Diss. ETH No. 25868

# De Novo Design of Natural Product Mimetics by Reaction-Based Molecule Assembly and Generative Deep Learning

LUKAS FRIEDRICH

2019

DISS. ETH NO. 25868

# De Novo Design of Natural Product Mimetics by Reaction-Based Molecule Assembly and Generative Deep Learning

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

LUKAS FRIEDRICH

MSc ETH in Chemistry,
ETH Zurich

born on 08.03.1991

Citizen of Germany

accepted on the recommendation of

**Prof. Dr. Gisbert Schneider - examiner**
**Prof. Dr. Jonathan Hall - co-examiner**

2019

Cover picture was created with the help of Dr. Joachim Schnabl. The background photograph of *Phyllanthus engleri* was taken by Bart Wursten.

*For my partner Louisa, my parents Martina and Volker and my sister Anabell.*

*"Study nature, love nature, stay close to nature. It will never fail you."*

Frank Lloyd Wright

# Publications

Parts of this thesis are based on manuscripts which are in preparation to submit, submitted or published in:

- **L. Friedrich**, M. Stöckli, P. Schneider, A. Koeberle, O. Werz, G. Schneider, "Automated *De Novo* Design of Natural Product Mimetics: An Application to Galanthamine", *MedChemComm*, submitted on February 6$^{th}$ 2019.
- **L. Friedrich**, V. Bobinger, D. Merk, A. Koeberle, O. Werz, G. Schneider, "Fully Computer-guided "Design-Make-Test Cycle": Computational *De Novo* Design of Novel Bioactive (±)-Marinopyrrole A Mimetics", in preparation.
- **L. Friedrich**, R. Byrne, M. Mederos-Schnitzler, A. Treder, I. Singh, C. Bauer, T. Gudermann, U. Storch, G. Schneider, "Shape Similarity by Fractal Dimensionality: An Application in *De Novo* Natural Product Mimetics", *Communications Chemistry*, submitted on February 17$^{th}$ 2019.
- D. Merk, **L. Friedrich**, F. Grisoni, G. Schneider, "*De Novo* Design of Bioactive Small Molecules by Artificial Intelligence", *Molecular Informatics*, **2018**, *37*, 1700153.
- D. Merk, F. Grisoni, **L. Friedrich**, G. Schneider, "Tuning artificial intelligence on the *de novo* design of natural-product-inspired retinoid X receptor modulators", *Communications Chemistry*, **2018**, **1**, 68.

# Contents

# Acknowledgements

This thesis, which is a result of the past three years at the ETH Zurich, would not have been possible without the help and support of many people. At this point, I would like to give many thanks to all my colleagues, friends and family members who accompanied me on my path and assisted me during the last years.

First and foremost, I owe my deepest gratitude to my supervisor Prof. Dr. Gisbert Schneider for giving me the opportunity to conduct my research in his computer-assisted drug design group at ETH Zurich. Your devotion to science and your inspiring visions of the future of drug discovery has had a lasting effect on my own scientific thinking. The freedom you gave me to conduct my research and all the fascinating discussions about drug design have shaped my perspective as scientist in a sustainable way. Thank you Gisbert.

I would like to thank Prof. Dr. Jonathan Hall and Prof. Dr. Gunnar Jeschke for taking the time to co-examine this thesis.

My special thanks goes to my colleagues and friends Dr. Daniel Merk and Dr. Francesca Grisoni for all their contributions and fruitful discussions about artificial intelligence in drug discovery. It was a great honor to work with you. I would also like to thank my colleague Ryan Byrne for his support regarding chemoinformatics and the application of his fractal dimension descriptor. Moreover, I would like to thank Dr. Petra Schneider and my former colleague Marco Stöckli for their contributions to the galanthamine project. I would like to express my extraordinary gratitude to my two former students Veronika Bobinger and Arpad Dunai for all their computational and experimental work.

I am particularly grateful to my collaborators Dr. Andreas Koeberle and Prof. Oliver Werz from the university of Jena for their efforts regarding the cyclooxygenase activity measurements and their scientific contributions to the galanthamine and marinopyrrole A projects. I would like to thank Prof. Michael Mederos y Schnitzler, Dr. Ursula Storch, Aaron Treder, Inderjeet Singh and Thomas Gudermann for contributing to the englerin project and giving me insights into the fascinating research area of transient receptor potential ion channels.

I would like to express my gratitude to Dr. Jan Hiss for his valuable and honest feedback to during all of the scientific presentations and his invaluable help with the laboratory equipment. My warm thanks go to Sarah Haller for her technical support and assistance in the chemistry laboratory. I would like to thank Dr. Christian Steuer and his team for their advice in terms of the analytic methods. Many thanks go to Dr. Bernhard Pfeiffer and Philipp Waser for their help regarding the NMR experiments.

I would like to thank Dr. Joachim Schnabl for his support in creating the cover picture and Bart Wursten for allowing me to use his photograph of *Phyllanthus engleri* as background picture (www.zimbabweflora.co.zw).
I am more than grateful to all present and former members of the computer-assisted drug design group at ETH Zurich for all the unforgettable moments over the past three years. It was a great pleasure to work, discuss, suffer and celebrate together. The last three years were a wonderful time that I wouldn't have wanted to miss. Especially, I would like to thank the former group members Dr. Arndt Finkelmann, Dr. Daniel Reker, Dr. Tiago Rodrigues and Dr. Christoph Bauer for all their support regarding the computational and synthetic challenges. I would like to express my great, personal gratitude to my former colleagues and friends Dominique Bruns, Claudia Neuhaus, Jens Alexander Fuchs and Dr. Alex Müller for their friendship, all the great discussions about science and the future of drug discovery, and all the memorable moments from other areas of life.
I would like to show my greatest appreciation to all my friends from the Zurich area and Germany who accompanied and supported me in the last years: Thanks to all of you. I am extremely grateful to be surrounded by such a supporting family. Special thanks go to Jutta and Reinhold, Rebecca and Marc, and my grandmother Leni for all of their endless support over the past years.
Finally, I would like to express my deepest heartfelt appreciation to my partner Louisa, my parents Martina and Volker and to my sister Anabell. Your unconditional love, your unlimited encouragement and your belief in me have made me the person I am today. Without you, I would not be where I am now.

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **°C** | degree Celsius |
| **1D** | one-dimensional |
| **2D** | two-dimensional |
| **3D** | three-dimensional |
| **4D** | four-dimensional |
| **Å** | angstrom [$10^{-10}$ m] |
| **AcOH** | acetic acid |
| **ADMET** | absorption, distribution, metabolism, excretion and toxicity |
| **AI** | artificial intelligence |
| **ANN** | artificial neural network |
| **BIOS** | biology-oriented synthesis |
| **CADD** | computer-assisted drug design |
| **CATS** | chemically advanced template search |
| **ChEMBL** | chemical database of bioactive molecules of the European Molecular Biology Laboratory |
| **DAG** | diacylglycerol |
| **DCC** | dicyclohexylcarbodiimide |
| **DLS** | dynamic light scattering |
| **DMAP** | 4-(dimethylamino)pyridine |
| **DNP** | Dictionary of Natural Products |
| **DOGS** | design of genuine structures |
| **DOS** | diversity-oriented synthesis |
| **DTS** | diverted total synthesis |
| **$EC_{50}$** | Half maximal effective concentration |
| **ECFP** | extended-connectivity fingerprint |
| **EDC** | 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide |
| **EtOAc** | ethyl acetate |
| **EtOH** | ethanol |
| **FBDD** | fragment-based drug discovery |
| **FD** | fractal dimension |
| **FDA** | Food and Drug Administration |
| **FOS** | function-oriented synthesis |
| **FP** | fingerprint |
| **h** | hour |
| **HOBt** | benzotriazol-1-ol |
| **HPLC** | high-performance liquid chromatography |
| **HTS** | high-throughput screening |
| **$IC_{50}$** | half maximal inhibitory concentration |
| **IUPAC** | International Union of Pure and Applied Chemistry |
| **$K_D$** | dissociation constant |
| **$K_i$** | dissociation constant of an inhibitor |
| **LBVS** | ligand-based virtual screening |
| **LE** | ligand efficiency |

| | |
|---|---|
| **LogP** | n-octanol-water partition coefficient |
| **LSTM** | long short-term memory |
| **M** | molar [g/mol] |
| **MACCS** | molecular access system |
| **min** | minute |
| **ML** | machine learning |
| **MMFF94** | Merck Molecular Force Field |
| **MOE** | molecular operating system |
| **MW** | molecular weight |
| **NCE** | new chemical entity |
| **$NEt_3$** | triethylamine |
| **NMR** | nuclear magnetic resonance spectroscopy |
| **NOESY** | nuclear overhauser enhancement and exchange spectroscopy |
| **NP** | natural product |
| **PCA** | principal component analysis |
| **QSAR** | quantitative structure-activity relationship |
| **RNN** | recurrent neural network |
| **Ro5** | Rule-of-Five |
| **S.E.M.** | standard error of mean |
| **SCONP** | structural classification of natural products |
| **SMILES** | simplified molecular-input line-entry system |
| **SPiDER** | self-organizing map-based prediction of drug equivalence relationships |
| **TBDPS** | tert-butyldiphenylsilyl |
| **Tc** | Tanimoto coefficient |
| **TFA** | trifluoroacetic acid |
| **THC** | tetrahydrocannabinol |
| **THF** | tetrahydrofuran |
| **µw** | microwave irradiation |
| **VS** | virtual screening |

# Summary

Natural products (NPs) offer an almost inexhaustible source of inspiration for the discovery of new drug-like molecules. Natural products and NP-related molecules represent a major fraction of approved small-molecule drugs. However, complex organic syntheses and limited supply often provide insufficient quantities for in-depth structure-activity relationship studies and further drug development. To overcome these obstacles, synthesis-driven concepts have utilized natural products as templates to design structurally simplified small molecules while retaining biological activity. A considerable synthetic effort and knowledge about activity-related structural features are necessary to obtain such mimetic structures.

Computational methods, in particular *de novo* design concepts, have been developed to design synthetically accessible small molecules with desirable biological activities. These methods utilize various molecular representations (descriptors) encoding predefined structural features. Together with similarity metrics, molecular descriptors have been employed to computationally evaluate the molecular similarity of virtual compounds to known bioactive templates. *De novo* design and *in silico* similarity assessment represents a powerful computational approach to identify novel bioactive molecules.

Although these computational concepts coupled with drug-like reference compounds have shown promising results, only few studies have employed natural products as templates in computational *de novo* design. Therefore, in this thesis we attempted at generating synthetically feasible bioactive natural product mimetics by computational *de novo* design.

In our first study, we applied the reaction-based molecule assembly method DOGS (Design Of Genuine Structures) to generate bioactive new chemical entities (NCEs) from the natural product galanthamine. We assessed the two-dimensional topological pharmacophore similarity to their template by the CATS descriptor (Chemically Advanced Template Search). Thereafter, we predicted potential targets with the software SPiDER (Self-organizing map–based Prediction of Drug Equivalence Relationships). Six simplified mimetic structures were obtained with an one-step reaction. Their activity on known and computationally predicted molecular targets was characterized *in vitro*. Two inhibitors of acetylcholinesterase were identified, the known primary target of galanthamine. The designed compounds possessed further potent activities against several molecular targets from various protein families. This behavior was further corroborated by computationally predicting their target promiscuity.

In a further application we utilized the aforementioned *de novo* design concept to generate mimetic structures of the marine natural product (±)-marinopyrrole A. We employed this anticancer agent as the template of the software DOGS, and evaluated

the pharmacophore similarity of the designs to their reference structure. Five mimetics were synthesized following the computationally proposed synthetic routes. The bioactivity of these compounds was assessed against predicted and known molecular targets. The mimetics shared similar activity profiles with their natural product template and showed comparable potencies on these targets. In certain cases, the mimetics exceeded the activity of their parent structure. These findings corroborated the underlying principle of our design approach that similar structures possess similar bioactivities. The observed bioactivities showed that the mimetics have inherited the target spectrum of their template structure.

In order to extend our design workflow with a three-dimensional similarity assessment, we employed the concept of fractal dimensionality to capture the shape of a molecule. Fractal dimensionality, as a novel numerical description of the surface of a 3D molecular conformation, has been applied to estimate the shape similarity between two molecules. In a prospective application, a set of *de novo* designs mimicking the natural product (-)-englerin A were ranked according to their similarity in shape to the natural product. Two compounds were selected which were similar in shape and computationally predicted as active for the transient receptor potential (TRP) ion channels, the known cellular targets of (-)-englerin A. Synthesis of the two mimetics and subsequent cell-based intracellular calcium level measurements and electrophysiological whole-cell experiments of TRPC4 and TRPM8 channels revealed potent inhibitory effects of one of the computer-generated compounds.

Finally, we handed the structure generation over to artificial intelligence by incorporating a generative deep learning model as novel *de novo* design method. In our first proof-of-principle study, we applied such a model to generate new drug-like molecules with desired bioactivities. In a two step procedure, a recurrent neural network (RNN) was first trained on the string representations (SMILES) of known bioactive molecules to capture their constitution. Second, this model was re-trained on a small set of active molecules sharing a similar molecular target set enabling us to generate a target-focused compound library. As fine-tuning set, we selected 25 known modulators of the retinoid X receptors (RXR) and/or peroxisome proliferator-activated receptors (PPARs). After sampling from the generative model, computational target prediction with SPiDER and similarity assessment were employed to rank the generated designs. We picked five high-ranked compounds for synthesis and *in vitro* characterization. Four designs possessed nano- to micromolar potencies on the intended molecular targets.

After this proof-of-principle study, we utilized a deep learning model to design natural product mimetics. We reused our model trained on drug-like compounds and fine-tuned it on six natural products with known agonistic effects on RXRs. Sampled designs were computationally ranked by the same two-step procedure leading to four natural product mimetics, which were then synthesized and characterized *in vitro*. Two

compounds showed micromolar activity on at least one RXR subtype. The results of both studies revealed that the model learned essential structural features from a set of bioactive molecules.

In summary, the results of this thesis highlight the capabilities of reaction-based molecule assembly and generative deep learning for *de novo* design of bioactive natural product mimetics. Coupled with *in silico* similarity assessment and computational target prediction, both *de novo* methods proved to be useful for identifying synthetically accessible small molecules with desired biological activities. However, generating synthetically feasible compounds and identifying bioactive compounds with computational tools in prospective applications are remaining challenges of both *de novo* drug design methods. Despite these limitations, the combination of computational *de novo* design and natural products in a knowledge-driven approach holds promise for future natural product-inspired drug discovery.

# Zusammenfassung

Naturstoffe sind eine nahezu unerschöpfliche Inspirationsquelle für die Entdeckung neuartiger Wirkstoffe. Ein grosser Teil der zugelassenen Wirkstoffe hat seinen Ursprung in Naturstoffen und naturstoffähnlichen Molekülen. Dennoch führen komplexe chemische Synthesen und eingeschränkte Versorgung durch Extraktion dazu, dass nicht genügend Material dieser Stoffe zur Verfügung steht, um umfangreiche Struktur-Wirkungsbeziehnung studien durchzuführen. Um diese Hürden zu meistern, sind verschiedene synthetische Methoden entwickelt worden, die Naturstoffe als Vorlagen nutzen, um strukturell vereinfachte Moleküle unter Beibehaltung ihrer gewünschten biologischen Aktivität zu entwerfen. Für solche Konzepte ist jedoch ein nicht zu unterschätzender synthetischer Aufwand nötig. Ausserdem ist es notwendig, dass aktivitätsrelevante strukturelle Elemente zuvor bekannt sind.

Unterschiedliche computerbasierte Methoden, insbesondere *de novo* Designkonzepte, sind entwickelt worden, um synthetisch zugängliche Moleküle mit wünschenswerten Bioaktivitäten zu konstruieren. Verschiedene molekulare Deskriptoren sind entwickelt worden, um die strukturellen Eigenschaften von Substanzen zu beschreiben. Zusammen mit Ähnlichkeitsmetriken, werden diese Deskriptoren benutzt, um virtuell erzeugte *de novo* Designs anhand ihrer Ähnlichkeit zur Vorlage auszuwählen.

Obwohl diese computergestützten Designmethoden mit wirkstoffähnlichen Strukturen als Vorlage bereits erste vielversprechende Resultate geliefert haben, sind erst wenige Studien bekannt, die Naturstoffe als Vorlagen verwendeten. Das Ziel meiner Arbeit war es, synthetisch zugängliche und biologisch aktive Naturstoffmimetika mit Hilfe von computerbasiertem *de novo* Design zu entwerfen.

In einer ersten Studie verwendeten wir die *de novo* Design Software DOGS (Design Of Genuine Structures), welche Moleküle auf Grundlage von implementierten Reaktionen virtuell zusammensetzt. Diese Software erlaubte es uns, Mimetika des Naturstoffs Galanthamin zu erzeugen. Um die Ähnlichkeit dieser Mimetika zu ihrer Vorlage zu berechnen, nutzten wir den Pharmacophor-basierten Deskriptor CATS (Chemically Advanced Template Search). Ausserdem verwendeten wir die Software SPiDER (Self-organizing map–based Prediction of Drug Equivalence Relationships) um mittels computergestützten Target-Vorhersagen potentiell aktive Moleküle auszuwählen. Sechs vereinfachte Strukturen wurden in einer einstufigen Synthese hergestellt und anschliessend auf ihre Aktivität in unterschiedlichen *in vitro* Untersuchungen getestet. Wir identifizierten zwei schwach aktive Inhibitoren der Acetylcholinesterase, eines der bekannten Haupttargets von Galanthamine selbst. Die vom Computer erzeugten Moleküle besassen noch weitere Aktivitäten gegenüber diversen Proteinen aus verschiedenen Proteinfamilien.

In einer zweiten Anwendung dieses Designkonzepts generierten wir Mimetika des marinen Naturstoffes (±)-Marinopyrrol A. Diese antikanzerogene Substanz wurde als

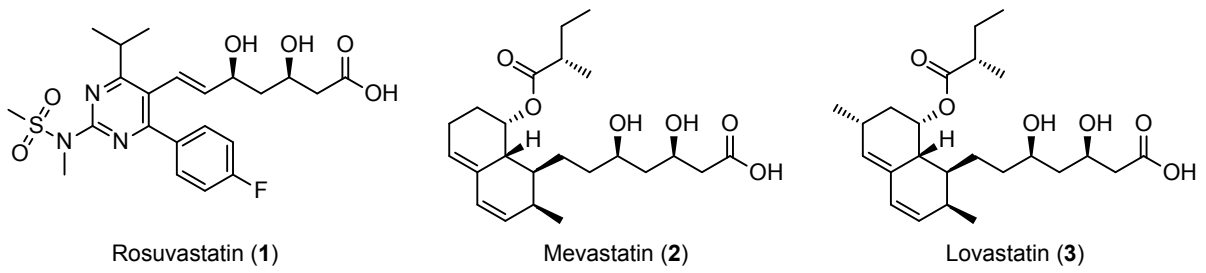
Vorlage für DOGS verwendet und die vom Computer erzeugten Strukturen wurden wie zuvor anhand ihrer Ähnlichkeit zur Vorlage sortiert. Fünf Moleküle konnten auf Grundlage der vorhergesagten virtuellen Synthese hergestellt werden. Die Bioaktivität dieser Substanzen wurde im Bezug auf bereits bekannte und vorhergesagte Targets gemessen. Die Mimetika zeigten ähnliche Aktivitätsprofile wie ihre Vorlage in einem vergleichbaren Aktivitätsbereich. In einigen Fällen übertrafen die gemessenen Aktivitäten der Mimetika die des Naturstoffes. Diese Beobachtungen untermauerten das zugrundeliegende Prinzip dieses Designkonzept, welches besagt, dass ähnliche Strukturen ähnliche Aktivitäten besitzen. Ausserdem zeigten die Resultate, dass die Mimetika das Targetspektrum ihrer Vorlage erbten.
Um in unserem Designprozess die Ähnlichkeit von Molekülen im dreidimensionalen Raum zu einzubauen, verwendeten wir das Konzept der fraktalen Dimension als neuartigen Deskriptor, welcher die dreidimensionale Form eines Moleküls beschreibt. Der auf fraktaler Dimension basierte Deskriptor ist eine numerische Beschreibung der Oberfläche eines Objekts und drückt aus, wie dieses Objekt den Raum um sich ausfüllt. Wir verwendeten diesen Deskriptor in einer prospektiven Studie, um Moleküle aus einer Bibliothek von computer-generierten Mimetika des Naturstoffes (-)-Englerin A anhand ihrer dreidimensionalen Ähnlichkeit auszuwählen. Ausserdem wurden mittels Target-Vorhersagen von SPiDER die Strukturen priorisiert, die als aktiv für die Transient Receptor Potential (TRP) Ionenkanäle vorhergesagt wurden. Zwei Moleküle wurden hergestellt und ihre Effekte auf TRPC4 und TPM8 wurden in elektrophysiologischen Experimenten getestet. Es zeigte sich, dass ein Molekül hemmende Wirkungen gegenüber diesen beiden Targets besitzt.
In zwei weiteren Studien benutzten wir Modelle aus dem Bereich der künstlichen Intelligenz zur Erzeugung von *de novo* Strukturen. In einer ersten Machbarkeitsstudie nutzten wir ein solches Modell, welches zuerst die Konstitution von wirkstoffähnlichen Molekülen lernte. Danach wurde das Modell auf einer Auswahl von Molekülen, welche ein gemeinsames Target besitzen, trainiert. Dieses sogenannte "Transfer Lernen" ermöglichte es uns, eine target-fokussierte Bibliothek von *de novo* Designs zu erzeugen. Wir verwendeten 25 Moleküle mit bekannten agonistischen Effekten auf Retinoid X Rezeptoren (RXRs) und/oder Peroxisomen-Proliferator-aktivierte Rezeptoren (PPARs). Die vom Modell generierten Strukturen wurden anhand ihrer Target-Vorhersagen und molekularen Ähnlichkeit zu bekannten Modulatoren ausgewählt. Fünf Moleküle wurden synthetisiert, wobei vier Designs nennenswerte Aktivitäten gegenüber den zwei Targets aufwiesen.
In einer Folgestudie nutzten wir das gleiche Modell, um Naturstoffmimetika zu generieren. Das bereits auf wirkstoff-ähnlichen Molekülen trainierte Modell wurde auf sechs Naturstoffen erneut trainiert. Diese sechs Naturstoffe wurden auf Grund ihres gemeinsamen Targets ausgewählt. Nach dem Design der Moleküle wurden diese wie

zuvor ausgewählt. Vier Naturstoffmimetika wurden hergestellt und zwei dieser Moleküle konnten als Agonisten von mindestens einem RXR Subtyp identifiziert werden. Die Resultate dieser beiden Studien zeigten, dass das verwendete Modell die relevanten strukturellen Eigenschaften von aktiven Molekülen gelernt hat und darauf aufbauend neue bioaktive Moleküle erzeugen konnte.
Die Resultate dieser Arbeit haben die Möglichkeiten von computer-gestütztem *de novo* Design gezeigt. Durch ein reaktionsbasiertes Konzept und eine Methode aus dem Bereich der künstlichen Intelligenz, zusammen mit der Beurteilung von molekularer Ähnlichkeit und computer-basierter Target-Vorhersage, war es möglich, neue synthetisch leichter zugängliche und bioaktive Substanzen zu erzeugen. Die Erzeugung von synthetisch zugänglichen Molekülen und die Identifizierung deren Bioaktivitäten mittels computerbasierten Methoden stellt vor allem in prospektiven Anwendungen weiterhin eine Herausforderung dar. Dennoch besitzt der Ansatz, computerbasierte Konzepte mit Naturstoffen zu kombinieren, das Potential, ein fester Bestandteil von zukünftigem naturstoff-inspirierten Wirkfstoffdesign zu werden.

# 1 Introduction

Natural products (NPs) provide an almost unlimited repository of inspirations for new drug-like molecules.[1–3] Various strategies to harness the full potential of natural products have been developed in the history of humankind. Both extraction from natural sources and synthesis of bioactive natural products have enabled the discovery of novel natural product-based drugs. Besides these two strategies, several synthetic and computational methods have been applied to obtain natural product-inspired bioactive molecules by simplifying natural product structures with retaining necessary structural features relevant for their bioactivity. These simplification strategies further aim to overcome synthetic difficulties, improve unfavorable properties and design mimetic structures of natural products with comparable or better biological activities.[4] The discovery of rosuvastatin (**1**), a drug for treating high levels of cholesterol,[5] represents an example, in which a drug molecule resembles two natural products, more precisely the natural products mevastatin (**2**) and lovastatin (**3**) (Figure 1.1). In case of natural products with unknown targets, most of the structural simplification methods"*mainly include step by step dissection of the complex structure, the elimination of redundant chiral centers, the reduction of the number of rings, and scaffold hopping*", as depicted by Wang and co-workers.[4] Especially the concept of "scaffold hopping", described as the "*identification of isofunctional molecular structures with significantly different molecular backbones*",[6] enables the development of simplified natural product mimetics by computer-assisted drug discovery.

This introductory chapter covers concepts and examples from the fields of natural product-inspired drug discovery and computer-assisted drug design (CADD). Various concepts combining natural products and computational methods, which aim to capture relevant features and design bioactive new chemical entities (NCEs) from natural product templates, will be highlighted.




Rosuvastatin (**1**) Mevastatin (**2**) Lovastatin (**3**)

**Figure 1.1:** Structures of the drug rosuvastatin (**1**) and the structurally related natural products mevastatin(**2**) and lovastatin(**3**).

## 1.1 Natural Products in Drug Discovery

The bitter tasting natural product quinine (**4**, Figure 1.2) is not only recognized as a flavor component in tonic water, it is also well-known in medicine as antimalarial drug. Although quinine was isolated in 1820 [7] and synthesized by Woodward and Doering in 1945,[8] it is still being investigated by researchers. Quite recently, scientists had identified a common binding target of quinine which will help to fully understand the mode of actions of this old natural antimalarial agent.[9] This latest example showcases that natural products are still an essential part of current research in the field of drug discovery and development dealing with the identification of novel drug candidates.
From the beginning of human history, mankind has explored nature's substances to treat all types of diseases. Relying on inherited knowledge from ancestors, the application of traditional medicine was common practice until the 19th century.[10] The oldest strategy to isolate naturally occurring medicinal substances was based on extraction. This traditional strategy included almost exclusively plant extracts in various forms, e.g. tinctures, soups or infusions. But most of the "ancient" drugs could not fulfill the modern definition of a drug including isolation and characterization of its chemical structure and its interactions with living organisms.[11] Morphine (**5**, Figure 1.2) is considered as one of the first isolated bioactive compound from a plant extract. In the early 1800s, Friedrich Sertürner isolated morphine as one pharmacologically active substance from the opium poppy *Papaver somniferum*.[12, 13] A few decades later, the German chemist Georg Merck discovered papaverine (**6**), another active compound in opium poppy extracts.[14] Further natural product drugs like atropin (**7**), codeine(**8**) or salicin(**9**) were isolated during the 19th century (Figure 1.2).[15] Due to major technical advances during the past decades, isolating pharmacologically active compounds from natural sources remains a prevalent part in the era of modern drug discovery.[16]

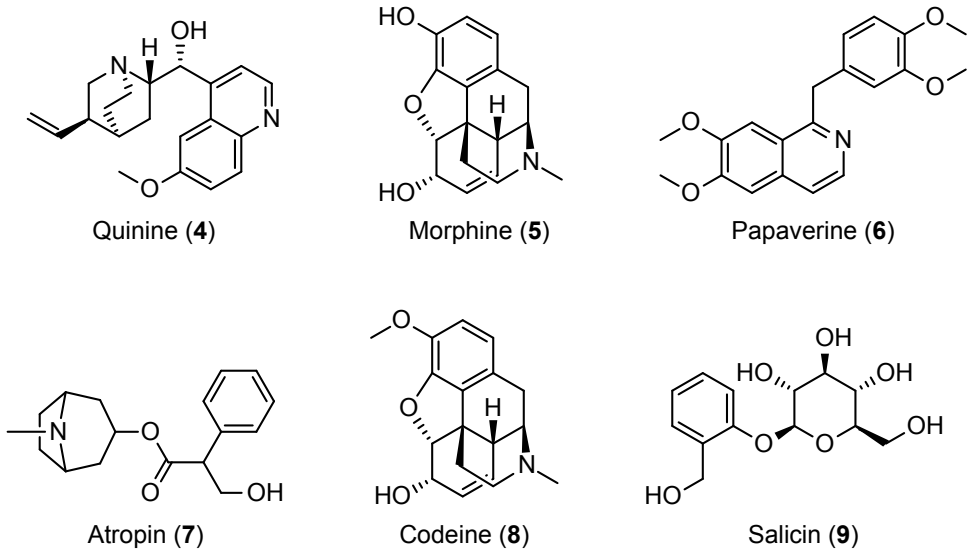



**Figure 1.2:** Several natural products isolated in the 19th century.

Apart from isolating active ingredients from plant extracts, chemical syntheses of natural products has become another common strategy to discover new drugs. Acetylsalicylic acid, an acetylated form of the natural product salicylic acid, was among the first synthesized natural product-derived drugs. In 1853, Charles Frederic Gerhardt synthesized acetylsalicylic acid,[17] and a few years later, Karl Kraut was able to obtain acetylsalicylic acid in a purer form.[18] At the end of the 19th century, the German company Bayer began to produce acetylsalicylic acid and marketed the natural product derivative under the trademark Aspirin©.[19] Since these first successful synthetic achievements, natural product synthesis has grown into a powerful tool in drug discovery, and is still nowadays one of the most important sources of natural-product drugs.[20, 21]

A third major category to harness natural products in drug discovery includes numerous simplification strategies.[4] Many synthesis-driven approaches have been utilized to overcome various challenges of natural products as drug molecules, like their limited supply from extraction and synthesis, insufficient structural properties or unfavorable pharmacological profiles. In many cases, the simplified natural product-inspired molecules comprise parts of their parent natural product or they contain relevant structural elements of the natural product pharmacophore.

To evaluate the importance of natural products in drug discovery, Newman and Cragg have investigates the origins of approved drugs from 1981 to 2014.[22] In their report, they analyzed all therapeutic agents, approved by the Food and Drug Administration (FDA) of the United States of America, to classify the different originating sources of these drugs. Newman and Cragg took into account all approved small-molecule drugs in this time period ($N$ = 1211), over 396 drugs (33%) are natural products, natural product derivatives or botanical defined mixtures. Further 395 approved molecules (32%) consist of either natural product pharmacophores or resemble the natural ligand of a protein. The remaining 420 drugs (35%) are solely synthetic compounds originating for example from high-throughput screening (HTS) campaigns. The proportion of natural products or NP-related drugs increases if only anticancer agents are taken into account. From over 136 approved anticancer drugs, 113 out of 136 drugs (83%) are natural products, derivatives or mimetics thereof. Only 23 compounds (17%) belong to the category of synthetic drug molecules.

## Natural Product, NP-Derivatives and NP-Mimetics as Drugs

To exemplify the three different categories, in particular natural products, NP-derivatives and NP-mimetics, illustrative examples of each category are mentioned.[22] The anticancer natural product paclitaxel (**10**), trade-marketed as Taxol©, has been approved in 1993 by FDA. Paclitaxel enhances tubulin polymerization and stabilizes microtubules against depolymerization resulting in an overall microtubules stabilization

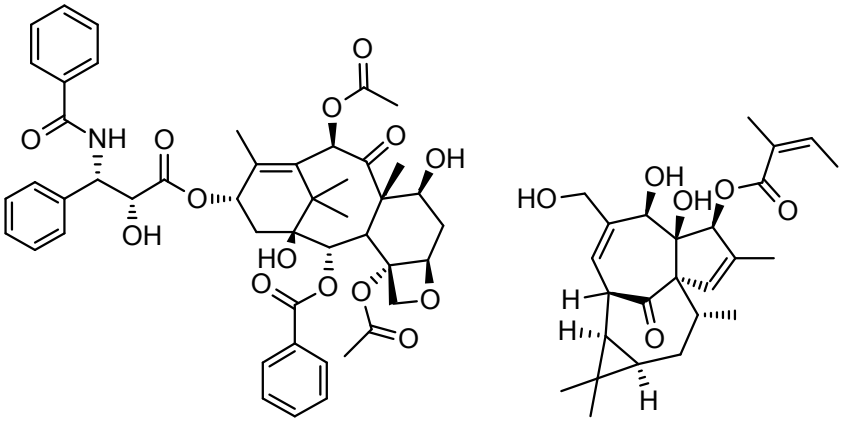



**Figure 1.3:** Examples of approved natural-product drugs. (*Left*) The natural product paclitaxel (**10**) has been approved in 1993 as chemotherapeutic agents. (*Right*) The plant-derived natural product ingenol mebutate (**11**), approved in 2012, is used to treat actinic keratosis.

(Figure 1.3, *left*).[23] In 2012, the natural-product drug ingenol mebutate (**11**) has been approved as a gel formulation for the treatment of actinic keratosis, a pre-cancerous area of skin (Figure 1.3, *right*).[24] Although the exact mechanism of action is unknown, it has been suggested that ingenol mebutate interacts with several protein kinase C (PKC) subtypes[25] leading to an activation of the related signaling pathway and finally inducing cell death.[26]

The proteosome inhibiting natural product epoxomicin (**12**)[27] served as starting point for the development of Carfilzomib (**13**), a selective proteosome inhibitor as well, representing an example from the category of natural product-derived drugs (Figure 1.4).[28] A third class of NP-related drugs are synthetic compounds mimicking a natural product or a natural substrate, whereas 334 out of 1211 approved drugs (28%) fall into this category.[22] As an example, natural occurring ligand 9-cis-retinoic acid (**14**, Figure 1.5), a form of vitamin A, activating the nuclear hormone receptors retinoid X receptor (RXR) and retinoic acid receptor (RAR)[29] has been exploited as template for the development of bexarotene (**15**, Figure 1.5), a selective RXR agonist with antineoplastic activity.[30] To capture the potencies and important features of bioactive natural products, further strategies have been developed in natural-product research. Crane

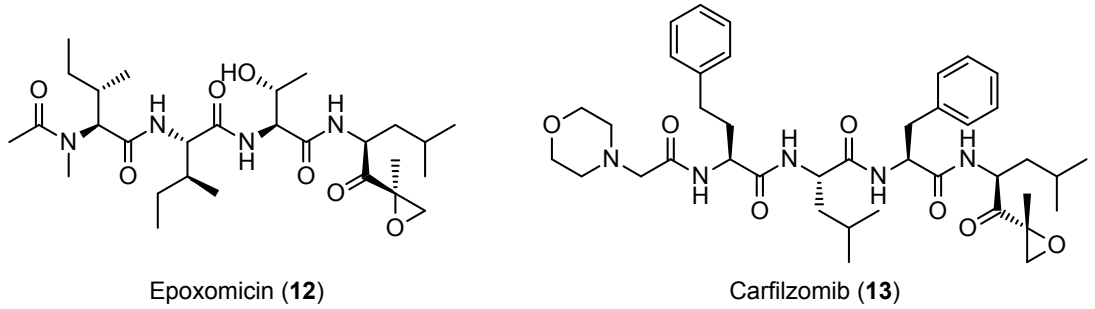
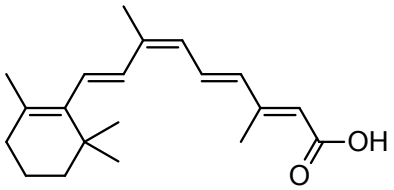
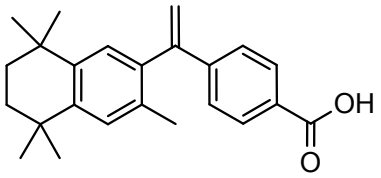



**Figure 1.4:** (*Left*) The proteosome inhibiting natural product epoxomicin (**12**) and (*right*) its FDA approved derivative carfilzomib (**13**).

and Gademann published a collection of case studies highlighting complex natural products which had been reduced to natural product fragments mimicking parts of their natural product template structure, while maintaining or increasing the pharmacological activity.[31] One such instance is (+)-migrastatin (**16**, Figure 1.6), a natural product which inhibits cell migration in cancer cells .[32] Besides the total synthesis of (+)-migrastatin,[33] the Danishefsky group investigated the activity of further natural product analogs (**17** and **18**) and identified natural product fragments with increased potencies compared to its template natural product (Figure 1.6).[34–36] The underlying concept exploited has been coined "diverted total synthesis" (DTS), introduced by Danishefsky.[37] This strategy of "molecular editing" allows the exploration of natural product analogs, starting from a synthetically advanced intermediate and transforming this intermediate into analogs with different complexity levels compared to the original natural product. This synthetic principle offers possibilities to make natural product analogs which are often inaccessible from the natural product itself because of missing synthetically feasible transformations. Mimicking natural products with chemical synthesis can also be achieved with designing functional rather then structural analogs. The main goal is to obtain isofunctional structures containing pharmacophoric features of their template structures. The so-called "function-oriented synthesis" (FOS) approach comprises of the design of NCEs with a focus on their function combined with synthetically economic methods.[38, 39] Wender *et al.* applied this synthetic strategy to design novel inhibitors of PKC by harnessing pharmacophoric features of phorbol esters, a set of

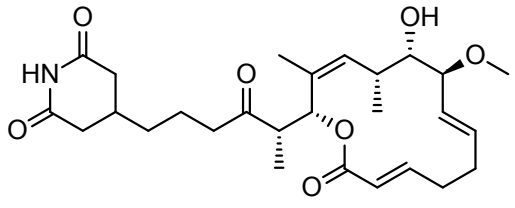



9-*cis*-Retinoic acid (**14**)

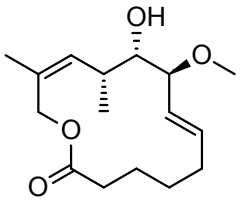



Bexarotene (**15**)

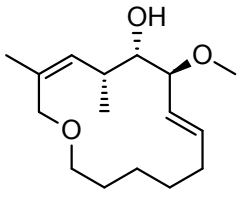
**Figure 1.5:** (*Top*) The natural ligand 9-cis-retinoic acid (**14**) has served as template for the mimetic bexarotene (**15**) (*bottom*).




(+)-Migrastatin (**16**)




Migrastatin analog (**17**)




Migrastatin analog (**18**)

**Figure 1.6:** (*Left*) Originating from (+)-Migrastatin (**16**), a cell migration inhibitor isolated from Streptomyces plantensis bacteria, more potent analogs (**17**, *middle*) and **18**, *right*) have been identified by reducing the original structural complexity.

natural carcinogenic agents.[40] The inclusion of the functional information gained from the natural ligand of PKC, diacylglycerol (DAG), resulted in the identification of new PKC modulators with competitive inhibition of phorbol ester binding. The Wender group investigated further natural products modulating PKC, in particular the marine natural product bryostatin 1 (**19**, Figure 1.7).[41] After publication of structural information about bryostatin 1,[42] analogs have been designed computationally by applying a similar protocol as before.[43] Bryostatin 1 and DAG were analyzed to identify shared pharmacophoric features. The first generation analog **20** (Figure 1.7) had comparable activity and could be obtained in less than 30 steps,[44] a shorter synthetic route than the first total synthesis of bryostatin 1 by Keck *et al.* which required 59 steps.[45] Further bryostatin analogs (**21** and **22**, Figure 1.7) were optimized for function (2nd generation analog **21**) and synthetic step-economy (3rd generation analog **22**).[41] Recently, Wu and Dockendorff utilized FOS and computational methods to mimic the antifungal natural agent sordarin with the goal to obtain accessible novel scaffolds which can be easily modified to improve their molecular properties.[46]

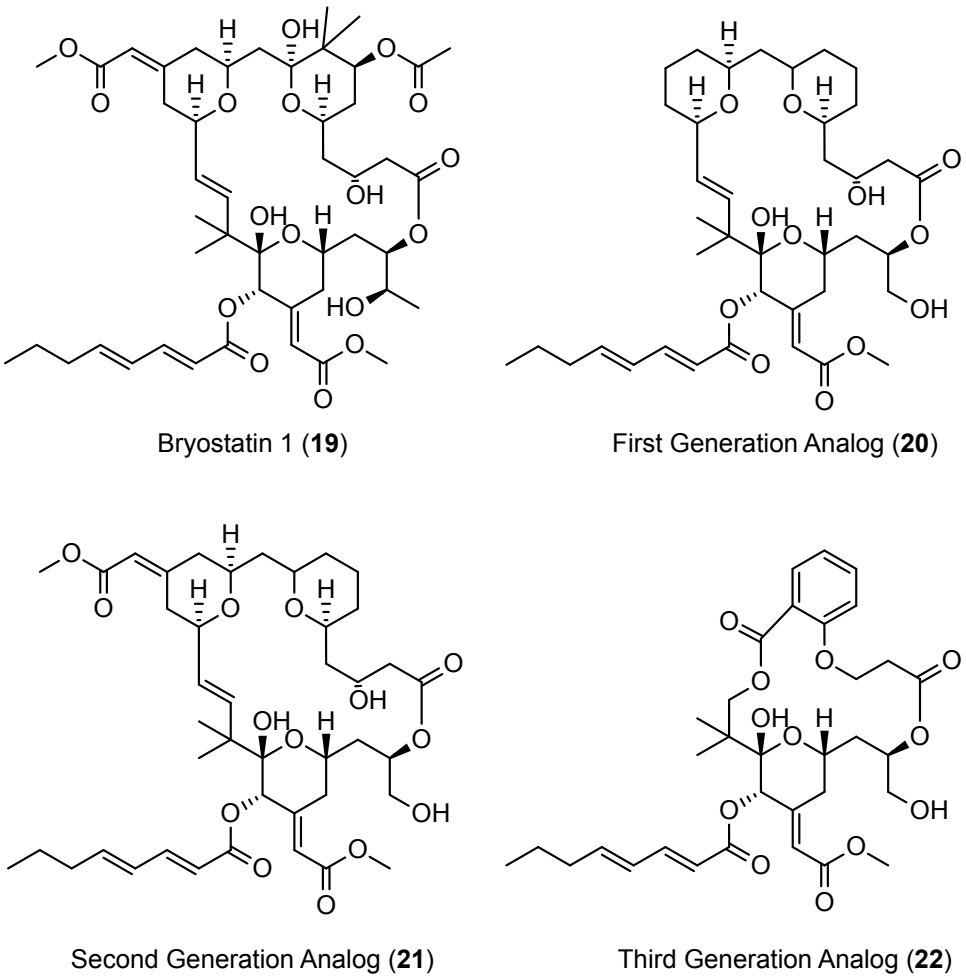



**Figure 1.7:** (*Top left*) The original natural product lead bryostatin 1 (**19**) and its functional analog structures of the first ( **20**, *top right*), second (**21**, *bottom left*) and the third generation (**22**, *bottom right*).

Even though not all described application studies revealed isofunctinal natural product mimetics, function-oriented synthesis as drug design strategy has the potential to mimic complex natural products by synthetically accessible chemical entities with similar or even superior functionalities. Therefore, natural products cannot be considered solely as biologically interesting and often complex structures, but can also be seen as sources of inspiration for making new isofunctional chemical entities. Furthermore, the basic principle of FOS as described by Wender *et al.*[38] states that "*the function of a biologically active lead structure can be recapitulated, tuned, or greatly enhanced with simpler scaffolds designed for ease of synthesis*" which can also be applied to the generation of natural product-inspired libraries. These collections had become an important source for selecting and finding new bioactive compounds in early drug discovery.[47]

## Synthetic Compound Collections

Screening large libraries of synthetic compounds are an essential part of identifying novel active compounds in medicinal chemistry. These high-throughput screenings (HTS) are among the most important sources of novel clinical candidates.[48] Both pharmaceutical companies and initiatives at the interface of academia and industry rely on HTS campaigns for small molecule lead discovery.[49–52] Nevertheless, HTS results have met only mixed success in the past decades.[53] The success of these lead discovery platforms depends strongly on the quality of the screened libraries.[54] Important library features are collection size and individual compound quality like purity and drug-likeness.[55] Furthermore, screened compounds should have appropriate physicochemical properties[56] and ought to have a certain molecular diversity.[57]

## Molecular Properties and Structural Diversity of Screening Libraries

In a computational analysis from 2001, Lee and Schneider compared these synthetic screening libraries to marketed drugs and natural products regarding their molecular properties and diversity.[58] The purely synthetic compound collection studied had a lower number of chiral centers along with an enriched ratio of aromatic atoms to ring atoms in comparison to marketed drugs and natural products. The analysis led to the assumption that trade drugs and natural products have a larger complexity and diversity than synthetic screening libraries. Additional chemoinformatic analyses came to similar conclusions.[59–61] In particular, molecules with larger complexity showed on average a higher bioactivity than less complex molecules.[62] However, Selzer *et al.* pointed out that too complex structures had an increased probability of pharmacokinetic problem and therefore "complexity must be balanced with other molecular properties". Almost ten years later, Stratton *et al.* examined structural and physicochemical characteristics of drugs which had either a natural or solely synthetic origin.[63] Their work confirmed the previous results. Moreover, their analysis indicated that drugs

emerged from a natural product pharmacophore already revealed desirable molecular features like lower hydrophobicity and a larger content of chiral centers. Although many natural products possess suitable physicochemical properties, several approved and orally bioavailable natural product drugs violated common rules,[64] like "Lipsinki's rules-of-five" (Ro5)[56] and "Veber's rules"[65] that guided and biased small-molecule drug discovery programs over the last two decades.[66–68] This led to the conclusion that natural products were considered exceptions of these rules. Nevertheless, an analysis of the all entries from the Dictionary of Natural Products (DNP) revealed that 60% of the 126,140 unique natural products had did not violate the Ro5.[69] An explanation of these observations was given by Ganesan, as he indicated that "*Lipinski did not look at* all *compounds in medicinal chemistry programs*".[64] Consequently, those criteria should not be viewed as strict rules but rather as guidelines for drug discovery programs, in particular in connection with natural products.

To further assess the impact of molecular complexity on the successful outcomes of clinical trials, Lovering *et al.* analyzed the properties of molecules from all stages in the drug discovery and development process (lead compounds, clinical candidates and approved drugs).[70] In this often-cited and frequently debated study it was demonstrated that the complexity of a molecule expressed as the fraction of $sp^3$-hybrdized carbon atoms ($Fsp^3$) and the amount of chiral carbon centers can be correlated to a successful drug development. Additionally, the saturation of drug compounds ($Fsp^3$ and number chiral carbon atoms) could be associated with their solubility, an experimental property which can be linked to the successful transition from lead structure to approved drug. In 2013, Lovering reported that an increased molecular complexity can be related to a reduction in promiscuity and toxicity in terms of associated cytochrome P450 (CYP) isozymes inhibtion.[71]

Focusing on their structural complexity, screening libraries, drugs or natural products can be described and analyzed in terms of molecular scaffolds. A well-established scaffold definition was given by Bemis and Murcko in 1996.[72] They devided a molecule into four structural elements: ring systems, linker atoms, atoms connecting ring systems, side chain atoms, which are all non-ring and non-linker atoms, and frameworks. A framework was defined "*as the union of ring systems and linkers in a molecule*".[72] This approach was considered as one of the first concepts "*to classify the crude shapes of molecules in terms of their cyclic frameworks*" or scaffolds as declared by Brown.[73] Grabowski *et al.* evaluated the complexity and molecular diversity in terms of molecular scaffolds and frameworks of natural products and drug-like molecules.[74, 75] The results confirmed previous findings that natural products consist of a remarkable structural diversity together with desirable physicochemical properties and provide a large number of distinct novel scaffolds which were not present in the investigated synthetic libraries. Out of the large variety of natural product scaffolds, several molecular frameworks obtained from natural products were labeled as "privileged scaffolds",[76, 77] a

term that was coined by Evans in 1988.[78] These *"privileged structures are capable of providing useful ligands for more than one receptor*" as Evans phrased it. These privileged scaffolds plus promising novel natural product-derived scaffolds [79] can be utilized to bias existing synthetic screening libraries towards new emerging target classes.[80] A representative set of privileged scaffolds derived from natural products but also found in drugs is shown in Figure 1.8.[76, 81] These include nitrogen-containing heterocycles like indoles (**23**), quinolines (**24**) and isoquinolines (**25**) or heterocycles with nitrogen and oxygen as benzoxazole (**26**). Coumarins (**27**), chalcones (**28**), chromones (**29**) and benzofurans (**30**) are common scaffolds of the oxygen-containing heterocycle family. Beyond these examples, several different NP-derived scaffolds have been in-

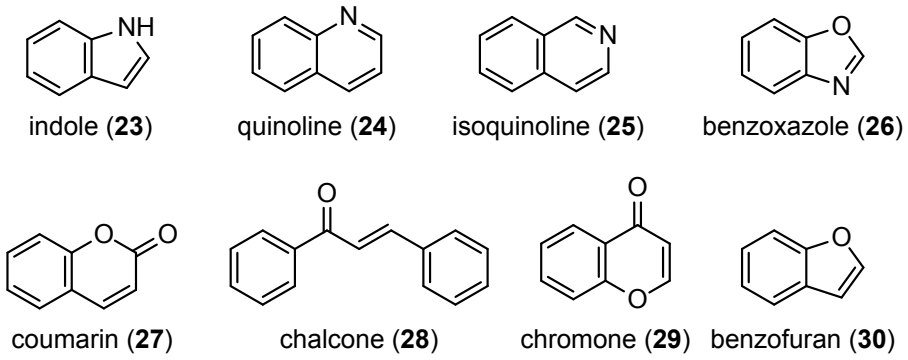



**Figure 1.8:** Representative set of privileged scaffolds occuring in natural products and drugs.

vestigated as suitable basis of identifying novel lead and drug structures. In 2015, a collection of various case studies using privileged scaffolds for drug discovery was published in "*Privileged Scaffolds in Medicinal Chemistry: Design, Synthesis, Evaluation*".[82] Notably, Cragg and Newman summarized instances of natural product-derived scaffolds in which these scaffolds served as root of the development of bioactive lead compounds.[83] In conclusion, natural products possess outstanding features like sufficient molecular complexity and diversity as well as desirable physicochemical properties. Moreover, numerous natural product scaffolds have been categorized as privileged scaffolds. Due to these beneficial characteristics of natural products, natural products have provided powerful starting points for generating natural product, NP-derived and NP-inspired screening libraries.

## Libraries of Pure Natural Products

Recent developments in separation and analysis methods as well as advanced organic synthesis strategies enabled the generation of natural product-based compound collections in the past decades.[47]

For example, the technical innovations in separating, isolating and elucidating of natural product structures from natural extracts led to a re-emergence of agents from natural source in drug discovery.[16] Just recently, a natural product screening library,

the so-called Canvass library, was composed for HTS against diverse target assays.[84] With contributions from academic and industrial laboratories, the natural products of this library consisted of molecular and structural features similar to drug molecules. The report mentioned highlights the initial efforts to extensively evaluate the potential bioactive natural products in an automated fashion. In 2018, the Kirchmair group summarized known and readily obtainable natural product libraries and analyzed their chemical properties.[85] They reported a high structural diversity of the available natural products and that these compounds covered biologically relevant areas in chemical space. Besides the facts that these screening libraries of pure natural products are suitable for HTS campaigns and populate desired areas in chemical space, the isolation and identification of natural products in sufficient amounts of material is challenging.

## NP-derived and NP-inspired Compound Libraries

Another strategy to build natural product-derived or natural product-inspired libraries counts on advanced organic synthesis concepts. Several synthetic approaches have been developed to construct synthesizable compound collections with favorable natural product-like features. Apart from the already described methods of diverted total synthesis (DTS) and function-oriented synthesis (FOS), diversity-oriented synthesis (DOS)[86] as well as biology-oriented synthesis (BIOS)[87] are further considerable library design methods.

Diversity-oriented synthesis aims to generate synthetic libraries with large molecular complexity and diversity from simple starting materials in a forward synthetic direction.[88] This is opposed to the more traditional approach of target-oriented synthesis starting from a target structure, e.g. a natural product, which will be dissected into fragments by retrosynthetic analysis until simple and available fragments are reached.[86] Linking DOS and natural products has the advantage to generate a huge variety of novel NP analogs and explore large, poorly covered areas in chemical space.[89] In one of the first studies applying DOS to design NP-inspired libraries, Tan *et al.* generated a compound library related to the the natural product shikimic acid (**31**, Figure 1.9),[90] a starting material of the synthesis of oseltamivir (Tamiflu).[91] The core fragments **32** and **33** of this library were immobilized on solid support to take advantage of a split-and-pool synthesis strategy[92] to increase the amount of diverse products. Based on these core fragments, over two million distinct encoded chemical entities could be obtained by using different terminal alkynes ($R_1$), carboxylic acids ($R_2$) and amines ($R_3$) (Figure 1.9). A similar approach was applied to create a NP library with prostaglandin $E_1$ (**34**) as leading natural product (Figure 1.10, left).[93] Over 26 prostaglandin $E_1$ analogs were synthesized in parallel within four steps containing the NP core fragment **35** (Figure 1.10, left) aiming to find active modulators of prostaglandin $EP_2$, $EP_3$ receptors. In 2000, Nicolaou and co-workers constructed

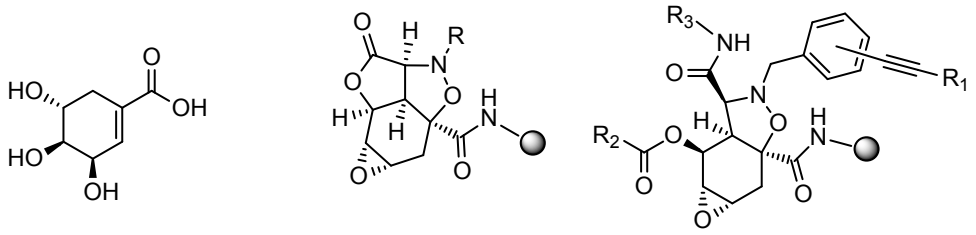



**Figure 1.9:** Shikimic acid (**31**) served as starting point to build a large synthetic compound collection from the shown core fragments (**32**,**33**) with different terminal alkynes ($R_1$), carboxylic acids ($R_2$) and amines ($R_3$).

extensive NP-like libraries founded on a privileged scaffold, for example present in the natural HIV protease inhibitor calanolide A (**36**, Figure 1.10, middle)[94]: The 2,2-dimethyl-2*H*-benzopyran scaffold (**37**, Figure 1.10, middle).[95–97] Repeat application of the split-and-pool synthesis strategy resulted in a collection of over 10,000 NP-like compounds. In a further study, Oguri *et al.* designed a NP-like library inspired by the natural antimalarial agent artemisinin (**38**, Figure 1.10, right),[98] which has been discovered in the early 1970s by the Chinese scientist and Nobel laureate Youyou Tu.[99] The generated compounds consisted of the NP-like elaborated core **39** (Figure 1.10, right) and led to the identification of NCEs with confirmed antitrypanosomal activities, also exhibited by artemisinin itself.

In 2006, the Waldmann group introduced the concept of biology-oriented synthesis (BIOS) as "*an efficient approach to the discovery of new compound classes for medicinal chemistry and chemical biology research*".[100] The underlying principle (Figure 1.11) is grounded

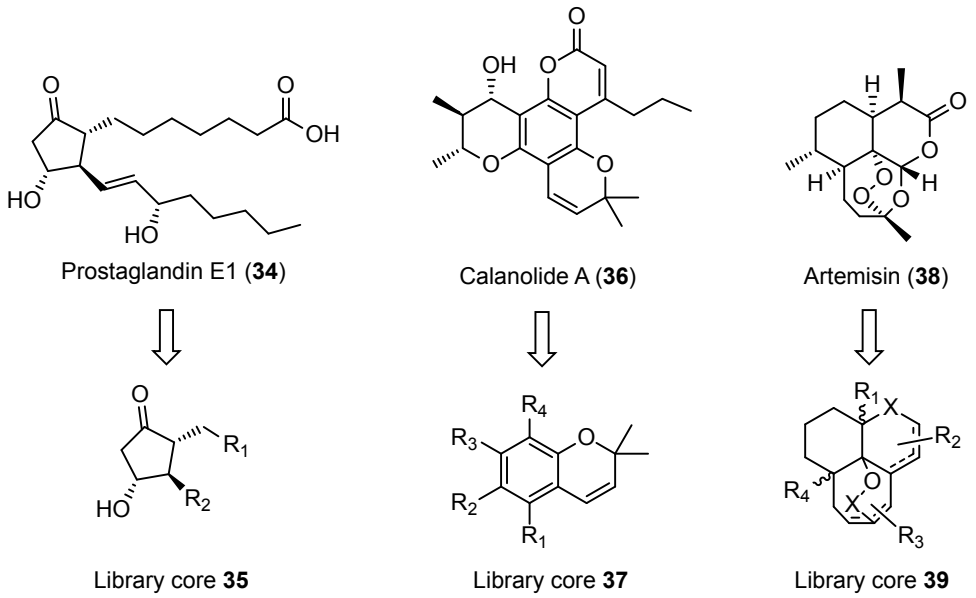



**Figure 1.10:** (*Left*) Based on prostaglandin $E_1$ (**34**), a NP library was constructed around the core fragment **35**. (*Middle*) The natural HIV reverse transcriptase inhibitor calanolide A (**36**) is one example containing the privileged scaffold 2,2-dimethyl-2H-benzopyran (**37**), which served as core for a NP library. (*Right*) The natural antimalarial agent artemisinin (**38**) was the source of inspiration for a NP-like compound collection with the library core **39**.

on the assumption that structural features of proteins and natural products have been conserved during evolution leading to preserved amounts of protein folds[101] and a limited number of natural product scaffolds.[102, 103] In contrast to this structural conservatism, variability within the amino acid sequences and the substitution patterns of natural products ensured a sufficient degree of diversity with the conserved structures. This unique combination of characteristics can be utilized to identify NCEs with NP-inspired scaffolds which may encode the interactions with conversed protein binding sites. A necessary requirement of BIOS is the identification and analysis of natural product scaffolds.[103]

For this purpose, Koch *et al.* extracted the scaffolds of all natural products from the Dictionary of Natural Products and classified them structurally to map the natural product chemical space.[104] The classification principle "Structural Classification of Natural Products" (SCONP) organizes extracted scaffolds in a tree-like structure and highlights connections between different hierarchical levels (Figure 1.12). The observation that not every scaffold branch can be reduced to a single ring structure, led to an extension of this concept by introducing "gap-filling" virtual scaffolds to achieve completion of all tree branches.[105] Varin *et al.* extended the scaffold tree algorithm to consider up to ten branches per hierarchical level and therefore generating a scaffold network with more information about scaffold relationships and potential scaffold "activity islands".[106] In the initial BIOS study, Nören-Müller *et al.* investigated synthetic compound collections derived or inspired by natural products and assessed their inhibitory effects on certain tyrosine phosphatases. They identified novel inhibitors of selected tyrosine phosphatases from the family of yohimbane alkaloids, *inter alia*, the natural product yohimbine (**40**).[100] Based on the scaffold of yohimbine (**41**), the application of SCONP revealed an assignment of this scaffold to the indole branch of the SCONP scaffold tree. Therefore, the indole-based structure **42** was employed to build a NP-inspired compound collection yielding in the discovery of a novel selective phosphatase inhibitor. Over *et al.* demonstrated that natural product-derived fragments are a valuable source of novel bioactive fragments for fagment-based drug discovery (FBDD).[107] They incorporated the extracted fragments and validated them in line with principles of fragment-based drug discovery.[108–110] They were able to

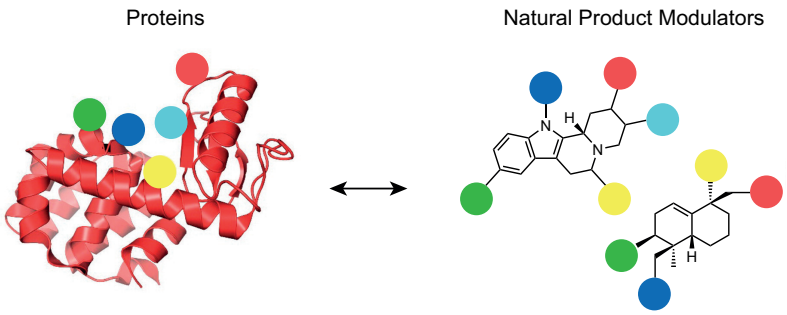



**Figure 1.11:** Proteins and natural products consist of conserved structural features of their folds and scaffolds. But both show a high degree of diversity in amino acid sequence and scaffold substituent patterns. Reprinted and adapted with permission from ref. [103]

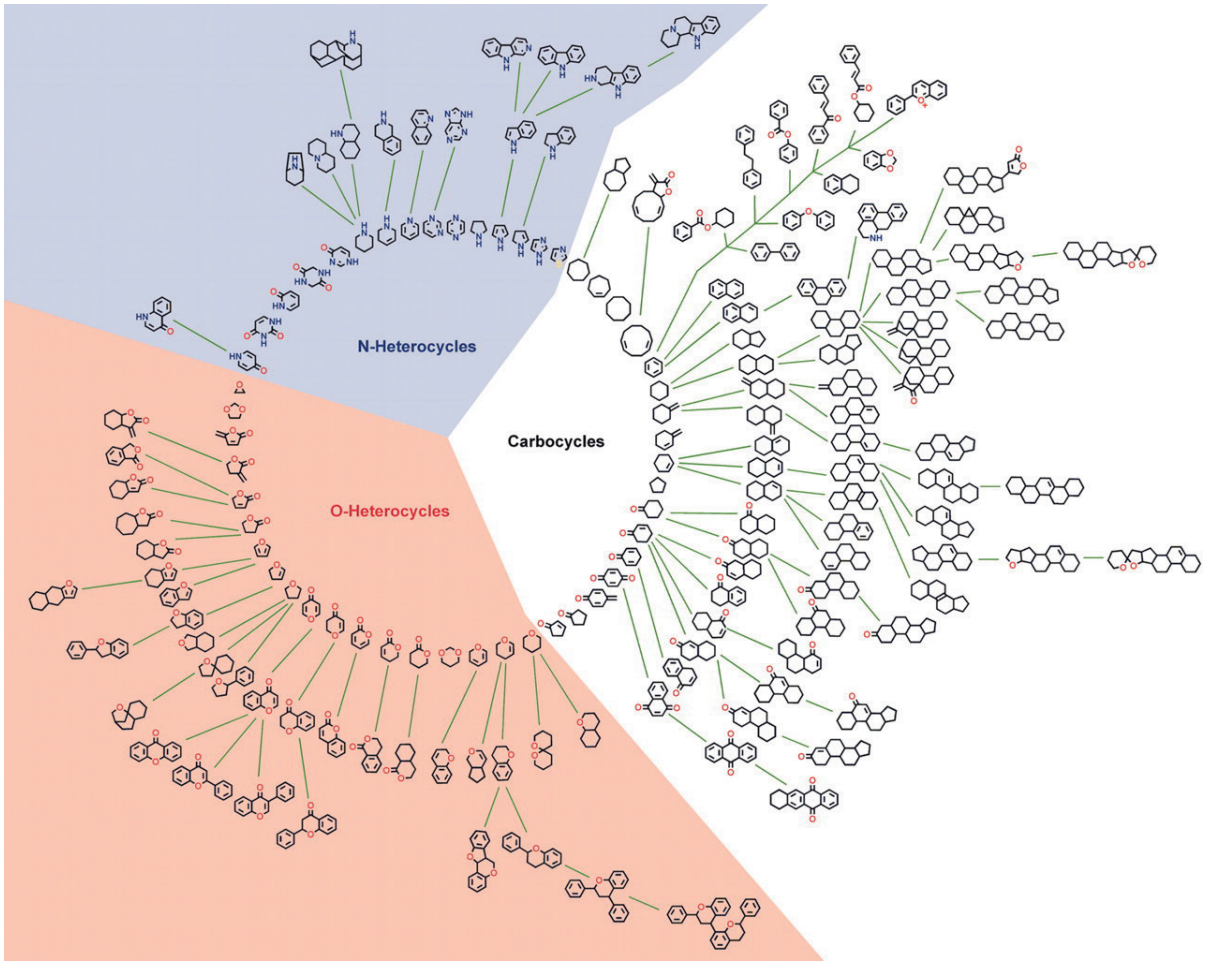



**Figure 1.12:** Tree-like representation of the different hierarchical levels (SCONP) of natural product scaffolds from the Dictionary of Natural Products. The scaffolds are broadly categorized into nitrogen- (N-heterocycles), oxygen- (O-heterocycles) and solely carbon-containing ring systems (carbocycles). Reprinted with permission from ref. [104]

find novel p38αMAP kinase modulators and phosphatase inhibitors with novel chemotypes. They concluded that exploiting NP-derived fragments to generate NP-inspired libraries may help *"to overcome limitations in the use of natural products in drug discovery due to lack of accessibility and synthetic tractability"*.[107] Recently, the Waldmann group

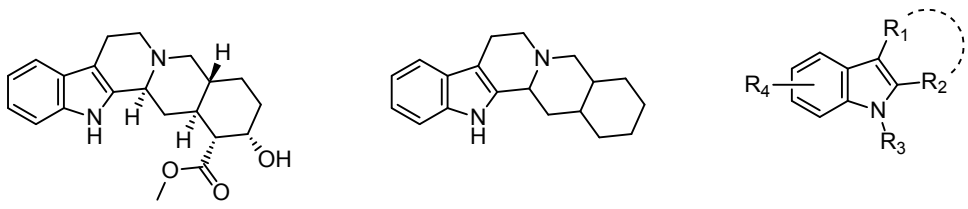



**Figure 1.13:** The natural tyrosine phosphatase inhibitor yohimbine **40** and its scaffold **41** served as starting point of a NP-inspired compound library with the indole core structure **42**.

combined the previous concepts of biology-oriented synthesis and natural product-derived fragments and obtained the so-called "pseudo natural products" crhomopynones with inhibitory effects on the two glucose transporters GLUT-1 and -3.[111] They described the associated limitations of both strategies, such as identified NPs covering only a small area in the total NP-like chemical space.[112] and focusing on known NP-derived scaffolds narrows down the search of less-explored biologically-relevant areas in chemical space.[113] They proposed the concept of "pseudo natural products" by merging NP-derived fragments into unprecedented structures following a *de novo* design approach. This molecular *de novo* design approach can be described as the design of "novel molecular structures with desired pharmacological properties from scratch", as outlined by Schneider and Fechner in 2005.[114] For selecting and designing these pseudo natural products, Karageorgis *et al.* followed a set of design guidelines.[111] First, they aimed to combine NP-derived fragments to create a novel three-dimensional scaffold which can be synthesized and diversified with manageable synthetic efforts. Second, the selected fragments should contain complementary heteroatoms (nitrogen and oxygen) to ensure sufficient structural diversity and adequate differences to the parent natural products. Finally, they proposed that fragments from natural products with diverse bioactivities as well as unrelated biosynthesis pathways "*will encode different structural parameters for binding to proteins*".[111] These guidelines led on the one hand to the selection of natural products, for example the anticancer agent catechin (**43**)[115] and the psychoactive natural agent $\Delta^1$-THC (**44**),[116] with a chromane fragment (**45**), which is closely related to the privileged benzopyran scaffold, as mentioned before.[76] On the other hand, the natural antibiotics THAN1057A/B (**46**) [117] and the antiproliferative natural products aplicyanins A-F (**47**)[118] shared the nitrogen-containing tetrahydropyrimidinone fragment (**48**). Both fragments (**45** and **48**) were fused into the chromopynones (**49**),a new pseudo natural product class. Based on this novel scaffold, a NP-inspired library of over 44 compounds was synthesized and its bioactivities were subsequently evaluated in several cell-based assays. The activity assessment revealed that 18 out of 44 compound had inhibitory effects on the glucose uptake and selectively targeted glucose transporters GLUT-1 and GLUT-3, upregulated targets in many cancer types.[119] The results of this proof-of-principle study showed for the first time that natural product-inspired libraries can be generated by combining the concepts of biology-oriented synthesis and fragment-based drug discovery. The outcome of these investigations can be considered as the next step towards the design of NP-inspired collections of NCEs by harnessing of the full potential of natural products.

Finally, natural products play still an important role in drug discovery today. Both their unique structural features as well as their beneficial physicochemical properties turn natural products into a highly valuable resource of novel drug-like compounds. Not only natural products themselves, but also their derivatives, mimetics and NP-inspired

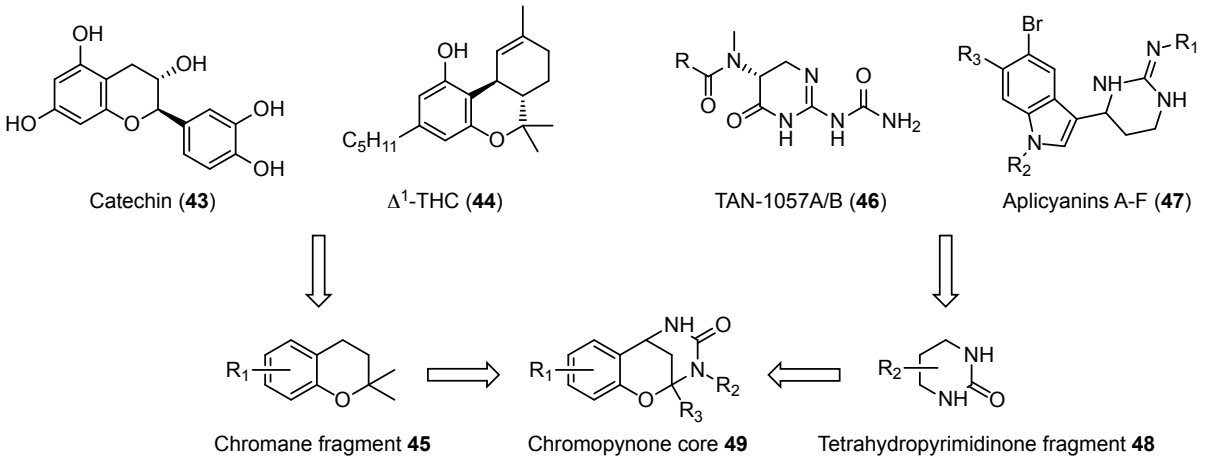



**Figure 1.14:** The natural products catechin (**43**) and $\Delta^1$-THC (**44**) are example consisting of a chromane scaffold (**45**). The natural agents THAN1057A/B (**46**) and aplicyanins A-F (**47**) are representatives containing a tetrahydropyrimidinone core (**48**). The two NP fragments **45** and **48** are merged into the "pseudo natural product" chromopynone **49**.

libraries are an endless source of inspiration for lead and structures in medicine. However, the supply of natural products from their natural origins is often limited and the syntheses of these agents and their related compounds are still challenging. The upcoming integration of chemoinformatic tools like molecular property calculation, pharmacophore comparison and scaffold analysis led to powerful advances in natural product-based drug discovery during the last years, in particular in the design of NP-like libraries. The combination of natural products, organic synthesis and chemoinformatics offers a great opportunity to promote and implement innovative strategies in a multi-disciplinary approach for future drug discovery. To this extent, the potential of computational methods in drug discovery will be elaborated in more detail in section 1.2.

## 1.2 Computer-Assisted Drug Design

The use of computational methods to support the identification of a novel lead or drug structure is increasingly common in modern drug discovery projects. Already in 2004, Jorgensen pointed out that *the systematic use of wide-ranging computational tools to facilitate and enhance the drug discovery process* has become an crucial part of successful drug discovery programs.[120] These tools include commonplace software such as chemical-drawing suits, two- and three-dimensional visualization tools as well as database management programs. Several computational methods have been developed to advance the identification of bioactive compounds and their transition into clinical drug candidates at each stage of the drug discovery process.[121] Physicochemical property calculation, energy minimization of molecular structures, prediction of potential interactions between ligand and protein, and absorption, distribution, metabolism, excretion and toxicity (ADMET) modeling are just some well-established examples of computer-based approaches in modern drug discovery.[122] Among these *in silico* approaches, gathered under the umbrella term computer-assisted drug design (CADD), several methods have been described to generate or identify new drug-like compounds. Here, various screening approaches and molecular design strategies will be discussed. Moreover, this section will highlight examples, where these computational concepts have employed natural products as templates to screen virtual libraries, and for the generation of novel bioactive small molecules. One of the foundations of CADD, especially the representation of molecular structures and the assessment of molecular similarity are explained in the next section.

