



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

New natural GPBAR-1 modulators
identified by combined in silico
- in vitro screening

verfasst von / submitted by

Benjamin Kirchweger

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magister der Pharmazie (Mag.pharm.)

Wien, 2017 / Vienna, 2017

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:

Pharmazie

Betreut von / Supervisor:

Univ.-Prof. Dr. Judith Maria Rollinger

“The main fuel to speed the world’s progress is our stock of knowledge, and the brake is our lack of imagination.”

Erik Brynjolfsson

Acknowledgments

First of all, I have to thank Judith Maria Rollinger. Her overview on the ongoing research projects, her encouragement, expertise and ambitious guidance made it a great privilege and joy to work under her supervision. I am very thankful for her trust and efforts in this project.

Secondly, I want to thank my first supervisor Jadel Müller Kratz, who supervised me during the first phase of my thesis, the modeling and writing. With his great expertise and his easy-going and nice personality he gave me perfect guidance and I am really honored having worked with him. I am also very grateful that Ulrike Grienke kindly filled in for Jadel after several months, always having an open ear for my questions and concerns.

Special thanks also to Dr. Angela Ladurner for her introduction and help at the cell culture laboratory, for her open ears for every question and her patient teaching. I am very thankful for all her efforts she had put into this work.

I am very grateful to Prof. Dirsch for giving me the opportunity to perform the experimental validation at her laboratory, to Prof. Langer for his valuable advice and for supporting me with licenses and knowledge in pharmacoinformatics.

I also want to express my thanks to Christine Elisabeth Maier, for her introduction to the phytochemistry laboratory and for her fancy triterpenes, Martin Groiss for helping me dealing with the customs, to Simone Latkolik, Verena Hiebl and Hortenzia Beres for their help at the cell culture, to Judith Fillip for her assistance at the substance database, and to Dr. Thomas Seidel for sharing his forthcoming manuscript. Further I thank all people, friends and colleagues at the department of pharmacognosy.

Last but most importantly I have to thank my friends and family, my sisters Christine and Katharina, my brother Valentin and my parents for their unceasing support all through this work, my studies and my life. Without them warranting me offset, backing me up financially and emotionally and encouraging me I would have not been able to do so.

Table of Contents

Table of Contents	1
1. Abbreviations:	4
2. Abstract.....	7
3. Zusammenfassung	9
4. Introduction	11
4.1. G protein-coupled bile acid receptor (GPBAR1)	11
4.1.1. The physiological role of GPBAR1	13
4.1.2. Endogenous ligands.....	18
4.1.3. GPBAR1 relevance in diseases and as drug target.....	19
4.1.4. Preclinical methods used for the evaluation of GPBAR1 modulation	22
4.1.5. Small molecule modulators of GPBAR1	24
4.1.6. Summary.....	33
4.2. Natural products and computational studies:	
A multidisciplinary approach in drug discovery	33
4.2.1. Natural products.....	33
4.2.2. Multidisciplinary approaches	36
4.2.3. Pharmacophore.....	38
4.2.4. Pharmacophore based virtual screening	43
5. Results and Discussion	46
5.1. The Workflow	46
5.2. Benchmark establishment - Classification of agonists from the literature.....	48
5.3. Benchmark establishment - Generation of decoys from DUDe	51
5.4. Conformational sampling.....	55
5.5. Theoretical model performance	55
5.5.1. The initial model collection	55
5.5.2. Theoretical performance of the models	60
5.6. Detailed model description	63
5.6.1. Cluster 7- Generation of BAMS22	63
5.6.2. Cluster 12- Generation of Model 12.3.8.....	69
5.7. Results of the prospective virtual screening	73
5.8. Results of hierarchical cluster analysis (HCA)	74
5.9. Hit selection	78
5.9.1. Hit selection for HCA group 1	78

5.9.2.	Hit selection for HCA group 2	79
5.9.3.	Hit selection for other HCA groups	81
5.9.4.	Additional compounds for bioactivity testing	83
5.10.	Results of the experimental validation	84
5.10.1.	Results on the experimental validation of group 1	87
5.10.2.	Determination of EC ₅₀ -values for microlobiden and farnesiferol B	88
5.10.3.	Results on the experimental validation of group 2	91
5.10.4.	Results on the experimental validation of groups 4, 5 and 6.....	92
5.10.5.	Results on the experimental validation of groups 7, 8 and 9.....	94
5.10.6.	Activity of extracts	95
5.10.7.	Results on the experimental testing of plant A's triterpenes	95
5.10.8.	Summary.....	96
6.	Materials and Methods	97
6.1.	Hardware information	97
6.2.	Software information	97
6.2.1.	Ligand Scout.....	97
6.2.2.	I:Con	98
6.2.3.	Marvin Sketch	98
6.2.4.	ChemGPS.....	98
6.2.5.	SIMCA	98
6.2.6.	ROCS	99
6.3.	Database sources	99
6.3.1.	External sources	99
6.3.2.	Internal data	100
6.4.	Database creation	100
6.5.	Pharmacophore modeling	108
6.5.1.	Clustering.....	108
6.5.2.	Training sets for modeling	111
6.5.3.	Prospective virtual screening.....	118
6.5.4.	Hit list validation	120
6.6.	Extract preparation	122
6.6.1.	<i>Phytolacca americana</i> root extract	122
6.6.2.	<i>Ferula assa-foetida</i> gum resign extracts.....	122
6.7.	Luciferase based reporter gene assay	123
6.7.1.	Chemicals	123
6.7.2.	Cell culture reagents	124

6.7.3. Cell-based luciferase reporter gene assay	125
6.7.4. Statistical Analysis	128
7. Summary and Conclusion.....	129
8. References.....	131
9. Appendix	151

1. Abbreviations:

11 β -HSD1	11 β -Hydroxysteroid dehydrogenase type 1
6 α -ECDCA	6 α -Ethylchenodeoxycholic acid
AChE	Acetylcholinesterase
AMP	Adenosine monophosphate
AUC	Area under the curve
BA	Bile acid
BAMS	Bile acid model shared features
CA	Cholic Acid
cAMP	Cyclic adenosine monophosphate
CAR	Constitutive androstane receptor
CAS	Chemical abstract services
CD95	Cluster of differentiation 95
CDCA	Chenodeoxycholic Acid
cDNA	Complementary deoxyribonucleic acid
CFTR	Cystic Fibrosis transmembrane conductance regulator
CHO	Chinese hamster ovary
cLogP	Concentration log partition-coefficient
CNS	Central nervous system
CoA	Coenzyme A
CRE	cAMP response element
CYP	Cytochrome P450
DC	Dendritic cell
DCA	Deoxycholic Acid
DCM	Dichlormethane
ddH ₂ O	Bidistilled water
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulfoxid
DTT	Dithiothreitol
EC50	Half maximal effective concentration
ECL	Extracellular loop
EDTA	Ethylenediaminetetraacetic acid
EF	Enrichment factor
EGFP	Enhanced green fluorescent protein
Epac	Exchange Protein directly activated by cAMP
FBS	Fetal bovine serum
FN	False negative
FP	False positive
FXR	Farnesoid X receptor
GFP	Green fluorescent protein
GLP-1	Glucagon like peptide 1
GPBAR1	G protein coupled bile acid receptor
GPCR	G protein-coupled receptor
GPCR19	G protein coupled receptor 19
GPR131	G protein coupled receptor 131
G-protein	Guanine nucleotide binding protein
GSK	GlaxoSmithKline
GTM	Generative topographic mapping
HBA	Hydrogen-bond acceptor
HBD	Hydrogen-bond donor
HBS	HEPES buffered saline
HCA	Hierarchical cluster analysis
HEK	Human embryonic kidney
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HERG	Human ether-à-go-go related gene

HTRF	High throughput resonance energy transfer
HUGO	Human genome organization
ICAM	Intercellular Cell Adhesion Molecule
ICL	Intracellular loop
IL	Interleukin
IP	Intellectual property
IUPAC	International union of pure and applied chemistry
KC	Kupffer cell
LCA	Lithocholic Acid
LDL	Low density lipoprotein
Luc	Luciferase
MAP	Mitogen-activated protein
M-BAR	Membrane-type receptor for bile acids
MOE	Molecular operating environment
MW	Molecular weight
NAFLD	Non-alcoholic fatty liver disease
NASH	Non alcoholic steatohepatitis
NCI	National cancer institute
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NO	Nitric oxide
NOS	Nitric oxide synthase
NP	Natural product
NS	Nervous system
OCA	Obeticholic acid
PAIN	Pan-assay interfering substances
PBC	Primary biliary cirrhosis
PBS	Phosphat buffer saline
PC	Prohormone kinase
PC	Principal component
PCA	Principal component analysis
PCA	Principal component analysis
PEG	Polyethylene glycol
Pen/Str	Penicillin G/ Streptomycin
PLC-ϵ	Phospholipase C- ϵ
PPAR	Peroxisome proliferator-activated receptor
PPARγ	Peroxisome proliferator-activated receptor γ
PSC	Primary sclerosing cholangitis
PXR	Pregnane X receptor
PYY	Peptide-Tyrosine-Tyrosine
QSAR	Quantitative structure activity relationship
RFU	Relative fluorescence unit
RLU	Relative luminescence unit
RLU/s	Relative luminescence unit per second
ROC	Receiver operating characteristic
Rup 43	Rumen undegradable protein
Se	Selectivity
SEC	Sinusoidal endothelial cell
Sp	Specificity
STAT3	Signal transducer and activator of transcription 3
SXR	Steroid and xenobiotic receptor
T/E	Trypsin/EDTA
T2D	Type 2 Diabetes
TC	Tanimoto combo
TCM	Traditional chinese medicine
TGR5	Takeda G-protein-coupled receptor 5
TLC	Taurolithocholic Acid
TMH	Transmembrane helices
TN	True negative
TNF	Tumor necrosis factor

TP	True positive
TPSA	Topological polar surface area
TR- FRET	Time resolved fluorescence resonance energy transfer
VCAM	Vascular Cell Adhesion Molecule
VDR	Vitamin D receptor
VS	Virtual screening
WHO	World Health Organization

2. Abstract

Since the discovery of the G protein coupled bile acid receptor (GPBAR1 or Takeda G-protein-coupled receptor 5 (TGR5)) 15 years ago, it has been recognized as a promising new target for the treatment of diverse metabolic diseases, including obesity, type 2 diabetes, fatty liver disease and atherosclerosis. Although, the potential use of GPBAR1 agonists seems to be promising, several classical preclinical hurdles have also been encountered in this field of research. The plethora of biological functions appears to be an obvious opportunity, but a major drawback at the same time. The need for novel and potent agonists is highly relevant, as metabolic diseases are on the rise and pharmacological unmet therapeutic needs are pervasive.

Natural product (NP) research has always been a promising approach to find new drug-candidates, but the traditional NP research is addressed as cost and time-intensive. That led to its recent downfall in pharmaceutical industry, although more than 50% of all approved drugs are derived from NP. Therefore, new concepts in the search for bioactive natural compounds have been piloted and this thesis is dedicated to one of them: The implication of in-silico methods to the straight-forward discovery of NP actives for the exemplified and notable GPBAR1 target.

A protocol was set up, starting with the creation of a comprehensive collection of structural information on known ligands, originated from different publications, patents and in-house data. This information was further used to create a selection of ligand-based pharmacophore models in LigandScout 4.0, which intended to comprise the whole structural diversity of yet known agonists. After theoretical validation, the two most promising models, namely BAMS22 and Model 12.3.8 were employed as virtual screening tools.

Several NP databases, beside the drugbank and SPECS database were evaluated and the received hits were forwarded to shape-matching experiments in ROCS 3.1.2.4. Secondly, the hitlist was subjected to principal component analysis with the aim to cluster them in 9 groups based on their physicochemical differences.

Candidates from every group were selected based on shape-matching and pharmacophore fit scores, novelty, availability and structural variance. Further, plant extracts, containing virtual hits were selected or generated using a standard protocol.

All 35 selected hits and three extracts were tested at least at two different concentrations in a reporter-gene based luciferase assay. The used HEK 293T cells were transfected with plasmids for the expression of GPBAR1 and a cAMP response element luciferase construct, and incubated together with compound dilutions. *In vitro* results were read out after 18 hours. The biological evaluation led to the identification of 17 new agonists with scaffolds ranging from benzothiazols to steroids, triterpenes and coumarins.

For two agonists, namely microlobiden and farnesiferol B, a concentration curve was evaluated, since their unique sesquiterpene-coumarin-scaffold is yet unknown to activate bile-acid receptors. The investigations concluded an EC_{50} of 14.64 μ M for microlobiden and a 19.16 μ M for farnesiferol B, respectively.

In conclusion, the two models have proven to be effective virtual screening queries able to identify hits aside from yet known and evaluated scaffolds. The established protocol proved to be a proficient approach for further NP drug discovery projects.

3. Zusammenfassung

Als der G-Protein gekoppelte Gallensäurerezeptor 1 (GPBAR1) vor 15 Jahren entdeckt wurde, erkannte man ihn bald als verheißungsvolles Ziel für neue Medikamente. Ihn zu aktivieren könnte zukünftig Relevanz in der Therapie von verschiedenen Stoffwechselerkrankungen, darunter Typ 2 Diabetes, Fettleber und Arteriosklerose haben. Zwar klingen die nützlichen Auswirkungen seiner Aktivierung sehr vielversprechend, dennoch wurde man auch mit einigen typischen präklinischen Hürden konfrontiert. Die Vielfalt seiner biologischen Funktionen erscheint offensichtlich auf der einen Seite große Chancen für die Therapie einer Vielzahl von pathologischen Zuständen zu bieten, aber andererseits könnte sich gerade diese Vielfalt auch als Nachteil entpuppen. Dennoch, Stoffwechselkrankheiten nehmen weltweit stark zu, ihre pharmakologische Therapie ist nicht lückenlos möglich und die Suche nach neuen, potenten Aktivatoren für den GPBAR1 erscheint in diesem Zusammenhang als höchst relevant.

Die Erforschung von Naturprodukten war immer schon ein von Erfolg geprägter Ansatz um neue Arzneikandidaten zu finden, aber ihre klassische Erforschung wird heute von vielen als zu kosten- und zu zeitintensiv betrachtet. Diese Sichtweise führte auch zum Rückgang ihrer Stellung in der pharmazeutischen Industrie, ein schwerer Irrtum, da mehr als die Hälfte aller zugelassenen Arzneistoffe ihren Ursprung oder Inspiration in der Natur hatten. Es werden daher neue Konzepte in der Naturstoffforschung gesucht und erprobt. Diese Arbeit kann als eine davon betrachtet werden: Die Einführung von computerbasierten Methoden um die Forschungskapazitäten zu fokussieren. Diese Arbeit kann also als multidisziplinärer Ansatz verstanden werden, um neue Liganden im Bereich der Naturstoffe für das mustergültige Ziel des GPBAR1 zu finden.

Als erster Punkt des Protokolls wurde eine umfassende Datenbank der strukturellen Information bereits bekannter Liganden erstellt. Grundlage dafür waren Publikationen, Patente und hauseigene Daten vorhergehender Projekte. Die Informationen wurden benutzt um ein Set von ligandenbasierten Pharmacophormodellen mit LigandScout 4.0 zu erstellen, welches zum Ziel hatte möglichst die gesamte strukturelle Diversität aller Liganden zu erklären. Die nach der theoretischen Validierung am Erfolg versprechendsten Modelle BAMS22 und Model 12.3.8. wurden als Filter für ein virtuelles Screening verwendet.

Es wurden mehrere virtuelle Naturstoffdatenbanken, aber auch die SPECS- und die drugbank Datenbank gescreened. Alle erhaltenen Treffer wurden zu einem virtuellen Formvergleichs-Screening mit ROCS 3.1.2.4. weitergeleitet. Weiters wurde die gesamte Trefferliste einer

Principal component analysis unterzogen, um sie basierend auf ihren physikalisch-chemischen Eigenschaften in 9 Gruppen aufzuteilen.

Soweit es die finanziellen Möglichkeiten gestatteten, wurden Treffer aus jeder der 9 Gruppen für Bioaktivitätsmessungen nominiert. Kriterien waren die Punktezahl des Formvergleichs und des pharmacophore-fits, Neuheit, Verfügbarkeit und strukturelle Vielfalt. Ebenfalls wurden Pflanzenextrakte hergestellt oder besorgt, welche mehrere Inhaltsstoffe als virtuelle Treffer beinhalteten.

Alle 35 ausgewählten Substanzen, sowie 3 Extrakte, wurden mindestens in 2 verschiedenen Konzentrationen in einem Reporter-gen basierenden Luciferase-Assay getestet. Die dabei benutzten HEK 293T Zellen wurden mit Plasmiden für die Expression von GPBAR1 und einem cAMP response element Luciferase Konstrukt transfiziert. Zusammen mit den Substanzverdünnungen wurden sie für 18 Stunden inkubiert und danach die Resultate ausgewertet. Es konnten so 17 neue Agonisten mit unterschiedlichsten Strukturgerüsten, von Benzothiazolen, zu Steroiden, Triterpenen und Cumarinen identifiziert werden.

Für zwei Agonisten, Farnesiferol B und Microlobiden, wurde eine Dosis-Wirkungs Kurve evaluiert, weil ihre spezielle Sesquiterpen-Cumarin Struktur bisher unbekannt war Gallensäurerezeptoren zu aktivieren. Es wurde ein EC50 von 14.64 μM für Microlobiden und 19.16 μM für Farnesiferol B ermittelt.

Zusammenfassend kann gesagt werden, dass beide Modelle (BAMS22 und Model 12.3.8.) effektive Werkzeuge für Virtual Screening sind, sie neue Liganden auch fernab bereits bekannter Strukturgerüste identifiziert haben. Das hierbei aufgestellte Protokoll stellt sich als fähiger Ansatz für weitere Projekte heraus.

4. Introduction

In the last 17 years the knowledge on the physiological role of bile acids (BAs) has considerably expanded. Traditionally, BAs were understood as simple emulsifying agents, acting in the solubilization of cholesterol, fatty acids, and liposoluble vitamins, and facilitating their absorption in the intestines¹. Only recently, the further activity of BAs as hormones and signaling molecules, playing a significant role in various physiological pathways, has been described². And not long after that, aroused the intention of developing novel drugs interacting in these pathways giving hope for the treatment of diverse diseases³.

4.1. G protein-coupled bile acid receptor (GPBAR1)

The radical change of view on the physiological role of BAs started in 1999 with the identification of the first receptor for BAs, the nuclear Farnesoid X Receptor (FXR)⁴. Only three years later, in 2002, Maruyama and coworkers described the first membrane G protein-coupled receptor specific for BAs, the G protein-coupled bile acid receptor 1 (GPBAR1)⁵. This receptor is also known as TGR5, GPCR19, GPR131, M-BAR, and Rup 43^{6,7}. Although, the TGR5 nomenclature is prevailed in the literature, during the course of this thesis the nomenclature GPBAR1 is used since it is recommended by the HUGO Gene Nomenclature Committee⁷.

GPBAR1 is a member of the G protein-coupled receptor (GPCR) superfamily. The gene is located at the chromosomal region 2q35 and the protein's encoding sequence consists of 993 base pairs resulting in a protein with 330 amino acids.

The cDNA sequence of the human GPBAR1 shows high conservation (>80% homology) with that of other mammalian species like bovine, rabbit or mouse, indicating an important role in mammalian physiology⁶.

Similar to other members of the GPCR superfamily, the structure includes seven-transmembrane domains, but has not yet been crystalized. It is further predicted to have three intracellular loops and three extracellular loops. While the extracellular loops may play an important role in the interactions with ligands, the intracellular domains mediate signal transmission to cellular signaling molecules⁸.

Since no crystal structure is available, several homology models had been published⁸⁻¹¹. **Figure 1** shows one of these approaches published in 2010⁸. According to these publications the

binding pocket of the BA endogenous ligands is likely to be located between the transmembrane helices (TMHs) ⁹⁻¹¹.

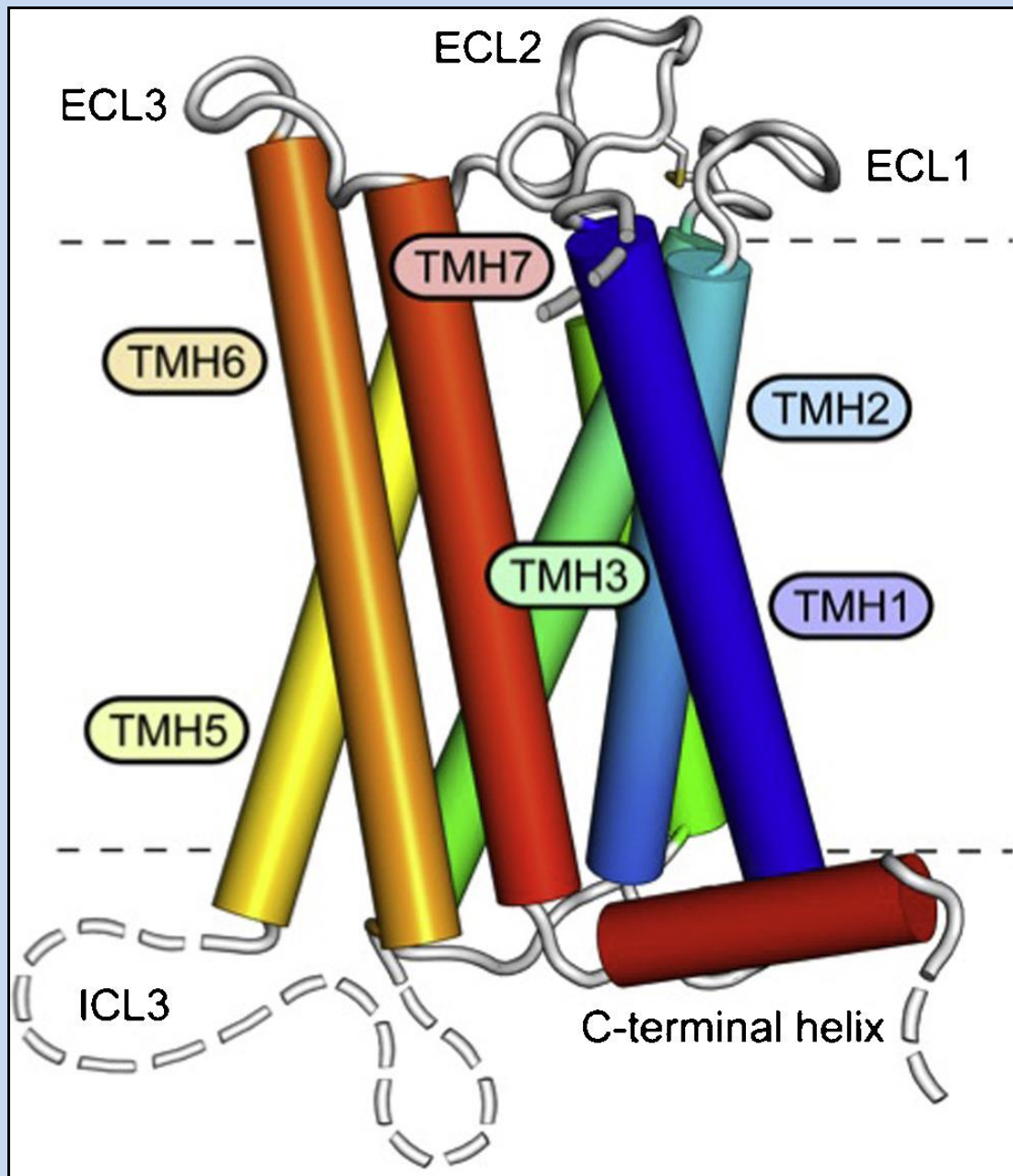


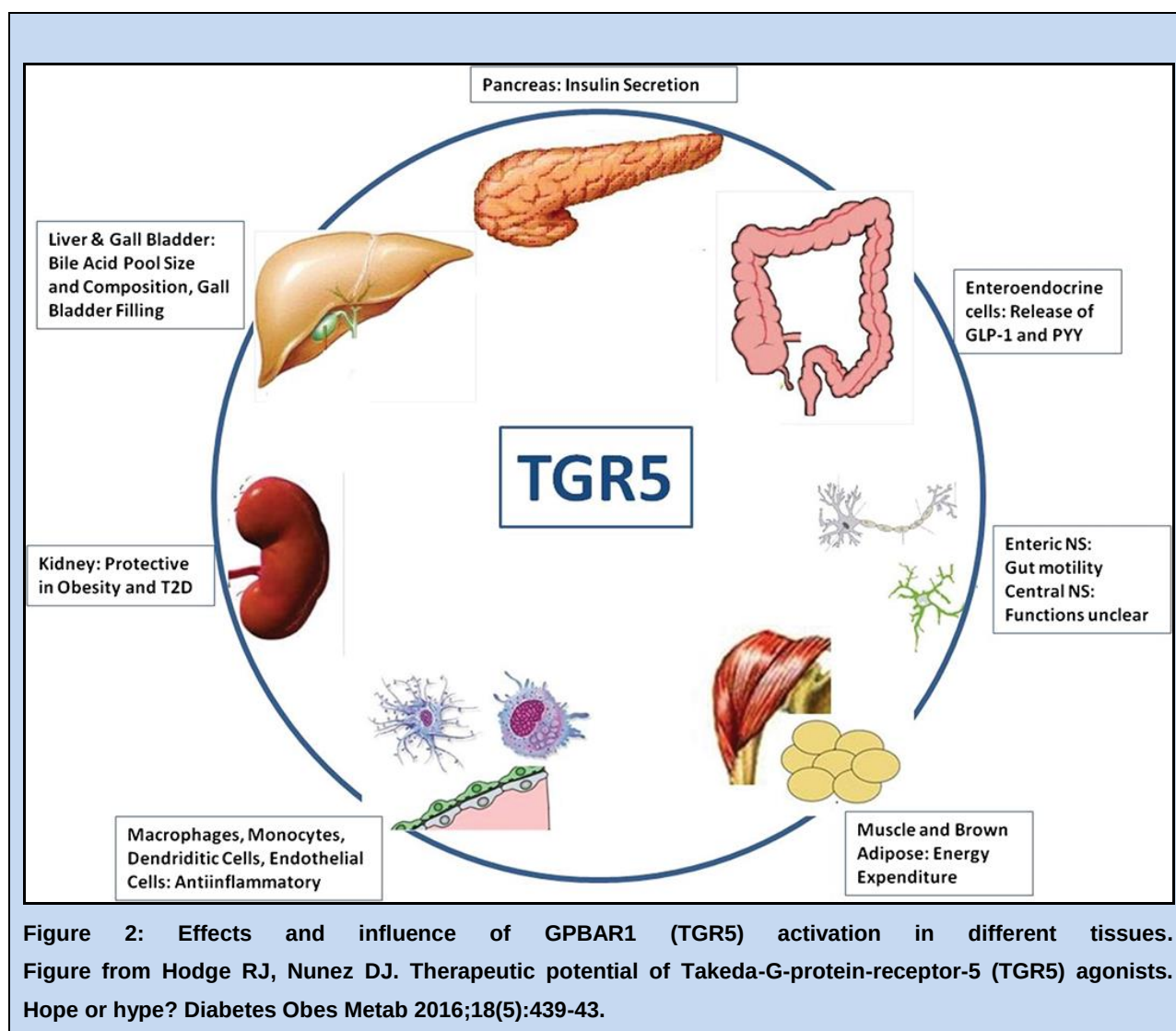
Figure 1: Predicted structure of GPBAR1: The structure consists of seven-transmembrane helices (TMH1-TMH7) with 3 extracellular loops (ECL1- 3) and 3 intracellular loops (ICL1- 3). The C- and N- terminal intracellular sections are like the ICL3 flexible and disordered (illustrated by dashed grey lines). According to the work of Macciarulo and coworkers¹⁰ and Gertzen and coworkers ⁹ the binding pocket is likely located between the TMHs.

Figure from PLOS one, Hov et al., Mutational characterization of the bile acid receptor TGR5 in primary sclerosing cholangitis. PLoS ONE. 2010;5(8):e12403 ⁸

4.1.1. The physiological role of GPBAR1

GPBAR1 is ubiquitously expressed in tissues and therefore has a vast influence on several physiological pathways, most important metabolic and anti-inflammatory ones. Expression in intestine, pancreas, liver, gallbladder, striated muscles, heart, spleen, kidney, brown adipose tissue, prostate, uterus, placenta, testis, mammary glands, macrophages, endothelial cells and certain areas of the central nervous system have been reported^{3,8,12-15}.

The activation of the receptor leads to a tissue-specific rise of intracellular cyclic adenosine monophosphate (cAMP), affecting different signaling pathways in the respective cells, resulting in multiple biological effects¹⁵, which are illustrated in **Figure 2**.



The most fundamental physiological role in the context of this thesis is the function as metabolic regulator with the two central actors, the intestine and the pancreas, in parallel with the effects on regulation of energy balance in muscles and brown adipocytes^{2,3,16}. These effects together make it an appealing target for metabolic disorders¹⁷.

4.1.1.1. Energy homeostasis and metabolic effects

BAs are key players in metabolic regulation and have a vast effect on energy homeostasis. One part is connected to the modulation of the nuclear receptor FXR and the other, which is getting more and more attention, to that of GPBAR1¹⁸.

In enteroendocrine L-cells of the gut, GPBAR1 modulation leads over guanine nucleotide binding protein subunit alpha (G-protein_{αs}) on a adenylatcyclase stimulating/ cAMP based pathway, involving a set of intracellular signaling molecules, namely Epac/PLC- ϵ /Ca²⁺, to a phosphatidylinositol-hydrolysis dependent increase of both GLP-1 and PYY secretion¹⁹. Firstly, peptide-tyrosine-tyrosine (PYY) secretion promotes appetite reduction, and therefore, trims food intake down. Secondly and more importantly, glucagon like peptide 1 (GLP-1) influences glucose homeostasis both through stimulation of insulin and inhibition of glucagon secretion in the pancreas. GLP-1 downregulates gut motility and shows appetite reducing effects similar to PYY²⁰. Independent of PYY and GLP-1 effects, activation of pancreatic GPBAR1 leads to a rise in insulin secretion in β -cells¹⁸ and to a switch via prohormone kinase (PC) from glucagon to GLP1 release in α -cells. This GLP-1 release affects the β -cells again and leads to insulin production and reinstates β -cell mass. Nevertheless, the complex function of GPBAR1 in the pancreas and the cross talk between α - and β - cells is not fully understood yet¹⁴.

BAs' physiological role via GPBAR1 modulation is also part of a specific energy expenditure adjustment in the body. In brown adipose tissue and skeletal myocytes, GPBAR1 controls the activation of the enzyme iodothyronine deiodinase and, consequentially, the conversion of inactive thyroxine into the active tri-iodothyronine. The result of GPBAR1 modulation is therefore higher net energy expenditure²¹.

4.1.1.2. Hepatic effects

GPBAR1 is not expressed in hepatocytes, but was found in sinusoidal endothelial cells (SECs)²², in Kupffer cells (KCs)²³, in cholangiocytes, in gallbladder smooth muscle cells, and in the epithelium of the gallbladder²⁴.

SECs are exceptional endothelial cells localized between the portal blood and the liver and they are playing an important role in immune response, liver inflammatory processes, detoxification²⁵ and liver metabolism. After high fat food intake they are confronted with large doses (up to 43 μ M) of BAs²⁶. The GPBAR1 activated signal pathway results on the one hand in phosphorylation of CD95, protecting SECs from BA induced apoptosis^{24,27}, and on the other hand in phosphorylation of nitric oxide synthase (NOS), providing induced liver microcirculation, for example after vast meals, via induced nitric oxide (NO) synthesis^{22,24}.

Depending on the food intake the alternating bile acid concentration in the blood serum (2 μ M- 15 μ M) may adapt the endothelial cells to the current nutrient availability²⁸.

Apoptosis protecting effects have also been reported in cholangiocytes and gallbladder endothelial cells. In cholangiocytes GPBAR1 seems to have, next to this apoptosis protection, also a major role in proliferation²⁹.

Moreover, GPBAR1 has the role of a bile acid sensor in the biliary tree and gallbladder, controlling gallbladder filling and bile flow. The intracellular cAMP- rise, after GPBAR1 activation, leads to activation of a specific chloride channel (CFTR). The resulting chloride and liquid flow leads to gallbladder filling combined with a GPBAR1-induced relaxation of the gall bladder smooth muscle cells. GPBAR1 activation at the gall bladder reduces therefore the bile flow^{30,31}.

In Kupffer cells, activation of GPBAR1 reduces cytokine production (IL1 α , IL1 β , IL6, TNF- α) and downregulates phagocytic processes. This may protect the liver from damage caused by inflammatory processes²³.

4.1.1.3. Anti-inflammatory effects

The anti-inflammatory BA-mediated effects have been already reported several years before GPBAR1 was discovered³², and are not exclusive to the mentioned hepatic Kupffer cells. In 2003, only one year after the discovery, Kawamata and coworkers described alveolar macrophage functions suppression, such as cytokine production and phagocytosis, through activation of the novel GPCR⁶. In macrophages, BAs normally inhibit, through an intracellular cAMP rise, the activation of the pro- inflammatory transcription factor NF- κ B. Pro inflammatory cytokine production is therefore downsized and the inflammatory processes, mediated by IL1 α , IL1 β , IL6 and TNF- α , are reduced (**Figure 3A**).

In a mice model the inhibition of NF- κ B also proved to reduce lipid loading of macrophages, therefore counteracting foam cell formation, which is an important mechanism involved in atherosclerosis development³³.

BAs also interact in the differentiation of monocytes to antigen presenting dendritic cells (DCs). Expressed in monocytes, activation of GPBAR1, and a subsequent intracellular cAMP rise, induce differentiation to special dendritic cells producing less IL12 and TNF- α when exposed to commensal bacterial antigens than others (although differentiated dendritic cells don't even express GPBAR1 anymore)³⁴. GPBAR1 has therefore a significant influence on early dendritic cell differentiation which is very critical to later inflammatory processes (**Figure 3B**).

Finally, GPBAR1 activation in vascular endothelial cells has anti-inflammatory effects on three levels. **(1)** Firstly, the intracellular cAMP rise after receptor activation induces NO production, through NOS phosphorylation, and can protect cells from possible injuries. **(2)** Secondly, the cAMP rise results in phosphorylation of CD95 protecting the endothelial cells from apoptosis. **(3)** Thirdly, and most importantly, the GPBAR1/Gs/cAMP/NO pathway leads, in both monocytes and macrophages, to an inhibition of the NF- κ B pathway, which per se reduces the expression of cell adhesion proteins, like the vascular cell adhesion molecule-1 (VCAM-1) or intercellular cell adhesion molecule (ICAM-1). Therefore, the monocyte adhesion is demoted (**Figure 3**).

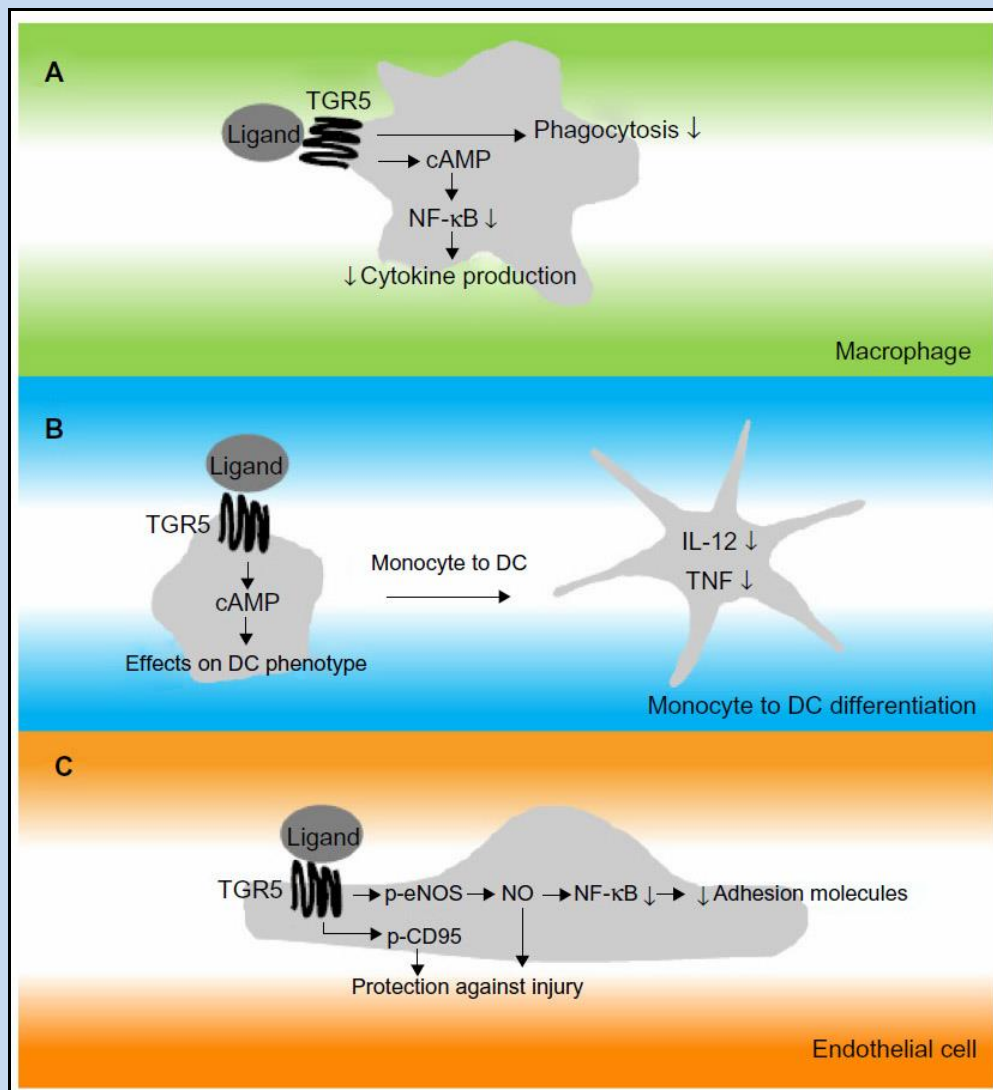


Figure 3: Overview on different anti-inflammatory GPBAR1 mediated effects and their pathways:
A: In macrophages and Kupffer cells GPBAR1 activation suppress functions like phagocytosis and cytokine production.

B: GPBAR1 activation in monocytes leads to differentiation into less inflammatory dendritic cells.

C: NF-κB inhibition in endothelial cells reduces the expression of adhesion molecules.

Figure from Pols T, Eggink H, Soeters M. TGR5 ligands as potential therapeutics in inflammatory diseases. *International Journal of Interferon, Cytokine and Mediator Research* 2014;27. ³⁵

4.1.1.4. Other investigated physiological effects

Remarkably the pleiotropic effects of GPBAR1 influence far more physiological and pathological settings than the three above stated major areas². GPBAR1 modulation also induces intestinal motility³⁶ and chloride efflux¹³. Furthermore it plays an important role in cell proliferation,

apoptosis and therefore tumorigenesis². Perhaps it may both activate the MAP-pathway in tumor cells³⁷ but also inhibit the STAT3 transcription factor³⁸. GPBAR1 is further expressed in the neurons and astrocytes of the central nervous system and may act as a novel neurosteroid receptor.³⁹ To sum up, GPBAR1 is nearly ubiquitously expressed in the body, some of its effects are yet well understood³.

4.1.2. Endogenous ligands

The strongest endogenous activators for GPBAR1 are the secondary BAs lithocholic acid (**1**, LCA) and deoxycholic acid (**2**, DCA), followed by the primary BAs chenodeoxycholic acid (**3**, CDCA) and cholic acid (**4**, CA). CA and CDCA are denominated primary BAs since they are generated in the liver via CYP7A1, while the secondary bile acids LCA and DCA are actually formed in the gut, via bacterial dihydroxylation².

In 2008, Sato and coworkers performed an extensive biological screening of a collection of natural occurring BAs, BAs derivatives and steroid hormones⁴⁰. The activity was evaluated via a CHO cells-based assay, and the results showed several outcomes: **(1)** taurine-conjugated BAs are more potent activators than glycine- conjugated BAs, or not conjugated BAs, **(2)** the ranking of potency is LCA>DCA>CDCA>CA (**Table 1**), and **(3)** showed that nearly all of the 14 steroidal hormones were active but with higher EC₅₀ values when compared to LCA or DCA⁴⁰.

Table 1: Main endogenous ligands of G protein coupled bile acid receptor (GPBAR1):

1 Lithocholic acid (LCA)	2 Deoxycholic acid (DCA)	3 Chenodeoxycholic acid (CDCA)	4 Cholic acid (CA)
 EC ₅₀ = 0.58 μM*	 EC ₅₀ = 1.25 μM*	 EC ₅₀ = 6.71 μM*	 EC ₅₀ = 13.6 μM*

*EC₅₀ values reported by Sato et al., 2008.⁴⁰

It is important to note that BAs are not exclusive endogenous ligands to the GPBAR1, and also target a broad set of nuclear receptors⁴¹, also involved in major metabolic pathways. BAs targeted FXR influences bile acid, lipid and carbohydrate homeostasis at a genomic level⁴². Besides, other nuclear receptors, such as the vitamin D receptor (VDR), the pregnane X receptor (PXR) also called steroid and xenobiotic receptor (SXR), and the constitutive androstane receptor (CAR), can be activated at physiological concentrations of BAs⁴¹.

4.1.3. GPBAR1 relevance in diseases and as drug target

Due to the broad number of effects and the overall expression of GPBAR1 in different cells and tissues, its major impact on homeostasis is clear. But its role in diseases, and putative implications in therapy, is still not fully understood and described³⁵. To avoid side effects in therapy when dealing with GPBAR1 agonists, the importance of selectivity seems clear, as described for the cross-talk among metabolic-relevant receptors.

Even though, numerous studies have highlighted the potential value of GPBAR1 agonists in the treatment of different diseases. **Table 2** summarizes putative cases where the activation of GPBAR1 may modulate specific human pathologies, together with the associated mechanisms and cell/tissue/organ.

4.1.3.1. Metabolic diseases

The world health organization calls obesity a global epidemic. An estimation of 1.7 billion obese people on the world is not exaggerated⁴³. The resulting metabolic disorders together with dyslipidemia, reduced insulin sensitivity and others often lead to serious diseases such as metabolic syndrome⁴⁴, leading to type 2 diabetes, arteriosclerosis with hypertension, non-alcoholic fatty liver disease (NAFLD)⁴⁵ and others.

The fact that the GPBAR1 pathway, also in interaction with the FXR and other pathways, is a central player in all this physiological conditions leading to these diseases suggests GPBAR1 agonists to be potential drugs^{2,3}. Therefore, treatment of metabolic diseases is, amongst others, a dominant field of recently evolving interest in GPBAR1 research.

The most important milestone for this approach was probably the discovery of Deiodinase 2 induction after GPBAR1 activation in skeletal myocytes and brown adipocytes in 2005. Watanabe and coworkers stated the interaction in the GPBAR1 pathway is '*a crucial mechanism for fine-tuning energy homeostasis that can be targeted to improve metabolic control*'²¹, highlighting the potential of the field.

4.1.3.2. Liver diseases

GPBAR1 is also connected to various liver diseases associated with metabolic syndrome, such as the non- alcoholic- fatty liver disease (NAFLD)⁴⁵. NAFLD, and the progressed more severe

form named non alcoholic steatohepatitis (NASH), are marked by hypertension, dyslipidemia, high blood glucose levels⁴⁶, hepatocyte injury, degeneration, apoptosis and inflammation⁴⁷ followed by fibrosis. It is a major risk factor for the development of diabetes, cardiovascular diseases, hypothyroidism, and, in the case of NASH, liver cirrhosis⁴⁵⁻⁴⁷. BAs showed anti-inflammatory and blood glucose decreasing effects through GPBAR1 pathways, as well as genomic effects through FXR for example induced PPAR- α and PPAR- γ . The modulation of both receptors results in improvement and recovery of NAFLD/NASH, and dual agonists are therefore broadly researched^{46,48-50}.

4.1.3.3. Inflammatory diseases

Primary sclerosing cholangitis (PSC), an inflammatory disease of the bile duct which also can lead to liver cirrhosis, is probably linked to GPBAR1, and, therefore, agonists may be beneficial against this disease too. Nevertheless, further studies are needed to verify this possibility^{8,51}. Important to mention is the possible role in arteriosclerosis, since this complex disease is strongly connected to metabolic disorders. The suppression of adhesion molecule expression, inhibition of “foam macrophages”, as well as the increased NO production in endothelial cells after activation of the GPBAR1/cAMP pathway makes it an attractive target also for arteriosclerosis²⁸.

Influences on other inflammatory diseases like colitis⁵², Morbus Crohn⁵³ and gastric inflammation⁵⁴ have also been reported. In 2013 Duboc and coworkers² stated: *‘The Anti-inflammatory effects, especially on TNF- α secretion in monocytes could lead to the development of new molecules to curtail inflammatory diseases, such as rheumatoid arthritis and inflammatory bowel disease.’*

4.1.3.4. Cancer

Whether GPBAR1 has protective effects against cancer development or maybe induces cancer or worsens cancer is yet controversial. While some researchers point out the potential protective effects in hepatocellular carcinoma connected to GPBAR1-related STAT3 inhibition⁵⁵, other authors refer to harming effects in gastric carcinoma³⁷ and esophageal carcinoma⁵⁶. Hodge and coworkers highlighted that the research in this subject is only documented in cell or animal models, and clearly under extreme conditions, making it unsatisfactory to forecast the involvement of GPBAR1 in cancer³.

Table 2: Overview on expression and physiological effects of GPBAR1 in different cells and organs and the resulting implications for therapy

Cell type	Tissue/organ	Effect of GPBAR1 activation	Physiological effect	Benefits and risks from GPBAR1 agonists in therapy	Source
Striated myocytes	Skeletal musculature	Induced T4 → T3 conversion via deiodinase phosphorylation	Higher Energy expenditure	Potential new therapeutics against Obesity, Metabolic Syndrome and Type 2 Diabetes	21
Brown adipocytes	Brown adipose tissue				
α cells	Pancreas	Glucagon to GLP1 secretion switch	Reduced blood sugar level		14
β cells		higher insulin excretion			18
Enteroendocrine cells	Intestine	PYY and GLP1 secretion	Appetite inhibition, improved insulin secretion and reduced Glucagon secretion		19,20
Sinusoidal cells	Liver	NO synthesis and CD95 phosphorylation	Apoptosis protection and induced blood flow in the liver	Induced liver microcirculation	22,24-27
Cholangiocytes	Biliary tree and gallbladder	Proliferation, apoptosis protection, Chloride and water secretion	Decreased bile flow with gallbladder filling	Protection against fibrosing cholangiopathies	12,24,29-31,57
Gallbladder epithelial cells					
Gallbladder smooth muscle cells		Relaxation			
Monocytes, macrophages	Immune system	Suppression of functions	Antiinflammation, reduced foam cell development	Diverse inflammatory diseases	6,23,28,32-35,58
Dendritic cells		Influence on differentiation		Improvement of inflammatory bowel diseases	34,52,53
Endothelial cells	Vascular system	NO synthesis Expression of VCAM1 and ICAM1 reduced			Improvement or protection against atherosclerosis
Gastric carcinoma cells	Cancer	Activation of MAP kinase pathway	Stimulates carcinogenesis	Potential risk of carcinogenesis in the upper GI tract	37,59
Liver carcinoma cells		STAT3 inhibition	inhibits carcinogenesis in the liver	Potential Protection against hepatocellular carcinoma	38

In summary, the search for novel agonists of GPBAR1 could be regarded as an interesting and promising endeavor for scientists looking to fulfill pharmacological unmet medical needs especially within the metabolic diseases field.

4.1.4. Preclinical methods used for the evaluation of GPBAR1 modulation

Due to the positive effects of GPBAR1 modulation, several *in vitro* models have been developed to study the receptor and also to allow rapid screening and optimization of novel agonists. The most frequent assays used in the discovery and preclinical development of drug candidates are luciferase-based cellular assays, mainly employing CHO^{40,60,61}, HEK-293⁶²⁻⁶⁴ or U2-OS cells^{65,66}. The NCI-H716 human cell line is also frequently used⁶⁷⁻⁶⁹, sometimes to verify the results obtained with CHO cells, originated from hamsters⁶⁷.

Nevertheless, the human embryonic kidney-293 (HEK293) cells are the most common human cell line in the literature on GPBAR1 modulators^{5,41,62-64,70-76}, combining robust and easy to handle cells with an uncomplicated transfection. Apart from GPBAR1 research, the HEK293 cell line is both the second most widely used cell line in biopharmaceutical production after CHO cells and also the second most widely used cell line in cell biology research after HeLa cells⁷⁷. In this thesis the bioactivity measurement was done on a daughter cell line, the HEK-293T cells, which incorporate mutations enhancing their protein expression⁷⁸.

The assays themselves can be seen as traditional luciferase reporter-gene assay systems. Plasmids, both for expressing GPBAR1 (target-protein) and the cyclic AMP response element (CRE)-luciferase construct, get transfected. The luciferase-enzyme, responsible for later quantification of the experiment, is under control of the cAMP response element. The CRE initiates luciferase gene transcription, when intracellular cAMP levels rise through activation of the GPBAR- G_s-protein-AC pathway.

In this work chemical transfection with calcium phosphate was possible, since it leads to the transfection of a big part of treated cells. It is a cheap and uncomplicated method⁷⁹. Normalizing variability in cell numbers was achieved by introducing the fluorescent EGFP also through transfection of a plasmid. The fluorescence acts as a normalization signal, correlating with the transfected cell quantity.

Figure 4 shows the process of the used luciferase-assay in a simplified form.

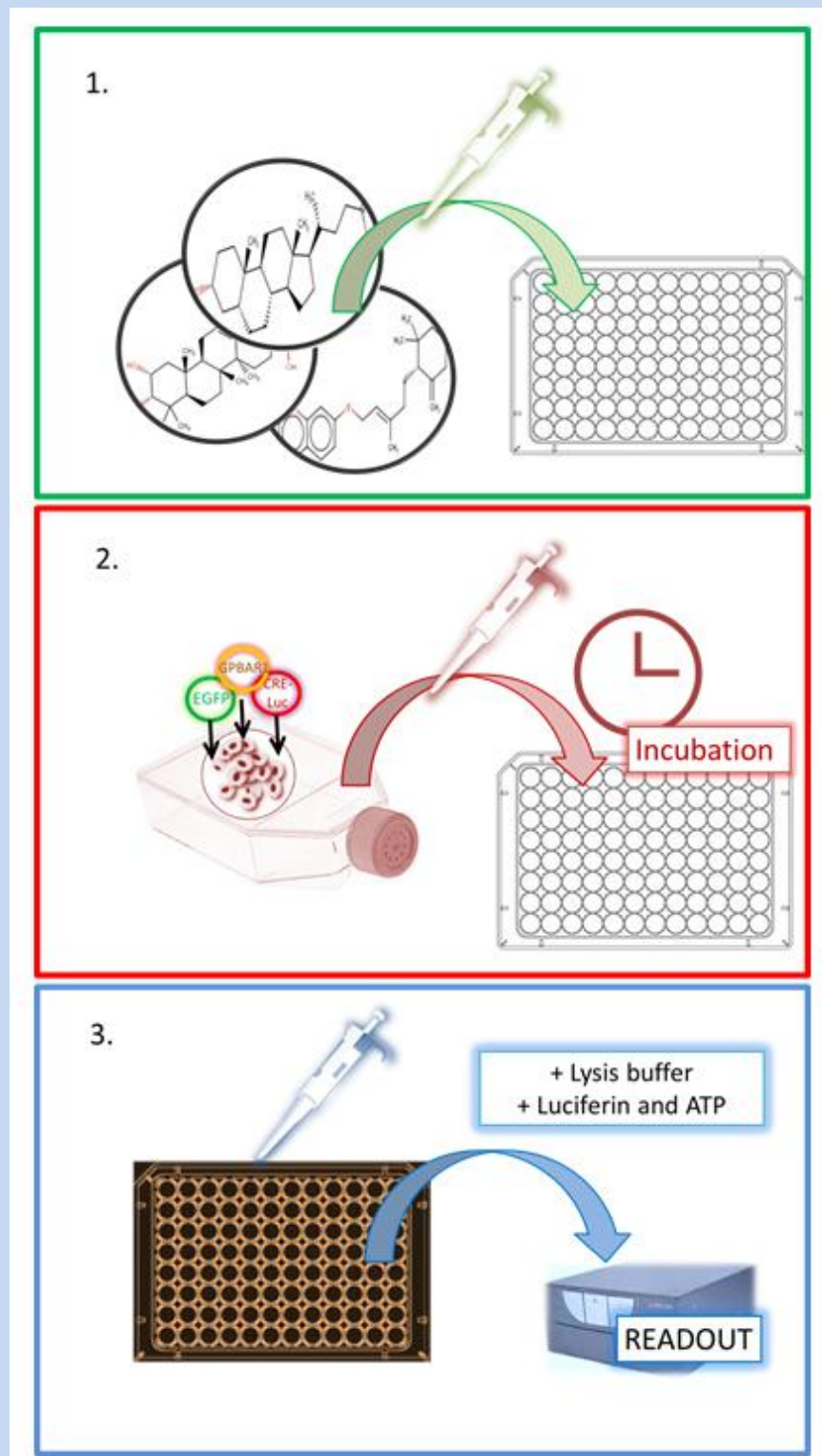


Figure 4: Simplified scheme of a GPBAR-CRE-driven luciferase assay: (1) the positive and negative controls (DMSO) as well as the test substances get filled into the wells. (2) The previously transfected cells are seeded into the wells and are incubated for 18 hours. Agonists of GPBAR1 lead to a rise of intracellular cAMP and therefore luciferase. (3) The cells are lysed, together with luciferin and ATP, the luciferase generates light, which is then quantified.

