universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„ Molecular modeling investigation of triketones as inhibitors of the herbicide target 4-hydroxyphenylpyruvate-dioxygenase “

verfasst von / submitted by

Ines Felicitas Werzinger

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

Magistra der Pharmazie (Mag.pharm.)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 449
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Diplomstudium Pharmazie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Thierry Langer

I Abstract

4-Hydroxyphenylpyruvate-dioxygenase (4-HPPD) is a target for chemical herbicides that has been researched for over 50 years. Triketones are one class of inhibitors of 4-HPPD, which have been examined. Using the 3D crystal structures of 4-HPPD of the crop Zea mays (PDB: 5YY6) and the weed Arabidopsis thaliana (PDB: 1SP8) the structural and sequential differences were elaborated with in silico methods. An in-house data set of 495 triketones was provided by BASF SE. These ligands were used to create pharmacophore models which were utilized for virtual screening to discover new bioactive herbicide compounds. The aim was to find a differentiation of selectivity of the enzyme source and to discriminate active from inactive compounds by in silico structure analysis combined with ligand-based pharmacophore modeling.

A structure-based distinction between Zea mays and Arabidopsis thaliana is not possible because the enzymes of both plants do not differ significantly. The distinction in both crystal structures is only in the different conformations of the C-terminal helices, which have a gate functionality. There should be more dynamic information about these helices, as these states are based on the rigid crystal structure to possibly solve the selectivity aim. Based on the ligand set a satisfying pharmacophore model could be generated which distinguishes between active and inactive compounds.

II Zusammenfassung

4-Hydroxyphenylpyruvat-Dioxygenase (4-HPPD) ist ein seit über 50 Jahren erforschtes Zielprotein für chemische Herbizide. Triketone sind eine Klasse von Inhibitoren für 4-HPPD, die in dieser Arbeit untersucht wurden. Mit Hilfe der 3D-Kristallstrukturen von 4-HPPD der kultivierten Pflanze Zea mays (PDB: 5YY6) und des Unkrauts Arabidopsis thaliana (PDB: 1SP8), wurden die strukturellen und sequentiellen Unterschiede mithilfe von *in-silico* Methoden erarbeitet. Ein BASF-interner Datensatz von 495 Triketonen wurde zur Verfügung gestellt. Diese Liganden wurden verwendet, um Pharmakophore zu erstellen, die für das virtuelle Screening zur Entdeckung neuer, bioaktiver Herbizidverbindungen verwendet wurden. Ziel war es, Selektivitätsunterschiede basierend auf den beiden Enzymsequenzen und -strukturen zu finden und aktive von inaktiven Verbindungen durch *in-silico* Strukturanalyse in Kombination mit ligandenbasierten Pharmakophormodellen zu unterscheiden.

Eine strukturbasierte Differenzierung zwischen 4-HPPD von Zea mays und Arabidopsis thaliana ist nicht möglich, da sich die Enzyme beider Pflanzen nicht wesentlich unterscheiden. Die Unterschiede der beiden Kristallstrukturen zeigen sich nur in den spezifischen Konformationen der C-terminalen Helices. Diese sind für das Öffnen und Schließen der Bindetasche verantwortlich. Demzufolge müssen mehr Informationen über die Dynamik dieser Helices gesammelt werden, da die statischen Zustände der Kristallstrukturen, keine ausreichende Bewertung zulassen. Basierend auf dem Satz an Liganden, konnte ein zufriedenstellendes Pharmakophormodell generiert werden, das zwischen aktiven und inaktiven Verbindungen unterscheiden kann.

III Acknowledgements

My thanks are first and foremost to Prof. Dr. Klaus-Jürgen Schleifer, who enabled me to do my diploma thesis at BASF SE in Ludwigshafen am Rhein. I greatly appreciate his expertise and have always enjoyed discussing the topic of my work. Thank you very much for giving me the opportunity to work on this exciting topic and to present it repeatedly in the modeling sessions.

I would also like to thank Dr. Michael Betz, who supported me on site. Every time I was confronted with a problem, he encouraged me to solve it independently in order to achieve the greatest possible learning effect. I would like to take this opportunity to thank him for the open ear he has always given me and his advice from day one to the last. His understanding and empathic nature made for a very pleasant working atmosphere.

I would also like to thank Univ.-Prof. Mag. Dr. Thierry Langer, who fully supported the diploma thesis as a cooperation between the University of Vienna and BASF SE. Thanks to the free license of LigandScout from InteLigand I was able to finish my project wonderfully and could learn a lot with the software.

Many thanks to Frank Sattel. He is a great colleague and helped me a lot in the beginning when everything was new and I had to get used to everything. With a lot of patience and his enjoyment of the work, the beginning was much easier for me. I have always appreciated his help.

Of course, I would also like to thank my family for the emotional and financial support during my studies.

I would like to emphasize my grandmother Anja, because she has strengthened my back for years and supported me 100% in all areas of my life. She was a special pillar of support throughout my studies and was literally my companion at the beginning and end of this phase of my life.

I would also like to thank all my friends, whom I cannot name here, for the wonderful years of our friendship, who have accompanied me through many years. On the one hand, I would like to thank you for the scientific support, but also for the attendance in

times when you always stood behind me, as well as for the common learning hours in the library.
Many, many thanks!

IV Contents

V Abbreviations

United States Department of Agriculture	USDA
Mode of action	MOA
Herbicide Resistance Action Committee	HRAC
Global Crop Protection Federation	GCPF
Weed Science Society of America	WSSA
Computer-aided drug design	CADD
Structure-activity relationships	SAR
Three-dimensional	3D
Arabidopsis thaliana	ARBTH
Zea Mays	ZEAMX
European and Mediterranean Plant Protection Organization	EPPO
4-hydroxyphenylpyruvate dioxygenase	4-HPPD
Kilodalton	kDa
Protein Data Bank	PDB
4-hydroxyphenylpyruvate	HPP
Molecular Operating Environment	MOE
Scientific Vector Language	SVL
Two-dimensional	2D
Grid-based Ligand Docking with Energetics	Glide
Library in local Database	ldb

VI List of Figures

VII List of Tables

I have made every effort to locate all the owners of the image rights and obtained their consent to use the images in this work. If a copyright infringement should become known nevertheless, I ask for report to me.

1 Introduction

1.1 Herbicides

Over the past 50 years, herbicides have been used extensively in agriculture to control weeds and ensure crop growth to ultimately provide safe harvests. In order to meet the challenges of modern farming, extensive research on new herbicides is essential.[1] In addition to economic aspects (e.g. production costs), next generation herbicides should also cover a wide range of weeds and deal with possible resistance developments.[1,2] The aim is to discover new chemotypes and modes of action of new, environmentally friendly, and efficient herbicides.[2,3]

Farmers from different regions of the world have different strategies for dealing with weeds. It is an interplay of crop rotation, fertilization and the use of herbicides to increase crop yields,[2] resulting in improved productivity.[4] Crops and weeds compete for water, light, physical space and nutrients. By reducing weeds, all crops such as maize can achieve a higher crop yield.[3,5] The quality of the crop and yield also suffers from competition for resources.[5] It is known that weeds reduce the global crop yields by up to 34 percent, in case if they are not treated with herbicides.[6] Nowadays, people are increasingly confronted with food shortages. There are several reasons for this, including the increase in the world's population.[2,4] Also, the increased spread of weeds must be mentioned.

Thus, weed management techniques play a key role in feeding the world’s population sustainably.[4] A sustainable farming depends on the development of new herbicides as well as plant production systems. The United States Department of Agriculture (USDA) defines three main aspects of sustainable agriculture. Their aims are to generate a long-term profit for environmental protection. This includes the sustainability of land, water, air and the quality of life of farmers and their community.[7]

By now, there are at least 315 weed biotypes and 183 weed species worldwide that have already developed resistance against a number of herbicides, making it necessary to search for new chemotypes possibly with a new mode of action (MOA).[8] However, the extensive use of herbicides has led to the adaptation of weeds to substances resulting progressive resistance to herbicides.[3] Resistance in plants can be defined as the absence of sensitivity of the organism to the standard dose of the

herbicide, which is associated with a genetic response based on an analogous MOA.[9] To address the emerging resistance issue, the Herbicide Resistance Action Committee (HRAC) was founded by international agrochemical companies in 2010.[4,10] The aim of HRAC is to protect crop yields and to prevent weed resistance to herbicides worldwide. The HRAC supports both agrochemical companies and farmers, by providing processed information on weed resistance and managing communication with regulatory authorities.[4,11] As part of the Global Crop Protection Federation (GCPF), this cooperation developed an extensive classification system for all available herbicides according to their biological target structure.[11] Next to the MOA, the classification system includes the chemical family, and the public name of the herbicide. It is used as an advantageous tool in the cultivation of crops to prevent the selection of herbicide-resistant weeds.[10,11] MOA of an herbicide is the biochemical or physical mechanism by which it disrupts or alters one or more of a plant's metabolic processes resulting in the death of the plant.[11] The focus of this thesis is on herbicides having bleaching MOA due to inhibition of pigment synthesis. In the appendix the Table 5 shows the classification of herbicides by HRAC. The first column shows the mode of action. The next column contains the code consisting of one letter. This is a symbol of a common group of herbicides targeting the same biological target. The third column shows the proteins on which the herbicide acts and the last column the commercial compound.

The HRAC classification scheme is applied worldwide. However, the Weed Science Society of America (WSSA) classification and the Australian Code System are comparable systems that find regional use. Each of these individual systems should be consulted by users of herbicides. Moreover it should describe the application of herbicides with the aim of getting the best results in weed control without causing resistance in countries that rely on particularly intensive agriculture cropping systems.[11]

New challenges to herbicides have emerged in recent years. In addition to the resistance problem, there is often low selectivity resulting in toxicity for other organisms and the extremely high costs of discovering, developing and approving new effective substances.[2]

Successful herbicide research is attentive to safety issues and selectivity. Selective substances usually bind to individual plant enzymes. Binding to the active site of the target protein achieving selective inhibitory effect in undesirable weeds. To meet the

safety aspect, strict toxicity regulations must be followed to protect humans, animals and the environment from adverse effects.[12]

In the last 40 years more than 300 herbicide active substances with diverse modes of action have been found that bind to different known and unknown proteins organisms.[9]

Earlier, new substances were discovered whose target structure could be assigned afterwards. In the meantime, computational chemistry is also being used to advance the discovery of new active substances and to further optimize them.[9]

1.2 Computer-aided Chemistry

The knowledge of folk medicine goes back thousands of years. Most of the active ingredients of plants were found by coincidence and were tested for their specific effects over this time by trial and error. Centuries later, new knowledge would be established through animal experiments. However, not everything can be measured on animals, for example emotions. Additional factors, such as bioavailability and selectivity, are essential properties that cannot be sufficiently tested using an animal model. Ethics in relation to animal experiments are another reason why progress in research was necessary. In vitro tests and in silico methods were developed..[13] The advantages of *in vitro* testing are improved control and time savings as fewer resources can be consumed. Unfortunately, no whole organism can be simulated *in vitro*. Many studies show that the transfer from laboratory to animal and human fails.[14]

In silico modelling is a computer-aided method that takes pharmacological and physiological aspects into account. It is an extension of the controlled *in vitro* tests and has the advantages of the two experimental methods already mentioned above. It is the result of an increased development of computing power and increasing cost savings. In *in silico* experiments, researchers have access to a variety of parameters that make the results useful for a holistic organism.[14] Computer-aided drug design (CADD) is also widely used in plant protection research. Attempts are made to predict the toxicity of substances at an early stage, analyze and simulate substances, which are usually based on a chemical structure.[6,15,16] Software tools are used to create models based on experimental data and scientific knowledge. Structure-

activity relationships (SAR) are also utilized.[15] Structure activity relationship which is the qualitative relationship between a chemical substance and its bioactivity. Hereby, there is a correlation between frequently occurring substructures in molecules characterized by predefined descriptors and their bioactivity.[16]

Toxicological prevention is a scientific challenge in the field of economics, ethics and technology in crop protection. More emphasis is placed on moral sensitivity, which should correlate with economic interests. Computational chemistry has several other advantages in various situations. In particular, it supports research when limited material is available for testing, toxicological test data is not available, complex test material cannot be produced, and a high-throughput approach is faster and cheaper. *In silico* approaches can exclude many chemicals a priori, due to the prediction of toxicity and based on the properties of the chemical.[9,15] Molecules that cannot be synthesized can also be eliminated.[9]

In recent decades, research on descriptors of agrochemicals has increased and the design of bioactive substances has been rationalized. Workflows have been created to classify and identify pesticides intended for the discovery of novel leads. These leads are intended to have the broadest possible spectrum of activity and to circumvent the resistance problem. Extensive screening of the databases with an enormous number of compounds is performed, which can be filtered according to various properties. This method supports drug research by finding bioactive substances for a specific target protein much faster.[9]

The aim of CADD is to combine knowledge from publicly accessible and internal databases to infer ligand target interactions. Thus, certain predictions are to be made regarding the properties of molecules. Finally, hypothesis on the MOA of new substances could also be developed.[9]

1.3 Goals and Motivation

The motivation of this thesis was to discover novel highly active nontoxic herbicide lead compounds targeting weeds rather than crops. The aim of this work was to make a selectivity difference between the weed Arabidopsis thaliana and the crop Zea mays. Another aim was to distinguish between active and inactive compounds,

whereas only triketone compounds were investigated, which are explained in more detail in chapter 2.3.

To achieve this goal, different computer-based approaches were used and combined. As described in the previous chapter, computational chemistry has many advantages, which is why some methods are used in this thesis. In a short time, a large set of data can be effectively aggregated, three-dimensional (3D) representations of the target under investigation with its ligands can be shown and SARs can be described. In order to understand the selectivity difference from the target protein 4-hydroxyphenylpyruvate dioxygenase (4-HPPD) and to achieve both objectives, pharmacophore models were generated based on the structure and ligands.

Pharmacophores are used in this thesis to represent molecules on the 3D level and to show the key elements of molecular recognition. Thus, the most important functional groups can be identified whose influence on the development of new drugs cannot be avoided (Chapter 3). Pharmacophore models have been gaining more popularity in recent years.[17,18] That is why I was particularly enthusiastic about being able to use this method and thus accompany an important step in research.

Through BASF I had the chance to get to know a variety of computational chemistry tools. In this way, the advantages of the individual tools could be combined to achieve the greatest possible output for my project.

2 Background

2.1 Enzyme Source

A distinction is made between the two plants Arabidopsis thaliana (ARBTH) and Zea mays (ZEAMX). ARBTH and ZEAMX are the official EPPO codes for those plants created by European and Mediterranean Plant Protection Organization (EPPO) and can be abbreviated in this thesis.[19] ARBTH is one of the weeds that compete for the resources of plants such as corn. HPPD is an enzyme that occurs in both plants and in mammals. This thesis deals with the enzyme 4-HPPD.

2.1.1 Arabidopsis thaliana

ARBTH (thale cress or mustard weed) shown in Figure 1 belongs to the family Brassicaceae, is a small, self-pollinating weed and can bear flowers.[20,21] It became an essential model system in plant research at an early stage.[20–22] Extensive research has led to the discovery of fundamental biological processes in plants in general. ARBTH has become one of the most investigated plants due to the generation of various available information and tools, such as the entire genome sequence, molecular genetic markers, and the simple generation of transgenic plants.[22,23] For example, biological interactions at the level of organisms have been studied.[20] Originally ARBTH was chosen because it has many advantages. These include a short generation time, a suitable size, and a high yield of the seed production of this plant.[24] ARBTH is a weed that is controlled by herbicides on agricultural land.

Figure 1: Arabidopsis thaliana.
It is a weed which grows around the world.[27]

2.1.2 Zea mays

ZEAMX from its phylogenetic viewpoint belongs to the family Poaceae. It is also known as corn and maize and is from genus level Zea with the species Zea mays.[25,26] Figure 2 shows a commercial maize field without weed. It can be assumed that herbicides were utilized to protect maize cultivation from weed.

Figure 2: Corn field.[27]

2.2 Enzyme 4-HPPD

Hypertyrosinemia is a hereditary disease in humans which is accompanied with a deficiency in 4-HPPD activity.[28] In humans the type II tyrosinemia occurs with mostly mild symptoms in contrast to type I tyrosinemia which shows fatal symptoms in case it is untreated. Tyrosinemia is caused by an accumulation of fumarylacetoacetate which results in a lack of fumarylacetoacetase. Reasons can be either that the enzyme is saturated or becomes decarboxylated to succinylacetoacetone or succinylacetone. This can cause liver cirrhosis, first stage liver cancer and anemia.[29]

4-HPPD presents a similar target in plants for which agrochemical companies have put research efforts in the last decades.[29,30] 4-HPPD-inhibiting herbicides have the advantage of broad activity against dicotes.[2,5,31] They are also effective against weeds that have already developed resistance to other substances.[2] These herbicides are highly effective so that good results can be achieved at low application rates, whether before or after sprouting. In addition, they have low environmental toxicity[32] and good plant selectivity. Many herbicides with this MOA have been presented already or are currently in the process of development.[5,29,30] 4-HPPD belongs to the category of dioxygenases as a member of the extradiol α-ketoacid-dependent group.[29,33] In most organisms these enzyme catabolize the amino acid tyrosine. It has been explored that 4-HPPD is located in the cytosol and not exclusively in chloroplasts.[11,34] It is known that 4-HPPD in ARBTH is found only in the cytosol as it can be an enzyme both in eukaryotes and prokaryotes.[11,31,35] In contrast, the enzyme in ZEAMX is located almost entirely in chloroplasts.[11,31] Further, there are plants having the subcellular localization of the enzyme 4-HPPD in both the cytosol and chloroplasts for example soybean.[11,31,35]

4-HPPD is an important enzyme in drug development for human therapy and agrochemical research.[5,36] The following chapters will explain the role of HPPD in plants as a major herbicide target based on its integration in anabolic pathways[29] and hence importance for plant growth.[5]

2.2.1 Enzyme Structures

It was demonstrated by biochemical characterizations that eukaryotic 4-HPPDs act as homodimers, but prokaryotic enzymes instead act as homotetramers.[2,11,29,37] Bacterial and plant enzymes have the equivalent overall fold, but for oligomerizing both use orthogonal surfaces.[29] Until now, all examined 4-HPPDs of mammals act as homodimers with subunits between 43 to 49 kilodalton (kDa).[34] However, the sequence identity of eukaryotic and prokaryotic 4-HPPD enzymes shows a low similarity of about 30 percent, which is replicated by their diverse oligomeric state.[29] Interestingly, the enzyme 4-HPPD of ARBTH is similar to that of mammals. The homodimers of both comprise 48 and 49 kDa subunits and can be inhibited with similar compounds, namely triketones which will be described as inhibitors for ZEAMX and ARBTH in the following chapter 2.3.[28] In plants, the N-termini of 4-HPPD enzymes differ from minimum 30 amino acids in their prolongation in contrast to bacterial and mammalian orthologs.[29]

Several publicly accessible resolved 4-HPPD crystal structures from diverse organisms can be found in the Protein Data Bank (PDB). All crystallized eukaryotic 4-HPPD structures have a molecular weight of 45 kDa and a similar topology as well as overall fold.[29]

ARBTH 4-HPPD crystals contain one dimer per asymmetric component. The asymmetric unit of 4-HPPD crystals of the ZEAMX contains two dimers with 44.8 kDa. Because ARBTH and maize have an equal dimerization mode they both consequently imitate the physiological oligomerization state in a soluble environment. It is also known that no higher oligomers are formed in the crystals.[29] Comparison of the sequences indicate that the both 4-HPPD enzymes have more than 60% sequence similarity.[38]

The tertiary structures of monomers of those proteins are separated in two structural domains which show an open β-barrel at the N- and C-terminal (Figure 3). They consist of eight β-strands each. It is likely that the flexible N-termini have no direct function in the catalysis. In contrast, the conserved C-terminal domain has catalytic properties and contains the active site.[2,29]

Figure 3 shows the crystal structure of the A chain of 4-HPPD of ZEAMX. The yellow box shows the binding pocket of the enzyme for substrates and inhibitors. The mech-

anism of action (MoA) and the binding mode of these compounds will be discussed in the chapters 2.2.2, 2.2.3, and 2.3.

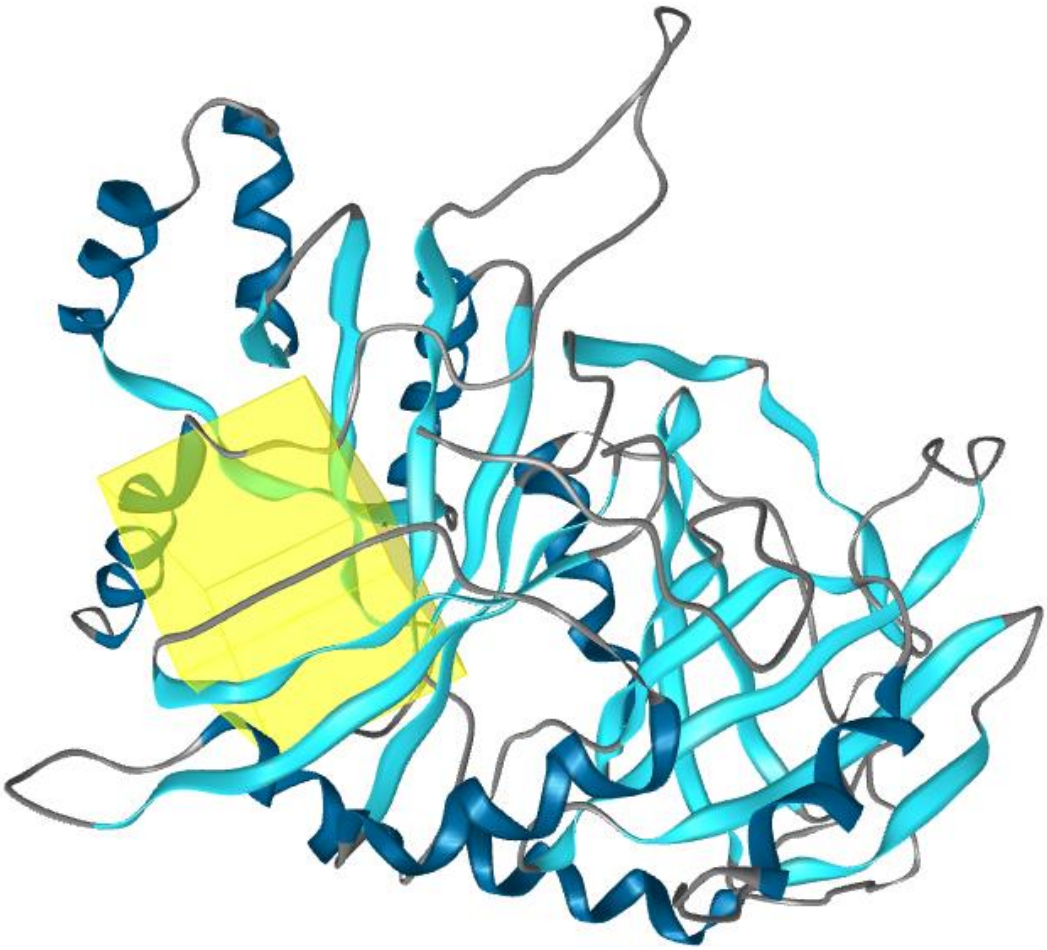


Figure 3: Crystal structure of the A chain of 4-HPPD from ZEAMX (PDB:1SP8).
The picture was generated with the tool LigandScout. The protein is shown with the render style snake. The tertiary structure with the binding pocket (yellow box) is automatically generated by LigandScout .

The Enzyme 4-HPPD contains a nonheme ferrous ion which is an important co-factor binding in the C-terminus of the protein.[5,12,31,37] It binds noncovalently to a 2-His-1-carboxylate motif and is positioned in the middle of the ß-barrel where the active site

is located. The C-terminal helix H11 completes the binding pocket at the exposed side of the open ß-barrel. On the one hand, this C-terminal helix of the 4-HPPD structures of ZEAMX shifts towards the active site and shields it from solvent.[11,29]
On the other hand, in the ARBTH structure this helix inclines by about 60° into the solvent enabling access to the active site. Consequently, it is presumed that helix 11 has the task of controlling the entry to the active side.[29] Figure 4 demonstrates the tertiary structure of the active binding site of the enzyme of ARBTH. Here the alpha-helix H11 is noticeable at the C-terminus, showing an open gate and thus an accessible binding site (PDB: 5YY6).[29] The X-ray crystal structure provides a substantial insight into the shape of the binding pocket.

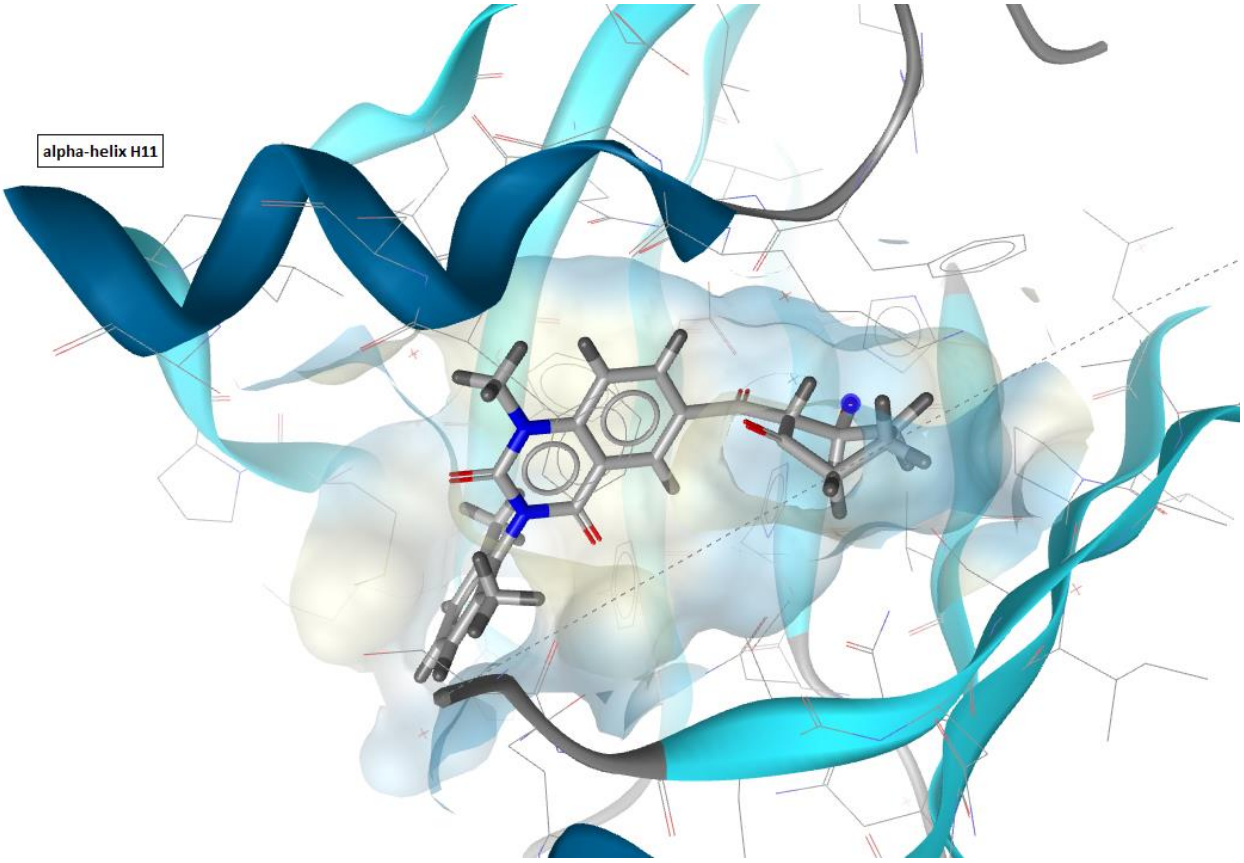


Figure 4: Active Site of the ARBTH 4-HPPD including a triketone inhibitor (PDB: 5YY6).
The surface from the binding pocket is shown around the compound which is located at the open ß-barrel next to the C-terminal alpha-helix H11. This helix has a gate regulating function to the active site. This figure is made in the software LigandScout. The style of the protein backbone is “snake” and the ligand bonds are shown as stick.

The active site ferrous iron is indispensable for the interaction between the enzyme and substrates.[28,29,34] It is mandatory for the catalytic activity and the stabilization of the binding. This binding imitates the substrate 4-hydroxyphenylpyruvate (HPP) with its α-keto acid fragment.[28]

Fe^{2+} is coordinated by different residues.[31]

Figure 5 illustrates the coordination of the nonheme ferrous ion coordinated by His308, His226, Glu394 and a water molecule.[29,37]

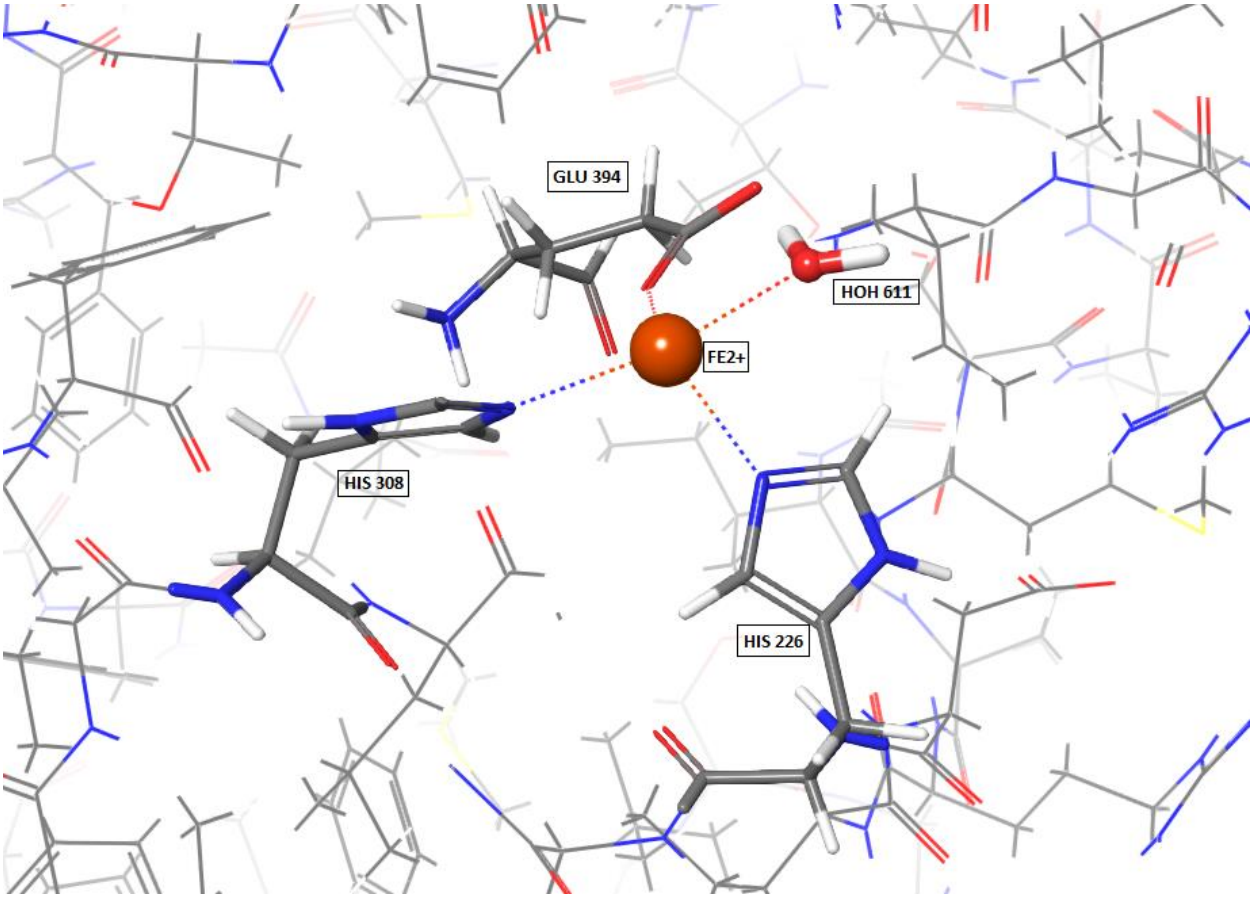


Figure 5: Active site architecture of 4-HPPD without its co-crystallized ligand (PDB: 5YY6).
The iron is coordinated with the protein side chains and an additional water molecule binding non-covalently to a 2-His-1-Glu motif. The Maestro software was used to generate the picture of the binding pocket.

