medikamente wie Penicillin wurden durch Zufall entdeckt [Fleming1929]. Die moderne Wirkstoffforschung setzt auf unterschiedliche Techniken [Keserü2006], um potentielle Leitstrukturen zu identifizieren. Virtual Screening ist eine davon. *Unter Screening (engl. to screen, „durchmustern“, durch Eliminierung auswählen) versteht man die mehr oder weniger gezielte Prüfung großer Zahlen von Substanzen* [Böhm1996]. Im Gegensatz zum Hochdurchsatz-Screening [Gordon1994] (engl. *high-throughput screening*, HTS), einem Verfahren, bei dem 10^4 - 10^6 Substanzen mit Hilfe von automatisierten Testsystemen auf gewünschte biologische Aktivität experimentell geprüft werden, reduziert das Screening *in silico*, auch Virtual Screening genannt, den Testaufwand auf 100-1000 Strukturen [Schneider2002].

Einen anderen Weg geht das *de novo* Design. Bei dieser Strategie wird versucht, eine Struktur mit gewünschter Aktivität inkrementell aus Strukturfragmenten oder gar Atomen aufzubauen [Schneider2005]. Als Referenz kann die durch Röntgenkristallographie oder Kernspinresonanzspektroskopie (engl. nuclear magnetic resonance, NMR-Spektroskopie) aufgeklärte Struktur des Rezeptors (strukturbasiertes *de novo* Design) dienen. Ist diese nicht vorhanden, unzureichend aufgelöst oder weist

induzierte-Passform (engl. *induced-fit*) Phänomene auf, wird das liganden-basierte *de novo* Design eingesetzt. Bei dieser Variante dienen eine oder mehrere Moleküle mit bekannter Aktivität als Referenz für das Design einer neuen Leitstruktur.

Virtual Screening avancierte seit den 90er Jahren zu einer Standardprozedur der Leitstrukturfindung und einer Alternative bzw. Ergänzung zu HTS [Klebe2006, Seifert2007]. Die Anzahl der Publikationen mit dem Stichwort „virtual screening“ auf der Webseite des Bibliothekssammlungsdienstes PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) steigt jährlich an (Abbildung 1). Hingegen verläuft die Entwicklung der *de novo* Design-Verfahren langsamer, die Anzahl der Publikationen bleibt seit 7 Jahren auf konstantem Niveau. Die Anzahl der prospektiv erfolgreichen Studien ist überschaubar [Schneider2005, Proschak2008].

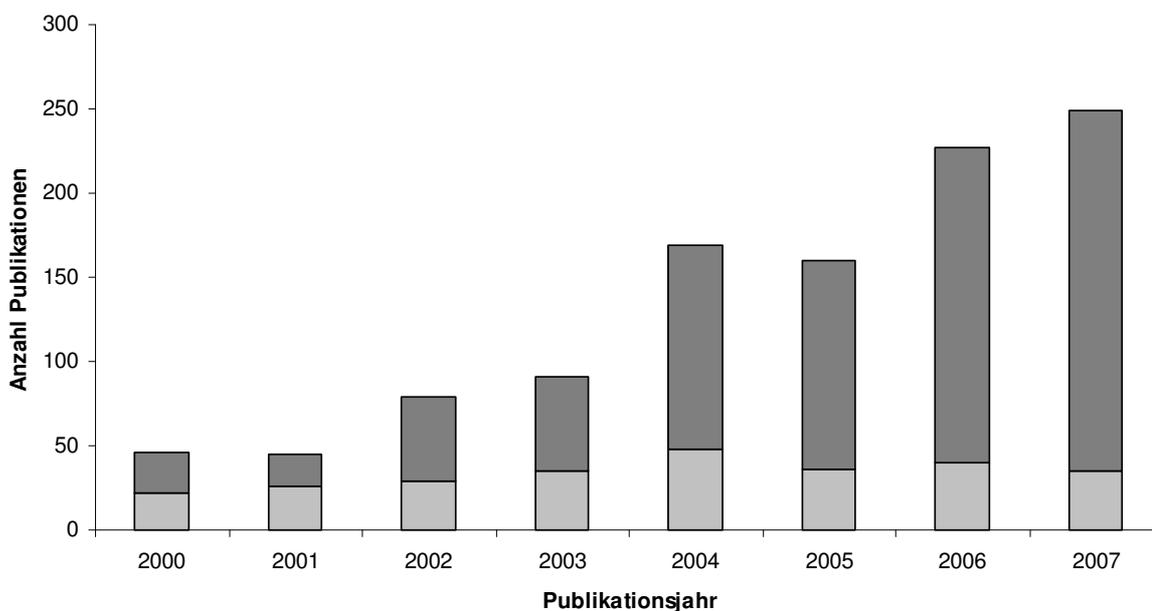


Abbildung 1: Anzahl der Publikationen mit Stichwort „de novo design“ (hellgrau) und „virtual screening“ (dunkelgrau) bei PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>).

Ein Ziel dieser Arbeit ist es daher, nicht nur ein Verfahren für *de novo* Design zu entwickeln, sondern auch experimentell, d.h. durch erfolgreiche Synthese und Testung die Anwendbarkeit in der Praxis zu zeigen.

PPARs – Peroxisom Proliferator Aktivierte Rezeptoren

PPAR ist ein Rezeptor aus der Familie der nukleären Rezeptoren. Es existieren drei Subtypen: PPAR α , PPAR δ und PPAR γ . PPAR spielen eine entscheidende Rolle bei der Lipid- und Glucosehomeostase [Rau2008]. Die Aktivierung von PPAR führt zur Dimerisierung mit dem Retinoid X Rezeptor (RXR) und der Expression von Genen, die beim Transport und der Metabolisierung von Glucose und Fettsäuren eine Rolle spielen [Desvergne1999]. Die Übersicht über die Funktionen wird in der Abbildung 2 gegeben [Steinhilber2005].

Die niedermolekularen Modulatoren von PPAR können je nach ihrer Wirkungsweise in Agonisten, partielle Agonisten und Antagonisten unterteilt werden. Die Messung der Aktivität kann im zellbasierten Reporter-Gen-Testsystem erfolgen [Rau2006]. Dabei führt die Zugabe eines PPAR-Agonisten zur Expression des Reportergens, z.B. der Luciferase. Partielle Agonisten aktivieren zwar die Genexpression, jedoch erreicht diese nicht das Level des Agonisten. Antagonisten hingegen blockieren die Expression des Reportergens, die durch einen Agonisten induziert wird.

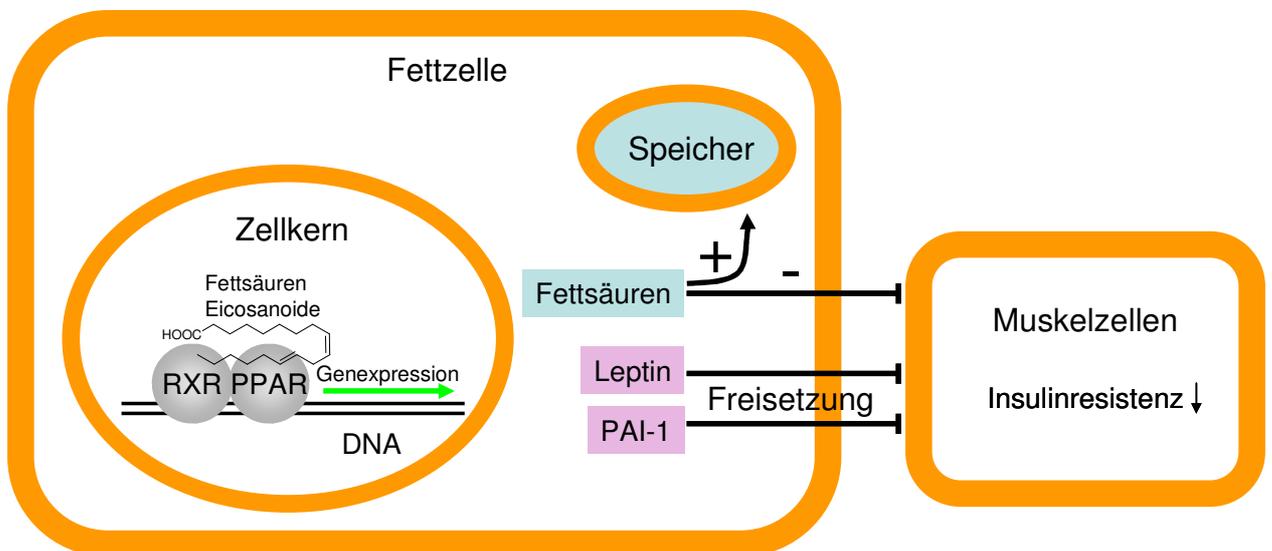


Abbildung 2. (aus [Steinhilber2005]): Die Funktion von PPAR am Beispiel der Adipozyten.

Die endogenen Liganden der PPARs sind Fettsäuren [Berger2002]. Die Regulation der PPARs spielt eine wichtige Rolle bei Erkrankungen, die mit Entzündungen einhergehen [Széles2007], Dyslipidämie [Rau2008, Duval2007], Diabetes Typ 2 [Gelman2007]. Die hohe Bedeutung von PPARs für therapeutische Anwendungen ist in zahlreichen Reviews dargelegt [Willson2000, Kersten2000, Desvergne2004].

PPARs bestehen im Wesentlichen aus einer N-terminalen A/B Domäne, der DNA-Bindedomäne (DBD), einer Hinge-Region und einer Liganden-Bindedomäne (LBD) [Sher1993]. Die A/B Domäne von PPAR α ist 100 Aminosäuren (AS) lang und ist für die Liganden-unabhängige Aktivierung verantwortlich. Die DBD umfasst 66 AS und enthält 2 Zink-Finger-Motive, die für die DNA-Erkennung und -Bindung verantwortlich sind. Die Hinge-Region ist 78 AS lang und dient als Scharnier (engl. *hinge*) zwischen der DBD und LBD. Die 224 AS lange LBD ist für die Ligandenabhängige Aktivierung verantwortlich. Die Struktur der LBD von PPAR α ist in Abbildung 3 gezeigt.

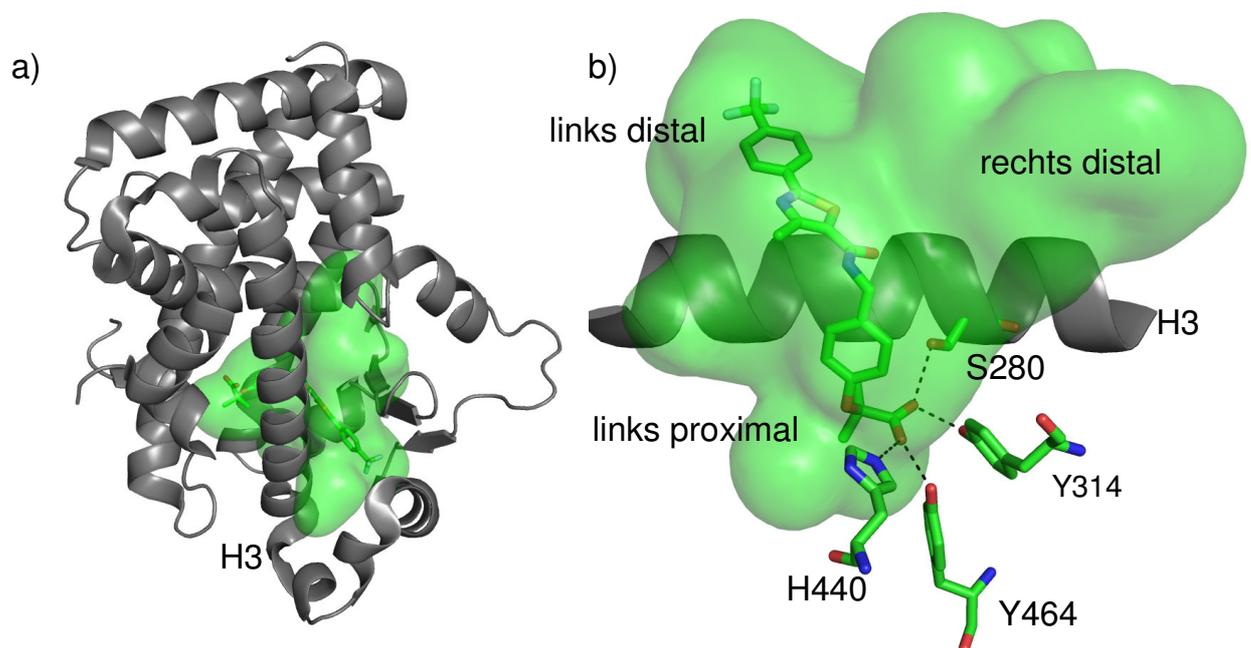


Abbildung 3: Röntgenkristallstruktur von PPAR α (PDB Struktur 2P54 [Sierra2007]). a) Die LBD von PPAR α , die tief vergrabene Bindetasche ist in grün dargestellt. b) Die schematische Unterteilung der Bindetasche. Die acide Kopfgruppe des Liganden interagiert mit 4 Aminosäuren (mit schwarzen Strichlinien gekennzeichnet).

Die Strukturen aller PPAR Subtypen wurden durch Röntgenkristallographie aufgeklärt. Im Mai 2008 waren 50 Strukturen in der Protein Data Bank (PDB) [Berman2000] verzeichnet. Die Bindetasche ist mit 1300-1400 Å³ sehr groß und vergraben [Zoete2007]. Das durchschnittliche Volumen der Bindetaschen, die von einem Liganden besetzt sind, liegt bei 610 Å³ [An2005]. Für eine bessere Übersicht wird die Bindetasche in mehrere Subtaschen unterteilt (Abbildung 3b) [Pirard2003]. Für die Aktivierung ist die Konformation der AF-2 Helix entscheidend (Abbildung 4), diese

wird in der aktiven Konformation durch die Interaktionen des Liganden mit vier Resten stabilisiert (S280, Y314, H440, Y464 in PPAR α [Sierra2007]; S289, H323, Y473, H449 in PPAR γ [Xu2001]). Diese Interaktionen werden von der meist aciden Kopfgruppe der Liganden eingegangen. Die Liganden gehen eine charakteristische U-Konformation ein, indem sie sich um die Helix H3 legen [Xu1999].

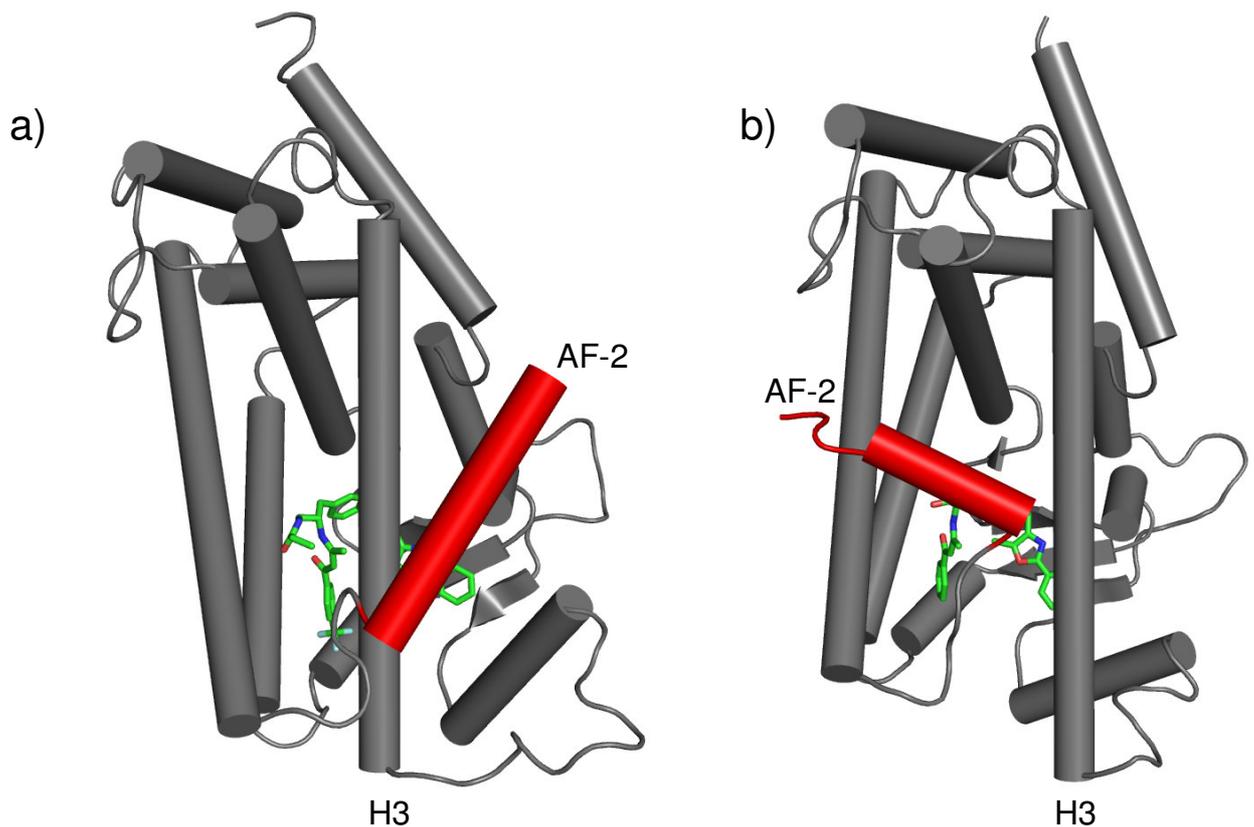


Abbildung 4: Konformationsänderung von PPAR α nach der Bindung eines Agonisten. Die Bewegung der AF-2 Helix (rot) ist wichtig für die Bindung an die DNA. a) Bindemodus des Antagonisten GW6471 (PDB Code 1KKQ [Xu2002a]) b) Bindemodus des Agonisten GW409544 (PDB Code 1K7L [Xu2001]).

Die meisten PPAR-Agonisten sind Fettsäureanaloge und zeigen den typischen Aufbau, bestehend aus einer aciden Kopfgruppe, einem aromatischen Kern, einem Linker und einem hydrophoben Rest [Kuhn2006]. Die Selektivität zwischen den Subtypen wird meist durch die Ausnutzung der Unterschiede in den linken proximalen und distalen Subtaschen erreicht [Markt2007]. Der Unterschied wird in der Abbildung 5 deutlich: die linke proximale Subtasche von PPAR α ist deutlich kleiner, Selektivität für PPAR γ wird durch den großen Benzophenon-Substituenten erreicht. Y314 ist bei PPAR γ durch

den kleineren Rest H323 ausgetauscht, dadurch sind große Kopfgruppen wie im Fall der Thiazolidindione selektiv für PPAR γ .

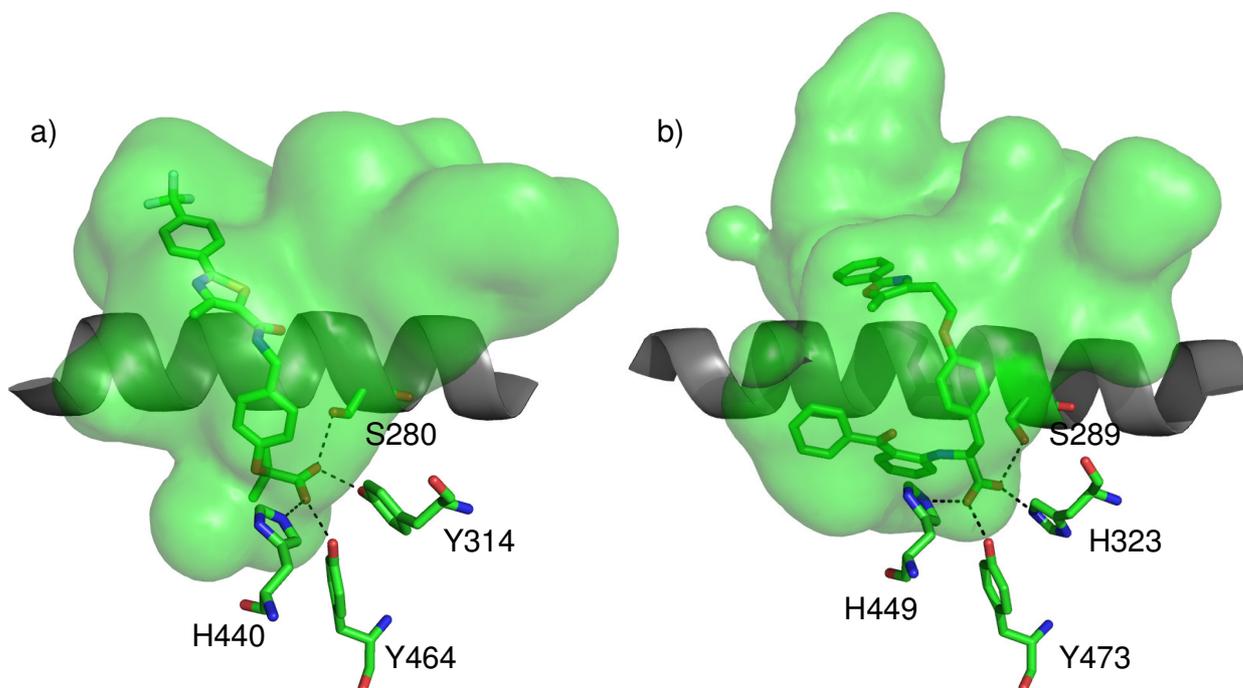
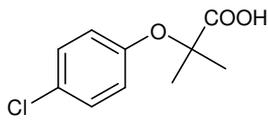
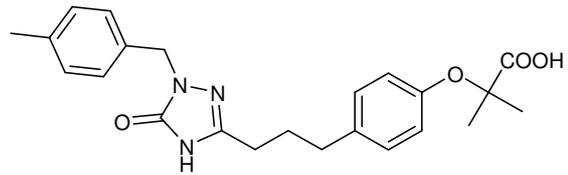


Abbildung 5: Unterschiede in der Bindetasche von PPAR α und PPAR γ . Die linke proximale Subtasche ist bei PPAR γ größer und kann Substituenten mit größeren Volumina aufnehmen. a) Gebundene Konformation des PPAR α -Agonisten GW590735 5 (PDB Code 2P54 [Sierra2007]) in der Bindetasche (grün). B) Gebundene Konformation des PPAR γ -Agonisten Farglitazar 8 (PDB Code 1FM9 [Gampe2000]).

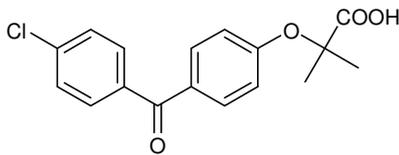
Typische PPAR α -Agonisten sind Derivate der Fibrinsäure. Sie zeichnen sich durch ein gemeinsames strukturelles Element, den 2-Hydroxy-2-Methylpropionsäure-Rest aus [Sierra2007]. Die Moleküle Clofibrat **1**, Fenofibrat **2** und Bezafibrat **3** sind seit den 60er Jahren als Triglycerid-Senker eingesetzt worden [Svoboda1966, Forman1997]. Die Fibrate **1-3** sind jedoch vergleichsweise schwache Agonisten von PPAR α [Willson2000]. Zwei der kürzlich entwickelten Fibrate LY518674 **4** [Singh2005] und GW590735 **5** [Sierra2007] erreichten die klinische Entwicklung, die jedoch für LY518674 nach der klinischen Phase II abgebrochen wurde [Rau2008]. Die Fibrate LY518674 und GW590735 zeigen hohe Potenz und große Selektivität gegenüber PPAR γ (EC_{50} =24,5 nM und 310fache Selektivität bei LY518674, EC_{50} = 4 nM und mind. 500fache Selektivität bei GW590735).



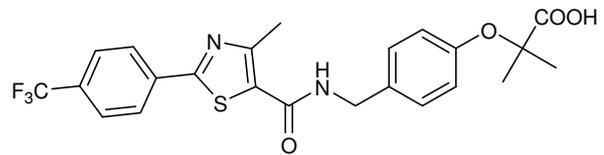
1, Clofibrate



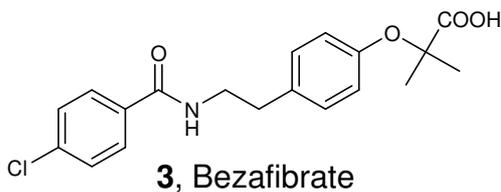
4, LY518674



2, Fenofibrate

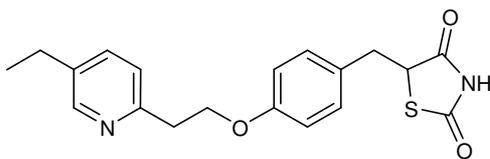


5, GW590735

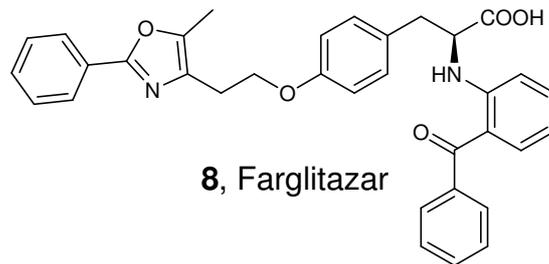


3, Bezafibrate

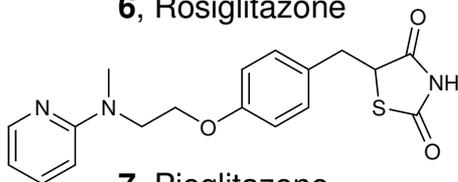
Typische PPAR γ -Agonisten sind Thiazolidindione [Lehmann1995] und Tyrosin-Analoga [Henke1998]. Die relevantesten Vertreter der beiden Klassen sind Rosiglitazon **6**, Pioglitazon **7** und Farglitazar **8**. Alle zeigen beträchtliche Potenz und Selektivität gegenüber PPAR α [Xu2001] (Rosiglitazon: EC_{50} = 18 nM und mind. 500fache Selektivität gegenüber PPAR α , Pioglitazon: EC_{50} = 280 nM mind. 35fache Selektivität gegenüber PPAR α , Farglitazar: EC_{50} = 0,2 nM und 1250fache Selektivität gegenüber PPAR α).



6, Rosiglitazone



8, Farglitazar



7, Pioglitazone

Die PPAR-Agonisten Fenofibrat, Gemfibrozil, Clofibrat, Bezafibrat, Ciprofibrat, Pioglitazon und Rosiglitazon sind bereits auf dem Markt. Die Nebenwirkungen der

bereits zugelassenen Substanzen, wie z.B. Gewichtszunahme, Wassereinlagerung, Myopathie, Homocysteinämie und Creatininämie machen jedoch die Erforschung neuer Substanzklassen notwendig [Rubenstrunk2007, Shearer2007].

Rechnergestützte Wirkstoffdesign-Methoden bei der Suche nach neuen PPAR-Agonisten

Die therapeutische Relevanz, eine vergleichsweise hohe Anzahl aufgelöster Röntgenkristallstrukturen, ein breites Spektrum bekannter Leitstrukturen und etablierte Testsysteme machen PPARs zu einem bevorzugten Modell-Target zur Evaluierung neuer Methoden für das rechnergestützte Wirkstoffdesign. Mehrere aktuelle Studien zeigen den erfolgreichen Einsatz von Virtual Screening bei der Suche nach neuen PPAR-Liganden [Seifert2007].

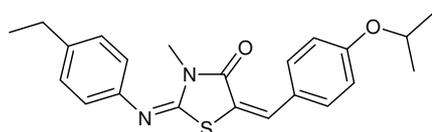
Derksen *et al.* [Derksen2006] stellten basierend auf unterschiedlichen zweidimensionalen Deskriptoren mit Hilfe maschineller Lernverfahren Modelle für PPAR-Agonisten auf und konnten dabei mehrere Hits aus der SPECS Datenbank (www.specs.net, SPECS, Delft, Niederlande) im zellbasierten Aktivierungsassay bestätigen. Das bei dieser Studie gefundene Molekül **9** aktivierte PPAR γ signifikant bei 100 μ M.

Scarsi *et al.* [Scarsi2007] bedienten sich der Röntgenkristallstruktur von PPAR γ und konnten die Klasse der N-Acylsulfonamide als potente PPAR γ -Agonisten identifizieren. Die durch Docking gefundene Struktur **10** aktiviert PPAR γ bei pEC_{50} von 5.

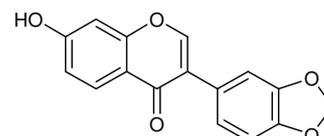
Lu *et al.* [Lu2006] wandten die Software Catalyst (Catalyst v. 4.7, Accelrys Inc, San Diego, USA) unter Berücksichtigung der Form des Liganden an [Hahn1997] und konnten in der Maybridge Datenbank (Maybridge Chemical Company, Tintagel, UK) potente selektive PPAR γ -Agonisten identifizieren. Die nachfolgende Suche der nächsten Nachbarn des ersten Hits führte in der zweiten Screening-Runde zu dem Sulfonamid-Derivat **11**, das mit einem IC_{50} von 22,7 nM an PPAR γ bindet.

Salam *et al.* [Salam2008] durchsuchten eine Naturstoff-Datenbank mit einem Docking-Verfahren und konnten dabei mehrere Flavonoide als PPAR γ -Agonisten identifizieren. Erwähnenswert ist hier *psi*-Baptigenin **12**, welches PPAR γ mit einem EC_{50} von 2,9 μ M aktivierte.

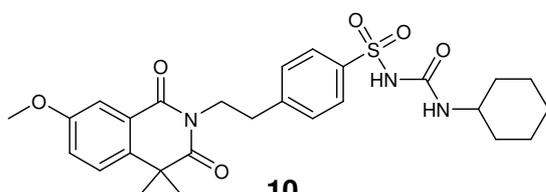
Li *et al.* [Li2007] führten eine Dockingstudie mit einer virtuellen kombinatorischen Bibliothek Chalkon-abgeleiteter Fibrate durch. Die ausgewählten und synthetisierten Verbindungen waren potente PPAR α -Agonisten, die beste Struktur **13** aktivierte PPAR α mit $EC_{50} = 60$ nM.



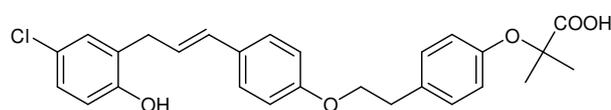
9



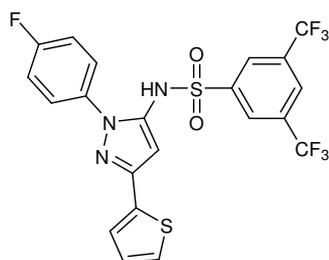
12, *psi*-Baptigenin



10



13



11

Das primäre Ziel dieser Arbeit ist die Entwicklung eines Verfahrens für das virtuelle Screening und das *de novo* Design. Im Vordergrund steht dabei die Anwendbarkeit des neuen Verfahrens in der Leitstrukturfindung neuer PPAR α -Agonisten. Wichtig für die Beurteilung der Methoden ist die bestätigte Aktivität der gefundenen Substanzen im zellbasierten Testsystem, welches im Arbeitskreis von Prof. Schubert-Zsilavec etabliert ist [Rau2006].