Summed up it can be stated that quantification is performed by measuring the cAMP rise via luminescence^{40,60-66}. This readout can then be expressed as EC₅₀ or % of activation in comparison with internal controls, such as LCA or CDCA.

There are several variations of luciferase-assays established by different working groups in the preceding research. A few publications illustrate for example that a doxycycline-induced GPBAR1 expression system in CHO cells leads to improved sensitivity in the assay. This approach also resulted in lower EC₅₀ values when compared to non-induced systems^{80,81}.

Other groups have also worked with diverse cell lines, using Jurkat cells⁸², melanophores⁶⁵ or 293T-cells⁸³. Differences among assays (especially in sensitivity) and the overall variability of *in vitro* screenings⁸⁴ raise concerns about direct data comparison among these assays. A deeper discussion on this topic is presented later in this thesis (**5.1 Benchmark establishment-Classification of Agonists from the literature**).

4.1.5. Small molecule modulators of GPBAR1

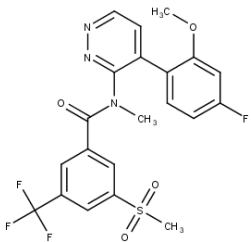
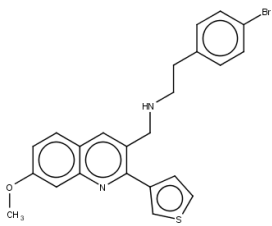
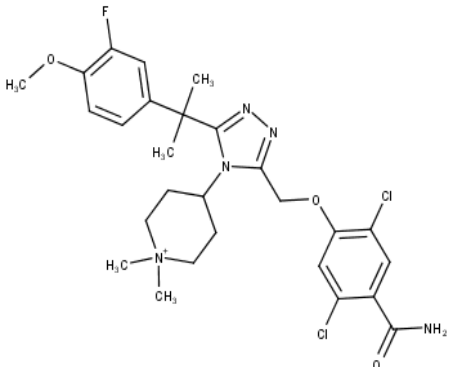
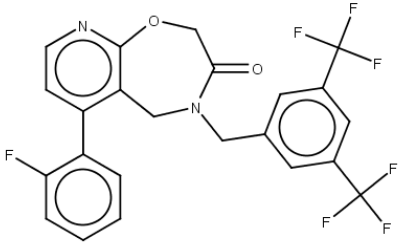
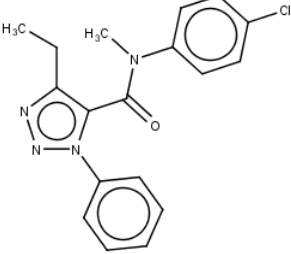
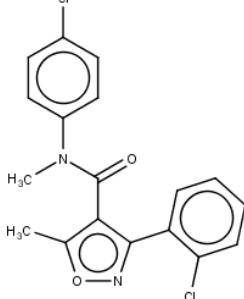
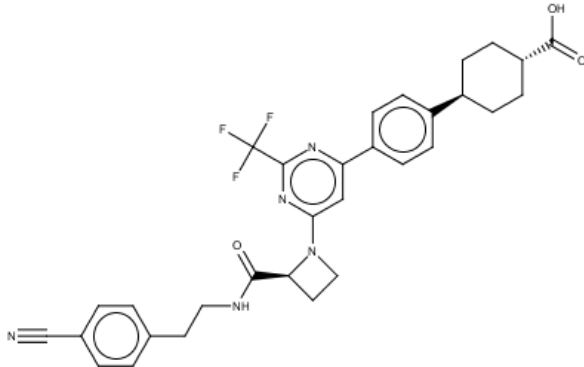
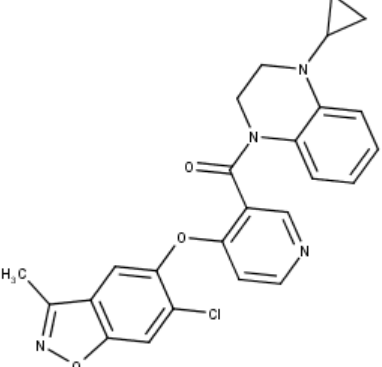
Apart from the endogenous ligands, a broad set of compounds have been evaluated in cellular assays for their GPBAR1 activity and published or patented as modulators of GPBAR1. A smaller number of compounds have also been evaluated *in vivo*, and only few candidates have been tried clinically. For a comprehensive review of published GPBAR1 ligands I refer to the work of Xu and coworkers¹.

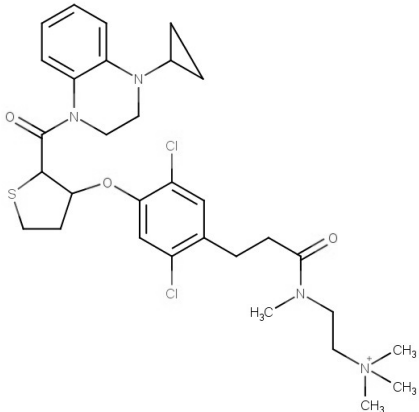
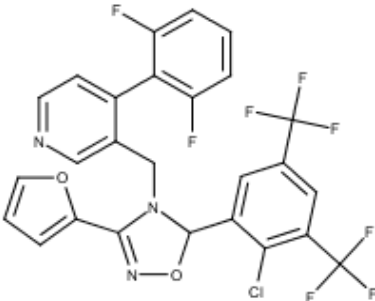
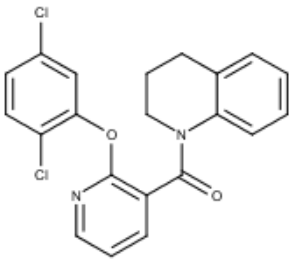
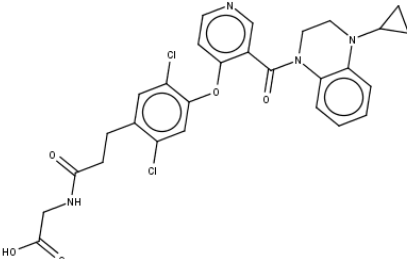
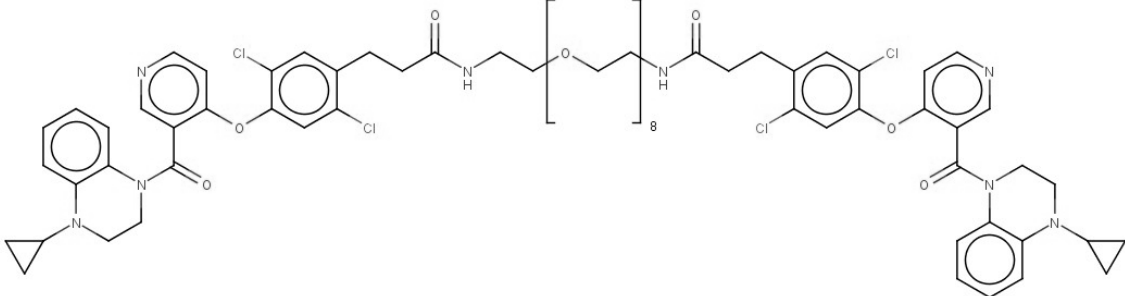
In general, GPBAR1 agonists can be classified as steroidal and non-steroidal scaffolds. Significant improvement in potency and selectivity when compared to primary BAs has been achieved by intensive medicinal chemistry efforts. **Table 3** presents a general overview on GPBAR1 ligands.

Additionally, a short discussion is given on selected leads representative of the progress in the research field.

Table 3: Selected GPBAR1 agonists reported in literature

Primary Bile Acids	
CDCA ^{5,40}	CA ^{5,40}
Secondary Bile Acids	
LCA ^{5,40}	DCA ^{5,40}
Semisynthetic Bile Acids	
INT 777 ^{85,86}	INT 767 ⁸⁶
Synthetic Compounds – Non-Steroidal	
RO-5527239 ⁶⁷	SB-756050 ⁸⁷

Roche®⁸⁸ 	Kalypsos®⁷⁰ 
Exelexis®⁸⁹ 	Takeda®⁶² 
GSK® patented triazol^{66,80,90} 	GSK® patented isoxazol^{65,66} 
Trifluoromethyl (pyrimidin-2-yl) azetidine-2-carboxamide⁶²	Dihydroquinoxaline⁹¹
	

Intestinal targeted GPBAR1 agonist with a quaternary ammonium ⁹²	3,4,5-trisubstituted 4,5-dihydro-1,2,4-oxadiazole ⁶³
	
2-Phenoxy-nicotinamide ^{62,93}	Intestinal targeted GPBAR1 agonist ⁶⁴
	
Intestinal Targeted GPBAR1 agonist with PEG-link ⁶⁴	
	

4.1.5.1. Tauroolithocholic acid (TLC)

In Sato's 2008 screening paper, tauroolithocholic acid (TLC, **Figure 5**) showed an EC₅₀ of 290 nM and was the most active natural BA⁴⁰. This compound is comprised by the bacterial product lithocholic acid conjugated with the aminosulfonic acid taurine. Both, the absence of hydroxylation in position 6, 7 and 12 and the presence of the sulfuric moiety result in the high potency on GPBAR1⁴⁰.

In another study, it showed also agonistic activity of the nuclear FXR, but far less than CDCA, the most potent natural BA FXR ligand⁹⁴.

Administered in a dog model, it showed positive modulation of PYY secretion⁹⁵ and the effects on GLP1 release, intracellular Ca^{2+} and cAMP rise was broadly investigated just recently⁹⁶. Therefore, this compound was also evaluated clinically on 10 diabetic obese male human subjects with rectal administration. Here, TLC led to a dose dependent release of GLP1 and PYY, enhancing fullness, and emphasizing the potential of GPBAR1 agonists⁹⁷.

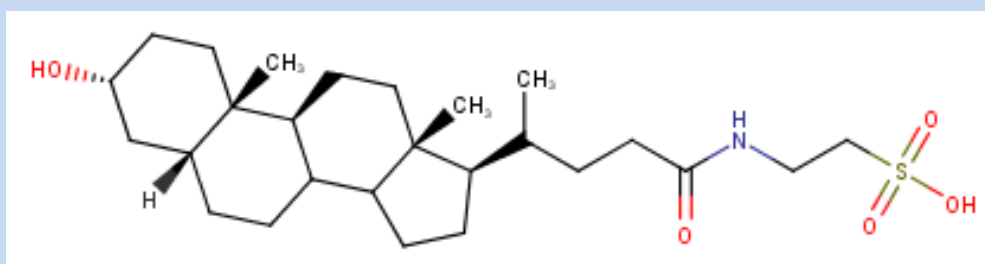


Figure 5: Structure of Tauroolithocholic Acid (TLC)

4.1.5.2. INT- 777

In 2009 Pellicciari and coworkers presented INT-777 (**Figure 6**), a potent and selective agonist on TGR5, as a result of a vast lead optimization process. They installed several modifications on the natural BA scaffold including: **(1)** the methyl-group in position (S)-23 lead to selectivity over FXR. **(2)** The three hydroxyl-groups in position 3, 7 and 12 (analogue to CA) ensure both better solubility and less activity on FXR. **(3)** The ethyl-group in position (S)-6 (analogue to 6 α -ECDCA) increases the agonistic activity⁸⁵.

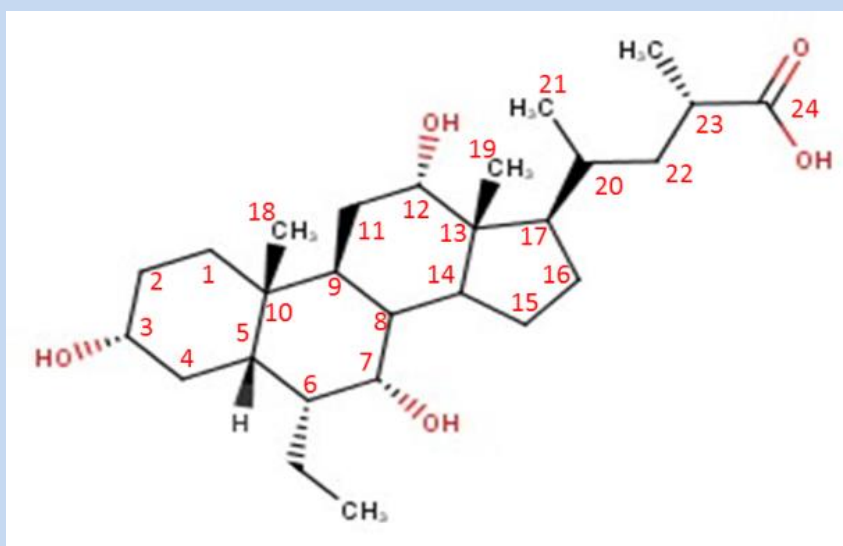


Figure 6 The high activity of INT777 is related to the α -ethyl group in position 6 and the high selectivity over FXR is related to the (S)- 23 methyl group

An *in vivo* study underlying the idea of GPBAR1 agonist efficacy on metabolic disorders was performed by Thomas and coworkers. Diabetes-induced mice were treated with the semisynthetic bile acid, named INT-777. The group observed the treatment with INT-777, which not only led to an enhanced glucose tolerance, but also increased energy expenditure and lead to a weight reduction of 15% in 10 weeks⁴⁸.

Besides the promising effects of INT-777, both *in vitro* and *in vivo*, a study published in 2015 in a mouse model noted way more critical results pointing out to the severe side effect of gallbladder extension, which cannot be undocked from the antidiabetic efficacy. Further, INT-777 showed little effect on GLP-1 secretion *in vivo*⁹⁸. INT-777 was never tested in humans.

4.1.5.3. SB-756050

Patented in 2007 by GSK's SB-756050 (**Figure 7**) or 1,4-bis ([3,4-bis(methoxy)phenyl]sulfonyl) piperazine is in contrast to the other molecules mentioned before not based on a BA scaffold and originated from a synthetic chemical library. As can be seen in **Figure 6** a central piperazine cycle is flanked by two aryl-sulfonyl-substitutes on each nitrogen-atom, resulting in a molecule with no hydrogen-bond-donating groups and two aromatic rings⁹⁹.

It was the most promising of the patented bis-aryl-sulfonyl-derivates, and showed potent activity in transfected melanophore cells and proved to increase GLP-1 levels in rodent models¹⁰⁰. It showed a selective activity on GPBAR1, a moderate *in vitro* activity with an EC₅₀ of 1.3 μ M and a favorable drug-profile in rat and dog models.

Therefore, in 2013 SB-756060 was tested in a first clinical trial. Unfortunately, the transferring of the positive *in vitro* and animal profile to humans in a phase 2 trial was not achieved. SB-756050 failed to meet the predetermined efficacy threshold, although it was well tolerated. The pharmacodynamic profile was puzzling and the effects were inconsistent³.

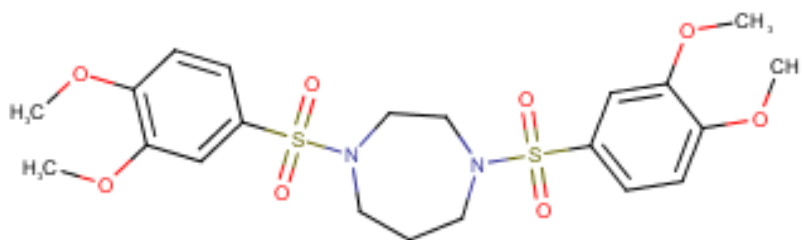


Figure 7: SB-756060

The few clinical experiments so far were manageable but the effects seem to connect well with the knowledge of the physiological role of the receptor and resulted in a better understanding of for example the structure-activity relationships of tested candidates. Obviously, the modulation of GPBAR1 via agonists seems still promising, although side effects and lack of efficacy have impaired the development of a drug until now³.

4.1.5.4. FXR agonists

More clinical research has been done on FXR modulators and also on dual FXR/GPBAR1 agonists, especially obeticholic acid (OCA) (**Figure 8**). OCA (or INT-747) can on the one hand be seen as a quite selective FXR modulator but on the other hand also shows activity on GPBAR1, with an EC_{50} of 15 μ M, compared to 7 μ M LCA and 33 μ M CDCA in the same study⁸⁶. OCA managed it into clinical research against liver diseases a few times with both promising and discouraging outcomes^{1,41}.

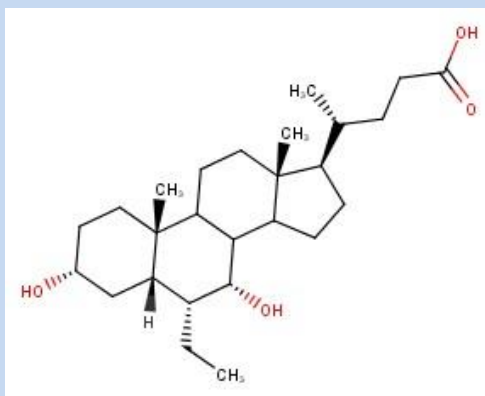


Figure 8: Obeticholic acid (OCA)

Firstly, OCA showed in a double-blind and placebo controlled phase 2 clinical trial on 64 diabetes patients with comorbid fatty acid disease a significant reduction of liver damage and a better insulin sensitivity, together with an induced weight loss¹⁰¹.

The second randomized and placebo-controlled study against non-cirrhotic, non-alcoholic steatohepatitis with INT-747 resulted in an improvement on several parameters, such as inflammation, but not in a regression of the disease. It was further noted that patients showed a problematic deterioration of serum cholesterol levels with higher total and LDL cholesterol and a decreased HDL cholesterol level. Moreover, subjects showed interestingly an increase of insulin resistance^{1,102}.

The third clinical study with obeticholic acid against primary biliary cirrhosis (PBC) INT-747 presented efficacy on nearly all biochemical parameters but exacerbated OCA dose dependent pruritus. The pathway causing this is not determined but GPBAR1 interaction was stated to be a possible cause^{103,104}.

During the completion of the thesis, the European Medicinal Agency granted marketing authorization for obeticholic acid (Ocaliva®) in December 2016 for the treatment of PBC¹⁰⁵.

But also GPBAR1 selective agonists showed severe side-effects, most importantly the above mentioned gallbladder-distension, due to smooth muscle relaxation and gallbladder filling⁹⁸ and maybe the observed itching is due to GPBAR1 interaction⁹¹. More intestinal targeted agonists, with reduced/limited intestinal permeability, can be regarded as a possible solution and in fact have shown good results in animal models⁶⁴.

A very interesting approach to guarantee more intestinal restricted drugs was published just recently by a Chinese working group. They equipped a known synthetic GPBAR1 agonist with a quaternary ammonium substitute and achieved both less gallbladder-related side effects and a robust GLP-1 mediated anti-hyperglycemic effect. Nevertheless, it was not fully possible to eliminate the gallbladder-filling effect⁹².

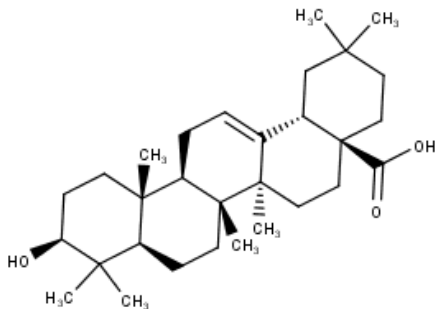
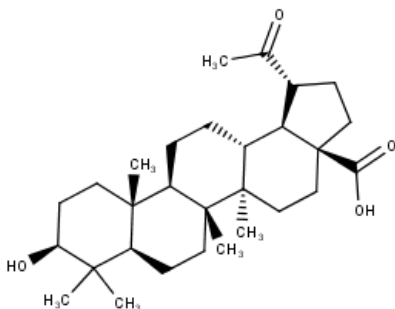
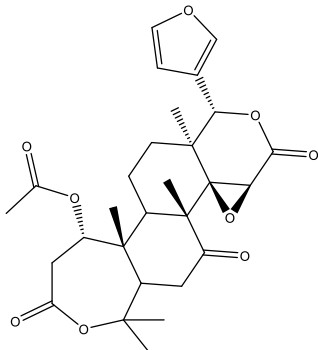
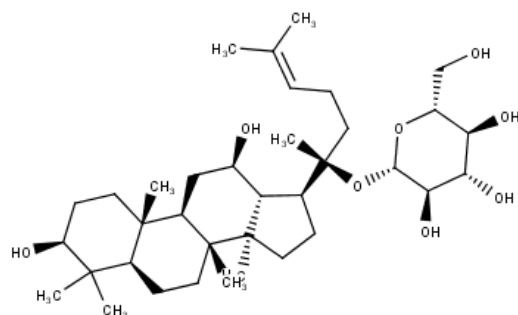
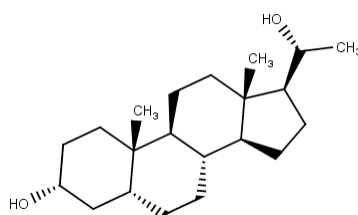
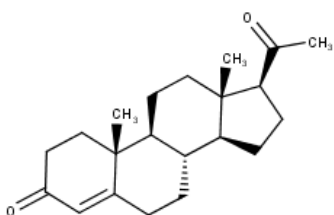
4.1.5.5. Natural products

In parallel with the classical medicinal chemistry efforts, some studies used phytochemistry approaches in the search for novel GPBAR1 modulators. In 2007, Sato and coworkers reported the GPBAR1 activating potential of *Olea europea* extracts, followed by the description of the *in vitro* and *in vivo* activity of the isolated triterpenoid oleanolic acid (**Table 4**)⁶⁰.

After this pioneer work, a completely new field of GPBAR1 research emerged focused on NPs. **Table 4** shows a selection of GPBAR1 modulators of natural origin. Other triterpenoids of natural origin, such as ursolic acid and betulinic acid, have been described as GPBAR1 modulators. Interestingly, these compounds not only showed potent activity on GPBAR1, but also selectivity over FXR⁶¹.

Further natural GPBAR1 modulators are neurosteroids³⁹ and other molecules with diverse non-BA steroidal scaffolds showed potent *in vitro* activity, for instance the androstane or pregnane derived pregnandione, 5 α - pregnandiol, etiocholanone and progesterone⁶⁰.

Table 4: Selected NP with GPBAR1 activity reported in literature

Secondary plant metabolites	
Oleanolic acid $EC_{50} = 1.42 \mu M$ ⁶⁰	Platanic acid $EC_{50} = 3.11 \mu M$ ⁶¹
	
Nomilin no EC_{50} reported ¹⁰⁶	Compound K $EC_{50} = 7.76 \mu M$ ⁶⁹
	
Steroid hormones and metabolites	
Pregnanediol $EC_{50} = 0.85 \mu M$ ⁴⁰	Progesterone $EC_{50} = 2.770 \mu M$ ⁴⁰
	

In fact, many extracts and isolated NPs have been reported to have antidiabetic and antihyperglycemic effects^{107,108}. In 1995, Marles and Farnsworth presented an overview on 1200 different species of plants reported to have antidiabetic effects¹⁰⁸. Later, the antidiabetic activity could in some cases be traced back to interactions with the GPBAR1. For instance, studies on compound K (**Table 4**), a metabolite of ginsenosides, revealed that the antidiabetic activity of the compound could be traced back to GPBAR1 interaction⁶⁹. Further, the highly oxygenated triterpenoid nomilin, found in citrus seeds, showed *in vitro* potent GPBAR1 activity and mediated antidiabetic activity in a high-fat diet mouse model in terms of preventing obesity and hyperglycemia¹⁰⁶.

4.1.6. Summary

Although the potential use of GPBAR1 agonists seems to be promising, several classical preclinical hurdles have also been encountered in this field of research. The plethora of biological functions appears to be an obvious opportunity, but a major drawback at the same time.³ The need for novel, potent, selective, and non-toxic leads on the one hand, and the further translational validation via clinical studies on the other hand, are paramount for making the therapeutic use of GPBAR1 modulators a reality in the future.¹

4.2. Natural products and computational studies: A multidisciplinary approach in drug discovery

4.2.1. Natural products

NPs for the treatment and prevention of diseases have been used for thousands of years and acted as the leading source for drug discovery with a switch in the twentieth century from complex therapeutic extracts/formulations to single molecule remedies¹⁰⁹. Next to the role as source of small molecule drugs in research and the traditional usage in industrialized countries, the importance in developing countries is even bigger. The World Health Organization (WHO) stated in 1985 that around 65% of the world population is still depended on plant-derived traditional medicines for their primary health care¹¹⁰.

Besides this major status, most of the big pharmaceutical companies have scaled down or even put an end to their research in NP for lead development during the last decades. The causes for this process are various¹¹¹.

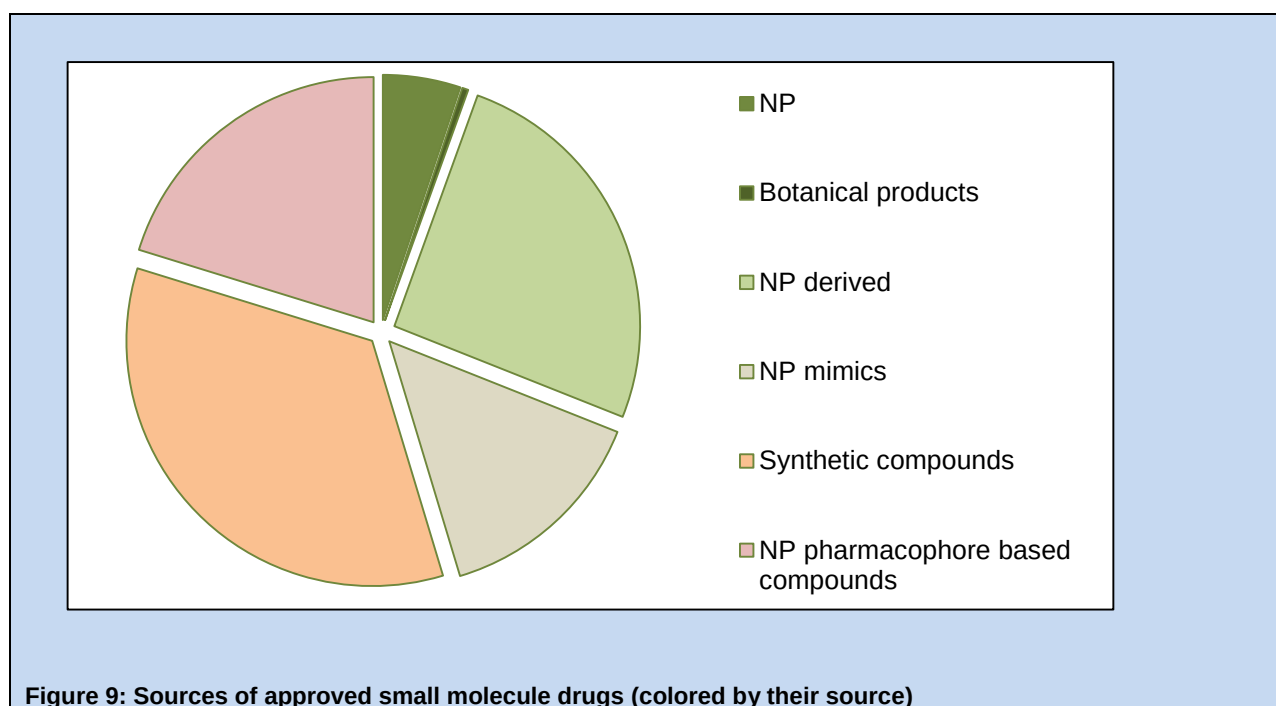
It has to be mentioned that, when compared to synthetic compounds with already known structures, research with NPs is both time and cost consuming to get from an NP assay hit to knowing its chemical structure. In addition, the complex NP leads are often difficult to optimize for medicinal chemists, because they are facing structural complexity, multiple hydroxyl moieties, ketones and chiral centers. With the development of robotized high throughput screening and combinatorial chemistry the drug discovery process has fast tracked and it got very difficult to combine this method with compounds from natural origin. Regarding all this disadvantages it may seem that lead structure development from NP will be overcome, by modern methods, but that is a fallacy¹¹².

When compared to synthetic compounds the structural diversity of NPs is in the first row surely a challenge, but on the other hand it covers some significant advantages over combinatorial

chemistry libraries. First of all, NPs cover a far wider range of chemical space and this range overlaps way more efficient with that of applied drugs^{111,113}. Moreover, approximately only 10% of all NPs violate the Lipinski's rule of 5 (correlated with oral bioavailability), which is the same amount as approved drugs¹¹⁴. This advantage indicates them as an interesting and effective source for drug discovery also in the future¹¹¹.

Borman and coworkers¹¹⁵ brought the conflict between huge combinatorial chemistry libraries with HTS and NP based methods to the striking point: *'A small collection of smart compounds may be more valuable than a much larger hodgepodge collection mindlessly assembled'*.

This idea is strengthened by recent statistics, which prove that NPs are a capable source for new drug discoveries. A recent paper by Newman and coworkers in 2016 lists all approved drugs between 1981 and 2014 by source (**Figure 9, Figure 11**): Out of the total 1328 approved drugs, 1045 had been small molecules. From all 1045 small molecules 31% was originated from natural sources (54 NPs, 4 botanical NPs, 268 NP derived) next to 14.3% of NP mimics, 34.4% total synthetic drugs and the remarkably 20.2% of all drugs, conceived after using the NP's pharmacophore (55 completely synthetic and 156 synthetic NP mimicking)¹¹⁶.



In the field of type 1 and type 2 diabetes the statistics are even more significant (**Figure 11**). Out of the 29 small molecules approved, only 4 were fully synthetic, 1 was from natural origin, 6

derived from NPs and 11 synthetic NP mimics. Finally, 7 molecules were approved after using NP's pharmacophores with only 1 not mimicking a NP¹¹⁶.

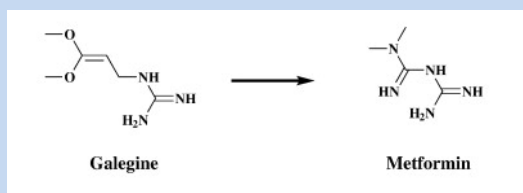


Figure 10: Probably the most commonly used antidiabetic metformin is based on galegine, a NP from *Galega officinalis* L. Figure from Cragg GM, Newman DJ. Natural products: a continuing source of novel drug leads. *Biochimica et biophysica acta* 2013;1830(6):3670-95.

That nature represents a good source for drugs and also for antidiabetic drugs is acknowledged. For example, the most commonly used anti-diabetic drug metformin is based on the discovery of galegine (**Figure 10**) from *Galega officinalis* and acarbose, an alpha glucosidase inhibitor, was discovered in a bacterium¹⁰⁶. Further, the novel field of incretin mimetics was based on the isolation of 4-exenedin from the lizard *Heloderma suspectum*'s venom^{117,118}.

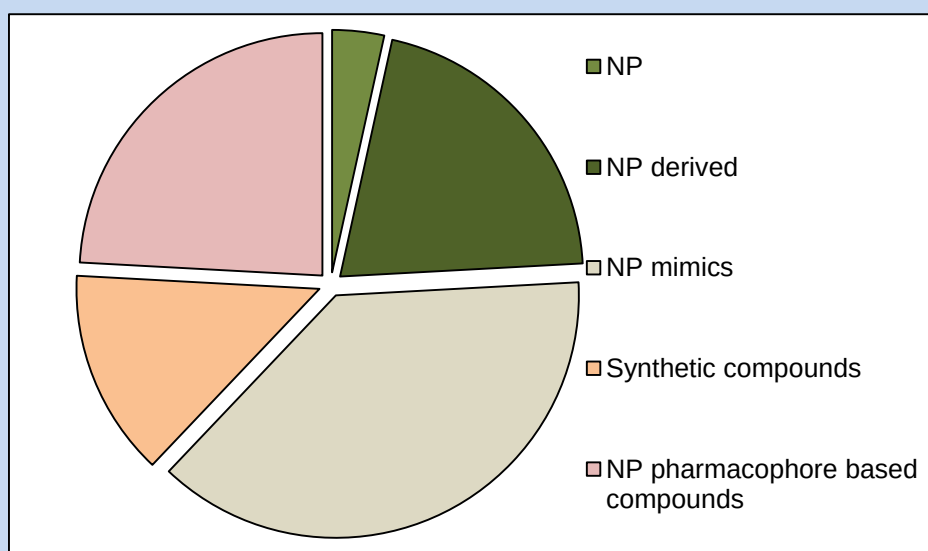


Figure 11: Sources of approved small molecule drugs on type 1 and 2 diabetes (colored by their source)

In the context of bile acid like agents it is worth to note that a vast amount of FXR modulators from natural sources, with unique selectivity and kinetics, have been described¹¹⁹⁻¹²⁶ but the field of GPBAR1 agonists seems, despite the high potential, underrepresented, appearing to be

just at the beginning. Sepe and coworkers⁴¹ stated therefore in 2015 '*The research in the field of triterpenoids and of neurosteroids, although still in its infancy, represents a promising strategy to selectively target GPBAR1-mediated disorders.*'

Even though NPs represent a valuable source of leads, a successful discovery process in this field can be time and resource consuming. In this view, a more rational, multidisciplinary and innovative approach, for instance by combining classical phytochemistry with computer-aided methods, could overcome some of the above mentioned difficulties and result in higher efficiency¹¹¹.

4.2.2. Multidisciplinary approaches

While the methods of classical photochemistry are bio-guided fractionation, ethno-pharmacological driven research, extract screening and structure activity relationships based on the retrieved chemotaxonomic knowledge, computer-aided methods have emerged as another very proficient approach to support research¹¹¹.

Considering all the data mankind has created over the years, and the recent increase in both speed and volume of data creation, there is a growing need for techniques that can extract knowledge from all this information. Computational methods, also termed *in silico* approaches, are regarded as one of the best solutions for this issue¹²⁷.

Several task areas for computer aided tools in NP research have emerged and the following just points out some of the potential application in pharmacognostic drug research with reputed examples. All of them share the strategic goal of focusing experimental efforts through predictions and rationalization of research¹²⁷.

Table 5: Exemplary studies using *in silico* methods in NP research

Aim of research	Example(s)	Citation
Exploitation of the chemical space occupied by natural compounds	Miyao and coworkers visualized through generative topographic mapping (GTM) of NP and synthetic drugs their chemical space and revealed with cluster analyzes new potential activities of NP and drug-like compounds	Miyao, Reker, Schneider, Funatsu and Schneider ¹²⁸
Understanding ligand-target interaction	Hegazy and coworkers showed the binding of sesquiterpenes from the wild artichoke <i>Cynara cornigera</i> to the AChE and explained their inhibitory effects through pharmacophore modeling and <i>in silico</i> docking experiments	Hegazy, Ibrahim, Mohamed, Shahat, El Halawany, Abdel-Azim, Alsaïd and Pare ¹²⁹
Identification of underlying causes (target identification) for reported effects	Gidaro and coworkers explained some of the health benefits by bergamot oil and red onions through target based screening against five carbonic anhydrase isoforms with the identification of eriocitrin and apigenin as potent inhibitors.	Gidaro, Alcaro, Carradori, Costa, Vullo, Supuran and Alcaro ¹³⁰
Understanding target patterns of pleiotropic plant extracts and making predictions on their potential therapeutic use	Goel et al. used <i>in silico</i> mining of phytochemical diversity and a programme that predicts biological activity to identify plants that contain constituents that have multiple mechanisms of action related to a desired therapeutic output.	Goel, Gawande, Lagunin, Randhawa, Mishra and Poroikov ¹³¹
Identifying potential health risks of NP through virtual screening (VS)	Kratz and coworkers identified <i>Carapichea ipecacuanha</i> alkaloids through pharmacophore based screening as HERG- channel blockers and showed their potential cardiotoxicity.	Kratz, Mair, Oettl, Saxena, Scheel, Schuster, Hering and Rollinger ¹³²
Pharmacophore based virtual screening (VS) of molecular NP databases with the aim to find new hits for exemplified targets	Vuorinen and coworkers found new 11 β -hydroxysteroid dehydrogenase type 1 (11 β -HSD1) inhibitors from <i>Pistacia lentiscus</i> oleoresin through pharmacophore based VS	Vuorinen, Seibert, Papageorgiou, Rollinger, Odermatt, Schuster and Assimopoulou ¹³³
	Grienke, Schuster and coworkers identified triterpenes from <i>Ganoderma lucidum</i> as agonists for FXR by VS against a target based pharmacophore model.	Grienke, Schuster et al. ^{122,134}
	Waltenberger and a multidisciplinary network of coworkers identified with pharmacophore based VS assistance novel anti-inflammatory compounds and underlined the efficacy of these multidisciplinary research.	Waltenberger, Atanasov, Heiss, Bernhard, Rollinger, Breuss, Schuster, Bauer, Kopp, Franz, Bochkov, Mihovilovic, Dirsch and Stuppner ¹³⁵
	Guasch et al discovered partial PPAR γ partial agonists of natural origin with novel scaffolds.	Guasch, Sala, Castell-Auvi, Cedo, Liedl, Wolber, Muehlbacher, Mulero, Pinent, Ardevol, Valls, Pujadas and Garcia-Vallve ¹³⁶

Pharmacophore based VS of molecular NP databases has led obviously to very successful results^{122,127,132-137}. In this context the integration of this approach into the search of new active substances on GPBAR1, is surely a straightforward and validated strategy, underlined by studies like the mentioned ones. The aim is hereby to discover new hits on a characterized

target and hopefully with yet unknown chemical scaffolds, with a cost and time effective approach¹³⁸.

For achieving that a viable method, a restrictive filter for a virtual high throughput screening, is necessary to reduce the number of compounds tested *in vitro* with a maximum count of received actives. A both well-established and very successful VS filter is a function-based pharmacophore¹³⁹.

4.2.3. Pharmacophore

The name pharmacophore is officially defined by the IUPAC¹⁴⁰ as:

‘A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response. A pharmacophore does not represent a real molecule or a real association of functional groups, but a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds toward their target structure. The pharmacophore can be considered as the largest common denominator shared by a set of active molecules.’

To sum up you can define a pharmacophore as a unique pattern of three dimensional attributes responsible for recognition between a set of active molecules and their biological target.¹⁴¹ Historically a representative and pioneer model has been created by Lloyd EJ, Andrews PR. and their coworkers in 1986¹⁴². Their CNS-pharmacophore, just describing 3D arrangements of a nitrogen atom, a phenyl-group and two lipophilic features, was able to predict very efficient potential CNS activity of numerous compounds. Far earlier, already in 1938 E.C. Dodds and W. Lawson realized that the estrane-scaffold isn't necessary to achieve estradiol like activity in mice and furthermore they proposed a pharmacophore like model, with features small molecules should be capable of, when testing on estradiol like activity¹⁴³. So the idea has been known for a long time but the broad use of the pharmacophore- concept had certainly its advent with computational methods coming on pharmaceutical research acting as a useful tool for drug discovery.

The popularity of the pharmacophore concept is simply explained by several advantages, compiled within this concept. Firstly, it can be built without any information on the target, when a

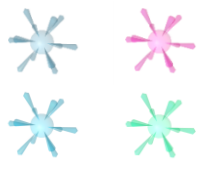
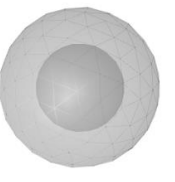
homology or experimental structure of the target protein is absent. This point is from major importance for membrane-bound proteins like ion channels (e.g. the HERG- channel¹³⁷) or in this case GPCRs like the GPBAR1. Secondly, their application as a VS filter is today known as a very effective approach^{144,145}. Thirdly, their ability to predict activity is from foremost importance in today's drug discovery process enabling focused and straightforward research. Fourthly, their capacity of reducing the immense complexity of ligand-target interactions to an uncomplicated but illustrating model enables both fast VS and easy understanding.¹⁴¹

A pharmacophore has to describe on the one hand the 3D location with necessary spatial orientation of the features and on the other hand the chemical characteristics of functional groups participating in ligand-target interactions. It therefore classifies different types of occurring interactions into a set of few 'pharmacophore features' like charged group features for ionic interactions, hydrogen bond donor and acceptor features for possible hydrogen bonds, aromatic features for π - π and π -cation interactions and finally lipophilic features for van-der Waal's interactions between receptor and ligand. Depending on the used software package further classes of 'pharmacophore features' are available for modeling but the most important features are named above¹⁴⁶. **Table 6** presents the most significant pharmacophore features, included in LigandScout 4.0.

There are different software applications for generating a model and VS. The most frequently used are Catalyst from Accelrys, MOE from Chemical Computing Group, Phase from Schrödinger and LigandScout by Inte:Ligand¹⁴⁷. A comparative VS experiment recognized a high variance between the hit lists obtained from the first three tools MOE, Catalyst and Phase due to different alignment algorithms¹⁴⁸.

Ligand Scout comprises two advantages. Firstly, there is no upper limit of features, resulting in higher selectivity. Secondly, an alignment through pattern recognition instead of graph matching algorithms results in no experimental growth with growing number of features^{149,150}.

Table 6: the most important available features for model generation in LigandScout 4.0

Symbol	Feature	Examples for fitting groups
	Positive ionizable feature	Basic nitrogen groups and cations
	Lipophilic feature	Lipophilic hydrocarbon groups for instance methyl, iso-propyl or phenyl residues
	Negative ionizable feature	Acid groups and anions
	Hydrogen bond acceptor feature	Ketones, alcohols, amines, amides, esters, ethers, fluorides, carbonic acids, sulfonic acids
	Hydrogen bond donor feature	Alcohols, amines, amides, carbonic acids
	Diverse metal atom binding locations	nitrogen and oxygen groups able to complex Fe, Mn, Mg or other ions
	Aromatic ring feature	Phenyl-, pyridyl-, thiophenylresidues amongst others
	Exclusion volume	maintaining steric restrictions in the model

Pharmacophore models can be basically generated with 2 approaches: A target based approach with available information both on the protein and its ligand(s) or when no target information can be used the ligand based approach representing still a reasonable method by only containing information on known ligands and no information on the target. The target based modeling approach founds on reliable material on the targeted protein like a high resolution x-ray model or a homology model, received after modeling against a very similar template¹¹¹.

In the case of GPBAR1 there was no crystal structure available and the published homology models⁹⁻¹¹ were not shared. Under such circumstances a completely ligand based pharmacophore model approach can also be very proficient. The principle of this attempt is the alignment of different conformations of dependable actives, called the training set, in order to receive the necessary steric and electrochemical features for the interaction between the ligands and the unknown binding pocket of the protein. The selection of the training set is hereby the kingpin of the whole method¹⁴⁷.

After building the model and before it is used for VS, it is validated by screening against a library of known actives, with same binding type and position, against a way larger library of inactives or/and decoys^{147,151,152}.

In contrast to inactives, which are reported to be not active at the target, decoys are molecules sharing large structural similarity with the active molecules, but have not been experimentally tested yet. This may result in obvious disadvantages in the validation process, because you cannot knock out possible bioactivity for all decoys. Unluckily, only few inactive molecules are published and scientists have to rely often on the biased actives vs. decoys screening as a benchmark^{147,151}. Nevertheless, several useful enrichment metrics can be calculated with the received hit lists, to compare models and enable their improvement¹⁴⁷.

The two fundamental values for this task are the true positive rate (Sensitivity) and the false positive rate (Specificity). While sensitivity describes the ability of the model to find actives, the specificity describes the ability of the model to discard inactives and both are independent from the used screening library with which they are retrieved or its size or ratio of actives/inactives^{139,153}.

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{\text{picked actives}}{\text{all actives}}$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{\text{all discarded inactives}}{\text{all inactives}}$$

Further very useful metrics are the accuracy and the balanced accuracy. Both the balanced accuracy as the mean value of specificity and selectivity¹⁵⁴ and the literal accuracy¹⁵⁵ are sometimes useful quality parameters¹⁴⁷.

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

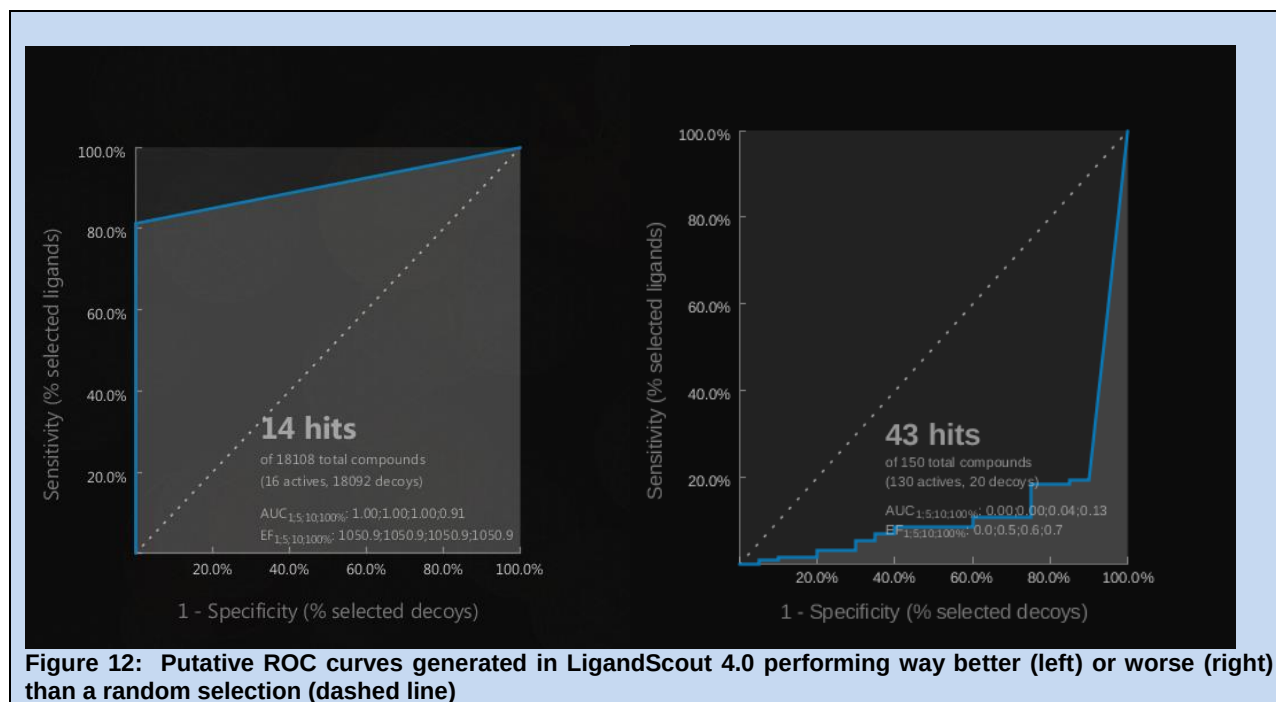
$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All Negatives}}$$

The enrichment factor (EF) is the ratio of true positives in the hit list in comparison to the ratio of all actives in the entire screening library¹⁴⁷. Summed up, the EF shows how many times more active molecules are retrieved by the pharmacophore model than by a random selection. The EF is surely the most illustrating numerical descriptor in terms of model refinement but has two disadvantages: **(1)** The EF is dependent on the ratio of active/decoys in the screening set, making it impossible to gain high values, if this ratio is high. Consequently EF values are just comparable, when the screening was done on the same screening set. **(2)** A ranking of the alignment scores is missing. For the second reason the receiver- operating system is favourable¹⁴⁷.

$$EF = \frac{\text{True Positives} / \text{All Positives}}{\text{All Actives} / \text{All Compounds}}$$

Finally, one of the most valuable and illustrating methods for hit list-validation is the receiver operating characteristic (ROC) curve analysis. After ranking the alignment scores for all hits, selectivity and specificity values for all possible score levels, fixing the numbers of active and inactives are computed. The hereby computed selectivity values are then plotted as a function of 1 - specificity, shown in **Figure 12**. A random selection of molecules would lead to a straight line, similar to the broken line in **Figure 12**. While good models have a large area under the

curve (AUC) with an intense rise at the beginning, indicating the capacity of the model to rank actives first, bad models have less increase at the beginning and a smaller AUC. Very bad models performing worse than random selection lie underneath the broken line, exposing a models' discriminatory power^{146,156,157}.



After model refinement a pharmacophore model is a superior tool for VS¹⁴⁵ with the windfall of unveiling possible new scaffolds or different functional groups than the originated molecules¹⁵⁸.

4.2.4. Pharmacophore based virtual screening

The tactical purpose of VS is the fast filtration of huge *in silico* compound libraries with the goal of discarding inactive, selecting active molecules and giving the possibility to focus prospective experimental efforts only on the most promising candidates to retrieve both a maximum amount of bioactive molecules in the experiments and reveal new bioactive scaffolds¹⁵⁹.

VS is knowledge based and can only be done with information either on the targeted protein (target based VS) or/and on a set of reliable ligands of the target protein (ligand based VS). One of the most common approaches of VS is the pharmacophore based VS, next to docking, QSAR and shape- matching methods¹⁶⁰.

A recent study by Karaboga and coworkers comparing different VS approaches showed that functional based pharmacophore VS may be superior over others¹⁴⁵. The opinions on whether pharmacophore or docking searches are the best VS approaches are contrary, although they are obviously two complementary approaches: Only around one third of all projects in pharmaceutical research can rely on X-ray structural target information and another third only on homology structural target information, making pharmacophores the primary tool in many projects¹⁶¹.

4.2.4.1. Advantages

Most importantly, the fact that a compound only has to exist virtual to get tested is an extremely resource effective approach, compared to the high costs and difficulties of maintaining and testing a real compound library. A substance can be tested in a straightforward manner on a large set of targets, neglecting common difficulties like accessibility of the compound, solubility, assay interferences and others. Finally, a model can be easily adapted to new knowledge, libraries can steadily be extended without large efforts and screenings can be easily repeated when needed^{147,159}.

4.2.4.2. Workflow

VS can be separated into library generation, query model generation (see chapter 4.2.3 pharmacophores) and finally the literal screening process, divided into pre-filtering and matching.

The screening libraries consist usually of previously generated 3D conformations for a large amount of molecules and can be generated in LigandScout via the implemented icon software. Libraries generated this way can be stored and used consistently. Another possibility would be to squeeze the conformation of the molecule “on-the-fly” in the pharmacophore fitting process, which is due to large time expenses unfavorable.

The following pre-filtering rapidly discards all molecules, which cannot be fitted into the query model and can speed up the screening very efficiently. Several filters depending on the software package are used and vary from very strict pre-filtering algorithms, which are able to accelerate screening speed very efficiently but at the price of discarding potential actives next to decoys and on the other side lossless filters like that applied in LigandScout, namely preliminary feature-count, two-point and three-point pharmacophore checks, which just discard molecules

unable to fit geometrically into the model resulting in time consuming but high qualitative pre-screening results.

After identification of all molecules which might match the query model in the pre-filtering, the conformations are tried to be aligned in the query model during the matching process. In this step the decision whether a molecule is retrieved as a hit or gets discarded is attained and is consequently a key process for the quality of the screening. While most software packages, like MOE, Phase or Catalyst realize that through searching in increasing n-point distances, LigandScout chose a different way, implementing a unique pattern recognition algorithm.

This pattern recognition algorithm considers a libraries' aligned pharmacophore only as matching, when **(1)** all the fitted feature points lie within the tolerance spheres of the mandatory query features, **(2)** no clashes between exclusion volumes or lipophilic spheres are obtained and **(3)** all limits for vector (i.e. hydrogen bond features) and plane (i.e. aromatic features) features are maintained. If all three points are achieved, the molecule is regarded as a virtual hit and the matching process for the next molecule starts.

The matching for a library molecule can just terminate when the first hit is retrieved or it can go on until the best matching conformation of the molecule is found. The latter one is surely time expensive but can be very important, because particularly flexible molecules can fit the query in multiple ways, which can lead to misleading assumptions^{146,149}.

4.2.4.3. Drawbacks

The VS success is highly dependent on the quality and the quantity of the used data¹⁶². The long list of successful approaches may hide the fact that a number of limitations and difficulties are adherent and should be taken into consideration.

Firstly, and most importantly a VS is a predictive method and one should not make the mistake of expecting to retrieve highly active molecules with it and a ranking of biological affinity is yet impossible to obtain. Secondly, especially for pharmacophore based VS, the awareness of very restrictive, high specific models will lead to hits with very similar structures without novel scaffolds, while unspecific and more selective models will retrieve a high number of false positives. Thirdly, there is a lack of information on the binding mode, just thinking of multiple binding pockets, which are not uncommon to drug targets. Fourthly, the selection of retrieved hits for experimental validation is a crucial step and biased picking is nearly inevitable.