Figure 6 shows the substrate HPP in the binding site. The two oxygen atoms chelate the iron ion in the binding pocket and thus form two further directed coordination. The driving force for the binding of ligands in the binding pocket is therefore the interaction of the described chelating moieties of the ligands with the cationic iron.[37]

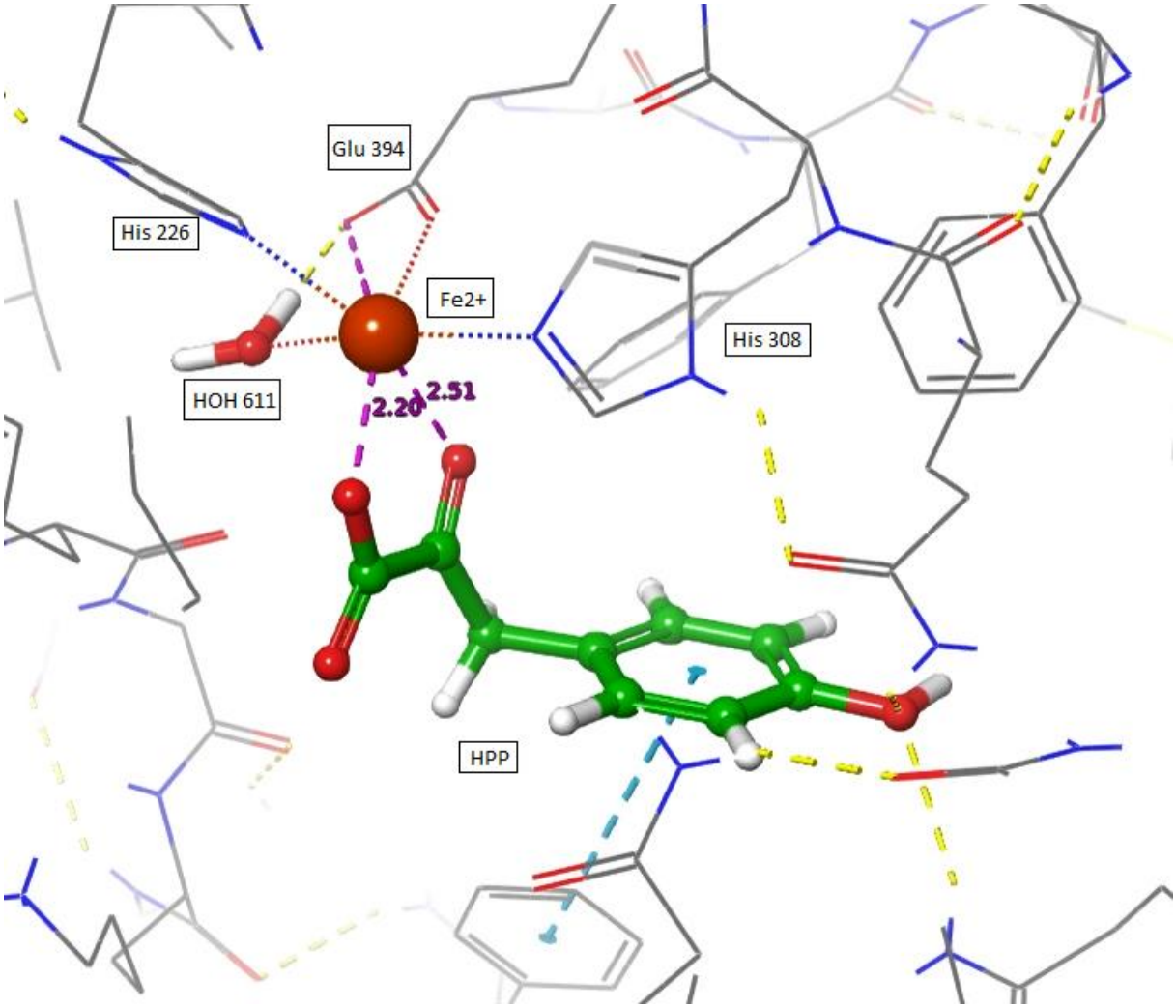


Figure 6: Active binding site of 4-HPPD of ARBTH (PDB: 5YY6) with the substrate HPP.
HPP appears in ball-and-stick presentation. The two oxygen atoms that form a chelating bond with the iron atom of the enzyme 4-HPPD can be seen. The measured distance between the oxygen atoms of the substrate and the cationic iron is 2.20 Å and 2.51 Å respectively. The iron ion has a total of 6 interactions in the enzyme-substrate complex. The edge-to-face-π stacking interaction starting from Phe 424 also plays an important role in the binding model. All amino acids are shown in wire style. This representation was created using Maestro.

In contrast to the small substrate HPP, most 4-HPPD inhibitors are bulky molecules, e.g. shown in Figure 7. Therefore, inhibitors can usually only enter the binding pocket when Helix 11 is in its open conformation.[37] The inhibitors bind similar to HPP at the same position by chelating the two oxygen atoms. In addition to the three amino acid residues interacting with the iron ion in the cavity (Figure 5), hydrophobic amino acids, which are dominated by their rigid secondary structural elements, are present in the binding pocket. Herbicides acting on target 4-HPPD bind similar to the substrate

HPP. They are competitive inhibitors and bind tightly to the enzyme (Figure 7) with the two keto enolate oxygens via the chelate bond to the iron ion and the π stacking to the phenylalanine Phe 424 in the active site.[39]

Figure 7 illustrates the active site from the crystallized 4-HPPD (PDB code: 5YY6) with a co-crystallized triketone inhibitor in the binding pocket. The ligand has the same binding motif as depicted in Figure 6.

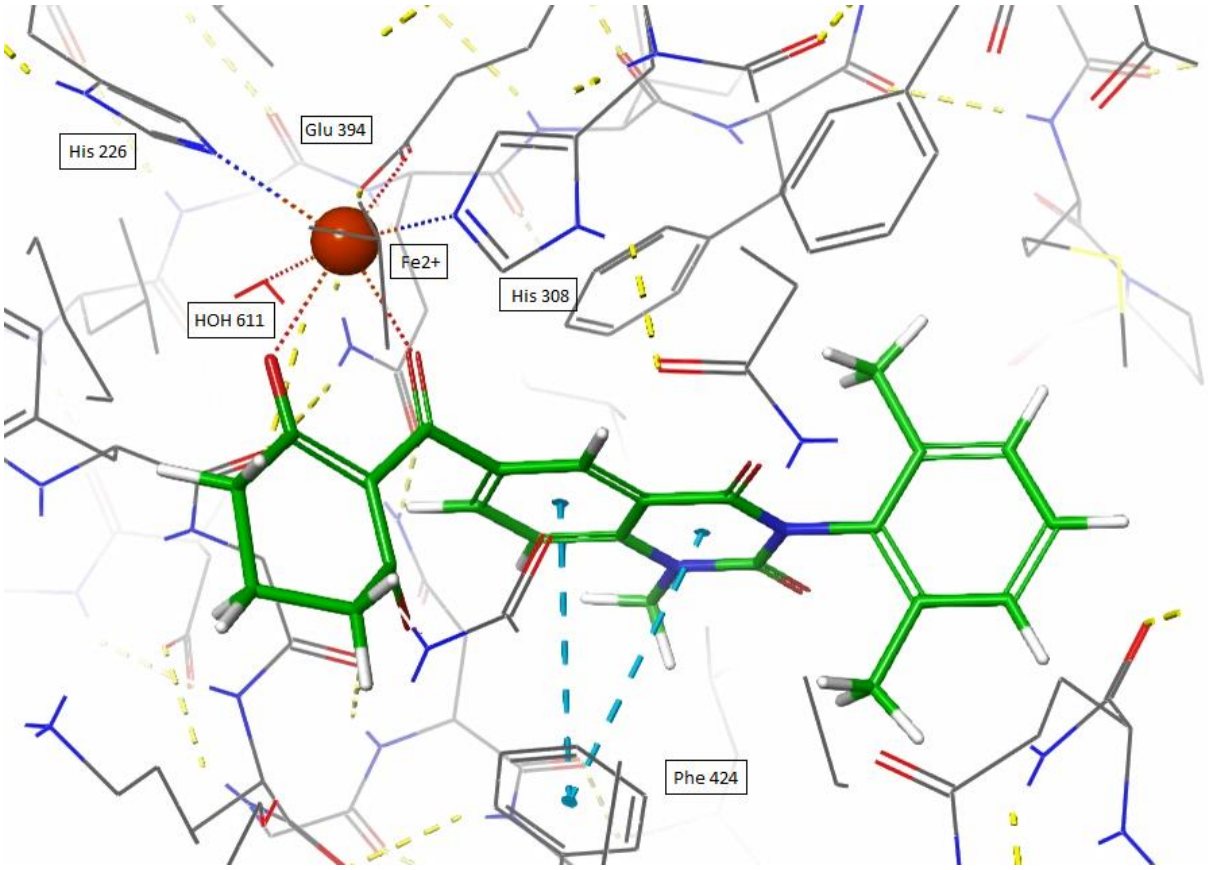


Figure 7: Active binding site of 4-HPPD of ARBTH (PDB: 5YY6) with an inhibitor of the class of triketones.

This figure shows a typical triketone inhibitor with the structure 3-(2,6-dimethylphenyl)-1-methyl-6-(2-oxidanyl-6-oxidanylidene-cyclohexen-1-yl)carbonyl-quinazoline-2,4-dione. This crystallized, rather bulky ligand with its oxygen atoms at the warhead also forms a chelate bond to Fe^{2+}. The figure created with Maestro represents the inhibitor in stick-and-ball style and the amino acids in wire style.

As described above, the C-terminal helix H11 has a gating function. It is located directly at the entrance of the binding pocket and is present in an open or closed conformation. In its closed form it covers the pocket from solvent exposure. For a compound to reach and bind into the binding pocket H11 must be present in its open conformation.

ARBTH is crystallized in an open conformation in contrast to ZEAMX, whose H11 is crystallized the closed state. ARBTH favors the open status due to higher configurational entropy and solvation of polar residues. H11 in ZEAMX in its closed form is more stable by forming additional hydrogen bonds.[38] The gating mechanism is differently triggered by the physiological compounds (i.e. substrate vs. product). In the unbound enzyme, the helix opens to allow the rapid entry of the substrate. As soon as the substrate binds, the helix locks the binding pocket and shields the substrate from the environment. The substrate is converted into the product and releases the product again via the temporarily opened helix.
The crystal structures of the two enzymes ARBTH and ZEAMX have so far only been crystallized in their respective more stable conformations.[29]

2.2.2 Metabolism

Dioxygenases in general play a decisive role in degrading aromatic amino acids.[29] In the second step of the catabolism of tyrosine 4-HPPD is the key enzyme which catalyzes the transformation of HPP and molecular oxygen to homogentisate (2,5-dihydroxyphenylacetate).[5,12,28,29,32,39–42] In chloroplasts the enzyme influences the anabolic synthesis of prenylquinone and catabolizes aromatic amino acids in the cytosol.[29,33] The substrate HPP is formed in the shikimate pathway and is derived from the transamination of the amino acid thyrosin.[11,41,43] For the catalysis of 4-HPPD hydroxyphenylpyruvate acts as an activating effector, leading to convert the first oxygenated intermediate.
This reaction from HPP to homogentisate is subdivided into 3 steps which follow directly one after each other. As illustrated in Figure 8, oxidative decarboxylation of the 2-oxo acid chain from the substrate HPP occurs first, followed by hydroxylation of the aromatic ring, and finally ortho-migration of the carboxymethyl chain.[28,41,44,45] During this reaction one oxygen molecule is consumed.[41]

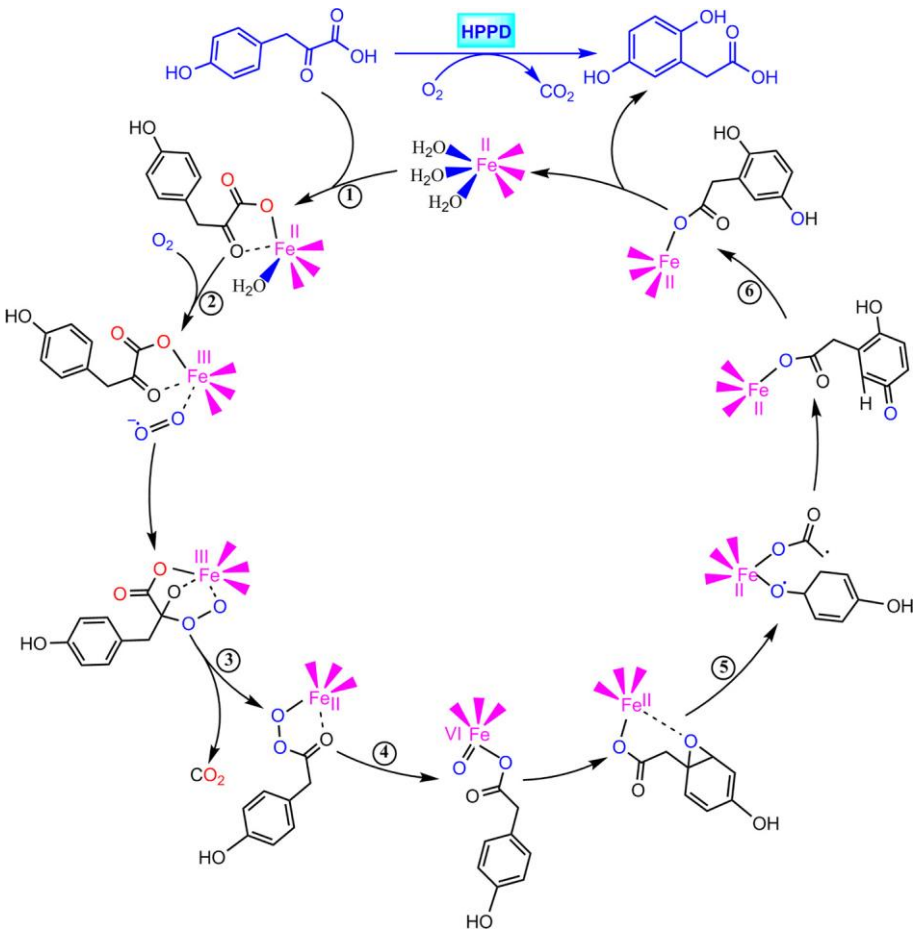


Figure 8: Catalytic mechanism of 4-HPPD.

It converse 4-hydroxyphenylpyruvate to homogentisate in one step.[2] This reaction is caused by an oxidative decarboxylation at the 2-ketoacid side chain of the substrate and thus a hydroxylation of the aromatic ring and rearrangement of the pyruvate side chain.

Homogentisate as the product of the reaction is an important aromatic precursor for prenylquinone in plants.[29,39] More precisely, it is a principal precursor for alpha-tocopherol and plastoquinone which are essential redox cofactors for isoprenoid bio-synthesis.[32,33,41] It was found that 4-HPPD is essential for the formation of both end products of this depicted reaction. Figure 9 shows the transformation from HPP to plastoquinone and alpha-tocopherol with homogentisate as the intermediate step.[29]

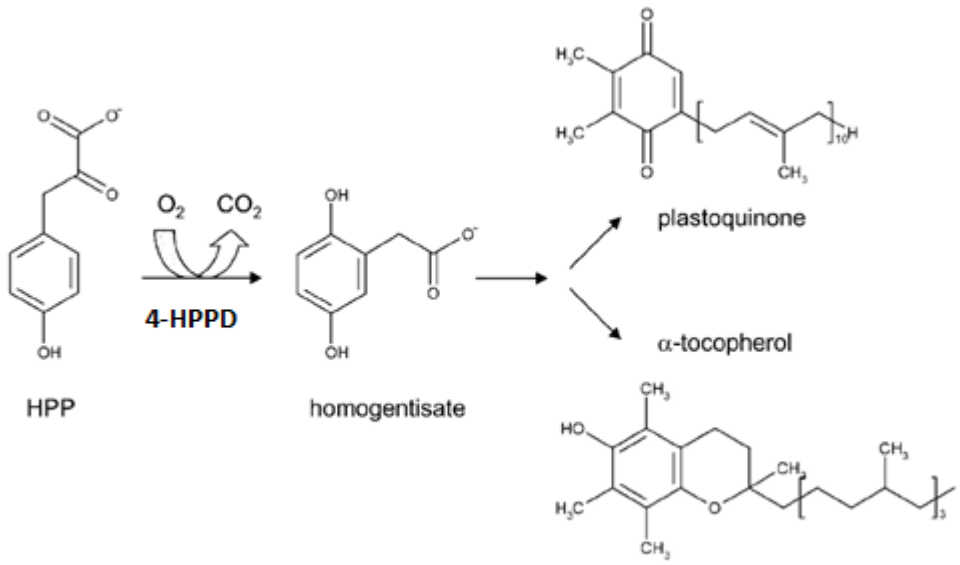


Figure 9: Reaction from HPP to plastoquinone and tocopherol catalyzed by 4-HPPD

The enzyme 4-HPPD converse the substrate HPP to homogentisate which is an intermediate step in the synthesis of plastoquinone and alpha-tocopherol. Modified from [**29**].

In higher plant chloroplasts alpha-tocopherol and plastoquinone are two main classes of lipid-soluble prenylquinone compounds[29,45] located in the inner chloroplast membrane.[46] Prenylquinones are only produced in plants and other aerobic organisms that are involved in photosynthesis. Alpha-tocopherol and plastoquinone are amphiphilic molecules with a hydrophilic head group and a hydrophobic residue consisting of isoprenoid moieties.[45] Both products are part of the antioxidative systems. Plastoquinone is necessary for the photosynthetic electron transport chain. It functions as an electron carrier between the photosystem two and the cytochrome b6/f complex. Furthermore, it is known as an electron carrier for NAD(P)H-plastoquinone oxidoreductase and as a cofactor for phytoene desaturase which is important for the carotenoid biosynthesis.[29,39,45] The antioxidant alpha-tocopherol reduces oxidative stress in plants and determines growth.[29] When an organism gets out of balance between prooxidative and antioxidant factors, it is called oxidative stress. As a result, free radicals are released that induce structural damage to various molecules.[42] It is also known as vitamin E and is a structural element of membranes.[29]

In the next chapter 2.2.3 Herbicidal Mode of Action, inhibition of 4-HPPD as an important herbicidal MOA will be described.

2.2.3 Herbicidal Mode of Action

Because the metabolism of the aromatic amino acid tyrosine plays an indispensable role in plants, inhibition of the enzyme became the focus of herbicide research several years ago.[29]

It has been discovered that molecules with multiple ketones inhibit 4-HPPD and operate as allelopathic agents.[29] 4-HPPD herbicides have the characteristic of being slightly acidic. This enables them to spread well in plants from the application on the leaf via the phloem to the shoot tips.[39] Their MOA is an irreversible inhibition of the target.[1,28] By inhibiting the enzyme 4-HPPD, the conversion of HPP into homogentisic acid is blocked, thus reducing the isoprenoid redox cofactors required for further carotenoid biosynthesis.[5,29,47] The consequence is a bleaching effect in plants caused by an accumulation of the carotenoid predecessor, phytoene. This was proven by an indirect inhibition of phytoene desaturase activity which is associated with the nonattendance of plastoquinone which is the indispensable cofactor for the desaturase.[29,34,41,47] This target takes part in the biosynthesis of photosynthetic pigments, the carotenoids.[29,47] It is associated with a reduction of carotenoids which sufficiently protects biological membranes from oxidative stress. Consequently, chlorophyll molecules are destroyed caused by singlet oxygen build by excessive light energy.[29,37,47] The photosynthetic system is no longer stable.[29,47] Obviously, inhibition of 4-HPPD leads to an increase in tyrosine levels.[41] In prone weeds the bleaching effect of these herbicides leads to plant necrosis, inhibited growth and death.[5,28,31,39]

If a plant treated with a bleaching herbicide with the described MOA is supplemented with homogentisate, the reverse case can be seen. The plant compensates for the homogentisate deficiency and stops the growth-inhibiting effects. This proves that this MOA blocks the conversion of HPP into homogentisate.[48]

These 4-HPPD class herbicides can be very effective for selective pre-emergence and in some cases post-emergence control against a broad range of broadleaf weeds and grasses in maize fields.[41,32]

This reaction and especially the target are of interest for the development of new herbicides.[29] 4-HPPD herbicides are mainly used in maize and monocotyledonous plants. Other crops such as soybeans and other dicotyledonous plants are more sensitive to this MOA.[31]This target can be blocked for example with triketones which are described in the following chapter.

Figure 10 shows the effect of triketone herbicides on ARBTH. In the top row untreated control plants are seen. The substance was sprayed only on the weed. Since the plant reacts very sensitively to 4-HPPD inhibitors, the leaf mass is already dead after six days. This picture shows how effective triketone herbicides are.

Figure 10: Visualization of the Results of retesting Triketones.

In the upper row the pots with the weed ARBTH are untreated. They are used as a reference against the treated plants in the bottom part. The picture was taken six days after applying a triketone compound in the greenhouse of BASF SE in Limburgerhof from a previous project.

2.3 Triketones

Many 4-HPPD inhibitors are potent herbicides and frequently used due to their numerous advantages.[5,8,49] In addition to a low application rate, they have a low toxicity and are broad-spectrum herbicides applied to unwanted weeds. Among the selective broadleaf herbicides that affect the enzyme 4-HPPD are triketones, pyrazoles, isoxazoles, diketone nitriles and benzophenones. Common representatives of triketones

are Sulcotrione, Mesotrione and Benzobicylone, which differ in chemical modification.[5,8] They are β-triketones with a 1,3-diketon moiety that mimics an alpha-keto acid group of HPP.[32,41] The triketones are chemically based on the natural phytotoxin Leptospermone, which occurs in the Californian bottle brush plant Callistemon citrinus.[1,2,6,11,32] The chemical structure of Leptospermone is drawn in Figure 11. In 1977, scientist noticed that this natural substance bleached the grasses in its environment.[11] Based on this property new 4-HPPD inhibitors have been developed.[2] Since the first active triketone inhibitor was found in 1982, attempts have been made to develop further inhibitors of this class. Thus Sulcotrione and Mesotrione were discovered.[2,11]

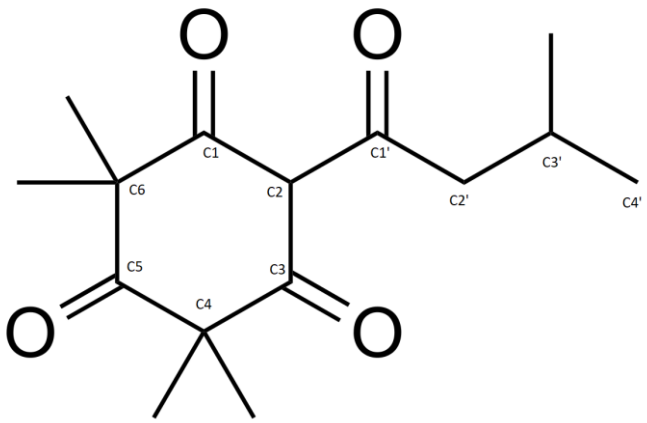


Figure 11: Leptospermone is a 4-HPPD inhibitor of natural origin.

Triketones belong to a very active family of bleaching herbicides. They act as tight-binding competitive inhibitors of the HPP substrate and only need a low dose to be effective.[28,40] The tight binding is due to the chelation of the protein-related Fe^{2+} with the enol tautomer of the 1,3-diketone unit.[40] The ketone at position C3 of the cyclohexane ring has a stabilizing function. Through the C3-carbonyl group it strengthens the inhibition of 4-HPPD. The presence of the C5-carbonyl group reduces overall planarity and thus prevents keto enol tautomerism, which would decrease inhibitory activity.[40] Triketones have the potential to initiate a keto-enol-tautomerization based on the vinylogy principle. In solution, more compounds are formed in the enol form.[2]

2.3.1 Triketone 4-HPPD Inhibitors

Triketones mimic the alpha-keto acid group of HPP and bind competitively to the Fe^{2+} at the active site of the enzyme causing its inhibition.[30] This binding is based on a metal chelating part with the enol tautomer of 1,3-diketone moiety that was explained in chapter 2.2.1 Structure.

Figure 11 shows the structure of the substrate HPP. In comparison, the structure of a triketone is shown. It is the co-crystallised ligand from the PDB with the code 5YY6. Both structures are similar.

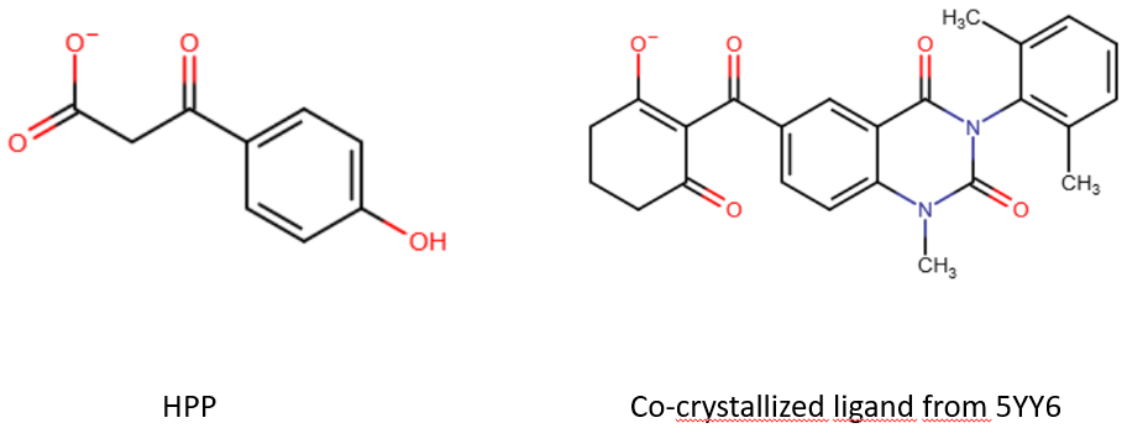


Figure 12: Structure of the substrate HPP and a triketone

Both structures have the same oxygen motif that can chelate the iron ion in the binding pocket.

2.3.2 Substitution on the Triketone Warhead

The structural feature of the 2-carbonyl group at the cyclohexane ring is essential for the activity of triketons.[40] This supposed triketone warhead can also be substituted. The structural change makes it difficult for plants to detoxify the triketone. In plants it is known that the hydroxylation at the 4-position on the cyclohexanedione takes place induced by cytochrome P-450 enzymes. Detoxification takes place both in weed and in maize. If this position is substituted and therefore occupied, the hydroxylation takes place at the chemically equivalent 6-position. If there is another methyl residue, a dimethyl residue, at the triketone warhead at 4-position, the metabolic process in plants slows down considerably and the activity as herbicides against weeds is high-

er. Some of the most active triketones are structurally similar to Leptospermone with the contained 2,2,4,4,-tetramethyl-cyclohexane-1,3,5-trione substructure. By the substitution at the triketone warhead the cytochrome p-450 metabolism slows down, but the selectivity decreases and the residence time in the soil increases.[11,49]

2.3.3 Commercial Triketone Herbicides for Corn

The most commonly used β-triketones in crop fields are Mesotrione and Sulcotrione,[1,49] which are used in therapies and agriculture.[2,29] As already mentioned in chapter 2.2.3, the classes of all 4-HPPD inhibitors have a bleaching effect on the plants, which is accompanied by an increase in the levels of tyrosine and phytoene. This causes a decrease in plastoquinone and vitamin E pools.[41]

Sulcotrione, with a 2-benzoylcyclohexanedione structure (Figure 13, a), was the first commercialized triketone herbicide in 1993,[11] which is used on maize fields mainly in South Africa and Europe. broad-leaved and grassy weed absorb Sulcotrione via leaves and roots.[41] It is only used for post-emergence control. The application rate of Sulcotrione ranges from 300 to 450g per hectare.[11]

Mesotrione (Figure 13, b) is a second generation triketone herbicide. It has a lower application rate of 100 to 225g per hectare for pre-emergence uses. Post-emergence application rates range from 70 to 150g per hectare.[41,49]

Both β -triketones are structurally very similar to the natural Leptospermone.[32]

Tembotrione (Figure 13, c) is an active triketone discovered in 1997. It is also used to control grass and broad-leaved weeds in corn with low application rates between 75 to 100g per hectare.[11]

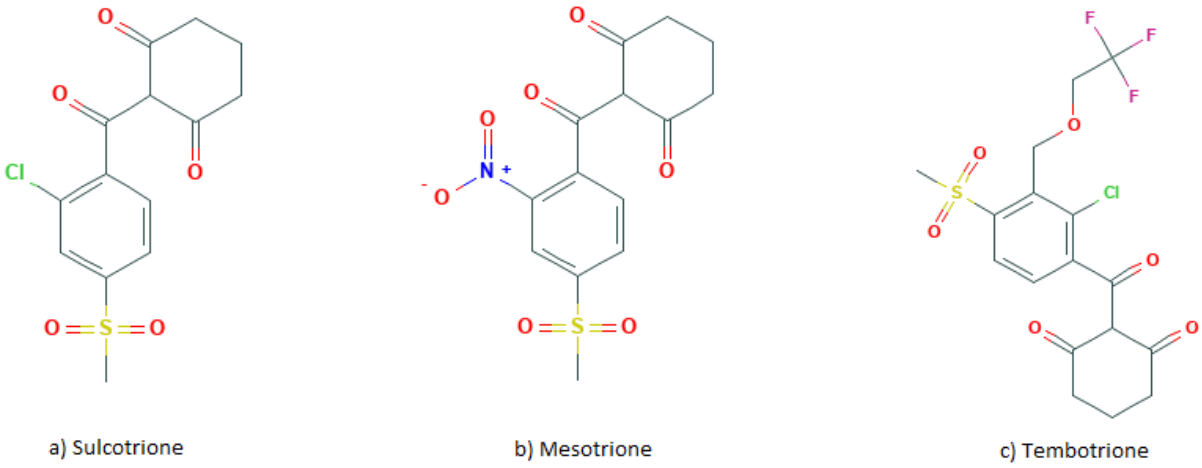


Figure 13: Structures of the most commonly used 4-HPPD inhibitors.

a) Sulcotrione with the (2-[2-chloro-4-methanesulfonylbenzoyl]) residue, b) Mesotrione with the (2-[4-methylsulfonyl-2-nitrobenzoyl]) residue, and c) Tembotrione with the 2-[2-chloro-4-methylsulfonyl-3-(2,2,2-trifluoroethoxymethyl)-benzoyl] residue is shown. These compounds were found at PubChem and find their use in weed control in corn. Modified from [50–52].

Further triketone herbicides are tefuryltriketone and benzobicyclone. They are mainly applied in rice cultivation.[11]

Triketones are defined as low risk compounds for the environment due to their low toxicity according to the U.S. Environmental Protection Agency.[32]

After 40 years of research, it is surprising that only 5 commercial products are on the market so far. Therefore, it is important to continue researching on triketones to find new and potent analogues.[11]

3 Pharmacophore Modeling

In the field of drug discovery and development, computer-based techniques are successfully applied and nowadays indispensable. Pharmacophore modeling is a powerful tool to find and optimize hits. Virtual screening as a conventional method can be used to search a large collection of filtered compounds. Pharmacophore models can improve experimental high-throughput screening. [5,12,54]

3.1 Definition

A pharmacophore according to IUPAC is defined as " an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response".[55] It is a popular word in medicinal chemistry. However, depending on the scientific background, the different medicinal chemistry groups attribute various meanings to this term.[18,56]

Consequently, a pharmacophore describes crucial, steric and electronic functional properties that are required for a possible interaction between a protein-ligand-complex and hence the biological bioactivity. The pharmacophore characterizes an abstract concept of common molecular interactions but not a particular structure or real functional groups. These common molecular interactions should be separated by a number of bioactive molecules which have comparable effects under comparable conditions in the active site.[18,54,56,57]

3.2 Overview

Amino acids in the active site and critical functional groups of inhibitors (and substrates) form explicit complementarities for binding. They can be reencoded as a pharmacophore feature in 3D and can define an area or a direction of interaction.[13,18,54,57]

There are various features that can be defined to optimize a pharmacophore with optimal protein-ligand-interaction. The specification of these properties is generally attributed to features like hydrogen bond donor and acceptor properties, hydrophobic groups and aromatic rings, cationic and anionic features[18,57] Figure 14 depicts some available pharmacophore features to choose in the software LigandScout. It is an example of features describing properties of ligands. The tolerance and weight of each feature can be adjusted. In addition, features can be set as optional during the matching procedure.[13] Pharmacophore queries are not only used to find active molecules, but can also be created as negative queries to filter out inappropriate ligands.[57]

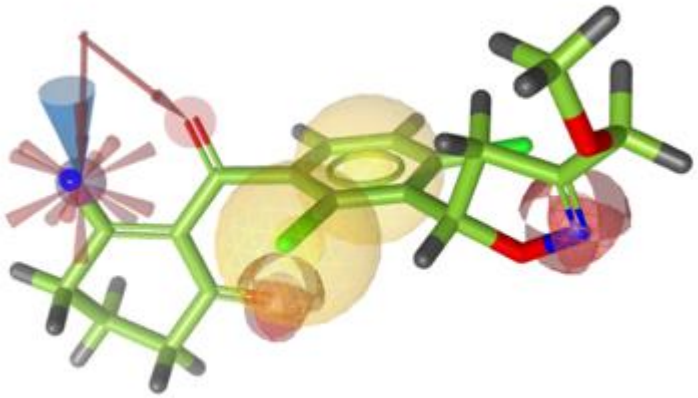


Figure 14: Selected features to define a pharmacophore.

The yellow spheres are hydrophobic interaction. Hydrogen acceptore features are shown as red spheres and can be directed (arrows), as visible at the triketone warhead. A negative ionizable area (star-shaped) and an iron binding feature (blue) are also shown at the warhead. The shown molecule is the co-crystallized ligand from 5YY6. These depicted features can be chosen from the software LigandScout.

Depending on whether the protein, ligand or protein-ligand complex is present, the application of pharmacophore modeling technology is different.[57]

The optimal scenario is to have a known protein structure and the structure of the bound ligand. Normally, pharmacophores reflect the key properties of small molecules that allow them to bind to receptors. But features can also be made from the active site. Ideally, the information of the protein is obtained via the crystal structure. If an experimental structure of the target protein is not available, homology models can be made. As long as an active ligand is known, the information derived from the

protein structure and the ligand interaction can be correlated in order to use a consensus approach to find new ligands.[57]

As soon as the structures of the active ligands are known, but not the structure of the protein, ligand-based pharmacophores can be created. This scenario is the most common case in drug design. However, if only one ligand and no structure is known, this method is not an option.[57]

If only the structural information of the target protein is available without a known inhibitor, it is possible to derive structure-based pharmacophore modeling. The chemical properties of the interactions of interest can be highlighted. There are a number of tools that convert the information into a 3D image.[54,57]

3.3 Pharmacophore Methods

Ligand-based and structure-based pharmacophore modeling are two general techniques that can be applied. Both have advantages and vulnerabilities that make them applicable to certain use cases. The choice between these two approaches depends mainly on the available information.[56–58] Figure 15 shows a decision tree. It simplifies the choice of the methodology available for selection.

