

Controlling Hyperglycemia: Discovery of Novel Small α -Amylase Inhibitors Using Structure-Based Virtual Screening

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im

Fachbereich Biologie, Chemie, Pharmazie

der Freien Universität Berlin

vorgelegt von

Jamil Al-Asri

aus Sana'a, Jemen

2014

Die vorliegende Arbeit wurde von März 2011 bis August 2014 unter der Leitung von Prof. Dr. Gerhard Wolber am Institut für Pharmazie der Freien Universität Berlin angefertigt.

1. Gutachter: Prof. Dr. Gerhard Wolber
2. Gutachter: Prof. Dr. Matthias F. Melzig

Disputation am 24. November 2014

Dedication

To my parents who keep unconditional supplication (*Dua*) for me. They did not only raise and nurture me but also taxed themselves dearly over the years for my education and intellectual development.

To my wife and my children (Amjd, Molhm, and Rfeef) who have been a constant source of motivation, support, and encouragement during the challenges of postgraduate studies and life.

To my brothers (Mahmoud, Yahya, and Ghailaan) who supported me in different aspects during my postgraduate studies.

Acknowledgements

I would like to express my deep gratitude to Prof. Dr. Gerhard Wolber, my research supervisor, for giving me a chance to pursue my PhD in his computer-aided drug design Lab in the Pharmaceutical and Medicinal Chemistry Department at the Institute of Pharmacy, Freie Universität Berlin. His guidance, encouragement and critiques of this research work were valuable and appreciated.

I am particularly grateful for biological testing carried out by research partner Dr. Gyöngyi Gyémánt from the University of Debrecen, Hungary. Many thanks for her valuable collaboration.

I would also like to thank Prof. Dr. Matthias F. Melzig for allowing me to perform a part of my biological experiments in the Pharmaceutical Biology Department, Freie Universität Berlin. Assistance provided by Mrs. Cornelia Görick during my work for biological experiments was greatly appreciated.

I would like to acknowledge my special thanks to Dr. Jérémie Mortier for his valuable and constructive suggestions and corrections on my thesis. His willingness to give me his time so kindly has been very much appreciated.

I would like to express my sincere thanks to Dr. Andrej Perdih, a former member of our group for his valuable advice and generously granting us some commercial compounds that positively contribute in this research.

I would like to thank all my colleagues in the computational Lab for their help and grateful assistance.

I am particularly thankful for the European program Erasmus Mundus for awarding me a doctoral scholarship for a period of 34 months.

Table of Contents

Abstract	viii
Zusammenfassung	ix
1 Introduction.....	1
1.1 Classification of α -amylases.....	1
1.2 Enzyme structure and activation mechanism	3
1.3 Diseases related to amylase and its applications	7
1.3.1 Dental caries	8
1.3.2 Obesity and type II diabetes.....	8
1.3.3 Industrial applications.....	9
1.4 α -Amylase inhibitors.....	9
1.4.1 Peptide-based α -amylase inhibitors	10
1.4.2 Carbohydrate-based α -amylase inhibitors	10
1.4.3 Polyphenols	13
1.4.4 2-aminobenzothiazole	14
1.4.5 Terpenoids	15
2 Objectives.....	17
3 Methods	19
3.1 Molecular modeling and virtual screening in drug discovery	19
3.1.1 Pharmacophore modeling	21
3.1.2 Molecular docking	22
3.1.3 Similarity search and clustering.....	23
3.1.4 Shape-based virtual screening.....	24
3.1.5 Statistical validation of model predictivity.....	25
3.2 Biochemical assay	27
4 Results	29
4.1 Structural analysis of different α -amylase isoforms and binding sites	29
4.1.1 Structural investigation of α -amylase.....	30
4.2 Docking using α -amylase inhibitors with previously unknown binding mode.....	33
4.2.1 Cyanidin derivatives	33
4.2.2 Steviol compounds.....	36
4.2.3 Verbascoside	39
4.3 Structure-Based virtual screening for α -amylase inhibitors	43
4.3.1 Pharmacophore-based virtual screening: Initial screening round.....	43

4.3.1.1	Development of 3D pharmacophore models	43
4.3.1.2	Virtual screening metrics for the developed pharmacophore model.....	47
4.3.1.3	Overall compound selection work-flow	48
4.3.2	Biological Testing	50
4.3.3	Attempt to rationalize differences in IC ₅₀ values	53
4.3.4	Docking studies with discovered α -amylase inhibitors	56
4.3.5	Similarity search for the best competitive inhibitor in the initial screening	62
4.3.6	Biological assay and SAR for analogues of the best competitive inhibitor	62
4.3.7	Docking studies of analogues of the best novel competitive inhibitor.....	66
4.3.8	Refined pharmacophore-based virtual screening	68
4.3.8.1	Development of 3D pharmacophores	68
4.3.8.2	Protocol of the refined virtual screening	74
4.3.9	Biological testing of hits selected by the refined virtual screening.....	76
4.3.10	Docking of novel inhibitors identified by the refined virtual screening	78
5	Discussion.....	86
5.1	Natural α -amylase inhibitors with unknown binding mode	87
5.2	Virtual screening	88
5.2.1	Initial virtual screening.....	88
5.2.2	Refined virtual screening.....	90
5.2.3	Evaluation of the performed virtual screening approaches	90
6	Conclusion and Prospective	96
7	Experimental part	98
7.1	Binding mode elucidation for inhibitors identified by collaboration partners	98
7.2	Structure-based virtual screening.....	98
7.2.1	Commercial chemical databases.....	98
7.2.2	Validation of developed pharmacophores.....	99
7.2.3	Target preparation for pharmacophore creation	101
7.2.4	Pharmacophore creation and virtual screening.....	101
7.3	Biological testing	103
7.3.1	Biological testing using the substrate <i>p</i> -NPG5 for PPA: First assay.....	103
7.3.2	Biological testing with GalG2CNP substrate for HSA: Second assay	105
7.4	Modeling differences in IC ₅₀ values in the initial virtual screening	105
7.5	Molecular Docking.....	106
7.5.1	Target preparation for molecular docking	106
7.5.2	Preparation of compounds for docking	106

7.5.3	Docking studies for inhibitors discovered by the initial virtual screening	106
7.5.4	Docking studies for inhibitors discovered by the refined virtual screening.....	107
7.6	Similarity search for the best competitive inhibitor in the initial screening	107
8	Summary.....	108
9	References.....	110
	Curriculum vitae	121
	Publications	123
10	Appendix.....	125
10.1	Macromolecules of pancreatic and salivary α -amylases (human and porcine)	125
10.2	Structural analysis.....	126
10.3	Biologically known active α -amylase inhibitors	127
10.4	Biologically known inactive α -amylase inhibitors	132
10.5	Steric overlap of substrates with tested compounds.....	141
10.6	Validation of docking experiments for α -amylase.....	143
10.7	Docking of analogues of the best competitive inhibitor in the initial screening	145
10.8	Validation of models developed in the refined virtual screening.....	146
10.9	Docking of the best competitive inhibitor in the refined virtual screening	148
10.10	Quality control performed by vendors for selected compounds	149
10.11	Commercial compounds selected by virtual screening approaches.....	151

Abstract

The enzyme α -amylase is secreted from salivary and pancreatic glands and hydrolyzes α -D-(1,4)-glycosidic linkage in carbohydrates such as starch. Its modulation represents the possibility to control postprandial hyperglycemia and is therefore considered an attractive strategy for the prevention or treatment of obesity or type II diabetes. However, only few drug-like α -amylase inhibitors without carbohydrate moieties exist and only sparse information about their mechanistic properties is available. The aim of this study was to discover novel small non-sugar α -amylase inhibitors and their binding modes using rational *in silico* methodology and biological experiments. To reach this goal, mechanistic 3D pharmacophore models were carefully developed and applied to several virtual screening experiments. Using this approach, about two million compounds could be computationally screened for potential inhibition of α -amylase resulting in the selection of 33 compounds in different virtual screening rounds, which were all biologically tested.

Our initial virtual screening resulted in the discovery of six inhibitors out of fourteen biologically tested compounds (IC_{50} range: 86 - 300 μ M). A subsequent analogue search using the most active and competitive newly identified inhibitor yielded twelve further compounds, out of which six showed slightly better inhibition up to an IC_{50} of 50 μ M. A final, refined virtual screening led to the identification of four improved binders out of seven tested molecules with an IC_{50} of up to 17 μ M. Overall, 50 % of the computationally suggested and selected virtual hits could be experimentally confirmed. Due to their small size, all identified binders show better ligand efficiency values than previously known inhibitors. Hence, these structures are ideal starting points for the design of novel α -amylase inhibitors. The discovered compounds have never been reported as α -amylase inhibitors before and represent novel scaffolds for this specific class of biological activity.

Zusammenfassung

Das Enzym α -Amylase wird in den Speicheldrüsen, sowie im Pankreas sezerniert und hydrolysiert die $\alpha(1-4)$ -Glykosidbindung in Kohlenhydraten. Die Modulation dieses Enzyms stellt eine Möglichkeit dar, postprandiale Hyperglykämien zu kontrollieren und wird deshalb als attraktive Strategie angesehen, Fettleibigkeit oder Typ II Diabetes vorzubeugen bzw. zu behandeln. Es gibt allerdings nur wenige verfügbare α -Amylase Inhibitoren, die nicht aus Kohlenhydratbausteinen oder deren Analoga aufgebaut sind. Desweiteren ist nur wenig über deren Bindungsmodus bekannt. Das Ziel dieser Arbeit war daher die Auffindung neuer α -Amylase-Inhibitoren ohne Kohlenhydratgerüst, die Analyse der entsprechenden Bindemodi durch *in silico* Modellierung und deren experimentelle Validierung. Zu diesem Zweck wurden mechanistische 3D Pharmakophormodelle entwickelt und als Basis für mehrere virtuelle Screening-Experimente verwendet. Mit diesem Ansatz konnten ca. zwei Millionen Moleküle auf potenzielle α -Amylase-Inhibition untersucht werden. 33 Verbindungen wurden in verschiedenen aufeinander folgenden Screening-Experimenten ausgewählt und anschließend biologisch getestet.

Das erste virtuelle Screening lieferte vierzehn ausgewählte Verbindungen, von denen sechs biologische Aktivität zeigten (IC_{50} : 86 - 300 μ M). Eine darauffolgende Suche nach Analoga des aktivsten kompetitiven Liganden lieferte zwölf weitere Verbindungen, von denen sechs bis zu einer mittleren Hemmkonzentration von 50 μ M biologisch aktiv sind. Ein weiteres, virtuelles Screening-Experiment führte zur Identifikation von vier Inhibitoren mit einer Aktivität von bis zu 17 μ M. Insgesamt zeigten 50 % der *in silico* ausgewählten Verbindungen biologische Aktivität und konnten somit erfolgreich experimentell validiert werden. Aufgrund ihres geringen Molekulargewichts, zeichnen sich alle neu entdeckten Inhibitoren durch eine höhere *ligand efficiency* aus als bisher bekannte α -Amylase-Hemmer. Daher eignen sich diese Verbindungen als Leitstrukturen für die Entwicklung neuer α -Amylase-Inhibitoren. Alle gefundenen Liganden repräsentieren neueartige Grundstrukturen für die Modulation der α -Amylase-Aktivität.

1 Introduction

In 1812 the German scientist Gottlieb Kirchoff discovered in St. Petersburg that starch can be converted into sugar by heating it in boiling water with sulfuric acid [1]. In 1831, α -amylase was the first enzyme catalyzing this reaction in saliva by the German scientist Erhard Friedrich Leuchs before it was determined in serum by Magendie in 1846, in pancreas by Alexander Jakulowitsch Danilewsky in 1862, and in urine by Cohnheim in 1863 [2, 3].

More advanced experiments at that time were performed by the French chemists Anselme Payen and Jean-François Persoz. They isolated an extract from germinating barley and used to convert gelatinized starch into sugars, mainly maltose. This extract was considered as an enzyme complex and termed “diastase”. Afterwards, a fermentation process called “Takadiastase” was developed by the Japanese scientist Takamine and is still used for the industrial production of fungal α -amylase [3, 4]. Since that time, α -amylase was thoroughly studied and different isoforms from different species have even been crystallized [5, 6].

The term “amylase” refers to an enzyme that hydrolyzes the O-glycosyl linkage in starch [7]. Starch composed of amylose (mainly α -1,4-glycosidic linkage) and amylopectin (α -1,4-linkage and branching α -1,6-linkage) is considered one of the main sources of energy in our nutrition because of the distinctive composition of polymers of glucose units. Hydrolysis of polysaccharides like starch into glucose is a crucial manifestation in metabolic diseases such as type II diabetes mellitus and obesity since it suddenly causes post-prandial hyperglycemia [8]. α -Amylase is one of the most attractive targets for the development of novel therapeutic agents and controlling the glucose level in the blood [9, 10]. A broad description of α -amylases will be given in chapter 1.1 while an overview of its structure and hydrolysis mechanism will be given in chapter 1.2. The role of α -amylase in metabolic diseases as well as its industrial applications will be summarized in chapter 1.3. In chapter 1.4, an overview of the so far α -amylase inhibitors described in literature will be introduced.

1.1 Classification of α -amylases

Since starch is the main source of glucose in human nutrition, a multitude of enzymes performing starch cleavage exists in living organisms. Starch-converting enzymes are mainly classified into four types [11]:

(i) Endoamylases, such as α -amylase (EC 3.2.1.1), which hydrolyze α -1,4-glycosidic linkage in the inner part of starch to produce oligomers. This enzyme is the focus of this work.

(ii) Exoamylases which mainly hydrolyze either α -1,4-glycosidic bond by β -amylase (EC 3.2.1.2) to produce β -maltose and β -dextrin or cleave both α -1,4- and α -1,6-glycosidic bonds to produce either β -glucose by glucoamylase (γ -amylase, EC 3.2.1.3) or α -glucose by α -glucosidase (EC 3.2.1.20), Figure 1.

(iii) Debranching enzymes, such as isoamylase (EC 3.2.1.68) and pullanase type I (EC 3.2.1.41) which hydrolyze α -1,6-glycosidic bonds in amylopectin to produce long polysaccharides.

(iv) Glycosyltransferases (E.C. 2.4) which cleave α -1,4-glycosidic bonds in a donor molecule and transfer the cleaved part into a glycosidic acceptor forming a new linkage as either α -1,4- or α -1,6-bond.

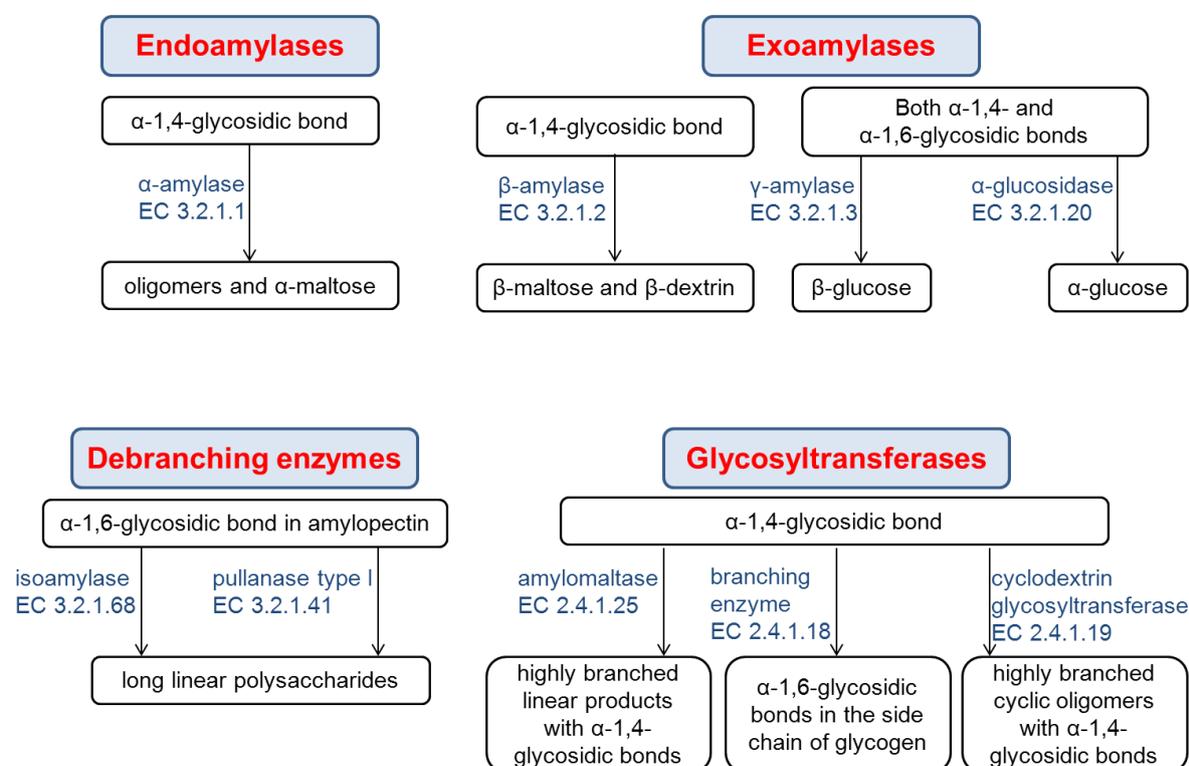


Figure 1: Starch-converting enzymes.

Recently, most of starch-converting enzymes were put into one group called α -amylase family [11-17], which belongs to glycosyl hydrolase group no. 13 (GH 13) according to the classification of glycosyl hydrolases by Hernissat (1991) that is based on sequence similarity. α -Amylases (1,4- α -D-glucan-4-glucanohydrolase, EC 3.2.1.1) are endoamylases that catalyze the hydrolysis of α -(1,4)-glycosidic linkages of polysaccharides such as starch and glycogen [3]. This family is characterized by four specific properties [11, 18]. These enzymes

(i) Hydrolyze α -glycosidic bonds to produce mono- or oligomers with preserved α -anomeric carbon conformation (C1 in the glucose unit) and create α -1,4-bonds (or α -1,6 bonds in case of transglycosylation), (ii) structurally consist of $(\alpha/\beta)_8$ TIM barrel structures containing the catalytic residues (one glutamic acid and two aspartic acids), (iii) contain highly conserved regions in the active site in β -strands (β 2: Asp, β 4: Arg, Asp, β 5: Glu, β 7: His, Asp) and (iv) provide an α -retaining double displacement mechanism, which is explained in the following chapter.

1.2 Enzyme structure and activation mechanism

α -Amylase is abundantly secreted from salivary glands and pancreas. In humans and other mammals, α -amylase is expressed by two genes: Gene AMY1 encodes salivary α -amylase and gene AMY2 synthesizes pancreatic α -amylase [19].

The primary sequence of human pancreatic α -amylase (HPA) and human salivary α -amylase (HSA) are very similar with 97% sequence identity and 92 % in the catalytic domains [20-22]. The overall differences between the sequences of HPA and HSA are 15 amino acids. Among them, two residues are in the active site region and could participate in substrate binding. These are Thr163 and Leu196 in HPA, which are substituted with Ser163 and Ile196 in HSA [21].

In contrast, there are 70 amino acid substitutions between HPA and porcine pancreatic α -amylase (PPA), which could exert functional differences between the two enzymes. In general, four regions of amino acid segments have significantly different conformations in HPA and PPA, two of them (segments 237-250 and 304-310) have an effect on substrate cleavage pattern [21].

Human pancreatic α -amylase (HPA) is a 56 kD protein composed of 496 amino acids as a single polypeptide chain folded into three domains, as illustrated in Figure 2. Domain A (residues 1-99 and 169-404) is the largest part and is composed of $(\alpha/\beta)_8$ -barrel structure where the barrel of eight β -strands are surrounded by eight α -helices [21, 23]. The active site located as a V-shaped cleft in the catalytic domain A binds to the substrate [24, 25]. Carboxylic amino acid residues Asp197, Glu233 and Asp300 are located in the active site and constitute the essential catalytic triad required for hydrolysis of the glycosidic bond [25]. A chloride ion is located close to the active site ($\sim 5 \text{ \AA}$) in between Arg195, Asn298 and Arg337. It works as an activator and modulates the optimum pH and maximal activity of α -amylase since it shields the positive charge of Arg337 from decreasing the pKa of the catalytic Glu233 [26, 27].

Domain B (residues 100-168) emerged from domain A is located between the third β -strand and the α -helix of the β -barrel of the catalytic domain and maintains the conformation of the enzyme. It contains a calcium ion, which stabilizes the structure of domain B. Domain C (residues 405-496) is an anti-parallel β -sheet of domain A and loosely associated to other domains [21]. It plays a role in folding and stability of the protein [23].

Binding subsites for glycosyl hydrolases like α -amylase were enumerated from negative numbers ($-n$) to positive numbers ($+n$) according to a proposal published by G. J. Davies *et al.* [28].

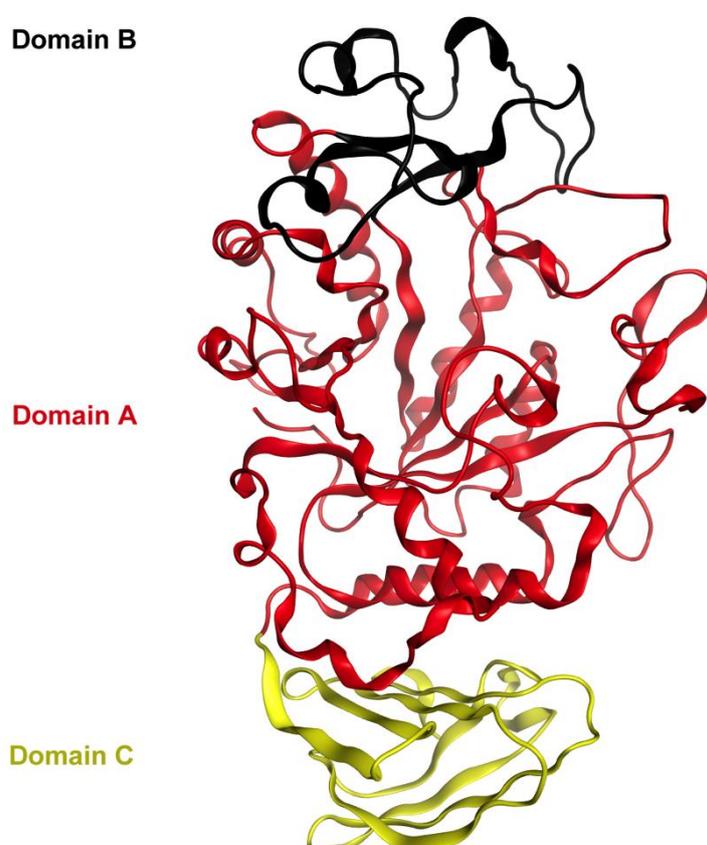


Figure 2: α -amylase (PDB entry 3OLE).

While glycone binding subsites are labeled with $-n$ on the left to accommodate the non-reducing end of a substrate, aglycone binding subsites are represented with $+n$ on the right to hold the reducing end. The site of cleavage of the scissile bond is between subsites -1 and $+1$ and every subsite is occupied with one glucose unit of the substrate [23].

Naturally, glycosidases, like α -amylase, catalyze the hydrolysis through a mechanism called double displacement mechanism [29] (Figure 3). Firstly, Glu233 acts as acid-catalyst and

protonates the glycosidic oxygen attached to the bond susceptible for cleavage (the oxygen that links the two glucose units at subsites -1, +1). Then, Asp197 acts as a nucleophile and attacks the anomeric carbon atom (C1 of the glucose at subsite -1) forming an oxocarbenium ion-like transition state, which is converted to a covalent glycosyl-enzyme intermediate. As a result, the glucose unit at subsite +1 leaves the active site with a proton as a glycosyl unit ($R'OH$). This process is called glycosylation of the enzyme (the first three steps are shown in Figure 3). Secondly, Glu233 acts as a base-catalyst by deprotonating a water molecule (in case of hydrolysis) or a new glucose unit (in case of condensation) to produce a nucleophilic hydroxyl group. This new nucleophile attacks the anomeric carbon atom of the previously formed glycosyl-enzyme intermediate to form again an oxocarbenium ion-like transition state. Then, the oxygen of this new nucleophile replaces the oxocarbenium bond (linking the glucose at subsite -1 and Asp197) forming a new hydroxyl group (in case of hydrolysis) or a new glycosidic bond (in case of condensation). Consequently, a glycosyl product with α -anomeric configuration is formed and leaves the binding site. This process is called deglycosylation of the enzyme (the next three steps are depicted in the Figure 3).

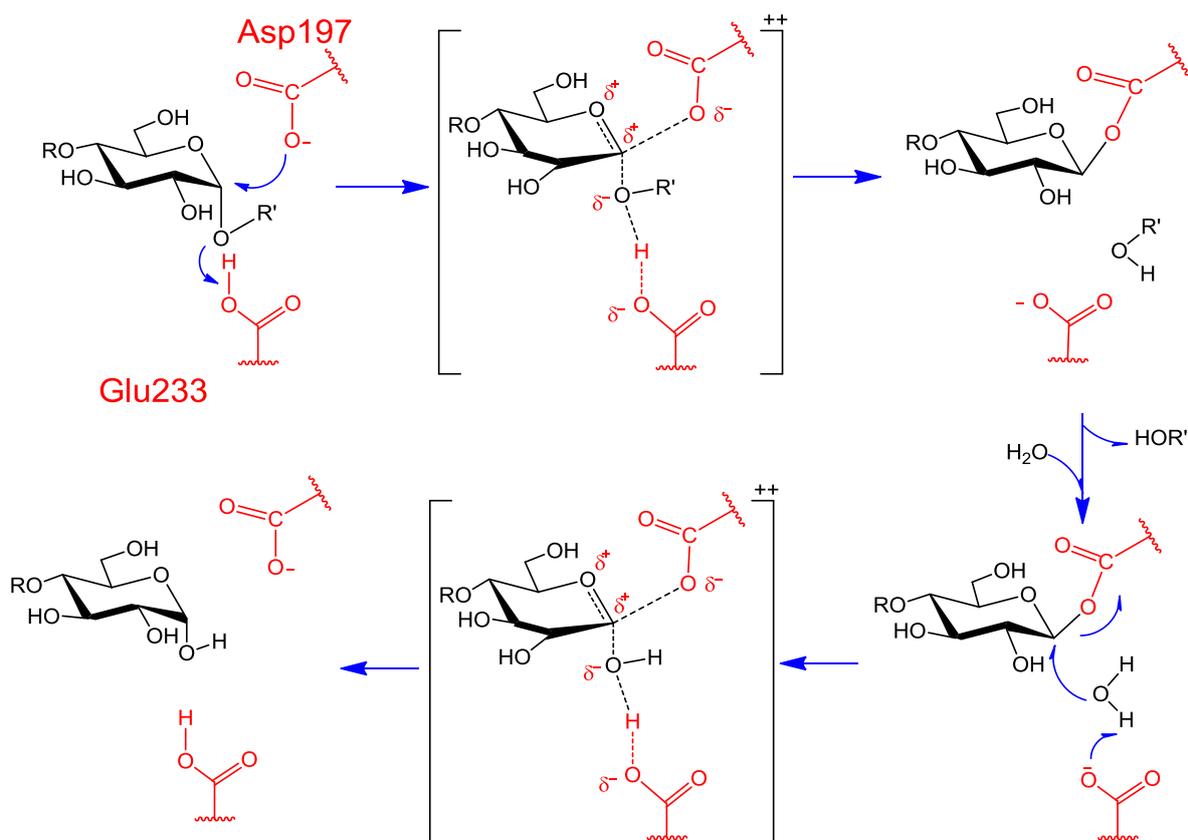


Figure 3: Catalytic mechanism of α -amylase.

For an optimal cleavage rate, at least two residues are required as leaving groups in the positive subsites [30]. Both processes occur through oxocarbenium ion-like transition state [3, 29, 31, 32]. The role of Asp300 is less definite, but it has been proposed to play role in stabilizing the conformation of the bound substrates and also in sustaining Glu233 in the right state of protonation for activity [32, 33].

The interaction of known biologically active α -amylase inhibitors were used to identify the important amino acids in the active site required for the activity based on their inhibitory potencies.

Mutation of catalytic residues (Asp197, Glu233, and Asp300) located in subsites -1 and +1 shows significant reduction in the enzyme activity; more than 10^5 -fold drop in case of Asp197 and up to 10^3 -fold drop with Glu233 or Asp300 substitutions; indicating their essential role in the hydrolysis process of α -amylase [34, 35].

Residues Gly306 and Glu240 at subsite +3 show modest effects in enzyme activity. This becomes clear for acarviosatin I03 ($K_i = 1.254 \mu\text{M}$), which occupies subsites -3 to +3 compared to pseudopentasaccharide acarbose ($K_i = 2.59 \mu\text{M}$) that occupies subsites -3 to +2 [36]. The neutral nature of subsite +3 in the electrostatic surface of HPA and its proximity to the solvent region are the cause of this modest effect on potency. Asp147 and the near Thr163 located at subsite -4 play important roles in case of long substrates spanning from subsites -4 to +3 [36].

Mutation of Arg195 causes 450-fold reduction in α -amylase activity. Arg195 directly interacts with the chloride ion that is essential for the enzymatic activity and forms direct H-bonds to catalytic residues Asp197 and Glu233. It has been reported that Arg195 along with His299 and Asp300 form H-bonds with the hydroxyl group of a sugar unit of a substrate that is subjected to nucleophilic attack by Asp197 in subsite -1 during the hydrolysis process [27]. This also points out the importance of His299, which has been reported with the same interaction with iso-acarbose [32]. During the binding of a ligand at subsites -1 and +1, the loop 304-310 anchored by Asp300 is re-oriented and allows the side chain of His305 to move inwards into the binding subsite -2 to be interacted [30, 32]. Trp58 positioned in the vicinity of subsite -2 shows critical importance for the activity and its mutation causes 150-fold drop in hydrolysis activity of HSA [37]. π -Stacking interactions between the bound glucose moieties and aromatic residues Tyr151 in subsite +2 and Trp59 and Tyr62 in subsite -2 have been reported to play a crucial role in enzymatic activity [38, 39].

Lipophilic residues Leu162, Leu165 and Ile235 in the active site form hydrophobic contacts with hydrophobic groups of the interacted inhibitors.

An overview of all interactions identified as crucial for an optimal α -amylase inhibition is given in Figure 4 that showing acarviosatin II03 bound to HPA (PDB: 3OLE).

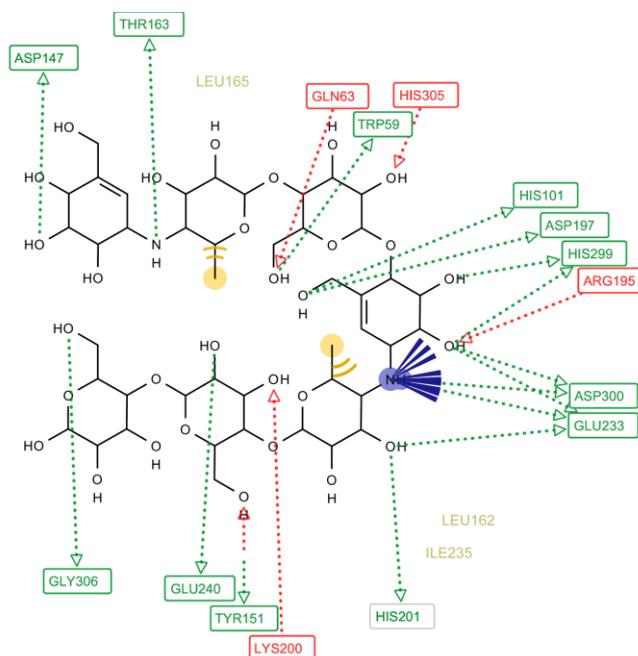


Figure 4: The final product of AI03 inhibitor (PDB: 3OLE). Green and red arrows represent H-bond donors and acceptors respectively, yellow spheres indicate hydrophobic contacts, and blue stars show positively ionizable interactions.

1.3 Diseases related to amylase and its applications

Metabolic diseases like obesity and type II diabetes are characterized by high levels of blood glucose [8]. Naturally, glycosidases, like α -amylase, catalyze the hydrolysis of polysaccharides like starch, resulting into sudden postprandial hyperglycemia. The total number of patients diagnosed with diabetes is predicted to increase from 171 million in 2000 to 366 million in 2030 and at least 300 million in the world are considered obese [40, 41]. These high numbers of people point out the importance of controlling these chronic disorders. One of the attractive targets for controlling the blood glucose level and the development of new therapeutic agents is α -amylase [9, 10].

Here, we briefly addressed the clinical manifestations and their consequences related to the hydrolysis process by α -amylase and its industrial applications.

1.3.1 Dental caries

The initiation of starch breakdown in the oral cavity by salivary α -amylase is not the only function of this enzyme, but it is also involved in oral diseases such as dental caries and periodontal diseases. It occurs when α -amylase is adsorbed to the tooth enamel [42, 43] and binds to oral *streptococci* species referred as α -amylase-binding *streptococci* (ABS) [44]. This binding; which occurs in sites other than the catalytic active site [45]; assists the colonization of these bacteria in the oral cavity [46, 47] where salivary α -amylase hydrolyses the dietary starch and produces oligomers as nutrients for these bacteria. Consequently, the production of lactic acid, which is a bacterial metabolite, is considered as a critical step in dissolving the tooth enamels and development of dental caries [45, 48, 49].

1.3.2 Obesity and type II diabetes

Obesity is defined as a condition of abnormal or excessive fat accumulation in adipose tissue to the extent that health may be impaired. Statistical figures in the United States propose that about 61% of type II diabetes and 17% of both coronary heart diseases (CHD) and hypertension can be attributed to obesity [50].

Therefore, obesity is accompanied with different medical consequences that seriously affect health. Some of these co-morbidities, in addition to type II diabetes and CHD, are dyslipidemia, cardiovascular diseases, sleep-disordered breathing and others [41, 51].

It is known that one of the main sources of energy in our diet is carbohydrates. The metabolic pathways of glucose in our body mainly includes: oxidation to produce energy, glycogen production in the liver and muscles, and lipid synthesis in the liver [52]. Excessive consumption of dietary carbohydrates is easily converted into fat [53] or even reduces the oxidation of fats [52] in the body and hence leads to fat accumulation. Therefore, carbohydrate is believed to manipulate substantial function in the development of obesity.

Starch as the main carbohydrate in human nutrition is degraded in different stages into glucose units to be absorbed [54]. α -Amylase secreted from salivary and pancreatic glands triggers the hydrolysis of starch into oligomers. These oligosaccharides pass through the mucus layer of the brush border membrane to be subjected for further digestion by α -glucosidase and produce glucose units. Excessive absorption of glucose increases the post-prandial blood glucose levels and lowers insulin resistance [55].

1.3.3 Industrial applications

In addition to its clinical importance, the whole amylase family has its own applications in other fields such as food, detergents, textiles, paper industry and biotechnology [56, 57].

Bio-ethanol production is generated from starch as a renewable source of fuel and works through fermentation by microorganisms. This process requires high temperature to solubilize starch granules. Hence, starch hydrolyzing enzymes can greatly reduce the cost of bio-ethanol production by using cold starch hydrolysis [58]. Also, amylases from *Bacillus* or *Aspergillus* are used in detergent industry to remove tough stains from starchy foods such as potatoes, gravies, custard, chocolate etc. The reason is that these enzymes can work at lower temperatures and alkaline pH which means higher stability under detergent conditions [59]. In textile industry, starch is used as strengthening agent during weaving process and sizing agent to yarn before fabric production to warrant fast and secure weaving process. Therefore, amylases from *Bacillus* strains remove the size without attacking the fibers. Concerning paper manufacturing, some commercial products of α -amylases are commercially available (Amizyme[®], Termamyl[®], BAN[®], and G9995[®]) and used for starch modification to make the surface of the paper smooth and strong to improve the writing quality of the paper [58]. One of the daily uses of α -amylases is its application in food industry such as baking process. The addition of this enzyme to the dough leads to improve the fermentation process and partially produces smaller dextrins and sugars to enhance the taste of the product such as bread [58]. Additionally, the shelf life of the baking products is improved when using α -amylase due to its anti-staling effect and it augments their softness.

1.4 α -Amylase inhibitors

As explained in the previous sections, α -amylase plays a role in postprandial hyperglycemia and hydrolyzing polysaccharides such as starch into oligomers. It is considered an attractive target for controlling metabolic diseases such as type II diabetes and obesity. Therefore, α -amylase inhibitors can regulate the breakdown of dietary starch into smaller sugars with the purpose of delaying glucose absorption to be in a rate the body can deal with and hence decreasing the post-prandial hyperglycemia [9, 60, 61]. This is valuable for patients who need lower amount of α -amylase to keep glucose levels in normal range.

While all known co-crystallized α -amylase inhibitors were peptide- or carbohydrate-based derivatives at the start of this work, limited drug-like inhibitors were reported before and during this work. Here, we are going to highlight classes that were identified so far as α -amylase inhibitors.

1.4.1 Peptide-based α -amylase inhibitors

Members of this class are large molecules with high molecular weight showing a strong inhibitory effect on α -amylase activity with inhibition constants (K_i) in the nanomolar range. As reported by Svensson *et al.* [62], seven types of proteinaceous α -amylase inhibitors are found in nature (in plants and *Streptomyces* species) and categorized based on similarities in sequence and 3D-structures. Tendamistat, one of these inhibitors, is co-crystallized with PPA (PDB code: 1BVN [63]) and composed of 75 amino acids. A total of 15 residues are involved in binding with a K_i of 9 nM [63]. This extreme specificity and binding affinity are due to the intricate interaction pattern into areas further away from the catalytic center [64]. However, peptide-based compounds are generally considered poor drug candidates because of their intrinsic physicochemical properties and pharmacokinetic profiles. The lower chemical stability of this class allows for rapid degradation by digestive system before they reach the required targets and hence limits their use by oral administration [65].

1.4.2 Carbohydrate-based α -amylase inhibitors

Most co-crystallized inhibitors are carbohydrate-derivatives such as acarbose (a well-known anti-diabetic drug) and belong to the trestatin family, which is a natural metabolite of *Streptomyces coelicoflavus* ZG0656 [66]. This family is characterized by the presence of a pseudotrisaccharide center called acarviosine-glucose illustrated in Figure 5.

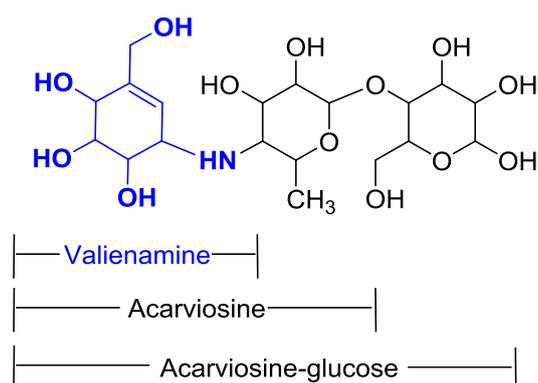


Figure 5: acarviosine-glucose center is composed of acarviosine (valienamine in blue color (left), 4,6-dideoxyglucose unit (middle)) and D-glucopyranose (right).

This center includes acarviosine unit linked by α -1,4-glycosidic bond to one glucose unit. Acarviosine itself is composed of an unsaturated cyclohexitol moiety (termed valienamine) attached to 4,6-dideoxy- α -D-glucose group by an N-glycosidic bond [67]. The reducing end

of this center is located in the D-glucose unit on the right side while the non-reducing end is present in the valienamine moiety on the left side. D-glucose units are attached to the reducing or non-reducing ends of that center and form what is called acarviosatin (A), followed by a Roman number and two digits. Roman numbers I, II, III, or IV stand for one, two, three or four acarviosine-glucose units. While the first digit that comes after the Roman number represents the number of glucose units at the non-reducing, the last digit represents the ones at the reducing ends [67], (Table 1).

Table 1: Acarviostatins with different numbers of sugar units and their inhibitory effect against α -amylase

inhibitor	K _i (μ M)	acarviosine- glucose units	glucose rings at		rings before binding	Ref.
			non-reducing end	reducing end		
Acarbose	2.593	1	0	1	4	[36]
AIII0(-1)	0.009	3*	0	0	8	[68]
AIII23	0.026	3	2	3	14	[68]
AI03	1.2549	1	0	3	6	[36]
AII03	0.0147	2	0	3	9	[36]
AIII03	0.0143	3	0	3	12	[36]
AIV03	0.0416	4	0	3	15	[36]
AII23	0.009	2	2	3	11	[67]
AII13	0.010	2	1	3	10	[67]

* It lacks one glucose unit at the reducing end.

One acarviostatins composed of three acarviosine-glucose centers, but lacking a glucose unit at the reducing end (called AIII0(-1)) was recently published in 2013 [68]. It comprises eight rings and it is the most potent acarviostatins inhibitor against α -amylase published so far compared to the ones with higher number of sugar rings [68, 69].

Some acarviostatins inhibitors such as AI03, AII03, AIII03, and AIV03 were co-crystallized with HPA and published in PDB with codes: 3OLD, 3OLE, 3OLG, and 3OLI, respectively [36]. The mechanism of inhibition of these molecules is particular in terms of final products observed inside the pocket. Indeed, hydrolysis, condensation and transglycosylation of acarviostatins inhibitors by HPA show rearranged binding products different from the initial structures used in biological testing. Interestingly, co-crystallized AII03, AIII03, and AIV03 inside the active site consist of seven carbohydrate units [36] (Figure 6).

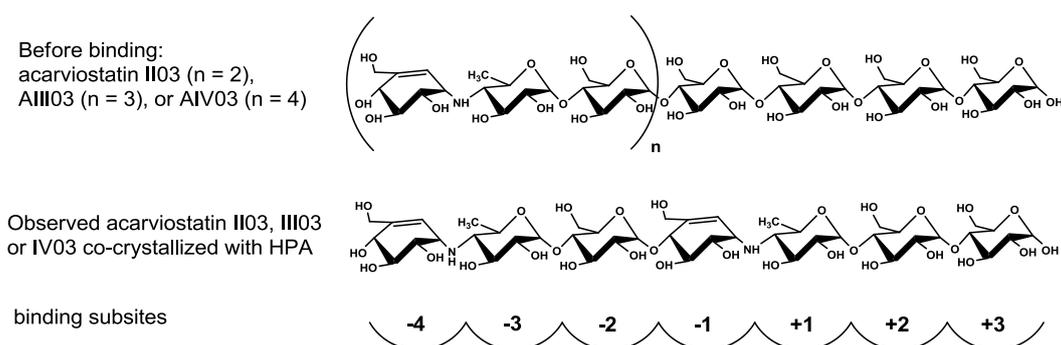


Figure 6: Chemical structures of acarviosatins (II03, III03, and IV03) before (above) and after (middle) interaction with human pancreatic α -amylase.

Likewise, acarbose (pseudotetrasaccharide) is subjected to a chemical rearrangement (Figure 7) that results into another binding product as pseudopentasaccharide (five rings) with HPA [32] or pseudohexasaccharide (six rings) with HSA [49].

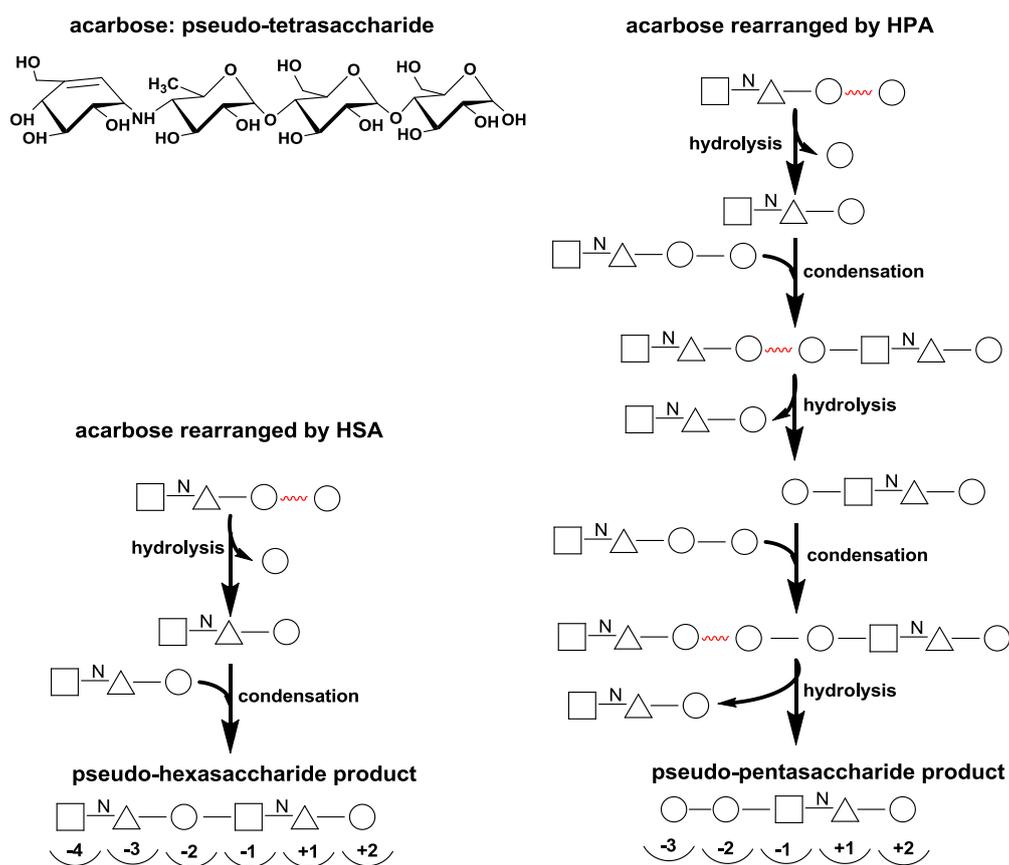


Figure 7: Suggested enzymatic rearrangement for acarbose by HPA [32] (right) and HSA [49] (left). ~~~ indicates the cleavage point. □ represents valienamine. △ symbolizes 4,6-dideoxy- α -D-glucose. ○ represents glucose unit. Negative and positive numbers denote the binding subsites.

This enzymatic rearrangement causes the acarviosine-glucose center to interact with amino acids located at subsites -1 to +2.

Acarviosine-glucose shows characteristic interactions in the active site due to the acarviosine fragment that is composed of valienamine and 4,6-dideoxy- α -D-glucose groups occupying subsites -1 and +1, respectively. The N-glycosidic linkage occupies the position between these two subsites where the breakdown of the natural substrate could occur and resists any further cleavage. This hydrolysis resistance attributes in strong inhibition of acarviostatins. Reasons behind that are (i) the valienamine moiety at subsite -1 has a half-chair conformation that imitates the transition state of the natural substrate, and (ii) the protonated nitrogen of the N-linkage forms strong electrostatic interactions with carboxyl groups of catalytic residues, Asp197, Glu233 and Asp300 [30].

1.4.3 Polyphenols

Polyphenols are widely distributed in plants as secondary metabolites and abundant in human nutrition. They are chemical structures that have at least one phenolic hydroxyl group attached to one or more benzene ring. Phenolic compounds are classified into 15 groups based on the number of carbons in the molecule [70-72]. General structures of important classes that contribute in α -amylase inhibition are depicted in Figure 8.