Fifthly, the selection of the molecule set, used for building the model, is the most central step in the whole process and a lot of concern should be taken into the reliability of the data. Data from databases like ChEMBL or PubChem are not reviewed and should therefore be compared extensively to the original publications. Data from different assays and working groups are hard to compare and a very restraining data assortment should be done. An even more fatal mistake is to select actual false positive compounds, like frequent hitters, interfering with the readout, or pan-assay interfering substances (PAINS), for model generation, which would overthrow all future efforts.

Next to the reliability of the active data set, a trustworthy inactive data set is nearly from comparable importance. When using a decoy set, it should be taken into account that decoys have not been tested on the target and this could impair the model refinement¹⁶¹.

Besides the pharmacophore's lacking ability to comprise the complete biophysical nature of drug interactions and the dependence on qualitative input data, feature based pharmacophore VS is broadly implemented in drug research and if all pitfalls are taken in account or can be eluded it is a superior and rational approach for hit identification^{146,159,161}.

5. Results and Discussion

5.1. The Workflow

The objective of this study was to find novel agonists on GPBAR1, through VS of databases using ligand based pharmacophores followed by experimental validation. The workflow can be broken down to 4 levels, presented in **Figure 13**.

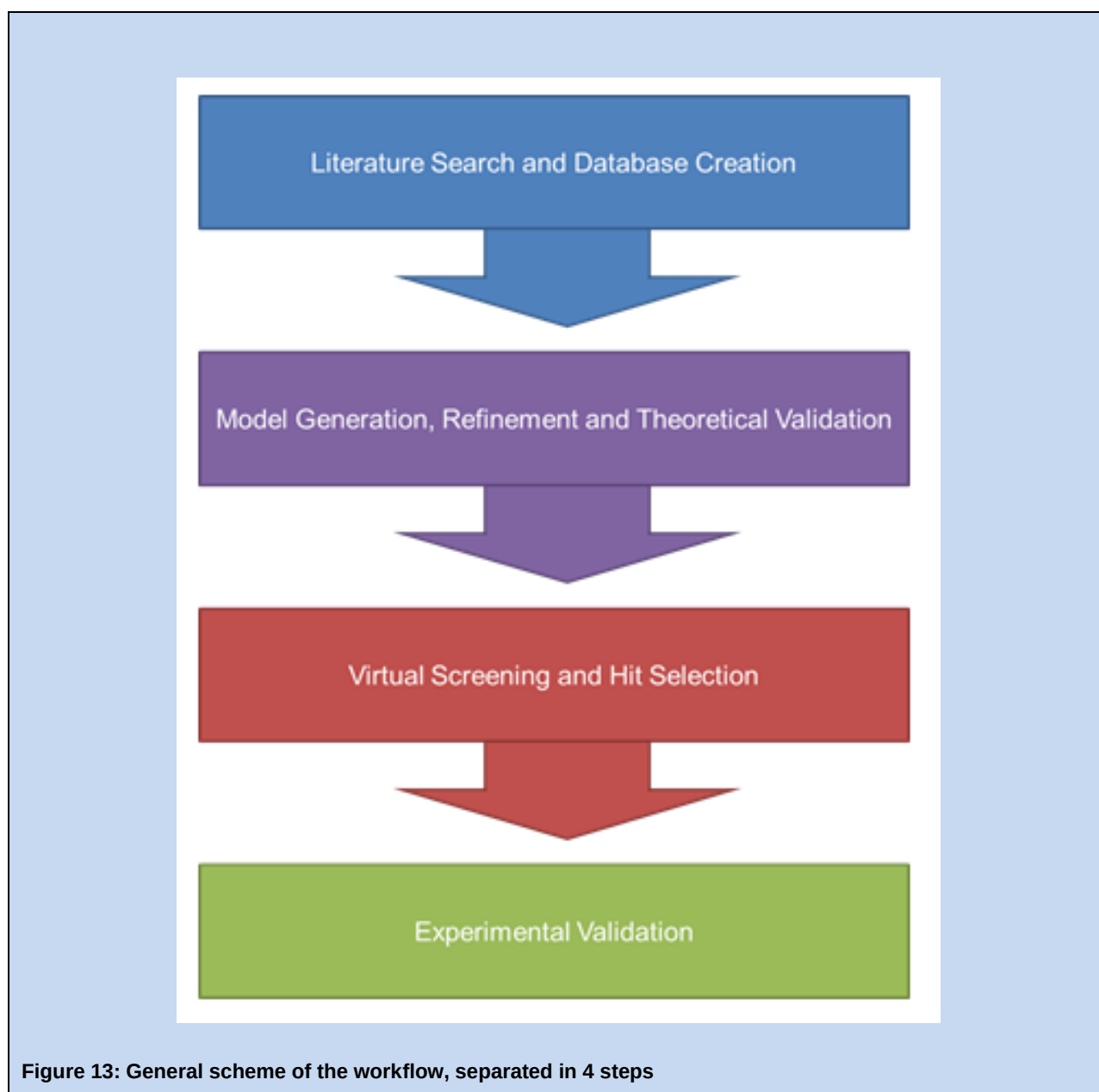
(1) Literature search and database creation both for the foundation of library generation and pharmacophore test and training sets.

(2) The generation of ligand based pharmacophore models with LigandScout 4.0 with subsequent theoretical validation and refinement, until favored theoretical performance was achieved.

(3) VS of diverse both NP and synthetic compounds databases using the models as queries. The hits for experimental testing were selected based on several criteria, namely novelty,

availability, shape-conformity, clustering of physicochemical parameters and stability. Clustering the VS-hits based on a ChemGPS-NP's PCA, with the aim of retrieving a chemical diverse hit selection, together with a subsequently ranking both through LigandScout's pharmacophore fit score and shape-matching scores, using the OpenEyes' ROCS, enabled a hitlist prioritization.

(4) The final experimental testing on HEK 293T-cells, transfected with GPBAR1-CRE driven luciferase plasmid.



5.2. Benchmark establishment - Classification of agonists from the literature

Comparing and classification of different agonists from literature in terms of their potency to activate GPBAR1 was one of the biggest challenges encountered in this work. It is of major importance that the benchmark molecules for the generation of pharmacophore models should be both very reliable and potent agonists. Particularly because the success of pharmacophore and generally VS approaches is connected to the quality of available data^{161,162}. Beyond that, a set of experimental solid inactive molecules is clearly beneficial^{147,161}. None or few of these molecules should turn up in the hitlist, when a VS run gets performed. They enable a proper and rational theoretical validation and are the key to a valuable model¹⁶¹. That is why the database of GPBAR1 has to be separated into reliable active and inactive molecules. Unfortunately, during this procedure several problems have been encountered. First of all, the EC₅₀ values were not only obtained by different working groups but also with different cellular assays, making it nearly impossible to separate them into activity classes. Differences among assays (especially in sensitivity) and the overall variability of *in vitro* screenings^{84,163} raised concerns about direct data comparison among these assays. For example, a variance range starting at 35 nM and ending at 6 µM for the endogenous ligand LCA was encountered (**Table 7**). Another very illustrative example for this was INT-777, a broadly tested agonist, with an even higher variance range from 177 nM to 13 µM (**Table 8**).

Table 7: Variance span of reported EC₅₀-values for lithocholic acid encountered in different studies

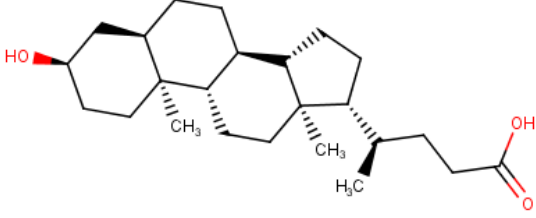
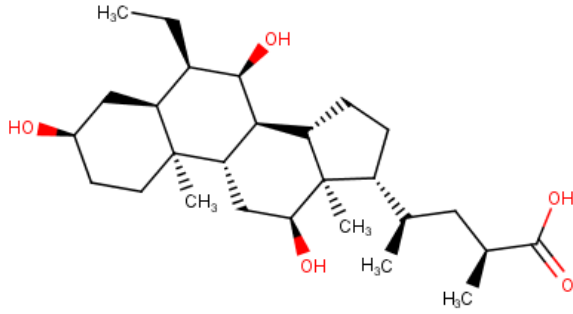
 Lithocholic acid			
Source	EC ₅₀ value for LCA	cell line	Assay Type
2002Maruyama	35nM	CHO	CRE driven Luciferase assay
Ref314Scif	180nM	HEK	CRE driven Luciferase assay
Ref176Scif	446nM	CHO	CRE driven Luciferase assay
Ref176Scif	571nM	CHO	TR- FRET assay
ChEMBL1144009	580nM	CHO	CRE driven Luciferase assay
Ref415Scif	580nM	CHO	CRE driven Luciferase assay
ChEMBL3399951	680nM	HEK	cAMP assay
ChEMBL3097990	730nM	CHO	CRE driven Luciferase assay
ChEMBL2062395	5000nM	NCI H716	TR FRET assay
ChEMBL1156761	5600nM	CHO	CRE driven Luciferase assay
Ref160Scif	6000nM	NCI H716	HTR- FRET assay

Table 8: Variance span of reported EC₅₀-values for INT-777 encountered in different studies

 INT-777			
Source	EC ₅₀ value for INT-777	cell line	Assay Type
CHEMBL2331085 (1)	166nM	CHO -doxycycline induced expression-	HTRF assay
ChEMBL3097990	180nM	CHO	CRE driven Luciferase assay
CHEMBL3352618	293nM	HEK	cAMP level assay
ChEMBL2062395	820nM	NCI H716	TR FRET assay
ChEMBL1154755	820nM	NCI H716	HTR FRET assay
Ref160Scif	900nM	NCI H716	HTR FRET assay
Ref6Scif	1226nM	HEK	CRE driven Luciferase assay
Ref201Scif	1226nM	HEK	CRE driven Luciferase assay
CHEMBL2434932	1230nM	HEK	CRE driven Luciferase assay
CHEMBL2331085 (2)	5540nM	CHO -uninduced expression-	CRE driven Luciferase assay
CHEMBL2406956 (3)	13000nM	NCI H716	cAMP level assay

Facing these problems an easy splitting of compounds into higher or lower actives by reported EC₅₀ values was not possible. The compounds had to be separated manually, relying on few assumptions.

(1) LCA was seen as a highly active agonist, due to the broad acceptance in the community as the most potent bile acid. Compounds showing lower or comparable EC₅₀-values in the same studies and the same assays than LCA can be seen therefore as **High Actives**. **(2)** DCA and CDCA can be assumed as intermediately active but in the μ M range; CA as low active in the 10 μ M range respectively. Compounds showing comparable EC₅₀ in same assays and studies can be assumed to be indeed active but not high active ones. Therefore, they were not admitted to the High Active group, instead they established the foundation for the **Actives** group. **(3)** All compounds tested clinically, broadly used as internal positive controls or had shown potent activity *in vivo* (i.e. INT-777, compare **Table 8**) were also accepted as **High Actives**. **(4)** The

few compounds explicitly stated as not active, showing no response or are known from internal data to be completely inactive are seen as **True Inactives**. **(5)** Compounds without activation in highly sensitive assays overexpressing GPBAR1 can be seen as really **True Inactives**. **(6)** Other compounds showing activity far lower than cholic acid (i.e. at 30 μ M) were assumed to be inactive but were not moved to the reliable true inactive database. Hence, they were moved to a fourth group, the **Inactives**. **(7)** All patent compounds, due to absence of information on assay type, experimental workflow (internal standard, three-point validation) and assumption of low quality HTS data, were seen as surely active but could not be assumed as **High Actives**. Consequently, they were added to the group **Actives**. **(8)** All studies reporting active compounds but lacking the report of an internal standard, three-point validation or have not given sufficient information on the assay type (besides other quality standards) were not moved into the **High Actives** or the **True Inactives** database, assuming them to be **Actives** or if stated as inactive into the **Inactives** database. **(9)** Studies, which showed a sufficient reported workflow, but had not a good standard for comparison were compared through duplicates in other studies, if this was possible, and were eventually added to the reliable **High Actives** or **True Inactives** databases. **(10)** All questionable studies and doubtful compounds were not shifted into the True Inactive or Highly Active database. **(11)** EC₅₀ values obtained in unusual assays were not used, because the data comparison and classification was impossible.

Severity and rigor selection were the first priorities in this step, leading to four databases (**Table 9**) of different size and with a strong shift into the direction of **High Actives** and **Actives**, due to the absence of negative results in the literature.

5.3. Benchmark establishment - Generation of decoys from DUDe

The quality assessment of pharmacophore models during their generation, their refinement and their theoretical validation is based on quality metrics, especially the Enrichment Factor. Enrichment can be seen as the quantification of a models ability to identify known ligands in a background of decoy molecules¹⁴⁷. Lucid enrichment measurement therefore not only relies on the quality of the agonist set but also on a vital selection of decoys¹⁶⁴. Although most of the experimental outcomes can be assumed as negative¹⁶⁵, the research showed a strong shift towards very active molecules. Almost no negative bioactivity data was available (**see chapter: 5.2 Benchmark establishment - Classification of agonists from the literature**). Therefore, this work had to resort to generated decoys.

The freely-available online-tool from the directory of useful decoys (DUDe) (<http://decoys.docking.org>) has been widely used^{166,167} and was a beneficial option for finishing the benchmark. The directory is based on 2950 active molecules on 40 target-proteins, based on data from literature. For every active molecule in the directory 36 decoys with same physical properties but divergent topology were included, resulting in a database of 95.316 decoys, all of them with drug-like properties¹⁵². In 2012 this database was enlarged reaching a total of 102 targets (but only 5 GPCR) and 22,886 ligands, with 50 decoys for each ligand (combined more than a million decoys)¹⁶⁴. Based on the physicochemical properties [i.e. MW, clogP, number of hydrogen-bond acceptors (HBA) or hydrogen-bond donors (HBD)] of the **High Actives** molecule dataset matches were appointed. 2D similarity fingerprints should minimize the topological resemblance of decoys and actives. This approach should guarantee both inactive and for the modeling process demanding (in terms of high physicochemical similarity) decoys¹⁶⁴.

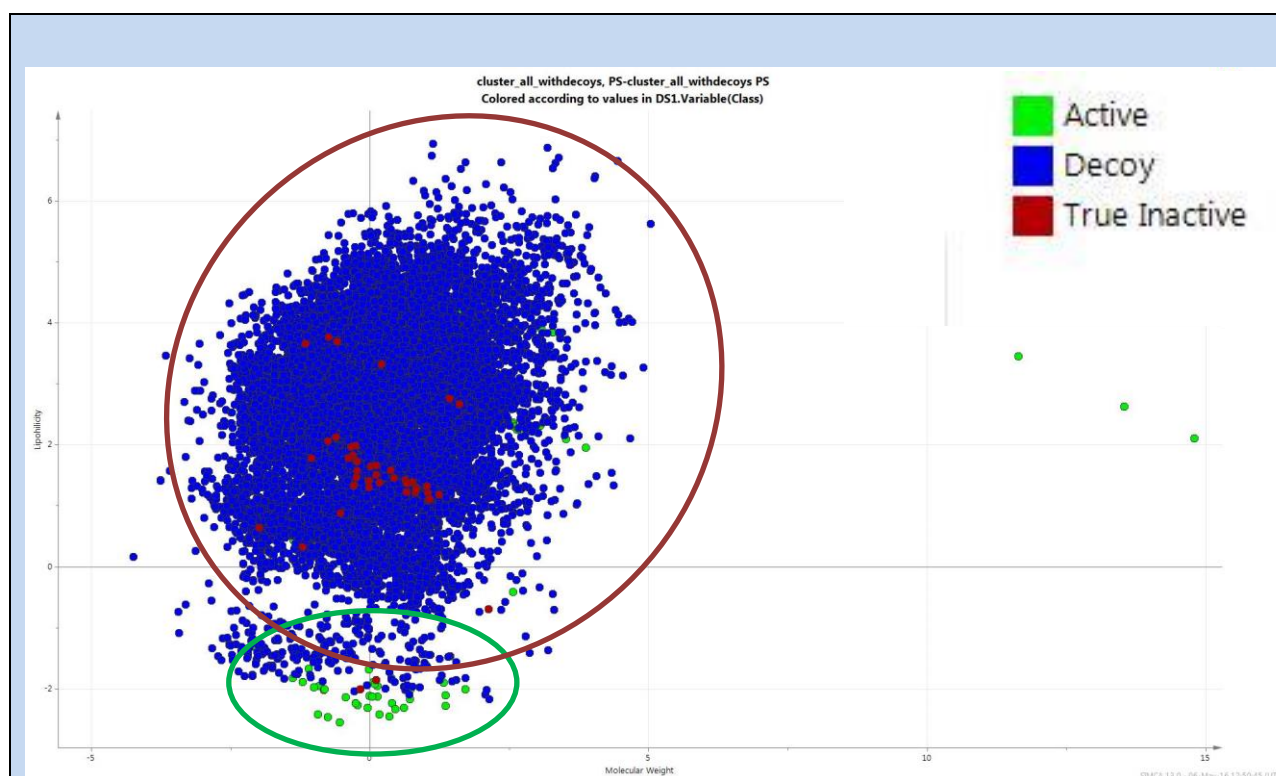


Figure 14: When the chemical space occupied by all molecules in the workbench is visualized in a very simplified form (plotted PC3 relating to lipophilicity against PC1 corresponding to properties like MW) one can see that the DUDe website performs very well to produce decoys (blue) with the same chemical space as the synthetic compounds (brown cluster) but struggles to provide as good decoys with the natural compounds

Nevertheless, skepticism is indicated, when working with benchmark molecules that have never been tested on your pinpoint target. That is why we performed a PCA of all molecules in the benchmark (DUDe Decoys, True Inactives and High Actives) through the chemGPS online tool, plotting the results in SIMCA both 2D (PC1 and PC2) and 3D (PC1-PC3) and compared the physicochemical space of the molecules. This procedure revealed the following characteristics: The DUDe-decoys are clearly very effective in straddling above the chemical space of agonists, derived from synthetic sources, but struggle to meet the chemical space of steroids and triterpenes. In fact, those decoys, which should match the NPs, are shifted a little bit into the higher PC2 (describing aromaticity, conjugated double bonds) area, which is already apparent in 2D plotting (**Figure 14**). There is also a good reason to believe that the magnificent theoretical enrichment metrics of those models, describing the pharmacophores of bile acids (model **BAMS22**) and triterpenes (**Model12.3.8.**), has to be interpreted with caution and suspicion.

A lot of effort has been undertaken to address the creation and quality assessment of decoy sets for docking and pharmacophore based VS. An exemplary and interesting approach was presented by Rohrer and Baumann in 2009¹⁶⁸: The Maximum Unbiased Validation data set (MUV) is a directory of real inactive molecules, derived from HTS data, unluckily with only 18 targets included. Wallach et al¹⁶⁹ presented their virtual decoy set, which is stated to be less biased on physical similarity than the DUDe database but the realizability of decoy synthesis is stated to be questionable¹⁷⁰. The GPCR ligand library (GLL) and GPCR Decoy Database (GDD), both freely available to the scientific community, address especially GPCR and would have been a competent alternative¹⁷¹. The DecoyFinder moreover tries to selectively create decoys for a specific target¹⁷² and is also freely available. The main focus of all these approaches is still docking, while application to ligand-based VS is limited, because it is commonly used, when no crystal structure is available, as in this work¹⁷⁰.

To sum up, it can be stated that there is a broad range of possibilities to create benchmark sets, but hopefully future works can be done with experimental derived inactive molecules and not with calculated decoys. Awareness for the requirement of negative results is already pervasive in the scientific community and extensive discussion on this topic has already been initiated¹⁶⁵. Until then it can be stated that the directory of useful decoys is a valuable source for decoy-collections but has to be used carefully.

The five classified or generated compound groups are described in **Table 9**, including the three databases used for ligand-based pharmacophore modeling, refinement and quality assessment.

Table 9: The five compound groups, including the three databases (*) used for ligand-based pharmacophore modeling, refinement and quality assessment with the requirements and scope.

Library	Content/ requirements	Size
HIGH ACTIVES*	Highly active compounds with higher or the same/comparable activity as LCA. Broadly <i>in vitro</i> tested compounds or compounds tested in clinics (i.e. SBSB756050, INT-777, INT767, RO5527239) Information on their reported activity was fully reliable, information on the experimental evaluation was conclusive and comparison to other study results had been possible. This database was used for model generation and improvement.	338
ACTIVES	Active Compounds with lower activity than LCA, comparable to other bile acids (DCA, CDCA, CA). Patent compounds, compounds obtained in unqualified assays and/or with inconclusive or questionable information on the protocol. Due to their inadequate quality, this database was not used for model generation or refinement.	295
INACTIVES	Inactive compounds and probably very low active compounds. The reliability of their inactivity was not given due to inconclusive comparability/lacking information/ inadequate assay type. Due to their insufficient quality, this database was not used for model generation or refinement.	133
TRUE INACTIVES*	Inactive Compounds retrieved from in- house data or from fully reliable studies. This database was used to improve the models via theoretical validation and acted as decoys	49
DUDE DECOYS*	Decoys from the DUDe website sharing the same physicochemical properties than the Highly Actives but with different 2D topology This database was used to improve the models via theoretical validation and acted as decoys.	18043

5.4. Conformational sampling

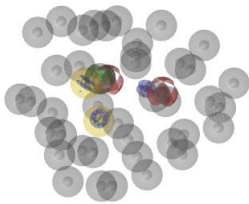
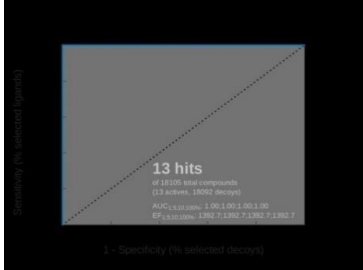
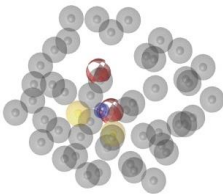
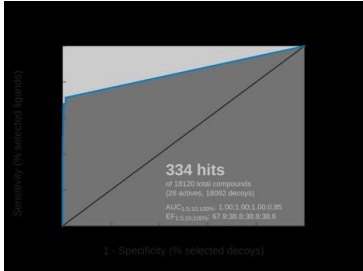
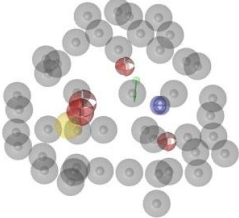
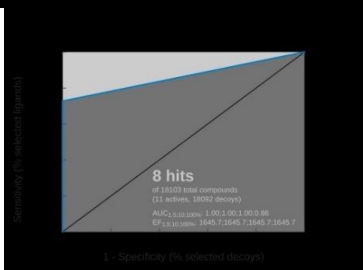
A crucial step in pharmacophore modeling and VS is the reliable generation of putative conformers that conceal satisfactorily the conformational space. A supposed molecule with only three rotatable bonds can be mentioned as an example: If every rotatable bond is sampled in 10° intervals, it would give rise to 46 656 (363) conformations. On the other hand and even more importantly, the generated conformers should include the molecule's bioactive conformation or conformations. Scior et al. describe several parameters necessary to achieve this goal¹⁶¹: (1) sampling algorithms and their specific parameters; (2) strain energy cutoffs; (3) maximum number of conformations per molecule; and (4) rmsd thresholds to remove duplicates. Probably one of the best validated and acknowledged conformer generators is Omega from OpenEye¹⁷³. It is stated that Omega is able to reproduce 83% of tested protein database (PDB) structures and 84% of Cambridge Structural Database (CSD) structures within 1Å rmsd of the experimental retrieved structures, when the default settings are used and 200 conformations are calculated. A proficient freely available alternate conformer generator is RDKit, which showed its capacity to generate 70.0% of the 160 PDB structures and 90.4% of the 469 CSD structures within 1 rmsd deviation from the crystal structures¹⁷⁴. However, the implemented conformer generator in LigandScout iCon just showed recently its ability to reproduce crystallized experimental structures, as well as Omega, for equivalent settings. Additionally, iCon achieved comparable results concerning the overall performance via a smaller number of generated conformers. This information was privately shared by Dr. Thomas Seidel and Univ.-Prof. Mag. Dr. Thierry Langer. Accordingly, the conformer generator implemented in LigandScout was used for this study.

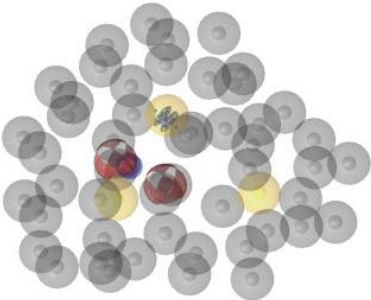
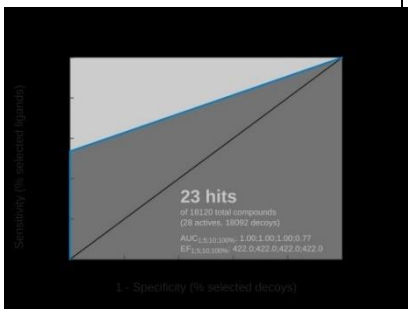
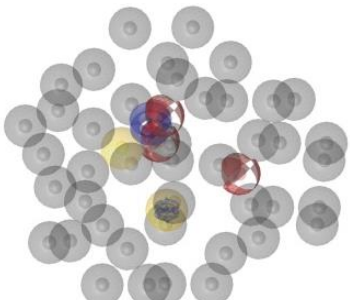
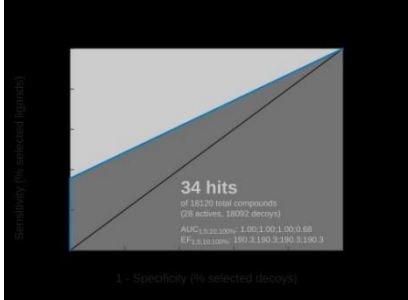
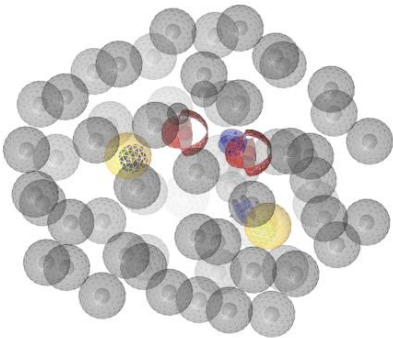
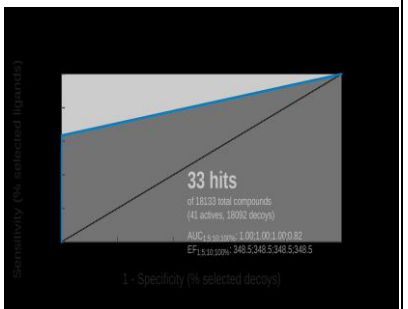
5.5. Theoretical model performance

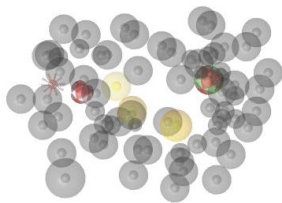
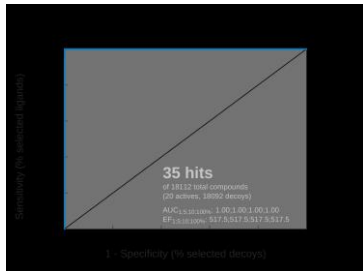
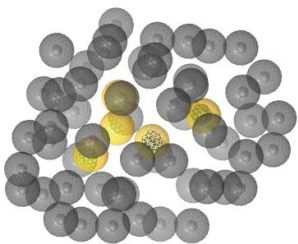
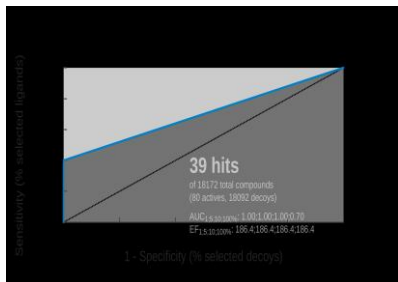
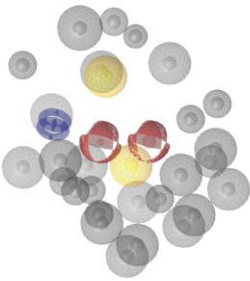
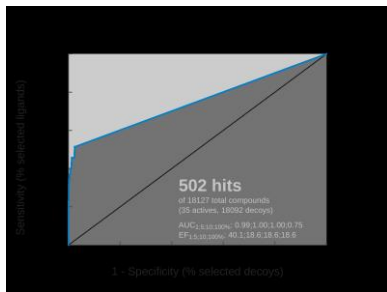
5.5.1. The initial model collection

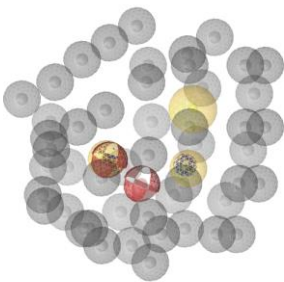
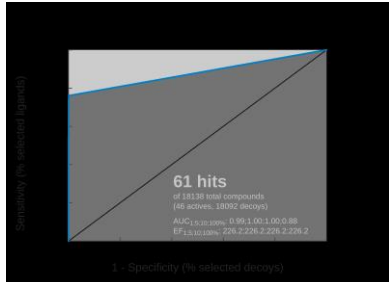
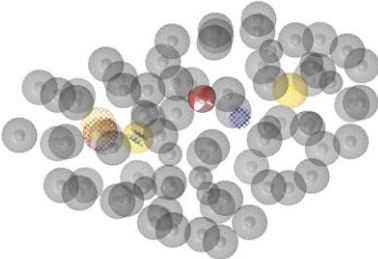
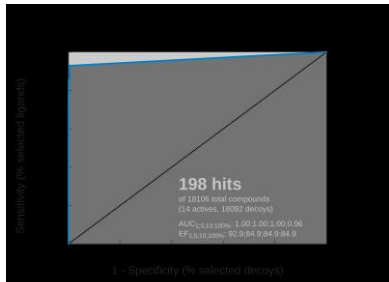
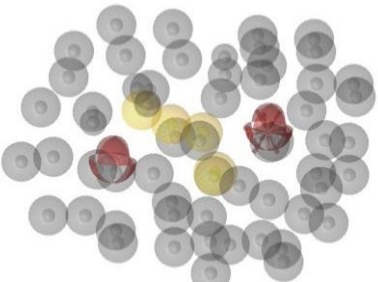
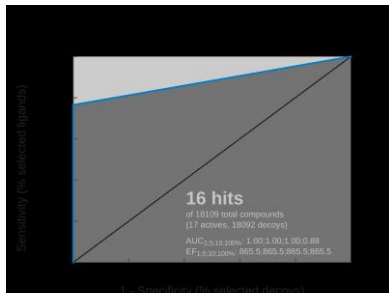
As described in detail in Material and Methods section **6.5.2 Clustering**, the high active molecules got clustered in LigandScout 4.0 into 12 clusters. Only cluster 1 was completely discarded, since it only consisted of a single molecule (CAS#1698050-57-5). A model set was generated, which aimed to describe the whole structural diversity of all GPBAR1 ligands in our **High Active** dataset, either consisting of one or more models. In **Table 10** the model set is described briefly with their templates, their pharmacophore features and two very important theoretical quality metrics the EF and the ROCS curve, retrieved from screening the cluster against the DUDE DECOYS and TRUE INACTIVE databases. A table of the template's chemical structures is presented in the **Material & Methods 6.5.2. Training Sets for Modeling**.

Table 10: Model set with brief description of the individual models generated in LigandScout 4.08

Model name Templates of the training set	Figure of the model pharmacophore's features	ROC EF
Cluster 2		
Model2best		
CHEMBL2407936 CHEMBL2407949 CHEMBL2407945 CHEMBL2407951 CHEMBL2408107	2 lipophilic features 2 hydrogen-bond acceptor features 1 hydrogen-bond donor feature 3 aromatic ring features	EF=1392.7
Cluster 3		
ModelC3SEL		
CHEMBL2435937 CHEMBL2435942 CHEMBL2435946 CHEMBL2435924 CHEMBL2435940 CHEMBL2435926 CHEMBL2435929 CHEMBL2435927	2 lipophilic features 2 hydrogen-bond acceptor features 1 aromatic ring feature	EF=67.9
Cluster 4		
Model CI4U		
CHEMBL3234568 CHEMBL3234571 CHEMBL3234575 CHEMBL3234563 CHEMBL3234582 CHEMBL3234583 CHEMBL3234572 CHEMBL3234569	1 lipophilic feature 4 hydrogen-bond acceptor features 1 hydrogen-bond donor feature 1 aromatic ring feature	EF=1645.7

Cluster 5		
Model CI5ISO		
Ref176Scif_TRC210258 ChEMBL3234564 ChEMBL3234561 ChEMBL3234562	3 lipophilic features 2 hydrogen-bond acceptor features 2 aromatic ring features	EF=422
Model P5		
ChEMBL3234564 ChEMBL3234554 ChEMBL3234552	2 lipophilic features 2 hydrogen-bond acceptor features 2 aromatic ring features	EF=190.3
Cluster 6		
Model 6.2		
Ref6Scif_31d ChEMBL3321758 Ref6Scif_30i	2 lipophilic features 2 hydrogen-bond acceptor features 3 aromatic ring features	EF=348.5

Cluster 7		
Model BAMS22		
CHEMBL270105 CHEMBL270287 CHEMBL1478-LCA CHEMBL410893 CHEMBL407717 CHEMBL567640-INT777	2 lipophilic <i>mandatory</i> features 1 lipophilic <i>optional</i> feature 2 hydrogen-bond acceptor features 1 hydrogen-bond donor <i>optional</i> feature 1 <i>optional</i> negatively ionizable feature	EF=516.2
Cluster 8		
Model 8M2		
Ref178Scif_29 Ref178Scif_17 Ref201Scif_32g Ref201Scif_27	5 lipophilic features 2 aromatic ring features	EF=186.4
Cluster 9		
Model 9-x_3		
CHEMBL601882 Ref135Scif_15 CHEMBL612138 CHEMBL3321849 CHEMBL3234579 CHEMBL3234868	2 lipophilic features 2 hydrogen-bond acceptor features 1 aromatic ring feature	EF=18.53

Cluster 10		
Model 10.z.		
CHEMBL2181228 CHEMBL2181221 CHEMBL3421896	3 lipophilic features 2 hydrogen-bond acceptor features 2 aromatic ring features	EF=226.2
Cluster 11		
Model 11.9		
CHEMBL2331645 CHEMBL2331654 Ref17Scif_79a	2 lipophilic <i>mandatory</i> features 1 lipophilic <i>optional</i> feature 1 hydrogen-bond acceptor features 1 hydrogen-bond donor optional feature 1 aromatic ring <i>mandatory</i> features 1 aromatic ring <i>optional</i> feature	EF=84.85
Cluster 12		
Model 12.3.8		
CHEMBL571592 CHEMBL583203 CHEMBL269277- BA CHEMBL169- UA	4 lipophilic features 2 hydrogen-bond acceptor features 1 negative ionizable group	EF=865.5

5.5.2. Theoretical performance of the models

Although showing fine metrics, very restrictive models do not have the capability to find novel hits with unknown structures, which can be seen as a major task of VS¹⁶¹. So, on the one hand restrictive models cause low hit-rates in prospective VS experiments, but on the other hand they ensure a low false-positive rate, a major advantage for later validation with expensive and rare NPs. Since all models describe quite specific chemical scaffolds most of the developed models can be seen as too restrictive. They selectively describe a ligand set but they would not be sufficient models for VS, since they would not retrieve many hits and especially are not capable to find hits with new scaffolds.

Model2best and Model CI4U are the most restrictive models of the set. They just describe simply the molecules from their own clusters. They achieve the best EFs but are not capable to find novel hits (or hits at all) in prospective VS. Model 12.3.8 shows also very restrictive metrics but proved to find novel hits in the experimental validation, although hits with divergent structures had been scarce. It can be described simply as a capable VS query to check a putative GPBAR1 activating ability of triterpenoid molecules. In this sense it is a valuable tool for further research (i.e. forthcoming activity profiling of plant extracts).

At the other end of the model set is Model 9-x_3, Model 11.9 and Model C3SEL. They find hits from different clusters. They not only describe their queries pharmacophores, as they seem to find 3D pharmacophore patterns synthetic active GPBAR1 ligands have in common. It is questionable if those patterns are responsible for their activity or just coincidence. Especially, ModelC3SEL and Model 9-x_3 have unrestricted exclusion volume spheres and their 3D pharmacophores are approachable to lots of molecules, as they both consist of 5 comparable tolerant features. For the purpose of this work, their bad metrics made them unfavorable for the application as prospective VS queries. Nevertheless, as models applied upstream HTS with the aim to perform scaffold-hopping those 3 models would be attractive.

The intermediate model BAMS22 proved to be a proficient VS filter, as the theoretical validation showed both good hit-rates and consisted of quite common pharmacophore features. The exclusion volume sphere is restrictive but the actual 3D pharmacophore 2 HBD features and 2 lipophilic features are quite common to lots of molecules. This fact and the fact that they cover the second big group of NP GPBAR1 agonists (those of BA and steroids) made them the most promising model for experimental validation. The other models with comparable metrics as BAMS22 would also be very interesting for theoretical validation but when applied to

prospective VS with natural product databases they have very low hit-rates, because their benchmark molecules inhabit a completely different space as NPs and this gets visible at pharmacophore based VS.

The 12 models cover complementary structural space and therefore found combined 81.85% of the known structures from the **High Active** database. As the whole benchmark-dataset (**High Actives, True Inactives, DUDE Decoys**) was screened the models in chorus achieved following metrics described in **Table 11**. As one can see an EF of 11 is not advantageous but this is firstly a cause of the unselective models, described previously. Secondly, it has to be taken in concern that a selection of 13 models covering a wide diversity of 3D pharmacophore spaces will unavoidable lead to high false positive hit-rates.

Table 11: Performance of the models in chorus, described in hit-rates and quality metrics (High Actives, True Inactives, DUDE Decoys)

High active hits	275
True inactive hits	4
Dude-Decoy hits	1033
Enrichment factor	11.22
Selectivity	0.81
Specificity	0.94

Table12: Performance of the individual models, described in hit-rates and quality metrics (High Actives, True Inactives, DUDE Decoys)

Model2best	
High active hits	15
True inactive hits	0
Dude-Decoy hits	0
Enrichment factor	53.53
Selectivity	0.04
Specificity	1
Model C3SEL	
High active hits	56
True inactive hits	0
Dude-Decoy hits	314
Enrichment factor	8.10
Selectivity	0.17
Specificity	0.98
Model CI5ISO	
High active hits	24
True inactive hits	0
Dude-Decoy hits	30
Enrichment factor	23.79
Selectivity	0.07
Specificity	1

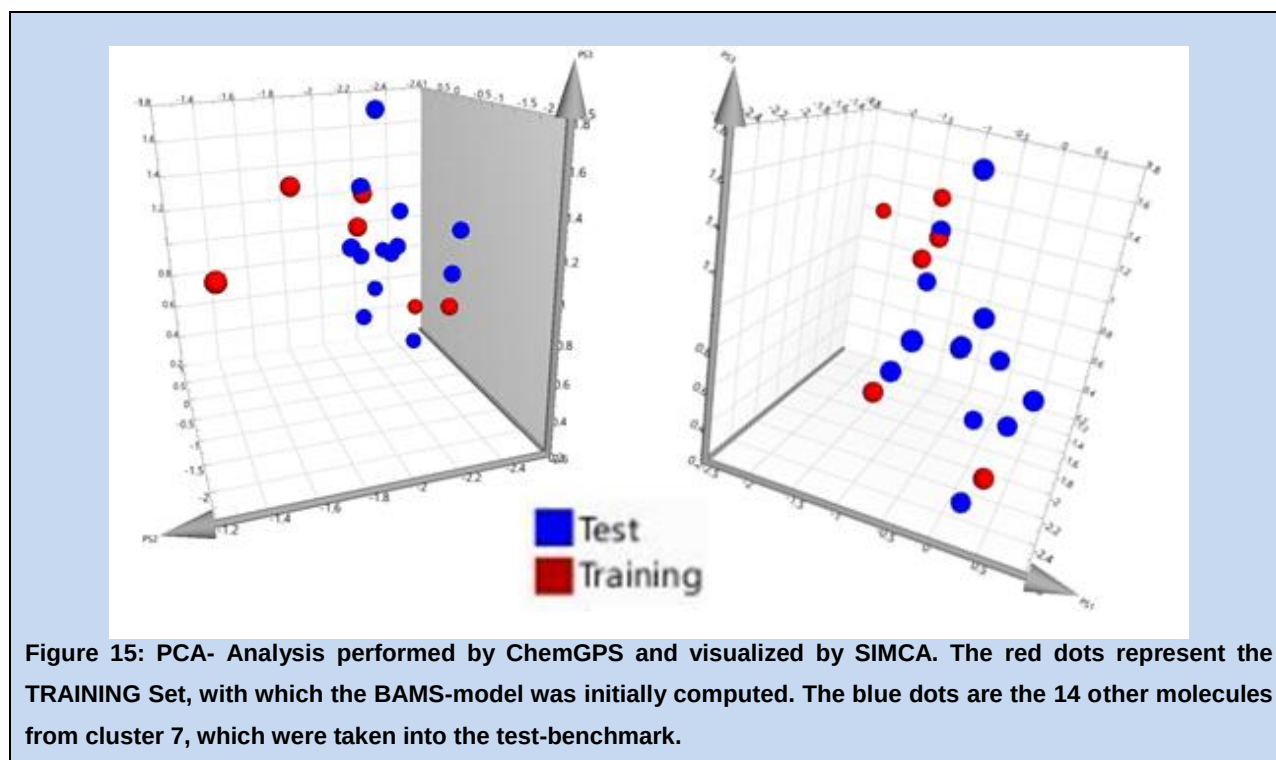
Model P5	
High active hits	12
True inactive hits	0
Dude-Decoy hits	24
Enrichment factor	17.84
Selectivity	0.04
Specificity	1
Model 6.2.	
High active hits	52
True inactive hits	1
Dude-Decoy hits	6
Enrichment factor	47.17
Selectivity	0.15
Specificity	1
Model BAMS22	
High active hits	21
True inactive hits	0
Dude-Decoy hits	15
Enrichment factor	42.42
Selectivity	0.06
Specificity	1
Model 8M2	
High active hits	42
True inactive hits	1
Dude-Decoy hits	10
Enrichment factor	42.42
Selectivity	0.12
Specificity	1
Model 9-x 2	
High active hits	56
True inactive hits	1
Dude-Decoy hits	483
Enrichment factor	5.55
Selectivity	0.17
Specificity	0.97
Model 10Z	
High active hits	25
True inactive hits	1
Dude-Decoy hits	26
Enrichment factor	25.73
Selectivity	0.07
Specificity	1
Model 11.9.	
High active hits	17
True inactive hits	0
Dude-Decoy hits	185
Enrichment factor	4.50
Selectivity	0.05
Specificity	0.99
Model 12.3.8.	
High active hits	13
True inactive hits	0
Dude-Decoy hits	3
Enrichment factor	0.04
Selectivity	1
Specificity	42.42

5.6. Detailed model description

The prospective VS was later performed with two models, describing bile acid and steroid-like structures (**BAMS22**) and triterpenes (**Model12.3.8**). The following chapter discusses generation and refinement for these two models, since they merely cover those scaffolds relevant for secondary metabolites from natural sources.

5.6.1. Cluster 7- Generation of BAMS22

The initial pharmacophore model for cluster 7 was built with six molecules as templates: CHEMBL270105, CHEMBL270287, CHEMBL1478-LCA, CHEMBL410893, CHEMBL407717 and CHEMBL567640-INT777. These templates had been selected on the one side for covering nearly the whole physicochemical space (**Figure 15**) and on the other side for incorporating all the structure-activity information contained in this set of molecules. For example, CHEMBL407717's ring A and B are configured *trans* rather than the other compounds, which cycles A/B are configured *cis*. This makes it impossible to congruently align CHEMBL407717 with the other molecules. For this reason, CHEMBL407717 had to be taken into the TRAINING-SET. Otherwise the information, saved in this deviation, would not be incorporated or would have to be laboriously inserted later in the model refinement step. CHEMBL567640-INT777 was taken into the Test-set, because of the three hydroxyl-moieties and the 6 α -ethyl-group. LCA and CHEMBL270287 were chosen, because their position in the chemical space is quite on an average compared to the whole molecule set. Thus they are likely to describe the cluster best. CHEMBL270105 and CHEMBL410893 were selected, since they fill up the openings in the chemical space best.



The initial models were generated with LigandScout 4.08 as shared or merged feature pharmacophores with different settings. After validating the models, Bile-Acid-Model-Shared feature pharmacophore (**BAMS1**) was chosen as a starting point for the model optimization and refinement. In **Table 13** one can see the settings used with which **BAMS1** was computed.

Table 13: Settings of the initial model generation for Cluster 7 (LigandScout 4.08)

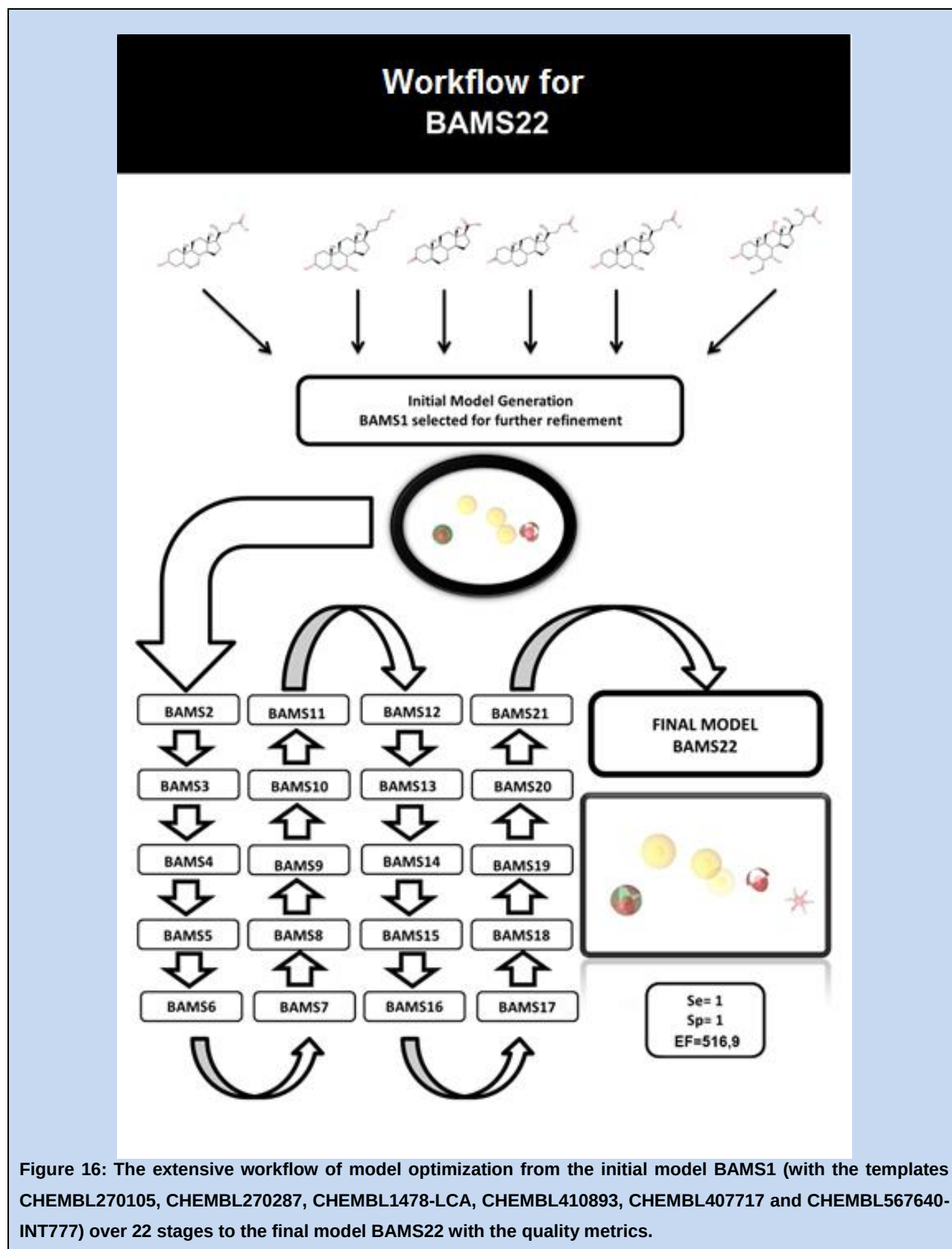
BAMS1 settings	
Pharmacophore Type	shared feature pharmacophore
Scoring Function	pharmacophore fit and atom overlap
Conformer Generator	iCon
Max. Number of conformations	200
Max Number of result Pharmacophores	3

BAMS1 consists of three mandatory hydrophobic features, two mandatory hydrogen-bond acceptors and one mandatory hydrogen-bond donor feature. It was primarily able to find 18 out of the 20 molecules (CHEMBL270105 and CHEMBL410893 were removed before) belonging to cluster 7 and further identified TLC from cluster 12 as ligand. It is very obvious that TLC shares the same pharmacophore characteristics, as the other molecules from cluster 7, not cluster 12. CHEMBL410893 and CHEMBL407717 have no 3-hydroxyl-group and therefore no hydrogen-

bond acceptor at position (-7.14/ 3.83/ 2.63). Both of them have a very high activity of EC_{50} = 240 nM for CHMEBL410893⁴⁰ and EC_{50} = 290 nM for CHEMBL407717, respectively⁴⁰. The hydrogen-bond accepting feature in this position is probably not that important for achieving high activity. BAMS1 only got four false positives (out of 18112) with an almost ideal specificity of nearly 1 and an EF of 740.95. Further optimization steps were intended to gain better selectivity, while keeping the high specificity. The whole workflow of **BAMS22**'s model optimization is depicted in **Figure 16**.

First of all, the hydrogen-bond donating group at (-7.14/ 3.83/ 2.63) was set to optional [BAMS2]. The true positive hit-rate rose to 19 but the false positive also from 4 to 29. Exclusion volumes at (-7.45/-5.78/6.69) with Weight 1 and Tolerance 2.05, at (-9.99/1.88/1.11) with Weight 1 and Tolerance 1 [BAMS3], (-6.49/6.29/1.37/Weight 1/Tolerance 1) [BAMS4], (-7.24/6.13/3.35/Weight 1/Tolerance 1) [BAMS5], (-8.2/5.12/0.29/Weight 1/ Tolerance 1.15) [BAMS6] were added to reduce the false positive rate from 29 to 19. The tolerance of the hydrogen-bond accepting group's tolerance sphere at (3.96/ -1.82/ -1.05) was reduced from 1.5 to 1.2 [BAMS7]. The lipophilic feature's tolerance at (-3.83/-0.01/5.26) was also reduced to 1.35 and another exclusion volume was added at (-9.89/0.47/4.25) [BAMS8]. The exclusion sphere was further restricted with exclusion volumes at position (1.08/ -4.51/ 1.25/ Weight 1/ Tolerance 1), (2.70/ -1.36/ 3.66/ Weight 1/ Tolerance 1), (-3.64/ -1.02/ 7.99/ Weight 1 and Tolerance 1) and (-7.44/ 2.75/ 0.85/ Weight 1/ Tolerance 0.7) [BAMS9]. An optional negative ionizable group was added at position (4.51/ -5.00/ -4.49/ Tolerance 1.3/ Weight 1), further three exclusion volumes at position (-2.15/ -4.02/1.98/ Tolerance 0.7/ Weight 1), (-9.89/ 3.05/4.27/ Tolerance 1/Weight 1) and (-9.63/ 1.89/ 6.34/ Tolerance 1/ Weight 1) [BAMS10 and BAMS11]. The tolerance of the hydrophobic feature at position (-3.83/ -0.01/ 5.26) was scaled up from 1.35 to 1.65 [BAMS12] and the hydrophobic feature at position (1.75/ 0.96/ 0.52) was set from a mandatory to an optional feature [BAMS13]. Afterwards, further exclusion volumes at position (-6.69/3.69/ 7.44/ Tolerance 1/ Weight 1) [BAMS14], at position (2.29/-3.31/0.83/ Tolerance 1/ Weight 1) [BAMS 15], at position (-5.70/ 4.48/ 0.38/ Tolerance 1/ Weight 1) [BAMS 16], at position (3.37/ 0.10/-5.09/ Tolerance 1/ Weight 1) [BAMS 17], at position (3.79/ 3.49/ -1.68/ Tolerance 1/ Weight 1) [BAMS 18], at position (-8.12/ 0.72/ 6.15/ Tolerance 1/ Weight 1) [BAMS 19], at (-4.87/ 0.55/ 8.14/ Weight 1/ Tolerance 1) [BAMS20], at (-5.03/ 4.86/ 6.79/ Weight 1/ Tolerance 1) [BAMS21], and finally at position (5.58/ -0.83/-4.91) were added in order to increase the specificity again, which was lost when removing the mandatory hydrophobic feature at BAMS13. The final model BAMS22 trialed against the set of decoys/true inactives (n=18.112) and cluster 7 (without CHEMBL270105 and CHEMBL410893) showed Sp=1(0.998785) and Se=1 achieving an EF=823.3. Next to the molecules from cluster 7 the

CHEMBL408701-TAURO-LITHOCHOLIC ACID from cluster 12 were also found and ranked high. **Figure 17** shows the final pharmacophore model with all pharmacophore features and the restrictive exclusion volume coat. **Table 14** and **Figure 18** show the hit-rates and quality metrics of **BAMS22**, when **Cluster7** is screened against **DUDE DECOYS** and **TRUE INACTIVES**.