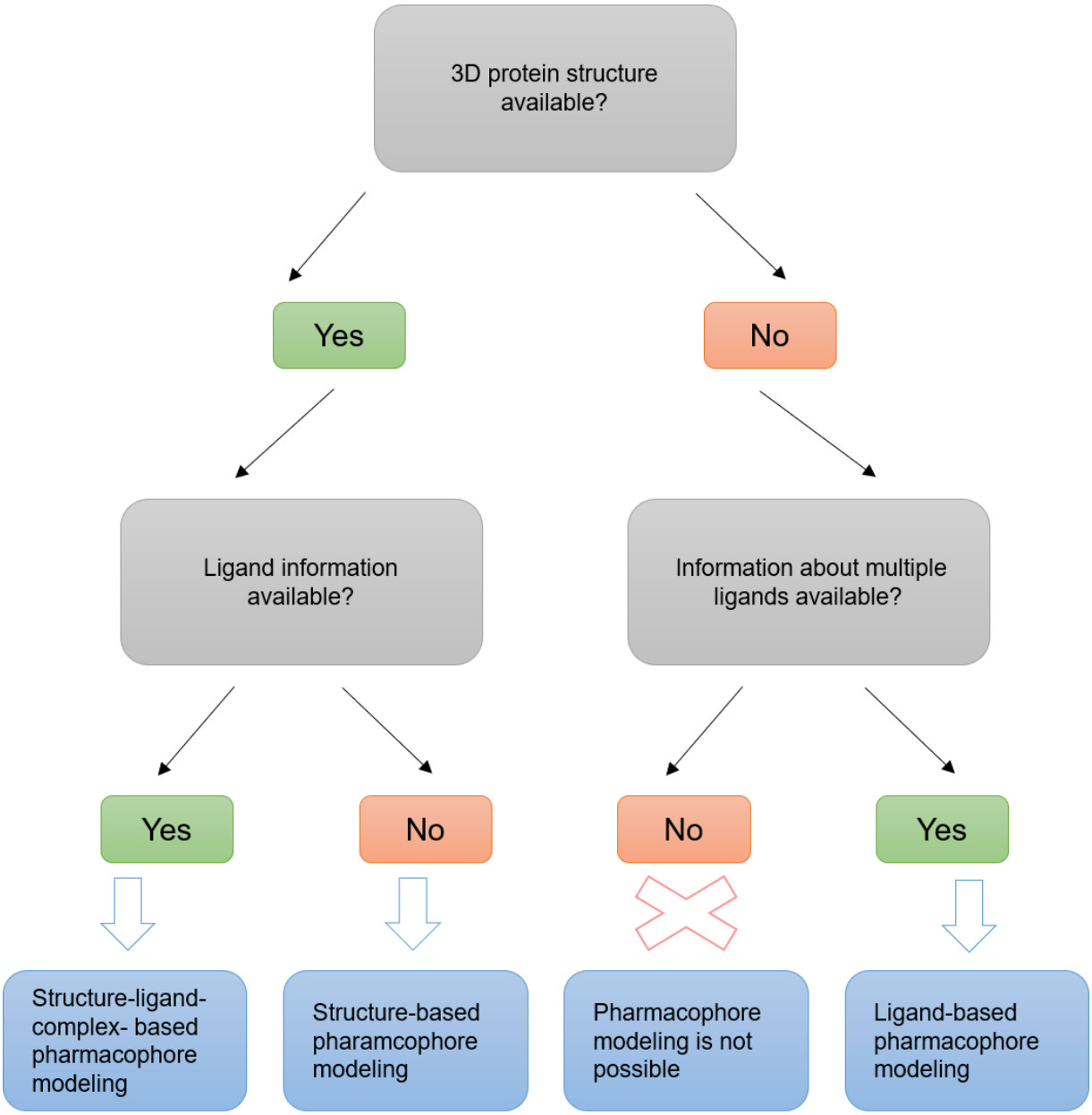


Figure 15: Decision tree for selecting the right pharmacophore method

Depending on the information available the pharmacophore method can be chosen.

3.3.1 Structure-based Pharmacophore Modeling

If the structure of a protein is known, potential ligand-protein interactions can be predicted using the 3D coordinates to prepare structure-based or structure-ligand-complex-based pharmacophore models.[17,56] The interactions emanating from the amino acids of the binding pocket must be complementary to the properties of the ligands in order to achieve optimal interactions.[18,56] The most commonly used method for structure-based drug design is to dock selected small molecules into the bind-

ing pocket. There are a number of docking tools that can be used for this purpose. The aim here is to dock several poses per ligand and evaluate each pose to identify the most energetically pose. The best pose can be determined from the docking scores and the molecules can be compared with each other. Scoring the poses based on the pharmacophore helps to validate the docking result in the context of the reliability of virtual screening for new lead structures in drug design.[59] Virtual screening, comparison of binding sites and ligand binding pose prediction are applications of structure-based pharmacophore modeling.[57]

3.3.2 Ligand-based Pharmacophore Modeling

Ligand-based pharmacophore modeling is used when the structure of the target protein is unknown.[13,58] Although a lot of 3D protein structures are known, many targets are not crystallographically resolved, which is why structure-based drug design is not possible.[57,58] In such case a data set of active ligands is required to derive the ligand-based pharmacophore model. Since the molecules are flexible and the bioactive conformation is not known, conformations have to be generated first.[13,18,57] Then the individual ligands are superimposed with the aid of computer tools in order to identify structural differences and similarities.[13,17,18,57] A ligand-based pharmacophore is created to cover the potential interaction points of all ligands in a 3D space. All crucial structural characteristics should be geometrically superimposed.[13,18,57] The predictive performance of the pharmacophore model depends on the superposition of small molecules and the 3D pharmacophores.[18] The comparison of molecule volumes can provide information on the shape and size of the binding pocket.[13,18,57] Thus, the structure of the unknown binding pocket can be gradually adumbrated.[13]
A set of inactive ligands can help to identify unfavorable features or a sterically unfavorable features that should be excluded.[57]

3.4 Criteria for a satisfying Pharmacophore Model

Based on the quality pharmacophore models are useful tools to identify new actives. As explained in chapter 3.1 and 3.2, the pharmacophore describes functional groups involved in target interaction, including their distances between charged atoms . A good pharmacophore discriminates stereoisomers. The receptor-ligand binding is stereospecific and therefore, stereoisomerism is important for affinity and selectivity of inhibitors. It should also differentiate between active and inactive molecules. There is a possibility that several pharmacophores can be combined to fulfill different conditions. This aims screening large libraries for new chemical starting points in a short time.[18,56]

In summary, an optimal pharmacophore model should cover all potential interactions between a protein and its ligands.[54]

3.5 Pharmacophore Modeling in Virtual Screening

Virtual screening plays an important role among computer techniques for the discovery and development of new actives in crop protection.[5,17,60] Pharmacophore models are mostly used as a baseline technology to identify new hits in databases. [5,59] A novel hit molecule will be found as soon as the molecular structure of a ligand from the database fits the pharmacophore spheres of the predefined pharmacophore model.[5,57] Via this approach, novel compounds can be found and new insights into SAR can be gained.[60]

There is no difference between a ligand- or a structure-based model for using it to search for hits. The number of found hits depends on the information provided in a database and the model used to filter them out.[13] The accuracy of the results depends on the interaction of the ligands and their targets also on the quantitative evaluation technique to the interaction result.[59] In general, databases can be screened using pharmacophore models.[13]

In most cases, ligands with only one 3D conformation are stored in the databases. Since each molecule can assume different conformations depending on its degree of freedom, it is very unlikely that the one conformation stored in the database is the correct one for the optimal ligand-protein interaction.[13,60] Consequently, several con-

formations must first be generated in order to find the correct spatial structure of the ligands that explicitly fits the given model. Especially in databases with a very high number of potential hits, the calculation of conformations takes relatively long time. Therefore, it is helpful to filter the databases in advance using other approaches.[13,18,57,60] For example, the ZINC database is publicly accessible and has pre-adjusted filters, as molecular weight or logP, so that several unsuitable structures are eliminated from several million molecules that have been matched to the search.[13,18] High-throughput screening can thus accelerate the search for first hits.[13,18,57] These hits serve researchers as a source of ideas for new lead structures.[13] The found hits are tested and evaluated experimentally in *in vivo* assays,[13] while conventional hits are tested for their intrinsic activity in *in vitro* assays. Most of the herbicidal compounds have lower Log-P values and fewer hydrogen donors than pharmaceutical drugs. While in herbicide research heteroaromatic cycles, carboxylic acids and sulfonamides are abundant, the functional groups amines and hydroxyl residues are less frequent. Since the requirements for ligands in plant protection are different, the databases must be adapted to the properties of suitable ligands. However, there is no public database with the requirements of herbicide research that can be used for virtual screening. Therefore, the hits from drug research databases must be thoroughly checked, since not every substance found can be applied to an agrochemical solution.[60]

4 Material and Methods

This chapter describes various tools used in this thesis. The focus of this work is on ligand-based pharmacophore modeling to discriminate active from inactive triketone herbicides. In this part, the methodology is explained gradually how the ligand-based pharmacophore modeling has historically grown from the structure-based data.

4.1 Software

There are a variety of software programs that have been used to complete the project. The partly licensed, partly open accessible tools can be used to create pharmacophore models. This allows, for example, to distinguished between active and inactive compounds. Since all programs have both advantages and disadvantages, only certain tools were used for certain specific tasks, as described in this thesis in the following chapters. This includes the dataset analysis, alignment methods, docking and scoring function, and the pharmacophore search with generating conformations, feature definition, and database creation including virtual screening. The tools utilized in this work are listed in this chapter. Companies and academic working groups worldwide use these software for simulation, molecular modeling and machine learning.[61,62]

4.1.1 MOE

MOE (Molecular Operating Environment) distributed by Chemical Computing Group is a commercial modeling platform. Various groups of chemists and modelers have access to computational tools and use the different applications such as protein modeling, SAR analysis, structure-based and ligand-based design and description of pharmacophores for their research.[56,61]

These applications are translated in Scientific Vector Language (SVL) which is designed for an enormous amount of chemical and molecular data.[56,62] This language is easy to read, write and parallelize for the software.[62]

4.1.2 Maestro

Maestro is an all-purpose drug discovery platform provided by Schrödinger and, equally with all other software, is used by research groups. This platform is applied in food, fragrances and plant science.[63]

Schrödinger provides a variety of licensed programs. Phase is that module which is used for the generation of pharmacophores. It utilizes the Maestro visualization program with its interface, that chemical structures can be drawn.[56] Glide is another module from Schrödinger that can be applied for computational docking and scoring. Further applications are de novo design, virtual screening, structure refinement and preparation, biologics modeling, and more.[63] All common molecular file formats are supported by Maestro.[56]

4.1.3 LigandScout

LigandScout is a software platform for computer-aided molecular design designed primarily for the creation and use of 3D pharmacophores. It was founded by the company InteLigand. The user-friendly tool can generate feature-based pharmacophores from structure data of protein-ligand complexes based on crystal structures or from training and test kits of multiple ligands. This tool includes all the common features needed to describe a pharmacophore in order to find the best hits via virtual screening. Many chemists and modelers worldwide use this tool for filtering, searching and prioritizing of molecules.[17,64]

4.1.4 KNIME

KNIME is a graphical, Java-based workflow editor that was developed at the University of Konstanz in 2004. The purpose was to analyze and apply huge amounts of data. In the meantime, it is managed by the Zurich company Knime.com AG. KNIME is a widely used software. The publicly accessible tool is used by bioinformatics, computer science and pharmaceutical research in particular, although it was original-

ly developed for the purpose of data mining.[65–68] The graphical user interface allows interactive visual workflows to be created instead of having to resort to command line programs.[65] This means that even non-experts in programming can rapidly and simply analyses their data using extensions. These "nodes" are developed by KNIME itself.[66] In order to compete with the state of the art in science, KNIME enables its users to make their own nodes available to the public.[66,68,69] The nodes have been developed in various programming languages (C++, C#, Java, Python2x/3x). For advanced applications there are additional Java snippets based on RDKit and Phython scripts.[67,68]

The principle is based on the fact that the exit of one node becomes the entrance of the next node. In this way, many nodes can be linked together, and an entire workflow can be created.[67,68] The data is displayed in tabular form within KNIME and thus forwarded to the next node. After each step it is possible to check how the data has been changed by the executing node.[68]

The core competencies of this tool are data input/output, data mining, machine learning, data and table processing, and visualization. For example, in the field of cheminformatics, there are a large number of nodes for the analysis and manipulation of chemical structures, calculations of molecular weight and determination of similarity.[66]

An entire created workflow can be saved and exported as an archive file with and without data and thus be used by other users.

4.1.5 DataWarrior

DataWarrior is a license-free platform in cheminformatics for the visualization and analysis of chemical data. The tool can filter the rows of a table with this data using chemical intelligence. Various diagrams of chemical data can be displayed graphically. Chemical structures can be viewed in 3D.[70]

This tool can filter by substructures and scaffolds and shows different chemical descriptors. Structures can be grouped, correlations can be displayed, and physicochemical properties can be calculated. Therefore, different view measures can be made visible as required. Finally, new knowledge can be generated through different correlations or existing knowledge can simply be checked.[70]

4.2 Data collection and preparation using KNIME

4.2.1 First Dataset

The first data set was composed of 2085 Smiles codes of in-house ligands, which had to be analyzed for further use. This data set was stored in a tabular form as Excel file. For each Smiles code, a LS-Core consisting of few digits was specified, which is a numerical identification of the chemotype. The bioactivity data were given for each of these molecules in the form of IC_{50} values after being tested on the plants ARBTH or ZEAMX. The IC_{50} values are a measure of the effective dose at which a half-maximal inhibitory enzyme effect is observed. They are given in power notation with the unit mol/L. The substances from the list were tested on ARBTH and ZEAMX with an enzyme assay between 1994 and 2009. The assay is explained in the following chapter 4.3. for 4-HPPD Activity. The property of the enzyme source is also shown in the table. Additional columns with information on the timestamp, assay name and comment are filled in for each listed substance. It was initially unclear whether duplicates were present. It had to be controlled that each structure has its own LS-Core.

4.2.1.1 Preparing the first dataset of ligands

The KNIME workflow imports the Excel list and converts it into a valid format. Furthermore, the Smiles codes are converted into 2D structures and duplicates are removed using the LS-Cores. Afterwards, all columns are removed that are not relevant for the further steps. This workflow is shown in Figure 16, which describes the first steps.

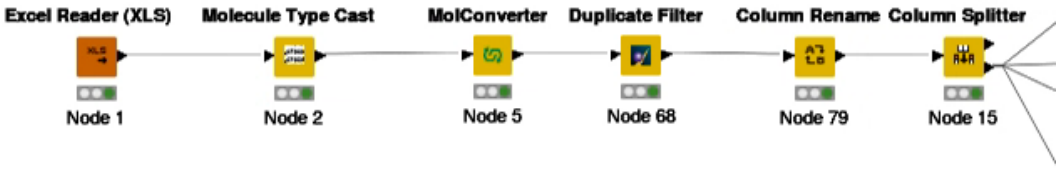


Figure 16: first step of every workflow

This workflow is important to prepare the data so that it can be used in further workflows.

4.2.1.2 Separation of substituted and non-substituted Triketones depending on the enzyme source

As described in detail in chapter 2.3.2, there are triketone herbicides which have additional substituents at the 1,3,5-triketon warhead. In the following workflow described in Figure 17, the substituted compounds are split from the non-substituted compounds by a SMARTS query. At the same time these ligands are divided by their enzyme source. This step requires preparation of the data as in Figure 16.

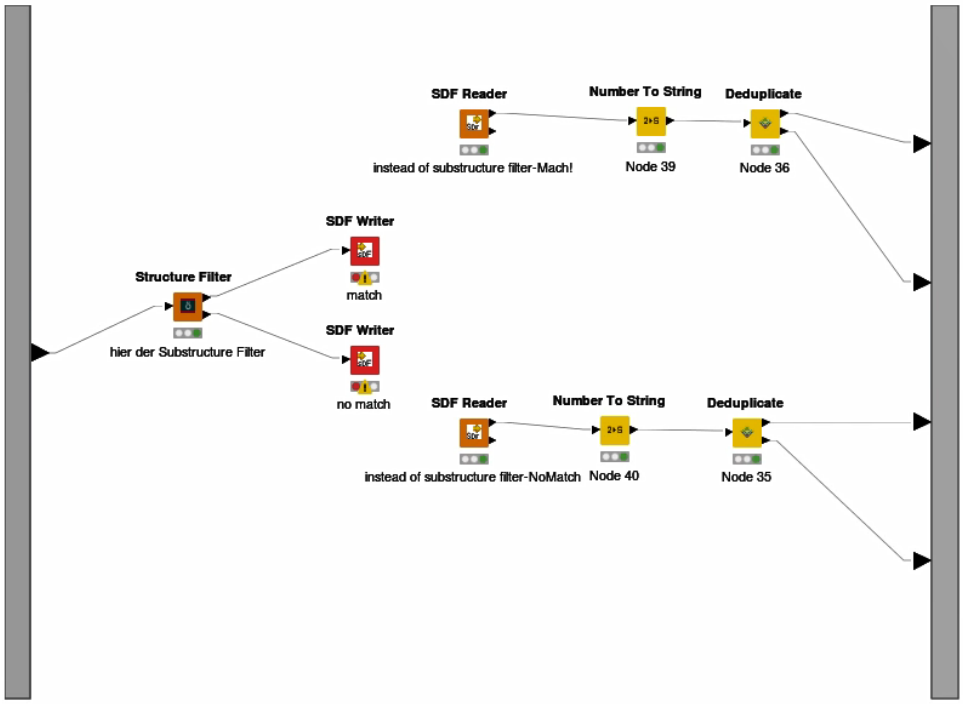


Figure 17: Metanode for better visualization of filtering the substructure of triketones.

In this metanode all required nodes to complete the filtering of the triketone warhead are wrapped.

In the next steps, the enzyme source is separated as presented in Figure 18.
First, the ligands are divided according to substituted or unsubstituted triketones warhead. Then they are sorted according to their enzyme source and ranked in descending order of their bioactivity. Thus, there are six different files with ligands, that are well sorted at the end and can be better evaluated.

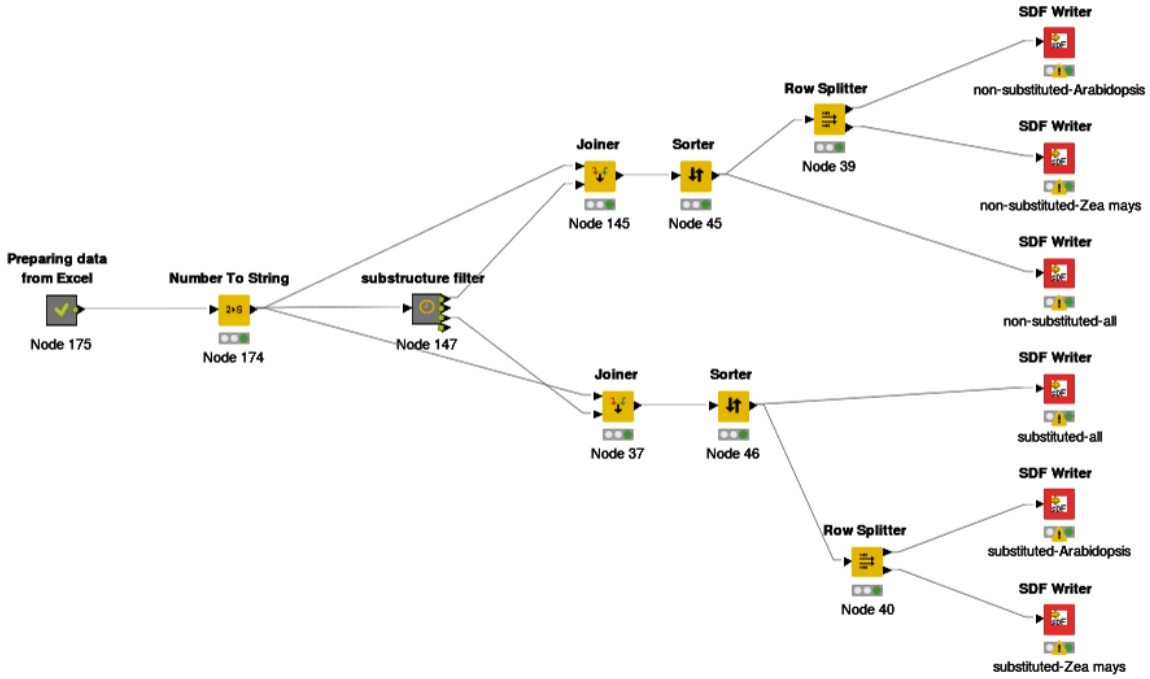


Figure 18: Workflow to separate compounds tested in different enzyme sources.
These triketone herbicides are filtered in substituted and unsubstituted triketone warheads.

4.2.1.3 Duplicates

In the next workflow, which is shown in Figure 17, the duplicates are prepared to be analyzed. The basis of this workflow is again the prepared Excel table with its 2085 Smiles codes, which was described in the previous chapter 4.2.1.1. The duplicates eliminated in the previous workflows (Figure 16) which are now used and merged with the structures retained in the previous workflow (Figure 18). Then, the duplicates are sorted by LS-Core to make them easy to find. Finally, an sdf file can be written and a pdf table can be created.

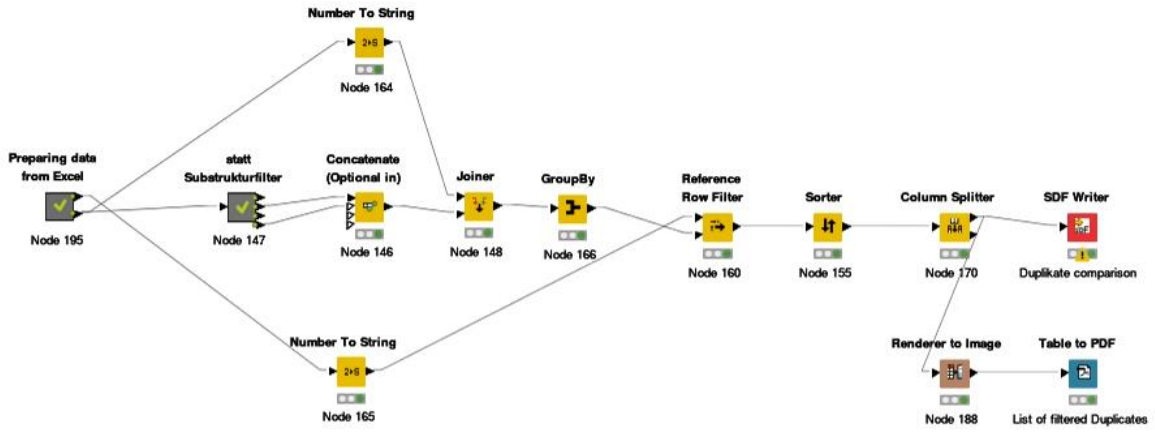


Figure 19: Workflow to analyze the duplicates

This workflow can analyze the duplicate with a substituted warhead and write them out into a pdf file in a tabular form.

4.2.2 Second Dataset

The second data set consists of 517 triketones. Most of these structures are contained in the first Excel file, which is why this data set was prepared, separated and analyzed by the KNIME workflows as before. The 517 in-house structures are stored in the sdf file as a smiles column. Assay name, enzyme source, bioactivity data with IC_{50} values (mol/L), timestamp and comment are given with the corresponding LS-Cores. The deposited substances were tested on ARBTH and ZEAMX with an enzyme assay from 1994 to 2003.

These structures are exclusively non-substituted 1,3,5-cyclohexane-triones, which is the focus of this thesis.

4.2.2.1 Preparing the second dataset of ligands

Figure 20 shows that the workflow is used to read the Excel file and convert it to a valid format. The duplicates were removed. This workflow serves as the basis for the further steps and is similar to the workflow described in chapter 4.2.1.1.

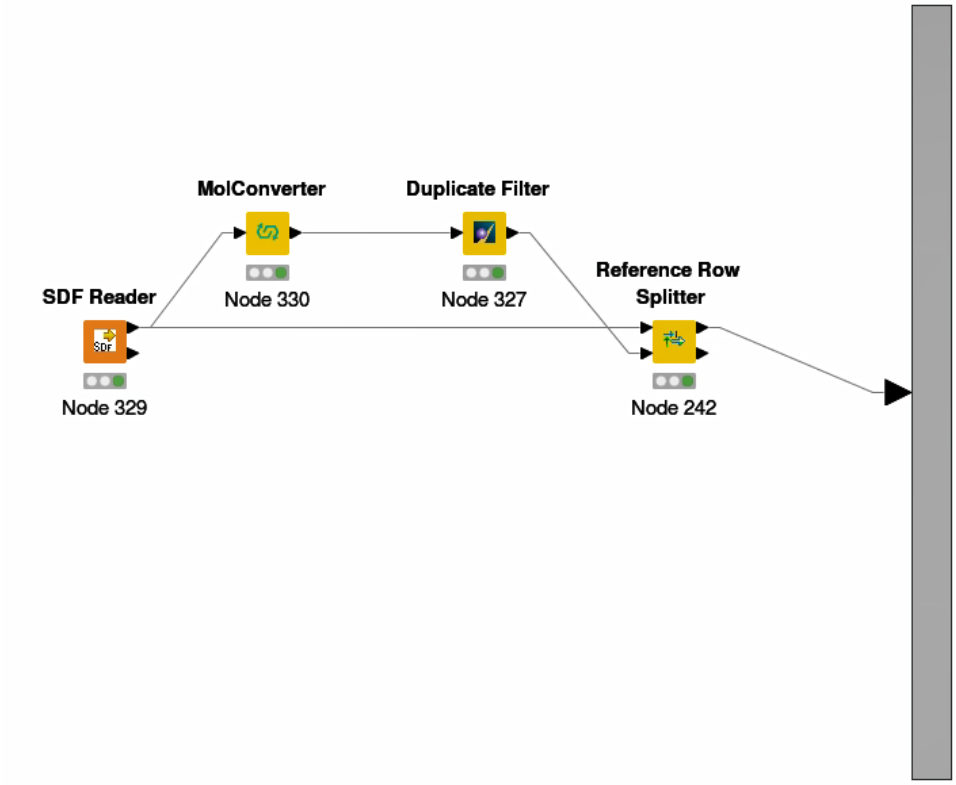


Figure 20: Metanode that includes the first 4 nodes for every following workflow with this dataset.

This metanode is the base and prepares the in-house data stored in a sdf file.

4.2.2.2 Duplicates

This dataset was investigated for duplicates in the same way. With this workflow, which is shown in Figure 21, the structures that were present at least twice in the data record can be written out and sorted according to their LS core and time stamp. Finally, the data is saved in tabular form as a pdf file.

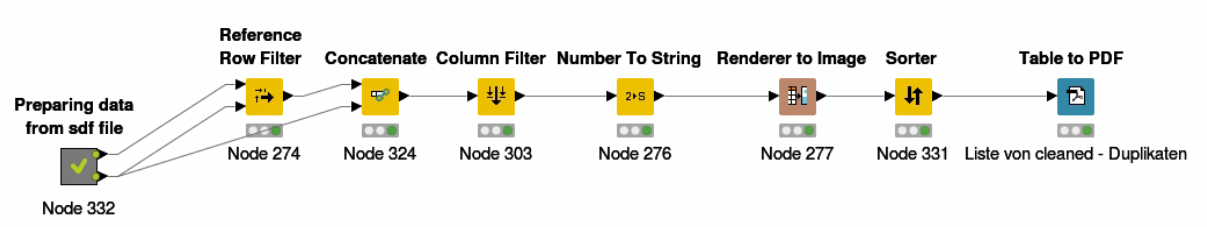


Figure 21: Workflow to analyze the duplicates of the non-substituted warhead.

This workflow can analyze the duplicate with a non-substituted warhead and write them out into a pdf file in a tabular form.

4.2.2.3 Differentiation of the Bioactivity regarding the Enzyme Source

This workflow, which is shown in Figure 22, involves splitting the data according to their bioactivity, which has been tested on ARBTH or ZEAMX. Their IC_{50} values are the benchmark against which bioactivity is measured. After the initial sdf file as described in chapter 4.2.2.1 has been processed, the workflow can be continued. After the columns that are not important for this workflow have been removed, the data is sorted according to their bioactivity in descending order. The ranges are defined which describe inactive (IC_{50} value below 1.0E-6), moderate active (IC_{50} value between E-5 and E-6) and active (IC_{50} value of minimum 1.0E-7) compounds. In the next step, the ligands are split according to their enzyme source ARBTH or ZEAMX and then written out.

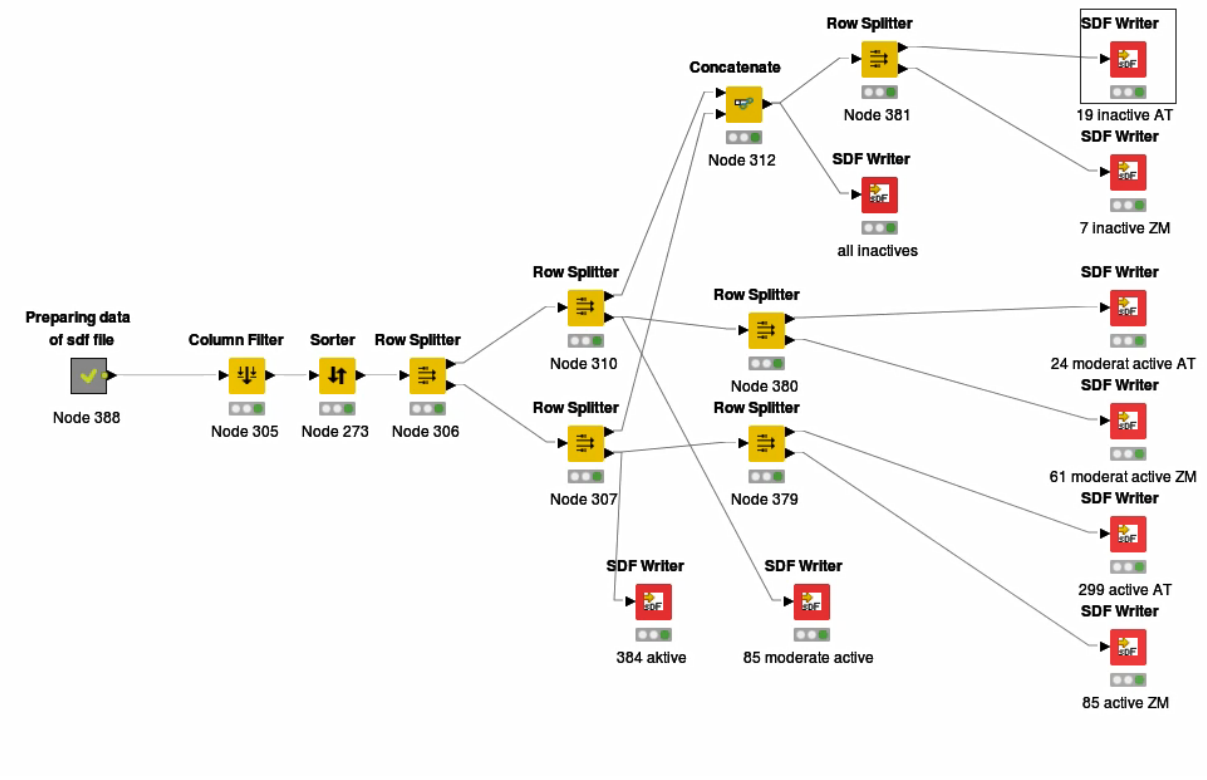


Figure 22: Distribution of non-substituted triketones according to their bioactivity.
Further can be divided according to the enzyme source depending on the EC-50 values.

4.3 Assay for 4-HPPD Activity

The first and second ligand data sets were tested with an assay noted before the age of digitization. The enzyme assay protocol is no longer traceable and therefore there is no information on how the assay was performed in the past. In the meantime, the assay has improved and therefore only the current assay is described under which the retesting of some compounds was performed.