### Molecular Descriptors and Molecular Similarity

A central aspect in identifying and designing novel chemical entities is the relationship between chemical structures and their experimentally-determined properties.[123] This relationship between chemical structure and function, e.g. pharmacological activity is shown in a simplified schema in Figure 1.15. In one direction, molecular structures can be linked to one or more biological functions. Quantitative structure-activity relationship (QSAR) models have been developed to extract rules and correlations from these links and can be applied subsequently to predict the biological functions of novel compounds. Already in 1962, Hansch *et al.* postulated the basis concepts of QSAR modeling.[124] In their seminal work, they found that the biological activity of phenoxyacetic acid derivatives can be correlated with slight changes in their structures. Since then, QSAR modeling have been well established in medicinal chemistry as a powerful means to analyze and predict bioactivities of new compounds, especially in lead optimization.[125] In the opposite direction, the connection of a biological function to chemical structure(s), has been described as an "inverse" QSAR approach[126, 127]

and solutions of this "inverse" QSAR are closely related to molecular *de novo* design approaches aiming to generate novel chemical entities based on mapping functions (Figure 1.15).[128]

Generally, the representation of data can have a large influence of the efficacy of computational methods. Therefore, a requirement to utilize computational methods in drug design is an appropriate representation of molecular structures, and their desired biological activities. The common visualization of molecules with balls and sticks contains adequate information to get an intuitive description of the molecular architecture. However, these visualizations are generally insufficient to adequately describe the relationship between molecular structures and their corresponding functional characteristics.[123] For this purpose, more than thousand different types of molecular descriptors have been developed and applied to model certain properties of chemical structures.[129]

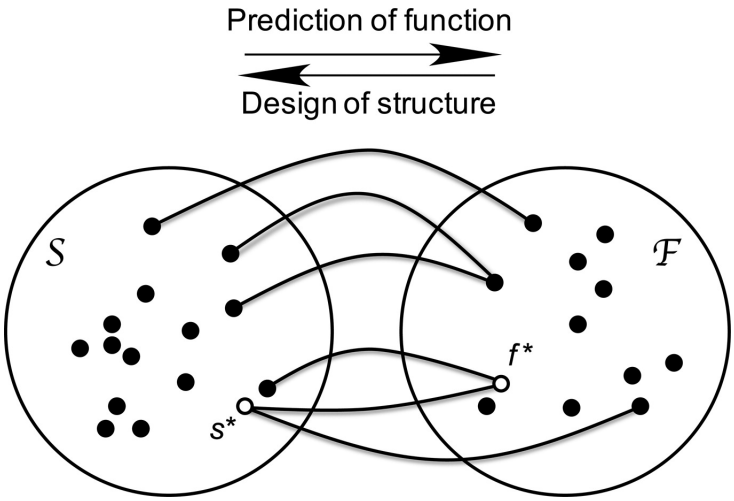



**Figure 1.15:** Simplistic representation of the relationships between structural chemical space *S* and functional biological space *F*. Ligands in *S* and protein targets in *F* are shown as black dots. Quantative structure activity relationship (QSAR) models correlate structure with activity (left to right) while *de novo* design links activity to structures (right to left). Certain molecules like *s** can bind several targets ("promiscuous binders"), and some targets, e.g. *f**, have multiple associated ligands ("promiscuous targets"). Reprinted with permission from ref. [128]

### Molecular Descriptors

One definition of a molecular descriptor was given by Todeschini and Consonni: *"The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment."*[130] Based on this definition, molecular descriptors can be classified as descriptions of experimental values, like the determination of physicochemical properties, or as theoretical descriptions of different kinds of molecular representations.[129] The first category consists of results from experimental measurements like solubility (e.g. Log*P*), or molar refractivity. Second, theoretical molecular descriptors can be separated according to complexity, also known as descriptor "dimensionality".[131, 132] The different levels of descriptor dimensionality are examplified for ibuprofen (**50**) in Figure 1.16. Molecular descriptors with dimen-

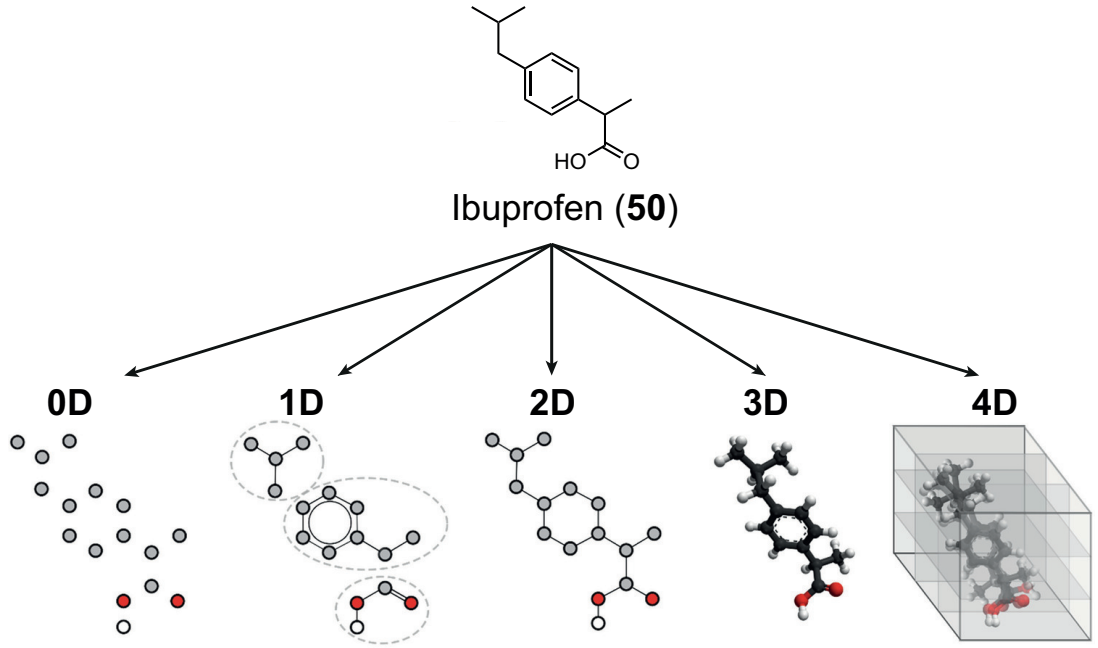



**Figure 1.16:** The two-dimensional structure of ibuprofen (**50**) shown with different levels of descriptor complexity, also known as "dimensionality" of a descriptor. Reprinted and adapted with permission from ref. [131]

sionality "0" are obtained from the simplest representation of a molecule: the chemical formula. These descriptors purely rely on atom types and numbers, and do not consider connectivity or different bond types. Prominent examples are molecular weight and certain atom counts (e.g. number carbon atoms). One-dimensional (1D) descriptors take into account the presence or absence of certain substructures in a molecular representation. These descriptors are often encoded binary representations (e.g. fingerprints) or as occurrence frequencies. The 2D descriptors contain information about connections between atoms and their bond types. The resulting higher complexity can be used to describe molecules as graphs, in which vertices represent atoms and edges display bonds. This representation allows for the incorporation of topological properties of a molecule together with the associated chemical features of the atoms.
An even higher level of descriptor complexity can be achieved by considering a molecule as a geometrical object. The introduction of spatial coordinates, e.g. obtained from crytsal structures or computer-based conformation generation and optimization, increases the information content of a particular presentation.[132] These types of descriptors are widely used in medicinal chemistry and drug discovery.[133] However, there are still certain issues when considering three-dimensional structures of molecules. On one hand, the generation of three-dimensional molecular conformations depends on the computational methods employed to calculate these structures,[134–136] resulting in multiple energetically-favorable conformers for particularly flexible molecules. These ensembles of similar conformers have been utilized in a clustered format to

identify bioactive compounds.[137, 138] On the other hand, the energetically most favorable conformation of a molecule can be different to its bioactive conformational state.[139] Keeping these aspects in mind, the application of 3D-based descriptors in drug discovery is a trade-off between the limitations and benefits of these content-rich representations which requires case-by-case analysis.[140]
A four-dimensional descriptor includes information about molecular interactions with active site(s) of protein(s) with the goal of quantifying and characterizing their potential bioactivities.[132] To highlight molecular presentations with different complexity levels, the following paragraph comprises examples of several molecular descriptors.

### Examples of Molecular Descriptors

Among the 1D molecular descriptors, fingerprints have become the standard representations for similarity calculations in computational drug discovery.[141] Two frequently employed fingerprints are the molecular access system structural key fingerprint (MACCS)[142] and radial fingerprints like the extended connectivity fingerprints (ECFPs)[143] or Morgan fingerprints.[144] MACCS keys, a 166 dimensional fingerprint, include a set of 166 structural fragments which results in a "1" (on bit) at the position of the corresponding fragment if it is present in a compound structure, and in "0" otherwise. Radial fingerprints as ECFPs considering circular atom neighborhoods and representing the presence of certain substructures.[143] In the beginning, the ECFP calculation assigns an integer identifier to each non-hydrogen atom in the query compound. The circular neighborhood of each atom is then iteratively investigated with an increasing diameter, which is a freely selectable parameter and commonly set to 4, corresponding to a topological radius of 2, as depicted in Figure 1.17. This leads to a list of existing substructures together with associated integer identifiers which can be mapped ("hashed") to a binary string representation of fixed-length. In Figure 1.17, all unique derived substructures of a template molecule are shown and

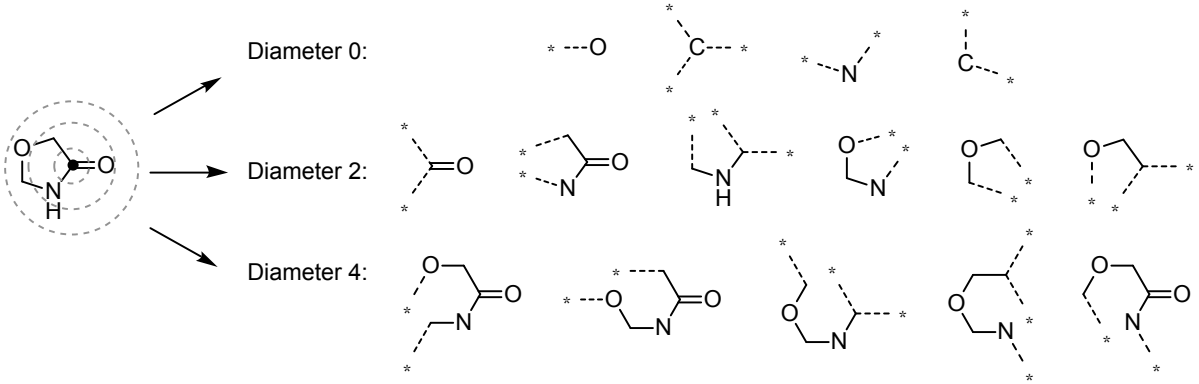



**Figure 1.17:** Schematic illustration of generating substructures from a template compounds with different diameters for extended connectivity fingerprints. Bonds to neighboring atoms outside the considered diameter (*) are drawn as dashed lines. Illustration adapted from ChemAxon documentation.

classified after considering the circular neighborhoods with different bond distances indicated as the diameters 0, 2 and 4. Each substructure receives an integer identifier which can be "hashed" into a binary bit string (only "0" and "1") with a pre-defined lengths (e.g. 1024 bit) the so-called folding.[143] Recently, the Keiser group the developed a three-dimensional fingerprint (E3FP) as an extension of ECFP fingerprints with three-dimensional information.[145] Their fingerprint representation encoded information about the 3D characteristics of molecules without requiring prior computational alignments. They identified and confirmed experimentally novel drug-to-target binding predictions which were inaccessible with the equivalent 2D fingerprint.
2D Descriptors are suitable to describe the topology of molecular graphs and the correlation of certain features within these graphs. For instance, Schneider *et al.* introduced a two-dimensional topological descriptor called CATS (chemically advanced template search) which includes pre-defined feature types for each pharmacophore which are assigned to each atom.[6] These different feature types (hydrogen-bond donor, hydrogen-bond acceptor , positively charged, negatively charged, lipophilic) represent the pharmacophoric characteristics of a molecule. A pharmacophore is "*an 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*", as defined by Wermuth *et al*.[146] The CATS descriptor performs a pairwise pharmacophore assignment within edge distances of 1 to 10 following the shortest path between two vertices in the molecular graph. The sum of the individual feature pair counts was employed to scale the final descriptor values.[6] The concept of the CATS descriptor was revisited and extended by Reutlinger *et al.* in 2013.[147] In the reported CATS2 descriptor they distinguished between "lipophilic" and "aromatic" atom types, whereas the original CATS descriptor did not differentiate between these two atom types. A simplistic representation of the CATS2 descriptor calculation is shown in Figure 1.18. It has to be mentioned that not all vertices of a molecular graph have been assigned with an pharmacophoric feature type. The coarse-grained atom-typing and feature pair correlation within the CATS2 descriptors led to certain level of "fuzziness" of the CATS2 descriptor,[147] whereby "*fuzziness of a molecular representation defines the degree to which compounds are considered as similar*" as Klenner and co-workers described it.[148] These less specific atom-typing schemes are suitable for the identification of new bioactive structures with novel chemotypes.[149]
An early example of a three-dimensional descriptor was reported by the Gasteiger and co-workers.[150, 151] In their 3D molecular descriptor called "Molecule Representation of Structures based on Electron Diffraction" (3D-MoRSE) molecules were encoded by a fixed number of values which were derived from the three-dimensional coordinates of the molecule. They were able to successfully differentiate between dopamine D1 and D2 agonists utilizing their 3D-MoRSE code. A common way to evaluate 3D similarity of molecules is by means of rapid overlay of chemical structures (ROCS) software, as

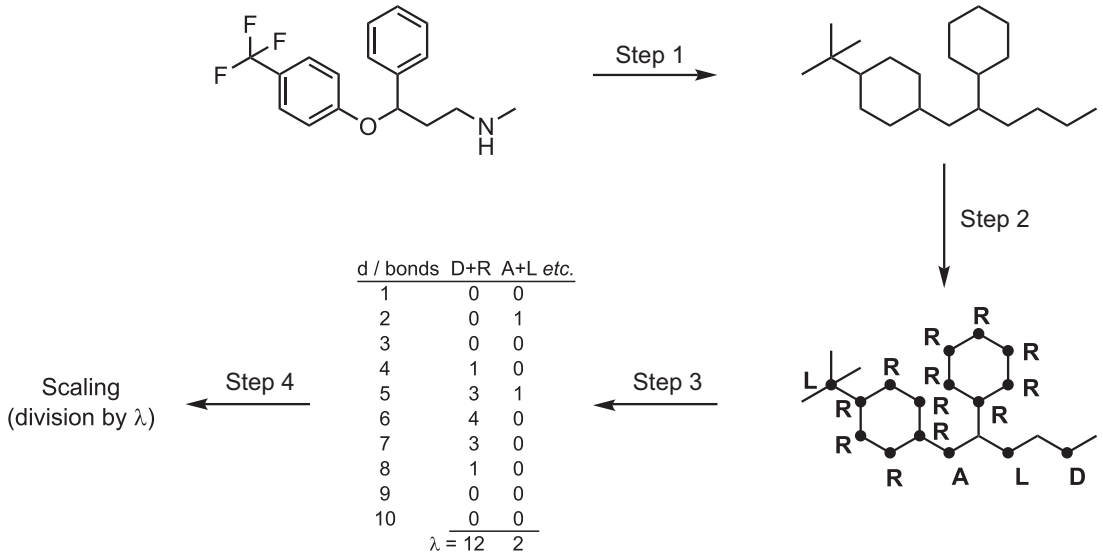



**Figure 1.18:** Schematic overview of the calculation of the CATS2 descriptor. First, a 2D molecular representation is reduced to a molecular graph (Step 1). Second, the pharmacophoric features are assigned to the related vertices (Step 2, A = hydrogen-bond acceptor, D = hydrogen-bond donor, L = lipophilic, R = aromatic). In a third step, atom feature pairs over bond distances from 1 to 10 are counted and summed up (Step 3) and finally they are scaled by the sum of the individual feature pair λ (Step 4). Reprinted and adapted with permission from ref. [147].

developed by OpenEye.[152] To compare 3D molecular shapes, this approach tries to find the maximum overlap between the conformers of two molecules by treating each compound as a mixed Gaussian model, allowing for partial, non-hard matching, which allowed for partial matches and the integration of colour, i.e. electrostatic, information. In one of the first applications of ROCS, Rush *et al.* identified weak inhibitors of bacterial protein-protein interactions with novel scaffolds.[152] Furthermore, Hawkins and co-workers proposed that ROCS as a shape-based ligand-centric concept can be considered as an alternative to receptor-focused approaches like molecular docking.[153] However, ROCS needs a computationally expensive conformer alignment which renders the applicability to larger datasets more difficult. An alternative approach to assess shape similarity is emphasized in section 4.3, in which this novel 3D similarity calculation has been utilized to select computationally designed mimetics of the complex natural product (-)-englerin A. One example of a 4D descriptor is the comparative molecular field analysis (CoMFA) introduced by Cramer *et al.* in 1988.[137] This analysis relies on grid-based representations which place active molecules into a three-dimensional grid of equally distributed grid points. By scanning each grid point in terms of steric and electrostatic interactions with embedded molecules, a map of favorable and unfavorable interactions can be generated from these scans. Such a map can be utilized subsequently to predict potential interactions of novel compounds.
Molecular descriptors encode the content information of molecular structures in different levels of complexity. The reference work from Todeschini and Consonni collates information on more than 3,300 available molecular descriptors from various types.[129]

All these molecular descriptors have their own limitations and benefits and no descriptor is applicable in every case. Therefore a careful descriptor selection is necessary for each individual study. The most prominent application of descriptors is in capturing structural features relevant for the bioactivity of molecules. Of particular note is the comparison of a template structure with query compounds, as it is an essential *de novo* molecular design. The comparisons can be performed under several similarity metrics. The concept of molecular similarity and chosen similarity metrics will be introduced and certain conceptual aspects will be discussed.

**Molecular Similarity**

The term molecular similarity is defined as a "*measure of the coincidence or overlap between the structural and physicochemical profiles of compounds*" as described by the "Glossary of Terms used in Medicinal Chemistry".[154] The concept of molecular similarity has been widely exploited in medicinal chemistry and computational drug discovery.[141, 155, 156] It relies on the basic assumption that structural similarity correlates with similar biological activity.[157] To assess this correlation and determine molecular similarity between to chemical structures, a suitable molecular representation (molecular descriptors) has to be chosen which encodes relevant features of a molecule.[156] Additionally, weighing molecular features in different ways induces a bias towards certain characteristics. Finally, similarity metrics (or similarity coefficients) translate the descriptor information of compared molecules into a similarity value, usually ranging from "0" to "1", where "1" indicates fully identical molecular representations. The general principle of similarity encompasses numerous diverse approaches to describe similarity from various perspectives. In a similar manner to the variation in conceptual complexity seen in the molecular descriptors, similarity-determining approaches vary in their theoretical basis and ready-interpretability.[156] On the level of chemical formulas and connectivities in molecular structures, chemical similarity deals mainly with physicochemical property comparisons (e.g. molecular weight, solubility, dipole moments, etc.), and molecular similarity outlines generally structural features like topologies, scaffolds or common substructure motifs in compounds. Another important aspect of similarity assessment is the distinction between a global or local similarity.[156] A global perspective focuses on the entire molecular structure, whereas a local similarity takes only parts of the molecules into account. The previously mentioned structural fingerprints are examples of global similarity methods. In contrast, pharmacophore-based methods process information from a small fraction of the total molecule and are examples of assessing local similarity.[158]
Many similarity metrics have been developed to translate similarity into numerical values.[159] The Tanimoto coefficient (Tc) is one of the most commonly applied similarity metrics in drug discovery.[160, 161] The Tanimoto coefficient of two compounds

($A$ and $B$) is defined by

$$Tc(A,B) = \frac{c}{a+b-c} \tag{1.1}$$

where $a$ and $b$ represents the number of on-bits("1") of compound $A$ and $B$, respectively. The variable $c$ is the number of shared on-bits of $A$ and $B$ and consequently Tc quantifies the ratio of common features of $A$ and $B$ to the total amount of features from both compounds corrected by the common features. The Tanimoto coefficient can be implemented easily and the calculations are fast which explains the large popularity of the tanimoto coefficient in drug discovery.[156]
Calculating the Euclidean distance between two molecular descriptors is an alternative method to measure their similarity. For example, the topological pharmacophore similarity, as described by the CATS descriptor,[6, 147] can be assessed by calculation of the Euclidean distance $D$ between vector representations (CATS/CATS2 descriptor) of two molecules $A$ and $B$:

$$D(A,B) = \sqrt{\sum_{i=1}^{p}(v_i^A - v_i^B)^2} \tag{1.2}$$

where $v_i^A$ and $v_i^B$ are the CATS description vectors and $p$ is the number of numerical elements contained in each molecular vector (CATS is a 150 and CATS2 a 210 dimensional vector).[6, 147]

Despite a long history, and a widespread acceptance and usage in medicinal chemistry, the assessment of similarity still remains a challenging task. Several thousand molecular descriptors can be combined with many similarity metrics to solve this task, and each combination has its own benefits and drawbacks. Additionally, computed similarity can deviate from the intuitive human notion of similarity, such as in the visual judgement of dissimilar objects.[156] Nevertheless, molecular descriptors and similarity concepts are at the core of two active research areas in computational drug discovery, namely virtual screening and *de novo* molecular design. Together with these two technologies, *in silico* target prediction, whose foundation is built on these concepts, enables the identification of as-yet unknown macromolecular targets of compounds of interest. In the following sections, the basic concepts of these methods will be discussed and examples of their applications to natural products will be highlighted.

## Virtual Screening

Complementary to traditional library screening approaches like high-throughput screenings (HTS), virtual screening (VS) is a computational method "*capable of automatically evaluating very large libraries of compounds*" as Walters, Stahl and Murcko wrote in their review from 1998.[162] This process is meant to filter the large chemical space with its almost inconceivably-large number of potential chemical structures to find a much more manageable subset.[163, 164] To perform this library filtering procedures, compounds have to be selected either by computationally predicting their binding to a macromolecular target or calculating similarities to already known active ligand(s).[165, 166] Because of that two general types of virtual screening can be distinguished. On the one hand, structure-based approaches which require prior knowledge about ligand-protein interactions in the binding site(s) of the target of interest, and there ligand-based methods which build on molecular similarity assessments between library compounds and known active ligands. Natural products undertake an important role in virtual screening studies, either as screening compounds or as reference ligands.[167] Guided by a 3D pharmacophore model derived from interactions between the natural product galanthamine and the enzyme acetylcholinesterase (AChE), Rollinger *et al.* discovered two natural products with inhibitory effects on AChE in a structure-based virtual screening of over 110,000 natural products.[168] In 2014, Liu and co-workers performed molecular docking techniques to discover novel inhibitors of the transcription factor family STAT (signal transducer and activator of transcription).[169] They virtually screened a library of over 90,000 natural products and natural product-like molecules which furnished 14 bioactive compounds, of which one structure showed considerable activity and selectivity. The basis of ligand-based virtual screening (LBVS) is the assessment of molecular similarity between query and reference compounds.[170] The assumption that similar molecular structures possess similar activities[171, 172] can be utilized to select compounds from virtual libraries which are similar to known active molecule(s) without any further prerequisite information about ligand-protein interactions. Many reported examples have confirmed the "similar property principle"[157] leading to selections of novel active structures.[173] In particular, natural products or endogenous ligands serve as valuable reference structures in ligand-based virtual screening studies.[167] In 2017, Karhu and co-workers employed a similarity screen between a set of 19 antichlamydial reference natural products and a natural product library to identify novel natural products with antichlamydial activities.[174] Their 2D similarity search resulted in 53 virtual hits. Six compounds showed considerable inhibitory effects in combination with novel chemotypes and one of the active natural products had an $IC_{50}$ value of 0.3 µM. Recently, Grisoni *et al.* identified novel natural product-inspired inhibitors of cyclooxygenase-2 in a prospective LBVS by utilizing topological matrix-based descriptors.[175] These

Matrix-based descriptors were derived from several descriptor metrics encoding the topological and chemical information of a molecule. Their implementation in a prospective LBVS of natural product-derivatives led to the finding of four bioactive molecules out of 22 experimentally tested compounds.
Over the last decade, virtual screening has evolved as a screening technology complementary to widely used high-throughput screenings. The essentially infinite size of virtual chemical space along with cost-efficient computational methods have turned virtual screening into a powerful and common approach in identifying novel bioactive structures. Moreover, virtual screening benefits from the current availability of large databases, in particular natural product libraries,[85] which can be searched to discover bioactive NCEs with novel chemotypes.

### *De Novo* Molecular Design

From the ever-expanding collection of computational methods in drug discovery, several tools can be utilized to deal with the design and generation of new drug-like molecules.[121] This field of CADD, often described as computer-based molecular "*de novo* design, was introduced in the 1990s as an instrument to generate *novel molecular structures with desired pharmacological properties from scratch*", as Schneider and Fechner depicted it in 2005.[114] As with virtual screening approaches, molecular *de novo* design can be defined as either structure/receptor-driven or ligand-guided design strategies.[114] These two distinct principles differ in how they gather essential information about the interactions between receptor and ligand. These "primary target constraints"[114] can be retrieved either from interaction spots within the binding site of a receptor structure (receptor-based) or from assessing similarity to molecular ligand structures which interact with the target(s) of interest. Over the past few decades, several different receptor- and ligand-scoring methods have been developed to computationally assess the quality of NCEs. *De novo* design methods are further defined in terms of their structure generation principles (structure sampling).[114, 123, 176] Hereinafter, selected *de novo* design software with different scoring and sampling concepts will be described. Moreover, the potential of combining state-of-the-art *de novo* design tools with natural products to design novel bioactive natural product mimetics will be addressed.
In the early days of computational *de novo* design, the main methods employed were those belonging to the sub-family of structural approaches.[123] As mentioned previously, these methods require a 3D structure model of a target protein in order to evaluate potential interactions between the newly designed entity and the binding site. One of the first instances of a receptor-based *de novo* design program was the software HSITE developed in 1989.[177, 178] The software considered potential hydrogen bond interaction sites of the receptor which then guided the 2D rule-based

automated *de novo* design by fitting and clipping the molecular skeletons of new ligands. One year later, an extension of this method included also the 3D information of molecules.[179] In his seminal work, Böhm developed the software tool LUDI which included two of the most established generation methods in *de novo* design, namely linking or growing molecular fragments into novel ligand structures.[180] LUDI consisted of an empirical scoring function which derived from a weighted sum of all individual ligand-receptor interactions. The software took into account hydrogen-bond, electrostatic and hydrophobic interactions. To obtain weights for each single interaction, a regression analysis of free-energy contributions from known receptor-ligand complexes was performed. This dependence on known structures with determined binding affinities, limits the applicability of these scoring functions to certain ligand and receptor types.[114] However, in one of the first successful prospective applications of *de novo* design, Babine *et al.* generated novel bioactive compounds with the

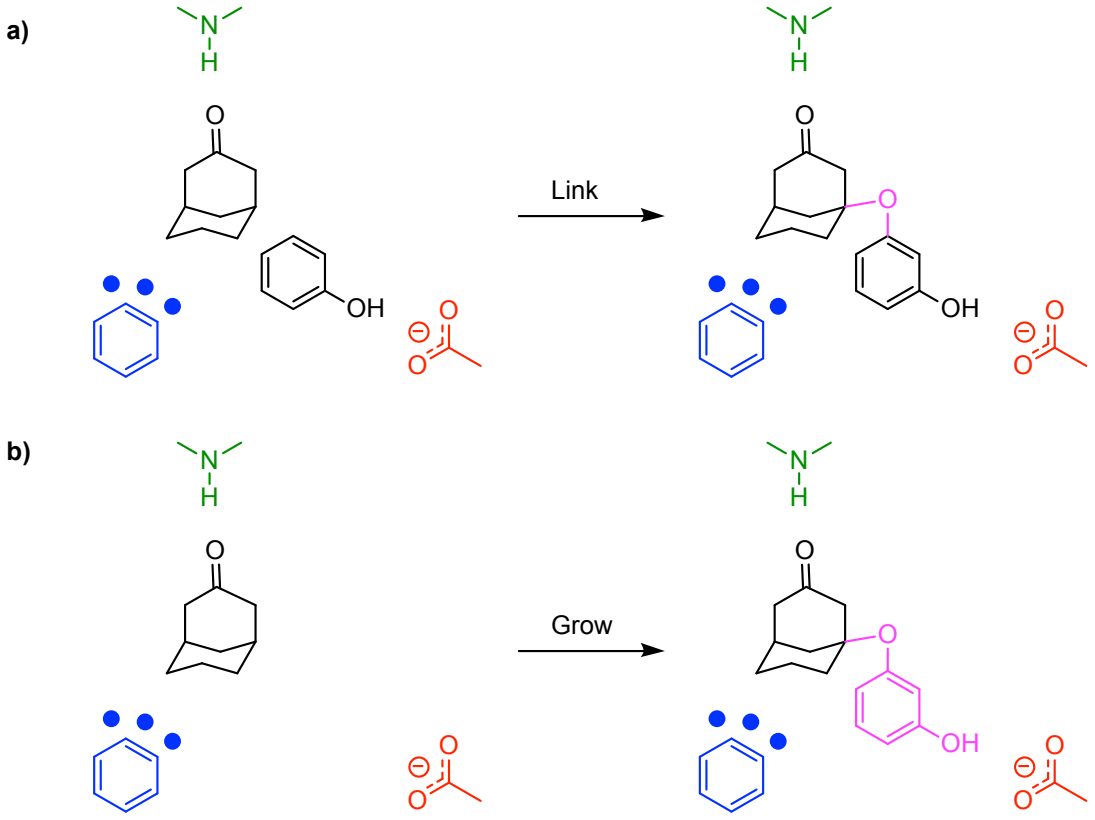



**Figure 1.19:** Overview of two widely used molecule construction methods in *de novo* design. **a)** In a fragment linking approach, relevant fragments are placed into the binding pocket considering potential interaction sites (green = hydrogen-bond donor, red = hydrogen-bond acceptor, blue = hydrophobic interactions) within the receptor which have been derived from a 3D protein structure. Subsequently, the two fragments are linked together (magenta). **b)** Fragment-growing starts with placing a first fragment into the binding site followed by an extension of the molecular structure of this start fragment with other fragments (magenta) to gather more interactions with the receptor. Reprinted and adapted with permissions from ref. [114].

software LUDI.[181] They designed micromolar inhibitors of the FK506-binding protein (FKBP12) with a fragment linking method (Figure 1.19a). The outcome and success of both generation strategies strongly depends on the quality of the underlying fragment libraries. The selected fragment set influences on one side the complexity of generated *de novo* designs, and by the same token it determines the synthesizability of the computationally generated compounds. An example from Brancale *et al.* highlighted the importance of selecting an appropriate fragment library (Figure 1.20).[182] They utilized the receptor-based software LigBuilder[183] to construct *de novo* designs from two different sets of fragments to find novel hepatitis C virus (HCV) helicase inhibitors. The derived design structures **51** and **52** had a huge difference in terms of complexity and synthetic feasibility. The second less complex design was further converted into a potent synthetically accessible inhibitor **53** with activity in the nanomolar range. Receptor-based *de novo* design concepts are reliant on structural information about the target protein and its binding site, and these methods are limited to cases in which this information is given. Because of this limitation, ligand-based tools have become very popular over the last two decades. The prerequisite of ligand-based *de novo* design is the identification of one or more bioactive ligands which serve as template to design similar molecular structures. Most ligand-based *de novo* design tools employ topological graph-based representations of molecules to generate novel structures.[114] The early software tools searched novel compounds with the help of evolutionary algorithms.[114] These algorithms iteratively optimize designs created temporarily *in silico* to increase their individual scores, and, after a certain number of design iterations, resulting in a candidate with a high score. For instance, Schneider and co-workers implemented the software TOPAS (topological assigning system), a ligand-based *de novo* design tool, based on an evolutionary algorithm, for finding novel similar structures.[184] In a retrospective application, they showed that starting from randomly assembled molecules (e.g. **54**) TOPAS was capable of modifying the intermediate structures (e.g. **55**) into a final molecule (**56**) with a high similarity to the original reference compound (**57**). (Figure 1.21).

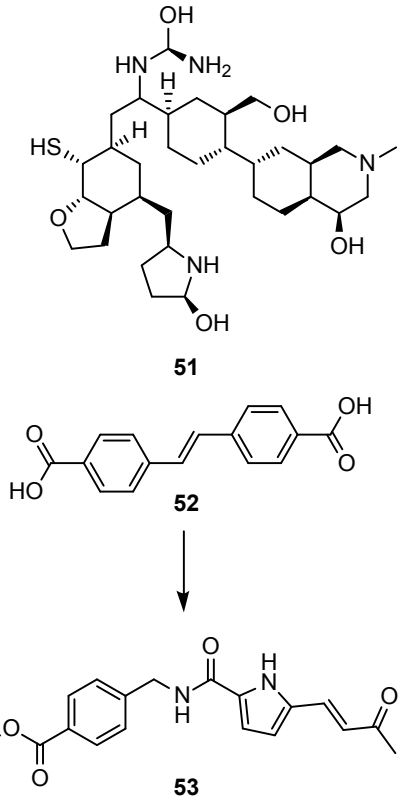



**Figure 1.20:** Various sets of fragments lead to *de novo* designs **51** and **52** with different complexity and synthetic feasibility. **52** was modified to a potent hepatitis C virus helicase inhibitor (**53**). Reprinted and adapted with permissions from ref. [114]

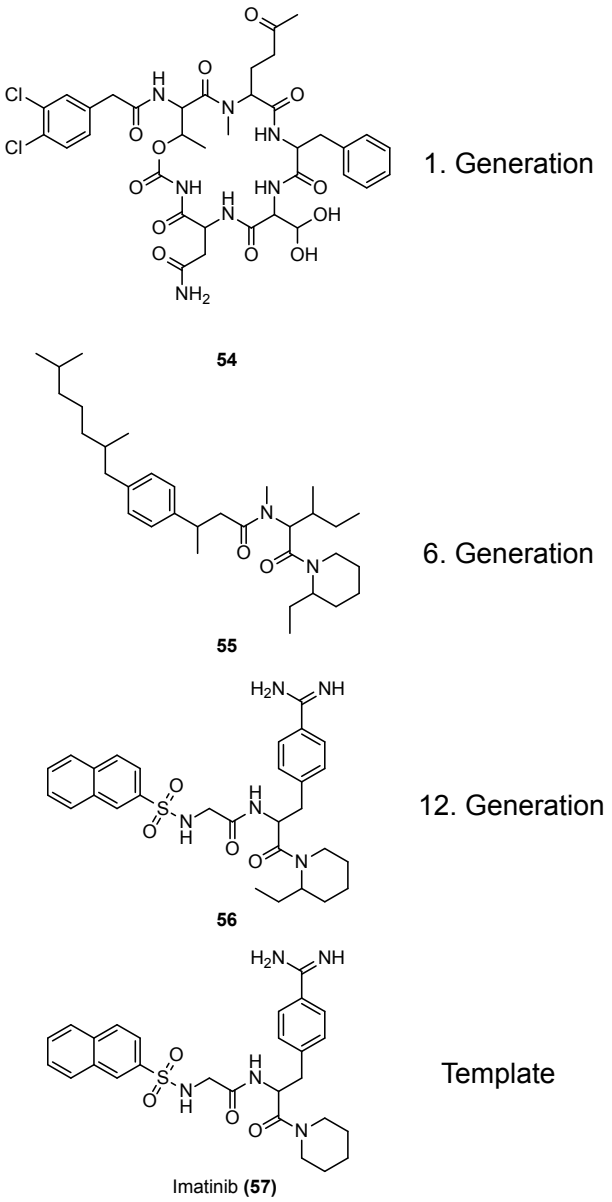



**Figure 1.21:** The optimization of an initial randomly assembled molecule **54** into designs with increasing similarity (**55** and **56**) to the reference compound imatinib ((**57**)) by the software TOPAS. TOPAS is based on an evolutionary algorithm to obtain compounds of higher similarity in an iterative process.

In a prospective study, Schneider *et al.* designed NCEs with TOPAS starting from a known potassium ion channel blocker.[185] They identified two new compounds with comparable bioactivity to their template structure. Still, the activity of the design was around two orders of magnitude less active then the template.

During the last 15 years, several *de novo* design software tools have been published which took further design constraints into account. In order to assess a design compound's bioactivity, it first must be synthesized. Therefore, constraints to ensure synthesizability of *de novo* designs have been recognized as important and necessary prerequisites of successful *de novo* design.[114] For instance, the aforementioned TOPAS software used a small set of pre-defined virtual retro-synthetic reactions to generate the building blocks, which were assembled in a forward virtual reaction to produce the design compound.[184] Vinkers *et al.* developed a *de novo* design software called "SYNOPSIS", which was built on a database of commercially available building blocks and a virtual reaction library of over 70 different organic reactions.[186] In the end, the software proposed a synthetic route of each generated design molecule based on the incorporated reactions. Similarly, Hartenfeller *et al.* implemented the *de novo* design tool DOGS (desgin of genuine structures) as a reaction-driven strategy to virtually assemble NCEs from a given template structure.[187] Based on 83 implemented reactions and a commercially available building block library of 25,144 entries, DOGS constructs *de novo* designs with a virtual synthesis approach and calculates the similarities between each computational entity and the respective template molecule by using the iterative similarity optimal assignment kernel (ISOAK) method (Figure 1.22).[188, 189] This algorithm relies on a graph representation of

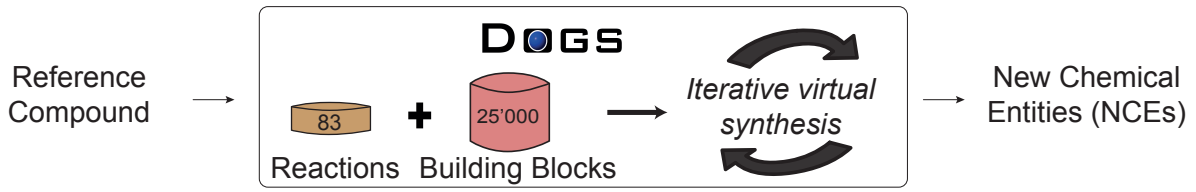



**Figure 1.22:** Schematic representation of the *de novo* design software DOGS (design of genuine structures). The tool employs a single ligand as reference structure and constructs de novo designs by an iterative virtual synthesis based on 83 implemented reaction and pre-defined 25,000 building blocks.

molecules with atoms as graph vertices and bonds as edges. The algorithm aims to *"optimally assigns the vertices between graphs based on pairwise vertex similarity"* as described by Rupp *et al.*[189] The implemented parameter $\alpha$, ranging from 0 to 1, can be used to regulate the influence of the graph neighborhood. Low $\alpha$ values lead to a direct and vertex-based graph similarity, while high $\alpha$ values results in an increased impact of the neighbor vertices (neighborhood-based).[190] This adaptable parameter can be defined by the user to either generate more structural analogs of the template molecule with smaller $\alpha$ values, or to design compounds that are structurally more distinct compared to their template. To generate structurally more distinct molecules while maintaining a certain degree of molecular similarity is associated with the concept of "scaffold-hopping",[176] a term which can be described as the *identification of isofunctional molecular structures with significantly different molecular backbones*.[6] For the calculation of similarity, molecules are represented as either molecular or reduced graphs. The algorithm starts by evaluating a set of promising start building blocks. From the fragment with the highest score, the software detects all feasible reaction steps which are applicable, followed by a scoring of each individual reaction. In a second step, reaction suitable building blocks for the top-ranked reactions are assessed and the reaction product with the highest score is considered as the intermediate for the next design cycle. To avoid construction of overly large molecules, the molecular mass of a final virtual product has to lie within pre-defined boundaries which are related to the template molecular mass. As a second stop criterion, the number of virtually performed reaction steps is limited to an user-defined value. In a proof-of-principle study, Spänkuch and co-workers applied the *de novo* design tool DOGS to generate novel inhibitors of the inactive human Polo-like kinase 1.[191] Employing the previously identified nanonmolar inhibitor **58** as template,[192] DOGS created over 218 virtual compounds from 100 promising fragments. Out of the top-ranked designs, they synthesized from **59** via intermediate **60** the selected original DOGS design **61** in two steps (Figure 1.23). They confirmed that **61** inhibited the inactive state of the human Polo-like kinase 1 in a similar degree to the template structure **58**. Several further studies have documented that the *de novo* design tool DOGS produces valuable active compounds as starting points for future drug discovery projects addressing a

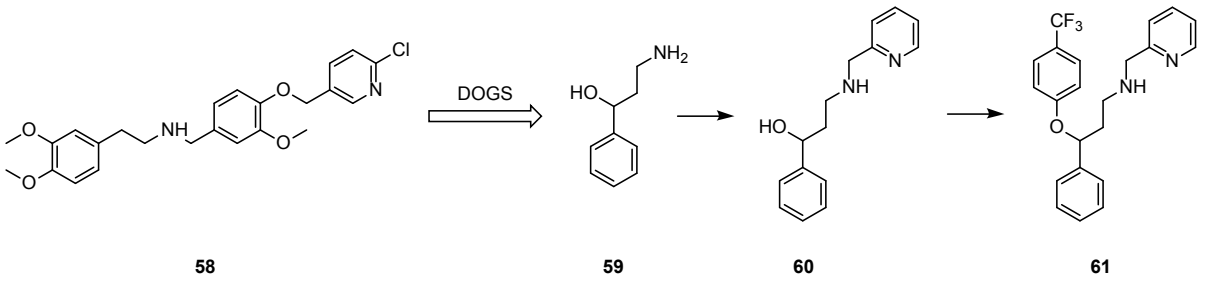



**Figure 1.23:** Inhibitor **58** of the inactive human Polo-like kinase 1 served as the template for the *de novo* design of novel chemical entities with the software tool DOGS (design of genuine structures). Starting from **59**, intermediate **60** was obtained from a reductive amination of **59** with 2-pyridinecarboxaldehyde. The second reaction, a O-arylation under Mitsunobu conditions yielded the original design **61**.[191]

broad range of diseases.[193–195] In all of these studies, synthetic drug-like molecules served as the reference compound in these *de novo* design studies. In 2016, the Schneider group extended the *de novo* design principle to natural products.[196] They selected the anticancer sesquiterpene (-)-englerin A (**62**),[197] a known activator of the transient receptor potential canonical channel C4/C5 (TRPC4/5),[198, 199] as a natural product template. The 903 designs generated from DOGS we further ranked according to their topological pharmacophore similarity using the previously described CATS descriptor.[147] Two virtual products (**63** and **64**) out of the 50 top-ranked designs were selected and further investigated. To ease the synthesis, the two designs **63** and **64** were manually simplified to the mimetics **65** and **66**, which were synthesized within two and three reaction steps, respectively (Figure 1.24). Both mimetics showed inhibitory effects on the transient receptor potential melastatin 8 channel (TRPM8), a known off-target of (-)-englerin a.[199] Mimetic **66** inhibited TRPM8 in a concentration range comparable to the activity of the parental natural product. However, it showed

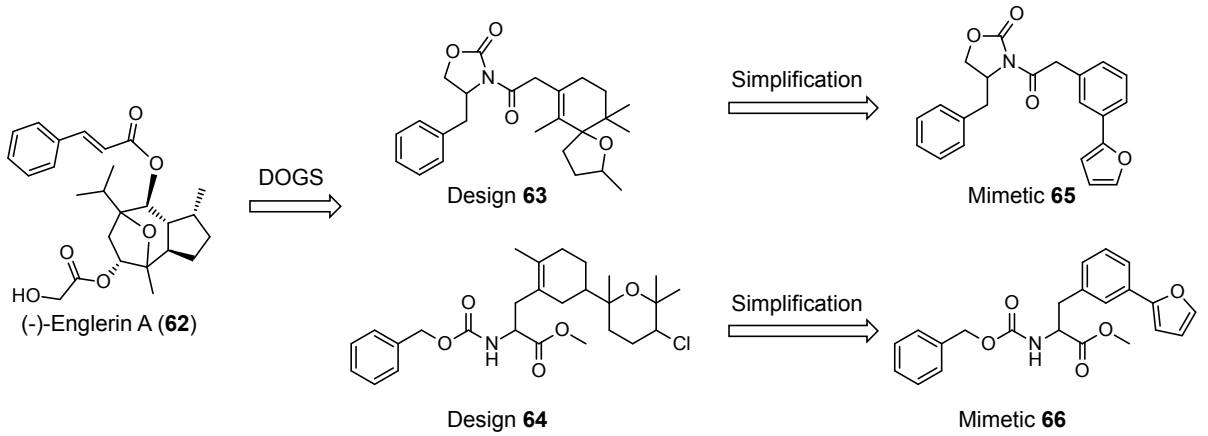



**Figure 1.24:** (-)-Englerin A (**62**) was employed as a reference natural product for the *de novo* design tool DOGS. After assessing their topological pharmacophore similarity, the two chosen designs **63** and **64** have been simplified to the synthetically-accessible mimetic structures **65** and **66**.

no activation of the TRPC4/5 ion channels, the main target of (-)-englerin a (unpublished data). Moreover, **66** was inactive in a screen against 60 human cancer cell lines (NCI-60).[200] Nevertheless, this study represented one of the first combinations of natural products and computational *de novo* designs in which synthetically-accessible and bioactive NCEs mimicking a structurally-complex natural product were produced.
Recently, Merk and co-workers integrated virtual screening and *de novo* design into a workflow which aimed to identify novel bioactive natural products from a large natural product database. Natural products selected from a virtual screen served as templates for the *de novo* design of mimetic compounds.[201] Three novel natural products with agonistic activity on the nuclear hormone receptor retinoid X receptor (RXR) were discovered from virtual screening. Together with two already known RXR active natural products, these five natural products served as the templates for five independent *de novo* design runs with DOGS. *De novo* designs were ranked computationally and selected according to their individual ranks and synthesizability. Eight structure were synthesized and characterized *in vitro*, resulting in the identification of NCEs with potencies in the low micromolar range.
The last two studies emphasize the huge potential of automated *de novo* design methodologies in finding small molecules mimicking bioactive natural products. These tools provide access to synthetically feasible structures which contain important inherited pharmacophoric features of the parent natural product. The introduced de novo concept has been applied to natural products in further studies as depicted in section 4.1, 4.2 and 4.3.
Assessing the feasibility of proposed synthetic pathways requires a certain level of human intervention and expertise. However, the rapidly-emerging trend towards the automation of many processes in drug discovery will help to overcome some of the challenges and limitations encountered in past attempts to make these systems useful in a practical, commonplace manner.[202]
An example where recent developments in the last years led to a more computationally automated process, is the research field of *in silico* target prediction.[203] In the following section, the underlining concepts and different methods of computational target prediction will be introduced. Selected examples will highlight the utility of target prediction efforts in the development of *de novo* design concepts in the context of natural products.

## Computational Target Prediction

*In silico* target prediction, virtual screening and computational *de novo* design share a common foundation: All three tools rely on the similarity principle stating that structurally similar molecules have similar biological effects.[171] The two previous sections described the utility of molecular similarity to identify and design novel bioactive molecules. However, *in silico* target prediction directly relates molecular structure to biological effects and therefore considers both structural similarity and estimated similarity in terms of bioactivity.[203] As with virtual screening and *de novo* design, *in silico* target prediction methods can be broken down into two main categories: ligand-based and receptor-based target prediction. Underlying concepts, developed tools and applications of both categories have been extensively summarized in the last few years.[203–206] Here, ligand-based target prediction will be discussed in more detail with a focus on its applicability to natural products and natural product mimetics.
The central assumption that similar ligands binds the same or similar targets requires a ligand description which captures important features of the structure-activity relationship and a metric that can accurately distinguishes between different bioactive ligands. Both requisites also apply for virtual screening, and suitable *de novo* design methods, and have been pre-discussed at length in section 1.2. Keeping in mind that multiple descriptors together with several similarity metrics can be applied in an almost endless number of combinations, many target prediction tools have been developed which utilize one of these approaches.[203] For instance, the Shoichet group implemented the *in silico* target prediction tool SEA (similarity ensemble approach) which predicts potential macromolecular targets from a 2D-fingerprint representation and a Tanimoto similarity calculation.[207] Known bioactive ligands and their related annotated targets were extracted from several databases (inter alia ChEMBL)[208] and integrated as reference structures. In a case study from the same group, Laggner *et al.* employed computational target prediction to identify the potential targets of 586 compounds, which showed activity in zebrafish behavior assays.[209] 11 out of 20 tested compounds displayed activity against predicted targets with potency values between 1 and 10,000 nM. They further specified that the predictions from SEA do not explicitly consider potency in their similarity assessment and therefore the predictions only indicate the likelihood of a possible ligand-target interaction at an appropriate concentration. In 2014, Gfeller and co-workers released the *SwissTargetPrediction* software implemented as a free webserver tool.[210] The prediction tool combines 2D and 3D ligand properties and uses a curated subset of ChEMBL as its reference database.[208] The software further allows for the target prediction of five different organisms, human ,mice, rat, cow and horse. Another ligand-based target prediction software is SPiDER (self-organizing map based prediction of drug equivalence relationships), as developed by

Reker and co-workers.[211] In this prediction method, molecular structures were represented in two different ways, by means of the topological pharmacophore descriptor CATS,[147] and by a collection of physicochemical descriptors which were part of the Molecular Operating Environment (MOE) software. Both types of representation were applied to describe query molecules and reference structures obtained from the manually annotated data set COBRA (collection of bioactive reference analogs).[212] The reference ligands were clustered and projected onto a 2D map using a self-organizing map algorithm, a variant of artificial neural networks which reduces the high dimensional space of the reference ligands to a two-dimensional map.[213] Each query compound is projected onto this map, assigned to exactly one cluster, preserving topological characteristics of the relationships between molecules in that higher-dimensional space. Only the targets of reference ligands from this cluster are counted. The Euclidean distance were calculated between all reference compounds in this cluster and the query molecule and transformed into probability values ($p$ values). Additionally, a background distribution as false-positive estimation is computed by considering the Euclidean distances between all pairs of reference compounds which were annotated to not bind to the identical target. Averaging all probability values provides a good estimation of the likelihood of a significant target assignment. In a prospective application, Reker *et al.* predicted the macromolecular targets of the anticancer natural product archazolid A.[214] They manually fragmented the complex macrolide into four segments (natural product-derived fragments) and predicted potential targets for each segment and the original natural product. They observed a high number of predicted proteins related to arachidonic acid pathway. They then screened archazolid A against eight predicted targets and measured *in vitro* effects on four out of eight investigated targets. They concluded that computational ligand-based target prediction is capable of identifying novel targets for natural products or natural product-derived fragments without any explicit information about protein structure and binding site. The SPiDER prediction tool has been applied in further studies to predict targets of natural products and natural product-related compounds.[196, 201, 215, 216]
These results substantiate the fact that *in silico* target prediction offers an efficient way to select potential protein targets for further experimental characterizations. Together with virtual screening and *de novo* design methods, target prediction plays an important role in the development of a fully computer-assisted and automated design-make-test cycle.[202] The prediction tools discussed above are knowledge-driven methods, meaning that they depend strongly on already discovered bioactive ligands combined with information about their related targets. These tools have the advantage that they do not rely on information about the protein structures, relevant binding pockets and related potential interaction points. What might seems like a disadvantage of these tools at first glance can be of great benefit in future drug discovery, in particular in the field of natural product-inspired drug discovery with technical advances in areas like

high-throughput assays and metabolomics.[16]
Novel technologies from the field of artificial intelligence (AI) and machine learning (ML) have entered the field of drug discovery and medicinal chemistry. Some of these methods, which have been applied over the last 50 years, turned into standard tools in certain areas.[217] Other concepts like deep learning or generative models are rather new in drug discovery.[218] In the closing section of the introduction, some machine learning concepts will be highlighted which can applied in the *de novo* design of drug-like molecules.

## Artificial Intelligence in *De Novo* Design

In the last few years, artificial intelligence (AI) and machine learning (ML) experience a renaissance in drug discovery. Although some of the currently applied machine learning tools have been developed already in the 1960s,[219] the impressive growth of accessible computational power and the remarkable increase of available data have caused a growing interest in these technologies.[220] In computer science, AI is commonly described as a rational agent which acts "*so as to achieve the best outcome or, when there is uncertainty, the best expected outcome*", as indicated by Russell and Norvig.[221] In other words, algorithms which utilizing AI have the objective to find the best solution for a certain task based on knowledge about known solutions. ML is a field in artificial intelligence that employing statistical methods for pattern recognition for all kinds of data.[222] The identified patterns form the basis of further predictions of previously unseen data. A widely used class of ML algorithms are artificial neural networks (ANNs) which are computational networks mimicking simplified version of information processing within the brain.[223] As the name implies, these networks consist of artificial neurons which are a simplified mathematical description of a biological neuron (Figure 1.25).[224]

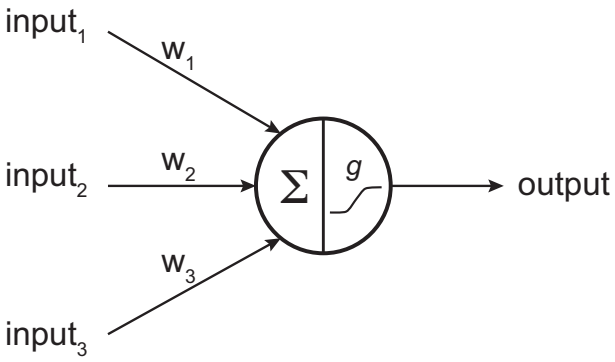



**Figure 1.25:** Simplified representation of an artificial neuron. The different input values are individually weighted ($w_1$ - $w_3$) indicating their specific importance. These weighted inputs are summed up ($\Sigma$), translated via an activation function ($g$) to an output value and this output is transmitted to the next neuron.

This concept dates back to 1943, when McCulloch and Pitts reported that the activity of nerve cells in the human nervous system can be described by a simplistic mathematical model.[225] Such an artificial neuron receives input values which have individual weights (e.g. $w_1$ - $w_3$) representing the importance of each input. These weighted inputs are summed up ($\Sigma$) and then sent to an activation function ($g$) which translates

the sum into an output.[224] This activation function is in analogy to the "firing" rate of action potentials in biological neurons.[226] The derived output is further transmitted to the next neuron(s), where the output values of the previous neurons become the new input values. This simplest neural network, a so-called "perceptron", consists of only a single neuron. It was mentioned the first time by Rosenblatt in 1957.[227] The combination of several neurons, at least three, is called a multilayer perceptron or feed-forward neural network. The neurons of these networks are grouped into layers. Each layer can be described according to its function in such a network (Figure 1.26). An input layer transmits the machine-readable input values to the next layer, the so-called hidden layer. The input values, along with the corresponding weights, are then transformed into the activation values of each hidden layer neuron. These activation values then serve as the inputs for the output which produces the final output values.[123] In contrast to a feed-forward network, which only consist of connections to the next layer (Figure 1.27, left), a recurrent neural network (RNN) allows links between neurons of the same layer (Figure 1.27, middle). Neural networks with multiple hidden layers are described as "deep" neural networks (Figure 1.27, right), which form the basis of "deep learning" and its successful applications in various fields.[228]

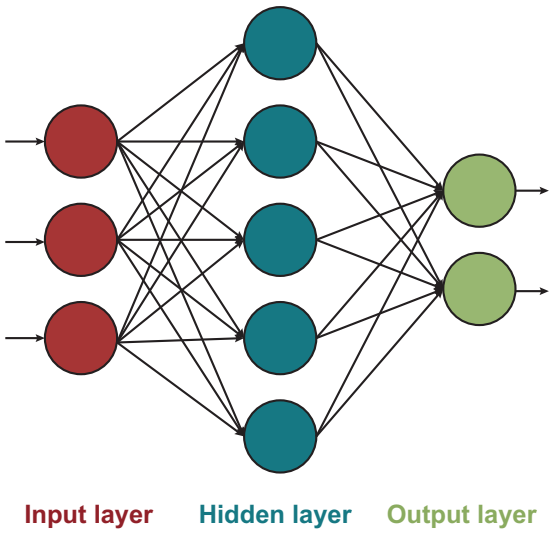



**Figure 1.26:** In a feed-forward neural network neurons are organized in layers, whereof each layer fulfills a certain function. The input layer sends the machine-readable input values to the hidden layer, which transforms and translates these data to the output layer that produces the predicted output values.

Within those network architectures the activation function plays an essential role in finding solutions to a given task. In 1969, Minsky and Papert proved that perceptrons with linear activation functions can only solve tasks which are linearly separable.[229] Roughly twenty years later, Cybenko showed that by using nonlinear activation functions, one can approximate any continuously valued function.[230] The introduction of nonlinear activation functions (e.g. sigmoid or tanh functions) extended the scope of these feed-forward artificial networks, especially in the area of chemistry and drug design.[231–233] Similar to virtual screening, *de novo* design and target prediction (section 1.2), ANNs and other machine learning methods require machine-readable representations of input data, for example molecular structures and biological activity. The selection of relevant features (feature selection) from these representations is a crucial step in solving tasks with machine learning algorithms.[234] In an early example from 1994, Tetko and co-workers trained a neural

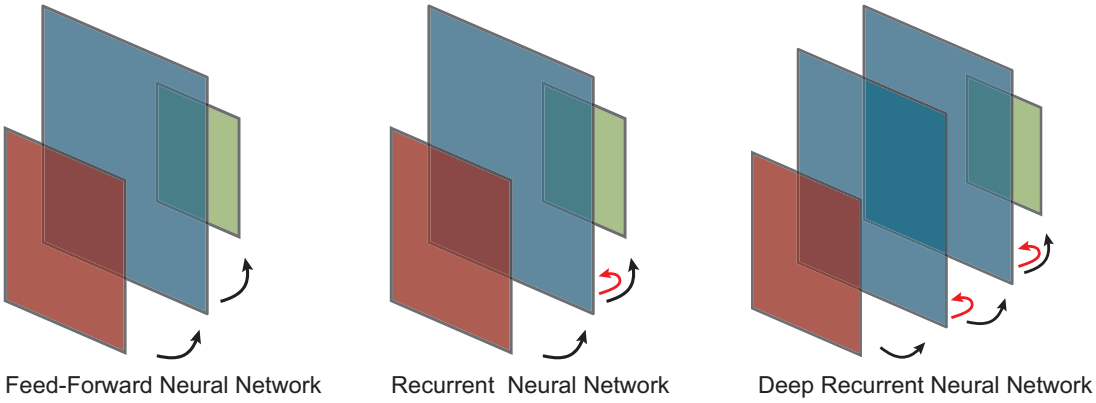



**Figure 1.27:** Schematic overview of different neural networks. Within each schema, every plane represents a layer colored according to its function (red = input layer, blue = hidden layer, green = output layer). All connections between (black) and within (red) of these layers are indicated by arrows. (left) A feed-forward network allows only connections to the next layers, whereas recurrent neural networks also contain links within neurons of the same layer (middle). (right) Neural networks with more than one hidden layer are called "deep" neural networks independent of the allowed connections.

network on a set of molecules with known inhibitory effects on the human immunodeficiency virus type 1 (HVI-1) reverse transcriptase. transcriptase. The model was trained to predict the bioactivities of a control set, with known activity values, and a set of new molecules, with unknown biological effects.[235] They confirmed that a molecule predicted "active" showed inhibitory effects on the HIV reverse transcriptase. Further QSAR modeling studies have been conducted in medicinal chemistry applying ANNs.[219]

In the last few years deep learning models have increased their popularity in drug discovery.[218, 220, 236, 237] In 2015, Ma *et al.* demonstrated that a deep learning model can be prospectively applied to predict bioactivity and ADMET properties.[238] In particular with large data sets, these deep learning methods outperformed other, more conventional, machine learning algorithms. Despite these advancements in QSAR modeling,[218] the generation of novel bioactive compounds by *de novo* based deep-learning, still remained unresolved until 2017. In their seminal work, Segler and coworkers developed a generative RNN model that had been trained on the string representations of molecules called SMILES (simplified molecular-input line-entry system).[239] This line notation of molecule encodes information about atom and bond types, branching points and ring system, as well as stereochemistry.[240] The model generate a probability distribution of the SMILES string characters. By sampling from this distribution, the model was able to generate entirely new molecules represented as SMILES strings. These generative RNNs have been applied before in various areas, for example in natural language processing.[241] In natural language processing, RNNs are used to generate text by predicting the next word in a sentence while considering the information from the previous words. In 1997, Hochreiter and Schmidhuber

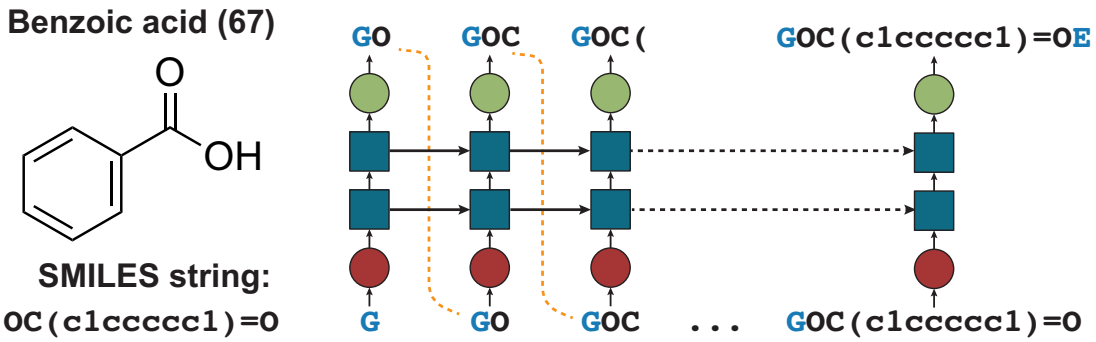



**Figure 1.28:** The structure of benzoic acid (**67**) and its SMILES (simplified molecular-input line-entry system) representation are shown (left). This string representation contains information about atom and bond types, branching points and ring systems, and stereochemistry. This SMILES string serves as input sequence of the LSTM-based RNN, which comprises an input layer (red), two hidden layers (blue) and an output layer (green). The pre-defined characters "G" and "E" indicate the beginning and the end of a given sequence. In a sequential manner, the network learns the probability of the next character in a given SMILES strings. The previous output character becomes the next input sequence as indicated by the orange dashed lines.

introduced a certain RNN architecture, the so-called long short-term memory (LSTM) cells.[242] These LSTM-based models are suitable to processing and predicting from sequences as for example in natural language processing.[243] Similar to words in a sentence, the SMILES string of a molecule is also represented by a sequence of characters. Segler *et al.* utilized such a LSTM-based network to generate novel molecules.[239] An example is depicted in Figure 1.28. The SMILES representation of benzoic acid (**67**) "OC(c1ccccc1)=O" serves as an input sequence for a LSTM recurrent neural network which learns the probability of the next character while considering information about the previous ones. This "remembering" of former characters in a string is a necessary requirement for SMILES generation. For example, parentheses indicate ring systems, and in the example of benzoic acid (**67**), the two parentheses are separated by several characters. Therefore, the model has to "remember" that at some point in the generation process a second parenthesis is needed to obtain a valid SMILES representation. The entire process of generating novel compounds with LSTM-based models can be separated into two steps. First, the model is trained on a large data set of molecules in order to learn the probability distribution of an appearing SMILES character given a set of previous characters. During this learning procedure, the error between the predicted and true output values is calculated. The weights within the network are gradually updated to decrease the estimated error by iterating multiple times over all training examples. After several iterations of minimizing this error, the model generates new compounds by sampling step-wise the characters of a SMILES string according to the learned probability distribution over all characters. This sampling ensures that the model not only memorizes structures from the training set, but also produces molecules which are not part of this set. A simplified overview of these two steps is shown in Figure 1.29. Recently, Gupta and co-workers demonstrated that this type of

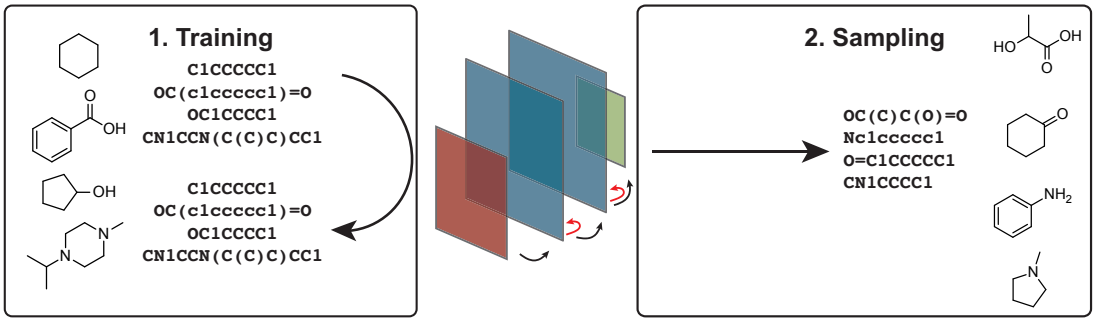



**Figure 1.29:** Visualization of the two-step design process to generate *de novo* designs by a generative deep learning model. The generative model comprises of an input layer (red), two hidden layers (blue), and an output layer (green). (*Left*) In a first step (**1. Training**), the model learns in an iterative procedure the probability distribution of all SMILES characters of a given training set. On the basis of this learned probabilities, the model is capable of sampling (**2. Sampling**) from this distribution and generating SMILES strings of molecular structures.

generative model is capable of creating sets of drug-like designs with a high proportion of valid SMILES representations.[244] They trained the model with over 500,000 highly active compounds extracted from ChEMBL.[208] They showed that sampled molecules possessed similar physicochemical properties compared to the training molecules from ChEMBl without explicitly specifying these properties during the training process. They further introduced a "fine-tuning" procedure which means that an already trained model is re-trained on a smaller dataset (fine-tuning set) to create a target-focused compound library with similar properties to the fine-tuning set This fine-tuning step, also called "transfer learning", enables a model to solve a new but still related task based on the information gathered previously.[245] An additional feature of their method is that it provides the possibility to define a structure as the starting point of the sampling process. This means that the model can generate new structures from this common starting fragment. Based on the promising theoretical results of the studies described above,[239, 244] two prospective applications have been conducted using generative deep learning models for de novo design of bioactive molecules (section 4.4).[246, 247] Besides the established screening techniques like high-throughput screening and virtual screening, these generative deep learning models as alternative knowledge-driven tools have the potential to identify novel pharmacologically active compounds in early drug discovery.

# 2 Aims of this Thesis

Natural products are one of the most inspiring sources for identifying novel bioactive molecules in drug discovery. Several limitations like synthetic difficulties, unfavorable physicochemical properties or insufficient potencies, render the development of potential drug candidates from natural products a challenging task. As computer-assisted drug design and in particular computational *de novo* design methods have matured over the past decades, they play an increasing role in discovering novel bioactive molecules.

Although these design approaches have been successfully applied to design active compounds from drug-like reference molecules, the combination of computational *de novo* design strategies deducting natural products as reference structures has not been thoroughly examined so far. Despite my first preliminary results,[196] a more comprehensive exploitation of computational *de novo* design relying on natural products as templates is still pending. To explore the potential of this combination, the following working hypothesis has been proposed: computer-assisted drug design methods can facilitate the design of novel bioactive chemical entities mimicking natural products (I) by employing reaction-driven computational *de novo* design and (II) by prospectively applying generative deep learning models .

With regard to (I), this thesis tests the working hypothesis by:

- Constructing virtual molecular structures with a reaction-driven and ligand-based *de novo* design tool using bioactive natural products as template structures;
- Picking computational designs by molecular similarity assessment between the virtual products and their template structures with two- and three-dimensional descriptors;
- Employing *in silico* target prediction to further prioritize designs that obtained predictions matching the known targets of the template, and to explore novel molecular targets of the template and its mimetic compounds;
- Synthesizing these computationally designed natural product mimetics and characterization of their *in vitro* biological activities on identified and predicted targets.

In the context of the stated hypothesis, the suitability of generative deep learning models as design methods (II) to construct novel bioactive natural product mimetics is verified. This part of the thesis claims to:

- Identify novel bioactive small molecules designed by a generative deep learning model which has been trained on molecular structures of drug-like synthetic compounds;
- Experimentally confirm computationally designed natural product mimetics.

Via the approaches (I) and (II) this thesis generally aims to apply computational *de novo* design methods to identify bioactive new chemical entities mimicking natural products.

# 3 Materials and Methods

Relevant experimental and computational methods for all studies are summarized in this chapter.

## 3.1 Laboratory Methods

### Chemical Synthesis

#### Nomenclature

Chemical structures were converted into their corresponding IUPAC name using the software ACD/ChemSketch (v2018.1.1).

#### Reagents

All reagents were purchased in the highest available purity from several suppliers (ABCR, Acros Organics, Alfa Aesar, Apollo, Aldrich-Fine Chemicals, Fluorochem, Tokyo Chemical Industry) and were used without further purifications.

#### Sovlents

Absolute and anhydrous solvents for reactions were purchased from commercial suppliers and used without further purification. Solvents of technical grade were used for extractions and flash chromatography.

#### Reactions

All reactions were carried out in oven-dried glassware (110°C). Reactions under air- and water-free conditions were performed under a nitrogen or argon atmosphere.

#### Microwave-asssisted Reactions

The microwave-assisted reactions were performed on a Biotage Initiator Classic 2.5 (Biotage AB, Uppsala, Sweden) with a robot Eight vial rack.

#### Solvent Evaporation

Solvent evaporations in vacuo were performed at 10-850 mbar at 40-60°C.

### Yields

Yields were reported for spectroscopically and chromatographically pure compounds after drying under high vacuum (<$10^{-2}$ mbar).

## Purification Methods

### Flash chromatography (FC)

Purifications by flash chromatography or plug filtrations were carried out with silica gel from SiliCycle (40-63 µm, (230-400 mesh), SiliCycle Inc., Quebec, Canada).

### Thin-Layer chromatography (TLC)

Thin-layer chromatography (TLC) was used to monitor reactions by using Merck silica gel ($SiO_2$) 60 F254 plates. UV light (254 nm or 366 nm) or staining with a potassium permanganate solution (3 g of $KMnO_4$, 20 g of $K_2CO_3$ and 0.25 g of NaOH in 400 mL $H_2O$) was used to visualize TLC spots.

## Analytical Methods

### Thin-Layer chromatography mass spectrometry (TLC-MS)

Mass spectra of TLC spots were recorded on an Advion Expression$^s$ compact mass spectrometer (CMS) (Advion, Ithaka, NY, USA) connected with an Advion Plate Express TLC plate reader (Advion) using electron spray ionization (ESI).

### Analytical High-Performance Liquid Chromatography

The purity of compounds for biological evaluations were determined either on a VWR Hitachi LaChrom Elite 2000Series HPLC (VWR Internationonal, Lutterworth, Leicestershire, United Kingdom) equipped with a Nucleodur C18 Htec analytical column (150mm x 3mm ,5µm, 110 Å) (Macherey & Nagel, Dueren, Germany) or on a Shimadzu LCMS-2020 with a Nucleodur C18 HTec analytical column (150×3 mm, 5 µm, 110 Å). Products were eluted with different linear gradients of acetonitrile (+ 0.1% formic acid) in double-destilled water (+ 0.1% formic acid) over 16 or 25 min with a flow rate of 0.5 ml/min. All compounds had a purity of higher than 95%.

### $^1$H and $^{13}$C nuclear magnetic resonance spectroscopy (NMR)

Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker AV 400 or on a Bruker AV 500 spectrometer (Bruker Corporation, Billerica, MA, USA). Chemical shifts ($\delta$) are given in ppm relative to tetramethylsilane (TMS). The multiplicity of each

proton signal is indicated (s = singlet, d = duplet, t = triplet, m = multiplet, and combinations thereof) and the corresponding coupling constants (J) are reported in Hertz (Hz).

### High-resolution mass spectra (HR-MS)

High-resolution mass spectra (HR-MS) was performed by the MS service of the Laboratorium für Organische Chemie ETH Zürich. High resolution electrospray-ionization (HR-ESI) spectra were measured on a Bruker maXis -ESI-Qq-TOF-MS spectrometer (Bruker Corporation, Billerica, MA, USA). High resolution matrix-assisted laser desorption/ionization (HR-MALDI) spectra were measured on a Bruker solariX 94 spectrometer (Bruker).

### Optical Rotation

Optical rotation [α] was measured on an MCP 300 Polarimeter (Anton Paar, Radnor, PA, USA) equipped with Anton Paar 100 mm cuvette (CL.0.01, 3 mm, Ni Alloy, 320157-35). The specific rotation (20°C, λ = 589 nm) is reported as $[\alpha]_D^{20}$ at a concentraction (c) [g/100 ml] in the solvent referred.

### Melting point (MP)

Melting points (MP) were measured on a Büchi Melting Point M 560 and were reported in degree Celsius.

### X-Ray Structure Analysis

X-ray structures were determined by the ETHZ Small Molecule Crystallography Center.

### Dynamic Light Scattering

Dynamic light scattering (DLS) was performed on a 90Plus Particle Size Analyzer (Brookhaven Instruments Corp., Holtsville, NV, USA) to determine aggregation potential of molecules. Particle sizes were measured at 25°C with default settings for water. The dust filter parameter was set to 50. A stock solution of test compound was prepared in DMSO at a concentration of 5 mM. This stock solution was diluted with double-distilled water to a concentration of 500 µM before measurement.The solution was further diluted up to a concentration, where no aggregation was observed over a time period of 60 min.