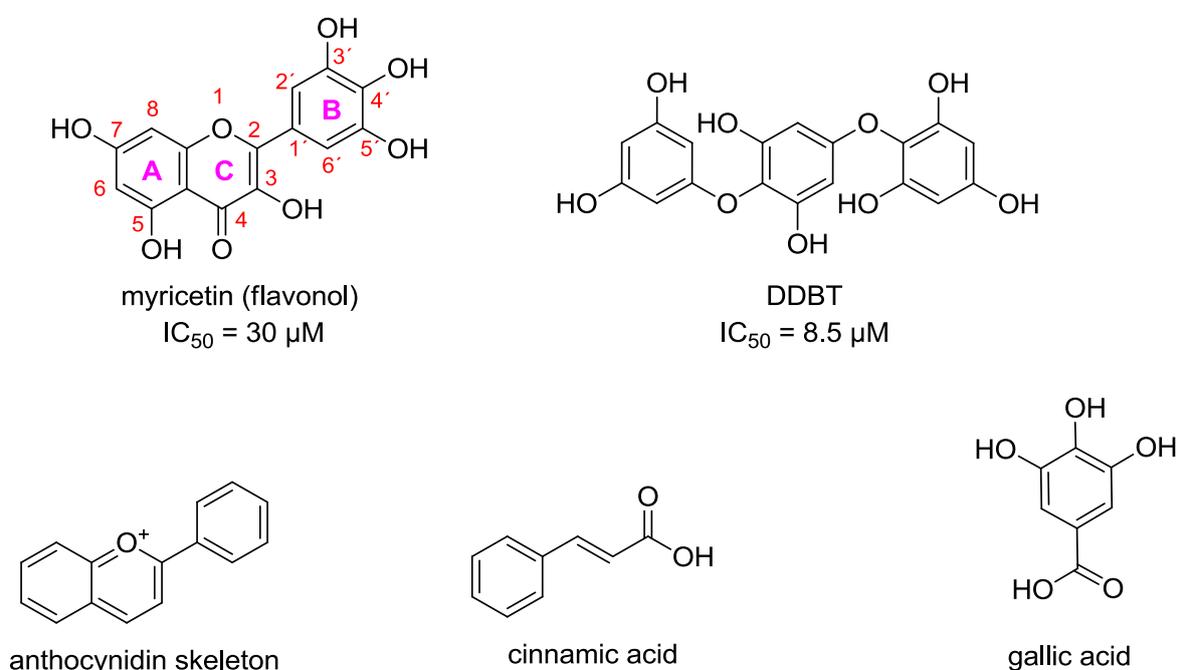


Figure 8: Phenolic compounds that showing α -amylase inhibition.

Flavonoids are the major class of phenols and grouped into six subclasses based on their substitution. These are flavonols, flavones, isoflavones, flavanones, flavanols, and anthocyanidins [72]. They are composed of three rings termed A, C, and B. Flavonoids have been studied as α -amylase inhibitors and the mode of action of the first five subclasses has been discussed in literature [20, 73-78]. Recently, myricetin ($IC_{50} = 30 \mu M$) has been co-crystallized with HPA (PDB code: 4GQR) and published by Williams *et al.* in 2012 [25]. The overall effects of substitutions with different functional groups or fragments are perfectly collected in a recent published review [71]. Briefly, hydroxylation of ring C and/or galloylation in C3 have resulted in better inhibition while methylation, methoxylation of ring C, glycosylation of ring A, or hydrogenation of the C2=C3 double bond have decreased the inhibition potency [71].

Anthocyanidines, the sixth flavonoid subclass, lack the carbonyl group at C4 in ring B and represent the aglycone part of anthocyanins (glycosidic anthocyanidins) [79-81]. This subclass has not been discussed from the point of view of its structural mechanism and its behaviour with α -amylase. Therefore, a study in chapter 4.2.1 will analyze the mode of action of this subclass using molecular modeling.

Another class that shows α -amylase inhibition and belongs to the group of phenolic compounds is the one with a cinnamic acid skeleton, such as ethyl caffeate [82]. The latter shows non-competitive α -amylase inhibitor ($K_i = 1.3 \text{ mM}$ [82]) and has been recently co-crystallized with HPA (PDB: 4GQQ) by Williams *et al.* in 2012 [25]. More details will be given in the study of verbasacoside in chapter 4.2.3.

Tannins which constitute a large group of phenolic compounds have diverse structures such as the phlorotannin 2-(4-(3,5-dihydroxyphenoxy)-3,5-dihydroxyphenoxy) benzene-1,3,5-triol (DDBT) that has been recently discovered in 2012 as a potent α -amylase inhibitor [83]; and gallic acid derivatives such as 1,2,3,4,6-pentagalloyl- β -D-glucose [84]. Theaflavins resulting from co-oxidation of flavanols [85] and containing galloyl fragments show inhibitory effect against α -amylase. A recent study reported by Miao *et al.* in 2013 has shown that hydroxyl groups of theaflavins form H-bonds with catalytic residues in the active site and galloyl moieties form aromatic interact with Tyr2 and Phe335 [86]. Chemical structures and activities of the previously mentioned examples are listed in Appendices 10.3 and 10.4.

1.4.4 2-aminobenzothiazole

In the late of 2012, a study published by Patil *et al.* reported a series of synthetic molecules that show inhibition against PPA [87]. The compound that displaying the most potent α -

amylase inhibition is [2-(4-Bromophenyl)-4H-benzo[d]pyrimido[2,1-b][1,3]thiazol-4-ylidene]acetonitrile with an IC_{50} value of 15.26 μ M (Figure 9). This series will be used in the refined virtual screening in chapter 4.3.8. Interestingly, the same group published in 2013 another series of 2-aminobenzothiazole as PPA inhibitors and the most potent ligand is (E)-1-(4-Bromophenyl)-3-(6-ethoxybenzo[d]thiazol-2-ylamino)-3-(methylthio)prop-2-en-1-one with an IC_{50} value of 15.87 μ M [88].

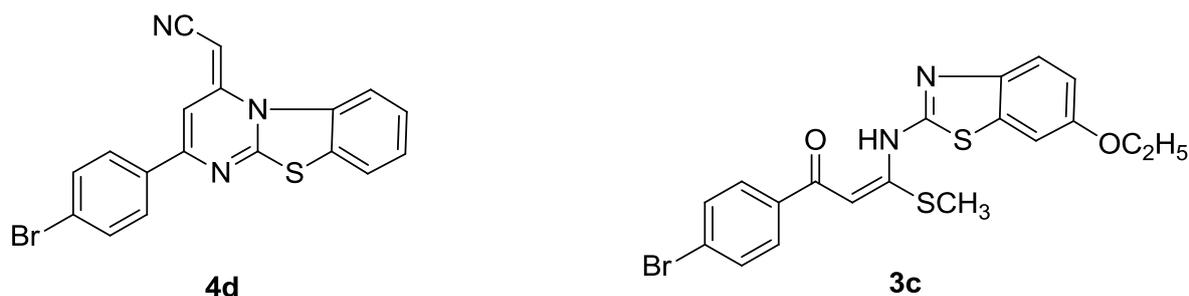


Figure 9: Recently published α -amylase inhibitors with unspecified mode of action [87, 88].

In addition to PPA inhibition, these series show comparable inhibition potencies against murine glucosidases (from pancreas, intestine and liver). However, none of these studies investigated the mode of action of these synthesized compounds.

1.4.5 Terpenoids

Terpenoids, also known as isoprenoids, are a class of natural compounds found in plants and derived from isoprene units (C_5H_8) which are then condensed to form diverse chemical structures with several functions in plant and animals [89]. Of these, an extract from *Phyllanthus amarus* plant containing triterpenoids oleanolic acid, ursolic acid and lupeol (Figure 10) show dose-dependent α -amylase inhibitory effect with an IC_{50} value of 4.41 μ M [90].

So far and to the best of our knowledge, structural mechanism of terpenoids as α -amylase inhibitors has not been investigated. As a result, a molecular modeling study will be conducted in chapter 4.2.2 for diterpenoid steviol (Figure 10) compounds (IC_{50} = 187 μ M for the extract) that show competitive α -amylase inhibition.

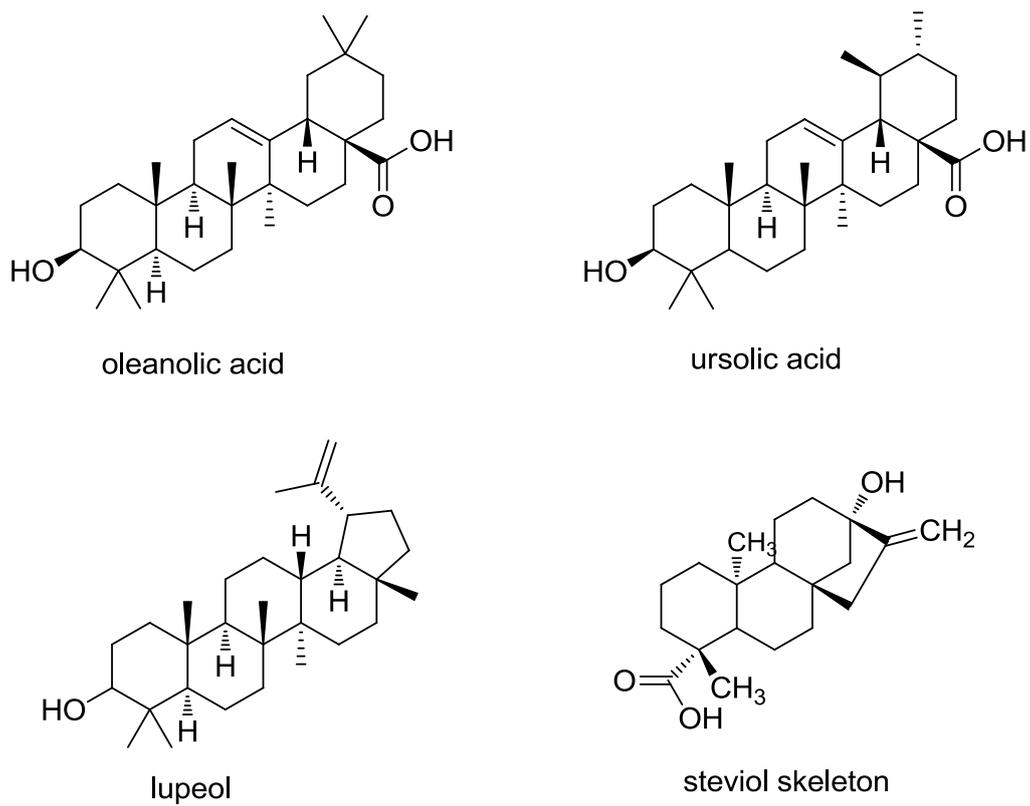


Figure 10: Terpenoids showing α -amylase inhibition.

2 Objectives

α -Amylase enzyme secreted from salivary and pancreatic glands triggers the hydrolysis of polysaccharides such as starch, the main source of energy in human nutrition. This controls the fast elevation of blood glucose levels after meals. Post-prandial hyperglycemia is not only a crucial manifestation in metabolic disorders such as type II diabetes and obesity, but is a direct risk factor for cardiovascular disease (CVD) in type II diabetes [91]. Therefore, α -amylase is considered an attractive target for controlling sudden raise in blood glucose levels after meals. For that purpose, we aimed in this research at identifying novel small drug-like α -amylase inhibitors using molecular modeling and virtual screening by following the objectives below:

1. To analyze binding conformations of known natural α -amylase modulators in order to explain the mechanism of inhibition of these binders. 3D models and docking studies shall be used for this purpose.
2. To explore X-ray structures of α -amylase available in the PDB to select templates for this project and obtain an overview on amino acids important for inhibition. For that, superposition of those structures will be performed and analyzed.
3. To generate structure-based 3D pharmacophores using available co-crystallized sugar-based α -amylase inhibitors. Careful analysis and statistical validation shall yield a predictive model to be used for virtual screening of commercial chemical databases. This step shall lead to identification of novel small drug-like α -amylase inhibitors (initial screening round).
4. To carry out experimental biological testing for hits selected in this work in order to evaluate their inhibitory potency against α -amylase. Kinetic assays will be performed in-house and in collaboration with Dr. Gyöngyi Gyémánt at the University of Debrecen, Hungary.
5. To refine models for inhibitors identified in this work and recently published during this research. The purpose is to elucidate the binding mode of these binders. Docking and 3D pharmacophores shall be used to achieve this goal. A further aim is to utilize those developed models in second virtual screening campaign for the identification of novel small α -amylase inhibitors (refined screening round).
6. To search for analogues similar to the best inhibitors identified in this work. 2D similarity search from a chemical database shall be performed aiming at the discovery of novel small α -amylase binders with improved inhibition. A subsidiary goal

is to analyze the structure activity relationship (SAR) for better understanding their inhibitory mechanism.

These approaches should give new insights into the discovery of small α -amylase inhibitors using an efficient and cost effective computer-aided drug design methodology.

A better understanding shall be gained about the role of sugar units in natural α -amylase inhibitors. Incorporating multi-step approaches as 2D and 3D filters in virtual screening campaigns shall improve the enrichment of true positives among selected virtual hits.

Previously mentioned objectives will be reached by combining both computational and biological approaches. Detailed descriptions for the used methods are provided in the following chapters.

3 Methods

3.1 Molecular modeling and virtual screening in drug discovery

The objective of drug discovery is the identification of novel lead compounds that can be developed into drug candidates [92]. Together with *in vitro* and *in vivo* traditional pharmacology tools, computational methods (*in silico*) have been increasingly developed to create models for the drug discovery and optimization [93]. Computational technologies such as computer-aided drug design (CADD) have the advantages of speed, efficiency and cost effectiveness. They have therefore become an indispensable strategy for pharmaceutical industry [94].

CADD is broadly classified into structure-based drug design (SBDD), also known as direct design, and ligand-based drug design (LBDD) named indirect design [94, 95]. While SBDD uses known crystal structures of a target to investigate interactions of a ligand bound to a binding site, LBDD relies solely on ligands known to interact with a specific target at a defined binding site. LBDD is typically employed if no crystal structure of the target is available [94].

Since high-throughput screening (HTS), which experimentally tests several thousands of compounds against a specific drug target [96], is extremely expensive with massive trial and error basis [97], a computational analog called virtual screening (VS) was developed [92]. As one of the most widespread applications of CADD, virtual screening is used to screen virtual libraries of compounds aiming at identifying lead structures. This leads to improved hit rates, higher efficiency and saves money and time [96].

Structure-based virtual screening requires a known crystal structure of a drug target, which can be elucidated by several tools such as X-ray crystallography and NMR. The largest public collection of 3D crystal structures of macromolecules is the Protein Data Bank (PDB) [5] and some of these structures are deposited as a protein-ligand complex. Thus, a binding site for a co-crystallized ligand is defined and can be used in drug discovery approaches. The most popular strategies of SBDD are molecular docking of ligands into a specific binding site [98] and structure-based 3D pharmacophores [99].

Compared to SBDD, ligand-based approaches lack the crystal structure of a target and hence biological and chemical properties of ligands are correlated to predict target affinity. The most prominent ligand-based methods are molecular similarity, 2D and 3D pharmacophore models, and quantitative structure activity relationship (QSAR) [96]. Similarity search depends on a concept that similar compounds are supposed to exert similar

effects [100]. It is achieved by encoding known bioactive ligands into molecular fingerprints to be used for a discovery of bioactive analogues [97, 101].

In case of 3D pharmacophore models, common chemical features of a set of bioactive compounds of a specific target are identified by flexible, three-dimensional overlay. These 3D models are applied in virtual screening aiming at identifying novel scaffolds with potential activity. For QSAR models, the third strategy of LBDD, a statistical correlation between physicochemical and structural properties of known bioactive compounds and their biological activity is derived for predicting potencies of novel compounds [102].

An overview of SBDD and LBDD as well as combination of both methods is summarized in Figure 11.

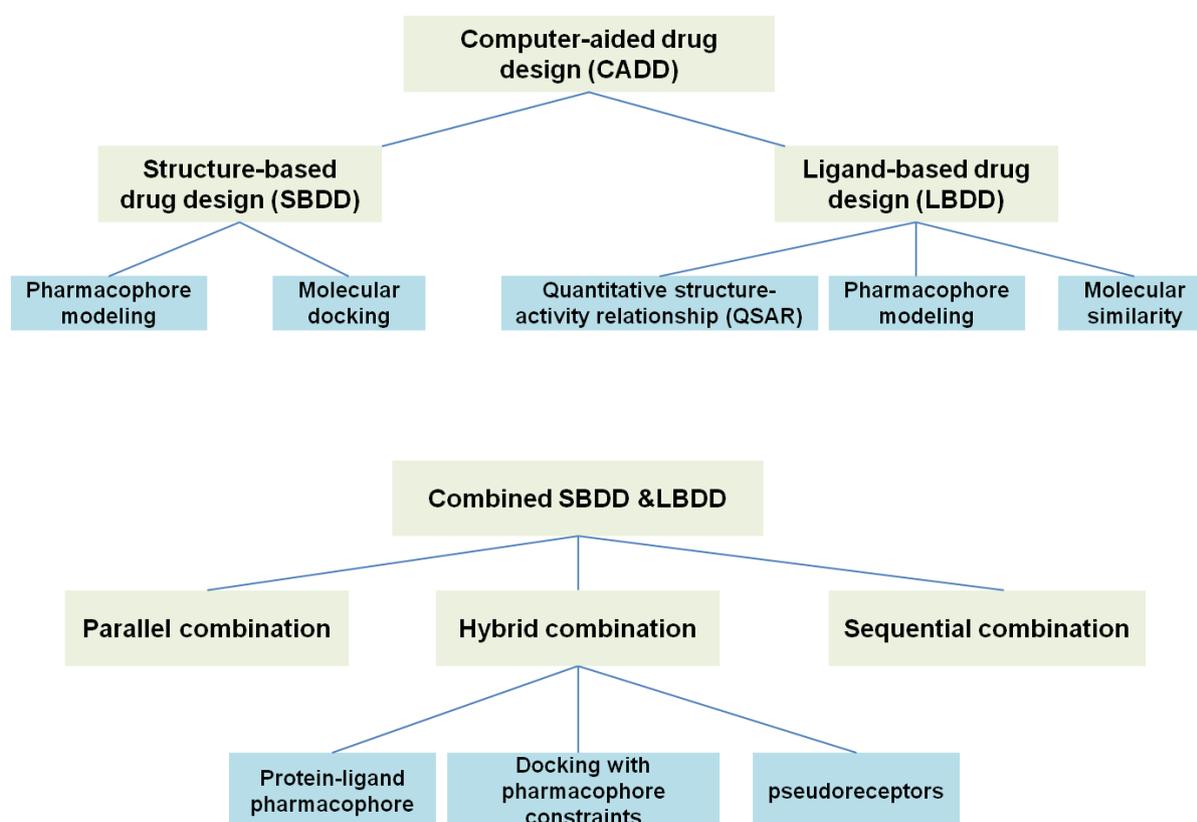


Figure 11: A general view of approaches used in computer-aided drug design.

Combining structure- (SB) and ligand-based (LB) methods in virtual screening is beneficial because all possible information of both protein and ligand are integrated to enhance the consistency, quality, and performance of computer-aided drug-design approaches [102, 103]. Combination of SB- and LB-methods can take place either in sequential, parallel or hybrid approach. Sequential manner uses different SB- and LB-filters consecutively in virtual

screening until a small manageable number of hits are attained for biological experiments. Parallel methods apply several SB- and LB-approaches independently and then the top hits from each approach are biologically tested [102]. Hybrid approaches represent a true combination of SB- and LB-methods since information of both a target and a ligand is applied in the same time as one approach. A good example of a hybrid approach is protein-ligand 3D pharmacophores that are created based on a target structure co-crystallized with a ligand as key interactions, which is then optimized using known ligand data [102].

In the next chapters, methods exploited in this work will be addressed. In chapter 3.1.1, the protein-ligand pharmacophore approach and virtual screening will be explained while molecular docking and its scoring functions will be introduced in chapter 3.1.2. Then, similarity search (2D approach) will be highlighted in chapter 3.1.3 while the concept of shape- and feature-based similarity search (3D approach) will be given in chapter 3.1.4. Assessment and validation strategy for 3D pharmacophore virtual screening will be explained in chapter 3.1.5 to evaluate the effectiveness of virtual screening techniques. More consideration will be applied for tools and software packages that were used in this study.

3.1.1 Pharmacophore modeling

In 1909, the term “pharmacophore” was first used by Ehrlich to represent a molecular framework that carries the essential features (phoros) responsible for the biological activity of a drug (pharmacon) [104, 105]. In 1998, IUPAC recommended a more precise definition for a pharmacophore which is stated as: “A *pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response*” [106]. Notes accompanying the previous definition clearly stated that a pharmacophoric feature does not represent a chemical fragment or functional group but chemical functionality. This model implies the shared molecular interaction capacities of a group of compounds with a set of complementary sites on the biological target [107].

Chemical features included in a 3D pharmacophore model are hydrogen bonding interactions, lipophilic areas, aromatic interactions (π stacking), and Coulomb charge interaction. Because pharmacophores are intuitively understandable and have been increasingly successful in computational drug discovery, they are considered one of the main tools used to find compounds with similar binding modes or even similar structures [108].

Software packages available for pharmacophore elucidation are numerous such as: Catalyst [109], Phase [110, 111], GASP [112], MOE [113] and LigandScout [114, 115].

3D Pharmacophore models are developed either in a ligand-based mode by extracting common chemical features that are necessary for activity of a set of bioactive compounds, or structure-based manner by deriving possible interaction points between a target and a bound ligand [105].

In this project, pharmacophore modeling will be processed as a protein-ligand pharmacophore since 3D X-ray structures of α -amylase co-crystallized with inhibitors are available. Because it automatically allows for extracting a co-crystallized ligand with its relevant protein environment, the software package LigandScout 3.1 will be used to develop 3D pharmacophores in this project. These 3D models are universal but still selective enough to describe the protein-ligand interaction for a complex in the PDB. Before automatic generation of a pharmacophore model, LigandScout can also be used to interpret incomplete geometric information available in the PDB regarding hybridization states and bond types of the bound ligand [114].

Another advantage of using LigandScout is its pharmacophore-based alignment algorithm that can efficiently superimpose molecules on a basis of their pharmacophoric features. Also, LigandScout can be used to create a shared feature pharmacophore that represents several structure-based pharmacophores [115].

After developing a 3D pharmacophore model, a multi-conformational library of compounds can be virtually screened in LigandScout aiming at finding novel chemical structures with potential biological activity for the relevant macromolecule [116].

3.1.2 Molecular docking

Docking represents a frequently used approach in structure-based drug design since it requires a 3D structure of a target. It can be used (i) to predict the favorable conformation of a ligand within the target binding site, and (ii) to roughly estimate the strength of this binding using a specific scoring function [117]. Briefly, molecular docking generates several ligand conformations/orientations (also called poses) at the active site, and subsequently ranks them by using a scoring function. Generating this binding hypothesis helps in understanding of its structural mechanism and can be invested to design more efficient ligands [118].

A large number of docking programs are currently available as well as newly ones are being developed and differ in their aspects such as pose prediction, type of algorithm, and scoring function [119]. As briefly described in [120], the software FlexX [121] utilizes incremental construction approaches by positioning a base fragment in an optimal location and then rebuilding the ligand with efficient optimization methods. FRED [122] uses exhaustive search to rigidly locate a generated multi-conformer library in the active site. Glide [123-125] puts a

ligand on a grid and uses energy terms to perform a systematic search of the conformational, orientational, and positional space of the docked ligand. LigandFit [126] is a shape-based docking tool combined with a Monte Carlo conformational search for creating ligand poses compatible with the shape of the binding. The program GOLD [127] uses a genetic algorithm for exploring the conformational space of a ligand within the active site.

Several scoring functions are developed for protein-ligand interactions with different accuracies and computational efficiencies [118]. There are three commonly-used scoring functions: force-field-, empirical-, and knowledge-based scoring functions. The first type is established based on physical atomic interactions such as van der Waals and electrostatic interactions and implemented in programs such as GOLD [127]. Empirical-based scoring function estimates the binding affinity of a protein-ligand complex based on a set of energy terms and is applied in programs such as FlexX [121] and Glide [123]. The third scoring function is also referred as statistical-potential-based and uses energy potentials derived from the structural information established in experimental structures. It is obtained from the repeated occurrence of atom pairs in a database. This type gives a better balance between accuracy and speed compared to the force-field- and empirical-based scoring functions [128-132].

Every scoring function has its pros and cons, and hence none of them is perfect in terms of accuracy and general application [118]. Therefore, in this project, the experimental binding mode of a co-crystallized ligand will be reproduced and evaluated based on values of root mean square deviation (RMSD) between the docked pose and the experimental binding mode as a matter of validation for docking experiments. Additionally, a custom 3D pharmacophore based scoring function is developed using LigandScout.

The software GOLD 5.1 (Genetic Optimization for Ligand Docking) [127] will be used in this work, which is based on a genetic algorithm and performs complete flexible ligand conformation with partial protein flexibility. Similar to natural evolution process, GOLD manipulates a collection of data structures (chromosomes) and possible poses are created. The force-field-based scoring function implemented in GOLD called GoldScore will be applied in this approach as a fitness score [133].

3.1.3 Similarity search and clustering

Based on the concept that chemically similar compounds should have similar biological properties, several similarity searches have been developed. These approaches can be performed as 2D and 3D searching [100]. While 2D similarity approaches derive information from a molecular graph (2D structural representation), 3D similarity approaches compare

molecular conformations and related properties [100]. When bioactive conformations of compounds are unknown, 2D molecular graphs can give robust results in SAR analysis and activity prediction [100, 134].

Similarity search can be performed using a fingerprint that maps 2D molecular structure information as a sequence of zero and one digits (bit string) to express the absence and presence of structural information, respectively. Several similarity coefficients are commonly used in chemical information to quantify the degree of similarity. Among these, the Tanimoto coefficient (also called Jaccard coefficient) is widely used in 2D similarity searches and can be calculated using the following simplified equation: $S_{A, B} = c / (a + b - c)$, where a and b indicate the number of features (bits) in objects A and B respectively, while c represents the common features (bits) in both A and B. The range between zero (low similarity) and one (high similarity) gives a simple measure to represent the degree of similarity between A and B [135, 136]. In addition to its application in 2D similarity approaches, the Tanimoto coefficient can also be applied for 3D shape similarity search [137].

Database clustering is an important application of similarity metrics. It aims at subdividing a group of objects (chemical molecules) into different clusters (groups) that have a high degree of similarity. Clustering provides low-cost significant information in drug discovery [135, 138]. Available clustering algorithms include hierarchical (iterative way of data analysis and is represented as tree diagram) and nonhierarchical methods (each cluster represents a single partition of compounds). The clustering software used in this thesis is JKlustor 5.8.0 [139], which implements the most important clustering methods described above.

Substructure searching is another application of similarity searches, in which all molecules containing a specific query substructure can be retrieved from a database. It is considered a valuable tool for accessing chemical structure databases [135]. This technique is incorporated in many cheminformatics programs such as Instant JChem [140].

3.1.4 Shape-based virtual screening

Virtual screening using shape-based molecular similarity approaches is a widespread ligand-based technique that leads to successful discovery drug design [141].

In this work, the platform Rapid Overlay of Chemical Structures (ROCS) will be used to perform 3D similarity ranking based on steric overlap [142-144]. The concept of this method is the use of a bioactive reference ligand in a single conformation as a query. One advantage of ROCS is that the knowledge of the experimental conformation of a bioactive molecule does not necessarily enhance the performance of the shape-based similarity search. Therefore, a low-energy conformation such as a docking pose can be used as a query and

lead to promising results. ROCS uses Gaussian spheres to determine the volume overlap between the heavy atoms of a query ligand and a screened molecule. To overcome the limitation that steric volume (shape) does not encode chemical functionality, a combination of both shape similarity and simplified pharmacophoric similarity pattern is used (Tanimoto ComboScore). [141].

3.1.5 Statistical validation of model predictivity

As explained in the previous sections, a model can be used for virtual screening and lead to the discovery of novel potential bioactive compounds. Such a model should have the ability to select as many bioactive compounds as possible and exclude a large fraction of inactive molecules [145]. To assess the predictive power of a model, a validation step is required using a set of molecules with known activity for the regarded target. In most cases, only limited numbers of inactive compounds are published and therefore available for molecular modeling. In such a situation, a random selection of compounds with similar structural and physicochemical properties and assumed biological inactivity can be used (decoys) [146].

Several methods are applied to estimate the predictive quality of a model: When screening a database that contains N molecules, a model retrieves n compounds that fit the pharmacophoric features but not all of them are necessarily biologically active.

Known bioactive molecules that fit the model and are therefore predicted correctly are called true positives (TP), while the ones which are inactives are called false positives (FP). The known biologically actives that do not match the model are false negatives (FN), while the ones that have no activity and do not match the pharmacophore are true negatives (TN) [145]. A simple representation of these four categories of compounds selected by a model is given in Figure 12.

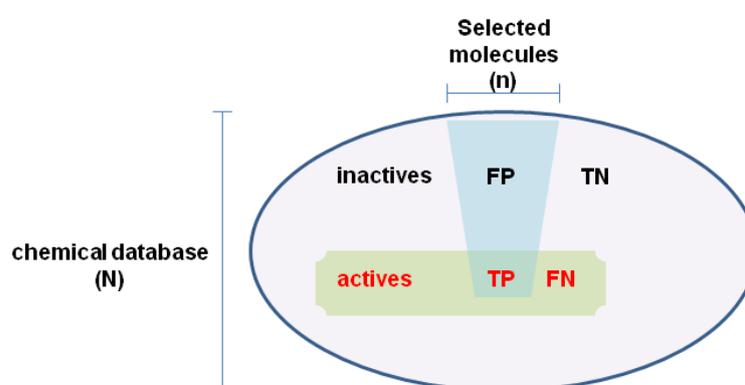


Figure 12: Selection of n molecules (FP+TP) by a model from a validation database N (actives and inactives) [146].

Those previous assessments can be represented as virtual screening validation metrics, which include the true positive rate (*sensitivity*, Se) and false positive rate (*specificity*, Sp), Table 2. Se deals with actives and is calculated by dividing the number of selected bioactive compounds over all bio-actives in the database. Sp deals with inactive molecules and is quantified by the ratio between inactives discarded by the pharmacophore and the total inactives in the database [145, 147].

Table 2: Enrichment metrics used in validation of virtual screening

Metric	Equation	
Sensitivity	$Se = \frac{\text{selected actives}}{\text{all actives}}$	$Se = \frac{TP}{TP + FN}$
Specificity	$Sp = \frac{\text{discarded inactives}}{\text{all inactives}}$	$Sp = \frac{TN}{TN + FP}$
Enrichment factor	$EF = \frac{\text{selected actives} \div \text{all selected}}{\text{all actives} \div \text{all molecules in database}}$	$EF = \frac{TP \div n}{A \div N}$
Area under the curve	$AUC = \sum_i [(Se_{i+1})(Sp_{i+1} - Sp_i) \div 2]$	

The enrichment factor (EF) is an important metric that describes how many times the list of compounds predicted as active is richer in true positives as the original database. When the EF value equals one, it means that a model has no predictive power, but simply performs a random selection. The model is considered useful if the EF value is substantially greater than one [145, 146, 148].

Because not all compounds matching a model during virtual screening can biologically be validated, only a cut-off of top-ranked hit list should be chosen for experiments.

To ensure that the picked cut-off list contains the best possible ratio of actives to inactives, a validation method is carried out using the receiver operating characteristic (ROC) curve [147]. This visual method scores compounds that match a pharmacophore and shows molecules with the highest scores at the top list. It uses Se values as a function of $1 - Sp$. The ROC curve (Figure 13) represents a random selection when a straight diagonal line from the origin to the upper right corner ($Se = 1 - Sp$) is plotted.

Comparatively, the ideal pharmacophore-matching is represented as a straight vertical line between the origin and the upper left corner until selection of all actives. Then a continuous line from the upper left corner to the upper right corner is plotted for scored inactive compounds. The ROC curve plotted between ideal and random indicates pharmacophore

performance that can retrieve actives better than random selection while the curve below the random line represents even lower predictiveness than a random selection [146, 147, 149].

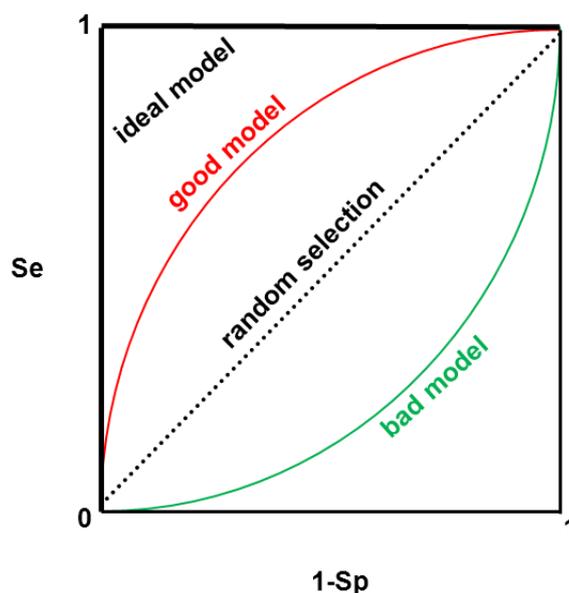


Figure 13: ROC curves representing the quality of a virtual screening experiment [147].

When aiming at validation of several models, the area under the ROC curve (AUC) for each model can be used as a single value. This value is ranged between zero (when all inactives are ranked first) and one (when all actives are ranked first). A convenient AUC value for virtual screening is the one, which is greater than the random selection ($AUC > 0.5$). When limited testing capacity is available, attention has to be paid to the beginning of the function, which reflects the early recognition of actives [147].

3.2 Biochemical assay

Generally, two main approaches can be developed to experimentally measure enzymatic activities: continuous (kinetic) and discontinuous (end-point) methods. While a kinetic assay uses continuous measurement of the enzyme activity with respect to time, end-point methods stop the enzymatic reaction by a specific reagent and subsequently measure product concentration after a specific time [150].

Different experimental procedures have been developed for the determination of the activity of α -amylase. Among them, the following methods are reported the most often: (1) Saccharogenic types: measuring the amount of reduced sugar released from the substrate; (2) Amyloclastic types: measuring the disappearance of a starch substrate using the specific

reaction between iodine and residual starch; (3) Chromogenic types: measuring the amount of the released chromogenic group linked to the substrate using spectrophotometric detection [151, 152].

The lack of standard assay, standard substrate, or even standard reaction conditions led to the development of these various methods for amylase activity. Therefore, comparing results obtained from different methods or laboratories is very difficult [152].

Hence, using different variables (such as chloride and calcium concentrations, substrate type, and incubation time) can lead to different inhibitory potency. Acarbose, a commercially available anti-diabetic drug is given as an example to show the effect of using different conditions on the reported inhibitory values (Table 3).

Table 3: Different assay conditions for acarbose with α -amylase showed diverse IC_{50} and K_i values

IC_{50}	K_i	Assay type	Substrate	Enzyme	Conditions	Ref.
41 μ M	-	End-point	amylopectin	HSA, HPA	sodium phosphate buffer, pH 7.0	[83]
4.89 mg/ml (7.58 mM)	-	End-point	starch	PPA	sodium phosphate buffer, pH 6.9	[153]
0.996 μ M	-	End-point	potato starch	HSA	sodium phosphate buffer, pH 6.0	[20]
-	2.6 μ M	End-point	amylose	HPA	potassium phosphate buffer, pH 6.0	[36]
-	3.7 μ M	End-point	amylose	HSA	Na_2HPO_4 , KH_2PO_4 , pH = 7.0	[154]
-	0.7 μ M	Kinetic	GalG2CNP	HSA	MES buffer, pH = 6.0	[154]
23 μ M	-	Kinetic	<i>p</i> -NPG5	PPA	HEPES buffer, pH = 7.1	[75]

HSA: human salivary, HPA: human pancreatic, PPA: porcine pancreatic α -amylases

The rearrangement of acarbose results in bound products with a different number of rings depending on the type of enzyme used, as already described in chapter 1.4.2. In Table 3, diverse IC_{50} and K_i values are reported, ranging from 0.7 μ M to 7.58 mM for one single inhibitor compound, acarbose in different conditions.

The kinetic assays developed in this work are based on previously described assays using two different chromogenic substrates: *p*-Nitrophenyl- α -D-maltopentaoside (*p*-NPG5) [75] and 2-chloro-4-nitrophenyl-4-O- β -D-galactopyranosylmaltoside (GalG2CNP) [154]. These assays are described in sections 7.3.1 and 7.3.2, respectively.

4 Results

As explained earlier, postprandial hyperglycemia is considered one of the main manifestations of metabolic disorders such as diabetes type II and obesity. Since α -amylase initiates the hydrolysis of polysaccharides such as starch, as a main source in human nutrition, into oligomers, it is considered a drug-target for controlling abnormal high levels of blood glucose.

In the following chapter 4.1, α -amylase crystal structures and their binding sites will be investigated. While unknown binding modes for α -amylase inhibitors discovered by our collaboration partners will be elucidated by molecular docking in chapter 4.2, perspective structure-based virtual screening approaches for α -amylase inhibitors will be explained in chapter 4.3.

4.1 Structural analysis of different α -amylase isoforms and binding sites

Crystal structures of α -amylase publically available in the PDB were investigated for the following reasons:

- i. To highlight the differences between human and porcine α -amylases, since the latter is widely used in many biological assays with α -amylase inhibitors.
- ii. To investigate conformational changes in the active site when co-crystallized with different inhibitors, and
- iii. To select templates for the modeling study based on type of a co-crystallized inhibitor, resolution of the x-ray structure, and inhibitory potency of the bound ligand.

In humans, α -amylase is secreted from pancreatic and salivary glands and represented as human pancreatic α -amylase (HPA) and human salivary α -amylase (HSA) which are publically available in the PDB [5, 6] with a total of 40 and 13 entries, respectively. Porcine pancreatic α -amylase (PPA, eight PDB entries) is also considered in this study.

In this project, macromolecules were divided into five categories:

- i. Co-crystallized with peptide-based inhibitors such as tendamistat [63]. This type was excluded due to complicated interaction pattern involving areas further away from the catalytic center [24, 64].
- ii. Mutated X-ray structures. This group was not taken into account except those that were bound to an inhibitor.

- iii. Co-crystallized with carbohydrate-based inhibitors. This group was analyzed and used in this work.
- iv. Co-crystallized structures recently published in 2012 with non-carbohydrate-, non-peptide-based small molecules [25]. This type was investigated and utilized in this study.
- v. Crystallized macromolecules without an inhibitor. This type was excluded from structural analysis.

All collected structures of HPA, HSA, and PPA are given in Appendix 10.1

4.1.1 Structural investigation of α -amylase

Conformational changes of α -amylase resulting from formation of a protein-inhibitor complex were investigated to highlight the conserved residues and deviated areas in different α -amylase structures. Four inhibitors are observed in more than one X-ray structure (Appendix 10.3). To avoid duplicate analysis of similar structures, inhibitor complexes were first visualized using LigandScout 3.1. The PDB code 3OLE was then chosen as a template in this study since it is co-crystallized at high resolution of 1.55 Å with the longest sugar-based inhibitor that comprises seven rings (chapter 1.4.2) termed acarviosatin II03 ($K_i = 0.014 \mu\text{M}$). Using the software package MOE 2010.10 [155], a sequence alignment between 3OLE as a reference structure and relevant α -amylase structures was conducted to establish the correspondence between the aligned amino acids (Table 4).

Table 4: Superposition of X-ray structures of α -amylase using the software MOE

Superposed macromolecules	RMSD of all 496 residues, Å	RMSD of pocket residues, Å	Number of pocket residues
3OLE, 1CPU	0.253	0.275	64
3OLE, 3OLD	0.086	0.083	65
3OLE, 1MFV	0.401	0.372	71
3OLE, 1XD1	0.328	0.494	64
3OLE, 1XCX	0.291	0.401	64
3OLE, 1XCW	0.275	0.325	64
3OLE, 3IJ9	0.255	0.402	74
3OLE, 3IJ7	0.239	0.365	64
3OLE, 4GQR	0.339	0.712	64
3OLE, 4GQQ	1.420	1.503	92
3OLE, 3OLI, 3OLD, 3OLG	0.263	0.232	75

Subsequently, the sequence-aligned 3D macromolecules were geometrically superpositioned by minimizing distances between alpha carbons (CA). The root mean square deviation (RMSD) was calculated for all residues and pocket residues (amino acids around the co-crystallized ligand).

The following HPA entry codes were investigated: 4GQR and 4GQQ [25] as models co-crystallized with small non-carbohydrate-based inhibitors, while codes 1CPU [30], 3OLE, 3OLI, 3OLG, 3OLD [36], 1XCW, 1XCX, and 1XD1 [32], 3IJ7, 3IJ9 [31] represent structures bound with carbohydrate-based inhibitors. PPA and HSA were represented by entries 1MFV [49] and 1OSE [156], respectively.

Results demonstrate that all superposed macromolecules co-crystallized with sugar-based inhibitors show only slight RMSD values below 0.5 Å in both states (the whole macromolecule and the active site region). These structures hold the same type of ligands and hence the high similar conformations can be explained. An example is given in Figure 14.

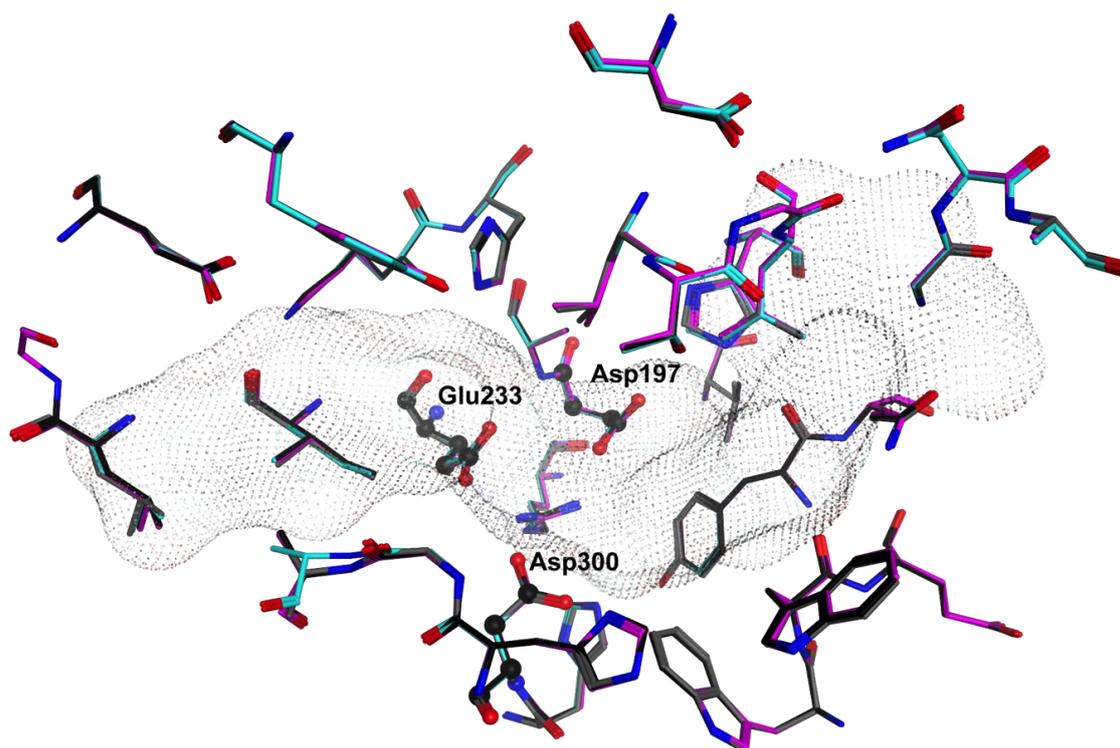


Figure 14: Superposition of α -amylase active site for PDB codes 3OLE (in black), 3OLG (gray), 3OLD (pink), and 3OLI (cyan) using the software MOE 2010.10. Catalytic residues are in balls and sticks. Molecular surface of acarviostatins II03 (from 3OLE) is depicted as grid dots.

Owing to the reasons previously mentioned about 3OLE, it was taken as one of the templates used for molecular modeling in this research. During this work, new crystal

structures have been published including 4GQR and 4GQQ [25]. These structures are the only ones that exhibit slightly higher RMSD values when superposed with 3OLE. 4GQR, the only structure that is co-crystallized with myricetin ($IC_{50} = 30.2 \mu\text{M}$ [20]), a small non-carbohydrate molecule inside the active site, was superposed with 3OLE and shows a RMSD value of 0.712 Å. This can be explained by the difference in binding mode for each type of co-crystallized ligands. The ligand of 4GQR is a flavonoid with three aromatic rings while the ligand of 3OLE is a trestatin with seven rings as carbohydrate-based inhibitor. This leads to a critical conformational change in the position of Asp300, one residue involved in the catalytic triad.

With 4GQR, the myricetin displaces the side chain of Asp300 that leads to be re-oriented in the direction of His305. Consequently, the loop 304-310 anchored by Asp300 moves and directs outwards from the active site (unlike with the trestatin family with whom this loop is directed toward the active site) [25]. This crystal structure will be used in the refined virtual screening described in chapter 4.3.8.

4GQQ, another crystal structure recently published by the same group is co-crystallized with ethyl caffeate ($K_i = 1.3 \text{ mM}$ [82]) in three allosteric sites away from the productive active site [25]. Superposition of 3OLE and 4GQQ shows high RMSD values (1.4 Å and 1.5 Å, respectively) for the whole structure and the pocket. This can be explained by different positions of binding of the co-crystallized inhibitors in both crystal structures as well as the structural difference between both inhibitors.

Three molecules of ethyl caffeate bound to the triad allosteric sites lead to disorder four segments of polypeptide chain loops (residues 51 - 56, 104 - 111, 298 - 315, and 343 - 358) located in the vicinity of HPA active site. Disordering the segment containing residues 304 - 310; which are normally anchored by Asp300 and oriented around the bound ligand; leads to dislocation of Asp300 and play a major role in ethyl caffeate inhibition [25]. This structure will be used to perform a docking study in chapter 4.2.3.

PDB entry codes 4GQR, 4GQQ [25], and 3OLE [36] were selected as templates for HPA. They were chosen because (i) they have high resolutions not exceeding 1.55 Å (ii) they are the most recently published structures of HPA; and (iii) 4GQR and 4GQQ are the only available structures that are co-crystallized with small molecule non-carbohydrate, non-peptide α -amylase inhibitors.

PDB codes 1MFV [49] and 1OSE [156] were chosen to represent HSA and PPA, respectively, since they represent wild type enzymes co-crystallized with acarbose. Structural comparisons between HPA and HSA, on one side, and between HPA and PPA, on the other side were described briefly in chapter 1.2.

4.2 Docking using α -amylase inhibitors with previously unknown binding mode

Our collaboration partners identified α -amylase activity for natural products with unknown binding mode and reaction mechanism. Therefore, we performed a molecular modeling study to elucidate how they exert their activity. Investigated inhibitors include: (i) cyanidin and steviol derivatives which were examined against HSA by our research partner Dr. Gyöngyi Gyémánt from the University of Debrecen, and (ii) verbascoside which was tested against PPA by Prof. Dr. Matthias F. Melzig at the Freie Universität Berlin.

Even though cyanidin derivatives were reported as α -amylase and α -glucosidase inhibitors [79, 80, 157], it is the first time their binding modes are investigated and discussed using molecular modeling. To the best of our knowledge, steviol derivatives and verbascoside have never been reported as α -amylase inhibitors before this study.

The scope of this computational study is to explain the inhibitory effect of these inhibitors and to investigate their mechanism of action using molecular docking and pharmacophore-based ranking and analysis of the docking poses.