The basis of the activity test for 4-HPPD compounds was based on the analysis of homogentisate by reverse phase HPLC. The assay mixture was prepared from a mixture of 150 mM potassium phosphate buffer with a neutral pH of 7.0, 100 µM catalase, 50 mM L-ascorbic acid, 1 µM FeSO4 and 0.2 units purified 4-HPPD enzyme. 1 unit corresponded to the amount of enzyme required to convert 1 mmol HGA per minute at 20°C. IC_{50} values, means the values of the concentration of an inhibitor necessary to block half of the enzyme 4-HPPD, were measured by a dilution series compared to the control treatment. For the experiment the ligands were dissolved in dimethyl sulfoxide and used in the highest concentration $5x10^{-6}$. To normalize the results, the uninhibited enzyme activity was set to 100 percent. The IC_{50} values were calculated by nonlinear regression. The mixture of the enzyme assay and the ligands was incubated for half an hour. Afterwards the 4-HPPD concentration was increased to the final concentration of 0.05 mM. Phosphoric acid was also added to the final concentration of 400 mM. 96 well plates were filtered with a PVDF filtration device with a pore size of 0.2 µM. Five microliters of the sample were analyzed on a UPLC HSS T3 column (particle size 1.8 µM, dimension 2.1 x 50 mm; Waters) by isocratic elution with 90% 20 mMNaH2PO4 (pH 2.2) and 10% methanol (v/v). Thus, HGA could be detected electrochemically at 750 mV and peak ranges could be quantified. The Michaelis constant (Km) and maximum reaction rates (Vmax) could be calculated by nonlinear regression, as well as k_{cat} values, which indicates a 100% purity of the enzyme preparation. Weighted mean values due to standard errors of the Km and IC_{50} Q10 values were calculated from a minimum of 3 independent experiments. The dissociation constant (Ki) was calculated by the Cheng-Prusoff equation. Using the tolerance index TI, the 4-HPPD enzymes could be compared and ranked. This index depends on the lack of sensitivity to the inhibitors and the respective protein activity.

4.4 Protein Preparation of 5YY6 and 1SP8

The PDB contains several crystallized protein structures of the enzyme 4-HPPD of ARBTH. The crystal structure with the 4-letter code 5YY6 (Figure 23) was finally selected because this structure binds a co-crystallized ligand of the triketone class in the binding pocket. It has the best resolution of 2.4 Å compared to the other 4-HPPD crystal structures with a co-crystallized triketone. The x-Ray crystal structure was loaded into the PDB and published in 2017. The perfect starting point is when the structure is present together with its ligand. Structure 1SP9 (Figure 23) is a protein crystal of ZEAMX, but it has no ligand crystallized in the binding pocket. Only the protein itself is available. This is the only stored crystal structure of ZEAMX. It has a resolution of 2 Å and was published in 2004. This thesis focuses on these two protein structures of 4-HPPD. The structure 5YY6 could not be crystallized with a bivalent iron ion. Therefore, a divalent cobalt was used for the crystallization. However, the protein is bioactive only with a Fe^{2+}. For this reason, this Co^{2+} must first be replaced by the Fe^{2+} in the interface of the selected software in order to obtain the correct binding between macromolecule and compound.

Figure 23 shows all 4-HPPD protein crystal structures located in the PDB. Among them are also the selected structures 5YY6 and 1SP9. As explained in chapter 2.1, the protein in plants is a homodimer. It is possible that a different number of chains are present, since crystals of different forms can be grown during crystallization, down to just one crystal.

Before continuing to work with the structure, this structure must be prepared. The software tools mentioned above can load the protein structure with the information stored in the PDB into their interface. However, there are some amino acids that cannot be displayed or appear in another orientations. The programs can recalculate the missing amino acids and display them in the interface in their probably correct form. Because hydrogen atoms cannot be visualized in an x-Ray crystal, they have to be added by the software first.

Maestro can use the Protein Preparation Wizard to prepare the 3D protein in 3 steps. First the refinement is started, then modifying and finally can be minimized. The MOE software automates the steps with the QuickPrep tool. First, the structure is prepared. Hydrogens are set and ionization states are assigned by the system and then modified. The potentially lowest energy for different states of the terminal function groups

is searched across the entire system. Finally, the protein structure is refined and prepared that it can be utilized for further modeling.

4-letter code	Enzyme Source	Chains in PDB	Resolution in °A	Inhibitor backbone	Compound 1
5YWG	Arabidopsis thaliana	2	2,6	Triketon	Co2+
5YWH	Arabidopsis thaliana	2	2,72	Pyrazol	Fe3+
5YY7	Arabidopsis thaliana	2	3,3	Triketon	Co2+
6ISD	Arabidopsis thaliana	2	2,4	Triketon	Co2+
6J63	Arabidopsis thaliana	4	2,62	Triketon	Fe3+
5YWI	Arabidopsis thaliana	1	2,58	Triketon	Co2+
5YWK	Arabidopsis thaliana	2	2,8	Triketon	Co2+
1TG5	Arabidopsis thaliana	1	1,9	Pyrazol	Fe2+
1TFZ	Arabidopsis thaliana	1	1,8	Pyrazol	Fe3+
5YY6	Arabidopsis thaliana	1	2,4	Triketon	Co2+
1SP9	Arabidopsis thaliana	2	3	/	Fe2+
5XGK	Arabidopsis thaliana	4	2,8	/	Fe3+
1SQD	Arabidopsis thaliana	1	1,8	/	Fe3+
3ISQ	Homo sapiens	1	1,75	/	Co2+
5EC3	Homo sapiens	1	2,1	/	Co2+
1CJX	Pseudomonas fluorescens	4	2,4	/	Fe2+
5HMQ	Pseudomonas putida	6	2,37	/	Mg2+
1SQI	Rattus norvegicus	2	2,15	Pyrazol	Fe3+
1T47	Streptomyces avermitilis	2	2,5	Triketon	Fe2+
3ZGJ	Streptomyces coelicolor	2	1,95	/	Co2+
1SP8	Zea mays	4	2	/	Fe2+

Figure 23: Selection of all x-Ray protein crystal structures of the enzyme 4-HPPD.
This figure shows a tabular list of all published 4-HPPD structures from the PDB. There is only one structure of ZEAMX with a resolution of 2.0 Å and was crystallized with the Fe^{2+}. In contrast to the selected ARBTH crystal structure 5YY6 it does not contain any ligands. 5YY6 has been crystallized with Co^{2+} and a triketone with a resolution of 2.4 Å.

4.5 Ligand Preparation and Alignment

Similar to the structure, the ligand set of non-substituted triketones must first be prepared. The set is loaded into the Maestro interface and modified using the tool LigPrep. All selected ligands are checked in the force field OPLS3e. The settings allow that tautomers may be generated, possible ionization states at target with pH between 5 and 9 are generated. In this way, any charged ligands are purified.
Furthermore, it can be adjusted how chirality, which influences the conformers generation, is handled. In total it is set that a maximum of 32 conformations are generated per ligand. After the calculations all prepared ligands are saved as sdf or maestro output file.
Now the alignment of the ligands can be started. Ligand 3D alignment can be difficult because the ligands are usually flexible. Therefore, it is important to determine a ref-

erence structure that is in the correct conformation that is needed. Once the reference structure is selected, the 3D alignment can be commenced. The tool Flexible Ligand Alignment from Maestro can be used. When the docking pose of the ligands is known through the crystallized protein-ligand complex, this conformation can be used as a reference. The ligand with the triketone warhead must be present in the enol tautomerism to form a chelate bond to the Fe^{2+}. The triketones are existing in this desired enol tautomeric in neutral pH, which was previously adjusted with the LigPrep panel. All ligands to be aligned must be selected and the reference structure must be at the top of the project table to signal to the system that it is the template. There are different methods for the settings. The best results are obtained with the common scaffold alignment method using a SMARTS pattern. All selected ligands orient themselves with the defined SMARTS pattern O=CC=C[O-] to the reference structure and superimpose. It describes the two keto enolate oxygens of triketones that interact with the bivalent iron.

4.6 Docking using Maestro

The docking of several ligands requires a ligand set and a 3D crystal structure of the enzyme 4-HPPD. For this purpose, the software Maestro was chosen. The prepared x-Ray crystal structure from the PDB with the code 5YY6 is retrieved. The ligands to be docked were split into different categories using KNIME. In order to find the best setting for the docking results, a redocking of the crystallized ligand from the crystallized protein-ligand complex was performed at the beginning. This ligand was extracted and stored in the Maestro interface. To define the binding pocket, a grid file must be created using the module Glide (Grid-based Ligand Docking with Energetics). A purple grid box automatically appears defines the binding pocket for the ligands to be docked in the future. In order to optimize the conditions, further settings are necessary. As shown in Figure 24, the grid box can be set under the tap Site that the ligand is centered in the grid box. The X, Y, and Z coordinates and the size of the box are displayed automatically as soon as the binding pocket has been defined. It is advantageous to redefine these parameters as integers that the compilation of the grid file can be better repeated and reproduced. With this setting shown in Figure 24

the box is large enough for the ligands to be docked and smaller than with the default.

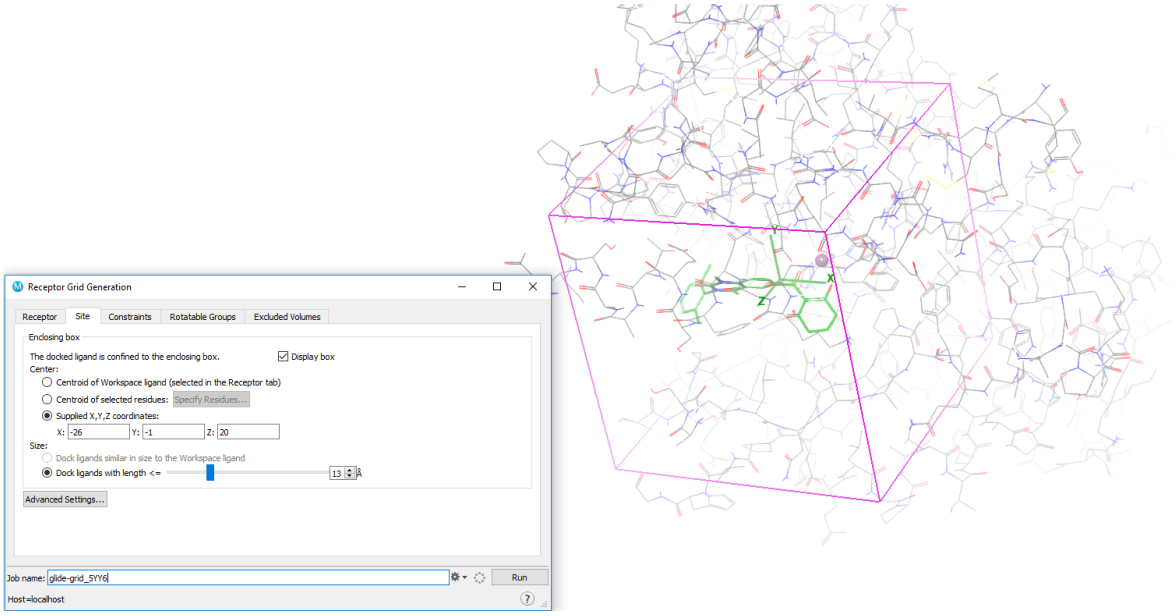


Figure 24: Grid box around the co-crystallized ligand with optimized settings.

The grid box specifies the binding pocket into which the potential ligands are docked. With the Receptor Grid Generation the conditions for the grid file can be improved. In the figure the settings for the active site are adjusted. The grid box should be as small as possible, but large enough to increase the docking results and to enable an optimal binding to the protein.

Using the tab Constraints, conditions can be set that are essential for the binding and must be adhered for the calculation of the docking. These constraints are part of the protein in the binding pocket and close to the ligand. Since the binding mode is detected, it is known that chelating to cationic iron is fundamental for bioactivity, a metal coordination constraint is additionally defined. The ligand should maintain an optimal coordination with the metallic atom at a given distance. The constraints spheres are centered at the iron ion and therefore directional to the ligand. Figure 25 shows the directed constraint spheres with the ideal position. The upper sphere in orange color is directed against the protein and is therefore not targeted during docking. The red sphere protrudes directly in the direction of the non-covalent ligand binding.

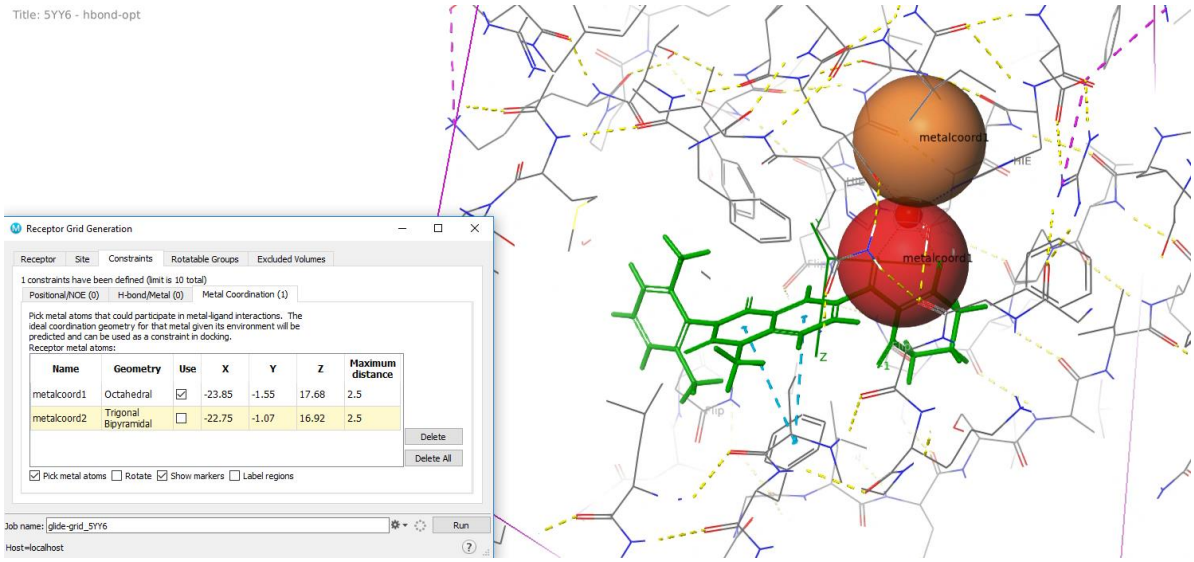


Figure 25: Metal Coordination Constraints.

The two-colored constraints spheres are directed around the iron ion. The upper sphere protrudes into the protein and is therefore irrelevant for docking. The red sphere is directed to the keto enolate oxygens of the ligand. The potential ligands with the same motif should be bound at the right distance to the iron ion under the constraint of a non-covalent metallic bond.

The ligands with a substituted or unsubstituted warhead filtered by KNIME were docked with the created grid file and the optimally adjusted properties. Ligands, depending on the potency of their bioactivity (inactive, moderate active and active) and ligands that differ in which enzyme they were tested, were also docked. The results of docking are explained and discussed in chapter 5.4 Docking Results.

4.7 Protein Structure Analysis

4.7.1 Protein Alignment

There are some deposited 4-HPPD crystal structures in the PDB. 13 of them are from ARBTH and a single protein crystal structure is from ZEAMX (Figure 23). Chapter 4.4 explains why the X-Ray crystal structures 1SP8 and 5YY6 were chosen as references. A protein alignment was performed with these two structures, which is described in this part. The crystal structure of 5YY6 has only one chain. However, 1SP8 has four chains. The two pdb files are imported into the Maestro interface of Schrödinger software and are prepared. Chapter 4.4 also describes how proteins are pre-

pared with the tool Protein Preparation Wizard that the structures can be further processed.

Then the four chains of 1SP8 must be extracted. The entry in the project table is selected and expanded. The chain to be extracted is selected and exported as a new entry. Consequently, there exist five entries to be aligned in the project table. For the next step the tool Protein Structure Alignment from Prime is needed, which is controlled via the Maestro interface.

First the ribbons of the entries are displayed, because the program uses matching of secondary structure elements for the alignment. For the reference structure to which the other chains are aligned, 5YY6 should be selected, because only one chain is crystallized. Only selected entries can be aligned, so all entries have to be selected at first. The residues of the reference enzyme to which the alignment refers can be specified as available atoms, residues, chains, molecules, entries and secondary structure and all atoms of the reference. All atoms of the entries of ZEAMX are superimposed on the chain of ARBTH and visualized directly in the interface. The amino acids of the respective chains of 1SP8 are represented by the amino acid sequence and the calculated alignment score is reported. An Alignment Score lower than 0.6-0.7 indicates a good alignment and the lower the value, the better the alignment. The superposition of the secondary structure residues and the alignment score can be used to decide which chain of ZEAMX should be used for further investigations of the protein structure.

4.7.2 Amino acid sequences and similarity analysis

The MOE software (Chapter 4.1.1.) prepares the proteins of 5YY6 and 1SP8 with the QuickPrep tool. Then the chains can be aligned. In the working space they are visually superimposed and the amino acids of the enzymes can be analyzed. The overall identity of the protein is calculated and the amino acid sequence of both aligned chains is given.

The similarity of the binding pocket is determined by the amino acid sequences. All amino acids interacting with the amino acids of the protein within a 4.5 A radius of the ligand are displayed.

The alpha helices H11, which are known to have been crystallized in closed conformation at ZEAMX (PDB: 1SP8) and in open conformation at ARBTH (PDB: 5YY6) (Chapter 2.2.1.), are investigated more closely with the use of protein alignment. The sequences are analyzed and the interactions between the helices and their protein are investigated. The sequence editor of MOE is used for this sequence analysis. Helices H11 Sequences.

4.8 Ligand-based pharmacophore modeling

Ligand-based pharmacophore modeling can be initiated out of 2 different starting situations. Often the exact binding mode is unknown and therefore a large number of conformers are generated per ligand in order to exhaust all possible modes. After the crystal structures of the 4-HPPD enzymes of ARBTH and ZEAMX are known and a docking of the non-substituted triketones has been performed, the docking poses created by the Schrödinger software can be utilized. The usage of docking poses, closest to the correct binding modes of the ligands, is the second possibility to start the ligand-based pharmacophore modeling. The software LigandScout which is described above was used to create pharmacophores.
In principle, pharmacophore modeling would have been possible only from the data of the ligands. However, in this work the ligand-based pharmacophore models were generated and related to the structure. Additionally, they are optimized by the knowledge gained from structural analysis. Since both crystal structures of ZEAMX and ARBTH were present, no pharmacophore models were created without utilizing the crystal structure.

4.8.1 Preparing 3D-pharmacophore generation

The docking poses of all unsubstituted triketones as shown in Figure 26 can be imported into the ligand-based perspective in LigandScout. The crystal structure 5YY6 with crystallized triketone and Co^{2+} are imported into the structure-based perspective.

First, in this perspective the Co^{2+} is replaced by a Fe^{2+}, which is important for the bioactivity of the enzyme.

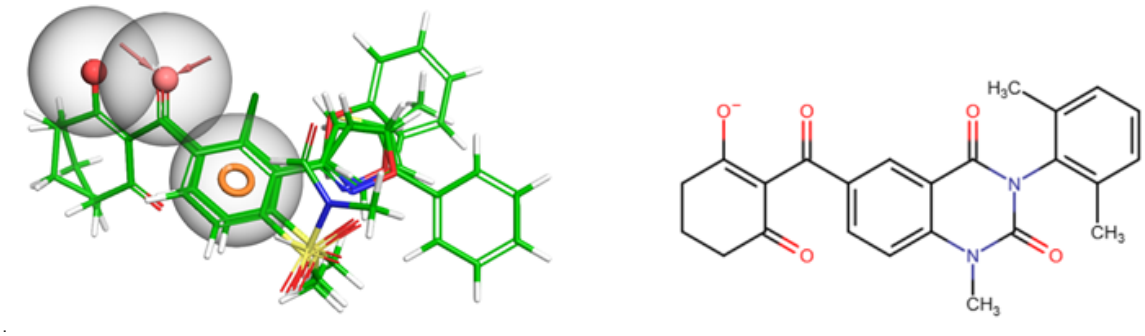


Figure 26: Selected docking poses

A few docking poses of unsubstituted triketones are displayed. Using anchor-based docking, accurate poses could be created that bind the same position as HPP. The pharmacophore spheres visualize that the identical warheads and the adjacent benzyl residue indicate essential pharmacophore features. Pharmacophore models from inactive compounds. In addition to the docking poses, a 2D structure of the co-crystallized triketone from ARATH (PDB: 5YY6) is shown for illustrative purposes.

4.8.2 Pharmacophore models from inactive compounds

The inactive compounds were also docked to get a docking pose which is in the correct binding mode. This has the advantage of not needing to generate many conformers when creating pharmacophore models, which would have been another option. There is a total of 26 inactive molecules from the data set with non-substituted triketones that are imported into the ligand-based perspective. First, a ligand is selected and replaced as a core molecule. The ligand is placed in the binding pocket in the desired conformation, i.e. the docking position. Figure 27 shows the four steps how a pharmacophore is generated and enhanced. A default for a pharmacophore can be generated automatically by the program (Figure 27, a)). Two hydrogen acceptor features are formed, which cause a chelation of the oxygen atoms of the 1,3,5-cyclohexane-triones. Because the triketone is present in its enol tautomerism, a negative feature is added. Further specific features for each ligand are also displayed. By the exchange in the bioactive Fe^{2+} additionally an iron binding feature can be displayed. (Figure 27, b)).

Since only those features are suggested that interact with the protein, it is necessary to verify which features are missing and whether all proposed features are useful. An

additional hydrogen acceptor feature is added to the last feature free ketone of the triketone warhead (Figure 27, c)). This feature is not essential for the binding but assists in subsequent screening to filter for triketones. Because these features appear in the same position in each triketone, the tolerance is reduced (Figure 27, d)). When screening databases the conformations with the correct binding modes are matched and optimizes the hit search.

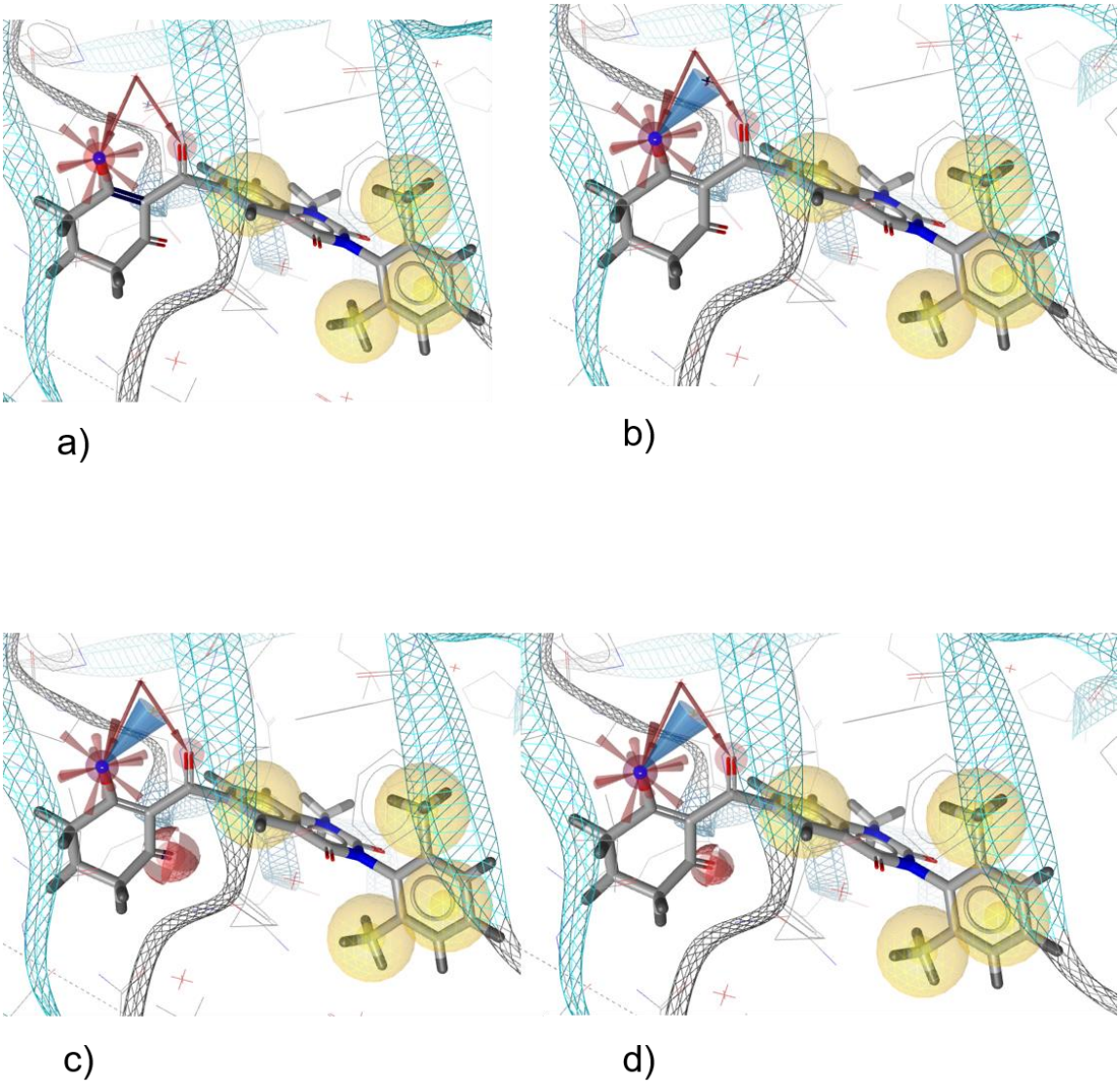


Figure 27: 4 steps of preparing a ligand-based-pharmacophore model.

a) the co-crystallized ligand of the protein crystal structure (PDB: 5YY6) is located in the binding pocket. An automatically generated pharmacophore model is displayed. There are 2 hydrogen acceptor features at the triketone warhead facing towards Co^{2+}. The first hydrophobic feature after the hydrogen acceptor features plays an important role in the π-π stacking between 2 phenylalanines. The other features are specific for this one ligand. b) Co^{2+} was exchanged into a Fe^{2+}. An additional feature for an iron binding is displayed. c) another hydrogen acceptor feature is added for the detection of triketones in virtual screening. d) the tolerance of the hydrogen acceptor feature is reduced to support the adherence to the docking poses when screening ligands with their conformations.

In addition, the exclusion volumes can be displayed which delimit the binding pocket to the protein. The first pharmacophore model is developed and can be sent to the screening perspective. This step is repeated for each of the 26 inactive ligands. In order to facilitate the work, the ligand set was clustered, and it was ensured that no conformation of the ligands was generated. There have been created 15 clusters of 26 molecules. However, because clustering did not make sense, this possibility was rejected.

All triketones show 2 hydrogen acceptor features at the oxygen atoms of the ligand. Additionally, there is a hydrophobic feature, which is related to the phenylalanines 381 and 424 via a π-π stacking interaction.

The aim was to generate pharmacophore models that match many inactive and no active ligands during virtual screening, which is explained in chapter 4.8.9.

4.8.3 Exclusion volumes

The data set of inactive ligands contains a few docking poses, whose size could be the limiting factor, because from a few ligands no docking pose could be generated. The tight binding pocket could lead to steric hindrance. Subsequently, these ligands were forced into the given binding mode to obtain a conformation for generating a pharmacophore models to classify inactive compounds. A pharmacophore model was created using exclusion volumes, with the aim that oversized ligands would not be found as a hit. Exclusion volumes are displayed spheres that define the binding pocket on the protein side. It therefore shows the maximum size of the binding pocket.

In the structure-based perspective the exclusion volumes can be indicated as soon as a pharmacophore has been created from a ligand embedded in the binding pocket. For a model that should not hit ligands that are too large, only docking poses of ligands of acceptable size were selected from the set. It was started with a first ligand. The docking pose of this ligand replaced the previous core molecule in the binding pocket. A pharmacophore is created and the spheres of the exclusion volume are made visible. Then only the pharmacophore without the preselected ligand is moved into the alignment perspective. This step is repeated with a representative number of further ligands of the ligand set. After twenty different pharmacophores are present in

the alignment perspective all pharmacophore features are rendered visible. Only those pharmacophore features are retained which respect the exclusion volumes and which define the triketone warhead features (Figure 28). The remaining features will be removed. The features are interpolated that all query features must be hit. As illustrated in Figure 28, a pharmacophore is created from several pharmacophores. The feature tolerance and feature weight can be increased and decreased and is thus optimized. The finished pharmacophore model of the exclusion volumes can be moved to the screening perspective and is available for virtual screening of data bases.

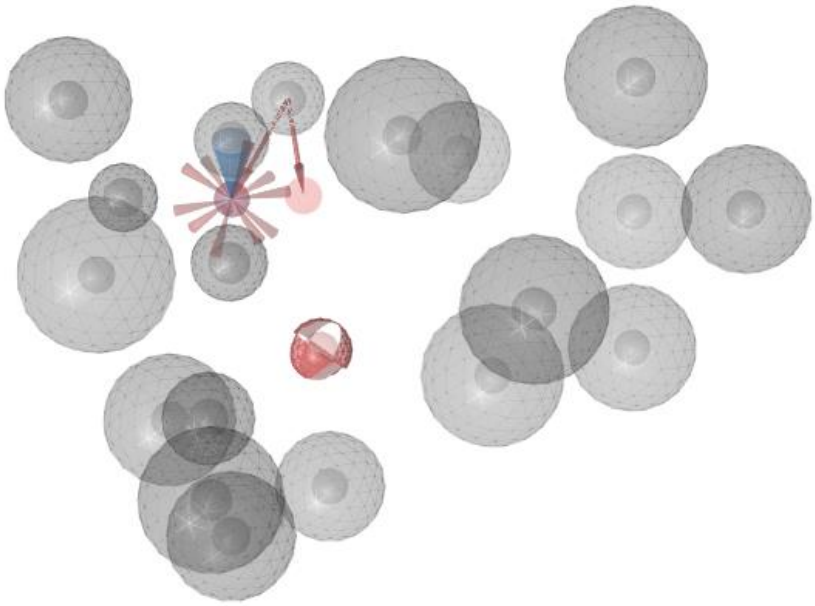


Figure 28: Exclusion volumes separate the protein from the binding pocket.
The size of the 4-HPPD binding pocket (PDB:5YY6) can thus be defined. The remaining pharmacophore features that define the triketone warhead are contained at the correct position.

4.8.4 Pharmacophore model of “Top 2 Bioactive”

After pharmacophore models of inactive compounds and exclusion volumes have been generated, pharmacophore models are created that hit active compounds. First, the two most bioactive docking poses were taken from the entire ligand set of non-substituted triketones. Pharmacophores of these two ligands were generated with the

LigandScout software according to the same principle as in the last chapters and shifted into the alignment perspective. Together with the exclusion volumes pharmacophore these three pharmacophores are merged by reference points. The result is a pharmacophore model that is supposed to find active triketones,

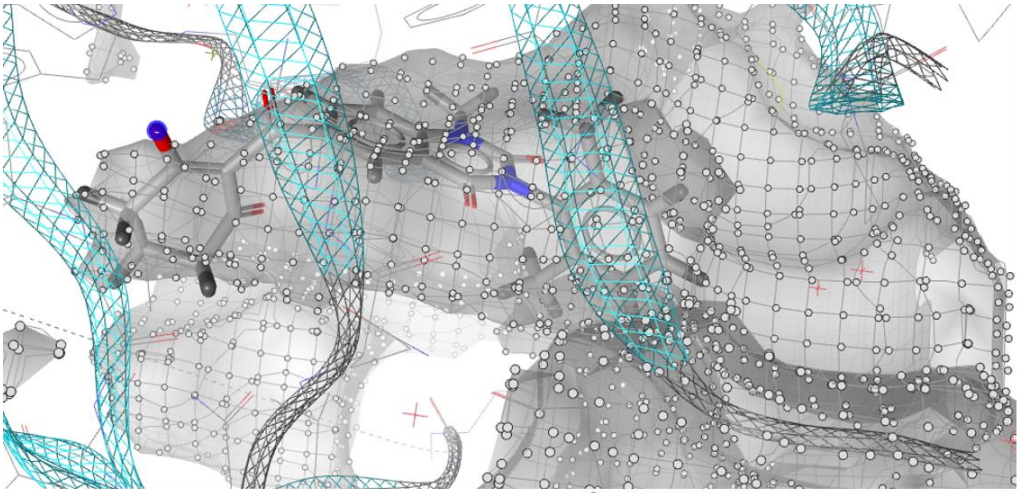
4.8.5 Pharmacophore models from the 10 most active compounds of Arabidopsis thaliana and the 10 most active compounds of Zea mays

For a further pharmacophore model only the 10 most active ligands of ZEAMX were selected. Their respective docking poses were used. From each single pose a pharmacophore model was generated and sent to the screening perspective, where these 10 pharmacophores were merged into a single pharmacophore by referent points.
The same procedure was applied to ARBTH. For each of ARBTH's 10 most bioactive docking poses, one pharmacophore was created, which was subsequently merged into one common pharmacophore.
This resulted in the two pharmacophore models "Top 10 Arabidopsis” for the best 10 substances tested on ARBTH and "Top 10 Maize” for the most bioactive 10 substances in the ZEAMX model. The two merged pharmacophores were transferred from the alignment perspective to the screening perspective in the program. There, data sets of ligands are screened using pharmacophore models. This allows the models to be tested and evaluated.

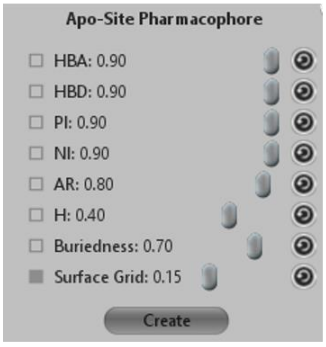
4.8.6 Apo Site Grid Pharmacophore modeling

With the Apo Side Grid tool from LigandScout all possible pharmacophore interaction of the binding pocket can be demonstrated. As can be seen in Figure 29 a), a 3D mesh appears that reflects the surface of the binding pocket. An Apo Site pharmacophore panel (Figure 29, b)) appears, which can be used to create a pharmacophore. In this panel the threshold values can be changed for different features. With the Apo Site Settings dialog (Figure 29, c)) the maximum number of features to be generated

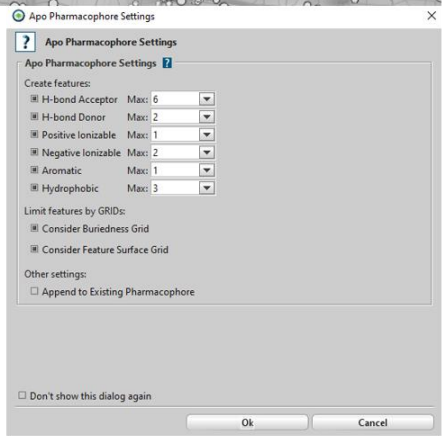
can be changed. Figure 29 shows the settings selected for 4-HPPD. The pharmacophore is automatically generated, and the maximum number of features appears (Figure 29, d)).