## 3.2 Computational Methods

### Data Processing, Data Analysis and Data Visualization

Data was processed and analyzed with Microsoft Excel for Mac (version 16.16.2), with the KNIME analytics platform (version 3.2.1)[248] or in Python (version 3.6) [249] with the packages `pandas` (version 0.23.4)[250], numpy (version 1.15.2),(REF) matplotlib (version 2.2.3)[251], seaborn (version 0.9.0),[252] scikit-learn (version 0.20.0)[253] managed with the Anaconda Software Distribution (version 4.5.11, www.anaconda.com) and written and executed in PyCharm Community (version 2017 and 2018.1) or in Jupyter Notebooks (version 1.0.0).[254] Molecular data was processed and analyzed with the *Molecular Operating Enviroment* (MOE) software (v2016.08, Chemical Computing Group, Montreal, Canada) and with *DataWarrior* (v4.7.1).[255] Bioactivity was visualized in Prism (version 7, GraphPad Software, San Diego, USA). Dynamic light scattering analysis was done in the BIC particle size software (version 5.28, Brookhaven Instuments Corporation, USA). Figures for publications were designed in Adobe Illustrator CS6 (v16.0.0) and the open-source software Inkspace (0.90).

### Molecular Property Calculations and Scaffold Analysis

Molecular properties were calculated using RDKit in Python (version 3.6) or RDKiT nodes in KNIME. Scaffolds were extracted as molecular scaffolds or as graph frameworks (ignoring atom and bond types) as described by Bermis and Murcko.[72] The scaffold extraction was executed in RDKit in Python (version 3.6) or RDKiT nodes in KNIME. Synthesizability scores[256] and the natural product likeness indices[257] were calculated with the scripts in Python (v 3.6) provided by RDKiT.

### *De Novo* Design by DOGS

*De novo* designs for a given template structure were generated by the software tool DOGS (desgin of genuine structures)[187]. Every DOGS run performed started from 200 fragments and nine independent construction runs with nine different $\alpha$-values, were performed in the range from 0.1 to 0.9 with a step size of 0.1.

### Topological Pharmacophore Similarity Assessment with CATS

To assess pharmacophore similarity between two molecules, depicted as 2D molecular graphs, we applied the topological pharmacophore descriptor CATS2 (Chemically Advanced Template Search, Version 2).[147] The Euclidean distance between two CATS vectors of two different molecules was used as a pharmacophore similarity estimation, wherein a lower distance indicates a higher degree of similarity. The CATS2 descriptor

was used as implemented as KNIME node. The Euclidean distance was calculated by the in-house developed KNIME node *DistanceCalculator*.

### Computational Target Prediction by SPiDER

The software tool SPiDER (Self-organizing map-based PredIction of Drug Equivalence Relationships)[211, 214] was applied to computationally predict macromolecular protein targets of a query molecule. The SPiDER software was used as implemented as KNIME node. Only target predictions with a $p$ value lower than 0.05 were considered as predicted active.

### Generative Deep Learning Models for *De Novo* Design

A generative recurrent neural network was applied to design new chemical entities as described by Gupta et al.[244] The model used consists of two LSTM layers[242] each with a hidden vector size of 256, which are regularized with a dropout function. The purpose of a dropout regularization is to ignore randomly selected neurons during the training procedure. This technique leads to a better generalization of the model and avoids overfitting.[258] Due to the fact that the input sequences were translated by one-hot encoding into a vectorized format the neural network comprises of a softmax function as final layer to transform these vectors to values ranging between 0 and 1, in which all values sum up to 1. This results in a probability distribution for all characters in the given sequences. To apply a the LSTM model to *de novo* design, a library of bioactive known compounds from ChEMBL22[208] with an annotated affinity below 1 µM (541'555 entries) were represented in their SMILES strings and these sequences were utilized to train the model. As a second step, this trained model was used as a basis for fine-tuning the neural network with a smaller data set by transfer learning. During the fine-tuning, the model keeps information from the initial training and tries to solve a different but related task.[245] Additionally, a pre-defined SMILES string can be employed as fixed starting fragment during the sequence generation process ("sampling") like in a fragment-based drug discovery approach.[110] The implementation of the LSTM model was done in Tensorflow (version 1.2)[259] and Keras (version 2.0)[260] in Python (version 3.6).[249] Molecules were converted into SMILES strings in RDKiT[261] in Python (version 3.6).

## 3.3 *In vitro* Biological Assessments

### Bioactivity Screening Assays

Selected biological assays were conducted by Eurofins Cerep SA (France) or Eurofins Panlabs (USA) on a fee-for-service basis. Detailed assay protocols can be found at

https://www.eurofinsdiscoveryservices.com . Initial screening measurements were carried out with two replicates at concentrations of 10, 30 or 50 µM. $IC_{50}$, ($K_i$) and $K_B$ values were determined by the service provider based on eight different concentrations with two replicates for each concentration. Log(concentration) response curves (four-parameter logistic curves) were plotted in Prism (version 7). The inhibition constants ($K_i$) were determined according to the Cheng Prusoff equation[262]

$$K_i = \frac{IC_{50}}{1 + (\frac{C}{K_D})} \quad , \tag{3.1}$$

where C is the concentration of the radioligand in the specific assay and $K_D$ the affinity of the radioligand for the receptor. For antagonists, the apparent dissociation constants ($K_B$) were calculated with the modified Cheng Prusoff equation[262]

$$K_B = \frac{IC_{50}}{1 + (\frac{C}{EC_{50.C}})} \quad , \tag{3.2}$$

where C is the concentration of control activator in the assay and $EC_{50.C}$ its $EC_{50}$ value.

## Cyclooxygenase (COX) Assays

Cyclooxygenase (COX) inhibition assays were performed by the research group of Prof. Oliver Werz at the Friedrich-Schiller-University Jena, Germany. Andreas Koeberle and Oliver Werz designed the experiments, analyzed and interpreted the assay results. All protocols for experiments with human blood and blood cells were approved by the ethical commission of the Friedrich-Schiller-University Jena.

### Activity assays of isolated cyclooxygenase-1 and -2 (COX-1 and -2)

Purified bovine COX-1 (Cayman Chemicals; 50 units) or human recombinant COX-2 (Cayman Chemicals; 20 units) in 100 mM Tris buffer pH 8, 5 mM glutathione, 5 µM hemoglobin and 100 µM EDTA were pre-incubated with test compounds for 5 min at 4°C followed by 1 min at 37°C. Then, arachidonic acid (AA, 2 µM for COX-2 and 5 µM for COX-1) was added and incubations were continued for another 10 min at 37°C. Formation of COX-derived 12(*S*)-hydroxy-5-cis-8,10-*trans*-heptadecatrienoic acid (12-HHT) from AA was analyzed by RP-HPLC as described.[263]

### Determination of COX-1 activity in washed platelets

Freshly isolated human platelets ($1\times10^8$) were pre-incubated with test compounds for 5 min at room temperature. Formation of COX-1-derived 12-HHT was initiated by addition of exogenous AA (5 µM). After 5 min at 37°C, 12-HHT was extracted and separated by RP-HPLC as described.[263]

### Isolation of monocytes from human blood

Monocytes were freshly isolated from human leukocyte concentrates, which were provided by the Institute for Transfusion Medicine of the University Hospital Jena (Germany) as described.[264] Venous blood was collected in heparinized tubes (16 I.E. heparin/ml blood) from fasted (12 h) adult (18–65 years) male and female registered healthy volunteers, with informed consent. These subjects donated blood every 8 to 12 weeks, had no apparent infections, inflammatory conditions, or current allergic reactions (according to prior physical inspection by a clinician) and had not taken antibiotics or anti-inflammatory drugs for at least 10 days prior to blood collection. Leukocytes were concentrated by centrifugation of the freshly withdrawn blood (4,000×g/20 min/20°C) and centrifuged on lymphocyte separation medium (LSM 1077, GE Healthcare, Freiburg, Germany). Monocytes were isolated from the peripheral blood mononuclear cell (PBMC) fraction by adherence to culture flasks (Greiner, Nuertingen, Germany) for 1.5 h at 37°C and 5% $CO_2$ in RPMI 1640 medium (Sigma-Aldrich) containing FCS (Sigma-Aldrich; 5%), *L*-glutamine (Sigma-Aldrich; 2 mM) and penicillin/streptomycin (GE Healthcare; 100 U/ml and 100 $\mu$g/ml) (monocyte medium).

### Analysis of prostaglandin formation of activated monocytes

Freshly isolated monocytes ($5 \times 10^6$) were stimulated with lipopolysaccharide (1 $\mu$g/ml) for 24 h at 37 °C and 5% $CO_2$. Cells were washed, preincubated with the test compounds for 15 min, and treated with 5 $\mu$M AA for 15 min. Then, the reaction was stopped and COX-derived prostaglandins were extracted as described.[265]

## Electrophysiological Experiments

Electrophysiological measurements were carried out in the research group of Dr. Ursula Storch and Prof. Michael Mederos y Schnitzler from the Walther Straub Institute at the Ludwig-Maximilians-University Munich (Germany). Ursula Storch and Michael Mederos y Schnitzler designed the electrophysiological experiments, analyzed and interpreted the experimental results. Aaron Treder and Inderjeet Singh performed electrophysiological measurements under the supervision of Ursula Storch and Michael Mederos y Schnitzler.

### Electrophysiology of Transient Receptor Potential Ion Channels

Human embryonic kidney cells (HEK293T, ATCC CRL-3216) were maintained in Earl's MEM (Sigma-Aldrich, Taufkirchen, Germany), with 100 units $ml^{-1}$ penicillin and 100 $\mu$g $ml^{-1}$ streptomycin supplemented with 10% (vol/vol) FCS (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and 2 mM glutamine. All cells were held at 37°C in

a humified atmosphere with 5% $CO_2$. Cells were seeded into 6-well dishes and transiently transfected at confluency of about 90% using GeneJuice (Merck Millipore, Billerica, MA, USA) according to the manufacturer's protocol. Conventional whole-cell recordings were carried out at room temperature 15 hours after transfection with the human TRPM8 (NP_076985) in pCAGGSM2-IRES-GFP expression vector or 24 hours after transfection of the cells with the rat TRPC4 isoform TRPC4-beta1 (NP_001076584) in pIRES2-eGFP expression vector. The following bath solution containing 140 mM NaCl, 5 mM CsCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$, 10 mM glucose, 10 mM HEPES (pH 7.4 with NaOH) and resulting in an osmolarity of 295-302 mOsm $kg^{-1}$ was used. The pipette solution for TRPM8 measurements contained 130 mM CsCl, 5.792 mM $MgCl_2$, 0.524 mM $CaCl_2$, 10 mM BAPTA (5.5 nM free $Ca^{2+}$), 1 mM HEDTA (3 mM free $Mg^{2+}$) and 10 mM HEPES (pH 7.2 with CsOH), resulting in an osmolality of 296 mOsm $kg^{-1}$. The pipette solution for TRPC4 measurements contained 120 mM CsCl, 9.4 mM NaCl, 0.2 mM $Na_3$-GTP, 1 mM $MgCl_2$, 3.949 mM $CaCl_2$, 10 mM BAPTA (100 nM free $Ca^{2+}$) and 10 mM HEPES (pH 7.2 with CsOH), resulting in an osmolality of 296 mOsm $kg^{-1}$. Patch pipettes made of borosilicate glass (Science Products, Hofheim, Germany) had resistances of 2.0-2.8 MΩ for the whole-cell measurements. Data were collected with an EPC10 patch clamp amplifier (HEKA, Lambrecht, Germany) using the Patchmaster software (HEKA). Current density-voltage relations were obtained from voltage ramps from –100 to +100 mV with a slope of 0.5 V $s^{-1}$ applied at a frequency of 2 Hz. Data were acquired at a frequency of 5 kHz after filtering at 1.67 kHz. For TRPM8 channel activation, 200 µM (-)-menthol was applied. 0.1, 1 and 10 µM compound concenctrations were applied in the presence of (-)-menthol. In some measurements (-)-menthol and applied compound were washed out and a second application of (-)-menthol caused second TRPM8 current increases. The maximal (-)-menthol-induced outward currents at +100 mV before application of selected compound were used for analysis. TRPM8-expressing cells which showed basal activity $\geq$2 nA/pF at +100 mV were excluded from further analysis. To determine $IC_{50}$ value, 0.3, 1, 2, 3, 10, 30 and 100 µM compound concentrations were applied. For TRPC4 channel activation, 50 nM (-)-Englerin A (Carl Roth, Karlsruhe, Germany) was applied two times. (-)-Englerin A was applied in the presence of different compound concentrations. The second (-)-Englerin A-induced current increase was used for normalization. For calculation of $IC_{50}$, maximal (-)-Englerin A-induced outward currents at +100 mV were used. For calculation of the percentage of maximal outward currents at +100 mV basal currents before application of the first stimulus were always subtracted. Dissociation constants ($K_i$) were calculated with the modified Cheng Prusoff equation ($K_i = IC_{50}[1+(C/EC_{50.C})]^{-1}$, where C is the concentration of control activator (-)-Englerin A (0.05 µM) in the assay and $EC_{50.C}$ its $EC_{50}$ value (0.0112 µM). Electrophysiological data was analyzed using Origin 7.5 software (OriginLab, Northampton, MA, USA). Data are presented as mean ± standard error of the mean (s.e.m.). For calculation of $IC_{50}$ value, concentration response

curve was fitted using Single Hill-equation until no reduction of Chi-square was notable.

## Nuclear Hormone Receptor Assays

In vitro activity determinations on mentioned nuclear hormone receptors (peroxisome proliferator-activated receptors (PPARs), liver X receptors (LXRs), farnesoid X receptor (FXR), retinoid X receptors (RXRs), retinoic acid receptors (RARs), vitamin D receptor (VDR), constitutive androstane receptor (CAR)) were conducted by Dr. Daniel Merk. Daniel Merk performed the assays, analyzed and interpreted the experimental results.

### Hybrid Reporter Gene Assays

Gal4 hybrid reporter gene assays were conducted as reported previously.[266–268] pFA-CMV-based constructs comprising the ligand binding domain of the human nuclear receptor in question were used as expression plasmids for the chimera receptors. pFR-Luc (Stratagene) served as reporter plasmid and pRL-SV40 (Promega) for normalization of transfection efficiency and cell growth. The assays were conducted in 96-well format in HEK293T cells that were cultured as described previously.[266–268] Transient transfection was carried out using Lipofectamine LTX reagent (Invitrogen) according to the manufacturer's protocol. After transfection and incubation with test compounds (12–14 h), cells were assayed for luciferase activity using Dual-Glo™Luciferase Assay System (Promega) according to the manufacturer's protocol. Luminescence was measured with an Infinite M200 luminometer (Tecan Deutschland GmbH). All hybrid assays were validated with reference agonists (PPAR$\alpha$: GW7647; PPAR$\gamma$: pioglitazone; PPAR$\delta$: L165,041; LXR$\alpha/\beta$: T0901317; FXR: GW4064; RXRs: bexarotene; RARs: tretinoin; VDR: calcitriol; CAR: CITCO) which yielded $EC_{50}$ values in agreement with literature. The assays were conducted in duplicates with at least two independent repeats and for active compounds repeated without hybrid receptor coding DNA for every test compound at the highest tested concentration to exclude unspecific effects.

# 4 Results and Discussion

## 4.1 Automated *De Novo* Design of Bioactive Natural Product Mimetics: An Application to Galanthamine

In line with our previous study,[196] a similar design strategy has been applied to the natural product galanthamine. The design concept includes a ligand-based reaction-driven *de novo* design approach with galanthamine as template structure, a computational pharmacophore similarity evaluation and *in silico* target prediction to design and select potential bioactive galanthamine mimetics. A set of mimetic structure containing a natural product-derived scaffolds has been synthesized and their activity has been characterized *in vitro* for known targets of galanthamine and computationally predicted targets.

### Publication Details and Contributions

This study was submitted to MedChemComm (Royal Society of Chemistry, United Kingdom, ISSN 2040-2511) on February 6$^{th}$ 2019. It is published in this thesis according to the permissions of authors to reuse their own materials (Licenses, copyright and permission information, Royal Society of Chemistry).

***Authors***
Lukas Friedrich[a,+], Marco Stöckli[a,+], Petra Schneider[a,b], Andreas Koeberle[c], Oliver Werz[c] and Gisbert Schneider[a]

[a] Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Vladimir-Prelog-Weg 4, 8093 Zurich (Switzerland)
[b] inSili.com LLC, Segantinisteig 3, 8049 Zurich (Switzerland)
[c] Department of Pharmaceutical/Medicinal Chemistry, Institute of Pharmacy, Friedrich-Schiller-University Jena, Philosophenweg 14, 07743 Jena (Germany)
[+] L.F. and M. S. contributed equally to this work.

***Author contributions***
L. F., M.S. and G.S. designed the study. M.S. generated and ranked the *de novo* designs. L.F. and M.S. selected, synthesized and characterized the mimetics. L.F. and

G.S. analyzed the *in vitro* data. A.K. and O.W. designed the experiments on cyclooxygenase inhibition, and analyzed and interpreted the results. L.F. and G.S. wrote the manuscript. All authors approved the final version of the manuscript.

***Acknowledgements***
S. Haller and Dr. T. Rodrigues are thanked for technical support and discussion. B. Schmalwasser, P. Wiecha, and H. Traber are thanked for their technical support with respect to the cyclooxygenase assays. Dr. M. D. Wörle and the ETH Small Molecule Crystallography Center performed the small molecule X-ray analysis.

***Reference***
L. Friedrich, M. Stöckli, P. Schneider, A. Koeberle, O. Werz, G. Schneider, "Automated *De Novo* Design of Bioactive Natural Product Mimetics: An Application to Galanthamine", *Med. Chem. Comm.*, **2019**, *submitted*.

***Source of funding***
This research was financially supported by the OPO-Foundation Zurich and ETH Zurich.

## Introduction

Natural products are an important source of inspiration for medicinal chemistry.[3] Drug discovery programs rely on natural products to be directly used as drugs or as starting points for synthetic druglike compounds.[22] The field of neurodegenerative disorders has particularly benefited from natural products as pharmacological leads.[269] For example, there are several natural-product-inspired acetylcholinesterase (AChE) inhibitors from natural sources among the prescribed treatments for Alzheimer's disease.[270] Among these, galanthamine (**68**) is a 3,4-fused dihydrobenzofuran alkaloid from the snowdrop flower *Galanthus caucasicus*.[271] It is a representative member of the *Amaryllidaceae* alkaloid family.[272] As a dual-acting AChE inhibitor ($IC_{50}$ = 227±16 nM in human-derived SH-SY5Y neuroblastoma cells)[273] and allosteric modulator of nicotinic acetylcholine receptors (nAChRs),[274] it possesses neuroprotective properties, and is used as drug for the treatment of mild to moderate Alzheimer's disease.[275] Several total syntheses of galanthamine have been described, motivated in part by the high costs of its isolation from limited natural sources. In 2001, Guillou and co-workers established an eight-step total synthetic route to (±)-galanthamine.[276] Recently, a new catalytic asymmetric total synthesis of the central scaffold has been described.[277] In this study, an automated *de novo* design approach has been applied to expedite the identification of natural-product-inspired new chemical entities (NCEs). Six computer-generated galanthamine mimetics were synthesized and evaluated for *in vitro* bioactivity against computationally predicted macromolecular targets.

## Methods

All methods applied in this study have been described in the general method section of this thesis (chapter 3). The following assays were performed by Eurofins Cerep on a fee-for-service basis (Table 4.1).

**Table 4.1:** Selected targets for *in vitro* characterization of galanthamine (**68**) and its mimetics **72 - 77**.

| Assay Description | Number | Reference |
|---|---|---|
| Acetylcholinesterase Human Enzymatic Assay | 0363 | [278] |
| nAChR (alpha7) Human Ion Channel Binding (Antagonist Radioligand) Assay | 3010 | [279] |
| nAChR (N muscle-type) Human Ion Channel Binding (Antagonist Radioligand) Assay | 0936 | [280] |
| M1 Human Acetylcholine (Muscarinic) GPCR Cell Based Agonist & Antagonist Calcium Flux Assay | G029 | [281] |
| 5-HT2B Human Serotonin GPCR Cell Based Agonist & Antagonist IP1 Assay, JAK1 Human TK Kinase Enzymatic LANCE Assay [Km ATP] | G183 | [282] |
| JAK2 Human TK Kinase Enzymatic LANCE Assay [Km ATP] | 2869 | [283] |
| JAK3 Human TK Kinase Enzymatic LANCE Assay [Km ATP] | 2905 | [284] |
| Src Human TK Kinase Enzymatic LANCE Assay [Km ATP] | 2907 | [285] |
| ERbeta Human Estrogen NHR Functional Antagonist Coactivator Assay | 311440-1 | [286] |
| Vascular Endothelial Growth Factor (VEGFR1) (Radioligand Binding) Assay | 4051 | [287, 288] |
| Human Cyclooxygenase Enzymatic Assay (COX-1) | 4173 | [289] |
| Human Lipoxygenase Enzymatic Assay (5-LOX) | 0772 | [290] |
| PDE4A1A Human Phosphodiesterase Enzymatic Assay | 4074 | [291] |
| PDE7A1 Human Phosphodiesterase Enzymatic Assay | 4078 | [291] |
| PDE3A Human Phosphodiesterase Enzymatic Assay | 4072 | [291] |

## Results and Discussion

We applied the ligand-based tool DOGS (*Design Of Genuine Structures*)[187] with galanthamine (**68**) as template to generate 1182 *de novo* designs, of which 395 structures were unique (33%). These computer-generated molecules were scored with the CATS (*Chemically Advanced Template Search*)[6, 147] method to determine their topological pharmacophore similarity to galanthamine. The five top-ranked designs with a CATS distance $\leq$1.2 (Euclidean distance of the CATS descriptors between design and galanthamine) were selected for further analysis (Figure 4.1). Designs **I** and **II** are derivatives of the mycotoxin (-)-altenuene (**69**) and were therefore omitted. The designs **III**–**V** were further investigated. Following the phenyl scaffold as the most frequent scaffold of the *de novo* designs, the coumarin core was the second most frequent scaffold, accounting for 11% of the unique designs. 12 compounds (24%) of the 50 top-ranked designs were based on a coumarin scaffold (Figure 4.2). The tetrahydro-6*H*-benzo[*c*]chromen-6-one core (scaffold **70**, Figure 4.3a) occurred in 5% of all the designs, and in 26% of the top-50. Since coumarins are known for anticholinesterase activity[292] and constitute a naturally occurring "privileged" structure,[293] scaffold **70** was manually modified to the coumarin-based core **71** (Figure 4.3a). This structural modification is inconsistent with the underlying concept of a reaction-driven design tool. Within the *de novo* design software DOGS, every virtual product is assembled in an iterative virtual synthesis based on a set of implemented organic reactions, a pre-defined building block library and a graph-based topological similarity assessment between reference structure and every virtual compound.[187] Therefore, every design compound owns theoretically a certain degree of synthetic feasibility. Nevertheless, the design software does not estimate different reactivities of competing functional groups, nor consider stereoeletronic

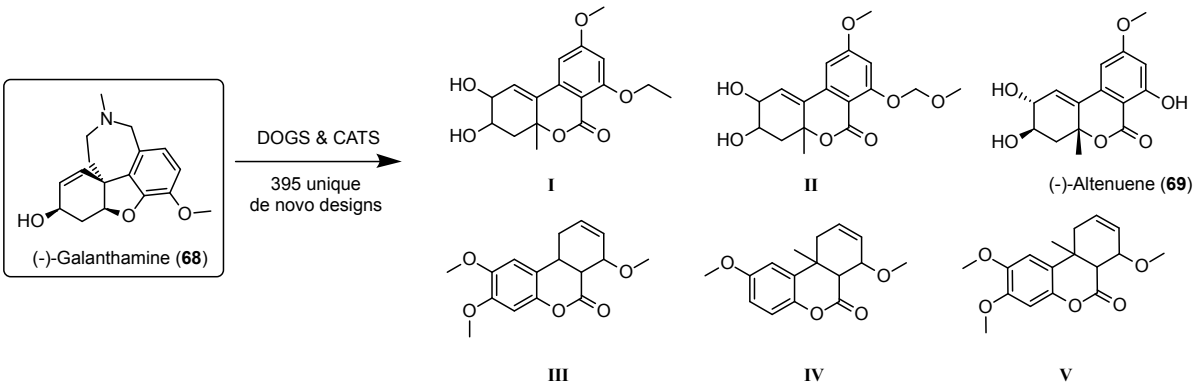



**Figure 4.1:** Chemical structures of the natural product galanthamine, a potent inhibitor of acetylcholinesterase (AChE), and its computer-generated mimetics. The DOGS software was used to generate new chemical entities, and the CATS descriptor was applied to rank the designs according to their topological pharmacophore similarity to galanthamine. The five top-ranked mimetics (CATS distance $\leq$1.2) are labeled according to their individual rank (**I**-**V**). Designs **I** and **II** are structurally similar to the mycotoxin (-)-altenuene (**69**).

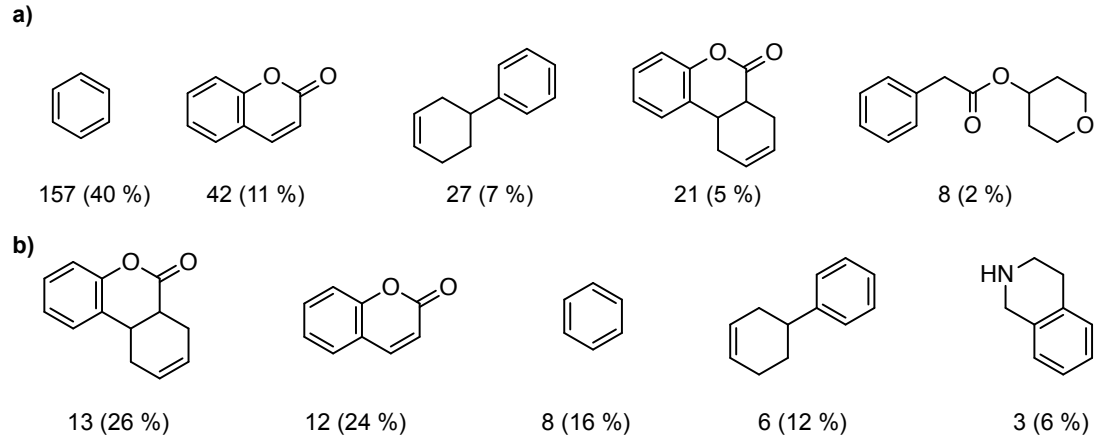



**Figure 4.2:** Most frequently occurring Murcko scaffolds among **a)** all unique *de novo* designs from DOGS, and **b)** from the fifty top-ranked designs according to their CATS distance.

effects of the reactants in certain reaction types, for example cycloaddition reactions. Thus, a manual validation of the synthesizability of each design is essential as long as the design tool does not include further synthetic accessibility predictions like reported synthesizability descriptors.[256, 294] However, this structural variation simplified the chemical synthesis. One-step microwave-assisted Pechmann reaction[295] (Figure 4.3b) led to six coumarin-based compounds (**72**-**77**, Figure 4.3c). SPiDER (Self-organizing map–based Prediction of Drug Equivalence Relationships)[211] target prediction software was employed to identify potential macromolecular targets of galanthamine and designs **72**–**77** (Table A.1). Galanthamine was not contained in the training data of the prediction model. We only considered targets with a high confidence score ($p$-value $\leq$0.05). In total, 37 protein target families were predicted for these seven compounds (**68** and **72**–**77**). 80% of the predicted protein targets were enzymes, G-protein coupled receptors (GPCR) or ion channels (Figure 4.4a). Ten different targets

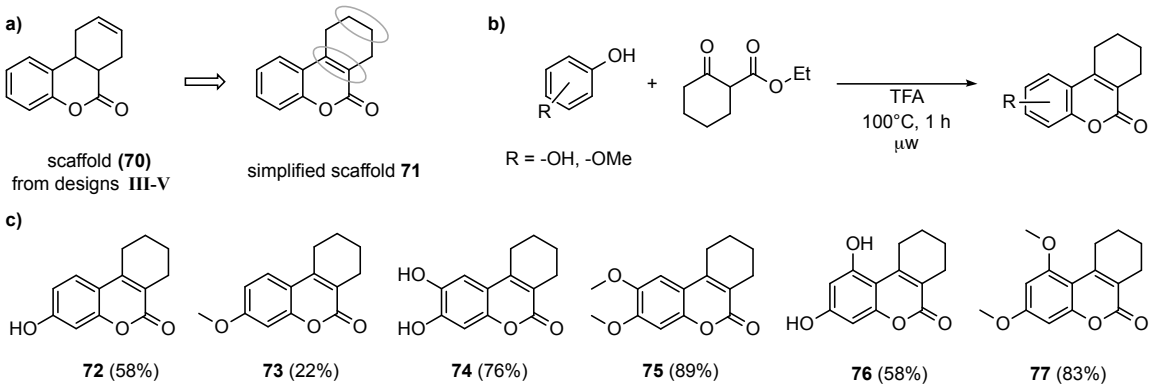
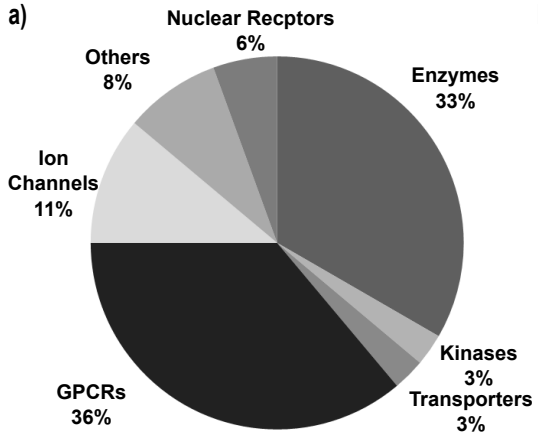



**Figure 4.3:** **a)** Scaffold **70** from designs **III-V** was structurally modified to obtain the simplified coumarin-based core structure **71**. **b)** Reaction scheme of the one-step Pechmann reaction under microwave conditions. **c)** Set of six synthesized coumarin-based mimetics (**72**-**77**). Numbers in parentheses give the yields obtained.

were predicted for at least four of the compounds (Figure 4.4b). A full list of all predicted target families shared among galanthamine and the mimetics is shown in Table A.2. Phosphodiesterases were the only target predicted for galanthamine and all *de novo* designs.




b)

| *Target* | *# compounds* |
|---|---|
| Phosphodiesterase | 7 |
| Tyrosine Kinase | 6 |
| Lipoxygenase | 5 |
| Sodium Neurotransmitter Symporter | 4 |
| Acetylcholine Receptor (Muscarinic) | 4 |
| Serotonin Receptor | 4 |
| Adrenergic Receptor | 4 |
| Acetylcholine Receptor (Nicotinic) | 4 |
| Cyclooxygenase | 4 |
| Phospholipase | 4 |

**Figure 4.4: a)** Protein classes of predicted targets for galanthamine **68** and mimetics **72**-**77**. **b)** Most frequently predicted targets and counts of associated compounds (galanthamine **68** and mimetics **72**-**77**).

Phosphodiesterase E3A inhibition in the single-digit micromolar range was corroborated in cell-free assays for galanthamine mimetics **72** ($IC_{50}$ = 2.6±1.4 µM), **74** ($IC_{50}$ = 2.1±1.2 µM), and **75** ($IC_{50}$ = 5.9±1.2 µM). Galanthamine was inactive in this assay. Notably, none of these designs were considered as AChE inhibitors with a high confidence, and even for galanthamine, AChE did not obtain the highest SPiDER target score. Mild AChE inhibition was confirmed for compounds **73** ($IC_{50}$ = 54±1 µM) and **75** ($IC_{50}$ = 16±1 µM). The prediction of the acetylcholine receptor (AChR) family as a potential biological target of compounds **73**, **75**, and **77** is in perfect agreement with the known activities of galanthamine. It was confirmed for **75** (M1 receptor, $IC_{50}$ = 15±1 µM, $K_i$ = 1.8 µM), but not for the designs **73** and **77**. In contrast to galanthamine, which activates the nAChR receptor at concentrations between 0.02-2 µM ($\alpha4/\beta2$ subtype),[296] *de novo* compound **75** was inactive in the nAChR receptor assay. This result does not come as a surprise, because galanthamine acts by an allosteric mode of action.[296] Compounds **74** and **76** inhibited Janus kinase 2 (JAK2) with $IC_{50}$ values of 59±1 µM and 13±1 µM, respectively, as predicted by the SPiDER software. Moreover, compounds **72**, **73** and **76** antagonized the predicted serotonin receptor 5-$HT_{2b}$ with $IC_{50}$ values in the low micromolar range (1.6±1.2, 2.0±1.4, 1.7±1.2 µM). Compound **76** was confirmed as a weak antagonist ($IC_{50}$ = 46±2 µM, $K_i$ = 9.5 µM) of estrogen receptor beta, as computationally predicted. The most potent activity was found for compound **76**. In contrast to galanthamine, this *de novo* designed molecule inhibited cyclooxygenase-1 (COX-1) in the submicromolar range ($IC_{50}$ = 0.37±1.08 µM, cell-free assay), which was confirmed by suppressed COX-1 activity in human platelets

(Figure A.10). Galanthamine and compound **76** did not inhibit isolated human recombinant cyclooxygenase-2 (COX-2), and COX-2-derived prostaglandin formation was neither observed for galanthamine nor for compound **76**. The identification of **7** as a potent and selective COX-1 inhibitor corroborates coumarin as privileged scaffold of COX inhibitors.[297]

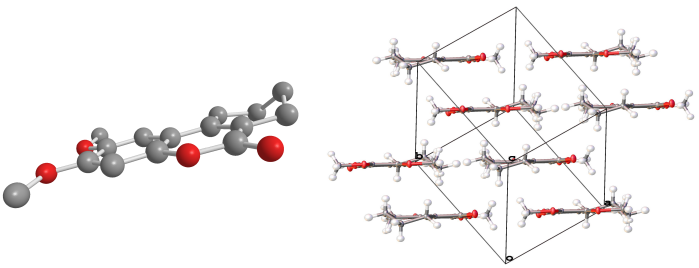


**Figure 4.5:** Crystal structure (*left*) and crystal packing (*right*) of compound **75**.

Because of the rather flat molecular structures of **72**–**77** (Figure 3), we measured their aggregation behavior in water by dynamic light scattering (Figure A.2 and Figure A.3). No aggregation was observed for compound **3** at a concentration <15 µM, for **73** <500 µM, and for **75**–**77** <62 µM. Compound **74** starts to aggregate at concentrations >31 µM. Thus, the measured JAK2 inhibition of compound **74** could be an artefact caused by unspecific inhibition by compound aggregates. Because of the ability to bind to targets from different protein families (receptors, enzymes, kinases) the mimetic compounds qualify as promiscuous ligands.[298] In fact, their common coumarin core is a privileged scaffold in medicinal chemistry.[76] We estimated the potential target promiscuity of galanthamine and the galanthamine-derived mimetics with a recently developed computational method.[298] Galanthamine and the structurally related compounds **72**–**77** displayed high promiscuity scores (>99%). Based on their properties (molecular weight: 216–260 g mol$^{-1}$, clog*P*: 1.72–2.34, number of hydrogen-bond donors: 0–2, number of hydrogen-bond acceptors: 3–4; Table A.3) and their ligand efficiency[299] (LE = 0.37–0.52), bioactive compounds **72**–**76** qualify as leadlike.[300]

## Conclusions

Even though the *de novo* design software DOGS and the pharmacophore similarity ranking by CATS have previously proven suitable to identify NCEs from a natural product template,[196] the galanthamine mimetics **72**-**77** had already been described and recorded in PubChem[301] (compound (PubChem CID): **78** (5322315), **79** (2274081), **80** (5453495), **81** (20683631), **82** (5397934), **83** (680696)). Among all mimetics, however, only compounds **72** and **76** have annotated bioactivity data in ChEMBL24[208](CHEMBL IDs: CHEMBL1486992 (**78**) and CHEMBL502843 (**82**)). The identification of coumarin-based molecules as promiscuous ligands derived from the natural product galanthamine demonstrates the potential of automated computer-assisted design of natural-product-inspired bioactive compounds for medicinal and biological chemistry. These compounds could serve as starting points for hit-to-lead optimization. The approach presented here could serve as a prototype for rapid conversion of structurally intricate natural products into synthetically accessible mimetics. It thereby extends the scope of scaffold-centered design methods, like biology-oriented synthesis (BIOS),[102] and has the additional potential to rediscover known chemical entities with novel bioactivities.

## 4.2 Fully Computer-Guided "Design-Make-Test" Cycle: *De Novo* Design of Novel Bioactive Marinopyrrole A Mimetics

A further application of computational *de novo* design has been conducted to generate natural product mimetics. In this second study, the marine natural product (±)-marinopyrrole A has served as template for a computational design strategy as introduced previously(section 4.1). Several new chemical entities have been obtained according to their virtually proposed synthetic routes and their bioactivty have been assessed against predicted molecular targets.

### Publication Details and Contributions

This study was not submitted yet. Further biological results are remaining and will be added before submission. Parts of the syntheses were published in V. Bobinger, "Natural Products in Ligand-Based *De Novo* Design", *Master Thesis*, University of Constance, **2017**.

***Authors***
Lukas Friedrich[a], Veronika Bobinger[a], Daniel Merk[a], Andreas Koeberle[b], Oliver Werz[b] and Gisbert Schneider[a]

[a] Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Vladimir-Prelog-Weg 4, 8093 Zurich (Switzerland)
[b] Department of Pharmaceutical/Medicinal Chemistry, Institute of Pharmacy, Friedrich-Schiller-University Jena, Philosophenweg 14, 07743 Jena (Germany)

***Author contributions***
L.F. and G.S. designed the study. L.F. and G.S. supervised V.B. conducting her master thesis. L.F. and V.B. generated and ranked the *de novo* designs. L.F. and V.B. synthesized and characterized the mimetics. L.F. performed the computational analysis. L.F. and G.S. analyzed *in vitro* data. D.M. designed and performed the experiments on nuclear receptor activation assays. D.M. analyzed and interpreted these results. A.K. and O.W. designed the experiments on cyclooxygenase inhibition, and analyzed and interpreted the results. L.F. and G.S. wrote the manuscript. All authors approved the final version of the manuscript.

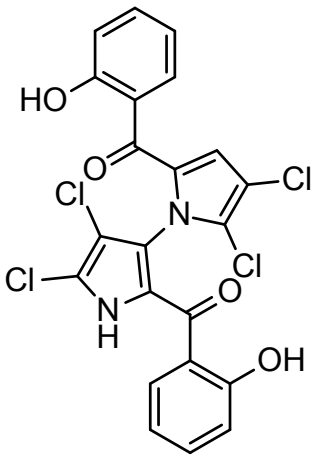
***Acknowledgments***
We thank Sarah Haller for technical support and Philipp Waser for NMR services. B.

Schmalwasser, P. Wiecha, and H. Traber are thanked for their technical support with respect to the cyclooxygenase assays.

***Source of funding***
This research was financially supported by ETH Zurich.

## Introduction

Natural products have been used as medicines for millennia and still are a key source of novel drugs and lead compounds in modern drug discovery.[1, 22, 302] Due to strong improvements in collecting and analyzing samples from marine organisms, marine natural products play an increasing as growing collection of novel bioactive compounds.[303] Oceans host millions of different species[304]producing secondary metabolites with an anticipated large structural diversity.[2, 112] Among recently discovered marine natural products, marinopyrroles were identified as potent antibacterial agents.[305] A member of this class, (±)-marinopyrrole A (**84**, Figure 4.6), was additionally reported to possess anticancer activity by selectively binding to and inducing degradation of myeloid cell leukemia 1 protein (Mcl-1), an anti-apoptotic protein of the B-cell lymphoma 2 (Bcl-2) family.[306] However, the putative selectivity of marinopyrrole A for Mcl-1 dependent cell lines is debatable[307, 308] based on recent results suggesting that marinopyrrole A inhibits protein translation leading to a decrease of Mcl-1 proteins.[309] In 2017, further molecular targets of marinopyrrole A were computationally predicted and confirmed biochemical assays.[310] We have now employed marinopyrrole A as template for computer-assisted *de novo* design. After *in silico* similarity comparison and virtual target prediction, computational designs were synthesized and characterized for their biological activity *in vitro*. This full design-make-test cycle successfully produced bioactive mimetics of the complex natural product marinopyrrole A.




(±)-Marinopyrrole A (**84**)

**Figure 4.6:** Structure of the marine natural product (±)-marinopyrrole A (**84**).

## Methods

All methods applied in this study are described in the general method section of this thesis (chapter 3). The following assays were performed by Eurofins Cerep on a fee-for-service basis (Table 4.2).

**Table 4.2:** Selected targets for *in vitro* characterization of marinopyrrole A (**84**) and its mimetics **85** - **87**.

| Assay Description | Number | Reference |
|---|---|---|
| Cyclooxygenase 1 (COX-1) Human Cyclooxygenase Enzymatic Assay | 4173 | [289] |
| EP1 Human Prostanoid GPCR Cell Based Antagonist Calcium Flux Assay | 2054 | [311] |
| EP2 Human Prostanoid GPCR Cell Based Antagonist Calcium Flux Assay | 1957 | [312] |
| EP3 Human Prostanoid GPCR Cell Based Antagonist Impedance Assay | 2578 | [313] |
| EP4 Human Prostanoid GPCR Cell Based Antagonist cAMP Assay | 1872 | [312] |
| CDK1 /CyclinB Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2875 | [314] |
| CDK2 /CyclinA Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2908 | [314] |
| CDK4 /CyclinD1 Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2876 | [315] |
| SAPK2A (p38alpha) Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2881 | [316] |
| ERK2 (MAPK1) Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2878 | [317] |
| JNK3 Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2916 | [318] |
| GSK3beta Human CMGC Kinase Enzymatic LANCE Assay [Km ATP] | 2879 | [319] |
| IKKalpha Human Other Protein Kinase Enzymatic LANCE Assay [Km ATP] | 2937 | [320] |
| IKKbeta Human Other Protein Kinase Enzymatic LANCE Assay [Km ATP] | 2938 | [320] |
| IKKepsilon Human Other Protein Kinase Enzymatic LANCE Assay [Km ATP] | 2587 | [321] |

*Continued on next page*

Table 4.2 – *Continued from previous page*

| Assay Description | Number | Reference |
|---|---|---|
| IRAK4 Human TKL Kinase Enzymatic LANCE Assay [Km ATP] | 2933 | [322] |
| PKCmu (PKD1) Human CAMK Kinase Enzymatic HTRF Assay [Km ATP] | 2204 | [323] |
| ROCK2 Human AGC Kinase Enzymatic LANCE Assay [Km ATP] | 2884 | [324] |
| CB1 Human Cannabinoid GPCR Cell Based Agonist cAMP Assay | 1744 | [325] |
| CB1 Human Cannabinoid GPCR Cell Based Antagonist cAMP Assay | 1745 | [325] |
| CB2 Human Cannabinoid GPCR Cell Based Agonist cAMP Assay | 1746 | [325] |
| CB2 Human Cannabinoid GPCR Cell Based Antagonist cAMP Assay | 1747 | [325] |
| CRF1 Human Corticotropin-Releasing Factor GPCR Cell Based Antagonist cAMP Assay | 505 | [326] |
| CCK2 (CCKB) Human Cholecystokinin GPCR Cell Based Antagonist cAMP Assay | 1879 | [327] |
| Human Glucocorticoid NHR Binding (Agonist Radioligand) Assay | 469 | [328] |
| OX1 Human Orexin GPCR Cell Based Antagonist Calcium Flux Assay | 2235 | [329] |
| OX2 Human Orexin GPCR Cell Based Antagonist Calcium Flux Assay | 2350 | [330] |

## Results and Discussion

For its attractive pharmacological activity and intricate molecular architecture, we applied the natural product (±)-marinopyrrole A (**84**) as template for ligand based *de novo* design using the tool "Design Of Genuine Structures" (DOGS).[187][14] This software is a virtual rule-based *de novo* design tool that constructs mimetics of a given template by fusing building blocks from a catalogue according to pre-coded reaction rules. The virtual products of such a reaction run are then evaluated for similarity to the given template and either saved as design or used as starting molecule for another construction run step until design with converging similarity to the template are retrieved. Similarity can be assessed in DOGS with two implemented molecular representations (molecular graph and reduced graph). We conducted three design runs with each molecular representation employing marinopyrrole A as template structure, which yielded a total set of 2300 *in silico* structures containing 802 unique molecular entities. With 405 designs (50.5%) resulted from the molecular graph representation and 397 structures (49.5%) from the reduced graph representation, both implemented scoring approaches were well balanced. All unique *de novo* designs were then further ranked according to their topological pharmacophore similarity (CATS distance)[6, 147] to the template marinopyrrole A. The designs resulting from molecular graph mode had a lower median CATS distance (2.07) than the designs from the reduced graph mode (2.20) (Figure 4.7). From the 100 top-ranked structures with the lowest CATS distance (CATS distance $\leq$ 1.8), 79 entities originated from molecular graph representation and 21 from reduced graph representation runs. Since reduced graph representations lead to a higher level of abstraction from the structural composition and constitution of the template molecule, this result seemed reasonable.[187]

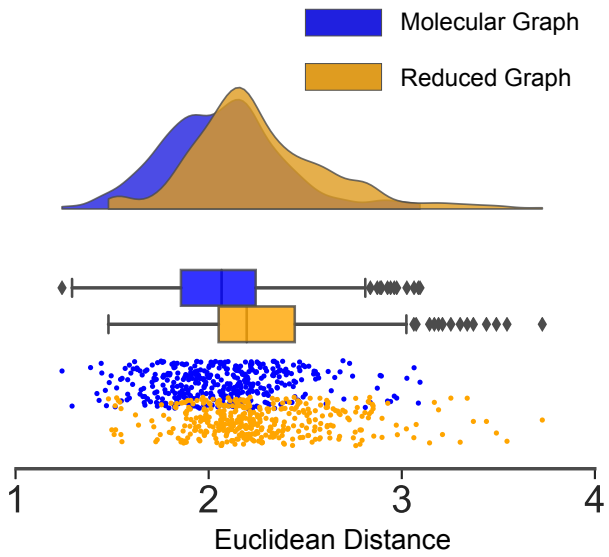



**Figure 4.7:** Distribution of CATS distances, a measure of pharmacophore similarity, from 802 unique *de novo* designed compounds. 405 designs (50.5 %) were generated by utilizing a molecular graph representation (blue). The remaining 397 *in silico* structures (49.5 %) were designed with reduced graph mode (orange). Designs from molecular graph mode had a median CATS distance of 2.07, whereas the compounds from a reduced graph representation had a median CATS distance of 2.20.

Closer inspection of the 802 unique designs revealed 334 unique Murcko scaffolds[72] (42%) and 38 unique scaffolds (38%) for the top-ranked 100 designs. 2,4,5-Triphenyl imidazole, known as lophine,[331]was the most frequent scaffold (122/802 entries, 15%, and 34/100 top-ranked designs, 34%, Figure A.11 and A.12). Thus, the 17 most similar designs of the 34 samples comprising a 2,4,5-triphenyl imidazole scaffold were further considered (Figure A.13). Compounds (**85**) and (**86**) ranking in position 1 and 17 were selected for synthesis and *in vitro* biological evaluation based on their favorable similarity to the template marinopyrrole A (**84**) and building block availability for the computationally proposed synthetic routes (Figure 4.8). The computationally proposed synthesic procedures contain an imidazole formation from a dicarbonyl compound, an aldehyde and ammonia, known as a Debus-Radziszewski reaction.[331, 332] To obtain compound **85** from the intermediate product **87** of this reaction, the software proposed an esterification of the phenol with an acyl halide. Imidazole synthesis was conducted for **87** and **86** as proposed by the software in good yields. Cyclization of **87** from the building blocks **88** and **89** was successful in batch whereas **86** was prepared from **88** and **90** under microwave irradiation in a sealed vial. To avoid side reactions in the software proposed esterification approach, this second synthetic step to **85** was achieved by a Steglich esterification using DCC/DMAP and a silyl-protected glycolic acid (Figure 4.9). Two additional derivatives (**85a** and **85b**) of compound **85** were synthesized via the same synthetic strategy. The real synthetic pathway deviated only slightly from the virtual synthetic route by an introduced protecting group. This small adaption was necessary to avoid potential side reactions of competitive functional groups. This small manual adjustments was necessary because the *de novo* design tool neglects both steric and stereoelectronic features of the virtual reactants which

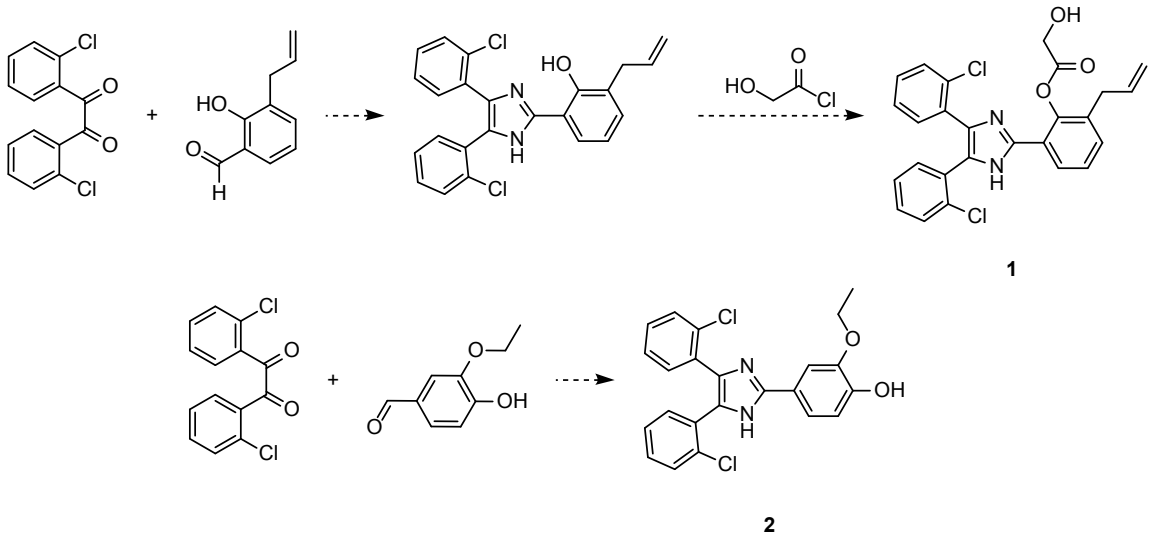



**Figure 4.8:** The computationally proposed synthetic routes to its mimetics **85** (*top*) and **86** (*bottom*) as obtained from the software DOGS. Both routes consist of an imidazole formation via Debus-Radziszewski reaction. To synthesize compound **85** from the imidazole intermediate **87**, esterification of the phenol group of the with 2-hydroxyacetyl chloride was proposed.

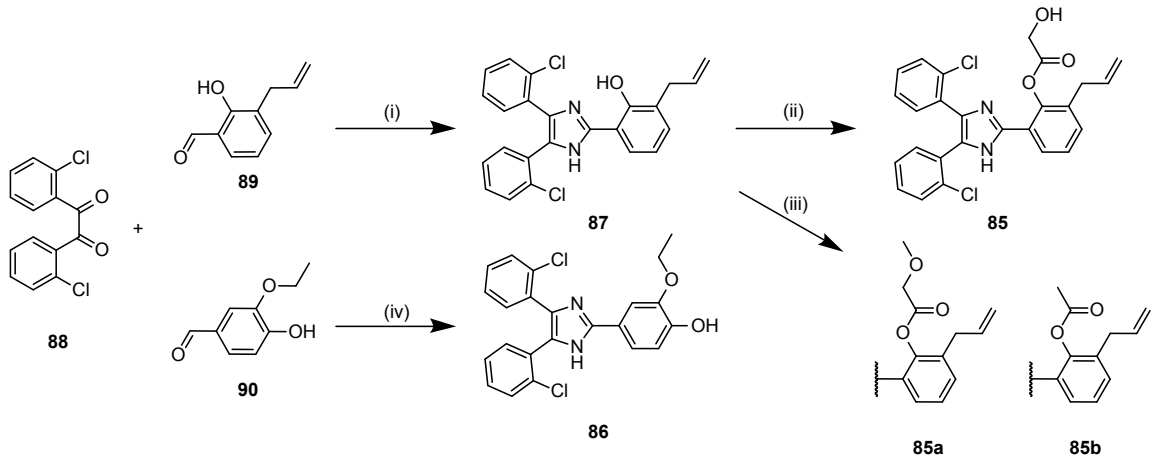



**Figure 4.9:** Synthesis overview of compounds **85**,**85a**,**85b**, and **86**. Reagents and conditions: (i) $NH_4OAc$, AcOH, reflux, 5 h, 66 % (**87**); (ii) (I) TBDPS-protected glycolic acid, DCC, DMAP, $CH_2Cl_2$, rt, 16 h, 38 %, (II) TBAF, AcOH, THF, 0 °Cto rt, 2h, 82 % (**85**); (iii) 2-methoxyacetic acid or acetic acid, DCC, DMAP, $CH_2Cl_2$, rt, 16h, 57 % (**85a**), 53 % (**85b**) (d) $NH_4OAc$, AcOH, µw irradiation, 180 °C, 5 min, 48 % (**86**).

strongly effects the outcome of a "real world" synthesis. Therefore, the software is unable to estimate the reactivities of different functional group. Due to the fact that the employed *de novo* design software did not include any kind of protection group, the final synthetic procedure has been considered as a results of autonomous virtual synthesis and its minor alterations based human expertise. To identify macromolecular targets of the template marinopyrrole A and its computationally designed mimetics **85**, **85a**, **85b**, **86** and **87**, we employed the target prediction software SPiDER (self-organizing map-based prediction of drug equivalence relationships)[211] which has been recently applied to predict targets of natural products and *de novo* designed natural product mimetics.[196, 201, 214] Only predictions with low false-positive error estimations ($p$-values<0.05)) were considered as potential molecular targets.We predicted the targets for the entire collection of marinopyrrole-inspired *de novo* designs and the 100 top-ranked compounds according to their pharmacophore similarity (CATS distance). The distribution of the corresponding protein classes of the 100 top-ranked designs resembled the protein class distribution of the target prediction of all *de novo* designs (Figure A.14). Figure 4.10 depicts the ten most frequently predicted target families for the top-ranking 100 designs and within these, 6 out of 10 also occurred in the most frequently predicted families for the entire collection of designs (Figure A.15).

Considering the target predictions of individual compounds, only six protein families were predicted active for the template marinopyrrole A (**91**) and at least one its mimetics (Table 4.3). Cyclooxygenases, prostanoid receptors and serine/threonine protein kinases were predicted for both marinopyrrole A and all mimetics. The cannabinoid receptor and the corticotropin-releasing factor receptor family were predicted active for the template marinopyrrole A and the mimetic **85** and the related derivatives **85a**,

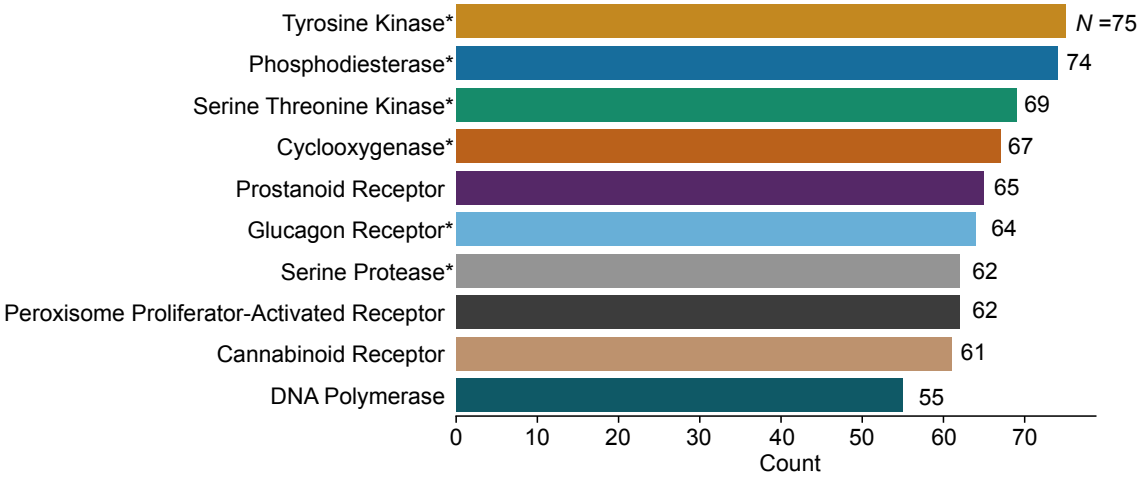



**Figure 4.10:** Ten most frequently predicted protein families of the 100 top-ranked designs according to their CATS distance. From the target prediction of these top-ranked designs, all annotated subtypes of each protein family were counted (*N*). *These protein families were among the ten most common predicted families of all *de novo* designs.

**85b**, and **87**, and the nuclear peroxisome proliferator-activated receptors were identified as potential target family of marinopyrrole A and compounds **85**, **85a**, and **85b**. To assess their bioactivity, we screened the compounds against selected subtypes of the predicted target families (Table A.4).

The inhibitory effects of marinopyrrole A and all mimetics on cyclooxygenase 1 (COX-1) and 2 (COX-2) were investigated in cell-free assays (Table 4.4 and Figure A.16). Marinopyrrole A inhibited COX-1 and COX-2 in the double-digit micro-molar range. Compound **85a** and intermediate **87** were markedly more potent COX-1 inhibitors with nanomolar (**85a**, $IC_{50}$ = 0.2±0.0 µM) and low micromolar (**87**, $IC_{50}$ = 1.8±0.5 µM) activity. Both compounds were almost inactive against COX-2 (>100 µM, 93.4±53.3 µM). Among known COX-1 and COX-2 inhibitors (7911 COX-1 and 9648 COX-2 inhibitors in ChEMBL24, January 2019), the 2,4,5-triphenyl imidazole scaffold of **85a** and **87** is only found in seven compounds, which in contrast to **85a** and **87** comprise selective COX-2 inhibitors.[333]

From the prostanoid receptor family, we assessed the antagonistic effects on the four

**Table 4.3:** Protein families predicted for marinopyrrole A (**84**) and the mimetic compounds (**85**, **85a**, **85b**, **86** and **87**).

| Protein Family | Compounds |
|---|---|
| Cyclooxygenase | **84**, **85**, **85a**, **85b**, **86**, **87** |
| Prostanoid receptor | **84**, **85**, **85a**, **85b**, **86**, **87** |
| Serine/threonine protein kinase | **84**, **85**, **85a**, **85b**, **86**, **87** |
| Cannabinoid receptor | **84**, **85**, **85a**, **85b**, **87** |
| Corticotropin-releasing factor receptor | **84**, **85**, **85a**, **85b**, **87** |
| Peroxisome proliferator-activated receptor | **84**, **85**, **85a**, **85b** |

**Table 4.4:** Effects of marinopyrrole A (**84**) and its mimetics **85**, **85a**, **85b**, **86**, and **87**, on isolated ovine cyclooxygenase 1 (COX-1) and human recombinant cyclooxygenase 2 (COX-2).

| compound | COX-1 | COX-2 |
|---|---|---|
| **84** | 16.6±2.6[1] | 45.2±21.3[1] |
| **85** | *n.d.*[2] | *n.d.*[2] |
| **85a** | 0.2±0.0[1] | n.i.[1,4] (up to 100 µM) |
| **85b** | >10[1] (71.0±2.6**,[3]) | n.i.[1,4] (up to 10 µM) |
| **86** | 29.7±18.1[1] | 53.1±15.1[1] |
| **87** | 1.8±0.5[1] | 93.4±53.3[1] |
| indomethacin | 18.4±2.3***,[3] | 31.9±4.6***,[3] |

[1] $IC_{50}$ values are given as mean ±S.E.M. of single determinations obtained in three,four (COX-2: cmpd 2) or five (indomethacin) independent experiments.
[2] n.d., not determined.
[3] Residual activities (% of control) at a compound concentration of 10 µM.
[4] n.i., no inhibition. (**) $P$ <0.01, (***) $P$ <0.001; student $t$-test.

prostaglandin $E_2$ receptors (EP1-EP4) as these were the most frequently predicted subtypes. Marinopyrrole A and the mimetics **85a**, **85b**, **86**, and **87** showed no significant inhibition of any of the four subtypes (<50%) at 50 µM (APPENDIX). Compound **85** was identified as an inhibitor of the EP1 receptor in the low micromolar range ($IC_{50}$ = 4.5 ± 1.3 µM, $K_i$ = 0.7 µM, Figure A.17). 2,4,5-triphenyl imidazole core of compound **85** represents a new scaffold for EP1 selective modulators compared to scaffolds from known modulators in ChEMBL24 (1599 modulators with annotated activity). To assess the kinase activity profile, we screened marinopyrrole A against a selection of subtypes of the serine/threonine kinase family (14 kinases) that were predicted by SPiDER (Table A.4). In particular, the kinases cyclin-dependent kinase 1 (CDK1),[334] mitogen-activated protein kinase-1 (MAPK1 or ERK2)[335] and glycogen synthase kinase 3 (GSK3),[335, 336] which are involved in the regulation of Mcl-1 in tumor cells.[337] As marinopyrrole A had no activities against the screened kinase panel (Table A.4), thus, the mimetics were not tested. Therefore, we did not further screen the mimetics against these kinases. We also studied marinopyrrole A for activation and inhibition of cannabinoid receptors 1 and 2 (CB1 and CB2). Marinopyrrole A (50 µM) weakly activated CB1 and CB2 at while mimetic **85** (50 µM) showed antagonistic effects on both cannabinoid receptors (Table A.4). Due to these contradictory effects, we did not further investigate the target class. On corticotropin-releasing factor receptor 1, compound **86** and **87** had antagonistic effects on corticotropin-releasing factor receptor 1 ($IC_{50}$ = 5.7±1.2 µM, 36 ±2.3 µM, Figure A.17), whereas all other compounds were inactive (Table A.4). Marinopyrrole A and its mimetics **85**, **85a** and **87**, were characterized on peroxisome proliferator-activated receptors, and a panel of closely related nuclear receptors (Figure A.18). Marinopyrrole A caused agonistic effects on the retinoic

**Table 4.5:** Bioactivity assessment of marinopyrrole A and mimetics **85**, **86**, and **87** on the glucocorticoid receptor (GR, binding competition), cholecystokinin receptor 2 (CCK2, antagonism), and the orexin receptors 1 and 2 (OX1 and OX2, antagonism).$IC_{50}$ curves are shown in Figure A.17.

| compound | GR | CCK2 | OX1 | OX2 |
|---|---|---|---|---|
| **84** | 0.7 (1.3) [1] | 1 (8.4)[1] | 0.3 (1.3)[1] | 0.6 (2.3)[1] |
| **85** | 0.6 (1.2±1.2) [2] | inactive [3] | inactive [3] | inactive [3] |
| **86** | 2.2 (4.3±1.2) [2] | 1.1 (5.1±2) [2] | 8.4 (40±1) [2] | inactive [3] |
| **87** | 0.7 (1.4 ±1.1) [2] | inactive [3] | inactive [3] | inactive [3] |

[1] Reported $K_B$ and $IC_{50}$ values in parentheses of marinopyrrole A in µM.[310]
[2] $IC_{50}$ values in parentheses are derived from eight different concentrations with two replicates for each concentration and were reported in µM.
[3] No activity in the initial screening at a concentration of 50 µM.

acid receptor α ($EC_{50}$ = 0.63±0.11 µM), the vitamin D receptor ($EC_{50}$ = 1.09±0.06 µM) and the liver X receptor β ($EC_{50}$ = 0.34±0.40 µM) (Figure A.19).
Moreover, we tested all mimetic structures against the cholecystokinin receptor 2 (CCK2), glucocorticoid receptor (GR) and orexin receptors 1 and 2 (OX1 and OX2), which were recently computationally predicted and biochemically confirmed molecular targets of marinopyrrole A (Table 4.5, and A.17).[310] Marinopyrrole A was identified as an active ligand of these four receptors with $K_i$ and $K_B$ values $\leq$1 µM.[310] Competitive binding to the glucocorticoid receptor was observed for **85** ($K_i$ = 0.6 µM), **86** ($K_i$ = 2.2 µM) and **87** ($K_i$ = 0.7 µM). Compound **86** also had antagonistic effects on CCK2 and OX1 with $K_i$ values in the low micromolar range ($K_i$ = 1.1 µM for CCK2, and 8.4 µM for OX1 respectively).
The bioactivity data gained in this project uncovered further informative instances in which either marinopyrrole was the only identified modulator of a certain target family (e.g. LXRβ, VDR, RARα) or, on the other hand, solely the mimetic structures exhibited activities on specific targets, for example on CRF1 or EP1. We also observed that marinopyrrole A activated the cannabinoid receptors CB1 and CB2, whereas one of its mimetics (**85**) showed antagonistic effects on both receptors at the initial screening concentration. Moreover, marinopyrrole A had no significant activity on a set of selected kinases at the screening concentration, but which were predicted as potential molecular targets.

## Conclusions

In this computational approach to discover natural product bioactivities and bioactive synthetic mimetics we have identified further molecular targets of marinopyrrole A and succeeded in designing synthetically feasible compounds that mimic the parent natural product's pharmacodynamic effects.The mimetic structures possess similar multi-target profiles confirming the assumption that key pharmacophoric features have been inherited from marinopyrrole A in computational *de novo* design. The identification of bioactie NCEs mimicking the marine natural product marinopyrrole A reveals the high potential of rule-based computational ligand design from a bioactive template. Our strategy involved the virtual construction of potential mimetics, computational prediction of likely synthetic procedures and *in silico* pharmacophore similarity ranking. The proposed synthesis routes turned out suitable to generate the desired compounds in good yields and *in vitro* biological evaluation proved a large proportion of predicted pharmacological activities correct. Applied on a structurally intricate natural product with desirable bioactivity, this approach results in synthesizable mimetic structures that inherit the target spectrum of the natural product template. Thereby, reaction-driven computational *de novo* design of natural product mimetics can accelerate innovation in early drug discovery.

## 4.3 Shape Similarity by Fractal Dimensionality: An Application in *De Novo* Design of Natural Product Mimetics

The former described studies (section 4.1 and 4.2) relied on two-dimensional methods to assess similarity between the natural product templates and their relative mimetics. In this study, a similarity metric has been incorporated into the design approach which estimates the three-dimensional shape similarity between molecules. In prospective application, this metric has been utilized to select *de novo* designs mimicking the natural product (-)-englerin A, which has already served as reference structure in a prior study.[196] Synthesis and biological evaluation has been performed to assess the activity of selected shape similar compounds.

### Publication Details and Contributions

This study was submitted to Communications Chemistry (Nature Publishing Group, London (Springer Nature), United Kingdom, ISSN 2399-3669) on February 17th 2019. It is published in this thesis according to the permissions of authors to reuse their own materials (reprints and permissions, American Chemical Society Publications).

***Authors***
Lukas Friedrich[a,+], Ryan Byrne[a,+], Michael Mederos y Schnitzler[b,c], Aaron Treder[b] Inderjeet Singh[b], Christoph Bauer[a], Thomas Gudermann[b,c,d], Urusula Storch[b,e] and Gisbert Schneider[a]

[a] Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Vladimir-Prelog-Weg 4, 8093 Zurich (Switzerland)
[b] Walther Straub Institute of Pharmacology and Toxicology, Ludwig Maximilians University of Munich, 80336 Munich (Germany)
[c] DZHK (German Centre for Cardiovascular Research), Munich Heart Alliance, Munich (Germany)
[d] Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich (Germany)
[e] Institute for Cardiovascular Prevention (IPEK), Ludwig Maximilians University of Munich, 80336 Munich (Germany)
[+] L.F. and R. B. contributed equally to this work.

***Author contributions***
G.S, L.F. and R. B. designed the study. R.B. implemented the fractal dimensionality

descriptor and analyzed its performance. L.F. ranked, selected, synthesized, characterized and analyzed the designs. C.B. performed the quantum chemical assignment. G.S. and L.F. analyzed the *in vitro* data. L.F., R.B and G.S. wrote the manuscript. U.S. and M.M.y.S. designed the electrophysiological experiments, analyzed the electrophysiological data and participated in writing the manuscript. I.S. and A.T. performed electrophysiological measurements under the supervision of U.S. and M.M.y.S. T.G. provided expertise and feedback and revised the manuscript. All authors approved the final version of the manuscript.

***Acknowledgements***
We thank the Small Molecule Crystallography Center of ETH Zurich for X-ray services, Dr. Bernhard Pfeiffer for NMR services, and Sarah Haller for technical support. We also thank Thomas Voets for providing human TRPM8 channel in pCAGGSM2-IRES-GFP expression vector.

***Reference***
L. Friedrich, R. Byrne, M. Mederos-Schnitzler, A. Treder, I. Singh, C. Bauer, T. Gudermann, U. Storch, G. Schneider, "Shape Similarity by Fractal Dimensionality: An Application in *De Novo* Natural Product Mimetics", *ACS Cent. Sci.*, **2019**, *submitted*.

***Source of funding***
This research was financially supported by ETH Zurich and received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Grant Agreement No. 675555, Accelerated Early staGe drug dIScovery (AEGIS). This work was supported by German Research Foundation (Deutsche Forschungsgemeinschaft projects no. 406028471 and TRR-152).

## Introduction

Virtual compound screening is one of the major techniques employed in the identification of novel bioactive molecules in drug discovery. The common underlying principle of the various computational approaches, ranging from compound database searching to automated *de novo* design, is the definition, quantification, and utilization of molecular similarity. Any 'similarity' approach must encapsulate some features correlated with aspects of interest to the chemist, and, ideally, extend that to provide useful perspectives on chemical space. Various methods have been proposed and utilized successfully to rank compound libraries predicated on physicochemical properties, topological indices, and higher-dimensional methods incorporating information on the distribution of such properties in Euclidean (or other) spaces. Each of these are to a greater-or-lesser extent correlated with the implicit similarity measure most often of interest for drug design; that molecules grouped by some measure should have recognizably related biological activities.[156, 176, 338, 339]
The shape of a molecule has been observed to correlate well with that of its binding pocket.[153, 340–342] One commonly-employed means of capturing this shape information is through the alignment-based methods. Popular approaches in this group are ROCS[343] and SHAEP[344], which utilize the maximum possible overlap in the volume of pairs of molecules to rank query molecules against a template. Alignment-independent methods, such as USR[345] and USRCAT[346], have been gaining in popularity, primarily owing to their speed. These provide a straightforward description of the distance distribution between heavy atoms in a molecule and a set of fixed reference points, but fail to consider molecular surface curvature and shape. We here aim to introduce and provide a proof-of-concept for a novel means to rapidly describe and compare shape representations of molecular objects, enabling quick filtering of large compound libraries. To this end, we expand on the concepts elucidated in our earlier work[347] in applying the concept of fractal dimensionality in fast, shape-based similarity-based virtual compound screening, with minimal sampling of ligand conformational space. This approach allows for the description of a molecule through shape analysis of its Connolly surface, and for comparison of molecules through a simple Euclidean distance measure.

## Methods

Additional methods, which are not part of the general laboratory and computational method section (chapter 3), are described below.

