



DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

Similarity search – Fingerprintanalyse und Konformerdatenbankanalyse

verfasst von / submitted by

Lisa-Maria Katzbeck

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Pharmazie (Mag.pharm.)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 449

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Diplomstudium Pharmazie

Betreut von / Supervisor:

Univ. Prof. mag. Dr. Gerhard Ecker

Mitbetreut von / Co-Supervisor:

Mag. Dr. Lars Richter

ABSTRACT

Das erste Ziel dieser Diplomarbeit war es, fünf Fingerprints (MACCS, GpiDAPH3, SHED, TGT, TGD) zu testen und ihre Anwendbarkeit und Leistung untereinander zu vergleichen und zu bewerten. Die Fragestellung war, ob ein bestimmter Fingerprint besonders gute Ergebnisse bei einer bestimmten Ähnlichkeitssuche liefert oder ob alle Fingerprints gleich gut geeignet sind.

Im Rahmen dieser Ähnlichkeitssuche wurden für insgesamt 30 Arzneistoffe jeweils 5 Fingerprints (MACCS, GpiDAPH3, SHED, TGT, TGD) pro Arzneistoff berechnet und diese mit den Fingerprints der Substanzen in ChEMBL - Datenbank 23 verglichen. Je nachdem wie ähnlich im Bezug auf seiner Fingerprint ein Arzneistoff einer Substanz in der ChEMBL - Datenbank war, desto vorangereiheter wurde er mit Hilfe des Tanimoto - Koeffizienten in einer Tabelle sortiert. Für die Analysen wurden dann die erstbesten dieser 1000, 10.000, 100.000 Ergebnisse herangezogen.

Die Auswahl der 30 Arzneistoffe kam so zustande, dass die 5 Arzneistoffe die bei einem bestimmten Target (ein spezifisches und ein unspezifisches Enzym, Ionenkanal und ein Transporter) am aktivsten getestet wurden, aus einer Liste ausgesucht wurden. Für diese Arzneistoffe derselben Liste wurde eine gewisse Ähnlichkeit untereinander angenommen, da sie am selben Target binden. Die Anzahl der Arzneistoffe dieser Liste, die dann auch in den bestgereihten Ergebnissen der Ähnlichkeitssuche mittels Fingerprints vorkommen, wurde für die Auswertung und Bewertung der einzelnen Fingerprints verwendet. Je größer die Anzahl, desto besser schnitt ein Fingerprint in der Bewertung ab. Das Ergebnis war, dass GpiDAPH3 am Besten abschnitt. Es stellte sich heraus, dass auch alle Fingerprints gemeinsam nur einen kleinen Teil der Arzneistoffe aus der Liste eines Targets finden konnten. Es konnte kein Zusammenhang der verschiedenen Targetklassen mit den Fingerprintergebnissen festgestellt werden.

Der zweite Teil dieser Diplomarbeit beschäftigt sich mit der Ähnlichkeit zwischen ausgewählten acht Arzneistoffen und deren verschiedenen Konformationen. Für die Arzneistoffe und deren Konformationen wurden die vsurf – Deskriptoren in MOE berechnet. Diese wurden dazu verwendet um die Distanzmatrix aller Arzneistoffe und Konformere zu berechnen und im Raum durch eine Hauptkomponentenanalyse darzustellen. Diese soll die Ähnlichkeit eines Arzneistoffes mit seinen Konformeren aufzeigen. Das Resultat zeigt, dass die einzelnen Konformere sich zum Arzneistoff sehr unterschiedlich anordnen.

In Excel wurde weiters nach jeder einzelnen Verapamilkonformation (insgesamt 145) ein Ähnlichkeitsabgleich mit den anderen Arzneistoffen und Konformeren anhand ihrer vsurf – Daten gemacht und nach Übereinstimmung sortiert. Das Ergebnis war, dass im Durchschnitt an Stelle 22 schon ein komplett anderer Arzneistoff einem Verapamil ähnlicher war als die restlichen Verapamilkonformere.

ABSTRACT

The first goal of this thesis was to test five fingerprints (MACCS, GpiDAPH3, SHED, TGT, TGD) and to compare and evaluate their applicability and performance. The question was, whether a certain fingerprint gives particularly good results in a particular similarity search or whether all fingerprints are equally suitable.

Within the scope of this similarity search, 5 fingerprints (MACCS, GpiDAPH3, SHED, TGT, TGD) per drug were calculated for a total of 30 drugs and compared with the fingerprints of the substances in ChEMBL database 23. Depending on how similar a drug was to a substance in the ChEMBL database in terms of its fingerprint, the more advanced it was sorted in a table using the Tanimoto coefficient. For the analyses, the first best of these 1000, 10,000, 100,000 results were used.

The selection of the 30 drugs was made in such a way that the 5 drugs that were most actively tested in a specific target (a specific and a non-specific enzyme, ion channel and transporter) were selected from a list. A certain similarity was assumed for these drugs on the same list, as they bind to the same target. The number of drugs in this list, which then also appear in the best ranked results of the similarity search using fingerprints, was used for the evaluation and evaluation of the individual fingerprints. The larger the number, the better a fingerprint performed in the evaluation. The result was that GpiDAPH3 scored best. It turned out that all fingerprints could find only a small part of the drugs from the list of a target. No correlation between the different target classes and the fingerprint results was found.

The second part of this thesis deals with the similarity between selected eight drugs and their different conformations. The vsurf descriptors in MOE were calculated for the drugs and their conformations. These were used to calculate the distance matrix of all drugs and conformers and to represent them in space by a principal component analysis. This should show the similarity of a drug with its conformers. The result shows that the individual conformers arrange themselves very differently to the drug.

In Excel, after each individual verapamil conformation (145 in total), a similarity comparison was made with the other drugs and conformers on the basis of their vsurf data and sorted according to agreement. The result was that, on average, in place 22, a completely different drug was already more similar to a verapamil than the other verapamil conforming drugs.

Inhaltsverzeichnis

1	Allgemeiner Teil	12
1.1	Einleitung Fingerprintanalyse	12
1.1.1	Ähnlichkeit und Ähnlichkeitssuche	14
1.1.2	Fingerprints	15
1.1.3	MACCS.....	16
1.1.4	GpiDAPH3	16
1.1.5	Shannon Entropy – Based Fingerprint alias SHED	17
1.1.6	TGT und TGD	17
1.1.7	Tanimoto - Koeffizient.....	17
1.2	Einfluss der Konformation auf Ausprägung der vsurf - Deskriptoren.....	18
1.2.1	Konformationsanalyse	19
1.2.2	Vsurf - Deskriptoren	19
1.2.3	Euklidische Distanz	19
1.2.4	Hauptkomponentenanalyse – PCA (Principal Component Analysis).....	20
2	Methoden.....	21
2.1	Ähnlichkeitssuche Fingerprints.....	21
2.1.1	Fingerprints	21
2.1.2	Targets und Targetdatenbank	21
2.1.3	Arzneistoffe.....	22
2.1.4	Vergleichsdatenbank.....	22
2.1.5	Ähnlichkeitssuche.....	23
2.2	Einfluss der Konformation auf vsurf - Ausprägung	24
2.2.1	Arzneistoffe.....	24
2.2.2	Konformationsuche	24
2.2.3	Verapamilsuche	26
3	Ergebnisse.....	27
3.1	Ergebnisse der Ähnlichkeitssuche mittels Fingerprints	29
3.2	Konformationsanalyse	43
4	Diskussion	45
4.1	Ähnlichkeitssuche Fingerprints.....	45
4.2	Konformationsanalyse und Verapamilsuche.....	45
4.3	Conclusio	45
4.4	Appendix – Scripts	46
4.4.1	Script to calculate partial charges	46
4.4.2	Script to calculate the Tanimoto Similarity.....	47
5	Literaturverzeichnis.....	48

Abbildungsverzeichnis:

ABBILDUNG 1 VERSCHIEDEN ÄHNLICHKEITSDEFINITIONEN (2) . ABBILDUNG ENTNOMMEN AUS REF (2)	14
ABBILDUNG 2 TORSIONSWINKEL (8)	18
ABBILDUNG 3 PARAMETER SETTING FÜR MOE WASH	23
ABBILDUNG 4 SUCH SETTINGS FÜR STOCHASTISCHE KONFORMATIONSANALYSE IN MOE	25
ABBILDUNG 5 WELCHER FINGERPRINT HAT WELCHE PLÄTZE BELEGT?	36
ABBILDUNG 6 . MACCS – NOTEN JE TARGET	37
ABBILDUNG 7. GPIDAPH3 – NOTEN JE TARGET	38
ABBILDUNG 8 SHED – NOTEN JE TARGET	38
ABBILDUNG 9. TGT – NOTEN JE TARGET	39
ABBILDUNG 10 TGD – NOTEN JE TARGET	39
ABBILDUNG 11 BEISPIEL DULOXETIN – HAUPTKOMPONENTENANALYSE EINZELNE FINGERPRINTWERTUNG.	41
ABBILDUNG 12 BEISPIEL CERIVASTATIN – HAUPTKOMPONENTENANALYSE GESAMTWERTUNG FINGERPRINTS	41
ABBILDUNG 13 BEISPIEL RITONAVIR – HAUPTKOMPONENTENANALYSE GESAMTWERTUNG FINGERPRINTS	42
ABBILDUNG 14:KONFORMERDATENBANK – HAUPTKOMPONENTENANALYSE. DIE KÜRZEL IN ABBILDUNG 11 STEHEN FÜR PROPAFENONDERIVAT (GPROP), PROPAFENON (PROPA), WARFARIN (WAR), DIAZEPAM (DIA), AMIODARON (AMIO), VERAPAMIL (VERA), FLECAINID (FLEC) UND ENCAINID (ENC)	43
ABBILDUNG 15: AN WELCHER STELLE KOMMT WIE OFT DER ERSTE ARZNEISTOFF, DER KEIN VERAPAMIL IST. DIE X- ACHSE AUS, AN WELCHER STELLE DAS 1. „NICHT VERAPAMIL“ KOMMT. DIE Y – ACHSE GIBT AN WIE HÄUFIG EINE BESTIMMTE STELLE VORKOMMT	44
ABBILDUNG 16 WIE OFT KOMMT EINE BESTIMME ANZAHL AN VERAPAMIL VON 100 MÖGLICHEN VERAPAMILS VOR ? DIE X – ACHSE DIE SUMME DER VERAPAMILS VON DEN ERSTEN 100 ERGEBNISSEN BEI EINER SUCHE NACH EINEN BESTIMMTEN VERAPAMIL. DIE Y – ACHSE GIBT AN WIE OFT EINE BESTIMMTE ANZAHL AN VERAPAMILMOLEKÜLEN VORKOMMT	44

Tabellenverzeichnis:

3.1 TABELLE NODIUM CHANNEL PROTEIN TYP 5 ALPHA SUBUNIT – FINGERPRINTSUCHE ERGEBNISSE	29
3.2 TABELLE SEROTONINTRANSPORTER – FINGERPRINTSUCHE ERGEBNISSE	30
3.3 TABELLE HMG – COA REDUKTASE – FINGERPRINTSUCHE ERGEBNISSE	31
3.4 HERG – KANAL – FINGERPRINTSUCHE ERGEBNISSE	32
3.5 P - GLYCOPROTEIN – FINGERPRINTSUCHE ERGEBNISSE	33
3.6 CYP3A4 – FINGERPRINTSUCHE ERGEBNISSE	34
3.7 NODIUMCHANNEL – BENOTUNG FINGERPRINT	35
3.8 SEROTONINTRANSPORTER – BENOTUNG FINGERPRINT	35
3.9 HMG – COA REDUKTASE INHIBITOR – BENOTUNG FINGERPRINT	35
3.10 HERG - KANAL – BENOTUNG FINGERPRINT	35
3.11 P - GLYKOPROTEIN – BENOTUNG FINGERPRINT	35
3.12 CYP3A4 – BENOTUNG FINGERPRINT	35
3.13 GESAMTWERTUNG DER FINGERPRINTS; ERGIBT SICH AUS TABELLE 3.7 – 3.12	36
4.1 SKRIPT FÜR DEN TANIMOTOWERT	47

1 Allgemeiner Teil

1.1 Einleitung Fingerprintanalyse

In der Pharmaindustrie wird heutzutage sehr viel Geld investiert um potentielle neue Wirkstoffe für den medizinischen Zweck zu entwickeln. Diese Neuerschaffung ist aber nicht nur kostspielig, sondern auch sehr zeitaufwendig. Das Durchlaufen der bekannten klinischen Phasen dauert mehrere Jahre.(1) Auf Grund dessen ist es für die Pharmafirmen sehr wichtig, die richtige molekulare Verbindung, die später als Arzneimittel verpackt werden soll, möglichst früh zu definieren und alle anderen Kandidaten, die auch in der Auswahl für die Weiterentwicklung stehen und die zum Beispiel nicht potent genug oder nicht löslich genug oder sogar toxisch sind, auszuschließen.

Ein wichtiger Begriff in diesem Zusammenhang ist die „chemische Ähnlichkeit“ dieser Moleküle. Hierbei wird nach dem Prinzip gearbeitet, dass man davon ausgeht, dass Moleküle, die einander chemisch ähnlich sind, sich auch in einem biologischen System ähnlich verhalten und somit beim Menschen ein ähnliches pharmakologisches Profil zeigen(1).Hierbei liegt das Hauptaugenmerk auf der chemischen Struktur, sprich den Aufbau des Moleküls mit den speziellen funktionellen Gruppen, die bestimmte physikochemische Eigenschaften tragen. Programme die gut differenzieren können, ob ein Molekül zu einem anderen Molekül ähnlich ist, können somit kostspielige experimentelle Versuchsreihen minimieren und Neuentwicklungen von Arzneistoffen antizipieren. Dieser Punkt wäre besonders wichtig wenn man an das toxikologische Profil von Arzneistoffkandidaten denkt. Wenn man während der Arzneistoffentwicklung bemerkt, dass der potentielle Arzneistoff toxisch für den menschlichen oder tierischen Körper ist und man die toxische Komponente nicht eliminieren kann ohne das Wirkprofil zu beeinflussen, führt dies meist zu einem Drop – out des Wirkstoffes in der klinischen Phase. Daher ist es sehr wichtig, potentielle toxische Wirkungen so früh wie möglich ausschließen zu können.

Obwohl dieses Prinzip, zwei Verbindungen einfach zu vergleichen, sehr einfach klingt, ist es methodisch sehr schwer zu erarbeiten. Wenn man einen Ähnlichkeitsabgleich zweier Verbindungen mit Hilfe spezieller Computerprogramme macht, findet man zwar Verbindungen die nahezu genau aufeinander passen oder eine Vielzahl von Verbindungen, die einen gleichen Strukturteil, aber unterschiedliche Reste, haben.

Es ist unmöglich mit absoluter Sicherheit zu sagen, dass sich zwei Strukturen ähnlich sind und wie sehr sie sich ähnlich sind.(2) Genauso wäre dies das Problem, würde man zwei Menschen vergleichen. Man könnte auch hier das Augenmerk z.B. auf die Haarfarbe, den Charakter oder klinische Parameter legen. Jedoch ist eine perfekte Methode, die alle möglichen zu vergleichenden Parameter einschließt und entsprechend eine Bewertung für die Wichtigkeit zuteilt, auch hier bisher noch nicht gefunden.