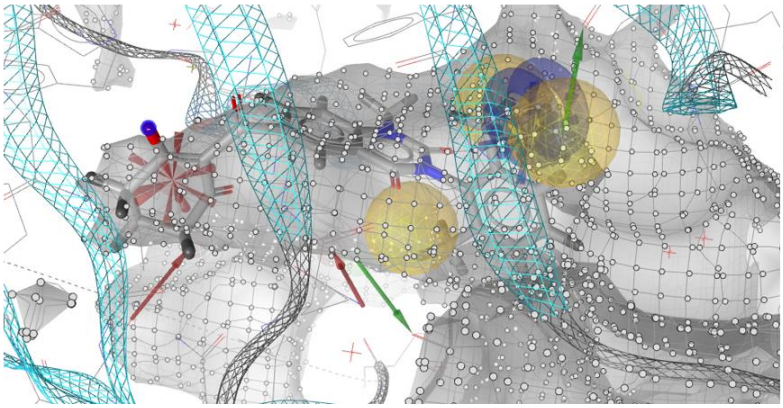
a) 3D surface of the binding pocket



b) Apo Site Pharmacophore Widget



c) Apo Pharmacophore settings



d) Created pharmacophore

Figure 29: Apo Site Grid pharmacophore generating.

a) With the Apo Site Grid tool, the binding pocket is calculated and represented by a 3D surface. Depending on the properties of the protein, each point on the grid can be an interaction point and a feature can be attached manually. b) At the same time, the Apo Site Pharmacophore panel opens. It shows different properties of the binding pocket. The Create button can be used to create a pharmacophore. c) The Apo Pharmacophore setting panel appears. The maximum permitted features can be

adjusted in the settings in order to generate the best possible pharmacophore. d) The automatically generated pharmacophore is displayed in the working space with the adjusted settings. The co-crystallized ligand of ARBTH (PDB: 5YY6) in the binding pocket is present for visualization purposes only. It does not participate in the creation of the pharmacophore model.

The Apo Side Grid tool calculates the pharmacophore model independently of a ligand in the binding pocket and is therefore optimal for structure-based pharmacophore modeling. After the protein-ligand-complex has been crystallized and therefore the binding mode is known, the knowledge of the ligand-based pharmacophore model from this chapter 4.8 can be utilized to optimize the Apo Site Grip pharmacophore.
In combination with the protein-ligand interaction fingerprint (Figure 44) and the automatically created Apo Site pharmacophore models, the pharmacophore models "Apo Site Donor" and the corresponding model "Apo Site Acceptor" are created, which has a hydrogen acceptor feature instead of the hydrogen donor feature. Gln 307 has the ability to form an interaction to a hydrogen acceptor or hydrogen donor. These two pharmacophore models can be seen in Figure 30. The second directed hydrogen acceptor feature interacts with Gln 293.

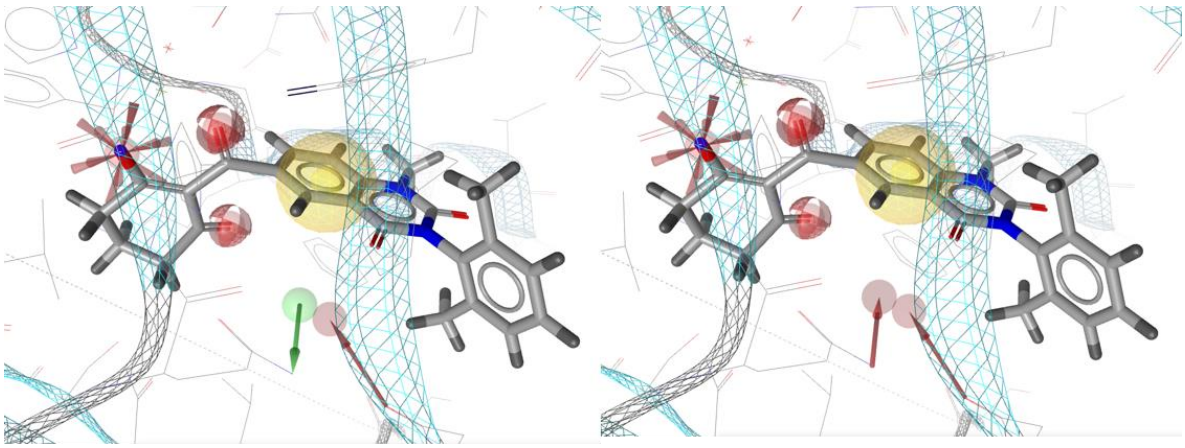


Figure 30: Pharmacophore model "Apo Site Donor" and "Apo Site Acceptor". The pharmacophore models have been created independently of the ligand in the binding pocket. The features have been refined by obtained knowledge of the ligands and the structure. The hydrogen donor feature interacts with Gln 307, which can also interact with a hydrogen acceptor feature. The presented ligand is the co-crystallized ligand of the PDB structure 5YY6

4.8.7 Further Pharmacophore Models

Generation of the pharmacophore model "Top 10 Arabidopsis" was explained in chapter 4.8.5. It does not match all ligands when running virtual screening of self-created libraries (Chapter 4.8.8). The 15 most bioactive ligands, which are not detected by this pharmacophore model, were selected to generate further satisfying pharmacophore models. One compound was used as a reference, on which the other 14 ligands can orientate and superimpose themselves without generating conformations. Thus, it was guaranteed that the docking poses with a right binding mode were kept. Via this overlay structural similarities and dissimilarities could be identified. The ligands have been clustered. 2 clusters were formed, one cluster containing thirteen molecules and the other cluster containing two ligands. For each cluster a pharmacophore model was created and finally the two pharmacophores were merged to a common pharmacophore called "15 leftovers". This model was then optimized according to the same principle as (Chapter 4.8.9).

4.8.8 Generating libraries

To screen pharmacophore models, libraries must be utilized. These libraries were created from the subsistent non-substituted triketones (Chapter 4.2.2). For a first database, the docking poses of the inactive triketones were chosen. The ligand set is imported for the ligand-based perspective. Now conformers have been generated up to 200 conformers per ligand were generated with the iCon algorithm. The complete library was converted to and saved as local Database (ldb) file. This format is supported by LigandScout and is the format imported for databases in the Screening perspective. The first multi-conformational database is created.

The same procedure was used to create more databases. To validate our previously created pharmacophores, it was necessary to create databases containing either moderately active or very active triketones. In addition, two libraries of i) the 10 highest active and ii) 10 lowest active triketones of ARBTH and ZEAMX have been created. Finally, a total of seven different databases from the different ligand sets was saved (Figure 31) and used for following virtual screening purposes.

4.8.9 Virtual Screening

After the pharmacophore models have been created and the libraries prepared, the models can be evaluated and tested by using virtual screening in the screening perspective.

Figure 31 shows the seven multi-conformational libraries. Decoys can be marked with a red box in front of the database if they are selected for screening. Not selected libraries have an empty circle as marker and libraries that contain active ligands and are to be added for screening are marked with a green box.

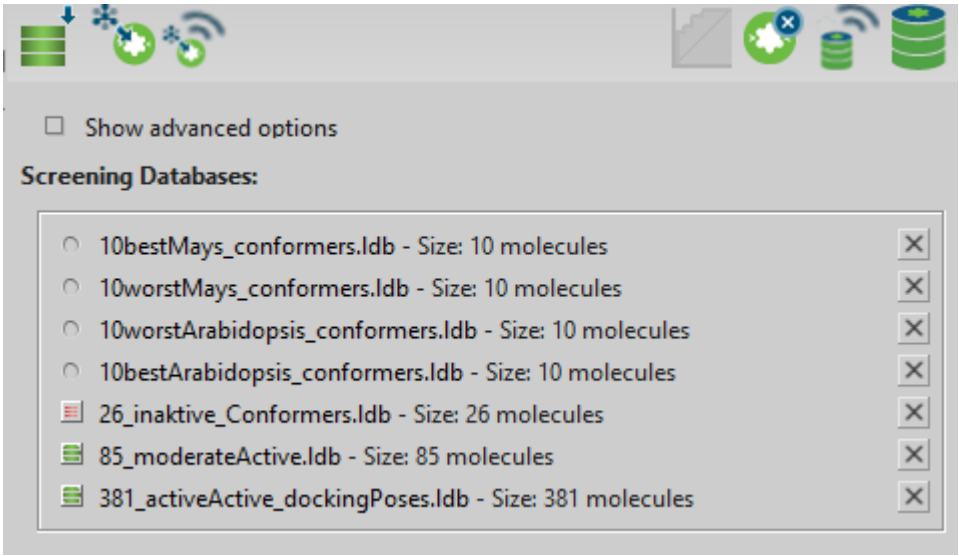


Figure 31: Databases available for virtual screening.

As explained in chapter 4.8.8, these seven libraries were created. They are multi-conformational libraries whose starting position was the respective docking pose. With these databases the pharmacophore models can be tested.

For virtual screening, the pharmacophore model and the prepared databases have to be selected. In the advanced options the scoring function and the retrieval mode can be specified. In this case, Pharmacophore-Fit is adopted from the standard settings and all query features have to be matched. The best conformation of the hit including the exclusion volumes has to be written out. This increases the calculation time but better results are expected. The screening process can now be run.

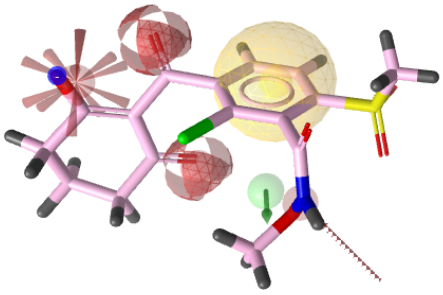
As soon as the virtual screening is finished, a hit list appears. In addition to the found hits, the screening score and the matching characteristics are displayed in color. This score is calculated accordingly to the number and accuracy of the hit features.

The better the match to the pharmacophore, the higher the score. The screening score can be used to rank and select the best hits.

In virtual screening most pharmacophore models had to be readjusted. Often the tolerance of the pharmacophore spheres needed to be decreased or the feature coordinates had to be optimized to achieve better screening results. The adjusted pharmacophores were always cross-checked to the structure to verify the interactions to the amino acids of the protein in the binding pocket.

Since the most important features at the triketone warhead are known, the features have never been marked as optional features. In this way, the features that define a triketone were consistently hit. The features of the residues of triketones were partly set as optional. The maximum number of omitted features in the advanced options was set to zero. This reduced calculation time and delivered better results with optional features. Repeated screening with modified optional features of a pharmacophore as well as modified tolerance and weight of the spheres results in findings with which any pharmacophore can be optimized to achieve better results.

Figure 32 shows the refined pharmacophore "Apo Site Donor" (Chapter 4.8.6). The tolerance ranges were reduced and a hydrophobic feature was set as optional. The self-generated libraries were screened with the pharmacophore, as it can be seen in Figure 31. The hits from the screening were written to the hit list and ranked accordingly to the Pharmacophore-Fit Score.



Hits for Query »ApoSide_ Donor« Hitrate: 2.24% (11 of 492) | Filter:

	Mark	Name	#	Matching Features	Pharmacophore-Fit Score
1	☐	373662	10		67.06
2	☐	364835	6		66.23
3	☐	353279	4		66.23
4	☐	365729	7		65.53
5	☐	4085090	11		57.68
6	☐	376156	9		57.29
7	☐	348713	3		55.63
8	☐	4038319	1		55.57
9	☐	369316	8		55.12
10	☐	338678	2		54.88
11	☐	346942	5		53.35

Figure 32: Hit list appeared after the screening is finished.

The table contains all matched hits from the virtual screening. The matching features are marked in color according to their properties. Where there is a gap, the optional feature is not hit. Fixed features

had to be hit. The Pharmacophore-Fit Score is shown in the right column of the table. According to this score can be ranked. The higher the value, the better the structure of the hit fits. Above, the screened pharmacophore is displayed in the working space together with a selected hit from the hit list.

4.8.10 Boolean expression

With the Boolean expression a condition can be set if two or more pharmacophore models are to be combined with each other. For this purpose, the pharmacophore models to be combined have to be selected from the pharmacophore list in the screening perspective. Preconditions can be entered in the field for Boolean expression (Figure 33). The terms "and" and "or" are allowed. For example: "1 and 2" means that all ligands that are in the pharmacophore model 1 and 2 have been hitting and are written out. Usually less hits are found than if the two pharmacophore models are screened individually against a library. Another example is "1 or 2". All hits are found which are hit either by pharmacophore 1 or by pharmacophore 2. A valid expression is marked by a green dot behind the description (Figure 33).
The pharmacophores “Top 10 Arabidopsis” (Chapter 4.8.5) and “15 leftovers” (Chapter 4.8.7) were combined with each other. The Boolean expression was defined with the condition "1 or 2 ". The results of this hit list with this combination is explained and discussed in chapter 5.8.

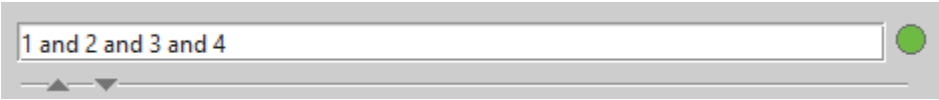


Figure 33: Example of valid Boolean Expression.
With this statement, four selected pharmacophore models can be combined with each other. This is just an example. In this work only two pharmacophore models were combined.

5 Results and Discussion

5.1 Data analysis from Data Filtering using KNIME

There were originally two data files with in-house data split using KNIME workflows. Both tables contained only triketone herbicides that are supposed to inhibit ARBTH's enzyme 4-HPPD. The triketones should act as selective as possible on ARBTH and not inhibit the enzyme of the crop ZEAMX. How many of the in-house data have been tested on the enzymes, how many duplicates were contained, which tautomers are present and bioactive, will be analyzed in the next chapters 5.1.1 and 5.1.2 and the data discussed critically.

5.1.1 Data analysis of the first dataset

In the first file with in-house data, 2084 triketones and their properties were described (see chapter 4.2.1). After removal of 136 duplicates, 1948 singular structures remained.

The triketones were separated according to the enzyme source and additionally according to their substitution, whether the structures are substituted or not at the 1,3,5-cyclohexantrione of the triketones. As shown in Table 1, 374 of the 526 unsubstituted triketones have been tested in ARBTH, 152 have been tested in ZEAMX. From 1442 non-substituted triketones, 963 were tested in ARBTH and 459 in ZEAMX. There were therefore 1337 substances tested for ARBTH and 611 molecules tested in ZEAMX.

Table 1: filtered data for substitution state according to the enzyme source.

	ARBTH	ZEAMX	In total
non-substituted	374	152	526
substituted	963	459	1442
	1337	611	1948

By removing duplicates (see chapter 4.2.1.3) 136 structures have been filtered out. With the workflow shown in Figure 19 the duplicates with the corresponding structures were merged A total of 250 structures was stored in the pdf table. After 136 structures were filtered out before, 114 structures remained in the output table, which then ran through the other workflows. As a result, these 114 structures occur at least twice. Further analysis of the duplicates revealed that 26 structures were tested in ARBTH and ZEAMX. The list with the 26 different substances was stored in a pdf table which is shown in the appendix in Table 6. On closer inspection, it was noticed that six compounds show different activities in ZEAMAX and ARBTH (Figure 34). These differences are larger than one order of magnitude for four substances. Smaller differences are statistically not relevant. The substances with LS-Cores 297300, 328115, 337128 and 382526 have this larger fluctuation in the IC_{50} value. The substance with the LS-Core 321416 has identically measured IC_{50} values at ARBTH and ZEAMX. The sixth ligand (LS-Core: 298647) shows lowest activities not allowing to derive an IC50 value. Therefore, only the percentage of inhibition was given. The 20 triketones of the other duplicates tested on ZEAMX and ARBTH were not available in inventory and therefore could not be tested. No data points were available for this purpose. Figure 34 shows the IC_{50} values from the data set in tabular form. These six structures have been retested and will be discussed in chapter 5.9. The latest enzyme assay was designed to test whether the ligands actually differ selectively from ZEAMX and ARBTH in terms of bioactivity, as no significant differences in selectivity have been found in the past. The remaining duplicates were tested on the same enzyme source but at a different time

	ARBTH	ZEAMX
297300	1.1E-7 M	2.6E-6 M
328115	1.1E-7 M	1.18E-6 M
337128	1.8E-7 M	7.7E-6 M
382565	7.78E-7 M	1.06E-6 M
298647	11% Inhib.	31% Inhib.
321416	5.32E-8 M	5.3E-8 M

„Old“ IC_{50} values

Figure 34: IC_{50} values from the dataset.

These 6 structures in this table can have incorrect measured IC_{50} values or they have a slightly larger range of variation in their measurement. These structures were selected for retesting as they were available. The results are described in chapter 5.9.

This dataset contains triketones with a substituted warhead. Since it was decided at a later stage that the focus was on non-substituted triketones, the remaining results were obtained with the ligand set of non-substituted triketones. An example for a substituted warhead is shown in Figure 35. The cyclohexane can only substitute small residues such as methyl or dimethyl moieties. This ligand has the LS-Core 297300 and is one of the ligands tested in both ZEAMX and ARBTH.

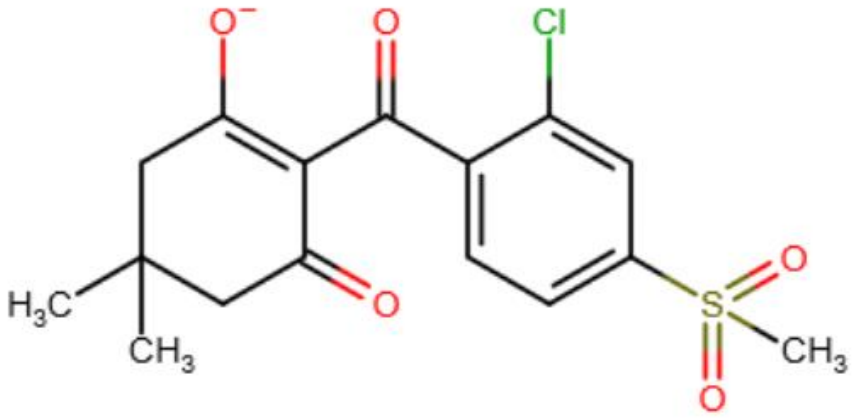


Figure 35: An Example of a substituted triketone
The ligand with the LS-Core 297300 is an example of a substituted triketone. The cyclohexane is dimethylated at C5.

5.1.2 Data analysis of the second dataset

The second set of ligands contains only non-substituted triketones. Already at the beginning of the thesis it turned out that the focus is on triketones without a substituted residue at 1,3,5-cyclohexantrion warhead. The warhead can be fixed in the binding pocket it can be focused on the different scaffolds of the ligands during pharmacophore modeling.
Then the duplicates were examined according to the same filter criteria as above. The total of 42 structures contain 20 individual LS-Cores, which means that these 20 structures were included in the data set at least twice. None of the 42 compounds were tested on both enzyme sources. They were included several times in the data

because they were repeatedly tested on the same enzyme source plant with different bioactivity or because the same structure is listed in different tautomers.

Table 2 visualizes the results of the workflow of chapter 4.2.2.3. The results were divided according to their bioactivity. 299 substances tested for ARBTH are highly active. 85 of the 384 highly effective substances were tested for ZEAMX. 85 substances are moderately effective, of which 61 are tested for ZEAMX and 24 for ARBTH. 7 of the 26 inactive triketones were tested in ZEAMX and 19 in ARBTH. The table shows that most non-substituted substances have only been tested for ARBTH. ARBTH is the weed on whose 4-HPPD the triketones are supposed to act in contrast to the crop ZEAMX.

Table 2: Separation of the bioactivity according to the enzyme source.

	Highly Efficacy ($IC_{50}=10^{-9}$-10^{-7})		Moderate Efficacy ($IC_{50}=10^{-6}$-10^{-5})		"Inactive" ($IC_{50}<10^{-5}$)	
495 Triketones	384		85		26	
Enzyme Source	85 ZEAMX	299 ARBTH	61 ZEAMX	24 ARBTH	7 ZEAMX	19 ARBTH

Figure 36 visualizes activity distribution of all considered compounds at ARBTH (left) and ZEAMX (right).

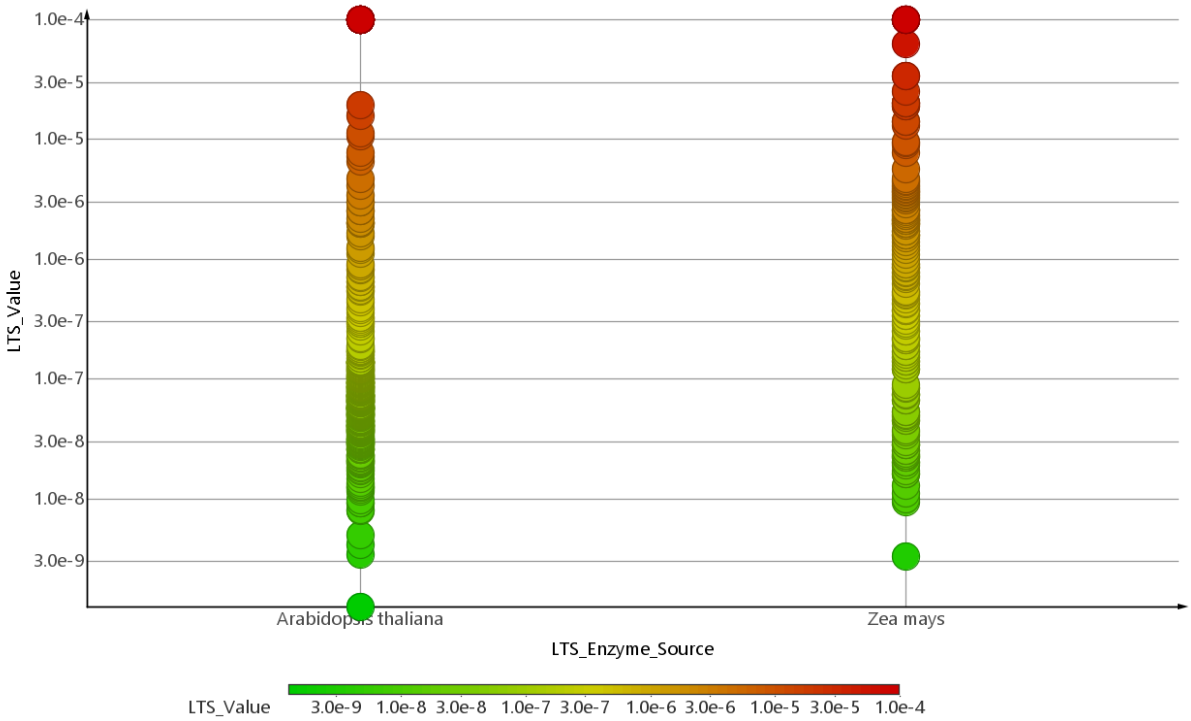


Figure 36: Visualization of the non-substituted triketones by their bioactivity and their enzyme source.

This 2D diagram illustrates the bioactivity in relation to the tested plants ARBTH (left) and ZEAMX (right). In this graph all 495 non-substituted triketones are included and tested in ARBTH or ZEAMX. They were divided on the X-axis. Along the Y-axis the bioactivity occurs on a logarithmic scale. This improves the visibility of the close bioactivities. Each point represents one tested triketone. The lower the IC_{50} value, the more bioactive the substance is and the greener the dot that symbolizes a substance is. The redder the ligands are colored, the less active and the more inactive they are. The IC_{50} values range from 1,0 E-4 to 3,0E-9. The chart was made with the DataWarrior.

5.2 Assay data

In the first data set it is noticeable that all ligands were tested for ZEAMX only from September 1994 to September 1996. From November 1996 to October 2009 all substances were tested only for ARBTH and no longer for ZEAMX.

The second data set was tested for ZEAMX from September 1994 to September 1996. From March 1997 to July 2003 the substances were tested only in ARBTH. No compounds were measured in both enzyme sources. It is known that the enzyme assay has developed over time and that the detection technique has made progress (Chapter 4.3). It is not known how this test was performed and what the differences

are to the current assay. Nowadays, the substances are measured with a 9th generation assay.

Therefore, the question arises to what extent the IC_{50} values are comparable over this long period. The IC_{50} values of ZEAMX are on average slightly higher than those of substances measured in ARBTH. This might be due to the improved assay, but also to the fact that ARBTH can react more sensitively to 4-HPPD inhibitors.

The 26 substances tested in both enzyme sources have also been investigated in another enzyme assay at different time points. This is the reason why six of these compounds were retested simultaneously in the current assay conditions (see chapter 4.3). These IC_{50} values should be more comparable.

5.3 Ligand Alignment

The aim of the ligand alignment was the overlay of the ligands with their non-substituted triketone warhead so that structural differences and commonalities of the ligands could be identified. Figure 36 shows one example of six ligands that are aligned. Beside the warhead they all have a benzoyl ring in common. The similarities are visualized by pharmacophore features. A pharmacophore was not created at this time. With only these 3 features the pharmacophore would be insufficiently selective. The residues of the core scaffold differ from each other.

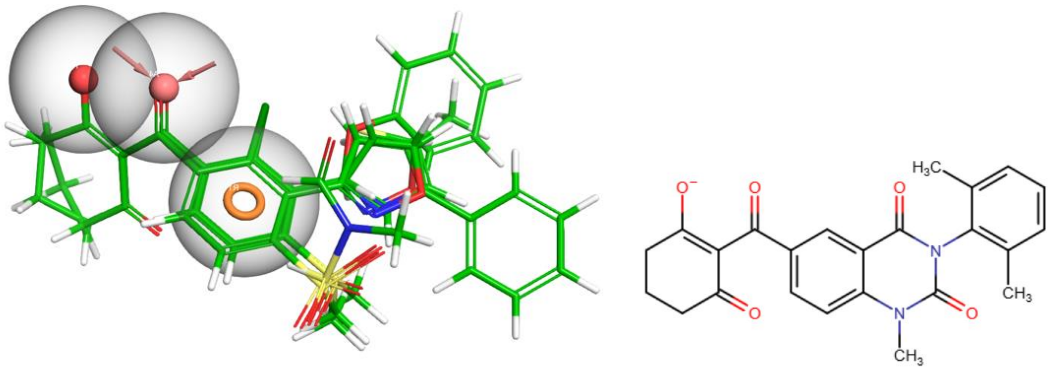


Figure 37: Ligand alignment of six active triketones.

The shown alignment was performed with Maestro. The non-substituted ligands are perfectly superimposed using the SMARTS pattern. The shared 1,3-dione moieties and the benzoyl ring are highlighted by the pharmacophore features. The ligands are represented in thin tube style. Next to it, a 2D structure of the co-crystallized triketone from the PDB with the code 5YY6 is represented for visualization.

When considering other ligands, it was seen that all compounds have an cyclic ring system after the 1,3,5-trione moieties. Aromatics, but also cyclopentanes are common residues. With the DataWarrior tool it was possible to simply select by scaffolds. 463 of 492 triketones have a benzoyl residue at the site. The remaining 29 triketones have heteroatoms in the benzoyl ring or there is a cyclopentene residue instead which can also be substituted. The quintessence of this is that a cyclic ring system is present following the 1,3,5-cyclohexane-trione. This information can be used for ligand-based pharmacophore modeling (Chapter 5.6).

5.4 Docking Results

The data sets were docked into the crystal structure of 5YY6, in order to gain initial insights into the binding modes of the ligands. The aim was to compare the individual data sets on the basis of their docking results. Further information should be derived by using active and inactive ligands as well as differently selectively compounds with regard to ZEAMX and ARBTH.

The prepared ligands (chapter 4.5) were present with the same binding mode in the binding pocket, which is forced by a metal coordination constraint.

Since the ligands are similar in their size and have the same binding modes, no significant structural difference is found. Eight ligands could not be docked due to their size. All other compounds had no steric hindrance and could be docked. The docking scores were considered. The more negative the score, the better is the docking. The best docked pose has a docking score of -12.4 and the worst docking score was -2.7. Even the docking pose with the highest score is passable in the binding pocket. The docking of the ligand with the worst docking score was therefore also successful.

Therefore, neither the docking score nor the docking pose could provide information about a possible selective differentiation of the bound ligands in ZEAMX and ARBTH. There is also no correlation between the docking scores of the ligands and their bioactivity. Therefore, the structure of the two selected crystal structures of 4-HPPD were investigated in more detail.

The docking poses are consistent with docking results from the literature (Chapter 2.2.1). The docked ligands in this thesis form the typical chelate complex with Fe^{2+}. The distance between the Fe^{2+} and the 1,3-diketone moiety could be kept.

5.5 Structure Analysis

5.5.1 Protein Alignment

After each chain of the 4-HPPD enzyme was aligned from ZEAMX (PDB:1SP8) to the A chain of ARBTH (PDB: 5YY6) (Chapter 4.4), it was decided by visual inspection which of the 4 chains should best be superimposed due to their secondary structure elements. Chain A of 1SP8 suited best structurally and is used as reference chain for all comparisons to the A chain of 5YY6. This chain had the lowest alignment score compared to the other chains of 1SP8. The alignment score is 0.092 and is shown in Figure 38. It is automatically calculated by Maestro. The alignment scores of each alignment were similar. As described in chapter 4.7.1, an alignment score below 0.6 counts as a good alignment, which is given in our case. RMSD calculates the average distance between atoms of superimposed proteins. The value is usually given in Angstrom. The smaller the value, the better is the model compared to the reference structure. Thus, the RMSD of 1.504 Å represents a good alignment.

```
Alignment Score:  0.092 (smaller is better)
           RMSD:  1.504 Angstrom
```

Figure 38: Alignment score of 5YY6 and 1SP8 A chain.
The Alignment Score is automatically calculated by Maestro after the alignment has been successfully completed. The lower the value, the better the alignment. The Alignment Score of 0.092 represents a good Alignment.

Figure 39 represents the two superimposed A chains. For visualization reasons amino acids are hidden and only secondary structural elements are indicated. The proteins appear to be structurally very similar. The green compound is the co-crystallized ligand from the 5YY6 binding pocket. Structurally the binding pockets do not differ from each other. The helices in the open (yellow) and closed (red) conformation of the crystal structures are significantly different and therefore highlighted. ARBTH is crystallized in an open conformation and ZEAMX is represented in a closed form. The yellow Helix H11 belongs to ARBTH and the pink Helix H11 to ZEAMX.

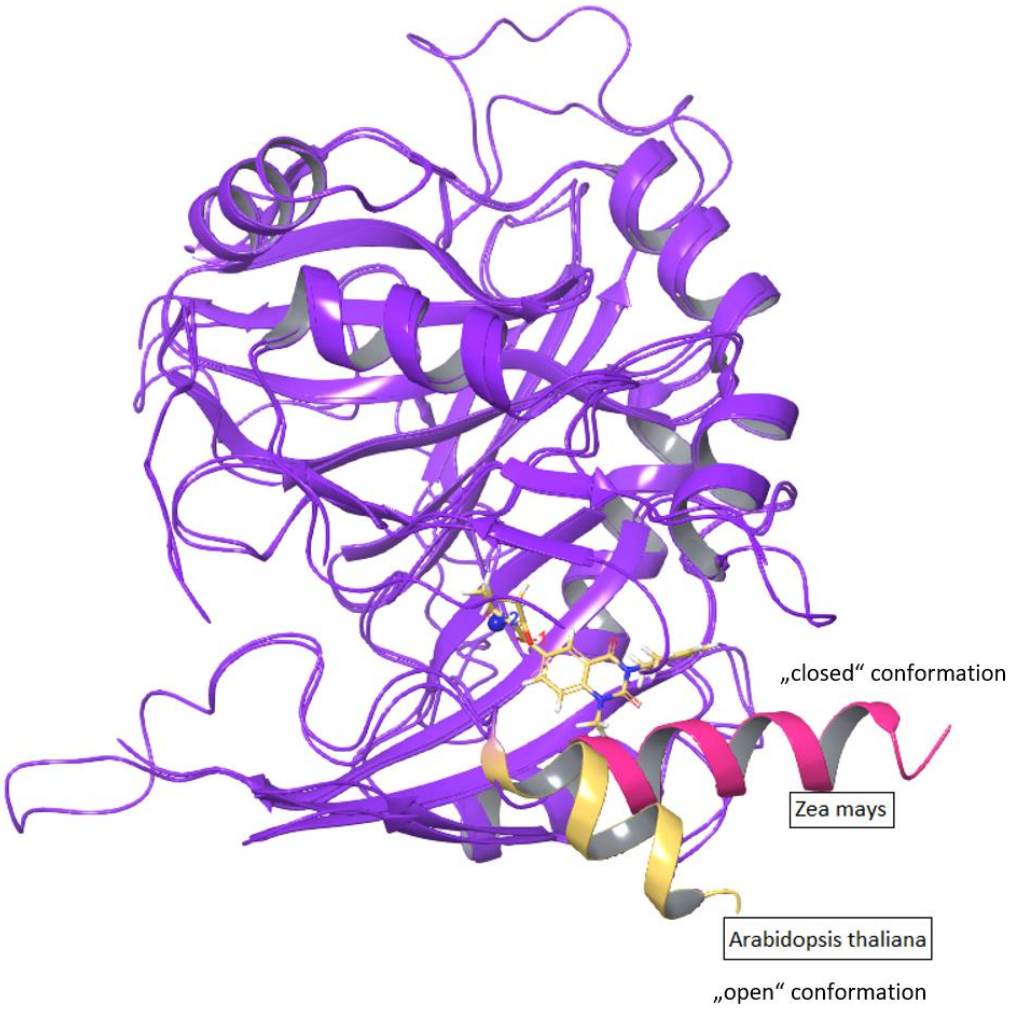


Figure 39: Protein Alignment of the A chains of Arabidopsis thaliana and Zea mays.
The A chains of 1SP8 and 5YY6 are aligned. The amino acids are hidden and only the secondary structure elements in ribbon style are visible in the Maestro interface. For visualization the chains were colored unicolored in purple. The helices H11 are colored that the different conformations are clearly recognizable. The protein with the pink Helix H11 in closed conformation is from ZEAMX. The other protein with the yellow helix H11 in open conformation is from ARBTH. The co-crystallized ligand is present in the binding pocket and is represented in ball-and-stick style in green color.