### Fractal Dimensionality Calculation

Intuitively, fractal dimensionality is a measure of the density of a molecular surface in a volume, and of how close the parts of that surface are to one another, describing the measured properties of an object as a function of the scale at which they are measured.[348] This approach benefits from a rigorous mathematical foundation and sensitivity to scale-dependency in shape description. Many natural objects and phenomena demonstrate self-repeating behavior, but are not true fractals, rendering an analytical solution inapplicable. A close approximation of this value can, however, be obtained by means of the application of various algorithms, such as the correlation exponent estimation approach adopted here.[349, 350]
For our 'global' fractal dimensionality approach (Figure 4.11), we first generate a single low-energy conformer per compound using the MMFF94 force-field provided in the RDKit (v. 2018.03.04) [261, 351] package for Python 3.6.5. Following this, we describe the solvent-accessible molecular Connolly surface[352] as a set of vertices using a ray-tracing approach, as implemented in NanoShaper (v.0.7.5.).[353] We determine the variation in an unbiased estimator of fractal dimensionality, the correlation sum $\hat{C}(\delta)$, with $\delta$, a distance in the set of distances in a Euclidean space, $E$, as follows (Equation 4.1, Figure 4.11):

$$\hat{C}(\delta) = \frac{2}{N(N-1)} \sum_{i<j} \theta(\delta - |x_i - x_j|) \tag{4.1}$$

where $N$ is the number of vertices in the surface representation, and $\theta$ represents the Heaviside step function, where a given pair of vertices ($x_i$ and $x_j$) is assigned a value of one if they are within distance $\delta$ of one another, and zero otherwise. The linear range of the $\hat{C}(\delta)/\delta$ relationship across the range of $E$ is extracted through a set of derivative-based rules, and we take its gradient, $v$, as an approximation of the correlation exponent using a Theil-Sen[354, 355] robust regression estimator. This value, $v$, strictly obeys the inequality $v \leq D$, where D is the canonical fractal dimension. In general, the two values are very similar ($R^2 = 0.99$),[349] therefore it provides an adequate approximation of D. Various methods for rendering this practicable in a reasonable time-frame are employed, primarily in reducing the number of comparisons to be made through heuristic optimizations, and through efficient matrix operations. This results in a numerical representation of shape, which can be used to rank the similarity of molecules based on the Euclidean distance of their fractal dimensionality to that of a given template. The properties of this similarity metric, and the efficiency of the correlation

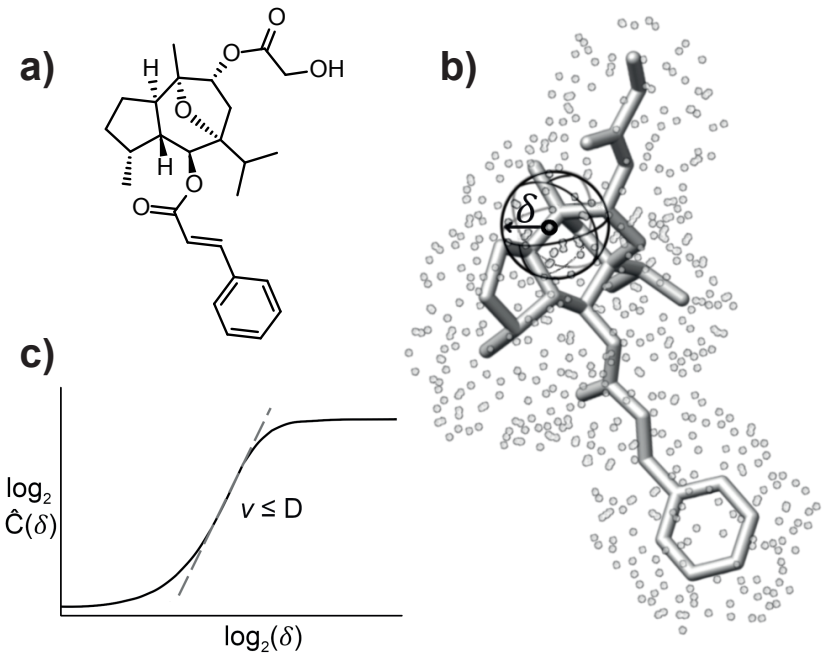



**Figure 4.11:** **a)** Molecular structure of the natural product (-)-Englerin A (**62**), **b)** a discrete representation of its Connolly surface (grey dots) for a generated conformer, and **c)** an illustration of the behavior of the point-inclusion sphere, and the calculation of the estimator, v, of fractal dimension, D. For each point in the surface representation, the relationship between the proportion of points ($\zeta(\delta)$) within a distance $\delta$ inclusion sphere and that distance is stored. These relationships are then combined, and described in terms of gradient, providing an unbiased estimation of the molecule's fractal dimensionality. V, or GFD, can be calculated for any small- or macro-molecule, and allows for shape-based screening based on a simple distance-from-template measure.

exponent estimation, allow for this method to be applied to the shape comparison of small molecules in a short time-frame.

### *In Vitro* Characterization

The following assays were performed by Eurofins Cerep on a fee-for-service basis.

**Table 4.6:** Selected targets for *in vitro* characterization of englerin A (**62**) and its mimetics **92** and **93**.

| Assay Description | Number | Reference |
|---|---|---|
| TRPM8 Human Transient Potential Ion Channel Cell Based Agonist and Antagonist Calcium Flux Assay | G185 | [356] |
| TRPA1 Human Transient Potential Ion Channel Cell Based Antagonist FLIPR Assay | CYL8066FL2 | |
| TRPV3 Human Transient Potential Ion Channel Cell Based Antagonist FLIPR Assay | CYL8065FL2 | |
| TRPV4 Human Transient Potential Ion Channel Cell Based Antagonist FLIPR Assay | CYL8064FL2 | |

## Results and Discussion

To determine the utility of this shape-only method in *de novo* drug design, we applied the fractal dimensionality description to identify computationally generated, small molecule mimetics with similar biological activities to the structurally intricate ('complex') natural product (-)-Englerin A (**62**) (Figure 4.11). The sesquiterpene (-)-Englerin A was described as an inhibitor of renal cancer cell proliferation.[357] This natural product acts as nanomolar activator of transient receptor potential canonical 4 and 5 (TRPC4/5) calcium-permeable cation channels which leads to selective growth inhibitions of cancer cell lines.[198, 199] Utilizing (-)-Englerin A as a template, we previously generated NCEs by ligand-based, chemical reaction-driven *de novo* design.[196] By topological pharmacophore-based scoring and manual refinement of the computational designs, we identified natural product mimetics inhibiting the TRP melastatin 8 (TRPM8) calcium permeable cation channel, also inhibited by (-)-Englerin A.[196, 199] We here extend this preliminary original study by introducing the FD shape similarity metric. Given that the previously employed design software tool (DOGS)[187] and the pharmacophore similarity metric (CATS)[147] each rely on two-dimensional molecular representations, we investigated the use of fractal dimensionality as an similarity ranking approach, to take the spatial disposition of molecules into account. This approach represented a ranking approach orthogonal to the CATS ranking. By omitting the proposed synthetic routes of the original designs, the library of 903 *in silico* structures employed in our previous study resulted in a set of 323 unique *de novo* designed small molecules. We ranked these computer-generated designs according to their Euclidean distance from (-)-Englerin A in terms of their global fractal dimensionality (GFD). To assess the potential of GFD as a shape-based descriptor for this target case, we conducted a comparative, retrospective, analysis of the chemical space retrieved by this method, against gold-standard fingerprint (ECFP4), moment (USR) and alignment-based shape (SHAEP) approaches. Given that we lack a ground-truth in this case, *i.e.* experimental activity data for each molecule in our compound library, our retrospective analysis adopts two approaches.

We begin with an analysis of three data sets; (i) the initial *de novo* design set, (ii) the thirty top-ranked compounds in terms of global fractal dimensionality distance (GFD distance), and (iii) the TOP30 compounds according to their topological pharmacophore similarity (CATS distance) to (-)-Englerin A.[6, 147] Set (iii) is included to compare the GFD ranking approach with the CATS approach described previously.[17] We extracted the molecular scaffolds ('Murcko scaffolds')[72] of these compounds and analyzed their molecular scaffold diversity by pairwise Jaccard-Tanimoto similarity coefficient (Tc) estimation based on Morgan structural fingerprints (*radius* = 2, 1024 bits), equivalent to ECFP4[143]. This approach allows us to compare the areas of chemical space retrieved by each method, a proxy for the likely on-target efficacy, necessitated

by the lack of ground-truth activity data for each library compound (Figure A.20). The 323 initial *de novo* designs consisted of 152 unique scaffolds (47%) with high diversity (Tc = 0.18). The 30 top-ranked molecules according to GFD distance contained 24 unique (80%) and diverse (Tc = 0.17) scaffolds, whereas the 30 top-ranked compounds by CATS distance comprised 19 unique scaffolds (63%) with slightly lower diversity (Tc = 0.24). Only two scaffolds were present in both top-ranking sets (Figure 4.12). Secondly, we employ an experimentally-validated target-prediction software developed in-house (SPiDER)[211, 214] to provide an estimate of the likelihood of a given compound being active against the target family 'Transient Receptor Potential Ion Channel'. The top 30 compounds retrieved by screening with the GFD, USR, SHAEP, and ECFP4 methods were analyzed to determine their predicted activity (number of compounds with an annotated $p < 0.05$) for the target family and the proportion and diversity of the unique molecular scaffolds for the predicted active compounds. GFD retrieved 10 compounds predicted as active, each with a unique scaffold (predicted actives = 10, proportion of unique scaffolds = 1.0, diversity of unique scaffolds (pairwise

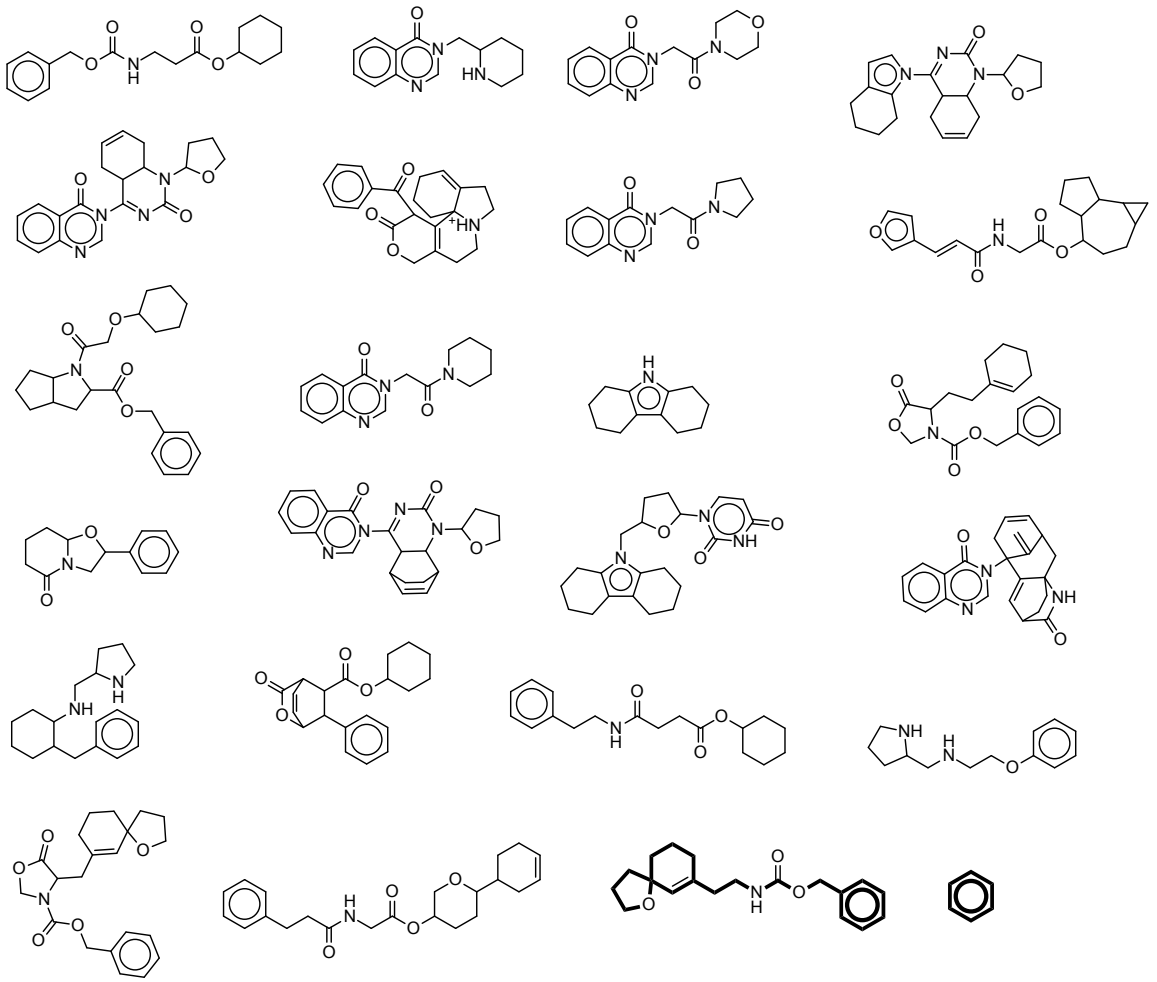


**Figure 4.12:** Unique Murcko scaffolds ($N$ = 24) of the thirty top-ranked designs according to fractal dimensionality. The two common scaffolds of the FD and CATS top-ranking designs are highlighted.

Tc) = 0.22). The SHAEP approach retrieved fewer predicted-active compounds, also all having unique scaffolds (6, 1.0, 0.21). USR retrieved the same number of predicted actives as the SHAEP approach, with fewer unique, but highly diverse, retrieved scaffolds (6, 0.66, 0.12). ECFP4 retrieved the same number of predicted actives as GFD, but with fewer, less diverse, unique scaffolds (10, 0.8, 0.33). Given that topological approaches were used in the processes of library generation and target prediction, it is corroborative that the GFD approach, which treats sub-structural information implicitly, achieved a similar predicted-active retrieval performance under evaluation with topological methods. We also performed activity prediction and diversity analysis for the library in its entirety (predicted actives = 0.25, scaffold diversity = 0.47). In summary, SHAEP and USR have slightly lower performance in terms of proportion of predicted actives in their top-ranked lists (0.2 for each), with variation in number and diversity of retrieved scaffolds. ECFP4 and GFD retrieve an identical number of predicted actives, with GFD having a higher, and highest-overall, number of unique molecular scaffolds in the predicted active compounds retrieved.
For our prospective application, we selected the thirty top-ranked compounds according to their GFD distance, and utilized computational target prediction to further refine our selection. Of these, nine had $p$-values lower than 0.05 for the target class 'TRP Ion

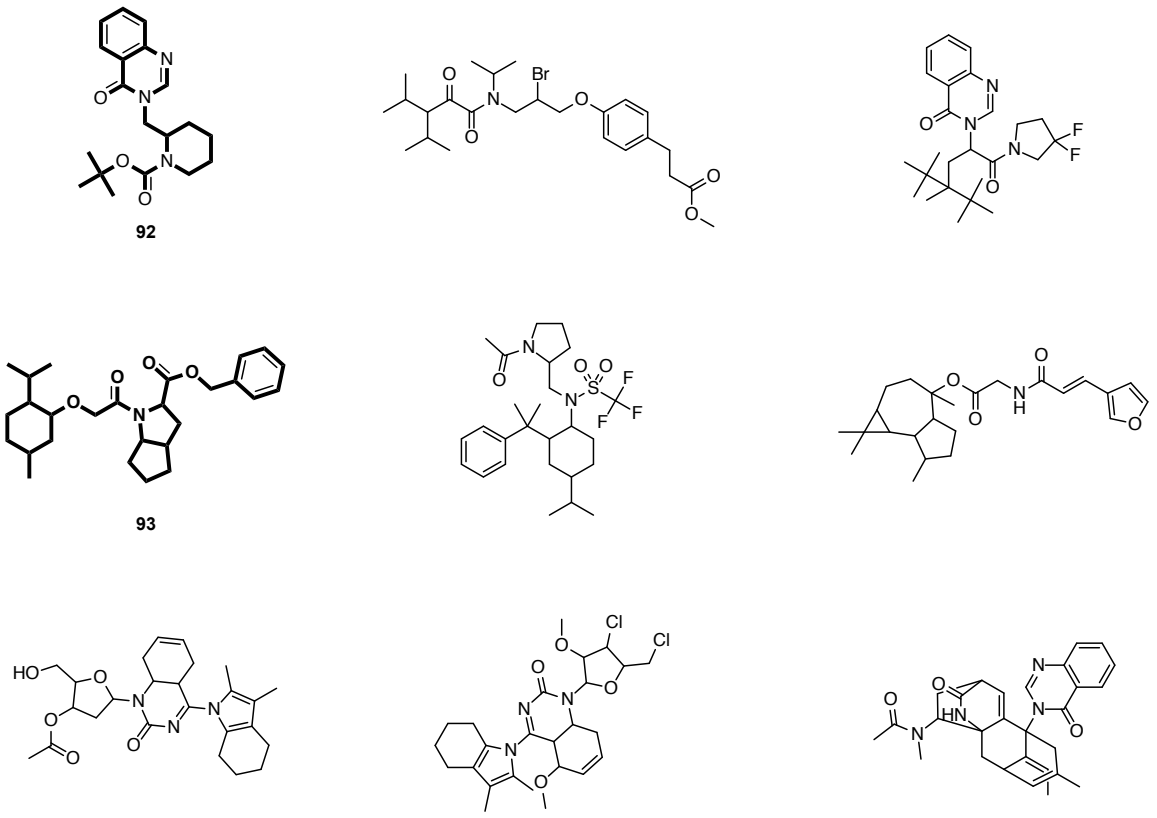



**Figure 4.13:** Nine top-ranked GFD compounds and their synthesizability score. Compound **92** and **93** were picked according to their low synthesizability score and the availability of building blocks.

Channel' (Figure 4.13). We selected compounds **92** and **93** for synthesis and bioactivity evaluation, considering their synthesizability[256] and building block availability (Figure 4.13).

Compound **92**, a piperidine extended derivative of the privileged quinazolinone scaffold,[358] is a rare structural entity. No similar entries were found in the current version of the ChEMBL database (ChEMBL24) by a substructure search. In a recent patent application, design **92** is present as a fragment in a series of novel phosphatidylinositol 3-kinase inhibitors.[359] Of note, compound **93** contains a menthol moiety, a common substructure for TRP ion channel modulators, especially for TRPM8,[360] and TRPA1[361]. To the best of our knowledge, no meaningful pharmacological interaction between menthol-containing compounds and the TRPC4 ion channel is known to date. The unmodified *in silico* structures **92** and **93** were synthesized in three steps each (Figure 4.14). Condensation of (*rac*)-(2-piperidinyl)methanamine (**94**) and isatoic anhydride (**95**) gave intermediate **96**. Cyclization of **96** with formic acid followed by Boc-protection gave final product **92**. Compound **93** was synthesized from L-menthol (**97**) and (*2S,3aS,6aS*)-octahydrocyclopenta[b]pyrrole-2-carboxylic acid (**98**). Intermediate **99** was obtained by a Williamson ether formation of **97** and 2-chloroacetic acid. Esterification of **98** with benzyl alcohol gave intermediate **100**. Amide coupling of **99** and **100** using EDC/HOBt afforded compound **93**.

To assess their bioactivity profiles, we tested compound **92** and **93** in several TRP assays, in which (-)-Englerin A showed activitis on TRPC4, TRPM8, TRPA1, TRPV3, and TRPV4.[199] Since (-)-Englerin A is a potent TRPC4 channel activator,[198] we

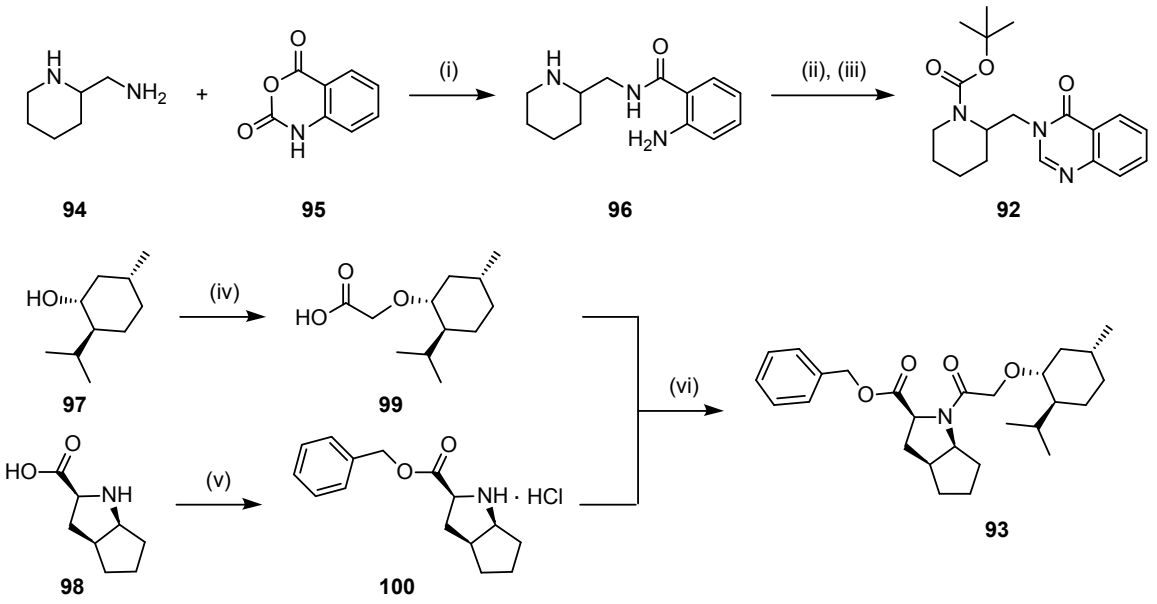



**Figure 4.14:** Synthesis of compound **92** and **93**. Reagents and conditions: (i) EtOH, reflux, 6 h, 37%; (ii) formic acid, molecular sieves (4 Å), reflux, 6 h, 83%; (iii) $Boc_2O$, $NEt_3$, $CH_2Cl_2$, 49%; (iv) Cl-CH2-COOH, NaH, KI, THF, 0°C to reflux, 16 h, 18%; (v) HCl (in dioxane, 4M), $C_6H_5-CH_2-OH$, $SOCl_2$, 0°C to rt, 70%; (vi) EDC, HOBt, $NEt_3$, THF, 0°C to rt, 86%.

analyzed the modulatory effects of compound **92** and **93** on TRPC4 channels. Compound **92** had only a weak inhibitory effect of $\leq$20% on TRPC4 currents at a concentration of 100 µM performing electrophysiological whole-cell measurements with TRPC4 over-expressing HEK293 cells (Figure A.22). In contrast, compound **93** displayed inhibitory effects on TRPC4 channels in the same electrophysiological assay (Figure 4.15a,b). (-)-Englerin A was used to elicit maximal TRPC4 currents. Application of stepwise increasing compound **93** concentrations in the presence of (-)-Englerin A decreased the (-)-Englerin A-induced TRPC4 currents (Figure A.23). As a control, (-)-Englerin A was applied for a second time inducing maximal TRPC4 currents which were used for normalization. The summary of the maximal outward currents induced by (-)-englerin A in the presence of compound **93** reveals an $IC_{50}$ for compound **93** of 5.1±0.8 µM ($K_i$ = 0.9 µM) (Figure 4.15b). Thus, we could identify compound **93** as a novel TRPC4 channel blocker, and the first validation of an interaction of that channel with a menthol-containing compounds.

The modulatory effects of both compounds were tested in a cell-based intracellular calcium assay for TRPM8. Compound **92** did not modulate ion channel TRPM8 at a concentration of 10 µM, whereas compound **93** showed an inhibitory effect in low micromolar concentration on TRPM8 ($IC_{50}$ = 1.8±1.1 µM, $K_i$ = 0.3 µM, Figure A.24). Electrophysiological whole-cell measurements with TRPM8 over-expressing HEK293 cells confirmed that compound **93** inhibits TRPM8 currents in a concentration dependent manner (Figure 4.15c,d). Maximal TRPM8 currents were induced by application of (-)-menthol. Application of 100 nM compound **93** in the presence of (-)-menthol reduced (-)-menthol-induced maximal TRPM8 currents by 1.5±0.6%, 1 µM compound **93** suppressed TRPM8 currents by 45±7%, and 10 µM compound **93** fully blocked (-)-menthol-induced maximal TRPM8 currents by 102±7% (Figure A.25). The summary of the maximal (-)-menthol-induced outward currents in the presence of indicated compound **93** concentrations suggests that the half maximal inhibitory concentration of compound **93** is about 1 µM which is in line with the results obtained with the intracellular calcium measurements. Compound **92** had no inhibitory effects on TRP ion channels TRPA1, TRPV3, and TRPV4 up to a concentration of 100 µM (Figure A.26). Compound **93** weakly inhibited TRPV4 ($IC_{50}$ = 39±1 µM) and increased intracellular free calcium concentration in TRPA1 expressing cells during the compound preincubation. Therefore, the measured inhibition of TRPA1 was likely caused by target desensitization (Figure A.26).

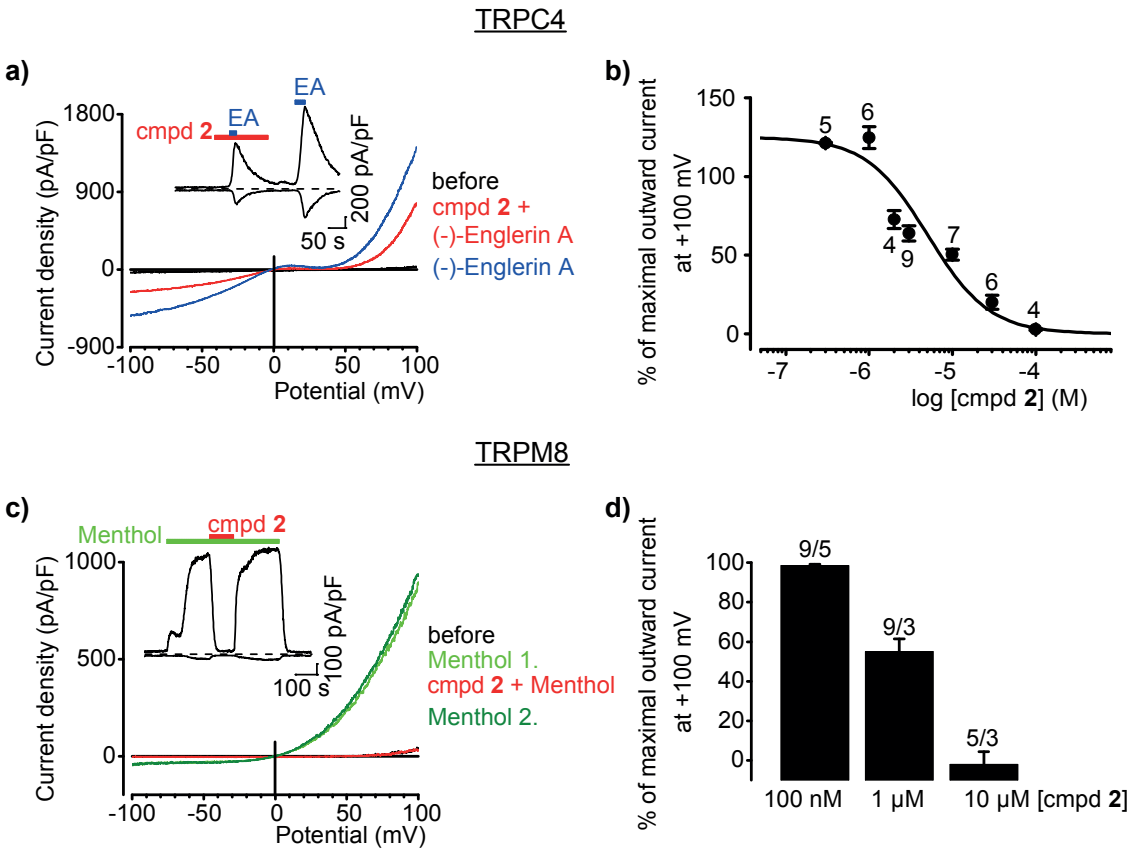



**Figure 4.15:** Electrophysiological whole-cell measurements of TRPC4 (**a**-**b**) or TRPM8 (**c**-**d**) overexpressing HEK293 cells. **a)** Representative current density-voltage curves before application of the first stimulus (black traces), during the first application of 50 nM (-)-Englerin A ('EA') in the presence of indicated concentration (10 µM) of compound **93** ('cmpd **93** + EA', red traces) and during the second application of 50 nM (-)-Englerin A (blue traces). **b)** Summary of maximal (-)-Englerin A-induced outward currents in the presence of indicated concentrations of **93**. Numbers indicate the number of measured cells from at least 3 independent experiments. To determine IC50 value the percentage of the maximal outward current at +100 mV elicited by the second application of 50 nM (-)-Englerin A was used. Insets show current density time courses at holding potentials of ±100 mV with indicated applications of compound **93** (red), (-)-menthol (green, a) and (-)-Englerin A (blue, b). **c)** Representative current density-voltage curves before application of the first stimulus (black traces), during first application of 200 µM (-)-menthol ('Menthol 1.', light green traces) and during application of indicated concentration (10 µM) of compound **93** in the presence of (-)-menthol ('cmpd **93** + (-)-menthol', red traces). The dark green line indicates current density-voltage curve after washout of **93** in the presence of (-)-menthol ('Menthol 2.', dark green traces). **d)** Summary of maximal (-)-menthol induced outward currents in the presence of indicated concentrations of compound **93**. Numbers indicate the numbers of measured cells and the numbers of independent experiments. The percentage of the maximal outward current at +100 mV elicited by the first application of 200 µM (-)-menthol was used for normalization.

## Conclusions

Comparing the computational method employed here to existing shape-based screening approaches, our approach offers a molecular shape representation rooted in a well-established field of mathematics, which captures information about local surface curvature as well as volumetric information, and that can be rapidly calculated for large compound libraries. That said, an assessment of the merits and limitations of this approach on a broader basis is necessary before we can draw firm conclusions as to its general applicability. The experimental results of this proof-of-principle study reveal the potential of fractal dimensionality as a shape-based descriptor in ligand-based virtual screening and *de novo* molecular design. Given a limited sampling of ligand conformational space, we were able to identify a compound with potent inhibitory properties and a comparable activity profile to our template (-)-Englerin A, based on the similarity in shape and local curvature of a library of small molecules generated through a *de novo* approach to our template (-)-Englerin A. The combination of a novel shape-based approach, together with a *de novo* design tool and a target prediction software, proved useful in this instance. This extent to which this applies to other natural products, and to larger sets of small molecules, is a matter for further investigation.

## 4.4 From Synthetic and Natural Compounds to Bioactive New Chemical Entities by Generative Deep Learning

Inspired by two theoretical studies,[239, 244] generative deep learning models have been investigated as novel molecular *de novo* design tools to generate bioactive small molecules. In two prospective applications, a model has been trained in a two step process to create a target-focused libraries based solely on molecular structures of already known ligands. In a proof-of-principle study, we evaluated the ability of such a model to generate bioactive compounds from a set of drug-like synthetic ligands. In a second approach, we evaluated the potential of generative deep learning models to design bioactive natural product mimetics from a set of bioactive natural products targeting the same protein family.

### Publication Details and Contributions

The first study was published in Molecular Informatics (WILEY-VCH Verlag GmbH & Co. KFA, Weinheim, Germany ISSN 1868-1751) with DOI 10.1002/minf.201700153.


***Authors***
Daniel Merk[a], Lukas Friedrich[a], Francesca Grisoni[a,b] and Gisbert Schneider[a]

[a] Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Vladimir-Prelog-Weg 4, 8093 Zurich (Switzerland)
[b] Department of Earth and Environmental Sciences University of Milano-Bicocca P.za della Scienza, 1, 20126 Milan (Italy)

***Author contributions***
G.S. designed the study. D.M. ranked, selected, synthesized, characterized and tested the designs. L.F. fine-tuned the model, generated and analyzed the designs. F.G. ranked and analyzed the designs. D.M. and G.S. analyzed the *in vitro* data. D.M. and G.S. supervised the project. D.M. and G.S. wrote the manuscript; all authors approved the final version of the manuscript.

***Acknowledgements***
We thank P. Schneider for compiling the subsets of the ChEMBL database and A. T. Müller for technical support.

***Reference***

D. Merk, L. Friedrich, F. Grisoni, G. Schneider, *Mol. Inf.* **2018**, *37*, 1700153.

***Source of funding***

This research was financially supported by the Swiss National Science Foundation (grant no. IZSEZ0_177477). D. M. was supported by an ETH Zurich Postdoctoral Fellowship (grant no. 16-2 FEL-07).

The second study was published in Communications Chemistry (Nature Publishing Group, London (Springer Nature), United Kingdom, ISSN 2399-3669) with DOI 10.1038/s42004-018-0068-1. 

***Authors***

Daniel Merk[a], Francesca Grisoni[a,b], Lukas Friedrich[a] and Gisbert Schneider[a]

[a] Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Vladimir-Prelog-Weg 4, 8093 Zurich (Switzerland)
[b] Department of Earth and Environmental Sciences University of Milano-Bicocca P.za della Scienza, 1, 20126 Milan (Italy)

***Author contributions***

D.M. and L.F. fine-tuned and analyzed the model and sampled the computational designs; D.M. and F.G. computationally ranked the sampled designs and analyzed the chemical space of the sets; D.M. selected, synthesized and characterized the designs and tested the designs *in vitro* for RXR modulatory activity; D.M. and G.S. supervised the project and wrote the manuscript. All authors approved the final version of the manuscript.

***Reference***

D. Merk, F. Grisoni, L. Friedrich, G. Schneider, *Commun. Chem.* **2018**, *1*, 68.

***Source of funding***

This research was financially supported by the Swiss National Science Foundation (Grant IZSEZ0_177477). D.M. was supported by an ETH Zurich Postdoctoral Fellowship (Grant 16–2 FEL-07)

## Introduction

Recent examples from the field of artificial intelligence (AI), for example deep learning, and the availability of large chemical and biological datasets enable the development of innovative concepts in drug discovery and development.[220] Recently, deep learning methods have been applied to overcome various challenges in the drug design process.[218, 236, 362, 363] In 2018, a *de novo* design method has been proposed which relies on generative AI.[239, 244] It consists of two essential steps: the deep neural networks learns the constitution of known bioactive compounds and then autonomously generates novel molecules with intrinsically inherited characteristics, e.g. bioactivity and synthesizability. Such a deep learning model comprises of a recurrent neural network (RNN) with long short-term memory (LSTM) cells,[242] which have proven to be effective generative models in various fields, for example in natural language processing.[243] The construction of new molecules can be perceived as forming new sentences by adding the most likely next word to a given sequence in consideration of all previous existing words. The associated probability of each new word correlates to the obtained probability distribution of all possible words from training the model with a large set of sentences. In analogy to a natural language, molecules can be represented as SMILES ("simplified molecular input line entry specification") strings, in which every character indicates either an atom type, bond type, branching point, ring system, or stereochemical information.[240] These structural representations can be employed in RNN with LSTM cells to learn the probability of each single SMILES character in a given of set molecule sequences. Based on the learned probability distributions, these novel concepts in *de novo* design are supposed to correctly construct chemical structures without the need of explicitly included building block libraries and predefined construction rules. Until now, generative AI has been applied to *de novo* design only retrospectively. Here, we describe a first prospective application of generative AI to design bioactive new chemical entities(NCEs).

Furthermore, we exploit natural products as source of inspiration in generating *de novo* designs by AI. Natural products have always been a major sources of new lead compounds for drug discovery.[1, 22] Natural products outperform synthetic small molecules in terms of their unique structural features and scaffolds.[77] Still, the synthesis of structurally complex natural products remains challenging which may hinder following structure-activity relationship studies. To overcome these limitations, computer-assisted *de novo* design constitutes a strategy to reduce synthetic efforts and create natural product-inspired molecules. It has been applied recently to generate synthesizable and bioactive natural product mimetics with rule-based computational *de novo* design methods.[196, 201] However, these rule-based approaches suffer from several limitations. These methods can only construct molecules which are accessible

from a fixed set of building blocks and pre-defined reaction pathways.[201, 364] Therefore, we address these challenges by applying the described generative AI model[244] to computationally design natural product mimetics.

## Methods

Additional methods, which are not part of the general laboratory and computational method section (chapter 3), are described below.

### Similarity searching with holistic molecular descriptors

The similarity between the unique and valid molecules generated by the machine-learning model and the sets of RXR and PPAR actives was calculated using Weighted Holistic Atom Localization and Entity Shape (WHALES) descriptors.[365] In a holistic way, the WHALES descriptor incorporates relevant information about the molecular shape, geometric interatomic distances, and atomic properties of chemical structures. Molecular geometry was optimized using the MMFF94[366] force field with 1000 iterations and 10 starting conformers for each compound; the minimum energy conformation was chosen for descriptor calculation. WHALES descriptors were computed with in-house software, using Gasteiger-Marsili[367] partial charges as weighting scheme. This scheme has shown to be the best compromise between scaffold hopping and enrichment for WHALES.[368] RXR and PPAR query structures were retrieved from ChEMBL: (1) RXR binding: the 12 most potent ligands from ChEMBL; (2) RXR agonism: the top-4 agonists according to their EC50 value; (3) PPAR agonism: the top-4 agonists annotated in ChEMBL according to EC50 for each of the PPAR subtypes. For each dataset, every compound was used as query to perform similarity ranking on the basis of their Euclidean distance on Gaussian-normalized WHALES values. The results of the individual virtual screenings on each compound were merged according to the sum of their reciprocal ranks.[369] The similarity search by WHALES and its implementation was done by Dr. Francesca Grisoni.

### Chemical Space Visualization

Principal component analysis (PCA)[370] and auto-scaling was conducted with python package scikit-learn (v0.19.1). Molecular datasets (ChEMBL library used for training, ChEMBL annotated RXR agonists, ChEMBL annotated PPAR agonists, fine-tuning compounds, sampled molecules and selected designs) were described by eight physicochemical features (molecular weight, total polar surface area , clogP, fraction of $sp^3$-hybridized carbon atoms, number of H-bond donors, number of H-bond acceptors, number of heavy atoms, number of heteroatoms) and calculated with RDKiT (version 2017.09) in Python (version 3.6). Descriptor values were auto-scaled (Gaussian normalization) and used for PCA calculation.

**Chemical Syntheses**

Syntheses of all selected designs(see Figure 4.17 and Figure 4.24) were conducted by Daniel Merk. All detailed synthetic protocols and analytical data can be found in the corresponding publications.

## Results and Discussion

### *De Novo* Design of Bioactive Small Molecules by Artificial Intelligence

We utilized a deep recurrent neural network (RNN) with long short-term memory cells (LSTM) as generative model.[244] In a first step, this LSTM based model was trained on a dataset of bioactive molecules from ChEMBL22 [208] (541'555 entities, p(Affinity) $\geq$ 6) to learn the syntax of molecules string representations (SMILES strings) and consequently the constitution of drug-like molecules. After this learning step, the model generates chemical molecules by sampling stepwise SMILES characters according to the learned probability distribution over all characters. From a set of 1000 sampled structures, a large amount of designs was valid (94%) and a high percentage of these valid molecules were unique (98%). Additionally, most of the unique structures were novel (93%) which means the generated entities were found in the training set. In a second step, the generic model was fine-tuned by transfer learning to enable the generation of *de novo* designs in a target-focused approach. Therefore, we employed a set of 25 fatty acid mimetics[371] with known modulatory activities on two nuclear receptors, namely retinoid X receptors (RXRs)[372] and/or peroxisome proliferator-activated receptors (PPARs).[373] We sampled 1000 designs from this fine-tuned model by using fragment growing from a carboxylic acid fragment ("-COOH"). The samples contained a high proportion of valid (93%) entries, whereof a large proportion represented unique structures (90%). None of the sampled molecules were part of training or fine-tuning set. To rank and select samples for synthesis and bioactivity testing, we applied computational target prediction (SPiDER software)[211] and determined the similarity to known bioactive ligands by employing molecular shape and partial charge descriptors (WHALES).[365] We merged the individual rankings and obtained a final set of 49 high-scoring designs. Five compounds (**101**-**105**) were picked for further investigations based on their individual *in silico* ranks and building block availability. None of these selected samples were present in the ChEMBL,[208] PubChem,[301] SureChEMBL,[374], Reaxys [375] (accessed July 2018) and SciFinder[376] (accessed July 2018) databases indicating their novelty. To visualize the occupied chemical space of the generated designs, we performed a principal component analysis of 8 physicochemical properties (Figure 4.16).

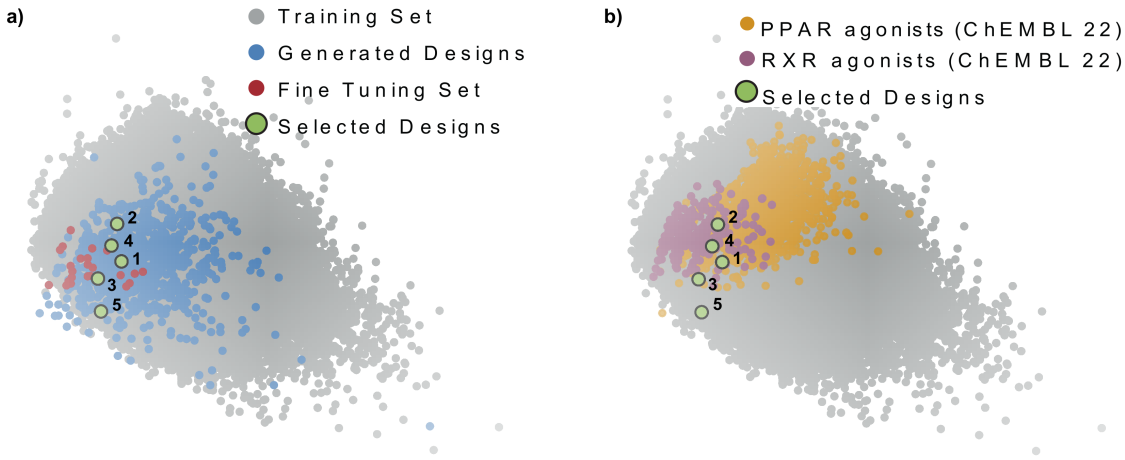



**Figure 4.16:** Chemical space visualization by principal component analysis (PCA) of 8 physicochemical properties. **a)**ChEMBL molecules (grey) represent the chemical space of the training set. The generated designs (blue) occupied an area around the 25 molecules used for fine-tuning (red). The five selected designs (**101-105**) (green) are in close proximity to the fine-tuning set. **b)** The same chemical space of the training molecules (grey) is shown. Known RXR (purple) and PPAR agonists (yellow) are plotted to visualize the target-focused areas in chemical space of these two nuclear receptor families. The selected designs **101-105** can be found in this target-focuses area.

Utilizing a PCA, this set of potentially correlated physicochemical features was transformed into two linear uncorrelated principal components which were the axes of the two-dimensional chemical space representation. The training molecules from ChEMBL (grey) are the basis of chemical space visualization. The generated samples (blue) are in proximity to the fine-tuning compounds (red) but cover a much broader area in chemical space (Figure 4.16a). The five selected designs (green) are part of the chemical space of RXR (purple) and PPAR agonists (yellow) extracted from ChEMBL (Figure 4.16b). The selected compounds **101-105** were obtained in two to four synthetic steps (Figure 4.17). **101-105** were subsequently characterized for their agonistic effects on the nuclear receptor RXR$\alpha/\beta/\gamma$ and PPAR$\alpha/\gamma/\delta$.[266] Four of the compounds showed agonistic effects on RXR and PPAR subtypes (Table 4.7). Moreover, we identified at least one agonist for every receptor subtype. Compound **101** and **102** were dual agonists of RXRs and PPAR$\gamma$, while design **103** and **104** turned out to activate two PPAR subtypes but had no activity on RXRs. Design **105** showed not activity on the nuclear receptors studied. The potency of **101-104** ranged from double-digit nanomolar $EC_{50}$ values of **101** on RXRs, to double-digit micromolar activity for compound **104** on PPAR$\delta$. Compound **104** turned out as the least potent design with weak agonistic activities on PPAR$\gamma$ and PPAR$\delta$.

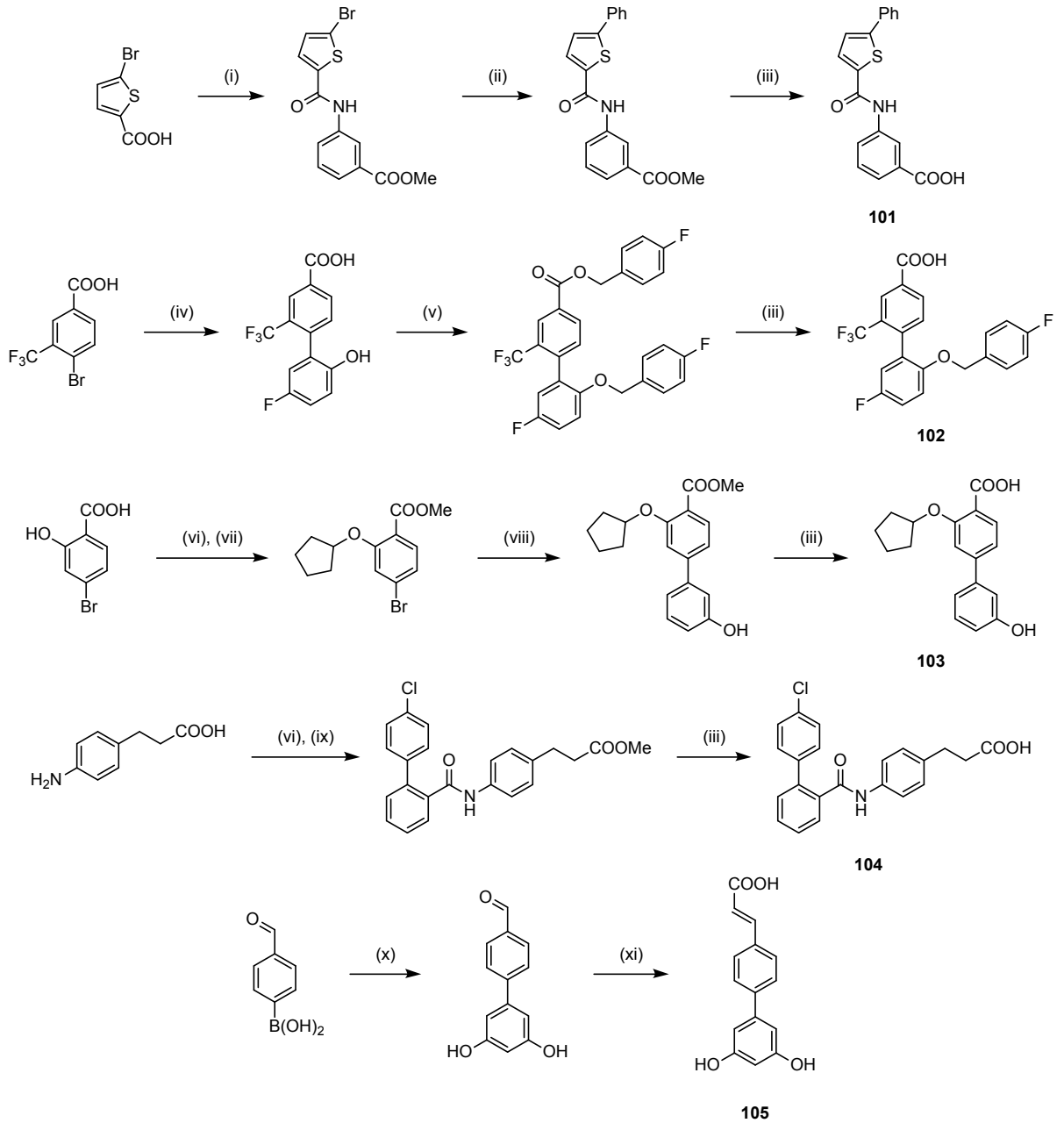



**Figure 4.17:** Synthesis of designs **101-105**. Reagents and conditions: (i) $H_2N-C_6H_4-COOH$, EDC, 4-DMAP, THF, reflux, 4 h; (ii) $C_6H_5-B(OH)_2$, $Pd(PPh_3)_4$, $Cs_2CO_3$, dioxane, 100 °C, 16 h; (iii) KOH, MeOH/THF/$H_2O$, µw, 70 °C, 30 min; (iv) $HO-C_6H_3F-B(OH)_2$, $Pd(PPh_3)_4$, $Cs_2CO_3$, toluene/EtOH, 100 °C, 20 h; (v) $F-C_6H_4-CH_2-Br$, $K_2CO_3$, DMF, µw, 100 °C, 120 min; (vi) MeOH, conc. $H_2SO_4$, reflux, 4 h; (vii) $C_5H_9Br$, $K_2CO_3$, DMF, µw, 100 °C, 6 h; (viii) $HO-C_6H_4-B(OH)_2$, $Pd(PPh_3)_4$, $Cs_2CO_3$, toluene/EtOH, 100 °C, 16 h; (ix) $C_6H_4Cl-C_6H_4-COOH$, EDC, 4-DMAP, $CHCl_3$, reflux, 12 h; (x) $C_6H_3Br(OH)_2$, $Pd(PPh_3)_4$, $Cs_2CO_3$, dioxane/DMF, reflux, 4 h; (xi) malonic acid, pyridine/piperidine, µw, 100 °C, 30 min.

**Table 4.7:** *In vitro* activity of designs **101-105** on RXRs and PPARs. Reported $EC_{50}$ values represents mean ± SEM [µM] of at least two independent experiments in duplicates.

| | $EC_{50}$ values ±SEM [µM] | | | | | |
|---|---|---|---|---|---|---|
| Compound | RXRα | RXRβ | RXRγ | PPARα | PPARγ | PPARδ |
| **101** | 0.13±0.01 | 1.1±0.3 | 0.06±0.02 | >50 | 2.3±0.2 | >50 |
| **102** | 13.0±0.1 | 9±2 | 8.0±0.7 | >50 | 2.8±0.3 | >50 |
| **103** | >50 | >50 | >50 | 4.0±1.0 | 10.1±0.3 | >50 |
| **104** | >50 | >50 | >50 | >50 | 9±3 | 14±2 |
| **105** | >50 | >50 | >50 | >50 | >50 | >50 |
| Templates[1] | | 0.024 - 43 | | | 5.2 - 54 | |
| Ref. agonist[2] | 0.033±0.002 | 0.024±0.004 | 0.025±0.002 | 0.06±0.002 | 0.6±0.1 | 0.5±0.1 |

[1] Template activities were reported as given in literature.[201, 377–379]
[2] reference agonists, literature data: bexarotene[380] for RXRs, GW7647[381] for PPARα, pioglitazone[**Willson2000**] for PPARγ, L-165041[**Willson2000**] for PPARδ.

Overall, five out of 49 top-ranked designs were selected considering their indiviual ranking, synthesizability and building block availability. An extended search in several database revealed that none of these five molecules were not reported before. After synthesizing these five designs in two to four steps, four of these five showed activity on at least one subtype of the intended targets. Besides the individual activities of compounds (**101 - 104**), the activity ranges of these four *de novo* designs on both subtypes are in the potency ranges of fine-tuning structures. These experimental results illustrated that generative deep learning models provide an innovative approach in computational *de novo* design. After this first successful proof-of-principle study, we employed the model to design chemical entities mimicking bioactive natural products.

**Tuning artificial intelligence on the *de novo* design of natural-product-inspired retinoid X receptor modulators**

We further expanded the prospective applications of this AI based generative model by tuning the model to generate novel bioactive natural product mimetics with modulatory effects on RXRs. We employed again the same generic model trained on bioactive molecules retrieved from ChEMBL.[208] In this study, six natural products with known activities on RXRs served as fine-tuning set for transfer learning (Figure 4.18).

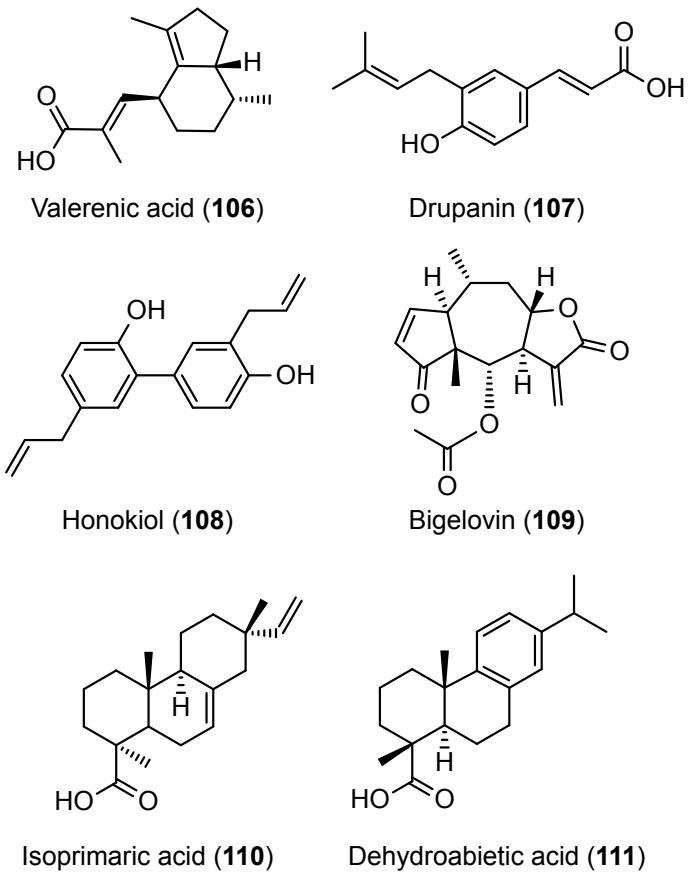



**Figure 4.18:** Selected RXR modulators from natural sources. All six natural products (**106**-**111**) have agonistic effects on RXRs in the micromolar range.[201, 382–384]

We investigated three cases with different numbers of natural products for fine-tuning. As a first approach, the generic model was fine-tuned on the natural product valerenic acid (**106**) and 1000 designs were sampled. Only 25% of sampled SMILES strings were valid chemical structures. These valid designs contained over 55% unique entities (14% with respect to all samples) whereof almost all unique compounds were novel (99%). Further analysis of the unique designs revealed that a high percentage (22%) of the designs were close molecular analogs of valerenic acid with only slight structural differences to the template (e.g. addition of methyl groups or variations in ring size) (Figure 4.19a).

Moreover, the model generated high proportions of chemically instable structures (36%, e.g. carbonic acid monoesters, anhydrides, imines, acetals, antiaromatic structures), very small molecules (21%, molecular weight below 150 g/mol) as well as linear fatty acids and nonfunctionalized hydrocarbons (14%). Out of the remaining compounds (7%), none was predicted as RXR modulator by SPiDER target prediction software.[211] Thus, no design was chosen for synthesis and bioactivity testing.

To increase number and quality of the generated designs, we extended the fine-tuning set by the natural products drupanin (**107**) and honokiol (**108**). Both natural compounds represent less intricate structures compared to other RXR activating natural

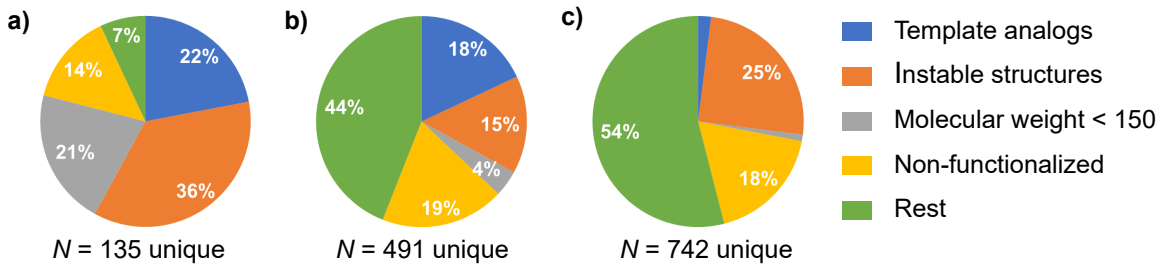
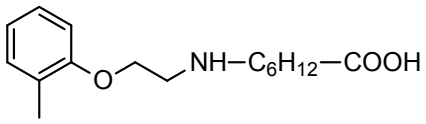


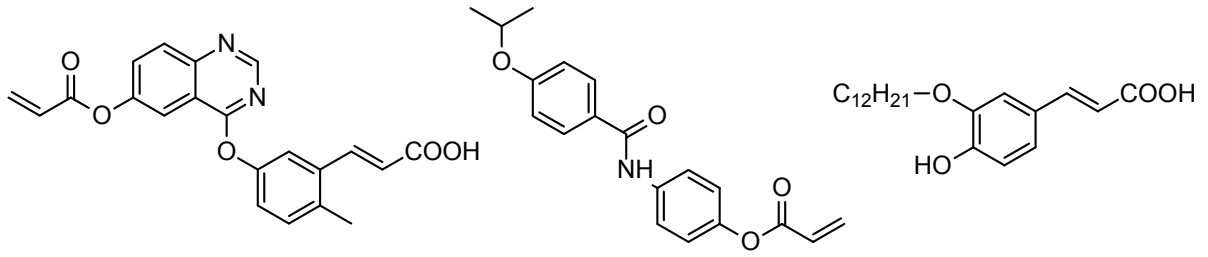

**Figure 4.19:** Characteristics of sampled designs from three distinctly fine-tuned AI models. **a)** After fine-tuning with the single natural product valerenic acid (**106**), a large proportion of the generated designs comprised of template analogs, chemically instable structures, non-functionalized fatty acids and very small molecular entities. **b)** With expanding the fine-tuning set to three natural products (valerenic acid (**106**), drupanin (**107**), honokiol (**108**)), more than half of the designs showed still unsuitable characteristics. **c)** Utilizing all six natural products (**106-111**) for fine-tuning led to reduced numbers of close template analogs and very small molecules. Unfunctionalized designs and instable structures still represented a considerable proportion of all generated samples.

products. The assumption behind the selection was that a set of structurally less complex template will increase the number and the quality of the generated designs. 1000 designs were sampled consisting of 79% valid SMILES strings from which 62% were unique (49% with respect to all samples). Over 99.5% of these unique designs represented novel compounds with respect to the training and fine-tuning set. Compared to the first experiment with a single fine-tuning template, the number of close analogs (18%), instable structures (15%), and fragment-like compounds (4%) considerably decreased with three templates for transfer learning (Figure 4.19b). The amount of sampled pure hydrocarbons remained high (19%). Out of the 215 residual designs (44%), 26 compounds were predicted as potential RXR modulators by SPiDER ($p<0.1$). The predicted actives were further ranked by determining the similarity to known RXR ligands with the already used WHALES descriptor,[365] which have been successfully applied to find novel RXR ligands.[**Merk2018b**, 201] The top-ranked design **112** (Figure 4.20) was selected for synthesis and bioactivity assessment. In a third experiment, we added three additional natural products, namely bigelovin (**109**), isoprimaric acid (**110**) and dehydroabietic acid (**111**), to the fine-tuning set. We sampled again 1000 SMILES strings from this fine-tuned model. A rise in valid sampled designs (79%) was not observed with regard to the fine-tuning with three templates. However, the amount of valid




**112**

**Figure 4.20:** The top-ranked design **112** originated from fine-tuning the LSTM model with three natural product templates.




**113** **114** **115**

**Figure 4.21:** From fine-tuning the generative deep learning model with six natural products, the three high-ranked natural product mimetics **113**, **114** and **115** were selected for synthesis and bioactivity assessment.

unique structures did increase markedly (94%, 74% with respect to all samples) in combination with a large proportion of novel compounds (99.8%). Within the unique samples, the fractions of close analogs (2%) and very small molecules (<1%) dropped noticeably (Figure 4.19c). Despite the considerable number of chemically instable structures (25%), and non-functionalized compounds (18%), over 491 of the generated designs (54%) were considered suitable for further investigations (Figure 4.19c). The 201 structures of these 491 preferred compounds, which were predicted as active for RXR modulation by p<0.1 with SPiDER, were ranked using the WHALES descriptors. The compounds **113**-**115** (Figure 4.21) were selected for synthesis and *in vitro* characterization from the 50 top-ranked designs according to their individual synthesizability and building block availability. We evaluated the natural-product-likeness[257] of the generated designs in comparison to the training molecules obtained from ChEMBL and natural products retrieved from the Dictionary of Natural Products (DNP)[385] (Figure 4.22). This score describes a similarity measure between the fragments derived from the query molecule and natural product-derived fragments from a reference database. The *de novo* designs possessed a higher natural-product-likeness score than ChEMBL molecules but were less natural-product-like than the natural product entries from the dictionary of natural products. Furthermore, a scaffold analysis was performed to further assess the structural novelty of generated *de novo* designs. The molecular frameworks[72](graph scaffold) of the generated compounds were compared to the frameworks of known RXR modulators from ChEMBL ($EC_{50}/IC_{50}$ < 50 µM, $N$ = 521) and the natural product templates utilized for fine-tuning (Figure 4.23). Independent from the fine-tuning set size, the most frequently occurring molecular frameworks differed from frameworks of the natural templates and of known RXR binders. This observation indicated that structural features of both synthetic and natural bioactive molecules were maintained by the training and fine-tuning step of the generative model.

The computational designs **112**-**115** were synthesized as described in Figure 4.24. All designs were obtained in one to four step(s) from commercially available building blocks with established organic reactions. Compounds **112**-**115** were subsequently

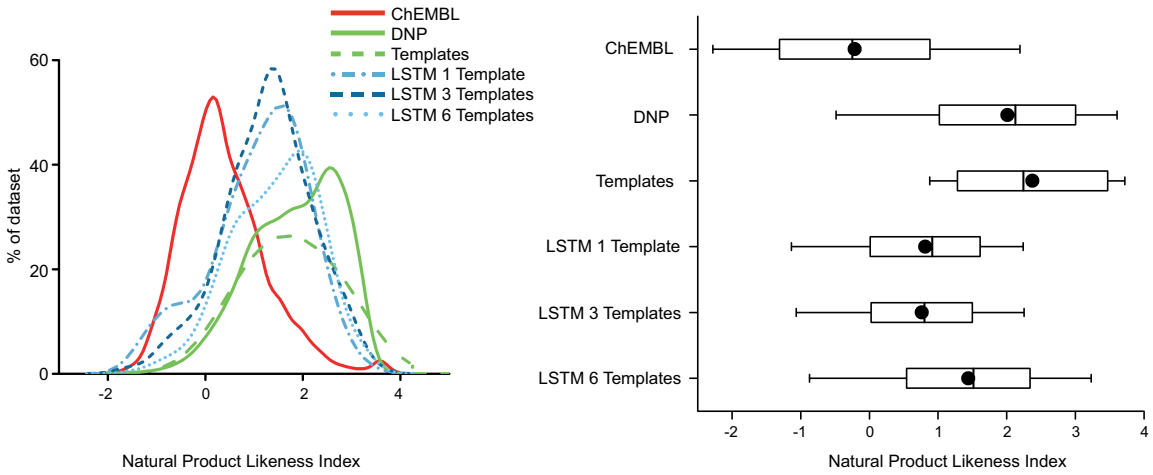



**Figure 4.22:** Comparison between the distribution of natural-product-likeness index[257] of the training molecules from ChEMBL[208], natural compounds from the Dictionary of Natural Products (DNP),[385] the natural product templates, and the sampled designs of the three different fine-tuned LSTM models. **a)** NP likeness index distribution of all compound sets and **b)** the corresponding boxplot representation. *De novo* designs showed a significantly higher natural-product-likeness than the ChEMBL compounds but are less natural-product-like than the DNP entries ($p$<0.001, Kruskal-Wallis test (H-Test) with post hoc Bonferroni correction with 30'000 randomly selected entries from ChEMBL and DNP, respectively). Box plots indicate mean (circle), median (solid line), standard deviation (box) and 1st/99th percentile (whiskers).

characterized *in vitro* for their modulatory effects on three RXR subtypes (Table 4.8). Design **112** and **113** showed no activities on RXRs at compound concentrations of 50 µM. **114** and **115** were confirmed as RXR agonists, whereof design **114** revealed potency on all three RXR subtypes in the double-digit micromolar range. Design **115** had agonistic activity on RXRα and RXRβ with low micromolar $EC_{50}$ values, whereas no activity of **115** on RXRγ was measured at 50 µM.

We analyzed the novelty of designs **112-115** by calculating their structural similarity to known RXR binders extracted from ChEMBL ($EC_{50}/IC_{50}$ < 50 µM) which were already used for scaffold analysis. We determined the maximum and average Jaccard-Tanimoto similarity index between each selected design and the known RXR modulators. The index ranges from 0 to 1 while greater values indicate higher molecular similarity. The designs as well as the RXR binders from ChEMBL were represented by four fingerprint descriptors commonly used in virtual screening[175] (AtomPairs fingerprints,[386] RDKiT fingerprints,[261] Morgan fingerprints,[144] and MACCS keys.[142] Low similarities were determined between designs **112-115** and known RXR modulators (Table 4.9). Especially the low similarity in terms of the presence of branched atom-centered fragments, as encoded by Morgan fingerprints, suggests structural novelty of

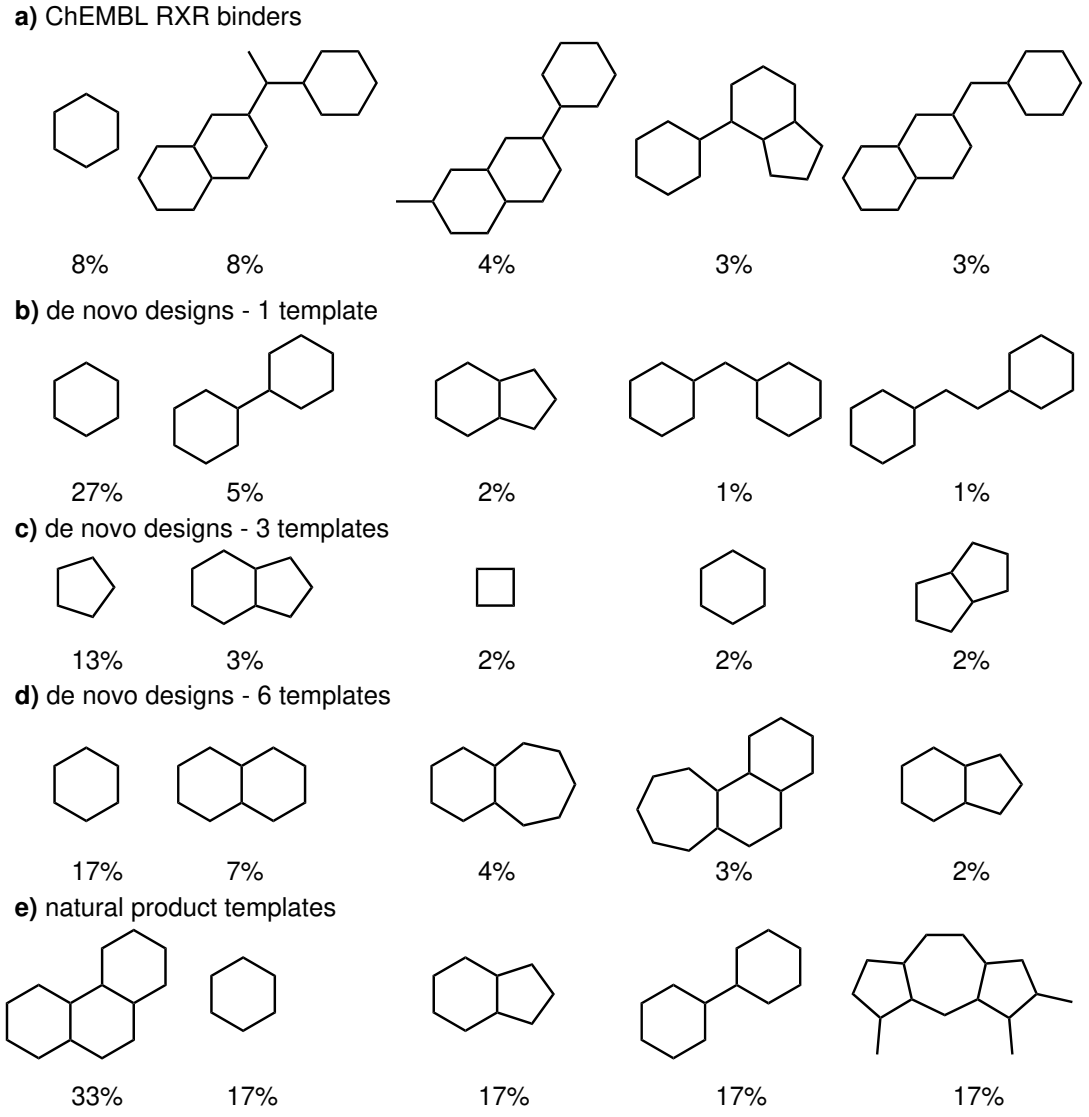



**Figure 4.23:** Molecular framework analysis of different compound sets. Most frequent occurring frameworks of known RXR binders from ChEMBL ($EC_{50}/IC_{50}$<50 µM, 521 compounds), the sampled designs generated by differently fine-tuned AI-models, and of the natural product templates. Percentage indicates the framework frequency within the considered compound set.

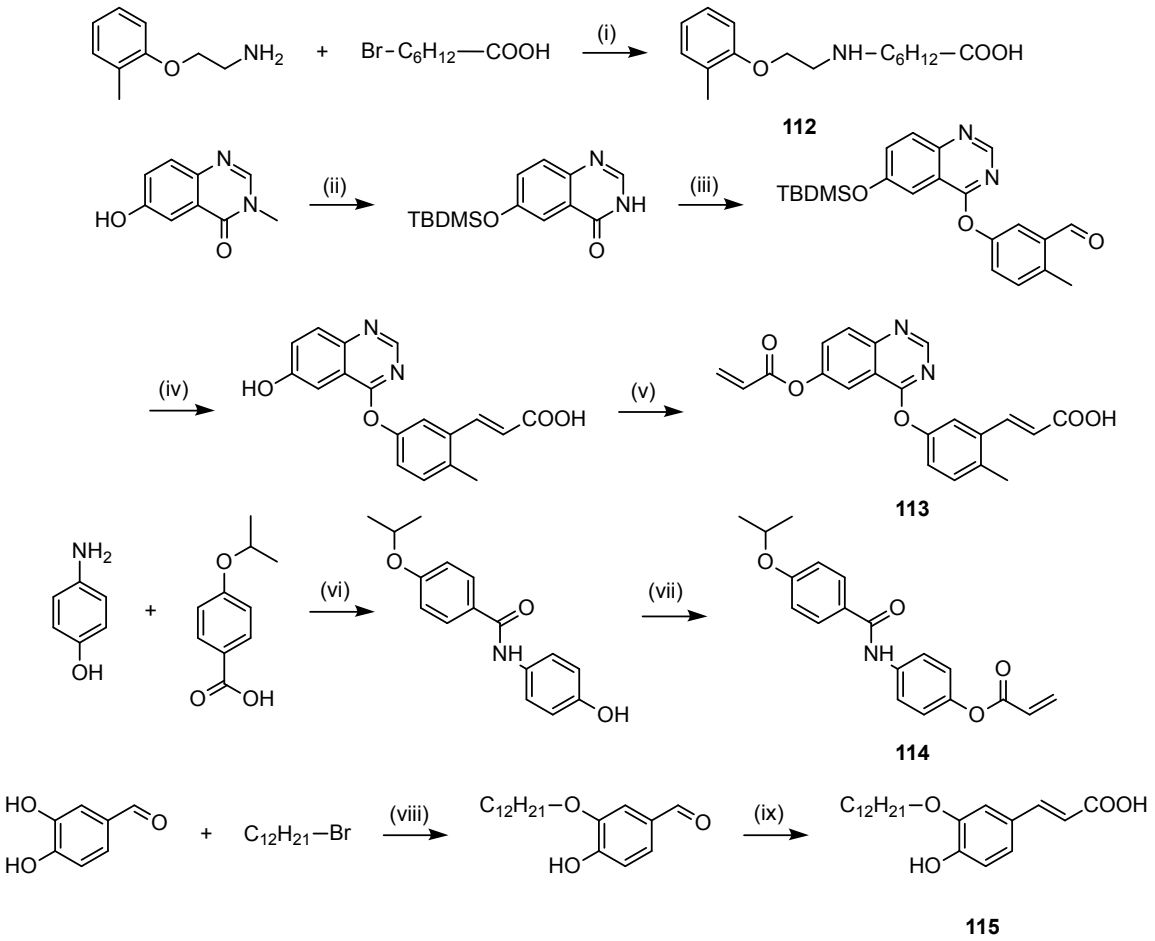



**Figure 4.24:** Synthesis of designs **112-115**.Reagents and conditions: (i) DMF, $NEt_3$, µw, 80 °C, 2 h, 18%; (ii) TBDMS−Cl, DMF, $NEt_3$, room temperature, 24 h, 74%; (iii) 3-formyl-4-methylphenylboronic acid, $Cu(OAc)_2$, 4 Å-molecular sieve, $CH_2Cl_2$, $NEt_3$, room temperature, 4 h, 94%; (iv) malonic acid, pyridine/piperidine, µw, 100 °C, 30 min., 96%; (v) acryloyl chloride, $CHCl_3$/DMF, $NEt_3$, room temperature, 2 h, 23%; (vi) EDC · HCl, 4-DMAP, $CHCl_3$, reflux, 16 h, 79%; (vii) acryloyl chloride, THF, pyridine, r.t., 2 h, 66%; (viii) DMF, $K_2CO_3$, room temperature, 4 h, 39%; (ix) malonic acid, pyridine/piperidine, µw, 100 °C, 30 min., 51%.