Ein Grund dieser Problematik ist, dass Ähnlichkeit sehr vielfältig definiert werden kann. Je nach Betrachter und Kontext, werden Moleküle als ähnlich/unähnlich eingestuft. Zum Beispiel betrachtet ein physikalischer Chemiker vor allem die 3D – Form und Verteilung von

physikochemischen Eigenschaften über diese Form. Ein Chemiker hingegen, der organische Synthese betreibt, fokussiert sich eher auf die Struktur und die funktionellen Gruppen.(1) Weiters ist zu bedenken, dass auch Verbindungen, die strukturell sehr unterschiedlich sind, auch ähnliche Eigenschaften teilen können. Ebenso können Verbindungen, die sich strukturell ähnlich sind, komplett andere Eigenschaften haben.(2)

Somit stößt man mit den bekannten Computerprogramme und Methoden bei der Ähnlichkeitssuche früher oder später an Grenzen, was schlussendlich dazu führt, dass die letzte Entscheidung, welchen der letzten übrig gebliebenen Wirkstoffe man weiterentwickelt und welchen nicht, eine individuelle Entscheidung bleibt. In der medizinischen Chemie spielt die Erfahrung und der „kognitive Algorithmus den sich Chemiker zu nutzen machen eine wichtige Rolle der Entscheidungsfindung. Allerdings hat sich gezeigt, dass diese Entscheidungen je nach Kontext und Reihung unterschiedlich ausfallen und diese sich nur auf einen Bruchteil der Parameter für die Ähnlichkeitsbeurteilung stützen.(2)(3)

Dies liefert also nicht immer das gleiche Ergebnis wie die Ähnlichkeitsbeurteilung mittels Computer. Besonders an dieser Stelle wäre es daher gut, wenn man sich mehr auf empirische Daten von den bereits untersuchten ähnlichen Molekülen als nur auf das individuelle Empfinden verlassen könnte. Eine Möglichkeit um die Effizienz eines Algorithmus objektiv zu beurteilen ist die Verwendung einer Validierungs – Datenbank. Die Performance eines Algorithmus kann abgeschätzt werden indem man eine Referenzsubstanz heranzieht, welches ein spezifisches Wirkprofil besitzt. Nun rechnet man die Ähnlichkeit der Referenzsubstanz gegen alle Moleküle der Validierungs-Datenbank und sortiert diese nach Größe der Ähnlichkeit. Falls der Algorithmus, der für die Ähnlichkeitsberechnung herangezogen wurde, gut funktioniert hat, sollten Moleküle mit einem ähnlichen Wirkprofil wie die Referenzsubstanz, ganz oben in der gereihten Datenbank zu finden sein.

Heutzutage kann man mit dem nötigen Know-how eine Vielzahl dieser Vortestungen in silico, sprich am Computer, ablaufen lassen. Weiters hat sich aber auch gezeigt, dass die Beurteilung der Aussagekraft der Computerentscheidung nicht immer leicht ist. Oft sind die Ergebnisse schwer interpretierbar, da man auch von diversen kognitiven Aspekten beeinflusst wird. Man tut sich leichter, eigene Testungen und deren Ergebnisse zu bewerten als einem Computerprogramm zu glauben, welches die Ergebnisse strikt nach dem jeweiligen Algorithmus liefert.(2)

Wie aussagekräftig das Ergebnis einer dieser Testungen ist, ist auch sehr von der verwendeten Methode, den vorhandenen Datensätzen und der Erfahrung des Wissenschaftlers abhängig. Es ist daher auch wichtig, die verschiedenen Methoden untereinander zu vergleichen und die Stärken und Schwächen aufzuzeigen, damit dann für die jeweilige Problemstellung die beste Methode gefunden werden kann. Es kann auch sein, dass es nicht nur eine beste Methode, um eine Ähnlichkeitssuche durchzuführen, gibt, sondern dass eine Kombination mehrerer Methoden das Beste ist. Um eine Ähnlichkeitssuche durchzuführen und Strukturen vergleichen zu können, muss der Computer Eigenschaften eines Moleküls berechnen, diese Eigenschaften werden im Fachjargon Deskriptoren genannt. Welche Deskriptoren sich besonders gut für Ähnlichkeitsanalyse eignen, ist eine Aufgabe dieser Diplomarbeit.

1.1.1 Ähnlichkeit und Ähnlichkeitssuche

Chemical similarity		Mol. weight	LogP	Rotatable bonds	Aromatic rings	Heavy atoms
	A	341.4	5.23	4	4	26
	B	463.5	4.43	4	5	35
Molecular similarity						
2D similarity						
3D similarity						
Biological similarity		Vascular endothelial growth factor receptor 2	Tyrosine-protein kinase TIE-2			
	A	active	inactive			
	B	active	active			
Global similarity						
Local similarity						

Abbildung 1 Verschieden Ähnlichkeitsdefinitionen(2) . Abbildung entnommen aus Ref (2)

Ähnlichkeit und die daraus folgenden funktionellen Eigenschaften eines Moleküls kann über verschiedene Arten definiert werden. Dies ist in der Abbildung 1 illustriert. Welche Ähnlichkeitseigenschaften nun wichtig sind, hängt auch immer von der Fragestellung ab. (2)

Die Ähnlichkeitssuche basiert darauf, Moleküle aus einer Datenbank, die ähnlich mit einer bestimmten ausgewählten Verbindung sind zu identifizieren. Die Effektivität so einer Suche kann sehr unterschiedlich ausfallen. Die Information der molekularen Strukturen wird auf der Basis von molekularen Deskriptoren oder Fingerprints, die zum Beispiel aus einer binären Folge von 0 und 1 bestehen für das jeweilige Computerprogramm abgespeichert. Diese Attribute und können je nach verwendeter Methode auf Basis von 2D oder 3D Merkmalen berechnet werden. Der Begriff molekularer Deskriptor umfasst zerlegte Fingerprints, binäre Fingerprints, strukturelle Parameter, verschiedene Pharmakophor – basierende Fingerprints und skalare Werte. Für die Kalkulation dieser Deskriptoren sind verschiedene Berechnungsmethoden verfügbar.

Jede Ähnlichkeitsberechnungsmethode besteht aus einem mathematisch berechneten Teil, also aus einem Graph, Vektor, Funktion oder einem Datensatz, welcher die relevante molekulare Information widerspiegelt und aus einer Art Ähnlichkeitsindex oder Koeffizient. (1)

1.1.2 Fingerprints

Wie schon in den oberen Kapiteln erwähnt, stellen Fingerprints eine Deskriptorenuntergruppe dar. Fingerprints werden schon eine lange Zeit verwendet, um am Computer ein virtuelles Screening durchzuführen und die Arzneistoffentwicklung zu ermöglichen. Fingerprints sind eine sehr populäre Methode, da diese keine besonders aufwendige Kalkulation benötigen und schnell berechnet sind.

Zwei Verbindungen zu vergleichen eröffnet eine Vielzahl von Möglichkeiten. Chemiker haben daher auf Grund von Erfahrungswerten definiert, welche Substrukturen oder welcher Aspekt des Moleküls für eine bestimmte Fingerprintberechnung wichtig sind. Dies dient dazu, die Komplexität eines Moleküls zu vereinfachen. Anhand dieser Substrukturen oder einer bestimmten Eigenschaft wird ein bestimmter Vektor berechnet. Am häufigsten werden molekulare Fingerprints benutzt. Hierbei stellt der berechnete Vektor am Ende eine einfache Bitfolge dar, der nun dazu dient, Moleküle leicht miteinander zu vergleichen.

Dieser Vektor charakterisiert nun die Verbindung und ermöglicht dem Computer diese Moleküle z.B. untereinander in Gruppen zu teilen oder sie zu vergleichen. Somit kann dann eine Aussage getroffen werden, wie ähnlich sich zwei Verbindungen sind(4). Es gibt sehr viele unterschiedliche Fragestellungen um Verbindungen zu vergleichen. Das führt zu vielen verschiedenen Fingerprints, berechnet über diverse Algorithmen, die man für ein Molekül kalkulieren kann. Das wirft nun die Frage auf, wie gut funktionieren bestimmte Fingerprint Methoden und welche liefert die besten Ergebnisse für eine bestimmte Fragestellung? Gibt es sehr starke Unterschiede zwischen den jeweiligen Methoden oder sind die jeweiligen Ergebnisse überall gleich repräsentativ?

Um diese Fragen beantworten zu können habe ich fünf Fingerprints während dieser Diplomarbeit untereinander verglichen. Diese sind MACCS, GpiDAPH3, SHED, TGT, TGD.

Diese ausgewählten Fingerprints werden in den folgenden Unterkapiteln (1.3.1 – 1.3.4) erklärt.

1.1.3 MACCS

Der MACCS – Fingerprint zählt zu den bekanntesten Fingerprints. Er beinhaltet Daten der 2D – Struktur und benötigt für seine Berechnung gewisse Substrukturen einer Verbindung. Mathematisch stellt der MACCS – Fingerprint einen Bitvektor dar. Distanzen zwischen zwei Molekülen (Bitvektoren) wird über ein Distanzmaß berechnet. Ein sehr populäres Distanzmaß für Bitvektoren stellt der Tanimotokoeffizient dar. Dieser ist näher im Unterkapitel 1.3.6 erklärt. (1)

Die chemische Struktur eines Moleküls wird mittels verschiedener Fragen, zum Beispiel ob ein Sauerstoff - Atom, eine Disulfid – Bindung oder ein Tetracyclus, etc. vorhanden ist, ermittelt. Diese Fragen werden mit Ja oder Nein beantwortet und somit ein binärer Bitvektor aufgestellt. Die Länge der Bitfolge richtet sich hierbei nach der Anzahl der Fragen. Diese Bitfolge kann dann sehr leicht mit den Bitfolgen einer Moleküldatenbank verglichen werden.

Diese Art von Fingerprints ist dann empfehlenswert, wenn eine große Anzahl an abgefragten Substrukturen in der gewünschten Vergleichsverbindung vorhanden ist. (4)

1.1.4 GpiDAPH3

GpiDAPH3 – Fingerprint (graph-based 3-point pharmacophore) basiert ebenso auf einem Graph. Jedem Atom wird eine von acht berechneten Eigenschaften zugewiesen. Diese werden von drei möglichen Atomeigenschaften („in Pi system“, „ist ein Donor“ und „ist ein Akzeptor“) berechnet. Die Abstände zwischen den Atomen werden in acht verschiedene Bindungsdistanzen geteilt.(5)(6)

1.1.5 Shannon Entropy – Based Fingerprint alias SHED

Der SHED – Fingerprint (Shannon entropy descriptor) stellt eine Alternative zu den kommerziell verwendeten bekannten Fingerprints dar. Hierbei geht man davon aus, dass sich die Entropie beim Hinzufügen einer Verbindung zu einer Datenbank ändert. Eine Verbindung, die den Referenzmolekülen einer Datenbank in der Fingerprint – Erscheinung sehr unähnlich ist, wird zu einem Anstieg der Entropie führen. Während eine Verbindung, die sich ähnlich ist, kaum eine Entropieänderung im Vergleich zu vorher bewirken wird.(7)

1.1.6 TGT und TGD

TGT (typed graph triangles) und TGD (typed graph distances) sind zwei – Punkt und drei – Punkt pharmakophorbasierte-Fingerprints. Ihre Berechnung erfolgt so wie der Maccs – Fingerprint mit Hilfe gewisser Substrukturen. Das Augenmerk liegt hier auf den Eigenschaften als Donor, Akzeptor, Anion, Kation und der Polarität und Hydrophobie. Sie werden von einem 2D molekularen Graph berechnet. TGD beinhaltet 735 Bit. Mittels 7 Atom - Eigenschaften und dem Abstand von 15 Bindungen werden Atompaaare verschlüsselt.

TGT – Fingerprint hat eine Bitfolge von 13,842 und benutzt drei Distanzen eines Graphs um 3 von 4 Atomeigenschaften zu verschlüsseln. (4)

1.1.7 Tanimoto - Koeffizient

Der Tanimoto - Koeffizient ist eine von vielen möglichen Methoden um molekular Fingerprints zu vergleichen und eine Bewertung zu geben. Dieser Koeffizient wird sehr häufig genommen und besteht aus den übereinstimmenden Merkmalen im Verhältnis zu den nicht übereinstimmenden Merkmalen der einzelnen Variablen. Der Ergebnisbereich spielt sich immer zwischen 0 und 1 ab, unabhängig wie kurz oder lang ein Fingerprint ausfällt. Die Formel für den Tanimoto koeffizient, der auch Jaccard- Koeffizient genannt wird ist folgende: $\frac{c}{a+b-c}$. Vergleicht man nun zwei Objekte, stellt a die Menge an 1 einer Bitfolge des ersten Objektes, und b die Menge Einser in der Bitfolge des zweiten Objekts dar. Die Variabel c stellt die Zahl der in beiden Molekülen übereinstimmend auf 1 gesetzten Bits dar.(4)

1.2 Einfluss der Konformation auf Ausprägung der vsurf - Deskriptoren

Ein Molekül kann auch über seine Konformation im dreidimensionalen Raum beschrieben werden. Die Konformation gibt eine Aussage darüber, welche Winkel die frei drehbaren Bindungen [Abbildung 2] einnehmen und definiert damit die räumliche Ausdehnung des Moleküls. Zu der Konformation werden Informationen über die Konfiguration und die Konstitution gezählt. Die Konfiguration gibt Aufschluss über die intramolekularen Atombindungen, ohne dass man Drehungen um Einfachbindungen berücksichtigt. Die Konstitution gibt etwas über die Anzahl und Art der Bindungen eines Moleküls an.

Der Torsionswinkel gibt die Drehung um Einfachbindungen wieder. In Abbildung 2 ist dies ersichtlich, wenn man sich die verschiedenen Zustände der Einfachbindungen anschaut. Diese können gestaffelt oder ekliptisch vorliegen, was ein unterschiedliches Energieniveau der Bindung zur Folge hat. Die gestaffelte Anordnung (staggered) hat das niedrigere Energieniveau und stellt damit die günstigere Konformation dar.