5.5.2 Amino Acid Sequence Analysis of 1SP8 and 5YY6

The sequence identity of both protein chains is shown in Figure 40. 5YY6 behaves identical with 62.9 % to the A chain of 1SP8. A chain of ZEAMX is with 63.4 % identical to the protein of ARBTH. These different percentage sequence identities are dif-

ferent because the amino acid sequences have different lengths and because the counting starts at different amino acids. To calculate it, the number of identical residues of both chains is divided by the number of amino acids of the respective chain. The first amino acid of 5YY6 that was dissolved is Met 31. In the Maestro and MOE interface the first visible amino acid is Lys 35. In the crystal structure of ZEAMX Ala 17 is the beginning of counting.

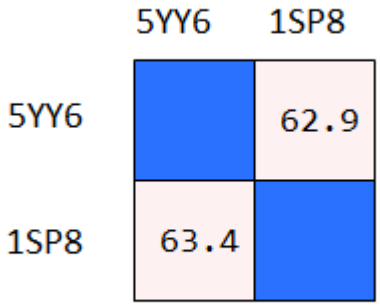


Figure 40: Overall identity of A chain of 1SP8 and the chain of 5YY6.
The A chains of ARBTH and ZEAMX of the homodimer 4-HPPD have an overall identity with over 60%. This graphic was created with the Similarity Monitor from MOE and the values were calculated with it.

The exact amino acid sequence is shown in Figure 41. Each protein starts with the start codon Met1. The sequences shown start with those residues whose positions were resolved clarified. The numbering is based on the full-length protein. The first visible amino acid in this sequence is Arg 36. As seen in the alignment, the corresponding amino acids lay on top of each other. Glu432 is the last structurally observed amino acid at the C-terminus of 4-HPPD from ARBTH. 13 further amino acids at this location cannot be imaged because their exact 3D position is not known. In 1SP8 Leu431 is the last amino acid that can be observed at the C-terminus. This is a deficit of one amino acid compared to ARBTH. If the chain B or D would have been chosen as an outcome of the alignment, Glu432 would have been additionally detected. 4-HPPD of 1SP8 is longer than the enzyme from 5YY6.

```
5YY6.A
...............mfvrknpksdkfkvkrfhhiefwcgdatnvarrfswglgmrfsak
1SP8.A
asaaeqaafrlvghrnfvrfnprsdrfhtlafhhvelwcadaasaagrfsfglgaplaar


5YY6.A
sdlstgnmvhasylltsgdlrflftapyspslsageikptttasipsfdhgscrsffssh
1SP8.A
sdlstgnsahaslllrsgslsflftapyahgada......ataalpsfsaaaarrfaadh


5YY6.A
glgvravaievedaesafsisvangaipssppivlneavtiaevklygdvvlryvsykae
1SP8.A
glavravalrvadaedafrasvaagarpafgpvdlgrgfrlaevelygdvvlryvsyp.d


5YY6.A
dtekseflpgfervedassfpldygirrldHaVgnvpelgpaltyvagftgfhqfaeFta
1SP8.A
gaagepflpgfegv..aspgaadyglsrfdHiVgnvpelapaaayfagftgfhefaeFtt


5YY6.A
ddvgtaesgLnSaVlasndemvllPiNepvhgtkrksQiqtylehnegaglQHlalmsed
1SP8.A
edvgtaesgLnSmVlannsenvllPlNepvhgtkrrsQiqtfldhhggpgvQHmalasdd


5YY6.A
ifrtlremrkrssiggfdfMpsppptyyqnlkkrvgdvlsddqikeceelgiLvdrddqg
1SP8.A
vlrtlremqarsamggfefMapptsdyydgvrrragdvlteaqikecqelgvLvdrddqg


5YY6.A
tllQiFTkplgdrptiFiEiiqrvgcmmkdeegkayqsggcggFGKgNFSELFkSIeeye
1SP8.A
vllQiFTkpvgdrptlFlEiiqrigcmekdekgqeyqkggcggFGKgNFSQLFkSIedye
```

Figure 41: Amino acid sequence of both chains of 5YY6 and 1SP8.

The amino acid sequence of both enzymes was generated with the MOE Sequence Monitor. The corresponding amino acids lay directly on top of each other. The small letters stand for no side chain interaction to the ligand in the binding pocket within 4,5 Å. The capital letters, here highlighted in green, interact with the co-crystallized ligand. In red the amino acids are marked which are located in the binding pocket, but which differ from the corresponding amino acid.

Figure 41 shows the protein similarity report. The amino acid sequences of the A chains of 4-HPPD of ARBTH and ZEAMX are depicted. The upper line shows the sequence of ARBTH (PDB: 5YY6) and directly below the amino acid sequence of ZEAMX (PDB: 1SP8). The corresponding amino acids are thus clearly located among each other. The amino acids that are written in capital letters in the one-letter code are amino acids in the binding pocket that have a side chain interaction with the ligand. The amino acids in red color differ from the corresponding amino acids. Except for one amino acid at the C-terminus, all red colored amino acids of 5YY6 are close to the binding pocket or stand between two interacting amino acids but are not themselves involved in the binding of the ligand. They have no side chain but a backbone interaction. Glu 426 of 5YY6 at the C-terminus is located in the helix H11 and can interact with the ligand.

Since no distinction can be made based on the alignment of the full protein, the binding pockets are examined more closely (Chapter 5.5.3). Furthermore, the C-terminal helices of both crystal structures are examined and analyzed (Chapter 5.5.4) to find indications to separate active and inactive as well as selective ligands, (ARBTH vs ZEAMX).

5.5.3 Analysis of the Binding Pocket

The similarity of the binding pockets of ARBTH and ZEAMX is shown in Figure 42. The two 4-HPPD binding pockets have an 96.6 % sequence identity.

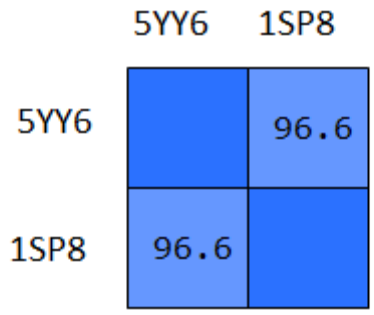


Figure 42: Similarity of the Binding Pocket regarding to the Amino Acid Sequequence of 5YY6 and 1SP8.

The two crystallized binding pockets of ARBTH and ZEAMX have a similarity of 96.6%. The values were calculated and graphically created with the Similarity Monitor from MOE.

In 1SP8, the Gln 419 corresponding to Glu 426 of 5YY6 cannot interact with the ligand. Two lysines in the binding pocket have the potential to interact. Both corresponding amino acids of ARBTH are Ile and do not interact with the ligand.
The interactions are illustrated in Figure 43.
The aligned binding pockets of both proteins are shown. The co-crystallized ligand of 5YY6 is located in the binding pocket. The amino acids interacting with the ligand within a radius of 4.5 Å are shown in this graph. The green colored amino acids are identical in both enzymes. If corresponding amino acids differ from each other, they are colored red. The red colored amino acid side chains turn away from the binding pocket towards the protein. At the entrance of the binding pocket, close to the C-terminus, Glu 426 from 5YY6 interacts with the ligand. This amino acid is the only distinction of the binding pockets between ARBTH and ZEAMX. Instead of Glu 426 1SP8 contains the amino acid Gln 419, which is not located in the binding pocket and cannot interact with the ligand. This explains the similarity of 96.6% of the binding pocket.

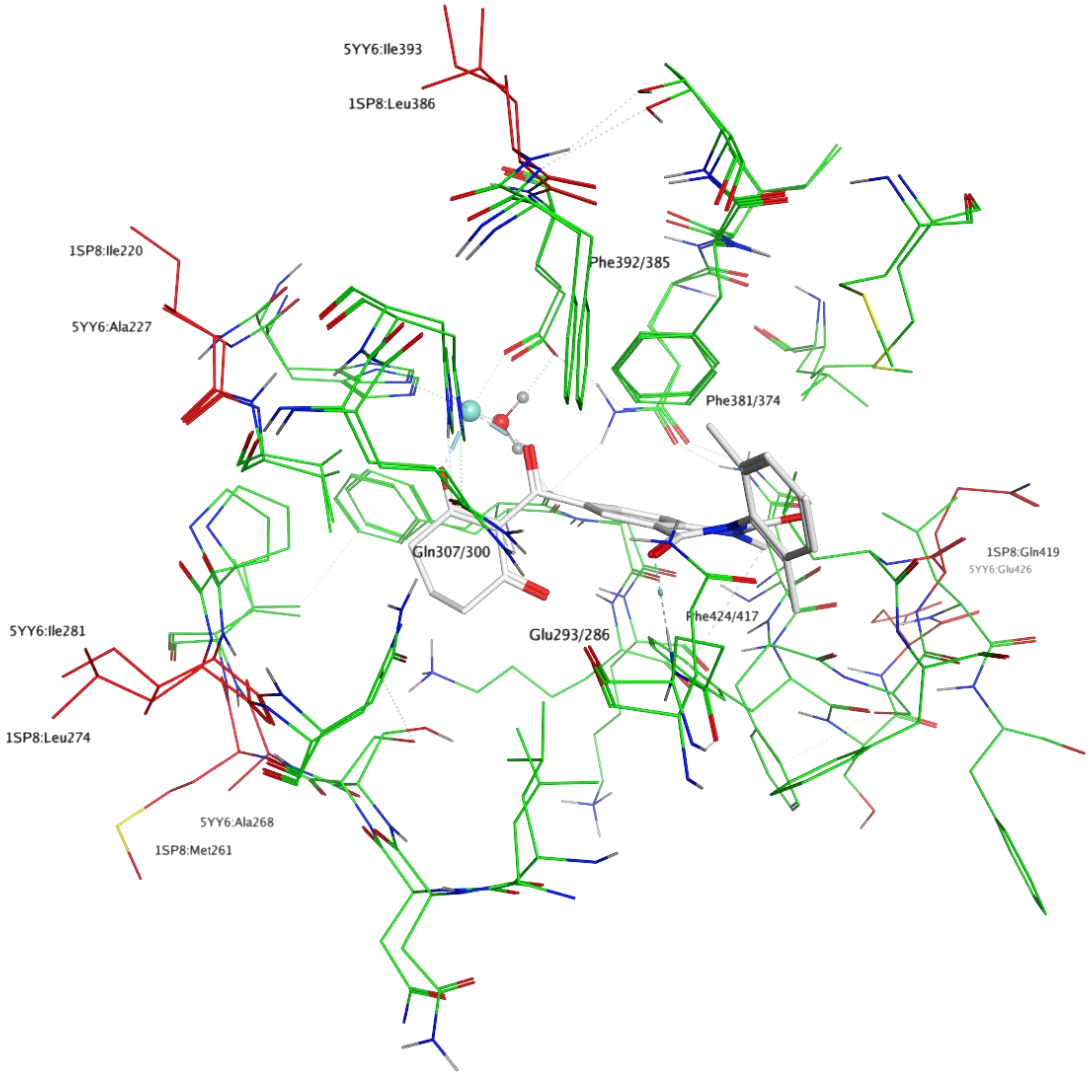


Figure 43: Aligned binding pocket of Zea mays and Arabidopsis thaliana.

All amino acids shown are located in the ZEAMX and ARBTH binding pocket and interact with the ligand within a radius of 4.5 Å. The co-crystallized ligand of 5YY6 is embedded in the middle of the binding pocket. The amino acids are green and do not differ from each other. Shown in red are the amino acids that differ from their corresponding amino acids. The amino acids that are important for the protein-ligand complex interaction or differ from their corresponding amino acids are labeled. The figure was created with MOE.

As can be observed in Figure 41 and Figure 43, the binding pocket is relatively hydrophobic due to its many apolar amino acid residues. Figure 43 also shows that the binding pocket is quite tight. Towards the C-terminus the binding pocket expands.

Figure 44 shows a protein-ligand interaction fingerprint. For each interaction between one of the 495 non-substituted triketones and an amino acid from the binding pocket of 5YY6, a black line is drawn. At the bottom of the fingerprint all interacting amino acids are indicated. Gln 307, Pro 336, Phe 281, Phe 392, Phe 419, Gly 420, and Phe

424 each have two columns which can be marked by an interaction. The left columns symbolize weak and markers in the right column indicate strong interactions. Gln 307 is an amino acid that can act as hydrogen acceptor and as hydrogen donor with its side chain. Therefore, there are two columns for this both cases. Fe^{2+} forms two interactions to the oxygens of the 1,3-diketon moiety of triketones. There are again two columns for each hit oxygen. Hence, in total there are four columns for Fe^{2+} interactions. The metal coordination forced during docking causes each column of Fe^{2+} to be hit in the fingerprint. There is one outlier, which instead of two strong interactions, forms one strong and one light interaction.

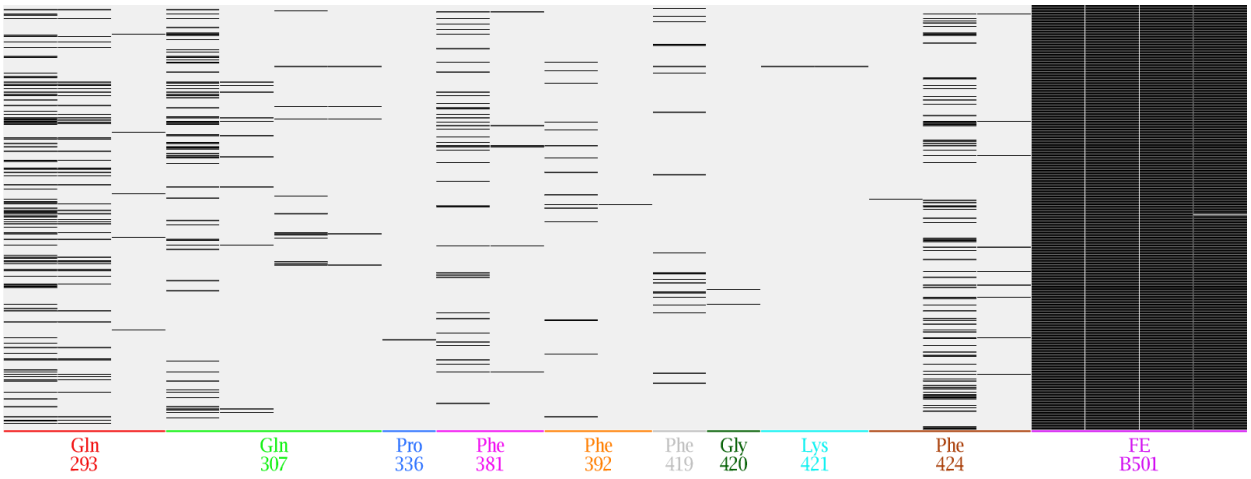


Figure 44: Protein-Ligand Interaction Fingerprint.

This fingerprint marks each interaction with a black line. On the x-axis the amino acid is named, which can interact with the ligand within 4.5 Å. On the right side is the iron ion, which must chelate with two oxygens of the triketone. Each specified amino acid has two columns. The left column for weak interactions and a right column for strong interactions. Gln 307 can act as hydrogen acceptor or as hydrogen donor and has two extra columns. The enzyme 4-HPPD from ARBTH (PDB: 5YY6) was chosen for fingerprinting in MOE.

The corresponding amino acids for 1SP8 are Gln 286, Gln 300, Pro 329, Phe 374, Phe 385, and Phe 417 from left to right as shown in the figure. Since the binding pocket has a sequence similarity of 96.6 %, it is assumed that a very similar interaction pattern could be expected.

This fingerprint is used to determine which amino acids are important for the binding of ligands. This information is useful for generating pharmacophore models (Chapter 4.8).

According to Fu et al [8] hydrophobic and π-π-interactions between the phenylalanines were shown. Proline can also interact. An interaction with methionine, aspara-

gine and arginine described in the literature could not be reproduced in this work, but further interactions of glutamine, glycine and lysine were shown.

5.5.4 Analysis of the Helices H11 of Arabidopsis thaliana and Zea mays

Since the amino acid Glu 424 of ARBTH differs from the equivalent amino acid Gln 419 of ZEAMX in the binding pocket (Figure 43), they were examined more closely. The two amino acids are already part of the C-terminal helix of the 4-HPPD enzymes, which have been crystallized in a rigid conformation. If the amino acid sequences of the different helices are analysed, it becomes apparent that the helices are of the same length based on their sequences (Figure 45). However, they are not shown for the same length in the interface, which is due to the fact that they cannot be shown crystallographically (Chapter 5.5.2). The helix of ZEAMX is longer. In its sequence glutamate (Glu) 424 in ARBTH differs from the equivalent amino acid Glutamine (Gln) 419 in ZEAMX. Glutamate has acidic properties, whereas glutamine is a neutral but polar amino acid. Figure 45 shows the sequences of the helices. The H11 starts with lysine and ends with the C-terminus on the right side. A black line in the sequence indicates to where the H11 can be displayed in the interface. Thus Glu 430 is the last amino acid in ARBTH and Leu 431 in ZEAMX. Since two further equivalent amino acids are different, but were not visible in the interfaces, they were neglected. These are the amino acids glutamate 431 and threonine 435 from ARBTH and their respective amino acids from ZEAMX. Due to the unclear location of these residues, which were crystallographically refined, neither interactions nor correlations to the two helices could be investigated. Crystallographically, the differentiation of these two amino acids glutamine and glutamate is very difficult to impossible.

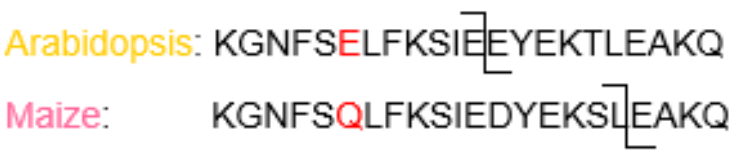


Figure 45:Amino Acid Sequences of the C-terminal Helices of ARBTH and ZEAMX

In the upper line the helix of ARBTH is positioned and below is the sequence of the helix of ZEAMX. The amino acid sequence is shown in one letter code. The distinction of Glu 426 from ARBTH and Gln 419 from ZEAMX is highlighted in red. The other different amino acids have not been studied as they have not been shown in the interface. A black line shows the last visible amino acid of each chain. Glu 432 in ARBTH and Leu 431 in ZEAMX can still be depicted in the interface.

Figure 46 shows the sequences of the two helices coloured with respect to their amino acid's properties (e.g. polarity and acidity). This coloured scheme continues in Figure 47. The amino acids with a blue background represent rather hydrophilic amino acids. As in the previous figure, it is only the crystallographically defined H11 is visible in the interface. The remaining amino acids of H11 have not been defined in the PDB.

Figure 46: Amino acid sequence of H11

The sequence is identical to Figure 45, but the polarities and acidities are separated by colour. The upper line corresponds to the sequence of ARBTH and the lower line is the sequence of ZEAMX. This image was created by the MOE Sequence Viewer.

The sequences of ARBTH and ZEAMX depicted as helical wheel representations (Figure 47) allow comparison of the 3D arrangement of H11. Both are classical alpha helices. The coloration is identical to the coloured amino acids in Figure 47. The left helical wheel is from ZEAMX and the right one is from ARBTH. Glu 432 of 1SP8 and Leu 431 of 5YY6 are the last amino acids of the C-terminus that could be shown. In Figure 47 they are represented in the way that a top view is obtained. The lower the digit of the amino acid, the closer the amino acid is to the binding pocket. Further examination reveals that the polar, blue residues look inside and thus to the side exposed to water. The apolar side chains of the two helices look towards the outside. Consequently, the hydrophobic residues are directed to the side of the protein. The arrangement of amino acids with these properties is a typical pattern for alpha helices.

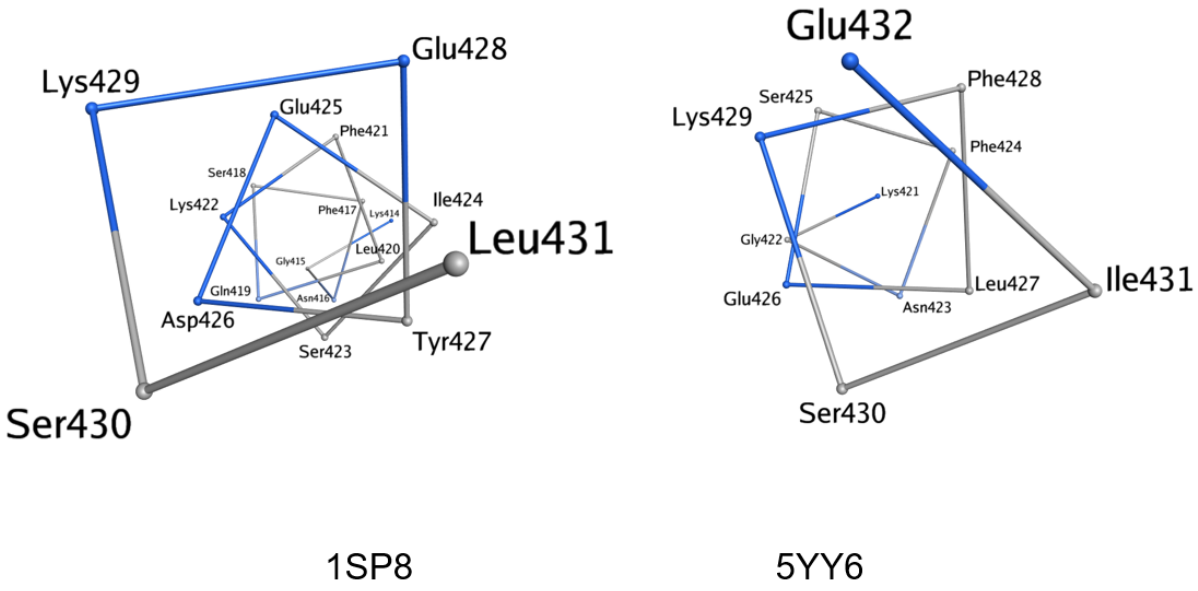


Figure 47: Helical wheel representation of the H11 helices

The left figure shows the C-terminal Helix H11 from ZEAMX and the H11 from ARBTH on the right. The coloration of the amino acids is identical to the coloration in Figure 46 and related to the polarity. The helices are 3.6 residue alpha helices. The figure was shown with MOE and only the amino acids traces were displayed.

Figure 48 shows the two helices described from the “back-side view”. The H11 of ZEAMX in the closed state is shown in pink. The open conformation helix of ARBTH is shown in yellow. The triketone shown is the co-crystallized ligand of 5YY6. If H11 is in its closed state, a triketone herbicide in the binding pocket would clash. The reason for this is leucine (Leu 420), whose side chain protrudes too deep into the binding pocket. However, if H11 is in the open conformation, as shown at ARBTH, the binding pocket is large enough for a triketone to accommodate. The particular leucine (here Leu 427), which in the closed state blocks part of the pocket, turns outwards. The closed helix shows a strong Lys 422 to Asp 334 interacting. An ionic hydrogen bond is formed. In the open conformation Asparagine 423 forms a hydrogen bond with Tyr 342. Due to these interactions, the respective state is stabilized.

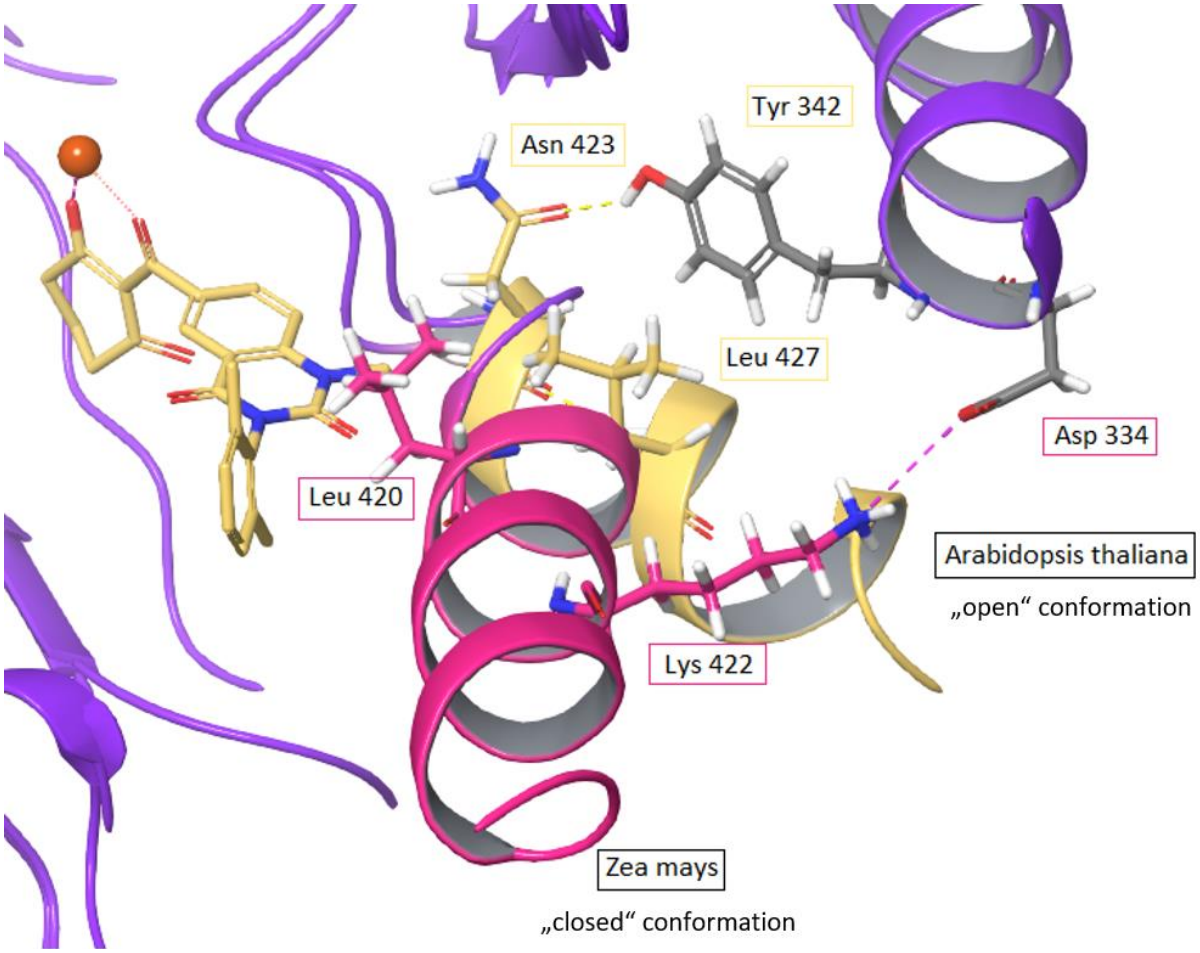


Figure 48: Interactions of Helices with their Proteins

Helices of ARBTH in open conformation (yellow) and ZEAMX in the closed conformation (pink) are shown. In the binding pocket is the co-crystallized yellow ligand of 5YY6. The interactions between the helices and their proteins can also be seen. This picture was made with Maestro.

Since the helices with the more hydrophilic, polar residues are on the water-facing side, the hypothesis occurs that the two rigid crystal structures 5YY6 and 1SP8 show only two types of conformations of many different H11 orientations. Only the 1SP8 from ZEAMX is stored in the PDB. All stored crystal structures of ARBTH have been crystallised in the same open conformation. Therefore, the 4-HPPD crystal structures of Homo sapiens and Rattus norvegicus have also been aligned and analysed. It was noticed that the H11 helices had approximately the same length as ZEAMX, but the conformations of the two helices were located between ZEAMX and ARBTH. The helices of Homo sapiens and Rattus norvegicus are both crystallized in an open conformation. Because of the different conformations of H11, it supports the hypothesis

that H11 can move dynamically and thus open and close the entrance to the binding pocket.

The B-factor indicates how flexible the structure of a protein is. It is a temperature factor that is calculated for each atom in the protein. The factor describes the extent to which atoms can be shifted from their position. The redder the colour of the secondary structure, the more flexible and dynamic it is. If a structural element is coloured white to bluish, it is a sign that the structure is not very flexible and can be rigid at this point. The B-factor of the two aligned proteins 1SP8 and 5YY6 from Figure 39 is displayed.

Figure 49 shows that the inner structural elements of the protein are relatively rigid. The C-terminal helices, by contrast, are strongly coloured red. This gives a further indication that the two C-terminal helices have a gate functionality and can thus open and close the binding pocket. It can be concluded that these two conformations of ZEAMX and ARBTH are only two rigid states that could move dynamically in the physiological milieu.

The hypothesis of the gating function of the C-terminal helix H11 is in accordance with Grossmann et al [48].

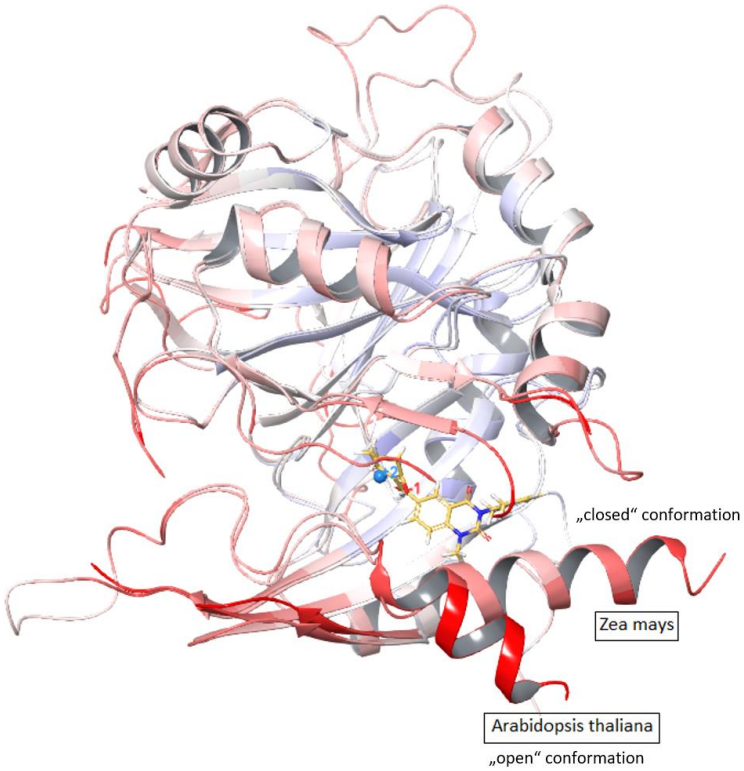


Figure 49: B-factor

The two aligned proteins in Figure 37 were coloured with respect to their B factor in the Maestro interface. The two H11 helices are intensely coloured red, which is an indication for high flexibility. There-

fore, it is assumed that the two conformations were only crystallized in these conformations, but that the helices behave dynamically and have a gate functionality towards the binding pocket.

Since the two 4-HPPD enzymes of ARBTH and ZEAMX are structurally almost identical and the binding pocket also has a similarity of over 96%, it can be assumed that the two enzymes will also behave identically. Probably these two proteins will have the same amino acid interactions between the H11 and its protein when they are in the same conformation. To prove this hypothesis, further dynamic information would have to be collected.

Since the analysis of the structure has not delivered sufficient differences, structure-based pharmacophore modeling was not suitable for this work. Nevertheless, all information obtained from the structure was further considered for the subsequent ligand-based pharmacophore modeling.

5.6 Ligand Analysis

The triketone shown in Figure XX is the most highly effectively measured triketone. Many of the highly active compounds have a common feature that one chlorine is in ortho and a second chlorine is at the para position of the benzoyl residue. The ligand loses activity when only the chlorine atom in ortho position is present. If a chlorine atom is only present in the para position, the ligand is less active compared to the ortho position. The analysis of the binding pocket has shown that the molecules have to be small and hydrophobic. In case the triketone warhead is substituted, only small residues, for example methyl residues, are allowed. Otherwise the molecule would become to large for the tight binding pocket.

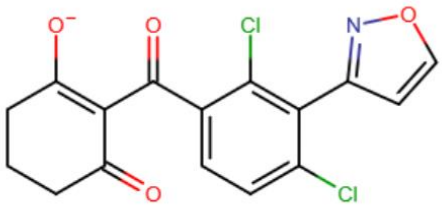


Figure 50: Scaffold of highly active triketones
This figure shows a triketone with 2 chlorine atoms at the benzoyl residue, It is a typical scaffold for highly active triketones.

5.7 Ligand-based-Pharmacophore Modeling

By use of the triketones data set (Chapter 5.1.2) ligand-based-pharmacophore models should be designed.