**Table 4.8:** *In vitro* activity of designs **112-115** on RXRs. Reported $EC_{50}$ values represent mean ± SEM [µM] and each concentration was tested in duplicates and each experiment was repeated two times for inactive and four times for active compounds.

| | $EC_{50}$ values ±SEM [µM] | | |
|---|---|---|---|
| Compound | RXRα | RXRβ | RXRγ |
| **112** | >50 | >50 | >50 |
| **113** | >50 | >50 | >50 |
| **114** | 29±5 | 27±1 | 19.1±0.1 |
| **115** | 16.9±0.6 | 15.7±0.8 | >50 |
| Natural product templates[1] | 2.1 - 42.0 | 4.6 - 42.0 | 7.0 - 42.0 |
| Ref. agonist[2] | 0.033±0.002 | 0.024±0.004 | 0.025±0.002 |

[1] Natural product template activities were reported as given in literature.[201, 382–384]
[2] reference agonists, literature data: bexarotene[380] for RXRα/β/γ.

the selected designs.

**Table 4.9:** The Jaccard-Tanimoto similarity index was computed for four types of molecular fingerprints (Atom Pairs, RdKiT, Morgan, MACCS keys) to quantify structural molecular similarity. Both the average and maximum similarity values to the ChEMBL RXR binders are reported for each *de novo* design.

| | AtomPairs | | RdKiT | | Morgan | | MACCS | |
|---|---|---|---|---|---|---|---|---|
| Compound | average | max | average | max | average | max | average | max |
| **112** | 0.26 | 0.37 | 0.25 | 0.34 | 0.14 | 0.25 | 0.31 | 0.52 |
| **113** | 0.39 | 0.49 | 0.49 | 0.63 | 0.16 | 0.23 | 0.35 | 0.59 |
| **114** | 0.32 | 0.42 | 0.35 | 0.41 | 0.15 | 0.24 | 0.32 | 0.52 |
| **115** | 0.28 | 0.43 | 0.32 | 0.41 | 0.17 | 0.33 | 0.36 | 0.70 |

## Conclusions

The results of the first study[246] validate the applicability of generative AI models to molecular *de novo* design. The computational design approach led to the identification of new agonists of nuclear receptors from known drug-like small molecules. The bioactive designs **101-105** possess considerable potency and diverse selectivity profiles on RXRs and PPARs. The designs were obtained by established chemical reactions from commercially available building blocks suggesting that the model intrinsically learned the chemical synthesizability. The results also indicate that a proper choice of molecules for model fine-tuning permits task-related *de novo* design by generative AI. With the second prospective study[247] we demonstrated that a small set of molecules can be sufficient to fine-tune an AI-based model by transfer learning. A set of natural products, sharing similar activities on the same target family, has been utilized as

template set for fine-tuning. The results illustrate the potential of this method to generate novel bioactive natural product mimetics located at uncharted areas in chemical space at the interface of bioactive natural products and drug-like synthetic molecules. Contrary to ruled-based *de novo* design approaches[128, 364] utilizing only one template, the applied deep learning concept is not capable designing mimetics of a single template. Under the currently used computational settings, the technique requires a set of templates with a common biological target. If a model is fine-tuned purely on a single template structure it will generate very often the template structure itself or close analogs due to the minimization of the error function of the neural network. This error function will reach a minimum if the template structure is reproduced. Therefore one cannot expect *de novo* design generation from a single template without optimizing the computational parameters of the deep neural network. On the basis of obtained results, even three templates are not sufficient to achieve the full potential of this method. However, the great benefit of this generative AI model is its ability to design bioactive NCEs from a set of selected bioactive molecules. The method was capable of designing bioactive compounds from fine-tuning templates with different origins, in particular synthetic drug-like molecules as well as natural products. Thus, generative AI for *de novo* design has the potential to enhance future computational drug discovery and especially natural product-inspired medicinal chemistry.

# 5 Conclusions and Outlook

Alternatively to traditional screening technologies for identifying novel bioactive molecules, computational *de novo* design holds promise for discovering NCEs with preferable biological activities by autonomous or semi-autonomous design approaches.[202] These *de novo* design methods have been mainly applied to identify pharmacologically active small molecules from drug-like reference compounds. In addition to such synthetic molecules, another source of inspirations for developing novel drug-like compounds are natural products. However, natural products have been rarely employed as reference structures in computational *de novo* design. Therefore, this thesis exploits the potential of the combination of natural products and computational *de novo* design. We employed the reaction-driven *de novo* design method as reported in my previous study[196] and extended our computational *de novo* design approach to generate bioactive NCEs from pharmacologically active natural products. In addition, generative deep learning models were evaluated in prospective applications. Such models present novel opportunities enabling the knowledge-driven design of molecules mimicking drug-like compounds and natural products.

In a first study, the neuroprotective 3,4-fused dihydrobenzofuran alkaloid galanthamine (**68**), an approved drug for treating mild to moderate Alzheimer's Disease[275], served as a template structure for reaction-driven *de novo* design tool DOGS (section 4.1). After assessing the pharmacophoric similarity (CATS) to galanthamine, three out of five designs with a CATS distance $\leq$1.2 contained the naturally occurring coumarin core. For practicability, the proposed mimetic compounds were structurally simplified. Such a manual modification step contradicts the basic principles of a reaction-driven design concept. Although DOGS generates virtual products according to established reactions, synthetic feasibility of each individual *in silico* design remains subject to evaluation by a chemist.

Six simplified coumarin-based mimetics were synthesized and evaluated in terms of their biological acitivities on computationally predicted molecular targets, whereof two mimetics weakly inhibited AChE, one of the main targets of galanthamine.[271] This initial finding indicates both the potential as well as the limitation of our design approach. On the one hand, the galanthamine mimetics inherited some of the essential structural features from their parent natural product for interacting with the known parental molecular target. On the other hand, these structures possessed only weak activities on AChE. This finding can be traced back to the similarity principle that similar molecular structures have similar biological activities.[157] The similarity assessment of two molecular structures varies by the choice of applied molecular descriptors and

similarity metric.[156] Therefore, the obtained results are in agreement with reported observations emphasizing that highly similar compounds do not inevitably possess similar biological activities.[171] The identified coumarin-based mimetics contained a natural product-derived "privileged" scaffold which is present in several natural products and drugs.[76] Privileged scaffolds are described as molecular frameworks which can be decorated with different side chains to generate ligands binding to various target proteins.[77] The bioactivity data gathered in this study substantiates this description of privileged scaffolds and classifies the synthesized compounds as "frequent hitters", a related concept designating structures which show activities in several biological assays.[**Roche2002a**] However, the proposed structures can serve as starting points for improving target specificity. Their molecular structures labeled as frequent hitters can be also employed in database filtering to exclude these promiscuous compounds from future screening collections and thereby increasing the efficiency of hit and lead identification in drug discovery. Overall, the findings of this study revealed that computational drug design has the potential of morphing a structurally intricate natural product into small molecules comprising a natural product scaffold which differs form the template scaffold. Contrary to the objective of generating NCEs, all mimetics were described before, whereas only two had annotated bioactivities. At first glance, the working hypothesis stating that our *de novo* design approach leads to novel bioactive molecules, has to be rejected. But on closer inspection the results of this study showcased that our design concept provides opportunities to re-discover chemical entities with readily available natural product-derived scaffolds. Also, hitherto unknown biological effects could be identified.

Based on the lessons learned from the aforementioned study, we applied our *de novo* design approach to identify novel mimetics of the marine natural product marinopyrrole A (section 4.2). The proposed target[306] of this anticancer agent was questioned and partially disproved in the last few years[307–309]. Further molecular targets of marinopyrrole A were computationally predicted and experimentally validated.[310] We virtually generated marinopyrrole A mimetics by the computational *de novo* design tool DOGS, synthesized selected NCEs and evaluated their activities on confirmed and other predicted molecular targets. The virtually constructed designs maintained certain levels of similarity regarding the two implemented ways to represent molecular structures. Both, molecular and reduced graph representation feature different levels of abstraction from the template's molecular constitution and composition aiming to generate molecules with higher or lower similarity to their template structure. Re-ranking the designs according to their pharmacophore similarity revealed that a lower level of abstraction implies a higher level of molecular similarity. Among the 100 generated designs with the highest pharmacophore similarity (CATS distance $\leq$ 1.8), 34 out of 100 designs contained a the 2,4,5-triphenyl imidazole scaffold. Selected structurally unmodified designs containing the 2,4,5-triphenyl imidazole scaffold were

synthesized according to the computationally proposed synthetic route. Opposite to the results of our first study, this outcome showed that reaction-driven *de novo* design tools are capable of proposing feasible synthetic routes to newly designed molecules only on the basis of implemented organic reactions. However, the current version of DOGS does not consider protecting groups. The proposed synthetic route was slightly adapted by introducing a proctection group. This type of manual optimization still requires human expertise. DOGS as rule-based tool is based on a pre-defined set of common organic reactions which limits the covered spaces to areas which are similar to those explored by medicinal chemists.[387] The question remains whether a larger set of implemented reaction scheme can be beneficial to explore promising areas of the biologically relevant chemical space. The fact that all synthesized mimetic structures were NCEs indicates that many "white spots" in apparently well-explored areas of the relevant chemical space still exist. Computational *de novo* design is well-suited to fill these gaps.

After synthesizing the marinopyrrole A mimetics, we examined their bioactivities on already known and potential novel targets as predicted by our computational target prediction software (SPiDER). Both the template and its mimetics had several predicted target families in common (Table 4.3). The *in vitro* characterization of these targets unveiled that in some cases the natural product and its mimetics displayed similar activities. In other cases no molecule had measurable effect at the compound concentration of the initial screening. When determining the inhibitory effects of marinopyrrole A and its mimetics on COX-1 and COX-2, we realized that both showed inhibition of at least one cyclooxygenase subtype (Table 4.4). But we also found that mimetic structures (**85a** and **87**) not only inherited necessary features from their parental natural product to inhibit the cyclooxygenases. they exceeded the activity of marinopyrrole A on COX-1 and showed a strong selectivity towards this subtype (Table 4.4). This results is a further indication that the molecular similarity concept holds true. Therefore, our design approach led to the discovery of mimetics surpassing the effects of their template structure. The characterized *in vitro* bioactivity of marinopyrrole A and its mimetics showcased that structurally similar compounds can possess similar bioactivity, but deviations from the similarity principle can be observed.

Overall, these results illustrated that our approach to select molecular targets with a computational ligand-based target prediction software like SPiDER relies again on the molecular similarity principle.[157] The fact that our approach is built on ligand information and their annotated targets but lacking structural information of the latter is both a blessing and a curse. A blessing because this concept can be applied to molecules with yet unknown targets like many natural products, or in cases where the protein structure has not been resolved yet. A curse because our approach covers a target space that is limited to the annotated targets of the associated ligand set. This excludes the identification of orphan targets. Methods which have been developed to

predict ligands for orphan targets[388], still need information about structurally similar protein structures as necessary prerequisite. Ligand-based concepts lack detailed knowledge about interactions between a ligand and its target. Especially precise information about regions of the ligands which are necessary for its binding to the target. It is reasonable to assume that ligand-based target prediction with "fuzzy" molecular descriptors will help to identify bioactive chemical entities with novel chemotypes rather than finding ligands with a high affinity to a molecular target. Besides the *in vitro* characterization of predicted targets, we tested our novel marinopyrrole A mimetics against previously recognized targets. In the case of the glucocorticoid receptor, both computationally retrieved designs **85** and **86** and the intermediate **87** exhibited both the ability to interact with the target and also exhibited comparable potencies in the performed binding competition assays. From a pharmacological point of view, these activities represent the most relevant findings of this study due to fact that the expression of the initial proposed target of marinopyrrole A, MCL-1, is directly regulated by the glucocorticoid receptor.[389] Further results from functional assays are necessary to characterize whether these structures agonize or antagonize glucocorticoid receptors and thus up- or downregulate the protein expression level of MCL-1 and other targets. Besides the glucocorticoid receptor, marinopyrrole A displayed activities with noteworthy potencies on CCK2 , OX1 and OX2, whereas only one mimetic compound (**86**) showed bioactivity on two out of three targets (CCK2 and OX1). Finally, the results obtained in this study provide further evidence that our ligand-based design approach possesses a huge potential to successfully design and pick mimetic compounds by *in silico* tools together with a computational selection of possible molecular targets. This study comprising of a fully computer-guided "design-make-test" cycle offers opportunities to discover novel active compounds from bioactive ligands of which no molecular targets have been disclosed yet, or a structure-based approach can be excluded due to the lack of resolved target protein structures.

The two previously mentioned studies comprised a computational design approach combining 2D molecular representation, similarity assessment and target prediction. As such, all kind of three-dimensional information was ignored. In principle, 2D-based descriptors can provide comparable and in some cases better results than the corresponding 3D descriptors.[131] A careful estimation of benefits and drawbacks of each selected descriptor is a necessary requirement.[204] Due to the fact that both molecular targets and ligands are 3D objects, the preferred spatial orientation of such a ligand is an essential aspect. 3D Descriptions of molecules have a higher information content and especially natural products with their unique frameworks displayed improved selectivity and frequency to biological targets.[390] Consequently, we extended our design strategy by incorporating a 3D similarity method to select molecules with similar shapes (section 4.3). The chosen method describes the molecular shape in terms of the

fractal dimensionality of its surface, expressed as a float number. The concept of fractal dimension has been applied before to identify surface patches of macromolecular structures which are involved in ligand-protein interactions.[347] As a case study, we reverted back to a previous example in which (-)-englerin A served as the template for our design approach.[196] The resulting mimetics displayed activities on some known targets of englerin A (TRPM8), but were inactive on TRPC4 activated by englerin A with nanomolar concentrations.[198, 199] We re-ranked the *de novo* design collection according to the shape similarity of each entry to (-)-englerin A. We further prioritized designs of the most similar compounds which had a predicted activity on the TRP ion channel family. 10 out of 30 structures were predicted as active and two compounds (**92** and **93**) were synthesized and characterized by their modulatory effects on TRP ion channels. Mimetic **92** showed weak activity on TRPC4 at the screening concentration, but was inactive on all other examined TRP channels. Molecule **93**, however, inhibited TRPC4 and TRPM8 at low micromolar concentrations, and had weak inhibitory effects on TRPV4. These observations illustrate that essential structural features have been translated from the parent natural product to the mimetic through our earlier design approach. But the results also indicate again the boundaries of our strategy. To computationally predict the different modes of action of a given ligand, information about the target structure, its flexibility and binding modes of other active ligands are an essential prerequisite. Given that this information is not included in our ligand-based approach, our strategy is unsuitable to intentionally design a modulator with a certain mode of action.

Taken together, the results from these three studies implicate that the combination of reaction-driven *de novo* design and natural products offers a promising concept to discover novel bioactive molecules for early drug discovery. Our concept provides a customizable computational framework to identify pharmacologically active NCEs from natural product templates. However, the major limitation of our design approach is the inability to incorporate structural information from different molecules. Integrating several reference structures with common molecular targets could help to recognize essential but non-obvious structural features conserved across different reference structures. A way to overcome this limitation would be to combine the results from independent design processes of each reference structure. In theory, joint chemical entities of all generated designs should have inherited some of the shared structural features from all reference structures. However, the probability to identify identical structures from distinct design runs is rather small. It remains to be examined to what extent this information is preserved during the entire design approach.

Generative deep learning models present an innovative solution to create new bioactive chemical entities by capturing crucial features from large sets of reference structures. In detail, we utilized RNNs containing LSTM cells[242] and trained these on

a large set of highly active compounds, retrieved from the ChEMBL database. These generative LSTM models have been shown to generate valid molecular structures from their SMILES representations.[244] In two prospective applications ((section 4.4)) we showed that these deep learning models were capable of designing new bioactive chemical entities from different sets of active ligands. In our first proof-of-principle study, we employed 25 molecules with known agonistic effects on two nuclear receptor types (PPAR and RXR) for transfer learning. The sampled designs from this fine-tuned model showed high proportion of valid and unique structures (90%) of which all generated molecules represented novel entities with novel meaning that they were neither identical to entries in the training nor to the fine-tuning set. The areas of chemical space populated by the generated designs were located within the training data domain and additionally covered target-related regions of the two investigated nuclear receptors (Figure 4.16). This finding implies that LSTM-based generative models are able to encompass structural features shared among several active ligands of a particular target by exclusively taking their SMILES representations into account. The possibility to define a structural element as a common starting fragment of all samples turned out to be a useful extension to obtain a target-focused library. Designs for synthesis and *in vitro* characterization were selected by molecular similarity assessment and *in silico* target prediction. The synthetic feasibility of the five selected designs is an indication that the model implicitly learned to generate synthesizable moelcules form training on drug-like compounds. The experimental results corroborate the theoretical observations that generative deep learning models are well suited to design bioactive NCEs on the basis different ligands with annotated activities. The outcome of this proof-of-principle study marks an important point in the field of generative deep learning as novel approach in molecular *de novo* design. This study confirmed experimentally the so far only theoretically proven capability to design novel bioactive molecules by generative deep learning. After this first successful application, we employed generative deep learning models to obtain mimetic structures from a selection of natural products interacting with the same protein target family. As in the previous study, a deep learning model, trained on drug-like molecules, was fine-tuned on up to six natural products. Fine-tuning on a single natural product would be a comparable approach to our previously applied reaction-driven *de novo* design concept. However, the findings revealed that the model was not able to create appropriate chemical entities with one fine-tuning template. The observed increased level of valid and unique structures showed that with six template structures, our deep learning model can generate an sufficient amount of *de novo* designs. Independent of the fine-tuning set size, all *de novo* designs possessed molecular frameworks differing from their templates and known RXR binders. This indicates that our model generates chemical entities with novel structural features allowing to discover yet under-explored areas in chemical space. In terms of their natural product likeness score,[257] the generated designs were less

natural product-like than entries of the DNP, but more natural product-like than compounds of the ChEMBL database. This reflects the fact that the model was initially trained on molecules retrieved from ChEMBL and then fine-tuned on a set of natural products. The model seemed to have captured structural features of both domains, natural products and synthetic compounds. After synthesis of the four selected compounds (**112-115**), *in vitro* characterization of these mimetics on the three RXR subtyes revealed that two out of four molecules showed agonistic effects on at least two subtypes of RXR with $EC_{50}$ values between 16.9 - 29.0 µM. Although the individual *in vitro* activities of the mimetics were lower in comparison to their natural product templates (2.1 - 42.0 µM), the overall activity range of the designs was similar to the range of their templates. Overall, our investigations confirm the hypothesis that novel bioactive chemical entities can be designed by the help of generative deep learning.
In this thesis, we have demonstrated that the combination of computational methods and natural products as a source of inspiration can lead to the identification of bioactive NCEs for early drug discovery. Reaction-driven molecule assembly and generative deep learning represent powerful tools which can utilize natural product templates to design synthetically more accessible and bioactive mimetics. The current limitations, like unfeasible proposed synthetic pathways, or the inability to work with more than one template structure, or the uncertainties in predicting biological activities, only to name a few, present valid starting points for future studies to enhance the full potential of these tools.

# Bibliography


[1] T. Rodrigues, D. Reker, P. Schneider, G. Schneider, "Counting on natural products for drug design", *Nature Chemistry* **2016**, *8*, 531–541.

[2] K.-H. Altmann, "Drugs from the Oceans: Marine Natural Products as Leads for Drug Discovery", *CHIMIA International Journal for Chemistry* **2017**, *71*, 646–652.

[3] S. Bernardini, A. Tiezzi, V. Laghezza Masci, E. Ovidi, "Natural products for human health: an historical overview of the drug discovery approaches", *Natural Product Research* **2018**, *32*, 1926–1950.

[4] S. Wang, G. Dong, C. Sheng, "Structural Simplification of Natural Products", *Chemical Reviews* **2019**, DOI: 10.1021/acs.chemrev.8b00504.

[5] F. McTaggart et al., "Preclinical and clinical pharmacology of rosuvastatin, a new 3-hydroxy- 3-methylglutaryl coenzyme A reductase inhibitor11CRESTOR is a trademark, the property of AstraZeneca PLC. Research discussed in this article was supported by AstraZeneca." *The American Journal of Cardiology* **2001**, *87*, 28–32.

[6] G. Schneider, W. Neidhart, T. Giller, G. Schmid, ""Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening", *Angewandte Chemie International Edition* **1999**, *38*, 2894–2896.

[7] J. Gay-lussac, N. L. Vauquelin, E. Supe, F. Achard in *Drug Discovery: A History*, John Wiley & Sons, Ltd, Chichester, UK, **2006**, pp. 88–105.

[8] R. B. Woodward, W. E. Doering, "The Total Synthesis of Quinine", *Journal of the American Chemical Society* **1945**, *67*, 860–874.

[9] J. M. Dziekan et al., "Identifying purine nucleoside phosphorylase as the target of quinine using cellular thermal shift assay", *Science Translational Medicine* **2019**, *11*, eaau3174.

[10] F. Chast in *The Practice of Medicinal Chemistry*, (Ed.: C. G. Wermuth), Elsevier, **2008**, pp. 1–62.

[11] In *IUPAC Compendium of Chemical Terminology*, IUPAC, Research Triagle Park, NC.

[12] F. Sertürner, "Auszüge aus Briefen an den Herausgeber. (a) Säure im Opium. (b) Ein deres Schreiben von Ebendenselben. Nachtrag zur Charakteristik der Säure im Opium", *J. Pharmazie für Ärzte Apotheker und Chemisten von D.J.B. Trommsdorff* **1805**, 29–30.

[13] F. Sertürner, "Darstellung der reinen Mohnsäure (Opiumsäure) nebst einer Chemischen Untersuchung des Opiums mit vorzüglicher Hinsicht auf einen darin neu entdeckten Stoff und die dahin gehörigen Bemerkungen ." *J. Pharmazie für Ärzte Apotheker und Chemisten von D.J.B. Trommsdorff* **1806**, 47–93.

[14] G. Merck, "Vorläufige Notiz über eine neue organische Base im Opium", *Annalen der Chemie und Pharmacie* **1848**, *66*, 125–128.

[15] H. Kaiser, "Von der Pflanze zur Chemie – Die Frühgeschichte der „Rheumamittel"", *Zeitschrift für Rheumatologie* **2008**, *67*, 252–262.

[16] A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, "The re-emergence of natural products for drug discovery in the genomics era", *Nature Reviews Drug Discovery* **2015**, *14*, 111–129.

[17] C. H. Gerhardt, "Untersuchungen über die wasserfreien organischen Säuren", *Annalen der Chemie und Pharmacie* **1853**, *87*, 149–179.

[18] A. Schröder, A. Prinzhorn, K. Kraut, "Über Salicylverbindungen", *Annalen der Chemie und Pharmacie* **1869**, *150*, 1–20.

[19] W. Sneader, "The discovery of aspirin: a reappraisal", *BMJ* **2000**, *321*, 1591–1594.

[20] K. C. Nicolaou, D. Vourloumis, N. Winssinger, P. S. Baran, "The Art and Science of Total Synthesis at the Dawn of the Twenty-First Century", *Angewandte Chemie International Edition* **2000**, *39*, 44–122.

[21] P. S. Baran, "Natural Product Total Synthesis: As Exciting as Ever and Here To Stay", *Journal of the American Chemical Society* **2018**, *140*, 4751–4755.

[22] D. J. Newman, G. M. Cragg, "Natural Products as Sources of New Drugs from 1981 to 2014", *Journal of Natural Products* **2016**, *79*, 629–661.

[23] J. J. Manfredi, S. B. Horwitz, "Taxol: an antimitotic agent with a new mechanism of action", *Pharmacology & Therapeutics* **1984**, *25*, 83–125.

[24] M. Lebwohl, N. Swanson, L. L. Anderson, A. Melgaard, Z. Xu, B. Berman, "Ingenol Mebutate Gel for Actinic Keratosis", *New England Journal of Medicine* **2012**, *366*, 1010–1019.

[25] N. Kedei, D. J. Lundberg, A. Toth, P. Welburn, S. H. Garfield, P. M. Blumberg, "Characterization of the Interaction of Ingenol 3-Angelate with Protein Kinase C", *Cancer Research* **2004**, *64*, 3243–3255.

[26] S. N. Freiberger et al., "Ingenol Mebutate Signals via PKC/MEK/ERK in Keratinocytes and Induces Interleukin Decoy Receptors IL1R2 and IL13RA2", *Molecular Cancer Therapeutics* **2015**, *14*, 2132–2142.

[27] L. Meng, R. Mohan, B. H. B. Kwok, M. Elofsson, N. Sin, C. M. Crews, "Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity", *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 10403–10408.

[28] H.-J. Zhou et al., "Design and Synthesis of an Orally Bioavailable and Selective Peptide Epoxyketone Proteasome Inhibitor (PR-047)", *Journal of Medicinal Chemistry* **2009**, *52*, 3028–3038.

[29] L. Nagy et al., "Activation of retinoid X receptors induces apoptosis in HL-60 cell lines." *Molecular and Cellular Biology* **1995**, *15*, 3540–3551.

[30] S. Thacher, J. Vasudevan, R. Chandraratna, "Therapeutic Applications for Ligands of Retinoid Receptors", *Current Pharmaceutical Design* **2000**, *6*, 25–58.

[31] E. A. Crane, K. Gademann, "Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis", *Angewandte Chemie International Edition* **2016**, *55*, 3882–3902.

[32] K. NAKAE et al., "Migrastatin, a New Inhibitor of Tumor Cell Migration from Streptomyces sp. MK929-43F1. Taxonomy, Fermentation, Isolation and Biological Activities." *The Journal of Antibiotics* **2000**, *53*, 1130–1136.

[33] C. Gaul, J. T. Njardarson, S. J. Danishefsky, "The Total Synthesis of (+)-Migrastatin", *Journal of the American Chemical Society* **2003**, *125*, 6042–6043.

[34] J. T. Njardarson, C. Gaul, D. Shan, X.-Y. Huang, S. J. Danishefsky, "Discovery of Potent Cell Migration Inhibitors through Total Synthesis: Lessons from Structure-Activity Studies of (+)-Migrastatin", *Journal of the American Chemical Society* **2004**, *126*, 1038–1040.

[35] D. Shan et al., "Synthetic analogues of migrastatin that inhibit mammary tumor metastasis in mice", *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 3772–3776.

[36] T. Oskarsson et al., "Diverted Total Synthesis Leads to the Generation of Promising Cell-Migration Inhibitors for Treatment of Tumor Metastasis: In vivo and Mechanistic Studies on the Migrastatin Core Ether Analog", *Journal of the American Chemical Society* **2010**, *132*, 3224–3228.

[37] R. M. Wilson, S. J. Danishefsky, "Small Molecule Natural Products in the Discovery of Therapeutic Agents: The Synthesis Connection", *Journal of Organic Chemistry* **2006**, *71*, 8329–8351.

[38] P. A. Wender, V. A. Verma, T. J. Paxton, T. H. Pillow, "Function-oriented synthesis, step economy, and drug design", *Accounts of Chemical Research* **2008**, *41*, 40–49.

[39] P. A. Wender, R. V. Quiroz, M. C. Stevens, "Function through Synthesis-Informed Design", *Accounts of Chemical Research* **2015**, *48*, 752–760.

[40] P. A. Wender, K. F. Koehler, N. A. Sharkey, M. L. Dell'Aquila, P. M. Blumberg, "Analysis of the phorbol ester pharmacophore on protein kinase C as a guide to the rational design of new classes of analogs." *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83*, 4214–8.

[41] P. A. Wender, A. C. Donnelly, B. A. Loy, K. E. Near, D. Staveness in *Natural Products in Medicinal Chemistry*, (Ed.: S. Hanessian), **2014**, pp. 473–544.

[42] G. R. Pettit, C. L. Herald, D. L. Doubek, D. L. Herald, E. Arnold, J. Clardy, "Isolation and structure of bryostatin 1", *Journal of the American Chemical Society* **1982**, *104*, 6846–6848.

[43] P. A. Wender et al., "Modeling of the bryostatins to the phorbol ester pharmacophore on protein kinase C." *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 7197–7201.

[44] P. A. Wender et al., "The practical synthesis of a novel and highly potent analogue of bryostatin." *Journal of the American Chemical Society* **2002**, *124*, 13648–9.

[45] G. E. Keck, Y. B. Poudel, T. J. Cummins, A. Rudra, J. A. Covel, "Total Synthesis of Bryostatin 1", *Journal of the American Chemical Society* **2011**, *133*, 744–747.

[46] Y. Wu, C. Dockendorff, "Synthesis of a novel bicyclic scaffold inspired by the antifungal natural product sordarin", *Tetrahedron Letters* **2018**, *59*, 3373–3376.

[47] S. Rizzo, V. Wakchaure, H. Waldmann in *Natural Products in Medicinal Chemistry*, (Ed.: S. Hanessian), **2014**, pp. 43–80.

[48] D. G. Brown, J. Boström, "Where Do Recent Small Molecule Clinical Development Candidates Come From?", *Journal of Medicinal Chemistry* **2018**, *61*, 9442–9468.

[49] T. Kogej et al., "Big pharma screening collections: more of the same or unique libraries? The AstraZeneca–Bayer Pharma AG case", *Drug Discovery Today* **2013**, *18*, 1014–1024.

[50] M. Follmann et al., "An approach towards enhancement of a screening library: The Next Generation Library Initiative (NGLI) at Bayer — against all odds?", *Drug Discovery Today* **2018**, DOI `10.1016/j.drudis.2018.12.003`.

[51] A. Karawajczyk et al., "Expansion of chemical space for collaborative lead generation and drug discovery: the European Lead Factory Perspective", *Drug Discovery Today* **2015**, *20*, 1310–1316.

[52] J. Besnard, P. S. Jones, A. L. Hopkins, A. D. Pannifer, "The Joint European Compound Library: boosting precompetitive research", *Drug Discovery Today* **2015**, *20*, 181–186.

[53] E. Perola, "An Analysis of the Binding Efficiencies of Drugs and Their Leads in Successful Drug Discovery Programs", *Journal of Medicinal Chemistry* **2010**, *53*, 2986–2997.

[54] H. O. Villar, M. R. Hansen, "Design of chemical libraries for screening", *Expert Opinion on Drug Discovery* **2009**, *4*, 1215–1220.

[55] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, "Quantifying the chemical beauty of drugs", *Nature Chemistry* **2012**, *4*, 90–98.

[56] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings", *Advanced Drug Delivery Reviews* **1997**, *23*, 3–25.

[57] J.-L. Reymond, "The Chemical Space Project", *Accounts of Chemical Research* **2015**, *48*, 722–730.

[58] M. L. Lee, G. Schneider, "Scaffold architecture and pharmacophoric properties of natural products and trade drugs: Application in the design of natural product-based combinatorial libraries", *Journal of Combinatorial Chemistry* **2001**, *3*, 284–289.

[59] T. Henkel, R. M. Brunne, H. Müller, F. Reichel, "Statistical investigation into the structural complementarity of natural products and synthetic compounds", *Angewandte Chemie International Edition* **1999**, *38*, 643–647.

[60] F. L. Stahura, J. W. Godden, L. Xue, J. Bajorath, "Distinguishing between Natural Products and Synthetic Molecules by Descriptor Shannon Entropy Analysis and Binary QSAR Calculations", *Journal of Chemical Information and Computer Sciences* **2000**, *40*, 1245–1252.

[61] M. Feher, J. M. Schmidt, "Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry", *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 218–227.

[62] P. Selzer, H.-J. Roth, P. Ertl, A. Schuffenhauer, "Complex molecules: do they add value?", *Current Opinion in Chemical Biology* **2005**, *9*, 310–316.

[63] C. F. Stratton, D. J. Newman, D. S. Tan, "Cheminformatic comparison of approved drugs from natural product versus synthetic origins", *Bioorganic & Medicinal Chemistry Letters* **2015**, *25*, 4802–4807.

[64] A. Ganesan, "The impact of natural products upon modern drug discovery", *Current Opinion in Chemical Biology* **2008**, *12*, 306–317.

[65] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, "Molecular Properties That Influence the Oral Bioavailability of Drug Candidates", *Journal of Medicinal Chemistry* **2002**, *45*, 2615–2623.

[66] B. C. Doak, B. Over, F. Giordanetto, J. Kihlberg, "Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates", *Chemistry & Biology* **2014**, *21*, 1115–1142.

[67] B. C. Doak, J. Zheng, D. Dobritzsch, J. Kihlberg, "How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets", *Journal of Medicinal Chemistry* **2016**, *59*, 2312–2327.

[68] D. A. DeGoey, H.-J. Chen, P. B. Cox, M. D. Wendt, "Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection", *Journal of Medicinal Chemistry* **2018**, *61*, 2636–2651.

[69] R. J. Quinn et al., "Developing a Drug-like Natural Product Library", *Journal of Natural Products* **2008**, *71*, 1792–1792.

[70] F. Lovering, J. Bikker, C. Humblet, "Escape from flatland: Increasing saturation as an approach to improving clinical success", *Journal of Medicinal Chemistry* **2009**, *52*, 6752–6756.

[71] F. Lovering, "Escape from Flatland 2: Complexity and promiscuity", *MedChemComm* **2013**, *4*, 515–519.

[72] G. W. Bemis, M. A. Murcko, "The Properties of Known Drugs. 1. Molecular Frameworks", *Journal of Medicinal Chemistry* **1996**, *39*, 2887–2893.

[73] N. Brown in *Scaffold Hopping in Medicinal Chemistry*, (Ed.: N. Brown), Wiley-VCH Verlag GmbH & Co. KGaA, **2013**, Chapter 1, pp. 1–14.

[74] K. Grabowski, G. Schneider, "Properties and Architecture of Drugs and Natural Products Revisited", *Current Chemical Biology* **2007**, *1*, 115–127.

[75] K. Grabowski, K. H. Baringhaus, G. Schneider, "Scaffold diversity of natural products: Inspiration for combinatorial library design", *Natural Product Reports* **2008**, *25*, 892–904.

[76] M. E. Welsch, S. A. Snyder, B. R. Stockwell, "Privileged scaffolds for library design and drug discovery", *Current Opinion in Chemical Biology* **2010**, *14*, 347–361.

[77] P. Schneider, G. Schneider, "Privileged Structures Revisited", *Angewandte Chemie International Edition* **2017**, *56*, 7971–7974.

[78] B. E. Evans et al., "Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists", *Journal of Medicinal Chemistry* **1988**, *31*, 2235–2246.

[79] M. Pascolutti, M. Campitelli, B. Nguyen, N. Pham, A.-D. Gorse, R. J. Quinn, "Capturing Nature's Diversity", *PLOS ONE* **2015**, *10*, (Ed.: B. Wagner), e0120942.

[80] J. Hert, J. J. Irwin, C. Laggner, M. J. Keiser, B. K. Shoichet, "Quantifying biogenic bias in screening libraries", *Nature Chemical Biology* **2009**, *5*, 479–483.

[81] E. J. Barreiro in *Privileged Scaffolds in Medicinal Chemistry: Design, Synthesis, Evaluation*, (Ed.: S. Bräse), 50, **2015**, Chapter 1, pp. 1–15.

[82] *Privileged Scaffolds in Medicinal Chemistry*, (Ed.: S. Bräse), Royal Society of Chemistry, Cambridge, **2015**.

[83] D. J. Newman, G. M. Cragg in *Privileged Scaffolds in Medicinal Chemistry: Design, Synthesis, Evaluation*, (Ed.: S. Bräse), Royal Society of Chemistry, **2015**, Chapter 13, pp. 348–378.

[84] S. E. Kearney et al., "Canvass: A Crowd-Sourced, Natural-Product Screening Library for Exploring Biological Space", *ACS Central Science* **2018**, *4*, 1727–1741.

[85] Y. Chen, M. Garcia de Lomana, N.-O. Friedrich, J. Kirchmair, "Characterization of the Chemical Space of Known and Readily Obtainable Natural Products", *Journal of Chemical Information and Modeling* **2018**, *58*, 1518–1532.

[86] S. L. Schreiber, "Target-Oriented and Diversity-Oriented Organic Synthesis in Drug Discovery", *Science* **2000**, *287*, 1964–1969.

[87] W. Wilk, T. J. Zimmermann, M. Kaiser, H. Waldmann, "Principles, implementation, and application of biology-oriented synthesis (BIOS)", *Biological Chemistry* **2010**, *391*, 491–497.

[88] M. D. Burke, G. Lalic, "Teaching Target-Oriented and Diversity-Oriented Organic Synthesis at Harvard University", *Chemistry & Biology* **2002**, *9*, 535–541.

[89] D. S. Tan, "Diversity-oriented synthesis: exploring the intersections between chemistry and biology", *Nature Chemical Biology* **2005**, *1*, 74–84.

[90] D. S. Tan, M. A. Foley, M. D. Shair, S. L. Schreiber, "Stereoselective Synthesis of over Two Million Compounds Having Structural Features Both Reminiscent of Natural Products and Compatible with Miniaturized Cell-Based Assays", *Journal of the American Chemical Society* **1998**, *120*, 8565–8566.

[91] J. C. Rohloff et al., "Practical Total Synthesis of the Anti-Influenza Drug GS-4104", *The Journal of Organic Chemistry* **1998**, *63*, 4545–4550.

[92] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, "A new type of synthetic peptide library for identifying ligand-binding activity", *Nature* **1991**, *354*, 82–84.

[93] D. R. Dragoli, L. A. Thompson, J. O'Brie, J. A. Ellman, "Parallel Synthesis of Prostaglandin E 1 Analogues", *Journal of Combinatorial Chemistry* **1999**, *1*, 534–539.

[94] Y. Kashman et al., "HIV inhibitory natural products. Part 7. The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical rainforest tree, Calophyllum lanigerum", *Journal of Medicinal Chemistry* **1992**, *35*, 2735–2743.

[95] K. C. Nicolaou, J. A. Pfefferkorn, A. J. Roecker, G.-Q. Cao, S. Barluenga, H. J. Mitchell, "Natural Product-like Combinatorial Libraries Based on Privileged Structures. 1. General Principles and Solid-Phase Synthesis of Benzopyrans", *Journal of the American Chemical Society* **2000**, *122*, 9939–9953.

[96] K. C. Nicolaou et al., "Natural Product-like Combinatorial Libraries Based on Privileged Structures. 2. Construction of a 10 000-Membered Benzopyran Library by Directed Split-and-Pool Chemistry Using NanoKans and Optical Encoding", *Journal of the American Chemical Society* **2000**, *122*, 9954–9967.

[97] K. C. Nicolaou, J. A. Pfefferkorn, S. Barluenga, H. J. Mitchell, A. J. Roecker, G.-Q. Cao, "Natural Product-like Combinatorial Libraries Based on Privileged Structures. 3. The "Libraries from Libraries" Principle for Diversity Enhancement of Benzopyran Libraries", *Journal of the American Chemical Society* **2000**, *122*, 9968–9976.

[98] H. Oguri et al., "Generation of Anti-trypanosomal Agents through Concise Synthesis and Structural Diversification of Sesquiterpene Analogues", *Journal of the American Chemical Society* **2011**, *133*, 7096–7105.

[99] X.-z. Su, L. H. Miller, "The discovery of artemisinin and the Nobel Prize in Physiology or Medicine", *Science China Life Sciences* **2015**, *58*, 1175–1179.

[100] A. Noren-Muller et al., "Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis", *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 10606–10611.

[101] Y. Zhang, I. A. Hubner, A. K. Arakaki, E. Shakhnovich, J. Skolnick, "On the origin and highly likely completeness of single-domain protein structures", *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 2605–2610.

[102] S. Wetzel, R. S. Bon, K. Kumar, H. Waldmann, "Biology-Oriented Synthesis", *Angewandte Chemie International Edition* **2011**, *50*, 10800–10826.

[103] H. Van Hattum, H. Waldmann, "Biology-Oriented Synthesis: Harnessing the Power of Evolution", *Journal of the American Chemical Society* **2014**, *136*, 11853–11859.

[104] M. a. Koch et al., "Charting biologically relevant chemical space: A structural classification of natural products (SCONP)", *Proceedings of the National Academy of Sciences* **2005**, *102*, 17272–17277.

[105] A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch, H. Waldmann, "The Scaffold Tree - Visualization of the Scaffold Universe by Hierarchical Scaffold Classification", *Journal of Chemical Information and Modeling* **2007**, *47*, 47–58.

[106] T. Varin, A. Schuffenhauer, P. Ertl, S. Renner, "Mining for Bioactive Scaffolds with Scaffold Networks: Improved Compound Set Enrichment from Primary Screening Data", *Journal of Chemical Information and Modeling* **2011**, *51*, 1528–1538.

[107] B. Over et al., "Natural-product-derived fragments for fragment-based ligand discovery", *Nature Chemistry* **2013**, *5*, 21–28.

[108] C. W. Murray, D. C. Rees, "The rise of fragment-based drug discovery", *Nature Chemistry* **2009**, *1*, 187–192.

[109] S. D. Roughley, R. E. Hubbard, "How Well Can Fragments Explore Accessed Chemical Space? A Case Study from Heat Shock Protein 90", *Journal of Medicinal Chemistry* **2011**, *54*, 3989–4005.

[110] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, "Twenty years on: the impact of fragments on drug discovery", *Nature Reviews Drug Discovery* **2016**, *15*, 605–619.

[111] G. Karageorgis et al., "Chromopynones are pseudo natural product glucose uptake inhibitors targeting glucose transporters GLUT-1 and -3", *Nature Chemistry* **2018**, *10*, 1103–1111.

[112] C. R. Pye, M. J. Bertin, R. S. Lokey, W. H. Gerwick, R. G. Linington, "Retrospective analysis of natural products provides insights for future discovery trends", *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 5601–5606.

[113] E. A. Crane, K. Gademann, "Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis", *Angewandte Chemie International Edition* **2016**, n/a–n/a.

[114] G. Schneider, U. Fechner, "Computer-based de novo design of drug-like molecules", *Nature Reviews Drug Discovery* **2005**, *4*, 649–663.

[115] M. J. Weyant, A. M. Carothers, A. J. Dannenberg, M. M. Bertagnolli, "(+)-Catechin inhibits intestinal tumor formation and suppresses focal adhesion kinase activation in the min/+ mouse." *Cancer Research* **2001**, *61*, 118–25.

[116] D. M. Lambert, C. J. Fowler, "The Endocannabinoid System: Drug Targets, Lead Compounds, and Potential Therapeutic Applications", *Journal of Medicinal Chemistry* **2005**, *48*, 5059–5087.

[117] N. Katayama, S. Fukusumi, Y. Funabashi, T. Iwahi, H. Ono, "TAN-1057 A-D, new antibiotics with potent antibacterial activity against methicillin-resistant Staphylococcus aureus. Taxonomy, fermentation and biological activity." *The Journal of Antibiotics* **1993**, *46*, 606–613.

[118] F. Reyes et al., "Aplicyanins A–F, new cytotoxic bromoindole derivatives from the marine tunicate Aplidium cyaneum", *Tetrahedron* **2008**, *64*, 5119–5123.

[119] C. C. Barron, P. J. Bilan, T. Tsakiridis, E. Tsiani, "Facilitative glucose transporters: Implications for cancer detection, prognosis and treatment", *Metabolism* **2016**, *65*, 124–139.

[120] W. L. Jorgensen, "The Many Roles of Computation in Drug Discovery", *Science* **2004**, *303*, 1813–1818.

[121] G. Sliwoski, S. Kothiwale, J. Meiler, E. W. Lowe, "Computational Methods in Drug Discovery", *Pharmacological Reviews* **2013**, *66*, 334–395.

[122] J. Bajorath, "Computer-aided drug discovery", *F1000Research* **2015**, *4*, 630.

[123] G. Schneider, *De novo Molecular Design*, (Ed.: G. Schneider), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2013**, pp. 1–551.

[124] C. Hansch, P. P. Maloney, T. Fujita, R. M. Muir, "Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients", *Nature* **1962**, *194*, 178–180.

[125] A. Cherkasov et al., "QSAR Modeling: Where Have You Been? Where Are You Going To?", *Journal of Medicinal Chemistry* **2014**, *57*, 4977–5010.

[126] L. B. Kier, L. H. Hall, J. W. Frazer, "Design of molecules from quantitative structure-activity relationship models. 1. Information transfer between path and vertex degree counts", *Journal of Chemical Information and Modeling* **1993**, *33*, 143–147.

[127] L. H. Hall, L. B. Kier, J. W. Frazer, "Design of molecules from quantitative structure-activity relationship models. 2. Derivation and proof of information transfer relating equations", *Journal of Chemical Information and Modeling* **1993**, *33*, 148–152.

[128] P. Schneider, G. Schneider, "De Novo Design at the Edge of Chaos", *Journal of Medicinal Chemistry* **2016**, *59*, 4077–4086.

[129] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2009**.

[130] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH Verlag GmbH, Weinheim, Germany, **2000**.

[131] F. Grisoni, D. Ballabio, R. Todeschini, V. Consonni in *Comprehensive Toxicology. Methods in Molecular Biology*, *Vol. 1800*, (Ed.: O. Nicolotti), Humana Press, New York, NY, **2018**, pp. 3–53.

[132] F. Grisoni, V. Consonni, R. Todeschini in *Computational Chemogenomics. Methods in Molecular Biology*, *Vol. 1825*, (Ed.: J. Brown), Humana Press, New York, NY, **2018**, pp. 171–209.

[133] A. Nicholls et al., "Molecular Shape and Medicinal Chemistry: A Perspective", *Journal of Medicinal Chemistry* **2010**, *53*, 3862–3886.

[134] D. Jung, J. Floyd, T. M. Gund, "A comparative molecular field analysis (CoMFA) study using semiempirical, density functional,ab initio methods and pharmacophore derivation using DISCOtech on sigma 1 ligands", *Journal of Computational Chemistry* **2004**, *25*, 1385–1399.

[135] I. Paster, M. Shacham, N. Brauner, "Investigation of the Relationships between Molecular Structure, Molecular Descriptors, and Physical Properties", *Industrial & Engineering Chemistry Research* **2009**, *48*, 9723–9734.

[136] M. Hechinger, K. Leonhard, W. Marquardt, "What is Wrong with Quantitative Structure–Property Relations Models Based on Three-Dimensional Descriptors?", *Journal of Chemical Information and Modeling* **2012**, *52*, 1984–1993.

[137] R. D. Cramer, D. E. Patterson, J. D. Bunce, "Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins", *Journal of the American Chemical Society* **1988**, *110*, 5959–5967.

[138] R. R. S. Pissurlenkar, V. M. Khedkar, R. P. Iyer, E. C. Coutinho, "Ensemble QSAR: A QSAR method based on conformational ensembles and metric descriptors", *Journal of Computational Chemistry* **2011**, *32*, 2204–2218.

[139] M. C. Nicklaus, S. Wang, J. S. Driscoll, G. W. Milne, "Conformational changes of small molecules binding to proteins", *Bioorganic and Medicinal Chemistry* **1995**, *3*, 411–428.

[140] J. H. Nettles, J. L. Jenkins, A. Bender, Z. Deng, J. W. Davies, M. Glick, "Bridging Chemical and Biological Space: "Target Fishing" Using 2D and 3D Molecular Descriptors", *Journal of Medicinal Chemistry* **2006**, *49*, 6802–6810.

[141] J. Bajorath in *Bioinformatics. Methods in Molecular Biology*, *Vol. 1526*, (Ed.: J. Keith), Humana Press, New York, NY, **2017**, pp. 231–245.

[142] MDL Information Systems Inc, "MACCS-II", San Leandro, CA, USA, **1987**.

[143] D. Rogers, M. Hahn, "Extended-connectivity fingerprints", *Journal of Chemical Information and Modeling* **2010**, *50*, 742–754.

[144] H. L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service." *Journal of Chemical Documentation* **1965**, *5*, 107–113.

[145] S. D. Axen, X.-P. Huang, E. L. Cáceres, L. Gendelev, B. L. Roth, M. J. Keiser, "A Simple Representation of Three-Dimensional Molecular Structure", *Journal of Medicinal Chemistry* **2017**, *60*, 7393–7409.

[146] C. G. Wermuth, C. R. Ganellin, P. Lindberg, L. A. Mitscher, "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)", *Pure and Applied Chemistry* **1998**, *70*, 1129–1143.

[147] M. Reutlinger et al., "Chemically advanced template search (CATS) for scaffold-hopping and prospective target prediction for 'orphan' molecules", *Molecular Informatics* **2013**, *32*, 133–138.

[148] A. Klenner, M. Hartenfeller, P. Schneider, G. Schneider, "'Fuzziness' in pharmacophore-based virtual screening and de novo design", *Drug Discovery Today: Technologies* **2010**, *7*, e237–e244.

[149] M. Sastry, J. F. Lowrie, S. L. Dixon, W. Sherman, "Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments", *Journal of Chemical Information and Modeling* **2010**, *50*, 771–784.

[150] J. H. Schuur, P. Selzer, J. Gasteiger, "The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity", *Journal of Chemical Information and Computer Sciences* **1996**, *36*, 334–344.

[151] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, "Chemical information in 3D space", *Journal of Chemical Information and Computer Sciences* **1996**, *36*, 1030–1037.

[152] T. S. Rush, J. A. Grant, L. Mosyak, A. Nicholls, "A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction", *Journal of Medicinal Chemistry* **2005**, *48*, 1489–1495.

[153] P. C. Hawkins, A. G. Skillman, A. Nicholls, "Comparison of shape-matching and docking as virtual screening tools", *Journal of Medicinal Chemistry* **2007**, *50*, 74–82.

[154] D. R. Buckle et al., "Glossary of terms used in medicinal chemistry. Part II (IUPAC Recommendations 2013)", *Pure and Applied Chemistry* **2013**, *85*, 1725–1758.

[155] A. Bender, R. C. Glen, "Molecular similarity: a key technique in molecular informatics", *Organic & Biomolecular Chemistry* **2004**, *2*, 3204.

[156] G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, "Molecular similarity in medicinal chemistry", *Journal of Medicinal Chemistry* **2014**, *57*, 3186–3204.

[157] G. M. Maggiora, M. A. Johnson in *Concepts and Applications of Molecular Similarity*, (Eds.: M. A. Johnson, G. M. Maggiora), John Wiley & Sons, Inc., **1990**, pp. 1–13.

[158] A. R. Leach, V. J. Gillet, R. A. Lewis, R. Taylor, "Three-Dimensional Pharmacophore Methods in Drug Discovery", *Journal of Medicinal Chemistry* **2010**, *53*, 539–558.

[159] G. M. Maggiora, V. Shanmugasundaram in *Chemoinformatics. Methods in Molecular Biology. Vol. 275*, (Ed.: J. Bajorath), 2, Humana Press, New York, NY, **2004**, pp. 1–50.

[160] T. T. Tanimoto, *An Elementary Mathematical Theory of Classification and Prediction*, International Business Machines Corporation, **1958**.

[161] P. Willett, J. M. Barnard, G. M. Downs, "Chemical Similarity Searching", *Journal of Chemical Information and Computer Sciences* **1998**, *38*, 983–996.

[162] W. Walters, M. T. Stahl, M. A. Murcko, "Virtual screening—an overview", *Drug Discovery Today* **1998**, *3*, 160–178.

[163] L. C. Blum, J.-L. Reymond, "970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13", *Journal of the American Chemical Society* **2009**, *131*, 8732–8733.

[164] J.-L. Reymond, "The Chemical Space Project", *Accounts of Chemical Research* **2015**, *48*, 722–730.

[165] G. Klebe, "Virtual ligand screening: strategies, perspectives and limitations", *Drug Discovery Today* **2006**, *11*, 580–594.

[166] G. Schneider, "Virtual screening: an endless staircase?", *Nature Reviews Drug Discovery* **2010**, *9*, 273–276.

[167] J. M. Rollinger, H. Stuppner, T. Langer in *Natural Compounds as Drugs Volume I. Progess in Drug Research, Vol. 65*, (Eds.: F. Petersen, R. Amstutz), Birkhäuser Basel, Basel, **2008**, pp. 211–249.

[168] J. M. Rollinger, A. Hornick, T. Langer, H. Stuppner, H. Prast, "Acetylcholinesterase Inhibitory Activity of Scopolin and Scopoletin Discovered by Virtual Screening of Natural Products", *Journal of Medicinal Chemistry* **2004**, *47*, 6248–6254.

[169] L.-J. Liu, K.-H. Leung, D. S. Chan, Y.-T. Wang, D.-L. Ma, C.-H. Leung, "Identification of a natural product-like STAT3 dimerization inhibitor by structure-based virtual screening", *Cell Death & Disease* **2014**, *5*, e1293–e1293.

[170] H. Eckert, J. Bajorath, "Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches", *Drug Discovery Today* **2007**, *12*, 225–233.

[171] Y. C. Martin, J. L. Kofron, L. M. Traphagen, "Do Structurally Similar Molecules Have Similar Biological Activity?", *Journal of Medicinal Chemistry* **2002**, *45*, 4350–4358.

[172] J. Boström, A. Hogner, S. Schmitt, "Do Structurally Similar Ligands Bind in a Similar Fashion?", *Journal of Medicinal Chemistry* **2006**, *49*, 6716–6725.

[173] P. Willett, "Similarity-based virtual screening using 2D fingerprints", *Drug Discovery Today* **2006**, *11*, 1046–1053.

[174] E. Karhu, J. Isojärvi, P. Vuorela, L. Hanski, A. Fallarero, "Identification of Privileged Antichlamydial Natural Products by a Ligand-Based Strategy", *Journal of Natural Products* **2017**, *80*, 2602–2608.

[175] F. Grisoni et al., "Matrix-based Molecular Descriptors for Prospective Virtual Compound Screening", *Molecular Informatics* **2017**, *36*, 1600091.

[176] G. Schneider, "De novo design – hop(p)ing against hope", *Drug Discovery Today: Technologies* **2013**, *10*, e453–e460.

[177] D. J. Danziger, P. M. Dean, "Automated Site-Directed Drug Design: A General Algorithm for Knowledge Acquisition about Hydrogen-Bonding Regions at Protein Surfaces", *Proceedings of the Royal Society of London. Series B Biological Sciences* **1989**, *236*, 101–113.

[178] R. A. Lewis, P. M. Dean, "Automated Site-Directed Drug Design: The Formation of Molecular Templates in Primary Structure Generation", *Proceedings of the Royal Society B: Biological Sciences* **1989**, *236*, 141–162.

[179] V. Gillet, A. Johnson, P. Mata, S. Sike, "Automated structure design in 3D", *Tetrahedron Computer Methodology* **1990**, *3*, 681–696.

[180] H.-J. Böhm, "A novel computational tool for automated structure-based drug design", *Journal of Molecular Recognition* **1993**, *6*, 131–137.

[181] R. E. Babine et al., "Design, synthesis and X-ray crystallographic studies of novel FKBP-12 ligands", *Bioorganic & Medicinal Chemistry Letters* **1995**, *5*, 1719–1724.

[182] S. Kandil et al., "Discovery of a novel HCV helicase inhibitor by a de novo drug design approach", *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 2935–2937.

[183] R. Wang, Y. Gao, L. Lai, "LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design", *Journal of Molecular Modeling* **2000**, *6*, 498–516.

[184] G. Schneider, M. L. Lee, M. Stahl, P. Schneider, "De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks." *Journal of Computer-Aided Molecular Design* **2000**, *14*, 487–94.

[185] G. Schneider et al., "Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design", *Angewandte Chemie* **2000**, *39*, 4130–4133.

[186] H. M. Vinkers et al., "SYNOPSIS: SYNthesize and OPtimize System in Silico", *Journal of Medicinal Chemistry* **2003**, *46*, 2765–2773.

[187] M. Hartenfeller et al., "Dogs: Reaction-driven de novo design of bioactive compounds", *PLoS Computational Biology* **2012**, *8*, (Ed.: J. M. Briggs), e1002380.

[188] M. Rupp, E. Proschak, G. Schneider, "Kernel Approach to Molecular Similarity Based on Iterative Graph Similarity", *Journal of Chemical Information and Modeling* **2007**, *47*, 2280–2286.

[189] M. Rupp, G. Schneider, "Graph Kernels for Molecular Similarity", *Molecular Informatics* **2010**, *29*, 266–273.

[190] G. Schneider, P. Schneider in *ACS Symposium Series*, **2016**, Chapter 8, pp. 143–158.

[191] B. Spänkuch et al., "Drugs by numbers: Reaction-driven de novo design of potent and selective anticancer leads", *Angewandte Chemie International Edition* **2013**, *52*, 4676–4681.

[192] S. Keppner, E. Proschak, G. Schneider, B. Spänkuch, "Identification and Validation of a Potent Type II Inhibitor of Inactive Polo-like Kinase 1", *ChemMedChem* **2009**, *4*, 1806–1809.

[193] T. Rodrigues et al., "De Novo Fragment Design for Drug Discovery and Chemical Biology", *Angewandte Chemie International Edition* **2015**, *54*, 15079–15083.

[194] T. Rodrigues et al., "Multidimensional De Novo Design Reveals 5-HT 2B Receptor-Selective Ligands", *Angewandte Chemie International Edition* **2015**, *54*, 1551–1555.

[195] A. M. Perna et al., "Fragment-Based De Novo Design Reveals a Small-Molecule Inhibitor of Helicobacter Pylori HtrA", *Angewandte Chemie International Edition* **2015**, *54*, 10244–10248.

[196] L. Friedrich, T. Rodrigues, C. S. Neuhaus, P. Schneider, G. Schneider, "From Complex Natural Products to Simple Synthetic Mimetics by Computational de Novo Design", *Angewandte Chemie International Edition* **2016**, *55*, 6789–6792.

[197] C. Sourbier et al., "Englerin A Stimulates PKC$\theta$ to Inhibit Insulin Signaling and to Simultaneously Activate HSF1: Pharmacologically Induced Synthetic Lethality", *Cancer Cell* **2013**, *23*, 228–237.

[198] Y. Akbulut et al., "Englerin a is a potent and selective activator of TRPC4 and TRPC5 calcium channels", *Angewandte Chemie International Edition* **2015**, *54*, 3787–3791.

[199] C. Carson et al., "Englerin A Agonizes the TRPC4/C5 Cation Channels to Inhibit Tumor Cell Line Proliferation", *PLoS ONE* **2015**, *10*, e0127498.

[200] Z. Wu, S. Zhao, D. M. Fash, Z. Li, W. J. Chain, J. A. Beutler, "Englerins: A Comprehensive Review", *Journal of Natural Products* **2017**, *80*, 771–781.

[201] D. Merk, F. Grisoni, L. Friedrich, E. Gelzinyte, G. Schneider, "Computer-Assisted Discovery of Retinoid X Receptor Modulating Natural Products and Isofunctional Mimetics", *Journal of Medicinal Chemistry* **2018**, *61*, 5442–5447.

[202] G. Schneider, "Automating drug discovery", *Nature Reviews Drug Discovery* **2017**, *17*, 97–113.

[203] R. Byrne, G. Schneider in *Systems Chemical Biology. Methods in Molecular Biology*, (Eds.: S. Ziegler, H. Waldmann), Methods in Molecular Biology, Humana Press, New York, NY, **2019**, pp. 273–309.

[204] J. L. Jenkins, A. Bender, J. W. Davies, "In silico target fishing: Predicting biological targets from chemical structure", *Drug Discovery Today: Technologies* **2006**, *3*, 413–421.

[205] A. Koutsoukas et al., "From in silico target prediction to multi-target drug design: Current databases, methods and applications", *Journal of Proteomics* **2011**, *74*, 2554–2574.

[206] A. Peón, C. C. Dang, P. J. Ballester, "How Reliable Are Ligand-Centric Methods for Target Fishing?", *Frontiers in Chemistry* **2016**, *4*, 1–10.

[207] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, B. K. Shoichet, "Relating protein pharmacology by ligand chemistry", *Nature Biotechnology* **2007**, *25*, 197–206.

[208] M. Davies et al., "ChEMBL web services: Streamlining access to drug discovery data and utilities", *Nucleic Acids Research* **2015**, *43*, W612–W620.

[209] C. Laggner et al., "Chemical informatics and target identification in a zebrafish phenotypic screen", *Nature Chemical Biology* **2012**, *8*, 144–146.

[210] D. Gfeller, A. Grosdidier, M. Wirth, A. Daina, O. Michielin, V. Zoete, "SwissTargetPrediction: a web server for target prediction of bioactive small molecules", *Nucleic Acids Research* **2014**, *42*, W32–W38.

[211] D. Reker, T. Rodrigues, P. Schneider, G. Schneider, "Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus", *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 4067–4072.

[212] P. Schneider, G. Schneider, "Collection of Bioactive Reference Compounds for Focused Library Design", *QSAR and Combinatorial Science* **2003**, *22*, 713–718.

[213] T. Kohonen, "Self-organized formation of topologically correct feature maps", *Biological Cybernetics* **1982**, *43*, 59–69.

[214] D. Reker et al., "Revealing the macromolecular targets of complex natural products", *Nature Chemistry* **2014**, *6*, 1072–1078.

[215] T. Rodrigues, D. Reker, J. Kunze, P. Schneider, G. Schneider, "Revealing the Macromolecular Targets of Fragment-Like Natural Products", *Angewandte Chemie International Edition* **2015**, *54*, 10516–10520.

[216] G. Schneider, D. Reker, T. Chen, K. Hauenstein, P. Schneider, K.-H. Altmann, "Deorphaning the Macromolecular Targets of the Natural Anticancer Compound Doliculide", *Angewandte Chemie International Edition* **2016**, *55*, 12408–12411.

[217] A. M. Jordan, "Artificial Intelligence in Drug Design—The Storm Before the Calm?", *ACS Medicinal Chemistry Letters* **2018**, *9*, 1150–1152.

[218] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, "The rise of deep learning in drug discovery", *Drug Discovery Today* **2018**, *23*, 1241–1250.

[219] J. Zou, Y. Han, S.-S. So in *Artificial Neural Networks. Methods in Molecular Biology*, (Ed.: D. J. Livingstone), Humana Press, New York, NY, **2008**, pp. 14–22.

[220] E. Gawehn, J. A. Hiss, G. Schneider, "Deep Learning in Drug Discovery", *Molecular Informatics* **2016**, *35*, 3–14.

[221] S. Russell, P. Norvig in *Artificial Intelligence - A Modern Approach*, Pearson Education, Inc., Upper Saddle River, New Jersey, **2003**, pp. 1–58.

[222] C. Bishop, *Pattern recognition and machine learning*, Springer, **2006**.

[223] A. Krogh, "What are artificial neural networks?", *Nature Biotechnology* **2008**, *26*, 195–197.

[224] D. Shiffman, *The Nature of Code: Simulating Natural Systems with Processing*, 1st Editio, (Ed.: S. Fry), **2012**.

[225] W. S. McCulloch, W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics* **1943**, *5*, 115–133.

[226] A. L. Hodgkin, A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve." *The Journal of Physiology* **1952**, *117*, 500–44.

[227] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review* **1958**, *65*, 386–408.

[228] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature* **2015**, *521*, 436–444.

[229] M. Minsky, S. Papert, *Perceptrons. An Introduction to Computational Geometry*, M.I.T. Press, Oxford, England, **1969**.

[230] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of Control Signals and Systems* **1989**, *2*, 303–314.

[231] J. Gasteiger, J. Zupan, "Neural Networks in Chemistry", *Angewandte Chemie International Edition* **1993**, *32*, 503–527.

[232] G. Schneider, P. Wrede, "Artificial neural networks for computer-based molecular design", *Progress in Biophysics and Molecular Biology* **1998**, *70*, 175–222.

[233] G. Schneider, "Neural networks are useful tools for drug design", *Neural Networks* **2000**, *13*, 15–16.

[234] A. L. Blum, P. Langley, "Selection of relevant features and examples in machine learning", *Artificial Intelligence* **1997**, *97*, 245–271.

[235] I. V. Tetko et al., "HIV-1 Reverse Transcriptase Inhibitor Design Using Artificial Neural Networks", *Journal of Medicinal Chemistry* **1994**, *37*, 2520–2526.

[236] L. Zhang, J. Tan, D. Han, H. Zhu, "From machine learning to deep learning: progress in machine intelligence for rational drug discovery", *Drug Discovery Today* **2017**, *22*, 1680–1685.

[237] J. Panteleev, H. Gao, L. Jia, "Recent applications of machine learning in medicinal chemistry", *Bioorganic & Medicinal Chemistry Letters* **2018**, *28*, 2807–2815.

[238] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik, "Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships", *Journal of Chemical Information and Modeling* **2015**, *55*, 263–274.

[239] M. H. Segler, T. Kogej, C. Tyrchan, M. P. Waller, "Generating focused molecule libraries for drug discovery with recurrent neural networks", *ACS Central Science* **2018**, *4*, 120–131.

[240] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules", *Journal of Chemical Information and Modeling* **1988**, *28*, 31–36.

[241] P. Goyal, S. Pandey, K. Jain in *Deep Learning for Natural Language Processing*, Apress, Berkeley, CA, **2018**, pp. 1–74.

[242] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory", *Neural Computation* **1997**, *9*, 1735–1780.

[243] R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, Y. Wu, "Exploring the Limits of Language Modeling", *arXiv* **2016**, DOI 10.1109/NLPKE.2008.4906752.

[244] A. Gupta, A. T. Müller, B. J. Huisman, J. A. Fuchs, P. Schneider, G. Schneider, "Generative Recurrent Networks for De Novo Drug Design", *Molecular Informatics* **2018**, *37*, 1700111.

[245] M. E. Taylor, P. Stone, "Transfer Learning for Reinforcement Learning Domains: A Survey", *J. Mach. Learn. Res.* **2009**, *10*, 1633–1685.

[246] D. Merk, L. Friedrich, F. Grisoni, G. Schneider, "De Novo Design of Bioactive Small Molecules by Artificial Intelligence", *Molecular Informatics* **2018**, *37*, 1700153.

[247] D. Merk, F. Grisoni, L. Friedrich, G. Schneider, "Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators", *Communications Chemistry* **2018**, *1*, 68.

[248] M. R. Berthold et al. in Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007), Springer, **2007**.

[249] G. van Rossum, "The Python Language Reference", **2016**.

[250] W. McKinney in Proceedings of the 9th Python in Science Conference, (Eds.: S. van der Walt, J. Millman), **2010**, pp. 51–56.

[251] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", *Computing in Science & Engineering* **2007**, *9*, 90–95.

[252] M. Waskom et al., "mwaskom/seaborn", **2017**.

[253] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

[254] T. Kluyver, B. Ragan-Kelley, Jupyter Development Team and others in Positioning and Power in Academic Publishing: Players, Agents and Agendas, (Eds.: F. Loizides, B. Schmidt), IOS Press, **2016**, pp. 87–90.

[255] T. Sander, J. Freyss, M. von Korff, C. Rufener, "DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis", *Journal of Chemical Information and Modeling* **2015**, *55*, 460–473.

[256] P. Ertl, A. Schuffenhauer, "Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions", *Journal of Cheminformatics* **2009**, *1*, 8.

[257] P. Ertl, S. Roggo, A. Schuffenhauer, "Natural product-likeness score and its application for prioritization of compound libraries", *Journal of Chemical Information and Modeling* **2008**, *48*, 68–74.

[258] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research* **2014**, *15*, 1929–1958.

[259] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", **2015**.

[260] F. Chollet et al., "Keras", https://keras.io, **2015**.

[261] "The RDKit: Open-Source Cheminformatics Software", **2018**.

[262] Y.-C. Cheng, W. H. Prusoff, "Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction", *Biochemical Pharmacology* **1973**, *22*, 3099–3108.

[263] A. Koeberle et al., "SAR studies on curcumin's pro-inflammatory targets: Discovery of prenylated pyrazolocurcuminoids as potent and selective novel inhibitors of 5-lipoxygenase", *Journal of Medicinal Chemistry* **2014**, *57*, 5638–5648.

[264] C. Pergola et al., "ERK-mediated regulation of leukotriene biosynthesis by androgens: A molecular basis for gender differences in inflammation and asthma", *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 19881–19886.

[265] A. M. Schaible et al., "High capacity for leukotriene biosynthesis in peripheral blood during pregnancy", *Prostaglandins Leukotrienes and Essential Fatty Acids* **2013**, *89*, 245–255.

[266] J. Schmidt et al., "A Dual Modulator of Farnesoid X Receptor and Soluble Epoxide Hydrolase To Counter Nonalcoholic Steatohepatitis", *Journal of Medicinal Chemistry* **2017**, *60*, 7703–7724.

[267] P. Heitel, J. Achenbach, D. Moser, E. Proschak, D. Merk, "DrugBank screening revealed alitretinoin and bexarotene as liver X receptor modulators", *Bioorganic & Medicinal Chemistry Letters* **2017**, *27*, 1193–1198.

[268] D. Flesch et al., "Nonacidic Farnesoid X Receptor Modulators", *Journal of Medicinal Chemistry* **2017**, *60*, 7199–7205.

[269] P. Williams, A. Sorribas, M. J. R. Howes, "Natural products as a source of Alzheimer's drug leads", *Natural Product Reports* **2011**, *28*, 48–77.

[270] P. K. Mukherjee, V. Kumar, M. Mal, P. J. Houghton, "Acetylcholinesterase inhibitors from plants", *Phytomedicine* **2007**, *14*, 289–300.

[271] M. Heinrich, H. L. Teoh, "Galanthamine from snowdrop - The development of a modern drug against Alzheimer's disease from local Caucasian knowledge", *Journal of Ethnopharmacology* **2004**, *92*, 147–162.

[272] A. M. Takos, F. Rook, "Towards a molecular understanding of the biosynthesis of Amaryllidaceae alkaloids in support of their Expanding medical use", *International Journal of Molecular Sciences* **2013**, *14*, 11713–11741.

[273] M. F. Santillo, Y. Liu, "A fluorescence assay for measuring acetylcholinesterase activity in rat blood and a human neuroblastoma cell line (SH-SY5Y)", *Journal of Pharmacological and Toxicological Methods* **2015**, *76*, 15–22.

[274] K. Takata et al., "Galantamine-induced amyloid-$\beta$ clearance mediated via stimulation of microglial nicotinic acetylcholine receptors", *Journal of Biological Chemistry* **2010**, *285*, 40180–40191.

[275] A. Fatima et al., "From Nature to Market: Examples of Natural Products that Became Drugs", *Recent Patents on Biotechnology* **2014**, *8*, 76–88.

[276] C. Guillou, J. L. Beunard, E. Gras, C. Thal, "An efficient total synthesis of (±)-galanthamine", *Angewandte Chemie International Edition* **2001**, *40*, 4745–4746.

[277] L. Li, Q. Yang, Y. Wang, Y. Jia, "Catalytic Asymmetric Total Synthesis of (-)-Galanthamine and (-)-Lycoramine", *Angewandte Chemie International Edition* **2015**, *54*, 6255–6259.

[278] G. L. Ellman, K. Courtney, V. Andres, R. M. Featherstone, "A new and rapid colorimetric determination of acetylcholinesterase activity", *Biochemical Pharmacology* **1961**, *7*, 88–95.

[279] C. G. V. Sharples et al., "UB-165: A Novel Nicotinic Agonist with Subtype Selectivity Implicates the $\alpha 4\beta 2$* Subtype in the Modulation of Dopamine Release from Rat Striatal Synaptosomes", *The Journal of Neuroscience* **2000**, *20*, 2783–2791.

[280] R. J. Lukas, "Characterization of Curaremimetic Neurotoxin Binding Sites on Cellular Membrane Fragments Derived from the Rat Pheochromocytoma PC 12", *Journal of Neurochemistry* **1986**, *47*, 1768–1773.

[281] C. Sur et al., "N-desmethylclozapine, an allosteric agonist at muscarinic 1 receptor, potentiates N-methyl-D-aspartate receptor activity", *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 13674–13679.

[282] R. H. P. Porter et al., "Functional characterization of agonists at recombinant human 5-HT 2A , 5-HT 2B and 5-HT 2C receptors in CHO-K1 cells", *British Journal of Pharmacology* **1999**, *128*, 13–20.

[283] M. F. Brizzi, M. G. Aronica, A. Rosso, G. P. Bagnara, Y. Yarden, L. Pegoraro, "Granulocyte-Macrophage Colony-stimulating Factor Stimulates JAK2 Signaling Pathway and Rapidly Activates p93, STAT1 p91, and STAT3 p92 in Polymorphonuclear Leukocytes", *Journal of Biological Chemistry* **1996**, *271*, 3562–3567.

[284] Y.-J. Zhou et al., "Distinct tyrosine phosphorylation sites in JAK3 kinase domain positively and negatively regulate its enzymatic activity", *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 13850–13855.

[285] H. C. Cheng, H. Nishio, O. Hatase, S. Ralph, J. H. Wang, "A synthetic peptide derived from p34cdc2 is a specific and efficient substrate of src-family tyrosine kinases." *Journal of Biological Chemistry* **1992**, *267*, 9248–9256.

[286] J. D. Obourn, N. J. Koszewski, A. C. Notides, "Hormone- and DNA-binding mechanism of the recombinant human estrogen receptor", *Biochemistry* **1993**, *32*, 6229–6236.

[287] B. Cornelissen et al., "In vitro and in vivo evaluation of [123I]-VEGF165 as a potential tumor marker", *Nuclear Medicine and Biology* **2005**, *32*, 431–436.