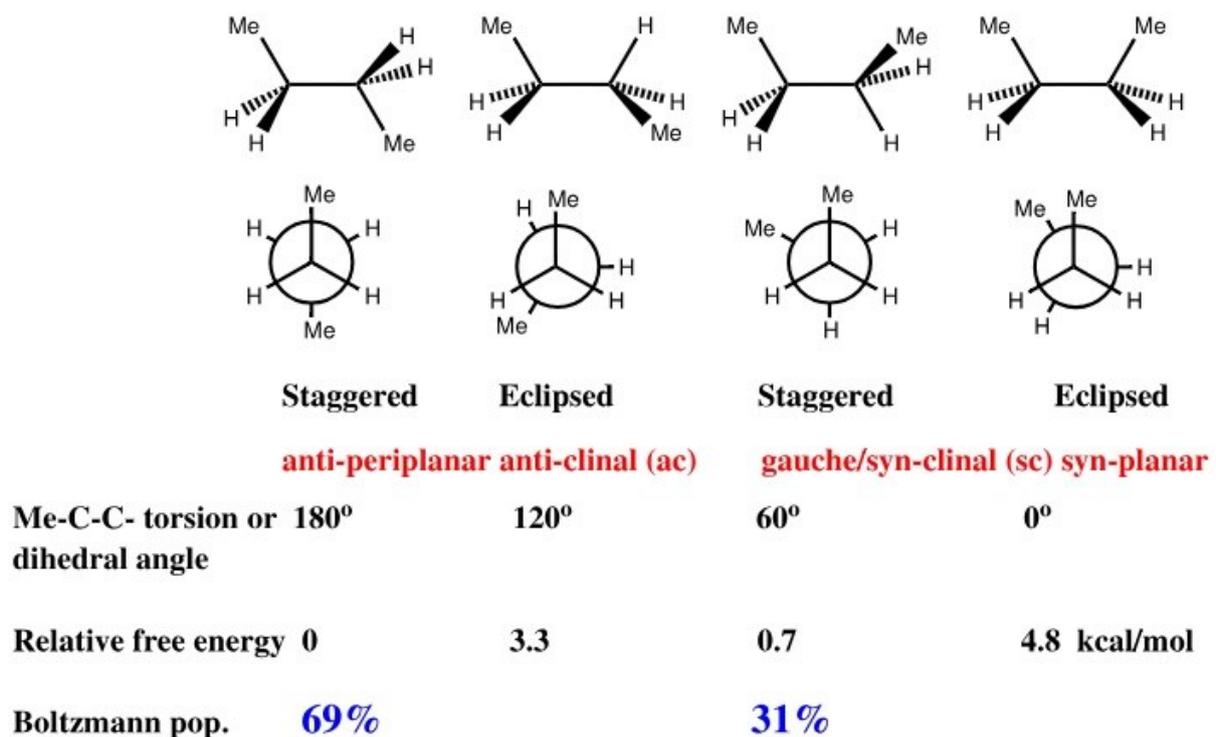


Abbildung 2 Torsionswinkel Entnommen aus Ref (8)

1.2.1 Konformationsanalyse

Da die bioaktive Konformation häufig nicht bekannt ist nutzt man bei der computerunterstützten Konformersuche das Prinzip der Vervielfältigung. Ein Molekül wird ausgewählt und mittels spezieller Algorithmen wird dasselbe Molekül mit unterschiedlichen Torsionswinkeln generiert und die resultierenden Konformationen werden in einer Datenbank abgespeichert. Die einzige Voraussetzung hierfür ist das Vorhandensein von mindestens einer frei-drehbaren Bindung. Werden Konformere zufällig erneut gebaut, die schon zuvor in der Datenbank abgespeichert wurden, werden diese automatisch verworfen, sodass die Datenbank am Ende aus den möglichst diversesten Konformeren besteht. Je nach Rigidität eines Moleküls erhält man eine unterschiedlich große Anzahl an Konformeren. (9)

Diese Technik der Vervielfältigung findet heutzutage große Bedeutung im „Docking“, „Pharmakophorebasierte Screening“, oder „shapebasiertes Screening“, da bioaktive Konformere näher an dem energetischen Minimum liegen als gedacht. (10)

1.2.2 Vsurf - Deskriptoren

MOE2015 bietet ein Set von 78 vsurf – Deskriptoren an, die vergleichbar mit den bekannten volsurf – Deskriptoren sind. Diese repräsentieren Eigenschaften der hydrophoben und hydrophilen Oberfläche, welche über ein molekulares Interaktionsfeld, welches Original von Cruciani et al. beschrieben wurde, berechnet werden. (11)

Vsurf – Deskriptoren können mit Hilfe der 3D Struktur eines Moleküls berechnet werden, in dem diese 3D Struktur zu einer 2D Struktur komprimiert wird. Diese ist dann vereinfacht zu interpretieren und wurde ursprünglich entwickelt um die pharmakokinetischen Eigenschaften eines Moleküls zu charakterisieren. Hervorzuheben ist, dass die volsurf wie auch die vsurf Deskriptoren konformationsabhängig sind. Das heißt, das gleiche Molekül kann unterschiedliche Deskriptorenwerte zeigen, abhängig von der jeweiligen 3D – Konformation. Ein Ziel dieser Diplomarbeit ist es diesen Effekt im Kontext von Ähnlichkeitsanalysen zu beleuchten.

1.2.3 Euklidische Distanz

Der euklidische Abstand gibt die Distanz zwischen zwei Objekten, die sich in einem n-dimensionalen Raum befinden, an. Quantifizierbar ist dieser Abstand durch die Länge der kürzesten Linie, im n-dimensionalen Raum, die diese zwei Objekte verbindet. Diesen kann man sowohl in der Ebene als auch im Raum mit Hilfe des Pythagoras - Satzes berechnen.

$$d(p, q)^2 = (q_1 - p_1)^2 + (q_2 - p_2)^2$$

Im Kontext der computerunterstützten Wirkstoffentwicklung stellen die Objekte Moleküle dar und die Dimensionen werden durch Deskriptoren aufgespannt.

Bei der Ähnlichkeitssuche ist die euklidische Distanz oft sehr hilfreich, vor allem wenn man einen Überblick über die vorhandenen Daten bekommen möchte. Sie kann bei jeglicher Art von Ähnlichkeitsmessungsmethoden verwendet werden. (12)

1.2.4 Hauptkomponentenanalyse – PCA (Principal Component Analysis)

Die Methode der Hauptkomponentenanalyse ist besonders wichtig für die explorative Datenanalyse und für die Anwendung in der multivariaten Statistik. Das Ziel ist hier schnell einen Überblick über die vorhandenen Daten zu bekommen.

Das Prinzip dieser Analyse ist es, eine Dimensionsreduktion vorzunehmen. Man erhält somit wenige Hauptkomponente, die die Originalvariablen beschreiben und zusammenfassen. Die erste Hauptkomponente (HK) fasst das Maximum an Informationen zusammen, während die Zweite das Maximum an fehlenden Informationen beinhaltet, die in der ersten HK nicht abgedeckt wird. Dieses Schema erfolgt für alle Variablen.(13)

2 Methoden

2.1 Ähnlichkeitssuche Fingerprints

Im Rahmen dieser Diplomarbeit wurden bereits bekannte und am Markt befindliche Arzneistoffe mit einer Liste aus anderen bekannten Molekülen, die der ChEMBL – Datenbank entnommen wurden, verglichen.⁽¹⁴⁾ Dieser Vergleich wurde durch eine Ähnlichkeitssuche mittels Deskriptoren, den Fingerprints, die für jeden Arzneistoff einzigartig berechnet worden waren, aufgestellt. Die genaue Vorgehensweise und die Methodik sollen in den nächsten Kapiteln erklärt werden.

2.1.1 Fingerprints

Für dieses Projekt wurden 5 verschiedene Fingerprintarten ausgewählt. Die Auswahl war abhängig davon, welche Fingerprints sich gut in die Software MOE, in welcher der Hauptteil dieses Projekts durchgeführt wurde, integrieren ließen. Ein weiterer Punkt war, dass sich die Fingerprints untereinander auch unterscheiden sollten. Hier sollte geprüft werden, ob unterschiedliche Ergebnisse mit den diversen Fingerprints erzielt werden, oder ob alle Fingerprints das gleiche Ergebnis liefern. Weiters wurden auch Fingerprints miteinbezogen, von denen bekannt ist, dass sie in diversen Anwendungen in der Praxis funktionieren und welche daher auch am Öftesten als Standard benutzt werden. Die Entscheidung fiel schlussendlich auf SHED (1.1.5), MACCS (1.1.3), und die auf Pharmakophor-basierenden Fingerprints TGT (1.1.6), TGD (1.1.6) und GpiDAPH3 (1.1.4).

Der SHED – Fingerprint war in MOE nicht vorhanden und wurde daher manuell integriert.⁽¹⁵⁾

2.1.2 Targets und Targetdatenbank

Für die Ähnlichkeitssuche eines bestimmten Arzneistoffes mit anderen Molekülen, wurden Arzneistoffe aus bestimmten Targetklassen ausgesucht. Da die Targetklassen recht unterschiedlich ausfallen sollten, fiel die Wahl auf insgesamt 6 Targets. Von diesen Sechs sollten drei zu eher spezifischen und drei zu unspezifischen Targets gehören. Je Dreierblock wurden je ein Enzym, ein Ionenkanal und ein Transporter ausgewählt. Für die spezifischen Klassen wurden ein Natriumkanal (Sodium channel protein typ 5 subunit alpha, Uniprot ID: Q14524), ein Serotonintransporter (Sodium-dependent serotonin transporter, Uniprot ID: P31645) und die HMG-CoA-Reduktase (3-hydroxy-3-methylglutaryl-coenzym A reductase, Uniprot ID: P04035) genommen. Für die unspezifischen Target wurden der hERG -Kanal (Potassium voltage-gated channel subfamily H member 2, Uniprot ID: Q12809), Pglykoprotein (Multidrug resistance protein 1, Uniprot ID: P08183) und Cyp3A4 (Cytochrome P450 3A4, Uniprot ID: P08684) ausgewählt.

Von ChEMBL wurden die jeweiligen Datensätze zu diesen Targets heruntergeladen. Für die Analysen wurden nur jene Substanzen genommen, für die ein K_i - oder ein IC_{50} - Wert vorhanden war.

2.1.3 Arzneistoffe

Insgesamt wurden 30 Arzneistoffe (5 Arzneistoffe/pro Target) für die Ähnlichkeitssuche genommen. Ausgewählt wurden diese auf Grund ihrer gemessenen K_i - Werte, bzw. IC_{50} - Werte am jeweiligen Target, die der ChEMBL-Datenbank entnommen wurden. Hier fiel die Wahl auf die Arzneistoffe, die in dieser Liste die höchsten K_i/IC_{50} - Werte bereitstellten. Für den Natriumkanal ergab die Literaturrecherche Propafenon, Flecainid, Quinidine, Verapamil und Mexiletin. Die ausgewählten Arzneistoffe für den Serotonintransporter sind Paroxetin, Clomipramin, Sertralin, Dulocetin und Fluoxetin. Cerivastatin, Atorvastatin, Lovastatin, Rosuvastatin und Simvastatin für die HMG-CoA-Reduktase. Für den hERG - Kanal wurden Astemizol, Sertindol, Pimozide, Dofetilid und Terfenadin ausgesucht. Verapamil, Itraconazol, Tarquidar, Elacridar und Cyclosporin wurden für den P-Glycoproteintransporter auserkoren. Für die CYP3A4- Enzymgruppe wurde Ketoconazol, Nefazodon, Ritonavir, Aprepitant und Diltiazem genommen. Alle Arzneistoffe wurden in MOE gezeichnet und energieminiert. Die ausgewählten Fingerprints wurden für jeden Arzneistoff berechnet.

2.1.4 Vergleichsdatenbank

Es wurde die gesamte ChEMBL Version 23 Datenbank mit 1,7 Millionen Molekülen heruntergeladen und in MOE für die Ähnlichkeitssuche vorbereitet. Für die Vorbereitung wurde MOE Wash verwendet (siehe Abbildung 3 für Parameter Setting) um Salze zu entfernen und die Moleküle zu neutralisieren. Desweiteren wurden größeren Moleküle (Massegewicht > 1000) gefiltert. Weiters wurden die PEOE Partialladungen in MOE (siehe 4.4.1) für die Moleküle berechnet. Am Ende verblieb eine Datenbank bestehend aus 1,6 Millionen Molekülen. Für jedes Molekül wurden die SHED, MACCS, TGT, TGD und GpiDAPH3 Fingerprints berechnet.

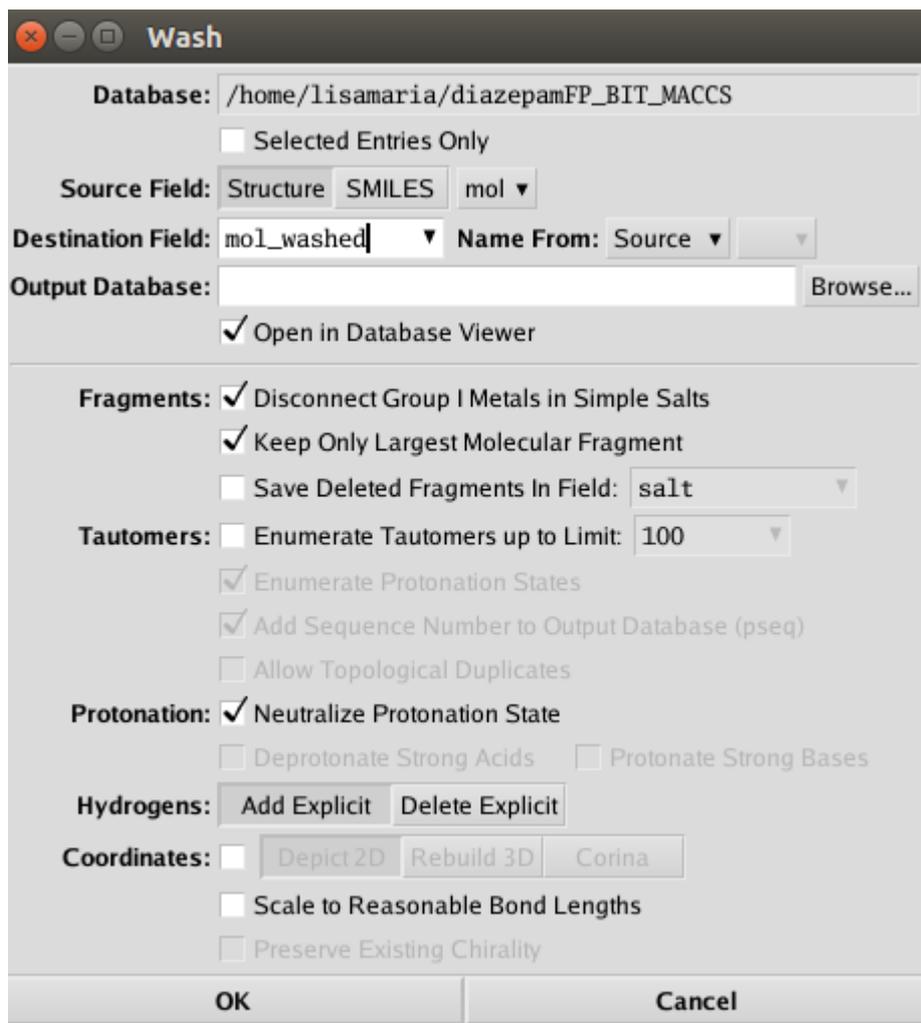


Abbildung 3 Parameter Setting für MOE Wash

2.1.5 Ähnlichkeitssuche

In MOE wurde für jeden Arzneistoff ein Ähnlichkeitsabgleich der berechneten Fingerprints mit den berechneten Fingerprints der ChEMBL – Datenbank ausgeführt. Als Distanzmaß wurde der Tanimoto – Koeffizient verwendet. Mittels integriertem Skript (siehe Appendix) wird jeder dieser 8 Millionen Versuche (1,6 Moleküle mal 5 Fingerprints) pro Arzneistoff ein Koeffizient zugewiesen, wie sehr ein Fingerprint von dem Arzneistoff und einem Molekül in der Datenbank übereinstimmt. Die Ergebnisse werden nach höchster Übereinstimmung sortiert.

Davon werden die top - gereihten 1000, 10.000 und 100.000 Moleküle der ChEMBL - Datenbank für die weitere Auswertung herangezogen. In Excel wurde dann mittels einer *Formel* berechnet, wie viele Arzneistoffe einer Targetliste in den top - gereihten 1000, 10.000 und 100.000 Ergebnissen je Fingerprint vorkommen. Dies wird für jeden der 30 Arzneistoffe gemacht. Um die Resultate zu veranschaulichen, wurden sie in R eingespielt und für diverse PCAs verwendet.

2.2 Einfluss der Konformation auf vsurf - Ausprägung

Um den Einfluss verschiedener Konformere auf 3D-Deskriptoren zu beurteilen, wurde im Rahmen dieser Diplomarbeit eine Arzneistoffdatenbank, die hauptsächlich aus Konformeren bestimmter Arzneistoffen besteht, im Program MOE generiert und bewertet. Die genaue Vorgehensweise soll in den nachfolgenden Kapiteln beschrieben werden.

2.2.1 Arzneistoffe

Für die Konformationsanalyse in MOE wurden 8 Arzneistoffe gezeichnet und energieminiert. Unter diesen Arzneistoffen war Diazepam, um ein starres Molekül im Versuchssatz zu haben. Weiters zählten zu diesem Versuchssatz Propafenon und ein Propafenonderivat, sowie Arzneistoffe welche Ähnlichkeit zu diesem aufweisen indem sie am selben Ionenkanal oder Transporter wie Propafenon binden: Flecainid, Encainid, Verpamil und Amiodaron. Der achte Arzneistoff war Warfarin, um noch einen Arzneistoff in den Datensatz reinzubringen, der sich zu den andern in Wirkung und Struktur sehr unterscheidet.