Theoretische Methoden

Shape Deskriptoren und Shape Matching

Die sterische Komplementarität der Bindetasche und des Liganden ist eines der Grundprinzipien der molekularen Erkennung [Mezey1993]. Fast alle Docking-Ansätze berücksichtigen dies implizit. Was aber, wenn die Strukturinformation nicht vorhanden oder für das Docking unbrauchbar ist? Insbesondere die Beteiligung von Wassermolekülen an Protein-Ligand-Interaktionen, die Flexibilität des Proteins und die unzureichende Auflösung der Struktur erschweren den Einsatz von Docking-Software [Sousa2006]. Die liganden-basierten Virtual Screening- und *de novo* Design-Techniken setzen darauf, dass Moleküle mit ähnlichen Eigenschaften auch ähnliche biologische Aktivität aufweisen [Johnson1989, Willett1998]. Bezogen auf die sterische Komplementarität kann also daraus gefolgert werden, dass Moleküle mit ähnlicher Form sich in ihrer biologischen Aktivität ähneln.

Diese Ähnlichkeit der Form (engl. *shape similarity*) wurde in mehreren Ansätzen berücksichtigt. So beschreibt Arteca [Arteca1996] unterschiedliche Möglichkeiten, die Form mittels Shape-Deskriptoren zu kodieren. Ein Beispiel für solche Shape-Deskriptoren ist die Globularität (engl. *globularity*), definiert als das Verhältnis des ersten und des dritten Eigenwertes der Kovarianzmatrix der Atomkoordinaten [Grabowski2007]. Liegt die Globularität nah bei 1, ist das Molekül sphärisch geformt. Mehrere Verfahren beschreiben die Shape-Ähnlichkeit über den Überlapp der Volumen. Hahn [Hahn1997] schlägt vor, die Volumendifferenz in einem 3D-Gitter als Ähnlichkeitsmaß zu benutzen. Diese Methode wurde als Ergänzung der Catalyst Software [Greene1994] implementiert. Eine weitere Lösung wurde von Grant *et al.* [Grant1995] entwickelt. Hierbei wird das Volumen des Moleküls als die Summe Gauß'scher Funktionen dargestellt. Dadurch kann die Volumendifferenz analytisch berechnet und die Minimierung der Volumendifferenz wesentlich beschleunigt werden [Grant1996]. Diese Methode ist in der kommerziellen Software ROCS (<http://www.eyesopen.com/products/applications/rocs.html>, Openeye Scientific Software, Santa Fe, USA) verfügbar.

Einen anderen Ansatz verfolgten Mavridis *et al.* [Mavridis2007], indem sie die molekulare Oberfläche über sphärische harmonische Koeffizienten darstellen und vergleichen. Der Vorteil dieses Vorgehens ist die Reduzierung der Freiheitsgrade bei der Überlagerung von 6 (Rotation und Translation) auf 3 (ausschließlich Rotation).

Diese Verfahren wurden erfolgreich im Bereich des Virtual Screenings eingesetzt [Rush2005, Bologna2006, Boström2007]. Leider bieten diese Verfahren jedoch keine Möglichkeit, molekulare Fragmente mit einem Referenzmolekül sinnvoll zu überlagern und diese zu bewerten. Dies ist aber für fragment-basierten *de novo* Design zwingend notwendig. Zwei Ansätze, die solches Vorgehen ermöglichen würden, wurden bereits entwickelt. Cosgrove *et al.* [Cosgrove2000] haben die Krümmung an jedem Punkt der molekularen Oberfläche berechnet und Punkte mit ähnlicher Krümmung mit Hilfe eines Clustering-Algorithmus zusammengefasst. Hofbauer *et al.* [Hofbauer2004] entwickelten ein ähnliches Verfahren, mit dem Unterschied, dass die Punkte nicht zusammengefasst, sondern Maxima und Minima betrachtet werden. Beide Verfahren benutzen Graphen-Algorithmen für die Überlagerung der Moleküle und erlauben daher ein partielles Matching der Oberflächen, können also potentiell für die Überlagerung von Molekülfragmenten dienen. Diese Verfahren wurden aber als zeitintensiv beschrieben und sind für große Datenmengen ungeeignet. Deshalb wurde in der vorliegenden Arbeit ein Verfahren entwickelt, welches dem von Cosgrove und Hofbauer ähnlich ist, dieses jedoch in der Geschwindigkeit übertrifft.

Entwickelte Methoden

Shapelets

Oberflächengenerierung

Der erste Schritt der *Shapelets*-Methode ist die Generierung der molekularen Oberfläche. Das Molekül M wird dabei durch eine kontinuierliche Funktion im dreidimensionalen Raum beschrieben. Dabei wird das etablierte Konzept [Blinn1982, Grant1995] verwendet, das Molekül M durch eine Summe modifizierter Gauß'scher Funktionen zu modellieren (Formel 1). \vec{x} steht dabei für die räumlichen Koordinaten, N für die Anzahl der Atome im Molekül, \vec{c}_n für die Atomkoordinaten, und r_n für den van-der-Waals Radius des n -ten Atoms.

$$M(\vec{x}) = \sum_{n=1}^N e^{-\frac{2(\vec{x}-\vec{c}_n)^2}{r_n^2}}$$

Formel 1

Die Werte von $M(\vec{x})$ werden an jedem Punkt eines kubischen Gitters berechnet, in dem sich das Molekül M befindet. Dieses Gitter wird vom *Marching Cubes* Algorithmus [Lorensen1987] benutzt, um eine Isofläche mitsamt Normalen an jedem Oberflächenpunkt zu extrahieren.

Marching Cubes ist ein Algorithmus, der zur Volumenvisualisierung genutzt wird. Die Eingabe besteht aus einer Funktion im 3D-Raum $f : \mathfrak{R}^3 \rightarrow \mathfrak{R}$ und einem Isoflächenwert c (engl.: *cutoff*). Die Funktionswerte an allen Punkten eines dreidimensionalen Gitters (Abbildung 6a) werden berechnet. Nun iteriert der Algorithmus schichtweise über alle Zellen des Gitters. Für jede Zelle wird ein binärer Index berechnet. Liegt der Funktionswert an einem Gitterpunkt über dem Isoflächenwert c , so trägt er zum Index mit 1 bei, sonst mit 0 (Abbildung 6b). Dieser Index wird dazu verwendet, um in einer Tabelle nachzuschauen, welche Oberflächendreiecke erzeugt werden müssen. Durch Interpolation werden dann die Oberflächenpunkte erzeugt und gemäß dem gefundenen Tabelleneintrag durch Kanten verbunden (Abbildung 6c).

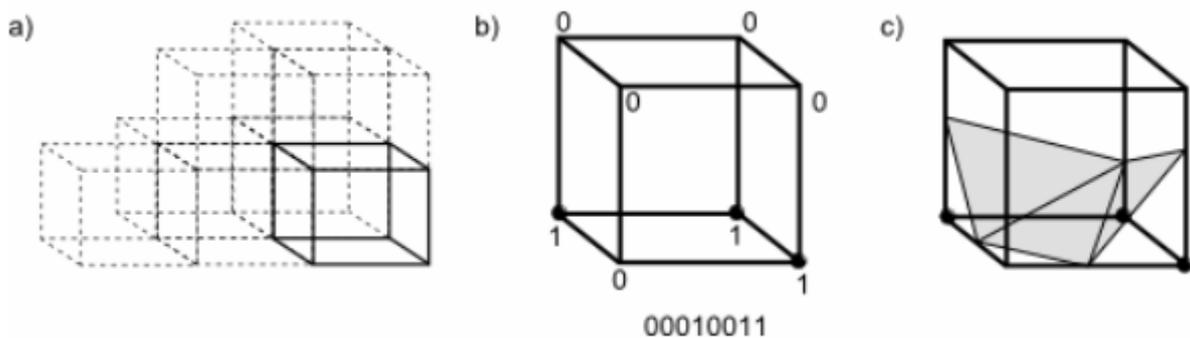


Abbildung 6: a) Ein 3-dimensionales Gitter, an dessen Eckpunkten die Funktionswerte berechnet werden. Einer der Zellen wird vom *Marching Cubes* Algorithmus herausgegriffen (fett). b) Die Eckpunkte der Zelle mit Werten über dem Schwellenwert c (umkreist) erhalten den Index 1, die übrigen 0. c) Die Indizierung wird zum Erzeugen der Oberflächendreiecke verwendet.

Halbiert man den Gitterabstand, verachtfacht sich die Rechenzeit. Der Gitterabstand von 0.5 \AA stellt dabei ein Kompromiss aus Rechenzeit und Glätte der generierten Oberfläche

dar. Der Marching Cubes Algorithmus extrahiert die Isofläche der Funktion $M(\vec{x})$ an jeder Gitterzelle, die einen bis sieben Eckpunkte mit Werten über dem Isoflächen-Wert von $c = e^{-2}$ besitzt. Diese Isofläche entspricht offensichtlich der van-der-Waals Oberfläche für einzelne Atome, ist an den Übergängen zwischen den Atomen jedoch glatter (Abbildung 7).

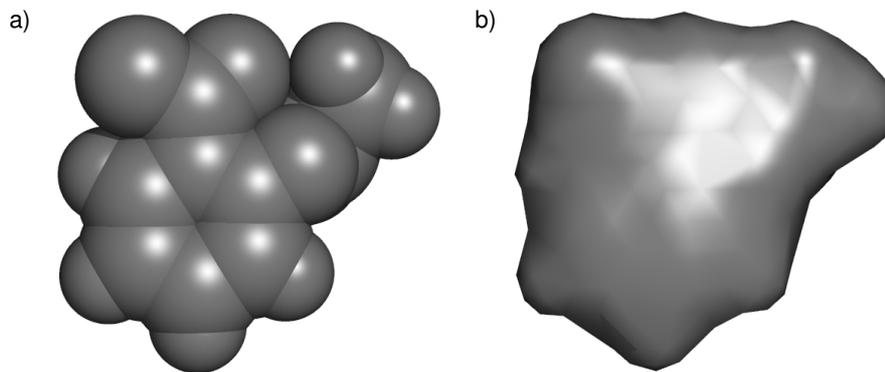


Abbildung 7: a) Die van-der-Waals Oberfläche und b) die vom Marching Cubes Algorithmus extrahierte Isofläche der Acetylsalicylsäure.

Ein Nachteil des Marching Cubes Algorithmus ist die Generierung einer großen Anzahl von Oberflächenpunkten, die teilweise ungleichmäßig verteilt sind [Schroeder1992]. Um diesen Nachteil auszugleichen, wurde ein einfacher Glättungsalgorithmus, genannt *Welding Vertices* [Dunn2002], eingesetzt. Dieses Verfahren fasst rekursiv Oberflächenpunkte mit einem Abstand, der geringer ist als eine gesetzte Untergrenze, zusammen. Wählt man die Untergrenze gemäß dem Abtasttheorem [Shannon1998] als die Hälfte des Gitterabstandes, also $0,25 \text{ \AA}$, reduziert sich die Anzahl der Oberflächenpunkte auf ca. 60%.

Oberflächendekomposition

Die in *Shapelets* implementierte Oberflächendekomposition ist ein deterministisches Verfahren, um die Oberfläche zu segmentieren. Für diesen Zweck wurde ein von Zachmann [Zachmann1992] vorgestelltes Verfahren für die Approximation der lokalen Oberflächenkrümmung adaptiert. Ein hyperbolischer Paraboloid wird dabei in einen Oberflächenausschnitt mit dem Radius r_S eingepasst. Die Wahl des Parameters $r_S = 2,5 \text{ \AA}$ wurde durch visuelle Inspektion der Dekompositionsergebnisse festgelegt. An jedem Punkt p_r der triangulierten Oberfläche wird der Oberflächenausschnitt als die Menge aller Punkte p_i mit einer Oberflächendistanz $d_M(p_r, p_i) < r_S$ festgelegt. Die

Oberflächendistanz wird als der kürzeste Weg zwischen den Punkten p_r und p_i entlang der Oberflächenkanten berechnet. Um den hyperbolischen Paraboloiden in den so definierten Oberflächenausschnitt einzupassen, wird jeder Punkt p_i zunächst in das sogenannte lokale paraboloidale Koordinatensystem transformiert. Dieses wird durch drei Einheitsvektoren e_n , e_u und e_v aufgespannt. e_n ist durch die Oberflächennormale am Punkt p_r definiert. e_u und e_v sind zwei beliebige Einheitsvektoren in der Tangentialebene (Formel 2), wobei p ein beliebiger Punkt des Oberflächenausschnitts ist.

$$e_u = \frac{e_n \times (p - p_r)}{\|e_n \times (p - p_r)\|} \quad \text{Formel 2}$$

$$e_v = e_n \times e_u$$

Das hyperbolische Paraboloid selbst ist durch die Formel 3 definiert, wobei u und v die Koordinaten im lokalen paraboloiden Koordinatensystem sind, und \mathbf{H} die 2×2 Hess'sche Matrix ist.

$$p(u, v) = \begin{pmatrix} u & v \end{pmatrix} \mathbf{H} \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{Formel 3}$$

Die Koeffizienten von \mathbf{H} können durch die Minimierung des quadratischen Fehlers E (Formel 4) erhalten werden, wobei n_i , u_i and v_i die Koordinaten jedes Punktes p_i im paraboloiden Koordinatensystem sind.

$$E(\mathbf{H}) = \sum_i (n_i - p(u_i, v_i, \mathbf{H}))^2 \quad \text{Formel 4}$$

Um den Paraboloiden an den Oberflächenausschnitt anzupassen, muss $E(\mathbf{H})$ minimiert werden. Das Minimum von $E(\mathbf{H})$ gibt die Güte der Anpassung wieder. Die beiden Eigenwerte der Hess'schen Matrix \mathbf{H} entsprechen der maximalen und minimalen lokalen Oberflächenkrümmung k_1 und k_2 . Die Oberflächenpunkte werden dann nach $E(\mathbf{H})$ sortiert. Der Algorithmus nimmt nun den Punkt mit dem geringsten $E(\mathbf{H})$ und entfernt den Oberflächenausschnitt um diesen Punkt aus der gesamten Oberfläche durch Löschen der Oberflächenpunkte aus der Datenstruktur. Dies wird rekursiv fortgesetzt,

bis keine Oberflächenausschnitte mit Radius $\geq r_s$ übrig sind. Die Mittelpunkte der Oberflächenausschnitte werden als charakteristische Punkte bezeichnet. Aus der maximalen und minimalen lokalen Oberflächenkurvatur k_1 und k_2 an den charakteristischen Punkten kann der von Duncan und Olson [Duncan1993] definierte „shape index“ SI (Formel 5) berechnet werden.

$$SI(p_r) = \text{atan}\left(\frac{k_1 + k_2}{k_1 - k_2}\right) \quad \text{Formel 5}$$

Abbildung 8 zeigt den Ablauf der Oberflächengenerierung und der Oberflächendekomposition im zweidimensionalen Raum. Das Ergebnis der Dekomposition ist somit eine Menge von charakteristischen Punkten mit den dazugehörigen Oberflächennormalen und den SI -Werten, welche für die Überlagerung von Molekülen zur Verfügung stehen.

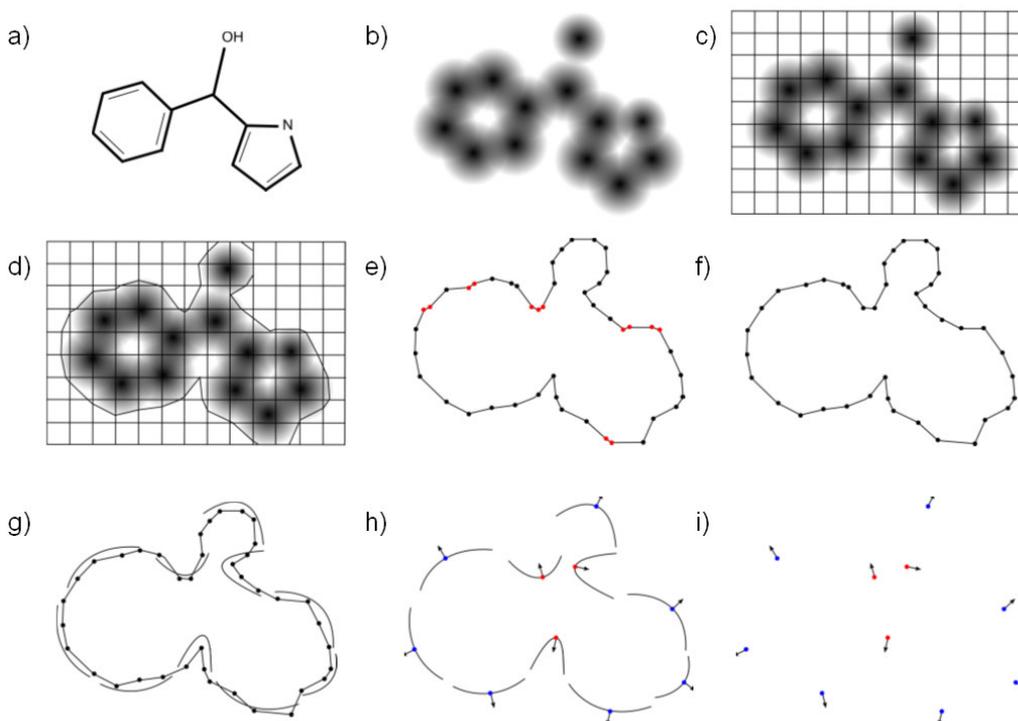


Abbildung 8: Die einzelnen Schritte der *Shapelets*-Berechnung. a) Das Ausgangsmolekül. b) Jedes Atom wird durch eine Gauß'sche Funktion approximiert. c) Die Funktionswerte werden auf ein Gitter projiziert. d) Die Isofläche wird mit dem *Marching Cubes* Algorithmus extrahiert. e) und f) Die Oberfläche wird mit dem *Welding Edges* Algorithmus geglättet. g) An jedem Punkt der

Oberfläche wird ein hyperboler Paraboloid an den lokalen Oberflächenausschnitt angepasst. h) und i) Die Mittelpunkte und die Normalen der ausgewählten Paraboloiden werden gespeichert.

Exner [Exner2002] definierte fünf Klassen von Oberflächenausschnitten abhängig vom Wert von SI (Abbildung 9): *gleichmäßig konvex* (Pfropf, $SI \approx 1.0$), *langgestreckt konvex* (Grat, $SI \approx 0.5$), *flach* (Fläche, $SI \approx 0.0$), *langgestreckt konkav* (Spalt, $SI \approx -0.5$), und *konkav* (Loch, $SI \approx -1.0$). In der Implementierung wurden nur Pfropfe ($SI = [0.9..1.0]$) und Flächen ($SI = [-0.1..0.1]$) verwendet. Spalten und Löcher tauchen in kleinen Molekülen selten auf, die Verwendung von Graten verursacht Rauschen, da diese besonders gut in kleinere Unebenheiten der Oberfläche passen. Bei der Dekomposition werden somit zwei Arten von Paraboloiden gespeichert: Flächen und Pfropfe. Die zu den charakteristischen Punkten gehörigen hyperbolen Paraboloiden werden „*Shapelets*“ genannt, namensgebend für das gesamte Verfahren zur Molekülüberlagerung.

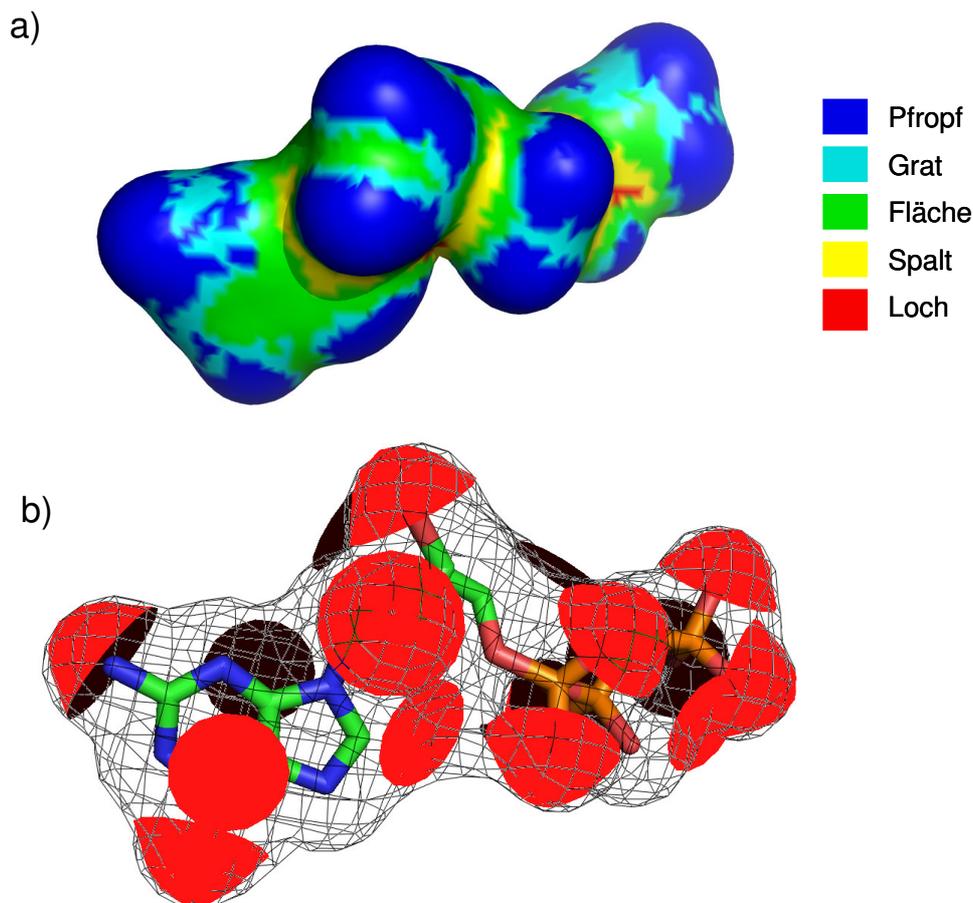


Abbildung 9: a) Die Oberfläche von Guanosinriphosphat (GTP) eingefärbt nach *Shape Index* (SI). b) Die *Shapelets*-Dekomposition der Oberfläche von GTP.

Überlagerung der Moleküle

Die aus der Dekomposition erhaltenen charakteristischen Punkte können nun für die Überlagerung zweier Moleküle basierend auf ihren Oberflächeneigenschaften benutzt werden. Eine Überlagerung zweier Moleküle sollte eine Vielzahl überlappender Oberflächenausschnitte ähnlicher Form aufweisen. Um eine solche Überlagerung zu finden, wird ein Verfahren aus der Graphentheorie, die Cliquendetektion auf dem Assoziationsgraphen, benutzt. Diese Methode wurde bereits von mehreren Forschungsgruppen zum Vergleich molekularer Oberflächen verwendet [Gardiner2000, Hofbauer2004]. Im Folgenden sollen einige grundlegende Definitionen aus der Graphentheorie angeführt werden, welche für das Verständnis der Methode notwendig sind.

Sei F_V die Menge der Knoteneigenschaften und F_E die Menge der Kanteneigenschaften. Ein **Graph** ist ein 4-Tupel $G = (V, E, \mu_V, \mu_E)$, wobei [Messmer1995]

- V eine endliche Menge der Knoten (engl. *vertices*) bezeichnet,
- $E \subseteq V \times V$ die Menge der Kanten (engl. *edges*) bezeichnet,
- $\mu_V: V \rightarrow F_V$ eine Funktion ist, die Knoten Eigenschaften zuweist,
- $\mu_E: V \rightarrow F_E$ eine Funktion ist, die Knoten Eigenschaften zuweist.

Ein Graph heißt **vollständig** oder **vollständig verbunden**, wenn jeder Knoten mit einem anderen Knoten verbunden (adjazent) ist [Gasteiger2003].

Sei $G = (V, E, \mu_V, \mu_E)$ ein Graph, dann heißt $G' = (V', E', \mu_{V'}, \mu_{E'})$ ein **Subgraph** von G , wenn

- $V' \subseteq V$,
- $E' \subseteq E$,
- $\mu_{V'} = \mu_V$ wenn $v \in V'$,
- $\mu_{E'} = \mu_E$ wenn $e \in E'$.

Die charakteristischen Punkte bilden dabei die Knoten V eines vollständig verbundenen Graphen $G = (V, E, \mu_V, \mu_E)$, mit den Knoteneigenschaften $F_V = SI$. Die Kanteneigenschaften F_E entsprechen den Abständen zwischen den charakteristischen Punkten (Abbildung 10a). Gegeben seien zwei vollständig verbundene Graphen $G = (V, E, \mu_V, \mu_E)$ und $G' = (V', E', \mu_{V'}, \mu_{E'})$. Nun bildet man den dazugehörigen Assoziationsgraphen $A = (V_a, E_a)$ with $V_a \subseteq V \times V'$ in folgender Weise:

- Für jedes Paar von Knoten $v \in V$ and $v' \in V'$ füge einen Knoten v_a hinzu, wenn SI von v and v' sich um weniger als c_{SI} unterscheiden.
- Zwei Knoten $v_a = (v, w)$, $v_b = (v', w') \in V_a$ sind durch eine Kante verbunden, wenn die Längen der dazugehörigen Kanten $(v, v') \in E$ und $(w, w') \in E'$ sich um weniger als c_D unterscheiden und außerdem gilt: $v \neq w$ und $v' \neq w'$.

Eine **Clique** ist ein vollständig verbundener Subgraph. Eine Clique in A entspricht einer räumlichen Übereinstimmung der charakteristischen Punkte der beiden Moleküle (Abbildung 10b). Das Problem der Cliquendetektion ist NP -vollständig [Hopcroft2001]. Trotzdem ist in der Praxis die Berechnung mit dem Bron-Kerbosch-Algorithmus [Bron1973] bei vorliegenden Graphgrößen hinreichend schnell [Brint1987, Gerhards1979]. Es handelt sich dabei um einen rekursiven Backtracking-Algorithmus, der alle maximalen Cliques in einem Graphen aufzählt. Im Wesentlichen operiert der BK-Algorithmus auf drei Mengen C , P und S .

Die Menge C (Clique) ist die Menge der Knoten, die bereits als komplett verbunden gefunden wurden, also eine Clique bilden. Die Menge P (Potential) enthält diejenigen Knoten, die zu dem zuletzt zu C hinzugekommenen Knoten adjazent sind, folglich also für die Erweiterung der Clique benutzt werden können. Die Menge S enthält all die Knoten, die nicht mehr für die Erweiterung von C benutzt werden können, weil alle Cliques, die diese Knoten enthalten, bereits gefunden wurden. Der Pseudocode aus [Koch2001] gibt die Funktionsweise des BK-Algorithmus wieder.

```

function find_all_cliques( $C, P, S$ )
//enumerates all cliques in an arbitrary graph  $G$  //
 $C$ : set of vertices belonging to the current clique
 $P$ : set of vertices which can be added to  $C$ 
 $S$ : set of vertices which are not allowed to be added to  $C$ 
 $N[u]$ : set of vertices which are adjacent to vertex  $u$  in  $G$ 
01: Let  $P$  be the set  $\{u_1, \dots, u_k\} \subseteq V$ ;
02: if  $P = \emptyset$  and  $S = \emptyset$ 
03:     then  $C$  is a maximal clique
04:     else for  $i \leftarrow 1$  to  $k$ 
05:         do  $P \leftarrow P \setminus \{u_i\}$ ;
06:              $P' \leftarrow P$ ;
07:              $S' \leftarrow S$ ;
08:              $N \leftarrow \{v \in V \mid \{u_i, v\} \in E\}$ ;
09:             find_all_cliques( $C \cup \{u_i\}, P' \cap N, S' \cap N$ );
10:              $S \leftarrow S \cup \{u_i\}$ ;
11:         end;
12:end;
end function

```

Der Algorithmus wird mit $C, S = \emptyset$ und $P=V$ initialisiert. Wenn die Mengen S und P leer sind, ist C eine maximale Clique und kann ausgegeben oder gespeichert werden (Zeile 03). Ist dies nicht der Fall, so wird jeder Knoten aus P in der Schleife (Zeilen 04-11) bearbeitet. Ein ausgesuchter Knoten u_i wird aus P entfernt (Zeile 05), P wird nach P' (Zeile 06) und S nach S' (Zeile 07) kopiert, um für den rekursiven Aufruf verwendet zu werden. Knoten, die mit u_i verbundenen sind, werden in N gespeichert (Zeile 08). Der rekursive Aufruf erfolgt mit der Menge $C \cup \{u_i\}$ (die bereits gefundene Clique wird um u_i erweitert), $P \cap N$ (nur die Knoten in P , die mit u_i verbunden sind, dürfen für die potentielle Erweiterung der Clique verwendet werden) und $S \cap N$ (alle Knoten, die bereits betrachtet wurden und mit u_i verbunden sind) (Zeile 09). Anschließend wird u_i

der Menge S zugefügt (Zeile 10). Der formale Korrektheitsbeweis ist in [Koch2001] aufgeführt.

Der BK-Algorithmus wurde bereits in vielen bioinformatischen Zusammenhängen verwendet und modifiziert [Grindley1993, Samudrala1998, Koch2001]. Die besondere Geschwindigkeit der sog. „Version2“ des BK-Algorithmus [Bron1973] liegt in einer Heuristik bei der Wahl des Knotens u_i aus P . Diese motiviert sich aus der Verwendung der Menge S , da man die Abbruchbedingung für die Ausgabe der Clique auch so formulieren kann: Es gibt einen Knoten s in S , der mit allen Knoten in P verbunden ist. Ein solcher Knoten s kann niemals aus S entfernt werden. Somit haben wir eine Möglichkeit, vorzeitig vorherzusagen, ob ein Ast im Backtrackingbaum zum Erfolg führt oder nicht. Allgemein nennt man diese Backtracking-Technik die „branch-and-bound“ Methode, das Abbruchkriterium selbst ist das „branch-and-bound“ Kriterium [Schöning2001]. Die „Version2“ des BK-Algorithmus sucht u_i aus P so aus, dass das „branch-and-bound“ Kriterium möglichst schnell erfüllt wird. Dies geschieht in folgender Weise: alle Knoten in S sind mit einem Zähler assoziiert, welcher die Anzahl der Kanten zu den Knoten in P wiedergibt. Verschiebt man nun einen Knoten p aus P nach S , so vergrößern sich die Zähler aller Knoten in S höchstens um 1 und p bekommt einen neuen Zähler. Das „branch-and-bound“ Kriterium kann nun folgendermaßen formuliert werden: Es gibt einen Knoten s in S mit Zähler $z = |P|$. Bron *et al.* [Bron1973] zeigen, dass man dieses Kriterium am schnellsten erreicht, wenn man u_i aus P so auswählt, dass u_i mit demjenigen Knoten s in S durch eine Kante verbunden ist, welcher den höchsten Zähler z hat. Dadurch erhöht man in jedem Schritt den Zähler von s um 1, das „branch-and-bound“ Kriterium wird auf dem schnellstmöglichen Wege erreicht.