[288] M. Simon et al., "Receptors of vascular endothelial growth factor/vascular permeability factor (VEGF/VPF) in fetal and adult human kidney: localization and [125I]VEGF binding sites." *Journal of the American Society of Nephrology : JASN* **1998**, *9*, 1032–44.

[289] P. Vanachayangkul, W. H. Tolleson, "Inhibition of Heme Peroxidases by Melamine", *Enzyme Research* **2012**, *2012*, 1–7.

[290] R. A. Pufahl et al., "Development of a fluorescence-based enzyme assay of human 5-lipoxygenase", *Analytical Biochemistry* **2007**, *364*, 204–212.

[291] D. H. Maurice, H. Ke, F. Ahmad, Y. Wang, J. Chung, V. C. Manganiello, "Advances in targeting cyclic nucleotide phosphodiesterases", *Nature Reviews Drug Discovery* **2014**, *13*, 290–314.

[292] B. R. Pinho, F. Ferreres, P. Valentão, P. B. Andrade, "Nature as a source of metabolites with cholinesterase-inhibitory activity: An approach to Alzheimer's disease treatment", *Journal of Pharmacy and Pharmacology* **2013**, *65*, 1681–1700.

[293] M. E. Welsch, S. A. Snyder, B. R. Stockwell, "Privileged scaffolds for library design and drug discovery", *Current Opinion in Chemical Biology* **2010**, *14*, 347–361.

[294] C. W. Coley, L. Rogers, W. H. Green, K. F. Jensen, "SCScore: Synthetic Complexity Learned from a Reaction Corpus", *Journal of Chemical Information and Modeling* **2018**, *58*, 252–261.

[295] M. Katkevičs, A. Kontijevskis, I. Mutule, E. Suna, "Microwave-promoted automated synthesis of a coumarin library", *Chemistry of Heterocyclic Compounds* **2007**, *43*, 151–159.

[296] M. Samochocki et al., "Galantamine is an allosterically potentiating ligand of the human alpha4/beta2 nAChR", *Acta Neurologica Scandinavica* **2000**, *102*, 68–73.

[297] H. M. Revankar, S. N. A. Bukhari, G. B. Kumar, H.-L. Qin, "Coumarins scaffolds as COX inhibitors", *Bioorganic Chemistry* **2017**, *71*, 146–159.

[298] P. Schneider, M. Röthlisberger, D. Reker, G. Schneider, "Spotting and designing promiscuous ligands for drug discovery", *Chemical Communications* **2016**, *52*, 1135–1138.

[299] A. L. Hopkins, C. R. Groom, A. Alex, "Ligand efficiency: a useful metric for lead selection", *Drug Discovery Today* **2004**, *9*, 430–431.

[300] M. Congreve, R. Carr, C. Murray, H. Jhoti, "A 'Rule of Three' for fragment-based lead discovery?", *Drug Discovery Today* **2003**, *8*, 876–877.

[301] S. Kim et al., "PubChem Substance and Compound databases", *Nucleic Acids Research* **2016**, *44*, D1202–D1213.

[302] E. Patridge, P. Gareiss, M. S. Kinch, D. Hoyer, "An analysis of FDA-approved drugs: Natural products and their derivatives", *Drug Discovery Today* **2016**, *21*, 204–207.

[303] C. Jiménez, "Marine Natural Products in Medicinal Chemistry", *ACS Medicinal Chemistry Letters* **2018**, *9*, 959–961.

[304] Y. M. Vaske, P. Crews in *Bioactive Compounds from Marine Foods*, (Eds.: B. Hernández-Ledesma, M. Herrero), John Wiley & Sons Ltd, Chichester, UK, **2013**, Chapter 1, pp. 1–26.

[305] C. C. Hughes, A. Prieto-Davo, P. R. Jensen, W. Fenical, "The marinopyrroles, antibiotics of an unprecedented structure class from a marine Streptomyces sp." *Organic Letters* **2008**, *10*, 629–631.

[306] K. Doi et al., "Discovery of Marinopyrrole A (Maritoclax) as a Selective Mcl-1 Antagonist that Overcomes ABT-737 Resistance by Binding to and Targeting Mcl-1 for Proteasomal Degradation", *Journal of Biological Chemistry* **2012**, *287*, 10224–10235.

[307] J. M. Eichhorn, S. E. Alford, C. C. Hughes, W. Fenical, T. C. Chambers, "Purported Mcl-1 inhibitor marinopyrrole A fails to show selective cytotoxicity for Mcl-1-dependent cell lines", *Cell Death & Disease* **2013**, *4*, e880–e880.

[308] K. Doi et al., "Maritoclax induces apoptosis in acute myeloid leukemia cells with elevated Mcl-1 expression", *Cancer Biology & Therapy* **2014**, *15*, 1077–1086.

[309] P. Gomez-Bougie et al., "The selectivity of Marinopyrrole A to induce apoptosis in MCL1 high BCL2 low expressing myeloma cells is related to its ability to impair protein translation", *British Journal of Haematology* **2018**, *180*, 157–159.

[310] P. Schneider, G. Schneider, "De-orphaning the marine natural product (±)-marinopyrrole A by computational target prediction and biochemical validation", *Chemical Communications* **2017**, *53*, 2272–2274.

[311] M. D. Ungrin, L. M. Singh, R. Stocco, D. E. Sas, M. Abramovitz, "An Automated Aequorin Luminescence-Based Functional Calcium Assay for G-Protein-Coupled Receptors", *Analytical Biochemistry* **1999**, *272*, 34–42.

[312] R. J. Wilson et al., "Functional pharmacology of human prostanoid EP2 and EP4 receptors", *European Journal of Pharmacology* **2004**, *501*, 49–58.

[313] G. Asbóth, S. Phaneuf, G. N. Europe-Finner, M. Tóth, A. L. Bernal, "Prostaglandin E2 activates phospholipase C and elevates intracellular calcium in cultured myometrial cells: involvement of EP1 and EP3 receptor subtypes." *Endocrinology* **1996**, *137*, 2572–9.

[314] L. Meijer et al., "Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5." *European Journal of Biochemistry* **1997**, *243*, 527–36.

[315] D.-M. Kim, K. Yang, B.-S. Yang, "Biochemical characterizations reveal different properties between CDK4/cyclin D1 and CDK2/cyclin A." *Experimental & Molecular Medicine* **2003**, *35*, 421–30.

[316] B. Frantz et al., "The Activation State of p38 Mitogen-Activated Protein Kinase Determines the Efficiency of ATP Competition for Pyridinylimidazole Inhibitor Binding", *Biochemistry* **1998**, *37*, 13846–13853.

[317] A. J. Bardwell, M. Abdollahi, L. Bardwell, "Docking sites on mitogen-activated protein kinase (MAPK) kinases, MAPK phosphatases and the Elk-1 transcription factor compete for MAPK binding and are crucial for enzymic activity." *The Biochemical journal* **2003**, *370*, 1077–85.

[318] J. Lisnock et al., "Activation of JNK3 alpha 1 requires both MKK4 and MKK7: kinetic characterization of in vitro phosphorylated JNK3 alpha 1." *Biochemistry* **2000**, *39*, 3141–8.

[319] L. Meijer et al., "GSK-3-selective inhibitors derived from Tyrian purple indirubins." *Chemistry & biology* **2003**, *10*, 1255–66.

[320] Q. K. Huynh et al., "Characterization of the recombinant IKK1/IKK2 heterodimer. Mechanisms regulating kinase activity." *The Journal of Biological Chemistry* **2000**, *275*, 25883–91.

[321] S. F. Eddy et al., "Inducible I$\kappa$B Kinase/I$\kappa$B Kinase $\epsilon$ Expression Is Induced by CK2 and Promotes Aberrant Nuclear Factor-$\kappa$B Activation in Breast Cancer Cells", *Cancer Research* **2005**, *65*, 11375–11383.

[322] S. Li, A. Strelow, E. J. Fontana, H. Wesche, "IRAK-4: A novel member of the IRAK family with the properties of an IRAK-kinase", *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 5567–5572.

[323] A. Hausser et al., "Protein kinase C mu is negatively regulated by 14-3-3 signal transduction proteins." *The Journal of Biological Chemistry* **1999**, *274*, 9258–64.

[324] M. S. Turner, Fen-Fen-Lin, J. W. Trauger, J. Stephens, P. LoGrasso, "Characterization and purification of truncated human Rho-kinase II expressed in Sf-21 cells." *Archives of Biochemistry and Biophysics* **2002**, *405*, 13–20.

[325] C. C. Felder et al., "Comparison of the pharmacology and signal transduction of the human cannabinoid CB1 and CB2 receptors." *Molecular Pharmacology* **1995**, *48*, 443–50.

[326] R. Chen, K. A. Lewis, M. H. Perrin, W. W. Vale, "Expression cloning of a human corticotropin-releasing-factor receptor." *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 8967–71.

[327] S. V. Wu, M. Yang, D. Avedian, M. Birnbaumer, J. H. Walsh, "Single amino acid substitution of serine82 to asparagine in first intracellular loop of human cholecystokinin (CCK)-B receptor confers full cyclic AMP responses to CCK and gastrin." *Molecular pharmacology* **1999**, *55*, 795–803.

[328] A. F. Clark, D. Lane, K. Wilson, S. T. Miggans, M. D. McCartney, "Inhibition of dexamethasone-induced cytoskeletal changes in cultured human trabecular meshwork cells by tetrahydrocortisol." *Investigative Ophthalmology & Visual Science* **1996**, *37*, 805–13.

[329] D. Smart et al., "SB-334867-A: the first selective orexin-1 receptor antagonist", *British Journal of Pharmacology* **2001**, *132*, 1179–1182.

[330] S. Ammoun, "Distinct Recognition of OX1 and OX2 Receptors by Orexin Peptides", *Journal of Pharmacology and Experimental Therapeutics* **2003**, *305*, 507–514.

[331] B. R. Radziszewski, "Untersuchungen über Hydrobenzamid, Amarin und Lophin", *Berichte der deutschen chemischen Gesellschaft* **1877**, *10*, 70–75.

[332] H. Debus, "Ueber die Einwirkung des Ammoniaks auf Glyoxal", *Annalen der Chemie und Pharmacie* **1858**, *107*, 199–208.

[333] A. Zarghi, S. Arfaei, R. Ghodsi, "Design and synthesis of new 2,4,5-triarylimidazole derivatives as selective cyclooxygenase (COX-2) inhibitors", *Medicinal Chemistry Research* **2012**, *21*, 1803–1810.

[334] R. Chu, D. T. Terrano, T. C. Chambers, "Cdk1/cyclin B plays a key role in mitotic arrest-induced apoptosis by phosphorylation of Mcl-1, promoting its degradation and freeing Bak from sequestration", *Biochemical Pharmacology* **2012**, *83*, 199–206.

[335] A. De Biasio et al., "N-terminal Truncation of Antiapoptotic MCL1, but Not G 2 /M-induced Phosphorylation, Is Associated with Stabilization and Abundant Expression in Tumor Cells", *Journal of Biological Chemistry* **2007**, *282*, 23919–23936.

[336] U. Maurer, C. Charvet, A. S. Wagman, E. Dejardin, D. R. Green, "Glycogen Synthase Kinase-3 Regulates Mitochondrial Outer Membrane Permeabilization and Apoptosis by Destabilization of MCL-1", *Molecular Cell* **2006**, *21*, 749–760.

[337] P. Juin, O. Geneste, F. Gautier, S. Depil, M. Campone, "Decoding and unlocking the BCL-2 dependency of cancer cells", *Nature Reviews Cancer* **2013**, *13*, 455–465.

[338] Y. Hu, D. Stumpfe, J. Bajorath, "Computational Exploration of Molecular Scaffolds in Medicinal Chemistry", *Journal of Medicinal Chemistry* **2016**, *59*, 4062–4076.

[339] M. Weisel, E. Proschak, J. M. Kriegl, G. Schneider, "Form follows function: Shape analysis of protein cavities for receptor-based drug design", *Proteomics* **2009**, *9*, 451–459.

[340] S. Kortagere, M. D. Krasowski, S. Ekins, "The importance of discerning shape in molecular pharmacology", *Trends in Pharmacological Sciences* **2009**, *30*, 138–147.

[341] A. Kahraman, R. J. Morris, R. A. Laskowski, J. M. Thornton, "Shape Variation in Protein Binding Pockets and their Ligands", *Journal of Molecular Biology* **2007**, *368*, 283–301.

[342] K. Yeturu, N. Chandra, "PocketMatch: A new algorithm to compare binding sites in protein structures", *BMC Bioinformatics* **2008**, *9*, 543.

[343] J. A. Grant, M. A. Gallardo, B. T. Pickup, "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.

[344] M. J. Vainio, J. S. Puranen, M. S. Johnson, "ShaEP: Molecular overlay based on shape and electrostatic potential", *Journal of Chemical Information and Modeling* **2009**, *49*, 492–502.

[345] P. J. Ballester, W. G. Richards, "Ultrafast shape recognition to search compound databases for similar molecular shapes", *Journal of Computational Chemistry* **2007**, *28*, 1711–1723.

[346] A. M. Schreyer, T. Blundell, "USRCAT: real-time ultrafast shape recognition with pharmacophoric constraints", *Journal of Cheminformatics* **2012**, *4*, 27.

[347] N. Todoroff, J. Kunze, H. Schreuder, G. Hessler, K.-H. Baringhaus, G. Schneider, "Fractal Dimensions of Macromolecular Structures", *Molecular Informatics* **2014**, *33*, 588–596.

[348] A. Steck, B. Steck, "Reflections on brain and mind", *Schweizer Archiv fur Neurologie und Psychiatrie* **2016**, *167*, 155–163.

[349] P. Grassberger, I. Procaccia, "Characterization of strange attractors", *Physical Review Letters* **1983**, *50*, 346–349.

[350] P. Grassberger, I. Procaccia, "Measuring the strangeness of strange attractors", *Physica D: Nonlinear Phenomena* **1983**, *9*, 189–208.

[351] P. Tosco, N. Stiefl, G. Landrum, "Bringing the MMFF force field to the RDKit: Implementation and validation", *Journal of Cheminformatics* **2014**, *6*, 37.

[352] M. L. Connolly, "Analytical molecular surface calculation", *Journal of Applied Crystallography* **1983**, *16*, 548–558.

[353] S. Decherchi, W. Rocchia, "A general and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale", *PLoS ONE* **2013**, *8*, (Ed.: C. Lorenz), e59744.

[354] H. Theil in *Henri Theil's Contributions to Economics and Econometrics*, **1992**, pp. 345–381.

[355] P. K. Sen, "Estimates of the Regression Coefficient Based on Kendall's Tau", *Journal of the American Statistical Association* **1968**, *63*, 1379–1389.

[356] H.-J. Behrendt, T. Germann, C. Gillen, H. Hatt, R. Jostock, "Characterization of the mouse cold-menthol receptor TRPM8 and vanilloid receptor type-1 VR1 using a fluorometric imaging plate reader (FLIPR) assay", *British Journal of Pharmacology* **2004**, *141*, 737–745.

[357] R. Ratnayake, D. Covell, T. T. Ransom, K. R. Gustafson, J. A. Beutler, "Englerin a, a selective inhibitor of renal cancer cell growth, from Phyllanthus engleri", *Organic Letters* **2009**, *11*, 57–60.

[358] E. Jafari, M. R. Khajouei, F. Hassanzadeh, G. H. Hakimelahi, G. A. Khodarahmi, "Quinazolinone and quinazoline derivatives: Recent structures with potent antimicrobial and cytotoxic activities", *Research in Pharmaceutical Sciences* **2016**, *11*, 1–14.

[359] S. Cai et al., "Phosphatidylinositol 3-Kinase Inhibitors", **2015**, US2015/0361068A1.

[360] M. J. Pérez De Vega, I. Gómez-Monterrey, A. Ferrer-Montiel, R. González-Muñiz, "Transient Receptor Potential Melastatin 8 Channel (TRPM8) Modulation: Cool Entryway for Treating Pain and Cancer", *Journal of Medicinal Chemistry* **2016**, *59*, 10006–10029.

[361] Y. Karashima et al., "Bimodal Action of Menthol on the Transient Receptor Potential Channel TRPA1", *Journal of Neuroscience* **2007**, *27*, 9874–9884.

[362] E. J. Griffen, A. G. Dossetter, A. G. Leach, S. Montague, "Can we accelerate medicinal chemistry by augmenting the chemist with Big Data and artificial intelligence?", *Drug Discovery Today* **2018**, *23*, 1373–1384.

[363] G. Hessler, K.-H. Baringhaus, "Artificial Intelligence in Drug Design", *Molecules* **2018**, *23*, 2520.

[364] G. Schneider, K. Funatsu, Y. Okuno, D. Winkler, "De novo Drug Design - Ye olde Scoring Problem Revisited", *Molecular Informatics* **2017**, *36*, 1681031.

[365] F. Grisoni et al., "Scaffold hopping from natural products to synthetic mimetics by holistic molecular similarity", *Communications Chemistry* **2018**, *1*, 44.

[366] T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94", *Journal of Computational Chemistry* **1996**, *17*, 490–519.

[367] J. Gasteiger, M. Marsili, "Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges", *Tetrahedron* **1980**, *36*, 3219–3228.

[368] F. Grisoni, D. Merk, R. Byrne, G. Schneider, "Scaffold-Hopping from Synthetic Drugs by Holistic Molecular Representation", *Scientific Reports* **2018**, *8*, 16469.

[369] B. Chen, C. Mueller, P. Willett, "Combination Rules for Group Fusion in Similarity-Based Virtual Screening", *Molecular Informatics* **2010**, *29*, 533–541.

[370] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space", *The London Edinburgh and Dublin Philosophical Magazine and Journal of Science* **1901**, *2*, 559–572.

[371] E. Proschak, P. Heitel, L. Kalinowsky, D. Merk, "Opportunities and Challenges for Fatty Acid Mimetics in Drug Discovery", *Journal of Medicinal Chemistry* **2017**, *60*, 5235–5266.

[372] P. Germain et al., "International Union of Pharmacology. LXIII. Retinoid X Receptors", *Pharmacological Reviews* **2006**, *58*, 760–772.

[373] L. Michalik et al., "International Union of Pharmacology. LXI. Peroxisome Proliferator-Activated Receptors", *Pharmacological Reviews* **2006**, *58*, 726–741.

[374] G. Papadatos et al., "SureChEMBL: a large-scale, chemically annotated patent document database", *Nucleic Acids Research* **2016**, *44*, D1220–D1228.

[375] *Reaxys*, Elsevier.

[376] *SciFinder*, Chemical Abstract Service.

[377] G. Allenby et al., "Retinoic acid receptors and retinoid X receptors: interactions with endogenous retinoic acids." *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 30–34.

[378] A. A. Santilli, A. C. Scotese, R. M. Tomarelli, "A potent antihypercholesterolemic agent: [4-chloro-6-(2,3-xylidino)-2-pyrimidinylthio]acetic acid (Wy-14643)", *Experientia* **1974**, *30*, 1110–1111.

[379] D. Merk, F. Grisoni, L. Friedrich, E. Gelzinyte, G. Schneider, "Scaffold hopping from synthetic RXR modulators by virtual screening and de novo design", *MedChemComm* **2018**, *9*, 1289–1292.

[380] M. F. Boehm et al., "Synthesis and Structure-Activity Relationships of Novel Retinoid X Receptor-Selective Retinoids", *Journal of Medicinal Chemistry* **1994**, *37*, 2930–2941.

[381] P. J. Brown et al., "Identification of a subtype selective human PPAR$\alpha$ agonist through parallel-array synthesis", *Bioorganic & Medicinal Chemistry Letters* **2001**, *11*, 1225–1227.

[382] K.-i. Nakashima, T. Murakami, H. Tanabe, M. Inoue, "Identification of a naturally occurring retinoid X receptor agonist from Brazilian green propolis", *Biochimica et Biophysica Acta (BBA) - General Subjects* **2014**, *1840*, 3034–3041.

[383] H. Kotani, H. Tanabe, H. Mizukami, M. Makishima, M. Inoue, "Identification of a Naturally Occurring Rexinoid, Honokiol, That Activates the Retinoid X Receptor", *Journal of Natural Products* **2010**, *73*, 1332–1336.

[384] H. Zhang, L. Li, L. Chen, L. Hu, H. Jiang, X. Shen, "Structure Basis of Bigelovin as a Selective RXR Agonist with a Distinct Binding Mode", *Journal of Molecular Biology* **2011**, *407*, 13–20.

[385] *Dictionary of Natural Products*, Taylor & Francis Group and CRC Press, Boca Raton, FL, U.S.A., **2011**.

[386] R. E. Carhart, D. H. Smith, R. Venkataraghavan, "Atom pairs as molecular features in structure-activity studies: definition and applications", *Journal of Chemical Information and Modeling* **1985**, *25*, 64–73.

[387] D. G. Brown, J. Boström, "Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone?", *Journal of Medicinal Chemistry* **2016**, *59*, 4443–4458.

[388] A. M. Wassermann, H. Geppert, J. Bajorath, "Ligand Prediction for Orphan Targets Using Support Vector Machines and Various Target-Ligand Kernels Is Dominated by Nearest Neighbor Effects", *Journal of Chemical Information and Modeling* **2009**, *49*, 2155–2167.

[389] J. T. Lynch, R. Rajendran, G. Xenaki, I. Berrou, C. Demonacos, M. Krstic-Demonacos, "The role of glucocorticoid receptor phosphorylation in Mcl-1 and NOXA gene expression", *Molecular Cancer* **2010**, *9*, 38.

[390] P. A. Clemons et al., "Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles", *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 18787–18792.

[391] P. F. Newton, G. H. Whitham, "trans-cycloalkenes. Part 9. Optically active trans-cyclo-octenes via resolution of $\beta$-hydroxyphosphine oxides. Proof of configuration of bicyclic trans-cyclo-octenes constrained in chair and twist conformations", *J. Chem. Soc. Perkin Trans. 1* **1979**, 3072–3076.

[392] G. C. Kondaiah et al., "Asymmetric synthesis of (S,S,S)-2-Aza-bicyclo-[3.3.0]-octane-3-carboxylic acid benzyl ester: Formal synthesis of ramipril", *Synthetic Communications* **2011**, *41*, 1186–1191.

# A Supplementary Information

## A.1 Supplementary Information to Section 4.1

### Computational Results

**Table A.1:** Target predictions by SPiDER for galanthamine (**68**) and the mimetics **72-77**. Only targets with a *p*-value $\leq$0.05 (false positive rate, rounded up to two decimal places) were reported.

| Cmpd | Target | *p*-value |
|---|---|---|
| **68** | Endopeptidase | 0.05 |
| | Dopamine Receptor | 0.04 |
| | Sodium:Neurotransmitter Symporter | 0.03 |
| | Phosphodiesterase | 0.04 |
| | Muscarinic Acetylcholine Receptor | 0.05 |
| | Opioid Receptor | 0.04 |
| | Sodium Channel | 0.04 |
| | Acetylcholinesterase | 0.03 |
| | Serotonin Receptor | 0.03 |
| | Adrenergic Receptor | 0.03 |
| | Sigma Receptor | 0.03 |
| | Nicotinic Acetylcholine Receptor | 0.02 |
| **72** | Cystic Fibrosis Transmembrane Conductance Regulator | 0.02 |
| | DNA Topoisomerase | 0.04 |
| | Phosphodiesterase | 0.02 |
| | Arachidonate 5-Lipoxygenase | 0.03 |
| | Tyrosine Kinase | 0.04 |
| | Cyclooxygenase | 0.03 |
| | Adenosine Receptor | 0.04 |
| **73** | Cannabinoid Receptor | 0.01 |
| | Sodium:Neurotransmitter Symporter | 0.00 |
| | Phosphodiesterase | 0.01 |
| | Muscarinic Acetylcholine Receptor | 0.01 |
| | Aromatase | 0.01 |
| | Nicotinicoid LGIC GABA Receptor | 0.02 |
| | Phospholipase | 0.03 |

*Continued on next page*

Table A.1 – *Continued from previous page*

| Cmpd | Target | *p*-value |
|---|---|---|
| | Metabotropic Glutamate Receptor | 0.01 |
| | Arachidonate 5-Lipoxygenase | 0.02 |
| | Melatonin Receptor | 0.02 |
| | Aggregation Inhibitor | 0.01 |
| | Tyrosine Kinase | 0.04 |
| | Serotonin Receptor | 0.00 |
| | 11beta-Hydroxysteroid Dehydrogenase | 0.02 |
| | Adrenergic Receptor | 0.01 |
| | Steroid 5-Alpha-Reductase | 0.00 |
| | Cyclooxygenase | 0.01 |
| | Prostanoid Receptor | 0.03 |
| | Androgen Receptor | 0.00 |
| | Nicotinic Acetylcholine Receptor | 0.00 |
| | Histamin Receptor | 0.04 |
| **74** | DNA Topoisomerase | 0.05 |
| | Phosphodiesterase | 0.03 |
| | Phospholipase | 0.03 |
| | Tyrosine Kinase | 0.02 |
| | Endopeptidase | 0.04 |
| | Hydroxycarboxylic Acid Receptor | 0.02 |
| | RNA Polymerase | 0.04 |
| | DNA Polymerase | 0.03 |
| | Adenosine Receptor | 0.03 |
| **75** | Cannabinoid Receptor | 0.01 |
| | Sodium:Neurotransmitter Symporter | 0.01 |
| | Phosphodiesterase | 0.01 |
| | Muscarinic Acetylcholine Receptor | 0.00 |
| | Aromatase | 0.01 |
| | Nicotinicoid LGIC GABA Receptor | 0.02 |
| | Phospholipase | 0.03 |
| | Metabotropic Glutamate Receptor | 0.01 |
| | Arachidonate 5-Lipoxygenase | 0.02 |
| | Melatonin Receptor | 0.02 |
| | Aggregation Inhibitor | 0.01 |
| | Tyrosine Kinase | 0.04 |
| | Serotonin Receptor | 0.00 |
| | 11beta-Hydroxysteroid Dehydrogenase | 0.03 |

*Continued on next page*

Table A.1 – *Continued from previous page*

| Cmpd | Target | *p*-value |
|---|---|---|
| | Adrenergic Receptor | 0.02 |
| | Steroid 5-Alpha-Reductase | 0.00 |
| | Cyclooxygenase | 0.02 |
| | Prostanoid Receptor | 0.03 |
| | Androgen Receptor | 0.00 |
| | Nicotinic Acetylcholine Receptor | 0.00 |
| | Histamin Receptor | 0.04 |
| **76** | Cystic Fibrosis Transmembrane Conductance Regulator | 0.02 |
| | DNA Topoisomerase | 0.05 |
| | Phosphodiesterase | 0.03 |
| | B-Cell Lymphoma | 0.04 |
| | Arachidonate 5-Lipoxygenase | 0.05 |
| | Tyrosine Kinase | 0.04 |
| | Estrogen Receptor | 0.01 |
| | Hydroxycarboxylic Acid Receptor | 0.05 |
| | RNA Polymerase | 0.04 |
| **77** | Cannabinoid Receptor | 0.01 |
| | Sodium:Neurotransmitter Symporter | 0.01 |
| | Phosphodiesterase | 0.01 |
| | Muscarinic Acetylcholine Receptor | 0.01 |
| | Aromatase | 0.02 |
| | Nicotinicoid LGIC GABA Receptor | 0.02 |
| | Phospholipase | 0.03 |
| | Metabotropic Glutamate Receptor | 0.01 |
| | Arachidonate 5-Lipoxygenase | 0.02 |
| | Melatonin Receptor | 0.02 |
| | Aggregation Inhibitor | 0.01 |
| | Tyrosine Kinase | 0.04 |
| | Serotonin Receptor | 0.00 |
| | 11beta-Hydroxysteroid Dehydrogenase | 0.02 |
| | Adrenergic Receptor | 0.02 |
| | Steroid 5-Alpha-Reductase | 0.00 |
| | Cyclooxygenase | 0.02 |
| | Prostanoid Receptor | 0.03 |
| | Androgen Receptor | 0.00 |
| | Nicotinic Acetylcholine Receptor | 0.00 |
| | Histamin Receptor | 0.04 |

**Table A.2:** Predicted Targets by SPiDER software with a p-value $\leq 0.05$. Entries were sorted according to compound counts (large to small) per predicted target.

| Predicted Target | Compound(s) | Count |
|---|---|---|
| Phosphodiesterase | **68, 72, 73, 74, 75, 76, 77** | 7 |
| Tyrosine Kinase | **72, 73, 74, 75, 76, 77** | 6 |
| Arachidonate 5-Lipoxygenase | **72, 73, 75, 76, 77** | 5 |
| Sodium:Neurotransmitter Symporter | **68, 73, 75, 77** | 4 |
| Muscarinic Acetylcholine Receptor | **68, 73, 75, 77** | 4 |
| Serotonin Receptor | **68, 73, 75, 77** | 4 |
| Adrenergic Receptor | **68, 73, 75, 77** | 4 |
| Nicotinic Acetylcholine Receptor | **68, 73, 75, 77** | 4 |
| Cyclooxygenase | **72, 73, 75, 77** | 4 |
| Phospholipase | **73, 74, 75, 77** | 4 |
| DNA Topoisomerase | **72, 74, 76** | 3 |
| Cannabinoid Receptor | **73, 75, 77** | 3 |
| Aromatase | **73, 75, 77** | 3 |
| Nicotinicoid LGIC GABA Receptor | **73, 75, 77** | 3 |
| Metabotropic Glutamate Receptor | **73, 75, 77** | 3 |
| Melatonin Receptor | **73, 75, 77** | 3 |
| Aggregation Inhibitor | **73, 75, 77** | 3 |
| 11beta-Hydroxysteroid Dehydrogenase | **73, 75, 77** | 3 |
| Steroid 5-Alpha-Reductase | **73, 75, 77** | 3 |
| Prostanoid Receptor | **73, 75, 77** | 3 |
| Androgen Receptor | **73, 75, 77** | 3 |
| Histamin Receptor | **73, 75, 77** | 3 |
| Cystic Fibrosis Transmembrane Conductance Regulator | **72, 76** | 2 |
| Adenosine Receptor | **72, 74** | 2 |
| Hydroxycarboxylic Acid Receptor | **74, 76** | 2 |
| RNA Polymerase | **74, 76** | 2 |
| Endopeptidase | **68** | 1 |
| Dopamine Receptor | **68** | 1 |
| Opioid Receptor | **68** | 1 |
| Sodium Channel | **68** | 1 |
| Acetylcholinesterase | **68** | 1 |
| Sigma Receptor | **68** | 1 |
| Endopeptidase | **74** | 1 |
| DNA Polymerase | **74** | 1 |
| B-Cell Lymphoma | **76** | 1 |
| Estrogen Receptor | **76** | 1 |

**Table A.3:** Calculated properties of compounds **72-77**. (MW = Total Molecular Weight (g/mol), cLogP = calculated logarithmic partition coefficient, HBA = count of hydrogen bond acceptors, HBD = count of hydrogen bond donors, PSA = polar surface area (Å), RB = count of rotatable bonds, LE = ligand efficiency = 1.4(-log$IC_{50}$)/$N$, where $N$ is number of non-hydrogen atoms)

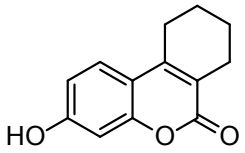
| Cmpd | MW | cLogP | HBA | HBD | PSA | RB | Assay | $IC_{50}$ [nM] | LE |
|---|---|---|---|---|---|---|---|---|---|
| **72** | 216.24 | 2.06 | 3 | 1 | 46.53 | 0 | PDE3A | 2600 | 0.48 |
| **72** | | | | | | | 5-$HT_{2b}$ (anta.) | 1600 | 0.50 |
| **73** | 230.26 | 2.34 | 3 | 0 | 35.53 | 1 | 5-$HT_{2b}$(anta.) | 2000 | 0.46 |
| **73** | | | | | | | AChE | 52000 | 0.35 |
| **74** | 232.23 | 1.72 | 4 | 2 | 66.76 | 0 | PDE3A | 2100 | 0.46 |
| **74** | | | | | | | JAK2 | 59000 | 0.34 |
| **75** | 260.29 | 2.27 | 4 | 0 | 44.76 | 2 | PDE3A | 6300 | 0.38 |
| **75** | | | | | | | $M_1$ (anta.) | 15000 | 0.35 |
| **75** | | | | | | | AchE | 16000 | 0.35 |
| **76** | 232.23 | 1.72 | 4 | 2 | 66.76 | 0 | COX-1 | 370 | 0.52 |
| **76** | | | | | | | 5-$HT_{2b}$(anta.) | 1700 | 0.47 |
| **76** | | | | | | | $M_1$(anta.) | 7100 | 0.42 |
| **76** | | | | | | | JAK2 | 13000 | 0.39 |
| **76** | | | | | | | ERß | 46000 | 0.35 |
| **77** | 260.29 | 2.27 | 4 | 0 | 44.76 | 2 | - | - | - |

## Chemistry

### General Procedure of a Microwave-Assisted Pechmann-Coumarin Synthesis

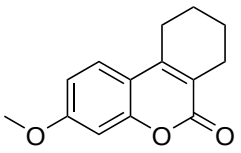
The coumarin derivatives were prepared following an updated literature procedure.[295] An oven dried microwave vial (0.5-2 ml) equipped with a Teflon-coated stirring bar was charged with phenol derivative (1.00 eq), ethyl 2-oxocyclohexane-1-carboxylate (1.00 eq) and TFA (0.5 ml/mmol) and closed using an aluminum open-top seal with a PTFE-faced septum. The reaction mixture was microwave-irradiated for 30 min at 100 °C. After cooling down to room temperature, the reaction mixture was added dropwise into intensely stirred (1000 rpm) ice cold water (50 ml). The formed precipitate was filtered and washed with ice cold water ( 3 x 10 ml) to afford the product. The solid product was dissolved in EtOH (60 °C) and allowed to cool down to room temperature. Deionized water was added after 24 h. The resulting crystals were collected by suction filtration and washed with ice-cold water (3 x 10 ml). The crystals were dried under high vacuum for at least 7 h.

**3-Hydroxy-7,8,9,10-tetrahydro-6*H*-dibenzo[*b*,*d*]pyran-6-one (72)**




Reaction of resorcinol (113.6 mg, 1.02 mmol), ethyl 2-oxocyclohexane-1-carboxylate (188.5 mg, 1.05 mmol) in TFA (0.51 ml) led to white crystals of **72** (0.132 g, 58 %).
**MP**: 205.8 °C(decmp.); $^{1}$**H-NMR** (400 MHz, DMSO-$d_6$) $\delta$ = 10.31 (s, 1H, O*H*), 7.50 (d, $J$ = 8.73 Hz, 1H, C$H_{ar}$), 6.76 (d, $J$ = 8.81 Hz, 1H, C$H_{ar}$), 6.67 (s, 1H, C$H_{ar}$), 2.71 (s, 2H, C$H_2$), 2.37 (s, 2H, C$H_2$), 1.85 - 1.60 (m, 4H, 2x C$H_2$); $^{13}$**C-NMR** (101 MHz, DMSO-$d_6$) $\delta$ = 161.42 , 160.30, 153.48, 125.48, 118.89, 113.12, 112.43, 102.38, 25.03, 23.95, 21.72, 21.35; **HR-MS** (ESI-TOF): calculated for $C_{13}H_{13}O_3$: *m/z* 217.0859, found: *m/z* 217.0858 $[M+H]^+$.

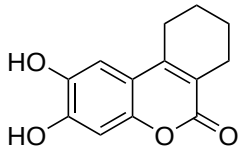
**3-Methoxy-7,8,9,10-tetrahydro-6*H*-dibenzo[*b*,*d*]pyran-6-one (73)**




Reaction of 3-methoxyphenol (0.120 ml, 1.05 mmol), ethyl 2-oxocyclohexane-1-carboxylate (0.171 ml, 1.01 mmol) in TFA (0.5 ml) led to neon-green crystals of **73** (50.8 mg, 22%).
**MP**: 107.1 °C; $^{1}$**H-NMR** (400 MHz, DMSO-$d_6$) $\delta$ = 7.61 (dd, $J$ = 0.75, 8.43 Hz, 1H, C$H_{ar}$),

6.98 - 6.88 (m, 2H, C$H_{ar}$), 3.84 (s, 3H, C*H*3), 2.80 - 2.72 (m, 2H, C$H_2$), 2.43 - 2.36 (m, 2H, C$H_2$), 1.82 - 1.65 (m, 4H, 2x C$H_2$); $^{13}$**C-NMR** (101 MHz, DMSO-$d_6$) $\delta$ =161.67, 161.34, 153.44, 148.02, 125.39, 119.86, 113.54, 112.36, 100.90, 56.23, 25.11, 24.00, 21.67, 21.30; **HR-MS** (ESI-TOF): calculated for $C_{14}H_{15}O_3$: *m/z* 231.1016, found: *m/z* 231.1021 $[M+H]^+$.

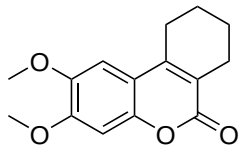
**2,3-Dihydroxy-7,8,9,10-tetrahydro-6*H*-dibenzo[*b*,*d*]pyran-6-one (74)**




Reaction of benzene-1,2,4-triol (300 mg, 2.35 mmol) and ethyl 2-oxocyclohexane-1-carboxylate (422 mg, 2.35 mmol) in TFA (1 ml) led to brownish crystals of **74** (0.416 g, 76 %).

**MP**: 230.0 °C(decomp.); $^1$**H-NMR** (400 MHz, methanol-$d_4$): $\delta$ = 7.00 (s, 1H, C$H_{ar}$), 6.71 (s, 1H, C$H_{ar}$), 2.80 - 2.70 (m, 2H, C$H_2$), 2.53 - 2.43 (m, 2H, C$H_2$), 1.91 - 1.75 (m, 4H, 2x C$H_2$); $^{13}$**C-NMR** (75 MHz, methanol-$d_4$) $\delta$ = 164.8, 150.3, 150.1, 147.9, 144.2, 120.2, 113.7, 109.0, 103.5, 26.4, 24.9, 22.8, 22.5; **HR-MS** (ESI-TOF): calculated for $C_{13}H_{12}O_4Na$ : *m/z* 255.0628, found: *m/z* 255.0632 $[M+Na]^+$.

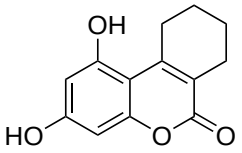
**2,3-Dimethoxy-7,8,9,10-tetrahydro-6*H*-dibenzo[*b*,*d*]pyran-6-one (75)**




Reaction of 3,4-methoxyphenol (117 mg, 0.74 mmol) and ethyl 2-oxocyclohexane-1-carboxylate (133 mg, 0.74 mmol) in TFA (0.37 ml) led to yellow crystals of **75** (172 mg, 89 %).

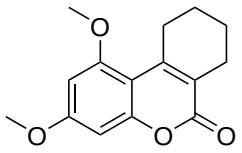
**MP**: 186.5 °C; $^1$**H-NMR** (400 MHz, chloroform-*d*): $\delta$ = 6.43 (d, *J* = 2.4 Hz, 1H, C$H_{ar}$), 6.29 (d, *J* = 2.4 Hz, 1H, C$H_{ar}$), 3.83 (d, *J* = 1.8 Hz, 6H, 2x OC$H_3$), 3.05 (tt, *J* = 3.0, 5.9 Hz, 2H, C$H_2$), 2.61 – 2.47 (m, 2H, C$H_2$), 1.73 (p, *J* = 3.4 Hz, 4H, 2x C$H_2$); $^{13}$**C-NMR** (101 MHz, chloroform-*d*) $\delta$ = 162.4, 151.5, 147.6, 147.1, 146.1, 121.1, 112.8, 104.3, 100.0, 56.5, 56.4, 25.6, 24.0, 21.8, 21.6; **HR-MS** (ESI-TOF): calculated for $C_{15}H_{16}O_4Na$ : *m/z* 283.0941, found: *m/z* 283.0944 $[M+Na]^+$.

**1,3-Dihydroxy-7,8,9,10-tetrahydro-6*H*-dibenzo[*b*,*d*]pyran-6-one (76)**




Reaction of benzene-1,3,5-triol (143 mg, 1.11 mmol) and ethyl 2-oxocyclohexane-1-carboxylate (186 mg, 1.04 mmol) in TFA (0.5 ml) and purified by flash chromatography (SiO2, $CH_2Cl2$/MeOH, 20:1 (v/v)) led to white crystals of **76** (149 mg, 58 %).
**MP**: 265 °C; $^1$**H-NMR** (400 MHz, DMSO-$d_6$) $\delta$ = 10.33 (s, 1H, O*H*), 10.08 (s, 1H, O*H*), 6.24 (d, *J* = 2.4 Hz, 1H, C$H_{ar}$), 6.14 (d, *J* = 2.4 Hz, 1H, C$H_{ar}$), 3.04 (dq, *J* = 2.8, 3.7, 5.6 Hz, 2H, C$H_2$), 2.34 (dt, *J* = 1.8, 4.1 Hz, 2H, C$H_2$), 1.74 – 1.58 (m, 4H, 2x C$H_2$); $^{13}$**C-NMR** (101 MHz, DMSO-$d_6$) $\delta$ = 161.30, 160.09, 157.78, 155.03, 150.49, 116.58, 102.51, 99.79, 94.69, 29.60, 24.51, 22.20, 21.34; **HR-MS** (ESI-TOF): calculated for $C_{13}H_{13}O_4$: *m/z* 233. 0808, found: *m/z* 233.0809 $[M+Na]^+$.

**1,3-Dimethoxy-7,8,9,10-tetrahydro-6*H*-dibenzo[*b*,*d*]pyran-6-one (77)**




Reaction of 3,5-dimethoxyphenol (200 mg, 1.26 mmol) and ethyl 2-oxocyclohexane-1-carboxylate (225 mg, 1.26 mmol), in TFA (0.6 ml) led to white crystals of **77** (271 mg, 83 %).
**MP**: 176.2 °C; $^1$**H-NMR** (400 MHz, chloroform-*d*): $\delta$ = 6.43 (d, *J* = 2.5 Hz, 1H, C$H_{ar}$), 6.29 (d, *J* = 2.5 Hz, 1H, C$H_{ar}$), 3.84 (s, 3H, OC$H_3$), 3.83 (s, 3H, OC$H_3$), 3.10 - 3.00 (m, 2H, C$H_2$), 2.59 - 2.49 (m, 2H, C$H_2$), 1.78 - 1.67 (m, 4H, 2x C$H_2$); $^{13}$**C-NMR** (101 MHz, chloroform-*d*) $\delta$ = 162.1, 161.6, 158.9, 155.1, 150.1, 119.3, 105.3, 95.8, 93.3, 55.8, 55.8, 30.0, 24.7), 22.5, 21.4; **HR-MS** (ESI-TOF): calculated for $C_{15}H_{17}O_4$: *m/z* 261.1121, found: *m/z* 261.1126 $[M+H]^+$.

## X-Ray Structures

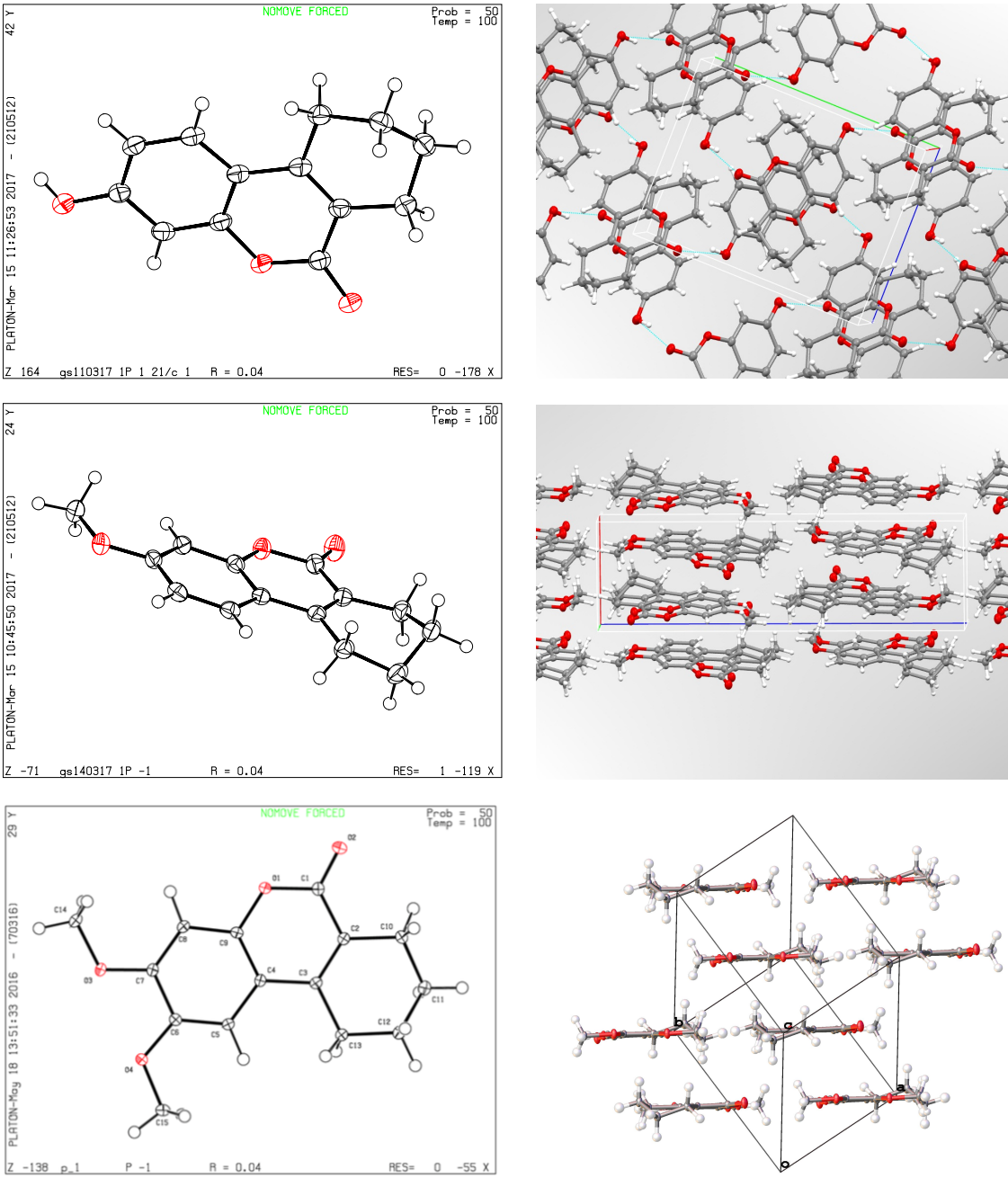



**Figure A.1:** X-ray structure and packing of compound **72** (*top*), **73** (*middle*) and **74** (*bottom*).

## Dynamic Light Scattering

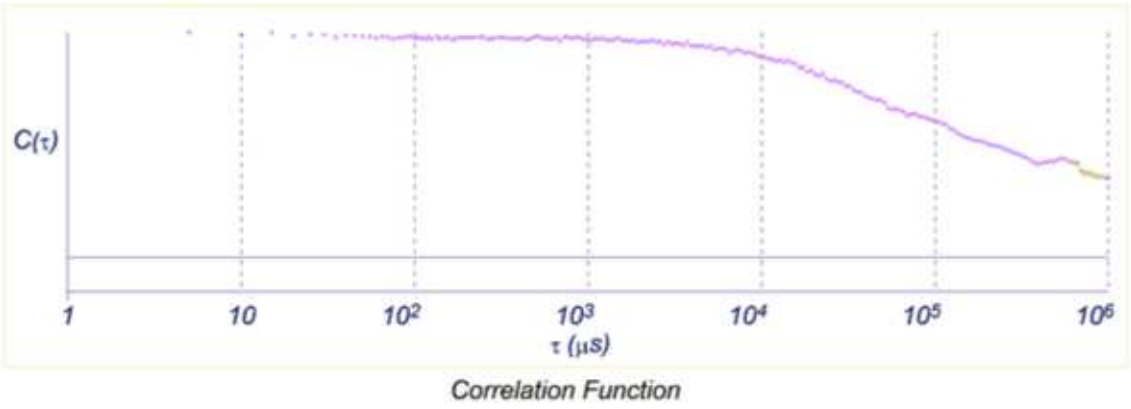



**Figure A.2:** Compound **74** at 62.50 µM showed aggregation behaviour (exponential decaying correlation function) in $H_2O$ after 0 min.

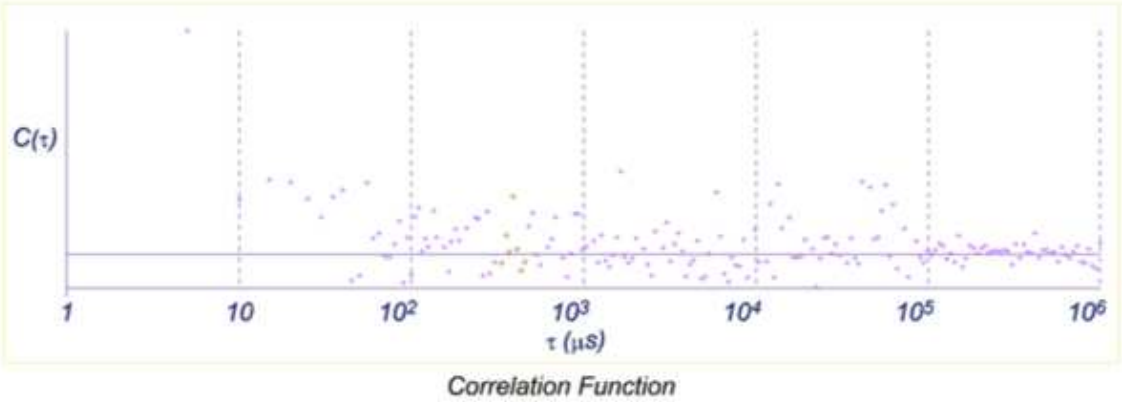



**Figure A.3:** No aggregation behaviour of compound **74** at 31.25 µM in $H_2O$ after 60 min was observed.

## *In Vitro* Characterization

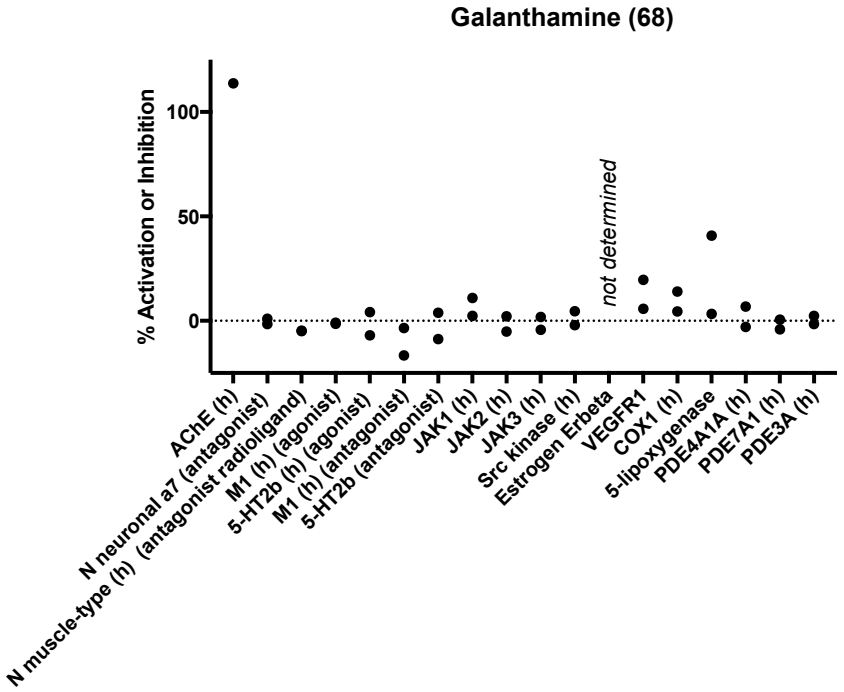



**Figure A.4:** Activation or inhibition of galanthamine (**68**) on selected targets in an initial screening at a concentration of 10 µM with two replicates. Galanthamine showed no or only weak activity on the investigated target set. The effect of galanthamine on estrogen ERβ was not determined.

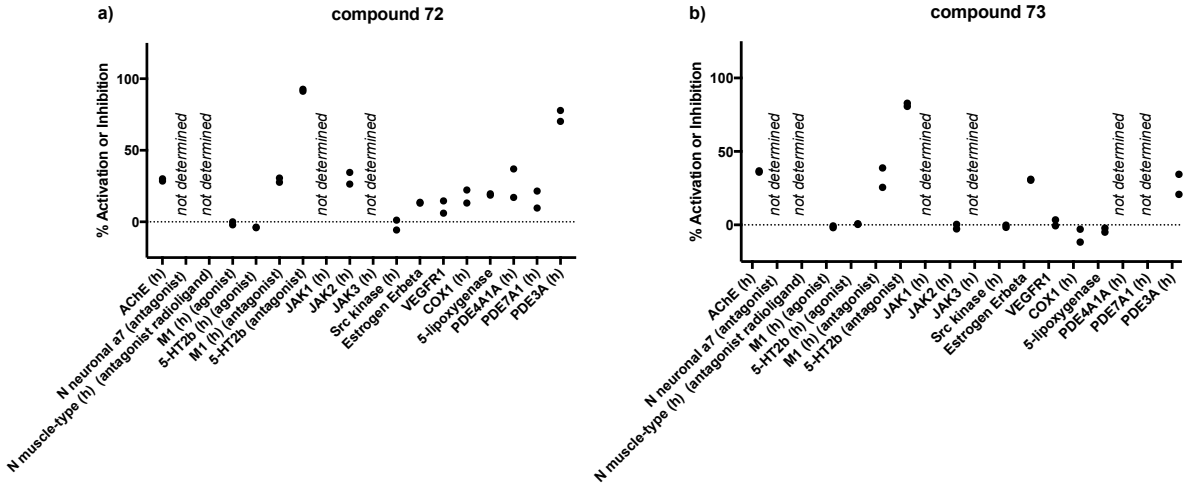



**Figure A.5:** Activation or inhibition of galanthamine mimetics **a) 72** and **b) 73** on selected targets in an initial screening at a concentration of 10 µM with two replicates. Galanthamine showed no or only weak activity on the investigated target set. The effects of the mimetics **72** and **73** were not determined on the several indicated targets.

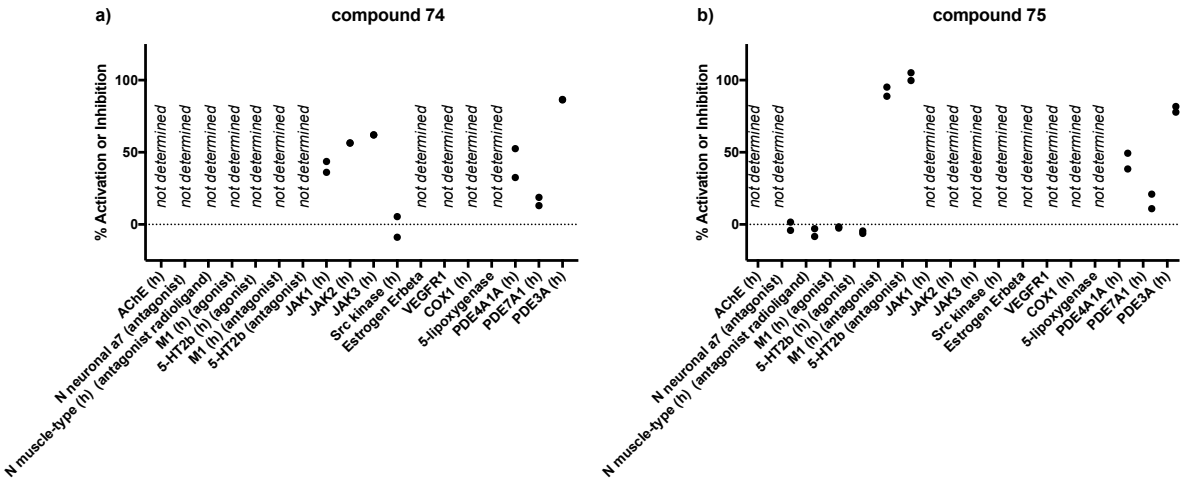



**Figure A.6:** Activation or inhibition of galanthamine mimetics **a) 74** and **b) 75** on selected targets in an initial screening at a concentration of 30 µM with two replicates. Galanthamine showed no or only weak activity on the investigated target set. The effects of the mimetics **74** and **75**were not determined on the several indicated targets.

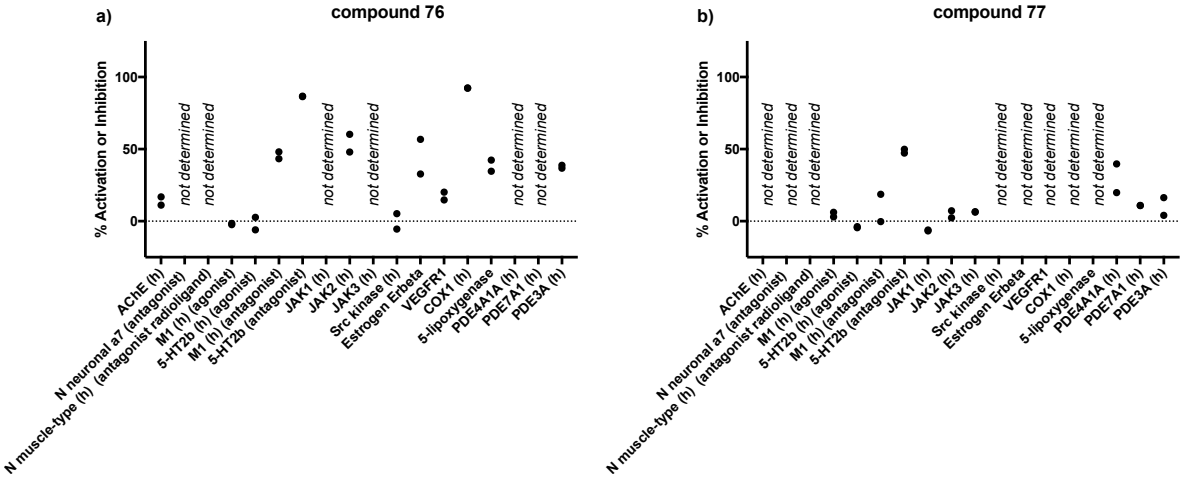



**Figure A.7:** Activation or inhibition of galanthamine mimetics **a) 76** and **b)77** on selected targets in an initial screening at a concentration of 10 (**76**) or 30 µM (**77**) with two replicates. Galanthamine showed no or only weak activity on the investigated target set. The effects of the mimetics **76** and **77**were not determined on the several indicated targets.

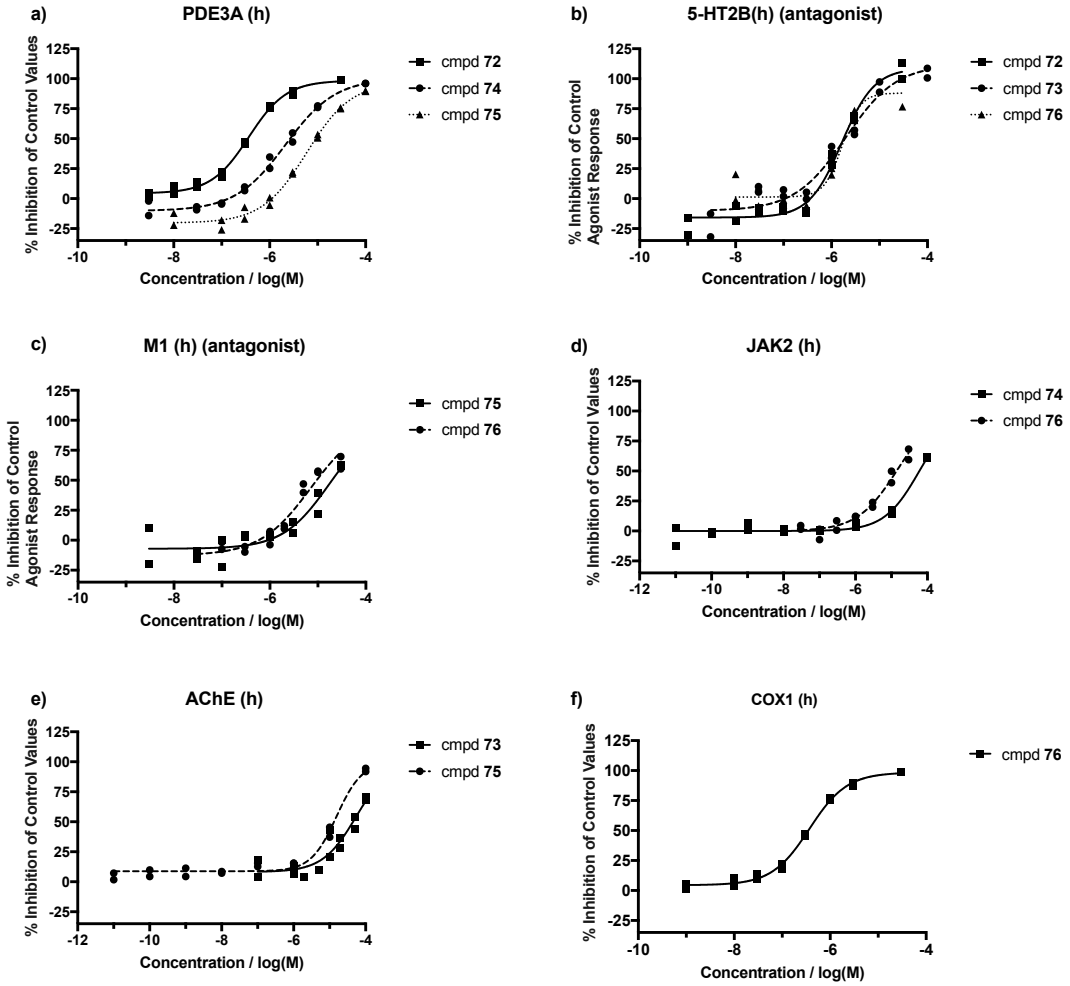



**Figure A.8:** In vitro activity data of mimetics **72-76** against the targets: **a)** Phosphodiesterase 3A (PDE3A), **b)** serotonin receptor ($5HT_{2b}$), **c)** choline receptor (muscarinic) (M1), **d)** janus kinase 2 (JAK2), **e)** acetylcholineesterase (AChE), **f)** cyclooxygenase (COX-1).

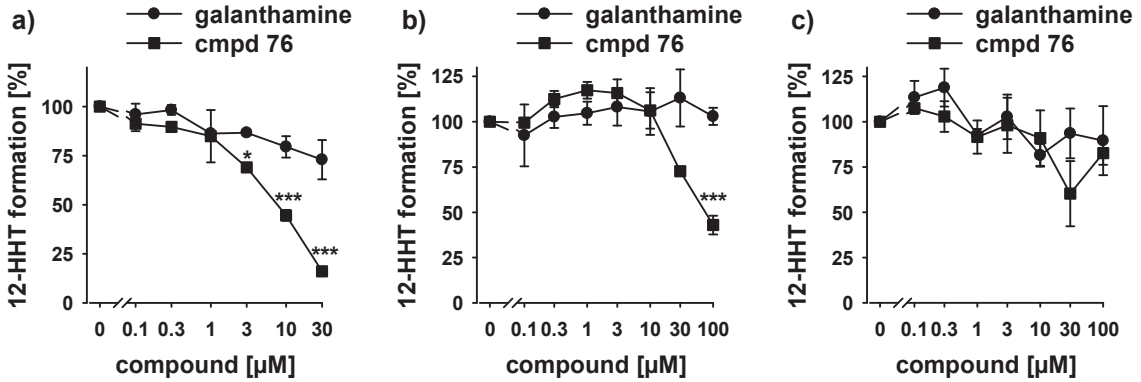



**Figure A.9:** Effects of galanthamine and compound **76** on **a)** COX-1 activity in human platelets, **b)** isolated bovine COX-1, and **c)** isolated human recombinant COX-2. Mean ± S.E. from n = 3 independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 vs. vehicle control. Repeated measures ANOVA+ Tukey HSD post hoc tests.

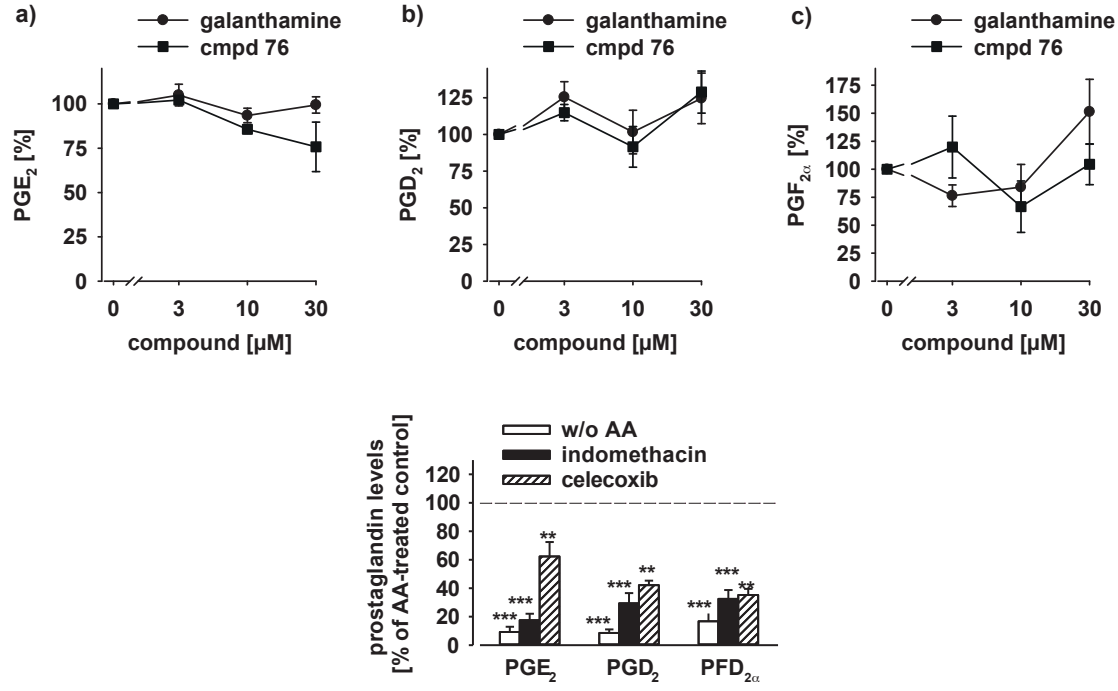



**Figure A.10:** Effects of galanthamine and compound **76** (a – c) and reference COX inhibitors (d) on COX-2-drived prostaglandin formation in human monocytes. Mean ± S.E. from n = 3 independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001 vs. vehicle control. Repeated measures ANOVA+ Tukey HSD post hoc tests.

# A.2 Supplementary Information to Section 4.2

## Computational Results

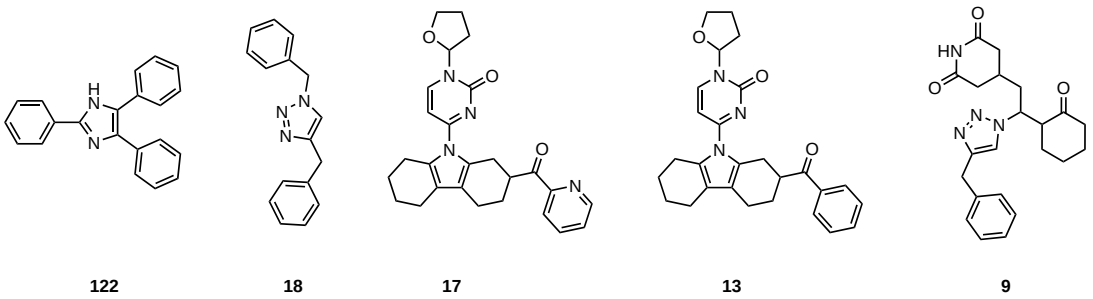



**Figure A.11:** Five most frequent scaffolds in the *de novo* design set ($N_{\text{total}}$ = 802). The 2,4,5-triphenyl imidazole scaffold was the most common scaffold.

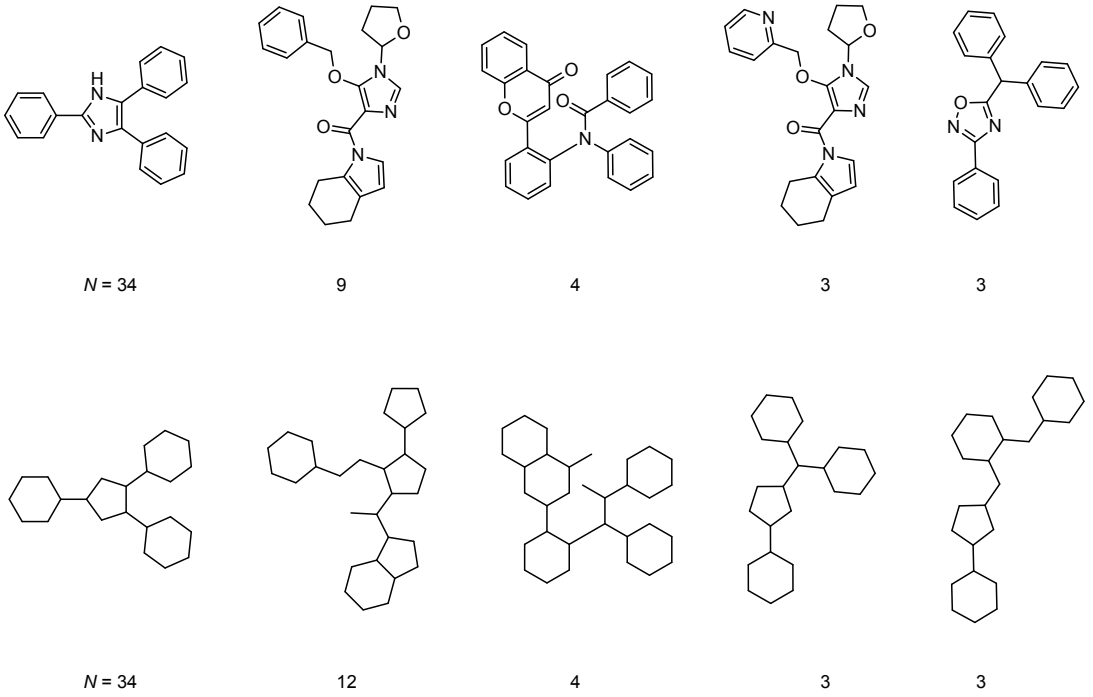



**Figure A.12:** Five most frequent scaffolds and frameworks 100 most similar designs ranked by the CATS descriptor distances ($N_{\text{total}}$ = 100). Again, the 2,4,5-triphenyl scaffold and its related framework was the most abundant scaffold and framework respectively.