2.2.2 Konformationssuche

Für jeden Arzneistoff wurde in MOE eine stochastische Konformationsanalyse durchgeführt (siehe Abbildung 4 für die Settings der Suche). Für das starre Diazepam wurden nur 2 Konformationen generiert, für Warfarin 14, für Encainid 40, für Amiodaron 60, für das Propafenonderivat 65, für Verapamil 145, für Propafenon 195 und für Flecainid 199. Aus 8 Arzneistoffen wurden somit insgesamt 720 Konformationen geschaffen.

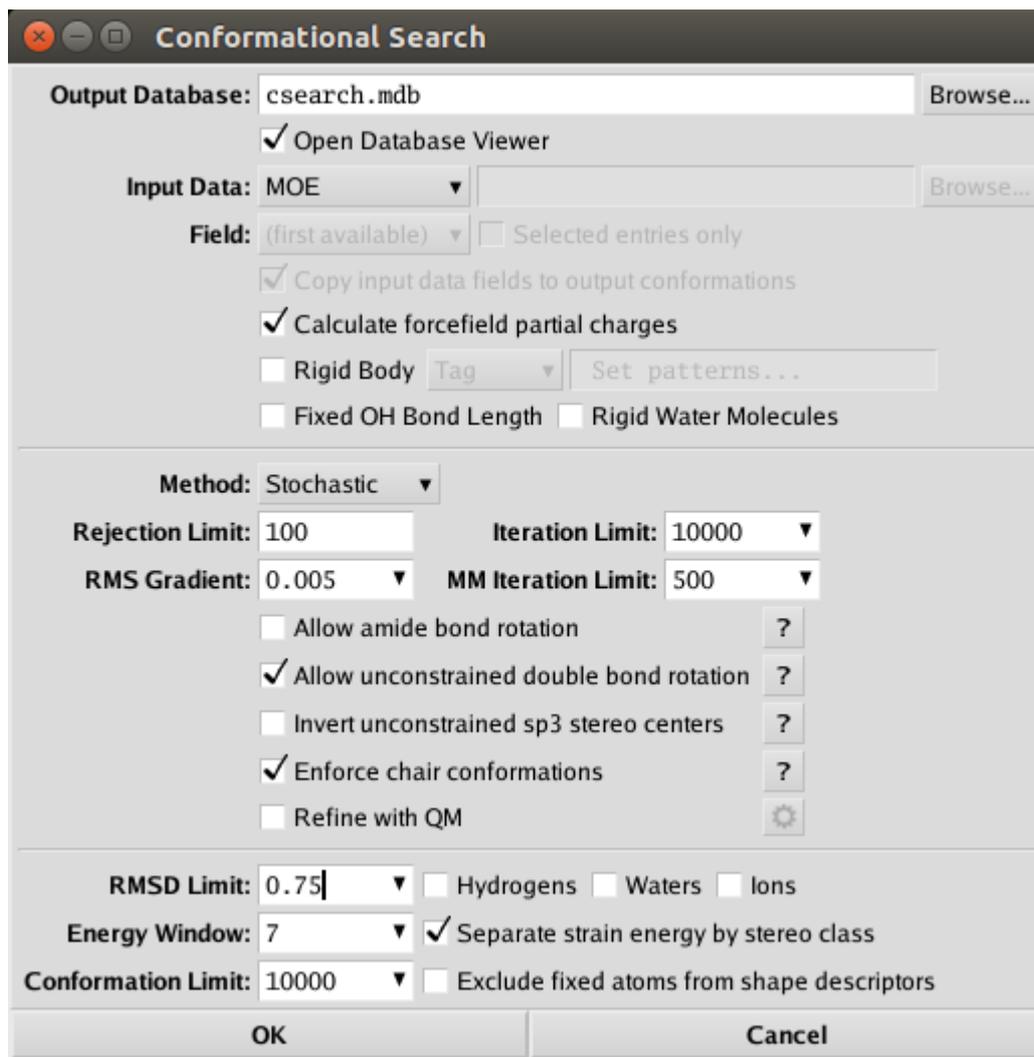


Abbildung 4 Such Settings für stochastische Konformationsanalyse in MOE

Für jede Konformation wurden die vsurf-Deskriptoren, die das Program MOE zu bieten hat, berechnet. Da die vsurf-Deskriptoren konformationsabhängig sind gibt es auch unterschiedliche Deskriptorenausprägungen für das gleiche Molekül. Der gesamte Datensatz wurde in R eingespielt und dort die Distanzmatrix berechnet. Wie nahe sich die einzelnen Arzneistoffe auf Grund ihrer Eigenschaften sind und welche Variabilität verschiedene Konformere erzeugen, wurde mittels Hauptkomponentenanalyse dargestellt. Die Hauptkomponentenanalyse wurde in R mittels des R package *factorextra* durchgeführt.

2.2.3 Verapamilsuche

Der ganze Konformationsdatensatz, der in dem Absatz zuvor beschrieben wurde, wurde in Excel eingespielt. Diese Liste bestand in ihrer Gesamtheit aus 720 Konformationen mit ihren vsurf – Deskriptorwerten und der zuvor berechneten euklidischen Distanzmatrix.

Im folgenden Schritt wurden nun die 145 Verapamilkonformationen nach einander herausgesucht und nach jedem einzelnen Verapamilkonformer des Datensatz hinsichtlich der Ähnlichkeit der vsurf – Deskriptoren unter allen Verbindungen sortiert. Von jeder dieser Sortierungen wurden die Reihung und die Anzahl an gefundenen Verapamil - Konformeren für die weiteren Berechnungen betrachtet. Es wurde analysiert, an welcher Stelle der 1. Arzneistoff in der Liste, der kein Verapamil ist, erscheint, und in eine Liste geschrieben. Davon wurde der Mittelwert gebildet und die ganzen 145 Reihungsergebnisse in einem Histogramm dargestellt.

Weiters wird mit einer anderen Formel (=COUNTIF) berechnet, wie viele Verapamilmoleküle sich unter den ersten 100 Molekülen dieser Reihung befinden und wie viele davon andere Arzneistoffe aus dem Konformationsdatensatz waren. Hier wurde ebenso ein Histogramm zur Veranschaulichung gemacht und der Mittelwert berechnet. Dies kann man in dem Kapitel „Ergebnisse“ in dieser Diplomarbeit einsehen.

3 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der Ähnlichkeitssuche mittels Fingerprints und mittels Distanzmatrix präsentiert.

In den folgenden Tabellen (3.1 – 3.6) kann man die Anzahl der gefundenen Arzneistoffe erkennen, die Ähnlichkeit mit dem Referenzarzneistoff mit dem die Ähnlichkeitssuche durchgeführt wurde, haben. Ähnlich sind sie dem Referenzarzneistoff in dem Sinne, dass sie am selben Target binden. .

In den Tabellen sind die Zählungen limitiert, was bedeutet, dass immer nur die ersten 1000, 10.000, 100.000 Ergebnisse der Ähnlichkeitssuche für den Vergleich mit der Targetdatenbank genommen wurden. Neben der Anzahl der Ergebnisse kann man links bei der Tabelle den jeweiligen zugewiesenen Platz erkennen. Finden zwei oder mehr Fingerprints dieselbe Anzahl an Ergebnissen, bekommen sie denselben Rang zugewiesen. Doppel/Mehrfachbeläge führen zu Auslassung der nächsten Ränge (z.B.: In Tabelle 3.1 werden unter den 485 erstbesten Ergebnissen von vier (MACCS, GpiDAPH3, TGT, TGD) Fingerprints 3 Substanzen gefunden, diese bekommen daher alle den Platz 1. Bei SHED sind es nur 2 Substanzen, daher bekommt dieser Fingerprint hier den 5.Rang.)

Theoretisch könnten alle Arzneistoffe, die an demselben Target binden gefunden werden. Aus diesem Grund wurde für die Auswertung die Anzahl der Verbindungen in der jeweiligen Targetliste genommen.

Referenzdatensatz von aktiven Substanzen

Für jede der selektierten Targets wurden aktive Substanzen in der ChEMBL gesucht. Als aktiv wurden jene Verbindungen mit EC50 und IC50-Werten erachtet die einen numerischen pChEMBL Wert besitzen. Die Suche ergab folgende Anzahl an Verbindungen per Protein Target:

Sodiumchannel: 485

Serotonintransporter: 4120

HMG – CoA – Reduktase – Inhibitor: 219

Herg – Kanal: 6192

P-Glykoprotein: 823

CYP3A4 – Inhibitoren: 3455

Diese Verbindungen wurden annotiert und dann gepoolt in die gesamte ChEMBL mit ihren 1.7 Millionen compounds. Anschließend wurden die Verbindungen gereiht nach Tanimoto-Similarity gegen die Target Arzneistoffe. Aus der gereihten Liste wurde immer eine Fraktion betrachtet (1000, 10.000 und 100.000) um die Wiederfindungsrate von aktiven zu ermitteln.

Der erste Platz wurde einem Fingerprint dann zugewiesen, wenn er im Vergleich zu den anderen Fingerprints die meisten Arzneistoffe gefunden hat. Der zweite Platz wird dem Fingerprint zugewiesen, der die zweit größte Summe an gefundenen Arzneistoffen liefert usw. Jeder Fingerprint bekam je Arzneistoff und je Bereich der zur Auswertung hinzugezogen wurde (insgesamt 4, für die ersten 1000, 10.000, 100.000 und die Anzahl der Arzneistoffe die in der Targetliste sind und gefunden werden könnten) eine Note von 1 bis 5.

3.1 Ergebnisse der Ähnlichkeitssuche mittels Fingerprints

SODIUMCHANNEL

Propafenon									
N outcomes:	<u>platz</u>	1000	<u>platz</u>	10000	<u>platz</u>	100000	<u>platz</u>	*485	**
MACCS	<u>3</u>	1	<u>3</u>	3	<u>2</u>	36	<u>1</u>	1	0,77/0,64
GpiDAPH3	<u>2</u>	2	<u>1</u>	26	<u>1</u>	63	<u>1</u>	1	0,51/0,39
SHED	<u>3</u>	1	<u>3</u>	3	<u>3</u>	32	<u>1</u>	1	0,87/0,71
TGT	<u>1</u>	3	<u>2</u>	7	<u>4</u>	28	<u>1</u>	1	0,89/0,82
TGD	<u>3</u>	1	<u>5</u>	1	<u>5</u>	17	<u>1</u>	1	0,89/0,88
Flecainid									
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	485	
MACCS	<u>1</u>	7	<u>1</u>	52	<u>1</u>	141	<u>1</u>	1	0,74/0,73
GpiDAPH3	<u>2</u>	1	<u>2</u>	1	<u>2</u>	46	<u>1</u>	1	0,53/0,52
SHED	<u>2</u>	1	<u>2</u>	1	<u>3</u>	41	<u>1</u>	1	0,87/0,84
TGT	<u>2</u>	1	<u>2</u>	1	<u>3</u>	41	<u>1</u>	1	0,87/0,82
TGD	<u>2</u>	1	<u>2</u>	1	<u>5</u>	33	<u>1</u>	1	0,89/0,81
Quinidine									
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	485	
MACCS	<u>1</u>	1	<u>1</u>	8	<u>1</u>	95	<u>1</u>	1	0,72/0,64
GpiDAPH3	<u>1</u>	1	<u>1</u>	8	<u>2</u>	78	<u>1</u>	1	0,53/0,52
SHED	<u>1</u>	1	<u>5</u>	1	<u>3</u>	30	<u>1</u>	1	0,89/0,87
TGT	<u>1</u>	1	<u>4</u>	2	<u>4</u>	28	<u>1</u>	1	0,89/0,89
TGD	<u>1</u>	1	<u>3</u>	7	<u>5</u>	17	<u>1</u>	1	0,93/0,91
Verapamil									
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	485	
MACCS	<u>1</u>	1	<u>2</u>	1	<u>1</u>	61	<u>1</u>	1	0,75/0,61
GpiDAPH3	<u>1</u>	1	<u>2</u>	1	<u>5</u>	6	<u>1</u>	1	0,57/0,47
SHED	<u>1</u>	1	<u>1</u>	2	<u>4</u>	37	<u>1</u>	1	0,88/0,84
TGT	<u>1</u>	1	<u>2</u>	1	<u>3</u>	46	<u>1</u>	1	0,96/0,92
TGD	<u>1</u>	1	<u>2</u>	1	<u>2</u>	50	<u>1</u>	1	0,91/0,85
Mexitilene									
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	485	
MACCS	<u>1</u>	3	<u>2</u>	5	<u>2</u>	21	<u>1</u>	3	0,63/0,87
GpiDAPH3	<u>1</u>	3	<u>4</u>	3	<u>1</u>	22	<u>1</u>	3	0,53/0,78
SHED	<u>2</u>	2	<u>4</u>	3	<u>4</u>	7	<u>5</u>	2	0,88/0,82
TGT	<u>1</u>	3	<u>2</u>	5	<u>3</u>	18	<u>1</u>	3	0,83/0,94
TGD	<u>1</u>	3	<u>1</u>	6	<u>1</u>	22	<u>1</u>	3	0,88/0,93

*: maximale Anzahl an Aktiven in der Datenbank.

**::Die erste Zahl gibt an, wie ähnlich das 1. Molekül in der Targetliste (Liste mit allen aktiven 485 Molekülen die an dem Target binden), das nicht der gesuchte Arzneistoff ist, im Vergleich mit dem Arzneistoff, ist. Die zweite Zahl gibt an, wie ähnlich die Verbindung an einer bestimmten Stelle in der ChEMBL – Datenbank zu dem Arzneistoff nach dem gesucht worden ist, ist. Diese Stellenzahl entspricht der Anzahl an möglichen gefundenen Arzneistoffen bzw. Anzahl der aktiven Verbindungen in einer Targetliste

3.1 Tabelle Sodium Channel Protein Typ 5 alpha subunit – Fingerprintsuche Ergebnisse

SEROTONINTRANSPORTER:**PAROXETINE**

N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	4120
MACCS	<u>1</u>	94	<u>1</u>	335	<u>1</u>	1011	<u>1</u>	205
GpiDAPH3	<u>2</u>	79	<u>2</u>	249	<u>2</u>	880	<u>2</u>	155
SHED	<u>5</u>	1	<u>5</u>	17	<u>5</u>	230	<u>5</u>	4
TGT	<u>4</u>	4	<u>3</u>	89	<u>3</u>	658	<u>3</u>	29
TGD	<u>3</u>	9	<u>4</u>	41	<u>4</u>	363	<u>4</u>	23

CLOMIPRAMIN

N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	4120
MACCS	<u>4</u>	34	<u>3</u>	165	<u>2</u>	1234	<u>3</u>	87
GpiDAPH3	<u>1</u>	56	<u>1</u>	400	<u>1</u>	1581	<u>1</u>	214
SHED	<u>5</u>	26	<u>5</u>	83	<u>4</u>	544	<u>5</u>	49
TGT	<u>2</u>	51	<u>2</u>	216	<u>3</u>	564	<u>2</u>	125
TGD	<u>2</u>	51	<u>4</u>	152	<u>5</u>	141	<u>4</u>	80

SERTRALIN

N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	4120
MACCS	<u>4</u>	95	<u>5</u>	254	<u>4</u>	1243	<u>3</u>	162
GpiDAPH3	<u>1</u>	269	<u>2</u>	718	<u>2</u>	1441	<u>1</u>	524
SHED	<u>3</u>	101	<u>1</u>	230	<u>5</u>	1080	<u>4</u>	156
TGT	<u>2</u>	182	<u>4</u>	298	<u>3</u>	1301	<u>2</u>	239
TGD	<u>5</u>	38	<u>3</u>	362	<u>1</u>	1648	<u>5</u>	113

DULOXETIN

N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	4120
MACCS	<u>5</u>	72	<u>4</u>	348	<u>5</u>	1152	<u>5</u>	199
GpiDAPH3	<u>4</u>	121	<u>5</u>	289	<u>4</u>	845	<u>4</u>	210
SHED	<u>3</u>	119	<u>3</u>	430	<u>3</u>	924	<u>3</u>	271
TGT	<u>1</u>	240	<u>2</u>	514	<u>1</u>	1291	<u>2</u>	309
TGD	<u>2</u>	257	<u>1</u>	613	<u>2</u>	1246	<u>1</u>	344