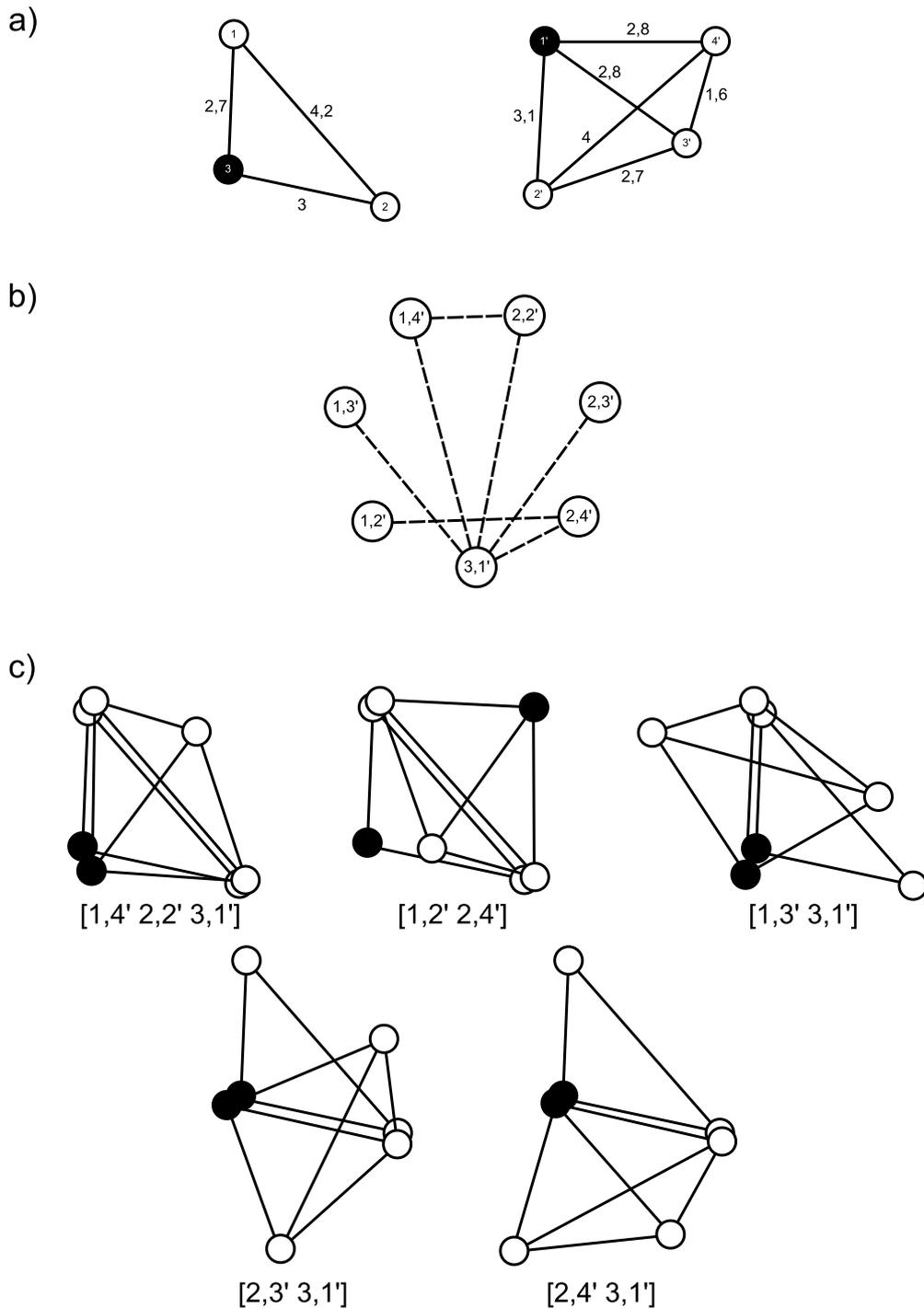


Abbildung 10: Prinzip der Cliquendetektion auf dem Assoziationsgraphen. a) Zwei Graphen, die es zu überlagern gilt. Die Knotentypen entsprechen der Einfärbung (schwarz und weiß), die Kanten sind mit dem geometrischen Abstand beschriftet. b) Der Assoziationsgraph, der aus den Graphen in a) erstellt wird. c) Alle Überlagerungen, die den Cliques des Assoziationsgraphen aus b) entsprechen, jede Clique besteht mindestens aus zwei Knoten.

Der Algorithmus findet alle Cliques in einem Graphen, daher sind mehr als eine Überlagerung möglich (Abbildung 10c). Ist eine Übereinstimmung gefunden, können die Moleküle anhand der korrespondierenden Punkte mittels der Minimierung der kleinsten Fehlerquadrate überlagert werden [Kabsch1976].

SQUIRREL

Die *Shapelets*-Methode ist in der Lage, unterschiedliche Überlagerungen zweier Moleküle herzustellen. Allerdings wird dabei nur die Form der Moleküle berücksichtigt. Zudem gibt *Shapelets* bereits bei zwei Molekülen mit jeweils einer Konformation zumeist mehrere Möglichkeiten einer Überlagerung aus. Eine Bewertungsfunktion soll nun eingeführt werden, welche die Durchführung folgender Szenarien ermöglicht:

- Gegeben seien zwei Moleküle in der bioaktiven Konformation mit gleichem Bindemodus. Finde die Überlagerung der beiden Moleküle, wie sie von der Bindetasche vorgegeben ist („rigid body alignment“).
- Gegeben seien ein Molekül in der bioaktiven Konformation und mehrere Konformere eines zweiten Moleküls, dessen bioaktive Konformation unbekannt ist. Finde diejenige Konformation des zweiten Moleküls, die potentiell der bioaktiven Konformation entspricht („semi-flexible alignment“).
- Gegeben seien zwei oder mehr Moleküle mit jeweils mehreren Konformationen. Finde die beste Überlagerung, welche im Idealfall den bioaktiven Konformationen in der Bindetasche entspricht („flexible alignment“ und „multiple flexible alignment“).
- Gegeben seien ein Molekül in der bioaktiven Konformation und eine Datenbank (10^2 - 10^7) von Molekülen mit mehreren Konformeren. Finde durch Überlagerung diejenigen Moleküle, die ähnliche biologische Aktivität wie das erste Molekül aufweisen („virtual screening“).
- Gegeben seien ein Molekül in der bioaktiven Konformation und molekulare Fragmente. Finde Überlagerungen von Teilen des Moleküls mit Fragmenten, die diese Teile ohne Verlust biologischer Aktivität ersetzen können („bioisosteric replacement“).

Die LIQUID (Ligand-based Quantification of Interaction Distributions) Bewertungsfunktion ist eine Pharmakophor-basierte Bewertungsfunktion [Tanrikulu

2007]. Eine Besonderheit von LIQUID ist die Unschärfe (engl. fuzziness) bei der Darstellung der potentiellen Pharmakophorpunkte (PPP). Im Folgenden werden die Einzelheiten der LIQUID-Bewertungsfunktion beschrieben.

Atomtypisierung

Der erste Schritt in der Generierung des Pharmakophor-Dichtefeldes ist die Zuweisung der Pharmakophortypen zu einzelnen Atomen. Folgendes Zuweisungsschema wird von der LIQUID Bewertungsfunktion benutzt:

- Als lipophil („L“) werden alle Kohlenstoffatome angesehen, die nicht an ein Heteroatom gebunden sind; außerdem Chlor-, Brom-, und Iodatome, sowie Schwefelatome, die nicht an Wasserstoffe, Stickstoffe oder Sauerstoffe gebunden sind.
- Als Wasserstoffbrückendonoren („D“) werden alle Sauerstoffe und Stickstoffe, gebunden an ein Wasserstoffatom, angesehen.
- Als Wasserstoffbrückenakzeptoren („A“) werden alle Stickstoffatome, die nicht an ein Wasserstoffatom gebunden sind; sowie alle Sauerstoffatome angesehen.

Zu beachten ist, dass einem Atom mehrere Pharmakophortypen zugewiesen werden können. So wird der Sauerstoff in einer Hydroxylgruppe sowohl als Wasserstoffbrückendonor als auch –akzeptor betrachtet. Im Gegensatz zu den Programmen LUDI [Böhm1992] und CATS [Schneider1999] wird nur zwischen 3 Atomtypen unterschieden. Die geringe Zahl der Atomtypen bringt zusätzliche Unschärfe in die Methode ein und ermöglicht das auffinden unterschiedlicher chemischer Grundgerüste (*scaffold-hopping*).

Zusammenfassen benachbarter PPPs

Um das Pharmakophor-Dichtefeld zu berechnen und die Unschärfe in die Bewertungsfunktion einzuführen, müssen benachbarte PPPs zunächst zusammengefasst werden. Dieser Vorgang wird auch als „Clustern“ oder „Clustering“ bezeichnet. Zunächst wird für jedes Atom die lokale Eigenschaftsdichte (engl. *Local Feature Density*, *LFD*) berechnet. *LFD* eines Atoms vom Pharmakophortyp *T* ist gegeben durch Formel 6, wobei *N* die Gesamtanzahl der Atome des Typs *T*, *D*₂ der Abstand der Atome im Raum und *r*_{*c*} ein vorgegebener Radius ist. Ist ein Atom *atom*_{*k*}^{*T*} von anderen Atomen

des gleichen Typs T umgeben tragen diese zu seiner LFD in Abhängigkeit vom Abstand bei. Liegen sie dabei außerhalb des Radius r_c , so ist dieser Beitrag 0.

$$LFD(atom_k^T) = \sum_{i=1}^N \max \left\{ 0, 1 - \frac{D_2(atom_k^T, atom_i^T)}{r_c} \right\} \quad \text{Formel 6}$$

Nun können die Atome anhand ihrer LFD zu Clustern zusammengefasst werden. Dazu wird bedient man sich einer *Union-Find-Strategie*.

```

INIT: each atom is a singleton
FOR each atom i of type T
  FOR each atom j of type T
    calculate Distance(i, j)
    IF Distance(i, j) < r_c THEN
      FIND maxLDF(Cluster_i)
      FIND maxLDF(Cluster_j)
      IF maxLDF(Cluster_i) ≤ maxLDF(Cluster_j) THEN
        UNION(Cluster_i, Cluster_j)

```

Am Anfang repräsentiert jedes Atom ein Cluster. Wird nun innerhalb des Radius r_c ein anderes Atom des gleichen Typs gefunden, wird die maximale LFD der Cluster gesucht, denen die Atome i und j angehören (Find). Ist die maximale LFD des Clusters, dem Atom i angehört größer oder gleich der maximalen LFD des Clusters dem Atom j angehört, werden diese Cluster vereinigt (Union).

Berechnen des Pharmkophor-Dichtefelds

Sind die Atome zu Clustern zusammengefasst, kann das Pharmkophor-Dichtefeld berechnet werden. Jeder Cluster trägt zum Dichtefeld mit einer trivariaten Gauss'schen Funktion (Formel 7) bei. Die Parameter Σ und $\bar{\mu}$ werden von der Größe, Form und der räumlichen Lage des Clusters bestimmt. $\bar{\mu}$ ist der geometrische Mittelpunkt, Σ die Covarianzmatrix des Clusters.

$$G(\vec{x}) = \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu})\right) \quad \text{Formel 7}$$

Das Pharmakophor-Dichtefeld setzt sich somit aus mehreren Gauss'schen Funktionen zusammen. Wird nun ein Molekül durch Überlagerung in das Pharmakophor-Dichtefeld gebracht, kann für jedes Atom mit den Koordinaten \vec{a} und vom Pharmakophortyp T eine Bewertung (engl. Score) $S(\vec{a}, T)$ nach Formel 8 berechnet werden.

$$S(\vec{a}, T) = \begin{cases} 0, & \text{wenn } T \text{ unbestimmt} \\ \sum G_L(\vec{a}) - \left(\sum G_D(\vec{a}) + \sum G_A(\vec{a})\right), & \text{wenn } T = L \\ \sum G_D(\vec{a}) - \sum G_L(\vec{a}), & \text{wenn } T = D \\ \sum G_A(\vec{a}) - \sum G_L(\vec{a}), & \text{wenn } T = A \end{cases} \quad \text{Formel 8}$$

Die Bewertungsfunktion wird negativ, wenn z.B. ein lipophiles Atom in der Mitte eines Akzeptor- oder Donorclusters zu liegen kommt und umgekehrt. Damit wird der ungünstige entropische Beitrag nicht abgesättigter Wasserstoffbrücken bei der Protein-Ligand Wechselwirkung implizit berücksichtigt.

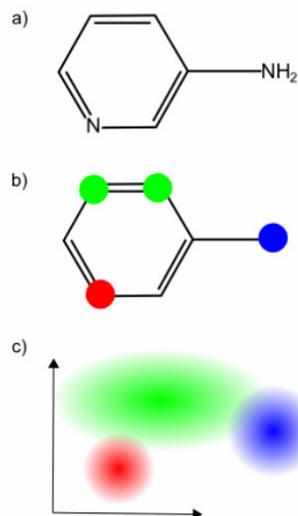


Abbildung 11: Die Berechnung der LIQUID-Pharmakophordichtefunktion. a) Das Ausgangsmolekül. b) Den Atomen werden Pharmakophortypen zugewiesen (rot: Wasserstoffbrückenakzeptor, blau: Wasserstoffbrückendonator, grün: lipophil). c) Die Cluster der räumlich benachbarten Atome gleichen Typs werden durch trivariate Gauß'sche Funktionen approximiert.

SQUIRREL – ein Programm für das virtuelle Screening

SQUIRREL (Sophisticated Quantification of InteRaction Relationship) kombiniert die *Shapelets*-Methode für die Überlagerung der Moleküle mit der LIQUID-Bewertungsfunktion. Ist ein Molekül mit bekannter biologischer Aktivität und in seiner aktiven Konformation gegeben, kann man eine Datenbank nach absteigender Bewertung sortieren und die am besten bewerteten Moleküle für die Durchführung biochemischer oder biologischer Tests (engl. *Assays*) bestellen. Abbildung 12 gibt den Ablauf einer solchen Prozedur wieder.

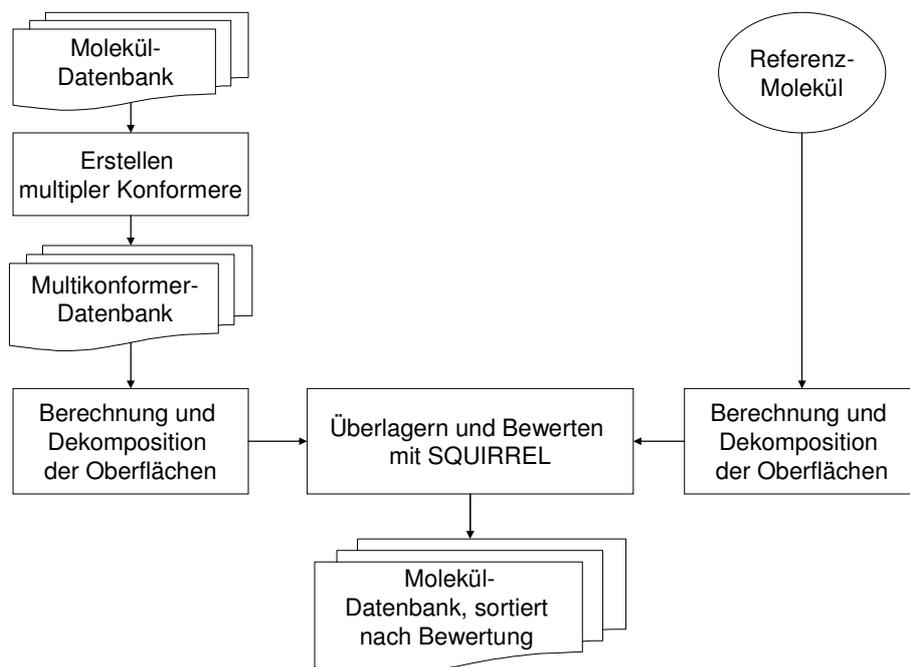


Abbildung 12: Berechnung der molekularen Überlagerung, dargestellt in einem Flussdiagramm.

SQUIRREL_{novo}

Die SQUIRREL-Methode kann auch dazu verwendet werden, molekulare Fragmente mit dem Referenz-Molekül zu überlagern, um Vorschläge für den bioisosteren Ersatz bestimmter Teile des Referenzmoleküls zu generieren. SQUIRREL_{novo} verwendet das Konzept der *pseudo*-retrosynthetischen Fragmentierung mit Hilfe der RECAP (*retrosynthetic combinatorial analysis procedure*) Retrosyntheseregeln [Lewell1998]. Um ein Molekül mit Hilfe der RECAP-Regeln retrosynthetisch zu zerlegen, werden zunächst die Bindungen, an denen gespalten wird, identifiziert. Dies geschieht mit Hilfe von MQL (Molecular Query Language) [Proschak2007], einer Programmibliothek zum Auffinden von Substrukturen auf molekularen Graphen (Tabelle 1).

Das Molekül wird an allen gefundenen Bindungen gespalten. Die entstehenden Fragmente werden samt der zum Fragment gehörenden charakteristischen Punkte der molekularen Oberfläche gespeichert. Abbildung 14 zeigt eine solche Aufspaltung. Entstehen bei der Spaltung Fragmente, die weniger als 4 charakteristische Punkte haben, d.h. nur im geringen Maße zur molekularen Oberfläche beitragen, werden diese nicht gespeichert, da bei 3 Dimensionen mindestens 4 Punkte notwendig sind, um eine eindeutige Überlagerung zu gewährleisten. Die Atome, deren Bindungen durch die RECAP-Regeln gespalten wurden, werden als *attachment points* bezeichnet und können später für die virtuelle Verknüpfung benutzt werden.

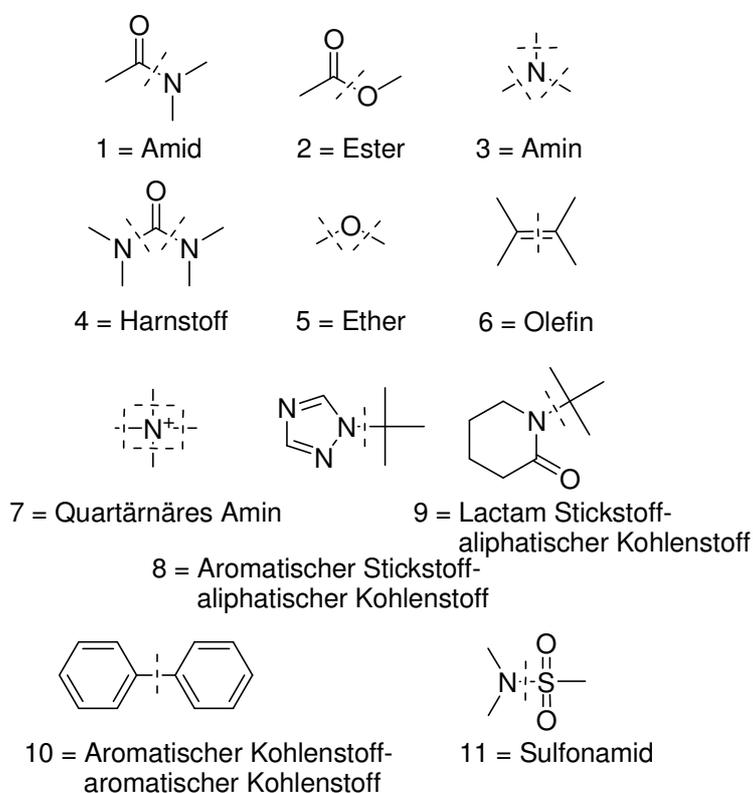


Abbildung 13: Die RECAP Retrosyntheseregeln [Lewell1998].

Tabelle 1: MQL-Ausdrücke für die verwendeten Retrosynthese-Regeln.

RECAP-Regel	MQL-Ausdruck
1 = Amid	C1[bound(=O)]-(!ring)N2
2 = Ester	C1[bound(=O)]-(!ring)O2
3 = Amin	C1-(!ring)N2(!bound(-C=O))
5 = Ether	C1-(!ring)O2[allHydrogens=0]
8 = Aromatischer Stickstoff – aliphatischer Kohlenstoff	N1[aromatic]-(!ring)C2[aliphatic]
9 = Lactam Stickstoff – aliphatischer Kohlenstoff	N1[ring](-C'[ring]=O)-(!ring)C2[aliphatic]
10 = Aromatischer Kohlenstoff – aromatischer Kohlenstoff	c1-(!ring)c2
11 = Sulfonamid	N1-(!ring)S2(=O)=O'

Auf die Implementierung der Reaktionstypen 4 = Harnstoff und 7 = quartäreres Amin wird verzichtet, da diese Bindungstypen implizit von den Reaktionstypen 1 = Amid und 3 = Amin abgedeckt werden. Der Reaktionstyp 6 = Olefin wird weggelassen, weil Fechner und Schneider gezeigt haben, daß die Verwendung dieses Reaktionstyps zu synthetisch schlecht zugänglichen Molekülen führt [Fechner2007].

Ist eine Datenbank mit Fragmenten entstanden, können diese wie bei der SQUIRREL-Methode mit dem Referenzmolekül überlagert werden. Jede Überlagerung wird mit Hilfe der LIQUID-Funktion bewertet. Die *attachment points* fließen nicht in die Bewertungsfunktion ein, da sie durch eine neue Verknüpfung einen anderen Pharmakophortyp erhalten können.

Die Datenbank der mit dem Referenzmolekül überlagerten Fragmente, sortiert nach der erhaltenen Bewertung, kann nun gefiltert werden. Die Filter sind vielfältig und können nach dem Ermessen des Experimentators eingesetzt werden, z.B. die Molekülmasse des Fragments, Anzahl der Schweratome, das Vorhandensein oder die Abwesenheit von funktionellen Gruppen, Anzahl, Lage und Art der „attachment points“ usw. Die anschließende Inspektion der vorgefilterten Datenbank erfolgt visuell mit dem Programm PyMOL (<http://pymol.sourceforge.net/>). Dabei kann der Experimentator die Fragmente zusammen mit dem Referenzmolekül betrachten und sich dann für einen

oder mehrere Möglichkeiten für den bioisosteren Ersatz eines Teils des Referenzmoleküls entscheiden.

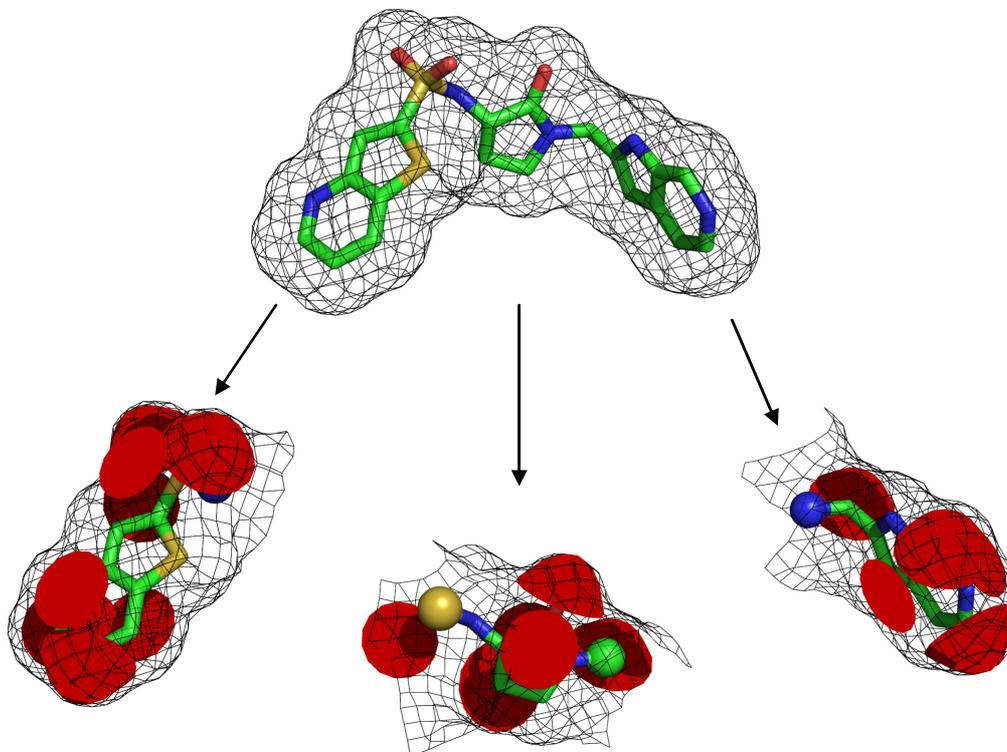


Abbildung 14: Die Zerlegung von RPR208707 [Maignan2000] in Fragmente entsprechend den RECAP-Retrosyntheseregeln. Die Kugeln symbolisieren die „attachment-points“, die durch die Spaltung der Bindungen entstanden sind. Die roten Paraboloiden sind den einzelnen Fragmenten zugewiesen und werden für die Überlagerung mit Hilfe der *Shapelets*-Methode verwendet.

Benutzte Methoden und Programme

Als Programmiersprache wurde Java (<http://java.sun.com/>) in der Version 5 benutzt. Zusätzlich wurden die frei verfügbaren Bibliotheken JGraphT (<http://jgrapht.sourceforge.net/>) in der Version 0.7.3 und Chemistry Development Kit (CDK) [Steinbeck2003, Steinbeck2006] (http://almost.cubic.uni-koeln.de/cdk/cdk_top) in der Version 20060714 verwendet. Die Visualisierung der Moleküle erfolgte mit PyMOL (<http://pymol.sourceforge.net/>) in der Version 0.99. Die statistische Auswertung der Ergebnisse wurde mit den Programmen Microsoft Excel 2002 und SigmaPlot 2001 durchgeführt.

Das Programm MOE v2006.08 (www.chemcomp.com/, Chemical Computing Group, Montreal, Canada) wurde unter anderem zur Vorbereitung der Moleküldatenbanken verwendet.

Selbstorganisierende Karten

Die selbstorganisierende Karten (engl. self-organizing maps, SOMs) wurden mit dem Programm molmap® [Schneider1998] erstellt.

SOMs wurden zuerst von Kohonen [Kohonen1982] entwickelt und gehören zu den nicht-überwachten Lernverfahren. Eine SOM besteht aus Neuronen $n \in N$, deren Gewichte die Position im Deskriptorraum wiedergeben und die mit anderen Neuronen entsprechend der Netztopologie verbunden sind. Das formale SOM-Training funktioniert entsprechend folgendem Pseudocode.

```
Initialize neurons  $N$  at random
while ( $t < t_{\max}$ )
    choose random sample  $x$ 
    determine winner neuron  $n_w$  with minimal  $d(n_w, x)$ 
    for each  $n \in N$ 
```

Die Neuronen werden also zufällig initialisiert. Ist die maximale Anzahl der Trainingsepochen t_{\max} nicht erreicht, wird zufällig ein Vektor x aus der Menge der Trainingsvektoren ausgewählt. Das „Gewinnerneuron“ n_w mit der minimalen Distanz $d(n_w, x)$ wird bestimmt. Die Wahl der Distanzmetrik ist von der Anwendung abhängig, in diesem Fall wurde die Manhattan-Distanz benutzt (Formel 9).

$$d(a, b) = \sum_n |a_n - b_n| \quad \text{Formel 9}$$

Nun werden die Gewichte aller Neuronen entsprechend Formel 10 angepasst. $\lambda(t)$ ist die zeitabhängige Lernrate und $v(n_w, n, t)$ die Nachbarschaftsfunktion, die bestimmt, wie stark die Gewichte des Neurons n in Abhängigkeit von der topologischen Distanz zum Neuron n_w adaptiert wird. Die topologische Distanz d_{topo} ist der kürzeste Abstand auf der Netztopologie.

$$\Delta n = \lambda(t) \cdot v(n_w, n, t) \cdot (x - n)$$

$$\lambda(t) = \lambda_{initial} \left(\frac{\lambda_{final}}{\lambda_{initial}} \right)^{t/t_{max}}$$

$$v(n_w, n, t) = \exp \left(\frac{-d_{topo}(n_w, n)^2}{2\sigma(t)^2} \right)$$

Formel 10

$$\sigma(t) = \sigma_{initial} \left(\frac{\sigma_{final}}{\sigma_{initial}} \right)^{t/t_{max}}$$

Eine SOM kann daher als eine nicht-lineare Projektion des Deskriptorraums, meist einhergehend mit einer Dimensionsreduktion, betrachtet werden. Es sind unterschiedliche Netztopologien denkbar: planar, toroidal, sphärisch usw. molmap® implementiert die toroidale Netztopologie.

Ist eine SOM trainiert, können die Neuronen nach Zugehörigkeit zu einer bestimmten Klasse eingefärbt werden. Dabei wird jeder Trainingsvektor genau einem Neuronen zugeordnet, nämlich dem mit dem geringsten Abstand. Es können aber auch weitere, nicht für das Training verwendete bzw. klassifizierte Punkte auf die Neuronen projiziert werden, vorausgesetzt sie befinden sich im selben Deskriptorraum. Fallen z.B. auf ein Neuron besonders viele aktive Moleküle, ist anzunehmen, dass unter den nicht untersuchten Molekülen, die auf dieses Neuron projiziert werden, auch vermehrt Aktive sind. [Hristozov2007]

CATS Deskriptor

Der CATS Deskriptor wurde von Schneider *et. al.* [Schneider1999] entwickelt und in vielen Applikationen erfolgreich eingesetzt [Derksen2006, Noeske2006, Renner2005, Roger-Evans2004]. Um den CATS Deskriptor eines Moleküls zu berechnen, wird jedem Atom ein Pharmakophortyp zugewiesen (vgl. SQUIRREL). Bei CATS werden die Pharmakophortypen Wasserstoffbrückendonator (D), Wasserstoffbrückenakzeptor (A), negativ geladen (N), positiv geladen (P) und lipophil (L) verwendet, Mehrfachzuweisungen sind möglich. Für alle möglichen Paare von Pharmakophorpunkten (DD, DA, DN, DP, DL, AA, AN, AP, AL, NN, NP, NL, PP, PL und LL) werden die kürzesten Wege im molekularen Graphen gezählt und die Häufigkeiten in einem Histogramm gespeichert. Die maximale Länge l_{max} der kürzesten

Wege bestimmt die Dimensionalität des Deskriptors, in dieser Arbeit wurde l_{max} auf 20 Bindungen gesetzt. Damit ergibt sich für die Dimensionalität des Deskriptors $15 \times 20 = 300$. Der Deskriptor wurde nicht skaliert, um implizit die Information über die Größe des Moleküls zu erhalten. Um den CATS Deskriptor zu berechnen, wurde die SpeedCATS Software (arbeitskreisinterne Version, implementiert von Dr. Ulrich Fechner) verwendet.

ROCS

ROCS ist eine *virtual screening* Methode, die Ligandenähnlichkeit über die Volumenüberlagerung definiert [Grant1996]. ROCS repräsentiert die Moleküle als eine Summe Gauß'scher Funktionen (Formel 1). Diese werden überlagert und das Differenzvolumen schrittweise minimiert. Die Erweiterung von ROCS, die Pharmakophorinformation der Liganden berücksichtigt, heißt ROCS-color. Die Software ROCS in der Version 2.3.1 (<http://www.eyesopen.com/>, Openeye Scientific Software, Santa Fe, USA) wurde im Rahmen einer Kooperation mit Boehringer-Ingelheim Pharma, Biberach a.d. Riss, zur Verfügung gestellt.

MOE 4-Punkt Pharmacophor Fingerprints

Die 4-Punkt Pharmakophor Fingerprints sind eine weit verbreitete Methode, die von Mason *et al.* [Mason1999] eingeführt wurde, um 3D-Konformationen von Molekülen zu beschreiben. Wie bei CATS wird jedem Atom ein Pharmakophortyp zugewiesen. Jedes 4-Tupel von Atomen bildet einen Tetraeder (Abbildung 15). Abhängig von Seitenlängen und Pharmakophortypen der Atome entspricht jeder Tetraeder einem Bit im Fingerprint. In dieser Arbeit wurden die Fingerprints aus MOE v2006.08 (www.chemcomp.com/, Chemical Computing Group, Montreal, Canada) benutzt.