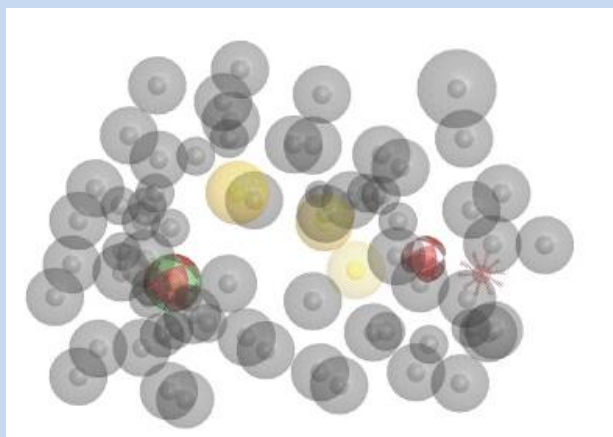


Figure 17: Model BAMS22 comprises a restrictive exclusion volume coat (grey), especially at the position of the HBA/HBD area (green/red) on the left side. It further consists of two mandatory lipophilic features (yellow), one optional lipophilic feature (yellow, reduced brightness), two HBA features (red), one optional HBD feature (green) and one optional positively ionizable feature (red star).

Table 14: Hit-rates and quality metrics of BAMS22, when Cluster7 is screened against DUDE DECOYS and TRUE INACTIVES

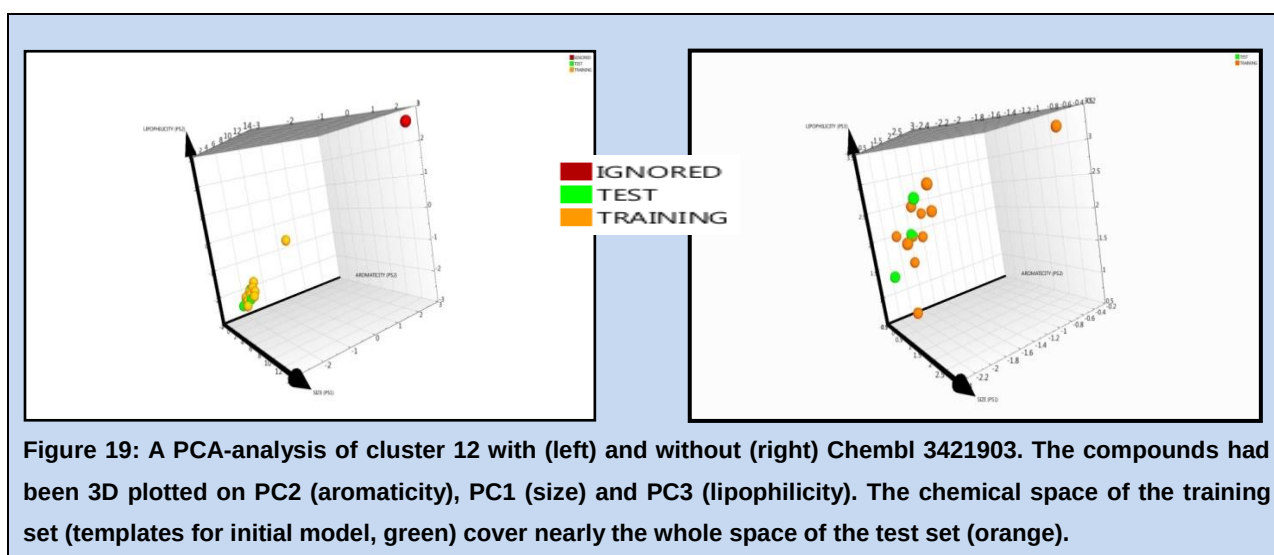
Enrichment Merics	
High active hits	20
True inactive hits	0
Dude-Decoy hits	15
Enrichment factor	516.91
Selectivity	1
Specificity	1



Figure 18: ROCs curve, when Cluster7 is screened against DUDE DECOYS and TRUE INACTIVES

5.6.2. Cluster 12- Generation of Model 12.3.8

The initial pharmacophore model for cluster 12 was built with 4 molecules as templates: CHEMBL583203, CHEMBL571592, CHEMBL269277- BETULINIC ACID and CHEMBL169- URSOLIC ACID. The templates had been selected on the one hand for covering the fast majority of the physicochemical space (**Figure 19**) and on the other hand for focusing on the average features occurring in this cluster and not having extravagant side chains, akin to CHEMBL569634 or CHEMBL571478. CHEMBL 3421903 was withdrawn from this cluster, because its structure with a large PEG-link was grouped incorrectly (**Figure 19**). CHEMBL408701 TAURO-LITHOCHOLIC ACID and CHEMBL260737 have been clustered into this cluster although they have a BA or BA like scaffold, probably because of their side chain size.



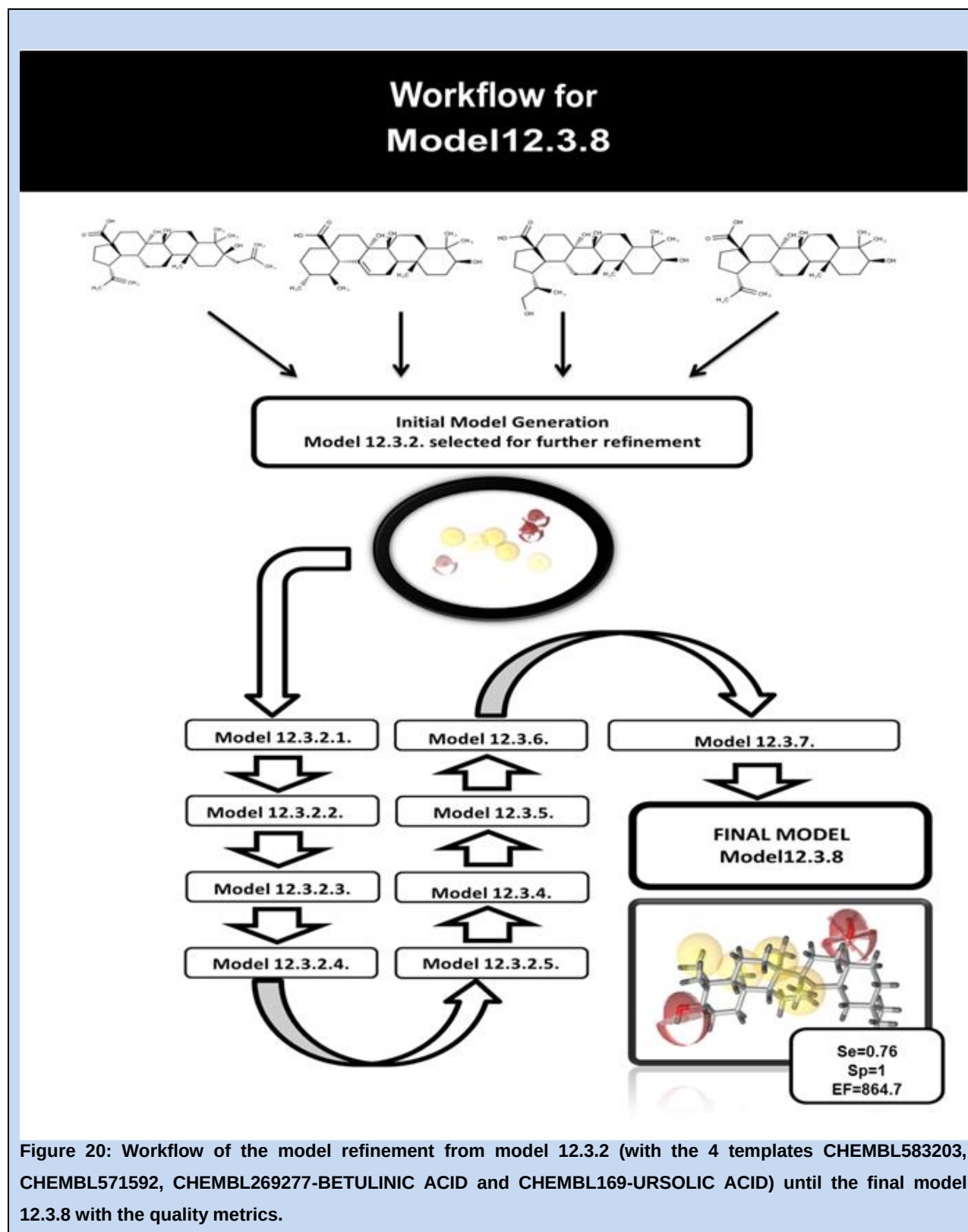
The initial models were generated with LigandScout 4.08 as merged feature pharmacophores with the settings represented in **Table 15**. After validating the 3 generated models, the **Model 12.3.2** in the function of a starting point for the model optimization and refinement was chosen. It was ranked second.

Table 15: Settings of the initial model generation for Cluster 12

Cluster 12.3.2 settings	
Pharmacophore Type	merged feature pharmacophore
Scoring Function	pharmacophore fit and atom overlap
Conformer Generator	iCon
Max. Number of conformations	200
Max Number of result Pharmacophores	3

Model 12.3.2 consists of 3 mandatory and 2 optional lipophilic features, 2 mandatory and 1 optional hydrogen-bond accepting feature and one mandatory negatively-ionizable group. The model was able to find 15 out of the 16 molecules belonging to cluster 12. Only betunolic acid was not found. The disadvantages of this initial model were firstly the relatively high decoy hit rate of 111, secondly the disability to not discard all True Inactives, since it showed a known bile acid like true inactive as active, and thirdly the relatively high number of features, which could result in overfitting of the model. The upsides are the initial very high EF of 121. The aim of this work was not to find a very selective model but more to generate a highly restrictive filter for VS. Therefore, high specificity and EFs were sought, while selectivity was only ranked in the second row. The main workflow with all intermediate models is described in **Figure 20**.

The first improvement was to set the hydrogen bond acceptor feature at (-0.04/-7.42/8.23) with tolerance 1.5 and weight 0.83 to mandatory [model 12.3.2.1] but further increased the tolerance to 1.8 [model 12.3.2.2]. The hydrophobic feature at (2.14/-0.58/-0.14) was deleted [model 12.3.2.3]. Four exclusion volumes at (0.96/4.16/-2.41 with tolerance 1.6 and weight 1), (-5.46/-3.26/9.52 with tolerance 1 and weight 1), (-5.83/-5.28/-1.19 with tolerance 1.3 and weight 1), (-8.3/-2.45/2.18 with tolerance 1 and weight 1), (-4.98/-8.18/5.24 with tolerance 1.15 and weight 1) were added. The lipophilic group weight and tolerances were optimized until they reached the best performance. The weight of lipophilic feature (-3.92/-6.27/6.73) was set from 0.92 to 1 and lipophilic feature (-2.6/-5.74/3.9) increased from 0.83 to 0.93. The tolerance of the lipophilic feature at (-0.05/-0.47/4.42) was increased from 1.5 to 1.65. The hydrogen-bond accepting feature at (-0.04/-7.42/8.23) was set from weight 0.83 to 0.73. [Model 12.3.2.4- 12.3.6]. These two hydrogen-bond accepting features flanking the negatively ionizable feature (-2.43/-0.3/-0.05/T1.5/W1) and (-4.09/1.38/-0.06/T1.5/W1) were interpolated to a single larger hydrogen-bond accepting feature (-3.26/1.38/-0.06 with T1.8 and W 1) [model 12.3.7]. As a last step, a last exclusion volume coat was added at (-2.69/8.93/9.07 with tolerance 1 and weight 1). This model was the most restrictive of all with an **EF=864.69** and **Selectivity>0.99**. It was still able to find 13 out of the 16 highly active molecules in cluster 12. Only the bile acid like scaffolds CHEMBL408701-TAURO-LITHOCHOLIC ACID, CHEMBL260737 and the triterpene betunolic acid were discarded. The final model is depicted in **Figure 21**. **Table 16** and **Figure 22** discuss the hit-rates, quality-metrics and ROCS-curves of Model 12.3.8.



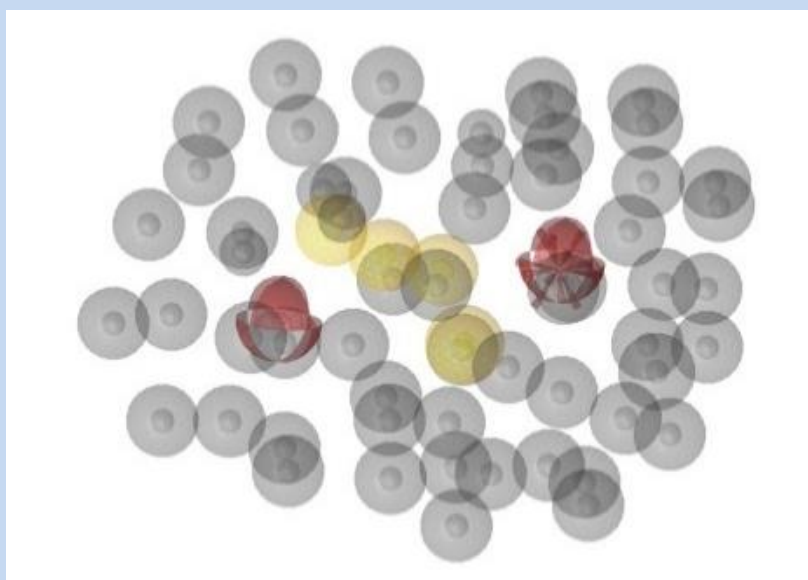


Figure 21: Model 12.3.8 with all exclusion volumes (grey), 4 lipophilic features (green), 2 HBD features (red) and one positively ionizable group (red star).

Table16: Hit-rates and quality metrics of Model 12.3.8., when HIGH ACTIVES database is screened against the two sets DUDE DECOYS and TRUE INACTIVES.

Enrichment Merics	
High Actives hits	13/17
True Inactives hits	0/49
DUDe Decoy hits	3/18043
Enrichment factor	864.69
Selectivity	0.76
Specificity	1

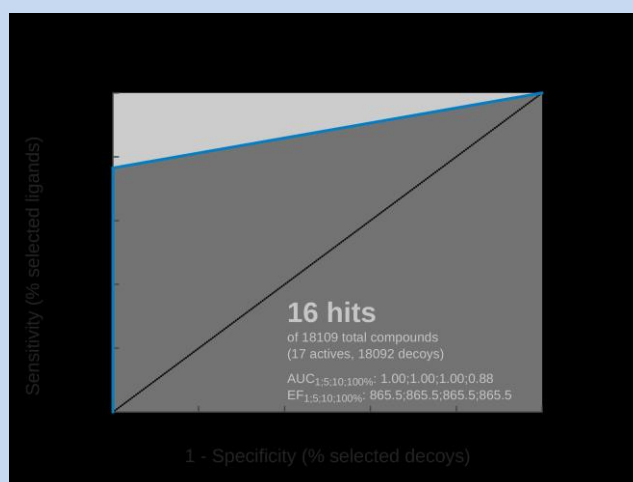


Figure 22: The ROC-curve of model 12.3.8 received after screening Cluster 12 against the TRUE INACTIVE and DUDE DECOYS. 13 out of 16 actives were found and only three decoys out of 18092 passed the filter.

5.7. Results of the prospective virtual screening

All databases (7.6.3. **Prospective Virtual Screening**) were virtually screened against the **BAMS22** model (Cluster 7) and **Model 12.3.8.** (Cluster 12). The output was saved as an .xls file with their smiles-codes, pharmacophore fit score, calculated physicochemical properties and appended information from the sources. Following hit-rates have been achieved:

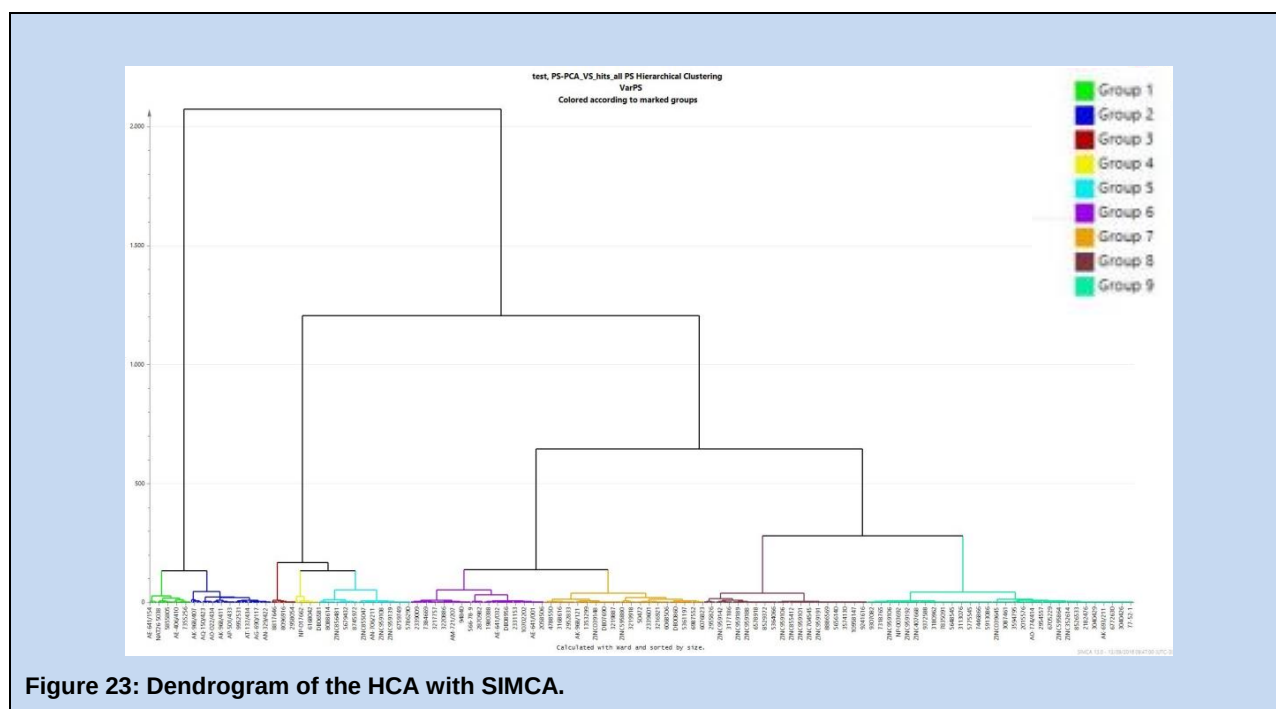
Table 17: Number of virtual hits; number of database structures (hit-rate) in the prospective VS

Database	Model Cluster 7	Model Cluster 12
Analyticon-MEGx/NATx	20 / 4355 (0.46%)	20 / 4355 (0.46%)
Analyticon-Triterpenes	0 / 409 (0.00%)	17 / 409 (4.16%)
Drugbank	23 / 6863 (0.34%)	1 / 6863 (0.01%)
NPDB	299 / 115275 (0.26%)	415 / 115275 (0.36%)
NuBBE	1 / 1628 (0.06%)	3 / 1628 (0.18%)
SPECS NP	8 / 871 (0.92%)	2 / 871 (0.23%)
SPECS SC	82 / 212446 (0.04%)	0 / 212446 (0.00%)
TCM-Taiwan	113 / 35993 (0.31%)	133 / 35993 (0.37%)
In-house database	10/1152 (0.87%)	6/1152(0.52%)

Possible duplicates within databases were removed by means of their same smiles-strings, which is one of the few possibilities to do so¹³⁷. When a duplicate was removed the one with the lower pharmacophore-fit score was removed, resulting in a hitlist of 1069 (83 duplicates were removed). The complete hitlist is added in the appendix.

5.8. Results of hierarchical cluster analysis (HCA)

To achieve a sufficient structural diversity in the final selection of hits a hierarchical cluster analysis on a level of 9 groups was implemented (**6.6.4. Hit List Validation**). For that reason, a PCA of the comprised hitlist with ChemGPS and a subsequent hierarchical cluster analysis in SIMCA was performed. The result is displayed in **Figure 23** in form of a dendrogram.



Group 1 and 2 distinguish on a very early hierarchical level from the other groups. The main structural difference of these two groups compared to the other ones is that they comprise synthetic compounds and NPs with aromatic rings and conjugated double-bonds, while group 3-9 consist of steroidal and/or triterpene structures. The main difference on the level of principal components is ChemGPS' PC2, which describes aromaticity and conjugated double bonds. As can be seen in **Figure 24**, group 2 has the highest PC2 scores, followed by group 1 and group 3. Group 3 has an interesting position with both high PC2 and PC1 (describes properties like MW, atom count) scores, since it is made of triterpenes with aromatic residues, such as cinnamic or caffeic acid. Group 4 inhabits a space with low PC2 and PC3 (describes properties like lipophilicity) but very high PC1 scores, because it is comprised by saponins with hydrophilic sugar moieties and a high-MW triterpene agenine part. The other groups (group 5-9) spread over a relatively similar chemical space, as they consist of steroidal structures, reaching from cardenolides, pregnanes, bile acids to steroids and triterpenes. The compounds of those groups mainly differ in their MW and their lipophilicity/hydrophilicity. Group 6 and 7 have low molecular

weight, have some double bonds and populate an area from the lipophilic to hydrophilic space (cardenolides, pregnanes, bile-acids and steroids). Group 5 and 9 have high PC1 scores, thus group 5 reaches into the more aromatic area (PC2). Group 5 in particular is made of acetylated triterpenes or other triterpene-esters or ethers, group 9 of triterpenes with an average number of alcohol and acid groups. Finally group 8 can be seen as quite the same as group 9 with a little shift into the low PC3 score area, because group 8's triterpenes and conjugated steroids incorporate more hydrophilic groups than those of group 9. **Table 18** addresses the differences of the HCA-groups and gives representative examples. The full hitlist can be seen in the appendix.

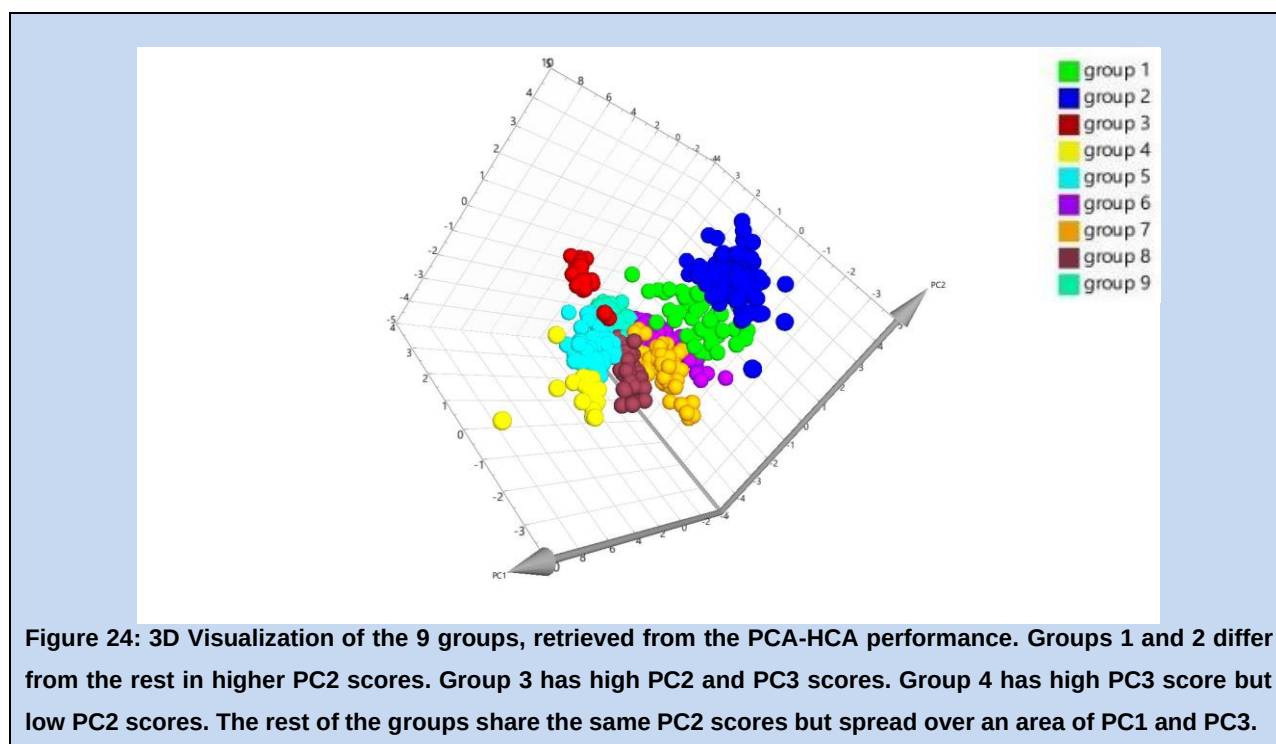
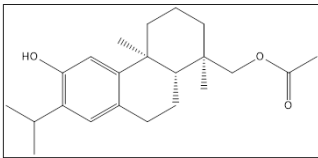
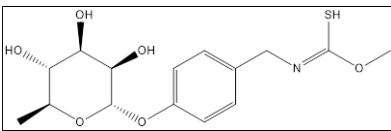
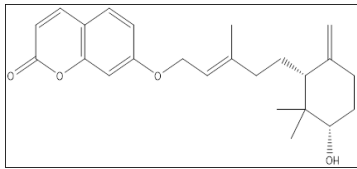
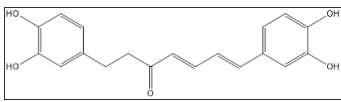
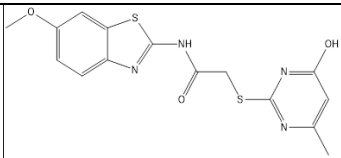
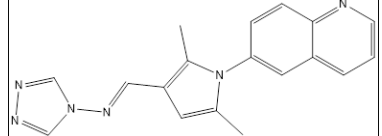
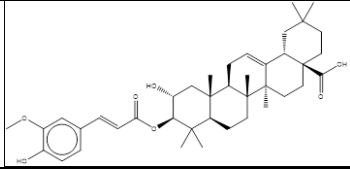
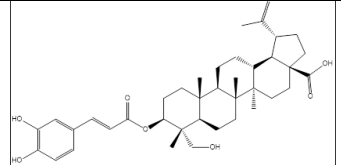
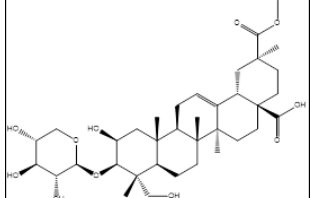
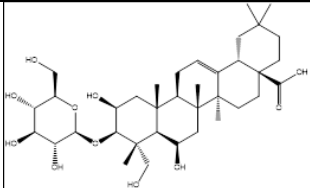
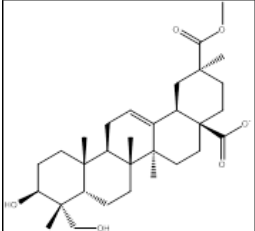
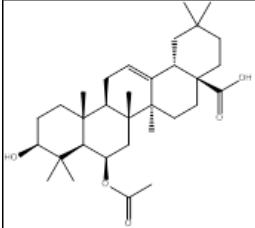
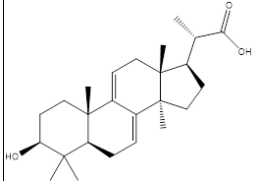
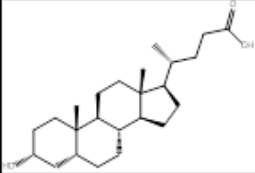
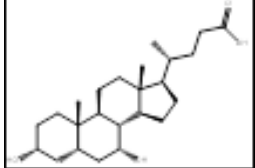
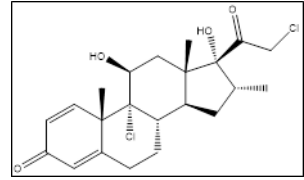
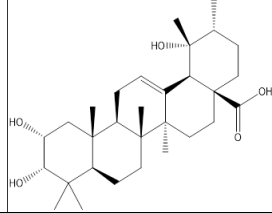
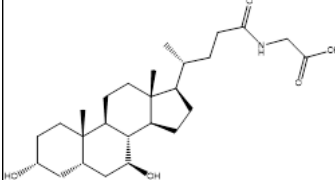
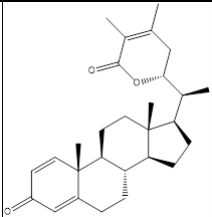
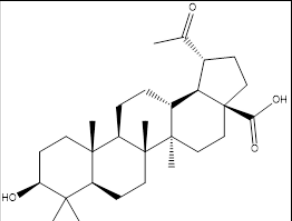


Table 18: The 9 derived HCA groups (number of enclosed compounds), a short description of the members and examples of them with their trivial names (in the case of NPs the organisms which they are derived from), their CAS registry-numbers, the databases they were retrieved from. A note if they had been tested is added.

HCA-group	description	exemplary members	
Group 1 (n=48)	Aromatic small molecules, mostly from NP origin		18- Acetoxyferruginol, aromatic diterpene from <i>Cryptomeria japonica</i> NPDB CAS# 145940-78-9 not available
			Thiocarbamat-glycoside from <i>Moringa olifera</i> NPDB CAS# 147821-57-6 not available
			Farnesiferol B, sesquiterpen-coumarin from <i>Ferula assa-foetida</i> In-house DB CAS# 54990-68-0 tested (active)
Group 2 (n=90)	Very aromatic, small molecules with high heteroatome count, mostly from synthetic origin		Curcuminoid from <i>Zingiber officinale</i> TCM-DB CAS#255065-27-1 not available
			Synthetic compound CAS#353253-76-6 SPECS-DB tested (active)
			Synthetic compound CAS# 432530-00-2 SPECS-DB tested (active)
Group 3 (n=27)	Triterpenes with aromatic residues		Eucalyptolic acid from <i>Eucalyptus camaldulensis</i> CAS# 189272-68-2 NPDB too expensive
			Triterpen-caffeate from <i>Betula pubescens</i> CAS# 159395-53-6 NPDB not available
Group 4 (n=24)	Saponins		Phytolaccoside B from <i>Phytolacca americana</i> CAS# 60820-94-2 NPDB tested (not active at 20µM)

				3-O-β-D-Glucopyranosyl protobassic acid from <i>Vitellaria paradoxa</i> CAS# 53840-47-4 Analyticon-Triterpenes not available
Group 5 (n=102)	Triterpenes with Ether and Ester groups (i.e. acetylated)			Phytolaccagenic acid from <i>Phytolacca americana</i> CAS# 54928-05-1 TCM-DB tested (active)
				Karachic acid from <i>Betula utilis</i> CAS# 56119-15-4 NPDB not available
Group 6 (n=144)	Lipophilic cardenolids, pregnanes, cholans and steroids			Pregnane from <i>Verticillium lecanii</i> CAS# 90522-85-3 NPDB not available
				Lithocholic acid TCM-DB benchmark molecule
Group 7 (n=167)	Corticosteroids, hydrophilic pregnanes, cholans, cardenolids and steroids			Chenodeoxycholic acid NPDB already tested
				Mometasone, inhalative corticosteroid against asthma bronchiale CAS# 105102-22-5 Drugbank too expensive
Group 8 (n=177)	Hydrophile triterpenes, conjugated bile acids and steroids			Euscaphic Acid from <i>Eriobotrya japonica</i> CAS# 53155-25-2 Analyticon-Megx tested (active)
				Glyco-chenodeoxycholic acid NPDB already tested

Group 9 (n=291)	Lipophilic triterpenes, conjugated bile acids and steroids		Minabeolide 1 from soft coral <i>Minabea</i> sp. CAS# 114820-24-5 NPDB not available
			Platanic acid NPDB benchmark molecule

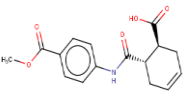
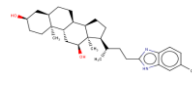
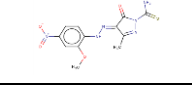
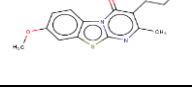
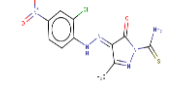
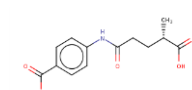
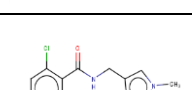
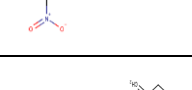
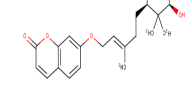
5.9. Hit Selection

As obvious from **Table 18**, most of the NPs were not available at common suppliers or were very expensive. Moreover, the most interesting molecules, in terms of structural diversity, are members of the HCA-group 2 and 1. SPECS is an affordable supplier of screening compounds and this is why these groups were extensively validated (also in terms of shape-comparison ranking). Still, NPs from these groups were difficult to get and were underrepresented in bioactivity-testing. In groups 6, 7, 8 and 9 some of the hits had also been benchmark molecules (i.e. LCA, platanic acid) or had been already tested. The following section gives a short overview on the selected hits.

5.9.1. Hit Selection for HCA group 1

In the HCA group 1, two molecules had been available in the in-house substance store (microlobiden and farnesiferol). Both compounds were derived from *Ferula assa-foetida*, a central Asian medicinal plant. Hence, two extracts of *Ferula assa-foetida* were also put up for the bioactivity testing (**6.7. Extract Preparation**). The other compounds were obtained from SPECS. Due to the affordable pricing of their screening compounds, it was possible to order compounds with high Tanimoto-Combo (TC) scores with CHEMBL270105 as a query. For this reason, all validated hits are retrieved from BAMS22, as CHEMBL270105 was one of BAMS22's benchmark molecules. **Table 19** gives an overview over the selected molecules with their trivial names (alternatively SPECS-ID), their CAS-numbers, their TC scores and pharmacophore fit (PF) scores. Further it is listed from which database they originate and from which model they were selected as hits.

Table 19: Selected hits from HCA group 1 with chemical structure, trivial name, CAS registry number, pharmacophore fit score (PF), tanimoto combo (TC) score calculated in OpenEye's ROCs with CHEMBL270105 as a query, the model from which they were hit and their origin database.

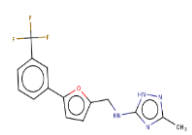
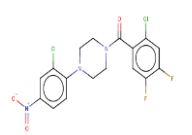
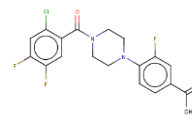
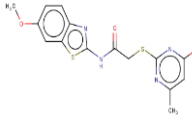
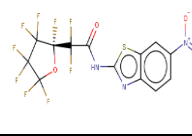
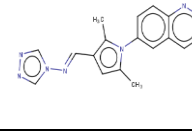
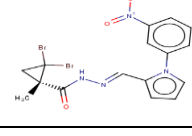
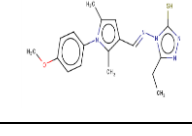
Group 1						
Structure	Trivial Name (SPECS-ID)	CAS	PF-score	TC-score	Model	Database
	(13) AK-918/37074025	352644-32-7	45.31	0.873	BAMS22	SPECS-SC
	(14) AE-641/15486021	500218-51-9	56.55	0.913	BAMS22	SPECS-SC
	(15) AN-329/40226744	314757-83-0	46.23	0.703	BAMS22	SPECS-SC
	(16) AE-406/41056941	380633-89-6	56.31	0.720	BAMS22	SPECS-SC
	(17) AN-329/40226746	26179-09-9	46.23	0.703	BAMS22	SPECS-SC
	(18) AL-281/42289098	664993-86-6	45.24	0.627	BAMS22	SPECS-SC
	(19) AK-968/41026577	525577-20-2	45.08	0.701	BAMS22	SPECS-SC
	(22) Farnesiferol B	54990-68-0	46.39	0.737	BAMS22	In-house
	(21) Microlobiden	89783-66-4	56.72	0.726	BAMS22	In-house

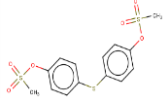
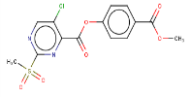
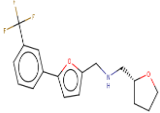
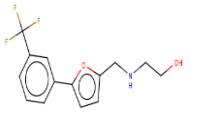
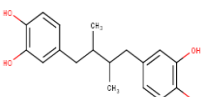
5.9.2. Hit Selection for HCA group 2

In HCA group 2, one molecule was available in the in house substance store (nordihydroguaretic acid). The other compounds were obtained from SPECS. The 12 best scoring compounds in terms of the TC score with CHEMBL270105 as a query were ordered. All of the compounds were retrieved from **BAMS22** but that is a result of the ranking according to their shape-matching with CHEMBL270105, a benchmark molecule of the model **BAMS22**.

Table 20 presents an overview over the selected molecules with their trivial names (alternatively SPECS-ID), their CAS-numbers, their TC scores and pharmacophore fit (PF) scores. Further it is listed from which database they originate and from which model they were selected as hits.

Table 20: Selected hits from HCA group 2 with chemical structure, trivial name, CAS registry number, pharmacophore fit score (PF), tanimoto combo (TC) score calculated in OpenEye's ROCs with CHEMBL270105 as a query, model with which they were fetched and their origin database.

Group 2						
Structure	Trivial Name (SPECS-ID)	CAS	PF-score	TC-score	Model	Database
	(1) AN-465/43411399	1019061-83-6	46.23	0.783	BAMS22	SPECS-SC
	(2) AG-205/37082045	303139-94-8	47.83	0.860	BAMS22	SPECS-SC
	(3) AG-205/12140277	330636-58-3	46.32	0.806	BAMS22	SPECS-SC
	(4) AI-020/37280139	353253-76-6	57.09	0.877	BAMS22	SPECS-SC
	(5) AK-968/11180151	353779-79-0	56.50	0.801	BAMS22	SPECS-SC
	(6) AO-081/15570035	432530-00-2	65.59	0.773	BAMS22	SPECS-SC
	(7) AK-968/15605196	444931-63-9	46.2	0.777	BAMS22	SPECS-SC
	(8) AO-081/15245166	496937-29-2	46.15	0.821	BAMS22	SPECS-SC

	(9) AO-548/43179669	791840-52-3	46.28	0.838	BAMS22	SPECS-SC
	(10) AO-022/43454373	902244-06-8	45,25	0.810	BAMS22	SPECS-SC
	(11) AN-465/43384126	915930-57-3	46.14	0.774	BAMS22	SPECS-SC
	(12) AN-465/43411358	932954-51-3	46.16	0.783	BAMS22	SPECS-SC
	(20) Nordihydroguaretic acid	89783-66-5	65.87	0.664	BAMS22	In-house

5.9.3. Hit Selection for other HCA groups

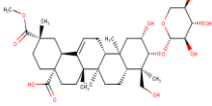
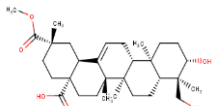
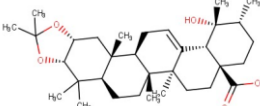
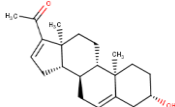
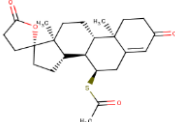
Due to unaffordable prices and missing access to the rare NPs of HCA group 3, we had to renounce on the validation of this cluster, although it included several very interesting compounds for structure-activity-relationship (SAR) of triterpenes. An investigation of those compounds, as far as somebody feasible is highly recommended.

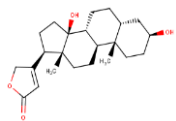
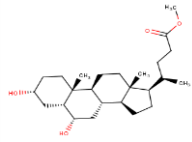
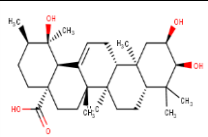
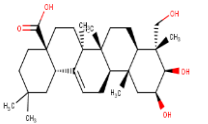
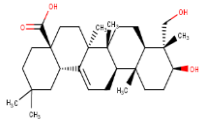
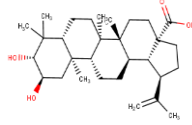
In HCA-group 4, including triterpene-saponins, at least the compound, phytolaccoside B, was evaluated. The genine of phytolaccoside B phytolaccagenic acid was also a member of the HCA-group 5 and represented therefore an interesting hit from this group. Because both hits were originated from the roots of *Phytolacca americana*¹⁷⁵, an extract of the plant-organs was prepared (**7.7. Extract Preparation**) and tested. The second hit 2,3-O-isopropylidenyl euscaphic acid, a triterpene just recently isolated from the seeds of *Rubus fruticosus*¹⁷⁶ was also selected from group 5. In group 6, two hits have been deployed hits for bioactivity testing: Firstly, 16-dehydropregnenolone, and secondly spironolactone, an aldosterone-receptor antagonist and the only approved drug tested in this work.

In group 7 two hits were nominated for testing, namely digitoxigenin, a cardenolide, and methylhyoxycholol, a bile acid which is known to modulate LXR. In group 8 and 9 five triterpene hits were selected: hederagenin and bayogenin, both largely distributed as plant secondary

metabolites, can be seen as derivate of oleanolic acid, which is a known GPBAR1 agonist, and were not tested on their GPBAR1 activity yet. Alphitolic acid was just recently isolated in the department and fitted the triterpene-model 12.3.8. It is a lupan-like triterpene. Finally euscaphic acid, derived from traditional Chinese herbal medicine plant *Eriobotrya japonica*, was reported to show hypoglycemic effects in an animal model¹⁷⁷. **Table 21** gives an overview over the selected molecules with their trivial names, their CAS registry-numbers, their TC scores and pharmacophore fit (PF) scores. Further it is listed from which database they are originated and from which model they were selected as hits.

Table 21: Selected hits from HCA group 4-9 with chemical structure, trivial name, CAS registry number, pharmacophore fit score (PF), *tanimoto combo (TC) score (calculated in OpenEye's ROCs with *Betulinic acid or **CHEMBL270105 as a query), model with which they were fetched and their screened database.

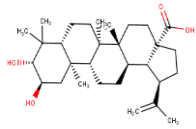
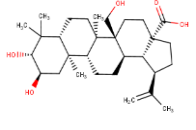
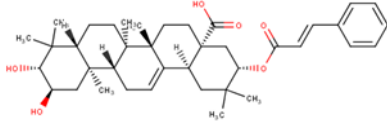
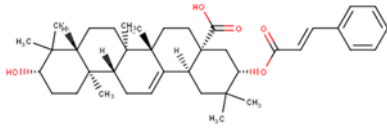
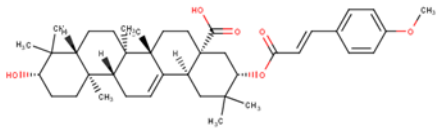
Group 4						
Structure	Trivial Name	CAS	PF-score	TC-score	Model	Database
	(33) Phytolaccoside B	60820-94-2	73.7 1	0.989*	Model12.3.8	NPDB
Group 5						
	(32) Phytolaccagenic Acid	54928-05-1	73.1 6	1.501*	Model 12.3.8.	TCM-DB
	(31) 2,3-O-Isopropyliden-euscaphic acid	220880-90-0	73.4 1	1.473*	Model 12.3.8.	NPDB
Group 6						
	(23) 16-Dehydropregnenolone	1162-53-4	55.9 3	1.222*	BAMS22	NPDB
	(24) Spironolactone	52-01-7	46.7 1	1.088*	BAMS22	drugbank

Group 7						
	(25) Digitoxigenin	143-62-4	56.19	1.014*	BAMS22	NPDB
	(26) Methylhyoxycholol	2868-48-6	68.54	1.611*	BAMS22	NPDB
Group 8						
	(27) Euscaphic Acid	53155-25-2	72.75	1.612*	Model12.3.8.	In-house
	(28) Bayogenin	6989-24-8	73.53	1.276*	Model12.3.8.	In-house
Group 9						
	(29) Hederagenin	53155-25-2	72.75	1.612*	Model12.3.8.	In-house
	(30) Alphitolic acid	19533-92-7	73.88	1.829*	Model12.3.8.	In-house

5.9.4. Additional Compounds for Bioactivity Testing

Alphitolic acid (see **Table 21**/ group 9) from our in house database was a predicted hit by Model 12.3.8. Since the compound was just recently isolated by Christine Elisabeth Mair together with four other triterpenes from a South African medicinal plant¹⁷⁸, we decided to test all the different sapogenins of the isolated triterpene saponins (**Table 22**). Whereas alphitolic acid (A10h) complied with all features of Model 12.3.8, A3h, A4h and A6h only fitted the model, when the exclusion volumes were ignored, and A8h not at all. Considering that the four other compounds were closely related to the known potent ligand oleanolic acid⁶⁰ testing was nevertheless attractive, since the testing of the three compounds, which incorporate a cinnamoyloxy-residue at position 21, extend the understanding of SAR of triterpenoid acids.

Table 22: Selected sapogenins derived from the triterpene saponins isolated from a South African medicinal plant¹⁷⁸

Plant A's triterpenes		
Structure	Trivial Name	Abbreviation
	(30) Alphitolic acid	A10h
	(34) 27-Hydroxyalphitolic acid	A8h
	(35) 21β- Cinnamoyloxymaslinic acid	A3h
	(36) 21β –Cinnamoyloxyoleanolic acid	A4h
	(37) 21β-p-Methoxycinnamoyloxyoleanolic acid 21β-cinnamoyloxyoleanolic acid	A6h

5.10. Results of the experimental validation

The experimental procedure can be seen in **7.8.Reporter-gene bioassay**. Compounds were seen as active, when they achieved 10% of the full receptor activation using a reporter gene assay. Therefore LCA at 10μM was always tested in parallel and acted as a positive control. The activation of LCA at 10μM was set to 100% activation. Vehicle control with a final dimethylsulfoxide (DMSO) concentration of 0.1% was set to 0% activation. Several compounds had shown no activity, although activity at higher concentrations can not be excluded (i.e. 30 μM). Therefore compounds, which showed at least an activation of more than 12% in one single experiment and a mean activity (retrieved of 3 experiments) of more than 8% could be seen as active, as this was the case with **(16)**, NDG, (6) and PB. **(14)**, spironolactone, digitoxigenin, bayogenin, alphitolic Acid, 27-hydroxyalphitolic acid, A3h, A4h and A6h showed at both 5 and 20 μM [digitoxigenin and **(14)**] or only at 20 μM a cytotoxic property. The results of the experimental validation are summarized in **Table 23** with % activation and Standard Error Mean (SEM).

Table 23: All tested compounds with their activity expressed as fold activation and % of activation (activation of LCA at 10µM is regarded as maximum activation). The green color marks active concentrations; blue color marks concentration of compounds very close to the threshold (10% activation); red color marks inactive concentrations. Data is given as mean $\pm$ SEM of single determinations obtained from at least three independent experiments. *These concentrations were only validated in two independent experiments

Compound	Fold Activation (5µM)	Fold Activation (20µM)	% Activity (5µM)	% Activity (20µM)
DMSO	1		0	
LCA (10µM)	19.35 ± 1.39		100%	
CDCA (50µM)	14.22 ± 0.59			
(1)	1.23 ± 0.12	1.28 ± 0.10	1.54% ± 0.91	1.65% ± 0.69
(2)	0.88 ± 0.07	1.19 ± 0.13	-0.60% ± 0.36	0.92% ± 0.67
(3)	0.95 ± 0.07	1.19 ± 0.33	-0.31% ± 0.38	0.88% ± 1.69
(4)	1.53 ± 0.06	3.89 ± 0.60	3.00% ± 0.58	19.45% ± 6.00
(5)	1.10 ± 0.07	1.18 ± 0.02	0.98% ± 1.13	1.00% ± 0.18
(6)	0.99 ± 0.03	2.09 ± 0.51	-0.02% ± 0.14	9.27% ± 2.62
(7)	0.98 ± 0.09	1.49 ± 0.02	-0.06% ± 0.44	3.10% ± 0.46
(8)	1.19 ± 0.02	1.06 ± 0.03	1.07% ± 0.10	0.34% ± 0.18
(9)	1.00 ± 0.11	1.58 ± 0.29	0.33% ± 0.63	3.87% ± 1.89
(10)	1.20 ± 0.42	1.63* 0.00	1.67% ± 2.54	3.18%* ± 0.01
(11)	1.56* ± 0.10	1.57 ± 0.36	2.63%* ± 0.69	2.45% ± 1.42
(12)	1.34 ± 0.37	1.37 ± 0.19	-0.13 ± 0.31	1.72% ± 0.90
(13)	1.55 ± 0.34	1.36 ± 0.27	2.34 ± 1.40	1.74% ± 1.36
(14)	Not evaluable neither at 5 µM nor 20 µM, due to cytotoxic properties			
(15)	1.32 ± 0.33	0.97 ± 0.08	1.42 ± 1.39	-0.06% ± 0.39
(16)	1.40 ± 0.24	2.43 ± 0.41	1.69% ± 1.01	11.00% ± 4.15
(17)	1.93 ± 0.30	2.27 ± 0.43	4.75% ± 1.51	6.12% ± 2.01
(18)	1.27 ± 0.09	1.17 ± 0.13	1.16% ± 0.35	0.70% ± 0.54
(19)	1.11 ± 0.14	1.41 ± 0.21	0.47% ± 1.82	0.58% ± 0.89
(20) Nordihydroguaretic acid	1.92 ± 0.72	2.00 ± 0.32	3.73% ± 2.99	4.35% ± 1.28

(21) Microlobiden	1.42 ±0.13	15.55 ±3.56	4.66% ±2.03	85.23% ±20.87
(22) Farnesiferol B	2.45 ±0.87	6.15 ±1.00	9.35% ±4.33	54.02% ±28.18
(23) 16-Dehydropregnenolone	2.25 ±0.62	6.19 ±0.52	5.20% ±2.42	22.58% ±0.74
(24) Digitoxigenin	Not evaluable neither at 5 µM nor 20 µM, due to cytotoxic properties			
(25) Spironolactone	2.12 ±0.23	Not evaluable at 20 µM, due to cytotoxic properties	12.80% ±6.60	Not evaluable at 20 µM, due to cytotoxic properties
(26) Methylhyoxcholate	1.29 ±0.19	6.47 ±1.54	1.13% ±0.77	23.22% ±5.29
(27) Phytolaccagenic acid	1.76 ±0.45	7.00 ±0.60	3.23% ±0.86	26.28% ±2.18
(36) Phytolaccoside B	1.22 ±0.18	1.62 ±0.40	3.76% ±1.67	8.78% ±3.87
(35) 2,3-O-Isopropylidenyl-euscaphic acid	1.61 ±0.78	4.35 ±2.78	3.45% ±4.66	22.19% ±15
(30) Euscaphic acid	1.23 ±0.12	1.88 ±0.26	1.99% ±1.07	12.08% ±6.51
(31) Hederagenin	3.20 ±1.26	6.58 ±1.25	12.69% ±5.33	35.95% ±2.37
(32) Bayogenin	0.84 ±0.15	Not evaluable at 20 µM, due to cytotoxic properties	-1.39% ±1.17	Not evaluable at 20 µM, due to cytotoxic properties
(33) Alphitolic acid (A10h)	12.53 ±5.680	Not evaluable at 20 µM, due to cytotoxic properties	29.79% ±10.79	Not evaluable at 20 µM, due to cytotoxic properties
(34) Hydroxyalphitolic acid (A8h)	0.83 ±0.02	Not evaluable at 20 µM, due to cytotoxic properties	-1.19% ±0.30	Not evaluable at 20 µM, due to cytotoxic properties
(35) 21β-Cinnamoyloxymaslinic acid (A3h)	15.19 ±1.12	Not evaluable at 20 µM, due to cytotoxic properties	99.62% ±21.53	Not evaluable at 20 µM, due to cytotoxic properties
(36) 21β-Cinnamoyloxy-Oleanolic acid (A4h)	10.51 ±1.11	Not evaluable at 20 µM, due to cytotoxic properties	64.91% ±10.64	Not evaluable at 20 µM, due to cytotoxic properties
(37) 21β-p-Methoxycinnamoyloxyoleanolic acid 21β-Cinnamoyloxy-oleanolic acid (A6h)	11.52 ±0.54	Not evaluable at 20 µM, due to cytotoxic properties	77.86% ±10.8	Not evaluable at 20 µM, due to cytotoxic properties

In the following chapters the results of every cluster are discussed in detail and experimental results including EC₅₀ values for the most active hits are added.

5.10.1. Results on the experimental validation of group 1

The experimental validation of virtual hits from group 1 showed 37.5% active molecules. Moreover, this cluster unraveled two completely new potent agonists with yet unknown scaffolds: microlobiden and farnesiferol B. Further SC16 concisely met the mean 10% preliminary threshold with 11.00% activation at 20 μ M. Farnesiferol B almost reached the threshold at 5 μ M. Compound **(14)** showed cytotoxic properties at both concentrations and therefore was discarded from this evaluation. **Figure 25** shows the % activation of group 1's compounds.