FLUOXETIN

N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	4120
MACCS	<u>3</u>	170	<u>3</u>	138	<u>4</u>	1142	<u>4</u>	200
GpiDAPH3	<u>1</u>	271	<u>2</u>	512	<u>3</u>	1121	<u>1</u>	311
SHED	<u>4</u>	114	<u>4</u>	281	<u>5</u>	808	<u>5</u>	139
TGT	<u>4</u>	216	<u>4</u>	281	<u>2</u>	808	<u>2</u>	264
TGD	<u>2</u>	185	<u>1</u>	568	<u>1</u>	1660	<u>3</u>	241

3.2 Tabelle Serotonintransporter – Fingerprintsuche Ergebnisse

HMG-CoA-Reduktase

Cerivastatin								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	219
MACCS	<u>3</u>	51	<u>2</u>	81	<u>4</u>	92	<u>3</u>	20
GpiDAPH3	<u>1</u>	101	<u>1</u>	135	<u>2</u>	150	<u>2</u>	22
SHED	<u>5</u>	28	<u>5</u>	42	<u>5</u>	90	<u>5</u>	7
TGT	<u>2</u>	56	<u>4</u>	63	<u>3</u>	115	<u>1</u>	31
TGD	<u>4</u>	35	<u>3</u>	69	<u>1</u>	151	<u>4</u>	9
ATORVASTATIN								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	219
MACCS	<u>2</u>	48	<u>3</u>	58	<u>3</u>	99	<u>3</u>	28
GpiDAPH3	<u>1</u>	99	<u>1</u>	117	<u>1</u>	148	<u>1</u>	45
SHED	<u>5</u>	7	<u>5</u>	29	<u>5</u>	77	<u>5</u>	7
TGT	<u>4</u>	33	<u>4</u>	48	<u>4</u>	97	<u>2</u>	31
TGD	<u>3</u>	36	<u>2</u>	66	<u>2</u>	109	<u>4</u>	24
LOVASTATIN								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	219
MACCS	<u>3</u>	16	<u>3</u>	19	<u>4</u>	49	<u>3</u>	13
GpiDAPH3	<u>2</u>	19	<u>2</u>	32	<u>3</u>	65	<u>1</u>	16
SHED	<u>5</u>	8	<u>5</u>	15	<u>5</u>	40	<u>4</u>	8
TGT	<u>4</u>	10	<u>4</u>	16	<u>2</u>	69	<u>5</u>	5
TGD	<u>1</u>	27	<u>1</u>	35	<u>1</u>	100	<u>2</u>	14
ROSUVASTATIN								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	219
MACCS	<u>4</u>	14	<u>4</u>	16	<u>4</u>	62	<u>4</u>	13
GpiDAPH3	<u>1</u>	110	<u>1</u>	144	<u>1</u>	173	<u>2</u>	19
SHED	<u>5</u>	1	<u>5</u>	8	<u>5</u>	16	<u>5</u>	1
TGT	<u>2</u>	56	<u>3</u>	62	<u>3</u>	109	<u>3</u>	18
TGD	<u>2</u>	56	<u>2</u>	69	<u>2</u>	147	<u>1</u>	27
SIMVASTATIN								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	219
MACCS	<u>3</u>	14	<u>3</u>	19	<u>5</u>	29	<u>2</u>	10
GpiDAPH3	<u>1</u>	20	<u>2</u>	30	<u>3</u>	56	<u>1</u>	15
SHED	<u>5</u>	10	<u>5</u>	15	<u>4</u>	41	<u>4</u>	8
TGT	<u>3</u>	10	<u>4</u>	16	<u>2</u>	69	<u>5</u>	5
TGD	<u>2</u>	17	<u>1</u>	37	<u>1</u>	102	<u>2</u>	10

3.3 Tabelle HMG – CoA Reduktase – Fingerprintsuche Ergebnisse

HERGKANALBLOCKER:

Astemizol								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	6192
MACCS	<u>3</u>	17	<u>3</u>	148	<u>2</u>	1415	<u>3</u>	60
GpiDAPH3	<u>1</u>	76	<u>1</u>	540	<u>1</u>	1911	<u>1</u>	310
SHED	<u>5</u>	7	<u>5</u>	53	<u>4</u>	642	<u>5</u>	15
TGT	<u>2</u>	34	<u>2</u>	167	<u>3</u>	740	<u>2</u>	112
TGD	<u>4</u>	9	<u>4</u>	72	<u>5</u>	626	<u>4</u>	42
Sertindol								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	6192
MACCS	<u>1</u>	49	<u>2</u>	236	<u>1</u>	1315	<u>1</u>	178
GpiDAPH3	<u>2</u>	32	<u>1</u>	250	<u>2</u>	1281	<u>2</u>	161
SHED	<u>5</u>	10	<u>4</u>	101	<u>5</u>	607	<u>5</u>	55
TGT	<u>3</u>	25	<u>5</u>	96	<u>4</u>	633	<u>4</u>	74
TGD	<u>3</u>	25	<u>3</u>	135	<u>3</u>	712	<u>3</u>	95
Pimozide								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	6192
MACCS	<u>1</u>	62	<u>2</u>	235	<u>2</u>	1284	<u>1</u>	121
GpiDAPH3	<u>2</u>	33	<u>1</u>	250	<u>1</u>	1451	<u>2</u>	105
SHED	<u>5</u>	9	<u>5</u>	39	<u>5</u>	416	<u>5</u>	12
TGT	<u>4</u>	25	<u>4</u>	96	<u>4</u>	633	<u>4</u>	66
TGD	<u>3</u>	25	<u>3</u>	139	<u>3</u>	730	<u>3</u>	87
Dofetilid								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	6192
MACCS	<u>3</u>	13	<u>5</u>	39	<u>5</u>	307	<u>5</u>	21
GpiDAPH3	<u>1</u>	24	<u>1</u>	91	<u>2</u>	743	<u>1</u>	54
SHED	<u>4</u>	11	<u>2</u>	80	<u>3</u>	662	<u>4</u>	26
TGT	<u>1</u>	24	<u>3</u>	74	<u>4</u>	585	<u>3</u>	32
TGD	<u>5</u>	7	<u>4</u>	69	<u>1</u>	725	<u>2</u>	38
Terfenadine								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	6192
MACCS	<u>2</u>	13	<u>2</u>	64	<u>2</u>	715	<u>4</u>	23
GpiDAPH3	<u>1</u>	16	<u>1</u>	67	<u>1</u>	810	<u>1</u>	38
SHED	<u>3</u>	7	<u>5</u>	49	<u>4</u>	473	<u>2</u>	28
TGT	<u>3</u>	7	<u>4</u>	52	<u>5</u>	471	<u>5</u>	21
TGD	<u>3</u>	7	<u>3</u>	61	<u>3</u>	564	<u>3</u>	26

3.4 Herg – Kanal – Fingerprintsuche Ergebnisse

P-GlycoproteinInhibitor

Verapamil									
N outcomes:									
	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>		823
MACCS	<u>2</u>	14	<u>3</u>	38	<u>5</u>	132	<u>2</u>		14
GpiDAPH3	<u>3</u>	13	<u>5</u>	27	<u>1</u>	215	<u>3</u>		13
SHED	<u>5</u>	7	<u>4</u>	36	<u>4</u>	148	<u>4</u>		12
TGT	<u>1</u>	39	<u>1</u>	105	<u>2</u>	204	<u>1</u>		37
TGD	<u>4</u>	10	<u>2</u>	44	<u>3</u>	177	<u>5</u>		9
Itraconazol									
N outcomes:									
	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>		823
MACCS	<u>1</u>	3	<u>5</u>	3	<u>4</u>	35	<u>2</u>		3
GpiDAPH3	<u>5</u>	2	<u>2</u>	5	<u>5</u>	30	<u>5</u>		2
SHED	<u>1</u>	3	<u>1</u>	37	<u>1</u>	210	<u>1</u>		17
TGT	<u>5</u>	2	<u>2</u>	6	<u>2</u>	151	<u>5</u>		2
TGD	<u>5</u>	2	<u>2</u>	5	<u>3</u>	97	<u>5</u>		2
Tarquidar									
N outcomes:									
	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>		823
MACCS	<u>2</u>	31	<u>3</u>	47	<u>3</u>	119	<u>2</u>		29
GpiDAPH3	<u>1</u>	51	<u>1</u>	77	<u>1</u>	130	<u>1</u>		50
SHED	<u>5</u>	9	<u>5</u>	22	<u>5</u>	93	<u>5</u>		12
TGT	<u>4</u>	22	<u>4</u>	44	<u>2</u>	120	<u>4</u>		17
TGD	<u>3</u>	27	<u>2</u>	69	<u>4</u>	115	<u>3</u>		26
Elacridar									
N outcomes:									
	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>		823
MACCS	<u>3</u>	30	<u>4</u>	42	<u>4</u>	108	<u>3</u>		29
GpiDAPH3	<u>1</u>	53	<u>3</u>	55	<u>1</u>	129	<u>1</u>		53
SHED	<u>5</u>	7	<u>5</u>	18	<u>5</u>	96	<u>4</u>		12
TGT	<u>4</u>	10	<u>2</u>	61	<u>2</u>	117	<u>5</u>		9
TGD	<u>2</u>	44	<u>1</u>	68	<u>3</u>	113	<u>2</u>		42
Cyclosporin									
N outcomes:									
	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>		823
MACCS	<u>1</u>	1	<u>1</u>	18	<u>2</u>	56	<u>1</u>		1
GpiDAPH3	<u>1</u>	1	<u>3</u>	1	<u>1</u>	63	<u>1</u>		1
SHED	<u>1</u>	1	<u>2</u>	3	<u>5</u>	15	<u>1</u>		1
TGT	<u>1</u>	1	<u>3</u>	2	<u>3</u>	23	<u>1</u>		1
TGD	<u>1</u>	1	<u>3</u>	1	<u>3</u>	22	<u>1</u>		1

3.5P - Glycoprotein – Fingerprintsuche Ergebnisse

CYP3A4Inhibitor

Ketoconazol								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	3455
MACCS	<u>1</u>	25	<u>2</u>	59	<u>1</u>	434	<u>2</u>	33
GpiDAPH3	<u>2</u>	24	<u>1</u>	63	<u>2</u>	325	<u>1</u>	27
SHED	<u>5</u>	5	<u>4</u>	22	<u>5</u>	242	<u>4</u>	13
TGT	<u>4</u>	4	<u>5</u>	19	<u>4</u>	267	<u>5</u>	5
TGD	<u>3</u>	16	<u>3</u>	50	<u>3</u>	287	<u>3</u>	25
NEFAZODON								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	3455
MACCS	<u>1</u>	16	<u>1</u>	66	<u>1</u>	554	<u>1</u>	33
GpiDAPH3	<u>3</u>	2	<u>5</u>	2	<u>5</u>	122	<u>5</u>	2
SHED	<u>5</u>	1	<u>4</u>	16	<u>2</u>	278	<u>4</u>	6
TGT	<u>2</u>	7	<u>3</u>	24	<u>4</u>	202	<u>2</u>	15
TGD	<u>3</u>	2	<u>2</u>	32	<u>3</u>	267	<u>3</u>	12
RITONAVIR								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	3455
MACCS	<u>1</u>	158	<u>1</u>	193	<u>3</u>	525	<u>1</u>	174
GpiDAPH3	<u>2</u>	76	<u>3</u>	93	<u>4</u>	459	<u>2</u>	85
SHED	<u>3</u>	49	<u>2</u>	111	<u>1</u>	720	<u>3</u>	81
TGT	<u>4</u>	20	<u>4</u>	84	<u>2</u>	551	<u>4</u>	29
TGD	<u>5</u>	13	<u>5</u>	54	<u>5</u>	401	<u>5</u>	25
APREPITANT								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	3455
MACCS	<u>1</u>	16	<u>1</u>	96	<u>1</u>	677	<u>1</u>	39
GpiDAPH3	<u>5</u>	1	<u>2</u>	28	<u>2</u>	367	<u>3</u>	5
SHED	<u>5</u>	1	<u>3</u>	18	<u>3</u>	259	<u>2</u>	9
TGT	<u>5</u>	1	<u>4</u>	7	<u>5</u>	164	<u>4</u>	3
TGD	<u>5</u>	1	<u>5</u>	5	<u>4</u>	165	<u>5</u>	2
Diltiazem								
N outcomes:	<u>Platz</u>	1000	<u>Platz</u>	10000	<u>Platz</u>	100000	<u>Platz</u>	3455
MACCS	<u>1</u>	4	<u>1</u>	38	<u>1</u>	404	<u>1</u>	17
GpiDAPH3	<u>5</u>	1	<u>5</u>	5	<u>3</u>	130	<u>5</u>	1
SHED	<u>2</u>	2	<u>4</u>	6	<u>5</u>	120	<u>4</u>	3
TGT	<u>5</u>	1	<u>2</u>	28	<u>2</u>	239	<u>2</u>	15
TGD	<u>2</u>	2	<u>3</u>	19	<u>4</u>	124	<u>3</u>	12

3.6CYP3A4 – Fingerprintsuche Ergebnisse

In den folgenden Tabellen (Tabelle 3.7-3.12) kann man eine Detailbewertung der Notenvergabe, wie die einzelnen Fingerprints bei den einzelnen Targets abschneiden, erkennen. Hierfür wurden von allen Arzneistoffen je Target und je Fingerprint alle Noten zusammen gezählt und durch die Anzahl an Wertungen dividiert.

Je niedriger die Note, desto mehr Ergebnisse / Moleküle hat der Fingerprint gefunden. Man kann erkennen, dass es kaum einen Unterschied macht, ob das Target spezifisch oder unspezifisch, ein Transporter, Enzym oder ein Ionenkanal ist, da die Fingerprints annähernd immer die gleiche Wertung haben.

Target: Natriumkanal:	Fingerprints	Notenvergabe von 1-5
	Maccs	1,4
	GpiDAPH3	1,65
	SHED:	2,5
	TGT:	2
	TGD	2,2

3.7 Natriumkanal – Benotung Fingerprint

Target: Serotonintransporter:	Fingerprints	Notenvergabe von 1-5
	Maccs	3,25
	GpiDAPH3	2,1
	SHED:	4,1
	TGT:	2,55
	TGD	2,85

3.8 Serotonintransporter – Benotung Fingerprint

Target: HMG – CoA Reduktase Inhibitoren:	Fingerprints	Notenvergabe von 1-5
	Maccs	3,25
	GpiDAPH3	1,5
	SHED:	4,85
	TGT:	3,2
	TGD	2,05

3.9 HMG – CoA Reduktase Inhibitor – Benotung Fingerprint

Target: Herg - Kanal:	Fingerprints	Notenvergabe von 1-5
	Maccs	2,5
	GpiDAPH3	1,3
	SHED:	4,25
	TGT:	3,45
	TGD	3,25

3.10 Herg - Kanal – Benotung Fingerprint

Target: P - Glycoprotein:	Fingerprints	Notenvergabe von 1-5
	Maccs	2,65
	GpiDAPH3	2,25
	SHED:	3,45
	TGT:	2,7
	TGD	2,85

3.11 P - Glykoprotein – Benotung Fingerprint

Target: Cyp3A4 - Inhibitor:	Fingerprints	Notenvergabe von 1-5
	Maccs	1,2
	GpiDAPH3	3,25
	SHED:	3,5
	TGT:	3,6
	TGD	3,7

3.12 Cyp3A4 – Benotung Fingerprint

In der Abbildung 5 kann man erkennen, wie oft ein Fingerprint eine Erreichung des 1. Platzes, geschafft hat.