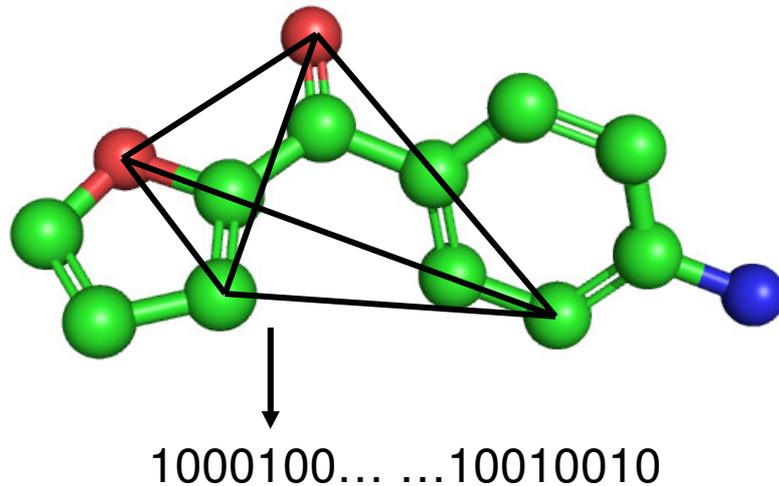


Abbildung 15: Beispiel für die Erzeugung des 4-Punkt-Pharmakophor-Fingerprints. Jedes Tetraeder mit Eckpunkten an vier Atomen mit potentiellen Pharmakophor-Eigenschaften trägt ein Bit zum Fingerprint bei.

Docking

Beim Docking wird versucht, die bioaktive Konformation des Liganden im Rezeptor zu berechnen. Da dies ein NP-vollständiges Problem ist [Kuhl1984], werden unterschiedliche Heuristiken eingesetzt, um die Berechnungen zu beschleunigen. Das Programm GOLD [Jones1995, Jones1997, Verdonk2003] in der Version 3.2 (http://www.ccdc.cam.ac.uk/products/life_sciences/gold, CCDC, Cambridge, UK) verwendet einen genetischen Algorithmus, um die Berechnungen zu beschleunigen. Jede Konformation, repräsentiert als die Menge der freien zu optimierenden Parameter (Translation, Rotation, Rotationswinkel der frei drehbaren Bindungen), wird auf virtuellen Chromosomen gespeichert. Die Menge der parallel existierenden Chromosomen wird als Population bezeichnet. Durch Mutation- und Crossover-Operationen werden die freien Parameter Generation für Generation optimiert.

Die Güte einer Konformation wird durch eine Bewertungsfunktion bestimmt, in diesem Fall Chemscore [Eldridge1997, Baxter1998]. Chemscore ist eine empirische Funktion, die aus 82 Ligand-Protein-Komplexen abgeleitet wurde, für die gemessene Bindeaffinitäten vorhanden waren. Chemscore approximiert die Bindungsenergie, indem es Wasserstoffbrücken, lipophile und Metallionen-Wechselwirkungen und Rotationsenthalpie berücksichtigt. Zusätzlich werden Strafterme für sterische Überlappungen eingeführt.

Bewertung der Güte der Molekülüberlagerung

Die Güte der Molekülüberlagerung, die von unterschiedlichen Methoden berechnet wurde, wird als mittlere quadratische Abweichung (engl. *root mean square deviation*, RMSD) angegeben. Sie wird nach der Formel 11 berechnet, wobei M und M' unterschiedliche Konformationen des gleichen Moleküls und $\bar{x}_{i,M}$ und $\bar{x}_{i,M'}$ die Koordinaten der desgleichen Atoms innerhalb der jeweiligen Konformation sind. N bezeichnet die Gesamtzahl der Atome innerhalb eines Moleküls und ist bei M und M' gleich.

$$RMSD(M, M') = \sum_{i=1}^N (\bar{x}_{i,M} - \bar{x}_{i,M'})^2 \quad \text{Formel 11}$$

Konformerengenerierung

Die Generierung multipler Konformere erfolgte mit MOE (www.chemcomp.com/, Chemical Computing Group, Montreal, Canada) in der Version v2006.08 und OMEGA (<http://www.eyesopen.com/products/applications/omega.html>) in der Version 2.2.1. Diese Software wurde im Rahmen einer Kooperation von Boehringer-Ingelheim Pharma, Biberach a.d. Riss, zur Verfügung gestellt.

MOE erzeugt die Konformere mit einer kraftfeldbasierten Methode [Kolossvary1996]. Dabei wird das *Merck Molecular Force Field* (MMFF) [Halgren1996] eingesetzt, um eine zufällig initialisierte Konformation zu minimieren. Aus dieser Konformation werden mit dem abgewandelten Random *Incremental Pulse Search* (RIPS) [Ferguson1989] Algorithmus der konformationelle Raum abgesucht. Bei dem originalen RIPS Algorithmus wurden die kartesischen Koordinaten einer Konformation zufällig verändert („perturbiert“), um Startpunkte für erneute Minimierungen zu erhalten, MOE hingegen wendet diesen Ansatz auf die Torsionswinkel an.

OMEGA ist eine regelbasierte Methode [Boström2001]. Im Unterschied zu der stochastischen Suche sind die Ergebnisse von OMEGA bei jeder Berechnung reproduzierbar. Die Moleküle werden dabei in Fragmente zerlegt, die durch drehbare Bindungen verbunden sind. Die Bindungslängen bzw. -winkel werden nicht optimiert. Durch die Überlagerung der Schweratome wird der RMSD berechnet, um die Generierung zu ähnlicher Konformere auszuschließen. Der RMSD Schwellwert wurde wie auch bei MOE auf 1.5 Å gesetzt. MMFF wird eingesetzt, um die Energie der

Konformationen abzuschätzen. Ausschließlich Konformere unter dem Energieschwellwert von 6 kcal/mol werden weiterverwendet.

Datenbanken

Für die retrospektive Analyse, das Training der SOMs und die Generierung der Fragmente wurde die COBRA6.1 [Schneider2003] Datenbank verwendet. Diese Version der Datenbank enthält 8331 Wirkstoffe und Wirkstoffkandidaten, klassifiziert nach Zielprotein, und wurde für die retrospektiven Analysen genutzt. Wertvoll ist die Benutzung der COBRA6.1 Datenbank für die Generierung der Fragmente, da die entstehenden funktionellen Gruppen Wirkstoffen entstammen und meist keine toxisch aktiven funktionellen Gruppen enthalten.

Praktische Methoden

PPAR-Assay

Die Fähigkeit von Substanzen, PPARs zu aktivieren, wurde in einem zellbasierten Reporter-Gen Assay in Cos7-Zellen [Gluzman1981] getestet. Die Cos7-Zelllinie wurde aus dem Nierengewebe der Grünen Meerkatze *Cercopithecus aethiops* gewonnen und stabil mit einer Mutante des SV-40 Virus transformiert. Der Assay wurde von Dr. Oliver Rau an der Universität Frankfurt etabliert und publiziert [Rau2006, Ulrich2006]. Der Assay wurde gemäß des veröffentlichten Protokolls durchgeführt:

Die DNA-Bindedomäne (DBD) des jeweiligen PPAR-Subtyps wurde durch die Gal4-DBD (*Saccharomyces cerevisiae*) ersetzt und als eukaryontisches Expressionsplasmid transfiziert, zusammen mit einem entsprechenden Reporter-Plasmid mit dem zugehörigen Response-Element (UAS). Zur Normalisierung wurde außerdem ein konstitutiv exprimiertes Kontrollgen transfiziert. Die in diesem Assay verwendeten Plasmide sind:

- pFA-CMV-PPAR-LBD codiert die Information für die Expression des Gal4-PPAR Fusionsproteins, bestehend aus der Liganden-Bindedomäne (LBD) eines der PPAR Subtypen und der DBD des Gal4-Proteins. Der Cytomegalovirus (CMV) Promotor sorgt für ein hohes konstitutives Expressionsniveau des Fusionsproteins (Abbildung 16a).
- pFR-Luc (Abbildung 16b) enthält eine dreifache Wiederholung des Gal4-Response-Elements, gefolgt von einer TATA-Box und einer aus *Photinus pyralis* gewonnenen Firefly-Luciferase-codierenden DNA-Sequenz [deWet1985, deWet1987].
- pRL-SV40 Vektor [Lorenz1991, Matthews1977] wurde für die interne Kontrolle benutzt. Er enthält die cDNA der Luciferase aus *Renilla reniformis*. Der SV40-Promotor bewirkt eine starke konstitutive Expression des *Renilla* Luciferase Proteins (Abbildung 16c).

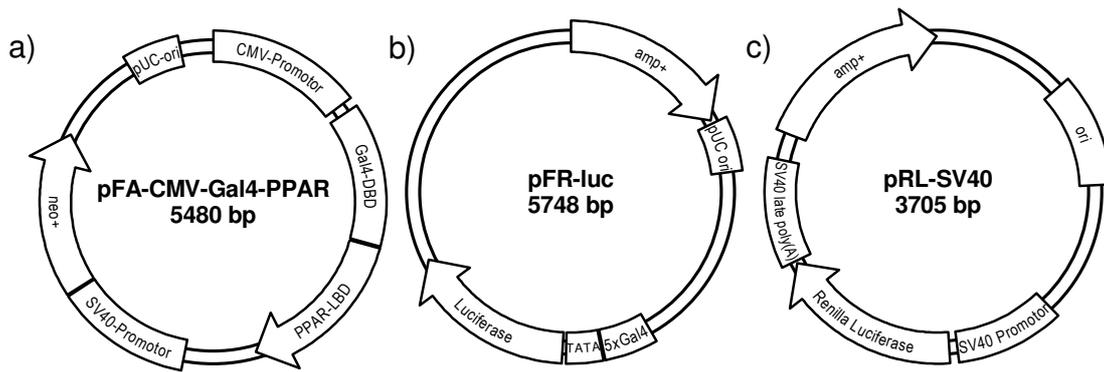


Abbildung 16: Verwendete Plasmide.

Der Assay funktioniert nach folgendem Prinzip (Abbildung 17): die transfizierten Zellen exprimieren das PPAR-LBD-Gal4-DBD Fusionsprotein. Werden nun die Zellen mit einer Substanz inkubiert, die PPAR aktiviert, kann das Reporterogen Firefly-Luciferase exprimiert werden. Werden die Zellen dann lysiert, kann durch die Zugabe des Substrats der Firefly-Luciferase – dem Luciferin **12**, der Expressionslevel der Luciferase durch Messung der Lumineszenz quantifiziert werden. Der Normierung dient die konstitutive Expression der *Renilla*-Luciferase, die durch die Zugabe von Coelenterazine **14** detektiert werden kann (Abbildung 18).

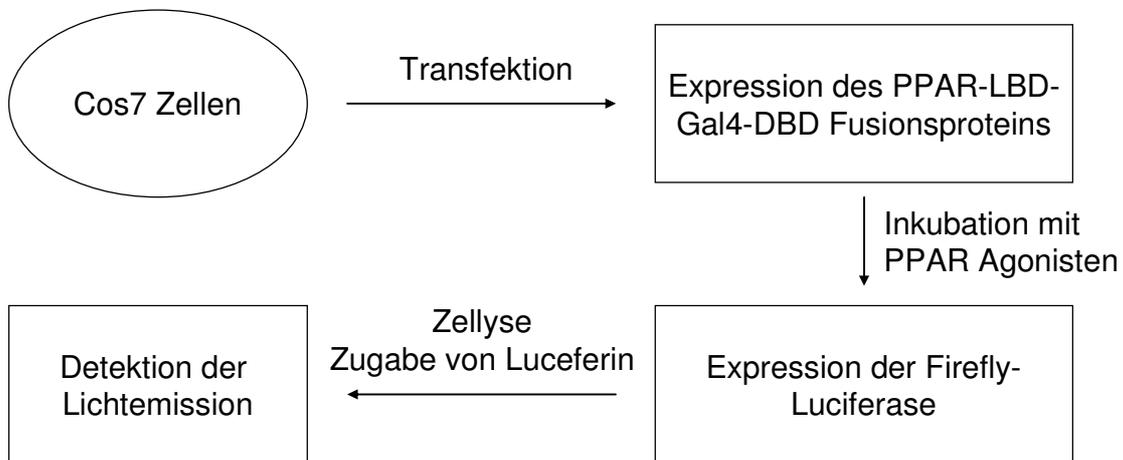


Abbildung 17: Schematischer Ablauf des zellbasierten PPAR Testsystems.

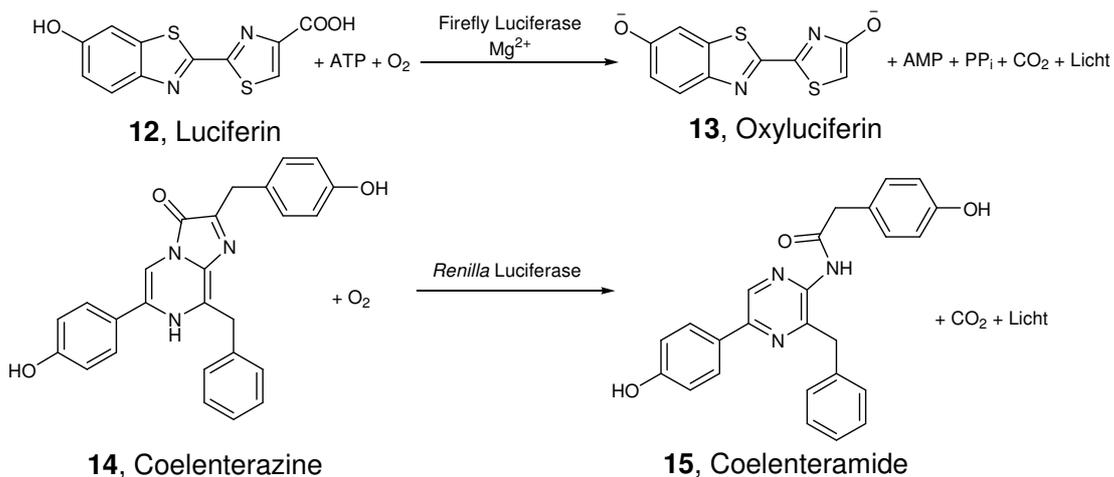


Abbildung 18: Reaktion der Firefly und *Renilla* Luciferase.

Reagenzien und Material

Fetales Kälberserum (*fetal calf serum*, FCS) und Dimethylsulfoxid (DMSO) stammen von Sigma; Dulbecco's modified Eagle's Medium (DMEM) und OptimumTM von Gibco (Carlsbad, CA, USA); Natriumpyruvat-Lösung, Glutamin-, Penicillin- und Streptomycin-Stammlösungen von PAA Laboratories GmbH (Pasching, Österreich); LipofectamineTM 2000 von Invitrogen (Carlsbad, CA, USA); DualGloTM Luciferase Assay System von Promega (Madison, WI, USA). Cos7 Zellen und Plasmide für die Transfektion wurden freundlicherweise von Dr. Oliver Rau zur Verfügung gestellt.

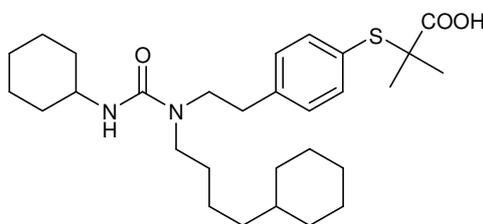
Zellkultur und Transfektion

Cos7 Zellen wurden in DMEM mit 10% FCS, 100 U/ml Penicillin, 100 µg/ml Streptomycin, 2 mM Glutamin und 1mM Natriumpyruvat bei 37°C und 5% CO₂ kultiviert, zur Anpassung des pH-Wertes des Mediums wurden dem Medium außerdem 20 mM HCl zugesetzt. Das Ablösen der Zellen vom Boden der Zellkulturflasche erfolgte mit 5 ml Trypsin/EDTA Lösung. Nach dem Ernten wurde die Zelldichte durch Auszählen unter dem Mikroskop bestimmt. Durch Zentrifugieren bei ca. 340g für 5 min, Absaugen des Überstandes und Resuspendieren der Zellen in berechneter Menge DMEM wurde die Zelldichte auf 300.000 Zellen/ml eingestellt. Die Zellsuspension wurde zu je 100 µl auf einer 96-well Platte ausgesät. Die Transfektion erfolgte mit 280 ng/well des Reporterplasmids pFR-Luc, 8 ng/well des Kontrollplasmids pRL-SV40, 14 ng/well des jeweiligen PPAR Expressionsplasmids pFA-CMV-PPAR-LBD für den zu

testenden PPAR-Subtyp (α oder γ), 0,8 μ l/well LipofectamineTM 2000 und 50 μ l/well Optimem Medium. Die Dauer der Transfektion betrug 4 Stunden.

Luciferase Assay

Die zu testenden Substanzen wurden in DMSO gelöst, die Konzentration der Stammlösung betrug 0,1 M. Von dieser Stammlösung ausgehend wurden Inkubationslösungen durch Verdünnen der Stammlösung mit FCS-freiem DMEM hergestellt. Die transfizierten Zellen wurden mit 50 μ l der Inkubationslösung behandelt und für 12-14 Stunden inkubiert. Für die Bestimmung der Basalaktivität wurde als Negativkontrolle DMEM mit 0,1% DMSO verwendet. Als Positivkontrolle (100% der Aktivität) diente bei PPAR α die Referenzsubstanz GW7647 **16** [Brown2001] (EC_{50} = 6 nM), bei PPAR γ die Referenzsubstanz Pioglitazon **7** (EC_{50} = 280 nM), jeweils in der Konzentration von 1 μ M. Pro Konzentration wurden 3 wells inkubiert, aus denen später der Mittelwert errechnet wurde. Nach Ende der Inkubationszeit wurden die Zellen mit 50 μ l DualGloTM (bestehend aus Lysepuffer und Substrat der Firefly-Luciferase) behandelt. Nach 10 min wurde die Lichtemission im GENiosPro Luminometer (Tecan, Männedorf, Schweiz) und 50 μ l StopGloTM (Substrat der *Renilla*-Luciferase) zugesetzt. Nach weiteren 10 min wurde die Lichtemission erneut gemessen.



16, GW7647

Berechnungen und Statistik

Die für die jeweilige Konzentration erhaltenen light-units (LU) wurden durch die bei der Messung des Kontroll-Gens, also der *Renilla*-Luciferase erhaltenen Werte geteilt, so dass sich die normierte Größe der relativen light-units ergab. Durch Dividieren der so erhaltenen RLUs durch den entsprechenden Wert der Positiv-Kontrolle ergibt sich die relative Aktivierung. Jedes Experiment wurde mindestens dreimal wiederholt und die Mittelwerte \pm Standardabweichung berechnet.

Für die Berechnung von EC_{50} Werten und der zugehörigen Standardabweichungen wurde die Funktion *Four Parameter Logistic Curve* der Software Sigma-Plot 2001 verwendet, auf Basis der jeweiligen Mittelwerte der Dreifachbestimmung. Zur Berechnung signifikanter Aktivierung im Vergleich zur Negativ-Kontrolle wurde ein einseitiger *t*-Test mit der Software Microsoft-Excel durchgeführt.

Synthese der Substanzen EP1 und EP5

Reagenzien und Geräte

Die Edukte (Prolinmethylester, 4-Formylbenzonnitril, 2-Naphtylchlorid, Ethyl-1*H*-pyrrol-2-carboxylat, und 4-(Bromomethyl)benzonnitril) stammen von Sigma-Aldrich und Alfa Aesar. Die Schmelzpunkte wurden mit Hilfe des Büchi 510 Schmelzpunktmessgeräts (Büchi, Schweiz) bestimmt und sind unkorrigiert. ^1H -NMR Spektren wurden am Bruker AMX 300 (300 MHz) Spektrometer (Bruker, Deutschland) aufgezeichnet. ^1H NMR Daten sind in folgender Reihenfolge aufgelistet: chemische Verschiebung (δ) in ppm bezogen auf den inneren Standard Tetramethylsilan (TMS); Multiplizität (s, Singulett; d, Dublett; dd, Dublett von Duppletts; t, Triplett; m, Multiplett); gerundete Kopplungskonstanten (*J*) in Hertz (Hz); Anzahl und Zuweisung der Protonen (pyrr, Pyrrol; pyrrolid, Pyrrolidin; naph, Naphthyl; ph, Phenyl).

^{13}C -NMR Spektren wurden auf einem Bruker AC 200 (50 MHz) Spektrometer (Bruker, Deutschland) aufgezeichnet.

Elektron-Spray-Ionisation-Massenspektroskopie (ESI MS) wurde auf der Fisons Instruments VG Platform II (Manchester, Großbritannien) bei positiver Ionisierung ausgeführt. Aufgelistet sind die Massenzahl($[\text{M}+\text{H}]^+$) und relative Intensität (%).

Die Elementaranalyse (C, H, N) wurde am CHN-Rapid (Heraeus, Deutschland) gemessen, die Werte liegen innerhalb der erlaubten Grenzen von $\pm 0.4\%$ der theoretischen Werte.

Präparative Säulenchromatografie wurde auf Säulen, beladen mit Kieselgel der Korngröße 63-200 μM (Merck, Deutschland) durchgeführt.

Die Hydrierungen wurden im Hochdruck-Haberautoklaven Model IV, Typ 50S, 500 mL (Roth, Deutschland) durchgeführt. Der benutzte Mikrowellenofen ist ein Biotage Initiator 2.0 (Biotage, Schweden).

Anmerkung: Aus Gründen der Übersichtlichkeit wurden bei der Interpretation der Analytikdaten die Dezimalkommata durch Dezimalpunkte ersetzt.

Synthese von EP1

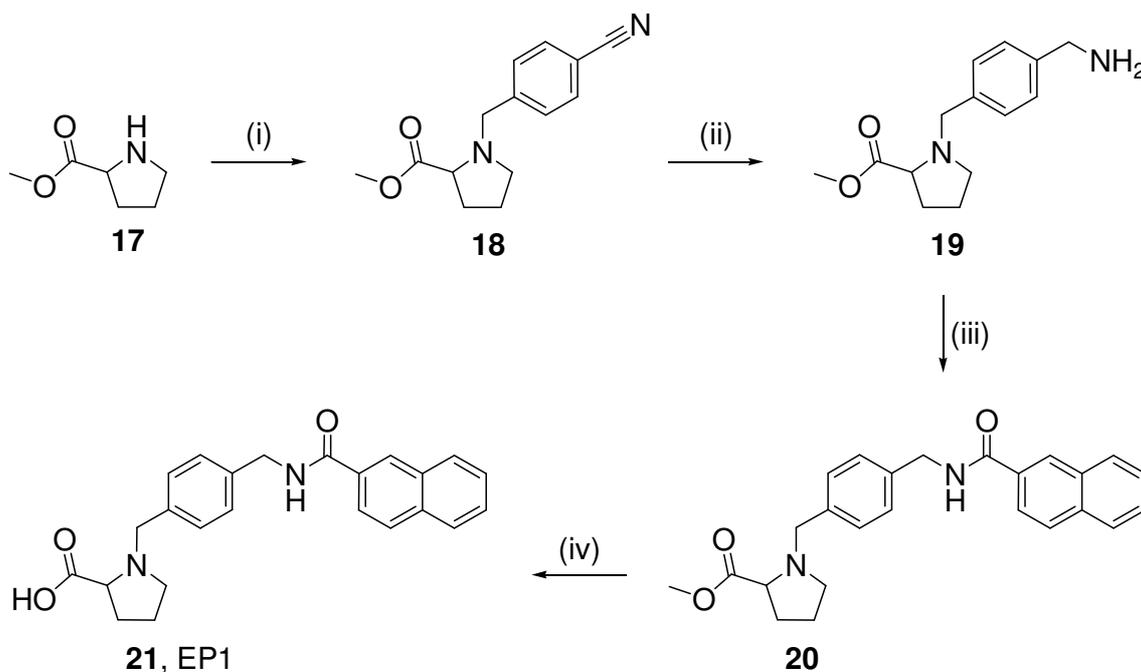


Abbildung 19 Übersichtsschema der Synthese der Substanz EP1. (i) 4-Formylbenzonitril, TEA, NaBH(OAc)₃, DCE; (ii) Raney-Ni, NH₃, MeOH; (iii) 2-Naphthylchlorid, DCE, μ w; (iv) KOH, H₂O/MeOH/THF, μ w.

Synthese von Methyl-1-(4-cyanobenzyl)pyrrolidin-2-carboxylat (18)

Für die reduktive Aminierung wurden 2,0 g (12 mmol) Prolinmethylester-hydrochlorid **17** mit 1,6 g (16 mmol) Triethylamin (TEA) und 1,57 g (12 mmol) 4-Formylbenzonitril in 50 ml 1,2-Dichlorethan (DCE) bei Raumtemperatur und unter Argon-Schutzgasatmosphäre im Rundkolben gerührt. Nach 2,5 h wurde 3,56 g (16,8 mmol) Natriumtriacetoxyborhydrid hinzugegeben und weitere 30 min gerührt. Die Reaktion wurde durch Zugabe von 50 ml wässriger gesättigter Natriumhydrogencarbonat-Lösung abgebrochen und das Produkt durch dreimaliges Waschen mit Ethylacetat extrahiert. Die organische Phase wurde mit gesättigter Kochsalzlösung gewaschen, über Natriumsulfat getrocknet und am Rotationsverdampfer eingeeengt. Die Aufreinigung erfolgte durch Säulenchromatographie. Als Fließmittel diente Ethylacetat/Hexan im Verhältnis 1:1.

Analytik: Ausbeute: 1,3 g (44%); $^1\text{H-NMR}$ ($\text{MeOH-}d_4$): δ 7.68 (d, $J = 8.2$ Hz, 2H, ph-3,5H), 7.54 (d, $J = 8.2$ Hz, 2H, ph-2,6H), 3.98 (d, $J = 13.4$ Hz, 1H, pyrrolid- CH_2 -ph), 3.64 (s, 3H, CH_3), 3.63 (d, $J = 13.4$ Hz, 1H, pyrrolid- CH_2 -ph), 3.34-3.29 (m, 1H, pyrrolid-2H), 3.00-2.94 (m, 1H, pyrrolid-5H), 2.47-2.39 (m, 1H, pyrrolid-5H), 2.20-2.14 (m, 1H, pyrrolid-3H), 1.94-1.82 (m, 3H, pyrrolid-3,4,4H). ESI-MS: 244.9 (100).

Synthese von Methyl-1-(4-(aminomethyl)benzyl)pyrrolidin-2-carboxylat (19)

1,25 g (5,1 mmol) Nitril **18** wurden mit 5 g frisch aktiviertem Raney-Nickel in 100 ml ammoniakalischem Methanol im Autoklaven bei 10 bar Wasserstoff über einen Zeitraum von 24 Stunden hydriert. Das gelöste Produkt wurde durch Filtration über Celite von Katalysatorresten befreit, das Lösungsmittel am Rotationsverdampfer bei Unterdruck entfernt.

Analytik: Ausbeute 1,12g, (88%); $^1\text{H-NMR}$ ($\text{MeOH-}d_4$): δ 7.19 (s, 4H, ph-2,3,5,6H), 3.75 (d, $J = 16.0$ Hz, 1H, pyrrolid- CH_2 -ph), 3.68 (s, 2H, ph- CH_2 - NH_2), 3.53 (s, 3H, CH_3), 3.45 (d, $J = 16.0$ Hz, 1H, pyrrolid- CH_2 -ph), 3.25 (s, 2H, NH_2), 3.15-3.11 (m, 1H, pyrrolid-2H), 2.90-2.85 (m, 1H, pyrrolid-5H), 2.34-2.25 (m, 1H, pyrrolid-5H), 2.09-2.00 (m, 1H, pyrrolid-3H), 1.84-1.68 (m, 3H, pyrrolid-3,4,4H). ESI-MS: 248.9 (100).

Synthese von Methyl-1-(4-((2-naphthamido)methyl)benzyl)pyrrolidin-2-carboxylat (20)

1,1 g (4,4 mmol) Amin **19** und 0,84 g (4,4 mmol) 2-Naphtylchlorid wurden in 10 ml 1,2 Dichlorethan gelöst. Das Reaktionsgemisch wurde 15 min bei eingestellter Temperatur von 90° C im Mikrowellenofen gerührt. Nach dem Einengen am Rotationsverdampfer wurde das Produkt durch Säulenchromatographie aufgereinigt. Als Laufmittel diente Ethylacetat/Hexan im Verhältnis 1:1.

Analytik: Ausbeute 1,7 g (96%); $^1\text{H-NMR}$ ($\text{MeOH-}d_4$): δ 8.40 (s, 1H, naph-1H), 7.98-7.89 (m, 4H, naph-3,4,5,8H), 7.62-7.53 (m, 2H, naph-6,7H), 7.36 (d, $J = 8.2$ Hz, 2H, ph-3,5H), 7.31 (d, $J = 8.2$, 2H, ph-2,6H), 4.62 (s, 2H, ph- CH_2 -NH), 3.84 (d, $J = 12.6$ Hz, 1H, pyrrolid- CH_2 -ph), 3.60 (s, 3H, CH_3), 3.56 (d, $J = 12.6$ Hz, 1H, pyrrolid- CH_2 -ph), 3.25-3.20 (m, 1H, pyrrolid-2H), 3.01-2.96 (m, 1H, pyrrolid-5H), 2.44-2.36 (m, 1H, pyrrolid-5H), 2.16-2.10 (m, 1H, pyrrolid-3H), 1.89-1.76 (m, 3H, pyrrolid-3,4,4H). ESI-MS: 403.2 (100).

Synthese von 1-(4-((2-Naphthamido)methyl)benzyl)pyrrolidin-2-carbonsäure (21)

1,7 g (4,2 mmol) Ester **20** und 0,7 g (12,6 mmol) Kaliumhydroxid wurden in 10 ml des Lösungsmittelgemischs Tetrahydrofuran/ Methanol/ H₂O im Verhältnis 1:1:2 gelöst und im Mikrowellenofen bei 70° C und 35 W für 15 min gerührt. Organische Lösungsmittel wurden am Rotationsverdampfer entfernt. Durch tropfenweise Zugabe von 10-N Salzsäure bis zum Einstellen des pH 7 fiel das Produkt aus. Die Aufreinigung erfolgte durch Umkristallisation in Ethanol.

Analytik: Ausbeute 1,4 g (86%); ¹H-NMR (DMSO-*d*₆): δ 9.23 (t, 1H, NH), 8.51 (s, 1H, naph-1H), 8.03-7.96 (m, 4H, naph-3,4,5,8H), 7.63-7.55 (m, 2H, naph-6,7H), 7.35 (s, 4H, ph-2,3,5,6H), 4.53 (d, 2H, ph-CH₂-NH), 4.07 (d, *J* = 12.9 Hz, 1H, pyrrolid-CH₂-ph), 3.78 (d, *J* = 12.9 Hz, 1H, pyrrolid-CH₂-ph), 3.38-3.33 (m, 1H, pyrrolid-2H), 3.11-3.06 (m, 1H, pyrrolid-5H), 2.64-2.61 (m, 1H, pyrrolid-5H), 2.13-2.06 (m, 1H, pyrrolid-3H), 1.89-1.67 (m, 3H, pyrrolid-3,4,4H). ¹³C-NMR (DMSO-*d*₆): 171.9 (COOH), 166.2 (CO-NH), 139.3 (naph-2C), 134.1 (ph-1,6C), 132.1 (naph-10C), 131.6 (naph-9C), 129.4 (naph-CH), 128.8 (ph-2,6C), 127.9 (naph-CH), 127.6 (naph-CH), 127.5 (ph-3,5C), 127.3 (2C, naph-CH), 126.7 (naph-CH), 124.2 (naph-CH), 65.8 (pyrrolid-2C), 56.9 (pyrrolid-CH₂-ph), 52.8 (pyrrolid-5C), 42.5 (ph-CH₂-NH), 28.5 (pyrrolid-3C), 22.6 (pyrrolid-4C). ESI-MS: 389.0 (100). Schmelzpunkt: 190 °C. Elementaranalyse (C₂₄H₂₄N₂O₃ · H₂O): C, H, N.