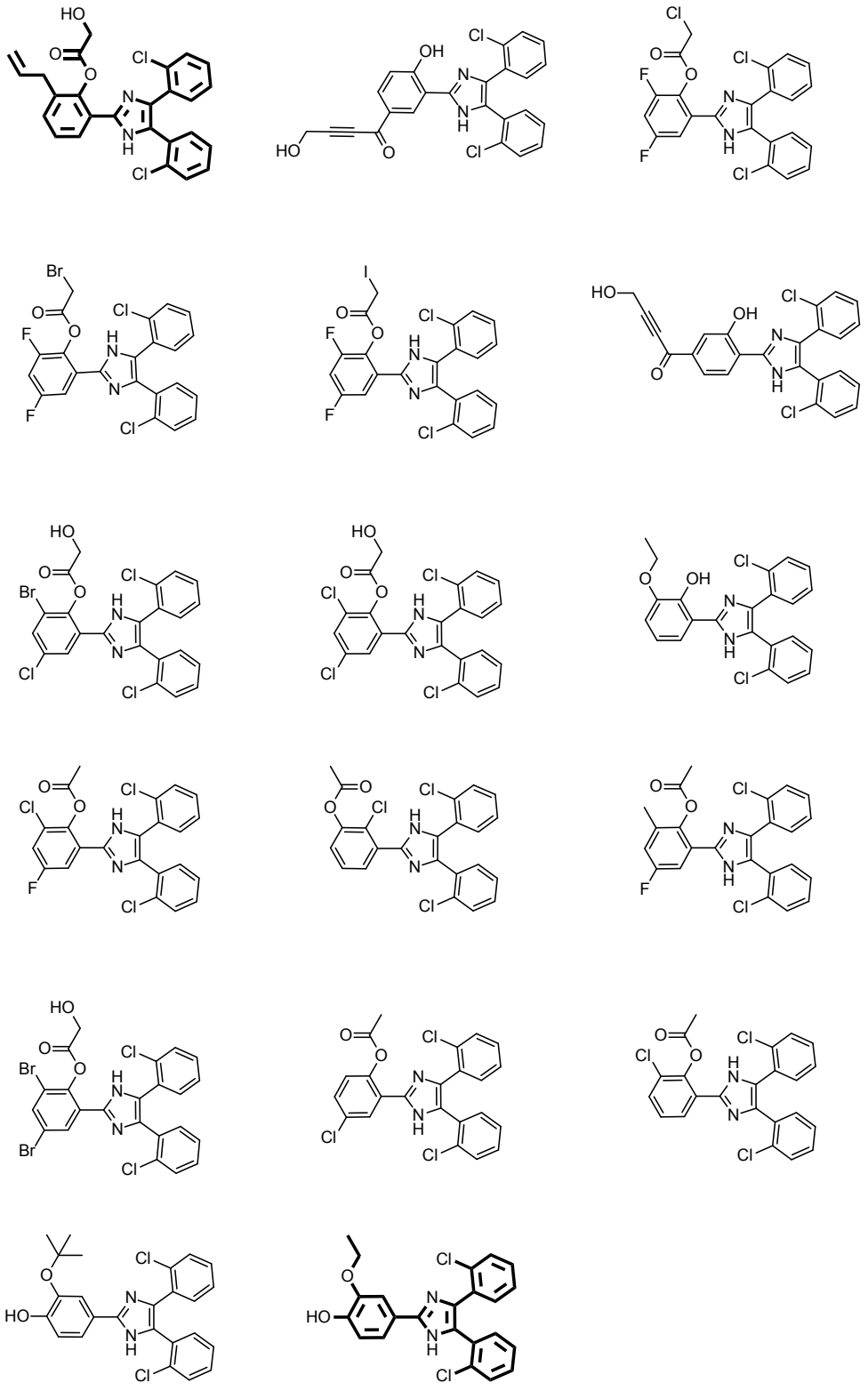
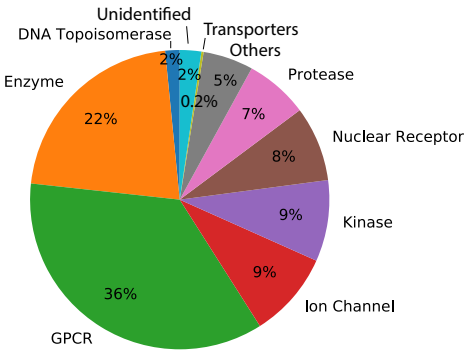
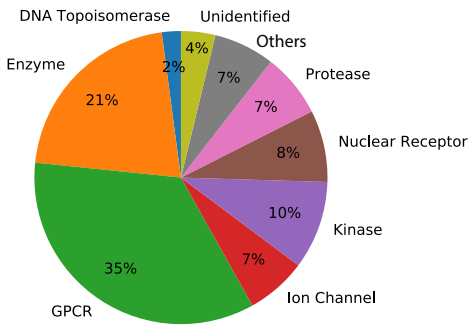


**Figure A.13:** The first 17 designs of the high ranked compounds (CATS distance) with an 2,4,5-triphenyl imidazole scaffold, whereof 34 top-ranked designs contained this scaffold. Designs **85** and **86** with their individual ranks 1 and 17 (bold) were selected according to the feasibility of the computationally proposed synthetic route and the building block availability.




**(a)** Proportion of protein families of predicted molecular targets of the 802 designs from DOGS ($N_{\text{total}}$ = 9418).




**(b)** Proportion of protein families of predicted molecular targets of the 100 most similar designs according to their CATS distance ($N_{\text{total}}$ = 1664).

**Figure A.14:** Proportion of protein families of predicted molecular targets of **(a)** all designs obtained from DOGS (802 entities) and **(b)** of the 100 most similar designs according to their CATS distance.

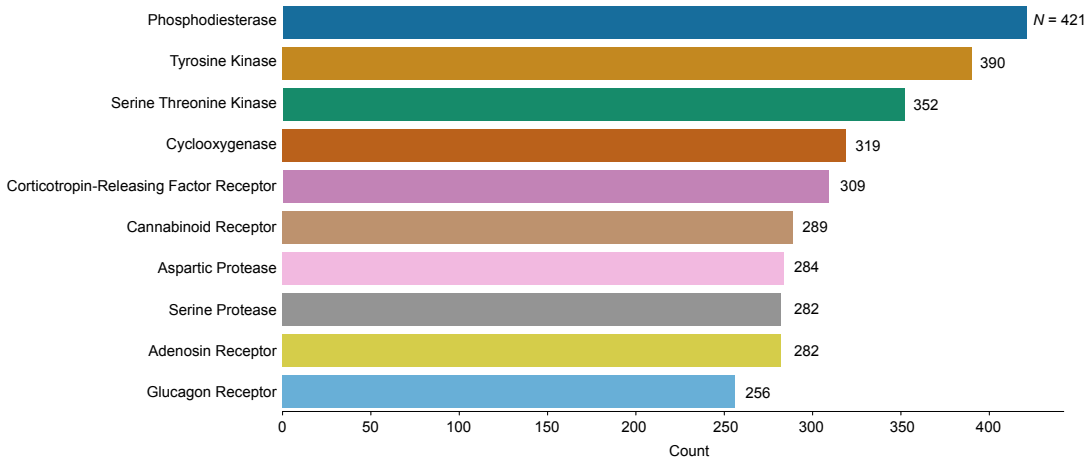
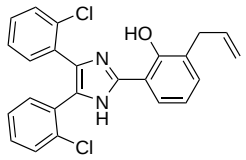



**Figure A.15:** Ten most frequent predicted protein families of all DOGS designs. From the target prediction of these top-ranked designs, all annotated subtypes of each protein family were counted ($N_{\text{total}}$ = 9418).

## Chemistry

**2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenol (87)**



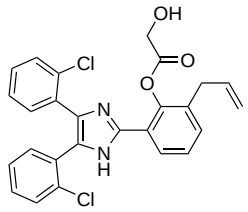

1,2-bis(2-Chlorophenyl)ethane-1,2-dione (**88**, 516 mg, 1.8 mmol, 1.0 eq), 3-allyl-2-hydroxybenzaldehyde (**89**, 300 mg, 1.8 mmol, 1.0 eq) and ammonium acetate (1.5 g 18.4 mmol, 10.2 eq) were dissolved in glacial acetic acid (10 ml) and heated for 5 minutes to 180 °Cunder microwave irradiation. The mixture was then allowed to cool to room temperature and added dropwise to a cold ammonium hydroxide solution (25%, 150 ml). A yellow precipitate was filtered off and the solid was washed with cold water. The crude solid was purified by column chromatography using hexane/$CH_2Cl_2$ (6:1 + 5% MeOH) to $CH_2Cl_2$ with 5% MeOH as mobile phase to yield the title compound **87** as colorless solid (210 mg, 27%).
**MP**: 70 °C; $^1$**H NMR** (400 MHz, DMSO-$d_6$): $\delta$ = 13.28 (s, 1H, N*H*), 13.09 (s, 1H, O*H*), 7.85 (dd, *J* = 7.9, 1.6 Hz, 1H, C$H_{ar}$), 7.60 – 7.52 (m, 1H,C$H_{ar}$), 7.51 – 7.25 (m, 7H, C$H_{ar}$), 7.15 (dd, *J* = 7.4, 1.4 Hz, 1H, C$H_{ar}$), 6.90 (t, *J* = 7.6 Hz, 1H, C$H_{ar}$), 6.01 (ddt, *J* = 16.7, 10.1, 6.6 Hz, 1H, $CH_2$=C*H*), 5.12 – 5.00 (m, 2H, CH=C$H_2$), 3.48 – 3.36 (m, 2H, $C_{ar}CH_2$); $^{13}$**C-NMR** (100 MHz, DMSO-$d_6$): $\delta$ = 154.5, 145.1, 136.7, 134.0, 133.2, 132.5, 132.3, 131.6, 130.6, 130.3, 129.9 , 129.8, 129.4, 129.3, 127.4, 127.3, 127.0, 126.1, 122.8, 119.5, 118.6, 115.6, 112.3, 33.6; **HR-MS** (ESI-TOF) calculated for $C_{23}H_{19}Cl_2N_2O_2^+$: *m/z* 421.0869, found: *m/z* 421.0871 $[M+H]^+$.

**2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl-2-((*tert*-butyldiphenylsilyl)oxy) acetate**

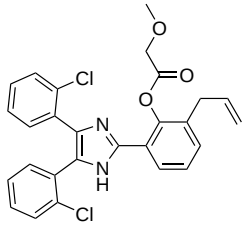
2-((*tert*-Butyldiphenylsilyl)oxy)acetic acid (130 mg, 413.00 µmol, 1.16 eq) was dissolved in $CH_2Cl_2$ (2 ml). 4-(Dimethylamino)pyridine (6.20 mg , 50.00 µmol, 0.14 eq) and **87** (150 mg, 356 µmol, 1 eq) were added to the reaction mixture and cooled to 0 °C. After 15 min, N,N'-dicyclohexylcarbodiimide (44 mg, 214 µmol, 1.50 eq) in CH2Cl2 (1 ml) was added slowly and the mixture was stirred for 1 h at 0 °C. The reaction solution was allowed to warm up to room temperature and stirred for further 16 h. White precipitates was filtered off, and the filtrate was diluted with EtOAc (20 ml). The organic solution was washed with water (25 ml) and brine (25 ml), dried over MgSO4, and concentrated under reduced pressure. The crude product was purified by column chromatography using hexane/EtOAc (7:1 to 3:1) as mobile phase to yield compound as colorless solid (96 mg, 38%).
$^1$**H NMR** (400 MHz, acetone-$d_6$): $\delta$ = 11.77 (s, 1H, N*H*), 7.99 (dd, *J* = 3.0, 6.5 Hz, 1H,

C$H_{ar}$), 7.76 – 7.71 (m, 4H, C$H_{ar}$), 7.52 (dd, *J* = 1.8, 7.5 Hz, 1H, C$H_{ar}$), 7.50 (d, *J* = 1.3 Hz, 1H, C$H_{ar}$), 7.48 (d, *J* = 1.4 Hz, 1H, C$H_{ar}$), 7.46 (t, *J* = 1.6 Hz, 1H, C$H_{ar}$), 7.44 (t, *J* = 1.0 Hz, 1H, C$H_{ar}$), 7.43 (t, *J* = 1.1 Hz, 1H, C$H_{ar}$), 7.41 (t, *J* = 1.6 Hz, 1H, C$H_{ar}$), 7.38 – 7.33 (m, 3H, C$H_{ar}$), 7.30 (dd, *J* = 1.8, 7.3 Hz, 1H, C$H_{ar}$), 7.28 – 7.23 (m, 2H, C$H_{ar}$), 7.21 (dd, *J* = 1.4, 7.4 Hz, 1H, C$H_{ar}$), 5.89 (ddt, *J* = 6.7, 10.1, 16.8 Hz, 1H, (ddt, *J* = 16.7, 10.1, 6.6 Hz, 1H, $CH_2$=C*H*), 5.07 – 4.97 (m, 2H, CH=C$H_2$), 4.77 (d, *J* = 9.7 Hz, 2H, C(=O)C$H_2$, 3.27 (dd, *J* = 1.5, 6.7 Hz, 2H, $C_{ar}$C$H_2$), 1.05 (s, 9H, 3x C$H_3$). $^{13}$**C-NMR** (100 MHz, DMSO-$d_6$): $\delta$ = 210.1, 169.8, 136.3, 135.4, 132.4, 130.5, 130.0, 129.4, 128.4, 127.6, 127.2, 34.5, 26.9; **HR-MS** (ESI) calculated for $C_{42}H_{39}Cl_2N_2O_3Si^+$: *m/z* 717.2093, found: *m/z* 717.2093 $[M+H]^+$.

**2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl 2-hydroxyacetate (85)**



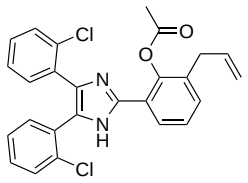

2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl-2-((*tert*-butyldiphenylsilyl)-oxy) acetate (40 mg, 56 µmol, 1 eq) was dissolved in anhydrous THF (1 ml) and cooled to 0 °C. Glacial acetic acid (65 µl, 1.1 mmol, 20.0 eq) was added and the mixture was stirred for 15 min at 0 °C. Tetra-n-butylammonium fluoride (1M solution in THF, 150 µl, 150.0 µmol, 2.7 eq) was added slowly to the reaction mixture. The solution was stirred for 30 min at 0 °Cand additional 90 min at room temperature. The reaction mixture was then quenched with saturated aqueous ammonium chloride solution (15 ml) and extracted with EtOAc (3x 15 ml). The combined organic layers were washed with brine (30 ml), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by column chromatography using hexane/EtOAc (4:1 to 1:1) as mobile phase to yield compound 2 as colorless solid (22 mg, 82%).
$^1$**H NMR** (500 MHz, DMSO-$d_6$): $\delta$ = 12.88 – 12.81 (m, 1H, N*H*), 7.90 (ddt, *J* = 1.9, 4.2, 5.8 Hz, 1H, C$H_{ar}$), 7.53 (d, *J* = 8.0 Hz, 1H, C$H_{ar}$), 7.47 (ddd, *J* = 1.8, 3.8, 7.4 Hz, 1H, C$H_{ar}$), 7.40 – 7.29 (m, 7H, C$H_{ar}$), 7.29 – 7.24 (m, 1H, C$H_{ar}$), 5.97 – 5.86 (m, 1H, $CH_2$=C*H*), 5.52 (dq, *J* = 2.0, 4.3 Hz, 1H, CH=C$H_2$), 5.13 (dt, *J* = 2.1, 17.2 Hz, 1H, CH=C$H_2$), 5.10 – 5.05 (m, 1H, O*H*), 4.41 (dd, *J* = 3.5, 4.5 Hz, 2H, C(=O)C$H_2$), 3.35 – 3.30 (m, 2H, $C_{ar}$C$H_2$). $^{13}$**C-NMR** (126 MHz, DMSO-$d_6$): $\delta$ = 172.0, 145.7, 143.0, 137.6, 136.4, 134.6, 134.4, 133.3, 132.6, 132.5, 130.7, 130.6, 130.4, 130.2, 123.0, 129.4, 127.6, 127.5, 127.4, 127.3, 126.9, 126.6, 123.7, 117.0, 60.7, 34.5; **HR-MS** (ESI) calculated for $C_{26}H_{21}Cl_2N_2O_3^+$: *m/z* 479.0924, found: *m/z* 479.0920 $[M+H]^+$.



**2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl 2-methoxyacetate (85**

2-Methoxyacetic acid (12.8 mg, 142.0 µmol, 1.0 eq) was dissolved in $CH_2Cl_2$ (1 ml). 4-(Dimethylamino)pyridine (1.7 mg, 1.4 µmol, 0.1 eq) and 4 (60 mg, 142 µmol, 1 eq) were added to the reaction mixture. The mixture was cooled to 0 °C, and *N*,*N*'-dicyclohexylcarbodiimide (44 mg, 213.0 µmol, 1.5 eq) was added. The reaction solution was stirred for 5 min at 0 °C, then allowed to warm to room temperature and stirred for 16 h. Formed precipitates were filtered off, and the filtrate was concentrated under reduced pressure. The residue was dissolved in $CH_2Cl_2$ (20 ml) and washed with aqueous hydrochloric acid solution (0.5 M, 2 x 15 ml) and saturated aqueous $NaHCO_3$ solution (2 x 15 ml). The organic layer was dried over $MgSO_4$, and concentrated under reduced pressure. The crude product was purified by column chromatography using hexane/EtOAc(10:1 to 3:1) as mobile phase to yield the title compound **85a** as colorless solid (40 mg, 57%).
**MP** = 215 °C; $^1$**H-NMR** (400 MHz, DMSO-$d_6$): $\delta$ = 12.90 (s, 1H, N*H*), 7.91 (dd, *J* = 7.5, 2.0 Hz, 1H, C$H_{ar}$), 7.52 (dd, 3*J* = 8.0, 1.3 Hz, 1H, C$H_{ar}$), 7.47 – 7.26 (m, 9H, C$H_{ar}$), 5.90 (ddt, *J* = 16.7, 10.0, 6.6 Hz, 1H, $CH_2$=C*H*), 5.16 - 5.03 (m, 2H, CH=C$H_2$), 4.37 (s, 2H, OC$H_2$), 3.33 (d, *J* = 7.6 Hz, 2H, $C_{ar}$C$H_2$), 3.13 (s, 3H, OC*H*3); $^{13}$**C-NMR** (100 MHz, DMSO-$d_6$): $\delta$ = 168.9, 145.0, 142.3, 136.0, 134.1, 133.9, 132.9, 132.3, 132.1, 130.2, 130.0, 129.7, 129.4, 129.0, 127.1, 126.8, 126.2, 126.0, 122.9, 116.5, 69.2, 58.2, 34.0; **HR-MS** (ESI-TOF): calculated for $C_{27}H_{23}Cl_2N_2O_3^+$: *m/z* 493.1080, found: *m/z* 493.1085 $[M+H]^+$.

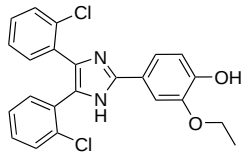
**2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl acetate (85b)**



Glacial acetic acid (8.6 mg, 142.0 µmol, 1.0 eq) was dissolved in $CH_2Cl_2$ (1 ml). 4-(Dimethylamino)pyridine (1.7 mg, 14.2 µmol, 0.1 eq) and **87** (60 mg, 142 µmol, 1 eq) were added to the reaction mixture. The mixture was cooled down to 0 °C, and *N*,*N*'-dicyclohexylcarbodiimide (44 mg, 214.0 µmol, 1.5 eq) was added. The reaction solution was stirred for 5 min at 0 °C, then allowed to warm up to room temperature and stirred

for 16 h. Formed precipitates were filtered off, and the filtrate was concentrated under reduced pressure. The residue was dissolved in $CH_2Cl_2$ (20 ml) and washed with aqueous hydrochloric acid solution (0.5 M, 2 x 15 ml) and saturated aqueous $NaHCO_3$ solution (2 x 15 ml). The organic layer was dried over $MgSO_4$, and concentrated under reduced pressure. The crude product was purified by column chromatography using hexane/EtOAc (10:1 to 3:1) as mobile phase to yield the title compound **85b** as colorless solid (35 mg, 53%).
**MP** = 225 °C; $^1$**H-NMR** (400 MHz, DMSO-$d_6$) $\delta$ = 12.83 (s, 1H, N*H*), 7.87 (dd, 3 = 7.3, 2.2 Hz, 1H, C$H_{ar}$), 7.52 (dd, *J* = 7.9, 1.4 Hz, 1H, C$H_{ar}$), 7.44 – 7.25 (m, 9H, C$H_{ar}$), 5.91 (ddt, *J* = 16.8, 10.0, 6.7 Hz, 1H, $CH_2$=C*H*), 5.16 – 5.04 (m, 2H, CH=C$H_2$), 3.35 – 3.29 (m, 2H, $C_{ar}CH_2$), 2.26 (s, 3H, C$H_3$); $^{13}$**C-NMR** (100 MHz, DMSO-$d_6$): $\delta$ = 169.1, 145.7, 142.5, 136.1, 134.2, 133.8, 132.8, 132.1, 132.0, 130.1, 130.0, 129.7, 129.5, 128.9, 127.1, 126.8, 126.3, 126.0, 123.2, 116.4, 34.1, 21.2; **HR-MS** (ESI-TOF): calculated for $C_{26}H_{21}Cl_2N_2O_2^+$: *m/z* 463.0975, found: *m/z* 463.0976 $[M+H]^+$.

**2-Allyl-6-(4,5-bis(2-chlorophenyl)-4,5-dihydro-1*H*-imidazol-2-yl)phenol (86)**




1,2-bis(2-chlorophenyl)ethane-1,2-dione (**88**, 140 mg, 0.5 mmol, 1.0 eq), 3-ethoxy-4-hydroxybenzaldehyde (**90**, 83 mg, 0.5 mmol, 1.00 eq) and ammonium acetate (385 mg, 5 mmol, 10 eq) were dissolved in glacial acetic acid (2.5 ml) and heated for 5 minutes to 180 °Cunder microwave irradiation. The mixture was allowed to cool to room temperature and added dropwise to a cold ammonium hydroxide solution (25%, 50 ml). The aqueous mixture was extracted with EtOAc (4 x 20 ml), the combined organic layers were washed with 50% (w/w) aqueous sodium bisulfite solution 6 x 30 ml). The organic layer was dried over $MgSO_4$, and concentrated under reduced pressure. The crude solid was purified by column chromatography using $CH_2Cl_2$ with 5% MeOH as mobile phase to yield the title compound as colorless solid (101 mg, 48%).
**MP** = 215 °C; $^1$**H-NMR** (400 MHz, DMSO-$d_6$): $\delta$ = 12.90 (s, 1H, N*H*), 7.91 (dd, *J* = 7.5, 2.0 Hz, 1H, C$H_{ar}$), 7.52 (dd, *J* = 8.0, 1.3 Hz, 1H, C$H_{ar}$), 7.47 – 7.26 (m, 9H, C$H_{ar}$), 5.90 (ddt, 3*J* = 16.7, 10.0, 6.6 Hz, 1H, $CH_2$=C*H*), 5.16 - 5.03 (m, 2H, CH=C$H_2$), 4.37 (s, 2H, OC$H_2$), 3.33 (d, *J* = 7.6 Hz, 2H, $C_{ar}CH_2$), 3.13 (s, 3H, C$H_3$); $^{13}$**C-NMR** (100 MHz, DMSO-$d_6$): $\delta$ = 147.1, 146.6, 145.9, 136.5, 134.2, 132.6, 132.1, 132.0, 131.9, 130.7, 129.6, 129.6, 129.5, 129.2, 128.6, 126.9, 126.5, 125.7, 121.6, 118.0, 115.5, 110.2, 63.7, 14.5. **HR-MS** (ESI-TOF): calculated for $C_{23}H_{19}Cl_2N_2O_2^+$: *m/z* 425.0818, found: *m/z* 425.0822 $[M+H]^+$.

## *In Vitro* Characterization

**Table A.4:** Initial Screening results of several assays provided from Eurofins (Cerep), France. Rounded values are reported as the mean effect (agonist assays) or inhibition (antagonist/binding assays) of two replicates [%]. *tbd* = to be determined; *n.d.* = not determined; (COX = cyclooxygenase; EP = prostaglandin E2 receptor; CDK = cyclin-dependent kinase; ERK/MAPK = mitogen-activated protein kinase; JNK = c-Jun N-terminal kinase; GSK = glycogen synthase kinase; IKK = inhibitor of nuclear factor kappa-B kinase; IRAK = interleukin-1 receptor-associated kinase; PKC = protein kinase C; ROCK = Rho associated coiled-coil containing protein kinase; CB = cannabinoid receptor; CRF = corticotropin-releasing factor; CCK = cholecystokinin receptor; GR = glutocorticoid receptor; OX = Orexin).

| Target (mode) | **84** | **85** | **85a** | **85b** | **86** | **87** |
|---|---|---|---|---|---|---|
| COX-1 (enzymatic) | 95 | *tbd* | 95 | 78 | 98 | 98 |
| EP1 (antagonism) | 86 | 73 | 3 | -4 | 19 | 11 |
| EP2 (antagonism) | 104 | 15 | -30 | 10 | 34 | 35 |
| EP3 (antagonism) | 84 | 46 | 20 | 13 | 41 | 13 |
| EP4 (antagonism) | 37 | 7 | -10 | -21 | -25 | 3 |
| CDK1/CyclinB (enzymatic, antagonistic) | 9 | | | *n.d.* | | |
| CDK2/CyclinA (enzymatic, antagonistic) | 22 | | | *n.d.* | | |
| CKD4/CyclinD (enzymatic, antagonistic) | -4 | | | *n.d.* | | |
| p38alpha kinase (enzymatic, antagonistic) | -4 | | | *n.d.* | | |
| ERK2/MAPK1 (enzymatic, antagonistic) | -2 | | | *n.d.* | | |
| JNK3 (enzymatic, antagonistic) | -10 | | | *n.d.* | | |
| GSK3beta (enzymatic, antagonistic) | 3 | | | *n.d.* | | |
| IKKalpha (enzymatic, antagonistic) | -9 | | | *n.d.* | | |
| IKKbeta (enzymatic, antagonistic) | -5 | | | *n.d.* | | |
| IKKepsilon (enzymatic, antagonistic) | 25 | | | *n.d.* | | |
| IRAK4 (enzymatic, antagonistic) | 2 | | | *n.d.* | | |
| PKCμ (enzymatic, antagonistic) | 16 | | | *n.d.* | | |
| ROCK2 (enzymatic, antagonistic) | 3 | | | *n.d.* | | |
| CB1 (agonism) | 43 | -4 | | *n.d.* | | |
| CB1 (antagonism) | -7 | 86 | | *n.d.* | | |
| CB2 (agonism) | 31 | -12 | | *n.d.* | | |
| CB2 (antagonism) | 14 | 59 | | *n.d.* | | |
| CRF1(antagonism) | 75 | 57 | 7 | 28 | 45 | 46 |
| CCK2 (antagonism) | 1 | 64 | 11 | 14 | 66 | 36 |
| GR (binding competition) | 98 | 85 | 27 | 22 | 59 | 86 |
| OX1 (antagonism) | 116 | 66 | 14 | 8 | 57 | 6 |
| OX2 (antagonism) | 103 | 63 | 15 | 9 | 37 | 11 |

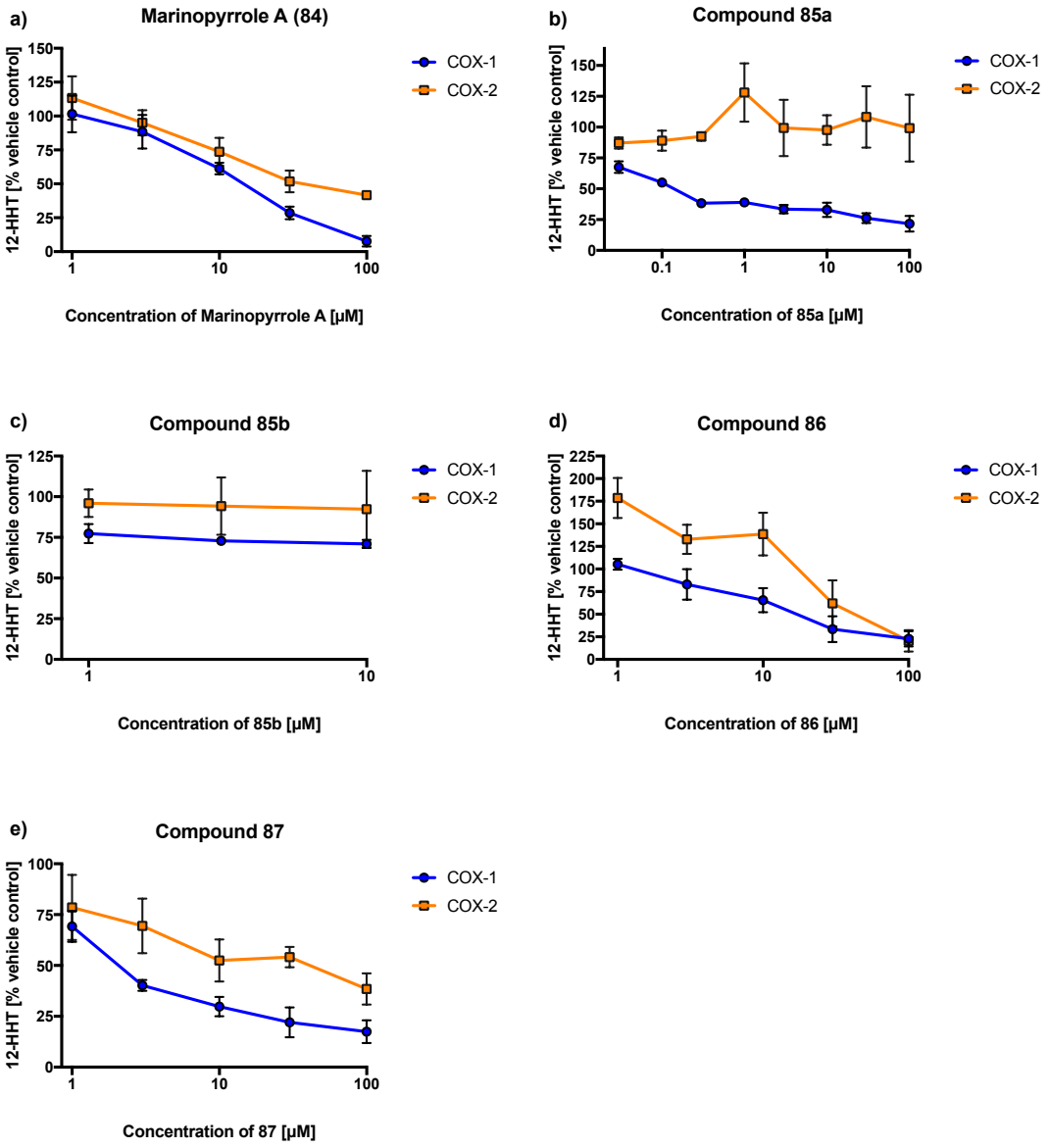



**Figure A.16:** Inhibition of isolated cyclooxygenase 1 (COX-1) and human recombinant COX-2 by marinopyrrole A (**84**), **85a**, **85b**, **86**, and **87**. **84** (COX-1: $IC_{50}$ = 16.6±2.6 µM, COX-2: $IC_{50}$ = 45.2±21.3 µM) and **86** (COX-1: $IC_{50}$ = 29.7±18.1 µM, COX-2: $IC_{50}$ = 53.1±15.13 µM) inhibited the activity of COX-1 and COX-2 in a concentration-dependent manner and slightly selective to COX-1 inhibition. **85a** (COX-1: $IC_{50}$ = 0.2±0.0 µM, COX-2: no inhibition up to 100 µM) and **87** (COX-1: $IC_{50}$ = 1.8±0.5 µM, COX-2: $IC_{50}$ = 93.4±53.3 µM) showed potent and selective COX-1 inhibition.

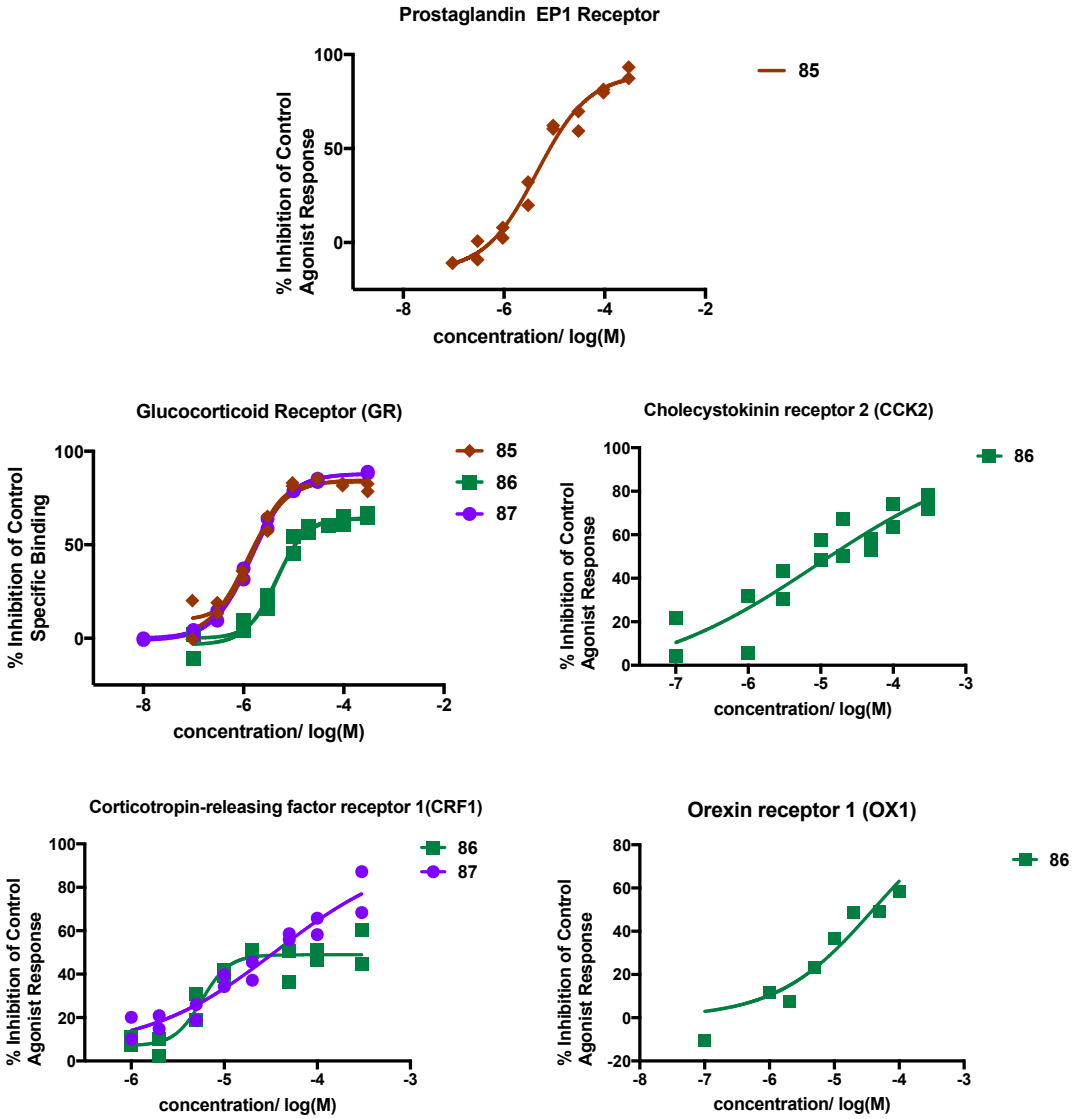



**Figure A.17:** *In vitro* activities of **85**, **86**, and **87** on glucocorticoid receptor (GR), cholecystokinin receptor 2 (CCK2), corticotropin-releasing factor receptor 1 (CRF1), and orexin receptor 1 (OX1). **a) 85** ($IC_{50}$ = 1.2±1.2 µM, $K_d$ = 0.6 µM), **86** ($IC_{50}$ = 4.3±1.2 µM, $K_d$ = 2.2 µM) and **87** ($IC_{50}$ = 1.4±1.1 µM, $K_d$ = 0.7 µM) binds to the glucocorticoid receptor (GR) in a competition binding assay. **b) 86** ($IC_{50}$ = 8.7±4.6 µM, $K_i$ = 1.1 µM) is an antagonist of the cholecystokinin B receptor (CCK2). **c) 86** ($IC_{50}$ = 4.7±1.2 µM, $K_i$ = 1.7 µM) and **87** ($IC_{50}$ = 40±2 µM, $K_i$ = 14 µM) have antagonistic effects on the corticotropin-releasing factor receptor 1 (CRF1). **d)** Mimetic **86** ($IC_{50}$ = 40±1 µM, $K_i$ = 8.4 µM) also antagonizes orexin receptor 1 (OX1).

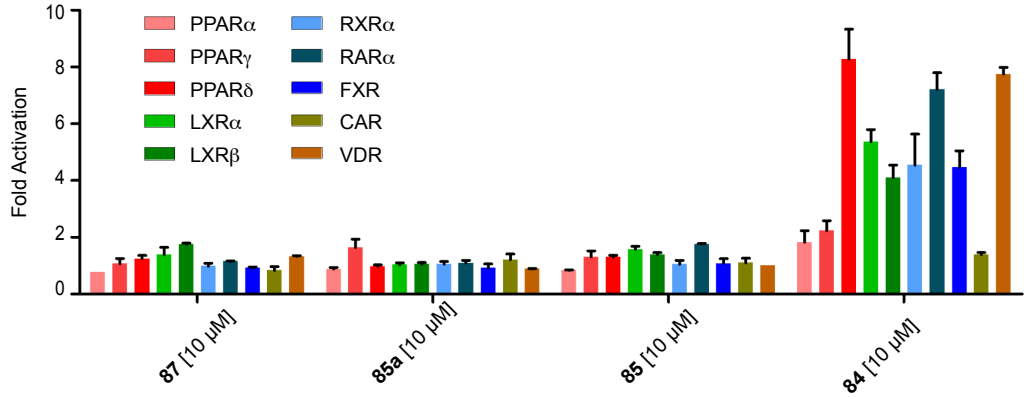



**Figure A.18:** In vitro characterization of marinopyrrole A (MP, **84**), **85**, **85a**, and **87** on nuclear receptors (PPAR = peroxisome proliferator-activated receptors; LXR = liver X receptors; RXR = retinoid X receptor; RAR = retinoic acid receptor; FXR = farnesoid X receptor; CAR = constitutive androstane receptor; VDR = vitamin D receptor). Marinopyrrole A (**84**) activated on several nuclear receptors with weak activation efficacy at a concentration of 10 µM.

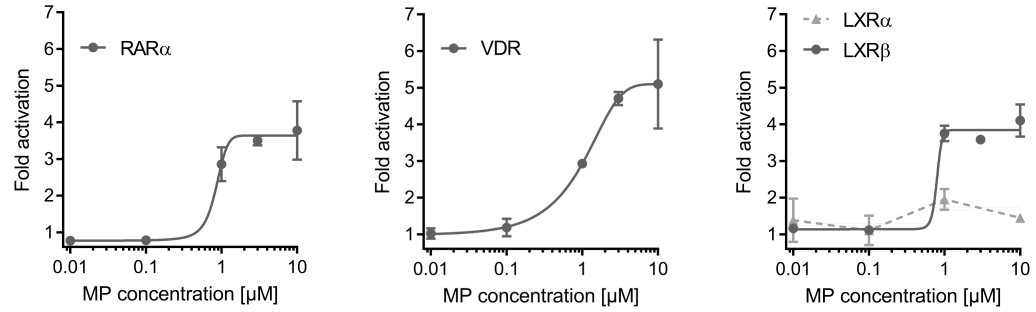



**Figure A.19:** Marinopyrrole A (**84**) activated RARα($EC_{50}$ = 0.63±0.11 µM), VDR ($EC_{50}$ = 1.09±0.06 µM and LXRβ ($EC_{50}$ = 0.37±0.40 µM) with low micromolar $EC_{50}$ values.

## A.3 Supplementary Information to Section 4.3

### Computational Results

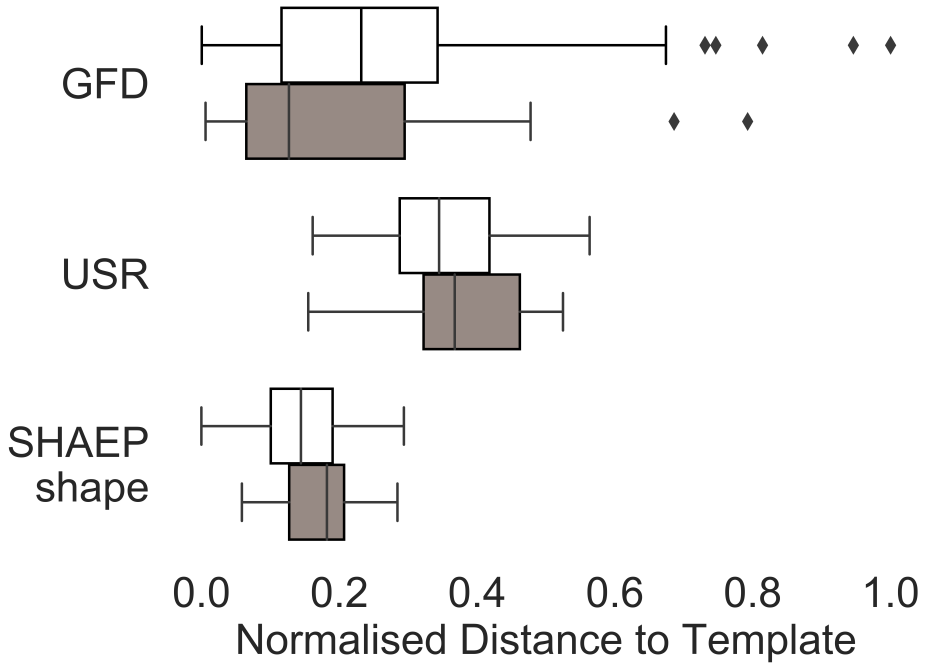
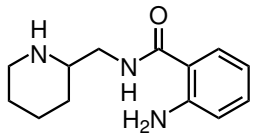



**Figure A.20:** Distance distributions for the set of unique *de novo* designs (n=323) against the (-)-englerin A template. ECFP4 (extended-connectivity fingerprint, *radius* = 2, 1024 bits) was used as the comparator. Here, global fractal dimension (GFD) represents max-scaled global FD distance, USR is the complement of the USR score[345], and SHAEP shape is the complement of the shape-comparison-only component of the SHAEP score.[344] For each description of the molecule set, the box-and-whisker diagram is separated into the top 10% (n=32) designs as ranked by ECFP4 similarity (in black), and the remaining molecules (in white). To quantify the enrichment, we performed a two-sided Mann-Whitney U (MWU) test with the null hypothesis in each instance that a value randomly chosen from the top 10% ECFP4 set has an equal likelihood of being higher or lower than a randomly chosen value from the bottom 90%. Calculated $p$ values are 0.03, 0.09, and 0.02 for GFD, USR, and SHAEP shape, respectively, so we reject the null hypothesis for GFD and SHAEP, indicating that the two distributions are separable.

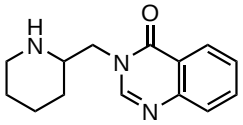
## Chemistry

**2-amino-*N*-[(piperidin-2-yl)methyl]benzamide (96)**



Piperidin-2-ylmethanamine (**94**, 0.65 ml, 5.19 mmol, 1.10 eq) was dissolved in ethanol (20 ml) and heated up to reflux. After the solution continuously reflux, 2H-3,1-benzoxazine-2,4(1H)-dione (**95**, 800 mg, 4.76 mmol, 1.00 eq) was added in one portion to the mixture. The mixture was stirred for 6 h after the mixture turned into a clear solution. After cool down to room temperature, the solution was acidified with aq. HCl solution (6 M, 20 ml) to pH 1 and extracted with EtOAc (3 × 50 ml). The acidic aqueous phase was alkalified with aqueous NaOH (1 M) until pH 9 was reached. The aqueous solution was extracted with EtOAc (5 × 200 ml), dried over $MgSO_4$, filtered and concentrated in vacuum. The crude product was purified by column chromatography using $CH_2Cl_2$/MeOH/$NEt_3$ (98:0:2 to 88:10:2) as mobile phase to yield the title compound **96** as yellowish solid (449 mg, 40%).
$^1$**H-NMR** (400 MHz, chloroform-*d*): $\delta$ = 7.36 (dd, *J* = 1.5, 7.9 Hz, 1H, C$H_{ar}$), 7.19 (ddd, *J* = 1.5, 7.2, 8.2 Hz, 1H, C$H_{ar}$), 6.70 - 6.58 (m, 3H, 2x C$H_{ar}$, C(=O)N*H*, 5.53 (s,2H, N$H_2$), 3.48 (ddd, *J* = 4.4, 5.6, 13.7 Hz, 1H, C(=O)NHC$H_2$), 3.25 (ddd, *J* = 5.4, 7.5, 13.6 Hz, 1H, C(=O)NHC$H_2$), 3.12 - 3.05 (m, 1H, NHC*H*), 2.79 (dddd, *J* = 2.7, 4.4, 7.3, 10.5 Hz, 1H, NHC$H_2$), 2.64 (td, *J* = 2.9, 11.9 Hz, 1H, NHC$H_2$), 1.96 - 1.76 (m, 3H, C$H_2$), 1.70 (dqd, *J* = 1.2, 3.2, 12.9 Hz, 1H, C$H_2$), 1.61 (dddd, *J* = 2.4, 3.9, 6.2, 7.5 Hz, 1H, C$H_2$), 1.48 - 1.31 (m, 2H, C$H_2$), 1.21 (tddd, *J* = 5.5, 6.7, 10.9, 12.5 Hz, 1H, C$H_2$); $^{13}$**C-NMR** (101 MHz, chloroform-*d*) $\delta$ = 132.2, 127.3, 117.3, 116.6, 55.9, 46.6, 45.2, 30.4, 26.4, 24.2; **HR-MS** (ESI-TOF): calculated for $C_{13}H_{20}N_3O$: *m/z* 234.1601, found: *m/z* 234.1603 $[M+H]^+$.

**3-[(piperidin-2-yl)methyl]quinazolin-4(3*H*)-one**

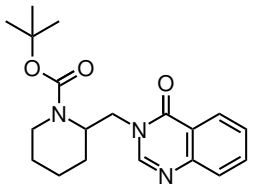


A suspension of **96** (119 mg, 0.51 mmol, 1.00 eq) in formic acid (5 ml, 129 mmol) was heated to 100 °Cfor 4 h. Molecular sieves (4 Å) were added and the solution was stirred for further 2 h. After cool down to room temperature, aqueous NaOH (1M, 25 ml) was added and the aqueous phase was extracted with EtOAc (3 × 15 ml). The combined organic layer was dried over $MgSO_4$, and concentrated under reduced pressure. The crude product was purified by column chromatography $CH_2Cl_2$/MeOH/$NEt_3$ (98:0:2

to 88:10:2) as mobile phase to yield the title compound as yellowish solid (103 mg, 83%).

$^1$**H-NMR** (400 MHz, chloroform-*d*): $\delta$ = 8.31 (dd, *J* = 1.4, 7.9 Hz, 1H, NC*H*$_{ar}$N), 8.05 (s, 1H, C*H*$_{ar}$), 7.81 - 7.67 (m, 2H, C*H*$_{ar}$), 7.50 (ddd, *J* = 1.5, 6.9, 8.2 Hz, 1H, C*H*$_{ar}$), 4.07 (dd, *J* = 4.8, 13.5 Hz, 1H, C(=O)NHC*H*$_2$), 3.77 (dd, *J* = 7.9, 13.4 Hz, 1H, C(=O)NHC*H*$_2$), 3.09 - 2.97 (m, 2H, N*H*,NHC*H*), 2.63 (q, *J* = 7.3 Hz, 2H, NHC*H*$_2$), 1.87 - 1.78 (m, 1H, C*H*$_2$), 1.74 (dtd, *J*= 1.5, 3.1, 14.3 Hz, 1H, C*H*$_2$), 1.60 (dtd, *J* = 1.5, 3.2, 3.7, 9.5 Hz, 1H, C*H*$_2$), 1.49 - 1.29 (m, 2H, C*H*$_2$), 1.26 - 1.13 (m, 1H, C*H*$_2$). $^{13}$**C-NMR** (101 MHz, chloroform-*d*) $\delta$ = 169.7, 148.7, 132.3, 127.7, 117.3, 116.6, 56.2, 46.1, 44.4, 29.5, 25.4, 23.8; **HR-MS** (ESI-TOF): calculated for $C_{14}H_{18}N_3O$: *m/z* 244.1444, found: *m/z* 244.1446 $[M+H]^+$.

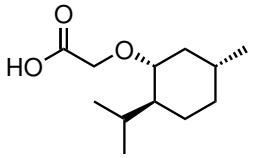
***tert*-butyl 2-[(4-oxoquinazolin-3(4*H*)-yl)methyl]piperidine-1-carboxylate (92)**



3-[(piperidin-2-yl)methyl]quinazolin-4(3*H*)-one (93 mg, 0.38 mmol, 1.00 eq) was dissolved in anhydrous $CH_2Cl_2$ (2 ml). $NEt_3$ (0.2 ml, 1.43 mmol, 3.75 eq) was added and the solution was stirred for 15 min at 0 °C. Di-*tert*-butyl dicarbonate (0.11 ml, 0.46 mmol, 1.20 eq) was added dropwise to the mixture over a period of 5 min. After 15 min the ice bath was removed, and the solution was stirred at room temperature for 5 h. The mixture was diluted with water (15 ml) and extracted with EtOAc (3 × 25 ml). The organic phases were combined, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified with column chromatography using EtOAc/hexane (83:17) as mobile phase to yield the title compound **92** as yellowish solid (64 mg, 49%).

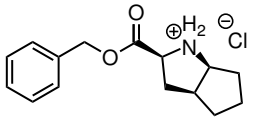
$^1$**H-NMR** (400 MHz, chloroform-*d*): $\delta$ = 8.30 (dd, *J* = 1.4, 7.9 Hz, 1H, NC*H*$_{ar}$N), 7.94 (s, 1H, C*H*$_{ar}$), 7.78 - 7.65 (m, 2H, C*H*$_{ar}$), 7.55 - 7.45 (m, 1H, C*H*$_{ar}$), 4.67 (m, 1H, NHC*H*), 4.36 - 3.91 (m, 3H, C(=O)NHC*H*$_2$, NHC*H*$_2$), 2.97 (m, 1H, NHC*H*$_2$), 1.81 - 1.53 (m, 5H, C*H*$_2$), 1.53 - 1.34 (m, 1H, C*H*$_2$), 1.07 (s, 9H, C*H*$_3$); $^{13}$**C-NMR** (101 MHz, chloroform-*d*) $\delta$ = 204.7, 161.1, 146.4, 134.2, 127.2, 126.8, 79.8, 27.9, 25.2, 19.5; **HR-MS** (ESI-TOF): calculated for $C_{19}H_{26}N_3O_3$: *m/z* 344.1969, found: *m/z* 344.1975 $[M+H]^+$.

**[(1*R*,2*S*,5*R*)-5-methyl-2-(propan-2-yl)cyclohexyl]oxyacetic acid (99)**



Sodium hydride (605 mg, 15.13 mmol, 3.00 eq), potassium iodide (113 mg, 0.67 mmol, 0.14 eq) in anhydrous THF (20 ml) were stirred at 0 °Cfor 15 min. (*1R,2S,5R*)-2-isopropyl-5-methylcyclohexan-1-ol (**97**) (1.87 g, 11.85 mmol, 2.40 eq) in anhydrous THF (15 ml) was added dropwise over 10 min to the slurry mixture. After 30 min, 2-chloroacetic acid (472 mg, 4.95 mmol, 1.00 eq) in THF (15 ml) was added slowly over 10 min to the slurry. The mixture was stirred vigorously and heated up to reflux for 16 h. The reaction was allowed to cool down to room temperature and the solvent was removed under reduced pressure. Water (20 ml) and toluene (20 ml) were added and separated. The aqueous layer was acidified to pH 2-3 and extracted with EtOAc (3 × 50 ml). The combined organic phases were washed with brine, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified with column chromatography using hexane/EtOAc (100:0 to 75:25) as mobile phase to yield the title compound **99** as colorless oil (191 mg, 18%).
$^1$**H-NMR** (400 MHz, chloroform-*d*): $\delta$ = 4.19 (d, *J* = 16.4 Hz, 1H, C(=O)C$H_2$O), 4.06 (d, *J* = 16.4 Hz, 1H, C(=O)C$H_2$O), 3.24 (td, *J* = 4.2, 10.7 Hz, 1H, OC*H*), 2.17 (m,1H, $CH_2CHCH_3$), 2.05 (dtd, *J* = 1.8, 3.8, 12.0 Hz, 1H, CHC*H*()$(CH_3)_2$), 1.66 (dtd, *J* = 3.1, 6.4, 10.5 Hz, 2H, CHCHC$H_2$), 1.45 - 1.22 (m, 2H), 1.06 - 0.82 (m, 9H), 0.79 (d, *J* = 7.0 Hz, 3H,C$H_3$); $^{13}$**C-NMR** (101 MHz, chloroform-*d*) $\delta$ = 81.0, 65.6, 48.1, 40.1, 34.4, 31.6, 25.9, 23.3, 22.3, 21.1, 16.3; **HR-MS** (ESI-TOF): calculated for $C_{12}H_{22}ONa$: *m/z* 237.1461, found: *m/z* 237.1461 $[M+Na]^+$; $[\alpha]_D^{20}$ : -94.3° (c = 2.00, ethanol), (lit. -95.1°).[391]

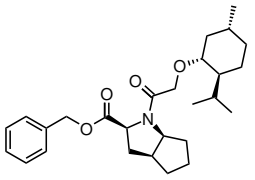
**(2*S*,3a*S*,6a*S*)-2-[(benzyloxy)carbonyl]octahydrocyclopenta[*b*]pyrrol-1-ium chloride (100)**



(*2S,3aS,6aS*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid (**98**, 86 mg, 0.52 mmol, 1.00 eq) was suspended in a hydrochloric acid solution ( 4 M in dioxane, 0.25 ml). Benzyl alcohol (2 ml, 19.13 mmol, 36.00 eq) was added and cooled down to 0-5 °C. Thionyl chloride (0.15 ml, 2.04 mmol, 3.90 eq) was added dropwise and the mixture was stirred for 16 h. The mixture was poured into methyl *tert*-butyl ether (50 ml) and title compound **100** was filtered off as white solid (103 mg, 70%). $^1$**H-NMR** (400 MHz, methanol-$d_4$): $\delta$ = 7.47 - 7.31 (m, 5H, C$H_{ar}$), 5.35 - 5.22 (m, 2H, $C_{ar}CH_2O$), 4.44 (dd, *J* = 6.9, 11.2 Hz, 1H, C(=O)C*H*NH), 4.12 (ddd, *J* = 2.9, 7.6, 8.8 Hz, 1H, NHC*H*), 2.97 (tdd, *J* = 2.9,

5.0, 8.6 Hz, 1H, NHCHC*H*), 2.64 (ddd, *J* = 6.9, 8.7, 13.5 Hz, 1H, CHC$H_2$CH), 2.05 - 1.81 (m, 2H), 1.83 - 1.65 (m, 4H), 1.62 - 1.46 (m, 1H); $^{13}$**C-NMR** (101 MHz, methanol-$d_4$) $\delta$ = 135.0, 128.4, 128.4, 67.8, 64.8, 61.1, 42.2, 33.7, 30.9, 29.7, 23.9; **HR-MS** (ESI-TOF): calculated for $C_{1520}NO_2$: *m/z* 246.1489, found: *m/z* 246.1490 $[M+H]^+$; $[\alpha]_D^{20}$ : -41.5° (c = 1.00, $H_2O$), (lit. -40.0°).[392]

**Benzyl(2*S*,3a*S*,6a*S*)-1-(2-(((1*R*,2*S*,5*R*)-2-isopropyl-5-methylcyclohexyl)oxy)-acetyl)-octahydrocyclopenta[*b*]pyrrole-2-carboxylate (93)**



**99** (40 mg, 0.19 mmol, 1.29 eq) and **100** (41 mg, 0.14 mmol, 1.00 eq) were suspended in THF (0.6 ml). After stirring for 5 min, $NEt_3$ (30 µl, 0.22 mmol, 1.49 eq) and 1H-benzo[*d*][1,2,3]triazol-1-ol hydrate (35 mg, 0.23 mmol, 1.59 eq) were added. The reaction mixture was cooled down to 0 °C. After 30 min 3-(((ethylimino)methylene)-amino)-*N*,*N*-dimethylpropan-1-amine (40.0 µl, 0.23 mmol, 1.58 eq) was added slowly to the reaction mixture. The mixture was allowed to warm up to room temperature and stirred for 16 h. The reaction solvent was removed under reduced pressure. The residue was redissolved in EtOAc (20 ml) and the organic phase was washed with aqueous hydrochloric acid solution (0.1 M, 3 × 15 ml), saturated $NaHCO_3$ (3 × 15 ml) and brine (25 ml). The organic phase was dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified with column chromatography using hexane/EtOAc (66:33) as mobile phase to yield the title compound **93** as white solid (55 mg, 86%). $^1$**H-NMR** (500 MHz, chloroform-*d*) (ratio of cis/trans rotamers 1.5:1) $\delta$ = 7.50 – 7.28 (m, 5H, C$H_{ar}$), 5.21 – 5.11 (m, 2H, $C_{ar}CH_2O$), 4.81 (dd, *J* = 4.3, 9.7 Hz, 1H, C(=O)C*H*NH), 4.40 (dt, *J* = 5.1, 8.1 Hz, 1H, C(=O)C*H*NH), 4.21 – 3.88 (m, 2H, C(=O)C$H_2$O), 3.14 (td, *J* = 4.1, 10.5 Hz, 1H, OC*H*), 2.81 – 2.52 (m, 1H, NHCHC*H*), 2.46 – 2.32 (m, 1H,C(=O)NHC$s_2$), 2.26 (m, 1H, $CH_2CHCH_3$), 2.18 – 1.99 (m, 3H, C(6)$H_2$, $CH_2CHCH_3$), C(6')$H_2$), 1.87 – 1.68 (m, 3H, C(6)$H_2$, C(5)$H_2$, C(4)$H_2$), 1.67 – 1.47 (m, 3H, C(6)$H_2$, C(3')$H_2$, C(4')$H_2$), 1.45 – 1.26 (m, 3H, C(5)$H_2$, C(4)$H_2$, OCHC*H*, CHC*H*$(CH_3)_2$), 1.20 – 1.06 (m, 1H, C(4)$H_2$), OCHC*H*), 1.01 – 0.81 (m, 9H, C(3')$H_2$), C(4')$H_2$), $CH_2CHCH_3$), $OCHCH_2$), 0.78 (d, *J* = 7.0 Hz, 2H, CHCH($CH_3$)$_2$), 0.70 (d, *J* = 7.0 Hz, 1H, C(6')$H_2$, CHCH($CH_3$)$_2$); $^{13}$**C-NMR** (125 MHz, chloroform-*d*) $\delta$ = 172.2, 168.6, 135.8, 128.6, 128.5, 128.1, 80.3, 79.3, 68.8, 68.0, 67.2, 66.8, 65.0, 63.1, 60.7, 59.6, 48.3, 48.1, 44.1, 40.9, 40.1, 39.4, 35.3, 34.4, 34.4, 33.6, 33.4, 33.3, 32.1, 31.5, 31.3, 25.7, 25.6, 25.4, 25.2, 23.2, 23.1, 22.3, 21.0, 16.2, 16.1; **HR-MS** (ESI-TOF): calculated for $C_{27}H_{40}NO_4$: *m/z* 442.2952, found: *m/z* 442.2950; $[\alpha]_D^{20}$ : -67.1° (c = 1.01, chloroform)

## X-ray Structure of 93

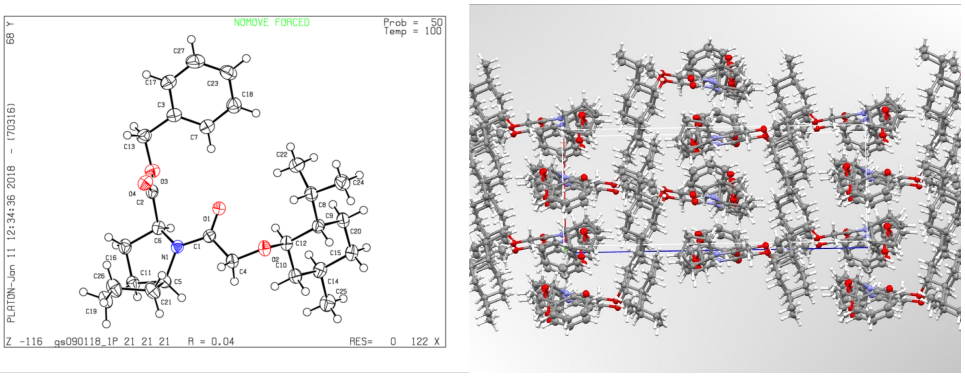



**Figure A.21:** X-ray structure (*left*) and crystal packing (*right*) of compound **93**.

## *In Vitro* Characterization

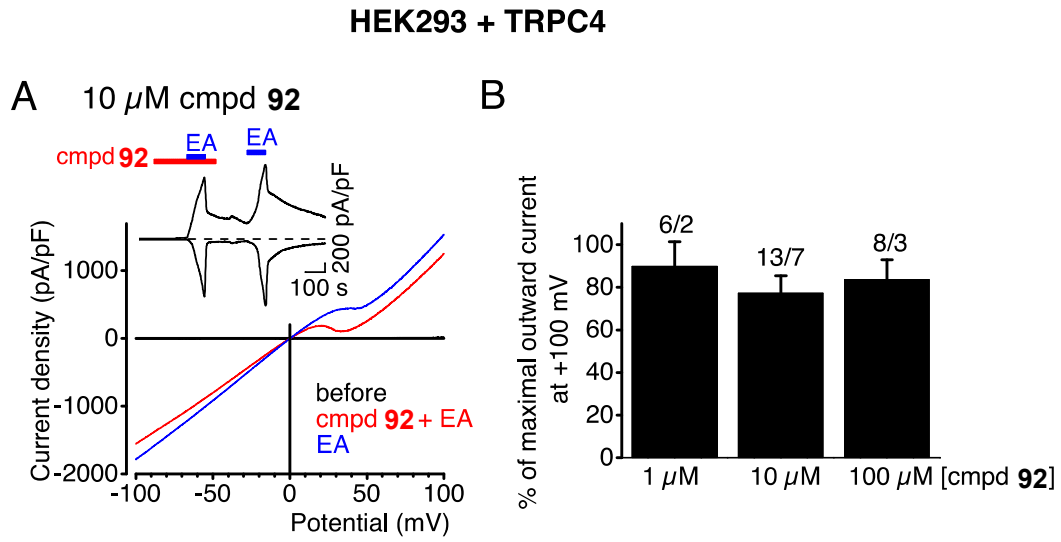



**Figure A.22:** Compound **92** is a very weak inhibitor of TRPC4 channels. Electrophysiological whole-cell measurements of TRPC4 expressing HEK293 cells. **A**, Representative current density-voltage curves before application of the first stimulus (black traces), during the first application of 50 nM (-)-Englerin A ('EA') in the presence of 10 µM compound **92** ('cmpd **92** + EA', red trace) and during the second application of 50 nM (-)-Englerin A (blue trace) are displayed. Insets show current density time courses at holding potentials of ±100 mV with indicated application of compound **92** (red bar) and of (-)-Englerin A (blue bars). **B**, Summary of maximal (-)-Englerin A-induced outward currents in the presence of indicated concentrations of compound **92**. Numbers indicate the number of measured cells and the number of experimental days. To percentage of the maximal outward current at +100 mV elicited by the second application of 50 nM (-)-Englerin A was used for normalization.

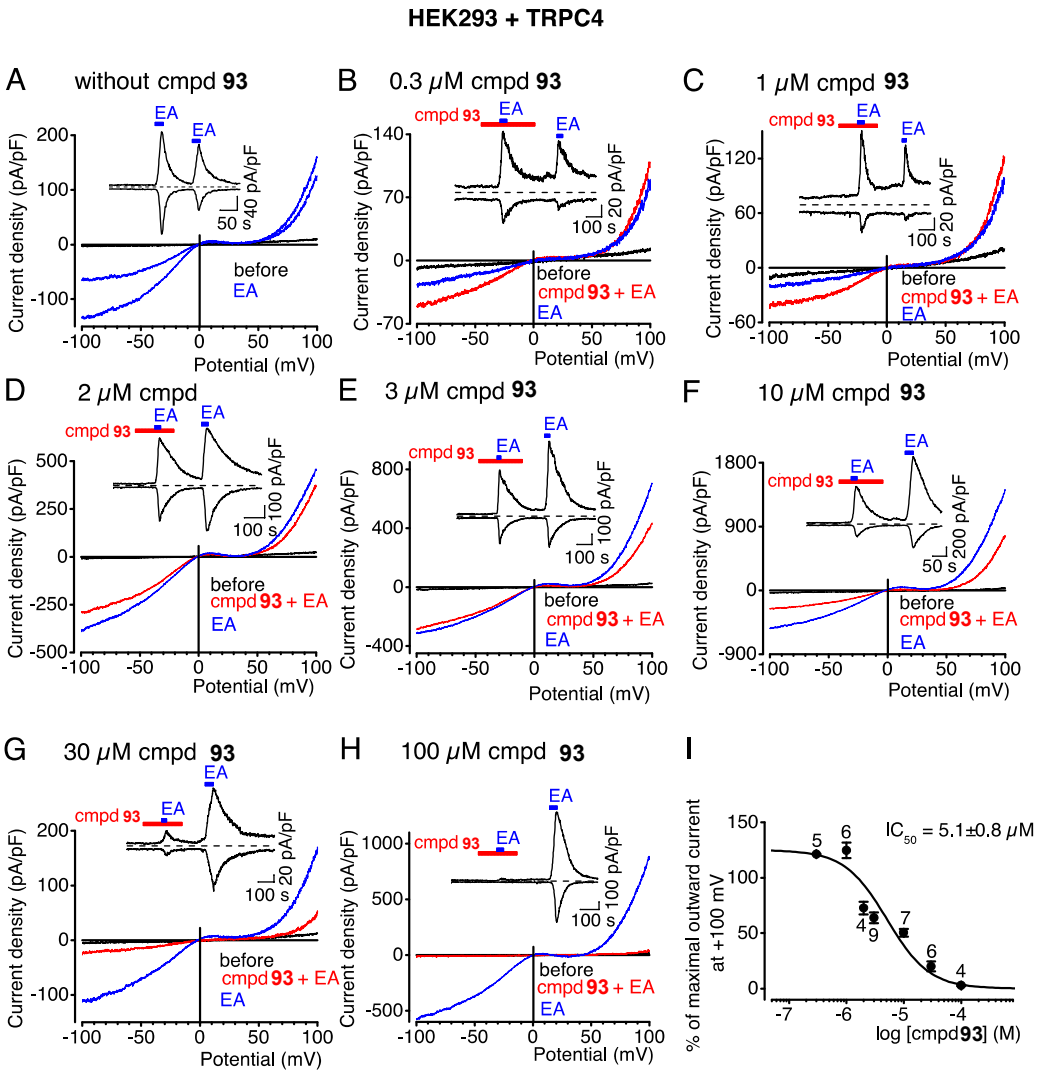



**Figure A.23:** Compound **93** is an inhibitor of TRPC4 channels. Electrophysiological whole-cell measurements of TRPC4 expressing HEK293 cells. (**A**-**H**) Representative current density-voltage curves before application of the first stimulus (black traces), during the first application of 50 nM (-)-Englerin A (**62**, 'EA') in the absence (blue traces) (A) or in the presence (**B**-**H**) of indicated concentrations of compound **93** ('cmpd **93** + EA', red traces) and during the second application of 50 nM (-)-Englerin A (blue traces) (**A**-**H**) are displayed. Insets show current density time courses at holding potentials of ±100 mV with indicated applications of compound **93** (red bars) and of (-)-Englerin A (blue bars). (**I**) Summary of maximal (-)-Englerin A-induced outward currents in the presence of indicated concentrations of compound **93**. Numbers indicate the number of measured cells from at least three independent experiments. To determine the $IC_{50}$ value, the percentage of the maximal outward current at +100 mV elicited by the second application of 50 nM (-)-Englerin A was used.

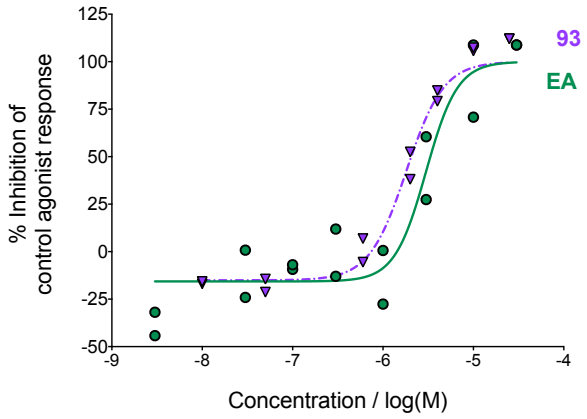



**Figure A.24:** Inhibitory effects of compound **93** and the natural product (-)-Englerin A (**62**, EA) on TRPM8 cation channels in an intracellular calcium assay. Compound **93** and EA showed inhibitory effects on TRPM8 at a screening concentration of 10 µM. In a follow-up characterization, the $IC_{50}$ values of **93** and EA were determined ($IC_{50}$ = 1.8±1.1 µM for **93** and ($IC_{50}$ = 3.0±1.2 µM for EA[196]). The dissociation constants were obtained as $K_i = IC_{50}[1+(C/EC_{50.C})]^{-1}$, where C is the concentration of control activator icilin (0.1 µM) in the assay and $EC_{50.C}$ its $EC_{50}$ value (0.016 µM).

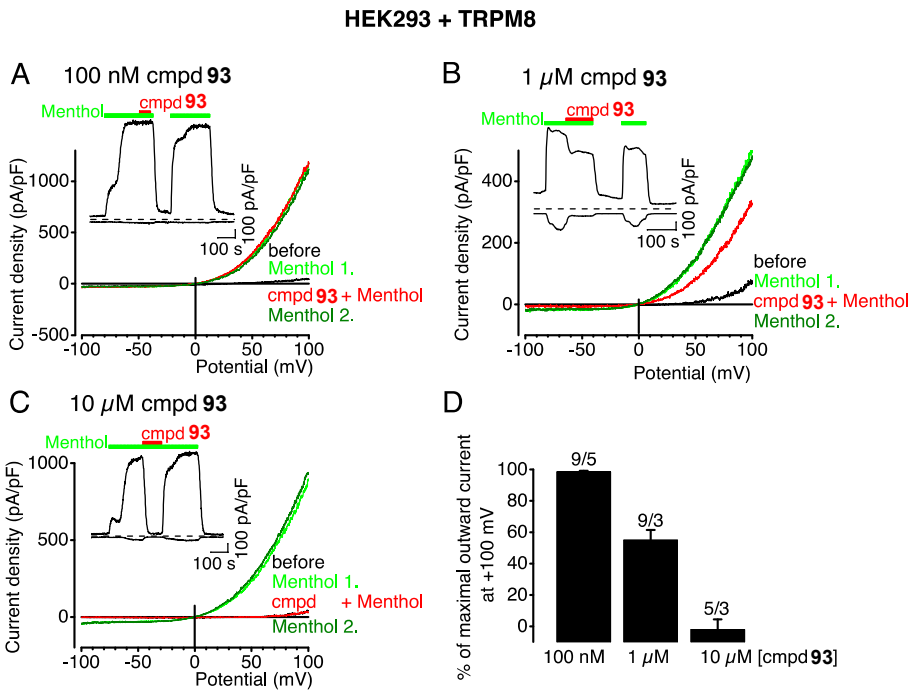



**Figure A.25:** Compound **93** blocks TRPM8 currents in a concentration dependent manner. Electrophysiological whole-cell measurements of TRPM8 expressing HEK293 cells. **A-C**, Representative current density-voltage curves before application of the first stimulus (black traces), during first application of 200 µM (-)-menthol (**97**, 'Menthol 1.', light green traces) and during application of indicated concentrations of **93** in the presence of (-)-menthol ('cmpd **93** + Menthol', red traces). **A**,**B**, The dark green lines indicate current density-voltage curves during second application of 200 µM (-)-menthol after washout of (-)-menthol and compound **93** ('Menthol 2.', dark green traces). **C**, The dark green line indicates current density-voltage curve after washout of compound **93** in the presence of (-)-menthol ('Menthol 2.', dark green traces). **A-C**, Insets show current density time courses at holding potentials of ±100 mV with indicated applications of compound **93** (red bars) and of (-)-menthol (green bars). **D**, Summary of maximal (-)-menthol induced outward currents in the presence of indicated concentrations of compound **93**. Numbers indicate the numbers of measured cells and the numbers of independent experiments. The percentage of the maximal outward current at +100 mV elicited by the first application of 200 µM (-)-menthol was used for normalization.