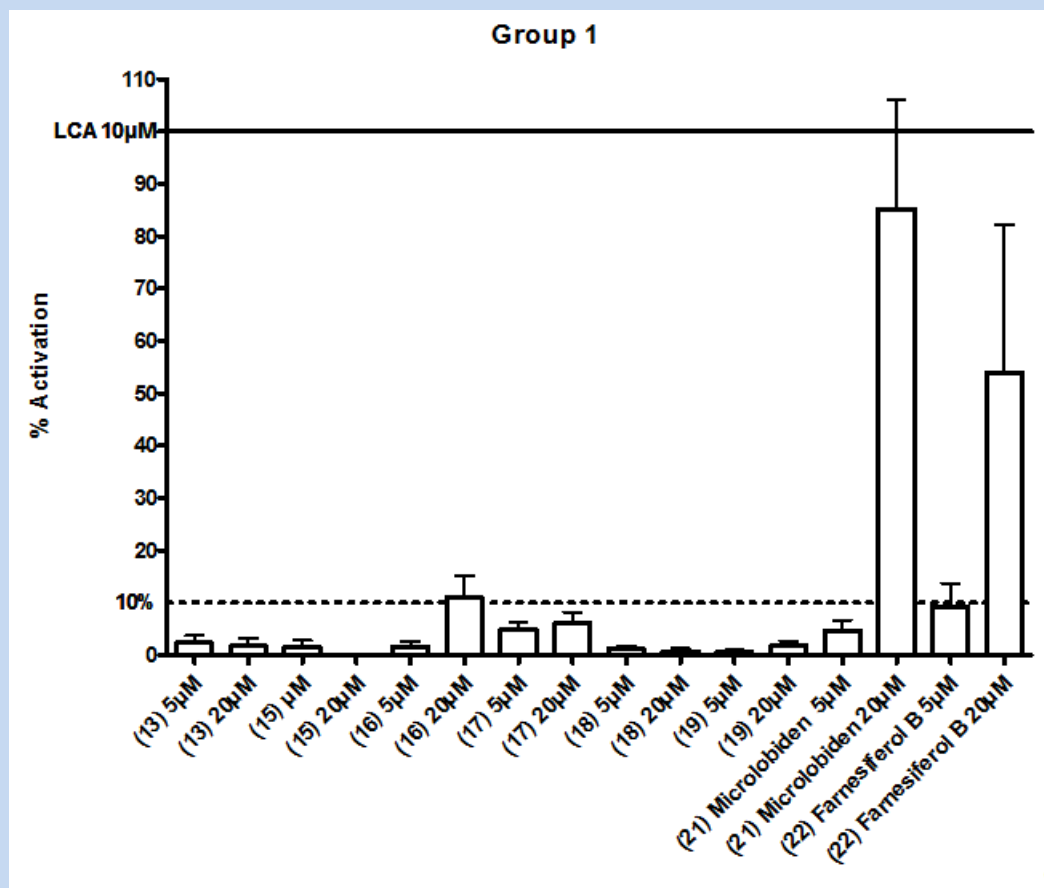


Figure 25: GPBAR1 activation of virtual hits from HCA-group 1 at 5 and 20 μ M using a luciferase reporter gene assay: Compound (14) showed severe cytotoxic properties in this assay and is therefore not depicted. The positive control LCA at 10 μ M was set to 100% activation (solid line). The negative control DMSO 0.1% (vehicle control) was set to 0% activation. The threshold was set to 10% (dotted line). Each value is the mean of at least 3 independent experiments and the variance is given as SEM.

5.10.2. Determination of EC₅₀-values for microlobiden and farnesiferol B

Because of their novelty, bioactivity at 5 and 20 μ M and importance for the validation of this work a concentration response curve was determined. Unfortunately, microlobiden showed at higher concentrations beginning cytotoxicity, farnesiferol B only slightly (see **Figure 27**). Nevertheless it was possible to test Microlobiden until a maximum concentration of 20 μ M (see **Figure 26**), which was also the maximum concentration tested for Farnesiferol B, whereas Farnesiferol B could have been tested at higher concentrations too. LCA showed a significant increase in relative fluorescence units (RFU), which corresponds with transfected cell mass and is in accordance with present literature, reporting bile acids as proliferation modulators¹⁷⁹⁻¹⁸¹.

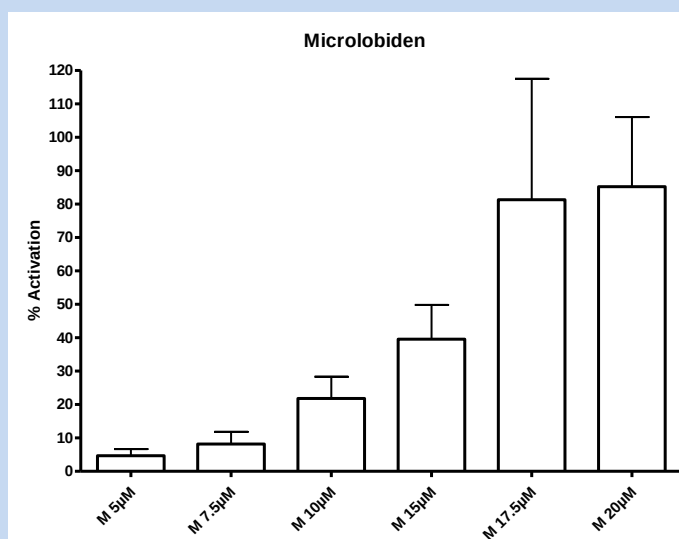


Figure 26: GPBAR1 activation at different concentrations of microlobiden with LCA 10 μ M set as maximum activation of 100%. Each value is the mean of at least 4 independent experiments and the variance is given as SEM.

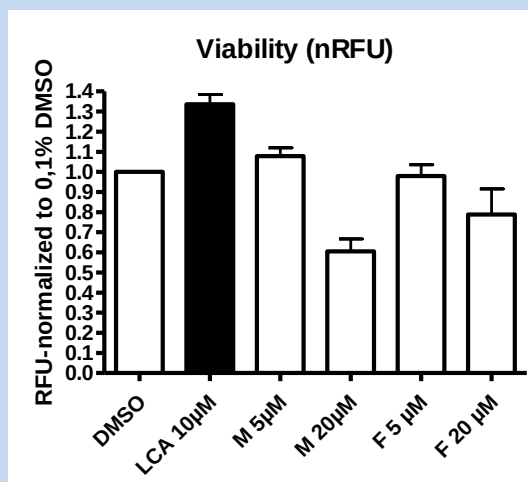


Figure 27: Relative fluorescence units normalized to the solvent DMSO 0.1% (nRFU), as an indicator for cell viability. LCA shows a significant effect on cell proliferation. While microlobiden (M) seems to do so at lower concentrations, it gets cytotoxic at 20 µM. Farnesiferol B (F) seems to share this properties but not as pronounced as microlobiden. Each value is the mean of at least 4 independent experiments and the variance is given as SEM.

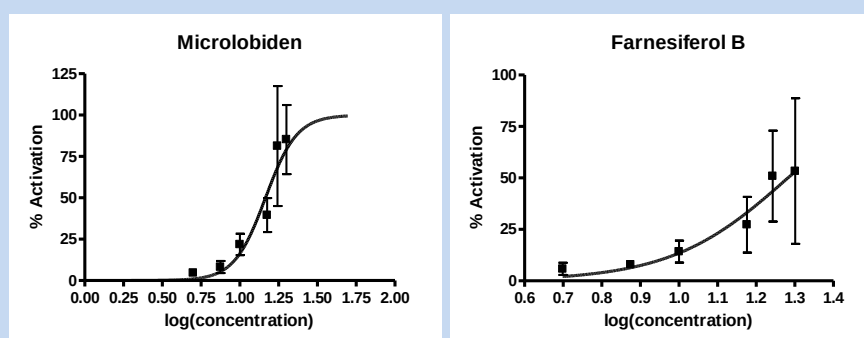


Figure 28: Concentration-response curves for microlobiden (left) and farnesiferol B (right); the curves were calculated with a restraint at 100% activation (activation of LCA 10 µM set as maximum). Each value is the mean of at least four independent experiments. The calculated EC_{50} values showed an $EC_{50} = 14.64$ for microlobiden and an $EC_{50} = 19.16$ for Farnesiferol B respectively. Each value is the mean of at least 4 independent experiments and the variance is given as mean $\pm$ SEM.

Microlobiden achieved a maximum fold activation at 20µM of 15.55 ± 3.56 and an EC_{50} of **14.64 µM $\pm$ 2.285**. Farnesiferol B showed a 6.15 ± 1.00 SEM fold activation at 20 µM and an EC_{50} of **19.16 µM $\pm$ 5.89**, although testing at higher concentrations would have been possible and a maximum fold activation was not evaluated. For comparison, Dr. Ladurner previously determined the EC_{50} values of the two potent bile acids, LCA and CDCA, in the same assay and

with the same statistical analysis. LCA showed an EC_{50} value of 10.5 μM with a 33 fold maximum activation. The moderate endogenous ligand CDCA had an EC_{50} of 33.2 μM with a maximum fold activation of 12 times. Microlobiden's activity can therefore be seen to be in the range of endogenous bile acids. The cytotoxic properties of sesquiterpene coumarins had been investigated previously. Farnesiferol B, among others, was reported to have cytotoxic properties on the human cancer-cell line HeLa¹⁸², while microlobiden was not described.

As lots of NPs are well known pan-assay interference substances (PAINS) or frequent hitters^{163,183}, the two GPBAR1-activating NP were further investigated. In order to prevent such unintentional false-positive results, the experiments have been repeated without transfecting the GPBAR1. EGFP and CRE-Luc plasmids have been transfected as usual with the same concentrations. In these control experiments none of the compounds showed a significant rise in the luminescence values, confirming a direct interaction on GPBAR1 (**Figure 29**).

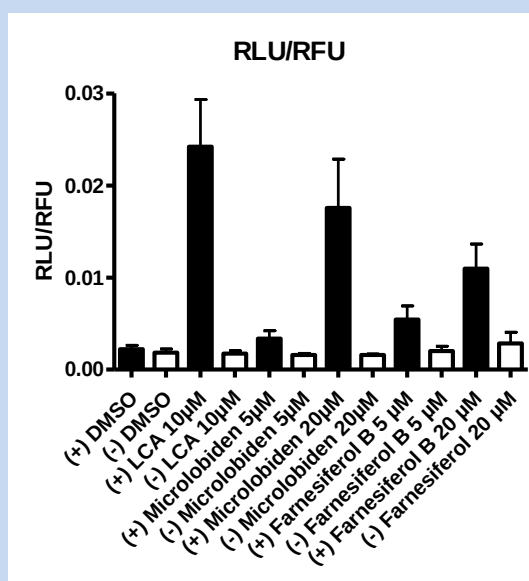
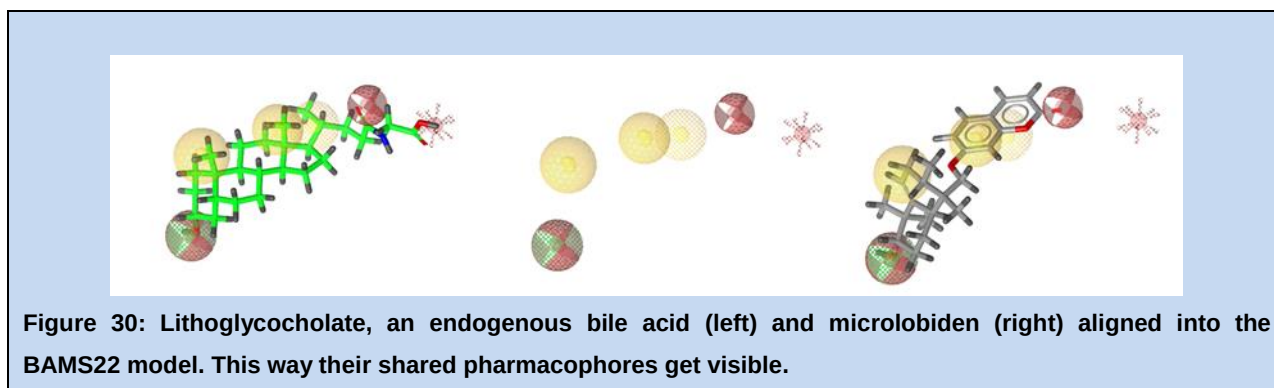


Figure 29: The relative luciferase units (RLU), corresponding to intracellular cAMP rise and therefore GPBAR1 activation, proportioned to the fluorescence units (RFU), related to the transfected cell count and viability. Neither DMSO, LCA, microlobiden nor farnesiferol B showed a significant rise of RLU/RFU, in GPBAR1non-transfected cells (-), as LCA, microlobiden and farnesiferol do in GPBAR1-transfected cells.

The comparison of the BAMS22-alignments of microlobiden and bile acids, suggest a proposed ligand-binding interaction of microlobiden to the GPBAR1. The lipophilic sesquiterpene coumarine scaffold aligns to the pharmacophore very similar to bile acids. The oxygen of the lactone-ester forms the same hydrogen-bond to the GPBAR1 as the ester at position 24 of bile

acids. The alcohol on the sesquiterpene part accepts or donates the same hydrogen bond as the alcohol on position 3 of bile acids. Between these two hydrogen bond anchors microlobiden's rigid lipophilic scaffold fills out the binding pocket very similar to bile acids. **Figure 30** shows the comparison of microlobiden's and lithoglycocholate's pharmacophore and underlines the capability of pharmacophore-based VS.



5.10.3. Results on the experimental validation of group 2

The validation of virtual hits included in group 2 revealed a comparable hit rate than that of group 1, although their chemical space differed more from the pharmacophore's benchmark-molecules. A potential cause could have been the fact those compounds were ordered, which achieved the best PF and TC scores. Interestingly compound **(4)** achieved the best TC score in the whole cluster, in the experimental validation. Compound **(4)** can be seen as the only active molecule meeting the predefined threshold. Compound **(6)** closely misses the bottom limit but it can be expected to meet it at higher concentrations. Kratz et al validated their pharmacophore model collection for hERG-blocker screening for example at 30 μM ¹³⁷. Testing compounds at 30 μM would have increased the experimental metrics further, but still several compounds would have not been able to investigate at those concentrations (i.e. microlobiden, spironolactone) due to cytotoxic issues. The results of the experimental validation of group 2 is illustrated in **Figure 31**.

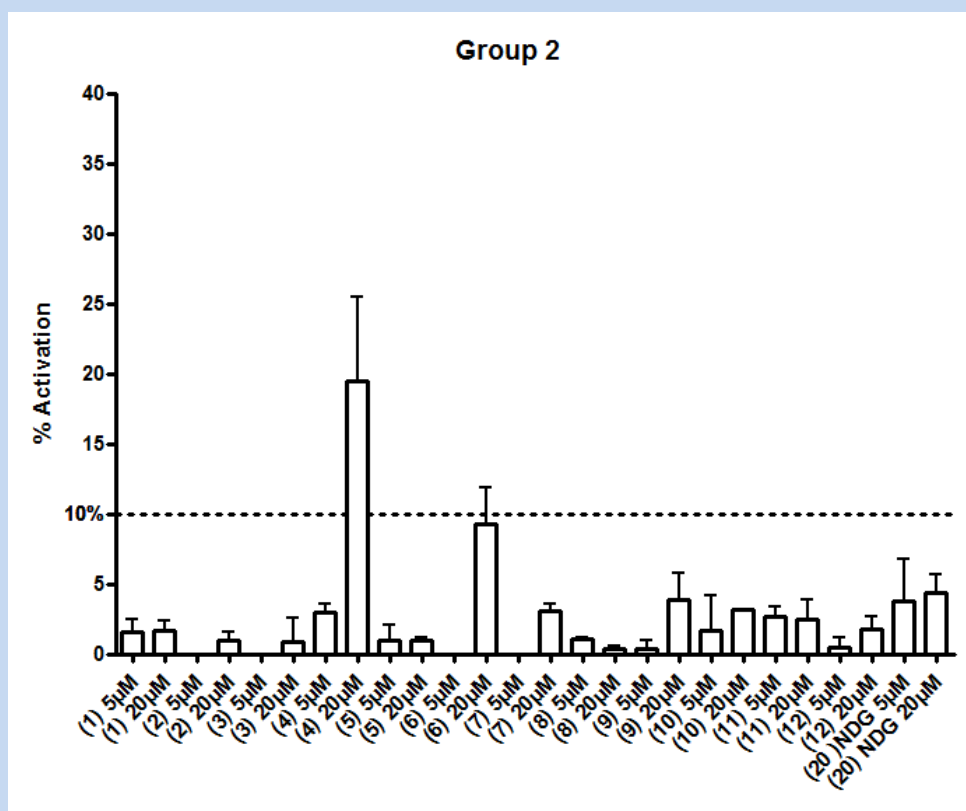


Figure31: GPBAR1 activation of virtual hits from HCA-group 2 at 5 and 20 µM using a luciferase reporter gene assay: Compound (4) showed a clear activation at 20 µM, (6) slightly failed to meet the threshold. The positive control LCA at 10 µM was set to 100% activation. The negative control DMSO 0.1% (vehicle control) was set to 0% activation. The threshold was set to 10% (dotted line).

5.10.4. Results on the experimental validation of groups 4, 5 and 6

The only compound tested belonging to the HCA group 4 **(36)** phytolaccoside B as mean of three independent experiments slightly missed the threshold. However it showed activation up to 18.9% in a single experiment. Testing at slightly higher doses would have identified phytolaccoside B probably as active. Both compounds from group 5 **(34, 35)** were significantly active with $22.19 \pm 15\%$ at 20 µM for **(34)** phytolaccagenic acid, $12.08 \pm 6.51\%$ at 20 µM for **(35)** 2,3-O-isopropylidenyl- euscaphic acid. In group 6, compound **(23)** 16- Dehydrogregnenolone achieved also a considerable activation of $22.58 \pm 0.74\%$ at 20 µM. Spironolactone showed cytotoxic properties, but an evaluation until 15µM was possible. In **Figure 33** you can see the cytotoxic properties of spironolactone. Nevertheless, at 5 µM spironolactone meets the threshold with an activation of $12.80 \pm 6.60\%$. All of the stated values are the mean of at least 3 independent experiments and the variance is given as SEM.

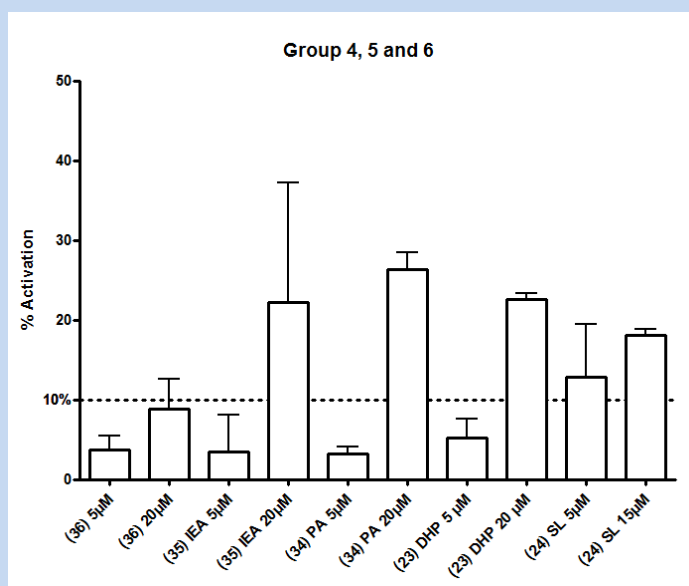


Figure 32: GPBAR1 activation of virtual hits from HCA-group 2 at 5 and 20 µM using a luciferase reporter gene assay: Compound (23), (24), (34) and (35) are clearly active, compound (24) spironolactone even at 5µM. As evaluation was only possible until 15 µM, spironolactone's highest concentration is therefore given at 15µM. Compound (36), the saponine phytolaccoside B only activated up to 8.78% ($\pm$ 3.87) and thus missed the threshold. All stated values are the mean of at least 3 independent experiments and the variance is given as SEM.

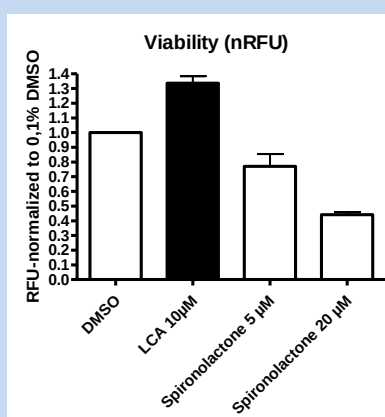


Figure 33: Relative fluorescence units normalized to the solvent DMSO 0.1% (nRFU), as an indicator for cell viability. LCA shows a significant effect on cell proliferation. Spironolactone even at 5 µM shows a reduced cell count (in terms of fluorescence). At 20 µM more than half of the cells were not viable and received values have not been used.

5.10.5. Results on the experimental validation of groups 7, 8 and 9

In group 7 the only available compound methylhyoxycholate **(26)** showed a GPBAR1 agonistic activity. In contrast digitoxigenin **(25)** was very cytotoxic and eradicated nearly all cells, even at low doses of 1 and 2 μM (data not shown). In fact the apoptotic effects through the Na^+/K^+ -ATPase sodium pump inhibition of cardenolides are well known ^{184,185}. In group 8 euscaphic acid **(27)** meets the 10% activation threshold at 20 μM , bayogenin **(28)** was inactive at 5 μM and cytotoxic at 20 μM . Hederagenin **(29)**, a largely distributed triterpene, showed activity and reached the 10% activation limit at 5 μM . Alphitolic acid **(30)** also appeared as a rather potent ligand with 30% receptor-activation at 5 μM . At 20 μM it was like most poly-hydroxylated triterpene acids (i.e. bayogenin) cytotoxic and not evaluable. The experimental results of the stated groups are presented in **Figure 34**.

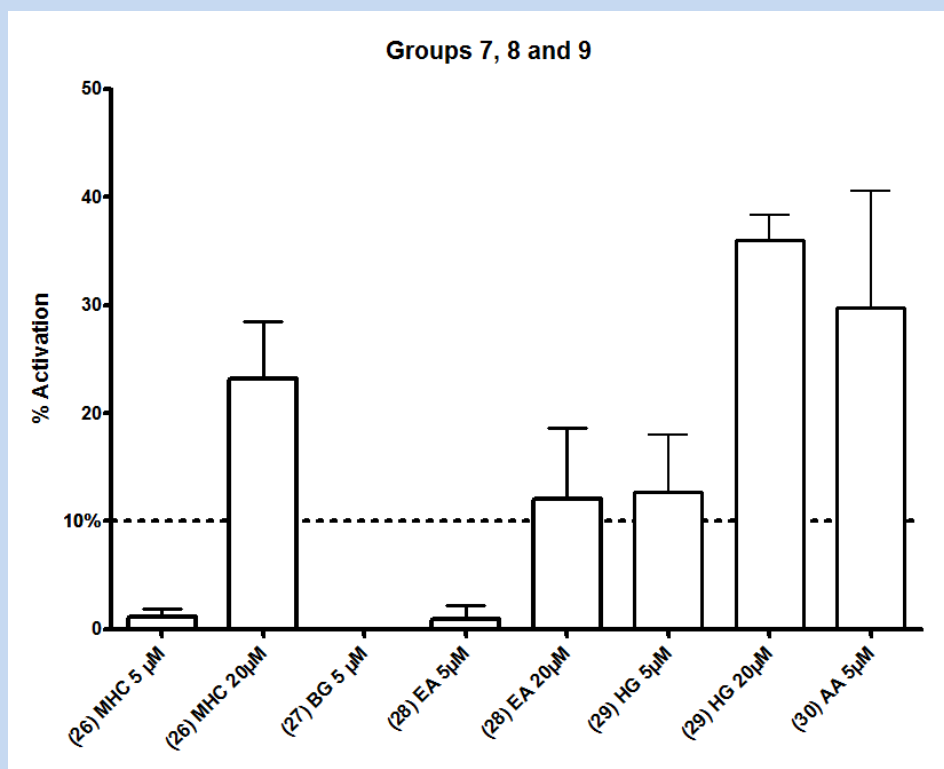


Figure 34: GPBAR1 activation of virtual hits from HCA-groups 8 and 9 at 5 and 20 μM using a luciferase reporter gene assay: Compounds **(26)**, **(28)**, **(29)** and **(30)** are clearly active, Compound **(25)** digitoxigenin was cytotoxic at either concentration. **(27)** Bayogenin was inactive at 5 μM and 20 μM . All stated values are the mean of at least 3 independent experiments and the variance is given as SEM.

5.10.6. Activity of extracts

Since the two sesquiterpene-coumarins farnesiferol B (**21**) and micolobiden (**22**) had been predicted as hits, two samples from the originated *Ferula assa-foetida* extracts (resin) were tested at 3 concentrations: 10 µg, 30 µg and 100 µg per ml. A third extract from *Phytolacca americana* roots was prepared, since phytolaccagenic acid and two saponins phytolaccoside B and esculentoside D had been virtual hits too, and all of them have been identified from *phytolacca americana* roots. For more information on extract description and preparation see chapter 6.7. **Extract Preparation.**

Interestingly, none of the extracts had been active at any concentration. The main reason for the fact that isolated compounds show activity in the assay but extracts, which contain the same compounds, do not, can be various. A plausible explanation is that the compounds are simply too low concentrated in the extracts. Other reasons can be assay interference or ingredients of the extract, which can antagonize the effects of the agonistic constituents. One thing is clear: It has to be stated that with a classical bio-guided fractionation the discovery of micolobiden or phytolaccagenic acid as GPBAR1 agonist would have not been possible and, thus, demonstrates the power of *in silico* screening.

5.10.7. Results on the experimental testing of Plant A's Triterpenes

All the triterpenes of the South African medicinal plant¹⁷⁸ showed activity, with one exception, compound (**34**) 27-hydroxyalphitolic acid. The results are presented in **Figure 35**. The triterpenes were only able to be evaluated at 5 µM, because as can be see in **Figure 35**, the cytotoxicity increases at higher concentration.

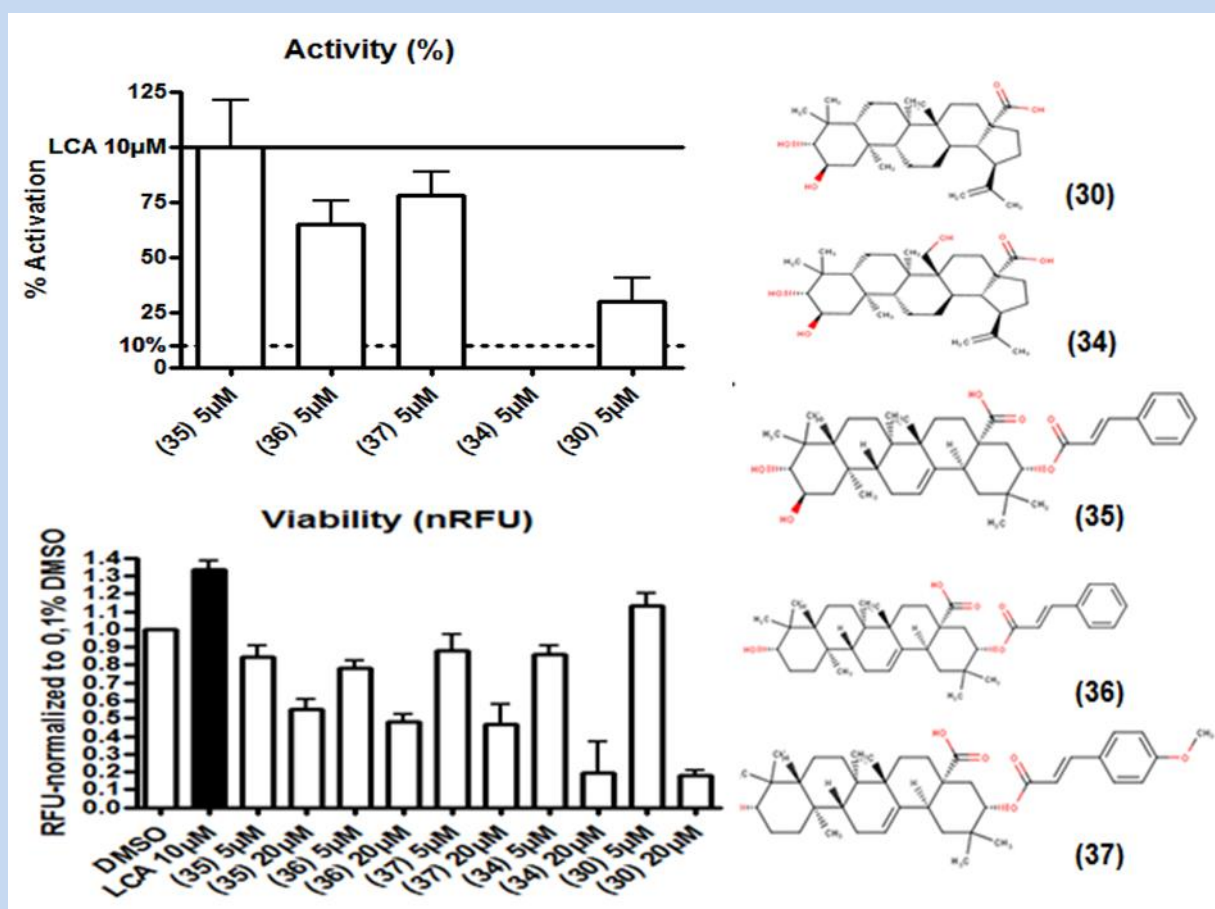


Figure 35: GPBAR1 activation and viability evaluation of the triterpenes of a South African medicinal plant¹⁷⁸ at 5 and 20 μM using a luciferase reporter gene assay:

Their activity on GPBAR1, expressed as % activation compared to 10 μM LCA (positive control, solid line) presented on the upper left side. The positive control LCA at 10 μM was set to 100% activation (solid line). The negative control DMSO 0.1% (vehicle control) was set to 0% activation. The threshold was set to 10% (dotted line).

Their effect on cell viability is expressed as RFUs normalized to the solvent 0.1% DMSO with 10 μM LCA and 0.1% DMSO as control is presented on the bottom left side.

Each value is the mean of at least 3 independent experiments and the variance is given as SEM.

5.10.8. Summary

Summed up it can be stated that the two validated models are very suitable to detect agonists within the range of their structural space, namely triterpenes and steroid-like structures as it was theoretically predicted before. The models performed poor on those groups, consisting of deviating structures (group 1 and 2). Nevertheless, the models were able to retrieve some hits

also in these areas and proved their ability to perform scaffold-hopping as hits like microlobiden (85.23% activation at 20 μ M) demonstrate. **Table 24** shows the real hit-rate of both models in canon. The models predicted 38.71% (12/31) compounds correctly in the experimental validation.

Table 24: Real hit-rate of the models BAMS22 and Model12.3.8. in canon of the validated hit-selection
***Digitoxigenin and SC14 were not evaluable at any concentration, due to their strong cytotoxic properties in this assay. They were therefore not counted as either active or inactive.**

Cluster 1	7.69% (1/13)
Cluster 2	37.50% (3/9)*
Cluster 3	Not validated
Cluster 4	0% (0/1)
Cluster 5	100% (2/2)
Cluster 6	100% (2/2)
Cluster 7	100% (1/2)*
Cluster 8	50% (1/2)
Cluster 9	100% (2/2)
Together	38.71% (12/33)*

6. Materials and methods

6.1. Hardware information

The work was done on 2 different systems. The libraries and models were generated by a Lenovo_BI_A3 Tower with Windows 7 professional, 64bit operating system with Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10 Ghz (x2) processor, 8.00 GB RAM, an NVIDIA Quadro 410 graphical device and a WDC WD10EZEX-00BN5A0 ATA device.

Model improvement and VS was done on a notebook with Windows 10 home, 64bit operating system capable of an Intel® Core™ i5-4310M CPU@ 2.70GHz (2x) processor, 8.00 GB RAM, an NVIDIA GeForce GT 745 graphical device and a Toshiba MQ01ABF050 hard drive.

6.2. Software information

6.2.1. Ligand Scout

The pharmacophore model generation, subsequent refinement and VS was performed with inte:ligand's LigandScout 4.08. Detailed information about LiganScout is outlined in the introduction section.

6.2.2. I:Con

I:Con is the conformer generator which is implemented in LigandScout. Its conformer generation algorithm is based on a systematic knowledge-based approach. It is a robust, fast, effective and recently extensively validated conformer generator. Compared to the probably best conformer generator Omega by OpenEye its performance is only slightly inferior with little differences depending on the used settings. The conformational libraries for both pharmacophore modeling and the subsequent VS process were created with I:Con.

6.2.3. Marvin Sketch

ChemAxon's Marvin Sketch is a suitable chemical structure drawing tool. 2D structures, which had been the starting point for model generation, were drawn with Marvin Sketch 16.1.18.0.

6.2.4. ChemGPS

A first priority of any database creation and pharmacoinformatics is to visualize data. ChemGPS-NP is a Principal Component Analysis (PCA) free web service adjusted for the assessment of a molecule's chemical space, particularly in the ranges of biologically active compounds. An eight dimensional (8D) mapping, established by chemical and physicochemical properties for a reference collection of structures, was performed. The first four magnitudes (PC1-PC4) incorporate an estimated data variance of 77% and can be interpreted as follows:

PC1 (=principal component one) or the first dimension of the ChemGPS map as a characterization of the size, shape and polarizability of the molecules.

PC2 describes aromatic and conjugation-related properties.

PC3 corresponds to lipophilicity, polarity, and H-bond capacity.

PC4 matches a molecule's flexibility or rigidity.

Utilizing interpolation in terms of PCA score prediction makes it possible to place every compound with a given 2D structure (smiles-code) onto this map. The emerging projection of the molecules leads to an easy chemical space comparison, clustering and trend analysis¹⁸⁶. Throughout the thesis ChemGPS-NP web service was used for visualization, comparison and clustering of molecule data.

6.2.5. SIMCA

Umetric's SIMCA is a multivariate analysis software that allows the user to easily visualize trends and clusters using the intuitive graphical interface.

6.2.6. ROCS

Open Eye's ROCS is a reputed VS software based on shape comparison¹⁸⁷. The application is based on a smooth Gaussian function to compute molecular volume. The subsequent alignment comparison of molecular volumes is based on the assumption that similar volumes lead to similar shapes¹⁸⁸. It is both a fast (20 – 40 compounds screened per second¹⁸⁹) and a sophisticated VS application¹⁹⁰. In this thesis it was used for shape comparison in order to evaluate the LigandScout's VS hitlist. As quality metric the Tanimoto combo (TC) score was used. It comprises both the Shape Tanimoto and the Color Tanimoto Score.

6.3. Database sources

Table 25: Database sources and number of extracted EC50 values

Source	EC50 values extracted
CHEMBL	623
SciFinder	293
Patents	66
Internal Data	43

6.3.1. External sources

6.3.1.1. ChEMBL-database

The ChEMBL database, maintained by the European Bioinformatics Institute, represents a platform combining information on protein targets and small molecules with their attached bioactivity data. Moreover, information on the used method for activity evaluation and information on the primary literature is available. It consists of 1,928,903 compounds with 13,967,816 connected activities on 11,019 targets (date 7.6.16) The ChEMBL database is accessible free of charge, making it a resourceful tool for bioinformatics¹⁹¹.

6.3.1.2. Scifinder

SciFinder® by the chemical abstract services (CAS) is probably one of the most important research applications for chemical information. Scifinder provides advanced search screen,

which provides access to the biggest and most respected database of references, substances and reactions in chemistry and related sciences, as well as synthesis options and commercial sources¹⁹².

6.3.1.3. Patent Databases

Finding and extracting useful information on the patented GPBAR1 modulators is more challenging than extracting it from ChEMBL or published papers. The PATENTSCOPE database from the World Intellectual Property Organization (WIPO) offers therefore a valuable alternative with access to plenty of patent documents. According to their website WIPO is the world's most comprehensive source of data on the intellectual property (IP) system, as well as of empirical studies, reports and factual information on IP.¹⁹³ Further patents.google.com¹⁹⁴ (including next to WIPO also USPTO, EPO, JPO, SIPO, DPMA, and CIPO patents), worldwide.espacenet.com and lens.org have been used as an external source of patent data.

Two reviews on patented GPBAR1/FXR agonists are highly recommended:

- (1) Gioiello A, Rosatelli E, Nuti R, Macchiarulo A, Pellicciari R. Patented TGR5 modulators: a review (2006 – present). *Expert Opinion on Therapeutic Patents* 2012;22(12):1399-414.
- (2) Xu Y. Recent Progress on Bile Acid Receptor Modulators for Treatment of Metabolic Diseases. *J Med Chem* 2016.

6.3.2. Internal data

Besides the extensive data from literature, we were able to resort to internal data from previous projects in our department.

6.4. Database Creation

The starting point for database creation was the bioactivity data provided by the ChEMBL database in form of a downloadable excel sheet. The available data for GPBAR1 was extracted on March 15th 2016. It consisted of 24 different publications with 623 attached EC₅₀ activity values (for some molecules there had been multiple/varying values reported; see **Table 7** and

8) on GPBAR1 (**Table 26**). The database was extended by extracting bioaffinity data on GPBAR1 found in various publications in Scifinder. Another 18 publications with 293 EC₅₀ values was proficient for extracting (**Table 27**). A very recent review¹ on known GPBAR1 agonists offered not only an overview on published ligands (already contained in the dataset) but also gave information on several patent claims on this topic. Accordingly, 23 compounds from those patents referred by Xu¹ were added. Several other patents lead to a final patent compound collection consisting of 66 further compounds. During this procedure, new scaffolds were of special interest for the database, which was not maintained yet (**Table 28**). Useful internet sources for that were patentscope, lens.org, espacenet and google-patents. (see **6.4.1.3 Patent databases**)

Table 26: Publications from the ChEMBL-database

	ChEMBL ID	Publication Year	Publication Journal	EC ₅₀ values extracted	Source
[1]	CHEMBL1144009	2008	J. Med. Chem.	68	Sato, Macchiarulo, Thomas, Gioiello, Une, Hofmann, Saladin, Schoonjans, Pellicciari and Auwerx ⁴⁰
[2]	CHEMBL1148757	2007	J. Med. Chem.	9	Pellicciari, Sato, Gioiello, Costantino, Macchiarulo, Sadeghpour, Giorgi, Schoonjans and Auwerx ¹⁹⁵
[3]	CHEMBL1154698	2009	J. Med. Chem.	5	Evans, Budzik, Ross, Wisnoski, Jin, Rivero, Vimal, Szewczyk, Jayawickreme, Moncol, Rimele, Armour, Weaver, Griffin, Tadeballi, Jeune, Shearer, Chen, Chen, Anderson, Becherer, De Los Frailes and Colilla ⁶⁵
[4]	CHEMBL1154755	2009	J. Med. Chem.	6	Pellicciari, Gioiello, Macchiarulo, Thomas, Rosatelli, Natalini, Sardella, Pruzanski, Roda, Pastorini, Schoonjans and Auwerx ⁸⁵
[5]	CHEMBL1156761	2010	J. Med. Chem.	56	Genet, Strehle, Schmidt, Boudjelal, Lobstein, Schoonjans, Souchet, Auwerx, Saladin and Wagner ⁶¹
[6]	CHEMBL1156896	2010	Bioorg. Med. Chem. Lett.	66	Budzik, Evans, Wisnoski, Jin, Rivero, Szewczyk, Jayawickreme, Moncol and Yu ⁶⁶

[7]	CHEMBL1255552	2010	Bioorg. Med. Chem. Lett.	31	Herbert, Siegel, Staszewski, Cayan, Banerjee, Dhamija, Anderson, Fan, Wang, Rix, Shiao, Rao, Noble, Heyman, Bischoff, Guha, Kabakibi and Pinkerton ⁷⁰
[8]	CHEMBL1764964	2011	Bioorg. Med. Chem.	4	Gioiello, Macchiarulo, Carotti, Filippini, Costantino, Rizzo, Adorini and Pellicciari ⁶⁸
[9]	CHEMBL2062395	2012	ACS Med. Chem. Lett.	8	Pellicciari, Gioiello, Sabbatini, Venturoni, Nuti, Colliva, Rizzo, Adorini, Pruzanski, Roda and Macchiarulo ¹⁹⁶
[10]	CHEMBL2177042	2012	J. Med. Chem.	40	Duan, Ning, Chen, Zou, Zhang, Feng, Zhang, Leng and Shen ⁶²
[11]	CHEMBL2331085	2013	ACS Med. Chem. Lett.	36	Piotrowski, Futatsugi, Warmus, Orr, Freeman-Cook, Londregan, Wei, Jennings, Herr, Coffey, Jiao, Storer, Hepworth, Wang, Lavergne, Chin, Hadcock, Brenner, Wolford, Janssen, Roush, Buxton, Hinchey, Kalgutkar, Sharma and Flynn ⁹⁰
[12]	CHEMBL2331106	2013	Bioorg. Med. Chem. Lett.	70	Londregan, Piotrowski, Futatsugi, Warmus, Boehm, Carpino, Chin, Janssen, Roush, Buxton and Hinchey ⁸¹
[13]	CHEMBL2406956	2013	Bioorg. Med. Chem. Lett.	44	Dehmlow, Alvarez Sanchez, Bachmann, Bissantz, Bliss, Conde-Knape, Graf, Martin, Obst Sander, Raab, Richter, Sewing, Sprecher, Ullmer and Mattei ⁶⁷
[14]	CHEMBL2434932	2013	Eur. J. Med. Chem.	32	Zhu, Ning, Guo, Zhang, Pan, Leng and Shen ⁷¹
[15]	CHEMBL3097990	2013	ACS Med. Chem. Lett.	6	Macchiarulo, Gioiello, Thomas, Pols, Nuti, Ferrari, Giacche, De Franco, Pruzanski, Auwerx, Schoonjans and Pellicciari ¹⁰

[16]	CHEMBL3132829	2013	MedChemComm	13	Futatsugi, Bahnck, Brenner, Buxton, Chin, Coffey, Dubins, Flynn, Gautreau, Guzman-Perez, Hadcock, Hepworth, Herr, Hinchey, Janssen, Jennings, Jiao, Laverne, Li, Li, Munchhof, Orr, Piotrowski, Roush, Sammons, Stevens, Storer, Wang, Warmus, Wei and Welford ⁸⁰
[17]	CHEMBL3232835	2014	J. Med. Chem.	44	Phillips, Gao, Yang, Zhang, Lerario, Lau, Jiang, Wang, Nguyen, Bhat, Trotter, Sullivan, Welzel, Landry, Chen, Joseph, Li, Gordon, Richmond, Johnson, Bretz, Bursulaya, Pan, McNamara and Seidel ⁸²
[18]	CHEMBL3259647	2014	J. Med. Chem.	2	D'Amore, Di Leva, Sepe, Renga, Del Gaudio, D'Auria, Zampella, Fiorucci and Limongelli ⁷²
[19]	CHEMBL3286383	2014	Eur. J. Med. Chem.	40	Zou, Duan, Ning, Liu, Feng, Zhang, Zhu, Leng and Shen ¹⁹⁷
[20]	CHEMBL3351323	2014	J. Med. Chem.	3	Festa, Renga, D'Amore, Sepe, Finamore, De Marino, Carino, Cipriani, Monti, Zampella and Fiorucci ⁴¹
[21]	CHEMBL3352137	2014	J. Med. Chem.	2	Sepe, Renga, Festa, D'Amore, Masullo, Cipriani, Di Leva, Monti, Novellino, Limongelli, Zampella and Fiorucci ¹⁹⁸
[22]	CHEMBL3352618	2014	Bioorg. Med. Chem. Lett.	14	Park, Ahn, Jung, Bang, Kim, Hong, Kim, Suh, Kim, Kim, Lee and Min ⁷³
[23]	CHEMBL3399951	2015	Bioorg. Med. Chem.	12	Yu, Sousa, Mattern, Wagner, Fu, Vaidehi, Forman and Huang ⁷⁴
[24]	CHEMBL3421628	2015	J. Med. Chem.	12	Duan, Ning, Zou, Ye, Feng, Zhang, Leng and Shen ⁶⁴

Table 27: Publications from the Scifinder-database

EC ₅₀ values extracted					
Doc_ID	Publication Year	Publication Journal	EC ₅₀ values extracted	Source	
[1]	REf135Scif	2014	ChemMedChem	15	Hogenauer, Arista, Schmiedeberg, Werner, Jaksche, Bouhelal, Nguyen, Bhat, Raad, Rauld and Carballido ¹⁹⁹
[2]	Ref136Scif	2014	Arch. Pharm. Research	4	Kim, Park, Lee, Rhyu and Kim ⁶⁹

[3]	Ref160Scif	2014	J. Pharmacol. Exp. Ther.	7	Roda, Pellicciari, Gioiello, Neri, Camborata, Passeri, De Franco, Spinozzi, Colliva, Adorini, Montagnani and Aldini ⁸⁶
[4]	Ref176Scif	2013	Diabetes, Metab. Syndr. Obes.: Targets Ther.	4	Zambad, Tuli, Mathur, Ghalsasi, Chaudhary, Deshpande, Gupta, Chauthaiwale and Dutt ²⁰⁰
[5]	Ref178Scif	2013	ChemMedChem	59	Martin, Bissantz, Gavelle, Kuratli, Dehmlow, Richter, Obst Sander, Erickson, Kim, Pietranico-Cole, Alvarez-Sanchez and Ullmer ⁹³
[6]	Ref17Scif	2015	Future Med. Chem.	1	Lamers, Merk, Gabler, Flesch, Kaiser and Schubert-Zsilavec ²⁰¹
[7]	Ref198Scif	2013	Expert Opin. Ther. Patents	10	Raval ²⁰²
[8]	Ref201Scif	2013	ChemMedChem	42	Zhu, Ye, Ning, Mandi, Feng, Zou, Kurtan, Leng and Shen ⁶³
[9]	Ref241Scif	2013	Clin. Pharmacol. Drug Dev.	1	Hodge, Lin, Vasist Johnson, Gould, Bowers and Nunez ⁸⁷
[10]	Ref27Scif	2015	PLOS ONE	2	Zheng, Zhou, Wang, You, Zhao, Yang, Wang, Luo, Chen, Liu and Chen ⁸³
[11]	Ref314scif	2011	Biol. Pharm. Bull	16	Iguchi, Nishimaki-Mogami, Yamaguchi, Teraoka, Kaneko and Une ⁷⁶
[12]	REf415Scif	2008	J. Med. Chem.	61	Macchiarulo, Gioiello, Thomas, Massarotti, Nuti, Rosatelli, Sabbatini, Schoonjans, Auwerx and Pellicciari ¹⁷
[13]	Ref429Scif	2007	J. Med. Chem.	9	Pellicciari, Sato, Gioiello, Costantino, Macchiarulo, Sadeghpour, Giorgi, Schoonjans and Auwerx ¹⁹⁵
[14]	Ref6Scif	2016	Chem. Pharm. Bull.	42	Chen, Ning, Zou, Cao, Ye, Leng and Shen ⁷⁵

[15]	Ref73Scif	2015	PLOS ONE	1	Briere, Ruan, Cheng, Siesky, Fitch, Dominguez, Sanfeliciano, Montero, Suen, Xu, Coskun and Michael ⁹⁸
[16]	Ref7Scif	2016	ACS Med. Chem. Lett	13	Agarwal, Patil, Aware, Deshmukh, Darji, Sasane, Sairam, Priyadarsiny, Giri, Patel, Giri, Jain and Desai ²⁰³
[17]	Sato	2007	Biochem. Biophys. Res. Commun.	2	Sato, Genet, Strehle, Thomas, Lobstein, Wagner, Mioskowski, Auwerx and Saladin ⁶⁰
[18]	Maruyama	2002	Biochem. Biophys. Res. Commun	4	Maruyama, Miyamoto, Nakamura, Tamai, Okada, Sugiyama, Nakamura, Itadani and Tanaka ⁵

Table 28: Publications from the Scifinder-database

	Application Number	Publication date	Inventors/ Applicants	Source
[1]	US 2006/0199795 A1	07.09.2006	Takeda Pharmaceuticals	Itoh, Hinuma, Kanzaki, Miki, Kawamata, Oi, Tawaraishi, Ishichi and Hirohashi ²⁰⁴
[2]	WO 2010/014739 A2	04.02.2010	Kalypsis Inc.	Smith, Payne and Hoffman ²⁰⁵
[3]	WO 2013/049585 A1	04.04.2013	Hengrui	Smith, FITZGERALD, Jiang, Atkinson, Veal and HUANG ²⁰⁶
[4]	WO 2013/096771 A1	27.06.2013	Ardelyx Inc.	Lewis, Reich, Chen, Jacobs, Charmot, Navre, FINN, Carreras and Spencer ²⁰⁷
[5]	WO 2013/164838 A1	07.11.2013	Cadila Healthcare Limited	Agarwal and Desai ²⁰⁸
[6]	WO 2013/054338 A1	18.04.2013	Cadila Healthcare Limited	Agarwal, Jain and Patel ²⁰⁹
[7]	WO 2013/102929 A1	11.7.2013	Cadila Healthcare Limited	Agarwal and Desai ²¹⁰
[9]	WO 2013/134527 A1	12.9.2013	Rigel Pharmaceuticals	Goff, Payan and Braselmann ²¹¹
[10]	WO 2010/117084 A1	14.10.2010	Banyu Parmaceutcal Co.	Sugimoto, Satoh and Nishimura ²¹²
[11]	WO 2010/117090 A1	14.10.2010	Banyu Parmaceutcal Co.	Arakawa, Sugimoto, Sasaki, Satoh and Nishimura ²¹³

[12]	WO 2013/057650 A1	25.04.2013	Torrent Pharmaceuticals Limited	Dutt, Chauthaiwale, Gupta, GHALSASI, Tuli, Deshpande, Chaudhari and ZAMBAD ²¹⁴
[8]	US 2012/0010190 A1	12.11.2012	Bissantz Caterina et al	Bissantz, Dehmlow, Erickson, Kim, Martin, Mattei, Sander, Pietranico-Cole, Richter and Ullmer ²¹⁵
[13]	WO 2013/060653 A1	02.05.2013	HOFFMANN-LA ROCHE INC.	Dehmlow, Martin, Mattei, Obst and Richter ²¹⁶
[14]	WO 2012/117000 A1	07.09.2012	HOFFMANN-LA ROCHE INC.	Bissantz, Dehmlow, Erickson, Karnachi, Kim, MARTIN, Mattei, Obst, Pietranico-Cole and Richter ²¹⁷
[15]	WO 2013/149977 A1	10.10.2013	HOFFMANN-LA ROCHE INC.	Dehmlow, Erickson, Mattei and Richter ⁸⁸
[16]	US 2013/0123267 A1	16.03.2013	HOFFMANN-LA ROCHE INC.	Dehmlow, Martin, Mattei, Sander and Richter ²¹⁸
[17]	US 2010/0105906 A1	29.04.2010	HOFFMANN-LA ROCHE INC.	Bissantz, Dehmlow, Martin, Sander, Richter and Ullmer ²¹⁹
[18]	WO 2012/082947 A1	21.6.2012	IRM, LLC	Gao, Lau, Pan, Phillips, Wang and Yang ²²⁰
[19]	WO 2010/093845 A1	19.8.2010	Exelexis Inc.	Bollu, Boren, Dalgard, Flatt, Haq, Hudson, Mohan, Morrissey, Pratt and Wang ²²¹
[20]	WO 2014/100021 A1	26.6.2014	Exelexis Inc.	Flatt and Mohan ⁸⁹
[21]	WO 2014/100025 A1	26.6.2014	Exelexis Inc.	Flatt and Mohan ²²²
[22]	WO 2012/149236 A1	01.11.2012	Bristol- Myers Squibb Company	Robl, Li, Kennedy, Walker, Wang, Washburn, Ahmad and Ngu ²²³
[23]	CN 102850321 A	28.06.2011	Chinese Publication	(⁹¹
[24]	CN 103864754 A	18.06.2014	Chinese Publication	(²²⁴

Finally the in-house data was used, because Dr. Ladurner from the department of pharmacognosy in Vienna already measured 43 compounds on their activity on GPBAR1 with an acknowledged CRE-driven luciferase assay in previous studies. This comparable small but due to a lot of reliable inactive compounds valuable bioactivity database was added.

The EC₅₀ values, the assays with which they were obtained, the smiles- codes and the 2D chemical structures were extensively double checked with the primary literature. Every molecule was drawn via Marvin Sketch in 2D with the information from the smiles codes, connected to the

bioaffinity data in ChEMBL, or from information given in the primary literature and saved as sdf-files. If mistakes were detected, they were corrected.

Hereupon the checked molecule set was classified into **(1)** High active compounds, **(2)** Active compounds, **(3)** Inactive and **(4)** True Inactive compounds. A more detailed elaboration on this classification can be found in the R&D section of this thesis. **(5.1.5. Benchmark establishment- Classification of Agonists from the literature)**

The basis for a favorable pharmacophore model generation is next to a reliable Highly Active database, a far larger set of decoys. Due to the common absence of published negative data and therefore only a small set of reliable inactive compounds, a set of decoys were generated via the **DUDE DECOYS** database. The molecules from the **High Active** database operated as templates for obtaining decoys from the DUDe website with at least the same 1D physicochemical properties but dissimilar 2D topology than the active compounds. This guaranteed, until a certain point, the rationality of the approach. It is of major importance that probably none of these molecules have ever been tested on their GPBAR1 activity and therefore there is a small chance of having some actual active compounds in the decoy-set. It would have been favorable to use a large set of inactive compounds, but the possibility was limited as it is based on the occurrence of such data in the literature or from extensive prior experimental investigation. A short discussion on this topic can be found in the R&D section of this thesis. **(5.1.6. Benchmark establishment- Generation of decoys from DUDe)**

To sum up, five databases were created; three of them were used for model generation. Due to the inconsistent reliability of The ACTIVES and INACTIVES conformational libraries, only the **HIGH ACTIVES** and the **TRUE INACTIVES** were further used together with the **DUDE DECOYS** for model building and refinement.