Es lässt sich herauslesen, dass GpiDAPH3 – Fingerprint am häufigsten den ersten bzw. besten Platz erreichte. Ebenso schneidet dieser Fingerprint mit einer Gesamtwertung von 2,01 in der Gesamtwertung auch am Besten (Tabelle 3.13) ab. MACCS folgt dicht mit 2,38 Die auf Pharmakophor – Struktur basierenden Deskriptoren (TGT, TGD) liefern annähernd gleichwertige Ergebnisse bei einer Wertung von 2,92 und 2,82. Mit gewaltigem Abstand belegt SHED mit einem Gesamtergebnis von 3,78 den letzten Platz. Diese Gesamtwertung ergibt sich durch die Addierung aller Ergebnisse aller Targets je Fingerprint und durch anschließender Teilung durch die Target Anzahl.

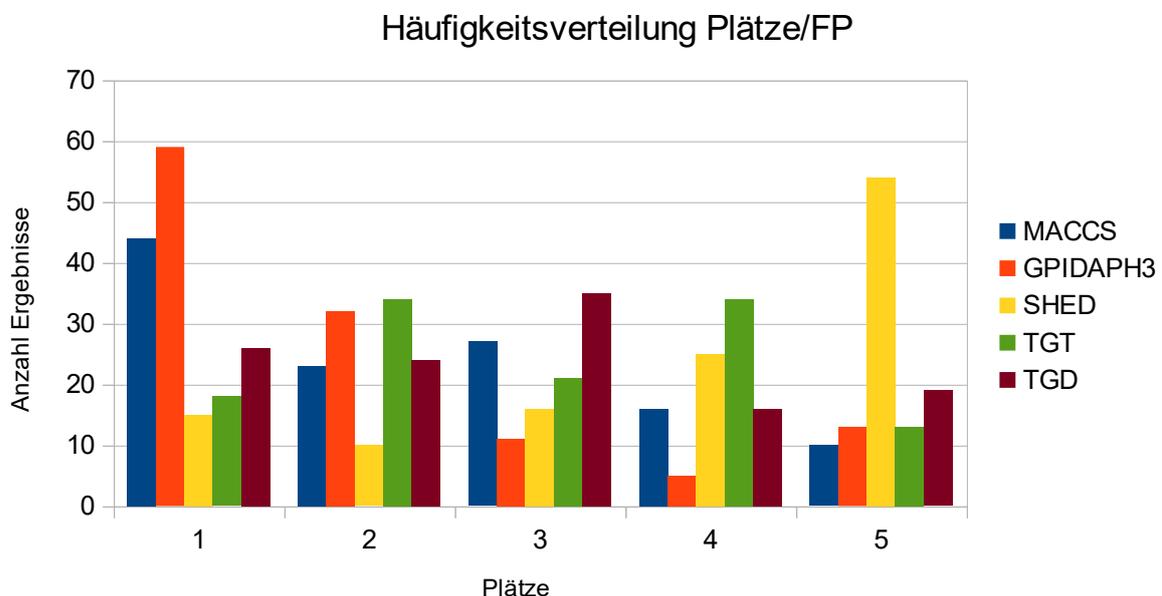


Abbildung 5 Welcher Fingerprint belegt welche Plätze.

Ranking

<u>Fingerprints</u>	<u>Note</u>	<u>Platz</u>
<u>Macacs:</u>	2,38	2.Platz
<u>GpiDAPH3</u>	2,01	1.Platz
<u>SHED:</u>	3,78	5.Platz
<u>TGT:</u>	2,92	4.Platz
<u>TGD</u>	2,82	3.Platz

3.13 Gesamtwertung der Fingerprints; ergibt sich aus Tabelle 3.7 – 3.12

In den folgenden Abbildungen (Abbildung 6, Abbildung 7, Abbildung 8, Abbildung 9 und Abbildung 10) lässt sich erkennen, bei welchem Target ein Fingerprint besonders gut oder schlecht abgeschnitten hat. Man merkt, dass der MACCS – Fingerprint immer eher eine bessere Wertung erzielt, allerdings kann man an den unterschiedlichen Balkenhöhen erkennen, dass die Ergebnisse von MACCS – Fingerprint je Target unterschiedlich sind.

Besonders gut schneidet er bei der Natriumkanal- und beim CYP3A4 – Inhibitorsuche ab, bei HMG – CoA Reduktase Inhibitor liefert er bei allen Arzneistoffen dieses Targets nur ein mittelmäßiges Ergebnis.

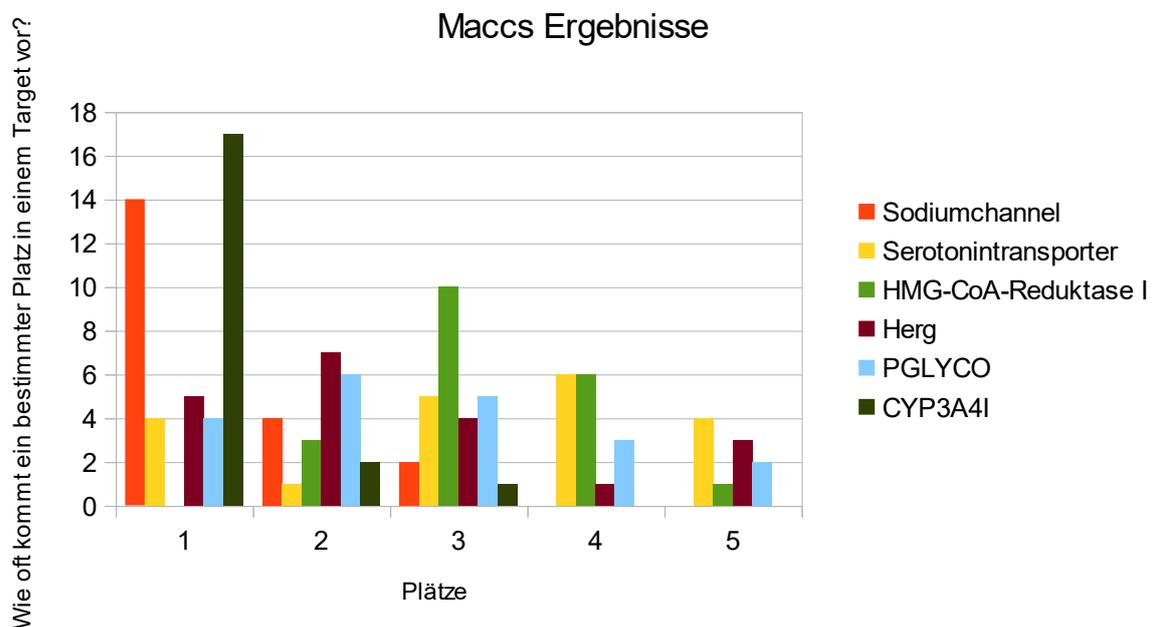


Abbildung 6. MACCS – Noten je Target

In der Abbildung 7 kann man erkennen, dass GpiDAPH3 sehr gute Ergebnisse, außer bei dem CYP3A4 – Target, erzielt.

GpiDAPH3

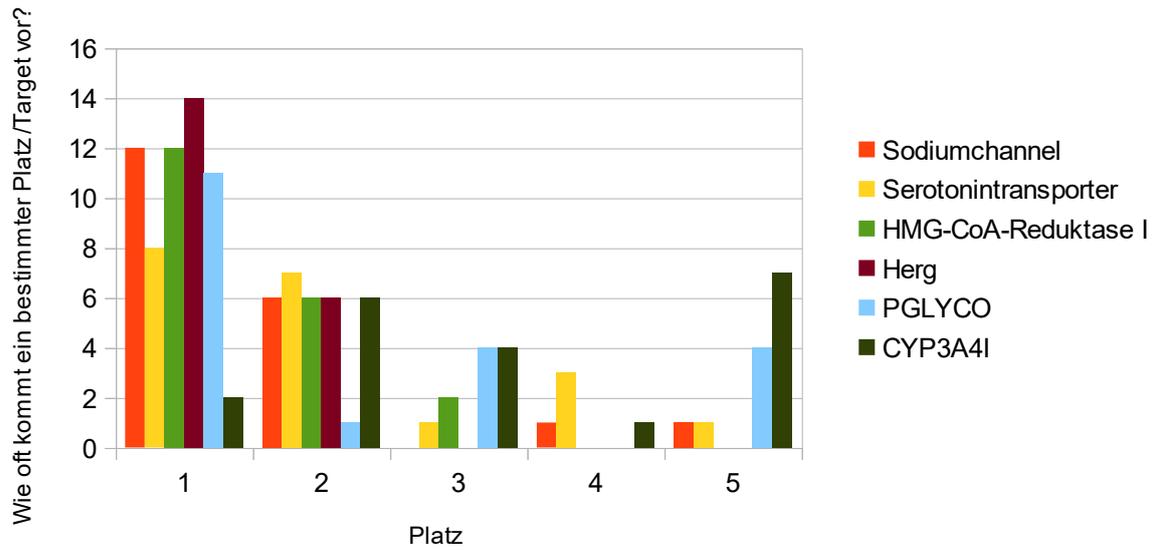


Abbildung 7. GpiDAPH3 – Noten je Target

SHED (Abbildung 8) zeigt insgesamt die schlechteste Performance und belegt häufig den 5. Platz.

SHED

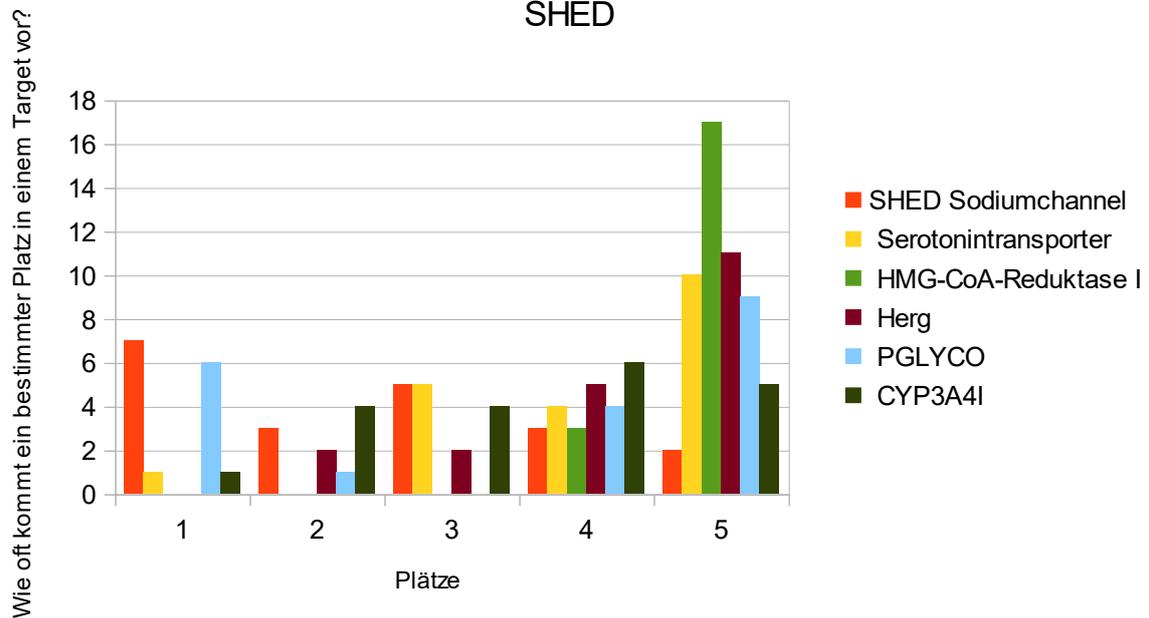


Abbildung 8 SHED – Noten je Target

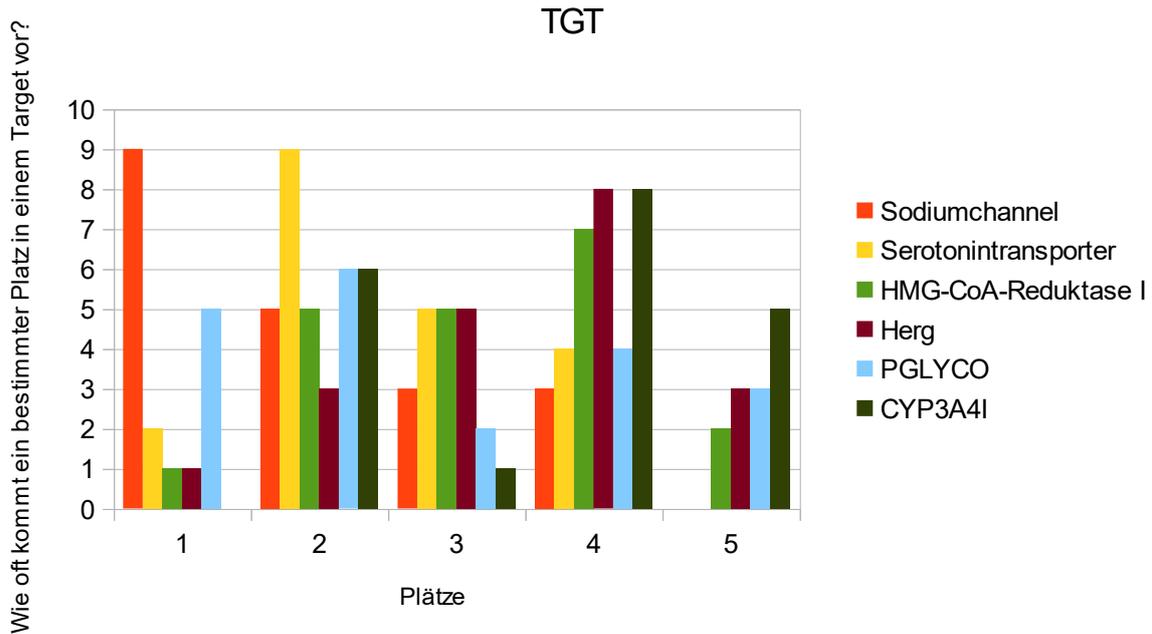


Abbildung 9. TGT – Noten je Target

Die Fingerprints TGT und TGD haben eine vergleichbare Gesamtwertung, allerdings kann man in der Einzelwertung (Abbildung 9 und Abbildung 10) dennoch einige Unterschiede im Muster erkennen. Jedenfalls haben sie trotzdem beide die besten Ergebnisse beim Sodiumchannel und am Schlechtesten haben sie (vor allem TGD) beim CYP3A4 – Target performt.