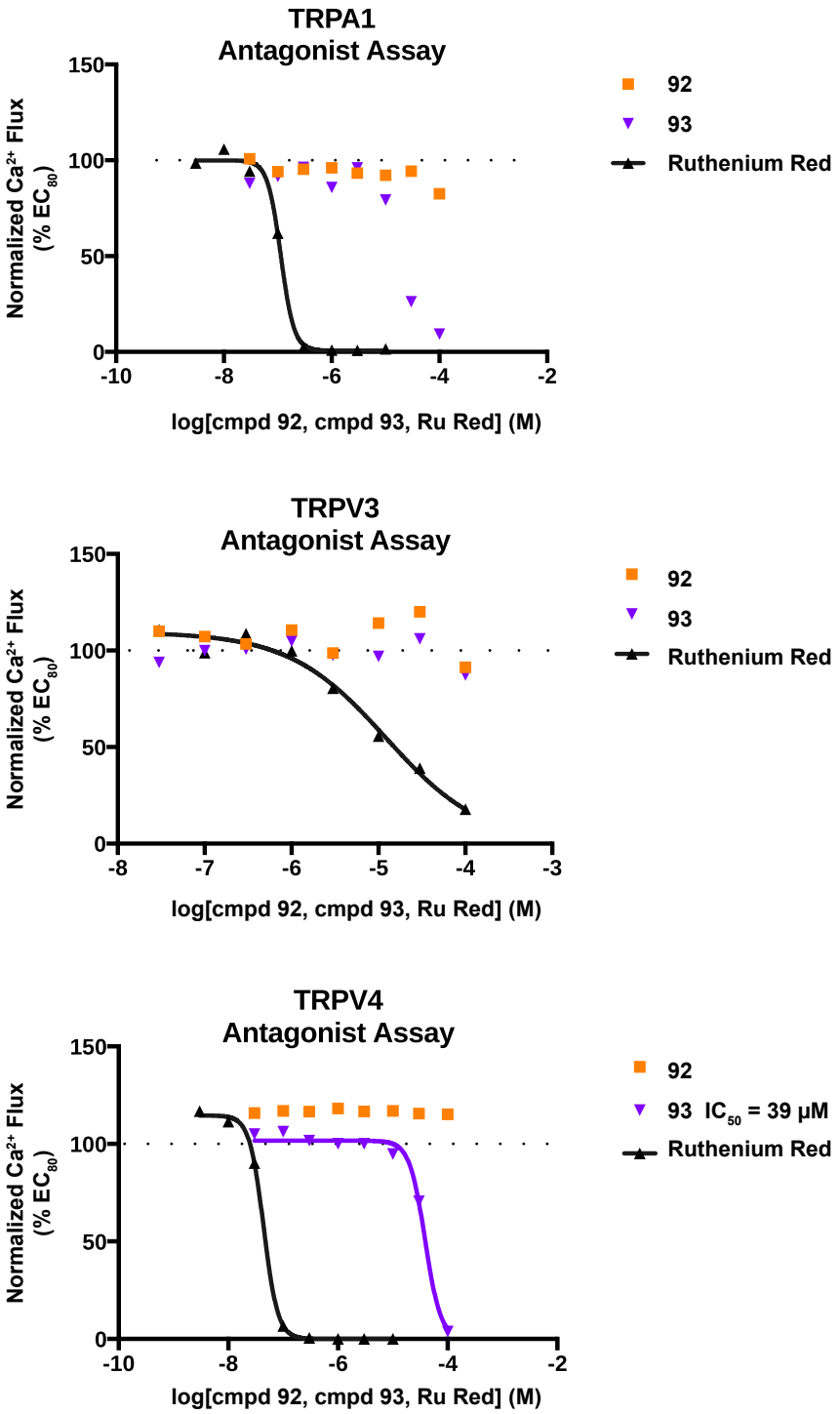



**Figure A.26:** Compound **93** is a weak inhibitor of TRPV4. Compound **92** and **93** were tested on their inhibitory effects in calcium assays of TRPA1 (*top*), TRPV3 (*middle*) and TRPV4 (*bottom*). Compound **92** showed no activity below 100 µM in all three assays. During the TRPA1 assay measurements, a calcium increase during the pre-incubation time was observed for compound **93** at concentrations 10, 30 and 100 µM. This desensitization of the receptor might impact the measures of the inhibitor assay of TRPA1. **93** inhibited TRPV4 in a concentration dependent manner with an $IC_{50}$ = 39±1 µM.

# A.4 $^{1}$H- and $^{13}$C-NMR spectra

## NMR Spectra of Compounds from Section 4.1

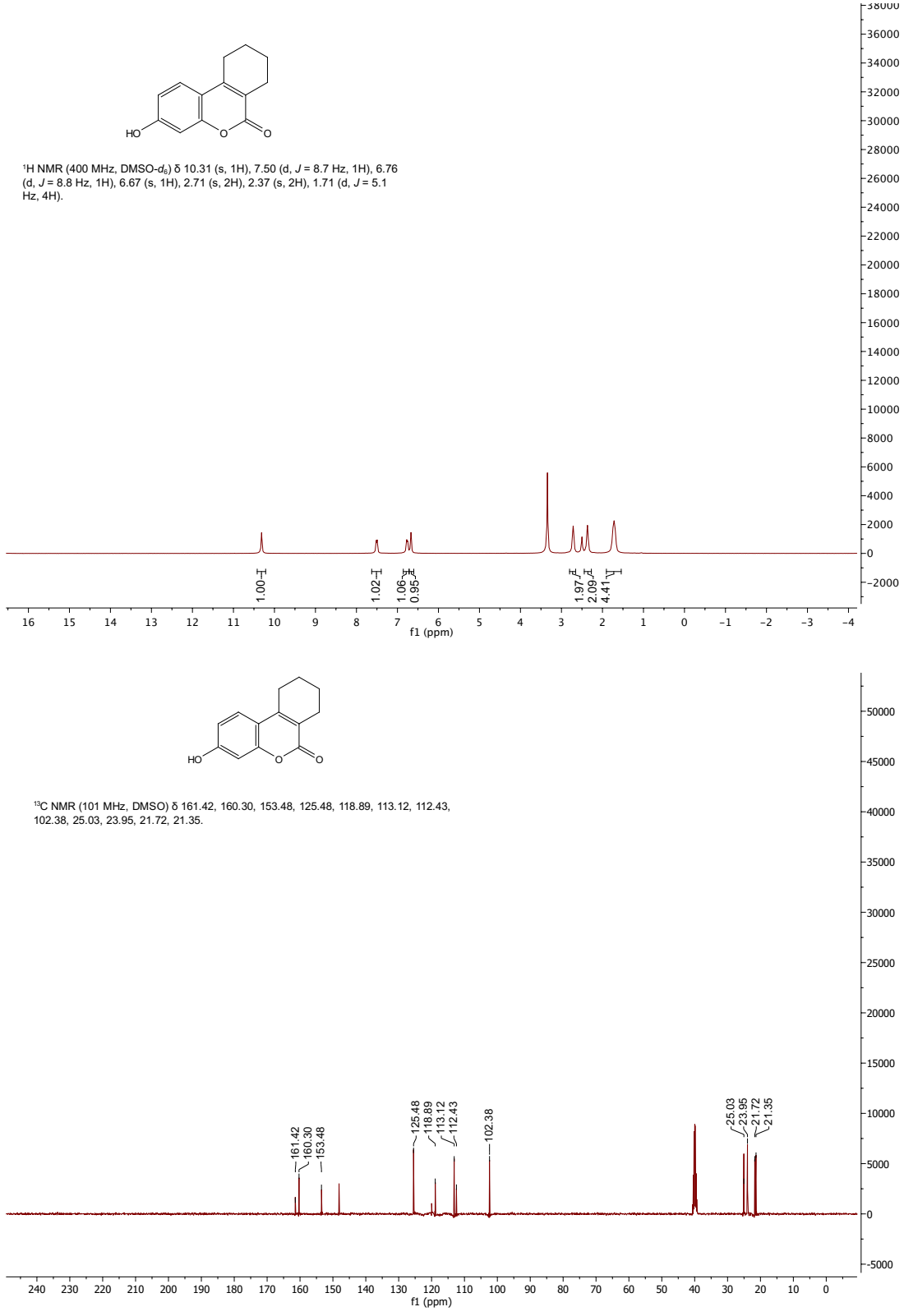



**Figure A.27:** (*Top*) $^{1}$**H-NMR** of compound **72**. (*Bottom*) $^{13}$**C-NMR** of compound **72**.

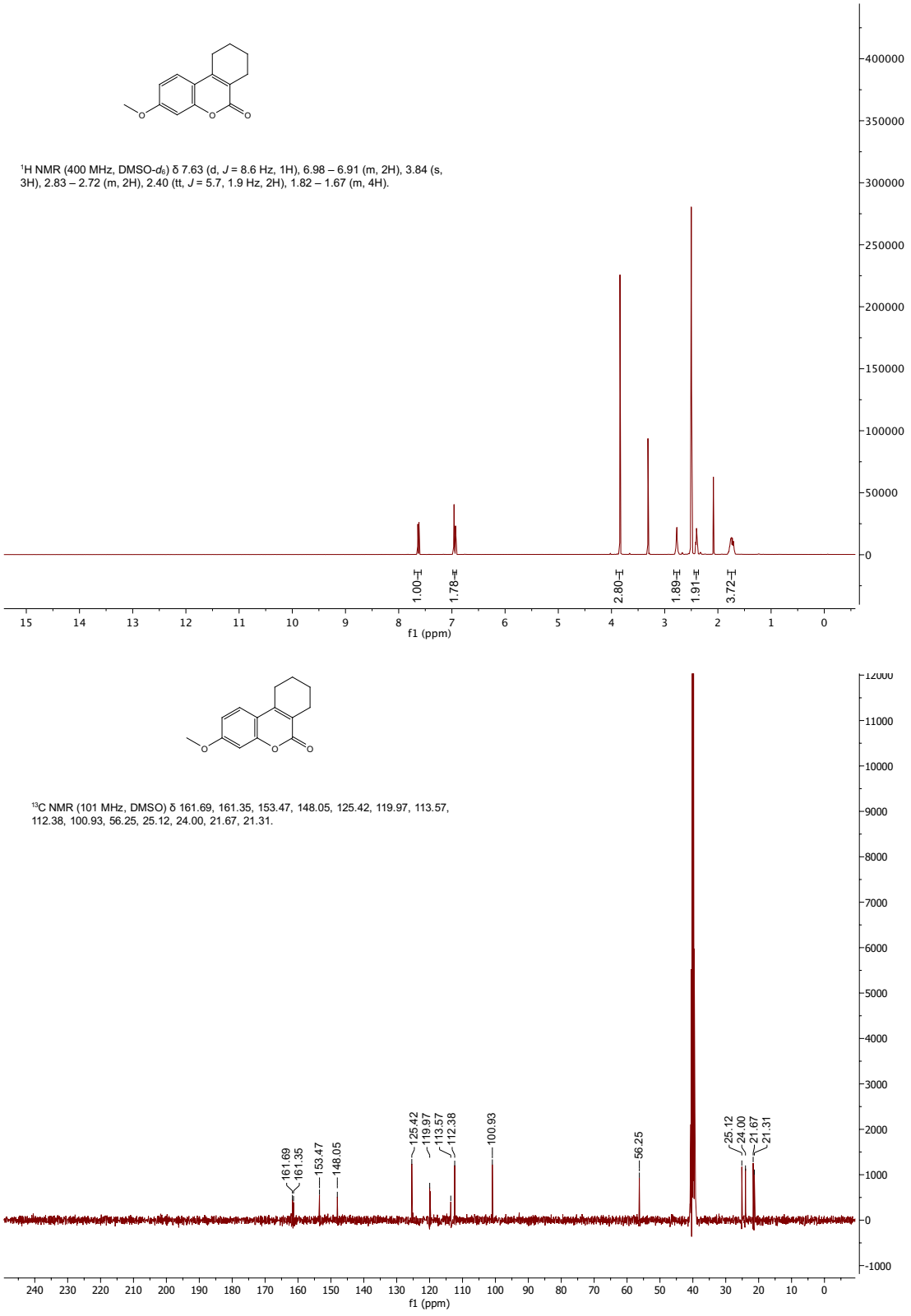



**Figure A.28:** (*Top*) $^{1}$**H-NMR** of compound **73**. (*Bottom*) $^{13}$**C-NMR** of compound **73**.

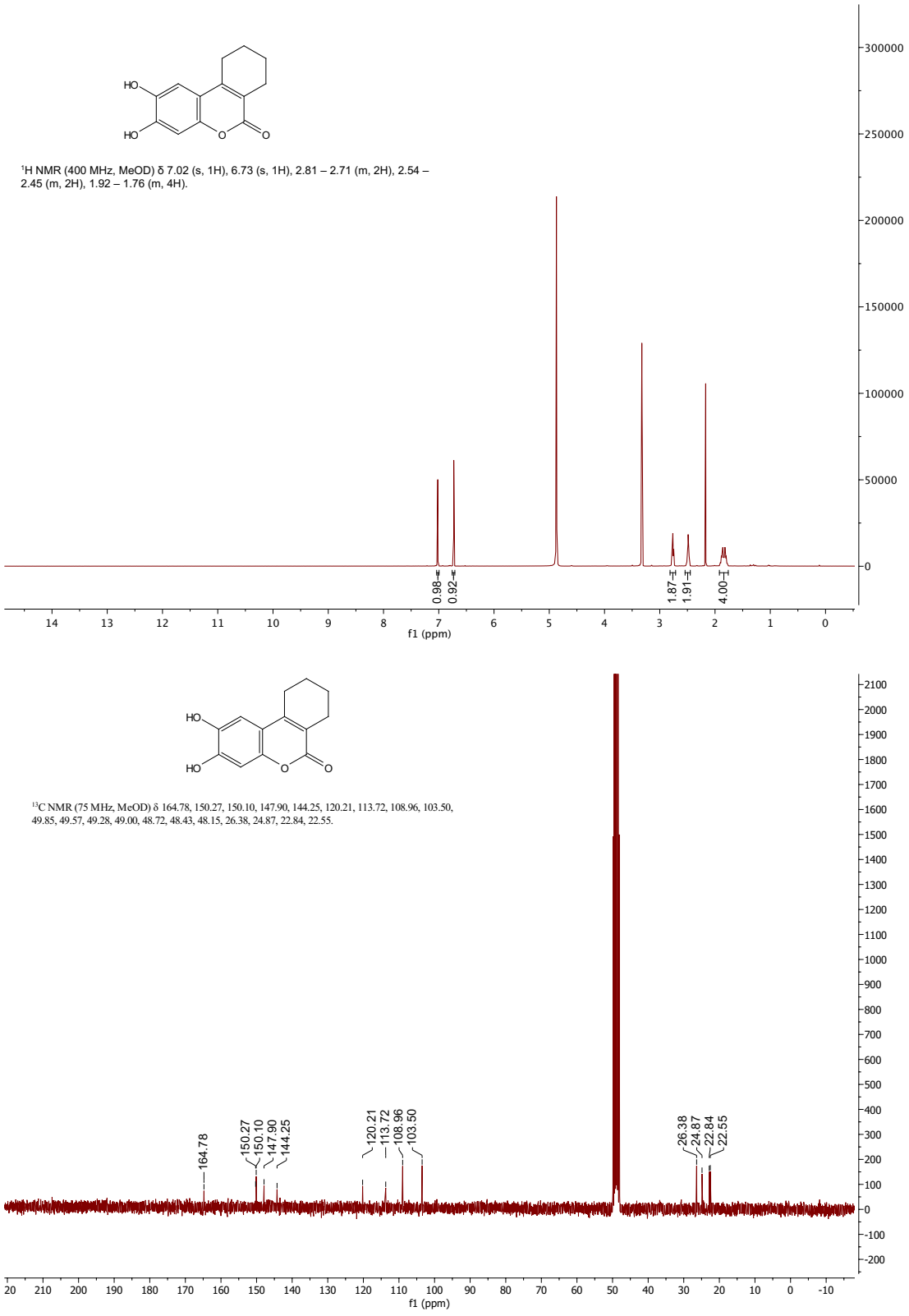



**Figure A.29:** (*Top*) $^{1}$**H-NMR** of compound **74**. (*Bottom*) $^{13}$**C-NMR** of compound **74**.


**Figure A.30:** (*Top*) $^{1}$**H-NMR** of compound **75**. (*Bottom*) $^{13}$**C-NMR** of compound **75**.

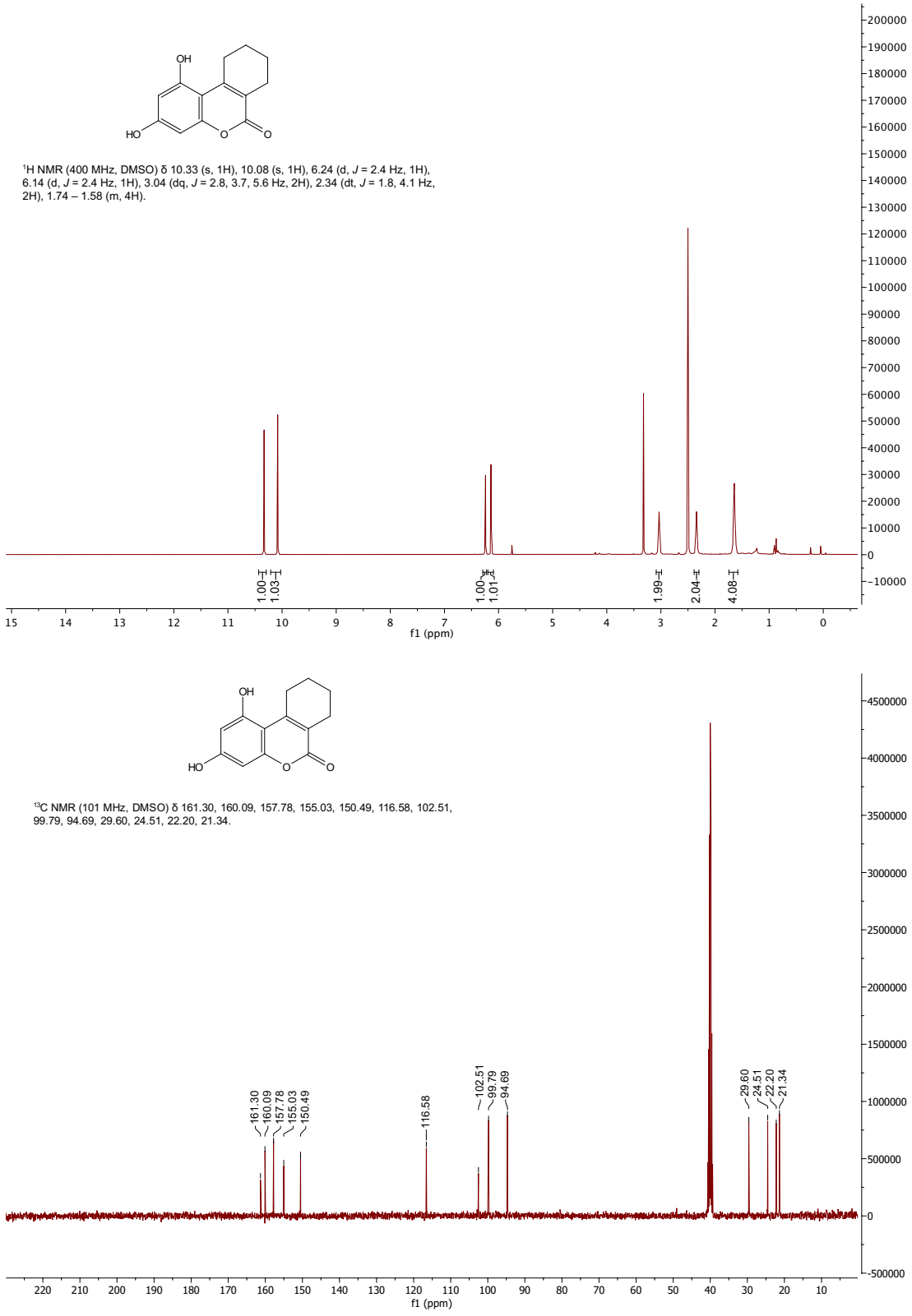



**Figure A.31:** (*Top*) $^{1}$**H-NMR** of compound **76**. (*Bottom*) $^{13}$**C-NMR** of compound **76**.

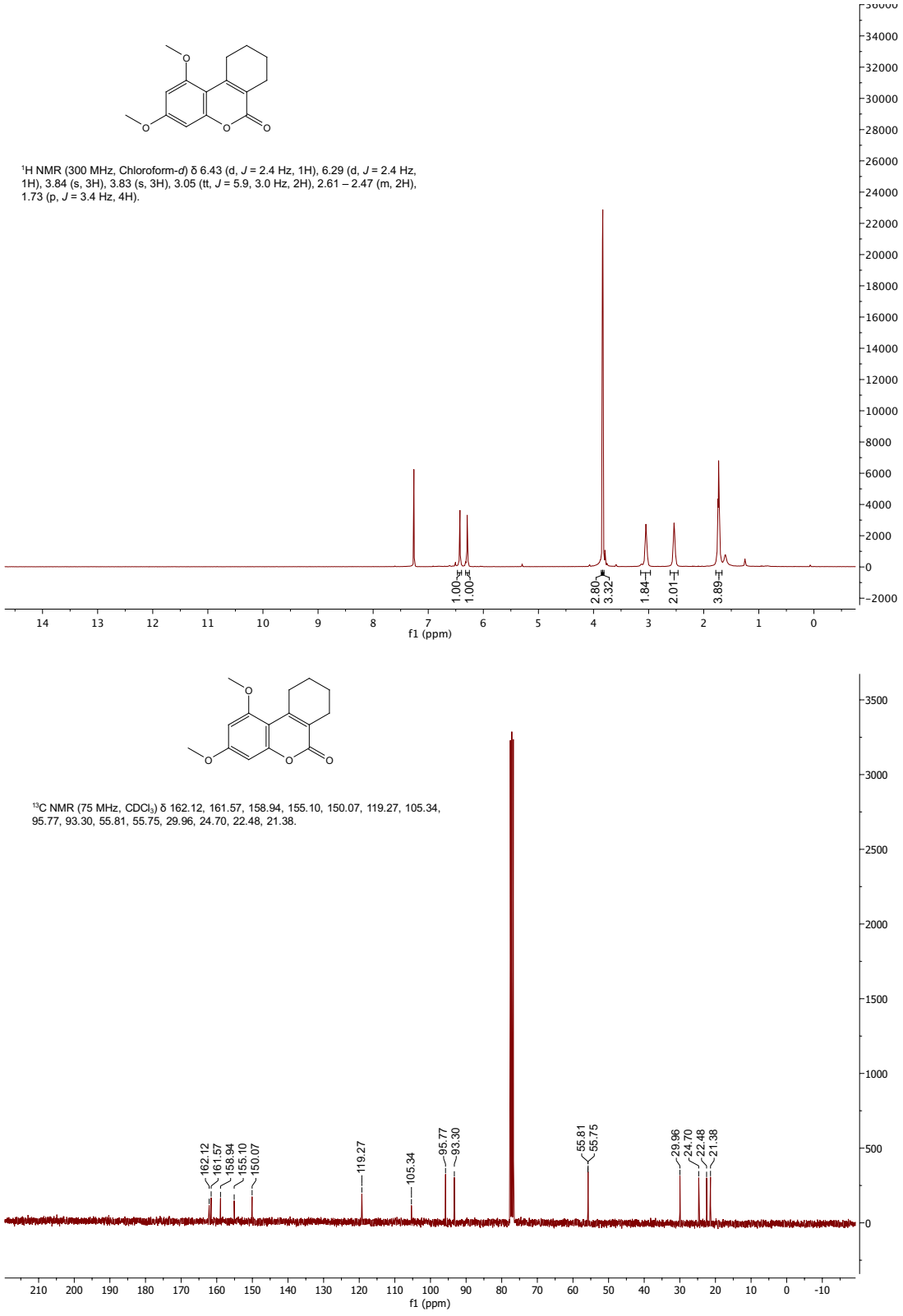



**Figure A.32:** (*Top*) $^{1}$**H-NMR** of compound **77**. (*Bottom*) $^{13}$**C-NMR** of compound **77**.

# NMR Spectra of Compounds from Section 4.2

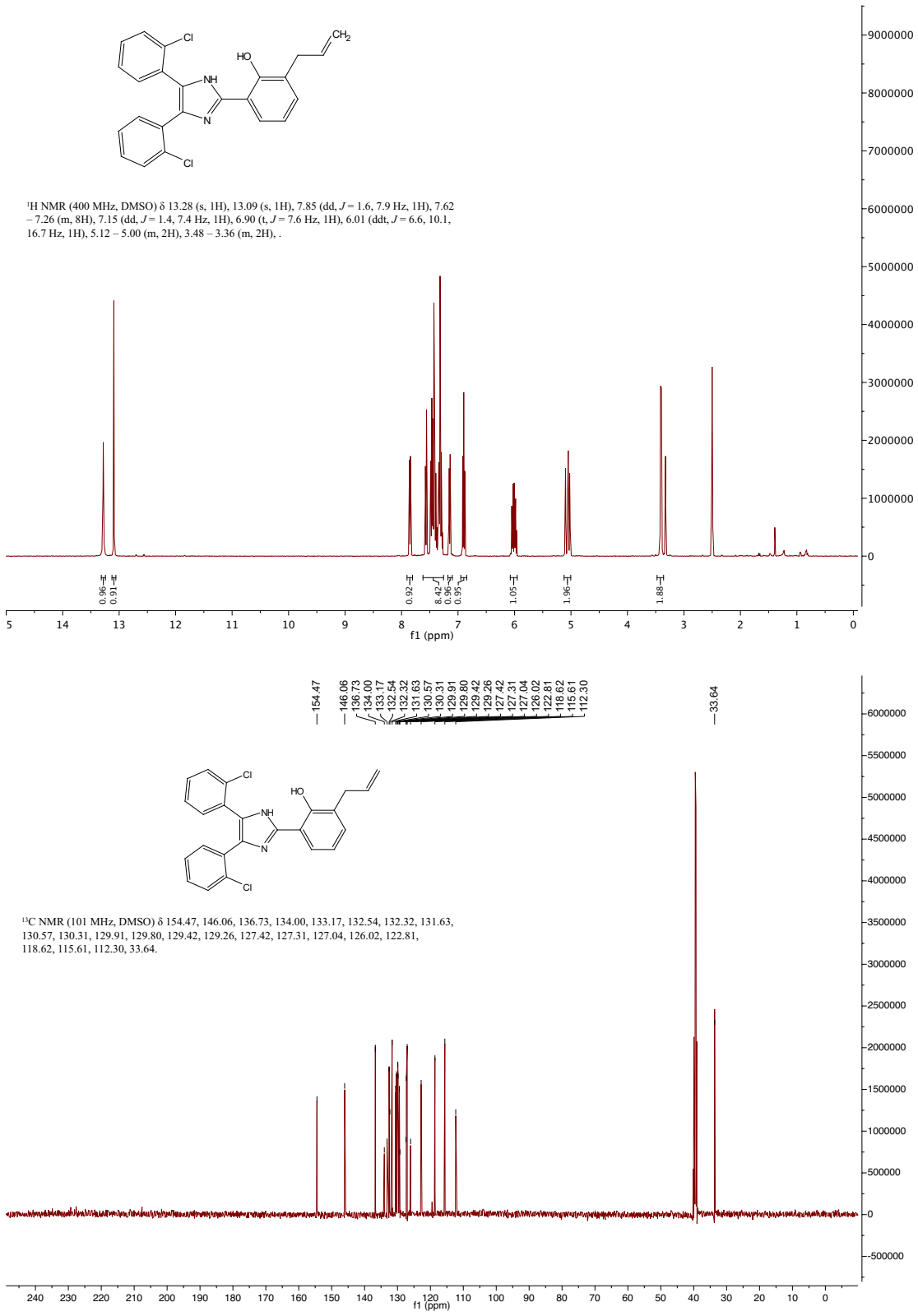



**Figure A.33:** (*Top*) $^{1}$**H-NMR** of compound **87**. (*Bottom*) $^{13}$**C-NMR** of compound **87**.

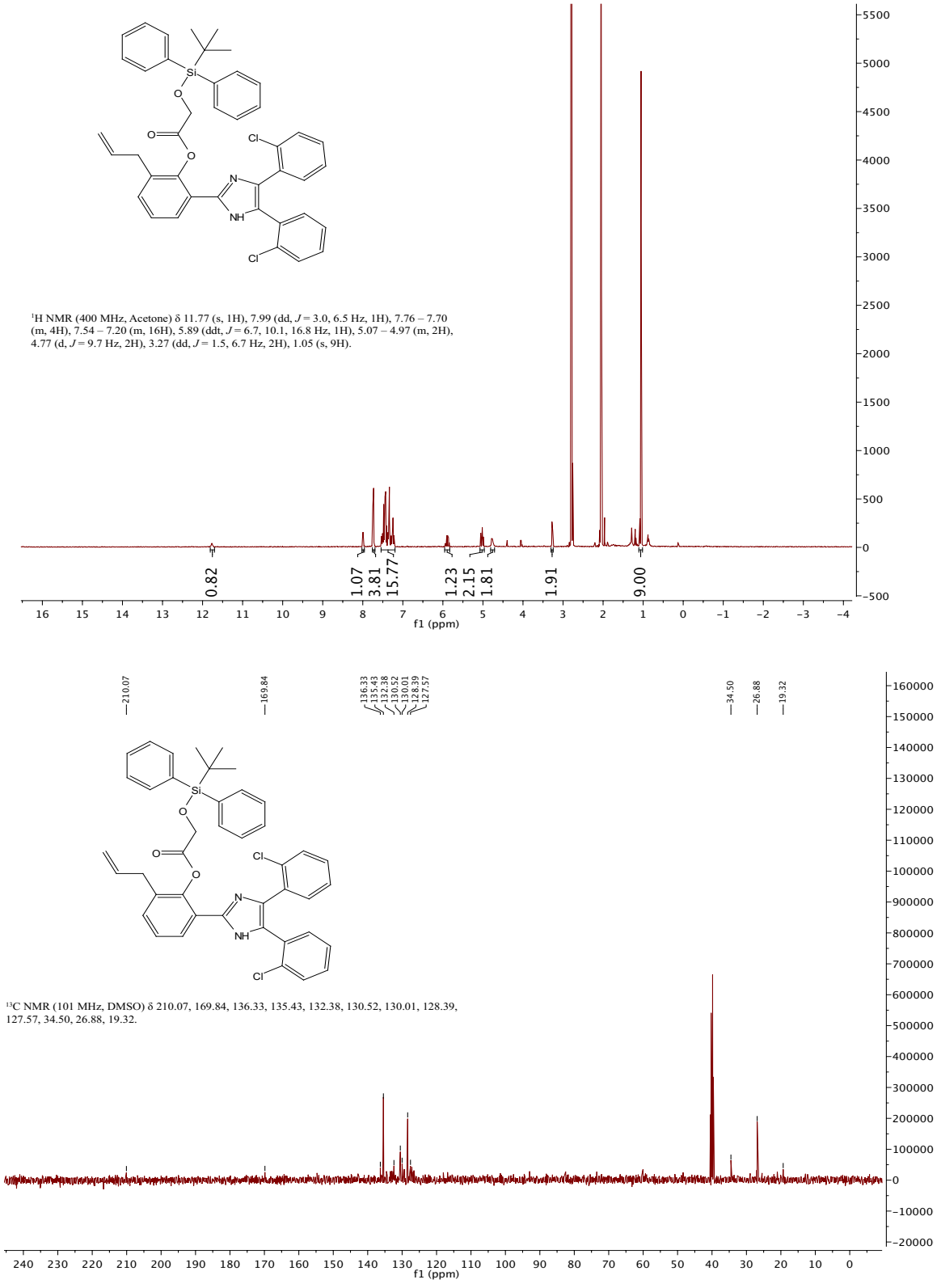



**Figure A.34:** (*Top*) **$^{1}$H-NMR** of 2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl-2-((*tert*-butyldiphenyl-silyl)oxy) acetate. (*Bottom*) **$^{13}$C-NMR** of 2-Allyl-6-(4,5-bis(2-chlorophenyl)-1*H*-imidazol-2-yl)phenyl-2-((*tert*-butyldiphenyl-silyl)oxy) acetate.

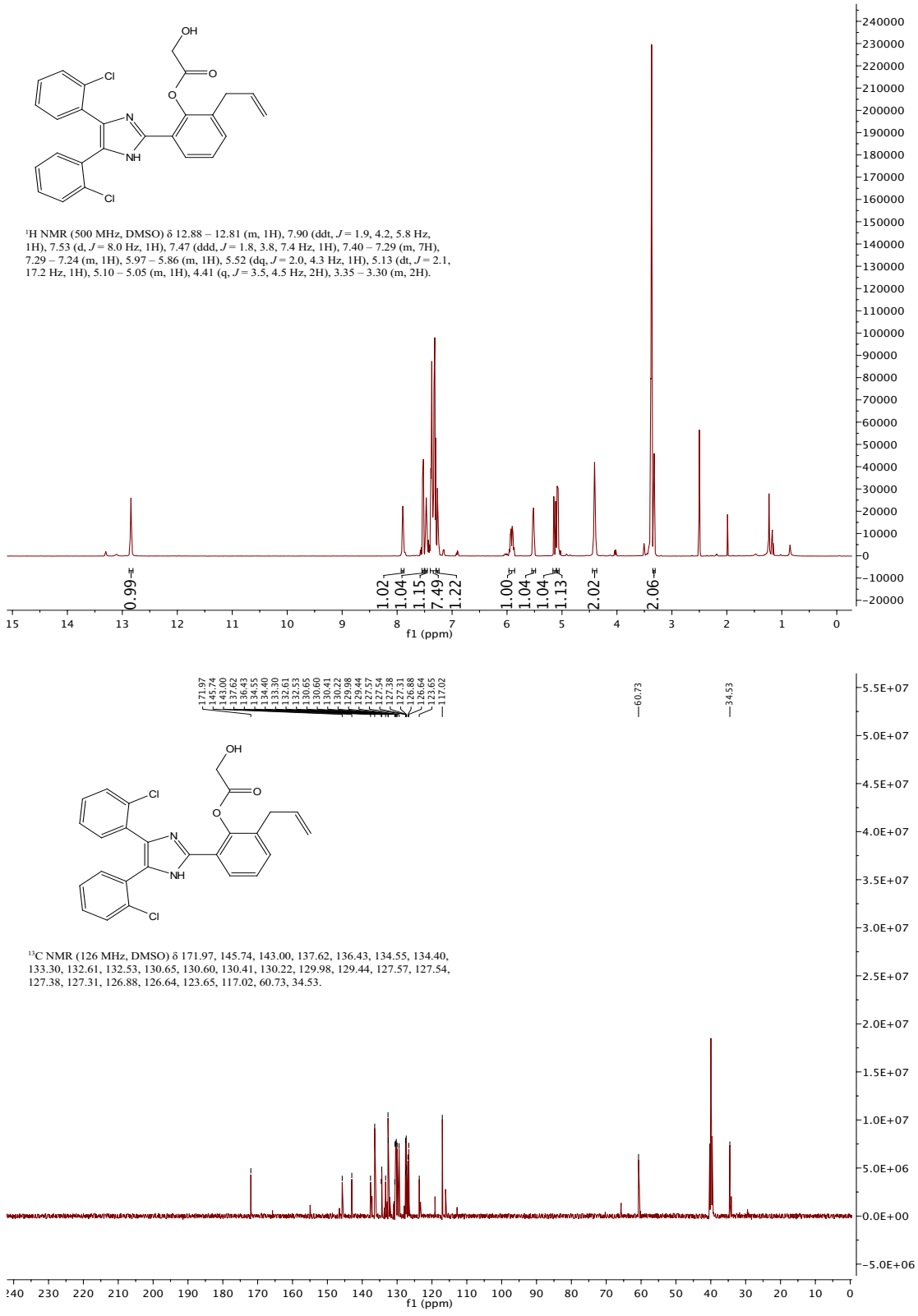



**Figure A.35:** (*Top*) $^{1}$**H-NMR** of compound **85**. (*Bottom*) $^{13}$**C-NMR** of compound **85**.

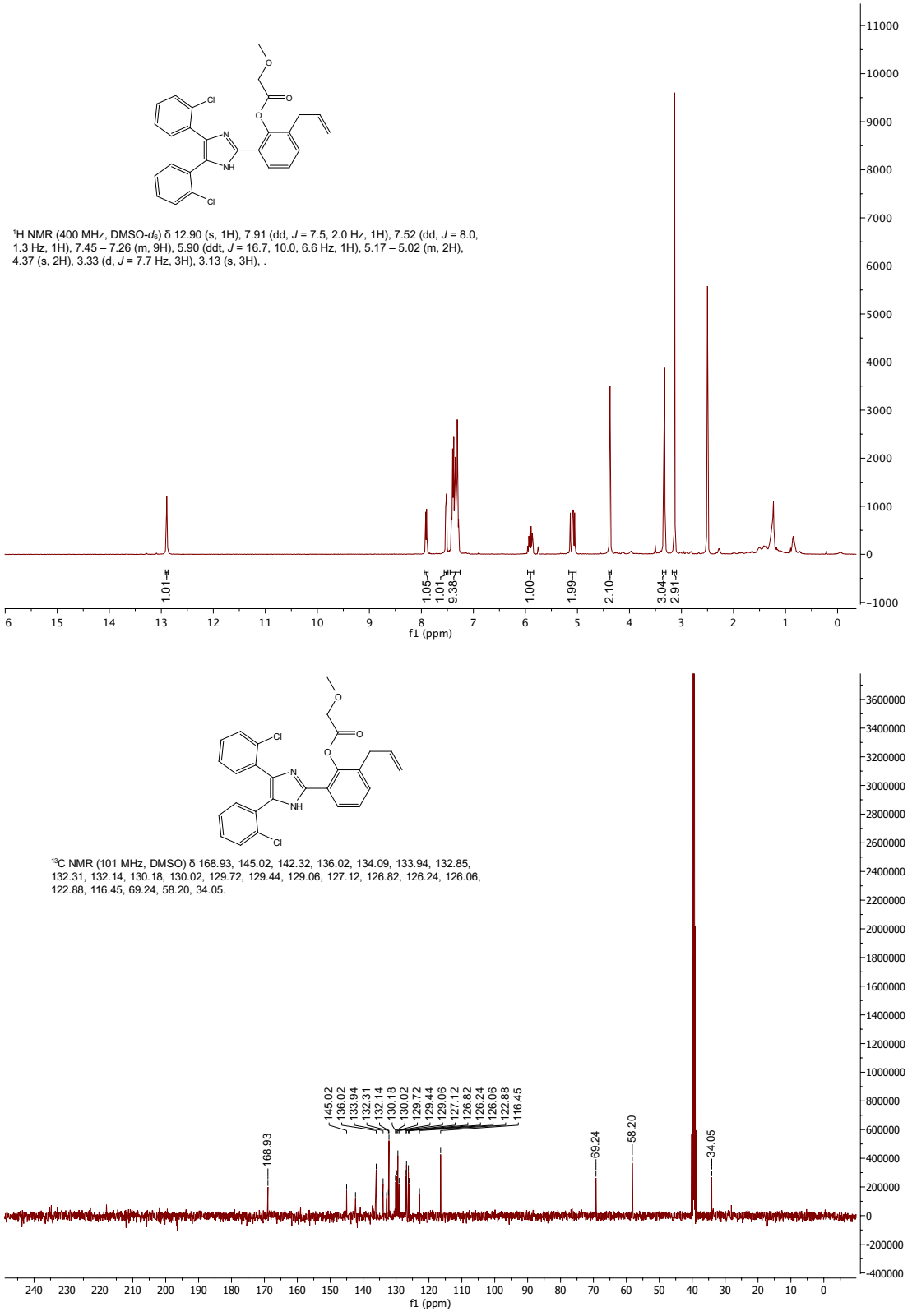



**Figure A.36:** (*Top*) $^{1}$**H-NMR** of compound **85a**. (*Bottom*) $^{13}$**C-NMR** of compound **85a**.

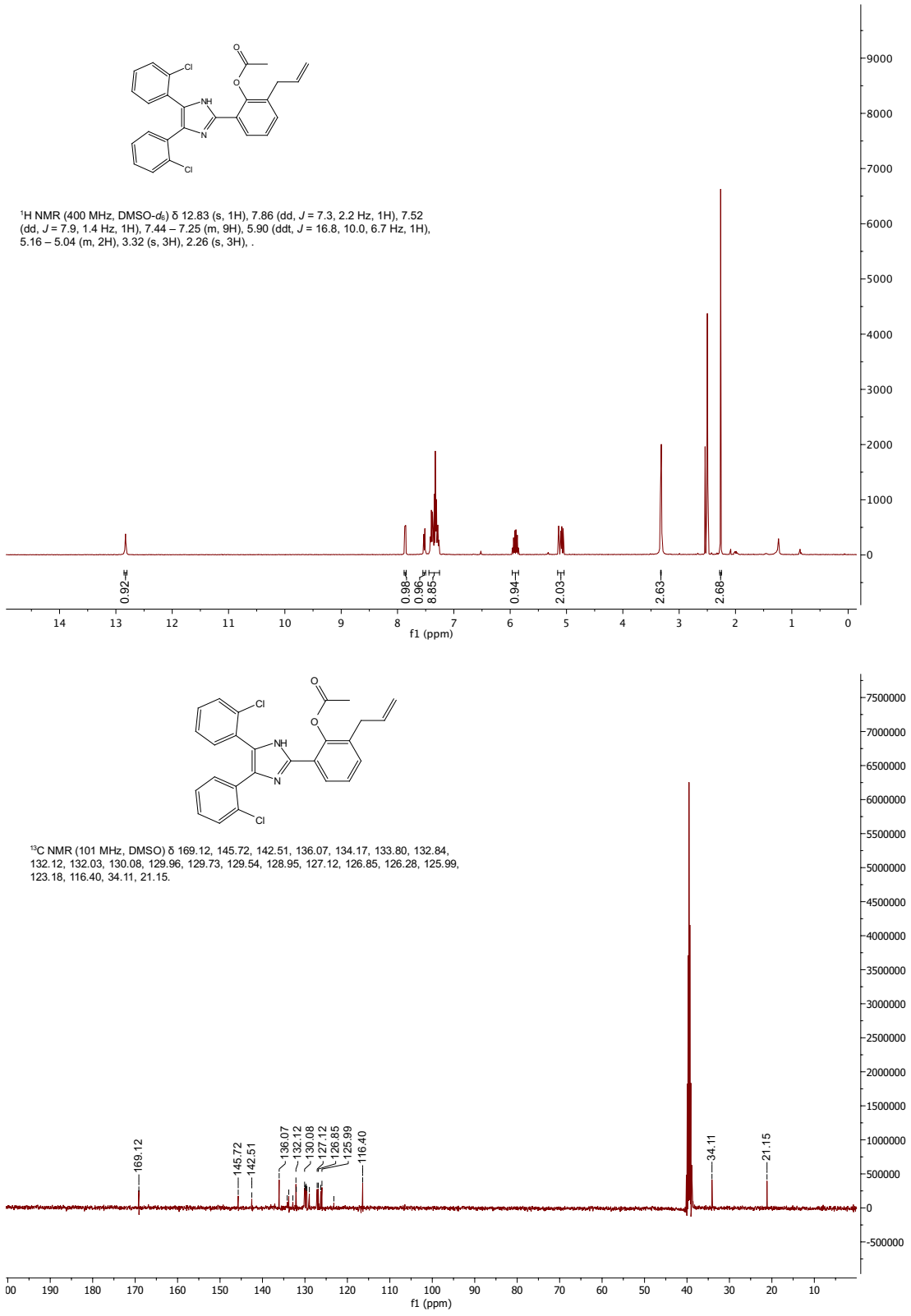



**Figure A.37:** (*Top*) $^{1}$**H-NMR** of compound **85b**. (*Bottom*) $^{13}$**C-NMR** of compound **85b**.

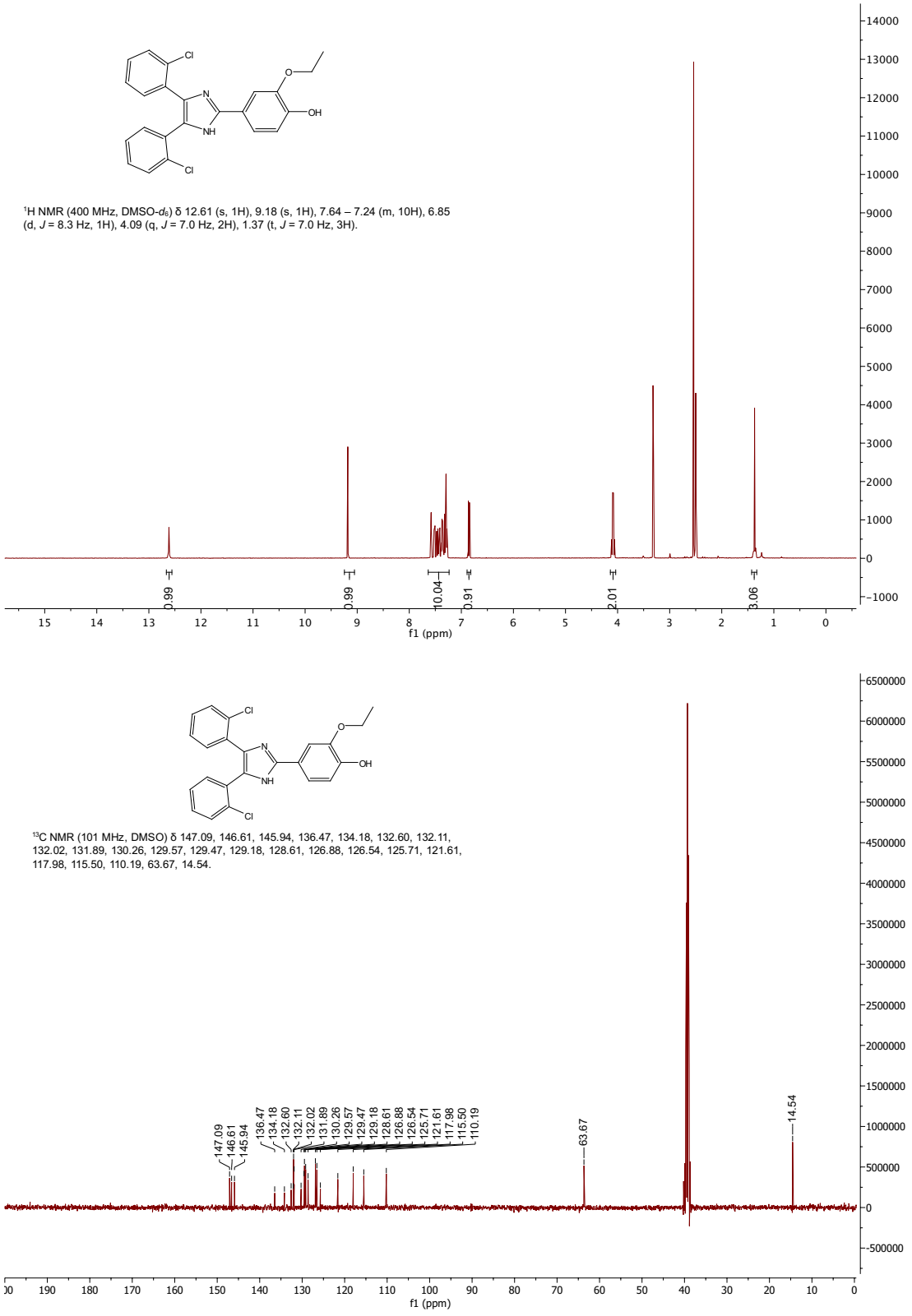



**Figure A.38:** (*Top*) $^1$**H-NMR** of compound **86**. (*Bottom*) $^{13}$**C-NMR** of compound **86**.

# NMR Spectra of Compounds from Section 4.3

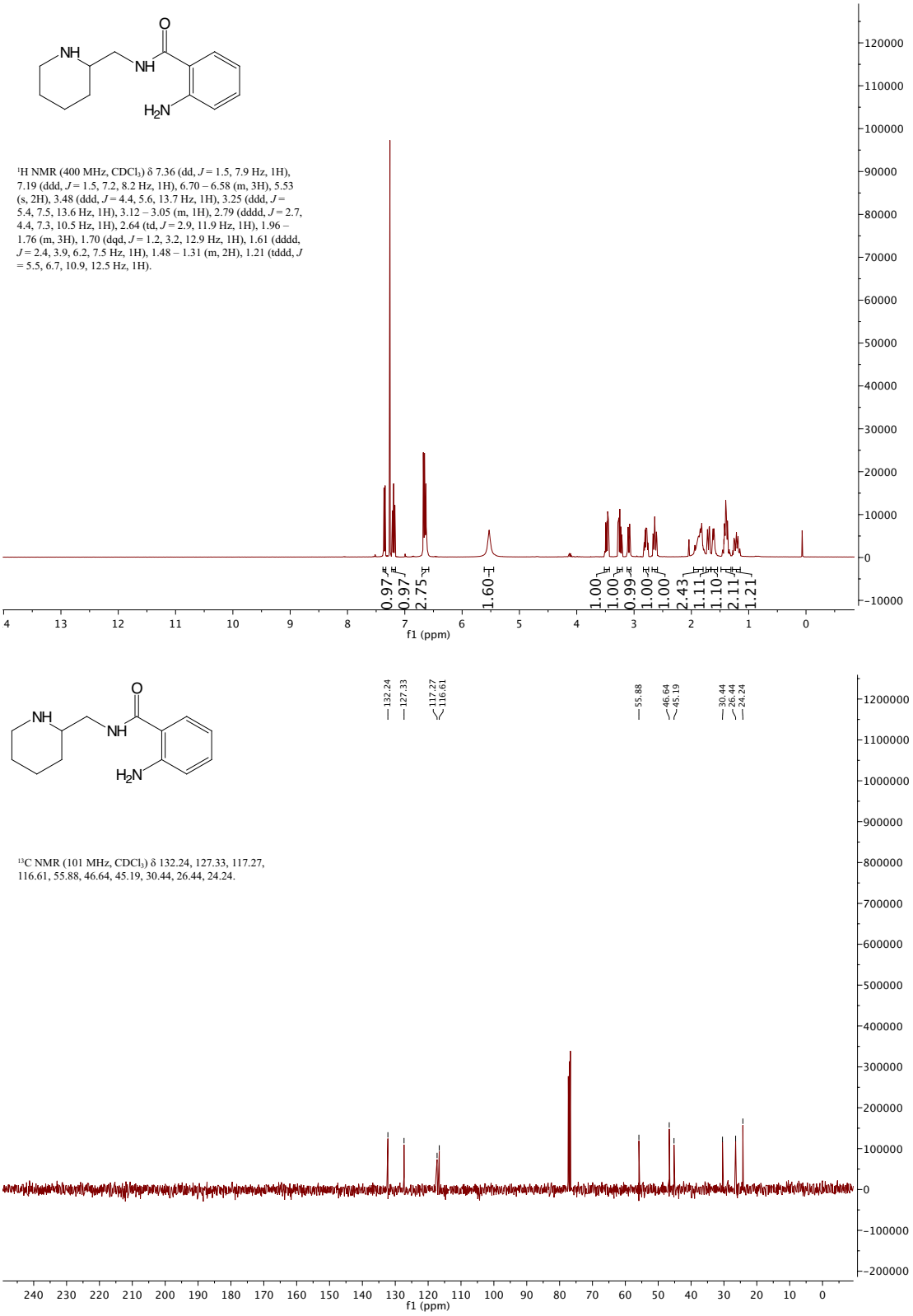



**Figure A.39:** (*Top*) **$^{1}$H-NMR** of compound **96**. (*Bottom*) **$^{13}$C-NMR** of compound **96**.

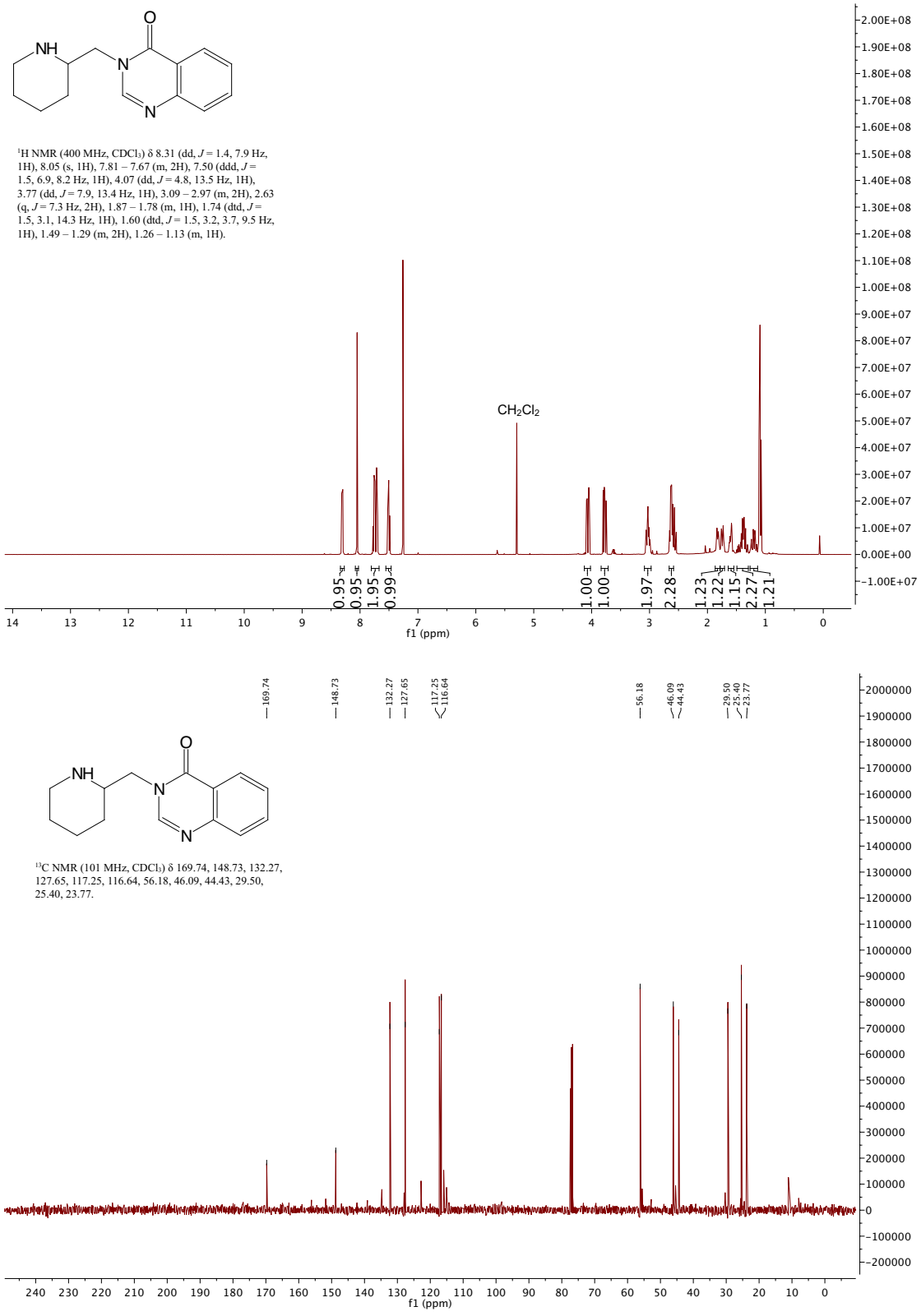



**Figure A.40:** (*Top*) **$^1$H-NMR** of 3-(piperidin-2-ylmethyl)quinazolin-4(3H)-one. (*Bottom*) **$^{13}$C-NMR** of 3-(piperidin-2-ylmethyl)quinazolin-4(3H)-one.

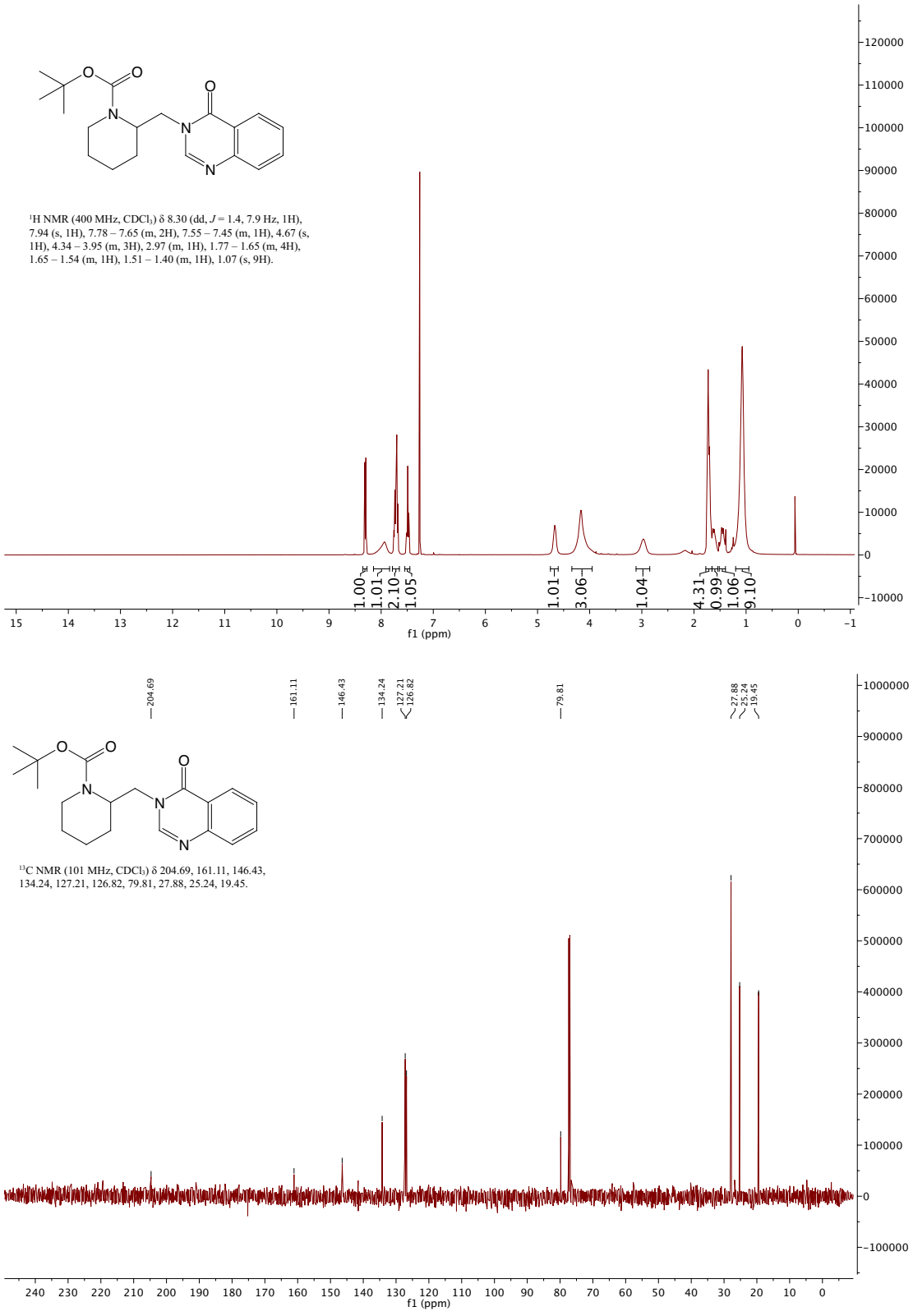



**Figure A.41:** (*Top*) $^{1}$**H-NMR** of compound **92**. (*Bottom*) $^{13}$**C-NMR** of compound **92**.

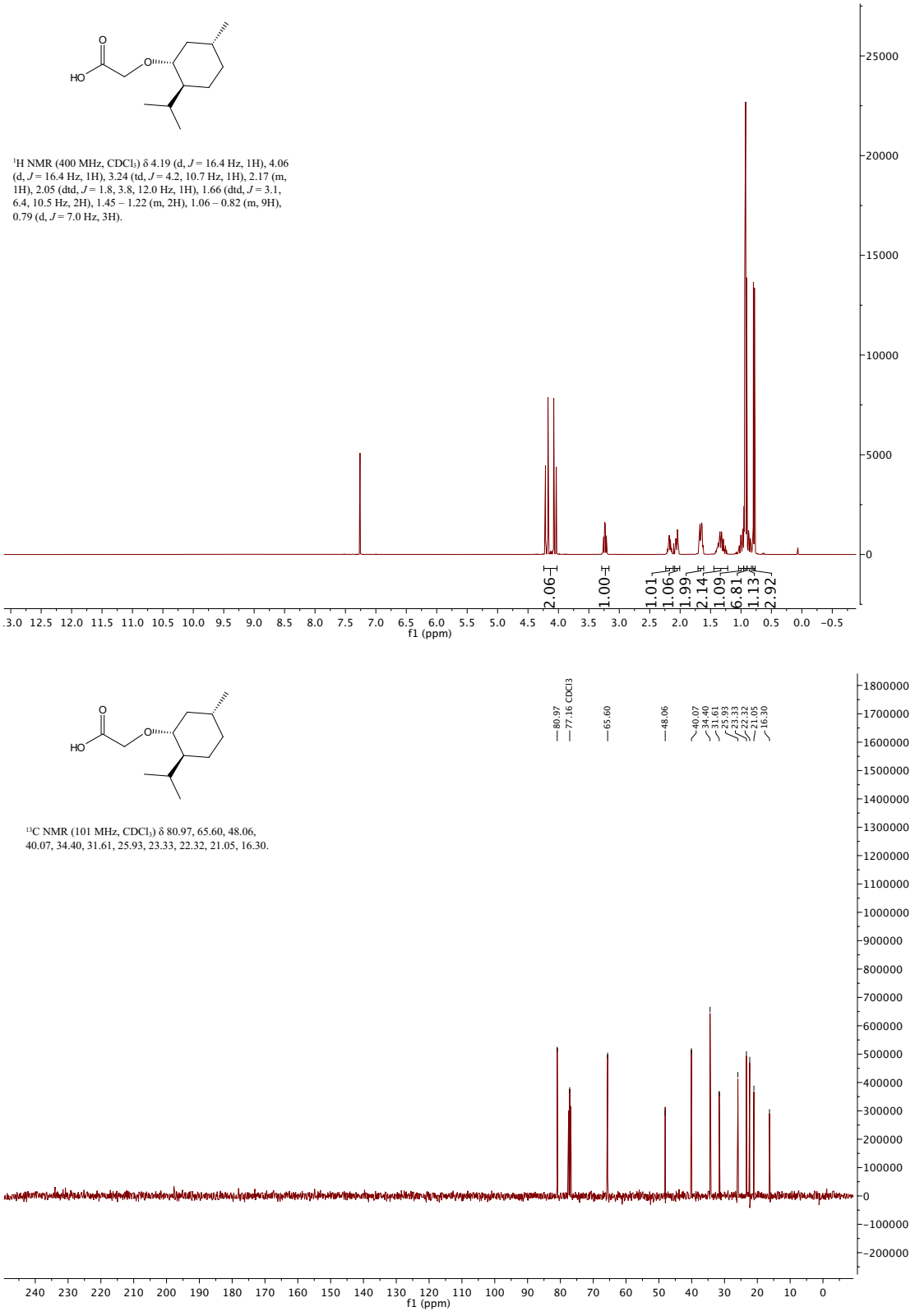



**Figure A.42:** (*Top*) $^1$**H-NMR** of compound **99**. (*Bottom*) $^{13}$**C-NMR** of compound **99**.

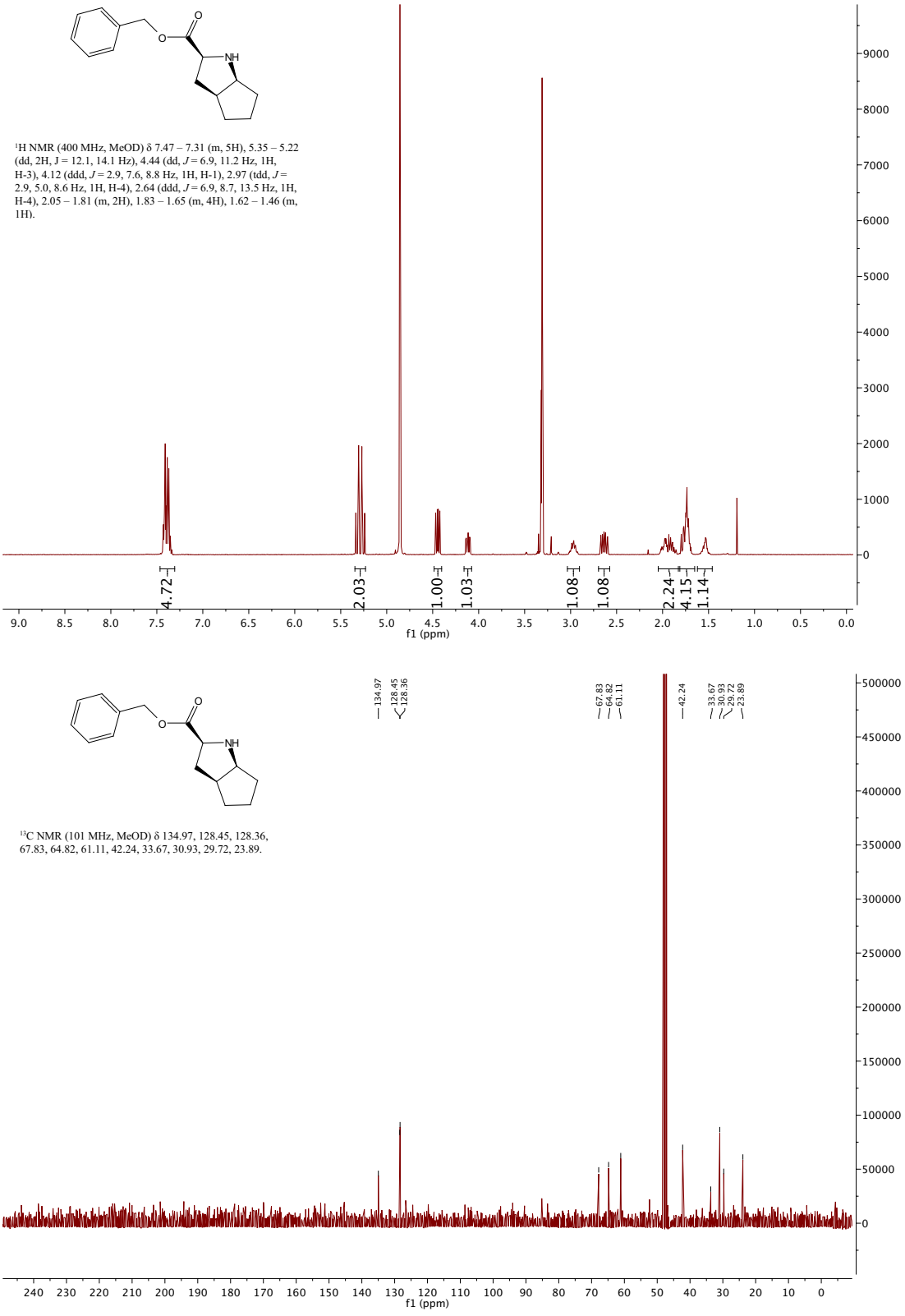



**Figure A.43:** (*Top*) $^{1}$**H-NMR** of compound **100**. (*Bottom*) $^{13}$**C-NMR** of compound **100**.

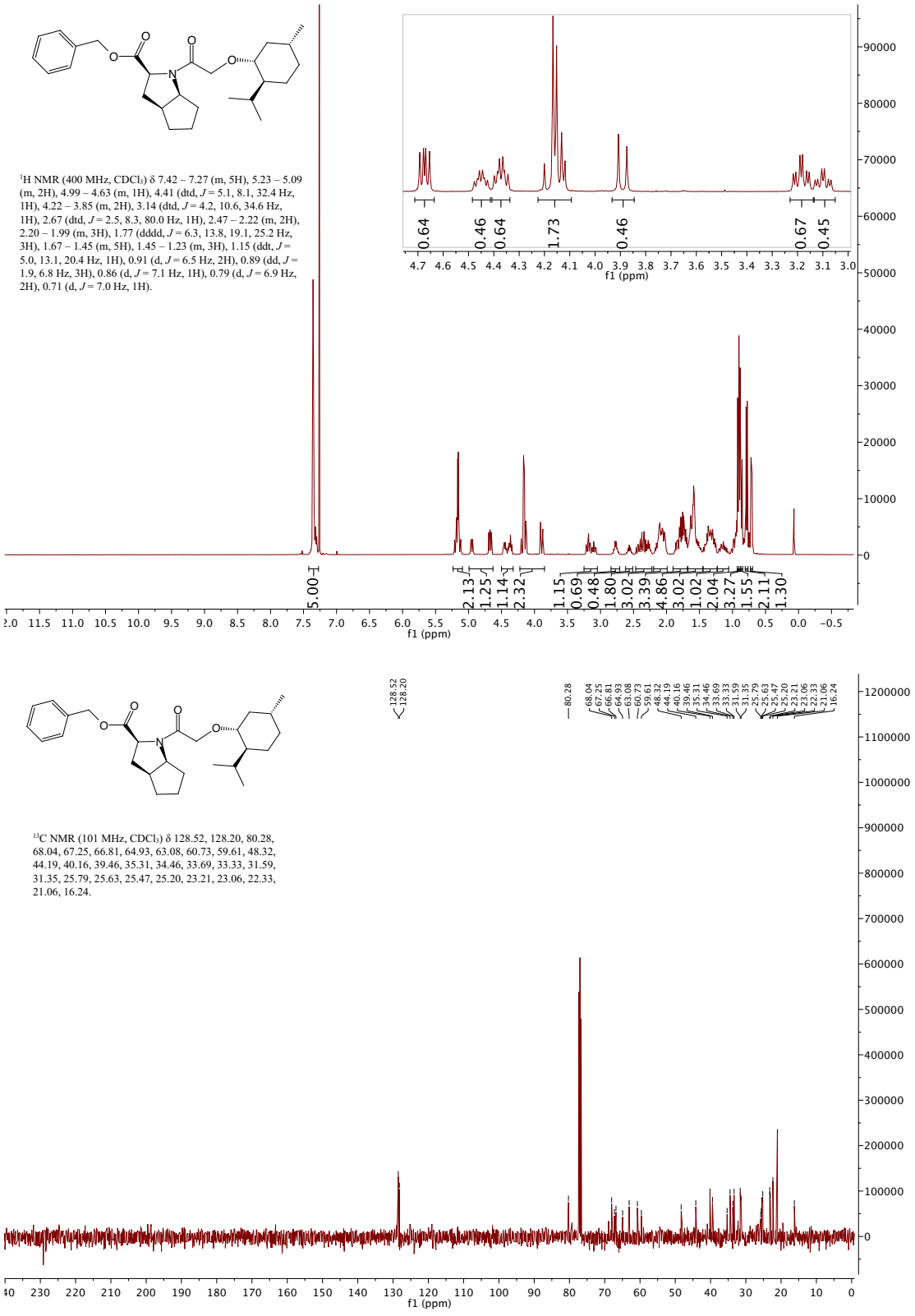



**Figure A.44:** (*Top*) [1]**H-NMR** of compound **93**. (*Bottom*) [13]**C-NMR** of compound **93**.

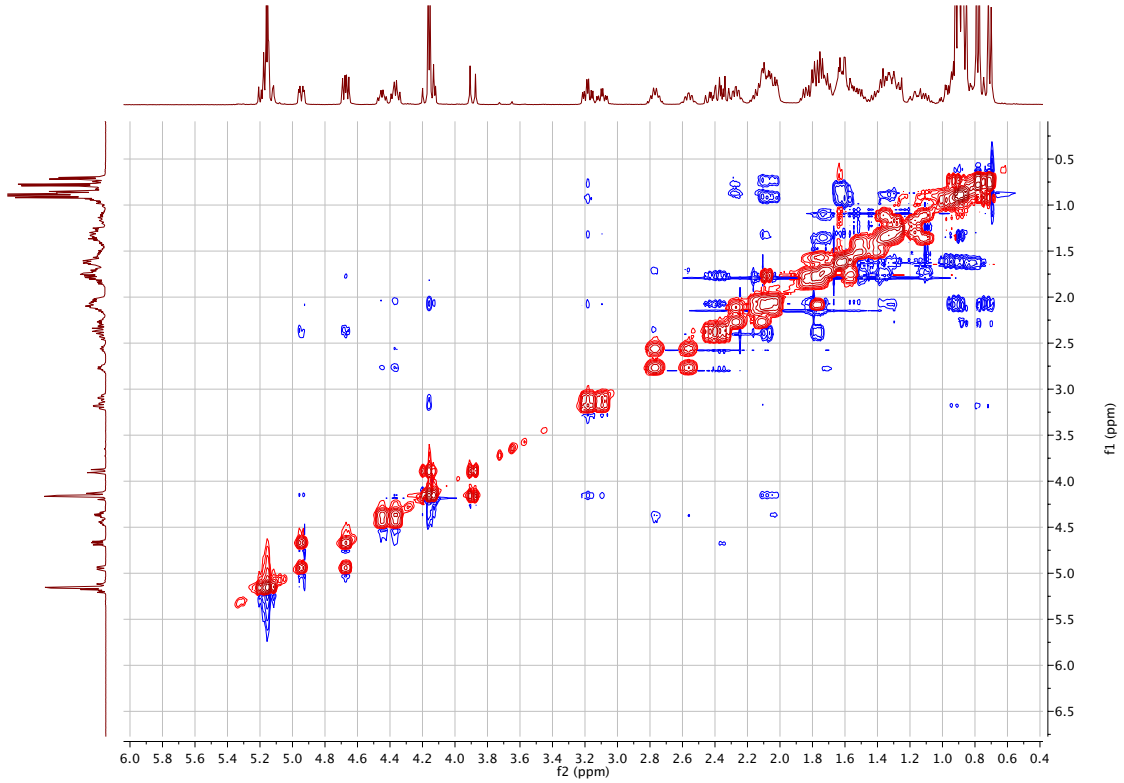



**Figure A.45: NOESY** spectrum of compound **93**.