The conformation libraries were created with iCon (**see chapter 6.2.2.**), integrated in LigandScout 4.0. The settings were chosen as follows: Timeout (sec): 600, RMS threshold: 0.8, energy window: 20.0, max. pool size: 4000, max fragment build time: 30 and for each molecule 200 conformations were generated and saved. These settings had been the default best settings in LigandScout.

6.5. Pharmacophore Modeling

6.5.1. Clustering

As described in the previous section **6.3** and **Table 9** the research relied on the best data from the highly active compounds. These 336 substances were clustered in LigandScout 4.0 with pharmacophore RDF-Code Similarity measure, with options distance 0.9 and MAXIMUM. 3 conformations for each molecule were considered.

The aim of this process was to group compounds with the same pharmacophore attributes into few clusters. Compounds with same pharmacophore features and characteristics are more likely to cover similar pharmacophores and this was a favorable starting position for the modeling process.

The RDF-Code Similarity measure is a fast 3D alignment process, grouping the compounds based on their similarity. The similarity distance 0.9 with the maximum distance calculation seemed beneficial after some trials. These 3 conformations are recommended in LigandScout, because in the end the algorithm combinatorially aligns all conformations of all compounds.

With this clustering setup the 336 molecules were separated into 12 clusters (**Figure 36**).

Cluster 1 was not used for model generation, because it existed of one single molecule and this molecule wasn't applicable to pharmacophore modeling in LigandScout 4.0. The synthetic compounds were classified into 9 clusters: Cluster 2 was comprised of 13, cluster 3 28, cluster 4 11, cluster 5 28, cluster 6 41, cluster 8 80, cluster 9 35, cluster 10 46, and cluster 11 14 molecules.

NPs were found in two clusters: In cluster 7 most of the bile acid- or steroid-like structures, i.e. LA, were categorized. All the triterpenes such as betulinic or oleanolic acid were classified in cluster 12. Consequently, the most pharmacophore modeling efforts have been done on those models describing cluster 7 and cluster 12.

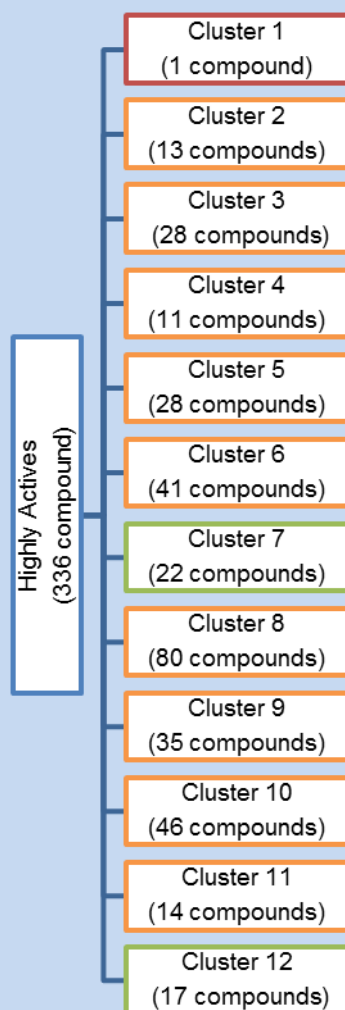


Figure 36: Scheme of the clustering process. 12 clusters were received with our setup.

Red boxes were not used for modeling.

Orange boxes constituted of mainly synthetic compounds.

Green boxes constituted of mainly natural compounds.

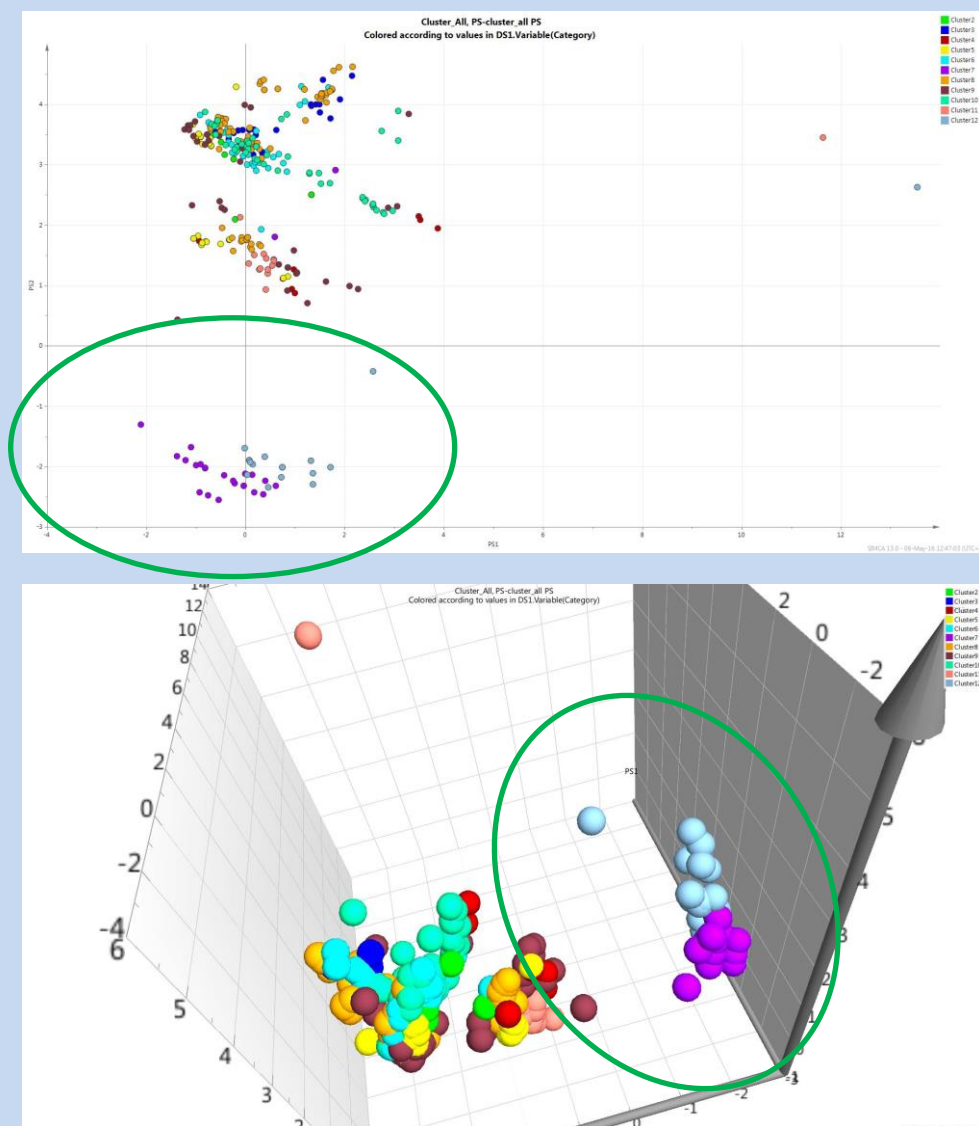


Figure 37: The High Actives plotted 2D (above PC1 and PC2) and 3D (PC1, PC2, PC3). The molecules from cluster 7 and 12 (green ring) differ from the other molecules mainly in properties relating to PC2.

6.5.2. Training sets for modeling

6.5.2.1. Training set for the modeling of BAMS 22(the bile acid model)

Cluster 7 consisted of 22 compounds (**Table 29**). The dominant scaffold is that of a steroid and most of the compounds are either NPs (i.e. Chembl1478-LCA, Chembl258818-Lithoglycocholol) or NP-derived (i.e. CHEMBL566315-INT-747). In contrast, CHEMBL2407934 and CHEMBL2407932 have a completely different scaffold with a synthetic origin and are similar to the molecules from cluster 2. Although they seem to share some pharmacophore-characteristic similarity with the other molecules, no proficient model was able to incorporate those obviously very different molecules. It is very likely that those molecules bind to GPBAR1 in a different way than bile acids or steroids like CHEMBL407717. A recent study suggested a different binding pose for nicotinamides and bile acids¹⁹⁹ and the fact that no model was able to incorporate both structure types strengthens the assumption that pyrimidine like molecules also have a different binding pose. As a result these two compounds were removed from the training set.

When the PC of Cluster 7's molecules are plotted PC1 against PC3, they seem to share the same physicochemical space (**Figure 38**). Although CHEMBL2407932 and CHEMBL2407934 largely differ in 3D, concerning(**Figure 39**).

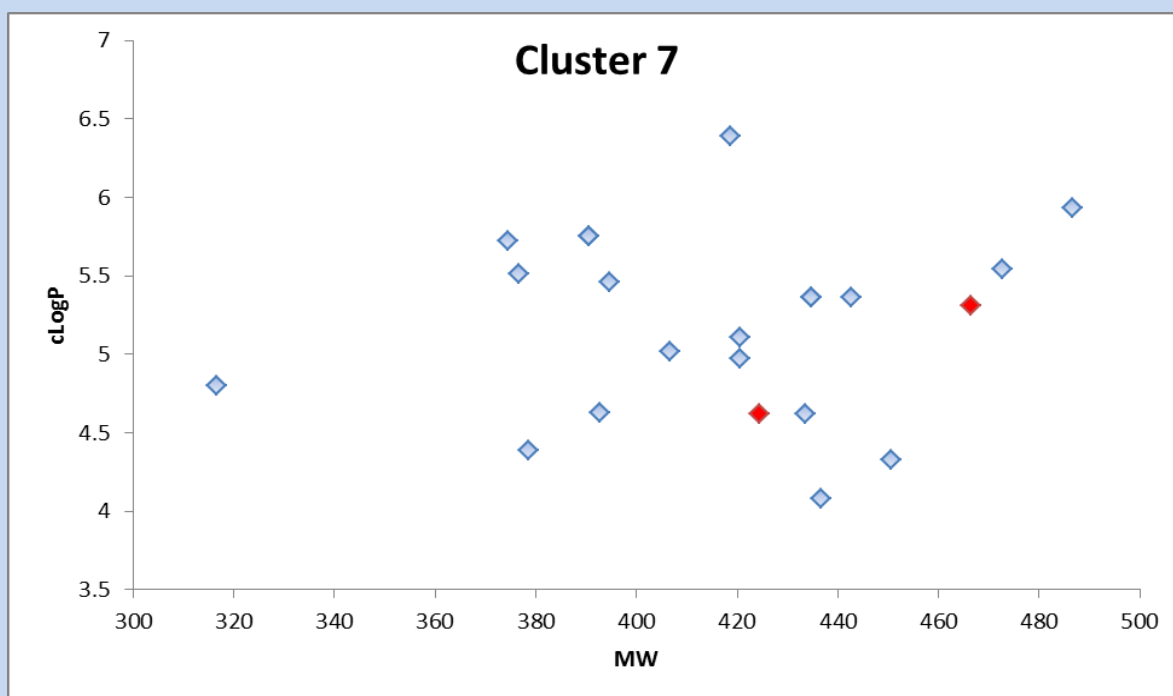


Figure 38: All the compounds contained in cluster 7 were plotted against their lipophilicity and their molecular weight in Excel. As you can see the compounds CHEMBL2407934 and CHEMBL2407932 seem to share next to their pharmacophore characteristics quite the same chemical space.

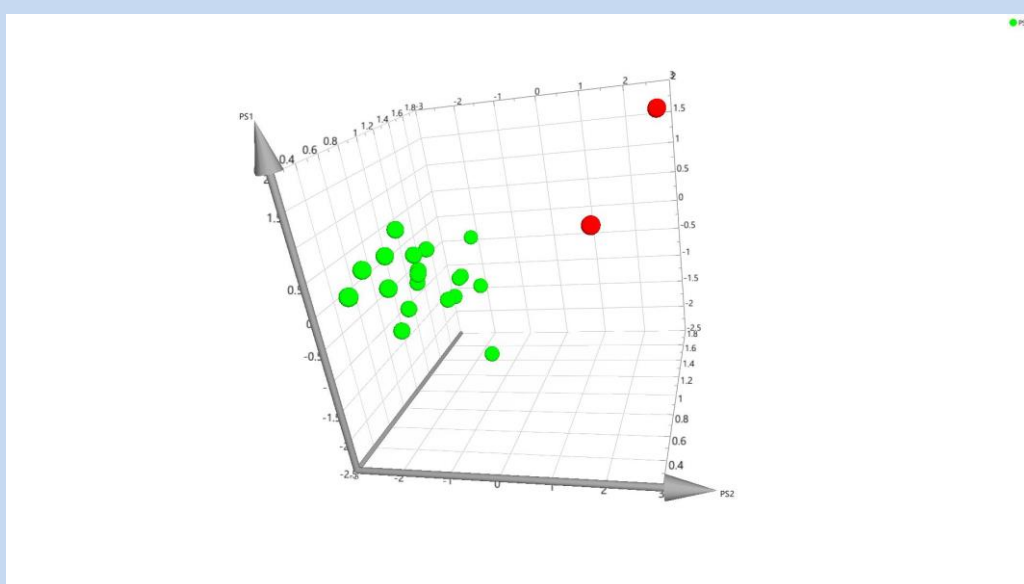
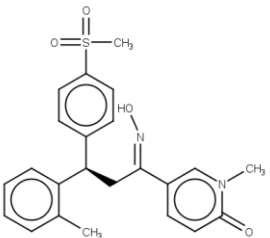
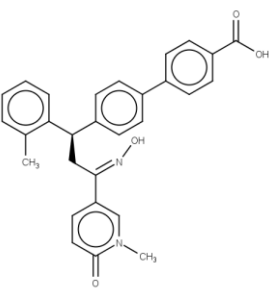
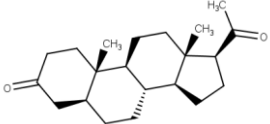
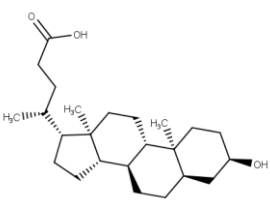
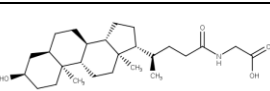
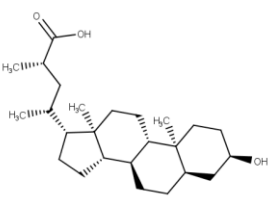
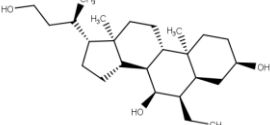
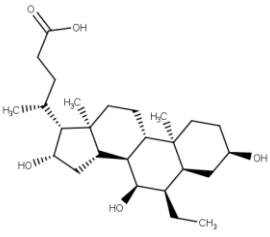
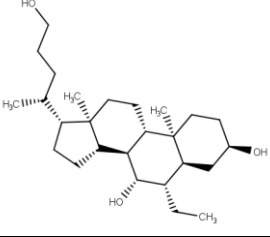
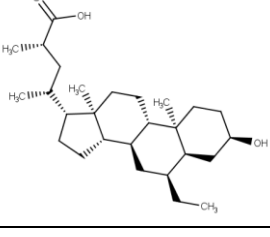
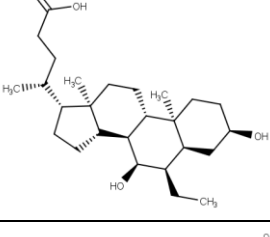
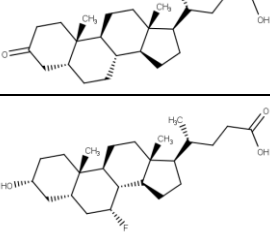
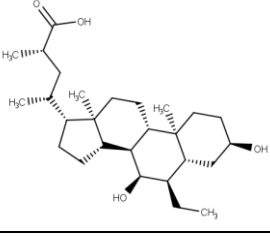
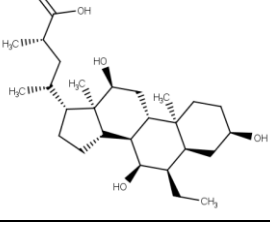
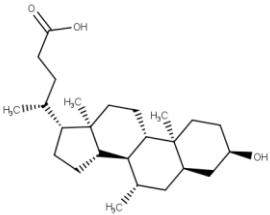
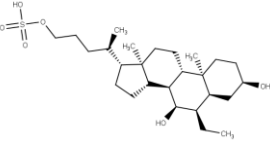
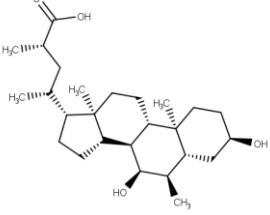
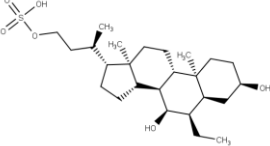
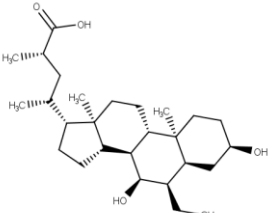
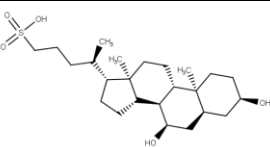
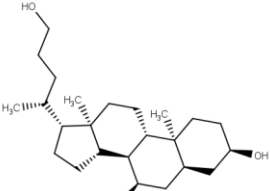


Figure 39: PCA of the chemical space occupied by the molecules classified in cluster 7. The red dots stand for CHEMBL2407934 and CHEMBL2407932. They differ largely from the other compounds in this cluster.

Table 29: All molecules clustered into Cluster 7 with their chemical structure, their ChEMBL ID, 3 physicochemical descriptors (*calculated in LigandScout 4.0), their source and their role in the modeling process. As you can see the first 2 molecules are synthetic-compounds unlikely to bind the same way as the rest of the cluster.

Structure	Name	MW*	cLogP*	Rel. TPSA*	Used for Building Final Model?	Source
	CHEMBL2407934	424.52	4.62	1.64	IGNORED	CHEMBL2406956 CHEMBL2406956
	CHEMBL2407932	466.54	5.31	1.51	IGNORED	CHEMBL2406956 CHEMBL2406956
	CHEMBL407717	316.48	4.8	0.62	YES	CHEMBL1144009
	CHEMBL1478 LCA	376.58	5.51	0.86	YES	Chembl1144009 Chembl1156761 Chembl2062395 Maruyama Chembl3097990 Chembl3399951 Ref160Scif Ref176Scif Ref314Scif Ref415Scif
	CHEMBL258818 LITHOGLYCO- CHOLATE	433.63	4.62	1.17	TRAINING	Chembl1144009 Ref415Scif
	CHEMBL3401351	390.61	5.75	0.82	TRAINING	CHEMBL3399951 REF415Scif (S)
	CHEMBL3342086 NOREDCOH	392.62	4.63	0.84	TRAINING	CHEMBL3351323

	CHEMBL2062554	436.63	4.08	1.31	TRAINING	CHEMBL2062395
	CHEMBL3342082 EUDCOH	406.65	5.02	0.81	TRAINING	CHEMBL3351323
	CHEMBL3401354	418.66	6.39	0.76	TRAINING	CHEMBL3399951
	CHEMBL566315 INT-747	420.63	5.11	1.05	TRAINING	CHEMBL1148757 CHEMBL1764964 ChEMBL3097990, Ref160Scif Ref429Scif
	CHEMBL410893	374.56	5.72	0.84	YES	CHEMBL1144009 Ref415Scif
	CHEMBL273087	394.57	5.46	0.86	TRAINING	CHEMBL1144009 Ref415Scif
	CHEMBL244785	434.66	5.36	1.01	TRAINING	CHEMBL1148757
	CHEMBL567640 INT-777	450.66	4.33	1.26	YES	ChEMBL1154755 ChEMBL2062395 CHEMBL2331085 CHEMBL2406956 CHEMBL2434932 ChEMBL3097990 CHEMBL3352618 Ref6Scif Ref160Scif Ref201Scif

	CHEMBL270105	390.61	5.75	0.82	YES	CHEMBL1144009 Ref415Scif
	CHEMBL3260994	486.71	5.93	1.32	TRAINING	CHEMBL3259647
	CHEMBL244784	420.63	4.97	1.05	TRAINING	CHEMBL1144009 Ref415Scif
	CHEMBL3306873 INT-767	472.68	5.54	1.37	TRAINING	Ref160Scif, CHEMBL3259647
	CHEMBL566985	434.66	5.36	1.01	TRAINING	CHEMBL1154755
	CHEMBL406923	442.66	5.36	1.32	TRAINING	CHEMBL1144009 Ref415Scif
	CHEMBL270287	378.6	4.39	0.88	YES	CHEMBL1144009 Ref415Scif

6.5.2.2. Training set for modeling of Model 12.3.8. (the triterpene model)

Cluster 12 consisted of 17 compounds (**Table 30**), which can be separated into 3 major scaffolds: 14 compounds were pentacyclic triterpenoid acids. 2 compounds can be seen as steroids with a taurinamide-appendix and a single PEGylated synthetic compound (Chembl3421903), inhibiting a completely different chemical space then the rest of the cluster,

as one can see in **Figure 41**. The reason for that is the simple incapability of the algorithm to cluster those large compounds with the very flexible PEG-link.

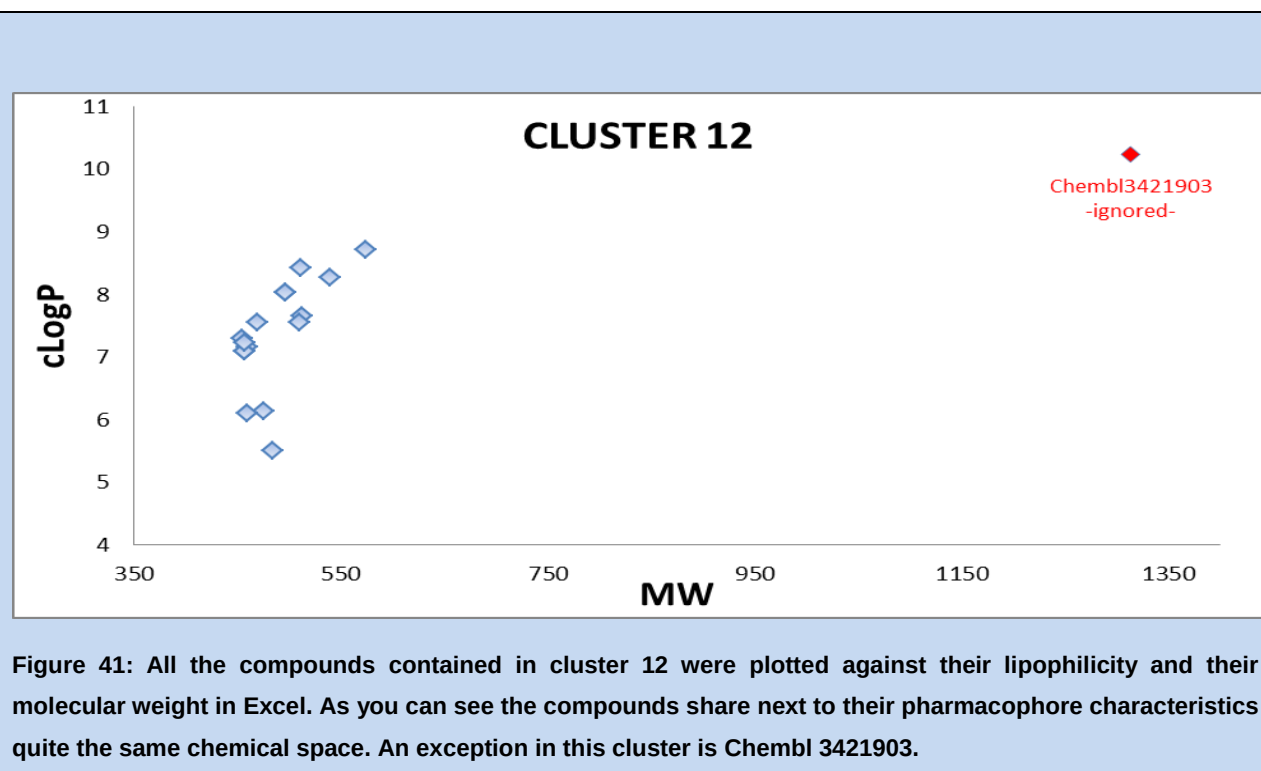


Figure 41: All the compounds contained in cluster 12 were plotted against their lipophilicity and their molecular weight in Excel. As you can see the compounds share next to their pharmacophore characteristics quite the same chemical space. An exception in this cluster is Chembl 3421903.

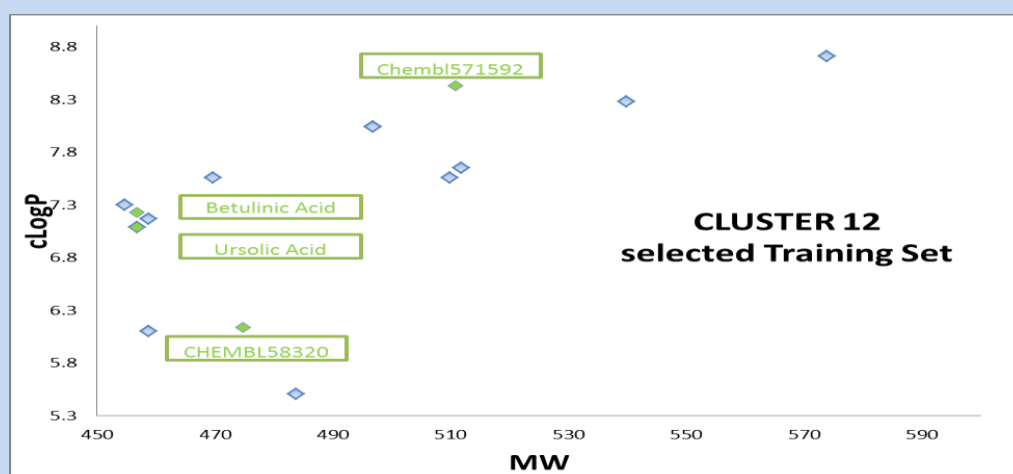
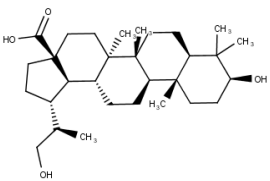
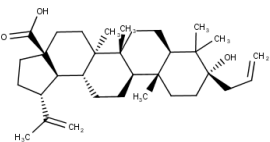
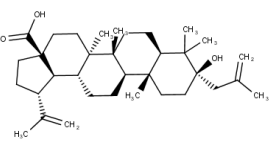
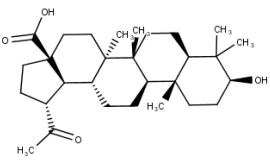
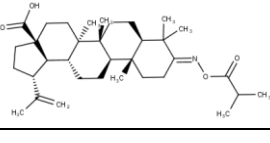
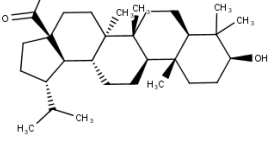
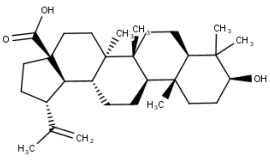
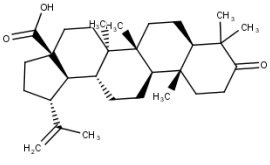
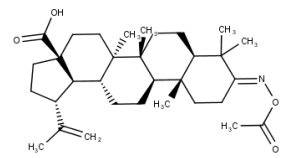
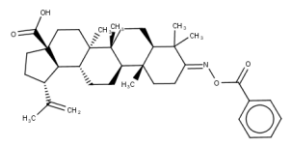
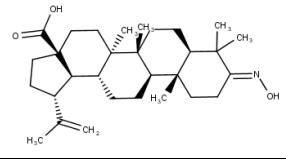
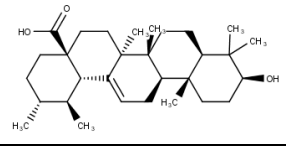
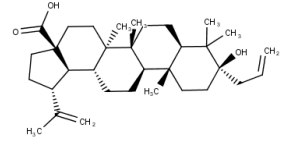
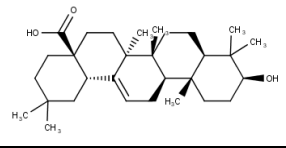
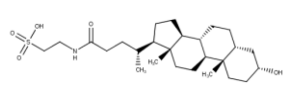
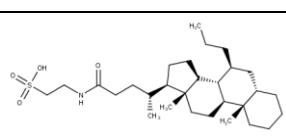
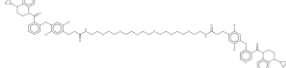


Figure 42: The chemical space of the 4 templates for the initial pharmacophore model and the whole cluster (training set). They noteworthy cover nearly the whole chemical space. Chembl 3421903 was discarded from the training set.

Table 30: All molecules belonging to Cluster 12 with their chemical structure, their ChEMBL ID, 3 physicochemical descriptors (*calculated in LigandScout 4.0), their source and their role in the modeling process. see the first 14 molecules are triterpenoid acids, while the two following molecules are steroids with a taurinamide substructure and finally a synthetic compound with a PEG-link. ChEMBL3421903 was discarded from this cluster, as it wasn't suited for proficient modeling

Structure	Name	MW*	cLogP*	Rel. TPSA*	Used for Building Final Model?	Source
	CHEMBL583203	474.73	6.14	0.93	YES	CHEMBL1156761
	CHEMBL568584	496.78	8.04	0.65	TRAINING	CHEMBL1156761
	CHEMBL571592	510.8	8.43	0.63	YES	CHEMBL1156761
	CHEMBL80460 PLATANIC ACID	458.68	6.1	0.94	TRAINING	CHEMBL1156761
	CHEMBL571493	539.8	8.28	0.83	TRAINING	CHEMBL1156761
	CHEMBL108505 DIHYDRO-BETULINIC ACID	458.73	7.17	0.69	TRAINING	CHEMBL1156761
	CHEMBL269277 BETULINIC ACID	456.71	7.09	0.71	YES	CHEMBL1156761
	CHEMBL431525 BETUNOLIC ACID	454.69	7.3	0.69	TRAINING	CHEMBL1156761

	CHEMBL569634	511.75	7.65	0.88	TRAINING	CHEMBL1156761
	CHEMBL571478	573.82	8.71	0.82	TRAINING	CHEMBL1156761
	CHEMBL296302	469.71	7.56	0.86	TRAINING	CHEMBL1156761
	CHEMBL169 URSOLIC ACID	456.71	7.09	0.71	YES	CHEMBL1156761
	CHEMBL585494	496.78	8.04	0.65	TRAINING	CHEMBL1156761
	CHEMBL168 OLEANOLIC ACID	456.71	7.23	0.71	TRAINING	2007Sato CHEMBL1156761
	CHEMBL408701 TAURO-LITHOCHOLIC ACID	483.71	5.51	1.33	TRAINING	Ref135Scif Ref415Scif CHEMBL1144009
	CHEMBL260737	509.79	7.56	0.97	TRAINING	Ref415Scif CHEMBL1144009
	CHEMBL342190 3	1313.1 8	10.24	1.25	IGNORED	CHEMBL3421628

6.5.3. Prospective virtual screening

Several databases were used for the VS. **Table 31** gives an overview on the screened databases with the amount of molecules, whether the database consists of NPs or synthetic compounds, and the source of the database. Like the workbench molecules, the conformational libraries were built with iCon, the implemented conformer generator in LigandScout 4.0. Conversely to the workbench molecules, the long computing times had to be considered for the way larger databases and therefore not only the best standard settings had been used.

The best settings in iCon computes a maximal number of 200 conformations with Timeout[sec]: 600, RMS threshold of 0.8, an energy window of 20.0, Max Pool Size of 4000 and a maximal fragment built time of 30. The workbench consisting of Highly Actives, True Inactives and Decoys next to the in House database of the department of pharmacognosy, the Nucleo de Bioensaios Biossintese e Ecofisiologica de Produtos Naturais (NuBBE)-database and the Megx- pure NPs database was built this way.

The other screening databases had to be computed with the fast settings of iCon with only 25 conformations each molecule, 600 Timeout [sec], an RMS threshold of 0.5, an energy window of 15.0, maximal pool size of 4000 and maximal fragment built time of 30. The SPECS and SPECS-NP database, the drugbank-database, analyticon's triterpenes-database, the NP database (NPDB) and the Traditional Chinese Medicine (TCM) Taiwan database had to be built.

Table 31: Screened Databases

Database	Number of Molecules	Type	Source
Highly Actives	336	Combination	In house Literature and internal data
True Inactives	49	Combination	In house Literature and internal data
Decoys	18043	Synthetic Compounds	Mysinger et al., 2012 http://dude.docking.org/
Analyticon-MEGx	4355	NPs	Analyticon http://www.ac-discovery.com/
Analyticon-Triterpenes	409	NPs	Analyticon http://www.ac-discovery.com/
Drugbank	6863	Approved Drugs	DrugBank 4.5 http://www.drugbank.ca/
In house database Department of Pharmacognosy University of Vienna	1152	NPs	In house Update 2016
NPDB	115275	NPs	Rollinger et al., 2004 ²²⁵
NuBBE	1628	NPs	Valli et al., 2013 http://nubbe.iq.unesp.br/portal/nubbedb.html
SPECS NP	871	NPs	SPECS http://www.specs.net/
SPECS SC	212446	Synthetic Compounds	SPECS http://www.specs.net/
TCM-Taiwan	35993	NPs	Chen et al., 2011 http://tcm.cmu.edu.tw/

6.5.4. Hit list validation

After the VS experiments only a small set of compounds out of 1128 virtual hits was able to get validated *in vitro*. On that account, an extensive prioritization of the hit list was done. The aim was to select both the most favorable and also structurally diverse molecules. Finally the availability and the price of generally quite expensive NPs played a crucial role for selecting the final shortlist, which evolved into the *in vitro* tests.

First of all, the whole Hit Library went through a second screening process with Open Eye's ROCS. ROCS is a screening software based on shape comparison with a query. For more information see section **6.2.6. ROCS**. The aim was to retrieve an additional quality parameter through a shape-based screening step. For every hit and every matching conformation the TC score was calculated and for molecules with more than one matching conformation the best TC score was chosen for further steps.

Three simple ROCS runs with the default force field, namely implicit mills dean, was performed with three different molecules as a query. The TC scores both against TLC, ChEMBL270105 and betulinic acid in their binding position for every molecule was evaluated.

Secondly, a PCA via the chemGPS tool was performed and a Hierarchical Cluster Analysis (HCA) with SIMCA facilitated us to assign 9 groups of molecules with different chemical space. For the Hierarchical Cluster Analysis the three first descriptors PC1, PC2 and PC3 (ChemGPS **6.2.4.**) were used and a hierarchy level of 9 groups was chosen.

The main goal was to appoint molecules out of every group for the *in vitro* testing. For that reason the molecules got ranked according to their TC scores or/and their pharmacophore fit score. The literature of the best ranked molecules was investigated in detail. From special interest was: **(1)** Has an activity on GPBAR1 been already stated? **(2)** Have the molecules shown anti-inflammatory or antidiabetic effects with unknown or an incoherent mechanism? **(3)** From which species had the NPs been isolated? Is an antidiabetic or anti-inflammatory use of this species reported? **(4)** For hits of the drugbank database, results of clinical or *in vivo* investigations was looked up, with special interest on effects, which could be traced back to GPBAR1 interaction i.e. effects on bile flow²²⁶. **(5)** How much research had been done on the molecules or NPs? Are they known to have a broad range of bioactivities?

Further some chemical descriptors in LigandScout were calculated and possible violations of Lipinski's rule of 5 were evaluated. Considering the four quality metrics (TC score TLC, TC score ChEMBL270105, TC score Betulinic acid, pharmacophore fit score), the knowledge from literature and calculated descriptors led to a small set of appointed molecules.

6.6. Extract preparation

6.6.1. *Phytolacca americana* root extract

The roots from *Phytolacca americana* was derived from the medicinal garden at the Department of Pharmacognosy of the University Vienna. The used extraction protocol was recently described by Schramm and coworkers²²⁷. 300mg of dried roots were defatted with 5ml n-hexane for 1min and consecutively extracted first with 7ml dichloromethane (DCM), second with 13ml methanol in an ultrasonic bath at room temperature for 15min each. The extracts were filtered, combined and dried. Reestablishment was done with 4ml methanol. The reconstituted extract was then loaded onto a solid-phase extraction cartridge, packed with a frit of 20 µM pore size and filled with polyamide gel (CC-6, particle size 0.05–0.16 mm, 900 mg) in order to remove tannins. The gel was washed with 4 ml of methanol twice; the washouts were combined and dried under vacuum. 100 mg of the dried extract was resolved in 1 ml DMSO as a preliminary stock solution, further aliquoted, and stored at –20°C until use.

6.6.2. *Ferula assa-foetida* gum resign extracts

The extracts were made by Kathrin Anrain in 2006 at the Institute of Pharmacy/Pharmacognosy and Center for Molecular Biosciences in Innsbruck in course of a project to find novel inhibitors of the human rhinovirus coat protein²²⁸. The grinded gum-resign had been initially defatted with petrol ether and then extracted with DCM. A liquid/liquid separation led to three partitions: (1) a petrol ether fraction, (2) a methanol/water fraction and a (3) DCM fraction. Fraction (1) and (2) has been tested, namely the DCM fraction and the petrol ether fraction. The last one comprehended several sesquiterpene coumarone ethers. Microlobiden, farnesiferol B and farnesiferol C amongst others had been successfully isolated from this fraction²²⁸. The extracts have been stored at -70°C. 100 mg of each extract were resolved in 1 ml DMSO as a preliminary stock solution, further aliquoted, and stored at –20°C until use.

6.7. Luciferase based reporter gene assay

6.7.1. Chemicals

Hederagenin (CAS#465-99-6) with purity above 97% and Bayogenin (CAS #80368) with purity above 94% were ordered from Phytol (Germany). 2,3-O-Isopropylidenyl-euscaphic acid (CAS #220880-90-0) with purity above 98% was ordered from Proactive Molecular Research (USA). Phytolaccoside B (CAS#60820-94-2) and Euscaphic acid (CAS#53155-25-2) was purchased from Cambridge Chemicals (USA), both with purity above 98%. Phytolaccagenic acid (CAS#54928-05-1, purity >95%) and 16-Dehydropregnenolone (CAS#1162-53-4, purity >98%) were obtained from Carbosynth (UK). Spironolactone (CAS#52-01-7, purity >95%) and Methylhyoxycholate (CAS#2868-48-6, purity<98%) were purchased from TCI Deutschland GmbH (Germany). The screening compounds (CAS#1019061-83-6, CAS#303139-94-8, CAS#330636-58-3, CAS#353253-76-6, CAS#353779-79-0, CAS#432530-00-2, CAS#444931-63-9, CAS#496937-29-2, CAS#791840-52-3, CAS#902244-06-8, CAS#915930-57-3, CAS#932954-51-3, CAS#352644-32-7, CAS#500218-51-9, CAS#314757-83-0, CAS#380633-89-6, CAS#26179-09-9, CAS#664993-86-6, CAS#525577-20-2) were obtained from SPECS (Netherlands) with a purity above 90%. Microlobiden (CAS# 89783-66-4) and Farnesifeol B (CAS# 54990-68-0) were provided by Prof. Rollinger and showed purity above 95%²²⁸. These two chemicals have been stored under light exclusion at -70°C, since 2006. Nordihydroguaretic acid (CAS# 500-38-9) was purchased from Fluka (Switzerland) and showed purity above 97%. The positive controls LCA and CDCA were obtained at Sigma Aldrich (Austria). Alphitolic Acid (CAS#19533-92-7), 27-Hydroxy-Alphitolic Acid, 21 β -Cinnamoyloxymaslinic acid, 21 β -Cinnamoyloxyoleanolic acid and a binary mixture of 21 β -p-Methoxy-cinnamoyloxyoleanolic acid and 21 β -Cinnamoyloxyoleanolic acid were obtained by hydrolysis with Killiani reagent from 5 different saponins isolated and discovered by Christine Maier at the department of pharmacognosy at university Vienna¹⁷⁸. The success of the hydrolysis was validated via thin layer chromatography and 13C-NMR.

The compounds were dissolved in DMSO, aliquoted, and stored at -20°C until use. An equal amount of DMSO (0.1%) was always tested under each condition in all experiments to assure that the solvent vehicle does not influence the results itself.

6.7.2. Cell culture reagents

Table 32: Composition of HEK293 growth and stripped medium

HEK293 complete growth medium	
DMEM white with 4.5 g/L glucose, glutamine, Pen/Str, 10% heat-inactivated FBS	
HEK293 stripped medium	
DMEM white with 4.5 g/L glucose, glutamine, Pen/Str, 5% stripped FBS	

Table 33: Composition of 5000ml PBS

PBS	
NaCl	36g
Na ₂ HPO ₄	7.4g
KH ₂ PO ₄	2.4g
ddH ₂ O	Ad 5000ml

Table 34: Composition of 1000ml T/E solution

T/E	
Trypsin	0.5g
EDTA	0.2g
PBS	Ad 1000ml

Table 35: Composition of the lysis buffer (for a single 96 well plate), mixed immediately before usage

Lysis buffer	
ddH ₂ O	4.3ml
Lysis buffer preliminary dilution*	1.2ml
CoA	6µl
DTT	6µl

*Lysis buffer preliminary dilution	
Tris-HCl (ph 7.8)	125mM
DTT	10.1mM
Cyclohexandiamintetraessigsäure	10.1mM
Glycerol	50% (v/v)
Triton X-100	5% (v/v)
CoA	270µM

Table 36: Composition of the (+) transfection mixture for 48 wells (1500µl)

Transfection Reagent (+)	
EGFP	5µg
CRE	5µg
TGR5	5µg
2xHBS	750.00µl
CaCl ₂ 2M	94.00µl
ddH ₂ O	Ad 1500µl

Table 37: Composition of the (-) transfection mixture for 48 wells (1500 µl)

Transfection Reagent (-)	
EGFP	5µg
CRE	5µg
2xHBS	750.00µl
CaCl ₂ 2M	94.00µl
ddH ₂ O	Ad 1500µl

Table 38: Composition of the luciferase buffer

Luciferase buffer	
ddH ₂ O	691.67ml
1M Tricine pH 7.8	16.54ml
10X Luciferin*	5µg
*10X Luciferin (Molecular Probes/Invitrogen: Cat\L2912)	

Table 39: Composition of the ATP assay buffer

ATP assay buffer	
ddH ₂ O	744.3ml
0.1M Tricine pH 7.8	16.50ml
0.5M MgCl ₂	35.561ml
0.1M ATP* (pH=7)	30.599ml
*ATP (Sigma ultra grade, disodium salt, F.M 5519)	

6.7.3. Cell-based luciferase reporter gene assay

HEK-293T cells were used for *in vitro* testing. For more information see the introduction section. All following steps were done under aseptic conditions in a laminar flow cabinet, all liquids were prewarmed to 37 °C in a water bath and cells were incubated at 37 °C and 5 % CO₂.

The HEK-293T cells were grown and maintained in complete growth medium, while the experiments were done in stripped medium. The composition of the solutions is shown in **Table 32**. Washing steps were done with phosphate buffered saline (PBS, **Table 33**). The adherent cells

were passaged by incubating them at room temperature for 2-3 minutes with T/E (**Table 34**). The trypsin activity was then stopped by adding complete growth medium. For measurement the cells have to be lysed, which was done with a lysis buffer explained in **Table 35**.

6×10^6 HEK-293T cells were seeded in 15cm cell culture dishes. Additionally, 4×10^6 cells, meant for an untransfected control, were seeded in 10 cm dishes in parallel. For every experiment two 15 cm dishes were needed. A 10 cm dish for un-transfected control was enough for two experiments. After incubation overnight and observation of the cells under a phase contrast microscope, the transfection mixture, explained in the next paragraph, is added dropwise.

Basically, the transfection mixture was prepared in three steps. Plasmids for CRE-Luc, GPBAR1 and EGFP (see **6.7.3.1. Used plasmids**) were added in a given ratio and brought to a volume of 656 μ l with purified water (ddH₂O). Then 750 μ l 2x HBS were added and the solution was mixed. As a last step 94 μ l 2M CaCl₂ was added and the solution was vortexed directly afterwards. The mixture was incubated at room temperature for 20 min, before adding them to the dishes. **Table 36** gives you an overview on the composition of the mixture.

For the control experiments the same amount of CRE-Luc and EGFP plasmid DNA was used but no GPBAR1 (**Table 37**).

After addition of the transfection mixture cells were incubated for 6 h. Meanwhile compound dilutions were added to the empty wells of a 96-well plate in two concentrations each to reach final concentrations of 5 μ M and 20 μ M. The two extracts were tested in three concentrations each: 10 μ g/ml, 30 μ g/ml and 100 μ g/ml. For every compound and concentration 4 wells were prepared in every single experiment and the mean of them was calculated. Another 4 wells on each plate were used as a positive control with LCA (10 μ M), 4 for a negative control purely with DMSO solution and 4 wells for the untransfected cells. Some experiments were done with CDCA (50 μ M) as a second positive control next to LCA.

The compound dilutions were prepared by adding 1 μ l of a compound stock solution (compounds dissolved in sterile DMSO) with a concentration of 5mM or 20mM to 249 μ l stripped medium. For the extracts 10 mg/ml, 30 mg/ml or 100 mg/ml extract stock solutions had to be diluted, respectively. In each well 50 μ l of these compound dilutions were added to each well. For the negative control 1 μ l of pure DMSO was diluted; for the positive control a 10 mg/ml LCA stock solution.

This way every well was filled up with 50 μ l solution, prior to the reseeding process. For reseeding, the medium with the transfection mixture was discarded, the cells were washed with PBS (**Table 33**) and 2 ml T/E per 15 cm dish and 1 ml T/E per 10 cm dish, respectively, were added. After an incubation-time of 3 min, 8 ml of complete growth medium was added to the 15 cm dishes, and 4 ml to the 10 cm dishes, respectively. The cell suspensions were transferred into falcon tubes. Persisting cells were washed off with another 10 ml complete growth medium per 15 cm dish or 5 ml per 10 cm dish. These cell suspensions were also transferred into the tubes. The falcon tubes were then centrifuged for 4 min at 1400 rpm. The supernatant was removed and the cell pellets were resuspended in 10 ml stripped medium. 1 ml of this cell suspension was then used to count the cells with Beckman Coulter's® VI-CELL XR cell viability analyzer.

50.000 cells per 150 μ l per well were then reseeded. This way the final concentrations of 5 μ M and 20 μ M were achieved. The plate was carefully shaken and incubated for 18 hours. After this incubation step the medium was removed, the plates immediately frozen at -80°C, kept frozen for at least on hour and measurement was done latest after ten days.

For the measurement the plates were thawed and 50 μ l lysis buffer was added into each well. The plates were shaken for 10 minutes on a plate shaker at room temperature. Afterwards, 40 μ l lysed cells from each well were transferred from the 96 well-plate into a black plate, which was necessary due to the luminescence based readout. The black plate was again shaken for 1 minute and measured in a Tecan Infinite® 200 PRO which added ATP solution (**Table 39**) and Luciferin (**Table 38**) and measured fluorescence and the emitted bioluminescence.

Following Settings had been used:

Fluorescence top reading with excitation wavelength of 485 nm, emission wavelength 520 nm, excitation bandwidth 9 nm, emission bandwidth 20 nm, integration time 1000 μ s, lag time 0 μ s, settle time 0 ms and Z-Position of 20 mm. The number of flashes was appointed 25. For Luminescence readout integration time was 2000 ms and settle time 50 ms.

6.7.3.1. Used plasmids

TGR5/GPBAR1: GPBAR1 (untagged)-Human G protein-coupled bile acid receptor 1 (GPBAR1), transcript variant 3, NM_170699.1 – bought from origene (OG-SC123315)

CRE-Luc: pGL4.29[luc2P/CRE/Hygro] Vector – bought from Promega (#E8471)

EGFP: pEGFP-N1 Vector Information PT3027-5 GenBank Accession #U55762 –bought from BDBiosciences (#6085-1)

6.7.4. Statistical analysis

Output was given in an excel sheet with RLU/s and EGFP/RFU. The output was searched for outliers, which were discarded. Experiments with more than one outlier were repeated. This way every score was derived from at least three values. The subsequent statistical analysis was done in GraphPad Prism 4.03 (GraphPad Software Inc., La Jolla, CA, USA). The non-linear regression was used to calculate EC_{50} with the sigmoidal dose response settings.

7. Summary and Conclusion

This work can be seen as an approach to utilize a broad palette of *in silico* methods, in specific ligand-based VS with descriptors as 3D pharmacophore models, into NP research on an exemplified target, the GPBAR1. To achieve this goal several ligand-based pharmacophore models with the intention to cover the whole structural diversity of yet known GPBAR1 agonists were set up in LigandScout4.0.

Each model explained quite specific scaffolds (i.e. steroids, 3-aryl-4-isoxazolecarboxamides) and in chorus they should complement each other. The ligand-based approach was chosen, as it has interestingly been regarded as superior in terms of enrichment to structure-based methods, but also because most trans-membrane proteins like GPCR lack on crystallized experimental structures and the published homology-models did not seem to achieve our needs.

The predictive power of two selected models had been experimentally validated. The bile acid shared-feature model 22 (BAMS22) and the triterpene Model (Model 12.3.8.) were selected, since they both had been very specific with good metrics in the theoretical. Additionally, they embrace scaffolds that are quite spread in nature as common secondary metabolites. Further, in particular these two well performing models consist of a relatively small number of mandatory pharmacophore features. Accordingly, they have the potential to retrieve further scaffolds with the ability to activate the GPBAR1. More than half a million compounds were considered. Obtained virtual hits were checked for their availability and ranked accordingly to their tanimoto combo scores, with three known benchmark-molecules as queries. A PCA of all virtual hits with a consequent hierarchical cluster analysis guaranteed structural diversity of the selected hit list.

As a result, 33 compounds were selected for experimental validation, which was performed at two concentrations (5 μ M and 20 μ M) in a reporter-gene based cellular assay. The 5 μ M concentrations had been chosen, since previous works from Dr. Angela Ladurner showed that several triterpenes develop severe cytotoxic properties in the established HEK293T-cell based assay starting at 6 μ M for corsolic acid and 10 μ M for maslinic acid. The 20 μ M concentration seemed to be a sufficient compromise between on the hand potential cytotoxicity and on the other hand being sensitive enough for less potent agonists. From the 33 compounds, only two were not able to get evaluated, as they showed cytotoxic properties at both concentrations. From the remaining 31 compounds 39% showed significant activity. Three (of four) additionally

tested triterpenes, resembling known ligands, also showed potent activity even at 5 μ M. A positive aspect was the discovery of sesquiterpene-coumarins (i.e. microlobiden) as potent activators of GPBAR1, which proved the scaffold-hopping competence of BAMS22 and depicts NP's ability to target different proteins, although they are made up from a very limited number of building blocks.

In conclusion, it can be stated that the pharmacophore-based VS was a successful method to complement the NP research on GPBAR1. The discovered activity of the virtual-hits and additionally tested triterpenes expands the understanding of their structure-activity-relationship. Both experimentally validated models proved to be viable tools for further screenings and underscore the potential benefits of other model's experimental validation. Regarding the intriguing activity of microlobiden, further evaluation of its activity on other bile-acid receptors (i.e. FXR) and its potential activity in obese animal models is recommended.