As described in the ligand alignment (Chapter 5.3), it was important to define the chelate complex as well as a hydrophobic feature after the 1,3,5-trione moieties. The analysis of the binding pocket (figure xx) and the protein-ligand interaction fingerprint (figure xx) provided knowledge that was essential for the creation and optimization of pharmacophore models.

This Figure 51 shows the pharmacophore model "Top 2 Bioactive" (Chapter 4.8.4) on which a triketone was added for visualisation. The chelate complex bond is defined by two hydrogen acceptor features, each directed to the iron ion of the protein. Additionally, there is a negative ionisable feature and a directed iron bonding feature. The negative ionisable feature is at the site of the triketone warhead where it would be expected due to the bioactive enolate form of the enol tautomer. At the triketone warhead there is a hydrogen acceptor feature which has been added additionally. This feature has no interaction pattern to the protein, but it defines a triketone more precisely and can therefore be searched more efficiently in huge databases. Furthermore, there is a hydrogen acceptor feature and two hydrophobic features, whereas one hydrophobic feature is ubiquitous for an active triketone.

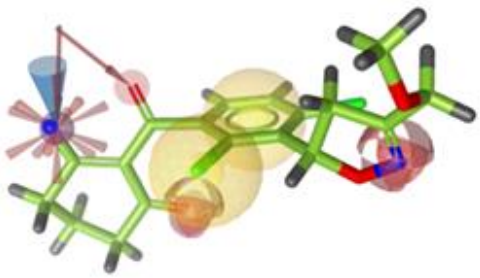


Figure 51: Triketone together with an exemplary Pharmacophore Model

A pharmacophore model defined for a triketone is depicted. It consists of several features, which should find an active triketone during the virtual screening. The yellow spheres are hydrophobic interaction. Hydrogen acceptor features are shown as red spheres and can be directed (arrows), as visible at the triketone warhead. A negative ionizable area (star-shaped) and an iron binding feature (blue) are also shown at the warhead. The shown molecule is the co-crystallized ligand from 5YY6. These depicted features can be chosen from the software LigandScout.

Figure 52 displays all pharmacophore models that have already been explained in Chapter 4.8 without a corresponding ligand but from the same angle as seen in Figure 51. All pharmacophore models have the essential features that define the triketone warhead and the hydrophobic feature after the chelate complex binding. The pharmacophore "Top 10 Arabidopsis" differs from the "Top 2 Bioactive" by an optional hydrophobic feature and by the fact that the hydrogen acceptor feature increases a higher tolerance sphere. This optional feature was important to increase the number of hits. "Top 10 Maize" has two hydrogen acceptor features in addition to "Top 10 Arabidopsis", one of which is optional. Since there has always been an intention to find a selectivity between ARBTH and ZEAMX and thus a selective triketone, the two pharmacophores "Top 10 Arabidopsis" and "Top 10 Maize" were generated.

The model "15 leftovers" has two hydrogen acceptor features, one of which is optional. It differs from the other models mentioned so far in that the 3D arrangement of the hydrogen acceptor features is different as before. Furthermore, this model has only one essential hydrophobic feature.

The pharmacophores "Apo Site Acceptor" and "Apo Site Donor" differ from each other only in the directional donor feature, which replaces the acceptor feature at the same position. Here it is possible to use the information obtained from the Protein-Ligand Interaction Fingerprint (Figure 44). The Gln 307 (ARBTH) or the equivalent Gln 293 (ZEAMX) can create acceptor or donor bindings. Since the basis for the generation of the pharmacophores was not a ligand, but the binding pocket, there are defined only essential features but not their directionalities. The iron bonding feature cannot be displayed either. The necessary hydrophobic feature was added manually as it is known to be important for the definition of the triketone (see Chapter 5.3). The optional hydrophobic feature is due to the many phenylalanines in the binding pocket. In general, better results were achieved during validation by adding instead of removing optional features.

Because the crystal structures are known, the size of the binding pocket could be reflected by exclusion volumes. The respective spars in Figure 52 were optimized and the model "Exclusion volumes" was created.

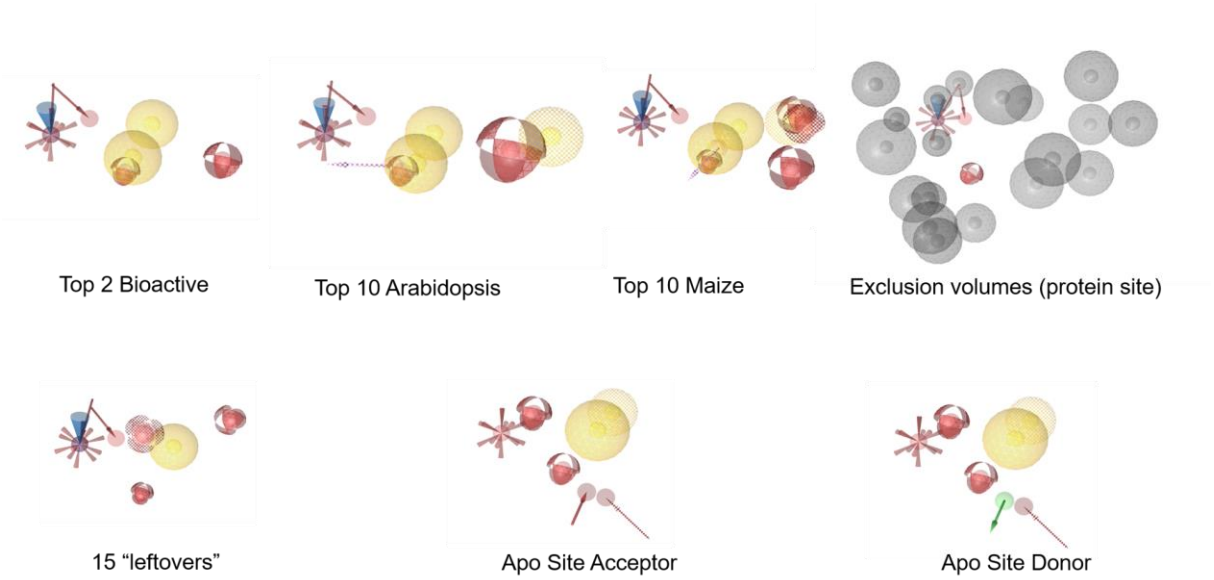


Figure 52: All representative Pharmacophore Models

The important and representative Pharmacophore models are illustrated, which were generated by LigandScout. The models have been created by selected ligands, merged, and subsequently optimized. The optimized pharmacophores were mapped back to the structure in order to match the structure and exclude sources of error.

However, it is uncertain, whether the triketones were classified according to their actual bioactivity. From the 10 most active compounds of each enzyme source a merged pharmacophore was created. Nine of the ten most bioactive compounds were tested with an enzyme assay for ARBTH, which were developed at a later stage. The problems of the different assays have already been explained in detail in Chapter 5.2. The enzyme assay has improved over time. It is therefore possible that technical progress and more sensitive measurement methods have led to better results. Variation of one order of magnitude is common in data sets coming from the same laboratory.

Based on the triketone data set and the additional knowledge of their target structure good pharmacophore models could be created and evaluated by virtual screening. The described pharmacophores are only selected models that were either satisfactory or important to investigate.

5.8 Virtual Screening

Different pharmacophores have been generated, as described in chapter 4.8.
In the Screening Perspective of LigandScout all pharmacophore models are stored. The goal of every pharmacophore is to match all active and no inactive ligands. Each pharmacophore model was screened against the created libraries (Chapter 4.8.8) and the number of hits carefully documented.
Most pharmacophores from the inactive ligands only hit themselves. The aim of pharmacophore modeling based on the inactive ligands was to generate a model that only hits inactive ligands and preferably no active ligands. A maximum of 3 out of 26 inactive ligands was found. This was accompanied by an increased number of active ligands found. None of the pharmacophores created from the inactive docking poses yielded a satisfying model.
Almost all ligands were detected as hits with this pharmacophore model, which was supposed to filter out the ligands that were too large. The problem was that ligands of each category of bioactivity were filtered out and therefore no precise statement can be made. This model finds too many hits (see Figure 27).
Generating pharmacophore models from the best active ligands was more successful. As described in chapter 4.8.5, the 10 most active ligands of the respective enzyme source were used to generate models. Pharmacophore models were developed, which mainly hit the active ligands during screening and if possible, no inactive ligands. They belong to the best models that could be created in this project and are shown in Figure 53.
Figure 53 shows a pharmacophore fingerprint. It visualizes the hits of the screenings by the corresponding pharmacophore models. The ligands are screened from the self-generated database. The triketones are sorted in descending order relative to their bioactivity data. For clarification, the classification of bioactivity was highlighted in color. The green section indicates the very active substances. The yellow border defines ligands with moderate activity and the last area is red for the inactive molecules. Thus, the most active substances are at the top of the column and the inactive ones at the bottom. The remaining six columns show the hits of the respective pharmacophore models in purple. If a ligand is hit by a pharmacophore model during virtual screening, the corresponding row for the hit is marked in the column of the

pharmacophore. The colorless lines do not indicate any hits for the respective model. In this way, the pharmacophore fingerprint is created, and can be easily illustrated.
Since there are 381 very active (IC_{50}: E-9 to E-7), 84 moderately active (IC_{50}: E-6 to E-5) and 26 inactive ligands, a maximum of 392 hits can be mapped in the fingerprint. In the first column of the figure, the ligands from the database are sorted according their bioactivity with a descending order.
The pharmacophore models "Apo Side Donor" and "Apo Side Acceptor" are unsatisfactory models, because they do not discriminate in terms of bioactivity. Therefore, these two models are not presented in this figure. However, it was important to create the Apo site grid and to generate the pharmacophore models from the protein site with the available information from the structure and the protein-ligand interaction fingerprint (Figure 44).
„Top 2 Bioactive“ is a pharmacophore consisting of the two most active triketone substances from the data set (Chapter 4.8.4). This pharmacophore hits 195 (51,18%) very active, 18 (21,18%) moderate active, and no inactive molecules.
The promising pharmacophore model „Top 10 Arabidopsis“ is derived from the 10 most active ligands of ARBTH (Chapter 4.8.5). Although this model hits one (3,85%) inactive ligand, there are 31 (36,47%) moderately active and 262 (68,77%) very active ligands. Many compounds are hit which are mainly in the upper rows due to their high bioactivity. 294 of 492 substances are hit by this pharmacophore model.
The next pharmacophore model „Top 10 Maize“ (Chapter 4.8.5) has a total of 99 hits, of which 90 (23,62%) are very active and 9 (10,59%) are moderate active. None of the inactive compounds was found, which is considered a positive result.
The pharmacophore „Exclusion volumes“ includes a shell of exclusion volumes (Chapter 4.8.3). The aim was to filter out the inactive triketones by limiting the potential binding site volume. With a total of 399 hits, this model hits almost all (81,09%) ligands of each category of bioactivity and is not a satisfying pharmacophore model.
The Model „15 leftovers“ (Chapter 4.8.7) has found a total of 245 hits. Of these results, 214 (56,17%) are very active, 29 (34,12%) moderate active and two (7,69%) inactive ligands. The most promising Pharmacophore „Top 10 Arabidopsis” did not find every active ligand. Therefore, the pharmacophore models „Top 10 Arabidopsis“ and „15 leftovers“ have been combined and are shown in the column “Combination” (Figure 50). Consequently, for a Boolean expression in LigandScout, the conditions were set that all ligands matching of at least one pharmacophore were written out.

Thus, the models find hits especially in the uppermost range of the most active ligands, but also three inactive substances are detected. In total, 297 (77,95%) highly active, 46 (54,12%) moderately active and three (11,54%) inactive hits were matched.

	Top 2	Top 10 AT	Top 10 ZM	excl. Vol	15 Leftovers	Combination
active	51,18%	68,77%	23,62%	82,15%	56,17%	77,95%
moderate	21,18%	36,47%	10,59%	76,47%	34,12%	54,12%
inactive	0,00%	3,85%	0,00%	80,77%	7,69%	11,54%

Figure 53: Pharmacophore Fingerprint.

The ligands to be hit are sorted in descending order according to their bioactivity. Several pharmacophores have been generated and are listed in columns. For each ligand hit by the pharmacophore, a purple bar is generated. The table below shows the hits in percent depending on the classification of the bioactivity into inactive, moderately active and very active.

In pharmacophore modeling, overfitting often occurs when combined pharmacophore models are adjusted too strongly to the data set to be screened. In the last pharmacophore from Figure 53, two different pharmacophore models are combined. Probably, the pharmacophore model „Top 10 Arabidopsis“ is sufficient to screen huge, publicly accessible databases of molecules and to obtain good active triketones.

The three inactive substances found via virtual screening are planned to be retested for their activity at BASF. Since in the first data set a few ligands were potentially misclassified, further misleading errors should be avoided. As the combination of the pharmacophore model of "Top 10 Arabidopsis" and "15 leftovers" finds many highly active ligands, a further experimental evaluation of the three inactive compounds would be very valuable.

During the screening, attention was given whether there were significant differences in selectivity between ARBTH and ZEAMX. The hits were not structurally different according their enzyme source. Thus, no relationship could be found between the generated pharmacophore models and the screened hits. The pharmacophores, which were created from the 10 most active triketones of each enzyme source, also found ligands tested in both ZEAMX and ARBTH.
As a result, no difference in selectivity can be made on the basis of the available data, neither structure-based nor via the ligands, which have only been tested in one enzyme source.

5.9 Retesting

As described in chapter 5.1.1, six molecules were selected for retesting. These six structures were tested in the past (Table 3, left) for ARBTH and ZEAMX. The reason for retesting was the significant difference of their activity data in the two assays (i.e. ARBTH vs. ZEAMX). Especially four compounds (LS-Cores 297300, 328115, 337128 and 382565) were indicated with a difference larger than one order of magnitude in their IC_{50} values. Although fluctuations in a range of one order of magnitude is in the normal range, these six compounds are conspicuous compared to the other 20 ligands, all of which have been tested on both plant targets.
One aim of this thesis was to find ligands that differ from ARBTH and ZEAMX due to their selectivity. This was a first ligand-based indication of a possible difference in selectivity. As Table 3 shows, the IC_{50} values of all four ligands have become lower. According to the new test they have a higher bioactivity. These four herbicides can therefore now be classified as more effective. The substance with LS-Core 298647 was tested in both plants. Only the extent of inhibition was indicated, as the compound had either a solubility problem (which meant that only a low concentration could become active) or a problem with detection. The substance was retested and the IC_{50} value could be measured. In ZEAMX the value of 1.13E-5 is in the moderate range, in ARBTH this ligand is in the lower range of the very active triketones with the IC_{50} value of 6.79E-7. This ligand can be classified as active instead of inactive.
The molecule with LS-Core 321416 had almost identical IC_{50} values for ZEAMX and ARBTH. Also, this should be retested. With the new assay, the data are almost iden-

tical and active as the previous IC_{50} values. Therefore, the ligand remains classified as very effective triketone.

The retesting results in the up-to-date assays for 4-HPPD are not surprising. the assay has been optimized over time and the measurement technique has improved (Chapter 4.3). The ligands can be classified more effectively or remain consistently effective. The fluctuations between the enzymes source are also decreased. As a result, there is no differentiation of IC_{50} values leading to the solution of the problem of selectivity of ARBTH and ZEAMX. It can be observed that triketones are a bit more active in ARBTH than in ZEAMX. Nevertheless, this result cannot be explained by the enzyme structural formula.

Table 3: Results of the retesting of six triketones which were tested on ARBTH and ZEAMX.

This table shows the old and new IC_{50} values of six molecules after retesting. The ligand, which had too low a concentration to be measured in ARBTH and ZEAMX, could be tested as an active ligand. The ligand with the nearly identical IC_{50} values in ARBTH and ZEAMX was measured correctly and the four ligands, which were differently active in both plants, are not significantly different according to the new measurements. All data were generated with the current assay (Chapter 4.3)

	ARBTH	ZEAMX
297300	1.1E-7 M	2.6E-6 M
328115	1.1E-7 M	1.18E-6 M
337128	1.8E-7 M	7.7E-6 M
382565	7.78E-7 M	1.06E-6 M
298647	11% Inhib.	31% Inhib.
321416	5.32E-8 M	5.3E-8 M

IC_{50} values before retesting

	ARBTH	ZEAMX
297300	1.5E-8 M	4.75E-7 M
328115	1.96E-8 M	9.67E-8 M
337128	1.15E-7 M	2.91E-6 M
382565	4.19E-8 M	3.32E-8 M
298647	6.79E-7 M	1.13E-5 M
321416	6.07E-8 M	5.54E-8 M

Retested IC_{50} values

These substances with LS-Cores 343069, 381048, and 4256292 are triketones classified as inactive in the provided dataset. However, these three compounds were hit by the ligand-based pharmacophore model "Top 10 Arabidopsis" or the model "15 leftovers", respectively the combination of both (Chapter 5.8). The pharmacophores should only hit bioactive triketones at virtual screening. For this reason, these three further substances were also retested. in order to confirm or to disprove it.

Table 4 shows the results of the retested compounds. When the activity was tested in the nineties, no IC_{50} values could be derived. Only the percentage of enzyme inhibition was indicated. LS-Core 343069 was originally measured only in ZEAMX, the other two triketones only in ARBTH.

In the right table the retesting results are given. All three substances were tested in both enzyme sources using the enzyme assay described in Chapter 4.3. Thus, IC_{50} values are available for both enzyme sources. Only for LS-Cores 343069 no IC_{50} value could be determined for ZEAMX, because it was not active enough. For ARBTH this substance is moderate active.

LS-Core 381048 is also moderately active at both enzyme assays, indicated by similar IC_{50} values (Δ= 4.15E-6 and 8.66E-6).

The third triketone with the LS-Core 4256292 was tested as very active in both plant enzymes. There is one order of magnitude difference. This substance is slightly more active in ARBTH according to the IC_{50} values. This is in general agreement with all other test results showing higher activity for the same compound at the Arabidopsis enzyme assay. Based on the retesting, it can be confirmed that the results with the most recent enzyme assay indicate higher activity. The created pharmacophore model finds many highly active triketones and does not hit any active ligand from the created databases.

Table 4:

This table shows the old value of inhibition (%) and new IC_{50} values of three triketones after retesting. The ligands, which had too low a concentration to be measured bevor in ARBTH and ZEAMX, could be retested as active triketones. These results demonstrate that the selected pharmacophore models are very satisfying and only active triketones are hit in virtual screening.

	ARATH	ZEAMX
343069		43
381048	46	
4256292	86	

Old value: inhibition in %

	ARATH	ZEAMX
343069	1.41E-6	-
381048	8.66E-6	4.15E-6
4256292	5.78E-8	5.69E-7

New IC_{50} values

6 Conclusion

The aim of this study was to find triketones that act selectively on the enzyme 4-HPPD in weeds, such as the model plant ARBTH, and not on the crop ZEAMX. Furthermore, it should be established that active triketones can be discriminated from inactive ones. Since the crystal structures of the two enzymes are very similar (sequence identity of over 62%) and the binding site topology did not give rise to any significant differences, no distinction in selectivity was found. The only structural difference exists in the C-terminal helix H11 of the enzymes. It was hypothesized that the "closed" and the "open" conformation are two different “frozen” conformations of these helices that physiologically can easily convert. Thus, H11 would have the gate function to open and close the binding pocket.

In ligand-based pharmacophore modeling, the knowledge derived from the co-crystallized ligand-enzyme X-ray structures helped to optimize the pharmacophore models. The model "Top 10 Arabidopsis" was created, which would probably be satisfactory on its own. If this model is combined with the second model "15 Leftovers", more active triketones could be hit. Among them are all very active hits but also three inactive ligands were found.

Since the three inactive hits could be turned to active via an initiated retesting campaign, the predictive quality of this model might be even better. Since the binding pockets have not been different, no solution to the problem of selectivity could be found. If differences would have been present, it might have been possible to determine selectivity based on structure. Also ligand-based, no selective discrimination could have been made because the measured data did not detect any significant differences in the enzyme assay. The results of the retesting also confirm the insight from the structure that the ligands do not differ significantly.

In retrospective, the procedure described in this thesis was correct, based on the available data. Due to the gained experience the same approach could be performed much faster today. Both the structure-based and the ligand-based methods should be applied. Based on the structure, the ligand-based pharmacophore models can be optimized. The intersection of both methods yields a satisfactory model.

7 Future Perspective

As discussed in this thesis, the C-terminal Helix H11 can move dynamically. Since dynamic information about the H11 helices is missing, it would be possible to simulate the dynamics in another project. In the binding pocket of the two enzymes, the amino acid glutamate differs from glutamine. These two amino acids belong to H11 with potential gate functionality. Therefore, this should be investigated in more detail. It could be that the different amino acids lead to a new interaction that can selectively differentiate between ARBTH or ZEAMX. Due to the rigid crystal structures, it was not feasible in the context of this diploma thesis. Based on molecular dynamic simulations, free energy calculations can be performed. In a later step, this can facilitate the synthesis decisions during lead optimization.

In addition, it can be investigated whether the complete C-terminal helix can be imaged by software tools. Thus, the different H11 helices can be analysed. Possibly a new 4-HPPD crystal structure can be crystallized by crystallography, whose resolution is better, if this is the solution to the problem. However, it should be evaluated before what exactly is the reason that the crystal structure cannot be displayed completely in the interface.

Triketones are only one class of inhibitors that act on 4-HPPD. In a further project the other chemotypes could be investigated. It can be researched whether there are already selective compounds and how they would structurally behave towards the triketones.

In plants, selectivity is often accompanied by the metabolism. It is therefore helpful to investigate the individual metabolic steps. It may be possible to find a mechanism that enables one of the analysed plants to metabolise drugs faster than the other. Finally, the existing triketones could also be tested for ZEAMX and ARBTH with the current enzyme assay, so that the bioactivity values are more comparable.

8 References

(1) Rocaboy-Faquet, E.; Noguer, T.; Romdhane, S.; Bertrand, C.; Dayan, F. E.; Barthelmebs, L. Novel Bacterial Bioassay for a High-Throughput Screening of 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors. *Applied Microbiology and Biotechnology* **2014**, *98* (16), 7243–7252. https://doi.org/10.1007/s00253-014-5793-5.

(2) Ndikuryayo, F.; Moosavi, B.; Yang, W.-C.; Yang, G.-F. 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors: From Chemical Biology to Agrochemicals. *J. Agric. Food Chem.* **2017**, *65* (39), 8523–8537. https://doi.org/10.1021/acs.jafc.7b03851.

(3) Sun, C.; Jin, Y.; He, H.; Wang, W.; He, H.; Fu, Z.; Qian, H. Two Novel Herbicide Candidates Affect Arabidopsis Thaliana Growth by Inhibiting Nitrogen and Phosphate Absorption. *Pesticide Biochemistry and Physiology* **2015**, *123*, 1–8. https://doi.org/10.1016/j.pestbp.2015.03.005.

(4) Classification of Herbicides According to Mode of Action https://hracglobal.com/.

(5) Fu, Y.; Sun, Y.-N.; Yi, K.-H.; Li, M.-Q.; Cao, H.-F.; Li, J.-Z.; Ye, F. 3D Pharmacophore-Based Virtual Screening and Docking Approaches toward the Discovery of Novel HPPD Inhibitors. *Molecules* **2017**, *22* (6), 959. https://doi.org/10.3390/molecules22060959.

(6) Fu, Y.; Zhang, S.-Q.; Liu, Y.-X.; Wang, J.-Y.; Gao, S.; Zhao, L.-X.; Ye, F. Design, Synthesis, SAR and Molecular Docking of Novel Green Niacin-Triketone HPPD Inhibitor. *Industrial Crops and Products* **2019**, *137*, 566–575. https://doi.org/10.1016/j.indcrop.2019.05.070.

(7) Soteres, J.; Glasgow, L.; Peterson, M.; Obrigawitch, T.; Evans, R.; Strek, H.; Pawlak, J.; Wilson, S. Herbicide Resistance Management and Sustainable AG: An Industry Perspective. In *Outlooks on Pest Management*; Research Information, **2013**; Vol. 24, pp 230–232.

(8) Fu, Y.; Sun, Y.-N.; Yi, K.-H.; Li, M.-Q.; Cao, H.-F.; Li, J.-Z.; Ye, F. Combination of Virtual Screening Protocol by in Silico toward the Discovery of Novel 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors. *Front Chem* **2018**, *6*, 14–14. https://doi.org/10.3389/fchem.2018.00014.

(9) Chiddarwar, R. K.; Rohrer, S. G.; Wolf, A.; Tresch, S.; Wollenhaupt, S.; Bender, A. In Silico Target Prediction for Elucidating the Mode of Action of Herbicides Including Prospective Validation. *Journal of Molecular Graphics and Modelling* **2017**, *71*, 70–79. https://doi.org/10.1016/j.jmgm.2016.10.021.

(10) Retzinger, E. J.; Mallory-Smith, C. Classification of Herbicides by Site of Action for Weed Resistance Management Strategies. *Weed Technology* **1997**, *11* (2), 384–393. https://doi.org/10.1017/S0890037X00043116.

(11) Edited by Krämer, W.; Schirmer, U.; Jeschke, P.; Witschel, M. *Modern Crop Protection Compounds*; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, **2012**; Vol. 1.

(12) Cho, J. E.; Kim, J. T.; Kim, E.; Ko, Y. K.; Kang, N. S. The Structure-Based Three-Dimensional Pharmacophore Models for Arabidopsis Thaliana HPPD Inhibitors as Herbicide. *Bull. Korean Chem. Soc.* **2013**. https://doi.org/10.5012/bkcs.2013.34.10.2909.

(13) Klebe, G. *Wirkstoffdesign-Entwurf Und Wirkung von Arzneistoffen*, 2nd ed.; Spektrum Akademischer Verlag Heidelberg, **2009**.

(14) Colquitt, R. B.; Colquhoun, D. A.; Thiele, R. H. In Silico Modelling of Physiologic Systems. *Best Practice & Research Clinical Anaesthesiology* **2011**, *25* (4), 499–510. https://doi.org/10.1016/j.bpa.2011.08.006.
(15) Myatt, G. J.; Ahlberg, E.; Akahori, Y.; Allen, D.; Amberg, A.; Anger, L. T.; Aptula, A.; Auerbach, S.; Beilke, L.; Bellion, P.; et al. In Silico Toxicology Protocols. *Regulatory Toxicology and Pharmacology* **2018**, *96*, 1–17. https://doi.org/10.1016/j.yrtph.2018.04.014.
(16) Amberg, A. In Silico Methods. In *Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays*; Vogel, H. G., Maas, J., Hock, F. J., Mayer, D., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2013**; pp 1273–1296. https://doi.org/10.1007/978-3-642-25240-2_55.
(17) Chen, Z.; Tian, G.; Wang, Z.; Jiang, H.; Shen, J.; Zhu, W. Multiple Pharmacophore Models Combined with Molecular Docking: A Reliable Way for Efficiently Identifying Novel PDE4 Inhibitors with High Structural Diversity. *J. Chem. Inf. Model.* **2010**, *50* (4), 615–625. https://doi.org/10.1021/ci9004173.
(18) Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-Pharmacophore Superpositioning and Pattern Matching in Computational Drug Design. *Drug Discovery Today* **2008**, *13* (1), 23–29. https://doi.org/10.1016/j.drudis.2007.09.007.
(19) EPPO codes https://www.eppo.int/RESOURCES/eppo_databases/eppo_codes (accessed Aug 15, 2019).
(20) Sivasubramanian, R.; Mukhi, N.; Kaur, J. Arabidopsis Thaliana: A Model for Plant Research. In *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology*; Bahadur, B., Venkat Rajam, M., Sahijram, L., Krishnamurthy, K. V., Eds.; Springer India: New Delhi, **2015**; pp 1–26. https://doi.org/10.1007/978-81-322-2283-5_1.
(21) Katam, R.; Panthee, D. R.; Bhattacharya, A.; Basha, S. M.; Kole, C. Arabidopsis. In *Wild Crop Relatives: Genomic and Breeding Resources: Oilseeds*; Kole, C., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2011**; pp 1–16. https://doi.org/10.1007/978-3-642-14871-2_1.
(22) Hayashi, M.; Nishimura, M. Arabidopsis Thaliana—A Model Organism to Study Plant Peroxisomes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2006**, *1763* (12), 1382–1391. https://doi.org/10.1016/j.bbamcr.2006.08.014.
(23) Holland, C. K.; Jez, J. M. Arabidopsis: The Original Plant Chassis Organism. *Plant Cell Reports* **2018**, *37* (10), 1359–1366. https://doi.org/10.1007/s00299-018-2286-5.
(24) Koornneef, M.; Meinke, D. The Development of Arabidopsis as a Model Plant. *The Plant Journal* **2010**, *61* (6), 909–921. https://doi.org/10.1111/j.1365-313X.2009.04086.x.
(25) Domínguez-Ramírez, L.; Díaz-Ruiz, G.; Wacher, C. Maize (Zea Mays l. Subsp. Mays) Fermentation. **2017**, No. Departamento de Alimentos y Biotecnología; Facultad de Química; UNAM, 177–215. https://doi.org/10.1201/9781315205359.
(26) Prasanna, B. M.; Vasal, S. K.; Kassahun, B.; Singh, N. N. Quality Protein Maize. *Current Science* **2001**, *81* (10), 1308–1319.
(27) Arabidopsis Thaliana, morphopedia.org/index.php?title=Image:Arabidopsis.jpg (accessed Jan 28, 2020) and Corn maze http://www.hathawayfarm.com/corn-maze/corn/ (accessed Aug 15, 2019).
(28) Garcia, I.; Job, D.; Matringe, M. Inhibition of P-Hydroxyphenylpyruvate Dioxygenase by the Diketonitrile of Isoxaflutole: A Case of Half-Site Reactivity. *Biochemistry* **2000**, *39* (25), 7501–7507. https://doi.org/10.1021/bi000135h.

(29) Fritze, I. M.; Linden, L.; Freigang, J.; Auerbach, G.; Huber, R.; Steinbacher, S. The Crystal Structures of Zea Mays and Arabidopsis 4-Hydroxyphenylpyruvate Dioxygenase. *Plant Physiol* **2004**, *134* (4), 1388–1400. https://doi.org/10.1104/pp.103.034082.
(30) Witschel, M. Design, Synthesis and Herbicidal Activity of New Iron Chelating Motifs for HPPD-Inhibitors. *Bioorganic & Medicinal Chemistry* **2009**, *17* (12), 4221–4229. https://doi.org/10.1016/j.bmc.2008.11.006.
(31) Siehl, D. L.; Tao, Y.; Albert, H.; Dong, Y.; Heckert, M.; Madrigal, A.; Lincoln-Cabatu, B.; Lu, J.; Fenwick, T.; Bermudez, E.; et al. Broad 4-Hydroxyphenylpyruvate Dioxygenase Inhibitor Herbicide Tolerance in Soybean with an Optimized Enzyme and Expression Cassette. *Plant Physiol.* **2014**, *166* (3), 1162. https://doi.org/10.1104/pp.114.247205.
(32) Meazza, G.; Scheffler, B. E.; Tellez, M. R.; Rimando, A. M.; Romagni, J. G.; Duke, S. O.; Nanayakkara, D.; Khan, I. A.; Abourashed, E. A.; Dayan, F. E. The Inhibitory Activity of Natural Products on Plant P-Hydroxyphenylpyruvate Dioxygenase. *Phytochemistry* **2002**, *60* (3), 281–288. https://doi.org/10.1016/S0031-9422(02)00121-8.
(33) Moran, G. R. 4-Hydroxyphenylpyruvate Dioxygenase. *Archives of Biochemistry and Biophysics* **2005**, *433* (1), 117–128. https://doi.org/10.1016/j.abb.2004.08.015.
(34) Garcia, I.; Rodgers, M.; Pepin, R.; Hsieh, T.-F.; Matringe, M. Characterization and Subcellular Compartmentation of Recombinant 4-Hydroxyphenylpyruvate Dioxygenase from Arabidopsis in Transgenic Tobacco. *Plant Physiol.* **1999**, *119* (4), 1507. https://doi.org/10.1104/pp.119.4.1507.
(35) Garcia, I.; Rodgers, M.; Pepin, R.; Hsieh, T.-F.; Matringe, M. Characterization and Subcellular Compartmentation of Recombinant 4-Hydroxyphenylpyruvate Dioxygenase from Arabidopsis in Transgenic Tobacco. *Plant Physiol.* **1999**, *119* (4), 1507. https://doi.org/10.1104/pp.119.4.1507.
(36) Lin, H.; Yang, J.; Wang, D.; Hao, G.; Dong, J.; Wang, Y.; Yang, W.; Wu, J.; Zhan, C.; Yang, G. Molecular Insights into the Mechanism of 4-hydroxyphenylpyruvate Dioxygenase Inhibition: Enzyme Kinetics, X-ray Crystallography and Computational Simulations. *The FEBS Journal* **2019**, *286* (5), 975–990. https://doi.org/10.1111/febs.14747.
(37) Ahrens, H.; Lange, G.; Müller, T.; Rosinger, C.; Willms, L.; van Almsick, A. 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors in Combination with Safeners: Solutions for Modern and Sustainable Agriculture. *Angewandte Chemie International Edition* **2013**, *52* (36), 9388–9398. https://doi.org/10.1002/anie.201302365.
(38) Schindler, C. E. M.; Hollenbach, E.; Mietzner, T.; Schleifer, K.-J.; Zacharias, M. Free Energy Calculations Elucidate Substrate Binding, Gating Mechanism, and Tolerance-Promoting Mutations in Herbicide Target 4-Hydroxyphenylpyruvate Dioxygenase. *Protein Science* **2019**, *28* (6), 1048–1058. https://doi.org/10.1002/pro.3612.
(39) Hawkes, T. R.; Langford, M. P.; Viner, R.; Blain, R. E.; Callaghan, F. M.; Mackay, E. A.; Hogg, B. V.; Singh, S.; Dale, R. P. Characterization of 4-Hydroxyphenylpyruvate Dioxygenases, Inhibition by Herbicides and Engineering for Herbicide Tolerance in Crops. *Pesticide Biochemistry and Physiology* **2019**, *156*, 9–28. https://doi.org/10.1016/j.pestbp.2019.01.006.
(40) Wu, C.-S.; Huang, J.-L.; Sun, Y.-S.; Yang, D.-Y. Mode of Action of 4-Hydroxyphenylpyruvate Dioxygenase Inhibition by Triketone-Type Inhibitors. *J. Med. Chem.* **2002**, *45* (11), 2222–2228. https://doi.org/10.1021/jm010568y.