4.2.1 Cyanidin derivatives

Cyanidin compounds are considered promising natural products that could be used as food additives with a potential role on human health. They are also responsible for pigmentations of widespread fruits such as berries and showing diverse colors depending on the pH. These natural compounds exert different biological properties such as antioxidants, anti-cancer, anti-obesity and anti-diabetes [158]. It was reported that these compounds have the ability to inhibit α -amylase [79, 80, 157] but none of the investigated articles has studied their mechanism of inhibition. Therefore, it was interesting to investigate how these compounds exert the inhibition against α -amylase.

The crystal structure of HSA (PDB entry 1MFV [49]) was used as a template in order to perform a docking study using CCDCs software GOLD 5.1 [127]. Docking poses were analyzed and compared to the co-crystallized ligand acarbose, the reference inhibitor in this study that shows IC_{50} value of 0.5 μ M. The docking conformations stabilized by similar interactions as acarbose in the HSA active site were selected and further investigated. Then, a strategy based on 3D pharmacophores was utilized for the analysis and discussion of the most plausible poses.

The modified acarbose inhibitor that is co-crystallized with HSA is spanning subsites -3 to +2 [38]. Using the crystal structure 1MFV [49], the binding mode of acarbose was analyzed with

LigandScout 3.1 [114, 115, 159] as illustrated in Figure 15. The key interactions of this inhibitor were compiled in a 3D pharmacophore model. One central interaction motif can be observed between the acarviosine moiety (valienamine and 4,6-dideoxy- α -D-glucose) and residues at subsites -1, +1 of HSA. The residues involved include Asp197, Glu233 and Asp300, which form electrostatic interactions with the positively charged N-atom of valienamine, on one side, and interact with the hydroxyl groups present in the acarviosine moiety, on the other side. Moreover, the large number of hydroxyl groups of the inhibitor allows for the formation of H-bonds with other amino acids present in the active site that enhance the binding efficiency of the ligand as depicted in Figure 15. Additional hydrophobic contacts, which stabilize the ligand, are formed between the methyl groups of 4,6-dideoxy- α -D-glucose moieties in the ligand with Ile235 and Leu165 at subsites +1 and -3, respectively.

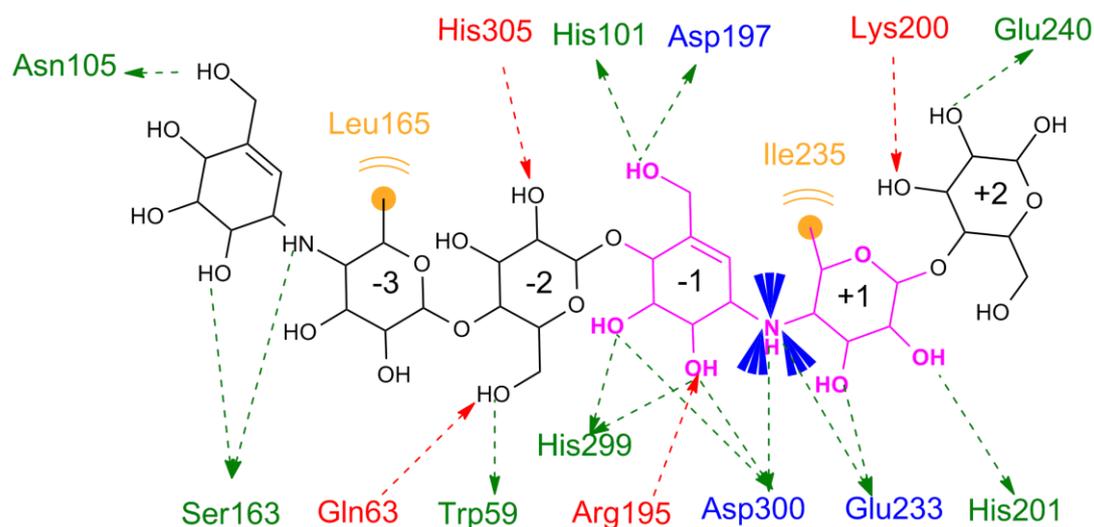


Figure 15: The pharmacophoric features of the co-crystallized modified acarbose inhibitor with HSA (PDB code: 1MFV) were derived using LigandScout 3.1. Yellow spheres indicate hydrophobic contacts, H-bonds are denoted as green and red arrows, and blue stars represent the positively ionizable interactions. Negative and positive numbers represent the active subsites and put inside the rings for clarity. Acarviosine moiety at subsites -1, +1 is depicted in pink.

The resulting conformations of the docked compounds were analyzed with LigandScout 3.1 to select the most plausible poses that fulfill similar interactions as the reference inhibitor. Then, a 3D pharmacophore was created for each selected pose to rationalize the activity of the compounds in question.

The selected conformations of cyanidin-3-*O*-glucoside (Cy-3-*O*-glc, $IC_{50} \sim 200 \mu\text{M}$) and cyanidin-3-*O*-rutinoside (Cy-3-*O*-rut, $IC_{50} = 200 \mu\text{M}$) have highly similar predicted binding modes

as illustrated in Figure 16. In Cy-3-O-glc, the 2-, 4-, and 6-hydroxyl groups of the glucose unit linked to ring C form H-bonds with His305, Asp300, His299 and Arg195. Likewise, in Cy-3-rut, the 2-, 3-, and 4-hydroxyl groups attached to this glucose unit form H-bonds with Arg195, Asp197 and Asp300, on one side, and Glu233 and the 5-hydroxyl group attached to ring A, on the other side.

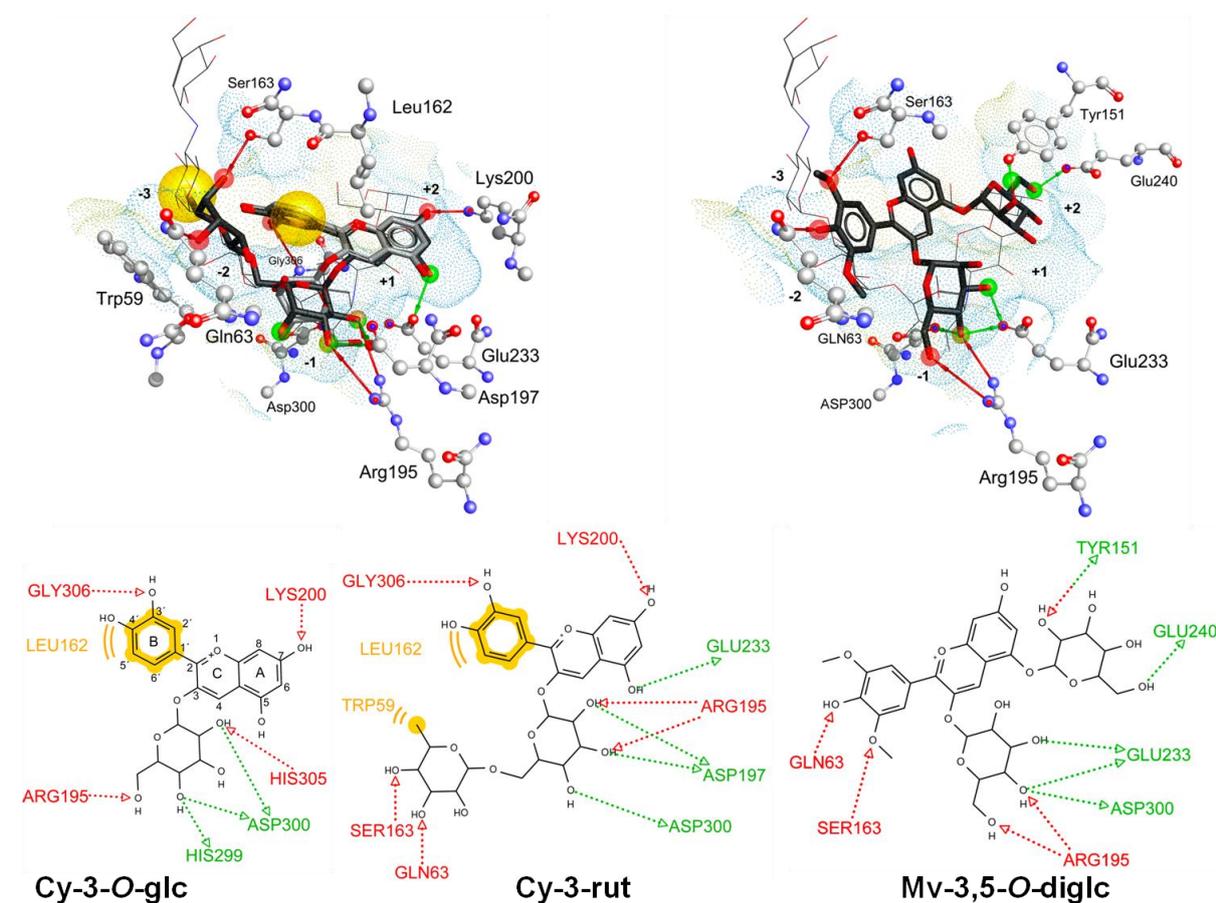


Figure 16: Proposed binding modes of the cyanidin compounds in HSA (PDB entry: 1MFV) depicted above in 3D [left for Cy-3-O-glc (gray stick) and Cy-3-rut (black stick) and right for Mv-3,5-O-diglc]. Acarbose is depicted as lines and the binding sub-sites are denoted as negative and positive numbers. Yellow spheres indicate hydrophobic contacts, green and red arrows represent donating and accepting H-bonds, respectively. Variable sizes of amino acid labeling correspond to their depth in the cavity. 2D (below) represents compounds Cy-3-O-glc, Cy-3-rut, and Mv-3,5-O-diglc.

Additionally, the cyanidin scaffold of compounds Cy-3-O-glc and Cy-3-rut is stabilized through hydrophobic contact between ring B and Leu162 and also by forming H-bond between Lys200 and the 7-hydroxyl group of ring A. Compared to Cy-3-O-glc, the 6-deoxyglucose unit of Cy-3-rut forms lipophilic contact between the 6-methyl group and the

Trp59 besides the formation of H-bonds between the 3-hydroxyl and 4-hydroxyl groups of this unit and Gln63 and Ser163.

Surprisingly, investigation of malvidin-3,5-diglucoside (Mv-3,5-O-diglc, $IC_{50} = 70 \mu\text{M}$) inside the pocket revealed that the docked ligand is flipped by about 60° with respect to the selected poses of compounds Cy-3-O-glc and Cy-3-rut, as illustrated in Figure 16. Ring B is directed toward subsites -2, -3 and the 4'-hydroxyl and 5'-methoxy groups form H-bonds with the side chains of Gln63 and Ser163, respectively.

Furthermore, this flipping directs the glucose unit linked to the ring C toward the subsite -1 to form H-bonds between 3-, 4-, and 6-hydroxyl groups with the side chains of Glu233, Asp300 and Arg195. The glucose unit linked to ring A is oriented toward subsite +2 to stabilize the compound by forming H-bonds between 2-hydroxyl and 6-hydroxyl groups with the side chains of Tyr151 and Glu240, respectively.

Thus, the flipping of Mv-3,5-O-diglc may explain why its binding efficiency is better than Cy-3-O-glc and Cy-3-rut and hence shows higher activity. This suggests that the orientation of ring B in cyanidin derivatives toward subsites -2 and -3 is important for α -amylase inhibition.

Generally, the main docking poses of cyanidin derivatives inhibiting HSA are similar to the interactions of acarbose at subsite -1. There, the glucose unit linked to ring C at position 3 of all cyanidin derivatives is located at the same position and superimposed on the valienamine moiety of acarbose. This allows the hydroxyl groups of that glucose unit to interact with the catalytic residues. In contrast, the lack of ionic interactions between cyanidin derivatives and the catalytic residues may explain their lower activity compared to acarbose. However, the unexpected orientation of Mv-3,5-O-diglc, which is not observed with Cy-3-O-glc and Cy-3-rut allows Mv-3,5-O-diglc to interact with the residues spanning from subsites -3 to +2. Thus, Mv-3,5-O-diglc fully occupies the large binding site of the enzyme, similarly to acarbose. This distinct orientation might explain the higher inhibitory potency of this molecule compared to Cy-3-O-glc and Cy-3-rut.

4.2.2 Steviol compounds

Steviol glycosides are natural products originally extracted from the South American plant called *Stevia rebaudiana*, which are responsible for the extensive sweet taste of the plant leaves (300 times more than sucrose). The aglycone part is called tetracyclic diterpenoic acid steviol and is structurally related to gibberellins [160]. As a calorie free sweetener, this *Steva* extracts are used as commercial sweeteners and food additives and can substitute sucrose [161]. It was reported that *Steva* extracts are able to enhance glucose uptake in an effect similar to insulin and hence hold beneficial action against diabetes and obesity [161]. The α -

amylase inhibition with an IC_{50} value of 187 μ M for an extract containing steviol, stevioside, and rebaudioside are given for the first time by collaborator partners in the University of Debrecen.

Here, we structurally investigated the three above-mentioned steviol derivatives as α -amylase inhibitors mechanistically, Figure 17. This should also give insights into the question, which compound is most likely to show α -amylase inhibiting effects.

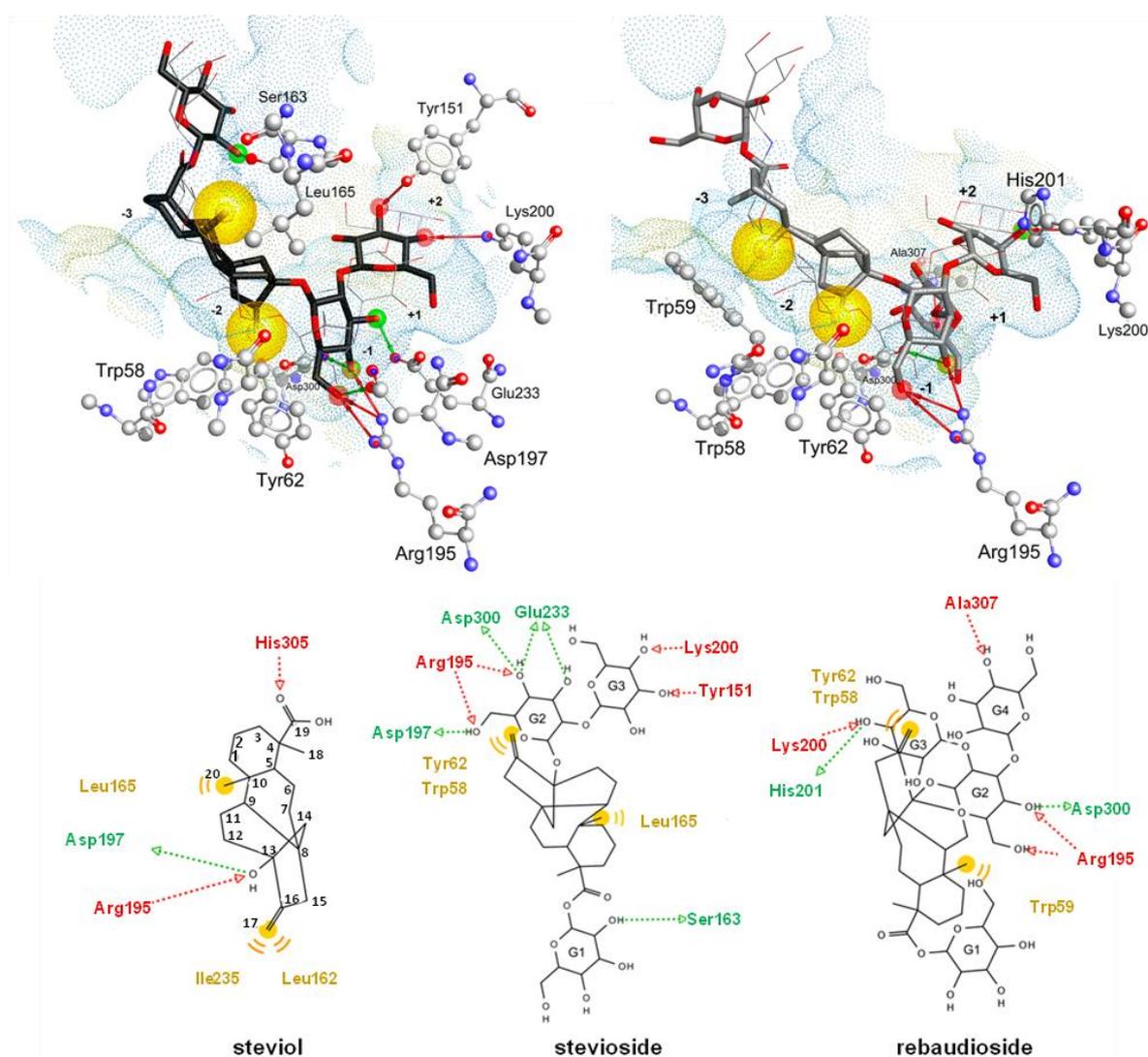


Figure 17: The most plausible binding poses of steviol compounds docked into HSA, PDB entry 1MFV. 3D (above) represents stevioside (black stick) at the left and rebaudioside (gray stick) at the right. The co-crystallized acarbose is illustrated as lines and the negative and positive numbers represent the binding subsites. Yellow areas stand for lipophilic contacts, green and red arrows are denoted for hydrogen bonding. 2D (below) represents steviol, stevioside, and rebaudioside and the glucose units are abbreviated as G1, 2, 3, and 4.

Three derivatives (steviol, stevioside and rebaudioside) were docked in HSA using the same approach as cyanidin derivatives. The most plausible poses were selected based on their ability to fulfill the 3D pharmacophore features of acarbose described before.

Subsequently, 3D pharmacophores were created for these selected poses using LigandScout 3.1.

None of the resulted poses of steviol fulfill the interactions compiled in the acarbose pharmacophore. In contrast, stevioside and rebaudioside show a more plausible orientation inside the binding site than steviol, and span from subsites -3 to +2 in a way that they have a similar orientation to the one of acarbose.

The hydroxyl groups of the glucose unit (G2) in stevioside and rebaudioside can form interactions through H-bonds with residues located near to subsites -1, +1. In stevioside, four H-bonds are observed between the 3-, 4-, and 6-hydroxyl groups of G2 and the side chains of the catalytic center (Asp197, Gl233 and Asp300). In contrast, the 4-hydroxyl group of G2 of rebaudioside shows only one H-bond with the catalytic residue Asp300.

Moreover, two H-bonds are formed between the 4- and 6-hydroxyl groups of G2 of both compounds with the side chain of Arg195. In addition, hydrophobic contacts are formed between the methylene group of both compounds with Trp58 and Tyr62 besides the ones that are formed between the methyl group at position 20 of stevioside with Leu165 and with Trp59 in case of rebaudioside.

As supporting interactions, the glucose unit (G1) of stevioside forms one H-bond with Ser163 at subsite -3, which is not observed with the other steviol derivatives. Additional interactions are also formed as H-bonds between the 3- and 4-hydroxyl groups of the glucose unit (G3) of stevioside with Tyr151 and Lys200; while the 4-hydroxyl group of G3 of rebaudioside forms H-bonds with Lys200 and His201. Furthermore, the 4-hydroxyl group in the distal glucose unit (G4) of rebaudioside accepts one H-bond from Ala307.

Based on this *in silico* study of steviol compounds, we hypothesize that filling up the large binding site and the orientation closer to the co-crystallized ligand might enhance the inhibitory effect of stevioside better than the other studied steviol compounds. Stevioside can accommodate all the essential interactions required to inhibit HSA compared to steviol and rebaudioside. It fits the binding pocket and interacts with the residues spanning from subsite -3 to +2. Most of the interactions of stevioside are H-bonds that are formed between the hydroxyl groups of G2 with all the catalytic residues at subsites -1, +1. Additionally, the 2-hydroxyl group of G1 donates one H-bond to Ser163 near to subsite -3; while the 3- and 4-hydroxyl groups of G3 accept H-bonds from Lys200 and Tyr151, which overall augment the

efficiency of binding. This may support our proposal that stevioside could mainly be responsible for HSA inhibition in the steviol extract that were tested against HSA.

4.2.3 Verbascoside

Verbascoside is a naturally occurring phenylethanoid glycoside. It is composed of one β -glucopyranose unit holding three fragments: phenylethanoid at C1; 4,6-dideoxy- α -D-glucose unit at C3; and caffeate at C4. It is also called acetoside and present in different plants and exerts biological properties such as antioxidant and anti-inflammatory [162, 163].

This compound has not been reported as α -amylase inhibitor before. Consequently, verbascoside had been biologically tested against α -amylase by our collaboration partner Prof. Dr. Matthias F. Melzig and shows inhibition with IC_{50} in upper micro-molar range.

Here, an *in silico* study was performed for this caffeate derivative, using docking to define the most plausible interactions of the inhibitor with α -amylase. Because verbascoside is a caffeate derivative, we decided to use the crystal structure (PDB code: 4GQQ [25]) of HPA co-crystallized with ethyl caffeate as a reference for this study. This PDB entry represents the only α -amylase X-ray structure to date that is co-crystallized with a caffeate derivative as noncompetitive inhibitor ($K_i = 1.3$ mM [82]).

Docking conformations of verbascoside were explored and compared to the co-crystallized ethyl caffeate. Those poses that demonstrate similar interactions to ethyl caffeate in HPA were selected for further analysis.

Briefly, three molecules of ethyl caffeate interact with three different HPA sites away from the catalytic center by about 20 Å [25]. Among them, two sites (sites **2** and **3**) are adjacent to each other. This inhibition is due to remote effect through disordering four segments with a total of 47 amino acids located within the elongated substrate binding site. Therefore, dislocations of those segments that also include Asp300 (one of catalytic residues in the productive active site) play a major role in the stability of the HPA polypeptide chain and enzymatic inhibition [25]. Using LigandScout 3.1 [114, 115], the co-crystallized ethyl caffeate was extracted from each binding site (PDB: 4GQQ) and the chemical features involved into ligand-enzyme interactions were compiled as 3D pharmacophores to explain the binding modes that are required for the inhibition. After successful re-docking experiments for the co-crystallized ethyl caffeate (Appendix 10.6), verbascoside was docked into each site (sites **1**, **2** and **3**) using the software GOLD 5.1 [127]. The resulted poses were prioritized using the pharmacophoric features of the relevant co-crystallized ethyl caffeate.

In site 1, the aromatic ring of ethyl caffeate inhibitor shows a hydrophobic contact with Val287. The hydroxyl group in *meta*-position donates one H-bond to Gly285 while the *para*-hydroxyl group forms two H-bonds with Asp236 and Lys257. The most plausible binding pose of verbascoside in site 1 that fulfills the above-mentioned pharmacophore was selected (Figure 18). This proposed binding mode is perfectly superimposed on the co-crystallized inhibitor and shows very similar interactions as with the caffeate moiety.

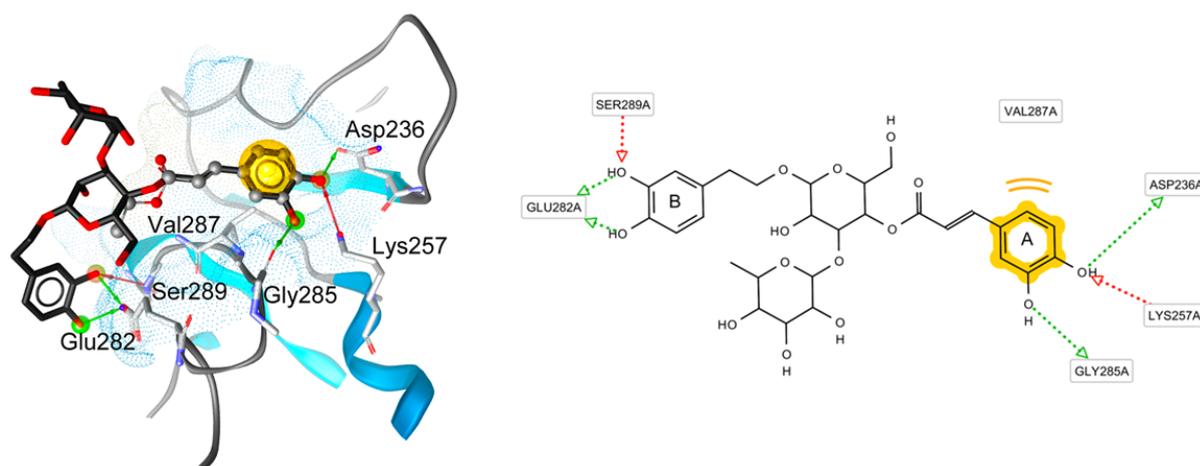


Figure 18: Superposition of ethyl caffeate (gray balls and sticks) and proposed binding mode of verbascoside (black sticks) in HPA (PDB code: 4GQQ) in site 1 in 3D (left). Yellow spheres indicate lipophilic contacts, green and red arrows indicate donating and accepting H-bonds, respectively.

The hydroxyl groups of the caffeate moiety (ring A) of verbascoside forms three H-bonds with Gly285, Lys257 and Asp236. Hydrophobic contacts are also formed between ring A and Val287. Additionally, the hydroxyl groups of the ethyl dihydroxybenzene fragment (ring B) forms three H-bonds with Glu282 and Ser289 that can stabilize the ligand in the pocket. At this site the glucose units do not show any interactions, and can be considered as linkers between main interacting fragments (rings A and B).

In site 2, the ethyl and ethenyl groups of the inhibitor ethyl caffeate form hydrophobic contacts with Leu26. H-bond donors are formed between the hydroxyl groups of the aromatic ring and Asp77 on one side, and between the carbonyl oxygen of the inhibitor and Arg85, on the other side. The most plausible pose fulfilling the caffeate-typical 3D pharmacophore features is illustrated in Figure 19.

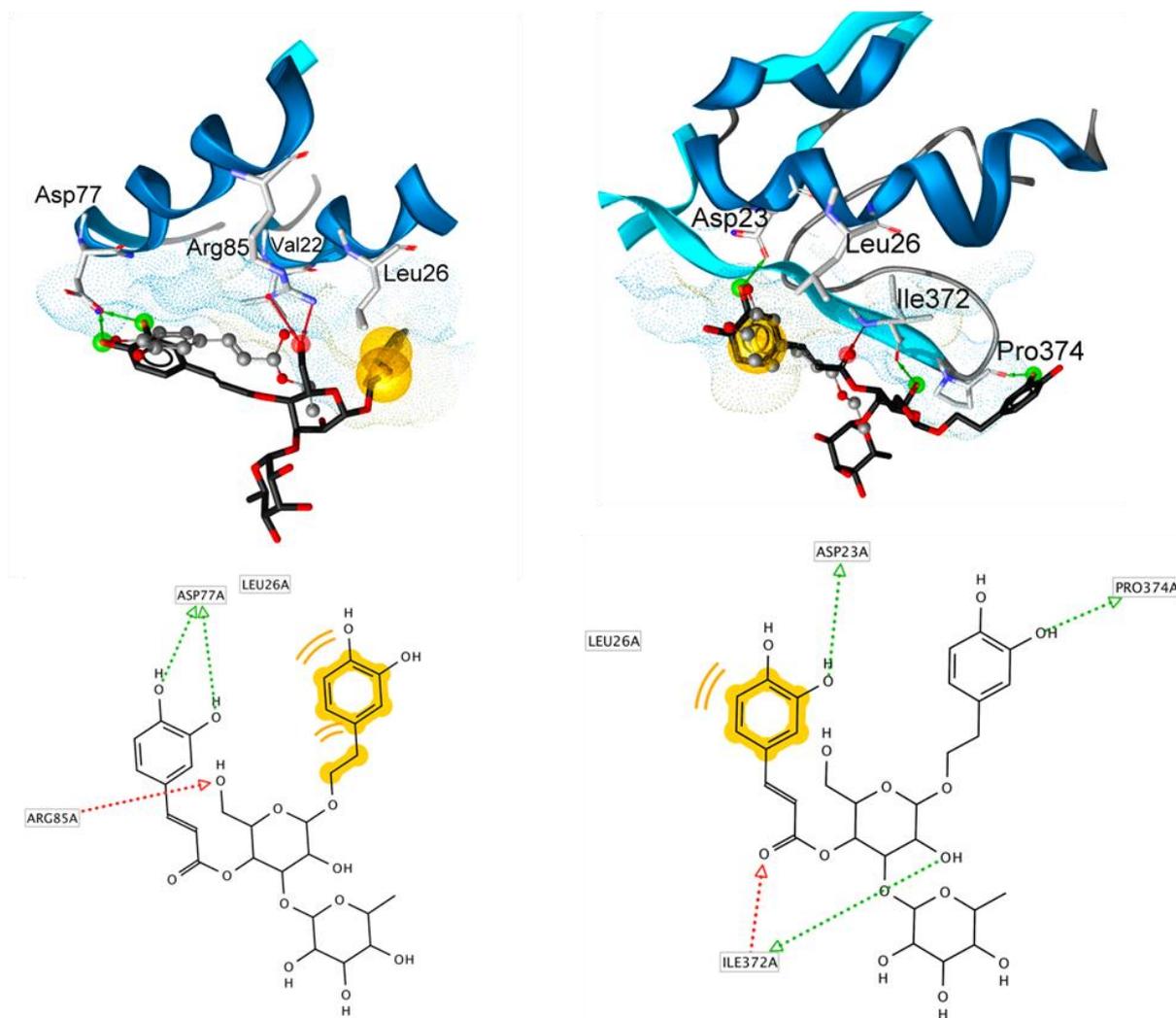


Figure 19: Selected poses of verbascoside in site 2 (left) and site 3 (right) interacted in HPA (PDB: 4GQQ) in 3D (above) and 2D (below). Yellow spheres indicate hydrophobic contacts while red and green arrows indicate H-bonding as acceptors and donors, respectively. Size of amino acid residues depends on their positions in the pocket.

The selected binding conformation of verbascoside in site **2** shows similar interactions to the co-crystallized inhibitor. However, the hydroxymethyl group in C6 of the glucose unit linked to caffeate forms additional H-bond acceptor with Arg85. Since site **2** is close to site **3**, and because verbascoside is larger than ethyl caffeate, the former might protrude into site **3**. Superposition of the docked verbascoside and the co-crystallized ethyl caffeate in site **2** reveal that ring B of verbascoside interacts in a similar fashion as the aromatic ring of the ethyl caffeate in site **3** through hydrophobic contact with Leu26. It can be surmised that verbascoside occupies both sites **2** and **3** and interacts with residues located in both sites.

In site **3**, the aromatic ring of the co-crystallized ethyl caffeate forms hydrophobic contacts with Leu26 and Ile372. While Asp23 forms one H-bond with *meta*-hydroxyl group of the aromatic ring, another H-bond is formed between the nitrogen atom of the backbone of Ile372 and the carbonyl oxygen of ethyl caffeate. The predicted binding conformation of verbascoside was superimposed with the docking reference in site **3**. It shows very similar interactions besides two additional H-bonds between *meta*-hydroxyl group in the phenylethanoid ring of verbascoside and Pro374 on one side, and between the 2-hydroxyl group of the glucose unit linked to ethyl caffeate moiety and the carbonyl oxygen of the backbone of Ile372, (Figure 19).

Docking results show that the presence of hydroxyl groups in caffeate moiety of verbascoside is essential for binding in addition to the hydrophobic contacts that are formed with the benzene ring of the same fragment. Also, the presence of the double bond in caffeate fragment stabilizes the ligand in the triad sites and decreases its flexibility leading to a better inhibitory potency of caffeate derivatives compared to dihydrocaffeate ($IC_{50} > 14$ mM) [75]. Sugar fragments act as linker in verbascoside without showing real important interactions except H-bonding with Arg85 in site **2** and with Ile372 in site **3**. Accordingly, sugar moieties mostly have insignificant effect on the inhibition. This supports our initial hypothesis that designing molecules without carbohydrate moieties could enhance the binding affinity of caffeate derivatives in the triad pockets.

4.3 Structure-Based virtual screening for α -amylase inhibitors

The main goal of performing virtual screening is to discover novel α -amylase inhibitors with drug-like properties. Since the crystal structure of the target is known, structure-based virtual screening approach is mainly employed in this project.

In subsequent chapters, an initial virtual screening experiment based on a shared-feature pharmacophore will be explained in chapter 4.3.1. A subsequent refined structure-based virtual screening using different 3D pharmacophores will be reported in chapter 4.3.8. Biological experiments for hits selected from both virtual screening approaches will be given in chapters 4.3.2 and 4.3.9. Binding mode analyses of discovered α -amylase inhibitors shall be elucidated in chapters 4.3.4 and 4.3.10. Additionally, an analogue search for the most promising α -amylase inhibitor identified by the initial virtual screening will be described in chapter 4.3.8. While biological testing and SAR analysis shall be given in chapter 4.3.6, docking studies for the identified analogues will be reported in chapter 4.3.7.

4.3.1 Pharmacophore-based virtual screening: Initial screening round

Structure-based virtual screening was performed using a 3D pharmacophore developed from available crystal structures of the HPA, HSA and PPA co-crystallized with carbohydrate-based inhibitors. Selected virtual hits were experimentally tested in-house in collaboration with Prof. Dr. Matthias F. Melzig and in collaboration with Dr. Gyöngyi Gyémánt at the University of Debrecen, Hungary.

In the following sections, 3D pharmacophore development (chapter 4.3.1.1), virtual screening metrics for the developed 3D models (chapter 4.3.1.2), and a workflow of selecting compounds for biological testing (chapter 4.3.1.3) will be described.

4.3.1.1 Development of 3D pharmacophore models

First, all co-crystallized inhibitors available from the PDB were analyzed in order to compile essential features for optimal interaction with the α -amylase catalytic pocket. Due to large interaction patterns involving areas further away from the active site, peptide-based inhibitors were excluded [24, 64]. Crystal structures (3OLD, 3OLE, 3OLG, 3OLI, 3IJ9, 3IJ7, 2QV4, 3BAJ, 3BAY, 1XH2, 1XD0, 1XD1, 1XH0, 1XCX, 1XCW, 1CPU, 1MFV, and 1OSE) of the HPA, HSA and PPA co-crystallized with sugar-based inhibitors deposited in the PDB [5, 6] were investigated using the platform LigandScout 3.1 [114, 115, 159]. Binding modes of co-crystallized inhibitors were analyzed in detail and show that a large number of amino acids in

subsites -4 to +3 in the active site cleft contribute in binding [5, 6]. This large pocket is challenging because the required compounds in this work are small molecules with drug-like properties able to inhibit the enzyme.

Chemical interactions of co-crystallized inhibitors were translated into 3D pharmacophore models (some examples are shown in Figure 20). The resulting pharmacophores were aligned and used as a basis to develop a unique shared-feature 3D query for the initial virtual screening.

Assuming that the cleavage site of α -amylase (at subsites -1 and +1) is a central domain for substrate binding, our starting hypothesis was that blocking this region is essential or at least important for inhibition. Thus, particular attention was put in interactions observed between the enzyme and the investigated ligands in this region.

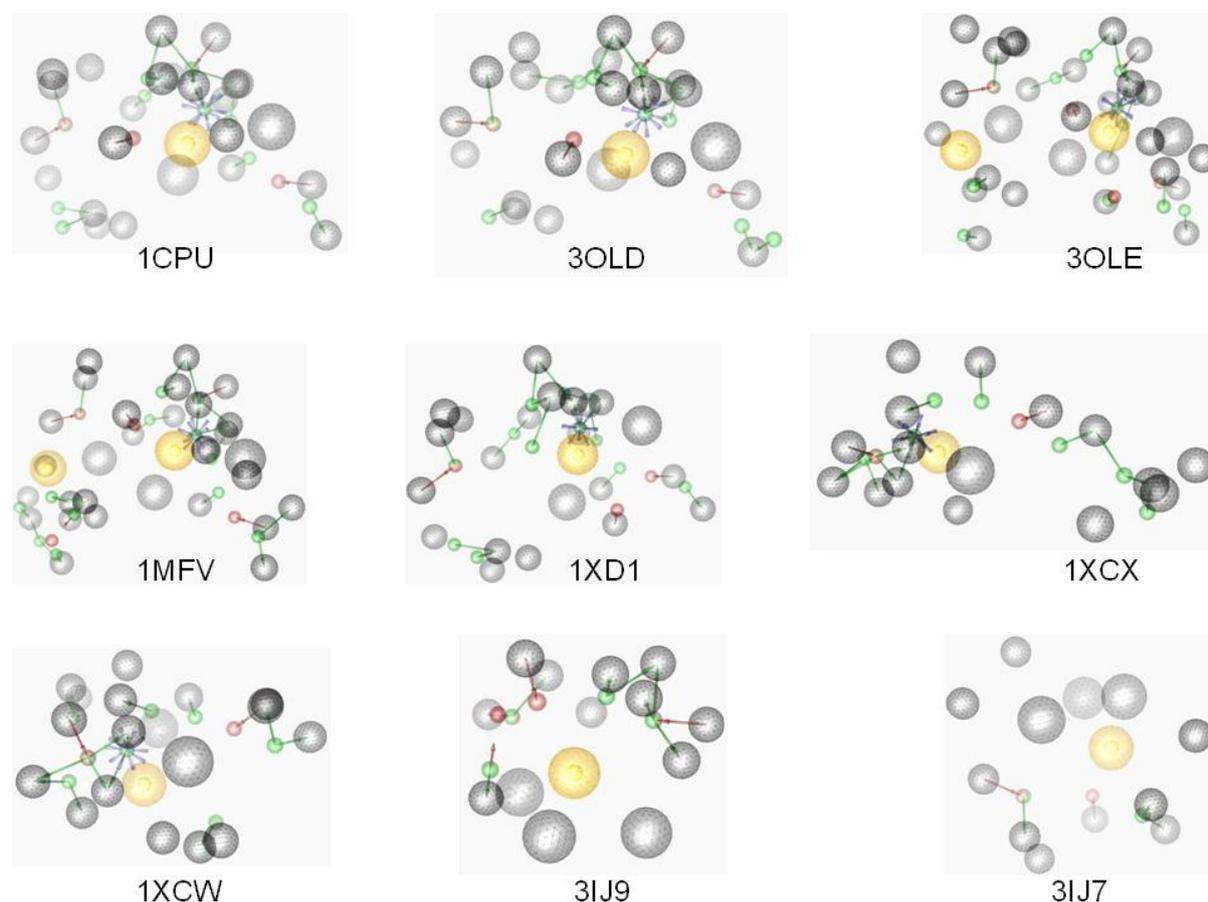


Figure 20: Examples of pharmacophore models generated for the known biologically active co-crystallized inhibitors denoted as their entries in the PDB. Chemical features are colored as: H-bond donors (HBD) green arrows, H-bond acceptors (HBA) red arrows, positive ionizable interactions (PI) blue stars, hydrophobic contacts (HYD) yellow spheres, excluded volumes (Xvols) gray spheres.

During pharmacophore development, the role of a particular water molecule (HOH-764) in the binding of ligands was studied. The software package MOE 2010.10 [155] was used to superpose crystal structures of HPA (codes: 3OLE, 1CXW, 1XCX), HSA (code: 1MFV), and PPA (code: 1OSE) (see Appendix 10.2). Analysis demonstrates that HOH-746 is located nearly at the same position in all investigated macromolecules, which indicates its important role in the binding of inhibitors. The HOH-746 was therefore taken into account in the pharmacophore model.

In order to assess and improve the performance of the developed models, a dataset of 19 active compounds and 55 known inactive compounds was assembled from literature and the ChEMBL database [164], (see Appendices 10.3 and 10.4). Multi-conformational virtual screening of these compounds was carried out and the signal-noise ratio was determined using Receiver Operating Characteristics (ROC) curve described in chapter 3.1.5.

Iterative development and validation yielded two pharmacophore models (Model A and B) with good early enrichment factor values ($EF_{1\%}$ 3.9), which mean that the retrieved hit list contains about four times more inhibitors than a random selection (Figure 21). It is also supported by the high area under the curve ($AUC_{100\%} = 0.80 - 0.81$) that measures the accuracy of the models to discriminate active from inactive compounds. Both models can retrieve 63 % (12 out of 19) of active compounds as true positive compounds, which indicates high sensitivity. From the point of view of selected inactives, model B shows the ability to retrieve only 1.8 % (one out of 55) of inactive compounds compared to 9 % (five out of 55) with model A (see Table 5). Hence, model B displays an excellent efficiency in retrieving maximum number of active compounds with the minimum inactive ligands. This model was selected for the next step of this work.

Table 5: Comparison and validation of the developed 3D pharmacophore models demonstrate high efficiency discrimination between true and false positives

	$EF_{1\%}$	$AUC_{100\%}$	Retrieved actives	Retrieved inactives
Model A	3.9	0.80	12 (63 %)	5 (9 %)
Model B	3.9	0.81	12 (63 %)	1 (1.8 %)

$EF_{1\%}$ means early enrichment factor, and $AUC_{100\%}$ means area under the curve.

The chemical features of the developed models compile mainly ligand-enzyme interactions detected at the center of the catalytic cavity, where Asp197, Glu233 and Asp300 are located.

Model A. This model consists of (i) three optional H-bond donors (HBD) with the nitrogen NE2 of His299, oxygen atom of water-746 and the carbonyl oxygen of the side chain of Glu233 and one essential HBD with the oxygen OE2 from the side chain of Glu233, (ii) one essential hydrophobic contact (HYD) with Ile235 and deeper in the active site with Leu162, (iii) one essential H-bond acceptor (HBA) with the oxygen of water-746, (iv) one optional positive ionizable interaction (PI) with the carboxyl groups of the catalytic triad (Asp197, Glu233 and Asp300), and (v) 25 exclusion volumes (Figure 21).

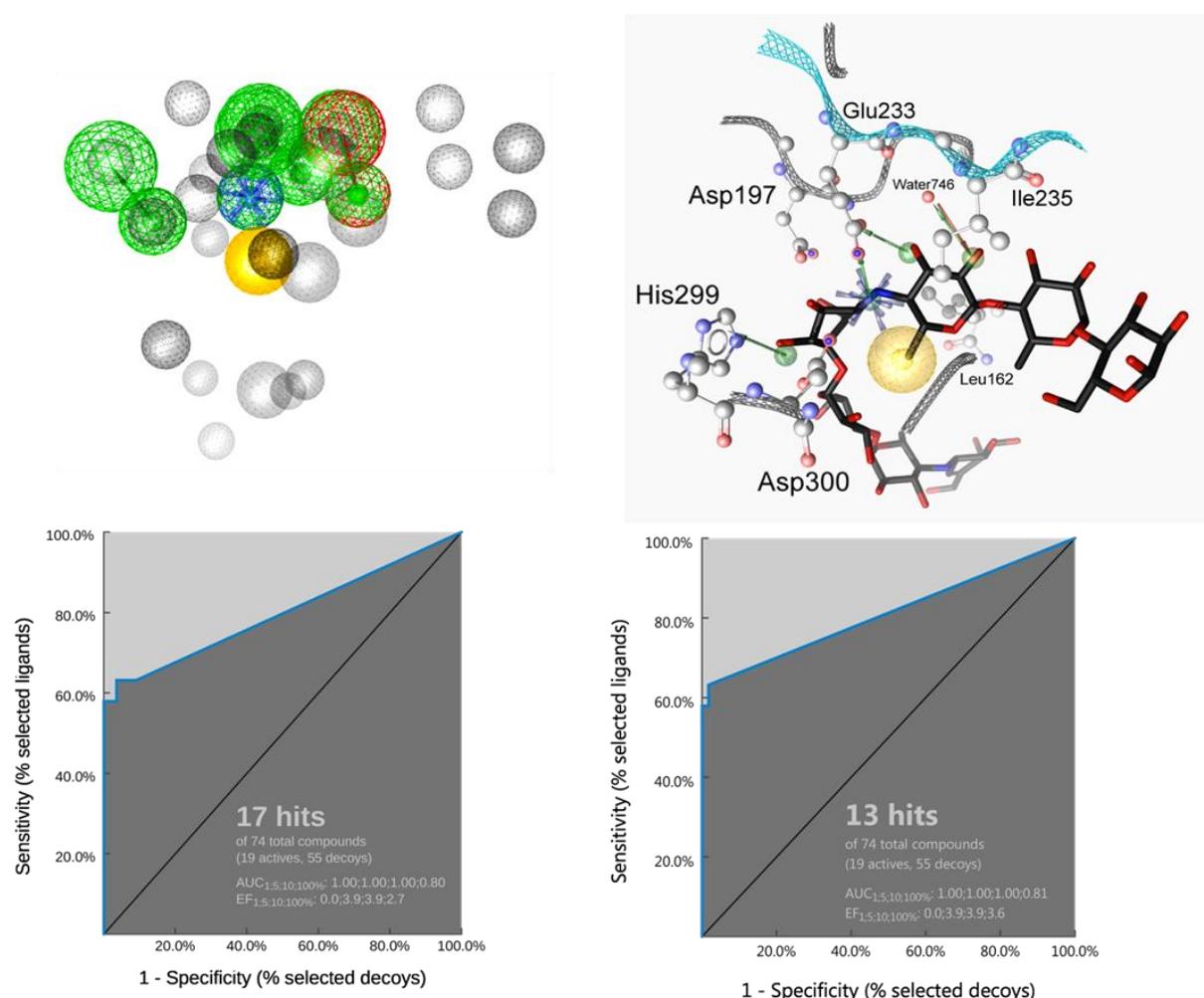


Figure 21: Above: 3D pharmacophore models A and B containing the following chemical features : three H-bond donors (green), one H-bond acceptor (red), one positive ionizable interaction (blue star), one Hydrophobic contact (yellow sphere), and 25 exclusion volumes (gray spheres). Acarviosatin II03 inhibitor in black ($K_i \sim 14$ nM, PDB: 3OLE) is aligned with the developed 3D model. H-Bond acceptor to water 746 is optional in model A. **Below: Receiver Operating Characteristics (ROC) curves:** model A (left) and the final model B (right).

During validation, model A was able to retrieve acarviosine scaffold (valienamine and 4,6-dideoxy- α -D-glucose) of the carbohydrate-based inhibitors with high pharmacophore fit score.

Additionally, five inactive compounds fit the model owing to the presence of hydroxyl groups in benzene rings as observed with rosmarinic acid ($IC_{50} = 1.4$ mM [75]) and epigallocatechin gallate ($IC_{50} = 1.4$ mM [165]) listed in Appendix 10.4.

Model B. This model consists of the same features as model A (Figure 21). The only difference is that the optional HBD to the oxygen atom of water-746 of model A was modified into an essential feature to have model B. Therefore, it was considered as the final model to be used for the next step.

4.3.1.2 Virtual screening metrics for the developed pharmacophore model

The final model was subjected to further validation by screening drugs and biological compounds available from the Derwent World Drug Index 2005 (WDI, www.thomsonreuters.com), MDDR2009 (www.symyx.com), and DrugBank [166] databases. There were three goals for this further validation: (i) to investigate the ability of the model to retrieve known drugs either with the same or different biological activity, (ii) to examine the efficiency of the model by ranking the 12 previously retrieved active compounds when screened altogether with drug databases (iii) to have an idea about the restrictiveness of the model by specifying the total number of the retrieved hits from the WDI2005, MDDR2009 and DrugBank databases. When screening the commercial biological databases alongside with the collected known α -amylase inhibitors, eleven out of 12 known active compounds previously retrieved during the first validation are ranked within the 100 top hits. Up to 17 % of the compounds retrieved and ranked within the first 100 hits belong to trestatin family possessing acarviosine scaffold. This means the developed model can retrieve active compounds and prioritize α -amylase inhibitors among others. This model is suitably considered restrictive because the overall recovered hits from WDI, MDDR and DrugBank were about ~ 1.4 % (1,969 out of 141,233 compounds). The main values from this validation step are summarized in Table 6.

Results of this screening campaign indicate that the 3D model can retrieve known active compounds from the screened databases with outstanding early enrichment factor values. $EF_{1\%}$ values reach up to 63 and good discriminating accuracy expressed as area under the curve (AUC) reaching 0.81, a value very close to the ideal accuracy one.