8. References

- (1) Xu, Y. Recent Progress on Bile Acid Receptor Modulators for Treatment of Metabolic Diseases. *J Med Chem* **2016**.
- (2) Duboc, H.; Tache, Y.; Hofmann, A. F. The bile acid TGR5 membrane receptor: from basic research to clinical application. *Dig Liver Dis* **2014**, *46*, 302-312.
- (3) Hodge, R. J.; Nunez, D. J. Therapeutic potential of Takeda-G-protein-receptor-5 (TGR5) agonists. Hope or hype? *Diabetes Obes Metab* **2016**, *18*, 439-443.
- (4) Wang, H.; Chen, J.; Hollister, K.; Sowers, L. C.; Forman, B. M. Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Molecular cell* **1999**, *3*, 543-553.
- (5) Maruyama, T.; Miyamoto, Y.; Nakamura, T.; Tamai, Y.; Okada, H.; Sugiyama, E.; Nakamura, T.; Itadani, H.; Tanaka, K. Identification of membrane-type receptor for bile acids (M-BAR). *Biochemical and biophysical research communications* **2002**, *298*, 714-719.
- (6) Kawamata, Y.; Fujii, R.; Hosoya, M.; Harada, M.; Yoshida, H.; Miwa, M.; Fukusumi, S.; Habata, Y.; Itoh, T.; Shintani, Y.; Hinuma, S.; Fujisawa, Y.; Fujino, M. A G protein-coupled receptor responsive to bile acids. *The Journal of biological chemistry* **2003**, *278*, 9435-9440.
- (7) Gene Symbol Report | HUGO Gene Nomenclature Committee. http://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=19680 (accessed 23.5.2016).
- (8) Hov, J. R.; Keitel, V.; Laerdahl, J. K.; Spomer, L.; Ellinghaus, E.; ElSharawy, A.; Melum, E.; Boberg, K. M.; Manke, T.; Balschun, T.; Schramm, C.; Bergquist, A.; Weismuller, T.; Gotthardt, D.; Rust, C.; Henckaerts, L.; Onnie, C. M.; Weersma, R. K.; Sterneck, M.; Teufel, A.; Runz, H.; Stiehl, A.; Ponsioen, C. Y.; Wijmenga, C.; Vatn, M. H.; Stokkers, P. C.; Vermeire, S.; Mathew, C. G.; Lie, B. A.; Beuers, U.; Manns, M. P.; Schreiber, S.; Schruppf, E.; Haussinger, D.; Franke, A.; Karlsen, T. H. Mutational characterization of the bile acid receptor TGR5 in primary sclerosing cholangitis. *PloS one* **2010**, *5*, e12403.
- (9) Gertzen, C. G.; Spomer, L.; Smits, S. H.; Haussinger, D.; Keitel, V.; Gohlke, H. Mutational mapping of the transmembrane binding site of the G-protein coupled receptor TGR5 and binding mode prediction of TGR5 agonists. *European journal of medicinal chemistry* **2015**, *104*, 57-72.
- (10) Macchiarulo, A.; Gioiello, A.; Thomas, C.; Pols, T. W.; Nuti, R.; Ferrari, C.; Giacche, N.; De Franco, F.; Pruzanski, M.; Auwerx, J.; Schoonjans, K.; Pellicciari, R. Probing the Binding Site of Bile Acids in TGR5. *ACS medicinal chemistry letters* **2013**, *4*, 1158-1162.

- (11) Sindhu, T.; Srinivasan, P. Exploring the binding properties of agonists interacting with human TGR5 using structural modeling, molecular docking and dynamics simulations. *RSC Adv.* **2015**, *5*, 14202-14213.
- (12) Vassileva, G.; Golovko, A.; Markowitz, L.; Abbondanzo, S. J.; Zeng, M.; Yang, S.; Hoos, L.; Tetzloff, G.; Levitan, D.; Murgolo, N. J.; Keane, K.; Davis, H. R., Jr.; Hedrick, J.; Gustafson, E. L. Targeted deletion of Gpbar1 protects mice from cholesterol gallstone formation. *Biochem J* **2006**, *398*, 423-430.
- (13) Casaburi, I.; Avena, P.; Lanzino, M.; Sisci, D.; Giordano, F.; Maris, P.; Catalano, S.; Morelli, C.; Ando, S. Chenodeoxycholic acid through a TGR5-dependent CREB signaling activation enhances cyclin D1 expression and promotes human endometrial cancer cell proliferation. *Cell cycle (Georgetown, Tex.)* **2012**, *11*, 2699-2710.
- (14) Kumar, D. P.; Asgharpour, A.; Mirshahi, F.; Park, S. H.; Liu, S.; Imai, Y.; Nadler, J. L.; Grider, J. R.; Murthy, K. S.; Sanyal, A. J. Activation of Transmembrane Bile Acid Receptor TGR5 Modulates Pancreatic Islet alpha Cells to Promote Glucose Homeostasis. *The Journal of biological chemistry* **2016**, *291*, 6626-6640.
- (15) Stepanov, V.; Stankov, K.; Mikov, M. The bile acid membrane receptor TGR5: a novel pharmacological target in metabolic, inflammatory and neoplastic disorders. *Journal of receptor and signal transduction research* **2013**, *33*, 213-223.
- (16) Vitek, L.; Haluzik, M. The role of bile acids in metabolic regulation. *The Journal of endocrinology* **2016**, *228*, R85-96.
- (17) Macchiarulo, A.; Gioiello, A.; Thomas, C.; Massarotti, A.; Nuti, R.; Rosatelli, E.; Sabbatini, P.; Schoonjans, K.; Auwerx, J.; Pellicciari, R. Molecular field analysis and 3D-quantitative structure-activity relationship study (MFA 3D-QSAR) unveil novel features of bile acid recognition at TGR5. *Journal of chemical information and modeling* **2008**, *48*, 1792-1801.
- (18) Kumar, D. P.; Rajagopal, S.; Mahavadi, S.; Mirshahi, F.; Grider, J. R.; Murthy, K. S.; Sanyal, A. J. Activation of transmembrane bile acid receptor TGR5 stimulates insulin secretion in pancreatic beta cells. *Biochemical and biophysical research communications* **2012**, *427*, 600-605.
- (19) Bala, V.; Rajagopal, S.; Kumar, D. P.; Nalli, A. D.; Mahavadi, S.; Sanyal, A. J.; Grider, J. R.; Murthy, K. S. Release of GLP-1 and PYY in response to the activation of G protein-coupled bile acid receptor TGR5 is mediated by Epac/PLC-epsilon pathway and modulated by endogenous H₂S. *Frontiers in physiology* **2014**, *5*, 420.
- (20) Woods, S. C.; D'Alessio, D. A. Central control of body weight and appetite. *J Clin Endocrinol Metab* **2008**, *93*, S37-50.
- (21) Watanabe, M.; Houten, S. M.; Matak, C.; Christoffolete, M. A.; Kim, B. W.; Sato, H.; Messaddeq, N.; Harney, J. W.; Ezaki, O.; Kodama, T.; Schoonjans, K.; Bianco, A. C.;

Auwerx, J. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **2006**, 439, 484-489.

(22) Keitel, V.; Reinehr, R.; Gatsios, P.; Rupprecht, C.; Gorg, B.; Selbach, O.; Haussinger, D.; Kubitz, R. The G-protein coupled bile salt receptor TGR5 is expressed in liver sinusoidal endothelial cells. *Hepatology (Baltimore, Md.)* **2007**, 45, 695-704.

(23) Keitel, V.; Donner, M.; Winandy, S.; Kubitz, R.; Haussinger, D. Expression and function of the bile acid receptor TGR5 in Kupffer cells. *Biochemical and biophysical research communications* **2008**, 372, 78-84.

(24) Keitel, V.; Haussinger, D. Perspective: TGR5 (Gpbar-1) in liver physiology and disease. *Clinics and research in hepatology and gastroenterology* **2012**, 36, 412-419.

(25) Knolle, P. A.; Limmer, A. Control of immune responses by scavenger liver endothelial cells. *Swiss medical weekly* **2003**, 133, 501-506.

(26) Wisse, E.; Braet, F.; Luo, D.; De Zanger, R.; Jans, D.; Crabbe, E.; Vermoesen, A. Structure and function of sinusoidal lining cells in the liver. *Toxicologic pathology* **1996**, 24, 100-111.

(27) Reinehr, R.; Haussinger, D. Inhibition of bile salt-induced apoptosis by cyclic AMP involves serine/threonine phosphorylation of CD95. *Gastroenterology* **2004**, 126, 249-262.

(28) Kida, T.; Tsubosaka, Y.; Hori, M.; Ozaki, H.; Murata, T. Bile acid receptor TGR5 agonism induces NO production and reduces monocyte adhesion in vascular endothelial cells. *Arteriosclerosis, thrombosis, and vascular biology* **2013**, 33, 1663-1669.

(29) Reich, M.; Deutschmann, K.; Sommerfeld, A.; Klindt, C.; Kluge, S.; Kubitz, R.; Ullmer, C.; Knoefel, W. T.; Herebian, D.; Mayatepek, E.; Haussinger, D.; Keitel, V. TGR5 is essential for bile acid-dependent cholangiocyte proliferation in vivo and in vitro. *Gut* **2016**, 65, 487-501.

(30) Keitel, V.; Haussinger, D. TGR5 in cholangiocytes. *Current opinion in gastroenterology* **2013**, 29, 299-304.

(31) Keitel, V.; Haussinger, D. TGR5 in the biliary tree. *Digestive diseases (Basel, Switzerland)* **2011**, 29, 45-47.

(32) Calmus, Y.; Guechot, J.; Podevin, P.; Bonnefis, M. T.; Giboudeau, J.; Poupon, R. Differential effects of chenodeoxycholic and ursodeoxycholic acids on interleukin 1, interleukin 6 and tumor necrosis factor-alpha production by monocytes. *Hepatology (Baltimore, Md.)* **1992**, 16, 719-723.

(33) Pols, T. W.; Nomura, M.; Harach, T.; Lo Sasso, G.; Oosterveer, M. H.; Thomas, C.; Rizzo, G.; Gioiello, A.; Adorini, L.; Pellicciari, R.; Auwerx, J.; Schoonjans, K. TGR5 activation inhibits atherosclerosis by reducing macrophage inflammation and lipid loading. *Cell metabolism* **2011**, 14, 747-757.

(34) Ichikawa, R.; Takayama, T.; Yoneno, K.; Kamada, N.; Kitazume, M. T.; Higuchi, H.; Matsuoka, K.; Watanabe, M.; Itoh, H.; Kanai, T.; Hisamatsu, T.; Hibi, T. Bile acids induce monocyte differentiation toward interleukin-12 hypo-producing dendritic cells via a TGR5-dependent pathway. *Immunology* **2012**, *136*, 153-162.

(35) Pols, T.; Eggink, H.; Soeters, M. TGR5 ligands as potential therapeutics in inflammatory diseases. *International Journal of Interferon, Cytokine and Mediator Research* **2014**, *27*.

(36) Alemi, F.; Poole, D. P.; Chiu, J.; Schoonjans, K.; Cattaruzza, F.; Grider, J. R.; Bunnett, N. W.; Corvera, C. U. The receptor TGR5 mediates the prokinetic actions of intestinal bile acids and is required for normal defecation in mice. *Gastroenterology* **2013**, *144*, 145-154.

(37) Yasuda, H.; Hirata, S.; Inoue, K.; Mashima, H.; Ohnishi, H.; Yoshida, M. Involvement of membrane-type bile acid receptor M-BAR/TGR5 in bile acid-induced activation of epidermal growth factor receptor and mitogen-activated protein kinases in gastric carcinoma cells. *Biochemical and biophysical research communications* **2007**, *354*, 154-159.

(38) Chen, W. D.; Yu, D.; Forman, B. M.; Huang, W.; Wang, Y. D. Deficiency of G-protein-coupled bile acid receptor Gpbar1 (TGR5) enhances chemically induced liver carcinogenesis. *Hepatology (Baltimore, Md.)* **2013**, *57*, 656-666.

(39) Keitel, V.; Gorg, B.; Bidmon, H. J.; Zemtsova, I.; Spomer, L.; Zilles, K.; Haussinger, D. The bile acid receptor TGR5 (Gpbar-1) acts as a neurosteroid receptor in brain. *Glia* **2010**, *58*, 1794-1805.

(40) Sato, H.; Macchiarulo, A.; Thomas, C.; Gioiello, A.; Une, M.; Hofmann, A. F.; Saladin, R.; Schoonjans, K.; Pellicciari, R.; Auwerx, J. Novel potent and selective bile acid derivatives as TGR5 agonists: biological screening, structure-activity relationships, and molecular modeling studies. *J Med Chem* **2008**, *51*, 1831-1841.

(41) Festa, C.; Renga, B.; D'Amore, C.; Sepe, V.; Finamore, C.; De Marino, S.; Carino, A.; Cipriani, S.; Monti, M. C.; Zampella, A.; Fiorucci, S. Exploitation of cholane scaffold for the discovery of potent and selective farnesoid X receptor (FXR) and G-protein coupled bile acid receptor 1 (GP-BAR1) ligands. *J Med Chem* **2014**, *57*, 8477-8495.

(42) Fiorucci, S.; Rizzo, G.; Donini, A.; Distrutti, E.; Santucci, L. Targeting farnesoid X receptor for liver and metabolic disorders. *Trends in Molecular Medicine* **2007**, *13*, 298-309.

(43) Joyce, S. A.; Gahan, C. G. Bile Acid Modifications at the Microbe-Host Interface: Potential for Nutraceutical and Pharmaceutical Interventions in Host Health. *Annual review of food science and technology* **2016**, *7*, 313-333.

(44) NCI Dictionary of Cancer Terms. **2016**.
<http://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=643126> (accessed 01.06.2016).

- (45) Stanimirov, B.; Stankov, K.; Mikov, M. Bile acid signaling through farnesoid X and TGR5 receptors in hepatobiliary and intestinal diseases. *Hepatobiliary & pancreatic diseases international : HBPD INT* **2015**, *14*, 18-33.
- (46) Adorini, L.; Pruzanski, M.; Shapiro, D. Farnesoid X receptor targeting to treat nonalcoholic steatohepatitis. *Drug discovery today* **2012**, *17*, 988-997.
- (47) Xiao, J.; Guo, R.; Fung, M. L.; Liong, E. C.; Tipoe, G. L. Therapeutic approaches to non-alcoholic fatty liver disease: past achievements and future challenges. *Hepatobiliary & pancreatic diseases international : HBPD INT* **2013**, *12*, 125-135.
- (48) Thomas, C.; Gioiello, A.; Noriega, L.; Strehle, A.; Oury, J.; Rizzo, G.; Macchiarulo, A.; Yamamoto, H.; Matak, C.; Pruzanski, M.; Pellicciari, R.; Auwerx, J.; Schoonjans, K. TGR5-mediated bile acid sensing controls glucose homeostasis. *Cell metabolism* **2009**, *10*, 167-177.
- (49) Cipriani, S.; Mencarelli, A.; Palladino, G.; Fiorucci, S. FXR activation reverses insulin resistance and lipid abnormalities and protects against liver steatosis in Zucker (fa/fa) obese rats. *Journal of lipid research* **2010**, *51*, 771-784.
- (50) Pean, N.; Doignon, I.; Garcin, I.; Besnard, A.; Julien, B.; Liu, B.; Branchereau, S.; Spraul, A.; Guettier, C.; Humbert, L.; Schoonjans, K.; Rainteau, D.; Tordjmann, T. The receptor TGR5 protects the liver from bile acid overload during liver regeneration in mice. *Hepatology (Baltimore, Md.)* **2013**, *58*, 1451-1460.
- (51) Keitel, V.; Reich, M.; Haussinger, D. TGR5: pathogenetic role and/or therapeutic target in fibrosing cholangitis? *Clinical reviews in allergy & immunology* **2015**, *48*, 218-225.
- (52) Cipriani, S.; Mencarelli, A.; Chini, M. G.; Distrutti, E.; Renga, B.; Bifulco, G.; Baldelli, F.; Donini, A.; Fiorucci, S. The bile acid receptor GPBAR-1 (TGR5) modulates integrity of intestinal barrier and immune response to experimental colitis. *PloS one* **2011**, *6*, e25637.
- (53) Yoneno, K.; Hisamatsu, T.; Shimamura, K.; Kamada, N.; Ichikawa, R.; Kitazume, M. T.; Mori, M.; Uo, M.; Namikawa, Y.; Matsuoka, K.; Sato, T.; Koganei, K.; Sugita, A.; Kanai, T.; Hibi, T. TGR5 signalling inhibits the production of pro-inflammatory cytokines by in vitro differentiated inflammatory and intestinal macrophages in Crohn's disease. *Immunology* **2013**, *139*, 19-29.
- (54) Guo, C.; Qi, H.; Yu, Y.; Zhang, Q.; Su, J.; Yu, D.; Huang, W.; Chen, W. D.; Wang, Y. D. The G-Protein-Coupled Bile Acid Receptor Gpbar1 (TGR5) Inhibits Gastric Inflammation Through Antagonizing NF-kappaB Signaling Pathway. *Frontiers in pharmacology* **2015**, *6*, 287.
- (55) He, G.; Yu, G. Y.; Temkin, V.; Ogata, H.; Kuntzen, C.; Sakurai, T.; Sieghart, W.; Peck-Radosavljevic, M.; Leffert, H. L.; Karin, M. Hepatocyte IKKbeta/NF-kappaB inhibits tumor promotion and progression by preventing oxidative stress-driven STAT3 activation. *Cancer cell* **2010**, *17*, 286-297.

- (56) Hong, J.; Behar, J.; Wands, J.; Resnick, M.; Wang, L. J.; DeLellis, R. A.; Lambeth, D.; Souza, R. F.; Spechler, S. J.; Cao, W. Role of a novel bile acid receptor TGR5 in the development of oesophageal adenocarcinoma. *Gut* **2010**, *59*, 170-180.
- (57) Li, T.; Holmstrom, S. R.; Kir, S.; Umetani, M.; Schmidt, D. R.; Kliewer, S. A.; Mangelsdorf, D. J. The G protein-coupled bile acid receptor, TGR5, stimulates gallbladder filling. *Molecular endocrinology (Baltimore, Md.)* **2011**, *25*, 1066-1071.
- (58) Greve, J. W.; Gouma, D. J.; Buurman, W. A. Bile acids inhibit endotoxin-induced release of tumor necrosis factor by monocytes: an in vitro study. *Hepatology (Baltimore, Md.)* **1989**, *10*, 454-458.
- (59) Lacaine, F.; Houry, S.; Huguier, M. Stomach cancer after partial gastrectomy for benign ulcer disease. A critical analysis of epidemiological reports. *Hepato-gastroenterology* **1992**, *39*, 4-8.
- (60) Sato, H.; Genet, C.; Strehle, A.; Thomas, C.; Lobstein, A.; Wagner, A.; Mioskowski, C.; Auwerx, J.; Saladin, R. Anti-hyperglycemic activity of a TGR5 agonist isolated from *Olea europaea*. *Biochemical and biophysical research communications* **2007**, *362*, 793-798.
- (61) Genet, C.; Strehle, A.; Schmidt, C.; Boudjelal, G.; Lobstein, A.; Schoonjans, K.; Souchet, M.; Auwerx, J.; Saladin, R.; Wagner, A. Structure-activity relationship study of betulinic acid, a novel and selective TGR5 agonist, and its synthetic derivatives: potential impact in diabetes. *J Med Chem* **2010**, *53*, 178-190.
- (62) Duan, H.; Ning, M.; Chen, X.; Zou, Q.; Zhang, L.; Feng, Y.; Zhang, L.; Leng, Y.; Shen, J. Design, synthesis, and antidiabetic activity of 4-phenoxy nicotinamide and 4-phenoxy pyrimidine-5-carboxamide derivatives as potent and orally efficacious TGR5 agonists. *J Med Chem* **2012**, *55*, 10475-10489.
- (63) Zhu, J.; Ye, Y.; Ning, M.; Mandi, A.; Feng, Y.; Zou, Q.; Kurtan, T.; Leng, Y.; Shen, J. Design, synthesis, and structure-activity relationships of 3,4,5-trisubstituted 4,5-dihydro-1,2,4-oxadiazoles as TGR5 agonists. *ChemMedChem* **2013**, *8*, 1210-1223.
- (64) Duan, H.; Ning, M.; Zou, Q.; Ye, Y.; Feng, Y.; Zhang, L.; Leng, Y.; Shen, J. Discovery of Intestinal Targeted TGR5 Agonists for the Treatment of Type 2 Diabetes. *J Med Chem* **2015**, *58*, 3315-3328.
- (65) Evans, K. A.; Budzik, B. W.; Ross, S. A.; Wisnoski, D. D.; Jin, J.; Rivero, R. A.; Vimal, M.; Szewczyk, G. R.; Jayawickreme, C.; Moncol, D. L.; Rimele, T. J.; Armour, S. L.; Weaver, S. P.; Griffin, R. J.; Tadepalli, S. M.; Jeune, M. R.; Shearer, T. W.; Chen, Z. B.; Chen, L.; Anderson, D. L.; Becherer, J. D.; De Los Frailes, M.; Colilla, F. J. Discovery of 3-aryl-4-isoxazolecarboxamides as TGR5 receptor agonists. *J Med Chem* **2009**, *52*, 7962-7965.

- (66) Budzik, B. W.; Evans, K. A.; Wisnoski, D. D.; Jin, J.; Rivero, R. A.; Szewczyk, G. R.; Jayawickreme, C.; Moncol, D. L.; Yu, H. Synthesis and structure-activity relationships of a series of 3-aryl-4-isoxazolecarboxamides as a new class of TGR5 agonists. *Bioorg Med Chem Lett* **2010**, *20*, 1363-1367.
- (67) Dehmlow, H.; Alvarez Sanchez, R.; Bachmann, S.; Bissantz, C.; Bliss, F.; Conde-Knape, K.; Graf, M.; Martin, R. E.; Obst Sander, U.; Raab, S.; Richter, H. G.; Sewing, S.; Sprecher, U.; Ullmer, C.; Mattei, P. Discovery and optimisation of 1-hydroxyimino-3,3-diphenylpropanes, a new class of orally active GPBAR1 (TGR5) agonists. *Bioorg Med Chem Lett* **2013**, *23*, 4627-4632.
- (68) Gioiello, A.; Macchiarulo, A.; Carotti, A.; Filipponi, P.; Costantino, G.; Rizzo, G.; Adorini, L.; Pellicciari, R. Extending SAR of bile acids as FXR ligands: discovery of 23-N-(carbocinnamyloxy)-3 α ,7 α -dihydroxy-6 α -ethyl-24-nor-5 β -cholan-23-amine. *Bioorg Med Chem* **2011**, *19*, 2650-2658.
- (69) Kim, K.; Park, M.; Lee, Y. M.; Rhyu, M. R.; Kim, H. Y. Ginsenoside metabolite compound K stimulates glucagon-like peptide-1 secretion in NCI-H716 cells via bile acid receptor activation. *Archives of pharmacal research* **2014**, *37*, 1193-1200.
- (70) Herbert, M. R.; Siegel, D. L.; Staszewski, L.; Cayan, C.; Banerjee, U.; Dhamija, S.; Anderson, J.; Fan, A.; Wang, L.; Rix, P.; Shiau, A. K.; Rao, T. S.; Noble, S. A.; Heyman, R. A.; Bischoff, E.; Guha, M.; Kabakibi, A.; Pinkerton, A. B. Synthesis and SAR of 2-aryl-3-aminomethylquinolines as agonists of the bile acid receptor TGR5. *Bioorg Med Chem Lett* **2010**, *20*, 5718-5721.
- (71) Zhu, J.; Ning, M.; Guo, C.; Zhang, L.; Pan, G.; Leng, Y.; Shen, J. Design, synthesis and biological evaluation of a novel class of potent TGR5 agonists based on a 4-phenyl pyridine scaffold. *European journal of medicinal chemistry* **2013**, *69*, 55-68.
- (72) D'Amore, C.; Di Leva, F. S.; Sepe, V.; Renga, B.; Del Gaudio, C.; D'Auria, M. V.; Zampella, A.; Fiorucci, S.; Limongelli, V. Design, synthesis, and biological evaluation of potent dual agonists of nuclear and membrane bile acid receptors. *J Med Chem* **2014**, *57*, 937-954.
- (73) Park, E. J.; Ahn, Y. G.; Jung, S. H.; Bang, H. J.; Kim, M.; Hong, D. J.; Kim, J.; Suh, K. H.; Kim, Y. J.; Kim, D.; Kim, E. Y.; Lee, K.; Min, K. H. Discovery of novel pyrimidine and malonamide derivatives as TGR5 agonists. *Bioorg Med Chem Lett* **2014**, *24*, 4271-4275.
- (74) Yu, D. D.; Sousa, K. M.; Mattern, D. L.; Wagner, J.; Fu, X.; Vaidehi, N.; Forman, B. M.; Huang, W. Stereoselective synthesis, biological evaluation, and modeling of novel bile acid-derived G-protein coupled bile acid receptor 1 (GP-BAR1, TGR5) agonists. *Bioorg Med Chem* **2015**, *23*, 1613-1628.

(75) Chen, Z.; Ning, M.; Zou, Q.; Cao, H.; Ye, Y.; Leng, Y.; Shen, J. Discovery and Structure-Activity Relationship Study of 4-Phenoxythiazol-5-carboxamides as Highly Potent TGR5 Agonists. *Chem Pharm Bull (Tokyo)* **2016**, *64*, 326-339.

(76) Iguchi, Y.; Nishimaki-Mogami, T.; Yamaguchi, M.; Teraoka, F.; Kaneko, T.; Une, M. Effects of chemical modification of ursodeoxycholic acid on TGR5 activation. *Biol Pharm Bull* **2011**, *34*, 1-7.

(77) Lin, Y. C.; Boone, M.; Meuris, L.; Lemmens, I.; Van Roy, N.; Soete, A.; Reumers, J.; Moisse, M.; Plaisance, S.; Drmanac, R.; Chen, J.; Speleman, F.; Lambrechts, D.; Van de Peer, Y.; Tavernier, J.; Callewaert, N. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature communications* **2014**, *5*, 4767.

(78) Thomas, P.; Smart, T. G. HEK293 cell line: a vehicle for the expression of recombinant proteins. *Journal of pharmacological and toxicological methods* **2005**, *51*, 187-200.

(79) Giger, E. V.; Puigmarti-Luis, J.; Schlatter, R.; Castagner, B.; Dittrich, P. S.; Leroux, J. C. Gene delivery with bisphosphonate-stabilized calcium phosphate nanoparticles. *Journal of controlled release : official journal of the Controlled Release Society* **2011**, *150*, 87-93.

(80) Futatsugi, K.; Bahnck, K. B.; Brenner, M. B.; Buxton, J.; Chin, J. E.; Coffey, S. B.; Dubins, J.; Flynn, D.; Gautreau, D.; Guzman-Perez, A.; Hadcock, J. R.; Hepworth, D.; Herr, M.; Hinchey, T.; Janssen, A. M.; Jennings, S. M.; Jiao, W.; Lavergne, S. Y.; Li, B.; Li, M.; Munchhof, M. J.; Orr, S. T. M.; Piotrowski, D. W.; Roush, N. S.; Sammons, M.; Stevens, B. D.; Storer, G.; Wang, J.; Warmus, J. S.; Wei, L.; Wolford, A. C. Optimization of triazole-based TGR5 agonists towards orally available agents. *MedChemComm* **2013**, *4*, 205-210.

(81) Londregan, A. T.; Piotrowski, D. W.; Futatsugi, K.; Warmus, J. S.; Boehm, M.; Carpino, P. A.; Chin, J. E.; Janssen, A. M.; Roush, N. S.; Buxton, J.; Hinchey, T. Discovery of 5-phenoxy-1,3-dimethyl-1H-pyrazole-4-carboxamides as potent agonists of TGR5 via sequential combinatorial libraries. *Bioorg Med Chem Lett* **2013**, *23*, 1407-1411.

(82) Phillips, D. P.; Gao, W.; Yang, Y.; Zhang, G.; Lerario, I. K.; Lau, T. L.; Jiang, J.; Wang, X.; Nguyen, D. G.; Bhat, B. G.; Trotter, C.; Sullivan, H.; Welzel, G.; Landry, J.; Chen, Y.; Joseph, S. B.; Li, C.; Gordon, W. P.; Richmond, W.; Johnson, K.; Bretz, A.; Bursulaya, B.; Pan, S.; McNamara, P.; Seidel, H. M. Discovery of trifluoromethyl(pyrimidin-2-yl)azetidine-2-carboxamides as potent, orally bioavailable TGR5 (GPBAR1) agonists: structure-activity relationships, lead optimization, and chronic in vivo efficacy. *J Med Chem* **2014**, *57*, 3263-3282.

(83) Zheng, C.; Zhou, W.; Wang, T.; You, P.; Zhao, Y.; Yang, Y.; Wang, X.; Luo, J.; Chen, Y.; Liu, M.; Chen, H. A Novel TGR5 Activator WB403 Promotes GLP-1 Secretion and Preserves Pancreatic beta-Cells in Type 2 Diabetic Mice. *PloS one* **2015**, *10*, e0134051.

- (84) Frye, S. V.; Arkin, M. R.; Arrowsmith, C. H.; Conn, P. J.; Glicksman, M. A.; Hull-Ryde, E. A.; Slusher, B. S. Tackling reproducibility in academic preclinical drug discovery. *Nat Rev Drug Discov* **2015**, *14*, 733-734.
- (85) Pellicciari, R.; Gioiello, A.; Macchiarulo, A.; Thomas, C.; Rosatelli, E.; Natalini, B.; Sardella, R.; Pruzanski, M.; Roda, A.; Pastorini, E.; Schoonjans, K.; Auwerx, J. Discovery of 6 α -ethyl-23(S)-methylcholic acid (S-EMCA, INT-777) as a potent and selective agonist for the TGR5 receptor, a novel target for diabetes. *J Med Chem* **2009**, *52*, 7958-7961.
- (86) Roda, A.; Pellicciari, R.; Gioiello, A.; Neri, F.; Camborata, C.; Passeri, D.; De Franco, F.; Spinozzi, S.; Colliva, C.; Adorini, L.; Montagnani, M.; Aldini, R. Semisynthetic bile acid FXR and TGR5 agonists: physicochemical properties, pharmacokinetics, and metabolism in the rat. *J Pharmacol Exp Ther* **2014**, *350*, 56-68.
- (87) Hodge, R. J.; Lin, J.; Vasist Johnson, L. S.; Gould, E. P.; Bowers, G. D.; Nunez, D. J. Safety, Pharmacokinetics, and Pharmacodynamic Effects of a Selective TGR5 Agonist, SB-756050, in Type 2 Diabetes. *Clinical pharmacology in drug development* **2013**, *2*, 213-222.
- (88) Dehmlow, H.; Erickson, S. D.; Mattei, P.; Richter, H.: 1,2- pyridazine, 1,6-pyridazine or pyrimidine - benzamide derivatives as gpbar1 modulators. Google Patents, 2013.
- (89) Flatt, B. T.; Mohan, R.: Tgr5 agonists: imidazole and triazole compounds containing a quaternary nitrogen. Google Patents, 2014.
- (90) Piotrowski, D. W.; Futatsugi, K.; Warmus, J. S.; Orr, S. T.; Freeman-Cook, K. D.; Londregan, A. T.; Wei, L.; Jennings, S. M.; Herr, M.; Coffey, S. B.; Jiao, W.; Storer, G.; Hepworth, D.; Wang, J.; Laverne, S. Y.; Chin, J. E.; Hadcock, J. R.; Brenner, M. B.; Wolford, A. C.; Janssen, A. M.; Roush, N. S.; Buxton, J.; Hinchey, T.; Kalgutkar, A. S.; Sharma, R.; Flynn, D. A. Identification of Tetrahydropyrido[4,3-d]pyrimidine Amides as a New Class of Orally Bioavailable TGR5 Agonists. *ACS medicinal chemistry letters* **2013**, *4*, 63-68.
- (91) 芳香氧基嘧啶甲酰胺或芳香氧基吡啶甲酰胺类化合物及其制备方法、药物组合物和用途. Google Patents, 2013.
- (92) Cao, H.; Chen, Z. X.; Wang, K.; Ning, M. M.; Zou, Q. A.; Feng, Y.; Ye, Y. L.; Leng, Y.; Shen, J. H. Intestinally-targeted TGR5 agonists equipped with quaternary ammonium have an improved hypoglycemic effect and reduced gallbladder filling effect. *Scientific reports* **2016**, *6*, 28676.
- (93) Martin, R. E.; Bissantz, C.; Gavelle, O.; Kuratli, C.; Dehmlow, H.; Richter, H. G.; Obst Sander, U.; Erickson, S. D.; Kim, K.; Pietranico-Cole, S. L.; Alvarez-Sanchez, R.; Ullmer, C. 2-Phenoxy-nicotinamides are potent agonists at the bile acid receptor GPBAR1 (TGR5). *ChemMedChem* **2013**, *8*, 569-576.
- (94) Yu, B. Z.; Kaimal, R.; Bai, S.; El Sayed, K. A.; Tatulian, S. A.; Apitz, R. J.; Jain, M. K.; Deng, R.; Berg, O. G. Effect of guggulsterone and cembranoids of *Commiphora mukul* on pancreatic phospholipase A(2): role in hypocholesterolemia. *J Nat Prod* **2009**, *72*, 24-28.

- (95) Izukura, M.; Hashimoto, T.; Gomez, G.; Uchida, T.; Greeley, G. H., Jr.; Thompson, J. C. Intracolonic infusion of bile salt stimulates release of peptide YY and inhibits cholecystokinin-stimulated pancreatic exocrine secretion in conscious dogs. *Pancreas* **1991**, *6*, 427-432.
- (96) Brighton, C. A.; Rievaj, J.; Kuhre, R. E.; Glass, L. L.; Schoonjans, K.; Holst, J. J.; Gribble, F. M.; Reimann, F. Bile Acids Trigger GLP-1 Release Predominantly by Accessing Basolaterally Located G Protein–Coupled Bile Acid Receptors. *Endocrinology* **2015**, *156*, 3961-3970.
- (97) Wu, T.; Bound, M. J.; Standfield, S. D.; Gedulin, B.; Jones, K. L.; Horowitz, M.; Rayner, C. K. Effects of rectal administration of taurocholic acid on glucagon-like peptide-1 and peptide YY secretion in healthy humans. *Diabetes Obes Metab* **2013**, *15*, 474-477.
- (98) Briere, D. A.; Ruan, X.; Cheng, C. C.; Siesky, A. M.; Fitch, T. E.; Dominguez, C.; Sanfeliciano, S. G.; Montero, C.; Suen, C. S.; Xu, Y.; Coskun, T.; Michael, M. D. Novel Small Molecule Agonist of TGR5 Possesses Anti-Diabetic Effects but Causes Gallbladder Filling in Mice. *PloS one* **2015**, *10*.
- (99) Szewczyk, J. R.; Laudeman, C. P.; Evans, K. A.; Li, Y. H.; Dock, S. T.; Chen, Z.: Use of sulfonamide derivatives in the treatment of disorders of the metabolism and the nervous system. Google Patents, 2007.
- (100) Gioiello, A.; Rosatelli, E.; Nuti, R.; Macchiarulo, A.; Pellicciari, R. Patented TGR5 modulators: a review (2006 – present). *Expert Opinion on Therapeutic Patents* **2012**, *22*, 1399-1414.
- (101) Mudaliar, S.; Henry, R. R.; Sanyal, A. J.; Morrow, L.; Marschall, H. U.; Kipnes, M.; Adorini, L.; Sciacca, C. I.; Clopton, P.; Castelloe, E.; Dillon, P.; Pruzanski, M.; Shapiro, D. Efficacy and safety of the farnesoid X receptor agonist obeticholic acid in patients with type 2 diabetes and nonalcoholic fatty liver disease. *Gastroenterology* **2013**, *145*, 574-582 e571.
- (102) Neuschwander-Tetri, B. A.; Loomba, R.; Sanyal, A. J.; Lavine, J. E.; Van Natta, M. L.; Abdelmalek, M. F.; Chalasani, N.; Dasarathy, S.; Diehl, A. M.; Hameed, B.; Kowdley, K. V.; McCullough, A.; Terrault, N.; Clark, J. M.; Tonascia, J.; Brunt, E. M.; Kleiner, D. E.; Doo, E. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. *Lancet (London, England)* **2015**, *385*, 956-965.
- (103) Hirschfield, G. M.; Mason, A.; Luketic, V.; Lindor, K.; Gordon, S. C.; Mayo, M.; Kowdley, K. V.; Vincent, C.; Bodhenheimer, H. C., Jr.; Pares, A.; Trauner, M.; Marschall, H. U.; Adorini, L.; Sciacca, C.; Beecher-Jones, T.; Castelloe, E.; Bohm, O.; Shapiro, D. Efficacy of obeticholic acid in patients with primary biliary cirrhosis and inadequate response to ursodeoxycholic acid. *Gastroenterology* **2015**, *148*, 751-761 e758.

- (104) Alemi, F.; Kwon, E.; Poole, D. P.; Lieu, T.; Lyo, V.; Cattaruzza, F.; Cevikbas, F.; Steinhoff, M.; Nassini, R.; Materazzi, S.; Guerrero-Alba, R.; Valdez-Morales, E.; Cottrell, G. S.; Schoonjans, K.; Geppetti, P.; Vanner, S. J.; Bunnett, N. W.; Corvera, C. U. The TGR5 receptor mediates bile acid-induced itch and analgesia. *The Journal of clinical investigation* **2013**, *123*, 1513-1530.
- (105) European Medicinal Agency - Human Medicines. http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/004093/human_med_002043.jsp&mid=WC0b01ac058001d124 (accessed 7.3.17).
- (106) Ono, E.; Inoue, J.; Hashidume, T.; Shimizu, M.; Sato, R. Anti-obesity and anti-hyperglycemic effects of the dietary citrus limonoid nomilin in mice fed a high-fat diet. *Biochemical and biophysical research communications* **2011**, *410*, 677-681.
- (107) Bailey, C. J.; Day, C. Traditional plant medicines as treatments for diabetes. *Diabetes care* **1989**, *12*, 553-564.
- (108) Marles, R. J.; Farnsworth, N. R. Antidiabetic plants and their active constituents. *Phytomedicine : international journal of phytotherapy and phytopharmacology* **1995**, *2*, 137-189.
- (109) David, B.; Wolfender, J.-L.; Dias, D. A. The pharmaceutical industry and natural products: historical status and new trends. *Phytochemistry Reviews* **2014**, *14*, 299-315.
- (110) Farnsworth, N. R.; Akerele, O.; Bingel, A. S.; Soejarto, D. D.; Guo, Z. Medicinal plants in therapy. *Bull World Health Organ* **1985**, *63*, 965-981.
- (111) Rollinger, J. M.; Langer, T.; Stuppner, H. Strategies for efficient lead structure discovery from natural products. *Current medicinal chemistry* **2006**, *13*, 1491-1507.
- (112) Strohl, W. R. The role of natural products in a modern drug discovery program. *Drug discovery today* **2000**, *5*, 39-41.
- (113) Koehn, F. E.; Carter, G. T. The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **2005**, *4*, 206-220.
- (114) Lee, M. L.; Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J Comb Chem* **2001**, *3*, 284-289.
- (115) Borman, S. T. U. ORGANIC LAB SPARKS DRUG DISCOVERY. *Chemical & Engineering News Archive* **2002**, *80*, 23.
- (116) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **2016**, *79*, 629-661.
- (117) Eng, J.; Kleinman, W. A.; Singh, L.; Singh, G.; Raufman, J. P. Isolation and characterization of exendin-4, an exendin-3 analogue, from *Heloderma suspectum* venom. Further evidence for an exendin receptor on dispersed acini from guinea pig pancreas. *The Journal of biological chemistry* **1992**, *267*, 7402-7405.

- (118) Cragg, G. M.; Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochimica et biophysica acta* **2013**, *1830*, 3670-3695.
- (119) Singh, R. B.; Niaz, M. A.; Ghosh, S. Hypolipidemic and antioxidant effects of *Commiphora mukul* as an adjunct to dietary therapy in patients with hypercholesterolemia. *Cardiovasc Drugs Ther* **1994**, *8*, 659-664.
- (120) Sinal, C. J.; Gonzalez, F. J. Guggulsterone: an old approach to a new problem. *Trends Endocrinol Metab* **2002**, *13*, 275-276.
- (121) Meng, Q.; Chen, X.; Wang, C.; Liu, Q.; Sun, H.; Sun, P.; Peng, J.; Liu, K. Alisol B 23-acetate promotes liver regeneration in mice after partial hepatectomy via activating farnesoid X receptor. *Biochem Pharmacol* **2014**, *92*, 289-298.
- (122) Grienke, U.; Mihaly-Bison, J.; Schuster, D.; Afonyushkin, T.; Binder, M.; Guan, S. H.; Cheng, C. R.; Wolber, G.; Stuppner, H.; Guo, D. A.; Bochkov, V. N.; Rollinger, J. M. Pharmacophore-based discovery of FXR-agonists. Part II: identification of bioactive triterpenes from *Ganoderma lucidum*. *Bioorg Med Chem* **2011**, *19*, 6779-6791.
- (123) Zou, J.; Jiang, J.; Diao, Y. Y.; Yang, L. B.; Huang, J.; Li, H. L.; Du, X.; Xiao, W. L.; Pu, J. X.; Sun, H. D. Cycloartane triterpenoids from the stems of *Schisandra glaucescens* and their bioactivity. *Fitoterapia* **2012**, *83*, 926-931.
- (124) Carter, B. A.; Taylor, O. A.; Prendergast, D. R.; Zimmerman, T. L.; Von Furstenberg, R.; Moore, D. D.; Karpen, S. J. Stigmasterol, a soy lipid-derived phytosterol, is an antagonist of the bile acid nuclear receptor FXR. *Pediatr Res* **2007**, *62*, 301-306.
- (125) De Marino, S.; Ummarino, R.; D'Auria, M. V.; Chini, M. G.; Bifulco, G.; D'Amore, C.; Renga, B.; Mencarelli, A.; Petek, S.; Fiorucci, S.; Zampella, A. 4-Methylenesterols from *Theonella swinhoei* sponge are natural pregnane-X-receptor agonists and farnesoid-X-receptor antagonists that modulate innate immunity. *Steroids* **2012**, *77*, 484-495.
- (126) Sepe, V.; Ummarino, R.; D'Auria, M. V.; Chini, M. G.; Bifulco, G.; Renga, B.; D'Amore, C.; Debitus, C.; Fiorucci, S.; Zampella, A. Conicasterol E, a small heterodimer partner sparing farnesoid X receptor modulator endowed with a pregnane X receptor agonistic activity, from the marine sponge *Theonella swinhoei*. *J Med Chem* **2012**, *55*, 84-93.
- (127) Rollinger, J. M.; Quinn, R. J. In silico Driven Pharmacognosy: Forth, Back and Reverse. *Planta Med* **2015**, *81*, 427-428.
- (128) Miyao, T.; Reker, D.; Schneider, P.; Funatsu, K.; Schneider, G. Chemography of natural product space. *Planta Med* **2015**, *81*, 429-435.
- (129) Hegazy, M. E.; Ibrahim, A. Y.; Mohamed, T. A.; Shahat, A. A.; El Halawany, A. M.; Abdel-Azim, N. S.; Alsaid, M. S.; Pare, P. W. Sesquiterpene Lactones from *Cynara cornigera*: Acetyl Cholinesterase Inhibition and In Silico Ligand Docking. *Planta Med* **2016**, *82*, 138-146.

(130) Gidaro, M. C.; Alcaro, F.; Carradori, S.; Costa, G.; Vullo, D.; Supuran, C. T.; Alcaro, S. Eriocitrin and Apigenin as New Carbonic Anhydrase VA Inhibitors from a Virtual Screening of Calabrian Natural Products. *Planta Med* **2015**, *81*, 533-540.

(131) Goel, R. K.; Gawande, D.; Lagunin, A.; Randhawa, P.; Mishra, A.; Poroikov, V. Revealing Medicinal Plants That Are Useful for the Comprehensive Management of Epilepsy and Associated Comorbidities through In Silico Mining of Their Phytochemical Diversity. *Planta Med* **2015**, *81*, 495-506.

(132) Kratz, J. M.; Mair, C. E.; Oettl, S. K.; Saxena, P.; Scheel, O.; Schuster, D.; Hering, S.; Rollinger, J. M. hERG Channel Blocking Ipecac Alkaloids Identified by Combined In Silico - In Vitro Screening. *Planta Med* **2016**.

(133) Vuorinen, A.; Seibert, J.; Papageorgiou, V. P.; Rollinger, J. M.; Odermatt, A.; Schuster, D.; Assimopoulou, A. N. Pistacia lentiscus Oleoresin: Virtual Screening and Identification of Masticadienonic and Isomasticadienonic Acids as Inhibitors of 11 β -Hydroxysteroid Dehydrogenase 1. *Planta Med* **2015**, *81*, 525-532.

(134) Schuster, D.; Markt, P.; Grienke, U.; Mihaly-Bison, J.; Binder, M.; Noha, S. M.; Rollinger, J. M.; Stuppner, H.; Bochkov, V. N.; Wolber, G. Pharmacophore-based discovery of FXR agonists. Part I: Model development and experimental validation. *Bioorg Med Chem* **2011**, *19*, 7168-7180.

(135) Waltenberger, B.; Atanasov, A. G.; Heiss, E. H.; Bernhard, D.; Rollinger, J. M.; Breuss, J. M.; Schuster, D.; Bauer, R.; Kopp, B.; Franz, C.; Bochkov, V.; Mihovilovic, M. D.; Dirsch, V. M.; Stuppner, H. Drugs from nature targeting inflammation (DNTI): a successful Austrian interdisciplinary network project. *Monatshefte Fur Chemie* **2016**, *147*, 479-491.

(136) Guasch, L.; Sala, E.; Castell-Auvi, A.; Cedo, L.; Liedl, K. R.; Wolber, G.; Muehlbacher, M.; Mulero, M.; Pinent, M.; Ardevol, A.; Valls, C.; Pujadas, G.; Garcia-Vallve, S. Identification of PPAR γ partial agonists of natural origin (I): development of a virtual screening procedure and in vitro validation. *PloS one* **2012**, *7*, e50816.

(137) Kratz, J. M.; Schuster, D.; Edtbauer, M.; Saxena, P.; Mair, C. E.; Kirchbner, J.; Matuszczak, B.; Baburin, I.; Hering, S.; Rollinger, J. M. Experimentally validated HERG pharmacophore models as cardiotoxicity prediction tools. *Journal of chemical information and modeling* **2014**, *54*, 2887-2901.

(138) Rollinger, J. M.; Langer, T.; Stuppner, H. Integrated in silico tools for exploiting the natural products' bioactivity. *Planta Med* **2006**, *72*, 671-678.

(139) Langer, T.; Wolber, G. Pharmacophore definition and 3D searches. *Drug Discov Today Technol* **2004**, *1*, 203-207.

(140) Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1998). **1998**.

- (141) History of 3D pharmacophore searching: commercial, academic and open-source tools. *Drug Discov Today Technol* **2010**, 7, e203-270.
- (142) Lloyd, E. J.; Andrews, P. R. A common structural model for central nervous system drugs and their receptors. *J Med Chem* **1986**, 29, 453-462.
- (143) Dodds, E. C.; Lawson, W. Molecular Structure in Relation to Oestrogenic Activity. Compounds without a Phenanthrene Nucleus. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **1938**, 125, 222-232.
- (144) Braga, R. C.; Andrade, C. H. Assessing the performance of 3D pharmacophore models in virtual screening: how good are they? *Current topics in medicinal chemistry* **2013**, 13, 1127-1138.
- (145) Karaboga, A. S.; Planesas, J. M.; Petronin, F.; Teixido, J.; Souchet, M.; Perez-Nueno, V. I. Highly specific and sensitive pharmacophore model for identifying CXCR4 antagonists. Comparison with docking and shape-matching virtual screening performance. *J Chem Inf Model* **2013**, 53, 1043-1056.
- (146) Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D pharmacophore-based virtual screening. *Drug Discovery Today: Technologies* **2010**, 7, e221-e228.
- (147) Markt, P.; Schuster, D.; Langer, T.: Pharmacophore Models for Virtual Screening. In *Virtual Screening*; Wiley-VCH Verlag GmbH & Co. KGaA, 2011; pp 115-152.
- (148) Spitzer, G. M.; Heiss, M.; Mangold, M.; Markt, P.; Kirchmair, J.; Wolber, G.; Liedl, K. R. One concept, three implementations of 3D pharmacophore-based virtual screening: distinct coverage of chemical search space. *Journal of chemical information and modeling* **2010**, 50, 1241-1247.
- (149) Wolber, G.; Dornhofer, A. A.; Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. *J Comput Aided Mol Des* **2006**, 20, 773-788.
- (150) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of chemical information and modeling* **2005**, 45, 160-169.
- (151) Irwin, J. J. Community benchmarks for virtual screening. *J Comput Aided Mol Des* **2008**, 22, 193-199.
- (152) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J Med Chem* **2006**, 49, 6789-6801.
- (153) Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Laggner, C.; Langer, T. Fast and efficient in silico 3D screening: toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *Journal of chemical information and modeling* **2007**, 47, 2182-2196.

(154) Weston, J.; Perez-Cruz, F.; Bousquet, O.; Chapelle, O.; Elisseeff, A.; Scholkopf, B. Feature selection and transduction for prediction of molecular bioactivity for drug design. *Bioinformatics* **2003**, *19*, 764-771.

(155) Jacobsson, M.; Liden, P.; Stjernschantz, E.; Bostrom, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J Med Chem* **2003**, *46*, 5781-5789.

(156) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem* **2005**, *48*, 2534-2547.

(157) Sullivan, P. M.: *The Statistical Evaluation of Medical Tests for Classification and Prediction*; Oxford University Press, 2003.

(158) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew Chem Int Ed Engl* **1999**, *38*, 2894-2896.

(159) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug discovery today* **2006**, *11*, 580-594.

(160) Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. Free resources to assist structure-based virtual ligand screening experiments. *Curr Protein Pept Sci* **2007**, *8*, 381-411.

(161) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing pitfalls in virtual screening: a critical review. *Journal of chemical information and modeling* **2012**, *52*, 867-881.

(162) Scior, T.; Bernard, P.; Medina-Franco, J. L.; Maggiora, G. M. Large compound databases for structure-activity relationships studies in drug discovery. *Mini reviews in medicinal chemistry* **2007**, *7*, 851-860.

(163) Bisson, J.; McAlpine, J. B.; Friesen, J. B.; Chen, S. N.; Graham, J.; Pauli, G. F. Can Invalid Bioactives Undermine Natural Product-Based Drug Discovery? *J Med Chem* **2016**, *59*, 1671-1690.

(164) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry* **2012**, *55*, 6582-6594.

(165) Teixeira da Silva, J. A. Negative results: negative perceptions limit their potential for increasing reproducibility. *Journal of Negative Results in Biomedicine* **2015**, *14*, 12.

- (166) von Korff, M.; Freyss, J.; Sander, T. Comparison of ligand- and structure-based virtual screening on the DUD data set. *Journal of chemical information and modeling* **2009**, *49*, 209-231.
- (167) Novikov, F. N.; Stroylov, V. S.; Zeifman, A. A.; Stroganov, O. V.; Kulkov, V.; Chilov, G. G. Lead Finder docking and virtual screening evaluation with Astex and DUD test sets. *J Comput Aided Mol Des* **2012**, *26*, 725-735.
- (168) Rohrer, S. G.; Baumann, K. Maximum Unbiased Validation (MUV) Data Sets for Virtual Screening Based on PubChem Bioactivity Data. *Journal of chemical information and modeling* **2009**, *49*, 169-184.
- (169) Wallach, I.; Lilien, R. Virtual decoy sets for molecular docking benchmarks. *Journal of chemical information and modeling* **2011**, *51*, 196-202.
- (170) Xia, J.; Jin, H.; Liu, Z.; Zhang, L.; Wang, X. S. An unbiased method to build benchmarking sets for ligand-based virtual screening and its application to GPCRs. *Journal of chemical information and modeling* **2014**, *54*, 1433-1450.
- (171) Gatica, E. A.; Cavasotto, C. N. Ligand and decoy sets for docking to G protein-coupled receptors. *Journal of chemical information and modeling* **2012**, *52*, 1-6.
- (172) Cereto-Massague, A.; Guasch, L.; Valls, C.; Mulero, M.; Pujadas, G.; Garcia-Vallve, S. DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics (Oxford, England)* **2012**, *28*, 1661-1662.
- (173) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *Journal of chemical information and modeling* **2010**, *50*, 572-584.
- (174) Ebejer, J. P.; Morris, G. M.; Deane, C. M. Freely available conformer generation methods: how good are they? *Journal of chemical information and modeling* **2012**, *52*, 1146-1158.
- (175) Wang, L.; Bai, L.; Nagasawa, T.; Hasegawa, T.; Yang, X.; Sakai, J.-i.; Bai, Y.; Kataoka, T.; Oka, S.; Hirose, K.; Tomida, A.; Tsuruo, T.; Ando, M. Bioactive Triterpene Saponins from the Roots of *Phytolacca americana*. *Journal of Natural Products* **2008**, *71*, 35-40.
- (176) Ono, M.; Yasuda, S.; Nishi, K.; Yamamoto, K.; Fuchizaki, S.; Higuchi, S.; Komatsu, H.; Okawa, M.; Kinjo, J.; Yoshimitsu, H.; Nohara, T. Two new triterpenoids from the seeds of blackberry (*Rubus fruticosus*). *Natural Product Research* **2016**, *30*, 904-911.
- (177) De Tommasi, N.; De Simone, F.; Cirino, G.; Cicala, C.; Pizza, C. Hypoglycemic Effects of Sesquiterpene Glycosides and Polyhydroxylated Triterpenoids of *Eriobotrya japonica*. *Planta Med* **1991**, *57*, 414-416.