Synthese von EP5

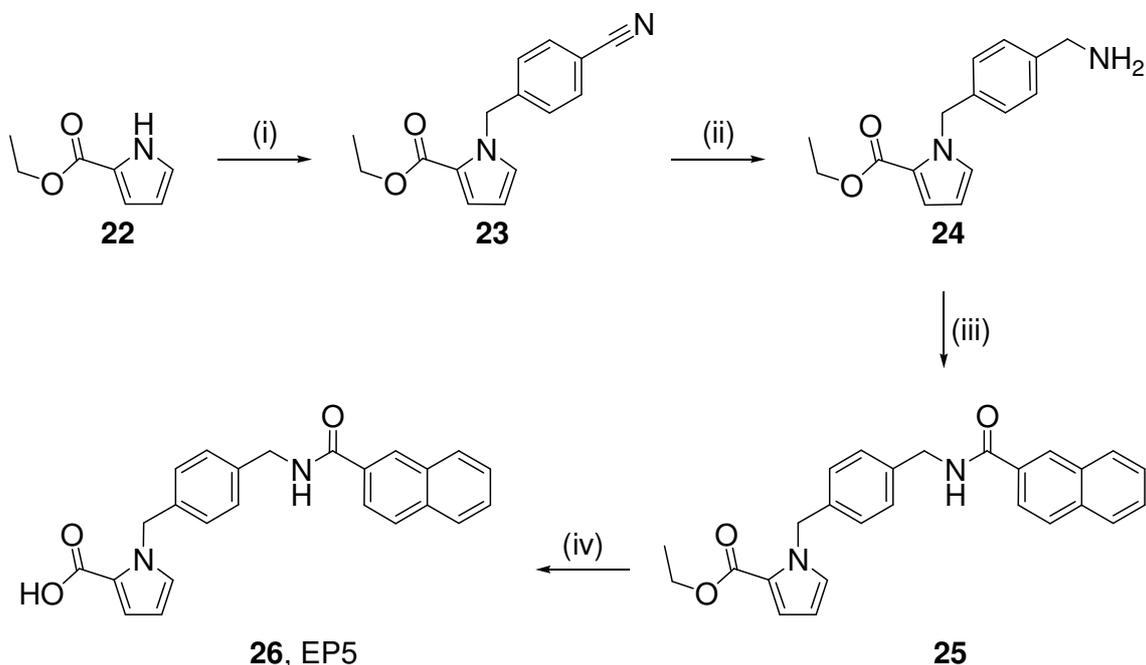


Abbildung 20: Übersichtsschema der Synthese der Substanz EP5. (i) 4-(Bromomethyl)benzotrile, TBAI, NaOH, DCM; (ii) Raney-Ni, NH₃, MeOH; (iii) 2-Naphthylchlorid, DCE, μ w; (iv) KOH, H₂O/MeOH/THF, μ w.

Synthese von Ethyl 1-(4-cyanobenzyl)-1H-pyrrol-2-carboxylat (23)

Die Kopplung des Stickstoffes in Ethyl-1H-pyrrole-2-carboxylat **22** an das 4-(Bromomethyl)benzotrile wurde mit Hilfe einer Phasentransferreaktion [Banwell2006] durchgeführt. 1,0 g (7,2 mmol) Ethyl-1H-pyrrole-2-carboxylat **22** wurde zusammen mit 0,5 ml einer wässrigen 60%igen TBAI-Lösung, 50 ml einer 50%igen Natriumhydroxid-Lösung und 50 ml Dichlormethan auf 0°C gekühlt. 1,55 g (7,9 mmol) 4-(Bromomethyl)benzotrile wurden der Reaktionslösung hinzugefügt. Das Reaktionsgemisch wurde 24 Stunden bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe von Salzsäure bis pH 1 beendet. Die wässrige Phase wurde dreimal mit 50 ml Dichlormethan gewaschen und die organische Phase über Natriumsulfat getrocknet. Das Lösungsmittel wurde am Rotationsverdampfer entfernt. Die Aufreinigung erfolgte durch Säulenchromatographie, als Laufmittel diente Etylacetat/Hexan im Verhältnis 1:1.

Analytik: Ausbeute 1,8 g (98%); ¹H-NMR (DMSO-*d*₆): δ 7.77 (d, *J* = 8.3, 2H, ph-3,5H), 7.32 (t, 1H, pyr-5H), 7.15 (d, *J* = 8.3, 2H, ph-2,6H), 6.93 (dd, 1H, pyr-3H),

6.21 (dd, 1H, pyrr-4H), 5.61 (s, 2H, pyrr-CH₂-ph), 4.10 (q, 2H, ethyl-CH₂), 1.16 (t, 3H, ethyl-CH₃). ESI-MS: 255.2 (100).

Synthese von Ethyl-1-(4-(aminomethyl)benzyl)-1H-pyrrol-2-carboxylat (24)

Die Reduktion des Nitrils **23** erfolgte unter den Bedingungen der katalytischen Hydrierung. 1,8 g (7,1 mmol) des Nitrils **7** wurden mit 5 g frisch aktiviertem Raney-Nickel in 100 ml ammoniakalischem Methanol im Autoklaven bei 10 bar Wasserstoff über 24 Stunden hydriert. Das gelöste Produkt wurde durch Filtration über Celite von Katalysatorresten befreit, das Lösungsmittel am Rotationsverdampfer bei Unterdruck entfernt.

Analytik: Ausbeute 1,6 g (89%). ¹H-NMR (DMSO-*d*₆): δ 7.26 (t, 1H, pyrr-5H), 7.23 (d, *J* = 8.1, 2H, ph-3,5H), 7.01 (d, *J* = 8.1, 2H, ph-2,6H), 6.88 (dd, 1H, pyrr-3H), 6.15 (dd, 1H, pyrr-4H), 5.49 (s, 2H, pyrr-CH₂-ph), 4.14 (q, 2H, ethyl-CH₂), 1.20 (t, 3H, ethyl-CH₃). ESI-MS: 258.9 (100).

Synthese von Ethyl-1-(4-((2-naphthamido)methyl)benzyl)-1H-pyrrol-2-carboxylat (25)

1,5 g (5,8 mmol) Amin **24** und 1,1 g (5,8 mmol) 2-Naphtylchlorid wurden in 10 ml 1,2-Dichlorethan gelöst. Das Reaktionsgemisch wurde 15 min bei eingestellter Temperatur von 90° C im Mikrowellenofen gerührt. Nach dem Einengen am Rotationsverdampfer wurde das Produkt durch Säulenchromatographie aufgereinigt. Als Laufmittel diente Ethylacetat/Hexan im Verhältnis 1:1.

Analytik: Ausbeute 2,2 g (93%). ¹H-NMR (DMSO-*d*₆): δ 9.12 (t, 1H, NH), 8.47 (s, 1H, naph-1H), 8.02-7.91 (m, 4H, naph-3,4,5,8H), 7.69-7.53 (m, 2H, naph-6,7H), 7.28 (d, *J* = 8.2, 2H, ph-3,5H), 7.25 (t, 1H, pyrr-5H), 7.05 (d, *J* = 8.2, 2H, ph-2,6H), 6.87 (dd, 1H, pyrr-3H), 6.15 (dd, 1H, pyrr-4H), 5.51 (s, 2H, pyrr-CH₂-ph), 4.47 (d, 2H, ph-CH₂-NH), 4.13 (q, 2H, ethyl-CH₂), 1.19 (t, 3H, ethyl-CH₃). ESI-MS: 413.3 (100).

Synthese von 1-(4-((2-Naphthamido)methyl)benzyl)-1H-pyrrol-2-carbonsäure (26)

2,1 g (5,1 mmol) Ester **25** und 0,85 g (15,1 mmol) Kaliumhydroxid wurden in 10 ml des Lösungsmittelgemischs Tetrahydrofuran/ Methanol/ H₂O im Verhältnis 1:1:2 gelöst und im Mikrowellenofen bei 70° C und 35 W für 15 min gerührt. Organische Lösungsmittel wurden am Rotationsverdampfer entfernt. Durch tropfenweise Zugabe von 10-N Salzsäure bis zum Einstellen des pH 7 fiel das Produkt aus. Die Aufreinigung erfolgte durch Umkristallisation in Ethanol (96%).

Analytik: Ausbeute 1,8 g (92%). $^1\text{H-NMR}$ ($\text{DMSO-}d_6$): δ 9.16 (t, 1H, NH), 8.48 (s, 1H, naph-1H), 8.02-7.93 (m, 4H, naph-3,4,5,8H), 7.63-7.55 (m, 2H, naph-6,7H), 7.28 (d, $J = 8.0$, 2H, ph-3,5H), 7.17 (t, 1H, pyrr-5H), 7.05 (d, $J = 8.0$, 2H, ph-2,6H), 6.82 (dd, 1H, pyrr-3H), 6.11 (dd, 1H, pyrr-4H), 5.52 (s, 2H, pyrr- CH_2 -ph), 4.23 (d, 2H, ph- CH_2 -NH). $^{13}\text{C-NMR}$ ($\text{DMSO-}d_6$): 166.2 (CO-NH), 161.9 (COOH), 138.5 (naph-2C), 137.7 (ph-C), 134.1 (ph-C), 132.1 (naph-10C), 131.7 (naph-9C), 129.1 (naph-CH), 127.8 (naph-CH), 127.54 (naph-CH), 127.50 (naph-CH), 127.45 (naph-CH), 127.39 (ph-2,6C), 126.7 (ph-3,5C), 126.3 (naph-CH), 124.2 (naph-CH), 122.4 (pyrr-C2), 117.6 (pyrr-C3), 107.9 (pyrr-C4), 50.5 (pyrr- CH_2 -ph), 42.5 (ph- CH_2 -NH). ESI-MS: 383.1 (100). Schmelzpunkt: 193 °C. Elementaranalyse ($\text{C}_{24}\text{H}_{20}\text{N}_2\text{O}_3 \cdot \text{H}_2\text{O}$): C, H, N.

Ergebnisse und Diskussion

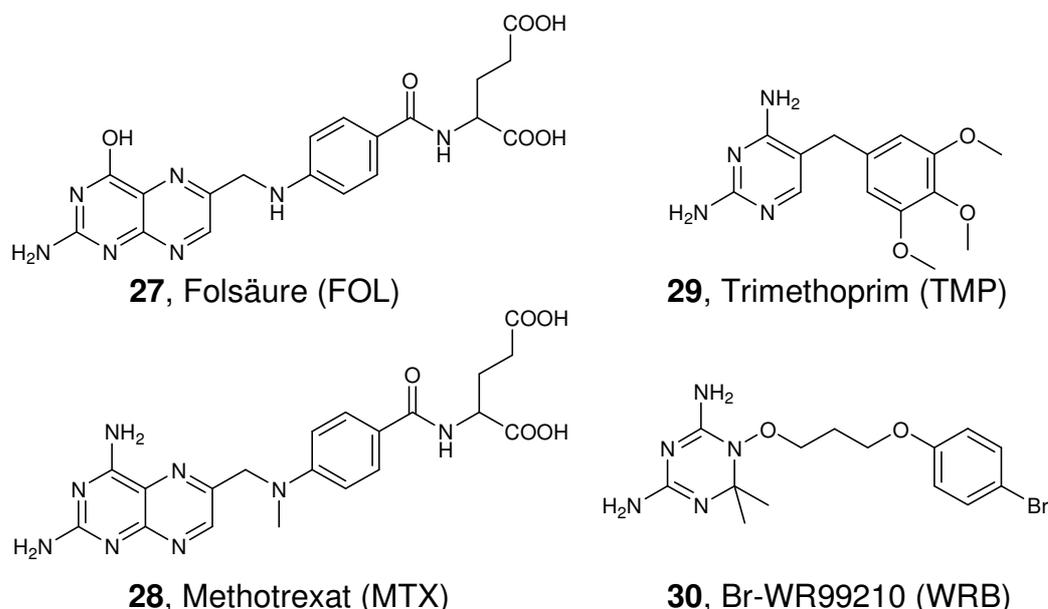
Im Rahmen dieser Arbeit wurden mehrere aufeinander aufbauende Experimente durchgeführt, welche die Zwischenstufen von der Entwicklung rechnergestützter Methoden bis zum Design und der Synthese eines neuen PPAR α -Agonisten dokumentieren. Im Wesentlichen sind die Experimente in folgende vier Gruppen zu unterteilen:

- Molekulares Alignment von Inhibitoren aus co-kristallisierten Röntgenstrukturen (sog. *rigid-body alignment* Experiment). Dieses Experiment sollte Aufschluss darüber geben, ob die Methode in der Lage ist, einander ähnliche Oberflächenbereiche zu erkennen und ob die Überlagerung der Moleküle anhand dieser Bereiche der bioaktiven Konformation entspricht.
- Retrospektives virtuelles Screening. Die Methode sollte im virtuellen Experiment eine Anreicherung der bioaktiven Moleküle erzielen.
- Prospektives virtuelles Screening. Die direkte Suche nach PPAR α -Agonisten in Substanzsammlungen kommerzieller Anbieter und die anschließende Testung im zellbasierten Assay sollte neuartige Strukturen mit gewünschter Aktivität liefern.
- Das *de novo* Design, Synthese und Testung eines PPAR α -Agonisten als endgültiges Ziel dieser Arbeit.

Überlagerung von Molekülen basierend auf Shapelets

Es wird postuliert, dass Moleküle, welche an die gleiche Bindetasche binden, eine ähnliche Oberflächenform besitzen, die komplementär zur Oberfläche dieser Bindetasche ist [Wilchek2006]. Dann müsste eine Methode, die Oberflächenformen vergleicht, eine sinnvolle Überlagerung dieser Moleküle berechnen können. Diese Annahme wurde in einem Starrkörper-Überlagerungsexperiment (engl. *rigid body alignment experiment*) untersucht. Die Protein Data Bank (PDB) [Berman2000] enthält durch Röntgenkristallographie aufgeklärte Strukturen von Proteinen, die mit ihren Liganden bzw. Inhibitoren co-kristallisiert wurden. Aus dieser Datenbank stammen drei Datensätze für das *rigid-body alignment* Experiment.

Der erste Datensatz bestand aus vier Röntgenkristallstrukturen der Dihydrofolatreduktase (DHFR): 1DHF (Ligand Folsäure **27**; FOL) [Davies1990], 1DDR (Ligand Methotrexat **28**; MTX), 1DG5 (Ligand Trimethoprim **29**; TMP) und 1DG7 (Ligand Br-WR99210 **30**; WRB) [Li2000]. Dieser Datensatz wurde bereits von Hofbauer *et al.* [Hofbauer2004] benutzt, um die Methode SURFCOMP zu untersuchen. SURFCOMP berechnet die kritischen Punkte der molekularen Oberfläche und projiziert auf sie zwei physikochemische Eigenschaften: das elektrostatische Potential (ESP) [Hofbauer2004] und das lipophile Potential (LP) [Heiden1993]. Die Überlagerung der Moleküle erfolgt dann mit Hilfe des Bron-Kerbosch-Cliquendetektion [Bron1973] auf dem Assoziationsgraphen, welcher jedoch im Vorfeld durch mehrere Filtertechniken reduziert wird.



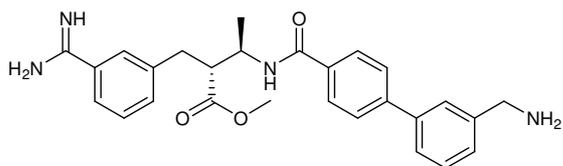
Für die Vorbereitung des Datensatzes wurden die vier Kristallstrukturen anhand der Positionen der C α -Atome mit dem Programm MOE überlagert. Aus dieser Überlagerung wurden paarweise zwei Moleküle gewählt. Eines der Moleküle diente als Referenz, das andere wurde mit Hilfe der *Shapelets*-Methode mit dem Referenzmolekül überlagert. Die Güte der Überlagerung wurde anhand des RMSD zu der Überlagerung der Kristallstrukturen berechnet. Angemerkt sei, dass es keine Rolle spielt, welches Molekül als Referenz diente, da es für die Erstellung des Assoziationsgraphen nicht von Bedeutung ist.

Die Ergebnisse des ersten Experiments sind in der Tabelle 2 zusammengefasst und waren mit den von SURFCOMP erzielten Werten vergleichbar.

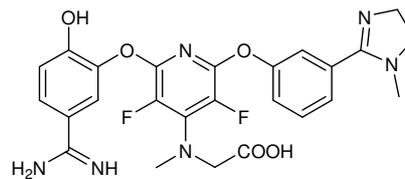
Tabelle 2: Ergebnisse der paarweise Überlagerung der Moleküle aus dem DHFR-Datensatz.

Referenzmolekül	Zielmolekül	SURFCOMP ESP	<i>Shapelets</i>
		RMSD [Å]	RMSD [Å]
FOL	MTX	1,2	1,2
	TMP	1,9	3,1
	WRB	1,3	1,1
MTX	TMP	1,6	2,1
	WRB	5,8	0,9
TMP	WRB	1,7	1,6
Median		1,7	1,4
Mittelwert		2,26	1,68
SD		1,762	0,840

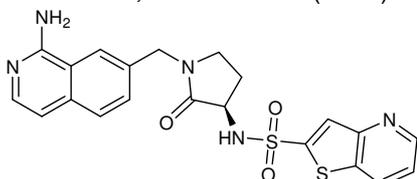
Um die Ergebnisse aus dem ersten Experiment zu untermauern, wurde ein zweiter Datensatz aus co-kristallisierten Faktor Xa-Liganden zusammengestellt. Die PDB-Röntgenkristallstrukturen 1EZQ (Ligand RPR128515 **31**, RPR), 1FOR (Ligand RPR208815 **32**, 815), 1FOS (Ligand RPR208707 **33**, PR2) [Maignan2000], 1FJS (Ligand **34**, Z34) [Adler2000] und 1KSN (Ligand FXV673 **35**, FXV) [Guertin2002] wurden auf dieselbe Weise wie die DHFR-Röntgenkristallstrukturen vorbereitet. Die Ergebnisse (Tabelle 3) auf diesem Datensatz waren vergleichbar mit denen, die auf dem DHFR Datensatz erzielt wurden.



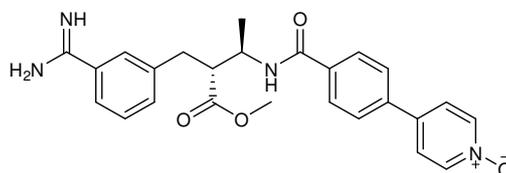
31, RPR128515 (RPR)



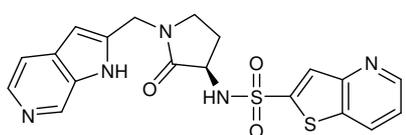
34, Z34



32, RPR208815 (815)



35, FXV673 (FXV)

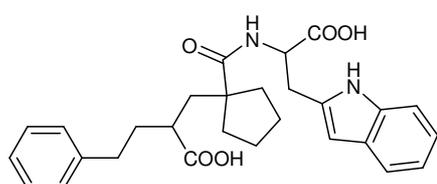


33, RPR208707 (PR2)

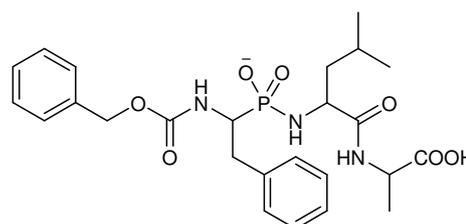
Tabelle 3: Ergebnisse der paarweise Überlagerung der Moleküle aus dem Faktor Xa-Datensatz.

		<i>Shapelets</i>
Referenzmolekül	Zielmolekül	RMSD [Å]
815	FXV	1,6
	PR2	0,5
	RPR	1,1
	Z34	2,0
FXV	PR2	1,4
	RPR	0,9
	Z34	1,8
PR2	RPR	1,3
	Z34	3,4
RPR	Z34	1,5
Median		1,5
Mittelwert		1,56
SD		0,774

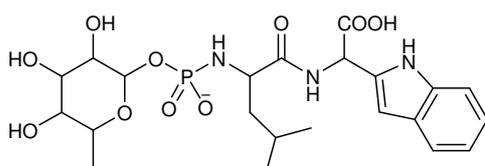
Eine andere Methode, die von Cosgrove *et al.* [Cosgrove2000] entwickelt wurde, benutzt die Zerlegung der Oberfläche in Ausschnitte mit annähernd gleicher Krümmung. Im Unterschied zu *Shapelets* benutzt Cosgrove einen Clustering-Algorithmus, um Punkte mit ähnlicher Krümmung zusammenzufassen. Die Validierung der Software von Cosgrove *et al.* erfolgte auf dem Thermolysin-Datensatz, bestehend aus acht Röntgenkristallstrukturen von Thermolysin, co-kristallisiert mit peptidischen Liganden. Die Röntgenkristallstrukturen 1THL **36** [Holland1994], 1TLP **37** [Tronrud1986], 1TMN **38** [Monzingo1984], 3TMN **39** [Holden1987a], 4TMN **40**, 5TMN **41** [Holden1987b], 5TLN **42** [Holmes1981] und 6TMN **43** [Tronrud1987] wurden auf dieselbe Art aufbereitet wie der DHFR- und Faktor Xa-Datensatz. Da SURFCOMP auch auf dem Thermolysin-Datensatz getestet wurde, bietet dieser eine Basis für den Vergleich der Methoden. Die Ergebnisse für alle Methoden sind in der Tabelle 4 zusammengefasst. Der Median der RMSD ist vergleichbar mit den Ergebnissen, die auf dem DHFR- und Faktor Xa-Datensatz erzielt wurden. Der Mittelwert ist jedoch mit 2,54 Å um 0,86 Å bzw. 0,98 Å schlechter.



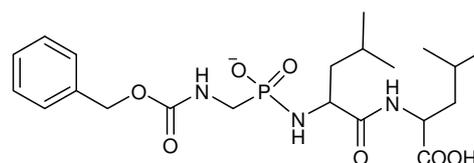
36, 1THL



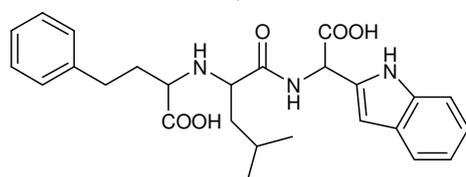
40, 4TMN



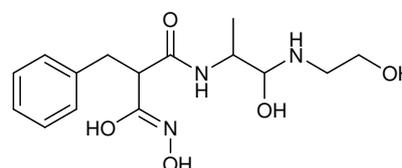
37, 1TLP



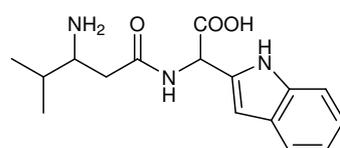
41, 5TMN



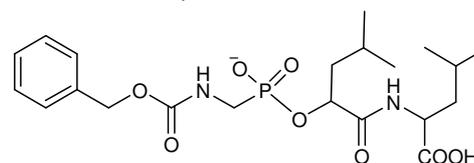
38, 1TMN



42, 5TLN



39, 3TMN



43, 6TMN

Tabelle 4: Ergebnisse der paarweise Überlagerung der Moleküle aus dem Thermolysin-Datensatz.

Referenz- molekül	Ziel- molekül	SURFCOMP	SURFCOMP	Cosgrove	
		ESP RMSD [Å]	LP RMSD [Å]	<i>et al.</i> RMSD [Å]	<i>Shapelets</i> RMSD [Å]
1THL	1TLP	0,6	1,0	0,8	1,1
	1TMN	0,4	0,3	0,4	0,2
	3TMN	0,3	0,6	0,5	0,7
	4TMN	1,2	1,0	0,4	0,5
	5TLN	5,7	5,1	6,8	4,4
	5TMN	1,2	7,1	1,1	1,2
	6TMN	0,6	0,7	5,3	1,2
1TLP	1TMN	0,5	1,4	0,9	0,9
	3TMN	0,4	1,2	0,4	0,7
	4TMN	1,3	6,0	9,7	0,9
	5TLN	0,6	1,2	6,2	6,8
	5TMN	0,6	1,2	1,4	1,1
	6TMN	0,6	1,0	9,2	1,5
1TMN	3TMN	1,0	0,8	1,1	1,0
	4TMN	1,0	0,8	0,9	1,1
	5TLN	5,1	6,3	7,3	4,7
	5TMN	0,9	0,8	0,9	1,4
	6TMN	1,0	0,9	0,7	1,9
3TMN	4TMN	1,4	5,5	7,7	8,4
	5TLN	2,9	6,9	7,2	4,2
	5TMN	1,5	1,2	9,2	6,0
	6TMN	4,3	0,7	5,7	9,2
4TMN	5TLN	5,8	6,2	8,2	7,1
	5TMN	0,4	0,5	2,0	0,9
	6TMN	0,6	0,5	2,1	0,8
5TLN	5TMN	2,3	3,4	9,7	1,6
	6TMN	5,8	1,2	9,9	1,5
5TMN	6TMN	0,1	0,1	0,1	0,2

Median	1,0	1,1	2,0	1,2
Mittelwert	1,78	2,47	4,14	2,54
SD	1,844	2,519	3,719	2,656

Aus den Ergebnissen der Überlagerung der DHFR- und Faktor Xa-Liganden ließ sich ableiten, dass *Shapelets* prinzipiell dazu geeignet ist, die Überlagerung zweier wirkstoffähnlicher Moleküle zu finden. Die Geschwindigkeit der Berechnungen war im Vergleich zu der SURFCOMP Implementierung hoch (4,5 s pro Überlagerung auf einer 3GHz Intel Pentium IV Workstation; im Vergleich zu 75 s pro Überlagerung auf einer 2,4 GHz Intel Xeon Workstation). Der Geschwindigkeitsvorteil der *Shapelets*-Methode lag in der vergleichbar geringen Anzahl der Paraboloiden (*Shapelets*: 8-21 Paraboloiden, SURFCOMP: 150-950 Oberflächenpunkte für ein wirkstoffähnliches Molekül), die für die Berechnung der Überlagerung hinzugezogen werden. Der Geschwindigkeitsvergleich mit dem Programm von Cosgrove *et al.* war nicht möglich, da seine Berechnungen auf einer R5000 SGI O2 Workstation durchgeführt wurden. Die Güte der Überlagerungen, die von Cosgroves Programm berechnet werden, lagen auf dem Thermolysin-Datensatz im Mittel hinter denen von *Shapelets* (Cosgrove *et al.*: 4,14Å, *Shapelets*: 2,54Å). Die Mediane der Ergebnisse von SURFCOMP und *Shapelets* auf dem DHFR- (SURFCOMP: 1,7Å, *Shapelets*: 1,4Å) und dem Thermolysin-Datensatz (SURFCOMP ESP: 1,0Å, SURFCOMP LP: 1,0Å, *Shapelets*: 1,2Å) zeigen, dass beiden Methoden in vielen Fällen die Molekülüberlagerung gut reproduzieren können. Da die Mittelwerte bei beiden Methoden auf dem Thermolysin-Datensatz wesentlich höher sind (SURFCOMP ESP: 1,78Å, SURFCOMP LP: 2,47Å, *Shapelets*: 2,54Å), gibt es in diesem Datensatz Ausreißer, die sich schlecht überlagern lassen. SURFCOMP ESP zeigt niedrigeren Mittelwert und Standardabweichung (1,78Å±1,84 Å) und scheint daher stabiler gegen Ausreißer zu sein.

Aus den Ergebnissen des *rigid body alignment* Experiments ließen sich folgende Schlussfolgerungen ableiten:

- *Shapelets* ist prinzipiell in der Lage, eine Überlagerung zweier Moleküle zu finden, die der Überlagerung bioaktiver Konformationen entspricht.

- Die Berechnungsgeschwindigkeit ist für Operationen auf Datenbanken bis zu einer Größenordnung von 10^5 Molekülen akzeptabel.
- Durch Einbindung weiterer Information, vergleichbar mit SURFCOMP ESP, könnte eine höhere Stabilität der Methode bei der Berechnung der Überlagerungen erreicht werden. Eine mögliche Erweiterung wäre die Einbindung von Pharmakophor-Informationen oder physikochemischen Eigenschaften.

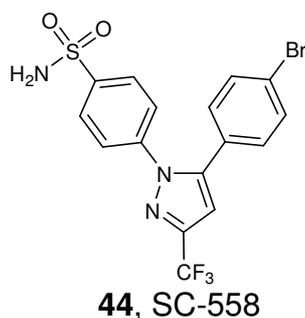
Oberflächen-basiertes virtual screening

Im vorherigen Experiment wurde gezeigt, dass es möglich ist, mit Hilfe der *Shapelets*-Methode zwei Moleküle zu überlagern. Die Voraussetzung dafür ist die Kenntnis der Konformation der Liganden in der Bindetasche. Aus diesen Ergebnissen ließ sich nun eine neue Hypothese ableiten: man kann anhand der Oberflächenform zwischen aktiven und inaktiven Molekülen unterscheiden. Um diese Hypothese zu verifizieren, wurde ein retrospektives virtuelles Screening-Experiment durchgeführt. Dazu wurden für die Moleküle aus der COBRA6.1 Datenbank multiple Konformere mit dem Programm MOE berechnet. Im Durchschnitt entstanden so 4,5 Konformere pro Molekül, deren Energie unterhalb des Schwellenwertes von 6 kcal/mol lagen. Drei affine Liganden, co-kristallisiert mit ihren Zielproteinen wurden in der bioaktiven Konformation als Referenzmoleküle für das virtuelle Screening verwendet. Das Molekül SC-558 ist ein selektiver, potenter COX-2 Inhibitor (IC_{50} : 9,3 nM). Die bioaktive Konformation wurde aus der PDB-Struktur 6COX [Kurumbail1996] entnommen. Die Oberflächen aller Molekülkonformationen aus der COBRA6.1 Datenbank wurden berechnet, mit der *Shapelets*-Methode zerlegt und mit SC-558 überlagert. Die Moleküle wurden sortiert nach Güte der Überlagerung ausgegeben. Der Anreicherungsfaktor (engl. *enrichment factor*, *EF*) [Xu2002b] im ersten Perzentil (**Formel 12**) der Datenbank wurde berechnet.

$$EF = \frac{\# \text{Aktive}_{\text{gefunden}}}{\# \text{Aktive}_{\text{erwartet}}}$$

Formel 12

Die Datenbank COBRA6.1 enthält 8311 Moleküle, darunter 136 COX-2 Inhibitoren. Im ersten Perzentil (83 Moleküle) der sortierten Liste waren 31 COX-2 Inhibitoren, dies entspricht einem *EF* von 23.



Dieses Experiment wurde mit dem DHFR Inhibitor Methotrexat aus der Röntgenkristallstruktur 1DDR und dem Faktor Xa Inhibitor FXV673 aus der Röntgenkristallstruktur 1KSN wiederholt. Die Ergebnisse sind in der Tabelle 5 zusammengefasst.