Table 6: Results of the computational validation of the final model using drugs and biological compounds databases

Validated databases	Size of a database	Retrieved hits from		EF _{1%}	AUC _{100%}
		databases	known actives		
WDI2005	64,255	1,141 (1.77%)	12	58.0	0.81
MDDR2009	72,383	752 (1.03%)	12	63.2	0.81
DrugBank	4,595	76 (1.65%)	12	58.1	0.81
Total	141,233	1,969 (1.39%)			

EF_{1%}: early enrichment factor, AUC_{100%}: area under the curve.

Based on these promising results regarding statistical validation of the final developed model, we decided to use it as a 3D query for virtual screening to identify novel α -amylase inhibitors with a relatively low rate of expected false positive compounds.

4.3.1.3 Overall compound selection work-flow

After validation of the 3D pharmacophore, the model was used as a query for virtual screening of 1,762,189 commercially chemical compounds available from different vendors (Figure 22).

The 3D model picked 5,748 hits, which correspond to 0.32 % of the database entries. Retrieved hits were ranked with pharmacophore fit score in LigandScout 3.1 and about 3,014 compounds were selected for the next step. Then, 2D descriptors were used (i.e. HBD \leq 2-5, HBA \leq 10, Mol.wt. \leq 700, and ClogP \leq 5) to rationally select 2,292 drug-like hits (chapter 7.2.4). In order to reduce the amount of inhibitors to handle prior to biological experiments, two different strategies were applied: virtual docking and structural clustering.

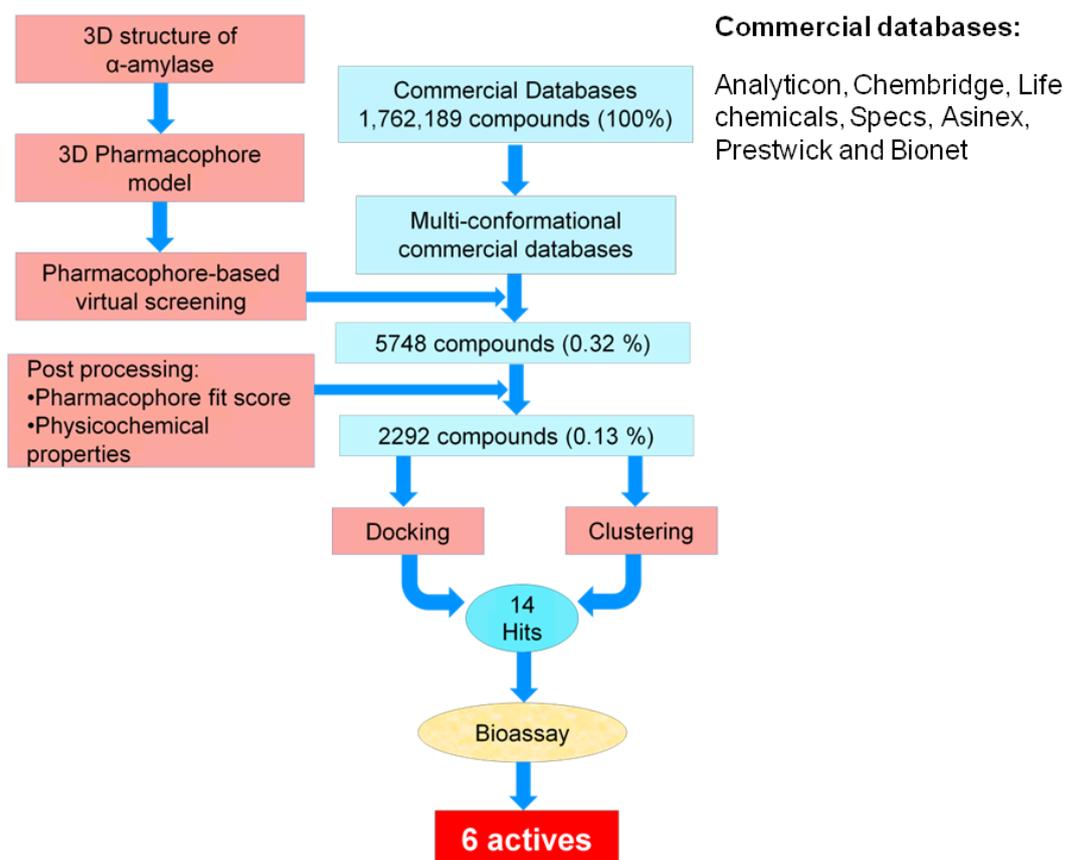


Figure 22: Workflow of the initial virtual screening of commercial chemical databases.

After successful re-docking experiments (chapter 7.5.3, Appendix 10.6), GOLD 5.1 [127] was used to dock the resulted virtual hits (2,292) into the α -amylase active site (PDB code: 3OLE). Using LigandScout 3.1, poses were minimized inside the enzymatic pocket with the force field MMFF94 and then prioritized based on their ability to geometrically fulfill the features compiled in the 3D pharmacophore. Careful visual inspection was conducted through compound conformations with the highest fit scores and the 30 most promising structures were selected for the next step.

All 2,292 virtual hits were additionally subjected to structural clustering with the software JKlustor 5.8.0 [139]. The 17 candidates with the highest structural diversity were selected for further investigation. Based on their commercial availability, 14 virtual hits were purchased for biological testing (Figure 23).

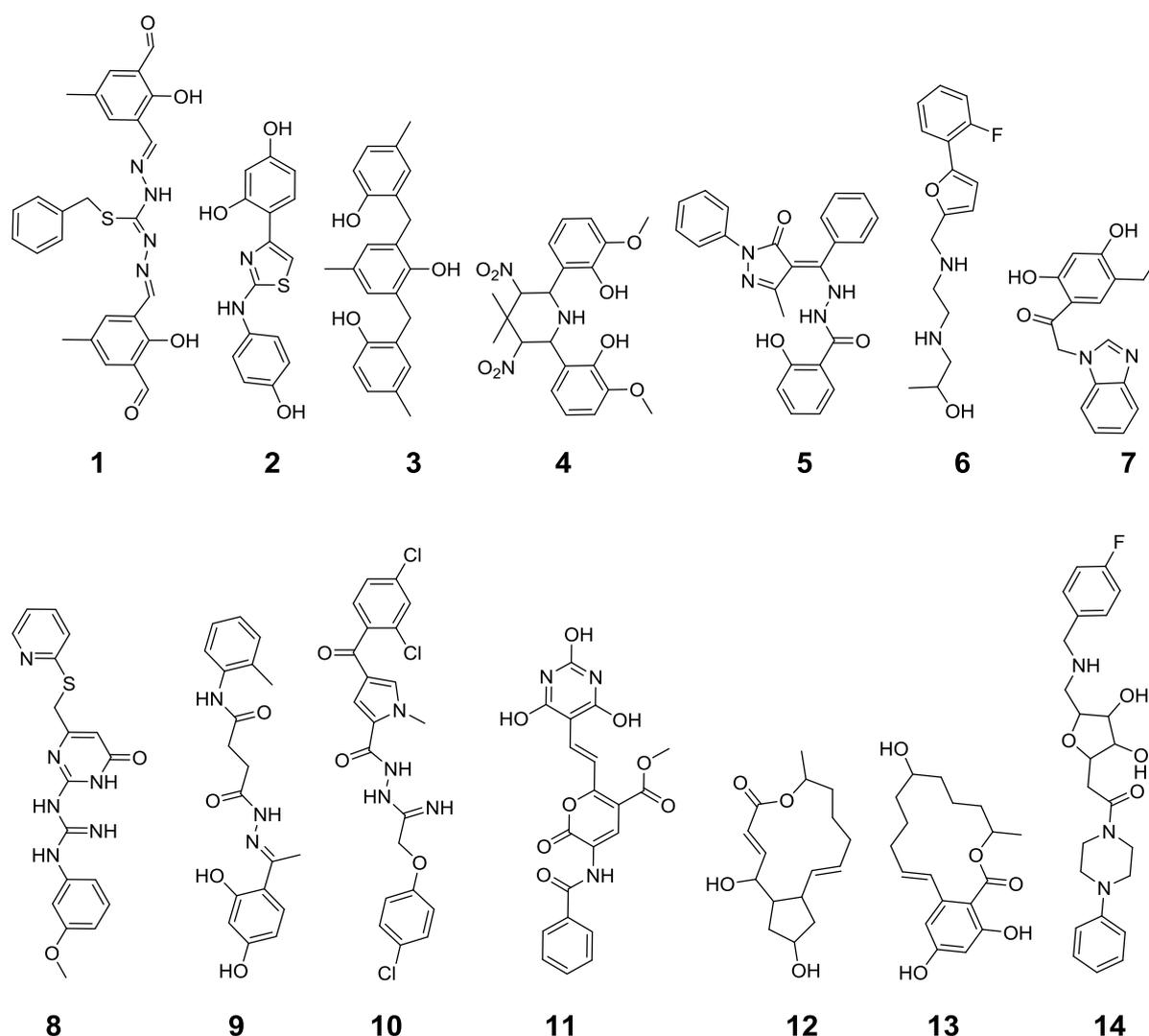


Figure 23: The virtual hits selected for biochemical assay.

4.3.2 Biological Testing

Kinetic assay was carried out with two different substrates, *p*-nitrophenyl- α -D-maltopentaoside (*p*-NPG5) for the first method and 2-chloro-4-nitrophenyl-4-O- β -D-galactopyranosylmaltoside (GalG2CNP) for the second one described in chapters 7.3.1 and 7.3.2, respectively. The purpose of using two assay methods was to analyze the behavior of tested compounds in different conditions. Compound 1 was excluded from the test due to low solubility.

In the first assay, PPA is used to hydrolyze the substrate *p*-NPG5 and release the chromogenic fragment *p*-nitrophenol (*p*-NP) which is continuously detected at 405 nm with respect to time. The residual 13 hits were prepared in the same manner like the positive control acarbose to evaluate their inhibitory effect against the enzyme. Of the tested

candidates, compounds **7**, **8**, and **9** exhibit inhibitory effects with IC_{50} values below 100 μM and comparable to acarbose that shows IC_{50} value of 97.06 μM (Table 7). This corresponds to a hit rate of 23 % (three out of 13). Compound **2** demonstrates weak inhibitory effect with IC_{50} value of 441.5 μM while the rest candidates do not reveal any effect in the first assay. Compound **7** showing the best inhibition with IC_{50} value of 86.6 μM is characterized by the presence of benzimidazole scaffold that is structurally different from carbohydrate-based inhibitors.

Table 7: Biological results and ligand efficiency values for hits selected by the initial virtual screening

Compound	Mol.wt	HA	<i>p</i> -NPG5, PPA		GalG2CNP, HSA	
			IC_{50} (μM)	LE	IC_{50} (μM)	LE
Bound acarbose*	948.92	65	97.06	0.08	0.5	0.13
1	489.56	35	NT		NT	
2	300.33	21	441.5	0.22	138	0.25
3	348.43	26	ND		200	0.19
4	447.43	32	ND		300	0.15
5	413.44	31	ND		ND	
6	294.36	21	ND		1000	0.19
7	296.32	22	86.62	0.25	ND	
8	382.44	27	97.375	0.20	ND	
9	355.38	26	93.58	0.21	ND	
10	479.74	31	ND		NT	
11	425.34	31	ND		NT	
12	280.36	20	ND		NT	
13	320.38	23	ND		NT	
14	444.53	32	ND		NT	

HA stands for number of heavy atoms, LE stands for calculated ligand efficiency, ND means inactive at 500 μM (with *p*-NPG5 and PPA) and at 200 μM (with GALG2CNP and HSA), NT means not tested due to insolubility.

* Bound acarbose (pseudohexasaccharide) in the active site of PPA (PDB: 1OSE [156]) and HSA (PDB: 1MFV [49]) was taken into account for LE calculations.

Also, compound **8** displays IC_{50} value of 97.4 μM and perfectly matches the value of acarbose. It seems that compound **8** with pyrimidinyl-guanidine scaffold is interacted inside the active site in a very similar pattern as acarbose compared to compound **9** that reveals IC_{50} value of 93.6 μM .

Different from the first assay, GalG2CNP is used as a substrate for HSA using MES buffer at pH 6.0 and the released chromogenic product chloronitrophenol (CNP) is continuously monitored at 400 nm with respect to time. Using this method, the effect of compounds **2** - **9** was evaluated (the other hits were unavailable at the time of the assay).

In this assay, compounds **2**, **3**, and **4**, out of the eight tested compounds reveal inhibition against HSA with $IC_{50} \leq 300 \mu\text{M}$ compared to acarbose that shows IC_{50} value of $0.5 \mu\text{M}$ (Table 7). This corresponds to a hit rate of 37.5 % (three out of eight).

Compounds **2** and **3** show competitive inhibitors with IC_{50} values $138 \mu\text{M}$ and $200 \mu\text{M}$, respectively (Figure 24).

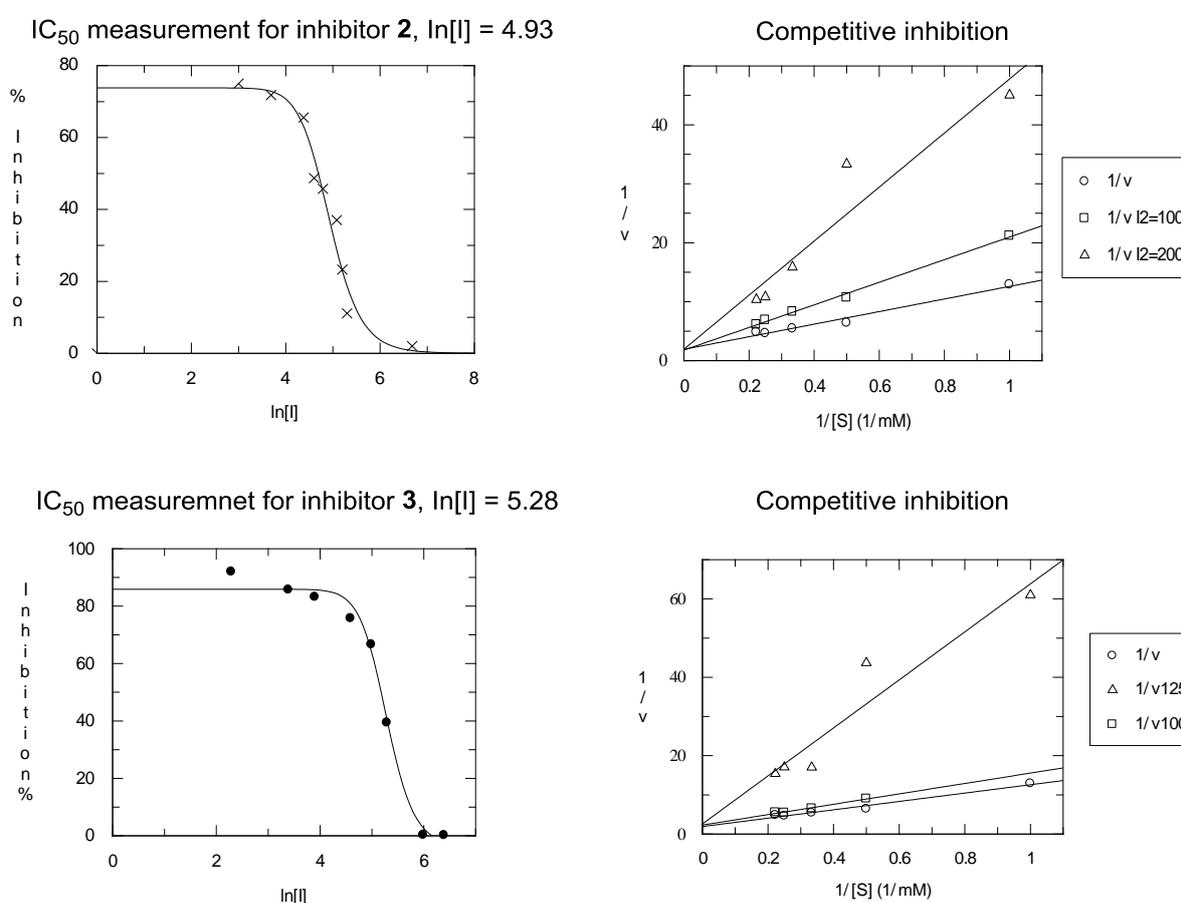


Figure 24: Concentration-effect curves (left) and investigation of inhibition types (right) of compounds **2 (above) and **3** (below) using GalG2CNP substrate.**

These inhibitors are characterized by three-ring-backbones. While compound **2** possesses *N*,4-diphenylthiazole-2-amine scaffold, compound **4** retains a dinitro-piperidine fragment and

both compounds are substituted with hydroxyphenyl rings. Compound **3** is composed of tri-*p*-cresol rings.

Although less potent than the reference inhibitor acarbose, these compounds bear a promising ligand efficiency (LE) due to their smaller size. Defined as the ratio between potency and their number of heavy atoms (HA), LE can be determined using the equation $LE = (1.37/HA) \times pIC_{50}$ [128, 167, 168]. The active compounds identified by virtual screening show higher LE values compared to the bound acarbose, more details will be discussed in chapter 5.2.3. These small molecules are therefore considered promising novel lead compounds and can be used as a structural query to develop potent α -amylase inhibitors.

Interestingly, compound **2** was published by Heng *et al.* in 2009 as allosteric inhibitor ($IC_{50} = 11 \mu M$) against fructose 1,6-bisphosphatase (FBPase) [169]. This enzyme plays a role in the production of glucose (gluconeogenesis) mainly in liver and kidney. Because of its small structure and its ability to inhibit two enzymes (α -amylase and FBase) involved in glucose production, compound **2** is a promising lead molecule that can be used to develop anti-diabetic agents.

During the development of this work, a natural phloroglucinol derivative (DDBT) was reported in 2012 as α -amylase inhibitor ($IC_{50} = 8.56 \mu M$) with a very analogous scaffold to compound **3**. This similarity confirms the rational of our strategy and the strength of our computer-aided methodology.

Performing more than one assay with different conditions (substrate type, enzyme origin, medium conditions and others) to test the same compound can reveal different results evoking a difficult comparison as explained in chapter 3.2. This case was noticed when using two assays in two different Labs to test virtual hits explained before. However, molecular modeling was carried out trying to elucidate differences in IC_{50} values of these tested compounds.

4.3.3 Attempt to rationalize differences in IC_{50} values

To explain variations in IC_{50} values and the inhibitory effect of compounds tested with two assays, molecular modeling tools were used. Therefore, compounds **2**, **3**, **4**, **6**, **7**, **8**, and **9** were investigated here while others not tested in second assay (using GalG2CNP substrate) were excluded.

All investigated compounds, the two substrates (*p*-NPG5 and GalG2CNP) and acarbose were docked into α -amylase active site (PDB: 3OLE) using the software GOLD 5.1 and the most plausible binding poses were selected as described in chapter 7.4. Subsequently, the

selected conformations were used for the calculation of the Gaussian shape similarity score. Calculations using Gaussian functions represent an established strategy for molecular similarity assessment [137]. Hence, the steric overlap of the substrates and the tested compounds could be calculated using Gaussian shape similarity score. This overlap is quantified into a value ranging between zero and one. In LigandScout 3.1, Gaussian shape similarity score is an analytical measurement for an overlap of Gaussian function representations of the molecular volume of a ligand which is aligned to a reference compound [159]. Our assumption was that Gaussian score could be used to explain the competitiveness of each tested compound with each substrate in the active site. In other words, the steric similarity can indicate how well the compound competes with the used substrate and might therefore be used to explain the differences in activity in the different assays.

The core molecule of the PDB code 3OLE was replaced in LigandScout 3.1 by each substrate as a reference for calculations. Then, the selected docking poses were inserted into the LigandScout 3.1 structure-based perspective to calculate their scores. Details are summarized in Appendix 10.5.

Analyzing the docking poses inside the active site shows that the substrate *p*-NPG5 demonstrates better orientation and overlapping than GalG2CNP.

It can fit into the glycone (subsites -3, -2, -1 in the left region) and the aglycone (subsites +1, +2, +3 in the right region) binding sides. GalG2CNP relatively shows good orientation in the active site and is more directed to the right side, whereas acarbose is directed in both sides of the active site from subsites -3 to +2.

The Gaussian scores of acarbose are higher than the ones of investigated compounds. This is not surprising because the carbohydrate nature of bound acarbose is highly similar to the substrates while the tested compounds are small non-carbohydrate molecules. Compounds **3**, **4**, **6**, **8** and **9** referred to *p*-NPG5 substrate show higher scores than the ones that referred to GalG2CNP (Figure 25). This could mean that the probability of the tested compounds to compete with *p*-NPG5 substrate in the active site is higher compared to GalG2CNP. This is supported by the findings of the Gaussian scores for the bound acarbose that shows values of 0.59 (referred to *p*-NPG5) and 0.34 (referred to GalG2CNP).

Additionally, bound acarbose (molecular weight 948.92) is similar in size to *p*-NPG5 (molecular weight 950) compared to GalG2CNP (molecular weight 660). Hence, the latter shows less competition in the active site. Consequently, a higher concentration of the bound acarbose is required to win the competition over *p*-NPG5 and fit in the pocket with almost full occupancy in the active site. Thus, the bound acarbose reveals a high IC₅₀ value of 97.06 μM

in the first assay (with *p*-NPG5) compared to the low IC₅₀ value of 0.5 μM in the second assay (with GalG2CNP).

When using GalG2CNP, investigated compounds show lower scores compared to the ones with *p*-NPG5. Only compound **7** shows higher score with GalG2CNP compared to *p*-NPG5. Even the low scores with GalG2CNP, compounds **2**, **3**, **4**, and **6** show inhibition properties. This could relatively indicate weak competition between GalG2CNP and those compounds. Thus, analysis of Gaussian scores reveals the following findings: (i) Compounds **2**, **8** and **9** do not fulfill the hypothesis that higher score could mean higher competition and then better inhibition (lower IC₅₀ value). Compound **2** shows identical scores with both substrates but lower IC₅₀ with GalG2CNP. Compounds **8** and **9** reveal lower scores with GalG2CNP but without any activity. (ii) Our hypothesis could be theoretically applicable for compounds **3**, **4**, **6** and **7**. Compounds **3**, **4** and **6** demonstrate lower scores with GalG2CNP and inhibitory activity (200 μM, ~ 300 μM, and 1 mM), which means higher concentrations of these inhibitors might be required to show activity with *p*-NPG5. However, increasing the concentrations of these compounds was not applicable during assay with *p*-NPG5 and precipitated in the well-plates.

Compound **7** shows higher similarity to GalG2CNP, which means a higher concentration could be required to show activity with GalG2CNP. It might be applicable if the concentration was more than 200 μM, the maximum used concentration during assay.

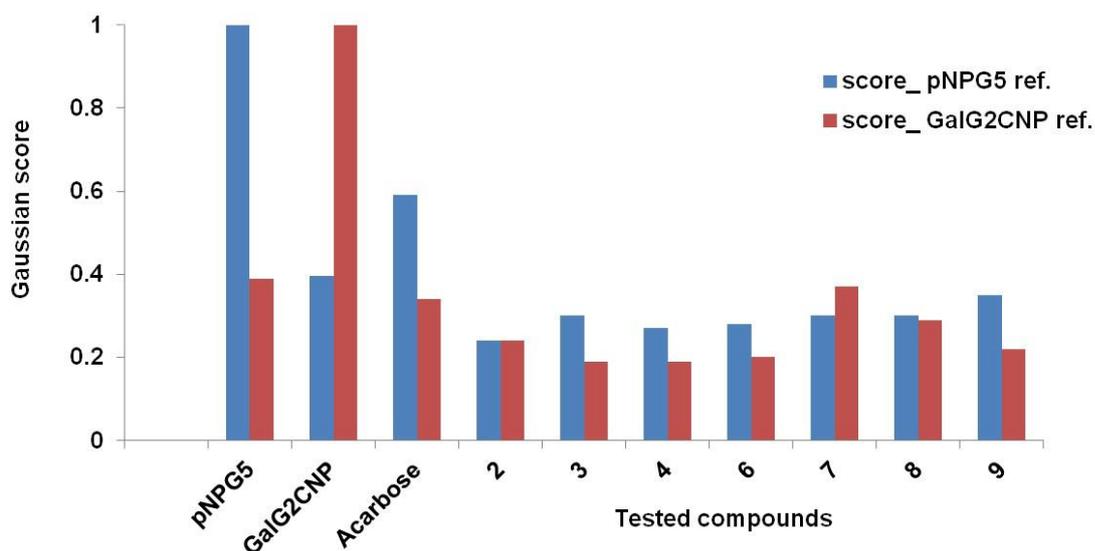


Figure 25: Diagram shows different values of Gaussian score calculated with LigandScout 3.1 using *p*-NPG5 and GalG2CNP as score references for acarbose and compounds showing α -amylase inhibition. Blue columns represent scores referred to *p*-NPG5 while intense red colors indicate scores referred to GalG2CNP.

In summary, the proposed reasons behind the differences in activity with the two assay methods could be due to the following points: Firstly, the longer substrate (*p*-NPG5) used in the first assay shows more interactions and hence better binding inside the active site than GalG2CNP. Hydrolysis of the substrate could occur more than once and the possibility of rearrangement and condensation could occur based on substrate concentration [170-174]. This leads to slow hydrolysis process of *p*-NPG5 and could explain the high amount of the PPA enzyme (5.680 μ M) and *p*-NPG5 substrate (25 mM) in the first assay compared to the amount of HSA enzyme (2 nM) and GalG2CNP (0.75 - 4 mM) in the second assay (chapter 7.3). Secondly, the substrate GalG2CNP used in the second assay was short and very sensitive to the enzyme which means the conversion rate and the affinity of GalG2CNP are high.

While our steric similarity approach could explain a few aspects of differing IC_{50} values, it could not be used for rationalizing all activity differences, mainly due to the complex kinetic processes that cannot be represented by a simple steric overlap. Consequently, comparing results from two different assays is considered challenging and many factors should be taken into account such as types of substrate, enzyme and concentrations as well as conditions used in the assay (chapter 3.2).

4.3.4 Docking studies with discovered α -amylase inhibitors

Re-docking the active compounds in the binding site of α -amylase identified in both assays was conducted to depict their hypothesized binding modes and compare them to the co-crystallized inhibitor acarviostatin 0113 ($K_i \sim 14$ nM, PDB: 3OLE [36]).

Compounds **7**, **8** and **9** show inhibition properties in the kinetic assay conducted with *p*-NPG5 substrate for PPA.

Docking was carried out with GOLD 5.1 as described in chapter 7.5.3. The most plausible poses were selected and prioritized based on the fulfillment of chemical features of the developed 3D model. Subsequently, a 3D pharmacophore was created for each selected pose to rationalize its activity as illustrated in Figure 26.

Superposition of compounds **7**, **8** and **9** with the docking reference acarviostatin 1103 reveals similar binding modes. They all occupy subsites -1 and +1 with interactions favorable for inhibition.

Investigating the selected poses shows that hydroxyl groups of compounds **7** and **9** and the NH-group in the pyrimidin-4-(3H)-one fragment of compound **8** donate H-bonds to Glu233. Asp197 forms H-bonds, on one side, with the hydroxyl group of compound **7** and, on the

other side, with the hydrazine fragment of compound **9**. The methyl group attached to the hydrazine fragment of compound **9** shows unfavorable orientation because it is located very near to the carboxyl group of Asp300 and impeded the formation of H-bond. However, this orientation helps the hydrazine group to form H-bond with Asp197. The basic nature of the guanidinium group allows compound **8** to better orient for interaction with the catalytic triad (Asp197, Glu233, and Asp300) and Tyr62 through H-bonds and positive ionizable interactions compared to compounds **7** and **9**. While the methoxy group of compound **8** accepts one H-bond from His305, the imidazole moiety of compound **7** and the hydroxyl group of compound **9** accept H-bonds from Lys200. Hydrophobic contacts are formed to stabilize the ligands in the enzymatic pocket.

Hydrophobic areas between lipophilic residues Leu162 and Ile235 with the resorcinol and the nearby methyl groups of compound **9** are favorable and similar to the ones with acarviostatin II03 in subsite +1. Although the hydrophobic nature could retard the interaction with Asp300, hydroxyl groups of **9** are directed toward Glu233 and Lys200 forming H-bonds. In compound **7**, the ethyl group forms hydrophobic contacts with Tyr62 and Trp58 and sterically hinders the interaction with Asp300 and His305 compared to the methoxy group of compound **8** that forms one H-bond at that location. While a hydrophobic contact is observed between compound **7** and Tyr151, compound **9** forms a lipophilic contact with Trp59.

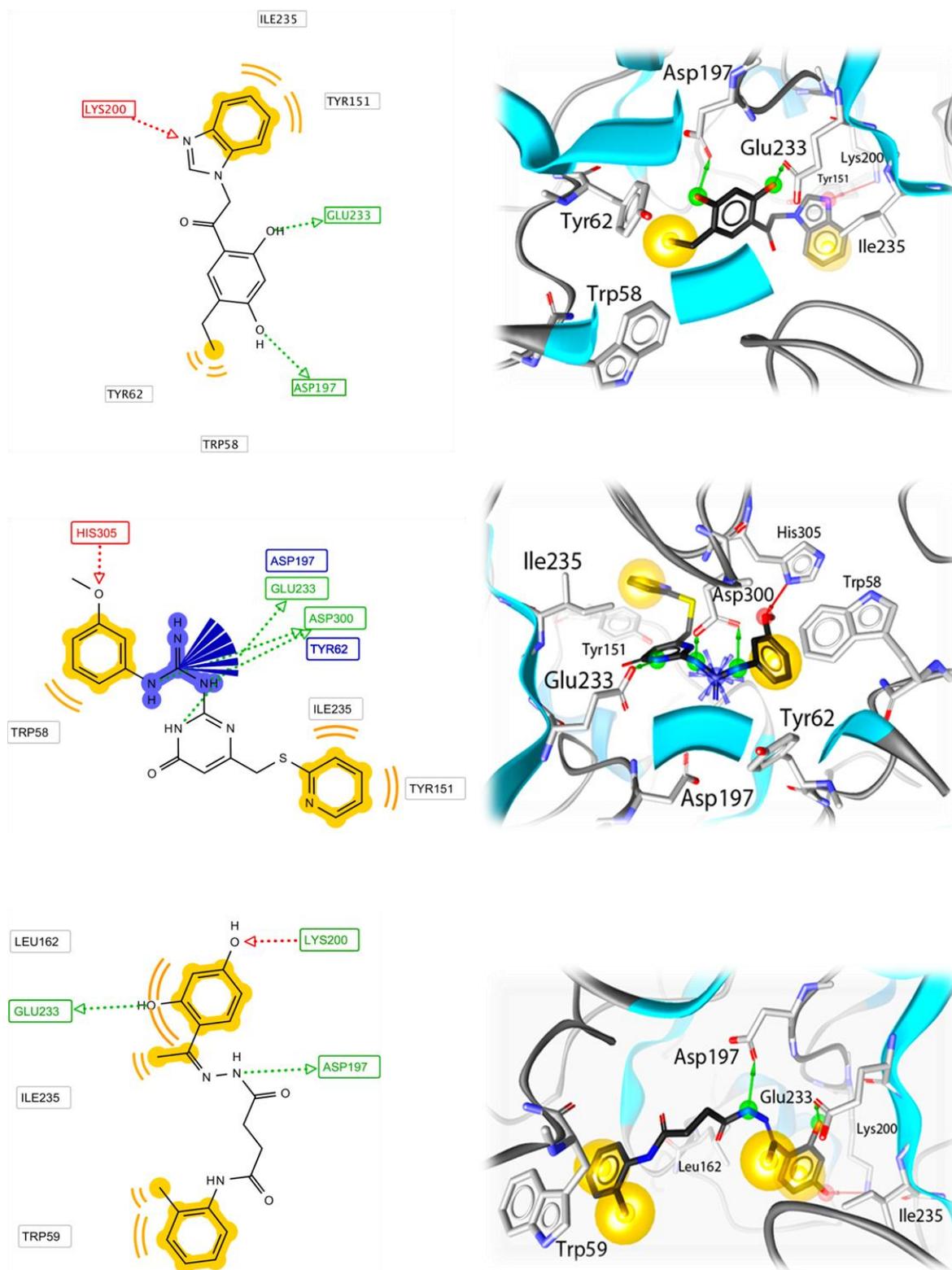


Figure 26: 2D (left) and 3D (right) depiction for the predicted conformations for compounds **7** (above), **8** (middle), and **9** (below) in HPA (PDB: 3OLE). Yellow spheres represent hydrophobic contacts; arrows illustrate H-bonds (green for donor and red for acceptor); Blue stars indicate positive ionizable interactions. Projection and labels of residues is based on their depth in the pocket.

Consequently, the hydroxyl groups of **7** and **9** and the guanidinium group of **8** play crucial roles for binding with catalytic residues. H-bonding and positive ionizable interactions are favorable for α -amylase inhibition.

In contrast, compounds **2**, **3**, **4** and **6**, show inhibitory activity against α -amylase (HSA) when using GalG2CNP as a substrate. Molecular docking was performed similarly as conducted with previously identified inhibitors. Investigating compounds **2**, **3**, and **4** as the most potent inhibitors identified by this assay reveals that three rings comprise their structural backbone. The proper accommodation in subsites -1 and +1 of compound **2** explains the inhibitory effect of this small molecule. The three ring-backbone of this inhibitor fully occupies subsites -2, -1, and +1 and partially +2 with similar orientation of acarviosine scaffold of acarviostatatin II03 (Figure 27).

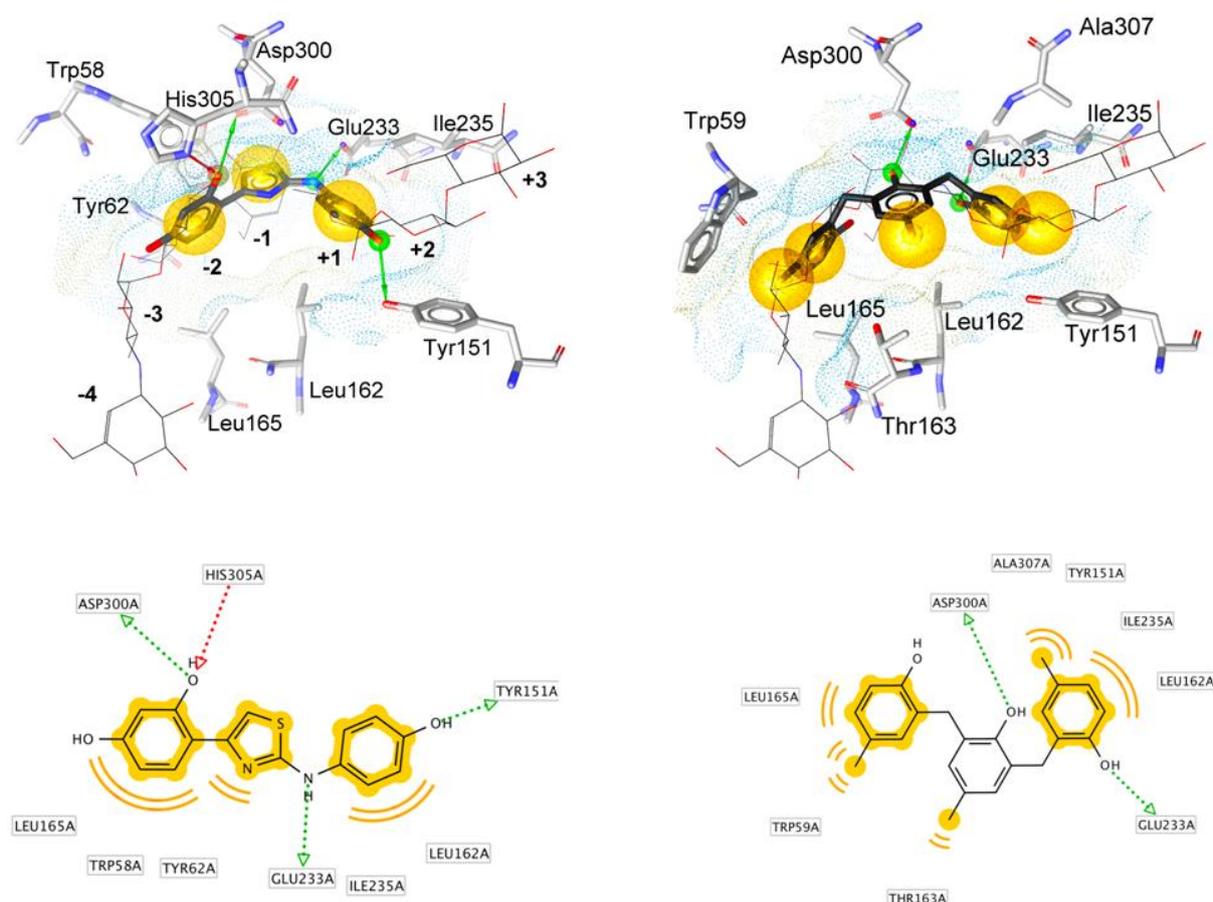


Figure 27: Plausible binding modes for compounds 2 (left) and 3 (right) as thick black sticks in 3D (above) superposed to the co-crystallized acarviostatatin II03 (PDB: 3OLE) as thin lines and in 2D (below).

This interaction with the catalytic core of the cavity is similar to the one of acarviosine. This essential scaffold of acarbose and other trestatin family inhibitors interacts with the central subsites -1 and +1. Our study shows that one hydroxyl group of resorcinol of compound **2** can form two H-bonds with His305 and the catalytic residue Asp300, in subsite -1. The side chain of Glu233, another catalytic residue in subsite +1, can form one H-bond with the amino group linked to the thiazole ring. An additional H-bond can be formed between the phenolic hydroxyl group and Tyr151 near subsite +2.

Interestingly, in the identified binding mode of compound **2**, the thiazole moiety occupies subsite -1, just like valienamine, and stabilizes the inhibitor via hydrophobic contact with the side chain of Tyr62. Additional lipophilic contacts are formed, on one side, between the resorcinol ring and Leu165 and Trp58 near subsite -2, and on the other side, between the phenolic ring and Ile235 and Leu162 near subsite +1. These interactions lead to a good stabilization of compound **2** in the catalytic core of α -amylase.

Similarly, compound **3** occupies the catalytic core (subsites -1 and +1) and protrudes to the nearby subsites in our docking study. Superposition with acarviostatin II03 shows, again, a similar binding mode. Predicted conformation of compound **3** suggests that two H-bonds are formed between the catalytic residues Asp300 and Glu233 and the hydroxyl groups of two *p*-cresol rings. The methyl group of the central *p*-cresol fragment of compound **3** can point to subsite -3 and form hydrophobic contact with the side chain of Thr163, stabilizing the inhibitor in the pocket. The other two *p*-cresol rings can form hydrophobic contacts in subsite -2 with Leu165 and Trp59, subsite +1 with Leu162 and Ile235, and near subsite +2 with Tyr151.

Although compound **3** is spanning from subsites -2 to +2, it shows weaker inhibitory potency ($IC_{50} = 200 \mu\text{M}$) than compound **2** ($IC_{50} = 138 \mu\text{M}$). This could be due either to (i) H-bonds playing a major role in inhibition and fewer could be detected with **3** compared to the binding mode of compound **2**, and/or (ii) the thiazole ring of compound **2** plays an important role for ligand stabilization in subsite -1 and is better oriented than *p*-cresol of compound **3** (Figure 27).

Compound **4** shows lower inhibitory potency ($IC_{50} \sim 300 \mu\text{M}$) even though it occupies subsites -3 to +2 in the docking study. The catalytic residue Asp300 can form a positive charge interaction with the amino group of the piperidine moiety, contributing to the stabilization of this ligand in the cavity. This inhibitor is stabilized inside the pocket by forming hydrophobic contact with lipophilic residues Thr163, Tyr62, Leu165, Leu162, and Ile235. Nevertheless, the presence of the nitro groups seems to affect the inhibitory potency. The

lack of adequate interactions with catalytic residues plays a crucial role in the weak inhibitory effect of compound **4**.

Surprisingly, compound **6** shows inhibitory activity and it might be due to the positively ionizable interaction between the amino groups with the catalytic triad. However, the flexibility of compound **6** could make these interactions less stable in the active site and hence it relatively shows weaker effect than compound **2**.

Superposition of the inactive compound **5** with the newly identified active compound **2** shows that even the phenolic groups in both ligands occupying subsite +1, the pyrazolone moiety of compound **5** is not located in subsite -1 compared to the second phenolic ring of compound **2**. This might suggest two possibilities: (i) pyrazolone fragment is not close enough to occupy subsite -1 or (ii) this fragment could be unfavorable for α -amylase inhibition. Additionally, the presence of benzene ring linked to pyrazolone moiety could augment the inactivity of compound **5** compared to thiazole ring in compound **2**.

4.3.5 Similarity search for the best competitive inhibitor in the initial screening

The most promising competitive α -amylase inhibitor identified in the previous initial virtual screening in both assays is compound **2**. This small fragment was used as a query, on one side, to find analogues possessing potential activity, and on the other side, to gain more information about the structure activity relationship. Detailed procedures will be described in the experimental section in chapter 7.6. This approach was conducted as 2D similarity search from an in-house chemical database obtained from commercial vendors for choosing 12 hits for biological testing.

4.3.6 Biological assay and SAR for analogues of the best competitive inhibitor

The previously described kinetic assay with GalG2CNP (chapter 7.3.2) was used for analogues of compound **2** that were selected from 2D similarity search (chapter 7.6). Interestingly, all the 12 compounds show activity against α -amylase (HSA) with IC_{50} values ranging from 50 μ M to 4 mM (Table 8).

Compound **2** with *N*,4-diphenylthiazole-2-amine as 6-5-6-membered rings was investigated using structurally related analogues. Compounds **22** - **27** were chosen to explore substitution effect on potency in ring A (Figure 28). Compounds **28** - **33** were selected with persistent 6-5-6-membered-rings pattern holding interesting scaffold types.

Replacement of the hydroxyl group in *para*-position of compound **2** (IC_{50} = 138 μ M) with the carboxyl group causes 5-fold decrease in inhibition as observed with compounds **22** and **23**. While substitution with *p*-sulfamoyl group tolerates the potency for compound **25** (IC_{50} = 165 μ M), *p*-dimethylamino group enhances the α -amylase inhibition of compound **26** (IC_{50} ~ 100). In contrast, substitution in *meta*-position by a hydroxyl group (compound **23**, IC_{50} = 642 μ M) or a carboxyl group (compound **24**, IC_{50} = 423 μ M) decreases the potency compared to compound **2**.

Indeed, the presence of both hydroxyl group in *meta*- and carboxyl group in *para*-positions or vice versa strongly diminishes the inhibitory potency (3- to 5-fold decrease) and hence their presence at the same time in ring A is unfavorable for α -amylase inhibition as observed with compounds **23** and **24**. The reason could be the formation of intramolecular H-bonds between hydroxyl and carboxyl groups and the resulting unavailability for potential interaction with residues in the binding site.

However, the presence of the carboxyl group in *meta*-position (compound **24**) seems to give better IC_{50} value than in *para*-position (compounds **22** and **23**).

Substitution in *meta*-position with an acetyl group interestingly enhances the inhibition of compound **27** ($IC_{50} \sim 100 \mu\text{M}$) compared to compound **2**.

Thus, substitution in ring A is clearly affect on the potency of the *N*,4-diphenylthiazole-2-amine scaffold. *Meta*-substitution with acetyl group (compound **27**), or *para*-substitution with hydroxyl group (compound **2**), dimethylamino-group (compound **26**) or sulfamoyl-group (compound **25**) are favorable for α -amylase inhibition. *Para*-substitution with carboxyl group appears unfavorable for inhibitory potency. An overview for SAR is given in Figure 28.

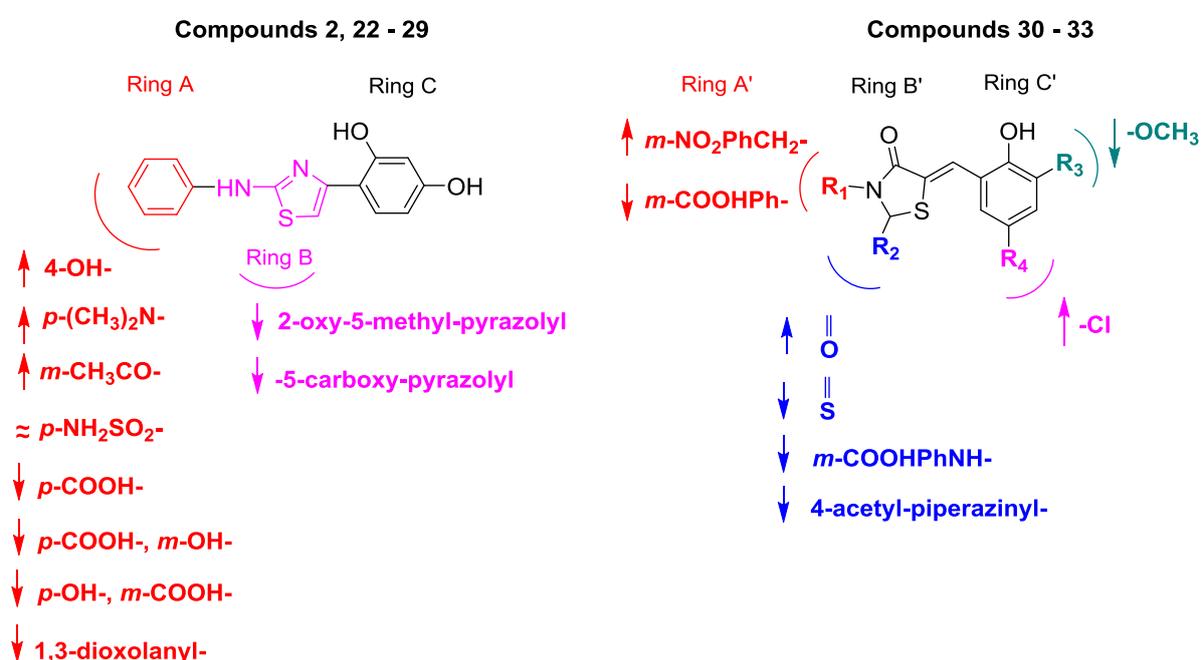


Figure 28: An overview for SAR for analogues of compound **2** against α -amylase.

Substitution of the ring B (thiazole ring) with 1*H*-pyrazole leads drastically to 29-fold decrease in the inhibition as observed with compounds **28** and **29** ($IC_{50} \sim 4 \text{ mM}$) and completely inactive compound **5** (from the initial virtual screening). This shows the importance of the 5-membered heterocyclic thiazole ring for α -amylase inhibition.

Compounds **30** - **33** are also 6-5-6-membered rings that share a scaffold called 5-benzylidenethiazolidin-4-one. Among compound **2** analogues, compounds **30** and **31** exert the most potent α -amylase inhibition with IC_{50} values of 50 μM and 62 μM , respectively.