(41) Matringe, M.; Sailland, A.; Pelissier, B.; Rolland, A.; Zink, O. P-Hydroxyphenylpyruvate Dioxygenase Inhibitor-Resistant Plants. *Pest Management Science* **2005**, *61* (3), 269–276. https://doi.org/10.1002/ps.997.
(42) Gálvez-Valdivieso, G.; Pineda, M.; Aguilar, M. FUNCTIONAL CHARACTERIZATION AND EXPRESSION ANALYSIS OF P-HYDROXYPHENYLPYRUVATE DIOXYGENASE FROM THE GREEN ALGA CHLAMYDOMONAS REINHARDTII (CHLOROPHYTA)1. *Journal of Phycology* **2010**, *46* (2), 297–308. https://doi.org/10.1111/j.1529-8817.2010.00812.x.
(43) Fiedler, E.; Soll, J.; Schultz, G. The Formation of Homogentisate in the Biosynthesis of Tocopherol and Plastoquinone in Spinach Chloroplasts. *Planta* **1982**, *155* (6), 511–515. https://doi.org/10.1007/BF01607575.
(44) Jefford, C. W.; Cadby, P. A. Evaluation of Models for the Mechanism of Action of 4-Hydroxyphenylpyruvate Dioxygenase. *Experientia* **1981**, *37* (11), 1134–1137. https://doi.org/10.1007/BF01989880.
(45) Tsegaye, Y.; Shintani, D. K.; DellaPenna, D. Overexpression of the Enzyme P-Hydroxyphenolpyruvate Dioxygenase in Arabidopsis and Its Relation to Tocopherol Biosynthesis. *Plant Physiology and Biochemistry* **2002**, *40* (11), 913–920. https://doi.org/10.1016/S0981-9428(02)01461-4.
(46) Soll, J.; Schultz, G.; Joyard, J.; Douce, R.; Block, M. A. Localization and Synthesis of Prenylquinones in Isolated Outer and Inner Envelope Membranes from Spinach Chloroplasts. *Archives of Biochemistry and Biophysics* **1985**, *238* (1), 290–299. https://doi.org/10.1016/0003-9861(85)90167-5.
(47) van Almsick, A. New HPPD-Inhibitors – A Proven Mode of Action as a New Hope to Solve Current Weed Problems. **2009**, *Volume 20* (Number 1), 27-30(4). https://doi.org/10.1564/20feb09.
(48) Grossmann, K.; Ehrhardt, T. On the Mechanism of Action and Selectivity of the Corn Herbicide Topramezone: A New Inhibitor of 4-Hydroxyphenylpyruvate Dioxygenase. *Pest Management Science* **2007**, *63* (5), 429–439. https://doi.org/10.1002/ps.1341.
(49) Beaudegnies, R.; De Mesmaeker, A.; Mallinger, A.; Baalouch, M.; Goetz, A. Design and Synthesis of Novel Spirocyclopropyl Cyclohexane-1,3-Diones and -1,3,5-Triones for Their Incorporation into Potent HPPD Inhibitors. *Tetrahedron Letters* **2010**, *51* (20), 2741–2744. https://doi.org/10.1016/j.tetlet.2010.03.047.
(50) NIH: U.S. National Library of Medicine https://pubchem.ncbi.nlm.nih.gov/compound/Mesotrione (accessed Jul 30, 2019).
(51) NIH: U.S. National Library of Medicine https://pubchem.ncbi.nlm.nih.gov/compound/Sulcotrione (accessed Jul 30, 2019).
(52) NIH: U.S. National Library of Medicine https://pubchem.ncbi.nlm.nih.gov/compound/11556911 (accessed Aug 1, 2019).
(53) Lygin, A. V.; Kaundun, S. S.; Morris, J. A.; Mcindoe, E.; Hamilton, A. R.; Riechers, D. E. Metabolic Pathway of Topramezone in Multiple-Resistant Waterhemp (Amaranthus Tuberculatus) Differs From Naturally Tolerant Maize. *Front Plant Sci* **2018**, *9*, 1644–1644. https://doi.org/10.3389/fpls.2018.01644.
(54) Hu, B.; Lill, M. A. Exploring the Potential of Protein-Based Pharmacophore Models in Ligand Pose Prediction and Ranking. *J. Chem. Inf. Model.* **2013**, *53* (5), 1179–1190. https://doi.org/10.1021/ci400143r.
(55) Wermuth C. G.; Ganellin C. R.; Lindberg P.; Mitscher L. A. Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1998). *pac* **1998**, *70* (5), 1129. https://doi.org/10.1351/pac199870051129.

(56) Edited by Langer, T.; Hoffmann, R. *Pharmacophores and Pharmacophore Search*; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, 2006.
(57) Qing, X.; Lee., X.; De Raeymaeker, J.; Tame, J.; Zhang., K.; De Maeyer, M.; Voet, A. Pharmacophore Modeling: Advances, Limitations, and Current Utility in Drug Discovery. *Journal of Receptor, Ligand and Channel Research* **2014**.
(58) Kutlushina, A.; Khakimova, A.; Madzhidov, T.; Polishchuk, P. Ligand-Based Pharmacophore Modeling Using Novel 3D Pharmacophore Signatures. **2018**. https://doi.org/10.3390/molecules23123094.
(59) Dong, G.; Deng, X. J.; Xiao, J. H. On the Integration of Pharmacophore Model and Molecular Docking Method. *Journal of Algorithms & Computational Technology* **2011**, *5* (1), 129–137. https://doi.org/10.1260/1748-3018.5.1.129.
(60) López-Ramos, M.; Perruccio, F. HPPD: Ligand- and Target-Based Virtual Screening on a Herbicide Target. *J. Chem. Inf. Model.* **2010**, *50* (5), 801–814. https://doi.org/10.1021/ci900498n.
(61) MOE. Chemical Computing Group https://www.chemcomp.com/ (accessed Aug 6, 2019).
(62) Deeth, R. J.; Fey, N.; Williams–Hubbard, B. DommiMOE: An Implementation of Ligand Field Molecular Mechanics in the Molecular Operating Environment. *Journal of Computational Chemistry* **2005**, *26* (2), 123–130. https://doi.org/10.1002/jcc.20137.
(63) Schrödinger Drug Discovery https://www.schrodinger.com/drug-discovery (accessed Aug 6, 2019).
(64) Inteligand LigandScout http://www.inteligand.com/ligandscout3/ (accessed Aug 6, 2018).
(65) Lee, M.-L.; Aliagas, I.; Feng, J. A.; Gabriel, T.; O'Donnell, T. J.; Sellers, B. D.; Wiswedel, B.; Gobbi, A. Chemalot and Chemalot_knime: Command Line Programs as Workflow Tools for Drug Discovery. *Journal of Cheminformatics* **2017**, *9* (1), 38. https://doi.org/10.1186/s13321-017-0228-9.
(66) Nicola, G.; Berthold, M. R.; Hedrick, M. P.; Gilson, M. K. Connecting Proteins with Drug-like Compounds: Open Source Drug Discovery Workflows with BindingDB and KNIME. *Database (Oxford)* **2015**, *2015*, bav087. https://doi.org/10.1093/database/bav087.
(67) Gally, J.-M.; Bourg, S.; Do, Q.-T.; Aci-Sèche, S.; Bonnet, P. VSPrep: A General KNIME Workflow for the Preparation of Molecules for Virtual Screening. *Molecular Informatics* **2017**, *36* (10), 1700023. https://doi.org/10.1002/minf.201700023.
(68) Wollmann, T.; Erfle, H.; Eils, R.; Rohr, K.; Gunkel, M. Workflows for Microscopy Image Analysis and Cellular Phenotyping. *Journal of Biotechnology* **2017**, *261*, 70–75. https://doi.org/10.1016/j.jbiotec.2017.07.019.
(69) Fillbrunn, A.; Dietz, C.; Pfeuffer, J.; Rahn, R.; Landrum, G. A.; Berthold, M. R. KNIME for Reproducible Cross-Domain Analysis of Life Science Data. *Journal of Biotechnology* **2017**, *261*, 149–156. https://doi.org/10.1016/j.jbiotec.2017.07.028.
(70) DataWarrior. Open molecules.org http://www.openmolecules.org/datawarrior/index.html (accessed Aug 6, 2019).

9 Appendix

Table 5:Classification system of herbicides created by HRAC.

In this table the commercial herbicides are classified. For each herbicidal compound the target protein, a code consisting of a letter and the mode of action are given. The mode of action can be abbreviated with this one-letter code.

Mode of action	Code	Class	Compound
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Alloxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Butroxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Clethodim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Cloproxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Cycloxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Profoxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Sethoxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Tepraloxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Cyclohexanediones (DIMs)	Tralkoxydim
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Clodinafop-propargyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Clofop
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Cyhalofop-butyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Diclofop-methyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Fenoxaprop-ethyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Fenthiaprop
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Fluazifop-butyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Haloxyfop-methyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Isoxapyrifop
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Metamifop
Inhibition of acetyl CoA carboxylase (ACCase)	A	Aryloxyphenoxy-propionates (FOPs)	Quizalofop-ethyl
Inhibition of acetyl CoA carboxylase (ACCase)	A	Phenylpyrazoline	Pinoxaden
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Bispyribac-sodium
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Pyribenzoxim (prodrug of bispyribac)
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Pyriftalid
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Pyriminobac-methyl
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Pyrithiobac-sodium
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Pyrimisulfan
Inhibition of acetolactate synthase (ALS)	B	Pyrimidinyl (thio) benzoates	Triafamone

Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Cloransulam-methyl
Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Diclosulam
Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Florasulam
Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Flumetsulam
Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Metosulam
Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Penoxsulam
Inhibition of acetolactate synthase (ALS)	B	Triazolopyrimidine	Pyroxsulam
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Amidosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Azimsulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Bensulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Chlorimuron-ethyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Chlorsulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Cinosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Cyclosulfamuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Ethametsulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Ethoxysulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Flazasulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Flucetosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Flupyrsulfuron-methyl-Na
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Foramsulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Halosulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Imazosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Iodosulfuron-methyl-Na
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Mesosulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Metazosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Metsulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Nicosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Orthosulfamuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Oxasulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Primisulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Propyrisulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Prosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Pyrazosulfuron-ethyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Rimsulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Sulfometuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Sulfosulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Triasulfuron
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Tribenuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Thifensulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Trifloxysulfuron-Na
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Triflusulfuron-methyl
Inhibition of acetolactate synthase (ALS)	B	Sulfonylureas	Tritosulfuron
Inhibition of acetolactate synthase (ALS)	B	Imidazolinones	Imazamethabenz-methyl
Inhibition of acetolactate synthase (ALS)	B	Imidazolinones	Imazamox
Inhibition of acetolactate synthase (ALS)	B	Imidazolinones	Imazapic

Inhibition of acetolactate synthase (ALS)	B	Imidazolinones	Imazapyr
Inhibition of acetolactate synthase (ALS)	B	Imidazolinones	Imazaquin
Inhibition of acetolactate synthase (ALS)	B	Imidazolinones	Imazethapyr
Inhibition of acetolactate synthase (ALS)	B	Sulfonylamino-carbonyl-triazolinones	Flucarbazone-Na
Inhibition of acetolactate synthase (ALS)	B	Sulfonylamino-carbonyl-triazolinones	Propoxycarbazone-Na
Inhibition of acetolactate synthase (ALS)	B	Sulfonylamino-carbonyl-triazolinones	Thiencarbazone-methyl
Inhibition of PS II	C1	Triazines	Atraton
Inhibition of PS II	C1	Triazines	Atrazine
Inhibition of PS II	C1	Triazines	Ametryne
Inhibition of PS II	C1	Triazines	Aziprotryne=aziprotryn
Inhibition of PS II	C1	Triazines	Chlorazine
Inhibition of PS II	C1	Triazines	CP 17029
Inhibition of PS II	C1	Triazines	Cyanazine
Inhibition of PS II	C1	Triazines	Cyprazine
Inhibition of PS II	C1	Triazines	Desmetryne
Inhibition of PS II	C1	Triazines	Dimethametryn
Inhibition of PS II	C1	Triazines	Dipropetryn
Inhibition of PS II	C1	Triazines	Eglinazine-ethyl
Inhibition of PS II	C1	Triazines	Ipazine
Inhibition of PS II	C1	Triazines	Methoprotryne=methoprotryn
Inhibition of PS II	C1	Triazines	procyazine
Inhibition of PS II	C1	Triazines	Proglinazine-ethyl
Inhibition of PS II	C1	Triazines	Prometon
Inhibition of PS II	C1	Triazines	Prometryne
Inhibition of PS II	C1	Triazines	Propazine
Inhibition of PS II	C1	Triazines	Sebuthylazine
Inhibition of PS II	C1	Triazines	Secbumeton
Inhibition of PS II	C1	Triazines	Simetryne
Inhibition of PS II	C1	Triazines	Simazine
Inhibition of PS II	C1	Triazines	Terbumeton
Inhibition of PS II	C1	Triazines	Terbuthylazine
Inhibition of PS II	C1	Triazines	Terbutryne
Inhibition of PS II	C1	Triazines	Trietazine
Inhibition of PS II	C1	Triazolinone	Amicarbazone
Inhibition of PS II	C1	Triazinones	Ethiozin
Inhibition of PS II	C1	Triazinones	Hexazinone
Inhibition of PS II	C1	Triazinones	Isomethiozin
Inhibition of PS II	C1	Triazinones	Metamitron
Inhibition of PS II	C1	Triazinones	Metribuzin
Inhibition of PS II	C1	Uracils	Bromacil
Inhibition of PS II	C1	Uracils	Isocil
Inhibition of PS II	C1	Uracils	Lenacil
Inhibition of PS II	C1	Uracils	Terbacil

Inhibition of PS II	C1	Phenlcarbamates	Chlorprocarb
Inhibition of PS II	C1	Phenlcarbamates	Desmedipham
Inhibition of PS II	C1	Phenlcarbamates	Phenisopham
Inhibition of PS II	C1	Phenlcarbamates	Phenmedipham
Inhibition of PS II	C1	Pyridazinone	Chloridazon (=pyrazon)
Inhibition of PS II	C1	Pyridazinone	Brompyrazon
Inhibition of PS II	C2	Ureas	Benzthiazuron
Inhibition of PS II	C2	Ureas	Bromuron
Inhibition of PS II	C2	Ureas	Buturon
Inhibition of PS II	C2	Ureas	Chlorbromuron
Inhibition of PS II	C2	Ureas	Chlorotoluron
Inhibition of PS II	C2	Ureas	Chloroxuron
Inhibition of PS II	C2	Ureas	Difenoxuron
Inhibition of PS II	C2	Ureas	Dimefuron
Inhibition of PS II	C2	Ureas	Diuron
Inhibition of PS II	C2	Ureas	Ethidimuron
Inhibition of PS II	C2	Ureas	Fenuron
Inhibition of PS II	C2	Ureas	Fluometuron
Inhibition of PS II	C2	Ureas	Fluothiuron
Inhibition of PS II	C2	Ureas	Isoproturon
Inhibition of PS II	C2	Ureas	Isouron
Inhibition of PS II	C2	Ureas	Linuron
Inhibition of PS II	C2	Ureas	Metobenzuron
Inhibition of PS II	C2	Ureas	Metobromuron
Inhibition of PS II	C2	Ureas	Methabenzthiazuron
Inhibition of PS II	C2	Ureas	Metoxuron
Inhibition of PS II	C2	Ureas	Monolinuron
Inhibition of PS II	C2	Ureas	Monuron
Inhibition of PS II	C2	Ureas	Neburon
Inhibition of PS II	C2	Ureas	Parafluron
Inhibition of PS II	C2	Ureas	Siduron
Inhibition of PS II	C2	Ureas	Tebuthiuron
Inhibition of PS II	C2	Ureas	Thiazafluron
Inhibition of PS II	C2	Amides	Chloranocryl=dicryl
Inhibition of PS II	C2	Amides	Pentanochlor
Inhibition of PS II	C2	Amides	Propanil
Inhibition of PS II	C3	Nitriles	Bromofenoxim
Inhibition of PS II	C3	Nitriles	Bromoxynil
Inhibition of PS II	C3	Nitriles	Ioxynil
Inhibition of PS II	C3	Phenyl-pyridazines	Pyridate
Inhibition of PS II	C3	Benzothiadiazinone	Bentazon
PS I electron diversion	D	Bipyridyliums	Cyperquat
PS I electron diversion	D	Bipyridyliums	Diquat
PS I electron diversion	D	Bipyridyliums	Morfamquat
PS I electron diversion	D	Bipyridyliums	Paraquat

Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Lactofen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Acifluorfen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Bifenox
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Chlornitrofen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Fomesafen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Fluorodifen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Fluoroglycofen-ethyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Fluoronitrofen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Nitrofen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Oxyfluorfen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Diphenyl ethers	Chlomethoxyfen
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Phenylpyrazoles	Pyraflufen-ethyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Oxadiazoles	Oxadiargyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Oxadiazoles	Oxadiazon
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Triazolinones	Azafenidin
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Triazolinones	Carfentrazone-ethyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Triazolinones	Sulfentrazone
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Thiadiazoles	Fluthiacet-methyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Pyrimidinediones	Butafenacil
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Pyrimidinediones	Saflufenacil
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	Oxazolidinediones	Pentoxazone
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	N-Phenyl-phthalimides	Chlorphthalim
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	N-Phenyl-phthalimides	Cinidon-ethyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	N-Phenyl-phthalimides	Flumiclorac-pentyl
Inhibition of protoporphyrinogfen oxygenase (PPO)	E	N-Phenyl-phthalimides	Flumioxazin
Inhibition of protoporphyrinogfen oxygenase	E	N-Phenyl-phthalimides	Flumipropyn

(PPO)			
Inhibition of phytoene desaturase (PDS)	F1	Other	Beflubutamid
Inhibition of phytoene desaturase (PDS)	F1	pyridine carboxamides	Diflufenican
Inhibition of phytoene desaturase (PDS)	F1	pyridine carboxamides	Picolinafen
Inhibition of phytoene desaturase (PDS)	F1	Other	Flurochloridone
Inhibition of phytoene desaturase (PDS)	F1	Pyridazinone	Norflurazon
Inhibition of phytoene desaturase (PDS)	F1	Other	Fluridone
Inhibition of phytoene desaturase (PDS)	F1	Other	Flurtamone
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Triketones	Mesotrione
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Triketones	Sulcotrione
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Triketones	Tembotrione
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Triketones	Tefuryltrione
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Triketones	Bicyclopyrone
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Triketones	Fenquinotrione
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Other	Benzobicyclon
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Pyrazoles	Benzofenap
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Pyrazoles	Pyrasulfotole
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Pyrazoles	Topramezone
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Pyrazoles	Pyrazolynate
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Pyrazoles	Pyrazoxyfen
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Pyrazoles	Tolpyralate
Inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)	F2	Isoxazoles	Isoxaflutole
Inhibition of 1-deoxy-D-xylose 5-phosphate synthase (DOXP)	F4	None	Clomazone
Inhibition of EPSP	G	Glycine	Glyphosate
Inhibition of glutamine synthase	H	Phosphinic acids	Glufosinate-ammonium
Inhibition of glutamine synthase	H	Phosphinic acids	Bialaphos/bilanafos
Inhibition of dihydropteroate synthase (DHP)	I	Carbamate	Asulam
Inhibition of microtubule assembly	K1	Dinitroanilines	Benefin=benfluralin
Inhibition of microtubule assembly	K1	Dinitroanilines	Butralin
Inhibition of microtubule assembly	K1	Dinitroanilines	Dinitramine
Inhibition of microtubule assembly	K1	Dinitroanilines	Ethalfluralin
Inhibition of microtubule assembly	K1	Dinitroanilines	Fluchloralin
Inhibition of microtubule assembly	K1	Dinitroanilines	Isopropalin
Inhibition of microtubule assembly	K1	Dinitroanilines	Nitralin

Inhibition of microtubule assembly	K1	Dinitroanilines	Prodiamine
Inhibition of microtubule assembly	K1	Dinitroanilines	Profluralin
Inhibition of microtubule assembly	K1	Dinitroanilines	Oryzalin
Inhibition of microtubule assembly	K1	Dinitroanilines	Pendimethalin
Inhibition of microtubule assembly	K1	Dinitroanilines	Trifluralin
Inhibition of microtubule assembly	K1	Pyridines	Dithiopyr
Inhibition of microtubule assembly	K1	Pyridines	Thiazopyr
Inhibition of microtubule assembly	K1	Phosphoroamidates	Butamifos
Inhibition of microtubule assembly	K1	Phosphoroamidates	DMPA
Inhibition of microtubule assembly	K1	Benzoic acid	Chlorthal-dimethyl=DCPA
Inhibition of microtubule assembly	K1	Benzamides	Propyzamide=pronamide
Inhibition of microtubule organization	K2	Carbamates	Barban
Inhibition of microtubule organization	K2	Carbamates	Carbetamide
Inhibition of microtubule organization	K2	Carbamates	Chlorbufam
Inhibition of microtubule organization	K2	Carbamates	Chlorpropham
Inhibition of microtubule organization	K2	Carbamates	Propham
Inhibition of microtubule organization	K2	Carbamates	Swep
Inhibition of cellulose synthesis (CBI)	L	Triazolocarboxamide	Flupoxam
Inhibition of cellulose synthesis (CBI)	L	Benzamides	Isoxaben
Inhibition of cellulose synthesis (CBI)	L	Alkylazines	Triaziflam
Inhibition of cellulose synthesis (CBI)	L	Alkylazines	Indaziflam
Inhibition of cellulose synthesis (CBI)	L	Nitriles	Dichlobenil
Inhibition of cellulose synthesis (CBI)	L	Nitriles	Chlorthiamid
Uncoupler (membrane disruption)	M	Dinitrophenols	Dinosam
Uncoupler (membrane disruption)	M	Dinitrophenols	Dinoseb
Uncoupler (membrane disruption)	M	Dinitrophenols	Dinitramine
Uncoupler (membrane disruption)	M	Dinitrophenols	Dinoterb
Uncoupler (membrane disruption)	M	Dinitrophenols	Etinofen
Uncoupler (membrane disruption)	M	Dinitrophenols	Medinoterb
Inhibition of cell division (VLCFA)	K3	Other	Cafenstrole
Inhibition of cell division (VLCFA)	K3	tetrazolinone	Fentrazamide
Inhibition of cell division (VLCFA)	K3	tetrazolinone	Ipfencarbazone
Inhibition of cell division (VLCFA)	K3	Other	Anilofos
Inhibition of cell division (VLCFA)	K3	Other	Piperophos
Inhibition of cell division (VLCFA)	K3	Others	Pyroxasulfone
Inhibition of cell division (VLCFA)	K3	Others	Fenoxasulfone
Inhibition of cell division (VLCFA)	K3	None/other	Indanofan
Inhibition of cell division (VLCFA)	Z?	None/other	Tridiphane
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Acetochlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Alachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Allidochlor=CDAA
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Dimethachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Butenachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Delachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Diethatyl-ethyl

Inhibition of cell division (VLCFA)	K3	Acetamides	Diphenamid
Inhibition of cell division (VLCFA)	K3	Acetamides	Naproanilide
Inhibition of cell division (VLCFA)	K3	Acetamides	Napropamide
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Dimethenamid
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Metazachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Metolachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Pethoxamid
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Pretilachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Propachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Propisochlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Prynachlor
Inhibition of cell division (VLCFA)	K3	Chloroacetamides	Thenylchlor
Inhibition of cell division (VLCFA)	K3	Oxyacetamides	Mefenacet
Inhibition of cell division (VLCFA)	K3	Oxyacetamides	Flufenacet
Inhibition of cell division (VLCFA)	N	Thiocarbamates	Butylate
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Cycloate
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Dimepiperate
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	EPTC
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Esprocarb
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Molinate
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Orbencarb
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Pebulate
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Prosulfocarb
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Thiobencarb (=Benthiocarb)
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Tiocarbazil
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Tri-allate
Lipid synthesis inhibition (VLCFA)	N	Thiocarbamates	Vernolate
Lipid synthesis inhibition (VLCFA)	N	Benzofurans	Benfuresate
Lipid synthesis inhibition (VLCFA)	N	Phosphorodithioate	Bensulide
Lipid synthesis inhibition (VLCFA)	N	Benzofurans	Ethofumesate
Synthetic Auxin	O	Pyridine-carboxylates	Picloram
Synthetic Auxin	O	Pyridine-carboxylates	Clopyralid
Synthetic Auxin	O	Pyridine-carboxylates	Aminopyralid
Synthetic Auxin	O	None	Halauxifen
Synthetic Auxin	O	None	Florpyrauxifen
Synthetic Auxin	O	Pyridine-carboxylates	Triclopyr
Synthetic Auxin	O	Pyridine-carboxylates	Fluroxypyr
Synthetic Auxin	O	Phenoxy-carboxylates	2,4,5-T
Synthetic Auxin	O	Phenoxy-carboxylates	2,4-D
Synthetic Auxin	O	Phenoxy-carboxylates	2,4-DB
Synthetic Auxin	O	Phenoxy-carboxylates	Clomeprop
Synthetic Auxin	O	Phenoxy-carboxylates	Dichlorprop
Synthetic Auxin	O	Phenoxy-carboxylates	Fenoprop
Synthetic Auxin	O	Phenoxy-carboxylates	Mecoprop
Synthetic Auxin	O	Phenoxy-carboxylates	MCPA

Synthetic Auxin	O	Phenoxy-carboxylates	MCPB
Synthetic Auxin	O	Benzoates	Dicamba
Synthetic Auxin	O	Benzoates	Chloramben
Synthetic Auxin	O	Benzoates	TBA
Synthetic Auxin	O	Quinoline-carboxylates	Quinclorac
Synthetic Auxin	O	Quinoline-carboxylates	Quinmerac
Synthetic Auxin	O	None	Aminocyclopyrachlor
Synthetic Auxin	O	Other	Benazolin-ethyl
Auxin transport inhibition (ATI)	P	Phthalamate	Naptalam
Auxin transport inhibition (ATI)	P	Semicarbazone	Diflufenzopyr-sodium
Unknown mode of action	Z	None	Cinmethylin
Unknown mode of action	Z		Bromobutide
Unknown mode of action	Z		Cumyluron
Unknown mode of action	Z		Difenzoquat
Unknown mode of action	Z		DSMA
Unknown mode of action	Z		Dymron=Daimuron
Unknown mode of action	Z		Etobenzanid
Unknown mode of action	Z		Fosamine
Unknown mode of action	Z		Methyldymron
Unknown mode of action	Z		Monalide
Unknown mode of action	Z		MSMA
Unknown mode of action	Z		Oleic acid
Unknown mode of action	Z		Oxaziclomefone
Unknown mode of action	Z		Pelargonic acid
Unknown mode of action	Z		Pyributicarb
Unknown mode of action	Z		Quinoclamine
Unknown mode of action	Z	Benzamide (not really)	Tebutam

Table 6: Duplicates written out in a pdf format using KNIME

This table contains the duplicates from the first dataset. The workflow in chapter 4.2.1.3 extracts the data and has written it into a pdf table. The 26 2D structures have been extracted and only the LS-Core serves as identification. The table includes substituted and non-substituted triketones which have been tested on ARBTH and ZEAMX.

The substances mentioned in chapter 5.1.1 have been retested. The results of the retesting are explained in chapter 5.9.

"LS-core"	"LTSEnzymeSource"	"LTSValueType"	"LTSValue"	"LTSTimestamp"
295179	Arabidopsis thaliana	IC50	6.19E-9	2011-08-31T04:00
295179	Arabidopsis thaliana	IC50	6.19E-9	2011-08-31T04:00
295179	Zea mays	EC50	4.3E-7	1994-11-22T00:00
295179	Zea mays	EC50	4.3E-7	1994-11-22T00:00
297300	Zea mays	EC50	2.6E-6	1994-11-23T00:00
297300	Arabidopsis thaliana	EC50	1.13E-7	1997-03-10T00:00
298647	Arabidopsis thaliana	I	11.0	1997-03-10T00:00
298647	Zea mays	I	31.0	1994-11-24T00:00
321411	Arabidopsis thaliana	I	38.0	1997-03-10T00:00
321411	Zea mays	EC50	5.0E-5	1994-09-22T00:00
321416	Arabidopsis thaliana	EC50	5.32E-8	2009-10-19T00:00
321416	Arabidopsis thaliana	EC50	5.32E-8	2009-10-19T00:00
321416	Zea mays	EC50	5.3E-8	1994-10-24T01:00
321416	Zea mays	EC50	5.3E-8	1994-10-24T01:00
328115	Arabidopsis thaliana	EC50	1.1E-7	1997-03-10T00:00
328115	Zea mays	EC50	1.8E-6	1995-04-19T00:00
329773	Arabidopsis thaliana	EC50	7.57E-8	2008-01-30T00:00
329773	Zea mays	EC50	1.6E-8	1995-05-29T00:00
332545	Zea mays	EC50	1.1E-6	1995-07-31T00:00
332545	Arabidopsis thaliana	EC50	6.53E-7	1997-03-10T00:00
337128	Arabidopsis thaliana	EC50	1.81E-7	1997-03-10T00:00
337128	Zea mays	EC50	7.7E-6	1995-12-20T00:00
337131	Arabidopsis thaliana	EC50	1.5E-6	2001-06-18T00:00
337131	Arabidopsis thaliana	I	113.0	2001-06-18T00:00
337131	Arabidopsis thaliana	EC50	1.51E-6	1997-03-10T00:00
337131	Zea mays	EC50	6.7E-7	1995-12-20T00:00
337133	Zea mays	EC50	1.2E-8	1995-12-20T00:00
337133	Arabidopsis thaliana	EC50	1.08E-7	1997-03-10T00:00
337134	Arabidopsis thaliana	EC50	1.07E-7	1997-03-10T00:00
337134	Zea mays	EC50	5.6E-9	1995-12-20T00:00
337135	Zea mays	EC50	5.3E-6	1995-12-21T00:00
337135	Arabidopsis thaliana	EC50	5.3E-6	2001-06-18T00:00
337135	Arabidopsis thaliana	I	101.0	2001-06-18T00:00
337139	Arabidopsis thaliana	EC50	8.12E-8	1997-03-10T00:00
337139	Zea mays	EC50	3.3E-8	1995-12-21T00:00
340241	Zea mays	EC50	9.1E-7	1996-03-19T00:00
340241	Arabidopsis thaliana	EC50	9.1E-7	2001-06-18T00:00
340241	Arabidopsis thaliana	I	108.0	2001-06-18T00:00
340242	Zea mays	I	45.0	1996-03-28T00:00
340242	Arabidopsis thaliana	I	68.0	2001-06-18T00:00
341393	Zea mays	I	32.0	1996-04-30T00:00
341393	Arabidopsis thaliana	I	28.0	1997-03-10T00:00
341402	Arabidopsis thaliana	EC50	1.1E-6	2001-06-18T00:00
341402	Arabidopsis thaliana	I	82.0	2001-06-18T00:00
341402	Zea mays	EC50	1.1E-6	1996-05-28T00:00
341403	Arabidopsis thaliana	EC50	3.7E-6	2001-06-18T00:00
341403	Arabidopsis thaliana	I	88.0	2001-06-18T00:00
341403	Zea mays	EC50	3.7E-6	1996-05-28T00:00
342405	Arabidopsis thaliana	EC50	7.81E-8	1997-03-10T00:00
342405	Zea mays	EC50	1.3E-7	1996-05-06T00:00
343069	Arabidopsis thaliana	EC50	2.0E-5	2001-08-20T00:00
343069	Arabidopsis thaliana	I	90.0	2001-08-20T00:00
343069	Zea mays	I	43.0	1996-05-23T00:00
343073	Arabidopsis thaliana	EC50	3.0E-6	2001-08-20T00:00
343073	Arabidopsis thaliana	I	106.0	2001-08-20T00:00
343073	Zea mays	EC50	3.0E-6	1996-05-23T00:00
344835	Arabidopsis thaliana	EC50	1.56E-7	1997-03-10T00:00
344835	Zea mays	EC50	1.0E-7	1996-07-02T00:00
346517	Zea mays	EC50	1.2E-7	1996-07-18T00:00
346517	Arabidopsis thaliana	EC50	2.57E-7	2008-01-31T00:00
382565	Arabidopsis thaliana	EC50	7.78E-8	1998-09-09T00:00
382565	Zea mays	EC50	1.06E-6	1998-03-03T00:00
414603	Arabidopsis thaliana	EC50	4.53E-8	1999-05-25T00:00
414603	Zea mays	I	95.0	1999-05-17T00:00