Tabelle 5: Ergebnisse des retrospektiven virtual screening-Experiments mit der *Shapelets*-Methode.

Referenzmolekül	Anzahl aktiver Strukturen in COBRA6.1	Anzahl aktiver Strukturen im ersten Perzentil	EF für das erste Perzentil
SC-558 44 , COX-2 Inhibitor	136	31	23
Methotrexat 28 , DHFR Inhibitor	64	3	5
FXV673 35 , Faktor Xa Inhibitor	228		0

Vergleicht man diese Ergebnisse mit veröffentlichten virtual screening Studien auf der COBRA6.1 Datenbank [Fechner2003], erkennt man, dass *Shapelets* effektiv nicht in der Lage war, aktive Moleküle von Inaktiven zu unterscheiden. Eine gute Anreicherung der COX-2 Inhibitoren war der Tatsache zuzuschreiben, dass ca. 70% der in COBRA6.1 enthaltenen selektiven COX-2 Inhibitoren die typische Struktur aus einem Fünfring, substituiert mit zwei Phenylgruppen besitzen [Schneider2003] Die Faktor-Xa und DHFR Inhibitoren besitzen unterschiedliche Grundstrukturen.

Folgende Schlussfolgerungen konnten aus den Ergebnissen des virtual screening Experiments gezogen werden:

- Die Oberflächeninformation alleine war wahrscheinlich nicht ausreichend, um mit *Shapelets* aktive von inaktiven Strukturen zu unterscheiden.
- Eine Verbesserung könnte durch die Hinzunahme zusätzlicher Informationen, wie z.B. Pharmakophor-Information erreicht werden.

Virtual screening mit SQUIRREL

Im rigid body alignment Experiment konnte gezeigt werden, dass die *Shapelets*-Methode in der Lage ist, die korrekte Überlagerung zweier aktiver Moleküle zu finden. Das virtual screening Experiment aus dem vorherigen Abschnitt wurde nun unter Zunahme von Pharmakophor-Informationen wiederholt, welches in die Bewertung der Überlagerung einfluss. Das überarbeitete Programm wurde SQUIRREL genannt. Die neuen Ergebnisse sind in Tabelle 6 zusammengefasst.

Tabelle 6: Ergebnisse des retrospektiven virtual screening-Experiments mit der SQUIRREL-Methode.

Referenzmolekül	Anzahl aktiver Strukturen in COBRA6.1	Anzahl aktiver Strukturen im ersten Perzentil	EF für das erste Perzentil
SC-558 44 , COX-2 Inhibitor	136	37	27
Methotrexat 28 , DHFR Inhibitor	64	13	20
FXV673 35 , Faktor Xa Inhibitor	228	18	8

Die Ergebnisse verbesserten sich gegenüber dem reinen Oberflächenvergleich, bei COX-2 um 17 %, bei DHFR um 300%. Bei Faktor Xa wurden erst durch die zusätzliche Pharmakophor-Information Aktive angereichert. Diese Ergebnisse lagen im Einklang mit der Studie von McGaughey *et al.* [McGaughey2007]. Diese Arbeit demonstrierte unter anderem in einem groß angelegten retrospektiven virtual screening Experiment, dass die Methoden ROCS [Grant1996] und SQW-shape [Miller1999] in fast allen Einstellungen des Experiments schlechtere Anreicherung lieferten als ihre Pendants mit zusätzlicher Pharmakophor-Information, ROCS-color und SQW. Folgende Schlussfolgerungen ließen sich aus diesem Experiment ziehen:

- Die Hinzunahme der Pharmakophorinformation führte zu einer erheblichen Verbesserung der Anreicherung. Es lässt sich also postulieren, dass die Oberflächenähnlichkeit ein notwendiges, aber kein hinreichendes Kriterium für die Aktivität ist.
- Die Güte der Ergebnisse schwankte stark in Abhängigkeit von dem Zielprotein. Diese Schlussfolgerung wurde von der Mehrheit der bisher durchgeführten Studien [Fechner2003, Hawkins2007, McGaughey2007] gezogen. Daher war

für weitere Arbeit eine retrospektive Validierung der Methode auf PPAR notwendig.

Retrospektive virtual screening Studie auf PPARs

Um die Anreicherung der SQUIRREL-Methode auf PPARs zu testen, wurden jeweils ein potenter, selektiver Agonist von PPAR α und PPAR γ aus der PDB extrahiert. Beim PPAR α -Agonisten handelte es sich um das Fibratderivat GW590735 **5** aus der Röntgenkristallstruktur 2P54 [Sierra2007]. GW590735 aktiviert PPAR α bei einem EC_{50} von 4 nM mit über 1000facher Selektivität gegenüber PPAR γ . Farglitazar **8** ist ein potenter, selektiver PPAR γ -Agonist [Xu2001] (EC_{50} : 0,2 nM auf PPAR γ , 250 nM auf PPAR α). Dieser wurde aus der Röntgenkristallstruktur 1FM9 [Gampe2000] entnommen und für die virtual screening Studie vorbereitet.

Die COBRA6.1 Datenbank enthält 44 selektive PPAR γ - und 16 selektive PPAR α -Liganden. Die multiplen Konformere wurden diesmal von der OMEGA-Software generiert. Zum Vergleich wurde neben SQUIRREL folgende Software eingesetzt: ROCS, ROCS-color, MOE 4-Punkt Pharmacophore Fingerprints (MOE-4P), LIQUID und CATS. LIQUID, CATS und MOE-4P sind Methoden, die nur Pharmakophor-Informationen benutzen, ROCS vergleicht ausschließlich die Form der Moleküle. SQUIRREL und ROCS-color verwenden sowohl die Form als auch die Pharmakophorinformation.

Abbildung 21 zeigt die Ergebnisse des Experiments. Die Anreicherungskurven in den ersten 2% der Datenbank ließen darauf schließen, dass die Methoden SQUIRREL und ROCS-color am besten für das virtuelle Screening auf PPARs geeignet sind. Die Kombination der Oberflächen- bzw. Volumenüberlagerung mit der Pharmakophorinformation ist für PPARs anscheinend besonders vorteilhaft.

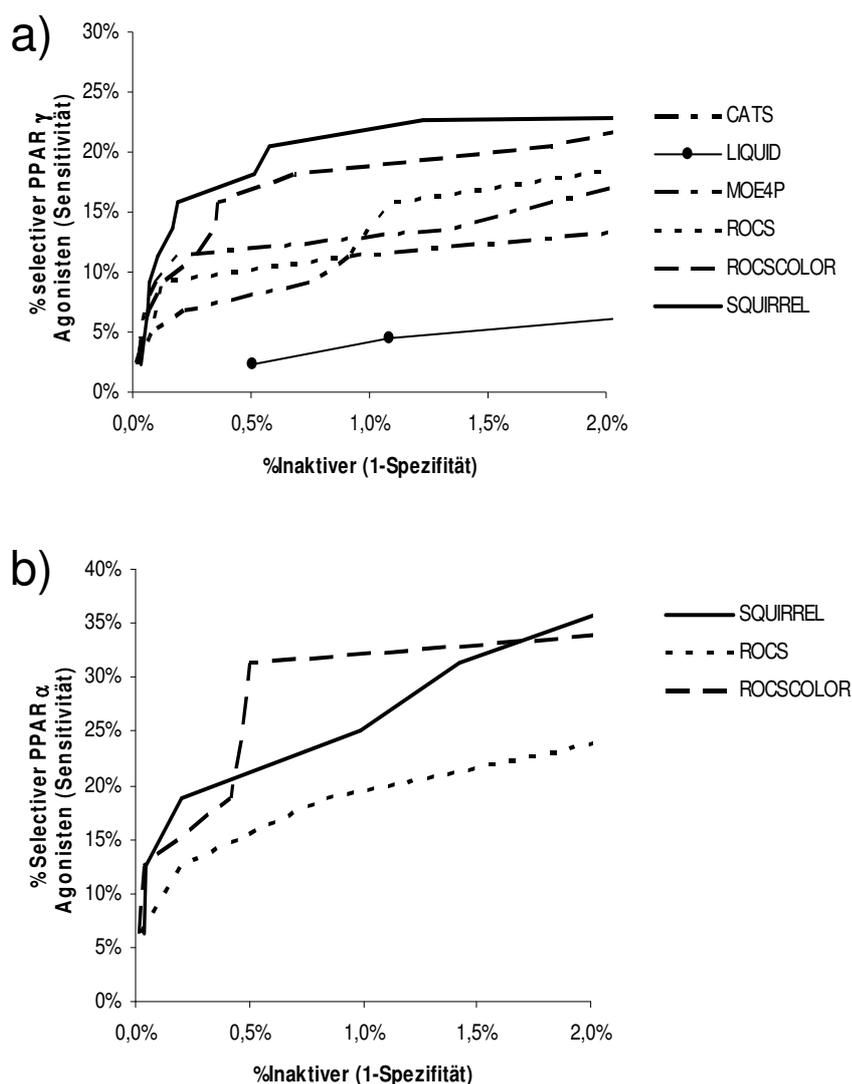


Abbildung 21: Ergebnisse der retrospektiven virtual screening Studien mit den Methoden CATS, LIQUID, MOE 4-Punkt-Pharmakophor-Fingerprints (MOE4P), ROCS, ROCS-color und SQUIRREL auf der COBRA6.1 Datenbank. a) Anreicherung der selektiven PPAR γ -Agonisten gegen die Referenzstruktur Farglitazar 8, angegeben als receiver-operator curve (ROC). b) Anreicherung der selektiven PPAR α -Agonisten gegen die Referenzstruktur GW590735 5, angegeben als receiver-operator curve (ROC). Die Methoden CATS, LIQUID und MOE4P haben in den ersten 2 Prozent der sortierten Datenbank keine PPAR α -Agonisten gefunden.

Eine mögliche Erklärung wurde durch die Betrachtung der Bindetasche der PPARs deutlich. Links unten auf der Abbildung 22 erkennt man, dass die Bindetasche von PPAR α tief im Inneren des Proteins vergraben ist. Der größte Teil der Oberfläche des Liganden interagiert mit der Oberfläche der Bindetasche, daher ist die

Komplementarität von Bindetasche und Ligand, und damit die Form des Liganden, von hoher Bedeutung. Andererseits sind die Interaktionen der aciden Kopfgruppe von GW590735 mit den Aminosäuren Ser280, Tyr314, His440 und Tyr464 wichtig für die Aktivierung des kernlokalisierten Rezeptors (Abbildung 22 oben rechts). Kombiniert man diese Information, ist es möglich, PPAR-Agonisten anhand der Ähnlichkeit zu erkennen.

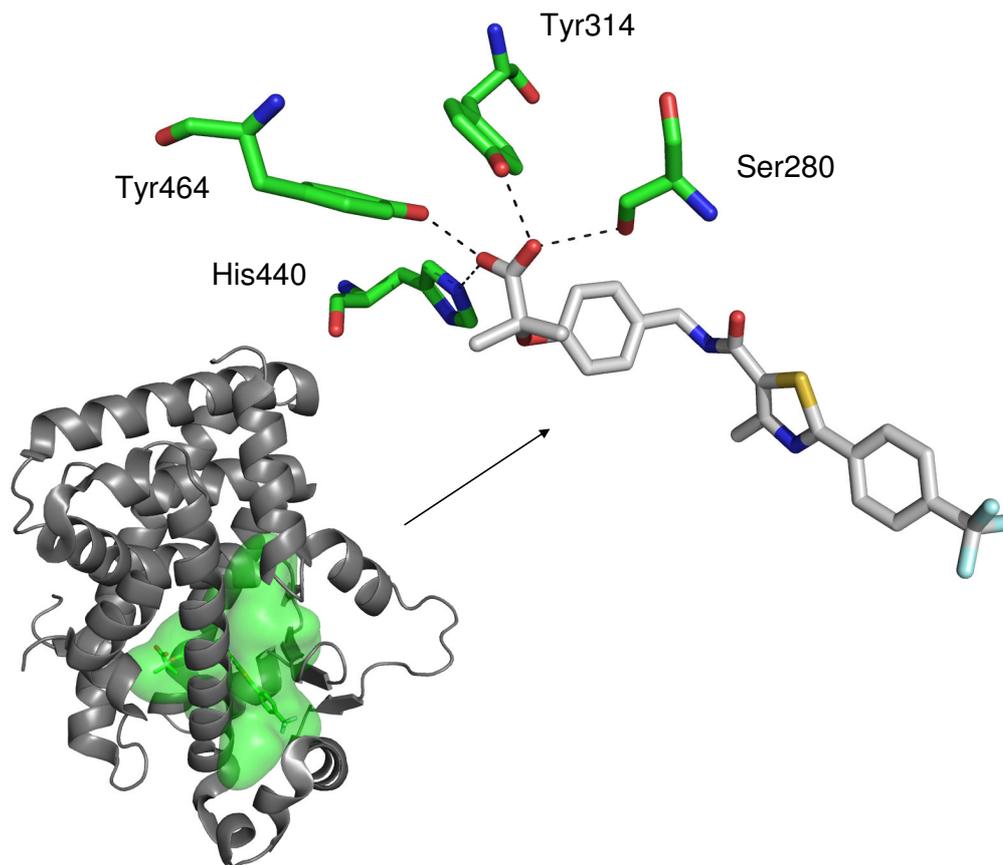


Abbildung 22: Bindemodus des selektiven PPAR α -Agonisten GW590735 5 in der Röntgenkristallstruktur 2P54 [Sierra2007]. Die Bindetasche ist vergraben, die für die Aktivierung wichtigen Interaktionen sind in der Vergrößerung gekennzeichnet.

Prospektive virtual screening Studie auf PPAR α

Die retrospektive Studie lieferte Anhaltspunkte darüber, dass SQUIRREL gut geeignet ist, PPAR-Liganden zu erkennen, da es die Form und die Pharmakophorinformation in die Suche einbezieht. Ausgehend von diesen Ergebnissen wurde eine prospektive Studie zum Auffinden neuer selektiver PPAR α -Agonisten gestartet. Die kommerzielle

Anbieterdatenbank SPECS (199272 Substanzen, v.06.2007; www.specs.net) wurde aufgrund der hohen Reinheit der Substanzen für die Studie ausgewählt. Ein mehrstufiges virtual screening Verfahren [Klebe2006] wurde durchgeführt, bestehend aus folgenden Schritten:

- Filtern der Datenbank nach relevanten physikochemischen Eigenschaften.
- Projektion der Datenbank auf eine SOM, um PPAR α -ähnliche Liganden auszuwählen.
- Erstellung multipler Konformere und anschließendes Sortieren nach Ähnlichkeit zu bekannten, potenten Referenzverbindungen.
- Manuelle Auswahl von Testkandidaten aus der sortierten Liste.

Aufgrund der Größe und Lipophilie der Bindetasche und der natürlichen Substrate der PPAR erschien es angemessen, Molekulargewicht (MW) und den berechneten Wasser-Oktanol Verteilungskoeffizienten (clogP) als wichtige physikochemische Eigenschaften anzusehen. Die Statistik über die in COBRA6.1 enthaltenen PPAR-Agonisten ergab ein mittleres Molekulargewicht von 430 ± 93 Da, der mittlere clogP befand sich bei $5,3 \pm 1,92$. Für die Filterung wurde die doppelte Standardabweichung oberhalb und unterhalb des Mittelwerts benutzt [Schneider2004]. Somit wurden nur Substanzen, deren Eigenschaften innerhalb der Filterintervalle ($244 < MW < 616$ Da; $1,44 < clogP < 9,12$) liegen, weiterverwendet.

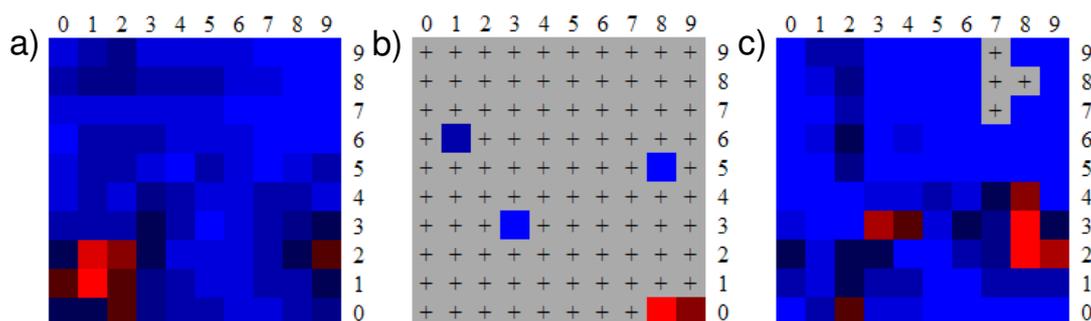


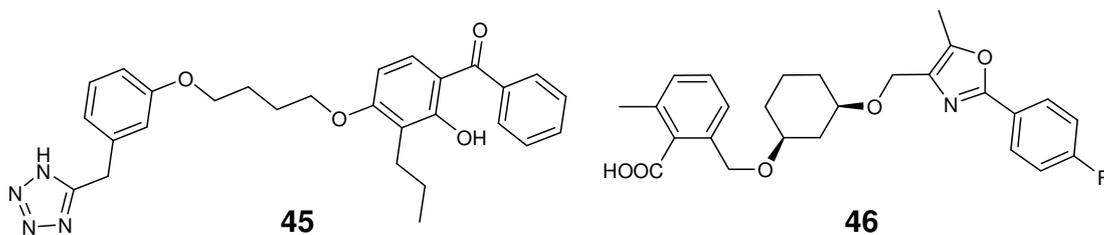
Abbildung 23: SOMs, trainiert auf der COBRA6.1 Datenbank mit dem CATS Deskriptor. Die relative Anzahl der Substanzen, die jedem Neuron zugeordnet sind, werden auf einer Farbskala von blau (wenig) nach rot (viel) codiert a) Verteilung aller COBRA6.1 Strukturen. b) Verteilung potenter und selektiver PPAR α -Agonisten. c) Projektion der Substanzen der SPECS Datenbank auf die SOM.

Im nächsten Schritt wurden die verbliebenen 176922 Substanzen auf eine SOM projiziert. Dazu wurde eine toroidale SOM mit 10×10 Neuronen auf der COBRA6.1 Datenbank trainiert, für die vorher der CATS Deskriptor berechnet wurde. Die Anzahl der Trainingsepochen betrug 100000, $\sigma_{initial} = 5$, $\lambda_{initial} = 1$ (Formel 10) und Manhattan-Abstandsmaß.

Abbildung 23a zeigt die Verteilung der COBRA6.1 Strukturen auf der SOM. Projiziert man 14 selektive potente PPAR α -Agonisten auf diese SOM, sieht man dass bestimmte Bereiche präferiert werden (sog. „Aktivitätsinseln“ [Schneider2003], Abbildung 23b). Um Substanzen auszuwählen, die PPAR α -Agonisten ähnlich sind, wird die SPECS Datenbank auf die SOM projiziert (Abbildung 23c). Die 1926 Substanzen, die auf PPAR α -Aktivitätsinseln fallen (mehrfach besetzte Neuronen 1/6, 8/0, und 9/0) wurden für die weiteren Berechnungen verwendet.

Von diesen PPAR α -ähnlichen Substanzen wurden multiple Konformere mit dem Programm MOE berechnet. Diese wurden für die Berechnung der Oberflächenfragmente und die Erstellung sortierter Listen mit der SQUIRREL-Methode benutzt.

Als Referenzverbindungen dienten GW590735 **5**, das Tetrazol-Derivat **45** und die Verbindung **46**. GW590735 **5** wurde mit PPAR α co-kristallisiert (PDB-Code 2P54 [Sierra2007]). Verbindung **45** aktiviert PPAR α mit $EC_{50} = 140$ nM und zwölfte Selektivität gegenüber PPAR γ [Etgen2003]. Die Struktur **46** ist ein außerordentlich potenter PPAR α -Agonist ($EC_{50} = 0,3$ nM) [Miyachi2004]. Die Referenzverbindungen wurden ausgewählt, um eine möglichst hohe Variabilität der für die PPAR-Subtypselektivität entscheidenden aciden Kopfgruppe zu gewährleisten. Da von den Verbindungen **45** und **46** keine Röntgenkristallstrukturen verfügbar waren, wurden die bioaktiven Konformationen dieser Substanzen durch Docking in die PPAR α -Kristallstruktur 2P54 mit GOLD 3.2 [Verdonk2003] modelliert. Die Einstellungen des Programms finden sich im Anhang. Für die Berechnung der Oberflächenfragmente wurden jeweils ein potentieller Bindemodus ausgewählt, der die Interaktion mit den Aminosäuren S280, Y314, H440 und Y464 zeigte und in der charakteristischen U-Konformation vorlag, die Ausgabe der Bewertungsfunktion wurde nicht berücksichtigt. Es wurden insgesamt drei Listen erstellt, jede sortiert nach SQUIRREL-Ähnlichkeit zu einer der Referenzverbindungen.



Diese Listen wurden von oben nach unten manuell durchsucht und jeweils 10 mg der 21 ausgewählten Verbindungen **47-67** bestellt. Dabei wurde insbesondere darauf geachtet, dass die Moleküle

- eine Gruppe enthielten, die als bioisosterer Ersatz der Säuregruppe dienen kann [Lima2005],
- unterschiedliche chemische Grundgerüste aufwiesen,
- zur Vermeidung der potentiellen Toxizität keine Aldehyd- oder Ketogruppen enthalten.

Abbildung 24 zeigt den schematischen Ablauf der virtual screening Strategie, die bei dieser Studie angewendet wurde.

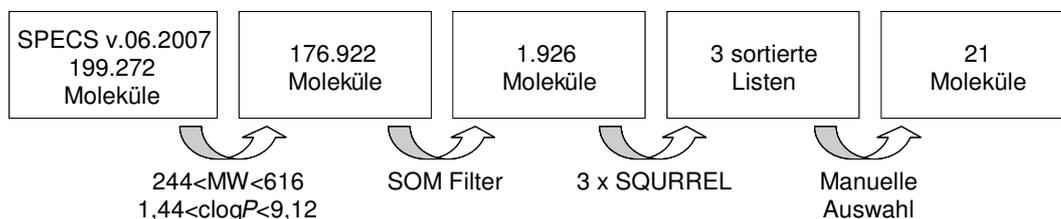


Abbildung 24: Schematischer Ablauf der mehrstufigen virtual screening Strategie.

Die Substanzen wurden dreimal im zellbasierten PPAR α - und PPAR γ -Assay in den Konzentrationen 3 μ M, 10 μ M, 30 μ M und 100 μ M getestet. Wenn eine hohe signifikante Aktivierung bei diesen Konzentrationen beobachtet wurde, wurden EC_{50} -Werte bestimmt. Tabelle 7 Aktivität der Substanzen 47-67 im zellbasierten Testsystem. Die maximale Konzentration der Substanzen 47-67 betrug 100 μ M. Die Aktivität ist als EC_{50} angegeben oder, falls EC_{50} nicht messbar war, als % Aktivierung bei Höchstkonzentration von 100 μ M. gibt die Ergebnisse dieser Untersuchungen wieder.

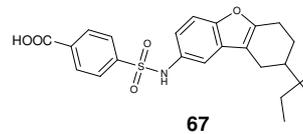
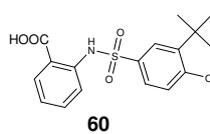
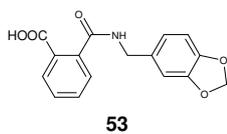
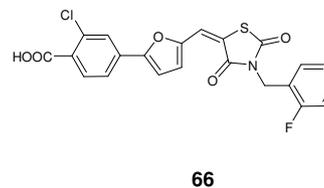
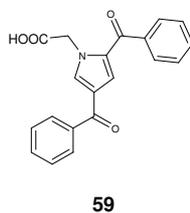
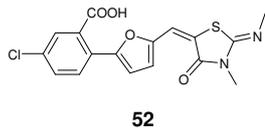
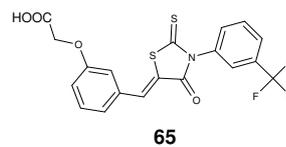
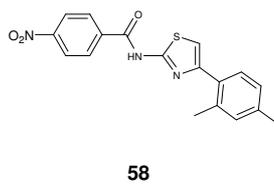
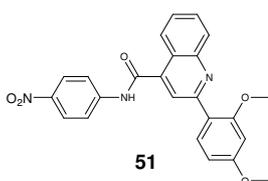
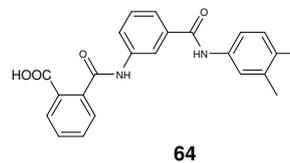
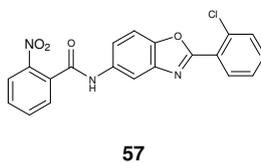
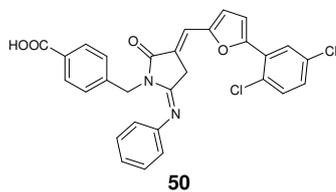
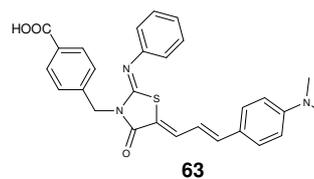
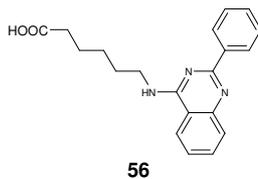
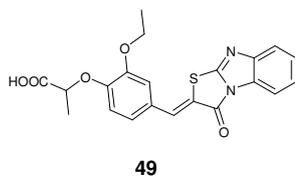
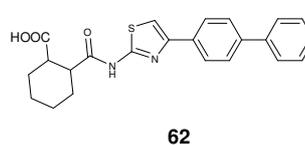
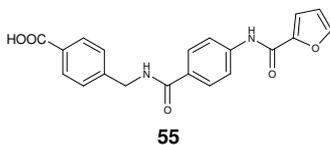
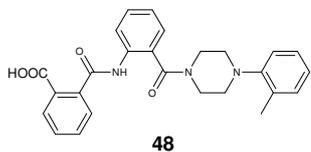
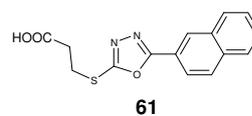
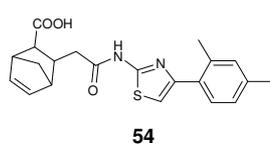
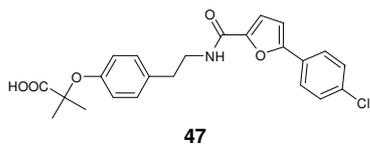


Tabelle 7 Aktivität der Substanzen 47-67 im zellbasierten Testsystem. Die maximale Konzentration der Substanzen 47-67 betrug 100 μM . Die Aktivität ist als EC_{50} angegeben oder, falls EC_{50} nicht messbar war, als % Aktivierung bei Höchstkonzentration von 100 μM .

Substanznummer	PPAR α Agonismus	PPAR γ Agonismus
47	1,75 \pm 0,24 μM	13 \pm 2,2 μM
48	inaktiv	inaktiv
49	inaktiv	34 \pm 18% bei 100 μM
50	80 \pm 17% bei 100 μM	inaktiv
51	inaktiv	inaktiv
52	63 \pm 6% bei 100 μM	106 \pm 37 μM
53	34 \pm 14% bei 100 μM	70,5 \pm 8,1 μM
54	inaktiv	47 \pm 12% bei 100 μM
55	inaktiv	inaktiv
56	inaktiv	7 \pm 0,5 μM
57	inaktiv	inaktiv
58	40 \pm 6% bei 100 μM	inaktiv
59	inaktiv	inaktiv
60	inaktiv	inaktiv
61	inaktiv	26 \pm 8% bei 100 μM
62	inaktiv	47 \pm 12% bei 100 μM
63	0,044 \pm 0,005 μM	4,9 \pm 0,4 μM
64	inaktiv	Inaktiv
65	inaktiv	56 \pm 28% bei 100 μM
66	inaktiv	10,5 \pm 0,7 μM
67	20,2 \pm 25,6 μM	19,5 \pm 0,1 μM

Sieben der 21 getesteten Moleküle haben im zellbasierten Assay signifikante Aktivierung von PPAR α gezeigt. Einige Strukturen sind besonders hervorzuheben: Molekül **47** ist von der Fibrinsäure [Grundy1987] abgeleitet und zeigte erwartungsgemäß hohe Potenz (EC_{50} = 1,75 \pm 0,24 μM) und siebenfache Selektivität gegenüber PPAR γ . Da jedoch die Fibrinsäurederivate bereits in hohem Maß als

Leitstruktur für die Aktivierung von PPAR α optimiert wurden [Sierra2007], ist diese Struktur nicht weiter interessant.

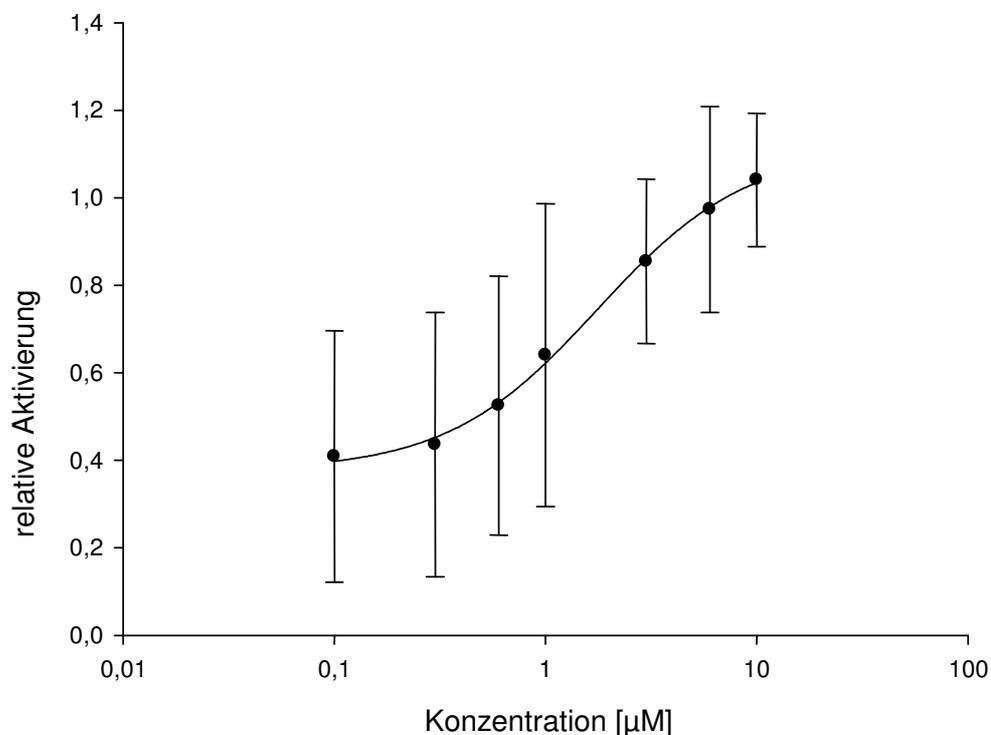


Abbildung 25: Aktivierung von PPAR α durch Substanz 47 in Abhängigkeit von der Konzentration.