Table 8: IC₅₀ measurements and ligand efficiency values for analogues of compound 2

Inhibitor	Structure	Mol.wt.	HA	IC ₅₀		LE	Comment
				µg/ml	µM		
22		328.24	23	223	680	0.19	-
23		344.34	24	221	642	0.18	-
24		344.34	24	144	423	0.19	-
25		363.41	24	60	165	0.22	-
26		327.4	23	33	~ 100	0.24	competitive
27		326.37	23	33	~ 100	0.24	mixed inhibition
28		326.3	24		~ 4000	0.14	-
29		368.34	27		~ 4000	0.12	-
30		388.78	26	19	50	0.23	competitive
31		357.4	24	22	62	0.24	
32		448.26	27		260	0.18	
33		331.39	23		~ 700	0.19	

From previous findings with compounds **22** - **24** and **28**, the carboxyl group does not enhance the activity. Thus, chlorine and 2-carbonyl substitutions in compound **30** seem to improve the inhibitory potency as compound **31** that lacks these substitutions and possesses a carboxyl group shows weaker activity. Compound **32** exerts IC_{50} value of 260 μ M weaker activity than compounds **30** and **31** (Table 8, Figure 28). The methoxy and carboxyl groups of compound **32** seem to exert 5-fold decrease in the inhibition compared to compound **30** which lacks these substitutions. Compound **33** shows 14-fold decrease in the inhibitory potency ($IC_{50} \sim 700 \mu$ M) compared to **30** due to the presence of piperazine moiety. This is also supported by the complete inactivity of compound **14** (a hit from the initial virtual screening) that possesses piperazine moiety. Thus, the latter seems unfavorable for α -amylase inhibition. More structural mechanism is given in docking studies in chapter 4.3.7. The dose-response inhibitory potency of the most interesting compounds is shown in Figure 29.

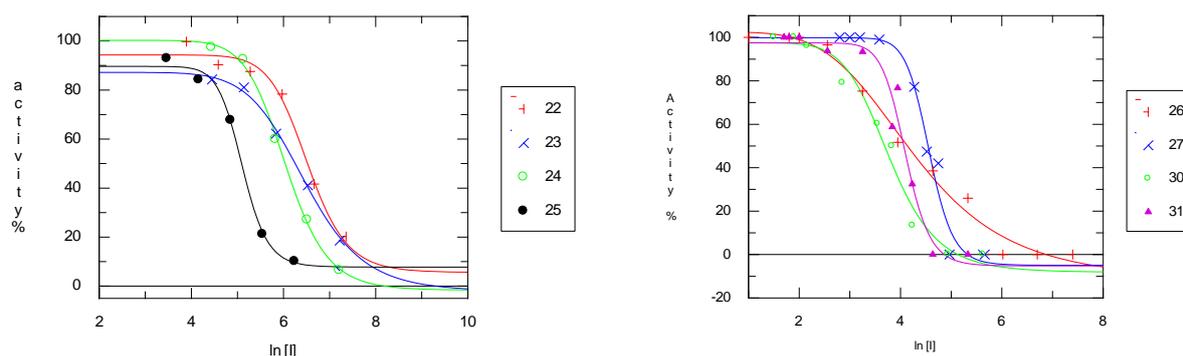


Figure 29: Dose-response curves for compounds 22 - 27, 30 and 31 using GalG2CNP as a substrate for HSA.

Ligand efficiency for the 12 compounds shows values better than the positive control ($LE = 0.13$), taking into account the structure of modified acarbose (pseudohexasaccharide) bound to the active site of HSA (PDB code: 1MFV [49]). Interestingly, compounds **26**, **27**, **30**, and **31** showing the most potent inhibition against α -amylase have the best LE among others (Table 8); more details will be discussed in chapter 5.2.3.

To understand how these compounds accomplish their potency, we investigated their binding conformation inside the α -amylase cavity using molecular docking.

4.3.7 Docking studies of analogues of the best novel competitive inhibitor

Molecular docking was carried out in the PDB code 3OLE using GOLD 5.1 (chapter 7.6). All compounds were docked into the catalytic active site. Plausible binding conformations were prioritized based on the pharmacophoric features of compound **2** and the quality of interactions with residues in the active site. The plausible poses of compounds **25**, **26**, **27**, **30**, and **31** as the most promising analogues of compound **2** are given in Figure 30.

Since compounds **22** - **24** share the same *N*,4-diphenylthiazole-2-amine scaffold present in compound **2**, they can form similar interactions. However, the presence of the strongly electron withdrawing carboxyl group of **22** - **24** near to the carboxylic amino acid residue Glu240 in subsite +2, causes weaker activity than **2**. The carboxyl group of **24** is located in *meta*-position and therefore, shows better inhibitory potency than **22** and **23** which possess *para*-carboxyl groups located far away from Glu240. Compounds **22** - **24**, **32**, and **33** that show weak inhibitory potency are depicted in Appendix 10.7.

Aromatic rings of compound **25** form hydrophobic contacts with Trp58, Tyr62, Ile235, and Leu162. The carbonyl oxygen of sulfamoyl fragment and the amino group linked to thiazole ring of **25** form two H-bonds with Lys200 and Glu233, respectively. The comparison of compounds **22** - **27**, which all have the same scaffold, suggests that hydrophobic contacts formed by *p*-dimethylamino substituent of **26** with Trp59, Thr163, and Leu165 in active subsites -2 and -3 are considered important for activity. Compounds lacking this fragment have lower activity than **26**. Among them, **27** shows activity same as **26** and this could be due to the formation of H-bonds with Asp300, His305 which are not observed with others and an additional H-bond with Lys200.

The comparison between compounds **30** - **33** that share 5-benzylidenethiazolidin-4-one scaffold suggests that H-bond between the 4-carbonyl oxygen of thiazolidin-4-one of **30** and Arg195 in subsite -1 seems important for inhibition, as **31** - **33** do not interact with this amino acid residue. The chlorine substituent of **30** forms a hydrophobic contact with Ile235 in subsite +1 that stabilizes the ligand binding. The carboxylic oxygen of compound **31** shows importance for potency by accepting one H-bond from the side chain of Gln63 in subsite -2, since **32** and **33** lack this interaction and hence show weaker inhibition. Also, **31** forms a hydrophobic contact with Thr163 in subsite -3 in addition to two H-bonds with Tyr151 in subsite +2. The carbonyl oxygen of thiazolidin-4-one of compound **32**, on one side, accepts one H-bond from Gly306 and forms a lipophilic contact between bromine atom and Thr163, on the other side (Appendix 10.7). In subsite +2, one H-bond and a negative ionizable

interaction can be formed between the carboxyl group of **32** and the side chain of Lys200 but the nearby acidic Glu240 might reduce the stability of this ligand in the active site and hence shows lower activity than **31**. Compound **33** shows weak inhibition and forms one H-bond between the carbonyl oxygen and the side chain of Lys200. Hydrophobic contacts are formed between aromatic rings of **33** and Trp58, Leu162, Ile235, and Ala198. The weak inhibitory potency of **33** seems due to the unfavorable orientation of piperazine ring that does not show any interactions with residues in subsite +1.

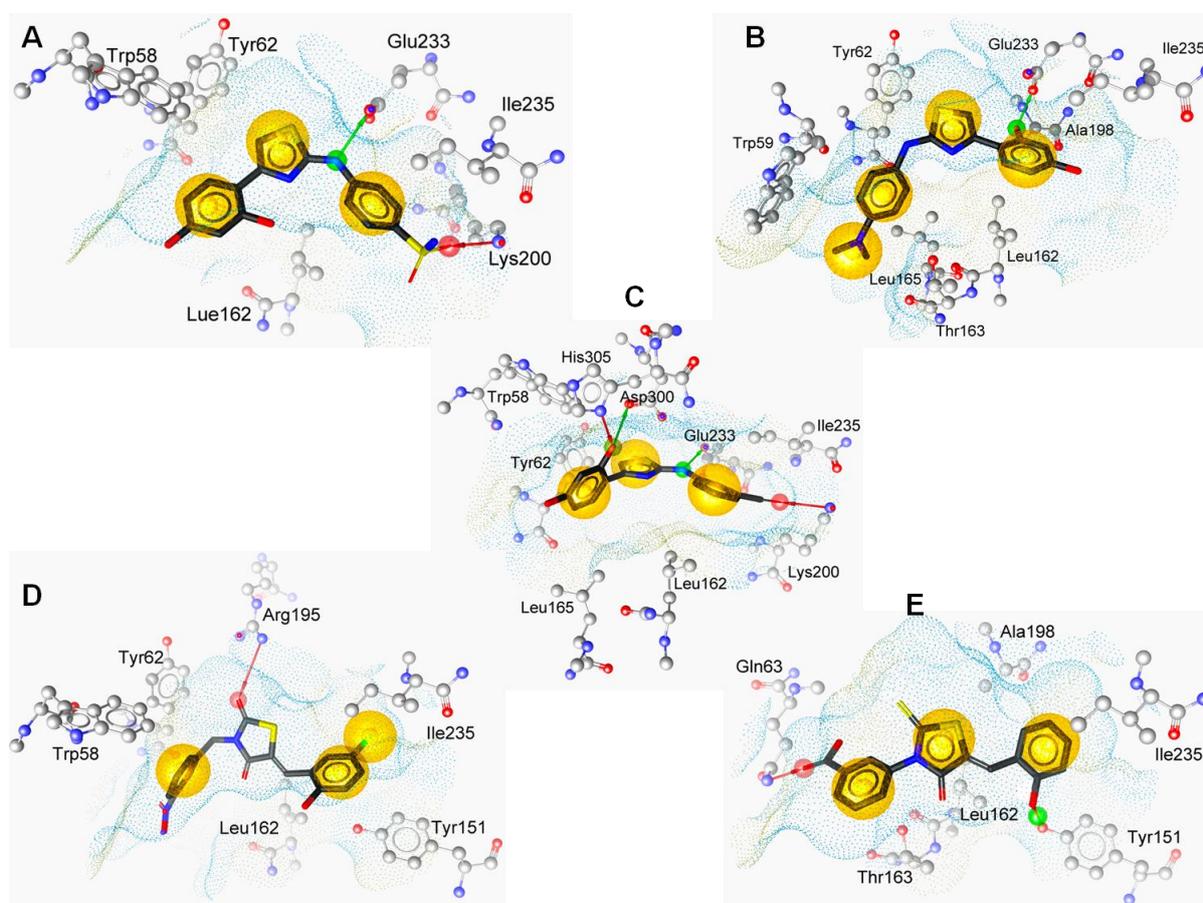


Figure 30: Predicted binding poses of compounds 25 (A), 26 (B), 27 (C), 30 (D), and 31 (E) in black sticks and protein residues as gray balls and sticks. Yellow spheres represent hydrophobic contacts, red and green arrows show H-bond acceptors and donors, and yellow and blue dots represent hydrophobic and lioppilic surfaces of the receptor binding pocket.

Investigating the predicted poses of **28** and **29** reveal that the pyrazole ring is located near to the charged subsite -1 (due to the presence of Asp197, Glu233, and Asp300) expecting to form favorable interactions. However, methyl substitution in case of **28** and carboxyl group in case of **29** destabilize both ligands at their sites and explain their weak inhibition.

4.3.8 Refined pharmacophore-based virtual screening

During this thesis, α -amylase has been co-crystallized with myricetin ($IC_{50} = 30.2 \mu\text{M}$ [82]) and published with PDB code 4GQR [25] as a first non-carbohydrate ligand bound to the HPA active site. The catalytic residue Asp300 in the active site and some other residues show different orientations from the ones observed with all other macromolecules bound to carbohydrate-based derivatives (chapter 4.1.1). Furthermore, novel non-carbohydrate-based α -amylase inhibitors have also been recently published. Hence this important new structural information was included in a refined virtual screening strategy. While common-based pharmacophore was applied in the initial virtual screening, individual 3D models derived from inhibitor clusters will be used in the refined one.

In the coming chapters, the development of 3D models based on this new information will be described as well as the update of validation databases (chapter 4.3.8.1). Virtual screening workflow using the developed pharmacophores and rational selection of hits will be described in chapter 4.3.8.2. While biological testing and novel identified α -amylase inhibitors will be addressed in chapter 4.3.9, their hypothesized binding modes will be given in chapter 4.3.10.

4.3.8.1 Development of 3D pharmacophores

As a first step in this refined virtual screening strategy, a new validation dataset was generated. The dataset used in the initial virtual screening was updated and comprised (i) 19 active compounds including most identified inhibitors recently published during this work (Appendix 10.3) (ii) 60 inactive molecules (Appendix 10.3 and 10.4), (iii) 254 decoys generated by a KNIME workflow (an in-house workflow called myDecoyFinder) based on the 19 active molecules, and (iv) 64,255 molecules (from WDI2005 database) and considered also as decoys. Preparation of validation datasets will be described in detail in the experimental part in chapter 7.2.2. Based on the scaffold type, the 19 biologically known active compounds were divided into six groups:

- i. Carbohydrate-based inhibitors, which include eight molecules with an acarviosine-glucose scaffold. This scaffold co-crystallized with HPA (PDB: 1XCW [32]) was chosen to represent this group.
- ii. Flavonoid-based inhibitors, which comprise five compounds including montbretin A, B and C. Myricetin was selected to represent this group since it is the only available co-crystallized flavonoid with HPA (PDB: 4GQR [25]).
- iii. Gallic acid derivatives which enclose 3 inhibitors: theaflavin monogallate, theaflavin

digallate, and 1,2,3,4,6-pentagalloyl- β -D-glucose [84, 165]. Here, the latter was used to build a model to represent this group since it is the most potent inhibitor among others.

- iv. Salacinol inhibitor with a tetrahydrothiophene fragment [175].
- v. 2-(4-(3,5-Dihydroxyphenoxy)-3,5-dihydroxyphenoxy) benzene-1,3,5-triol (DDBT), which has a phloroglucinol scaffold [83].
- vi. [2-(4-Bromophenyl)-4H-benzo[d]pyrimido[2,1-b][1,3]thiazol-4-yliden]acetonitrile inhibitor (ChEMBL2203334) which contains an acetonitrile moiety [87].

Individual 3D pharmacophore models for each group mentioned above were generated. Acarviosine-glucose and myricetin co-crystallized with α -amylase (PDB codes 1XCW and 4GQR, respectively) were used to build 3D models.

The other four models were obtained by docking inhibitors 1,2,3,4,6-pentagalloyl- β -D-glucose [84], salacinol [175], DDBT [83], and ChEMBL2203334 [87] into the HPA active site (PDB: 3OLE) using the software GOLD 5.1. Resulted poses were prioritized using the pharmacophoric features of the co-crystallized acarviostatin II03 and the quality of interaction with catalytic residues. Selected poses were used to create 3D pharmacophores (chapter 7.2.4).

The 3D model created from ChEMBL2203334 was excluded since it was too unspecific during validation and its early enrichment factor ($EF_{1\%}$) was null. It means that this model is not able to discriminate between true and false positives from the used validation dataset.

Validation results of the selected five models show high efficiency in retrieving active out of inactive compounds and decoys. This is translated into $EF_{1\%}$ and $AUC_{100\%}$. Validation results are summarized below in Table 9 and ROC curves are given in Figure 52 (Appendix 10.8).

Table 9: Overview of validation results of models developed for the refined virtual screening

Model	Retrieved ligands	Actives (19)	Decoys (60 + 254)	$EF_{1\%}$	$AUC_{100\%}$
1XCW-based	37	11	26	17.5	0.76
4GQR-based	12	7	5	11.7	0.68
Salacinol-based	27	6	21	17.5	0.63
Pentagalloyl-based	4	3	1	11.7	0.58
DDBT-based	27	2	25	11.7	0.52
All models	67	14	53	11.7	0.83

$EF_{1\%}$ stands for early enrichment factor and $AUC_{100\%}$ represents the area under the curve.

Consequently, the chosen models were further validated by incorporating the database WDI2005 as decoys added to the previous dataset. A summary of the statistical validation for the models used in the refined virtual screening is given in Table 10 while ROC curves are detailed in Appendix 10.8, Figure 53.

Table 10: Results of second validation for pharmacophores selected for the refined virtual screening

Model	Retrieved ligands	Actives (19)	Decoys (64569)	EF _{1%}	AUC _{100%}
1XCW-based	3,496	11	3,485	31.6	0.77
4GQR-based	794	7	787	36.9	0.68
Salacinol-based	4,194	6	4,188	5.3	0.63
Pentagalloyl-based	60	3	57	170	0.58
DDBT-based	3,733	2	3,731	10.5	0.53

EF_{1%} stands for early enrichment factor and AUC_{100%} represent the area under the curve.

Details of the chemical features of the five selected models are explained and illustrated below.

1XCW-based 3D pharmacophore model:

PDB code 1XCW containing acarviosine-glucose was taken as a starting point to develop this model to represent carbohydrate-based inhibitors. As illustrated in Figure 31, this model comprises six essential and optional H-bonds besides two optional hydrophobic contact (HYD) and positive ionizable interactions (PI). Essential H-bonds include: one HBA with Arg195, one HBD with the carboxylic oxygen OD1 of the side chain of Asp300, and one HBD with the carboxylic oxygen OE2 of the side chain of Glu233. Optional H-bonds include a HBD with the carboxylic oxygens OE1 of the side chain of Glu233, OD2 of the side chain of Asp300, and OD2 of the side chain of Asp197. While an optional positive ionizable interaction is formed with the catalytic triad (Asp197, Glu233, and Asp300), a hydrophobic contact is formed with Ile235 and Leu162. Thirteen exclusion volumes spheres were added to include information on the steric volume of the binding site. Sizes were tailored to accommodate most of the actives.

4GQR-based 3D pharmacophore model:

As depicted in Figure 31, this flavonoid-based model is composed of one essential HBA to water molecule 1244 (HOH-1244), three essential HBDs with the carboxylic oxygen OD1 of the side chain of Asp197, the nitrogen atom NE2 of His101, and HOH-1244, four optional HBDs with Trp59, Gln63, and OE1 of the side chain of Glu233, and created HBD at C7 in ring A of myricetin. Eleven exclusion volume spheres were included.

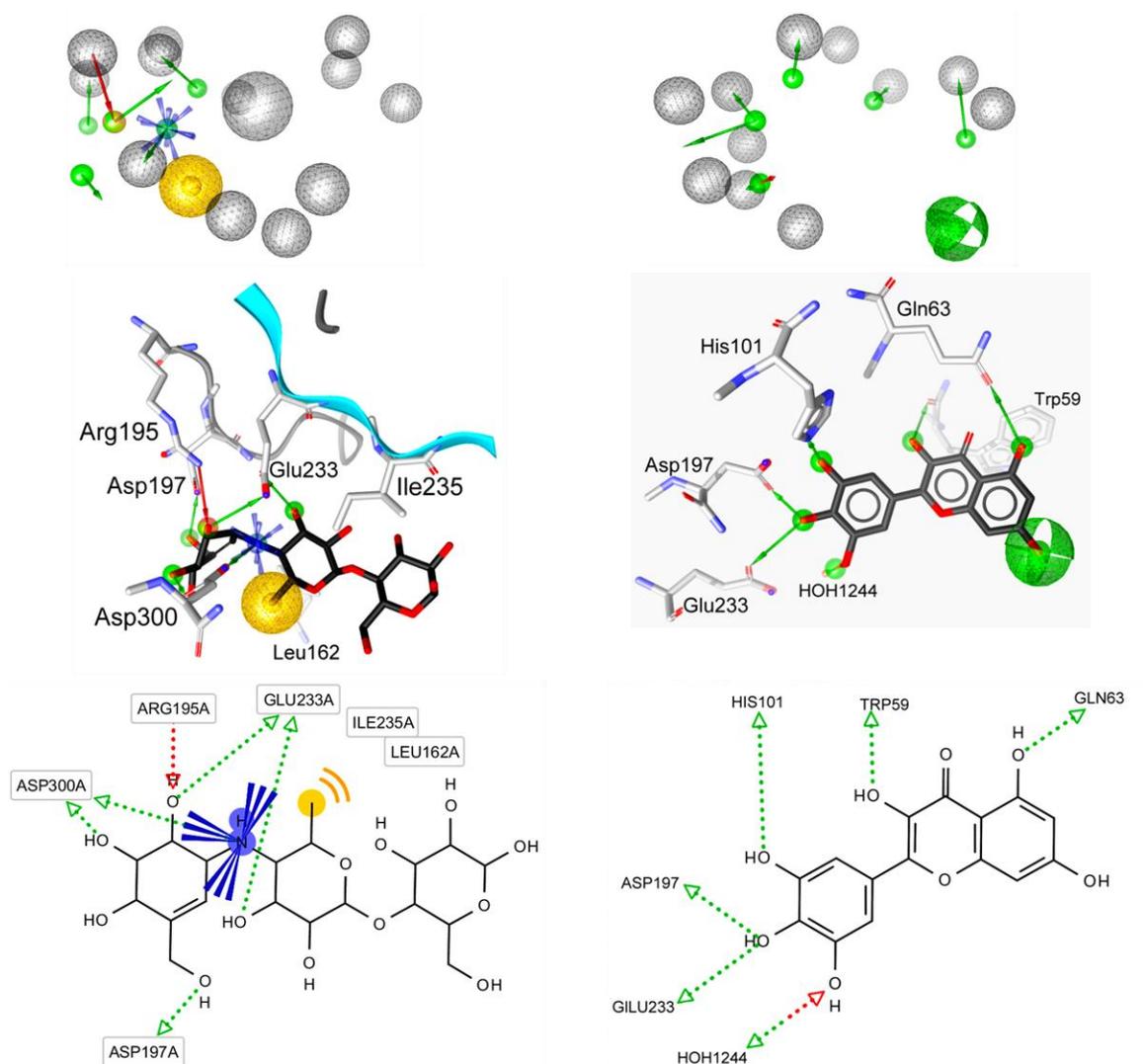


Figure 31: 3D pharmacophore models of 1XCW-based (left) and 4GQR-based (right) developed for the refined virtual screening. Above: 3D pharmacophore with exclusion volume spheres. Middle and below represent 3D and 2D interactions, respectively, in the α -amylase binding site for acarviosine-glucose (left) and myricetin (right). Color codes: green: H-bond donor, red: H-bond acceptor, yellow: hydrophobic contact, blue stars: positively ionizable interactions, gray spheres: exclusion volumes.

Salacinol-based 3D pharmacophore model:

As mentioned before, this model was created based on the docking of salacinol inhibitor into the catalytic pocket of the enzyme coded 3OLE (Figure 32). This model is composed of four H-bonds and one negative ionizable interaction (NI). Three essential HBDs are formed with the side chains of Asp197, Glu233, and Arg195, while the fourth feature is optional HBD with the side chain of Tyr151. The negative nature of the sulfate group of salacinol allows for NI interactions with the amino group of Lys200. Seven exclusion volume spheres were included to delimit the steric shape of the active site.

Pentagalloyl-based 3D pharmacophore model:

This model was developed to represent inhibitors containing gallic acid scaffold showing α -amylase inhibitor. 1,2,3,4,6-Pentagalloyl- β -D-glucose, a representative molecule of this group, was docked into 3OLE and the most plausible pose was selected in a similar manner as for salacinol. Iterative development and validation resulted into a 3D model illustrated in Figure 32.

This model includes five essential and two optional H-bonds. Three essential H-bond donors are formed with the carboxylic oxygens OD1 in the side chains of Asp197 and Asp300, and with the carboxyl group of Thr163 as well as one optional HBD with the nitrogen NE2 of His201. The guanidinium group of Arg195 accommodates two HBAs as essential and optional features, while another essential HBA is formed with the hydroxyl group of Tyr151. Exclusion volumes were defined as sixteen spheres.

DDBT-based 3D pharmacophore model:

This model was determined based on docking experiments of the recently published inhibitor 2-(4-(3, 5-dihydroxyphenoxy)-3,5-dihydroxyphenoxy) benzene-1,3,5-triol (DDBT) with IC_{50} value of 8.56 μ M [83]. The most plausible binding mode was selected similarly as salacinol. The developed model is given in Figure 32 and includes four H-bond interactions. Two essential H-bonds are formed with the carboxylic oxygen OD2 of the side chain of Asp300 as HBD and with the guanidinium group of Arg195 as HBA. The carboxylic oxygen OD2 of the side chain of Asp197 and the nitrogen NE2 of His201 accommodate two optional H-bond donors. Eleven exclusion volume spheres were added as steric constraints.

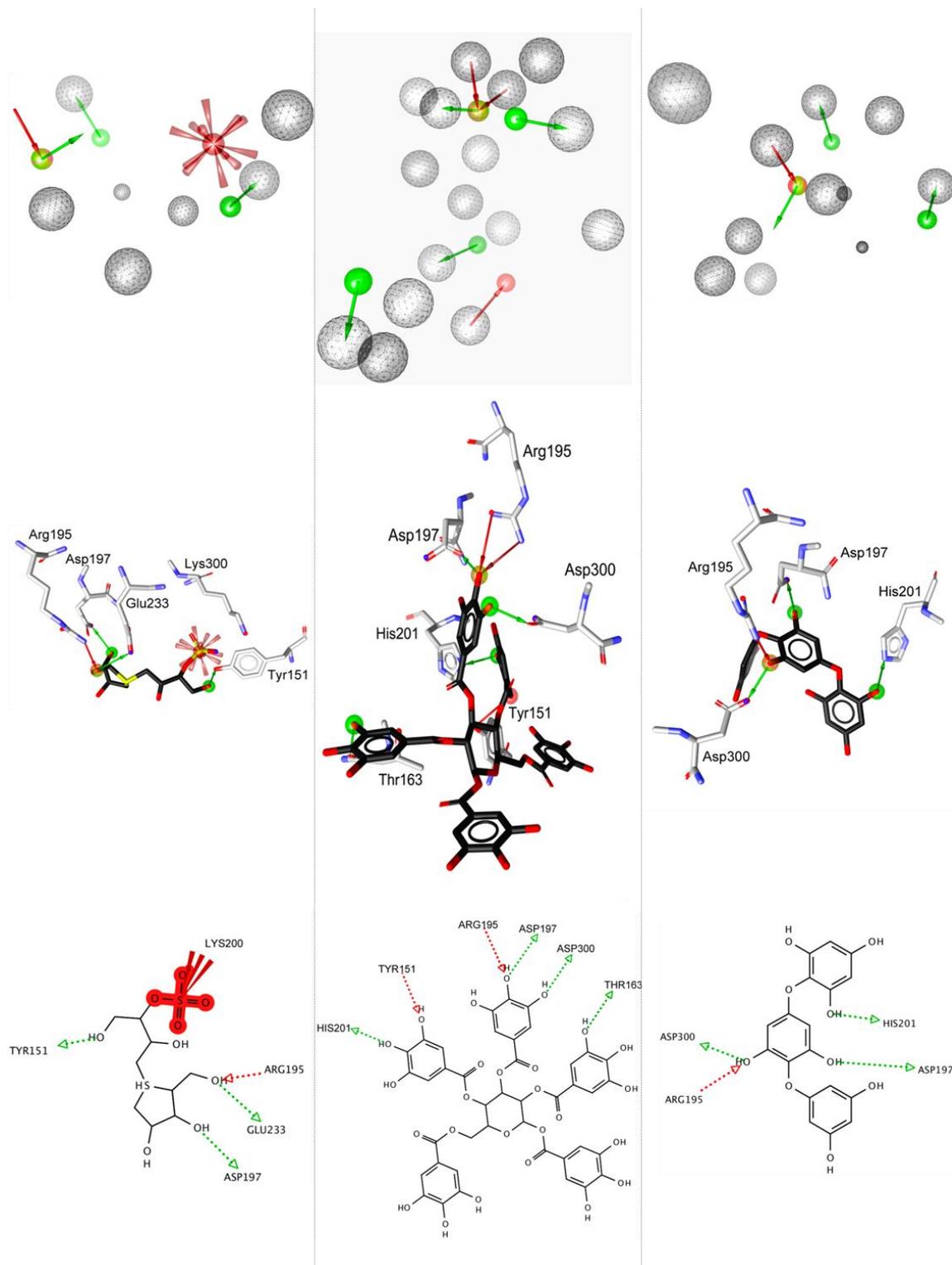


Figure 32: 3D pharmacophore models for salacinol-based (left column), pentagalloyl-based (middle column), and DDBT-based (right column) developed from docking salacinol, 1,2,3,4,6-pentagalloyl- β -D-glucose, and DDBT into α -amylase active site (PDB: 3OLE). Above: 3D pharmacophore with exclusion volume spheres. 3D and 2D ligands interactions are given in the middle and below, respectively. Color codes are similar to previous figures.

4.3.8.2 Protocol of the refined virtual screening

The previously defined 3D pharmacophore models were used individually for virtual screening of a library of 1,346,275 commercially available compounds from different vendors (Figure 33). LigandScout 3.1 was used to perform virtual screening and resulted into the selection of 128, 591, 2974, 57, and 1350 hits for 1XCW-based, 4GQR-based, salacinol-based, pentagalloyl-based and DDBT-based 3D pharmacophore models, respectively.

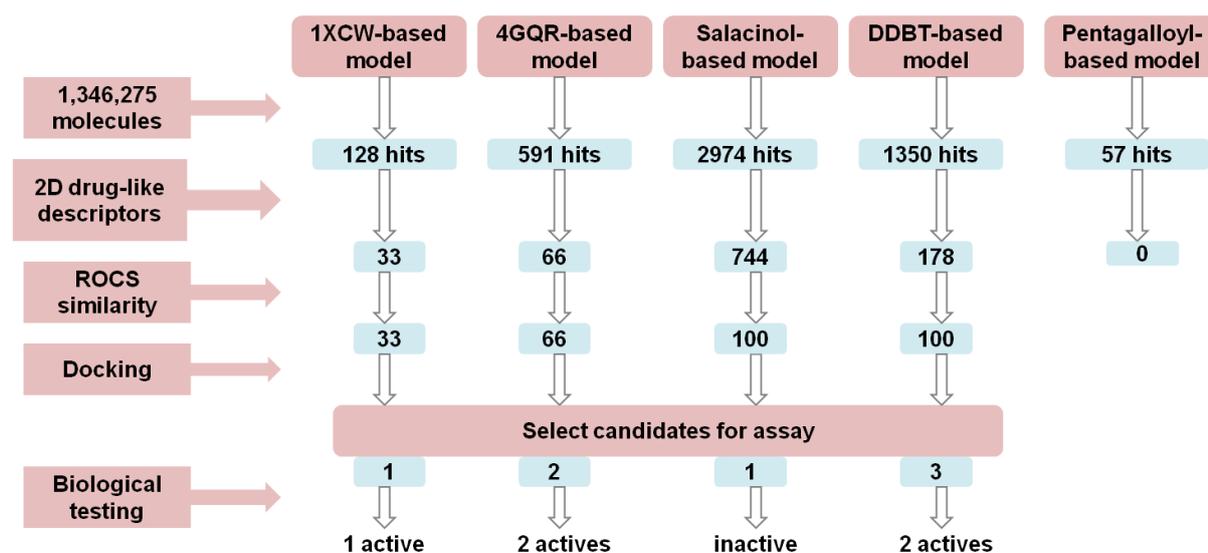


Figure 33: Workflow of the refined virtual screening.

Based on their physicochemical properties, virtual hits resulting from each 3D model were refined using the software FILTER 2.1.1 [176]. This step was conducted to exclude unstable, reactive and toxic molecules using 2D drug-like descriptors (chapter 7.2.4).

As a next step, hits with scaffolds similar to the known biologically inactive compounds (Appendix 10.4) or those with sugar fragments were excluded using substructure similarity search implemented in the tool Instant JChem 5.8.0 [140]. Consequently, previous 2D descriptors led to 33, 66, 744, and 178 hits belonging to 1XCW-, 4GQR-, salacinol-, and DDBT based-models, respectively. Hits from pentagalloyl-based model did not tolerate those 2D filters and hence were excluded from further steps.

Subsequently, molecules that are the closest to the lead queries in terms of shapes and features were selected. To do so, compounds remaining from each hit-list were transformed into multi-conformational structures using the software OMEGA 2.4.6 [177] to be used as

inputs for the software ROCS 3.1.2 [142-144, 178] aiming at ranking virtual hits with Tanimoto ComboScore as explained in chapter 3.1.4.

While conformations of acarviosine-glucose and myricetin queries were respectively derived from crystal structures coded as 1XCW and 4GQR, conformations of salacinol and DDBT queries were obtained from their selected docking poses (Figure 34).

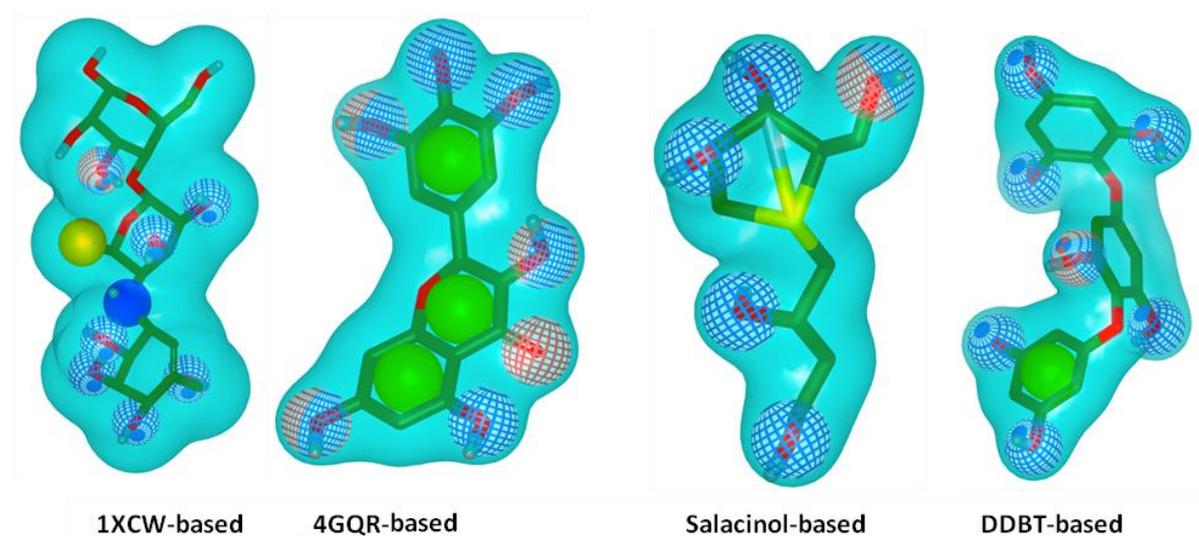


Figure 34: 3D Queries developed in ROCS and used to rank hitlists in the refined virtual screening. Color codes: Ligands are depicted as stick green; cyan represents shape of the query, preliminary chemical features are illustrated as green spheres for ROCS ring features, yellow sphere for hydrophobic contacts, red and blue nets represent H-bond acceptors and donors, respectively.

Corresponding chemical features defined in previous pharmacophore models were incorporated into each relevant query. Consequently, hits selected using 1XCW-based model, 4GQR-based model and the 100 best hits from salacinol-based model and DDBT-based model were kept for the next step.

Then, a total of 299 hits were docked into the α -amylase binding site (PDB: 3OLE) using the software GOLD 5.1. Next, a careful visual inspection of the docking poses was performed to select the best candidates for biological testing. Previously defined 3D pharmacophore models for each hit-list were used to prioritize the relevant docking poses with a particular emphasis on the interactions with the catalytic triad, drug-likeness, diversity of the chemical structures and availability for immediate purchase at the vendors. Therefore, seven molecules were selected for biological assay. Compound **16** was selected with the 1XCW-based model and hits **17** and **20** were chosen with 4GQR-based model. Compound **15** was

chosen by salacinol-based model while compounds **18**, **19** and **21** were chosen by DDBT-based model.

All selected virtual hits were tested using the kinetic assay described in chapter 7.3.2. Overall, five out of seven biologically tested compounds show inhibition potency against α -amylase, corresponding to a hit rate of 70 %. An overview of the results is shown in Table 11.

Table 11: An overview of virtual hits selected by the refined virtual screening for biological assay

Query	Selected compounds	Identified inhibitors	Hit rate
1XCW-based	1	1	100 %
4GQR-based	2	2	100 %
Salacinol-based	1	0	0 %
DDBT-based	3	2	66%
Overall	7	5	71% (5 out of 7)

4.3.9 Biological testing of hits selected by the refined virtual screening

Using the kinetic assay with GalG2CNP and HSA, virtual hits selected by the refined virtual screening were evaluated and their IC_{50} were determined and given in Table 12 (Figure 35). Compounds **15** and **19** were inactive at concentrations 1.8 mM and 560 μ M, respectively.

Table 12: Biological results of hits selected by the refined virtual screening and ligand efficiency (LE) for the identified inhibitors

Compound	Mol.wt	IC_{50} (μ M)	HA	LE
15	298.35	ND*	22	
16	286.29	298	21	0.23
17	341.15	~ 300	22	0.22
18	341.15	17	22	0.30
19	345.37	ND**	26	
20	457.49	27	34	0.18
21	381.41	407	27	0.17

ND* means inactive at 1.8 mM, ND** means inactive at 560 μ M (with GalG2CNP and HSA). HA represents heavy atoms, and ligand efficiency was calculated by the equation $LE = (1.37/HA) \times pIC_{50}$.

The type of inhibition for the most promising inhibitors (compounds **18** and **20**) was further investigated. Compound **18** shows non-competitive inhibition while compound **20** inhibits in a competitive manner.

While compounds **17** and **20** were selected with 4GQR-based pharmacophore, compounds **18** and **21** were identified by DDBT-based pharmacophore. These four compounds share (*E*)-*N'*-benzylideneacetohydrazide scaffold. To the best of our knowledge, it is the first time this scaffold is reported as a part of α -amylase inhibitors. The fifth identified compound that shows α -amylase inhibition is **16**, recognized with the 1XCW-based model and has 2*H*-pyrrolo[1,2-*a*]indole-3,9-dione scaffold.

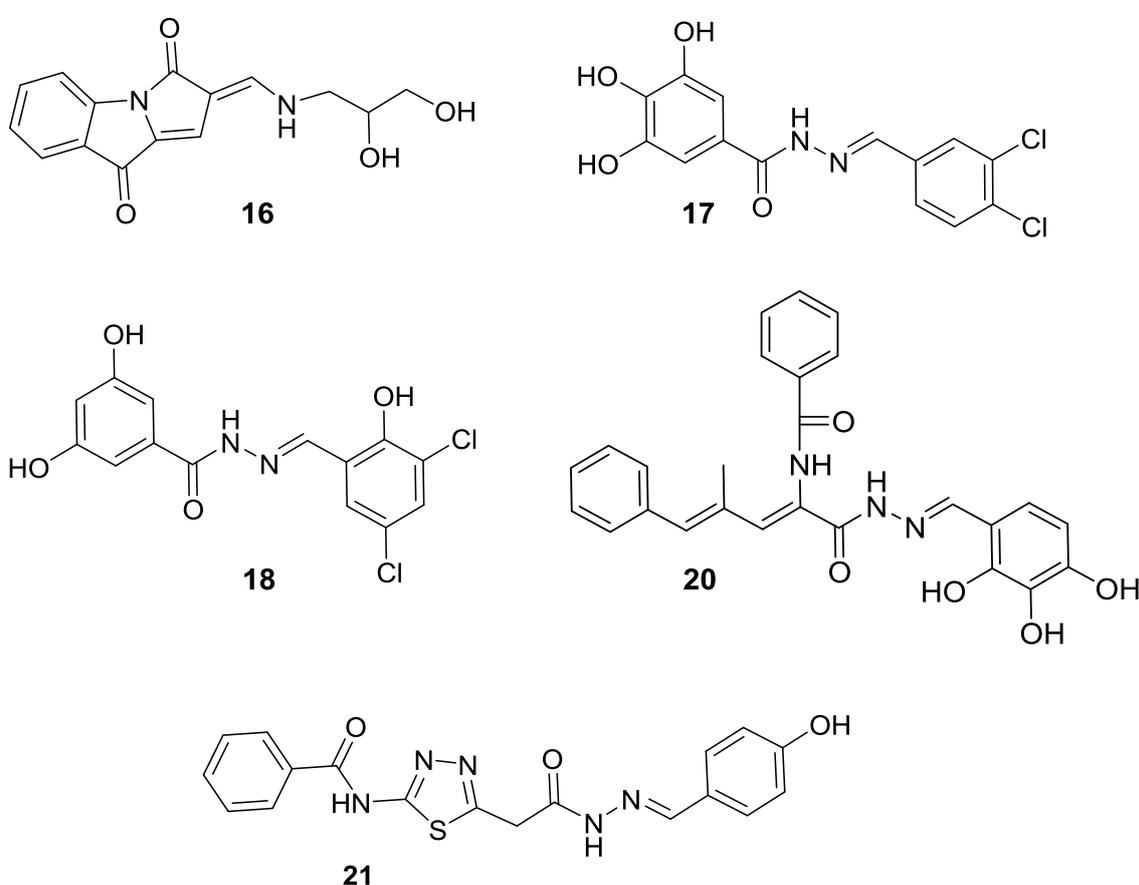


Figure 35: Novel α -amylase inhibitors identified by the refined virtual screening.

The *ortho*-hydroxyl substitution at the benzylidenehydrazine moiety seems to be essential for activity as observed with compounds **18** and **20** and compared to other identified inhibitors (**17** and **21**) that lack this substitution.

Compound **18** ($IC_{50} = 17 \mu M$) is the most potent inhibitor identified by the refined virtual screening and differs from compound **17** ($IC_{50} = 300 \mu M$) by its hydroxyl and chlorine

substitutions. On the benzylidenehydrazine scaffold, compound **18** has one hydroxyl group in *ortho*-position, which is absent in case of compound **17**. Chlorine substitutions in compound **18** are located in *meta*-positions compared to *meta*- and *para*-positions in compound **17**. Our results indicate that this substitution is important for the activity of **18**.

On the other hand, hydroxyl substitution on the benzoyl moiety of **17** and **18** seems to play a minor role in the activity of the identified inhibitors. At this moiety, compound **17** has three hydroxyl groups while only two are flanking compound **18**. In contrast, compound **20** ($IC_{50} = 27 \mu\text{M}$), which holds a distal benzoyl group lacking hydroxyl substitution.

Interestingly, all inhibitors identified by the refined virtual screening strategy show higher ligand efficiency values than acarbose, the positive control, which has $LE = 0.13$, (Table 12), more details will be discussed in chapter 5.2.3.

All other tested compounds show no activity, including compound **15** that has two pyrazole rings attached to benzene-1,4-diol and compound **19** that has imidazole rings attached to tetrahydroxybutane. The inactivity of **15** supports our previous findings concerning pyrazole ring as unfavorable for inhibition (compounds **5**, **28**, and **29** from the initial screening).

A detailed molecular modeling study of the binding modes and subsequent SAR of these inhibitors is reported in the following section.

4.3.10 Docking of novel inhibitors identified by the refined virtual screening

After the biological testing, binding modes of the identified inhibitors were re-investigated to rationalize the structural requirements for α -amylase inhibition. To reach this goal, the discovered inhibitors **16**, **17**, **18**, **20**, and **21** as well as inactive virtual hits were re-docked into α -amylase binding site using GOLD 5.1 (chapter 7.5.4).

The DDBT-based model was used to prioritize poses of compound **18** and a 3D pharmacophore was created for the proposed binding mode. Since it is the most potent identified compound in the refined virtual screening, pharmacophoric features of **18** were used to select plausible poses of other compounds.

Biochemical assay shows a non-competitive mechanism of inhibition for compound **18**. This means **18** has two binding possibilities: (i) It could bind to the active site at the same time with the substrate and interfere with hydrolysis, or (ii) it could bind to a site or several sites that are away from the catalytic center and remotely interfere with the substrate binding, similarly to the non-competitive ethyl caffeate ($K_i = 1.3 \text{ mM}$ [82], PDB: 4GQQ [25]). Both possibilities were investigated in this study.

To explore the first hypothesis, compound **18** was docked into α -amylase active site (PDB: 3OLE [36]). All investigated poses occupy subsites -3, -2, -1, and +1. The proposed binding mode of **18** shows five H-bond interactions as well as two lipophilic contacts (Figure 36). The *ortho*-hydroxyl group of benzylidenehydrazine donates one H-bond to the catalytic residue Asp300. Compounds with the same scaffold, but lacking this substituent, are considerably weak inhibitors (**17** and **21**) and the one that has this *ortho*-hydroxyl group shows good inhibitory potency (**20**). In addition, this hydroxyl group accepts one H-bond from His305. The catalytic residue Glu233 forms two H-bonds with the hydrazine group and the *meta*-hydroxyl group of the benzoyl ring. The other *meta*-hydroxyl group accepts one H-bond from Lys200. Interestingly, hydrophobic contacts are observed between the chlorine atoms of benzylidenehydrazine and Trp58, Lue162, Lue165, Tyr62 and Ala198 in subsites -3 and -2. To investigate the second hypothesis, the possible binding of compound **18** in allosteric sites located at the surface of HPA was evaluated by docking (using PDB code 4GQQ [25]).

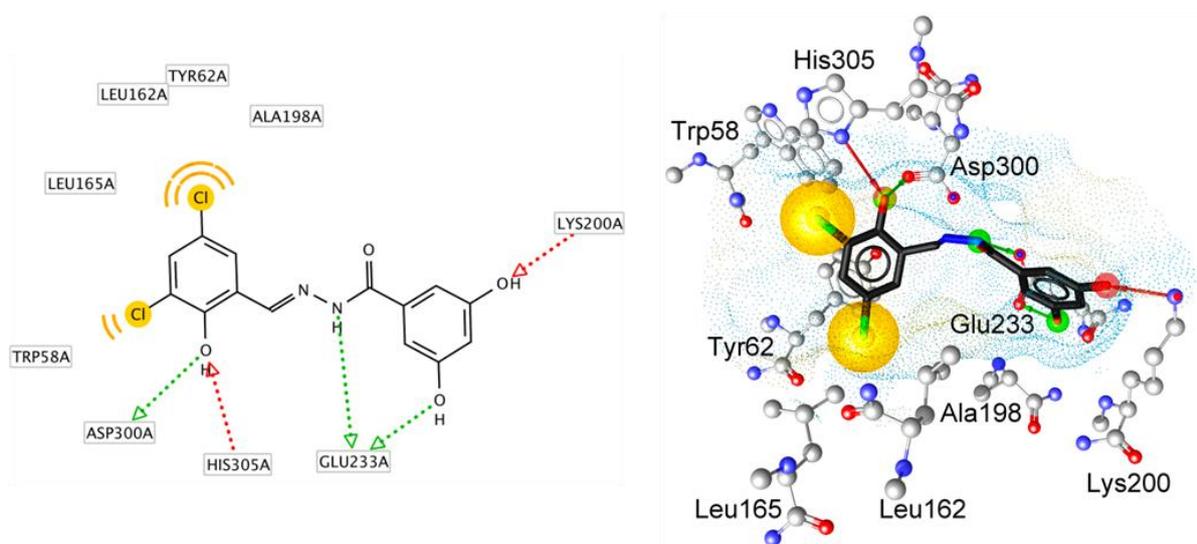


Figure 36: Proposed binding mode of compound 18 as 2D (left) and 3D (right) docked into α -amylase active site (PDB: 3OLE). Color codes: H-bond donors and acceptors as green and red arrows, respectively; and hydrophobic contacts as yellow spheres. Ligand is in black sticks and amino acids are in gray balls and sticks.

As explained earlier in chapter 4.2.3, three identified sites located at the HPA surface can accommodate three molecules of ethyl caffeate. This exerts inhibition through disordering four chain segments and re-orientation of the catalytic residue Asp300. Interactions of compound **18** at these three allosteric sites were investigated using molecular docking. Results are described below,

Site 1: Site 1 accommodates the co-crystallized ethyl caffeate which forms three H-bonds between *meta*- and *para*-hydroxyl groups with Gly285, Asp236, and Lys257, on one side, and a lipophilic contact between the aromatic ring and Val287, on the other side. These features will be used to select the proposed binding mode for compound **18** at this site.