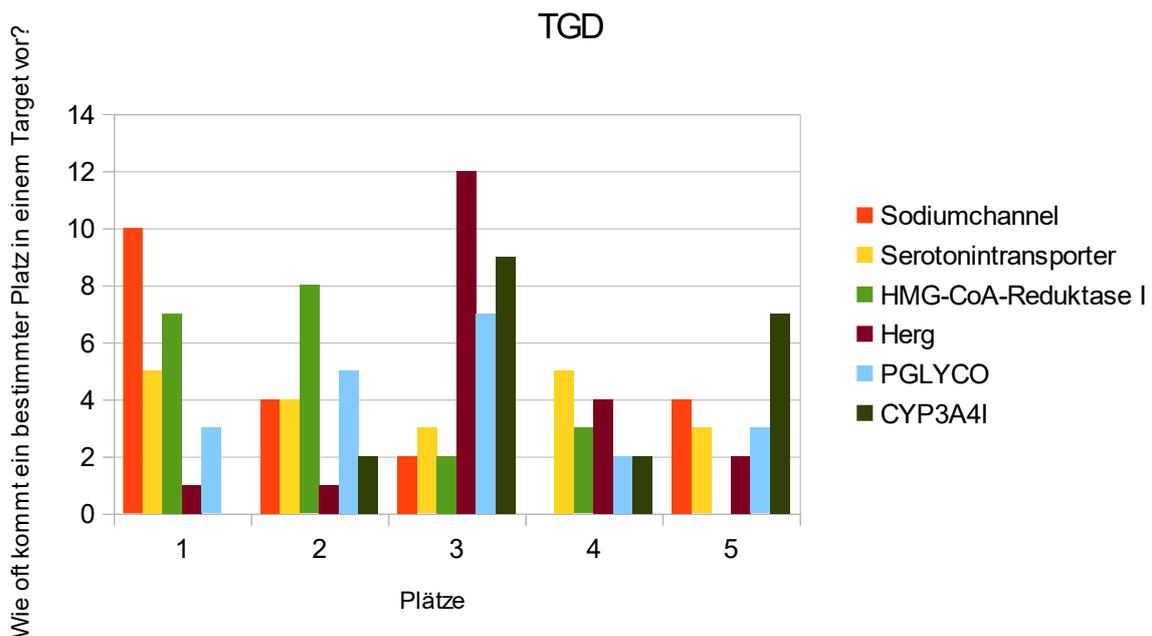


Abbildung 10 TGD – Noten je Target

Näher veranschaulicht wird der ganze Datensatz in einer Hauptkomponentenanalyse (PCA). Das Ziel dieser Analyse war es die Coverage eines Fingerprints hinsichtlich physikochemischer Parameter abzuschätzen. Für diese Analyse haben wir den Serotonin-Transporter Datensatz (Abbildung 8), den HMG-CoA - Reduktase Datensatz (Abbildung 9) und den CYP3A4 Datensatz (Abbildung 10) verwendet. Wir haben für alle Aktiven in den Datensatz folgende physikochemische Deskriptoren in MOE berechnet: *a_acc*, *a_don*, *b_IrotN*, *logP(o/w)*, *mr*, *PEOE_VSA_HYD*, *TPSA*, *vsa_acc*, *vsa_don*, *vsa_hyd* und *Weight*. Anschließend wurde eine Principal Component Analyse in R (prcomp) durchgeführt um eine Dimensionsreduktion auf 2 Achsen durchzuführen. Wir haben dann jene Verbindungen farblich in den PCA-Plots (Abbildung 8-10) hervorgehoben die von einem Fingerprint gefunden wurden.

Die Abbildungen zeigen den möglichen Bereich, den ein Fingerprint abdecken könnte (alle aktiven) und den Bereich, den er wirklich abdeckt. Somit kann auch ausgemacht werden, welcher Fingerprint am Besten ist und ob Verbindungen gefunden werden, die im selben Bereich sind oder willkürlich verstreut liegen. Es wurde bei jedem Arzneistoff eine PCA gemacht. Da allerdings die Abbildungen das gleiche Ergebnis zeigen, wird hier zur Veranschaulichung als zufälliges Beispiel Duloxetin bzw. Cerivastatin und Ritonavir herangezogen.

Anhand dieser Analysen lassen sich nun verschiedene Rückschlüsse ziehen. Der Fingerprint, der den Bereich am Besten abdeckt, ist so gut wie immer GpiDAPH3.

MACCS deckt zwar einen ähnlichen Bereich ab, jedoch ist dieser meist etwas kleiner. TGT, TGD und SHED decken oft einen anderen Bereich als MACCS und GpiDAPH3 ab. Dieses Muster zieht sich durch alle PCAs, daher lässt sich auch hier kein Unterschied zwischen den Targets, ob spezifisch oder nicht, feststellen.

In der Abbildung 11 stellen die grünen (bzw. in Abbildung 12 und Abbildung 13 die grauen Punkte) die Arzneistoffe der Serotonintargetliste (bzw. in Abbildung 12 und Abbildung 13 die jeweiligen anderen Targetlisten für diese Arzneistoffe) dar, die nicht erfasst worden sind. Alle Arzneistoffe dieser Targetliste wurden mit den Ergebnissen der Fingerprintsuche verglichen und analysiert ob sie da dabei waren.

Die blauen bzw. bunten Datenpunkte zeigen die Verteilung der gefundenen Moleküle, welcher Fingerprint welchen Bereich abdeckt und welchen Bereich sie insgesamt alle zusammen abdecken. Hier wird noch einmal veranschaulicht, dass auch alle Fingerprints gemeinsam nicht einmal annähernd den ganzen Bereich an möglichen Arzneistoffen finden.

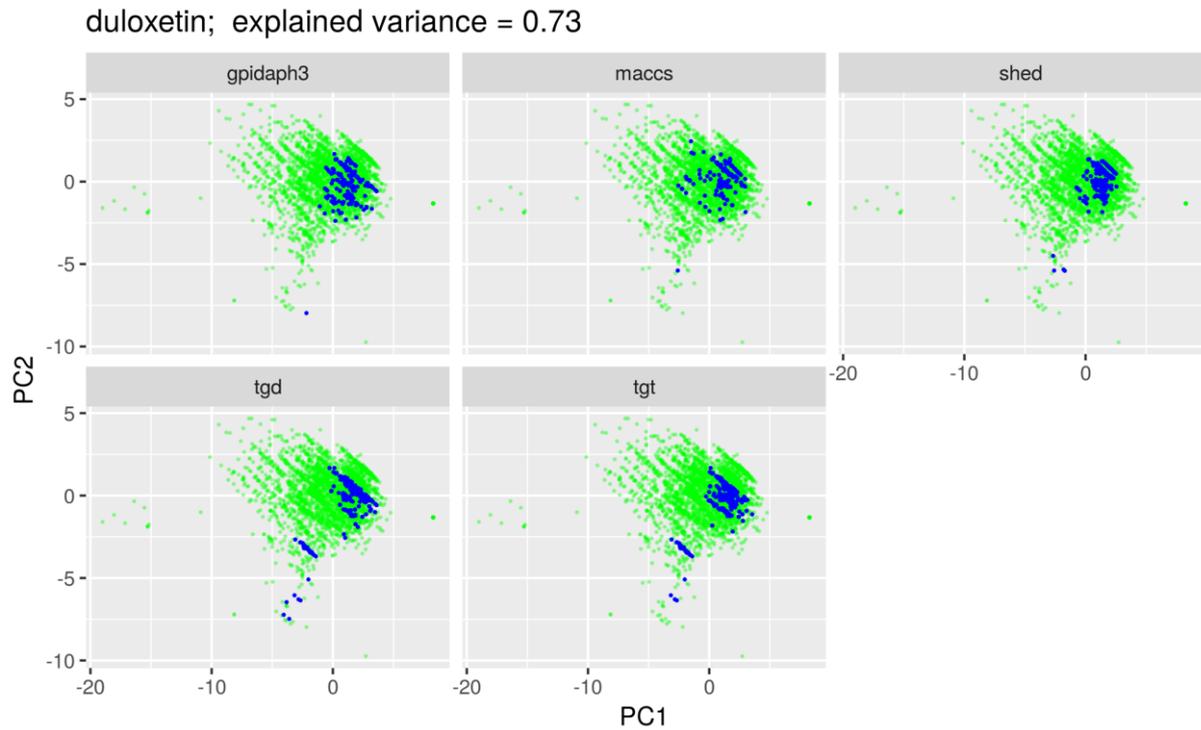


Abbildung 11 Beispiel Duloxetine – Hauptkomponentenanalyse einzelne Fingerprintwertung.

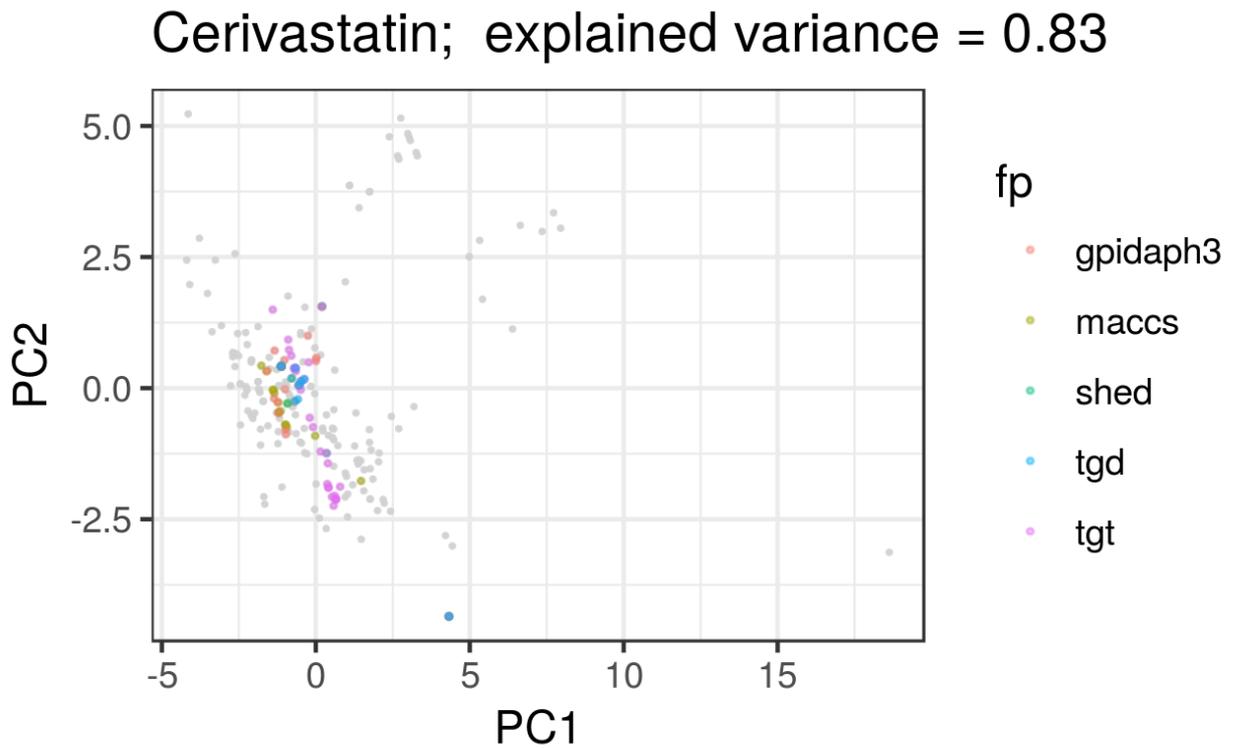


Abbildung 12 Beispiel Cerivastatin – Hauptkomponentenanalyse Gesamtwertung Fingerprints

Ritonavir; explained variance = 0.75

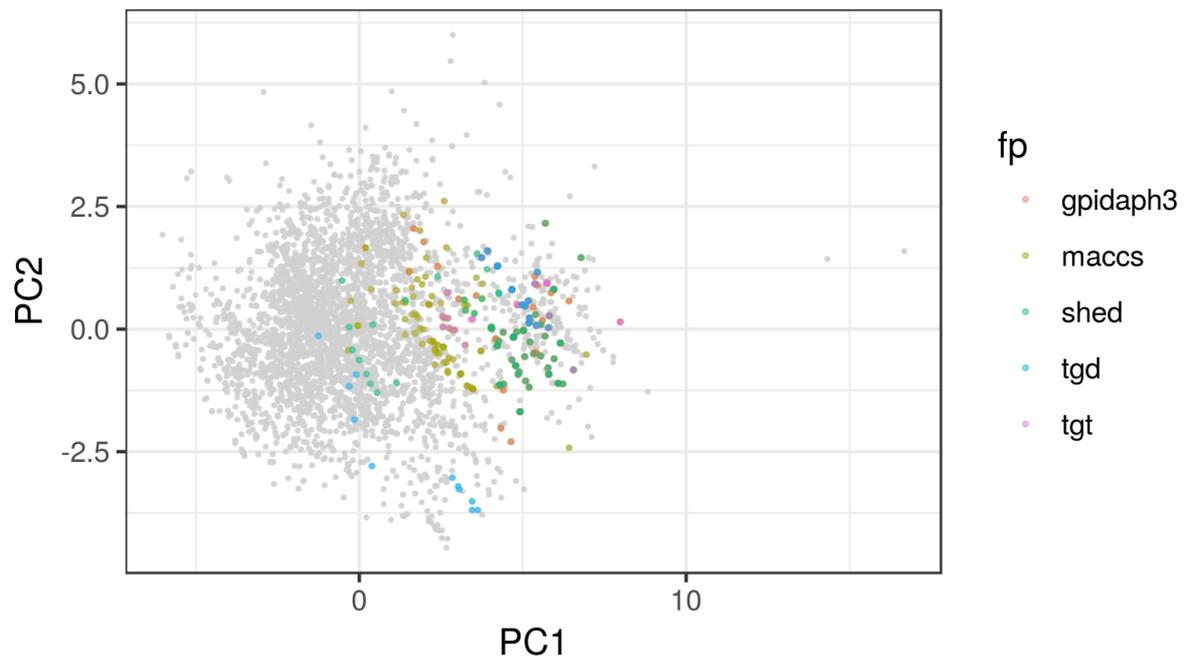


Abbildung 13 Beispiel Ritonavir – Hauptkomponentenanalyse Gesamtwertung Fingerprints

3.2 Konformationsanalyse

Um den Einfluss von Konformeren auf die vsurf – Deskriptoren abzuschätzen wurde eine Konformationsdatenbank (siehe Kapitel 2.2.2) aus 8 Arzneistoffen erstellt. Um die Deskriptorenausprägung in 2D darzustellen wurde eine PCA erstellt (Abbildung 14)- Aus der Abbildung ist ersichtlich, dass 7 Arzneistoffe einen ähnlichen physikochemischen Raum einnehmen und kein Verapamilkonformer mit anderen Arzneistoffen überlappt. Die Abbildung gibt auch einen guten Überblick wie stark die Streuung der Eigenschaften für gewisse Arzneistoffe ist.

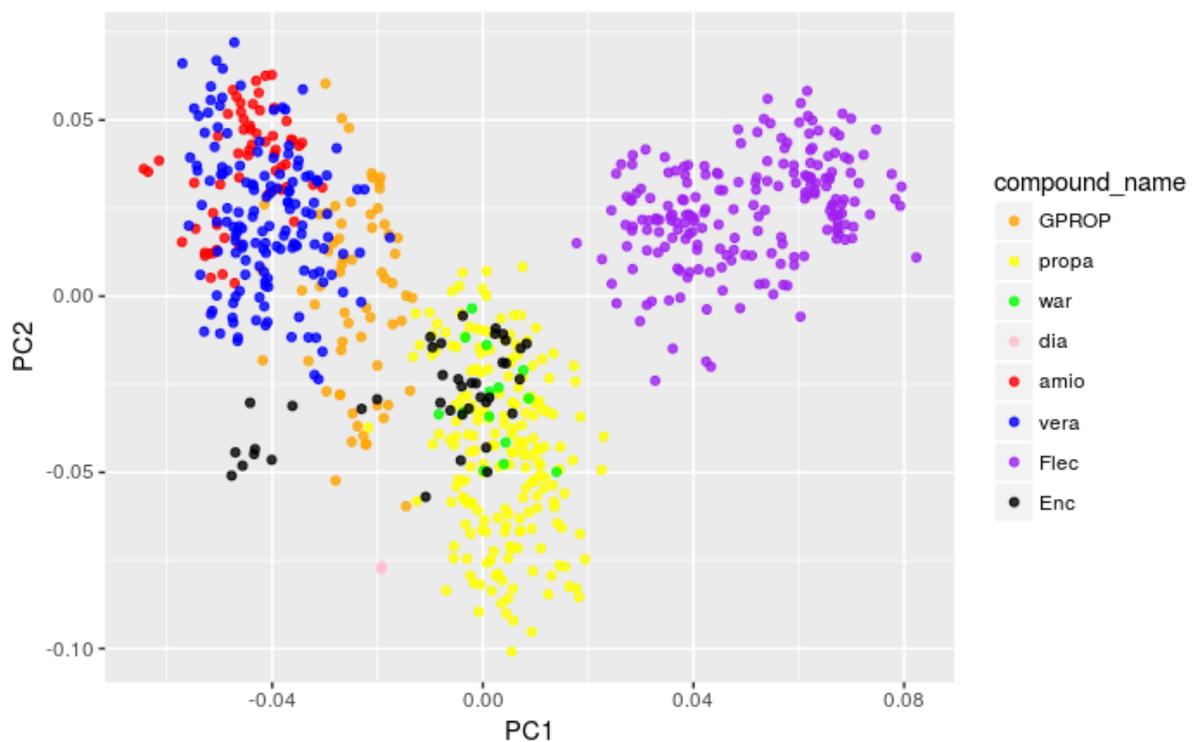


Abbildung 14:Konformerdatenbank –Hauptkomponentenanalyse. Die Kürzel in Abbildung 11 stehen für Propafenonderivat (GPROP), Propafenon (Propa), Warfarin (War), Diazepam (Dia), Amiodaron (Amio), Verapamil (Vera), Flecainid (Flec) und Encainid (Enc)

Die Ähnlichkeitssuche von Verapamil über die euklidische Distanz (basierend auf vsurf Deskriptoren) liefert ähnliche Ergebnisse. In der Abbildung 15 kann man erkennen, dass zum Beispiel nur einmal an Stelle 14 ein anderer Arzneistoff aus dem Konformerendatensatz als Verapamil als erstes gefunden wird. Nach Mittelwertberechnung von allen 145 Versuchen, kommt man zu dem Ergebnis, dass der 1. andere Arzneistoff durchschnittlich schon an 22. Stelle kommt.