Die restlichen Verbindungen zeigten unterschiedliche Chemotypen auf, was auf die Generalisierungsfähigkeit der SQUIRREL-Methode hindeutet. Der Wechsel zwischen den Chemotypen bei gleichzeitigem Erhalt der Aktivität, sog. „Scaffold-Hopping“ [Schneider1999] ist daher gelungen. Insbesondere zeigte Struktur **63** hohe Potenz ($EC_{50} = 44 \pm 5$ nM) und 111fache Selektivität gegenüber PPAR γ . Diese Struktur hat einen anderen Chemotyp als die Referenzstrukturen.

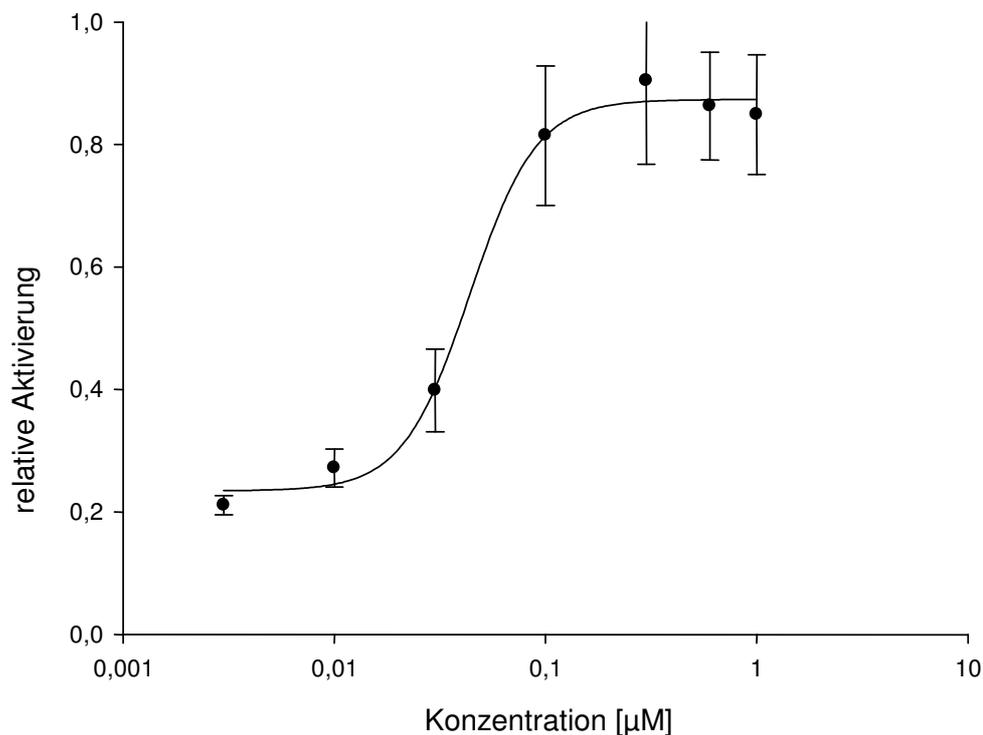


Abbildung 26: Aktivierung von PPAR α durch Substanz 63 in Abhängigkeit von der Konzentration.

Ein Docking-Experiment gab Hinweise auf die Potenz und Selektivität der Verbindung **63**. Dazu wurde aus der PDB-Röntgenkristallstruktur 2P54 [Sierra2007] der Ligand entfernt und hypothetische Bindemoden der Struktur **63** mit der Docking-Software GOLD 3.2 [Verdonk2003] erstellt. Die Einstellungen des Programms finden sich im Anhang. Die Kristallstruktur 2P54 wurde verwendet, weil der co-kristallisierte Ligand GW590735 **5** den linken proximalen Teil der Bindetasche nicht besetzt. In anderen Röntgenkristallstrukturen wird diese Subtasche von einem zu der Carboxylgruppe in der α -Position ständigem Substituenten okkupiert. Da Verbindung **63** nicht über einen solchen Substituenten verfügt, war 2P54 die geeignete Wahl für die Durchführung der Docking-Experimente. Der Bindemodus mit dem höchsten *Chemscore*-Wert (50) ist in Abbildung 27 gezeigt. In diesem Modell interagierte die Carboxylgruppe mit den Seitenketten der Aminosäuren S280, Y314, H440 und Y464. Diese Interaktionen sind für die Stabilisierung der AF-2 Helix und die Aktivierung des Rezeptors essentiell [Xu2001]. Zusätzliche Interaktionen gehen von der Phenylgruppe in der rechten distalen Tasche [Pirard2003] des Rezeptors aus. Der lange hydrophobe Rest besetzt die linke

distale Bindetasche. Vermutlich sind diese Interaktionen für die Selektivität gegenüber PPAR γ verantwortlich, da dort dieser Bereich vornehmlich von voluminösen Resten besetzt wird. Die linke proximale Tasche, die bei PPAR α kleiner ist als bei PPAR γ [Zeote2007], wurde nicht besetzt, auch dies sprach für die hohe Selektivität dieser Verbindung.

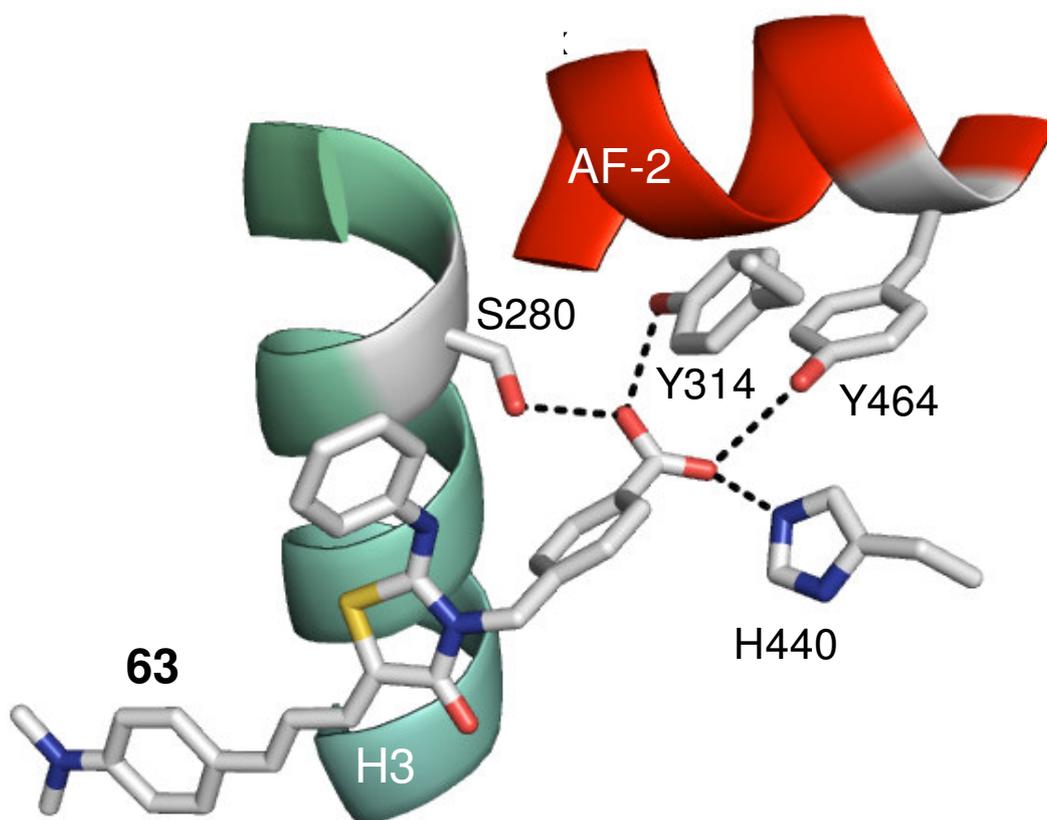


Abbildung 27: Potentieller Bindemodus der Verbindung 63, berechnet mit Docking-Software GOLD.

De novo Design und Synthese eines PPAR α -Agonisten

Einer der Vorteile der SQUIRREL-Methode war die Verwendung eines graphenbasierten Ansatzes, welcher auch partielle Überlagerungen erlaubt. Diese partielle Überlagerungen konnten dazu benutzt werden, Molekülfragmente mit bestimmten Teilen eines Moleküls zu überlagern und somit Vorschläge für bioisostere Ersatz dieser Teile zu generieren. In dieser Studie sollten diese Vorschläge dazu verwendet werden, einen PPAR α -Agonisten *de novo* zu erzeugen, zu synthetisieren und die biologische Aktivität zu bestätigen.

Dazu wurden multiple Konformere für die COBRA6.1 Datenbank mit der MOE Software erzeugt. Die Oberflächen wurden berechnet und mit der *Shapelets* Software in Paraboloiden unterteilt. Die Moleküle wurden nun mit den RECAP Retrosyntheseregeln [Lewell1998] in Fragmente zerlegt. Abbildung 28 zeigt die Zerlegung des PPAR α -Agonisten GW590735 **5**.

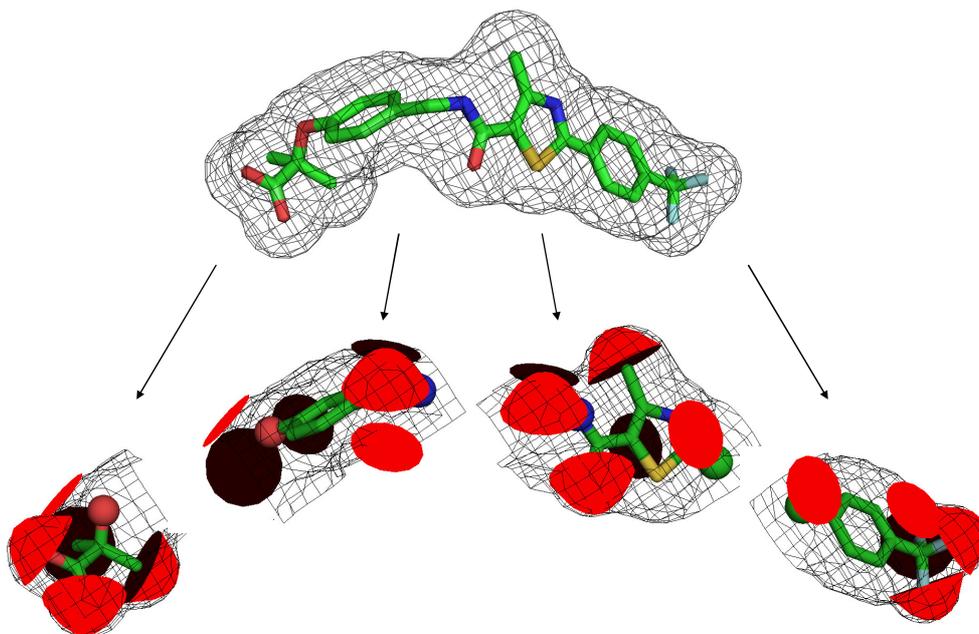


Abbildung 28: Die Zerlegung von GW590735 **5** [Sierra2007] in Fragmente entsprechend den RECAP-Retrosyntheseregeln. Die Kugeln symbolisieren die „attachment-points“, die durch die Spaltung der Bindungen entstanden sind. Die roten Paraboloiden sind den einzelnen Fragmenten zugewiesen und werden für die Überlagerung mit Hilfe der *Shapelets*-Methode verwendet.

Pro Molekül der COBRA6.1 Datenbank wurden im Durchschnitt 4,5 Konformere erzeugt. Nach der Zerlegung wurden 17.934 Fragmente erhalten. Ein Fragment kann, aus unterschiedlichen Mutterstrukturen entnommen, unterschiedliche Oberflächenform aufweisen, durchschnittlich gibt es von jeder Fragmentstruktur 2,3 Konformere.

Als Referenzverbindung diente wieder GW590735 **5**, welches nicht nur herausragende Potenz und Selektivität auf PPAR α aufweist, sondern auch über das etablierte und gut untersuchte Fibrat-Grundgerüst verfügt. In diesem Zusammenhang sollte untersucht werden, ob es möglich ist, durch bioisosteren Ersatz mehrerer Gruppen zu einer neuen Grundstruktur zu gelangen. Bezogen auf GW590735 wurden die Kopfgruppe und der hydrophobe Rest ersetzt, während ein Teil des Moleküls, der aromatische Kern mit dem Amid-Linker konstant bleiben sollte.

Die Überlagerung der Fragmente mit GW590735 **5** ergab 77 Vorschläge für die Kopfgruppe und 90 für den hydrophoben Rest (Listen im Anhang). Die Vorschläge wurden visuell inspiziert. Es wurden einige Kopfgruppen wiedergefunden, für die Wirksamkeit auf PPAR α bekannt ist. Abbildung 29 zeigt Beispiele für solche Bioisostere der 2-Hydroxy-2-methyl-propionsäure, der Kopfgruppe des Fibratgrundgerüsts.

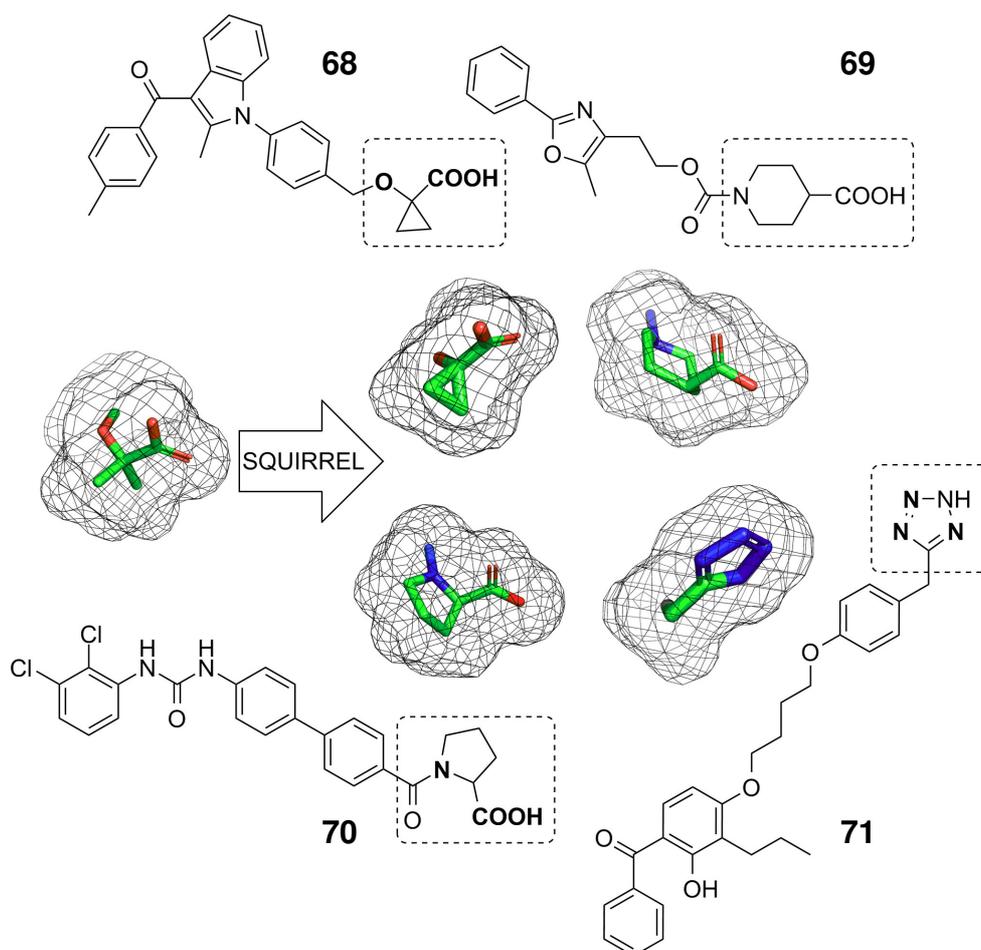


Abbildung 29: Beispiele für die von SQUIRREL_{nov} generierten Vorschläge für den bioisosteren Ersatz des 2-Hydroxy-2-methyl-propionsäure-Restes von Fibraten. Die Moleküle 68-71 wurden für die Aktivität auf PPAR α patentiert [68: JP 2004-202918, 20040709; 69: WO 2006075638; 70: WO 2007016538 71: WO030202269].

Aus den von SQUIRREL_{nov} generierten Vorschlägen wurde Prolin als bioisostere Ersatz der 2-Hydroxy-2-methyl-propionsäure und Naphthalen-2-carboxylat als Ersatz des hydrophoben Restes gewählt (Abbildung 30). Die Synthese der Verbindung **EPI** gelang durch reduktive Aminierung des Prolinesters mit 4-Formylbenzonnitril.

Anschließend folgte die katalytische Hydrierung der Nitrilgruppe, die Amid-Kopplung mit Naphtylchlorid und das Entschützen der Carboxylfunktion durch Verseifung. Bei der Testung von **EP1** im zellbasierten Assay in Konzentrationen 3 μM , 10 μM , 30 μM und 100 μM konnte keine signifikante Aktivierung von PPAR α oder PPAR γ festgestellt werden.

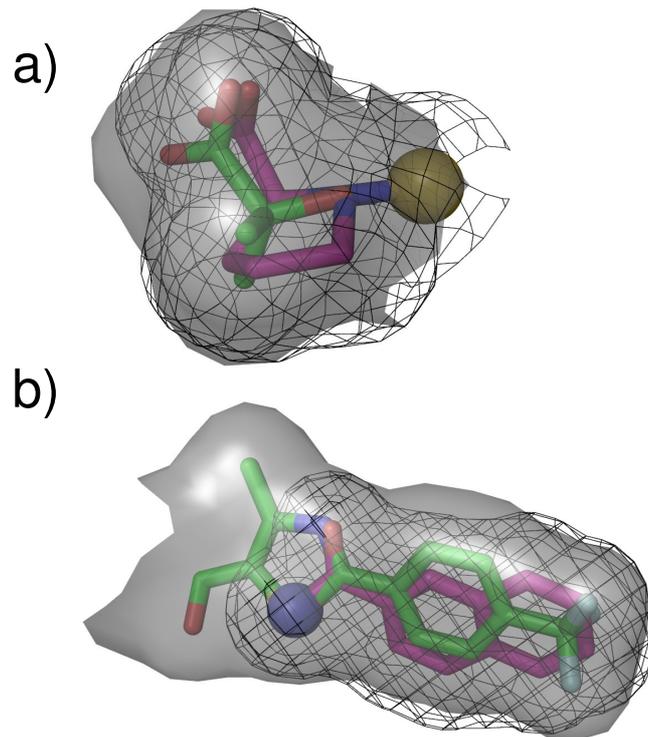


Abbildung 30: Vorschläge für bioisosteren Ersatz für die acide Kopfgruppe (a) und den hydrophoben Rest (b), die für den Design von EP1 verwendet wurden. Graues Volumen und grüne Atome entsprechen GW590735, Netzvolumen und magenta Atome EP1.

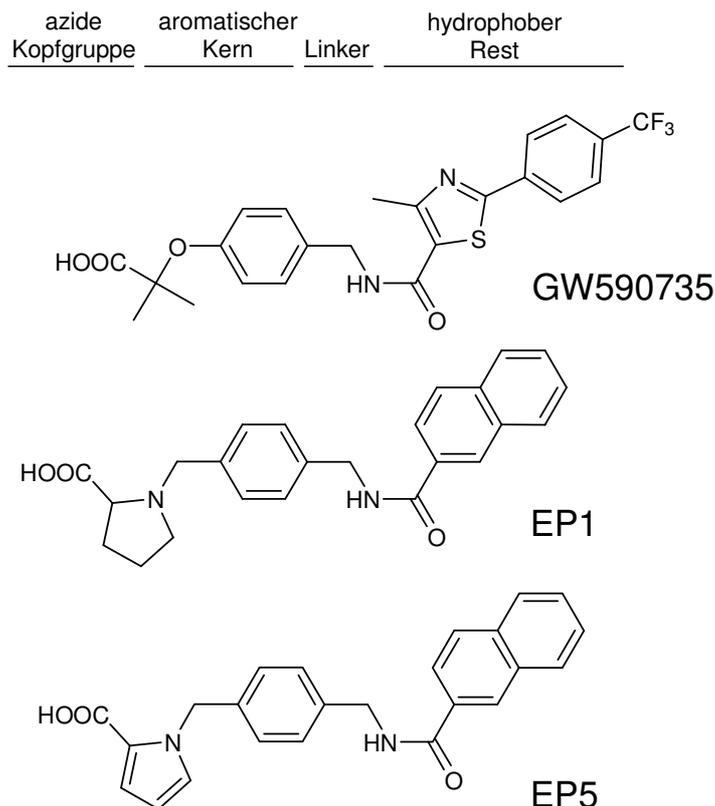


Abbildung 31: Schematische Aufteilung der Moleküle GW590735 5, EP1 21 und EP5 26.

Die Ursache für die fehlende Aktivität von **EP1** liegt wahrscheinlich in den physikochemischen Eigenschaften der Bindetasche der PPARs. Diese zeigt ein ausgeprägtes positives elektrochemisches Potential. Da der Amin-Stickstoff in der Prolin-Kopfgruppe von **EP1** wahrscheinlich unter Assay-Bedingungen protoniert und somit positiv geladen vorliegt, sind die repulsiven Coulomb'schen Wechselwirkungen zu stark, als dass eine stabile Bindung zustande kommen könnte.

Um diese Annahme zu prüfen, wurde die positive Ladung entfernt, indem der Pyrrolidin-Ring von Prolin durch Pyrrol ersetzt wurde. Die Verbindung **EP5** wurde synthetisiert, indem Pyrrol-2-Carbonsäureethylester in einer Phasentransferreaktion [Banwell2006] an 4-(Bromomethyl)benzonnitril gekoppelt wurde. Alle weiteren Schritte entsprechen der Synthese der Verbindung **EP1**. **EP5** wurde im zellbasierten Assay getestet und zeigte einen $EC_{50} = 0,51 \pm 0,12 \mu\text{M}$ für PPAR α , und $EC_{50} = 0,63 \pm 0,03 \mu\text{M}$ für PPAR γ . Dieses positive Ergebnis bestätigt die Hypothese zum Einfluss der positiven Ladung auf die Bindungseigenschaften des Liganden. Der Verlust der Selektivität ist

wahrscheinlich auf die relativ einfache strukturelle Beschaffenheit des hydrophoben Naphtyl-Rests zurückzuführen.

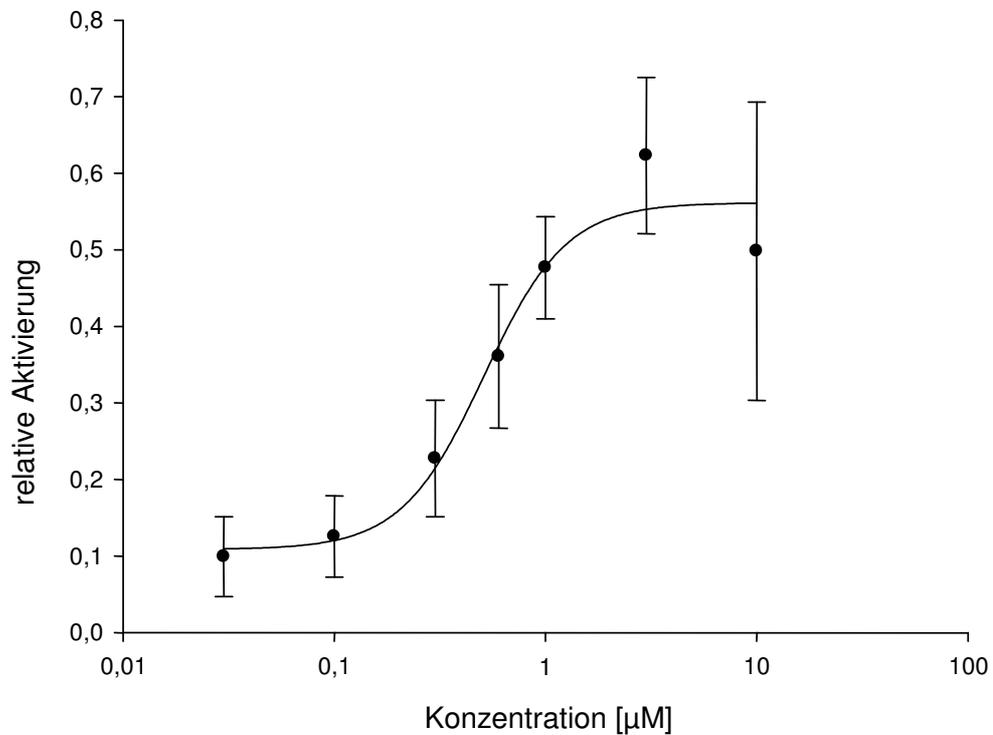


Abbildung 32: Aktivierung von PPAR α durch EP5 in Abhängigkeit von der Konzentration.

Die Synthese von **EP5** zeigte insgesamt, dass es möglich ist, ausgehend von computergenerierten Vorschlägen für den bioisosteren Ersatz potente Leitstrukturen *de novo* zu generieren.

Schlusswort und Ausblick

In dieser Arbeit wurde gezeigt, dass die Kombination der Methoden, die Oberflächenform und Pharmakophor-Informationen berücksichtigen, für das virtual screening und das *de novo* Design von PPAR-Agonisten geeignet sind. Um die Zielsetzung der Arbeit zu erfüllen, wurde die Methode *Shapelets* entwickelt, die eine deterministische Zerlegung molekularer Oberflächen ermöglicht. Es wurde gezeigt, dass die Oberflächeninformation zwar ausreichend ist, um die molekulare Überlagerung zweier Moleküle mit gleichem Bindemodus zu berechnen, jedoch nicht genügt, um aktive Strukturen von inaktiven zu unterscheiden. Unter Berücksichtigung zusätzlicher Information in Form des Pharmakophor-Dichtefelds (LIQUID) konnte die Anreicherung aktiver Strukturen verbessert werden. Durch virtual screening mit der neuen Methode konnten potente und selektive PPAR α -Agonisten identifiziert werden. Anschließend wurde eine neue Leitstruktur mit aktivierender Wirkung auf PPAR α synthetisiert, welche mit Hilfe von bioisosteren Ersatzstrukturen der *de novo* Design Methode SQUIRREL_{novo} entworfen wurde.

Jede der bearbeiteten Ziele hat neue Anregungen zur weiteren Bearbeitung geliefert. Hier werden einige davon aufgelistet:

- Die Methode SQUIRREL hat sich als besonders geeignet für das virtual screening von PPAR-Agonisten erwiesen. Ähnliche Studien wurden bereits von Hawkins *et al.* [Hawkins2007] und McGaughey *et al.* [McGaughey2007] für die einige virtual screening Programme durchgeführt. Eine weitreichende Untersuchung des Verhaltens der Methode SQUIRREL auf weiteren pharmakologischen Targets würde eine bessere Übersicht über die Stärken und Schwächen des Verfahrens liefern.
- Bei SQUIRREL wurde die Pharmakophor-Information als zusätzliches Kriterium für die Auswahl aktiver Substanzen gewählt. Das Verfahren SURFCOMP [Hofbauer2004] nutzt stattdessen das elektrostatische und das lipophile Potential. Auch Exner *et al.* [Exner2002] projizierten diese Eigenschaften auf molekulare Oberflächen. Die Einbindung des

elektrostatischen Potential, des lipophilen Potentials oder weiterer Eigenschaften wie die Polarisierbarkeit könnte die Methode vielseitiger machen.

- SQUIRREL ist für die Überlagerung zweier Liganden geeignet. Eine Erweiterung des Verfahrens, um die Überlagerung des Liganden mit einer geeigneten Repräsentation der Bindetasche könnte Aufschluß über die Bedeutung der Oberflächenkomplementarität bei den Protein-Ligand-Wechselwirkungen geben. Einen ähnlichen Ansatz verfolgten Norel *et al.* [Norel1999], wobei hier kritische Punkte der molekularen Oberflächen für Docking-Experimente verwendeten.
- Die Substanz **EP5** ist zwar ein vergleichsweise potenter PPAR α -Agonist, jedoch wäre eine Optimierung der Aktivität und vor allem der Selektivität gegenüber PPAR γ ein weiteres Ziel. Sierra *et al.* [Sierra2007] haben durch die Variation des hydrophoben Restes die Optimierung der Leitstruktur vorgenommen.

Diese Ziele werden weiterhin in den Arbeitskreisen von Prof. Dr. Schneider und Prof. Dr. Schubert-Zsilavec bearbeitet.

Literaturverzeichnis

A

[Adler2000] Adler, M., Davey, D.D., Phillips, G.B., Kim, S.H., Jancarik, J., Rumennik, G., Light, D.R., Whitlow, M. *Biochemistry* 2000, **39**, 12534-12542.

[An2005] An, J., Totrov, M., Abagyan, R. *Mol. Cell. Proteomics* 2005, **4**, 752-61.

[Arteca1996] Arteca G.A. *Rev. Comput. Chem.* 1996, **9**, 191-253.

B

[Banwell2006] Banwell, M.G., Hamel, E., Hockless, D.C.R., Verdier-Pinard, P., Willis, A.C., Wong, D. J. *Bioorg. Med. Chem.* 2006, **14**, 4627-4638.

[Baxter1998] Baxter, C.A., Murray, C.W., Clark, D.E., Westhead D.R., Eldridge M.D., *Proteins*, 1998, **33**, 367-382.

[Berger2002] Berger, J., Moller, D.E. *Annu. Rev. Med.* 2002, **53**, 409-35.

[Berman2000] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *Nucleic Acids Res.* 2000, **28**, 235-242.

[Blinn1982] Blinn J.F., *ACM Transactions on Graphics*, 1982, **1**, 235-256.

[Böhm1992] Böhm, H-J. *J. Comput. Aid. Mol. Des.* 1992, **6**, 61-78.

[Böhm1996] Böhm, H-J., Klebe, G., Kubinyi, H. *Wirkstoffdesign*, Spektrum Akademischer Verlag: Heidelberg, **1996**.

[Bologa2006] Bologa, C.G., Revankar, C.M., Young, S.M., Edwards, B.S., Arterburn, J.B., Kiselyov, A.S., Parker, M.A., Tkachenko, S.E., Savchuck, N.P., Sklar, L.A., Oprea, T.I., Prossnitz, E.R. *Nature Chem. Biol.* 2006, **2**, 207-212.

[Boström2001] Boström, J, *J. Comput.-Aid. Mol. Des.* 2001, **15**, 1137-1152.

[Boström2007] Boström, J., Berggren, K., Elebring, T. Greasley, P.J., Wilstermann, M., *Bioorg. Med. Chem. Lett.* 2007, **15**, 4077-4084.

[Brint1987] Brint, A. T., Willett, P., *J. Chem. Inf. Comput. Sci.*, 1987, **27**, 152-158.

[Bron1973] Bron, C., Kerbosch, J., *Communications of the ACM*, 1973, **16**, 575-577.

[Brown2001] Brown, P.J., Stuart, L.W., Hurley, K.P., Lewis, M.C., Winegar, D.A., Wilson, J.G., Wilkison, W.O., Ittoop, O.R., Willson, T.M., *Bioorg. Med. Chem. Lett.* 2001, **11**, 1225-1227.

C

[Cosgrove2000] Cosgrove, D.A., Bayada, D.M., Johnson, A.P. *J. Comput. Aided Mol. Des.* 2000, **14**, 573-591.

D

[Davies1990] Davies 2nd., J.F., Delcamp, T.J., Prendergast, N.J., Ashford, V.A., Freisheim, J.H., Kraut, J. *Biochemistry* 1990, **29**, 9467-9479.

[Derksen2006] Derksen, S., Rau, O., Schneider, P., Schubert-Zsilavec, M., Schneider, G., *ChemMedChem*, 2006, **1**, 1346-1350.