Compound **18** shows interactions similar to ethyl caffeate at this site (Figure 37). The hydroxyl group at *meta*-position of the benzoyl ring donates one H-bond to Asp236 and accepts one H-bond from Lys257 in a similar manner as with the *para*-hydroxyl group of ethyl caffeate. Likewise, the hydrazine group of **18** donates one H-bond to Gly285 in a similar way to the one with *meta*-hydroxyl group of ethyl caffeate. Both *meta*-hydroxyl groups of **18** form two H-bond acceptors with Ser244 and Ser245 and one H-bond is donated from the *ortho*-hydroxyl group of **18** to Glu282. However, **18** does not show any of the hydrophobic contacts observed with ethyl caffeate and the chlorine atoms of **18** are oriented toward the solvent. Nevertheless, our observations indicate that a binding of **18** could be possible in this site 1.

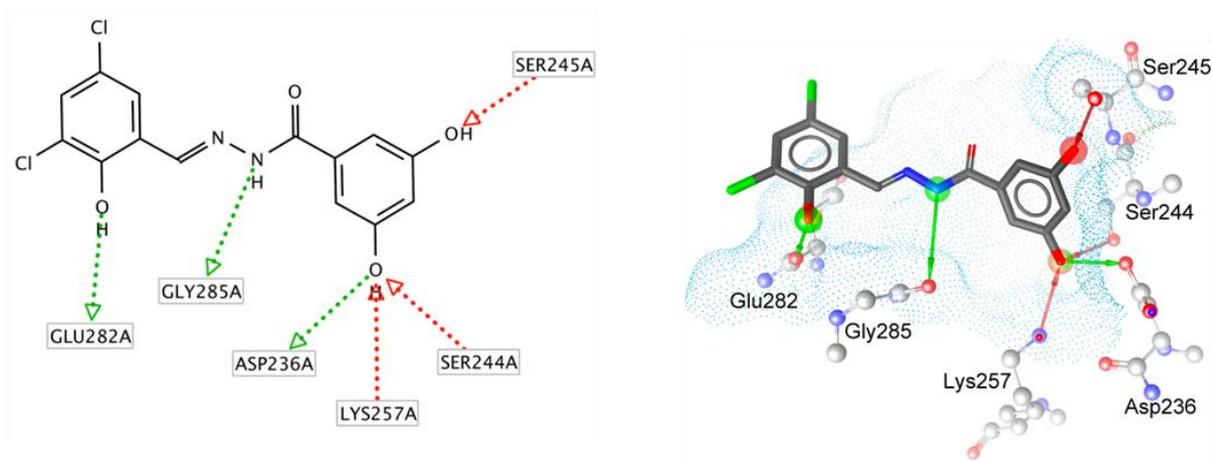


Figure 37: Suggested binding mode of compound 18 docked into allosteric site 1 (PDB: 4GQQ). Green and red arrows represent H-bond donors and acceptors, respectively.

Site 2: Ethyl caffeate forms three key interactions at site 2: a first H-bond between the *para*-hydroxyl group and Asp77, a second H-bond between the carbonyl oxygen and Arg85, and one lipophilic contact between the ethyl group and Leu26. Based on these features, the binding mode of **18** was investigated in this cavity. The *ortho*-hydroxyl and amino group of benzylidenehydrazine of **18** create two H-bonds with Asp77, on one side, and one chlorine atom forms a hydrophobic contact with Val22, on the other side. Hence, compound **18** can find a binding mode similar to ethyl caffeate in this site 2 (Figure 38). The other hydroxyl groups in the benzoyl ring of **18** are oriented toward the solvent, which can be in favor of binding and can stabilize the ligand in site 2.

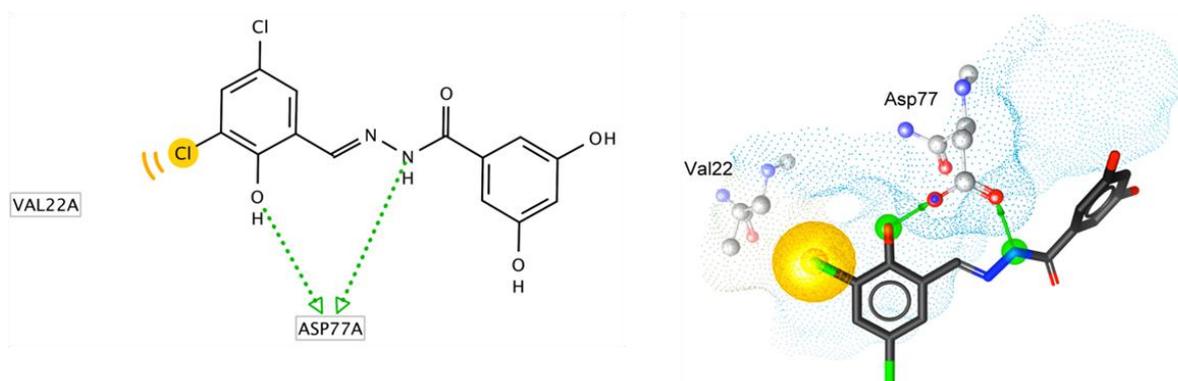


Figure 38: Proposed binding mode of compound 18 docked into allosteric site 2 (PDB: 4GQQ). Green arrows represent H-bond donors and yellow spheres symbolize lipophilic contacts.

Site 3: The *meta*-hydroxyl group of the co-crystallized ethyl caffeate creates one H-bond with Asp23 and so does the carbonyl oxygen of the ligand with Ile372. The aromatic ring and ethyl group of ethyl caffeate form hydrophobic contacts with Leu26 and Ile372, respectively. In the same manner as for the previous docking experiments, the most plausible binding pose of compound 18 at site 3 was selected based on its ability to fulfill the above-mentioned features.

In Figure 39, one hydroxyl group in *meta*-position of the benzoyl ring of compound 18, on one side, creates one H-bond with Asp23 in a similar fashion as ethyl caffeate and accepts two H-bonds from Arg30, on the other side.

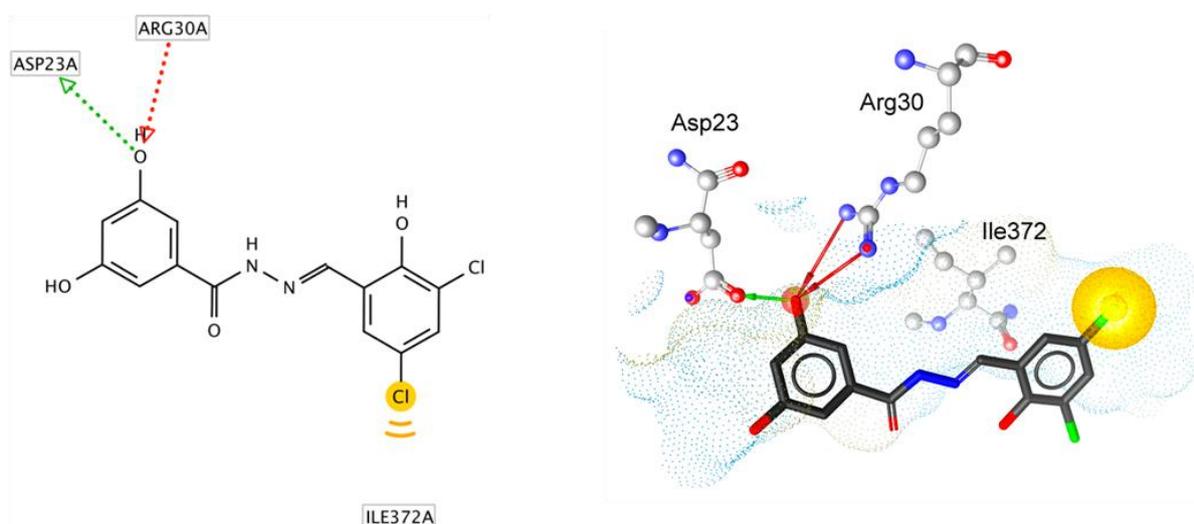


Figure 39: Predicted binding mode of compound 18 docked into allosteric site 3 (PDB: 4GQQ). Red arrows denote H-bond acceptors, green arrows show H-bond donors, and yellow spheres represent hydrophobic contacts.

One hydrophobic contact is formed between one chlorine atom of **18** and the side chain of Ile372. Thus, compound **18** shows more H-bonding but fewer hydrophobic contacts than ethyl caffeate in this site 3, which indicates that a binding is also possible in this allosteric site of the enzyme.

This docking study with the three allosteric sites of HPA shows that compound **18** can fit in these cavities in a similar way to ethyl caffeate, although not all functional groups of the ligand are involved into interaction with the enzyme. Nevertheless, based on this molecular modeling study, we surmise that compound **18** most likely binds to the catalytic center rather than the allosteric sites. This is supported by the following observations:

- i. The α -amylase active site is large in size (up to seven subsites) and both compound **18** and GalG2CNP substrate are small.
- ii. Compound **18** does not fulfill all chemical interactions identified between the co-crystallized ethyl caffeate and the HPA allosteric sites, usually creating fewer hydrophobic contacts.
- iii. Compound **18** shares the similar scaffold as competitive inhibitor **20**, which means they could share similar binding site. However, the small size of **18** allows a simultaneous binding of the inhibitor and the substrate, leading to non-competitive inhibition. Instead, the larger molecular size of **20** leads to an occupation of the catalytic site with a competitive inhibition mode.

As stated before, compound **20** shows competitive inhibition with an IC_{50} value of 27 μ M. This inhibitor was selected by virtual screening using the 4GQR-based model.

The most plausible binding mode of **20** shows interactions with the HPA catalytic sub-sites spanning from -3 to +2 (Figure 40). It can form five H-bonds: two HBDS between the *ortho*- and *meta*-hydroxyl groups of the benzylidinehydrazine fragment and His201, one HBA between the *para*-hydroxyl group and Lys200, and two H-bonds with the side chain of Thr163. The latter creates, on one side, one HBD with the amino group linked to the benzoyl moiety of **20** and, on the other side, another HBA is observed with the carbonyl oxygen attached to the hydrazine group of **20**.

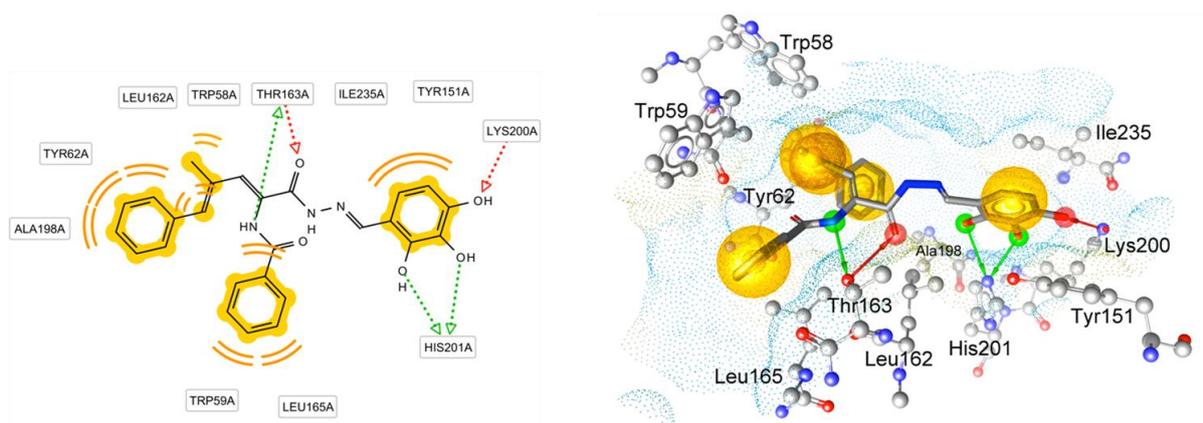


Figure 40: Proposed binding mode of compound 20 docked into α -amylase active site (PDB: 3OLE). Red and green arrows show H-bonding and yellow spheres represent hydrophobic contacts.

The benzoyl ring of **20** occupies subsite -1 and is stabilized by forming hydrophobic contacts with Leu165, Trp59 and Thr163. Other hydrophobic contacts are formed between the ligand and the lipophilic residues inside the active site. These residues include Tyr151, Ile235, and Leu162 in subsites +2 and +1, Tyr62 in subsite -2, and Thr163 and Leu165 in subsite -3.

Interestingly, the proposed binding mode of **20** in the active site of α -amylase crystal structure in PDB entry 3OLE does not show any interactions with the catalytic triad (Asp197, Glu233, and Asp300). Hence, compound **20** was docked into α -amylase active site of PDB code 4GQR to explore its binding with the triad residues. Based on the quality of interactions with the 4GQR-based model and with catalytic residues, a suggested pose was selected. Then, a 3D pharmacophore was created for this pose inside the active site of the enzyme coded 4GQR to investigate its binding mode. In brief, compound **20** is oriented in subsites -3 to -1 in 4GQR and shows interactions with Asp197 (more details are given in Appendix 10.9). The comparison between the docked binding mode of **20** in 3OLE with the one in 4GQR suggests that this inhibitor is better accommodated in the first than in the second, since it occupies sub-sites spanning from -3 to +2 of PDB entry 3OLE in a similar manner as for acarbose.

Compound **16** ($IC_{50} = 298 \mu\text{M}$) was selected by virtual screening with the 1XCW-based model. This model contains one H-bond interaction with Arg195, one H-bond donor interaction with the carboxylic oxygen of the side chain of Asp197, and a hydrophobic contact with Ile235. These features are confirmed through the proposed binding pose of **16** by its hydroxyl groups that create H-bonds as well as the aromatic benzene ring that forms a lipophilic contact (Figure 41). However, this binding pose does not fulfill all chemical features

of **18** and shares only one H-bond acceptor with Lys200. Additionally, the benzene ring forms other hydrophobic contacts with Tyr151 and Ala307.

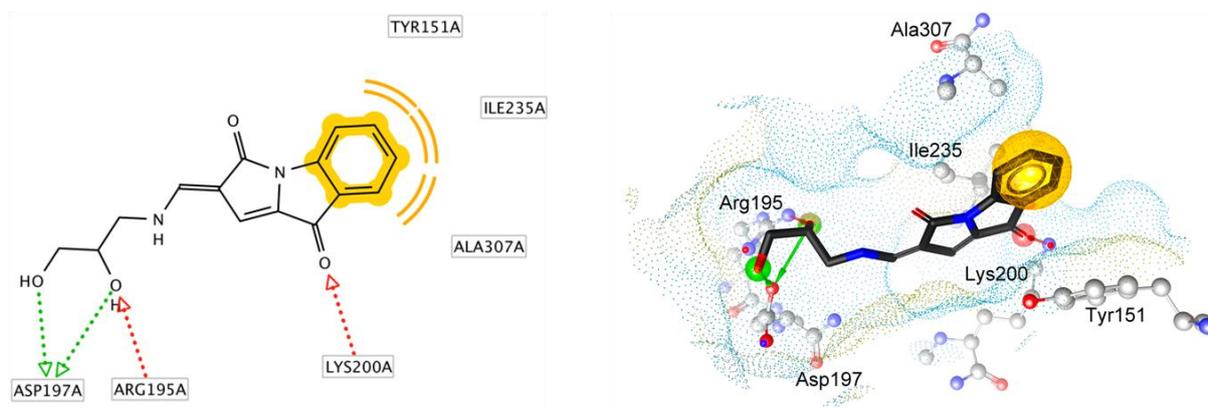


Figure 41: Suggested binding mode of compound 16 docked into α -amylase active site (PDB: 3OLE). Color codes: green and red arrows represent H-bond donors and acceptors, respectively; yellow spheres indicate hydrophobic contacts.

This chemical feature pattern can explain the lower inhibitory potency of **16** compared to **18**. This can also be due to two possibilities: (i) the high flexibility of the aminopropanediol fragment that allocates at sub-sites -1 and +1 and the resulting entropic penalty to binding, and/or (ii) all investigated conformations span from sub-sites -1 to +2 without protruding to subsite -2, which possesses a critical role in the enzymatic activity [37].

Likewise, compound **17** ($IC_{50} = 300 \mu M$) was identified by virtual screening with the 4GQR-based model. This model contains one H-bond donor interaction with Glu233 and this interaction was retrieved by the proposed binding pose of **17** (Figure 42). The high structural similarity between compounds **17** and **18** suggests that both compounds could have similar binding mode.

Hydroxyl groups on the benzoyl ring of compound **17** can form H-bonds with Glu233 and Lys200. Hydrophobic contacts are observed between chlorine substituents of benzylidene and Leu165, Thr163, and Trp59. However, the lack of an *ortho*-hydroxyl group in the benzylidene ring of **17** and the absence of the interaction between hydrazine group and Glu233 explain the weak inhibitory potency compared to **18**.

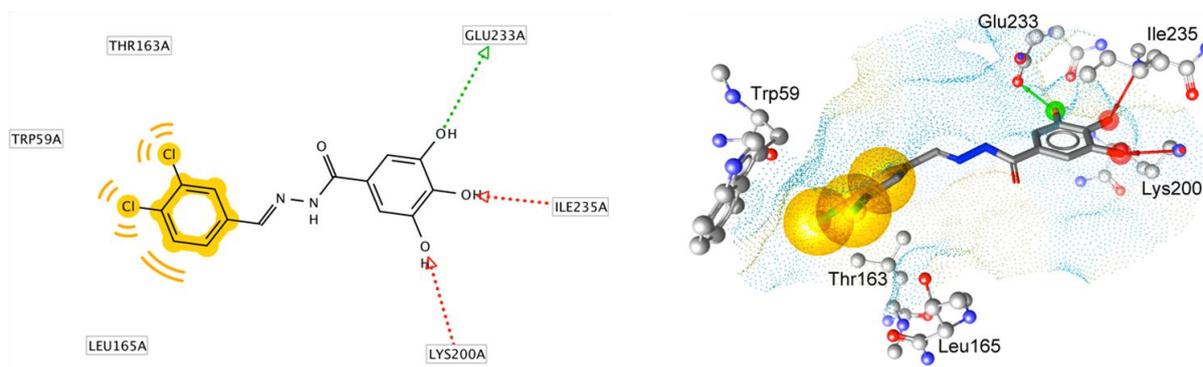


Figure 42: Predicted binding mode of compound 17 docked into α -amylase active site (PDB: 3OLE). Yellow spheres symbolize lipophilic contacts, red arrows represent H-bond acceptors, and green arrows denote H-bond donors.

Compound **21** ($IC_{50} = 407 \mu M$) was selected by virtual screening using the DDBT-based model, which includes H-bond donor interaction with Asp300. This feature is confirmed in the selected pose of compound **21** (Figure 43). This binding pose shows interaction of the *para*-hydroxyl group with Lys200 similar to the one with **18**. Nevertheless, **21** lacks the *ortho*-hydroxyl moiety linked to the benzylidene group of **18** and hence shows weaker inhibition.

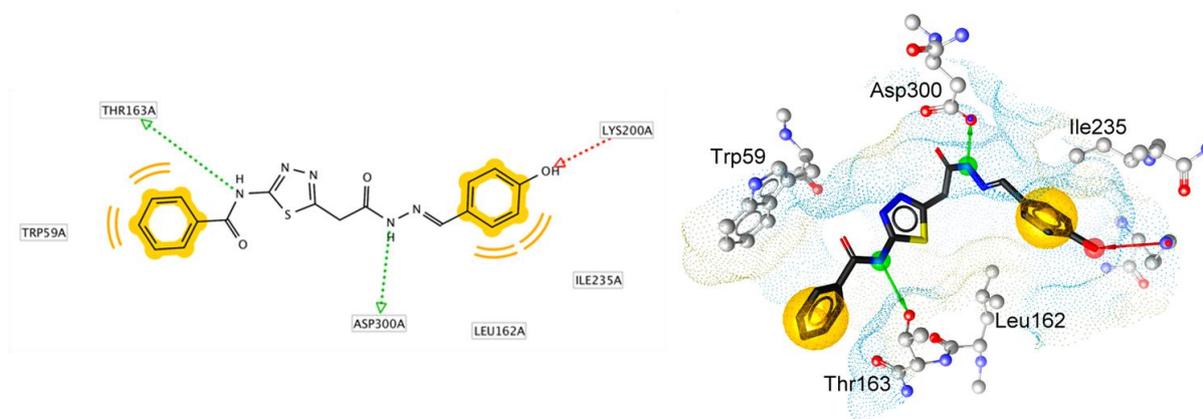


Figure 43: Plausible binding mode of compound 21 docked into α -amylase active site (PDB: 3OLE). Green arrows represent H-bond donors; red arrows denote H-bond acceptors; and yellow spheres depict hydrophobic contacts.

5 Discussion

Metabolic disorders such as type II diabetes and obesity are characterized by chronic hyperglycemia and unbalanced carbohydrate metabolism that causes other serious complications such as cardiovascular diseases. One of the main therapeutic approaches for treating these metabolic diseases is to control postprandial hyperglycemia. This can be achieved by delaying the absorption of glucose that resulted from carbohydrate digestion. Since α -amylase initiates the hydrolysis of polysaccharides such as starch, one of the main energy sources in human nutrition, this enzyme has been intensively studied as a promising target for controlling postprandial hyperglycemia.

The main aim of this study was the development of predictive structure-based 3D pharmacophore models that allow for the identification of novel α -amylase inhibitors with drug-like properties.

The search for non-sugar, non-peptide small molecules α -amylase inhibitors turned out to be challenging because of the following reasons:

- i. At the start of this project, only sugar-based co-crystallized α -amylase inhibitors were available with unfavorable properties for oral administration. Also, a limited number of non-sugar inhibitors with unknown binding modes were reported. Flavonoids were the only drug-like molecules reported with molecular mechanism for inhibition of HSA [20], PPA [179], and α -amylase from the bacterium *Bacillus subtilis* [78].
- ii. The catalytic pocket is a large binding cleft that can accommodate large carbohydrate substrates (up to seven binding subsites with HPA). This makes the binding of non-sugar small molecules a challenge, as they must compete with the sugar substrate.
- iii. Sugar-based structures undergo a series of enzymatic hydrolysis and condensation reactions in the α -amylase active site. α -Amylase hydrolyzes starch into a mixture of oligomaltosidic chains. These chains behave as substrates and will compete for the active site [172].

At the beginning of this work in 2011, a hypothesis using co-crystallized sugar-based inhibitors was proposed as a starting point to develop 3D models that integrate all structural features required for an optimal inhibitor binding. After validation, the best model was selected for the initial virtual screening to identify non-sugar and drug-like small molecules with high potential α -amylase binding affinity (chapter 4.3.1).

During this work, other research groups have published new information about the enzyme and hence these data have been used to conduct refined virtual screening campaigns. At that time, a recent study published by Williams *et al.* in 2012 [25] for myricetin as a first small flavonoid co-crystallized with HPA is generally in agreement with structural determinations proposed by Lo Piparo *et al.* [20].

Additionally another new small α -amylase inhibitor called DDBT as a natural phlorotannin compound was reported with a K_i value of 1.8 $\mu\text{g/ml}$ (= 4.8 μM) [83]. Another research group published synthetic small compounds with 2-amino-benzothiazole scaffolds as α -amylase inhibitors [87, 88].

First, our discovered binding modes of natural α -amylase inhibitors identified by collaboration partners will be discussed in chapter 5.1. Then, our virtual screening approaches will be discussed and compared in chapter 5.2 in terms of yields, effect of substrates on IC_{50} values, and ligand efficiency of the discovered inhibitors.

5.1 Natural α -amylase inhibitors with unknown binding mode

Natural α -amylase inhibitors (cyanidin derivatives, steviol extract and verbascoside) identified by our collaboration partners were investigated to elucidate their binding modes using molecular docking and 3D pharmacophore modeling. While cyanidins were reported as α -amylase inhibitors [80, 180], it is the first time that their binding modes are investigated using molecular modeling. Our findings in chapter 4.2.1 confirmed the role of the glucose unit linked to 3-*O*-position of cyanidin that was suggested by Akkarachiyasit *et al.* [80]. α -Amylase active subsite -1 accommodates this glucose unit of both Cy-3-*O*-glc and Cy-3-rut and forms H-bond interactions with catalytic residues similar to valienamine moiety of the bound acarbose (PDB: 1MFV [49]). The competitive inhibitions of both Cy-3-*O*-glc and Cy-3-rut against HSA with identical IC_{50} values can suggest two different findings: (i) their binding modes are very similar which was confirmed by docking (chapter 4.2.1), (ii) the number of sugar units does not necessarily improve the inhibition since Cy-3-rut has one unit more than Cy-3-*O*-glc. Among the cyanidins addressed in this study, malvidin-3,5-diglucoside (Mv-3,5-*O*-diglc) represents a novel α -amylase inhibitor, which has never been reported before. Interestingly, the binding mode predicted for this competitive inhibitor shows the importance of ring B orientation for the inhibition. This clarifies why Mv-3,5-*O*-diglc exerts 3-fold inhibitory potency compared to other cyanidins even it has two sugar units, similarly as Cy-3-rut.

The second group of natural α -amylase inhibitors studied in this work was a steviol extract containing steviol, stevioside and rebaudioside. These ligands have not been reported as α -

amylase inhibitors previously and their binding modes were investigated for the first time in chapter 4.2.2. The three compounds could collectively inhibit HSA. Based on our molecular modeling strategy, stevioside is the most probably molecule that exerts HSA inhibition in the extract. It can occupy better the binding site of the enzyme. However, further experimental investigation is required to evaluate the inhibition potency of these ligands individually.

Similarly, the binding mode of verbascoside was investigated in chapter 4.2.3. Our *in silico* study showed that binding in the triad allosteric sites, remotely located at the surface and far away from the catalytic active site of HPA, is in general agreement with interactions of the co-crystallized ethyl caffeate. The presence of the phenylethanoid fragment alongside with ethyl caffeate enhances its binding at the HPA surface and hence improves its inhibitory potency. This validates the suggestion of Williams L. *et al.* that decorating ethyl caffeate could enhance the binding at the HPA surface [25]. Sugar units linking ethyl caffeate and phenylethanoid in verbascoside do not significantly contribute to the inhibition according to molecular modeling. This validates our previous findings concerning sugars in cyanidin compounds.

5.2 Virtual screening

5.2.1 Initial virtual screening

Due to the scarce information available about small α -amylase inhibitors, a hypothesis based on the binding modes of co-crystallized sugar-based inhibitors was developed to identify non-sugar small molecule inhibitors using virtual screening. A predictive common feature-based 3D pharmacophore model focusing on the catalytic core was developed and used for virtual screening of more than 1.7 million commercial chemical compounds available from different vendors. Thus, 14 virtual hits (one was excluded due to insolubility) were selected and biologically tested using two kinetic assays.

The first assay, using *p*-NPG5 as a substrate for PPA, led to the discovery of three α -amylase inhibitors with IC_{50} values below 100 μ M (compounds **7**, **8**, and **9**) comparable to acarbose, the positive control. This yield corresponds to a hit rate of 23 % (three out of 13). The second assay was carried out for only eight out of 14 selected virtual hits (based on commercial availability at the time of assay) using GalG2CNP as a substrate for HSA. This assay method revealed other three compounds (**2**, **3**, and **4**) with IC_{50} values between 138 and \sim 300 μ M resulting in a hit rate of 37.5 % (three out of eight).

Consequently, our computer-aided approach using structure-based drug design was successful for the discovery of novel small drug-like non-carbohydrate α -amylase inhibitors. This validates our previous assumption about using sugar-based α -amylase inhibitors to discover non-sugar inhibitors. Using kinetic assay turned out to be advantageous since the type of inhibition for the best identified α -amylase inhibitors could be investigated.

The competitive inhibition mode of inhibitors **2** and **3** validates our proposal that focusing on the catalytic core in the large binding site can lead to block the hydrolysis function of α -amylase. A more detailed discussion for the identified inhibitors will be given in chapter 5.2.3. Among others, compound **2** holding the scaffold *N*,4-diphenylthiazole-2-amine is the most interesting inhibitor identified in the initial virtual screening because of the following:

- i. It inhibits both PPA ($IC_{50} = 441 \mu\text{M}$) and HSA ($IC_{50} = 138 \mu\text{M}$) compared to others tested in both assays.
- ii. It is a small fragment (molecular weight of 300) with the highest ligand efficiency value of 0.25 among others identified in the initial virtual screening.
- iii. It is kinetically validated as competitive HSA inhibitor.
- iv. It represents an allosteric inhibitor against fructose 1,6-bisphosphatase ($IC_{50} = 11 \mu\text{M}$ [169]), an enzyme that plays a role during gluconeogenesis. This increases its importance as a small fragment playing dual role (against two enzymes) to decrease glucose level in the blood. It is promising ligand with potential anti-diabetic activity.

Therefore, compound **2** was used as a similarity template to search for analogues with improved α -amylase inhibition and enrich the structure activity relationship.

A 2D similarity search approach was conducted for a commercial database available in-house leading to the selection of 12 hits (chapter 7.6). Interestingly, all hits demonstrate α -amylase inhibition with IC_{50} ranging from 50 μM to 4 mM. All identified analogues are novel and have not been reported as α -amylase inhibitors before.

Interestingly, the first assay method revealed different inhibitory profiles from the ones identified by the second method. This result triggers a question regarding the mode of action of these inhibitors when using different substrates at different conditions. Our steric similarity approach using Gaussian similarity score could partially rationalize the competition between inhibitors and a substrate (chapter 4.3.3). IC_{50} differences occur due to the complexity of kinetic processes using different conditions such as substrate, enzyme type, and medium (chapter 3.2).

5.2.2 Refined virtual screening

Known biological α -amylase inhibitors were updated and clustered based on their scaffold type (chapter 4.3.8). In contrast to the initial virtual screening, individual 3D pharmacophores were developed and validated for each cluster and then used in the refined virtual screening. This step allows for better incorporating more scaffold types to refine the 3D models. Each model was used for virtual screening of commercial chemical databases comprising 1.4 million molecules (chapter 4.3.8.2). All 3D models showed restrictive selection since each model did not select more than 0.22 % of totally screened compounds. Hits chosen by the pentagalloyl-based model could not tolerate the 2D descriptors used in filtering steps and therefore, were excluded from further process.

Five out of seven biologically tested compounds demonstrate α -amylase inhibition with IC_{50} values ranging from 17 μ M to 407 μ M. This corresponds to a hit rate of 70 %. Binders identified in this refined virtual screening have never been reported before this work as α -amylase inhibitors. Interestingly, the scaffold (*E*)-*N*-benzylideneacetohydrazide is the common substructure in four inhibitors and has never been reported as a backbone for α -amylase inhibition before.

Different aspects can be raised as a matter of comparison between the initial and the refined virtual screening approaches. In the initial approach, only one but unique common feature-based 3D pharmacophore model developed from co-crystallized sugar-based binders was used. In the refined approach, four individual models developed from docking poses alongside with co-crystallized small ligands were utilized. While the refined virtual screening exhibits better efficiency in terms of IC_{50} values and higher hit rate than the initial campaign, the latter allows for identifying more novel scaffolds than the refined approach (chapter 5.2.3). Therefore, both strategies are successful approaches in finding novel small α -amylase inhibitors.

5.2.3 Evaluation of the performed virtual screening approaches

Virtual screening approaches in this work were assessed taking the IC_{50} cutoff of 300 μ M. Novelty of scaffolds, hit rate percentage, and ligand efficiency values for initial discovered hits, as well as folds of improvement in activity (activity ratio) and improvement in ligand efficiency (LE ratio) for optimized hits will be analyzed as matters of evaluation for this research.

As stated in Table 13, six α -amylase inhibitors identified by the initial virtual screening using two assays demonstrate potencies ranging between 86 μ M and 300 μ M. This corresponds to

a hit rate of 46 % (six out of 13). Likewise, four novel α -amylase inhibitors were identified by the refined virtual screening using one assay method showing IC_{50} values ranging between 17 μ M and 300 μ M. This matches a hit rate of 57 % (four out of seven). In total, ten out of 20 biologically tested hits (50 % hit rate) exhibit α -amylase inhibition with IC_{50} values ranging between 17 μ M and 300 μ M. These yields in both strategies are highly acceptable outcomes for structure-based virtual screening compared to a typical hit rate for a prospective virtual screening that ranged between 1 % and 40 % [181]. Among the identified inhibitors, compound **3** that has tri-phenyl scaffold is similar to the natural inhibitor DDBT recently published in 2012 [83]. This proves the strength of our computational approach for the discovery of α -amylase inhibitors.

Novelty of our discovered scaffolds stresses on the importance of these fragments, which can be used individually as a starting point for identifying other hits with potential α -amylase inhibition and give space for optimization.

Even though compounds **17**, **18** and **20** share the same backbone, compound **20** still has larger substituents and hence considered with novel scaffold.

Table 13: Evaluation of inhibitors discovered by initial and refined virtual screening campaigns

virtual screening method	Assay code	IC_{50} (μ M)	type of assay	type of inhibition	HA	LE	novelty of scaffold
Initial	2	138	kinetic with GalG2CNP substrate	competitive	21	0.25	Yes
	3	200		competitive	26	0.19	No
	4	300		ND	32	0.15	Yes
	7	86	kinetic with <i>p</i> -NPG5 substrate	ND	22	0.25	Yes
	8	97		ND	27	0.20	Yes
	9	93		ND	26	0.21	Yes
Refined	16	298	kinetic with GalG2CNP substrate	ND	21	0.23	Yes
	17	300		ND	22	0.22	No
	18	17		non-competitive	22	0.30	Yes
	20	27		competitive	34	0.17	Yes

IC_{50} : half maximal inhibitory concentration; LE: ligand efficiency which is calculated by the equation $LE = (1.37 / HA) \times pIC_{50}$; ND: not determined; HA: heavy (non-hydrogen) atoms.

Ligand efficiency (LE) is an important and useful metric for evaluating virtual screening and the optimized compounds [181]. It becomes widely used in correlation with the potency of drug-like compounds in all stages of drug discovery. Based on a definition, LE is the

calculated binding free energy (ΔG^0) of a ligand normalized by its molecular size, a change in free energy of binding of -1.37 kcal/mol is equivalent to a 10-fold increase in affinity. This follows that $LE = \Delta G^0/HA$, which is practically approximated into the equation $LE = (1.37/HA) \times pIC_{50}$ [128, 167, 168]. The latter was used here for ligand efficiency calculations.

A mini-perspective study recently published by Zhu, T. *et al.* in 2013 [181] stated that an LE value of 0.3 for a drug-like compound (molecular weight of 500 or 35 HA) corresponds to about 10 nM activity is practically unrealistic for initial hit identification. The same reference stated that LE values of 0.32, 0.25, and 0.19 are recommended for compounds with $HA \leq 18$, $HA = 19 - 25$, and $HA = 26 - 35$, respectively.

The LE of the positive control in this work exhibits a value of 0.13, taking the bound form of acarbose (pseudohexasaccharide, $HA = 65$) for LE calculations. This rearranged acarbose product exerts the inhibition instead of the original acarbose (chapters 1.4.2) since this modified form is observed in the active site of PPA (PDB: 1OSE [156]) and HSA (1MFV [49]). All inhibitors listed in Table 13 display LE values higher than the one of the bound acarbose (Figure 44), and in agreement with LE values suggested by Zhu, T. *et al.* in 2013 [181].

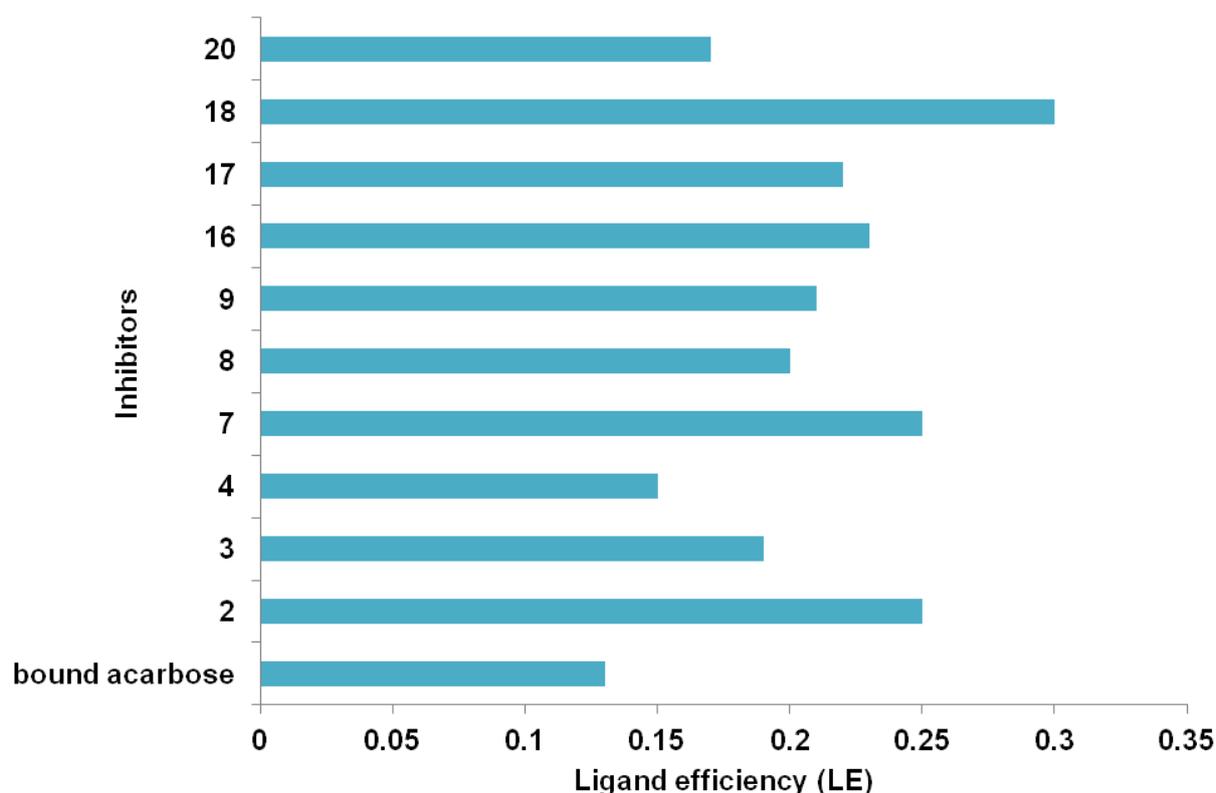


Figure 44: Diagram of ligand efficiency values for positive control and identified α -amylase inhibitors.

Inhibitors **2**, **7**, **16**, **17**, and **18** have either 21 or 22 heavy atoms (HA) that belong to the range of HA = 19 - 25 with a LE value of 0.25 to be considered interesting, according to the recommendations of Zhu, T. *et al.* [181]. While compounds **2** and **7** reach the recommended LE value, compound **18** has even better value of 0.3 than recommended (0.25). Compounds **16** and **17** that have LE values lower than 0.25 do not attain the proposed value since their potencies are weak. Likewise, Inhibitors **3**, **4**, **8**, **9**, and **20** have HA belonged to the range from 25 to 35 and this proposes LE values reaching 0.19 to be remarkable as initial hits in virtual screening. While inhibitor **3** attains the suggested LE value, compounds **8** and **9** display values of 0.20 and 0.21 better than recommended (0.19). Compounds **4** and **20** show LE values lower than recommended. Although compound **20** does not reach the minimum proposed LE value, it reveals a potent inhibitory potency ($IC_{50} = 27 \mu\text{M}$) against α -amylase. The effectiveness of our computational approaches for identifying novel α -amylase fragments/small molecules with drug-like properties is successfully proved. Our hypothesis of blocking the catalytic center for the development of α -amylase inhibitors is biologically validated by determining the competitive inhibition modes of **2**, **3**, **20**, **26**, and **30**. Nevertheless, binding modes discovered by docking studies for the identified α -amylase inhibitors should be further validated with experimental X-ray crystallography. In the same manner to the previous evaluation, IC_{50} cutoff between 50 μM and 260 μM was used to evaluate six out of 12 analogues of compound **2** (Table 14).

Table 14: Evaluation of compound 2 analogues identified with similarity search

initial inhibitor	analogue code	IC_{50} (μM)	activity ratio	inhibition type	HA	LE	LE ratio	novelty of scaffold
Compound 2 ($IC_{50} = 138 \mu\text{M}$; HA = 21; LE = 0.25)	25	165	0.83	ND	24	0.22	0.88	No
	26	100	1.38	competitive	23	0.24	0.96	No
	27	100	1.38	mixed	23	0.24	0.96	No
	30	50	2.76	competitive	26	0.23	0.92	Yes
	31	62	2.22	ND	22	0.24	0.96	No
	32	260	0.53	ND	27	0.18	0.72	No

IC_{50} : half maximal inhibitory concentration; ND: not determined; HA: heavy atoms; LE: ligand efficiency; Activity ratio = IC_{50} for compound **2** / IC_{50} of an analogue; LE ratio = LE of an analogue / LE of compound **2**.

Compounds **30** - **32** share similar scaffolds and hence the one of **30**; as the most active and competitive identified analogue among others; is elected as a novel backbone. Similarly, compounds **25** - **27** hold the same scaffold of compound **2**. However, all identified analogues

have never been reported as α -amylase inhibitors before this work. Assessment was carried out using parameters published by Zhu, T. *et al.* in 2013 [181]. They include the activity ratio and improvement in ligand efficiency (LE ratio).

Applying the previously suggested LE values as metrics for assessment, none of these analogues attain the minimum recommended values, except compound **30** that has 26 HA with a LE value of 0.23 higher than suggested (0.19). However, LE ratios for all analogues have values close to one (Figure 45).

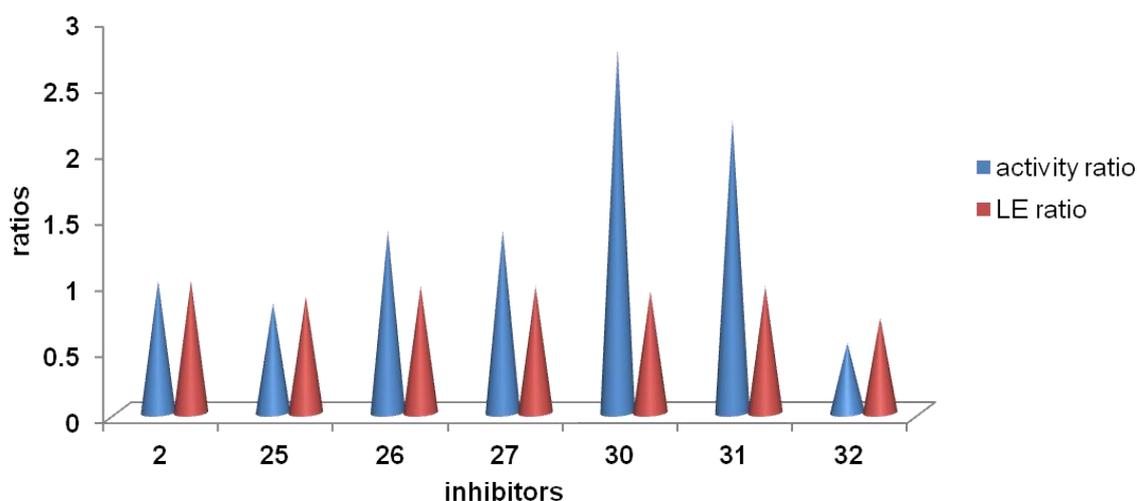


Figure 45: representations of improvements in ligand efficiency and inhibitory potency of compound 2 analogues.

This means that the optimization process preserves the drug likeness of the identified analogues without LE improvement. Furthermore, IC_{50} values of inhibitors **26**, **27**, **30**, **31** are 100 μ M, 100 μ M, 50 μ M, and 62 μ M, respectively, and much better than compound **2** that exhibits an IC_{50} of 138 μ M. The activity ratios for these analogues are mostly improved (Table 14). Among them, compound **30** reaches up to 3-fold improvement better than compound **2** (Figure 45). This assures the success of the optimization process that was performed for compound **2**.

Basically, analogues of compound **2** demonstrate interesting SAR with different substituents (chapter 4.3.6). Substitution by the *p*-dimethylamino or *m*-acetyl groups in ring A (*N*-phenyl ring) increases the inhibition by about 1.4-folds. In contrast, the presence of the *p*- or *m*-carboxyl group diminishes the potency up to 5-folds and the *p*-sulfamoyl group bears the

activity against the enzyme. Interestingly, a 29-fold decrease in activity is observed when ring B (thiazole ring) is substituted with the pyrazole ring.

Substitutions in the 5-benzylidenethiazolidin-4-one scaffold of compounds **30** - **33** exert different effects on α -amylase activity. The *m*-carboxyphenyl group of **31** decreases the potency compared to the *m*-nitrobenzyl group in **30**. However, compound **31** with an IC_{50} value of 62 μ M still remains better inhibitor than compound **2** (IC_{50} = 138 μ M). Also, substitutions in the previously mentioned scaffold with the *m*-carboxyphenylamino moiety in case of compound **32** and 4-acetylpiperazinyl ring in case of compound **33** show 5-folds and 14-folds decrease in α -amylase inhibition, respectively.

Consequently, the simplicity of these identified chemical structures as α -amylase inhibitors and their low molecular weight shall allow for using them as basis for developing novel binders. Decorating these fragments with functional groups identified from SAR can be achieved by chemical synthesis as one possibility to obtain more potent α -amylase inhibitors.

6 Conclusion and Prospective

The purpose of this research was to discover novel small α -amylase inhibitors using computer-aided drug design approaches. This necessity comes from the fact that available α -amylase inhibitors are mostly sugar-based ligands with unsuitable properties for oral administration. Known drug-like binders are still limited in number with scarce information about the structural requirements for α -amylase inhibition. The available data could be efficiently investigated using molecular modeling. This was implemented in two virtual screening campaigns as initial and refined approaches.

The catalytic core of the active site was exploited as a starting point to develop a validated predictive 3D pharmacophore for virtual screening of a library of commercial molecules. Among 13 rationally selected hits and biologically tested, six novel non-sugar α -amylase inhibitors were identified as small ligands with drug-like properties. Likewise, information obtained from our initial virtual screening and recent publications by other research groups during this project was compiled to conduct refined virtual screening campaigns. Using four pharmacophores originated from different scaffolds alongside with multi-step 3D refinements was clearly profitable owing to the discovery of four novel inhibitors out of only seven hits selected for biological testing. These promising results validate the ability of our hypothesized pharmacophores to select potential α -amylase binders.

Taking the IC_{50} cutoff of 300 μ M, ten out of 20 biologically tested molecules in both approaches exerted α -amylase inhibition (50 % hit rate) in mid- and low micro-molar range (the best IC_{50} = 17 μ M). The identified compounds were never reported as α -amylase inhibitors before. Interestingly, 8 novel scaffolds are stated for the first time as promising structures for α -amylase inhibition. These findings successfully show the quality of virtual screening approaches and confirm the power of our molecular modeling methodologies for identifying novel α -amylase inhibitors with drug-like properties.

A further step was carried out as a preliminary optimization for the most active and competitive inhibitor discovered in the initial virtual screening. This was achieved by 2D analogue-similarity search aiming at finding other potent α -amylase inhibitors and enriching the structure activity relationship. Six out of 12 biologically tested compounds exhibit inhibitory potency in low micro-molar range with IC_{50} values between 50 μ M and 260 μ M and about 3-fold improvement in inhibition compared to the used query inhibitor.

The discovered inhibitors were compared and discussed using the ligand efficiency metric (LE). Due to their small size, all inhibitors reported in this work display LE better than acarbose; a well-known anti-diabetic sugar-based drug; that was used as a positive control in this study.