Bei Betrachtung von Abbildung 16 kann man erkennen, wie viele Arzneistoffe von den ersten 100 Molekülen tatsächlich Verapamilkonformere sind. Bei 145 Versuchsreihen ergibt sich ein Mittelwert von 75.

Reihung Suchergebnisse von Verapamil (vsurf Desk.)

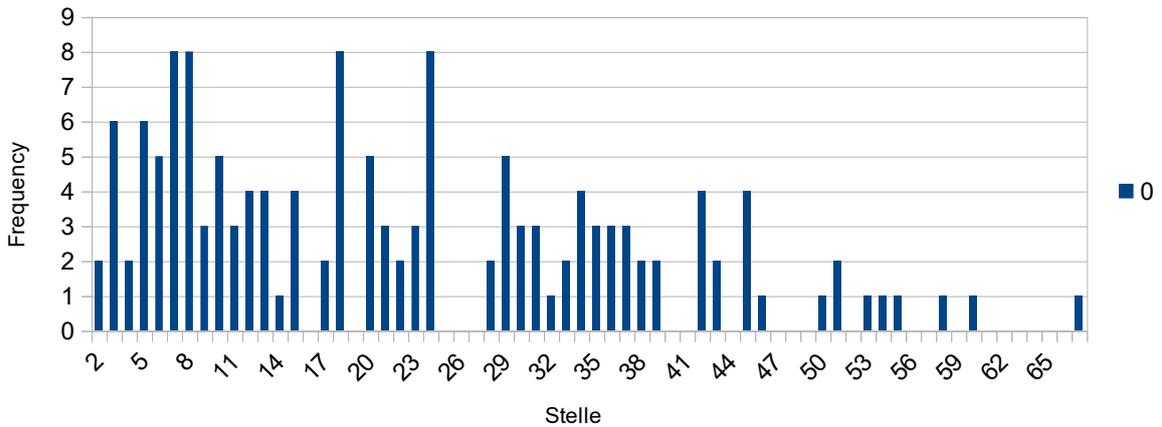


Abbildung 15: An welcher Stelle kommt wie oft der erste Arzneistoff, der kein Verapamil ist. Die x-Achse aus, an welcher Stelle das 1. „nicht Verapamil“ kommt. Die y-Achse gibt an wie häufig eine bestimmte Stelle vorkommt

Wie oft (bei 145 Versuchen) findet man eine bestimmte Anzahl an Vera?

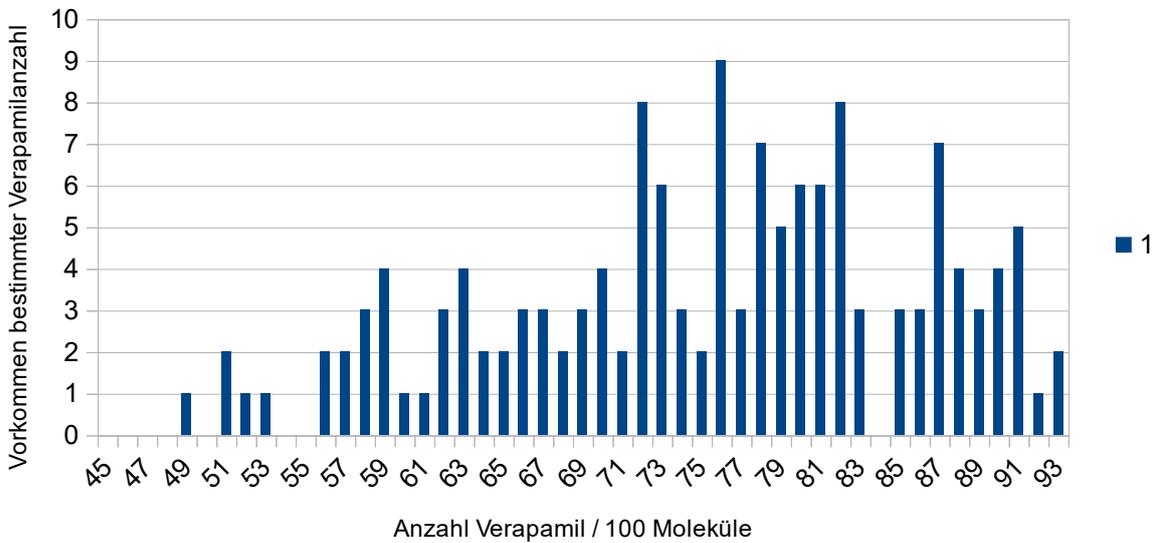


Abbildung 16: Wie oft kommt eine bestimmte Anzahl an Verapamilkonformeren von 100 möglichen Verapamilkonformeren vor? Die x-Achse die Summe der Verapamilkonformere von den ersten 100 Ergebnissen bei einer Suche nach einem bestimmten Verapamilkonformer. Die y-Achse gibt an wie oft eine bestimmte Anzahl an Verapamilkonformeren vorkommt

4 Diskussion

4.1 Ähnlichkeitssuche Fingerprints

Die im Rahmen dieser Diplomarbeit erhobenen Daten zeigen auf, dass die verschiedenen Fingerprints, verschieden gute Ergebnisse erzielen. In der Gesamtwertung steht GpiDAPH3 an der Spitze, wenn man jedoch die Einzelwertungen betrachtet, kann man erkennen, dass der MACCS – Fingerprint bei der Testreihe des Natriumkanals, sowie bei der Testreihe der CYP3A4 – Inhibitoren eine bessere Performance liefert. Allerdings ist hier die Einzelwertung von jedem der 5 Arzneistoffe je Target auch zum Teil recht unterschiedlich ausgefallen. Selbst das Gesamtergebnis aller Fingerprintsuchen zeigt auf, dass viel zu wenige Arzneistoffe, die auch tatsächlich ähnlich sind, gefunden werden.

4.2 Konformationsanalyse und Verapamilsuche

Die Analyse der Konformationsdatenbank (Kapitel 2) lässt erkennen, dass unterschiedliche Konformere deutlichen Einfluss auf die Ausprägung der vsurf – Deskriptoren haben. Durch PCA konnte diese Variabilität visualisiert werden. In der Analyse der 8 Arzneistoffe zeigten 7 überlappende Eigenschaften. Einzig Verapamil zeigte vsurf – Deskriptoren die mit keiner anderen Substanz überlappen.

Mittels Verapamilsuche konnte aufgezeigt werden, dass die generierten Konformere sich unterschiedlich verhalten. Durch die Berechnung der vsurf – Deskriptoren jedes Konformere konnte gezeigt werden, dass bei einer Ähnlichkeitssuche mittels Distanzmatrix nach einer Verapamilkonformation im Durchschnitt 25 Arzneistoffe von den ersten 100 gar keine Verapamilmoleküle sind. Das bedeutet, dass eine bestimmte Konformation von Amiodaron, die zum Beispiel an Stelle 22 kommt, ähnlichere Eigenschaften zu einer bestimmten Verapamilkonformation hat als 124 andere Verapamilkonformationen.

4.3 Conclusio

Die Ergebnisse dieser Diplomarbeit lassen erkennen, dass eine Fingerprintsuche als Ähnlichkeitssuche mit einem oder sogar mehreren Fingerprints eine dürftige Anzahl an gefunden Molekülen liefert.

Im Generellen deuten die Ergebnisse an, dass der pharmakophor – basierte Fingerprint GpiDAPH3 die beste Performance zeigt, wenn es darum geht Substanzen zu identifizieren, die ein ähnliches Wirkprofil haben wie die Referenzsubstanz. Die Erkenntnisse aus dieser Diplomarbeit liefern eine wichtige Grundlage für das Fernziel eines linear-kombinierten Ähnlichkeitsmaßes, welches über „Gewichtung“ verschiedene Fingerprint Distanzen kombiniert. Somit wäre es vielleicht möglich, durch eine Kombination mehrerer Methoden der Ähnlichkeitssuche ein besseres Ergebnis zu liefern.

4.4 Appendix – Scripts

4.4.1 // Script to calculate partial charges

```
1 #svl
2
3 #set main batch_partial_charge
4
5 function PartialChargeMDB, db_ImportSD, QuaSAR_DescriptorMDB;
6
7 function batch_partial_charge sd_file
8
9     local basename_token = fbase sd_file;
10    local mdb_file = tok_cat[basename_token, '.mdb'];
11    db_Open[mdb_file, 'create'];
12
13    print mdb_file;
14    db_ImportSD [ mdb_file,
15                  sd_file, 'mol',
16                  [],
17                  [],
18                  [],
19                  [append:1, add_hydrogens:0, start_entry:1, end_entry:[], file_field:1, no_fields:0, scan_file:[ ] ]
20                ];
21
22    PartialChargeMDB[mdb_file, 'PEOE', 'mol', 'mol', [esel: 0, add_h: 1]];
23    QuaSAR_DescriptorMDB [mdb_file, 'mol', ['vsurf_A', 'vsurf_CP', 'vsurf_CW1', 'vsurf_CW2'], [esel:0, verbose:0]];
24
25 endfunction
```

4.4.2 // Script to calculate the Tanimoto Similarity

Dieses Skript wurde in MOE (16) integriert um die Ergebnisse der Fingerprintsuche zu bewerten. Jede Verbindung in der ChEMBL – Datenbank erhielt damit eine Tanimotowertung zwischen 0 und 1 und konnte dann nach der höchsten Übereinstimmung mit dem Referenzmolekül sortiert werden.

```
1 #svl
2
3 #set main 'sim_calc'
4
5 function ph4_SimilarityMDB;
6
7 const panel = [
8     title : 'Fingerprint selection',      // shell title
9     Text : [
10         name : 'fp',                    // text field identifier
11         title: 'Enter Fingerprint (press <RETURN>)',
12         type : 'character',             // expected entered data type
13         onTrigger : 'return'           // make it a trigger object
14     ]
15 ];
16
17 global function makewin []
18     local values = WindowPrompt [ panel, [fp:"FP:NAME"] ];
19                                     // initial Text value 12
20     return values;
21 endfunction
22
23 function sim_calc []
24
25 local db = dbv_DefaultView[];
26 local entries = db_Entries db;
27 local fp_code = makewin[];
28 local out_field = tok_cat ['tanimoto_', fp_code.fp];
29 db_EnsureField[db, out_field, 'float'];
30
31 Warning 'The first entry in the database \n will be taken as the reference molecule';
32 local fp_query = cat db_ReadFields[db, entries(1), fp_code.fp];
33
34 print fp_query;
35
36 local tanimoto_sims = ph4_SimilarityMDB [db, entries, fp_code.fp, 'tanimoto', fp_query];
37
38 local i = 1;
39 local n = length entries;
40
41 for i = 1, n, 1 loop
42
43     db_Write[db, entries(i), tag[out_field, tanimoto_sims(i)]];
44
45 endloop
46
47 endfunction
```

4.1 Skript für den Tanimotowert

5 Literaturverzeichnis

1. **Bajorath, Jürgen.** *Cheminformatics: Concepts, Methods, and Tools for Drug Discovery*. Totowa, NJ : Humana Press, 2004. 978-1-59259-802-1.
2. **Maggiora, Gerald, et al.** Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry*. 57, 2014, Bd. 8, S. 3186–3204.
3. **Lajiness, Michael S, Maggiora, Gerald M und Shanmugasundaram, Veerabahu.** Assessment of the consistency of medicinal chemists in reviewing sets of compounds. *Journal of medicinal chemistry*. 47, 2004, Bd. 20, S. 4891-6.
4. *Molecular fingerprint similarity search in virtual screening*. **Cereto-Massagué, Adrià, et al.** 2015, *Methods*, Bd. 71, S. 58-63. ISSN: 1046-2023.
5. *How diverse are diversity assessment methods? A comparative analysis and benchmarking of molecular descriptor space*. **Koutsoukas, Alexios, et al.** 2014, *Journal of chemical information and modeling*, Bd. 54(1), S. 230-42.
6. *Reduction and Recombination of Fingerprints of Different Design Increase Compound Recall and the Structural Diversity of Hits*. **Nisius, Britta und Bajorath, Jürgen.** Oxford : s.n., 2010, *Chemical Biology & Drug Design* , Bd. 75(2), S. 152-160. ISSN: 1747-0277.
7. *Shannon entropy-based fingerprint similarity search strategy*. **Wang, Yuan, Geppert, Hanna und Bajorath, Jürgen.** 2009, *Journal of chemical information and modeling*, Bd. 49(7), S. 1687-91. ISSN: 1549-9596.
8. ChemWiki. *ChemWiki*. [Online] [Zitat vom: 14. März 2019.] <http://www.ch.ic.ac.uk/local/organic/conf/2.svg>.
9. **Mortimer, Charles E., Müller, Ulrich [VerfasserIn] und Beck, Johannes [Bearb.]**. *Chemie*. 11., vollst. überarb. Aufl. Stuttgart [u.a.] : Thieme. ISBN: 9783134843118; 9783131713315.
10. *Toward accurate relative energy predictions of the bioactive conformation of drugs*. **Butler, Keith T., Luque, F. Javier und Barril, Xavier.** 2009, *Journal of computational chemistry*, Bd. 30(4), S. 601-10.
11. *Molecular fields in quantitative structure—permeation relationships: the VolSurf approach*. **G, Cruciani, et al.** [Hrsg.] Elsevier. 1-2, Italien; Schweiz : *Journal of Molecular Structure: THEOCHEM*, 2000, Bd. 503, S. 17-30. Cruciani2000.
12. *Efficient similarity search in sequence databases*. **Agrawal, Rakesh, Faloutsos, Christos und Swami, Arun.** [Hrsg.] David B. Lomet. Berlin : Springer Berlin Heidelberg, 1993. *Foundations of Data Organization and Algorithms*. S. 69-84. ISBN: 978-3-540-48047-1.
13. *Principal component analysis*. **Bro, Rasmus und Smilde, Age K.** 9, s.l. : The Royal Society of Chemistry, 2014, *Anal. Methods*, Bd. 6, S. 2812-2831.
14. *The ChEMBL database in 2017*. **Gaulton, Anna, et al.** D1, England : s.n., 1 2017, *Nucleic acids research*, Vol. 45, pp. D945-D954.
15. *SHED: Shannon entropy descriptors from topological feature distributions*. **Gregori-Puigjané, Elisabet und Mestres, Jordi.** 2006, *Journal of Chemical Information and Modeling*, Bd. 46, S. 1615-1622. ISSN: 15499596.
16. Chemical Computing Group (CCG) | Research. [Online] [Zitat vom: 14. März 2019.] https://www.chemcomp.com/Research-Citing_MOE.html.