[Desvergne1999] Desvergne, B., Wahli, W. *Endocr. Rev.* 1999, **20**, 649–688.

[Desvergne2004] Desvergne, B., Michalik, L., Wahli, W. *Mol. Endocrinol.* 2004, **18**, 1321–1332.

[deWet1985] de Wet, J.R., Wood, K.V., Helinski, D.R., DeLuca M. *Proc Natl Acad Sci U S A.* 1985, **82**, 7870-7873.

[deWet1987] de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R., Subramani, S. *Mol Cell Biol.* 1987, **7**, 725-737.

[Duncan1993] Duncan, B.S., Olson, A.J., *Biopolymers*, 1993, **33**, 231-238.

[Dunn2002] Dunn, F., Parberry, I., *3D Math Primer for Graphics and Game Development*; Wordware Publishing: Plano; **2002**, 331.

[Duval2007] Duval, C., Müller, M. Kersten, S. *Biochim. Biophys Acta* 2007, **1771**, 961–971.

E

[Eldridge1997] Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini G.V., Mee R.P., *J. Comput.-Aided Mol. Des.*, 1997, **11**, 425-445.

[Etgen2003] Etgen, G.J., Mantlo, N. *Curr. Top. Med. Chem.* 2003, **3**, 1649-1661.

[Exner2002] Exner, T.E., Keil, M., Brickmann, J., *J. Comput. Chem.*, 2002, **23**, 1176-1187.

F

[Fechner2003] Fechner, U., Franke, L., Renner, S., Schneider, P., Schneider, G. *J. Comput. Aided Mol. Des.* 2003, **17**, 687-698.

[Fechner2007] Fecner, U., Schneider, G. *J. Chem. Inf. Model.* 2007, **47**, 656-667

[Ferguson1989] Ferguson, D.M., Raber, D.J., *J. Amer. Chem. Soc.* 1989, **111**, 4371-4378.

[Fleming1929] Flemming, A. *Br. J. Exp. Pathol.* 1929, **10**, 226–236.

[Forman1997] Forman, B.M., Chen, J., Evans, R.M. *Proc. Natl. Acad. Sci. U.S.A.* 1997, **94**, 4312-4317.

G

[Gampe2000] Gampe Jr, R.T., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.A., Willson, T.M., Xu, H.E., *Mol Cell.* 2000, **5**, 545-555.

[Gardiner2000] Gardiner, E.J., Willet, P., Artymiuk, P.J., *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 273-279.

[Gasteiger2003] Gasteiger, J., Engel T., *Chemoinformatics A Textbook*, Wiley-VCH, Weinheim, **2003**, 31ff.

[Gelman2007] Gelman, L., Feige, J.N., Desvergne, B. *Biochim. Biophys Acta* 2007, **1771**, 1094–1107.

[Gordon1994] Gordon, E.M., Barrett, R.W., Dower, W.J., Fodor, S.P.A., Gallop, M.A. *J. Med. Chem.* 1994, **37**, 1385-1401.

[Gerhards1979] Gerhards L., Lindenberg W., *Computing*, 1979, **21**, 295-322.

[Gluzman1991] Gluzman, Y., *Cell* 1981, **23**, 175-182.

[Grabowski2007] Grabowski, K., Schneider, G., *Curr. Chem. Biol.* 2007, **1**, 115-127.

[Grant1995] Grant, J.A., Pickup, B.T., *J. Phys. Chem.*, 1995, **99**, 3503-3510.

[Grant1996] Grant, J.A., Gallardo, M.A., Pickup, B.T., *J. Comp. Chem.* 1996, **17**, 1653-1666.

[Greene1994] Greene, J., Kahn, S., Savoj, H., Sprague, P., Teig, S. *J. Chem. Inf. Comput. Sci.* 1994, **34**, 1297-1308.

[Grindley1993] Grindley, H.M., Artymiuk, P.J., Rice, D.W., Willett, P., *J. Mol. Biol.* 1993, **229**, 707-721.

[Grundy1987] Grundy, S.M., Vega, G.L. *Am. J. Med.* 1987, **83**, 9-20

[Guertin2002] Guertin, K.R., Gardner, C.J., Klein, S.I., Zulli, A.L., Czekaj, M., Gong, Y., Spada, A.P., Cheney, D.L., Maignan, S., Guilloteau, J.P., Brown, K.D., Colussi, D.J., Chu, V., Heran, C.L., Morgan, S.R., Bentley, R.G., Dunwiddie, C.T., Leadley, R.J., Pauls, H.W. *Bioorg. Med. Chem. Lett.* 2002, **12**, 1671-1674.

[Guido2008] Guido, R.V., Oliva, G., Andricopulo, A.D. *Curr. Med. Chem.* 2008, **15**, 37-46.

H

- [Hahn1997] Hahn, M. *J. Chem. Inf. Comput. Sci.* 1997, **37**, 80–86.
- [Halgren1996] Halgren, T.A., *J. Comput. Chem.*, 1996, **17**, 490-519, 520-552, 553-586, 587-615, 616-641.
- [Hawkins2007] Hawkins, P.C.D., Skillman, A.G., Nicholls, A. *J. Med. Chem.* 2007, **50**, 74-82.
- [Heiden1993] Heiden, W., Moeckel, G., Brickmann, J. *J. Comput.-Aided Mol. Des.* 1993, **7**, 503 514.
- [Henke1998] Henke, B.R., Blanchard, S.G., Brackeen, M.F., Brown, K.K., Cobb, J.E., Collins, J.L., Harrington, W.W., Jr., Hashim, M.A., Hull-Ryde, E.A., Kaldor, I., Klierer, S.A., Lake, D.H., Leesnitzer, L.M., Lehmann, J.M., Lenhard, J.M., Orband-Miller, L.A., Miller, J.F., Mook, R.A., Noble, S.A., Oliver, W., Parks, D.J., Plunket, K.D., Szewczyk, J.R., Willson, T.M.N. *J. Med. Chem.* 1998, **41**, 5020 5036.
- [Hofbauer2004] Hofbauer, C., Lohninger, H., Aszódi, A., *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 837-847.
- [Holden1987a] Holden, H.M., Matthews, B.W. *J. Biol. Chem.* 1987, **263**, 3256-3260.
- [Holden1987b] Holden, H.M., Tronrud, D.E., Monzingo, A.F., Weaver, L.H., Matthews, B.W. *Biochemistry* 1987, **26**, 8542-8553.
- [Holland1994] Holland, D.R., Barclay, P.L., Danilewicz, J.C., Matthews, B.W., James, K. *Biochemistry* 1994, **33**, 51-56.
- [Holmes1981] Holmes, M.A., Matthews, B.W. *Biochemistry* 1981, **20**, 6912-6920.
- [Hopcroft2001] Hopcroft, J.E., Motwani, R., Ullman J.D., *Introduction to Automata Theory, Languages, and Computation*, Addison Wesley: Boston; **2001**, 179.
- [Hristozov2007] Hristozov, D., Oprea, T.I., Gasteiger, J. *J. Chem. Inf. Model.* 2007, **47**, 2044-2062

I

J

- [Johnson1989] Johnson, M., Lajiness, M., Maggiora, G. *Prog. Clin. Biol. Res.* 1989, **291**, 167-171.
- [Jones1995] Jones, G., Willett, P., Glen, R. C., *J. Mol. Biol.*, 1995, **245**, 43-53.
- [Jones1997] Jones, G., Willett, P., Glen, R. C., Leach A. R., Taylor, R., *J. Mol. Biol.*, 1997 **267**, 727-748.

K

- [Kabsch1976] Kabsch, W., *Acta Cryst.*, 1976, **A32**, 922-923.
- [Keserü2006] Keserü, G.M., Makara, G.M., *Drug Discov. Today* 2006, **11**, 741-748.
- [Kersten2000] Kersten, S., Desvergne, B., Wahli, W., *Nature* 2000, **405**, 421-424.
- [Klebe2006] Klebe, G. *Drug Discov. Today*, 2006, **11**, 580-594.
- [Koch2001] Koch I. *Theoretical Computer Science*, 2001, **250**, 1-30.
- [Kohonen1982] Kohonen, T. *Biol. Cybern.* 1982, **43**, 59-69.
- [Kolossvary1996] Kolossváry, I., Guida, W.C. *J. Am. Chem. Soc.* 1996, **118**, 5011-5019.
- [Kuhl1984] Kuhl, F.S., Crippen, G.M., Friesen, D.K., *J. Comp. Chem.* 1984, **5**, 24-34.
- [Kuhn2006] Kuhn, B., Hilpert, H., Benz, J., Binggeli, A., Grether, U., Humm, R., Märki, H.P., Meyer, M., Mohr, P. *Bioorg Med Chem Lett.* 2006, **16**, 4016-4020.
- [Kurumbail1996] Kurumbail, R.G., Stevens, A.M., Gierse, J.K., McDonald, J.J., Stegeman, R.A., Pak, J.Y., Gildehaus, D., Miyashiro, J.M., Penning, T.D., Seibert, K., Isakson, P.C., Stallings, W.C. *Nature* 1996, **384**, 644-648.

L

- [Lehmann1995] Lehmann, J.M., Moore, L.B., Simth-Oliver, T.A., Wilkison, W.O. Willson, T.M., Kliewer, S.A. *J. Biol. Chem.* 1995, **270**, 12953-12956.
- [Lewell1998] Lewell, X.Q., Judd, D.B., Watson, S.P., Hann, M.M., *J. Chem. Inf. Sci.*, 1998, **38**, 511-522.
- [Li2000] Li, R., Sirawaraporn, R., Chitnumsub, P., Sirawaraporn, W., Wooden, J., Athappilly, F., Turley, S., Hol, W.G. *J. Mol. Biol.* 2000, **295**, 307-323.
- [Li2007] Li, X.H., Zou, H.J., Wu, A.H., Ye, Y.L., Shen, J.H. *Acta Pharmacol. Sin.* 2007, **28**, 2040-52.
- [Lima2005] Lima, L.M., Barreiro, E.J. *Curr. Med. Chem.* 2005, **12**, 23-49
- [Lorensen1987] Lorensen, W.E., Cline, H.E., *Computer Graphics*, 1987, **21**, 163-169.
- [Lorenz1991] Lorenz, W.W. McCann, R.O., Longiaru, M., Cormier, M.J. *Proc. Natl. Acad. Sci. USA* 1991, **88**, 4438-4442.
- [Lu 2006] Lu I.L., Huang, C.F., Peng, Y.H., Lin, Y.T., Hsieh, H.P., Chen, C.T., Lien, T.W., Lee, H.J., Mahindroo, N., Prakash, E., Yueh, A., Chen, H.Y., Goparaju, C.M., Chen, X., Liao, C.C., Chao, Y.S., Hsu, J.T., Wu, S.Y. *J. Med. Chem.* 2006, **49**, 2703-2712.

M

[Maignan2000] Maignan, S., Guilloteau, J.P., Pouzieux, S., Choi-Sledeski, Y.M., Becker, M.R., Klein, S.I., Ewing, W.R., Pauls, H.W., Spada, A.P., Mikol, V. *J.Med.Chem.* 2000, **43**, 3226-3232.

[Markt2007] Markt, P., Schuster, D., Kirchmair, J., Laggner, C., Langer, T. *J Comput Aided Mol Des.* 2007, **21**, 575-590.

[Mason1999] Mason, J.S., Morize, I., Menard, P.R., Cheney, D.L., Hulme, C., Labaudiniere, R.F. *J. Med. Chem.* 1999, **42**, 3251-64.

[Mavridis2007] Mavridis, L., Hudson, B. D., Ritchie, D. W. *J. Chem. Inf. Model.* 2007, **47**, 1787-1796.

[Matthews1977] Matthews, J.C., Hori, K., Cormier, M.J. *Biochemistry* 1977, **16**, 85-91.

[McGaughey2007] McGaughey, G.B., Sheridan, R.P., Bayly, C.I., Culberson, J.C.; Kreatsoulas, C., Lindsley, S., Maiorov, V., Truchon, J.-F., Cornell, W.D. *J. Chem. Inf. Model.* 2007, **47**, 1504-1519.

[Messmer1995] Messmer, B., Dissertation, **1995**, Universität Bern, Bern, 11ff.

[Mezey1993] Mezey, P.G.; *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH Publishers New York, **1993**.

[Miller1999] Miller, M.D., Sheridan, R.P., Kearsley, S.L. *J. Med. Chem.* 1999, **42**, 1505-1514.

[Miyachi2004] Miyachi, H. *Expert Opin, Ther. Patents* 2004, **14**, 607-618.

[Monzingo1984] Monzingo, A.F., Matthews, B.W. *Biochemistry* 1984, **23**, 5724-5729.

N

[Noeske2006] Noeske, T., Sasse, B.C., Stark, H., Parsons, C.G., Weil, T., Schneider, G., *ChemMedChem*, 2006, **1**, 1066-1068.

[Norel1999] Norel, R., Wolfson, H.J., Nussinov, R. *Comb. Chem. High. Throughput. Screen.* 1999, **2**, 223-237.

O

P

[Proschak2007] Proschak E., Wegner J.K., Schüller A., Schneider G., Fechner U., *J. Chem. Inf. Model.*, 2007, **47**, 295-301.

[Proschak2008] Proschak, E., Tanrikulu, Y., Schneider, G. In: *Chemoinformatics: An Approach to Virtual Screening* (A. Varnek, A. Tropsha; Eds). RSC Publishing, Cambridge **2008**.

[Pirard2003] Pirard, B. *J. Comput. Aid. Mol. Des.* 2003, **17**, 785-796.

Q

R

[Rau2006] Rau, O., Wurglics, M., Paulke, A., Zitzkowski, J., Meindl, N., Bock, A., Dingermann, T., Abdel-Tawab, M., Schubert-Zsilavecz, M., *Planta Med*, 2006, **72**, 881-887.

[Rau2008] Rau, O., Zettl, H., Popescu, L., Steinhilber, D., Schubert-Zsilavecz, M. *ChemMedChem* 2008, **3**, 206-21.

[Renner2005] Renner, S., Noeske, T., Parsons, C. P., Weil, T., Schneider, G., *ChemBioChem*, 2005, **6**, 620-625.

[Roger-Evans2004] Rogers-Evans, M., Alanine, A.I., Bleicher, K.H., Kube, D., Schneider, G., *QSAR Comb. Sci.*, 2004, **23**, 426-430.

[Rubenstrunk2007] Rubenstrunk, A., Hanf, R. Hum, D.W. Fruchart, J-C., Staels, B. *Biochim. Biophys Acta* 2007, **1771**, 1065-1081.

[Rücker2006] Rücker C, Scarsi M, Meringer M. *Bioorg Med Chem.* 2006 Aug 1;14(15):5178-95.

[Rush2005] Rush, T.S., Grant, J.A., Mosyak, L., Nicholls, A., *J.Med.Chem.* 2005, **48**, 1489-1495

S

[Salam2008] Salam, N.K., Huang, T.H., Kota, B.P., Kim, M.S., Li, Y., Hibbs, D.E. *Chem. Biol. Drug Des.* 2008, **71**, 57-70.

[Samudrala1998] Samudrala, R., Moulton, J., *J. Mol. Biol.*, 1998, **279**, 287-302.

[Seifert2007] Seifert, M.H., Lang M. *Mini Rev. Med. Chem.* 2008, **8**, 63-72.

[Scarsi2007] Scarsi, M., Podvinec, M., Roth, A., Hug, H., Kersten, S., Albrecht, H., Schwede, T., Meyer, U.A., Rücker C. *Mol. Pharmacol.* 2007, **71**, 398-406.

[Schneider1998] G. Schneider, G., Wrede, P. *Prog. Biophys. Mol. Biol.* 1998, **70**, 175-222.

- [Schneider1999] Schneider, G., Neidhart, W., Giller, T., Schmid, G., *Angew. Chem. Int. Ed.*, 1999, **38**, 2894-2896.
- [Schneider2002] Schneider, G., Böhm, H-J. *Drug Discov. Today* 2002, **7**, 64-70.
- [Schneider2003] Schneider, P.; Schneider, G. *QSAR Comb. Sci.*, 2003, **22**, 713-718.
- [Schneider2004] Schneider, G., Schneider, P. In: *Chemogenomics in Drug Discovery* (H. Kubinyi, G. Müller, Eds.), Wiley-VCH, Weinheim, **2004**, 341-376.
- [Schneider2005] Schneider, G., Fechner, U. *Nature Rev. Drug Discov.* 2005, **4**, 649-663.
- [Schöning2001] Schöning, U., *Algorithmik*, Spektrum Akademischer Verlag GmbH, Heidelberg, **2001**, 231ff.
- [Schroeder1992] Schroeder, W.J., Zarge, J.A., Lorensen, W.E., *Computer Graphics*, 1992, **26**, 65- 70.
- [Shannon1998] Shannon, C.E. *Proc. IEEE* 1998, **86**, 447-457.
- [Shearer2007] Shearer, B.G., Billin, A.N. *Biochim. Biophys Acta* 2007, **1771**, 1082-1093.
- [Sher1993] Sher, T., Yi, H.F., McBride, O.W., Gonzalez, F.J. *Biochemistry* 1993, **32**, 5598-5604.
- [Sierra2007] Sierra, M.L., Beneton, V., Boullay, A.B., Boyer, T., Brewster, A.G., Donche, F., Forest, M.C., Fouchet, M.H., Gellibert, F.J., Grillot, D.A., Lambert, M.H., Laroze, A., Le Grumelec, C., Linget, J.M., Montana, V.G., Nguyen, V.L., Nicodème, E., Patel, V., Penfornis, A., Pineau, O., Pohin, D., Potvain, F., Poulain, G., Ruault, C.B., Saunders, M., Toum, J., Xu, H.E., Xu, R.X., Pianetti, P.M. *J. Med. Chem.* 2007, **50**, 685-695.
- [Singh2005] Singh, J.P., Kauffman, R., Bensch, W., Wang, G., McClelland, P., Bean, J., Montrose, C., Mantlo, N., Wagle, A. *Mol. Pharmacol.* 2005, **68**, 763-768.
- [Sousa2006] Sousa, S.F., Fernandes, P.A., Ramos, M.J. *Proteins* 2006, **65**, 15-26.
- [Steinbeck2003] Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E., Willighagen, E.L., *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 493-500.
- [Steinbeck2006] Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R., Willighagen, E.L., *Curr. Pharm. Des.*, 2006, **12**, 2111-2120.
- [Steinhilber2005] Steinhilber, D., Schubert-Zsilavec, M., Roth, H.J. *Medizinische Chemie*, Deutscher Apotheker Verlag, Stuttgart, **2005**.

[Svoboda1966] Svoboda, D.J., Azarnoff, D.L. *J. Cell. Biol.* 1966, **30**, 442-450.

[Széles2007] Széles, L., Töröcsik, D., Nagy, L. *Biochim. Biophys. Acta* 2007, **1771**, 1014-1030.

T

[Tanrikulu2007] Tanrikulu, Y., Nietert, M., Scheffer, U., Proschak, E., Grabowski, K., Schneider, P., Weidlich, M., Karas, M., Göbel, M., Schneider, G. *ChemBioChem* 2007, **8**, 1932-1936

[Tronrud1986] Tronrud, D.E., Monzingo, A.F., Matthews, B.W. *Eur. J. Biochem.* 1986, **157**, 261-268.

[Tronrud1987] Tronrud, D.E., Holden, H.M., Matthews, B.W. *Science* 1987, **235**, 571-574.

U

[Ulrich2006] Ulrich S., Loitsch, S.M., Rau, O., von Knethen, A., Brune, B., Schubert-Zsilavecz, M., Stein, J.M., *Cancer Res*, 2006, **66**, 7348-7354.

W

[Wilchek2006] Wilchek, M., Bayer, E.A., Livnah, O. *Immunol. Lett.* 2006, **103**, 27-32.

[Willett1998] Willett, P., Barnard, J.M., Downs, G.M. *J. Chem. Inf. Comput. Sci.* 1998, **38**, 983-996.

[Willson2000] Willson, T., Brown, P.J., Sternbach, D.D., Henke, B.R. *J. Med. Chem.* 2000, **43**, 527-550.

V

[Verdonk2003] Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray C.W., Taylor, R.D., *Proteins*, 2003, **52**, 609-623.

X

[Xu1999] Xu, H.E., Lambert, M.H., Montana, V.G., Parks, D.J., Blanchard, S.G., Brown, P.J., Sternbach, D.D., Lehmann, J.M., Wisely, G.B., Willson, T.M., Kliewer, S.A., Milburn, M.V. *Mol. Cell* 1999, **3**, 397-403.

[Xu2001] Xu, H.E., Lambert, M.H., Montana, V.G., Plunket, K.D., Moore, L.B., Collins, J.L., Oplinger, J.A., Kliewer, S.A., Gampe Jr., R.T., McKee, D.D., Moore, J.T., Willson, T.M. *Proc. Natl. Acad. Sci. U S A* 2001, **98**, 13919-13924.

[Xu2002a] Xu, H.E., Stanley, T.B., Montana, V.G., Lambert, M.H., Shearer, B.G., Cobb, J.E., McKee, D.D., Galardi, C.M., Plunket, K.D., Nolte, R.T., Parks, D.J., Moore, J.T., Kliewer, S.A., Willson, T.M., Stimmel, J.B. *Nature* 2002, 415, 813-817.

[Xu2002b] Xu, H., Agrafiotis, D.K., *Curr. Top. Med. Chem.* 2002, **2**, 1305-1320.

Y

Z

[Zachmann1992] Zachmann, C.D., Heiden, W., Schlenkrich, M., Bickmann, J., *J. Comput. Chem.*, 1992, **13**, 76-84.

[Zoete2007] Vincent Zoete, V., Grosdidier, A., Michielin, O. *Biochim. Biophys Acta* 2007, **1771**, 915-925.

Danksagung

Ich danke allen Personen, die es ermöglicht haben, dass diese Arbeit in dieser Form angefertigt wurde, insbesondere

Prof. Dr. Gisbert Schneider für die exzellente Betreuung dieser Arbeit,

Prof. Dr. Manfred Schubert-Zsilavecz für die Bereitstellung der Ausrüstung, Mittel und Räumlichkeiten zur Testung der Substanzen,

Prof. Dr. Holger Stark für die Bereitstellung der Geräte, Mittel und Räumlichkeiten für die Synthese,

Dr. Oliver Rau für die Einweisung und Betreuung bei der Durchführung des Assays und der Auswertung,

Frau Kerstin Sander für die Betreuung bei der Synthese,

Herrn Yusuf Tanrikulu für die Bereitstellung der LIQUID-Software und die Hilfe bei der Durchführung der *in silico* Studien,

Herrn Heiko Zettl für die Hilfe beim Assay und der Synthese,

Herrn Matthias Rupp für die Erklärungen zum theoretischen Teil der Arbeit,

Herrn Martin Weisel für die Berechnung der Bindetaschen und die Hilfe bei der Durchführung der *in silico* Studien,

Frau Swetlana Derksen für die Einweisung in die Docking-Experimente,

Herrn Norbert Dichter für die Einweisung in die Arbeit mit dem Rechnercluster,

Dr. Petra Schneider für die Bereitstellung der COBRA6.1 Datenbank,

Herrn Jan Kriegl und der Firma Boeringer Ingelheim für die Möglichkeit, die Software OMEGA und ROCS zu nutzen,

der Firma Chemical Computing Group für die Bereitstellung der MOE-Software-Lizenzen,

Ferner danke ich allen Mitgliedern der Arbeitskreise Schneider, Schubert-Zsilavecz und Stark, die das Anfertigen dieser Arbeit durch ihre hilfreiche Diskussionen und Tips um einiges erleichtert und durch wunderbares Arbeitsklima verschönert haben:

Andreas Schüller, Alexander Klenner, Bettina Hofmann, Heiko Zettl, Jan Hiss, Kathleen Isensee, Kerstin Sander, Kristina Grabowksi, Lutz Franke, Martin Weisel, Matthias Rupp, Markus Hartenfeller, Manuel Nietert, Martin Löwer, Matthias Wirth,

Michaela Dittrich, Michael Meissner, Michael Schmuker, Oliver Rau, Ramona Steri, Sarah Keppner, Swetlana Derksen, Tim Kottke, Uli Fechner, Volker Hähnke, Yusuf Tanrikulu und allen die ich vergessen habe.

Ich danke meinen Großeltern, meinen Eltern und meinem Bruder für die familiäre Unterstützung und Interesse an meine Arbeit.

Schließlich danke ich meinem Schatz Anna Swatski für die Geduld, den Rückhalt und den Glauben an meine Arbeit, auch an Tagen oder Monaten, in denen nichts geklappt hat.

Lebenslauf

Persönliche Daten

Name	Proschak Ewgenij
Geburtsdatum	29.11.1980
Geburtsort	St. Petersburg (Rußland)
Familienstaus	ledig

Schulische Ausbildung

1986-1992	Gesamtschule in St. Petersburg
1992-1993	Gymnasium St. Goarshausen
1993-2001	Hilda-Gymnasium Koblenz
	Abschluß: Abitur

Universitäre Ausbildung

2001-2006	Goethe-Universität Frankfurt Diplomarbeit angefertigt im AK von Prof. Dr. G. Schneider Abschluß: Diplom-Bioinformatiker
seit 2006	Promotion im AK von Prof. Dr. G. Schneider

Beruflicher Werdegang

1997-2001	Trainer bei CTG Koblenz
2002-2004	Nachhilfelehrer für Mathematik bei Schülerhilfe, Frankfurt
2004-2006	Programmierer bei Socratec R&D, Oberursel
seit 2006	wissenschaftlicher Mitarbeiter der Goethe Universität Frankfurt

eigene Publikationsliste

Im Rahmen dieser Arbeit sind folgende Publikationen entstanden:

Eingereichte Publikationen:

18. Keppner, S., Proschak, E., Schneider, G., Spänkuch, B. (2008) Structure-based virtual screening with a model of Polo-like kinase 1 (Plk1) in the inactive conformation produces an inhibitor of kinase activity and cell proliferation. *EMBO J.*, submitted

17. Proschak, E., Sander, K., Zettl, H., Tanrikulu, Y., Weisel, M., Kriegl, J.M., Schneider, P., Rau, O., Stark, H., Schubert-Zsilavecz, M., Schneider, G. (2008) From molecular shape to potent bioactive agents: Bioisosteric replacement and fragment-based de novo design, *J. Med. Chem.* submitted.

16. Reisen, F., Schneider, G., Proschak, E. (2008) Reaction-MQL: line notation for representation of chemical reactions. *J. Chem. Inf. Model.* submitted

15. Schneider, G., Hartenfeller, M., Reutlinger, M., Tanrikulu, Y., Proschak, E., Schneider, P. (2008) Voyages to the (un)known: Adaptive design of druglike compounds, *Trends in Biotech.*, submitted.

14. Tanrikulu, Y., Rau, O., Schwarz, O., Proschak, E., Siems, K., Müller-Kuhrt, L., Schubert-Zsilavecz, M., Schneider, G., (2008) Structure-based Pharmacophore Screening for Natural Product-derived PPARgamma Agonists, *ChemBioChem*, submitted.

Bereits erschienene Publikationen. Publikationen die mit gelb markiert sind, befinden sich im Anhang.

13. Weisel, M., Proschak, E., Kriegl, J., Schneider G. Form Follows Function: Shape Analysis of Protein Cavities for Receptor-based Drug Design, *Proteomics*, accepted

12. Hähnke, V., Hofmann, B., Grgat, T., Proschak, E., Steinhilber, D., Schneider, G., (2008) PhAST: Pharmacophore Alignment Search Tool. *J. Comput. Chem.* accepted.

11. Hofmann B, Franke L, Proschak E, Tanrikulu Y, Schneider P, Steinhilber D, Schneider G.(2008) Scaffold-hopping cascade yields potent inhibitors of 5-lipoxygenase. *ChemMedChem.* accepted.

10. Grabowski, K., Proschak, E., Baringhaus, K.-H., Rau, O., Schubert-Zsilavecz, M., Schneider, G. (2008) Bioisosteric Replacement of Molecular Scaffolds: From Natural Products to Synthetic Compounds. *Nat. Prod. Comm.*, **3**, 1355-1360.

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

8. Proschak, E., Tanrikulu, Y. and Schneider, G. (2008) Fragment-based de novo design of druglike compound libraries. In: *Chemoinformatics: An Approach to Virtual Screening* (A.Varnek, A. Tropsha; Eds). RSC Publishing, Cambridge, accepted.

7. Rupp M, Proschak E, Schneider G. (2007) Kernel approach to molecular similarity based on iterative graph similarity. *J. Chem. Inf. Model.* **47**, 2280-2286.

6. Tanrikulu Y, Nietert M, Scheffer U, Proschak E, Grabowski K, Schneider P, Weidlich M, Karas M, Göbel M, Schneider G.(2007) Scaffold Hopping by "Fuzzy" Pharmacophores and its Application to RNA Targets. *ChemBioChem* **8**, 1232-1236

5. Proschak, E., Rupp, M., Derksen, S. and Schneider, G. (2007) Shapelets: possibilities and limitations of shape-based virtual screening. *J. Comp. Chem.*, **29**, 108-114.

4. Schneider, G., Proschak, E., Grabowski, K., Schneider, P. and Tanrikulu, Y. (2007) More than colored blobs – “Fuzzy” descriptions of potential pharmacophore points. *Proceedings of the Beilstein Workshop 2006* (M. Hicks, C. Kettner, Eds), Logos-Verlag, Berlin.

3. Schmuker, M., Schwarte, F., Brück, A., Proschak, E., Tanrikulu, Y., Givehchi, A., Scheiffel, K. and Schneider, G. (2007) SOMMER: Self-organizing maps for education and research. *J. Mol. Model.* **13**, 225-228.

2. Weisel, M., Proschak, E. and Schneider, G. (2007) PocketPicker: Analysis of ligand bindingsites with shape descriptors. *Chem. Cent. J.* **1**, 7.

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

Anhang

- Einstellungen der Docking-Software GOLD
- Liste der von SQUIRRELnovο vorgeschlagenen aciden Kopfgruppen
- Liste der von SQUIRRELnovο vorgeschlagenen hydrophoben Reste
- Ausgewählte Publikationen

Parametereinstellungen des Programms GOLD

AUTOMATIC SETTINGS

autoscale = 2.00

POPULATION

popsiz = auto

select_pressure = auto

n_islands = auto

maxops = auto

niche_siz = auto

GENETIC OPERATORS

pt_crosswt = auto

allele_mutatewt = auto

migratewt = auto

FLOOD FILL

radius = 10

origin = 0 0 0

do_cavity = 1

floodfill_atom_no = 0

cavity_file = 735.mol2

floodfill_center = cavity_from_ligand

FLAGS

display = 0

internal_ligand_h_bonds = 0

n_ligand_bumps = 0

flip_free_corners = 0

flip_amide_bonds = 0

flip_planar_n = 1 flip_ring_NRR flip_ring_NHR

flip_pyramidal_n = 0

rotate_carboxylic_oh = flip

use_tordist = 1

TERMINATION

early_termination = 0

n_top_solutions = 3

rms_tolerance = 1.5

CONSTRAINTS

force_constraints = 0

COVALENT BONDING

covalent = 0

FITNESS FUNCTION SETTINGS

initial_virtual_pt_match_max = 3.0

relative_ligand_energy = 0

gold_fitfunc_path chemscore

score_param_file = DEFAULT