In brief, a total of 16 out of 32 biologically tested compounds display α -amylase inhibition in micro-molar range with good LE and inhibitory potency. Available information of sugar-based inhibitors were investigated computationally and resulted at the end in the discovery of novel non-sugar α -amylase inhibitors. Hence, small organic compounds can bind to the challenging large binding site with high inhibitory affinity. The competitive inhibition mode for these inhibitors validates our hypothesis that blocking the catalytic core can lead to the discovery of novel α -amylase inhibitors. Thus, our 3D pharmacophore-based virtual screening approaches can be used for the discovery of potent drug-like α -amylase inhibitors.

As a next step, an optimization approach will be conducted by compiling all information obtained from this work to perform a 3D shape- and feature-based similarity search. The best competitive α -amylase inhibitors will be used as queries for high throughput ligand-based virtual screening of commercial chemical library. Then, molecular docking for the best ranked hits will be conducted to investigate their binding modes. Careful visual inspection as well as the quality of interactions with residues inside the active site will be performed as prioritization criteria to select candidates for biological testing. Another perspective will be using information obtained from structure activity relationships of the identified inhibitors to synthesize chemical compounds with potential α -amylase inhibition.

7 Experimental part

7.1 Binding mode elucidation for inhibitors identified by collaboration partners

Natural α -amylase inhibitors identified by collaboration partners were prepared in MOE 2010.10 [155] and the 3D coordinates were generated using CORINA 3.4 [182, 183]. While cyanidin and steviol compounds were docked into HSA active site (PDB entry code 1MFV [49]), verbascoside was docked into the three allosteric sites of HPA (PDB code 4GQQ [25]). Docking studies were carried out using the software GOLD 5.1 [127, 184]. All water molecules as well as co-crystallized ligands were deleted and hydrogen atoms were added to protein structures before docking. No constraints were introduced and the default parameters were used in this study (binding site: within 10 Å around the co-crystallized ligand, scoring function: GoldScore [133], genetic algorithm: 100% search efficiency). 100 Poses were generated for each compound. Validation for the docking experiments by GOLD 5.1 was performed by reproducing the conformation of the relevant co-crystallized inhibitor. Thus, the modified acarbose co-crystallized with 1MFV was docked into the active site with RMSD 1.1 Å (Appendix 10.6, Figure 48) between heavy atoms of the original and docked conformations. Similarly, the co-crystallized ethyl caffeate was docked individually into allosteric sites 1, 2, and 3 of HPA codes 4GQQ and show respectively RMSD values of 0.5 Å, 1.07 Å, and 1.46 Å between heavy atoms of the cocrystals and docked poses (Appendix 10.6).

All residues within 7 Å from any of the outside atoms of the co-crystallized ligand were determined as active site using the software LigandScout 3.1 [114, 115, 159]. All docked poses were minimized with MMFF94 force field in LigandScout 3.1. The 3D pharmacophore of co-crystallized inhibitors and the quality of the superposition of each pose with the relevant co-crystallized ligand were used to prioritize the most relevant poses. LigandScout 3.1 was also used for analysis, pharmacophore creation, visualization and elucidation.

7.2 Structure-based virtual screening

7.2.1 Commercial chemical databases

3D pharmacophores developed in chapters 4.3.1 and 4.3.8 were used for virtual screening of commercially available chemical databases assembled from different vendors which

comprise 1,762,189 and 1,346,275 molecules in the initial and refined virtual screening campaigns, respectively (Table 15).

Table 15: Commercial chemical databases used in virtual screening approaches

Vendor	Amount	Used in	Reference
Analyticon	30,352	initial and refined virtual screening	[186]
Asinex	465,543	initial and refined virtual screening	[187-189]
Chembridge	646,018	initial and refined virtual screening	[190]
Specs	204,362	initial and refined virtual screening	[191]
Bionet	42,660	initial virtual screening	[192]
Life Chemicals	372,071	initial virtual screening	[193]
Prestwick	1,183	initial virtual screening	[194]

These libraries were computationally prepared using Standardizer 5.8.0 [195] to correct protonation states. Then, the command-line program idbgen was used to convert libraries into ligand-database format (ldb) to be used as inputs for virtual screening in LigandScout 3.1 [114, 115, 159]. The software Omega [196] with FAST parameters was used to generate up to 25 conformations per molecule. For shape- and feature-based similarity search by ROCS 3.1.2 [178], commercial libraries were transformed into 3D coordinates using CORINA 3.4 [183] and then the software OMEGA 2.4.6 [177] was used to create 25 conformations per molecule as oeb.gz format as inputs for ROCS 3.1.2.

7.2.2 Validation of developed pharmacophores

Initial virtual screening (round 1)

Pharmacophore models created and developed in chapter 4.3.1.1 were validated using 19 biologically known active and 55 inactive α -amylase inhibitors collected from ChEMBL database [164] and literature (Appendices 10.3 and 10.4). All these collected molecules were built and minimized using MOE 2010.10 [155]. 3D coordinates were generated using CORINA 3.4 [183]. The command-line tool idbgen was used to generate database files to be used as inputs for screening in LigandScout 3.1 [159]. The software Omega [196] was used to generate 25 conformations per compound. Based on the best early enrichment factor ($EF_{1\%}$) and $AUC_{100\%}$, one pharmacophore was selected.

A further validation step was performed in chapter 4.3.1.2 for the best chosen model using drugs and biological molecules commercially available in the Derwent World Drug Index 2005 (WDI, www.thomsonreuters.com), MDDR2009 (www.symyx.com), and DrugBank [166]

databases. These commercial databases were similarly prepared as previously mentioned libraries.

Refined virtual screening (round 2)

During this work, new small molecules have been published as α -amylase inhibitors (chapter 1.4). Therefore, validation dataset used in the initial virtual screening was updated to be used in the refined virtual screening. The active-list was re-constructed to: include all known compounds with IC_{50} or K_i less than 15 μ M (IC_{50} of CHEMBL2203334, the most recent active molecule published in 2012 [87]); exclude all compounds with molecular weight less than the one of scutellarein (molecular weight = 286, the best flavonoid published as α -amylase inhibitor with IC_{50} value of 9.6 μ M [20]); incorporate actives found by intensive search in literature to enrich the active-list.

Therefore, the updated dataset assembled from literature including 19 known active and 60 inactive α -amylase inhibitors (Appendices 10.3 and 10.4) were prepared as Idb format to be used as inputs for screening in LigandScout 3.1.

A set of decoys comprising 254 structures was created from ZINC database [197] using a Knime workflow called myDecoyFinder developed in-house by Dr. Susanne Dupré. The updated 19 known active inhibitors were used as a basis to create decoys with the following settings:

- Molecular weight (± 25 Da)
- Mannhold LogP (± 1.0)
- Number of H-bond donors (± 1)
- Number of H-bond acceptors (± 2)
- Number of free rotatable bonds (± 1)
- Tanimoto coefficient (< 0.75)

The output of this workflow was saved as smiles format (smi). Then, the database files for the updated active and inactive lists, as well as created decoys, were treated in the same manner with the tool idbgen and software Omega [196] to obtain input files for screening in LigandScout 3.1 [159]. Further validation was performed for the developed pharmacophores in the refined virtual screening by using commercial drugs and biological compounds available in WDI2005 database (64,588 compounds) as decoys.

7.2.3 Target preparation for pharmacophore creation

X-ray structures of α -amylase derived from the PDB used to build common-based pharmacophores (chapter 4.3.1.1) were 3OLD, 3OLE, 3OLG, 3OLI [36], 3IJ9, 3IJ7 [31], 3BAJ, 3BAY [198], 1XH2, 1XH0 [26], 1XD0, 1XD1, 1XCX, 1XCW [32], 1MFV [49], and 1OSE [156].

The software LigandScout 3.1 [159] was used for visualization and interpretation of geometries of co-crystallized inhibitors. Co-crystallized inhibitors as well as calcium and chloride ions were kept inside the active site while all other ligands such as MPD (protecting agent during crystallization [199]) were deleted from binding sites.

7.2.4 Pharmacophore creation and virtual screening

Initial virtual screening

Using LigandScout 3.1, a preliminary 3D pharmacophore was generated and analyzed for each co-crystallized inhibitor. Chemical features in each model were deleted or modified to have common features as many as possible. Therefore, the number of chemical features was reduced in order to have more specific models. Spatial restriction for 3D models was taken into account either by deleting or modifying the size of the exclusion volumes which prevent a ligand from occupying space restricted for amino acids of the active site.

All models were aligned to obtain common-based pharmacophore models. Based on iterative validations (chapter 7.2.2), an improved model was elected for virtual screening (chapter 4.3.1.3).

Virtual screening for commercially available chemical databases was performed for 1,762,189 molecules using LigandScout 3.1. Compounds matching the selected 3D model were further reduced using pharmacophore fit score and 2D descriptors (i.e. HBD \leq 2-5, HBA \leq 10, Mol.wt. \leq 700 and ClogP \leq 5). Then, 2292 resulted hits were transformed into 3D coordinate structures (sdf format) using CORINA 3.4 [182] for the next step.

Two parallel steps were performed for the resulted hits, docking and structural similarity clustering. First, GOLD 5.1 [184] was used to dock 2,292 virtual hits into α -amylase active site (PDB 3OLE [36]) using default settings and GoldScore [133] as scoring function to create 10 poses for each compound. In LigandScout 3.1, resulted poses were minimized with the force field MMFF94 and then prioritized using the developed pharmacophore. Second, the software JKlustor 5.8.0 [139] was used to perform structural similarity clustering for compounds obtained from virtual screening to select at the end 17 candidates based on

structural diversity and commercial availability. Visual inspections for best docking poses as well as diversity in chemical structures were used to select 14 hits for biological experiments.

Refined virtual screening

The 19 known active α -amylase inhibitors (Appendix 10.3) were divided into six groups based on their chemical diversity (chapter 4.3.8.1). A representative-structure for each group was selected to build a 3D pharmacophore. The first two models were built based on the co-crystallized acarviosine-glucose (PDB code 1XCW [32]) and myricetin flavonoid (PDB: 4GQR [25]) and were taken as representatives for carbohydrate- and flavonoid-based inhibitors, respectively.

Four other 3D pharmacophores were built from non co-crystallized active inhibitors (salacinol [175], 1,2,3,4,6-pentagalloyl- β -D-glucose [84], DDBT [83], and ChEMBL2203334 [87]). These inhibitors were docked into the active site of PDB code 3OLE [36] using GOLD 5.1 with default settings and Goldscore. All resulted poses were minimized with MMFF94 in LigandScout 3.1 and the best poses fulfilling the pharmacophore features of the co-crystallized acarviostatin II03 inhibitor were selected.

Iterative development and optimization were performed for each model using the previously validating databases. Model of ChEMBL2203334 was excluded because it fails to retrieve actives during validation and gives EF value of null. Thus, five pharmacophores were obtained for the refined virtual screening.

Developed pharmacophore models were separately used for virtual screening of a library comprising 1,346,275 commercial chemical molecules as described in chapter 7.2.1. Five hit-lists were obtained for the five models. First, the command-line tool FILTER 2.1.1 [176] was used to retain drug-like molecules. 2D descriptors derived from 141 best-selling, non-antibiotics, prescription drugs were implemented in FILTER 2.1.1 as default settings. These descriptors are basically physical properties (such as molecular weight and logP), atomic and functional group content (such as content of heteroatom), and molecular graph topology (such as number and size of ring system). Thus, default 2D descriptors were applied in this step. Second, substructure search using the software Instant JChem 5.8.0 [140] was used to exclude sugar-based compounds. The hit list derived from pentagalloyl-based pharmacophore could not pass the previous steps. Compounds derived from 1XCW-based model, 4GQR-based model, salacinol-based model, and DDBT-based model that passed the previous steps were ranked using the software ROCS 3.1.2 [178]. This was performed by using the software CORINA 3.4 to convert compounds into 3D coordinates and then, the software OMEGA 2.4.6 [177] to create 25 conformations per compound as oeb.gz format to

be used as inputs for the program ROCS 3.1.2. Queries were two types: (i) X-ray conformations of acarviosine-glucose and myricetin with their developed chemical features were used for hit-lists obtained from 1XCW-based and 4GQR-based hit-lists, respectively, and (ii) docking poses of salacinol and DDBT with their pharmacophoric features were used for salacinol-based and DDBT-based hit-lists, respectively.

Best hits ranked in the first 100 hits were kept and prepared for docking. A total of 299 hits resulted from the ranking step were transformed into 3D coordinates using CORINA 3.4 and then docked into the active site (PDB: 3OLE) using GOLD 5.1 as described before.

Resulted poses were minimized inside the active site with the force field MMFF94 implemented in LigandScout 3.1. Previously used pharmacophores were used to prioritize the relevant poses. Visual inspection, fulfillment of the relevant pharmacophore, chemical diversity, and commercial availability were used as basis for selecting seven candidates for biological testing.

7.3 Biological testing

7.3.1 Biological testing using the substrate *p*-NPG5 for PPA: First assay

This kinetic method was performed in-house according to procedures published by our collaboration partner Prof. Dr. Matthias F. Melzig, Freie Universität Berlin [75].

Reagents: HEPES was purchased from Lancaster (Mühlheim, Germany). Sodium chloride was obtained from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). DMSO and calcium chloride were obtained from Merck KgaA, Darmstadt, Germany. Acarbose was purchased from AK Scientific, Inc (Union City, CA, United States). Compounds **1** - **7** were purchased from Asinex Europe B.V. (Rijswijk, The Netherlands). Compounds **8** and **9** were obtained from Chembridge Corporation (San Diego, CA, United States). Compounds **10** and **11** were obtained from Key Organics (Camelford Cornwall, United Kingdom). Compounds **12** - **14** were granted from Analyticon (Potsdam, Germany).

Enzyme. α -Amylase (EC 3.2.1.1, PPA) from porcine pancreas (Type I-A) from Sigma-Aldrich Chemie GmbH (Steinheim, Germany) was used.

Substrate. *p*-Nitrophenyl- α -D-maltopentaoside (*p*-NPG5) was obtained from Megazyme (Bray, Co.Wicklow, Ireland). Liberation of *p*-nitrophenol (*p*-NP) from *p*-NPG5 was monitored continuously with an Infinite F200 Microplate Reader and Spectra-Flour Microplate reader, Tecan Group Ltd., Austria under temperature-controlled conditions (37°C) at 405 nm.

Kinetic analysis. Kinetic experiments were conducted in 96-well-plate at 37°C in 50 mM HEPES buffer pH 7.1 containing 70 mM NaCl and 1 mM CaCl₂.

The substrate and inhibitors were dissolved in DMSO and then diluted with the buffer. The maximum DMSO concentration in the well did not exceed 10% without affecting the experiments [200]. The substrate (50 μ l, 25 mM) and each tested hit (25 - 500 μ M) were mixed together and the reaction was initiated by adding 50 μ l of PPA (previously prepared as 1 part of enzyme to 100 parts of HEPES buffer, equivalent to 5.680 μ M) to the incubation medium. The total volume in the well was adjusted by buffer to 200 μ l. The 100 % activity of α -amylase was determined without using the inhibitor and used as a control. Blank solutions were measured and subtracted from the measurement of the sample to take into account the effect of the used reagents.

Calculations were obtained using the equation: Inhibition % = $1 - [(A_1 - A_2) / A_0] \times 100$, where A_1 is the absorbance of the sample (a hit, the substrate, buffer, and PPA enzyme), A_2 is the absorbance of the blank (a hit, the substrate and buffer), and A_0 is the absorbance of the control (the substrate, buffer, and PPA enzyme). Each assay was conducted at least two times with duplicate samples.

Measurement interval was 5.5 minutes with a total period of 55 minutes. The increase in the absorbance of *p*-NP liberated by PPA was measured with respect to time to give a progress curve for the reaction. The slope of the curve is decreased in the presence of an inhibitor.

A concentration-effect curve (inhibition % in Y axis and inhibitor concentration in X axis) was obtained for each inhibitor using linear regression analysis in Microsoft Excel to determine the IC_{50} .

Before conducting the assay, the linearity curve that represents the absorbance of the product (*p*-nitrophenol) with respect to time interval was determined using 25 mM of *p*-NPG5 substrate and PPA (1 part of enzyme : 100 total parts of HEPES buffer), Figure 46.

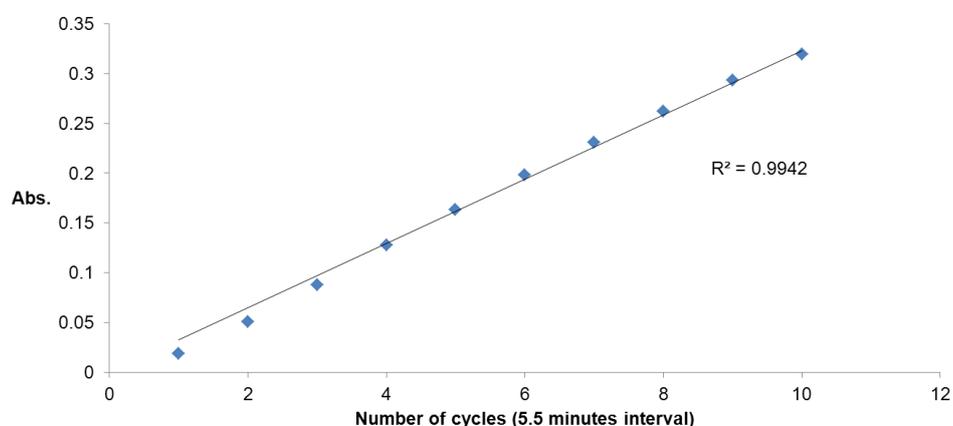


Figure 46: Linearity curve for *p*-nitrophenol resulted from the hydrolysis of *p*-NPG5 by PPA with respect to time.

7.3.2 Biological testing with GalG2CNP substrate for HSA: Second assay

This kinetic assay was carried out based on procedures previously described by collaboration partner Dr. Gyöngyi Gyémánt, University of Debrecen [154].

Enzyme. α -Amylase (EC 3.2.1.1, HSA) from human saliva (Type IXA) from SIGMA was used and it gave a single band on sodiumdodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and possessed no α - and β -glycosidase activity.

Substrate. 2-Chloro-4-nitrophenyl-4-O- β -D-galactopyranosylmaltoside (GalG2CNP) was purchased from Diagnosticum ZRt, Hungary. Liberation of chloronitrophenol (CNP) from GalG2CNP was monitored continuously with a UV-VIS spectrophotometer (JASCO V550) under temperature-controlled conditions (37°C) at 400 nm.

Kinetic analysis. Kinetic experiments were carried out at 37°C in 50 mM MES buffer pH 6.0 containing 5 mM of $\text{Ca}(\text{OAc})_2$, 51.5 mM of NaCl and 152 mM of NaN_3 . The substrate (0.75 - 4 mM) and inhibitor were mixed together and the reaction was initiated by adding HSA (2 nM) to the incubation medium (total volume of 500 μl).

The measurements without inhibitor were carried out in the presence of DMSO not exceeding 2.5 %. The increase of absorbance of CNP liberated by HSA was measured continuously at 400 nm using the Parallel Kinetics Analysis program of a JASCO V550 spectrophotometer. Concentration-response plots were used to determine the effects of the inhibitor on the enzymatic reaction and IC_{50} values of inhibitors. These experiments were performed at constant enzyme and substrate concentrations. Fractional activity (Y axis) was plotted as a function of inhibitor concentration (X axis). The data were fit using a standard four-parameter logistic nonlinear regression analysis of Grafit software. The type of inhibition was determined by the Lineweaver-Burk plot.

7.4 Modeling differences in IC_{50} values in the initial virtual screening

Substrates *p*-NPG5 and GalG2CNP, positive control as well as compounds **2**, **3**, **4**, **6**, **7**, **8**, and **9** were docked into α -amylase (PDB: 3OLE [36]) aiming at explaining the differences in potencies when using two different assays. Docking parameters of GOLD 5.1 [184] was used with its default parameters and scoring function GoldScore. Analysis and visualization for resulted poses were performed in LigandScout 3.1. The best plausible binding mode for each ligand was selected based on the fulfillment of the developed 3D pharmacophore and

interactions with the catalytic residues as well as the quality of superposition with the co-crystallized ligand. Gaussian similarity score implemented in LigandScout 3.1 was used to quantify the steric overlap between each substrate and tested compounds. For this purpose, each selected pose of substrates *p*-NPG5 and GalG2CNP was inserted individually into the active site (PDB code 3OLE) replacing the co-crystallized ligand to be used as a reference for calculations (chapter 4.3.3, Appendix 10.5).

7.5 Molecular Docking

7.5.1 Target preparation for molecular docking

All docking studies were performed using GOLD 5.1 [184]. α -Amylase structures with PDB codes 1MFV [49] and 4GQQ [25] were used for docking natural inhibitors identified by collaborations partners (chapter 4.2). Virtual hits and analogues selected in the initial virtual screening were docked into PDB code 3OLE [36]. Compounds selected from the refined virtual screening were docked into PDB codes 3OLE, 4GQQ and 4GQR [25]. All crystal structures used for docking were protonated, and water molecules and co-crystallized ligands were deleted in GOLD 5.1 before docking.

7.5.2 Preparation of compounds for docking

All compounds docked with GOLD 5.1 [184] were converted into 3D coordinates using CORINA 3.4 [183] before docking.

7.5.3 Docking studies for inhibitors discovered by the initial virtual screening

GOLD 5.1 was used for docking molecules into α -amylase active site (PDB: 3OLE [36]). The protein structure was prepared in GOLD 5.1 as stated in 7.5.1.

Validation of docking experiments was performed by reproducing the binding mode of the co-crystallized inhibitor showing RMSD of 0.8 Å between the original and docked poses (Appendix 10.6, Figure 48). All docking experiments were carried out without constraints. As described before, default settings were used to generate 100 docking poses for each compound unless it is stated. Then, all created poses were analyzed inside the pocket using LigandScout 3.1.

7.5.4 Docking studies for inhibitors discovered by the refined virtual screening

As mentioned before, α -amylase structures and docked compounds were first prepared as described in chapters 7.5.1 and 7.5.2, respectively.

Default settings and GoldScore were applied in GOLD 5.1 [184] to create 100 poses for each compound, unless it is stated. α -Amylase active site with PDB code 3OLE [36] was used as a template for docking. All created poses were imported into LigandScout 3.1 and minimized inside the binding pocket with the force field MMFF94.

Because compound **18** shows non-competitive inhibition, a second possibility was investigated by docking the ligand into allosteric binding sites 1, 2, and 3 of PDB code 4GQQ [25], previously described in chapter 4.2.3. The pharmacophore features of the co-crystallized ethyl caffeate were used to prioritize poses of compound **18**.

A second binding mode was obtained by docking compound **20** into the active site of the PDB code 4GQR. The docking experiment by GOLD 5.1 [127, 184] was validated for 4GQR and shows RMSD of 2.1 Å between the heavy atoms of the co-crystallized ligand and docked conformation (Appendix 10.6, Figure 50).

The crystal structure and compounds were prepared as described before. LigandScout 3.1 was used to visualize and analyze resulted poses.

7.6 Similarity search for the best competitive inhibitor in the initial screening

Aiming at finding new hits similar to compound **2**, similarity searching from an in-house chemical database originally obtained from commercial vendors was conducted. 12 compounds were selected for biological testing (Appendix 10.11, compounds **22 - 33**).

Chosen analogues as well as protein structure coded 3OLE were prepared for docking as described in chapters 7.5.2 and 7.5.1, respectively. Default settings of GOLD 5.1 and the scoring function GoldScore were applied to create 10 poses per ligand. Then, resulted poses were analyzed in LigandScout 3.1. All poses were minimized inside the pocket and prioritized using the pharmacophore features of compound **2**. The best pose for each hit was selected based on the highest pharmacophore fit scores.

8 Summary

α -Amylase, a digestive enzyme secreted from salivary glands and pancreas, initiates starch catabolism resulting into postprandial blood glucose spikes. Hyperglycemia is closely associated with metabolic diseases such as obesity and diabetes, as well as cardiovascular mortality [91, 201]. Therefore, a better control of α -amylase activity is a promising therapeutic strategy for controlling unbalanced starch degradation and related pathologic consequences. The objective of this work was the discovery of novel small α -amylase inhibitors.

To achieve our goal, a computer-aided drug design strategy was developed, in which two 3D pharmacophore-based virtual screening approaches were conducted. In the initial virtual screening, all available information on binding modes of co-crystallized carbohydrate-based α -amylase inhibitors were assembled and used to develop common feature-based 3D pharmacophore models. After validation with a dataset of biologically known active and inactive α -amylase inhibitors, the best generated model was used for virtual screening of commercial chemical databases comprising about two million compounds. In order to reach a reasonable amount of molecules for biological testing, virtual hits were filtered using 2D drug-like descriptors, docking and structural clustering. This initial virtual screening led to the final selection of 14 for biological testing. Six out of 13 biologically tested compounds (1 hit was excluded due to insolubility) display α -amylase inhibition with IC_{50} values between 86 μ M and 300 μ M.

In parallel and during this project, novel small α -amylase inhibitors have been published by other research groups, and hence their information was included into a second round of virtual screening. The validating dataset was updated and categorized into five clusters based on scaffold type. The most potent small inhibitor from each cluster was selected and five different pharmacophore models were generated. These models were validated and then used individually for virtual screening. More than one million commercial chemical molecules were virtually screened resulting in five hit-lists, one for each 3D model. One hit-list was excluded since its compounds could not tolerate one filtering step with predefined 2D descriptors. Then, each lead structure inhibitor used for pharmacophore development was employed to build a 3D query applied for ranking the relevant resulted hits. This was achieved by 3D similarity search to pick molecules that are the closest to the relevant query in terms of shape and chemical features. Then, the best ranked hits were docked and analyzed to select seven virtual hits for biological experiments. Four out of seven biologically

tested compounds reveal promising α -amylase inhibition with IC_{50} ranging from 17 μ M to 300 μ M.

An optimization step was performed for the most active and competitive α -amylase inhibitor discovered by the first round virtual screening aiming at finding more potent ligands and rationalizing their structure activity relationships. 2D similarity search was carried out using an in-house chemical database and 12 analogues were selected for biological testing. Six out of 12 biologically tested compounds demonstrate α -amylase inhibition with IC_{50} values ranging from 50 μ M to 260 μ M.[14]

At the end of this project, 33 compounds were selected using a rational computationally driven approach. In total, 16 novel non-sugar small compounds with α -amylase inhibition (IC_{50} range: 17 - 300 μ M) were identified using kinetic assays. The binding modes of all α -amylase inhibitors discovered in this work were elucidated using molecular docking, leading to the discovery of the structural requirements for an optimal α -amylase inhibition.

Moreover, natural α -amylase inhibitors (cyanidin derivatives, steviol extract, and verbascoside) identified by collaboration partners with previously unknown binding modes were investigated by docking and pharmacophore modeling and their mechanism of inhibition were elucidated.

9 References

1. Purich, D.L., *An Introduction to Enzyme Science*, in *Enzyme Kinetics: Catalysis & Control*, D.L. Purich, Editor. Elsevier: Boston, 2010. **1**. p. 1-51.
2. Nater, U.M., *Salivary alpha-amylase as a marker for stress*, in *Psychoneuroendocrinology Research Trends*, M.T. Czerbska, Editor. Nova Science Publishers, Inc.: New York, 2007. **1**. p. 117-148.
3. Kumari, A., K. Singh, and A.M. Kayastha, *alpha-Amylase: General Properties, Mechanism and Biotechnological Applications – A Review*. *Curr Biotech*, 2012. **1**(1): p. 98-107.
4. J., T., *Enzymes of Aspergillus oryzae and the application of its amylolytic enzyme to the fermentation industry*. *Ind Eng Chem Biodivers*, 1914. **6**(10): p. 824-28.
5. Berman, H.M., et al., *The Protein Data Bank*. *Nucleic Acids Res*, 2000. **28**(1): p. 235-42.
6. Berman, H., K. Henrick, and H. Nakamura, *Announcing the worldwide Protein Data Bank*. *Nat Struct Biol*, 2003. **10**(12): p. 980.
7. Kuriki, T. and T. Imanaka, *The concept of the alpha-amylase family: structural similarity and common catalytic mechanism*. *J Biosci Bioeng*, 1999. **87**(5): p. 557-65.
8. Low, L.C., *The epidemic of type 2 diabetes mellitus in the Asia-Pacific region*. *Pediatr Diabetes*, 2010. **11**(4): p. 212-5.
9. McCue, P., Y.I. Kwon, and K. Shetty, *Anti-diabetic and anti-hypertensive potential of sprouted and solid-state bioprocessed soybean*. *Asia Pac J Clin Nutr*, 2005. **14**(2): p. 145-152.
10. Subramanian, R., M.Z. Asmawi, and A. Sadikun, *In vitro alpha-glucosidase and alpha-amylase enzyme inhibitory effects of Andrographis paniculata extract and andrographolide*. *Acta Biochim Pol*, 2008. **55**(2): p. 391-8.
11. Van der Maarel, M.J., et al., *Properties and applications of starch-converting enzymes of the alpha-amylase family*. *J Biotechnol*, 2002. **94**(2): p. 137-55.
12. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics*. *Nucleic Acids Res*, 2009. **37**(1): p. 233-8.
13. Henrissat, B., *A classification of glycosyl hydrolases based on amino acid sequence similarities*. *Biochem J*, 1991. **280**(2): p. 309-16.
14. Henrissat, B. and A. Bairoch, *New families in the classification of glycosyl hydrolases based on amino acid sequence similarities*. *Biochem J*, 1993. **293**(3): p. 781-8.
15. Henrissat, B. and A. Bairoch, *Updating the sequence-based classification of glycosyl hydrolases*. *Biochem J*, 1996. **316**(2): p. 695-6.
16. Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases*. *Structure*, 1995. **3**(9): p. 853-9.
17. Henrissat, B. and G. Davies, *Structural and sequence-based classification of glycoside hydrolases*. *Curr Opin Struct Biol*, 1997. **7**(5): p. 637-44.
18. Takata, H., et al., *Action of neopullulanase. Neopullulanase catalyzes both hydrolysis and transglycosylation at alpha-(1----4)- and alpha-(1----6)-glucosidic linkages*. *J Biol Chem*, 1992. **267**(26): p. 18447-52.
19. Groot, P.C., et al., *The human alpha-amylase multigene family consists of haplotypes with variable numbers of genes*. *Genomics*, 1989. **5**(1): p. 29-42.
20. Lo Piparo, E., et al., *Flavonoids for controlling starch digestion: structural requirements for inhibiting human alpha-amylase*. *J Med Chem*, 2008. **51**(12): p. 3555-61.
21. Brayer, G.D., Y. Luo, and S.G. Withers, *The structure of human pancreatic alpha-amylase at 1.8 Å resolution and comparisons with related enzymes*. *Protein Sci*, 1995. **4**(9): p. 1730-42.

22. Ramasubbu, N., et al., *Structure of human salivary alpha-amylase at 1.6 Å resolution: implications for its role in the oral cavity*. Acta Crystallogr D Biol Crystallogr, 1996. **52**(3): p. 435-46.
23. Motyan, J.A., et al., *Computer-aided subsite mapping of alpha-amylases*. Carbohydr Res, 2011. **346**(3): p. 410-5.
24. Nahoum, V., et al., *Crystal structures of human pancreatic alpha-amylase in complex with carbohydrate and proteinaceous inhibitors*. Biochem J, 2000. **346**(1): p. 201-8.
25. Williams, L.K., et al., *Order and disorder: differential structural impacts of myricetin and ethyl caffeate on human amylase, an antidiabetic target*. J Med Chem, 2012. **55**(22): p. 10177-86.
26. Maurus, R., et al., *Structural and mechanistic studies of chloride induced activation of human pancreatic alpha-amylase*. Protein Sci, 2005. **14**(3): p. 743-55.
27. Numao, S., et al., *Probing the role of the chloride ion in the mechanism of human pancreatic alpha-amylase*. Biochemistry, 2002. **41**(1): p. 215-25.
28. Davies, G.J., K.S. Wilson, and B. Henrissat, *Nomenclature for sugar-binding subsites in glycosyl hydrolases*. Biochem J, 1997. **321**(2): p. 557-9.
29. Zechel, D.L. and S.G. Withers, *Glycosidase mechanisms: anatomy of a finely tuned catalyst*. Acc Chem Res, 2000. **33**(1): p. 11-8.
30. Brayer, G.D., et al., *Subsite mapping of the human pancreatic alpha-amylase active site through structural, kinetic, and mutagenesis techniques*. Biochemistry, 2000. **39**(16): p. 4778-91.
31. Zhang, R., et al., *Directed "in situ" inhibitor elongation as a strategy to structurally characterize the covalent glycosyl-enzyme intermediate of human pancreatic alpha-amylase*. Biochemistry, 2009. **48**(45): p. 10752-64.
32. Li, C., et al., *Acarbose rearrangement mechanism implied by the kinetic and structural analysis of human pancreatic alpha-amylase in complex with analogues and their elongated counterparts*. Biochemistry, 2005. **44**(9): p. 3347-57.
33. Uitdehaag, J.C., et al., *X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family*. Nat Struct Biol, 1999. **6**(5): p. 432-6.
34. Rydberg, E.H., et al., *Mechanistic analyses of catalysis in human pancreatic alpha-amylase: detailed kinetic and structural studies of mutants of three conserved carboxylic acids*. Biochemistry, 2002. **41**(13): p. 4492-502.
35. Rydberg, E.H., et al., *Cloning, mutagenesis, and structural analysis of human pancreatic alpha-amylase expressed in Pichia pastoris*. Protein Sci, 1999. **8**(3): p. 635-43.
36. Qin, X., et al., *Structures of human pancreatic alpha-amylase in complex with acarviolstatins: Implications for drug design against type II diabetes*. J Struct Biol, 2011. **174**(1): p. 196-202.
37. Ramasubbu, N., et al., *Human salivary alpha-amylase Trp58 situated at subsite -2 is critical for enzyme activity*. Eur J Biochem, 2004. **271**(12): p. 2517-29.
38. Ramasubbu, N., et al., *Structure-function relationships in human salivary alpha-amylase: role of aromatic residues*. Biologia, 2005. **60**(16): p. 47-56.
39. Mishra, P.J., C. Rangunath, and N. Ramasubbu, *The mechanism of salivary amylase hydrolysis: role of residues at subsite S2'*. Biochem Biophys Res Commun, 2002. **292**(2): p. 468-73.
40. Wild, S., et al., *Global prevalence of diabetes: estimates for the year 2000 and projections for 2030*. Diabetes Care, 2004. **27**(5): p. 1047-53.
41. Wilborn, C., et al., *Obesity: prevalence, theories, medical consequences, management, and research directions*. J Int Soc Sports Nutr, 2005. **2**(2): p. 4-31.
42. Orstavik, D. and F.W. Kraus, *The acquired pellicle: immunofluorescent demonstration of specific proteins*. J Oral Pathol, 1973. **2**(1): p. 68-76.

43. Al-Hashimi, I. and M.J. Levine, *Characterization of in vivo salivary-derived enamel pellicle*. Arch Oral Biol, 1989. **34**(4): p. 289-95.
44. Douglas, C.W., *The binding of human salivary alpha-amylase by oral strains of streptococcal bacteria*. Arch Oral Biol, 1983. **28**(7): p. 567-73.
45. Rangunath, C., et al., *Probing the role of aromatic residues at the secondary saccharide-binding sites of human salivary alpha-amylase in substrate hydrolysis and bacterial binding*. J Mol Biol, 2008. **384**(5): p. 1232-48.
46. Scannapieco, F.A., L. Solomon, and R.O. Wadenya, *Emergence in human dental plaque and host distribution of amylase-binding streptococci*. J Dent Res, 1994. **73**(10): p. 1627-35.
47. Nikitkova, A.E., E.M. Haase, and F.A. Scannapieco, *Taking the starch out of oral biofilm formation: molecular basis and functional significance of salivary alpha-amylase binding to oral streptococci*. Appl Environ Microbiol, 2013. **79**(2): p. 416-23.
48. Kandra, L., et al., *Inhibitory effects of tannin on human salivary alpha-amylase*. Biochem Biophys Res Commun, 2004. **319**(4): p. 1265-71.
49. Ramasubbu, N., C. Rangunath, and P.J. Mishra, *Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase*. J Mol Biol, 2003. **325**(5): p. 1061-76.
50. Antipatis, V.J. and T.P. Gill, *Obesity as a Global Problem*, in *International Textbook of Obesity*, P. Bjorntorp, Editor. John Wiley & Sons, Ltd.: Goteborg, Sweden, 2002. **1**. p. 1-22.
51. Lawrence, V.J. and P.G. Kopelman, *Medical consequences of obesity*. Clin Dermatol, 2004. **22**(4): p. 296-302.
52. Jequier, E., *Carbohydrates as a Source of Energy*. Am J Clin Nutr, 1994. **59**(3): p. 682-685.
53. Flatt, J.P., *Conversion of carbohydrate to fat in adipose tissue: an energy-yielding and, therefore, self-limiting process*. J Lipid Res, 1970. **11**(2): p. 131-43.
54. Lehmann, U. and F. Robin, *Slowly digestible starch - its structure and health implications: a review*. Trends Food Sci Technol, 2007. **18**(7): p. 346-355.
55. Kahn, S.E., R.L. Hull, and K.M. Utzschneider, *Mechanisms linking obesity to insulin resistance and type 2 diabetes*. Nature, 2006. **444**(7121): p. 840-6.
56. Westerfors, M., et al., *Structure-based discovery of a new affinity ligand to pancreatic alpha-amylase*. J Mol Recognit, 2003. **16**(6): p. 396-405.
57. Gupta, R., et al., *Microbial alpha-amylases: a biotechnological perspective*. Process Biochem, 2003. **38**(11): p. 1599-1616.
58. Rana, N., A. Walia, and A. Gaur, *alpha-Amylases from Microbial Sources and Its Potential Applications in Various Industries*. Natl Acad Sci Lett, 2013. **36**(1): p. 9-17.
59. Mitidieri, S., et al., *Enzymatic detergent formulation containing amylase from Aspergillus niger: a comparative study with commercial detergent formulations*. Bioresour Technol, 2006. **97**(10): p. 1217-24.
60. Preuss, H.G., et al., *Inhibition by natural dietary substances of gastrointestinal absorption of starch and sucrose in rats 2. Subchronic studies*. Int J Med Sci, 2007. **4**(4): p. 209-15.
61. Thilagam, E., et al., *alpha-Glucosidase and alpha-amylase inhibitory activity of Senna surattensis*. J Acupunct Meridian Stud, 2013. **6**(1): p. 24-30.
62. Svensson, B., et al., *Proteinaceous alpha-amylase inhibitors*. Biochim Biophys Acta, 2004. **1696**(2): p. 145-56.
63. Wiegand, G., O. Epp, and R. Huber, *The crystal structure of porcine pancreatic alpha-amylase in complex with the microbial inhibitor Tendamistat*. J Mol Biol, 1995. **247**(1): p. 99-110.

64. Machius, M., et al., *Carbohydrate and protein-based inhibitors of porcine pancreatic alpha-amylase: structure analysis and comparison of their binding characteristics*. J Mol Biol, 1996. **260**(3): p. 409-21.
65. Vlieghe, P., et al., *Synthetic therapeutic peptides: science and market*. Drug Discov Today, 2010. **15**(1-2): p. 40-56.
66. Geng, P., et al., *Four acarviosin-containing oligosaccharides identified from Streptomyces coelicoflavus ZG0656 are potent inhibitors of alpha-amylase*. Carbohydr Res, 2008. **343**(5): p. 882-92.
67. Geng, P. and G. Bai, *Two novel aminooligosaccharides isolated from the culture of Streptomyces coelicoflavus ZG0656 as potent inhibitors of alpha-amylase*. Carbohydr Res, 2008. **343**(3): p. 470-6.
68. Geng, P., et al., *Two novel potent alpha-amylase inhibitors from the family of acarviostatins isolated from the culture of Streptomyces coelicoflavus ZG0656*. Chem Biodivers, 2013. **10**(3): p. 452-9.
69. Geng, P., et al., *Taxonomy of the Streptomyces strain ZG0656 that produces acarviostatin alpha-amylase inhibitors and analysis of their effects on blood glucose levels in mammalian systems*. J Appl Microbiol, 2009. **106**(2): p. 525-33.
70. Vermerris, W. and R. Nicholson, *Families of Phenolic Compounds and Means of Classification*, in *Phenolic Compound Biochemistry*, W. Vermerris and R. Nicholson, Editors. Springer: Dordrecht, 2008. **1**. p. 1-34.
71. Xiao, J., et al., *A review on structure-activity relationship of dietary polyphenols inhibiting alpha-amylase*. Crit Rev Food Sci Nutr, 2013. **53**(5): p. 497-506.
72. Ghosh, D. and A. Scheepens, *Vascular action of polyphenols*. Mol Nutr Food Res, 2009. **53**(3): p. 322-31.
73. Tadera, K., et al., *Inhibition of alpha-glucosidase and alpha-amylase by flavonoids*. J Nutr Sci Vitaminol (Tokyo), 2006. **52**(2): p. 149-53.
74. Kim, J.S., C.S. Kwon, and K.H. Son, *Inhibition of alpha-glucosidase and amylase by luteolin, a flavonoid*. Biosci Biotechnol Biochem, 2000. **64**(11): p. 2458-61.
75. Funke, I. and M.F. Melzig, *Effect of different phenolic compounds on alpha-amylase activity: screening by microplate-reader based kinetic assay*. Pharmazie, 2005. **60**(10): p. 796-797.
76. Williamson, G., *Possible effects of dietary polyphenols on sugar absorption and digestion*. Mol Nutr Food Res, 2013. **57**(1): p. 48-57.
77. Li, Y., et al., *Study on the Interaction between 3 Flavonoid Compounds and alpha-Amylase by Fluorescence Spectroscopy and Enzymatic Kinetics*. J Food Sci, 2009. **74**(3): p. 199-203.
78. Kim, H., et al., *Docking and Scoring of Quercetin and Quercetin Glycosides against alpha-Amylase Receptor*. Bull Korean Chem Soc, 2010. **31**(2): p. 461-463.
79. Adisakwattana, S., et al., *Cyanidin-3-rutinoside alleviates postprandial hyperglycemia and its synergism with acarbose by inhibition of intestinal alpha-glucosidase*. J Clin Biochem Nutr, 2011. **49**(1): p. 36-41.
80. Akkarachiyasit, S., et al., *In vitro inhibitory effects of cyanidin-3-rutinoside on pancreatic alpha-amylase and its combined effect with acarbose*. Molecules, 2011. **16**(3): p. 2075-83.
81. Truong, V.D., et al., *Characterization of anthocyanins and anthocyanidins in purple-fleshed sweetpotatoes by HPLC-DAD/ESI-MS/MS*. J Agric Food Chem, 2010. **58**(1): p. 404-10.
82. Tarling, C.A., et al., *The search for novel human pancreatic alpha-amylase inhibitors: high-throughput screening of terrestrial and marine natural product extracts*. Chembiochem, 2008. **9**(3): p. 433-8.

83. Kawamura-Konishi, Y., et al., *Isolation of a new phlorotannin, a potent inhibitor of carbohydrate-hydrolyzing enzymes, from the brown alga Sargassum patens*. J Agric Food Chem, 2012. **60**(22): p. 5565-70.
84. Gyemant, G., et al., *Evidence for pentagalloyl glucose binding to human salivary alpha-amylase through aromatic amino acid residues*. Biochim Biophys Acta, 2009. **1794**(2): p. 291-6.
85. Sang, S.M., et al., *Enzymatic synthesis of tea theaflavin derivatives and their anti-inflammatory and cytotoxic activities*. Bioorg Med Chem, 2004. **12**(2): p. 459-467.
86. Miao, M., et al., *Elucidation of structural difference in theaflavins for modulation of starch digestion*. J Funct Foods, 2013. **5**(4): p. 2024-2029.
87. Patil, V.S., et al., *Synthesis and glycosidase inhibitory activity of novel (2-phenyl-4H-benzopyrimido[2,1-b]-thiazol-4-ylidene)acetonitrile derivatives*. Bioorg Med Chem Lett, 2012. **22**(23): p. 7011-4.
88. Patil, V.S., et al., *Synthesis, crystal structure and antidiabetic activity of substituted (E)-3-(Benzo [d]thiazol-2-ylamino) phenylprop-2-en-1-one*. Eur J Med Chem, 2013. **59**(2013): p. 304-9.
89. Roberts, S.C., *Production and engineering of terpenoids in plant cell culture*. Nature Chem Bio, 2007. **3**(7): p. 387-395.
90. Ali, H., P.J. Houghton, and A. Soumyanath, *alpha-amylase inhibitory activity of some Malaysian plants used to treat diabetes; with particular reference to Phyllanthus amarus*. J Ethnopharmacol, 2006. **107**(3): p. 449-455.
91. Ceriello, A., *Postprandial hyperglycemia and diabetes complications: Is it time to treat?* Diabetes, 2005. **54**(1): p. 1-7.
92. Tanrikulu, Y., B. Kruger, and E. Proschak, *The holistic integration of virtual screening in drug discovery*. Drug Discov Today, 2013. **18**(7-8): p. 358-364.
93. Ekins, S., J. Mestres, and B. Testa, *In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling*. Br J Pharmacol, 2007. **152**(1): p. 9-20.
94. Huang, H.J., et al., *Current developments of computer-aided drug design*. J Taiwan Inst Chem Eng, 2010. **41**(6): p. 623-635.
95. Mandal, S., M. Moudgil, and S.K. Mandal, *Rational drug design*. Eur J Pharmacol, 2009. **625**(1-3): p. 90-100.
96. Sliwoski, G., et al., *Computational Methods in Drug Discovery*. Pharmacol Rev, 2014. **66**(1): p. 334-395.
97. Reddy, A.S., et al., *Virtual screening in drug discovery - A computational perspective*. Curr Protein Pept Sci, 2007. **8**(4): p. 329-351.
98. Kontoyianni, M., L.M. McClellan, and G.S. Sokol, *Evaluation of docking performance: comparative data on docking algorithms*. J Med Chem, 2004. **47**(3): p. 558-65.
99. Seidel, T., et al., *Strategies for 3D pharmacophore-based virtual screening*. Drug Discov Today Technol, 2010. **7**(4): p. 221-228.
100. Maggiora, G., et al., *Molecular Similarity in Medicinal Chemistry*. J Med Chem, 2014. **57**(8): p. 3186-3204.
101. Lee, C.H., H.C. Huang, and H.F. Juan, *Reviewing Ligand-Based Rational Drug Design: The Search for an ATP Synthase Inhibitor*. Int J Mol Sci, 2011. **12**(8): p. 5304-5318.
102. Drwal, M.N. and R. Griffith, *Combination of ligand- and structure-based methods in virtual screening*. Drug Discov Today Technol, 2013. **10**(3): p. 395-401.
103. Wilson, G.L. and M.A. Lill, *Integrating structure-based and ligand-based approaches for computational drug design*. Future Med Chem, 2011. **3**(6): p. 735-750.
104. Ehrlich, P., *Über den jetsigen Stand der Chemotherapie*. Ber Dtsch Chem Ges, 1909. **42**(1): p. 17-47.