- (178) Mair, C. E. Bioactive natural products: targeted identification of antivirals and hERG blockers. Dissertation, Universitaet Wien, 2016.
- (179) Li, T.; Apte, U. Bile Acid Metabolism and Signaling in Cholestasis, Inflammation, and Cancer. *Advances in pharmacology (San Diego, Calif.)* **2015**, *74*, 263-302.
- (180) Kozoni, V.; Tsioulas, G.; Shiff, S.; Rigas, B. The effect of lithocholic acid on proliferation and apoptosis during the early stages of colon carcinogenesis: differential effect on apoptosis in the presence of a colon carcinogen. *Carcinogenesis* **2000**, *21*, 999-1005.
- (181) Dossa, A. Y.; Escobar, O.; Golden, J.; Frey, M. R.; Ford, H. R.; Gayer, C. P. Bile acids regulate intestinal cell proliferation by modulating EGFR and FXR signaling. *American journal of physiology. Gastrointestinal and liver physiology* **2016**, *310*, G81-92.
- (182) Dastan, D.; Salehi, P.; Ghanati, F.; Gohari, A. R.; Maroofi, H.; Alnajjar, N. Phytotoxicity and cytotoxicity of disesquiterpene and sesquiterpene coumarins from *Ferula pseudalliacea*. *Industrial Crops and Products* **2014**, *55*, 43-48.
- (183) Baell, J. B. Feeling Nature's PAINS: Natural Products, Natural Product Drugs, and Pan Assay Interference Compounds (PAINS). *Journal of Natural Products* **2016**, *79*, 616-628.
- (184) Schneider, N. F. Z.; Geller, F. C.; Persich, L.; Marostica, L. L.; Pádua, R. M.; Kreis, W.; Braga, F. C.; Simões, C. M. O. Inhibition of cell proliferation, invasion and migration by the cardenolides digitoxigenin monodigitoxoside and convallatoxin in human lung cancer cell line. *Natural Product Research* **2016**, *30*, 1327-1331.
- (185) Mijatovic, T.; Van Quaquebeke, E.; Delest, B.; Debeir, O.; Darro, F.; Kiss, R. Cardiotonic steroids on the road to anti-cancer therapy. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2007**, *1776*, 32-57.
- (186) Larsson, J.; Gottfries, J.; Muresan, S.; Backlund, A. ChemGPS-NP: Tuned for Navigation in Biologically Relevant Chemical Space. *Journal of Natural Products* **2007**, *70*, 789-794.
- (187) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* **2007**, *50*, 74-82.
- (188) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *Journal of Computational Chemistry* **1996**, *17*, 1653-1666.
- (189) ROCS information. <http://www.eyesopen.com/rocs> (accessed 08.06.2016).
- (190) Sheridan, R. P.; McGaughey, G. B.; Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J Comput Aided Mol Des* **2008**, *22*, 257-265.

- (191) ChEMBL database information homepage. <https://www.ebi.ac.uk/about> (accessed 07.06.2016).
- (192) CAS Homepage. <http://www.cas.org/products/scifinder/features> (accessed 07.06.2016).
- (193) patentscope. <http://www.wipo.int/patentscope/en/> (accessed 1.6.16).
- (194) Google. Google Patents Support. <https://support.google.com/faqs/answer/6261372> (accessed 07.06.2016).
- (195) Pellicciari, R.; Sato, H.; Gioiello, A.; Costantino, G.; Macchiarulo, A.; Sadeghpour, B. M.; Giorgi, G.; Schoonjans, K.; Auwerx, J. Nongenomic actions of bile acids. Synthesis and preliminary characterization of 23- and 6,23-alkyl-substituted bile acid derivatives as selective modulators for the G-protein coupled receptor TGR5. *J Med Chem* **2007**, *50*, 4265-4268.
- (196) Pellicciari, R.; Gioiello, A.; Sabbatini, P.; Venturoni, F.; Nuti, R.; Colliva, C.; Rizzo, G.; Adorini, L.; Pruzanski, M.; Roda, A.; Macchiarulo, A. Avicholic Acid: A Lead Compound from Birds on the Route to Potent TGR5 Modulators. *ACS medicinal chemistry letters* **2012**, *3*, 273-277.
- (197) Zou, Q.; Duan, H.; Ning, M.; Liu, J.; Feng, Y.; Zhang, L.; Zhu, J.; Leng, Y.; Shen, J. 4-Benzofuranyloxynicotinamide derivatives are novel potent and orally available TGR5 agonists. *European journal of medicinal chemistry* **2014**, *82*, 1-15.
- (198) Sepe, V.; Renga, B.; Festa, C.; D'Amore, C.; Masullo, D.; Cipriani, S.; Di Leva, F. S.; Monti, M. C.; Novellino, E.; Limongelli, V.; Zampella, A.; Fiorucci, S. Modification on ursodeoxycholic acid (UDCA) scaffold. discovery of bile acid derivatives as selective agonists of cell-surface G-protein coupled bile acid receptor 1 (GP-BAR1). *J Med Chem* **2014**, *57*, 7687-7701.
- (199) Hogenauer, K.; Arista, L.; Schmiedeberg, N.; Werner, G.; Jaksche, H.; Bouhelal, R.; Nguyen, D. G.; Bhat, B. G.; Raad, L.; Rauld, C.; Carballido, J. M. G-protein-coupled bile acid receptor 1 (GPBAR1, TGR5) agonists reduce the production of proinflammatory cytokines and stabilize the alternative macrophage phenotype. *J Med Chem* **2014**, *57*, 10343-10354.
- (200) Zambad, S. P.; Tuli, D.; Mathur, A.; Ghalsasi, S. A.; Chaudhary, A. R.; Deshpande, S.; Gupta, R. C.; Chauthaiwale, V.; Dutt, C. TRC210258, a novel TGR5 agonist, reduces glycemic and dyslipidemic cardiovascular risk in animal models of diabetes. *Diabetes Metab Syndr Obes* **2013**, *7*, 1-14.
- (201) Lamers, C.; Merk, D.; Gabler, M.; Flesch, D.; Kaiser, A.; Schubert-Zsilavecz, M. SAR studies on FXR modulators led to the discovery of the first combined FXR antagonistic/TGR5 agonistic compound. *Future Med Chem* **2016**, *8*, 133-148.
- (202) Raval, S. Compound and compositions as TGR5 agonists: WO2012082947. *Expert Opin Ther Pat* **2013**, *23*, 531-536.

- (203) Agarwal, S.; Patil, A.; Aware, U.; Deshmukh, P.; Darji, B.; Sasane, S.; Sairam, K. V.; Priyadarsiny, P.; Giri, P.; Patel, H.; Giri, S.; Jain, M.; Desai, R. C. Discovery of a Potent and Orally Efficacious TGR5 Receptor Agonist. *ACS medicinal chemistry letters* **2016**, 7, 51-55.
- (204) Itoh, F.; Hinuma, S.; Kanzaki, N.; Miki, T.; Kawamata, Y.; Oi, S.; Tawaraishi, T.; Ishichi, Y.; Hirohashi, M.: Receptor agonists. Google Patents, 2006.
- (205) Smith, N. D.; Payne, J. E.; Hoffman, T. Z.: Heterocyclic modulators of tgr5. Google Patents, 2010.
- (206) Smith, E. D.; FITZGERALD, M. E.; Jiang, R.; Atkinson, R. N.; Veal, J. M.; HUANG, K. H.: Arylsubstituted thiazolotriazoles and thiazoloimidazoles. Google Patents, 2013.
- (207) Lewis, J. G.; Reich, N.; Chen, T.; Jacobs, J. W.; Charmot, D.; Navre, M.; FINN, P.; Carreras, C.; Spencer, A.: Non-systemic tgr5 agonists. Google Patents, 2013.
- (208) Agarwal, S.; Desai, R. C.: Heterocyclic compounds and their use for treatment of diabetes, obesity or related disorders. Google Patents, 2013.
- (209) Agarwal, S.; Jain, M. R.; Patel, P. R.: 2-thio-imidazole derivatives as tgr5 modulators. Google Patents, 2013.
- (210) Agarwal, S.; Desai, R. C.: Novel compounds for treatment of diabetes, obesity or related disorders. Google Patents, 2013.
- (211) Goff, D.; Payan, D. G.; Braselmann, S.: Synthesis of polycyclic alkaloids and their use as tgr5 agonists. Google Patents, 2013.
- (212) Sugimoto, Y.; Satoh, A.; Nishimura, T.: Novel isoquinoline derivatives. Google Patents, 2010.
- (213) Arakawa, K.; Sugimoto, Y.; Sasaki, Y.; Satoh, A.; Nishimura, T.: Novel isoquinolinylloxymethyl heteroaryl derivatives. Google Patents, 2010.
- (214) Dutt, C.; Chauthaiwale, V.; Gupta, R. C.; GHALSASI, S.; Tuli, D.; Deshpande, S.; Chaudhari, A.; ZAMBAD, S.: Novel substituted imidazopyrimidines as gpbar1 receptor modulators. Google Patents, 2013.
- (215) Bissantz, C.; Dehmlow, H.; Erickson, S. D.; Kim, K.; Martin, R. E.; Mattei, P.; Sander, U. O.; Pietranico-Cole, S. L.; Richter, H.; Ullmer, C.: 1-hydroxyimino-3-phenyl-propanes. Google Patents, 2012.
- (216) Dehmlow, H.; Martin, R. E.; Mattei, P.; Obst, S. U.; Richter, H.: 1-cycloalkyl- or 1-heterocyclyl-hydroxyimino-3-phenyl-propanes. Google Patents, 2013.
- (217) Bissantz, C.; Dehmlow, H.; Erickson, S. D.; Karnachi, P. S.; Kim, K.; MARTIN, R. E.; Mattei, P.; Obst, S. U.; Pietranico-Cole, S. L.; Richter, H.: 3-amino-pyridines as gpbar1 agonists. Google Patents, 2012.
- (218) Dehmlow, H.; Martin, R. E.; Mattei, P.; Sander, U. O.; Richter, H.: 1-pyridazinyl-hydroxyimino-3-phenyl-propanes. Google Patents, 2013.

- (219) Bissantz, C.; Dehmlow, H.; Martin, R. E.; Sander, U. O.; Richter, H.; Ullmer, C.: Novel phenyl amide or pyridyl amide derivatives. Google Patents, 2010.
- (220) Gao, W.; Lau, T.; Pan, S.; Phillips, D. P.; Wang, X.; Yang, Y.: Compounds and compositions as tgr5 agonists. Google Patents, 2012.
- (221) Bollu, V.; Boren, B. C.; Dalgard, J. E.; Flatt, B. T.; Haq, N.; Hudson, S.; Mohan, R.; Morrissey, M.; Pratt, B.; Wang, T. L.: Triazole and imidazole derivatives for use as tgr5 agonists in the treatment of diabetes and obesity. Google Patents, 2010.
- (222) Flatt, B. T.; Mohan, R.: Tgr5 agonists having an imidazole or triazole core with substituent having a quaternary nitrogen. Google Patents, 2014.
- (223) Robl, J. A.; Li, J.; Kennedy, L. J.; Walker, S. J.; Wang, H.; Washburn, W. N.; Ahmad, S.; Ngu, K.: Novel bicyclic nitrogen containing heteroaryl tgr5 receptor modulators. Google Patents, 2012.
- (224) **五元唑类杂环化合物及其制备方法、药物组合物和用途.**
- (225) Rollinger, J. M.; Haupt, S.; Stuppner, H.; Langer, T. Combining ethnopharmacology and virtual screening for lead structure discovery: COX-inhibitors as application example. *Journal of chemical information and computer sciences* **2004**, *44*, 480-488.
- (226) von Bergmann, K.; Schwarz, H. P.; Paumgartner, G. Effect of phenobarbital, spironolactone and pregnenolone-16 α -carbonitrile on bile formation in the rat. *Naunyn-Schmiedeberg's Archives of Pharmacology* **1975**, *287*, 33-45.
- (227) Schramm, A.; Saxena, P.; Chlebek, J.; Cahlikova, L.; Baburin, I.; Hering, S.; Hamburger, M. Natural products as potential human ether-a-go-go-related gene channel inhibitors - screening of plant-derived alkaloids. *Planta Med* **2014**, *80*, 740-746.
- (228) Rollinger, J. M.; Steindl, T. M.; Schuster, D.; Kirchmair, J.; Anrain, K.; Ellmerer, E. P.; Langer, T.; Stuppner, H.; Wutzler, P.; Schmidtke, M. Structure-based virtual screening for the discovery of natural inhibitors for human rhinovirus coat protein. *J Med Chem* **2008**, *51*, 842-851.

9. Appendix

COMPLETE HITLIST								
SMILES-CODE	NAME	group	Source DB	Best Match	TC_CHEMBL27 0105	PF Score	MW	
O1C[C@H]5C[C@H](O)C1C[C@H]4C[C@H]1[C@@H]3[C@H]1[C@H]1C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10047885	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.737	73.7072	514.75	
O1C[C@H]4CC3C[C@H]1[C@@H]2[C@H]1[C@@H]1[C@@H]1(C=C=O)C=CC1)CC2(C)CC=3(C)CC4C	1009330	6	NPD08_150-700_iconfast.ldb	BAMS22	0.962	55.934273	314.47	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1[C@@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10134536	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.84	73.09075	472.71	
O1C[C@H]1[C@@H](O)C1C[C@H]2[C@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10136434	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.789	73.65759	474.68	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10225842	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.87	72.832436	472.67	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10274557	6	NPD08_150-700_iconfast.ldb	BAMS22	0.835	65.616844	280.32	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10303301	9	NPD08_150-700_iconfast.ldb	BAMS22	1.239	55.93392	398.63	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10317206	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.772	73.13624	488.71	
O3[C@H]1[C@@H]1[C@@H]1[C@@H]1(C=C=O)C=CC1)CC2(C)CC=3(C)CC4C	10489609	7	NPD08_150-700_iconfast.ldb	BAMS22	0.691	55.13058	344.36	
O=C1C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10498189	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.932	72.6352	452.68	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10506694	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.802	73.25457	488.71	
O1C[C@H]1[C@@H](O)C1C[C@H]2[C@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10508031	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.864	73.27148	504.71	
O1C[C@H]1[C@@H](O)C1C[C@H]2[C@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10508098	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.79	72.682655	506.72	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10622193	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.807	73.09496	470.74	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10623777	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.762	73.208534	474.73	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10624033	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.875	72.29958	472.71	
O=C1C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10624287	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.727	72.38457	474.73	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10628641	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.791	73.67866	486.69	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10629251	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.846	73.06671	502.69	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10702202	6	NPD08_150-700_iconfast.ldb	BAMS22	0.677	65.51597	306.49	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10703713[6575105]	2	NPD08_150-700_iconfast.ldb	BAMS22	0.615	56.042805	344.41	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10730782	7	NPD08_150-700_iconfast.ldb	BAMS22	0.895	66.14081	380.52	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10740229	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.861	73.03421	454.69	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10743039	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.81	73.727455	472.71	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10745886	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.823	73.680084	496.73	
O1C[C@H]1[C@@H](O)C1C[C@H]2[C@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10748936	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.734	73.21632	504.71	
O2[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10749818	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.724	73.076866	544.77	
O1C[C@H]1C=C1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10794806	6	NPD08_150-700_iconfast.ldb	BAMS22	0.72	66.134384	274.4	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10845267	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.6	73.570625	586.81	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10845645	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.78	73.67255	586.81	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10845735	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.636	72.66088	604.82	
O1C[C@H]5C[C@H]1[C@@H]1[C@@H]1[C@@H]1[C@@H]1(C)C1C[C@H]2C[C@H]1(C=O)O)CC1)CCC(C2)(C)C=CC3(C)(CC4)C(C5)C(C)C=C	10846217	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.756	73.44562	602.81	
O=CNC1C[C@H]1C=C1C=C1C(C)C1C=C1C								

[illegible]

Q[C@H]2C[C@H]1C@H(O)(C=O)C[C@H]3([C@H]1CC2)C)CC=C4[C@H]3(CC[C@H]5[C@H]4CC(C)(C)C)C)C	2924830	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.73	73.7095	514.75
O=C(O)C=C(C[C@H]2[C@H]1C@H([C@H]1CO)CC(C)C)CCC2=C)C)C	2947020	6	NPD08_150-700_iconfast.lib	BAMS22	0.66	46.582207	320.47
O=C(O)C[C@H]2([C@H]1C@H([C@H]1C)CC(C)C)C)C)C)C)C)C	2951105	6	NPD08_150-700_iconfast.lib	BAMS22	0.598	46.342033	362.51
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C	2951262	7	NPD08_150-700_iconfast.lib	BAMS22	1.263	67.43715	392.58
Q[C@H]4C[C@H]1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C	2952101	7	NPD08_150-700_iconfast.lib	BAMS22	1.241	67.484604	406.61
Q[C@H]2C[C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C)C)C	2952112	7	NPD08_150-700_iconfast.lib	BAMS22	1.611	68.53722	406.61
Q[C@H]4C[C@H]1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C	2952833	7	NPD08_150-700_iconfast.lib	BAMS22	1.234	67.40834	408.58
Q[C@H]4C=C1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C	2952884	7	NPD08_150-700_iconfast.lib	BAMS22	1.144	56.94163	404.59
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C	2954537	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.85	73.91911	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2954557	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.942	73.52145	442.68
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2954988	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	1.012	72.776054	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2954996	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.964	73.423256	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2955643	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.742	73.226814	472.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2955645	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.909	72.53716	472.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2955646	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.847	73.26791	472.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2955658	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.777	73.721275	472.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2955659	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.918	72.78426	472.71
Q[C@H]4C[C@H]1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2955826	8	NPD08_150-700_iconfast.lib	BAMS22	1.272	67.23802	465.63
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	2956487	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.994	72.34331	498.75
O=C1C(C=C1)C=C(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C	2958054	3	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.598	72.98306	618.85
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3001410	6	NPD08_150-700_iconfast.lib	BAMS22	1.764	68.09963	376.58
O=C(O)C=C([C@H]2C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3002306	6	NPD08_150-700_iconfast.lib	BAMS22	0.692	46.815456	348.48
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3015155	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.814	73.75139	454.69
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3020380	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.691	72.85472	488.71
Q[C@H]1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C)C	3021366	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.782	72.97712	504.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3022015	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.8	73.76288	500.72
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3022958	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.789	73.1561	514.75
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3033634	6	NPD08_150-700_iconfast.lib	BAMS22	0.696	56.588596	308.51
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3037754	7	NPD08_150-700_iconfast.lib	BAMS22	1.161	56.050655	348.48
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3037757	7	NPD08_150-700_iconfast.lib	BAMS22	1.254	56.690105	348.48
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3040426	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.989	72.78411	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3040428	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.888	73.51022	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3040429	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.847	73.35994	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3041211	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.776	73.50246	498.75
Q[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C)C	3076227	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.759	73.117294	488.71
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3086265	7	NPD08_150-700_iconfast.lib	BAMS22	0.949	56.111225	392.58
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3087145	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.847	72.74016	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3087421	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.798	73.58499	488.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3087461	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.754	72.18284	498.75
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3087486	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.816	73.4797	486.69
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3087568	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.821	73.72522	500.72
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3087721	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.786	73.460106	542.76
O=C(O)C=C([C@H]2C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3105005	6	NPD08_150-700_iconfast.lib	BAMS22	0.734	46.557194	348.53
O=C2([C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C	3108660	7	NPD08_150-700_iconfast.lib	BAMS22	1.003	46.323	344.45
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3108894	7	NPD08_150-700_iconfast.lib	BAMS22	1.593	68.1895	390.56
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3110754	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.835	72.4717	458.73
O=C(O)C=C([C@H]2C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3111782	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.849	72.9191	456.71
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3112749	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.829	73.73612	454.69
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3113076	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.776	73.24764	472.71
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3115162	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.792	73.25792	502.69
Q[C@H]1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C)C	3123826	8	NPD08_150-700_iconfast.lib	BAMS22	0.983	66.74086	478.63
Q[C@H]1C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C)C	3123858	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.776	73.1332	504.71
O=C1C(C=C1)C=C(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C	3164905	6	NPD08_150-700_iconfast.lib	BAMS22	1.09	47.749996	328.45
O=C4C=C3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C)C)C	3165221	9	NPD08_150-700_iconfast.lib	BAMS22	1	55.885014	400.65
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3168531	7	NPD08_150-700_iconfast.lib	BAMS22	1.642	67.71781	408.58
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3168616	7	NPD08_150-700_iconfast.lib	BAMS22	1.577	68.74398	408.58
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3171824	7	NPD08_150-700_iconfast.lib	BAMS22	1.319	57.340908	404.55
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3173096	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.771	73.581345	458.73
Q[C@H]5C([C@H]4C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3175245	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.868	73.61132	456.71
Q[C@H]4C[C@H]3C@H([C@H]1C@H([C@H]1C)C)C)C)C)C)C	3177154	8	NPD08_150-700_iconfast.lib	BAMS22	1.27	68.45918	466.66

[illegible]

[illegible]

0)CC4)CC(C@H)56)C							
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5362130	8	NPD08_150-700_iconfast.ldb	BAMS22	0.964	66.03426	460.65
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5364066	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.747	73.44927	504.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5384623	8	NPD08_150-700_iconfast.ldb	BAMS22	1.194	56.473106	450.66
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5449307	6	NPD08_150-700_iconfast.ldb	BAMS22	0.668	45.97687	362.51
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5457867	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.809	73.40771	440.67
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5458963	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.805	73.64002	454.69
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5460842	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.87	73.162964	498.75
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5461569	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.749	73.45227	454.65
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5461879	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.891	73.39619	486.69
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5465060	3	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.64	73.50982	558.8
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5466076	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.594	73.61882	588.83
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5466425	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.68	71.963425	604.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5475603	8	NPD08_150-700_iconfast.ldb	BAMS22	0.605	68.59636	548.74
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5482850	7	NPD08_150-700_iconfast.ldb	BAMS22	0.885	56.138634	354.49
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5486579	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.847	72.993164	488.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	54990-68-0	1	phkqg_db_big_20160621.ldb	BAMS22	0.737	46.390884	382.5
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5561841	6	NPD08_150-700_iconfast.ldb	BAMS22	0.696	56.02181	304.47
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5601356	8	NPD08_150-700_iconfast.ldb	BAMS22	1.228	67.6749	452.68
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5647230	9	NPD08_150-700_iconfast.ldb	BAMS22	1.125	56.378746	425.61
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5648319	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.854	73.92757	472.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5648545	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.822	73.88439	472.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	56536	7	NPD08_150-700_iconfast.ldb	BAMS22	1.071	58.084167	406.52
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5656140	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.867	73.54706	488.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	566-65-4	6	phkqg_db_big_20160621.ldb	BAMS22	0.984	45.38991	316.48
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	566-78-9	6	phkqg_db_big_20160621.ldb	BAMS22	0.92	56.767437	374.52
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5679432	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.604	73.68498	572.78
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5683709	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.565	72.89185	618.85
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5748683	6	NPD08_150-700_iconfast.ldb	BAMS22	0.734	55.74723	306.49
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5755882	6	NPD08_150-700_iconfast.ldb	BAMS22	0.76	56.4374	322.49
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5763069	8	NPD08_150-700_iconfast.ldb	BAMS22	1.312	67.93461	438.65
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5775545	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.782	73.40925	472.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5776880	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.747	72.770294	472.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5785804	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.631	72.72048	604.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5786287	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.689	73.095215	604.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5786290	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.671	73.65877	604.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5786581	3	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.627	73.76883	618.85
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5788176	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.661	72.803925	632.83
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5855805	1	NPD08_150-700_iconfast.ldb	BAMS22	0.792	54.951836	343.4
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5857474	9	NPD08_150-700_iconfast.ldb	BAMS22	1.041	65.39292	427.63
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5881513	1	NPD08_150-700_iconfast.ldb	BAMS22	0.757	45.652138	344.49
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5887671	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.86	73.5286	456.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5890616	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.786	72.670555	488.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5892345	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.75	73.36379	518.73
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5892984	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.719	73.23849	520.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5895318	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.687	72.7225	604.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5913086	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.947	73.16587	472.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5913087	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.798	73.204155	472.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	5913850	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.796	73.60523	504.71
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6002350	6	NPD08_150-700_iconfast.ldb	BAMS22	1.128	47.54568	372.46
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6004519	6	NPD08_150-700_iconfast.ldb	BAMS22	1.249	56.175217	400.56
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6009853	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.857	73.707054	454.69
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6013377	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.807	73.68697	470.69
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6016049	7	NPD08_150-700_iconfast.ldb	BAMS22	1.174	56.927933	400.47
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6029476	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.593	73.15343	604.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6030311	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.628	73.61635	618.85
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6030931	4	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.703	73.53919	620.82
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6045831	8	NPD08_150-700_iconfast.ldb	BAMS22	0.82	46.504177	522.61
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6068504	7	NPD08_150-700_iconfast.ldb	BAMS22	1.092	57.388	350.5
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6068505	7	NPD08_150-700_iconfast.ldb	BAMS22	1.032	57.4119	350.5
01C@H)[C@H]1[C@H]2[C@H](C(C)C)C[C@H]1[C@H]3[C@H]4[C@H](O)(C2)C[C@H]1[C@H]3[C@H]4[C@H](C(C)C)C[C@H]1[C@H]2[C@H](C(C)C)C	6068506	7	NPD08_150-700_iconfast.ldb	BAMS22	1.081	57.548344	350.5

[illegible]

01C@H5C1C@HMA4C@H3C@H1C1C@H1C@HMA2C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6773063	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.831	73.1911	498.75
01C@H5C1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6773162	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.879	73.21853	504.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6823016	6	NPD08_150-700_iconfast.ldb	BAMS22	1.167	57.438602	386.58
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6828615	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.837	71.55652	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6897646	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.589	73.6533	676.89
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6937385	2	NPD08_150-700_iconfast.ldb	BAMS22	0.652	56.71109	342.43
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6972236	6	NPD08_150-700_iconfast.ldb	BAMS22	0.634	65.43647	308.51
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6977482	6	NPD08_150-700_iconfast.ldb	BAMS22	0.612	56.48242	334.5
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6984760	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.853	73.233635	440.67
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6985809	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.791	73.31241	456.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6986951	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.826	73.532425	472.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6987368	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.865	73.7168	458.68
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6987552	7	NPD08_150-700_iconfast.ldb	BAMS22	1.059	58.044956	418.53
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6987585	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.769	73.70358	472.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6987617	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.858	72.66629	472.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6987990	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.836	73.30383	500.76
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6988220	7	NPD08_150-700_iconfast.ldb	BAMS22	0.988	57.922695	416.51
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6988830	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.818	73.58233	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6989222	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.76	73.90997	504.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6989554	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.804	73.16026	504.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6989568	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.766	72.664795	504.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6992544	3	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.678	73.52016	594.83
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6992862	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.73	73.82671	588.78
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	6993802	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.743	72.93403	616.79
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7002003	7	NPD08_150-700_iconfast.ldb	BAMS22	0.788	66.28819	448.64
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7002850	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.815	73.26581	456.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7002851	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.929	72.354355	456.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7002852	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.801	73.8919	456.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7003061	7	NPD08_150-700_iconfast.ldb	BAMS22	0.811	57.235752	418.53
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7003159	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.8	72.78211	488.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7016196	7	NPD08_150-700_iconfast.ldb	BAMS22	0.84	46.153183	336.47
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7038917	6	NPD08_150-700_iconfast.ldb	BAMS22	0.741	56.606045	348.53
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7061714	3	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.637	73.6698	602.86
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7062564	3	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.606	73.108185	618.85
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7063294	3	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.555	73.81088	634.85
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7078219	6	NPD08_150-700_iconfast.ldb	BAMS22	0.778	55.828606	304.47
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7093276	6	NPD08_150-700_iconfast.ldb	BAMS22	0.598	55.931007	245.23
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7103503	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.823	73.30523	456.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7157743	4	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.598	73.35494	650.85
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7158884	4	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.662	73.710976	664.83
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	718608	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.881	73.753975	456.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7269929	6	NPD08_150-700_iconfast.ldb	BAMS22	0.755	55.13217	304.47
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7274919	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.781	71.92934	458.73
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7277039	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.778	73.20289	488.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7277396	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.754	73.214195	504.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7316276	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.943	73.17839	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7318765	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.767	73.2962	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7324170	5	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.581	72.87877	604.82
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7342638	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.835	73.62794	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7342639	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.791	72.48858	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7342899	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.926	73.4817	472.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7355256	1	NPD08_150-700_iconfast.ldb	BAMS22	0.606	56.878212	338.4
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7357294	6	NPD08_150-700_iconfast.ldb	BAMS22	1.255	57.36129	372.5
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7359041	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.784	73.828285	472.67
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7384669	6	NPD08_150-700_iconfast.ldb	BAMS22	0.882	46.6337	360.54
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7385743	6	NPD08_150-700_iconfast.ldb	BAMS22	0.906	46.80253	358.52
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7395070	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.818	73.87242	470.69
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7395147	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.763	73.11637	488.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7446866	9	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.822	73.25296	472.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7447693	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.914	73.395645	488.71
0=C(O)C@H1C@HMA4C@H3C@H1C(=O)O)(CC=1)CCCC(C2)(C)(CC3)C(C4)C(C5)C(C6)C(C7)C(C8)C	7450409	8	NPD08_150-700_iconfast.ldb	Model12.3.2.8	0.741	72.77857	520.71

[illegible]

Q(CC2)C(C)CC3(C)CC4(C)CC5(C)C(C)C							
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8377177	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.723	72.23419	542.76
C2(C)C(=CC3(C)CC4(C)CC5(C)C(C)C(=O)CC(=O)O							
Q(C@C@H)1OC(C@H)O(C@H)O(C@H)1O(C@H)6C@H(C@C@H)5C@H(C@C@H)4C@H(C@C@H)2C(C@C@H)3C@H(C(=O)O)(CC2)CCC(C3)(C)C(=CC4(C)CC5(C)CC6(C)CO)C	8380218	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.614	73.78844	604.82
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C@C@H(C@C@H)1C(C(=O)O)C	8398618	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.823	73.862724	472.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C@C@H(C@C@H)1C(C(=O)O)C	8398712	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.785	73.76883	472.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C@C@H(C@C@H)1C(C(=O)O)C							
Q(C1ccc(cc1)C=C(C@C@H)2ccc(O)cc2)C=C(C)C	8417251/986694	2	NPD08_150-700_iconfast.lib	BAMS22	0.67	65.74699	266.34
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8460331	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.834	73.595276	456.71
Q(C@C@H)1C@C@H(O)C(C@C@H)3C@H(C@C@H)4C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8460710	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.746	73.049034	472.71
Q(C@C@H)1C@C@H(O)C(C@C@H)2C(C@C@H)1C(C(=O)O)C	8463332	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.865	73.48217	504.71
Q(C@C@H)2C(C@C@H)1C(C(=O)O)C	8512308	6	NPD08_150-700_iconfast.lib	BAMS22	0.811	55.60156	364.53
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8525799	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.78	73.61211	458.73
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8526428	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.75	72.42083	472.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8526533	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.837	73.61847	454.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8526955	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.816	73.46363	486.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8527153	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.829	73.63482	454.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8527724	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.755	73.14012	488.71
S(=O)(=O)O(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8527932/903063	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.756	73.14694	536.77
S(=O)(=O)O(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8528297/102278	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.714	73.62187	552.77
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8528787	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.875	73.48966	502.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8528943	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.742	72.69005	504.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8529098	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.826	73.208824	502.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8529372	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.846	73.2558	504.71
O=C(O)C(C@C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8598292	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.831	72.86453	470.69
O=C1ccc(cc1)C(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8604245	3	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.69	73.738594	592.82
Q(C1c(O)ccc(c1)C(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8604678	3	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.659	73.60799	622.84
O=C1ccc(cc1)C(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8641329	1	NPD08_150-700_iconfast.lib	BAMS22	0.548	46.20236	342.35
Q(C@C@H)2C(C@C@H)1C(C(=O)O)C	8655474	6	NPD08_150-700_iconfast.lib	BAMS22	0.819	55.97446	290.4
O=C1ccc(cc1)C(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8659892	7	NPD08_150-700_iconfast.lib	BAMS22	0.932	65.38923	416.65
O=C1ccc(cc1)C(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8660649	7	NPD08_150-700_iconfast.lib	BAMS22	0.936	56.910458	436.59
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8664418	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.879	72.55895	498.75
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8664524	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.789	73.72769	486.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8665089	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.902	71.77415	488.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8669858	4	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.589	73.82988	650.81
O=C2C(C@C@H)1C(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8714002	6	NPD08_150-700_iconfast.lib	BAMS22	0.775	56.03509	302.46
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8739750	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.906	73.78884	454.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8741698	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.822	73.690704	470.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8743238	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.778	73.6212	474.73
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8745972	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.725	72.96623	500.68
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8803375	7	NPD08_150-700_iconfast.lib	BAMS22	0.846	56.097946	364.53
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8815460	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.736	73.123985	504.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8815762	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.761	73.152695	486.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8815887	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.795	73.421745	488.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8817194	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.809	73.87523	504.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8817646	3	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.612	73.3272	586.86
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8819277	3	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.584	73.776215	618.85
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8896569	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.768	73.143776	486.69
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8958154	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.929	73.23148	486.74
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8962642	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.695	73.41113	570.81
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8963574	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.577	72.17371	586.81
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8964239	3	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.635	72.090225	634.85
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	89783-66.4	1	phkog_db_20160621.lib	BAMS22	0.726	56.72	382.5
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	8994366	6	NPD08_150-700_iconfast.lib	BAMS22	0.707	65.479416	274.4
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9020734	7	NPD08_150-700_iconfast.lib	BAMS22	0.878	46.931232	398.5
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9021569	6	NPD08_150-700_iconfast.lib	BAMS22	1.107	66.44118	388.59
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9025957	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.731	72.17782	472.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9028094	9	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.942	73.31788	472.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9028977	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.797	73.23197	488.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9029568	8	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.802	73.60658	488.71
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9029975	8	NPD08_150-700_iconfast.lib	BAMS22	0.837	66.02458	490.72
Q(C@C@H)5C(C@H)4C@C@H(C@C@H)3C@H(C@C@H)2C(C@C@H)1C(C(=O)O)C	9033527	5	NPD08_150-700_iconfast.lib	Model12.3.2.8	0.627	73.66069	604.82

[illegible]

	205/07900063						
<chem>O2c(c1ccc([N+](=O)[O-])ccc1)ccc2C=NNC(=O)CC#N</chem>	AG-205/07903006	2	Specs_SC_Apr2016.ldb	BAMS22	0.727	45.579437	298.26
<chem>Clc1c(cc(F)c(F)c1)C(=O)N3CCN(c2c(F)cc(cc2)C(=O)C)CC3</chem>	AG-205/12140277	2	Specs_SC_Apr2016.ldb	BAMS22	0.806	46.323673	396.8
<chem>Clc3c(N2CCN(C(=O)c1c(C)cc(F)c(F)c1)CC2)ccc([N+](=O)[O-])c3</chem>	AG-205/37082045	2	Specs_SC_Apr2016.ldb	BAMS22	0.86	47.828632	416.21
<chem>O(c1ccc(cc1)C(=O)C)C(=O)c2cc(OC)ccc2</chem>	AG-205/40022551	2	Specs_SC_Apr2016.ldb	BAMS22	0.686	45.34777	270.28
<chem>O=C(c1ccc(N=[N+]=[N-])cc1)c2ccc(N=[N+]=[N-])cc2</chem>	AG-690/11763435	2	Specs_SC_Apr2016.ldb	BAMS22	0.676	46.26647	264.25
<chem>O=[N+](=[O-])c1c(ccccc1)CNc2c(cc([N+](=O)[O-])cc2)C</chem>	AG-690/11765455	2	Specs_SC_Apr2016.ldb	BAMS22	0.709	46.33618	287.27
<chem>FC(F)F)c1ccc(cc1)CNCc2c(OC)c(OC)ccc2</chem>	AG-690/11822088	2	Specs_SC_Apr2016.ldb	BAMS22	0.754	45.08255	325.33
<chem>Clc1c(cc([N+](=O)[O-])cc1)C=Nlc2ccc(N(C(=O)C)C)cc2</chem>	AG-690/11961546	2	Specs_SC_Apr2016.ldb	BAMS22	0.724	46.494797	331.76
<chem>O=[N+](=[O-])c1c(c(ccc1)C)Oc2cc(OC)ccc2=O)C</chem>	AG-690/12136597	2	Specs_SC_Apr2016.ldb	BAMS22	0.762	45.44747	287.27
<chem>s1c(=S)n(c(=O)c1=C/c3oc(c2c(ccc([N+](=O)[O-])c2)C)cc3)CC=C</chem>	AG-690/36159012	2	Specs_SC_Apr2016.ldb	BAMS22	0.603	45.491932	386.45
<chem>O=[N+](=[O-])c4cc(Oc2cc1c(=O)n(c(=O)c1cc2)c3cnccc3)ccc4</chem>	AG-690/36880041	2	Specs_SC_Apr2016.ldb	BAMS22	0.664	45.442375	361.31
<chem>s1c(nc2c1cc([N+](=O)[O-])cc2)NC(=O)C(F)C(F)C(F)C(F)F</chem>	AG-690/36932002	2	Specs_SC_Apr2016.ldb	BAMS22	0.69	55.208645	391.22
<chem>O=[N+](=[O-])c2cnc(Oc1ccc(cc1)C=C(#N)C#N)cc2</chem>	AG-690/37099161	2	Specs_SC_Apr2016.ldb	BAMS22	0.687	46.49591	292.25
<chem>O[C@@]4([C@@]1([C@@H]([C@H]3[C@H]([C@@H](O)C1)[C@@]2(C=CC(=O)CC2)CC3)CC4)C(C(=O)COC(=O)C)C(=O)C</chem>	AH-034/06873032	7	Specs_SC_Apr2016.ldb	BAMS22	0.843	45.695194	404.5
<chem>O=[N+](=[O-])c1c(O)c(ccc1)CNC2c(OC)cc([N+](=O)[O-])cc2</chem>	AH-034/34967061	2	Specs_SC_Apr2016.ldb	BAMS22	0.595	46.548428	319.27
<chem>s1c(nc2c1cc(OC)cc2)NC(=O)CSc3nc(O)cc(n3)C</chem>	AI-020/37280139	2	Specs_SC_Apr2016.ldb	BAMS22	0.877	57.088604	362.43
<chem>Clc3c(Oc2cc1cc(sc1cc2)=O)c(C)cc([N+](=O)[O-])c3</chem>	AI-204/31754049	2	Specs_SC_Apr2016.ldb	BAMS22	0.691	46.889	358.15
<chem>s1sc3c(c1=S)c4c(N(C(=O)CSsc2sc(nn2)C)C3(C)C)ccccc4</chem>	AJ-292/13093072	2	Specs_SC_Apr2016.ldb	BAMS22	0.736	55.205914	437.65
<chem>o1c(nc2ncccc12)c4c(c(NC(=O)c3cc([N+](=O)[O-])ccc3)ccc4)C</chem>	AK-778/37026068	2	Specs_SC_Apr2016.ldb	BAMS22	0.719	45.762478	374.36
<chem>O(c1c(c(ccc1)C#N)C#N)c3ccc(Oc2c(c(ccc2)C#N)C#N)cc3</chem>	AK-918/13072013	2	Specs_SC_Apr2016.ldb	BAMS22	0.644	46.20976	362.35
<chem>O=C(OC)c2ccc(NC(=O)[C@@H]1[C@@H]1C(=O)O)CC=CC1)cc2</chem>	AK-918/37074025	1	Specs_SC_Apr2016.ldb	BAMS22	0.873	45.3101	303.31
<chem>O(c1c(c(ccc1)C#N)C#N)c2ccc(NC(=O)C)cc2</chem>	AK-918/41206569	2	Specs_SC_Apr2016.ldb	BAMS22	0.772	46.328426	277.28
<chem>s1c(nc2c1cc([N+](=O)[O-])cc2)NC(=O)C(F)C(F)C(F)C(F)C(F)C(F)C(F)F</chem>	AK-968/11180151	2	Specs_SC_Apr2016.ldb	BAMS22	0.801	56.499794	469.24
<chem>O([C@@]4([C@@]3([C@@H]([C@H]2[C@@H]([C@@H](O)C1(C=CC(=O)C)CC2)C)C(=O)C3)CC4)C(C(=O)COC(=O)C)C(=O)C</chem>	AK-968/12117376	7	Specs_SC_Apr2016.ldb	BAMS22	1.111	46.352074	442.51
<chem>s1c(nc2c1cc([N+](=O)[O-])cc2)NC(=O)CSc3nc(cc(n3)C)C</chem>	AK-968/15361425	2	Specs_SC_Apr2016.ldb	BAMS22	0.691	46.030205	375.43
<chem>BrC3(Br)[C@]([C](=O)NN=Cc1n(ccc1)c2ccc([N+](=O)[O-])ccc2)(C3)C</chem>	AK-968/15605196	2	Specs_SC_Apr2016.ldb	BAMS22	0.777	46.207684	470.13
<chem>Clc1c(sc2c1ccc([N+](=O)[O-])cc2)C(=O)Nc3cc(OC)ccc3</chem>	AK-968/15605660	2	Specs_SC_Apr2016.ldb	BAMS22	0.736	45.27977	362.79
<chem>O1[C@H]2[C@H]([C@@H]([C@H]1C=C2)C(=O)Nc3ccc(cc3)C(=O)OCC)C(=O)O</chem>	AK-968/37173264	1	Specs_SC_Apr2016.ldb	BAMS22	0.665	45.88379	345.35
<chem>Clc1c(NC(=O)C(F)C(F)C(F)C(F)C(F)C(F)cc([N+](=O)[O-])cc1</chem>	AK-968/40282009	1	Specs_SC_Apr2016.ldb	BAMS22	0.734	55.449028	418.6
<chem>Clc1c(nn(c1C)Cc2oc(cc2)C(=O)Nc3c(cc([N+](=O)[O-])cc3)C)[N+](=O)[O-]</chem>	AK-968/40707858	2	Specs_SC_Apr2016.ldb	BAMS22	0.661	47.26507	419.78
<chem>Clc1c(nn(c1C)Cc2oc(cc2)C(=O)NN=Cc1c(O)ccc(C)cc1)[N+](=O)[O-]</chem>	AK-968/40709719	2	Specs_SC_Apr2016.ldb	BAMS22	0.515	56.826668	438.23
<chem>Clc1c(sc2c1ccc([N+](=O)[O-])cc2)C(=O)Nc3c(OC)ccc(OC)c3</chem>	AK-968/40735036	2	Specs_SC_Apr2016.ldb	BAMS22	0.702	46.00433	392.82
<chem>Clc1c(cc([N+](=O)[O-])cc1)C(=O)NCC2cn(nc2)C</chem>	AK-968/41026577	1	Specs_SC_Apr2016.ldb	BAMS22	0.701	45.08394	294.7
<chem>o1c(ccc1COc2ccc(cc2)C#N)C=O</chem>	AK-968/41170437	2	Specs_SC_Apr2016.ldb	BAMS22	0.675	46.035206	227.22
<chem>Fc1c(F)c(F)c(F)c1F)Cn3nc(c(NC(=O)c2n(nc2)C)c3C)C</chem>	AK-968/41923428	1	Specs_SC_Apr2016.ldb	BAMS22	0.741	45.942772	399.32
<chem>O=C(OC)c1ccc(NC(=O)CC[C@@H]1C(=O)O)C)cc1</chem>	AL-281/42289098	1	Specs_SC_Apr2016.ldb	BAMS22	0.627	45.242447	279.29
<chem>o1c(=O)ccc(c1)[C@@H]5[C@@]14([C@@]([O]([C@H]3[C@@H]1[C@@]2([C@@H]([C@@H](O)CC2)C3)CC4)CC5)C</chem>	AM-721/20741010	6	Specs_NP_Apr2016.ldb	BAMS22	0.775	56.792778	386.53
<chem>s2c1nc(cc(c1c(N)c2C(=O)c3c(F)cc(F)cc3)C(F)F</chem>	AM-807/14146976	2	Specs_SC_Apr2016.ldb	BAMS22	0.706	55.890507	372.31
<chem>s1c(nc2c1cc(OC)cc2)NC(=O)C(C#N)=C/c3sccc3</chem>	AM-896/36458020	2	Specs_SC_Apr2016.ldb	BAMS22	0.746	55.913906	341.41
<chem>s1c([N+](=O)[O-])ccc1C=Nlc2cc(ccc2)C#N</chem>	AN-329/10733059	2	Specs_SC_Apr2016.ldb	BAMS22	0.64	46.524788	257.27
<chem>s1c([N+](=O)[O-])ccc1C=NNc2c([N+](=O)[O-])ccccc2</chem>	AN-329/11481787	2	Specs_SC_Apr2016.ldb	BAMS22	0.745	46.47216	292.27
<chem>S=C(n2nc(c=N1Nc1c(OC)cc([N+](=O)[O-])cc1)c2=O)C)N</chem>	AN-329/40226744	1	Specs_SC_Apr2016.ldb	BAMS22	0.703	46.227753	336.33
<chem>Clc2c(N1N=c11c(=O)n(ncl1C)C(S)N)ccc([N+](=O)[O-])c2</chem>	AN-329/40226746	1	Specs_SC_Apr2016.ldb	BAMS22	0.686	46.00647	340.75
<chem>Clc3c(N1N=c11c(=O)n(ncl1C)c2scc(n2)C(C)C)c(C)cc([N+](=O)[O-])c3</chem>	AN-329/40484290	2	Specs_SC_Apr2016.ldb	BAMS22	0.636	46.229275	455.32
<chem>S(=O)(=O)(N(C)C)c2ccc(NC(=O)Cc1ccc(OC)cc1)cc2</chem>	AN-329/42260884	2	Specs_SC_Apr2016.ldb	BAMS22	0.756	45.140068	348.42
<chem>O(c1ccc(cc1)CC(=O)Nc2ccc(cc2)C(=O)NCCC)C</chem>	AN-329/43448654	2	Specs_SC_Apr2016.ldb	BAMS22	0.715	45.195778	326.4
<chem>O=C(OC)c1ccc(cc1)C(=O)Nc2cc(ccc2)C(=O)C</chem>	AN-329/43449172	2	Specs_SC_Apr2016.ldb	BAMS22	0.708	45.552948	297.31
<chem>S(=O)(=O)(N2CCN(c1c(cc([N+](=O)[O-])cc1)C(F)C(F)CC2)c3ccccc3</chem>	AN-465/13571007	2	Specs_SC_Apr2016.ldb	BAMS22	0.715	46.046608	415.39
<chem>O(c1c(OC)cccc1)CNCc2ccc(cc2)C(=O)O)C</chem>	AN-465/42888866	2	Specs_SC_Apr2016.ldb	BAMS22	0.655	45.278004	301.34
<chem>FC(F)F)c3cc(c1oc(cc1)C)NCC[C@@H]2OCC2)ccc3</chem>	AN-465/43384126	2	Specs_SC_Apr2016.ldb	BAMS22	0.774	46.13495	325.33
<chem>FC(F)F)c2cc(c1oc(cc1)C)NCCO)ccc2</chem>	AN-465/43411358	2	Specs_SC_Apr2016.ldb	BAMS22	0.783	46.16368	285.26
<chem>FC(F)F)c3cc(c1oc(cc1)C)Nc2[nH]n(c(n2)C)ccc3</chem>	AN-465/43411399	2	Specs_SC_Apr2016.ldb	BAMS22	0.783	46.226334	322.29
<chem>Clc2c(cc(c1oc(cc1)C)NCCO)cc2)C(=O)O</chem>	AN-465/43479890	2	Specs_SC_Apr2016.ldb	BAMS22	0.772	46.466225	295.72
<chem>O=[N+](=[O-])c1n(c(nc1C)CC(=O)Nc3ccc(n2nnnc2)cc3</chem>	AN-648/15099007	2	Specs_SC_Apr2016.ldb	BAMS22	0.602	46.440475	328.29
<chem>s2c(n1nc(c1N)C#N)C)nc3c2cc(OC)cc3</chem>	AN-668/41692987	2	Specs_SC_Apr2016.ldb	BAMS22	0.763	45.888783	285.33
<chem>lc2c(OS(=O)(=O)c1ccc(NC(=O)C)cc1)c(OC)cc(c2)C=O</chem>	AN-698/40847893	2	Specs_SC_Apr2016.ldb	BAMS22	0.573	45.44247	475.25
<chem>O([C@H]5[C]([C@H]4[C@@]([C@@H]3[C@]([C@@H]1([C@@H]([C@H]2[C@@]([C(=O)O)1)CCC(C2)C)C</chem>	AN-	5	Specs_NP_Apr2016.ldb	Model12.3.2.8	0.715	73.518875	588.83

]](CC1)CC(C2)(C)C=CC3(C)(CC4(C)CC5(C))C/C						
0=C[O-]](C@)45(C@H)[(C@)3(C@@)[(C@H)2(C@)[(C@H)1(C@@)[(C@H)(C=O)CC1(C)(CC2(C)(CC3(C)(C4(C)C)CC(C5(C)C)C	ZINC32022435	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.826	72.75516 455.7
Q(C@H)5C(=O)C(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC(C@H)4C5(C)C)C/C	ZINC32023021	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.779	73.38298 485.68
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC32036901	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.939	73.29635 471.7
Q(C@H)2(C@H)3(C@H)1(C@H)[(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC33830320	7	TCM_DB_iconfast.lib	BAMS22	0.94	46.929604 360.49
C(C@H)1(C@H)2(C@H)3(C@H)4(C@H)5(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC33830686	6	TCM_DB_iconfast.lib	BAMS22	1.2	58.25777 388.55
o1c(=O)ccc4c1cc2ccc2c4O1C=C(C)[C@H]3O(C@H)3(C=C)C/C	ZINC33832269	2	TCM_DB_iconfast.lib	BAMS22	0.591	55.672173 352.39
Oc1ccc(cc1)C(Cc2ccc(O)c2)=C=C	ZINC33834000	2	TCM_DB_iconfast.lib	BAMS22	0.704	65.621414 252.31
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC33902655	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.9	73.63353 467.67
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC33986139	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.794	72.70326 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC35261408	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.812	73.557655 485.68
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC35261409	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.83	73.52078 467.67
Q(C@H)1(C@H)O[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC35570683	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.765	73.691635 487.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC37538785	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.796	73.01088 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38140514	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.819	72.69134 455.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38140516	9	NuBBE_All_ZINC_iconbest.lib	Model12.3.2.8	0.823	73.09325 455.7
Q(C@H)2(C@H)3(C@H)1(C@H)O[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38140520	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.79	72.846794 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38143792	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.907	73.19396 487.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38143794	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.825	73.23244 487.7
Q(C@H)1(C@H)O[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38143825	5	TCM_DB_iconfast.lib	Model12.3.2.8	0.639	73.23323 603.82
Q(C@H)1(C@H)O[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38154869	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.762	71.53457 503.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38349969	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.806	73.08323 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38454010	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.779	73.02791 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38545497	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.96	73.3449 455.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC38605859	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.849	73.65486 471.7
Oc1c(O)ccc6(c1)C=C(C=C)C(C)C2ccc(O)c2Occc2	ZINC38820666	2	TCM_DB_iconfast.lib	BAMS22	0.79	65.40056 326.35
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC39036211	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.716	72.82192 471.7
Q(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC39107940	9	TCM_DB_iconfast.lib	BAMS22	1.284	66.38156 440.71
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC39560074	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.842	73.03274 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC40164365	8	TCM_DB_iconfast.lib	BAMS22	1.233	67.57222 452.68
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC40766841	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.925	73.4678 471.7
Q(C@H)7(C@H)6(C@H)5(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC42834192	6	TCM_DB_iconfast.lib	BAMS22	1.125	56.199306 400.51
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC42852273	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.837	73.21337 503.7
Q(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC43352180	7	TCM_DB_iconfast.lib	BAMS22	1.203	56.928356 407.57
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC43552895	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.875	72.95514 487.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC43574732	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.758	73.16277 487.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC44281721	5	TCM_DB_iconfast.lib	Model12.3.2.8	0.703	73.64318 587.82
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC49832948	5	TCM_DB_iconfast.lib	Model12.3.2.8	0.66	71.607346 630.82
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC49841444	5	TCM_DB_iconfast.lib	Model12.3.2.8	0.689	73.15884 603.82
Q(C@H)6(C@H)5(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC56871284	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.806	72.955574 439.66
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59586359	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.84	73.01973 439.66
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59586429	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.887	73.63519 455.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59587701	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.914	73.41222 457.67
O2C(C@H)O(C@H)1(C@H)O[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59588060	7	TCM_DB_iconfast.lib	BAMS22	0.758	45.712124 386.44
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59588111	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.857	72.48298 423.62
O(c1ccc(cc1)C=C(C)C)C2ccc(O)c2C=C	ZINC59588240	2	TCM_DB_iconfast.lib	BAMS22	0.644	65.67165 266.34
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59589031	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.834	72.61668 455.66
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59589370	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.852	73.33588 455.66
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC59779517	5	TCM_DB_iconfast.lib	Model12.3.2.8	0.741	73.52775 529.74
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70450903	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.899	72.94411 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70450999	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.813	73.39775 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70451075	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.875	73.039 487.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70451082	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.864	72.75717 471.7
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70451085	9	TCM_DB_iconfast.lib	BAMS22	1.168	66.500694 462.76
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70454002	8	TCM_DB_iconfast.lib	Model12.3.2.8	0.801	72.90854 503.7
Q(C@H)4C(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70454261	7	TCM_DB_iconfast.lib	BAMS22	0.749	56.365463 362.51
Q(C@H)5C[(C@H)4(C@H)3(C@H)2(C@H)1(C@H)O[(C@H)CC(C@)1(C=O)O-]](CC2(C)C=CC3(C)(CC4(C)C)CC5(C)C/C	ZINC70454411	9	TCM_DB_iconfast.lib	Model12.3.2.8	0.766	72.25876 455.7

[illegible]