105. Yang, S.Y., *Pharmacophore modeling and applications in drug discovery: challenges and recent advances*. Drug Discov Today, 2010. **15**(11-12): p. 444-50.
106. Wermuth, G., et al., *Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)*. Pure Appl Chem, 1998. **70**(5): p. 1129-1143.
107. Leach, A.R., et al., *Three-dimensional pharmacophore methods in drug discovery*. J Med Chem, 2010. **53**(2): p. 539-58.
108. Wolber, G., et al., *Molecule-pharmacophore superpositioning and pattern matching in computational drug design*. Drug Discov Today, 2008. **13**(1-2): p. 23-9.
109. Güner, O., O. Clement, and Y. Kurogi, *Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances*. Curr Med Chem, 2004. **11**(22): p. 2991-3005.
110. Dixon, S.L., A.M. Smondyrev, and S.N. Rao, *PHASE: A novel approach to pharmacophore modeling and 3D database searching*. Chem Bio Drug Des, 2006. **67**(5): p. 370-372.
111. Dixon, S.L., et al., *PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results*. J Comput Aided Mol Des, 2006. **20**(10-11): p. 647-71.
112. Jones, G., P. Willett, and R.C. Glen, *A genetic algorithm for flexible molecular overlay and pharmacophore elucidation*. J Comput Aided Mol Des, 1995. **9**(6): p. 532-549.
113. Labute, P., et al., *Flexible alignment of small molecules*. J Med Chem, 2001. **44**(10): p. 1483-90.
114. Wolber, G. and T. Langer, *LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters*. J Chem Inf Model, 2005. **45**(1): p. 160-9.
115. Wolber, G., A.A. Dornhofer, and T. Langer, *Efficient overlay of small organic molecules using 3D pharmacophores*. J Comput Aided Mol Des, 2006. **20**(12): p. 773-88.
116. Langer, T. and G. Wolber, *Pharmacophore definition and 3D searches*. Drug Discov Today Technol, 2004. **1**(3): p. 203-207.
117. Yuriev, E. and P.A. Ramsland, *Latest developments in molecular docking: 2010-2011 in review*. J Mol Recognit, 2013. **26**(5): p. 215-239.
118. Huang, S.Y., S.Z. Grinter, and X.Q. Zou, *Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions*. Phys Chem Chem Phys, 2010. **12**(40): p. 12899-12908.
119. Cross, J.B., et al., *Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy*. J Chem Inf Model, 2009. **49**(6): p. 1455-1474.
120. Mortier, J., et al., *Computational Tools for In Silico Fragment-Based Drug Design*. Curr Top Med Chem, 2012. **12**(17): p. 1935-1943.
121. Rarey, M., et al., *A fast flexible docking method using an incremental construction algorithm*. J Mol Biol, 1996. **261**(3): p. 470-489.
122. McGann, M., *FRED Pose Prediction and Virtual Screening Accuracy*. J Chem Inf Model, 2011. **51**(3): p. 578-596.
123. Friesner, R.A., et al., *Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. J Med Chem, 2004. **47**(7): p. 1739-1749.
124. Halgren, T.A., et al., *Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening*. J Med Chem, 2004. **47**(7): p. 1750-1759.
125. Friesner, R.A., et al., *Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes*. J Med Chem, 2006. **49**(21): p. 6177-96.

126. Venkatachalam, C.M., et al., *LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites*. J Mol Graph Model, 2003. **21**(4): p. 289-307.
127. Jones, G., et al., *Development and validation of a genetic algorithm for flexible docking*. J Mol Biol, 1997. **267**(3): p. 727-48.
128. Shultz, M.D., *Setting expectations in molecular optimizations: Strengths and limitations of commonly used composite parameters*. Bioorg Med Chem Lett, 2013. **23**(21): p. 5980-5991.
129. Huang, N., et al., *Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening*. J Chem Inf Model, 2006. **46**(1): p. 243-53.
130. Bohm, H.J., *Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs*. J Comput Aided Mol Des, 1998. **12**(4): p. 309-23.
131. Thomas, P.D. and K.A. Dill, *An iterative method for extracting energy-like quantities from protein structures*. Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11628-33.
132. Koppensteiner, W.A. and M.J. Sippl, *Knowledge-based potentials - Back to the roots*. Biochemistry-Moscow, 1998. **63**(3): p. 247-252.
133. Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD*. Proteins, 2003. **52**(4): p. 609-623.
134. Brown, R.D. and Y.C. Martin, *The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding*. J Chem Inf Comp Sci, 1997. **37**(1): p. 1-9.
135. Willett, P., J.M. Barnard, and G.M. Downs, *Chemical similarity searching*. J Chem Inf Comp Sci, 1998. **38**(6): p. 983-996.
136. Albrecht, B., G.H. Grant, and W.G. Richards, *Evaluation of structural similarity based on reduced dimensionality representations of protein structure*. Protein Eng Des Sel, 2004. **17**(5): p. 425-432.
137. Good, A.C. and W.G. Richards, *Explicit Calculation of 3D Molecular Similarity*, in *3D QSAR in drug design: Ligand-protein interactions and molecular similarity*, H. Kubinyi, G. Folkers, and Y.C. Martin, Editors. Kluwer Academic Publishers: London, 1998. **2**. p. 321-338.
138. Downs, G.M. and J.M. Barnard, *Clustering methods and their uses in computational chemistry*, in *Reviews in Computational Chemistry*, K.B. Lipkowitz and D.B. Boyd, Editors. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2003. **1**. p. 1-40.
139. *JKlustor 5.8.0*, available from ChemAxon, Budapest, Hungary, www.chemaxon.com.
140. *Instant JChem 5.8.0*, available from ChemAxon, Budapest, Hungary, www.chemaxon.com.
141. Kirchmair, J., et al., *How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information*. J Chem Inf Model, 2009. **49**(3): p. 678-692.
142. Grant, J.A., M.A. Gallardo, and B.T. Pickup, *A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape*. J Comput Chem, 1996. **17**(14): p. 1653-1666.
143. Nicholls, A. and J.A. Grant, *Molecular shape and electrostatics in the encoding of relevant chemical information*. J Comput Aided Mole Des, 2005. **19**(9-10): p. 661-686.
144. Rush, T.S., et al., *A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction*. J Med Chem, 2005. **48**(5): p. 1489-1495.
145. Triballeau, N., H.-O. Bertrand, and F. Acher, *Are You Sure You Have a Good Model?*, in *Pharmacophores and Pharmacophore Searches*, T. Langer and R.D. Hoffmann, Editors. Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006. **1**. p. 325-364.

146. Markt, P., D. Schuster, and T. Langer, *Pharmacophore Models for Virtual Screening*, in *Virtual Screening: Principles, Challenges, and Practical Guidelines*, C. Sotriffer, Editor. Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2011. **1**. p. 115-152.
147. Triballeau, N., et al., *Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4*. J Med Chem, 2005. **48**(7): p. 2534-47.
148. Jacobsson, M., et al., *Improving structure-based virtual screening by multivariate analysis of scoring data*. J Med Chem, 2003. **46**(26): p. 5781-5789.
149. Markt, P., et al., *Pharmacophore modeling and parallel screening for PPAR ligands*. J Comput Aided Mol Des, 2007. **21**(10-11): p. 575-90.
150. Skorjanc, D. and D. Pette, *Kinetic versus endpoint measurement for quantitative histochemical determination of enzyme activity in muscle fibers*. J Histochem Cytochem, 1998. **46**(2): p. 275-6.
151. Yoo, Y.J., J. Hong, and R.T. Hatch, *Comparison of alpha-amylase activities from different assay methods*. Biotechnol Bioeng, 1987. **30**(1): p. 147-51.
152. Zakowski, J.J. and D.E. Bruns, *Biochemistry of human alpha amylase isoenzymes*. Crit Rev Clin Lab Sci, 1985. **21**(4): p. 283-322.
153. Mohamed, E.A., et al., *Potent alpha-glucosidase and alpha-amylase inhibitory activities of standardized 50 % ethanolic extracts and sinensetin from Orthosiphon stamineus Benth as anti-diabetic mechanism*. BMC Complement Altern Med, 2012. **12**(10): p. 176.
154. Kandra, L., et al., *Kinetic investigation of a new inhibitor for human salivary alpha-amylase*. Biochem Biophys Res Commun, 2005. **334**(3): p. 824-8.
155. MOE 2010.10, available from Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite # 910, Montreal, QC, Canada, www.chemcomp.com.
156. Gilles, C., et al., *Crystal structure of pig pancreatic alpha-amylase isoenzyme II, in complex with the carbohydrate inhibitor acarbose*. Eur J Biochem, 1996. **238**(2): p. 561-9.
157. Yilmazer-Musa, M., et al., *Grape Seed and Tea Extracts and Catechin 3-Gallates Are Potent Inhibitors of alpha-Amylase and alpha-Glucosidase Activity*. J Agric Food Chem, 2012. **60**(36): p. 8924-9.
158. Galvano, F., et al., *Bioavailability, antioxidant and biological properties of the natural free-radical scavengers cyanidin and related glycosides*. Ann Ist Super Sanita, 2007. **43**(4): p. 382-93.
159. LigandScout 3.1, available from Inte:ligand GmbH, Mariahilferstrasse 74B, 1070 Vienna, Austria, www.inteligand.com.
160. Ruddat, M., *Inhibition of the biosynthesis of steviol by a growth retardant*. Nature, 1966. **211**(5052): p. 971-2.
161. Rizzo, B., et al., *Steviol glycosides modulate glucose transport in different cell types*. Oxid Med Cell Longev, 2013. **2013**(2013): p. 348169.
162. Carrillo-Ocampo, D., et al., *Anti-Inflammatory Activity of Iridoids and Verbascoside Isolated from Castilleja tenuiflora*. Molecules, 2013. **18**(10): p. 12109-12118.
163. Gyurkovska, V., et al., *Anti-inflammatory activity of Devil's claw in vitro systems and their active constituents*. Food Chem, 2011. **125**(1): p. 171-178.
164. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery*. Nucleic Acids Res, 2012. **40**(1): p. 1100-7.
165. Koh, L.W., et al., *Evaluation of different teas against starch digestibility by mammalian glycosidases*. J Agric Food Chem, 2010. **58**(1): p. 148-54.
166. Wishart, D.S., et al., *DrugBank: a comprehensive resource for in silico drug discovery and exploration*. Nucleic Acids Res, 2006. **34**(1): p. 668-72.

167. Hopkins, A.L., et al., *The role of ligand efficiency metrics in drug discovery*. Nat Rev Drug Discov, 2014. **13**(2): p. 105-121.
168. Schultes, S., et al., *Ligand efficiency as a guide in fragment hit selection and optimization*. Drug Discov Today Technol, 2010. **7**(3): p. 157-162.
169. Heng, S., K.R. Gryncel, and E.R. Kantrowitz, *A library of novel allosteric inhibitors against fructose 1,6-bisphosphatase*. Bioorg Med Chem, 2009. **17**(11): p. 3916-22.
170. Ajandouz, E.H. and G.J. Marchis-Mouren, *Subsite mapping of porcine pancreatic alpha-amylase I and II using 4-nitrophenyl-alpha-maltooligosaccharides*. Carbohydr Res, 1995. **268**(2): p. 267-77.
171. Nielsen, M.M., et al., *Two secondary carbohydrate binding sites on the surface of barley alpha-amylase 1 have distinct functions and display synergy in hydrolysis of starch granules*. Biochemistry, 2009. **48**(32): p. 7686-97.
172. Seigner, C., E. Prodanov, and G. Marchis-Mouren, *On porcine pancreatic alpha-amylase action: kinetic evidence for the binding of two maltooligosaccharide molecules (maltose, maltotriose and o-nitrophenylmaltoside) by inhibition studies. Correlation with the five-subsite energy profile*. Eur J Biochem, 1985. **148**(1): p. 161-8.
173. Robyt, J.F. and D. French, *Multiple attack and polarity of action of porcine pancreatic alpha-amylase*. Arch Biochem Biophys, 1970. **138**(2): p. 662-70.
174. Robyt, J.F. and D. French, *The action pattern of porcine pancreatic alpha-amylase in relationship to the substrate binding site of the enzyme*. J Biol Chem, 1970. **245**(15): p. 3917-27.
175. Ghavami, A., et al., *Synthesis of nitrogen analogues of salacinol and their evaluation as glycosidase inhibitors*. J Am Chem Soc, 2001. **123**(26): p. 6268-71.
176. *FILTER 2.1.1*, available from OpenEye Scientific Software, Santa Fe, New Mexico, USA, www.eyesopen.com.
177. *OMEGA 2.4.6*, available from OpenEye Scientific Software, Santa Fe, New Mexico, USA, www.eyesopen.com.
178. *ROCS 3.1.2*, available from OpenEye Scientific Software, Santa Fe, New Mexico, USA, www.eyesopen.com.
179. Najafian, M., et al., *Trans-chalcone: a novel small molecule inhibitor of mammalian alpha-amylase*. Mol Biol Rep, 2011. **38**(3): p. 1617-20.
180. Akkarachiyasit, S., et al., *Inhibitory activities of cyanidin and its glycosides and synergistic effect with acarbose against intestinal alpha-glucosidase and pancreatic alpha-amylase*. Int J Mol Sci, 2010. **11**(9): p. 3387-96.
181. Zhu, T., et al., *Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis*. J Med Chem, 2013. **56**(17): p. 6560-6572.
182. Sadowski, J., J. Gasteiger, and G. Klebe, *Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures*. J Chem Inf Comp Sci, 1994. **34**(4): p. 1000-1008.
183. *CORINA 3.4*, available from Molecular Networks GmbH, Erlangen, Germany, www.molecular-networks.com.
184. *GOLD 5.1*, available from Cambridge Crystallographic Data Centre, Cambridge, UK, www.ccdc.cam.ac.uk.
185. Brvar, M., et al., *In silico discovery and biophysical evaluation of novel 5-(2-hydroxybenzylidene) rhodanine inhibitors of DNA gyrase B*. Bioorg Med Chem, 2012. **20**(8): p. 2572-80.
186. *MEGx: Natural Products Screening Compounds*. Analyticon Discovery GmbH, Potsdam, Germany, www.ac-discovery.com, downloaded in Sept. 15, 2011.
187. *Asinex Synergy Library*. Asinex Ltd., Moscow, Russia, www.asinex.com, downloaded in Sept. 15, 2011.

188. *Asinex Platinum Collection*. Asinex Ltd., Moscow, Russia, www.asinex.com, downloaded in Sept. 15, 2011.
189. *Asinex Gold Collection*. Asinex Ltd., Moscow, Russia, www.asinex.com, downloaded in Sept. 15, 2011.
190. *Chembridge Screening Library*. Chembridge Corporation, San Diego, CA, USA, www.chembridge.com, downloaded in Sept. 15, 2011.
191. *Specs Screening Compounds*. Specs, Delft, The Netherlands, www.specs.net, downloaded in Sept. 15, 2011.
192. *Bionet Screening Compounds*. Key Organics Ltd., Camelford, UK, www.keyorganics.net, downloaded in Sept. 15, 2011.
193. *Life Chemical Compound Collection for HTS*. Life Chemical Inc., Niagara-on-the-Lake, Ontario, Canada, www.lifechemicals.com, downloaded in Sept. 15, 2011.
194. *The Prestwick Chemical Library*. Prestwick Chemical, Illkirch, France, www.prestwickchemical.com, downloaded in Sept. 15, 2011.
195. *Standardizer 5.8.0*, available from ChemAxon, Budapest, Hungary, www.chemaxon.com.
196. Kirchmair, J., et al., *Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations*. J Chem Inf Model, 2006. **46**(4): p. 1848-61.
197. Irwin, J.J. and B.K. Shoichet, *ZINC - A Free Database of Commercially Available Compounds for Virtual Screening*. J Chem Inf Model, 2005. **45**(1): p. 177-82.
198. Maurus, R., et al., *Alternative catalytic anions differentially modulate human alpha-amylase activity and specificity*. Biochemistry, 2008. **47**(11): p. 3332-44.
199. Michaux, C., et al., *Protecting role of cosolvents in protein denaturation by SDS: a structural study*. BMC Struct Biol, 2008. **8**(29): p. 1-7.
200. Baumann, H., et al., *Rational design, synthesis, and verification of affinity ligands to a protein surface cleft*. Protein Sci, 2003. **12**(4): p. 784-93.
201. Butterworth, P.J., F.J. Warren, and P.R. Ellis, *Human alpha-amylase and starch digestion: An interesting marriage*. Starch-Starke, 2011. **63**(7): p. 395-405.
202. Numao, S., et al., *In situ extension as an approach for identifying novel alpha-amylase inhibitors*. J Biol Chem, 2004. **279**(46): p. 48282-91.
203. Ragunath, C., et al., *Structure-function relationships in human salivary alpha-amylase: role of aromatic residues in a secondary binding site*. Biologia, 2008. **63**(6): p. 1028-1034.
204. Fisher, S.Z., et al., *Structure of human salivary alpha-amylase crystallized in a C-centered monoclinic space group*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2006. **62**(2): p. 88-93.
205. Ramasubbu, N., et al., *Structural studies of a Phe256Trp mutant of human salivary alpha-amylase: implications for the role of a conserved water molecule in enzyme activity*. Arch Biochem Biophys, 2004. **421**(1): p. 115-24.
206. Desmyter, A., et al., *Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology*. J Biol Chem, 2002. **277**(26): p. 23645-50.
207. Bompard-Gilles, C., et al., *Substrate mimicry in the active center of a mammalian alpha-amylase: structural analysis of an enzyme-inhibitor complex*. Structure, 1996. **4**(12): p. 1441-52.
208. Funke, I., *Investigations on the Impact of Herbal Extracts and Polyphenolic Compounds on the α -Amylase Activity with particular regard to *Cynara cardunculus* L. ssp. *flavescens* Wikl. and its Ingredients*. Doctoral thesis, Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, 2007: p. 1-121.
209. Oka, H., et al., *Simple and conveniently accessible bi-fluorescence-labeled substrates for amylases*. Bioorg Med Chem Lett, 2010. **20**(6): p. 1969-71.

210. Abell, A.D., M.J. Ratcliffe, and J. Gerrard, *Ascorbic acid-based inhibitors of alpha-amylases*. *Bioorg Med Chem Lett*, 1998. **8**(13): p. 1703-6.
211. Gerrard, J.A., M.J. Prince, and A.D. Abell, *Kinetic characterisation of ene-diol-based inhibitors of alpha-amylase*. *Bioorg Med Chem Lett*, 2000. **10**(14): p. 1575-6.
212. Anon, *Food additives. Polycarbonate resins*. *Fed Regist*, 1970. **35**(93): p. 7414.
213. Agh-Atabay, N.M., B. Dulger, and F. Guzin, *Synthesis and investigation of antimicrobial activity of some bisbenzimidazole-derived chelating agents*. *Eur J Med Chem*, 2003. **38**(10): p. 875-881.
214. Brvar, M., et al., *In silico discovery of 2-amino-4-(2,4-dihydroxyphenyl)thiazoles as novel inhibitors of DNA gyrase B*. *Bioorg Med Chem Lett*, 2010. **20**(3): p. 958-62.
215. Perdih, A., et al., *Inhibitor Design Strategy Based on an Enzyme Structural Flexibility: A Case of Bacterial MurD Ligase*. *J Chem Inf Model*, 2014. **54**(5): p. 1451-1466.
216. Kreuzsch, A., et al., *Crystal structures of human HSP90 alpha-complexed with dihydroxyphenylpyrazoles*. *Bioorg Med Chem Lett*, 2005. **15**(5): p. 1475-1478.

Curriculum vitae

For reasons of data protection, the curriculum vitae is not published in the electronic version.

For reasons of data protection, the curriculum vitae is not published in the electronic version.

Publications

Articles:

1. J. R. Homoki; A. Nemes; E. F. Fazekas; G. Gyémánt; P. Balogh; F. G. Gál; **J. Al-Asri**; J. Mortier; G. Wolber; L. Babinszky: Anthocyanin composition, antioxidant efficiency, and alpha-amylase inhibitor activity of different Hungarian varieties of sour cherry (*Prunus cerasus* L.). *Submitted*.
2. L.A. Shemchuk, D.V. Levashov, **J. Al-Asri**, Y.A. Shenhof, P.S. Arzumanov, V.P. Chernykh: Studying the Grignard reaction of quinoxaline derivatives containing an ester group. *Journal of Organic and Pharmaceutical chemistry-Ukraine*, 2011. **3**(35): p. 36-39.
3. L.A. Shemchuk, **J. Al-Asri**, D.V. Levashov, Y.A. Shenhof, P.S. Arzumanov: Synthesis of a new quinoxaline derivatives based on 2-(4-oxo-3,4-dihydro-3-quinoxalinyloxy)acetohydrazide. *Bulletin of Pharmacy-Ukraine*, 2011. **3**(67): p. 30-32.

Conferences:

4. Poster: J. Al-Asri, G. Wolber: Discovery of novel α -amylase inhibitors using structure-based drug design, the 9th German Conference on Chemoinformatics, November 10-12, 2013; Fulda, Germany. *Journal of Cheminformatics* 2014, **6**(Suppl 1):P50.
5. Poster: J. Al-Asri, J. Mortier, G. Wolber: Pharmacophore development for the discovery of new alpha-Amylase inhibitors, the 22nd International Symposium on Medicinal Chemistry (EFMC-ISMIC 2012), September 2-6, 2012, Berlin, Germany.
6. Poster: J. Al-Asri, G. Wolber: Pharmacophore creation for the discovery of new alpha-amylase inhibitors, the 3rd Strasbourg Summer School on Chemoinformatics, June 25-29, 2012, Strasbourg, France.
7. Poster: J. Al-Asri, J. Mortier, G. Wolber: 3D pharmacophore models for the identification of novel human α -amylase inhibitors, the 2nd Berliner Chemie Symposium, April 3, 2012, Berlin, Germany.
8. Poster: J. Al-Asri, J. Mortier, G. Wolber: 3D pharmacophore models for the identification of novel human α -amylase inhibitors, the 7th International Workshop on New Approaches in Drug Design & Discovery, March 19-22, 2012, Marburg, Germany.

9. Poster: J. Al-Asri, D. V. Levashov, P.S. Arzumanov, V.P.Chernykh, L.A. Shemchuk: Synthesis of new 3, 4-dihydroquinazolin-4-ones based on anthranilic acid derivatives and glycine, the VII National Congress of Pharmacists of Ukraine, September 15-17 , 2010, Kharkov, Ukraine.
10. Poster: J. Al-Asri, D. V. Levashov, P.S. Arzumanov, L.A. Shemchuk: Synthesis of new quinazolinones containing glycine fragment and study their reactions with Grignard reagent, the IV All-Conference Dombrowski Chemical Reading, May 17-20, 2010, Lovo, Ukraine.
11. Oral presentation: J. Al-Asri, D. V. Levashov, P.S. Arzumanov, V.P.Chernykh, L.A. Shemchuk: Synthesis of new 3, 4-dihydroquinazolin-4-ones based on anthranilic acid derivatives and glycine, the Scientific Conference for Ukrainian Students "Chemical Karazinski Reading", April 19-22, 2010, Kharkov, Ukraine.
12. Oral Presentation: J. Al-Asri, D. V. Levashov, P.S. Arzumanov, V.P.Chernykh, L.A. Shemchuk: Synthesis of new quinazolinone derivatives based on anthranilic acid derivatives and glycine, the Practical Conference for Young Scientists, April 21-22, 2010, Kharkov, Ukraine. (*Awarded the first position*).

10 Appendix

10.1 Macromolecules of pancreatic and salivary α -amylases (human and porcine)

Crystal structures of HPA, HSA, and PPA publically available in the PDB are listed in Table 16.

Table 16: The PDB codes of human and porcine α -amylase crystal structures

PDB	Binding ligands, co-factors, mutation	Release date	Resolution Å	Ref.
Human pancreatic α-amylase (HPA)				
4GQR	myricetin	10.2012	1.2	[25]
4GQQ	ethyl caffeate	10.2012	1.35	[25]
3OLD	acarviosatin I03 (6 rings, valienamine # 3)	04.2011	2.00	[36]
3OLE	acarviosatin II03 (7 rings, valienamine # 1 and 4)	04.2011	1.55	[36]
3OLG	acarviosatin III03 (7 rings, valienamine # 1 and 4)	04.2011	2.30	[36]
3OLI	acarviosatin IV03 (7 rings, valienamine # 1 and 4)	04.2011	1.50	[36]
3IJ7	oligosaccharyl-5-F-glycosyl-F	10.2009	2.00	[31]
3IJ8	oligosaccharyl-5-F-glycosyl-F	10.2009	1.43	[31]
3IJ9	oligosaccharyl-5-F-glycosyl-F	10.2009	1.85	[31]
2QMK	nitrite	03.2008	2.30	[198]
2QV4	nitrite and acarbose (5 ring, valienamine # 2)	03.2008	1.97	[198]
3BAI	nitrate	03.2008	1.90	[198]
3BAJ	nitrate and acarbose (5 ring, valienamine # 3)	03.2008	2.10	[198]
3BAK	mutant N298S and nitrate	03.2008	1.90	[198]
3BAW	azide	03.2008	2.00	[198]
3BAX	mutant N298S , azide	03.2008	1.90	[198]
3BAY	N298S, nitrate and acarbose (5 rings, valinamine # 3)	03.2008	1.99	[198]
1XGZ	mutant, N298S	05.2005	2.00	[26]
1XH0	mutant/ acarbose (6 rings), No Cl. (valienamine # 3)	05.2005	2.00	[26]
1XH1	mutant, N298S	05.2005	2.03	[26]
1XH2	mutant and acarbose (5 rings, valienamine # 3)	05.2005	2.20	[26]
1XCW	acarviosine-glucose (Acarv-glu)	12.2004	2.00	[32]
1XCX	iso-acarbose	12.2004	1.90	[32]
1XD0	acarviosine-glucose with G3F (5 ring, valienamine # 3)	12.2004	2.00	[32]
1XD1	isoacarbose with G3F (6 rings, valienamine # 1)	12.2004	2.20	[32]
1U2Y	d-gluconohydroximino-1,5-lactam (GHIL)	09.2004	1.95	[202]
1U30	trisaccharide (G2-GHIL)	09.2004	1.90	[202]
1U33	trisaccharide (MeG2-GHIL)	09.2004	1.95	[202]
1KB3	mutant R195A	05.2002	2.10	[27]
1KBB	mutation of catalytic resdiues	05.2002	1.90	[34]
1KBK	mutation of catalytic resdiues	05.2002	1.90	[34]
1KGU	mutant R337A	05.2002	2.00	[27]
1KGW	mutant R337Q	05.2002	2.10	[27]

1KGX	mutant R195Q	05.2002	2.00	[27]
2CPU	mutation of catalytic residues	06.2001	2.00	[30]
3CPU	mutation of catalytic residues, maltose	06.2001	2.00	[30]
1B2Y	mutation, acarbose (5 rings, valienamine # 3)	02.2000	3.20	[24]
1CPU	modified acarbose (5 rings, valienamine # 3)	06.1999	2.00	[30]
1BSI	mutant, D197A, D197N	05.1999	2.00	[35]
1HNY		03.1996	1.80	[21]
Human salivary α-amylase (HSA)				
PDB	Binding ligands, co-factors, mutation	Release date	Resolution Å	Ref.
3DHP	Mutant/ acarbose (4 rings, valienamine # 2)	07.2008	1.5	[45]
3BLP	W388/ acarbose (5 rings, valienamine # 3)	11.2008	1.60	[203]
3BLK	W316A/ acarbose (5 rings, valienamine # 3)	11.2008	2.00	[203]
1Z32	mutant, Y151M (3 rings, valienamine # 2)	05.2005	1.60	[38]
1XV8	dimer	10.2005	3.00	[204]
1Q4N	mutant F256W	03.2004	2.07	[205]
1NM9	mutant W58A (3 rings)	01.2004	2.10	[37]
1MFU	mutant, residues 306-310 (7 rings, valienamine # 1 and 4)	11.2002	2.00	[49]
1MFV	modified acarbose, (6 rings, valienamine # 1 and 4)	11.2002	2.00	[49]
1JXJ		09.2001	1.99	[37]
1JXK	Role of mobile loop	09.2001	1.90	[49]
1C8Q	Recombinant HSA	06.2000	2.30	[205]
1SMD		07.1996	1.60	[22]
Porcine pancreatic α-amylase (PPA)				
1KXQ	proteinaceous amylase inhibitor	06.2002	1.6	[206]
1KXT	proteinaceous amylase inhibitor	06.2002	2	[206]
1KXV	proteinaceous amylase inhibitor	06.2002	1.6	[206]
1BVN	proteinaceous amylase inhibitor (tendamistat)	09.1998	2.5	[63]
1DHK	dimer	12.1997	1.85	[207]
1OSE	modified acarbose (6 rings, valienamine # 1 and 4)	04.1997	2.3	[156]
1PIF		12.1996	2.3	[64]
1PIG	pseudooctasaccharide V-1532,(6 rings, valienamine # 3)	12.1996	2.2	[64]

10.2 Structural analysis

Software package MOE 2010.10 [155] was used to superimpose crystal structures of α -amylase coded as 3OLE, 1CXW, 1XCX, 1MFV, and 1OSE with the purpose of investigating the position of water molecule 746 (HOH-746). Analysis showed that HOH-746 is located nearly at the same position in all investigated macromolecules (Figure 47). Distances between the oxygen atom of HOH-746 and the neighboring amino acid residues were

measured. The oxygen atoms from the side chain (OE1) and the backbone carbonyl of the catalytic residue Glu233 are respectively located at 3.32 Å and 2.82 Å from the oxygen atom of HOH-746. Also, the nitrogen atom in the backbone of Lys200 is found at 3.14 Å from the oxygen atom of this water molecule. With inhibitors belonging to the trestatin family, such as acarbose and acarviostatins, HOH-746 can form H-bonds with hydroxyl groups of the sugar unit accommodated in subsite +1. Imposing the pharmacophore model to retrieve this particular HBD feature led to excluding most of the inactive compounds during the validation process for model B that was then used for the initial virtual screening.

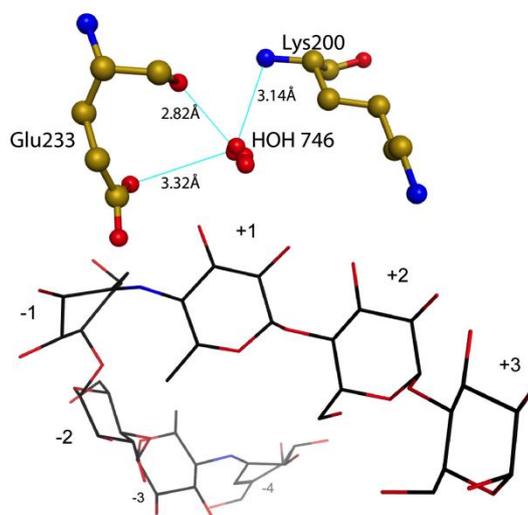


Figure 47: The water-746 is depicted as red spheres obtained by superposition of α -amylase structures coded in the PDB as 3OLE, 1CXW, 1XCX, 1MFV, and 1OSE (view with inhibitor acarviostatin II03). Positive and negative numbers represent active subsites.

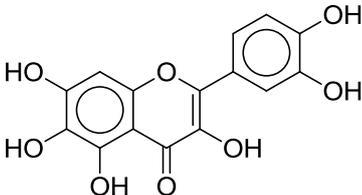
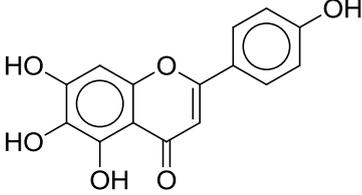
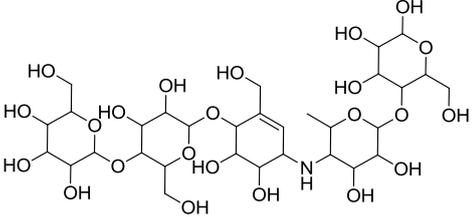
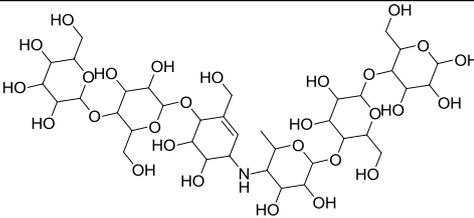
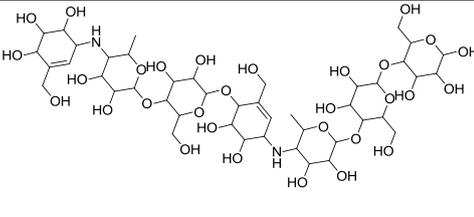
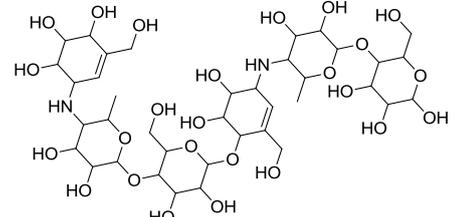
10.3 Biologically known active α -amylase inhibitors

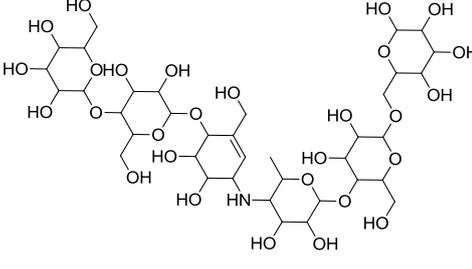
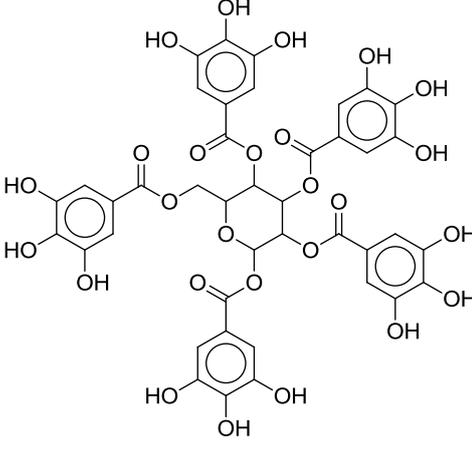
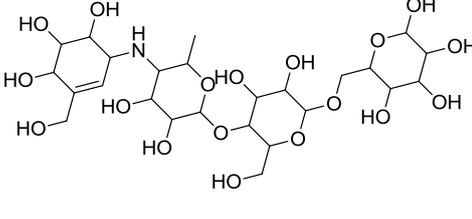
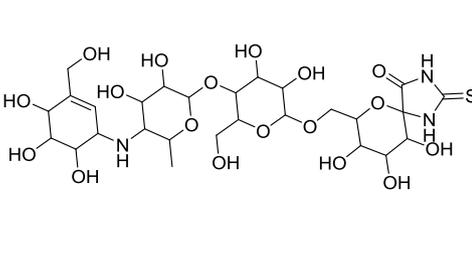
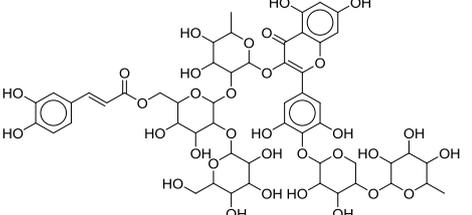
Compounds with biologically known active α -amylase inhibitors were assembled from ChEMBL database and literature to be used as validation dataset (Table 17).

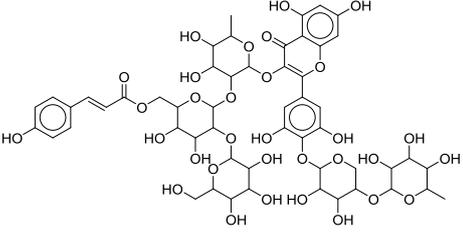
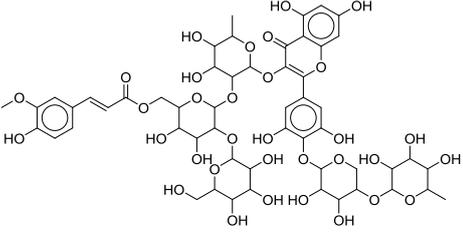
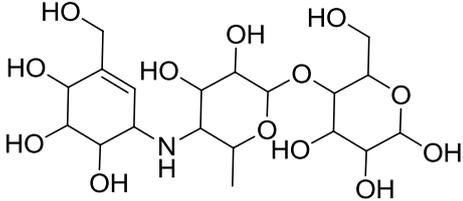
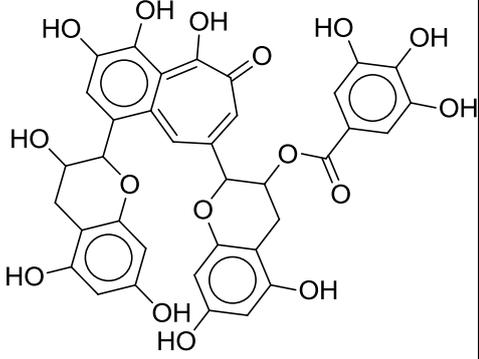
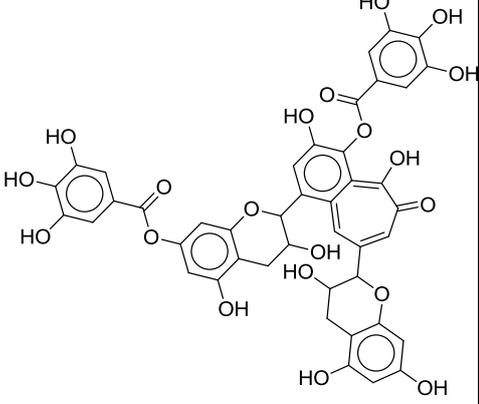
In the initial virtual screening: 19 compounds numbered from 1 to 14 and from 20 to 24 were used as actives during validation of the developed pharmacophores.

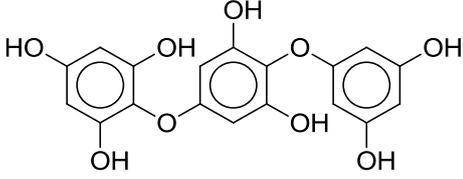
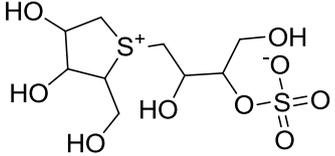
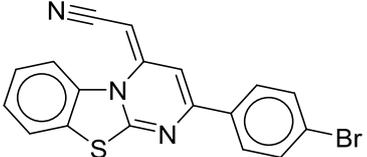
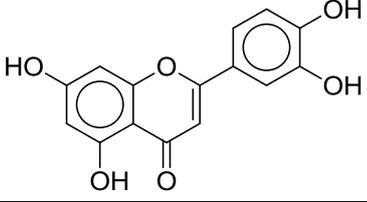
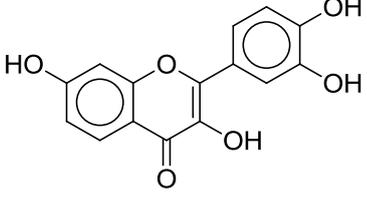
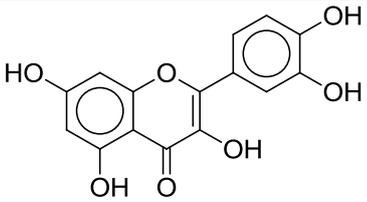
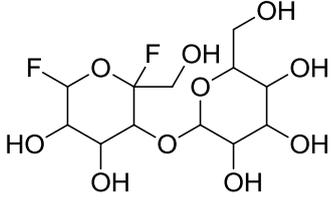
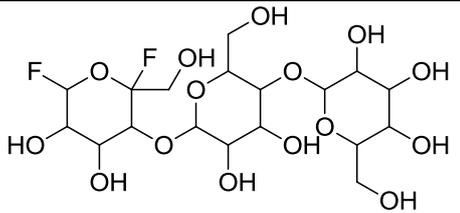
In the refined virtual screening: 19 compounds numbered from 1 to 19 were considered actives to validate the generated 3D models.

Table 17: Collected biologically known active α -amylase inhibitors

No.	Name	Structure	Ref.	Activity	PDB
1	Quercetagetin		[20]	$IC_{50} = 10.2 \mu M$	
2	Scutellarein		[20]	$IC_{50} = 9.64 \mu M$	
3	Product of rearranged acarbose, and elongation of acarviosine-glucose		[36] [198] [26] [32] [24] [30]	$IC_{50} = 2.59 \mu M$	3BAJ 3BAY 1XH2 1B2Y 1CPU 1XD0
4	Product of acarviostatin I03, and acarbose		[36] [26]	$K_i = 1.25 \mu M$	3OLD 1XH0
5	Product of acarviostatins II03, III03, and IV03		[36]	$K_i = 0.0147 \mu M$ $K_i = 0.0143 \mu M$ $K_i = 0.0416 \mu M$	3OLE 3OLG 3OLI
6	Product of rearranged acarbose		[154] [49] [156]	$K_i = 0.7 \mu M$	1MFV 1OSE

No.	Name	Structure	Ref.	Activity	PDB
7	Product of iso-acarbose elongated with G ₃ F		[32]	$K_i = 0.012 \mu\text{M}$	1XD1
8	1,2,3,4,6-pentagalloyl- β -D-glucose		[84]	$IC_{50} = 2.35 \mu\text{M}$ ($K_i = 2.6 \mu\text{M}$)	
9	Iso-acarbose		[32]	$K_i = 0.012 \mu\text{M}$	1XCX
10	α -Acarviosinyl-1,4- α -D-glucopyranosyl-1,6-D-glucopyranosylidene-spirothiohydantoin (PTS-G-TH)		[154]	$K_i = 0.19 \mu\text{M}$	
11	Montbretin A		[82]	$K_i = 0.008 \mu\text{M}$	

No.	Name	Structure	Ref.	Activity	PDB
12	Montbretin B		[82]	$K_i = 3.600 \mu\text{M}$	
13	Montbretin C		[82]	$K_i = 6.100 \mu\text{M}$	
14	Acarviosine-glucose		[32]	$K_i = 0.075 \mu\text{M}$	1XCW
15	Theaflavin monogallate		[165]	$IC_{50} = 5.5 \mu\text{M}$	
16	Theaflavin digallate		[165]	$IC_{50} = 2.9 \mu\text{M}$	

No.	Name	Structure	Ref.	Activity	PDB
17	2-(4-(3, 5-Dihydroxyphenoxy)-3,5-dihydroxyphenoxy) benzene-1,3,5-triol (DDBT)		[83]	IC ₅₀ = 8.56 μM, K _i = 4.8 μM	
18	Salacinol		[175]	K _i = 10 μM for PPA	
19	CHEMBL2203334		[87]	IC ₅₀ = 15.260 μM	
20	Luteolin		[20]	IC ₅₀ = 18.4 μM	
21	Fisetin		[20]	IC ₅₀ = 19.6 μM	
22	Quercetin		[20]	IC ₅₀ = 21.4 μM	
23	G ₃ F/5FIdoF		[31]	Inhibition = 90 %, 5FIdoF (25 mM) elongated with G ₃ F(40 mM)	3IJ9
24	MeG ₂ F/5FIdoF		[31]	Inhibition = 95 %, 5FIdoF (50 mM) elongated with MeG ₂ F (20 mM)	3IJ7

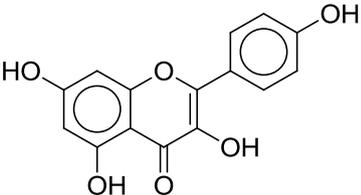
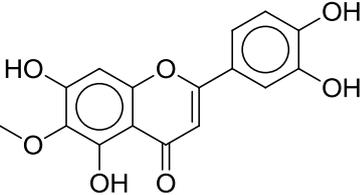
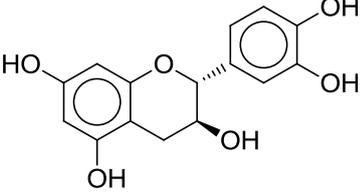
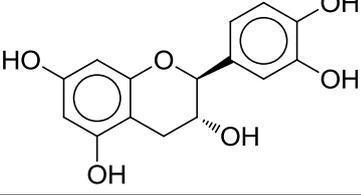
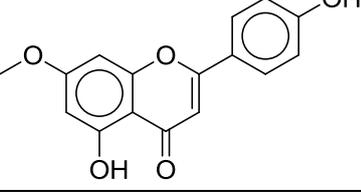
10.4 Biologically known inactive α -amylase inhibitors

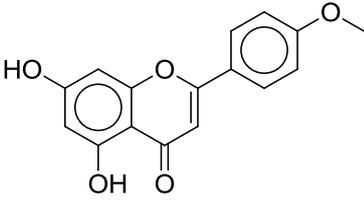
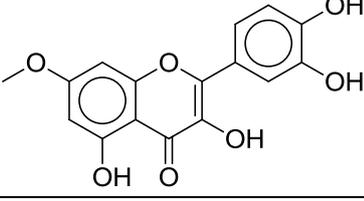
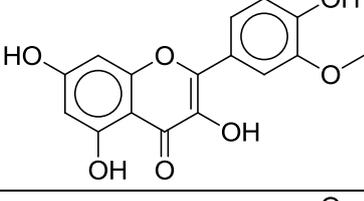
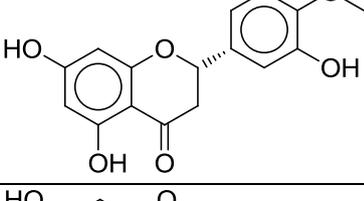
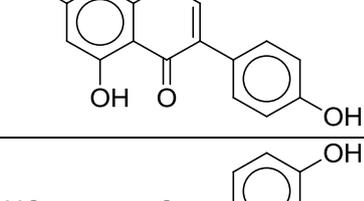
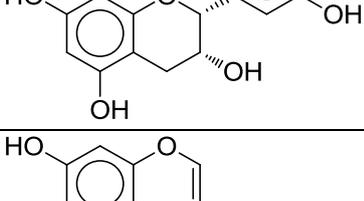
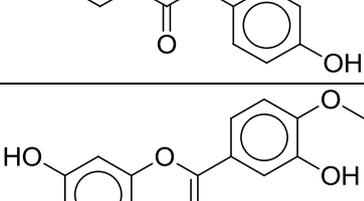
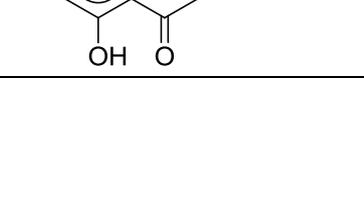
Compounds biologically tested against α -amylase with known weak or inactive inhibition were collected from literature and listed in Table 18 to be used in validation dataset.

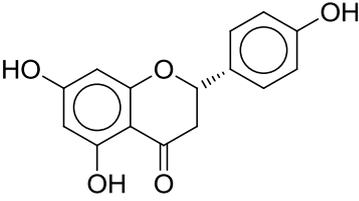
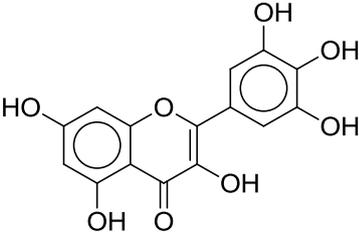
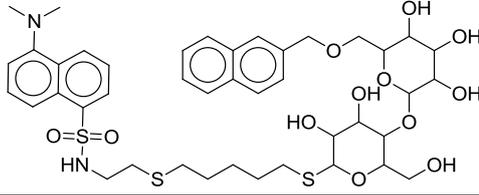
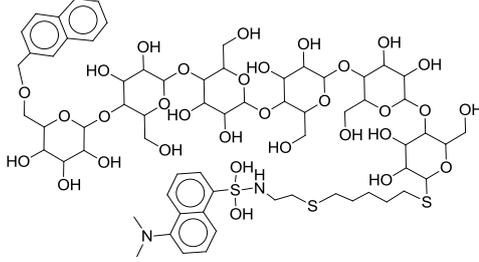
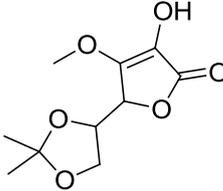
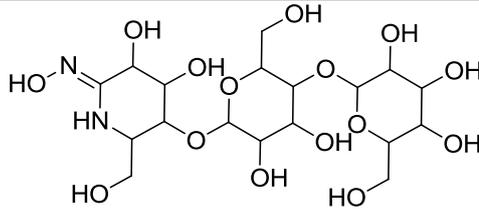
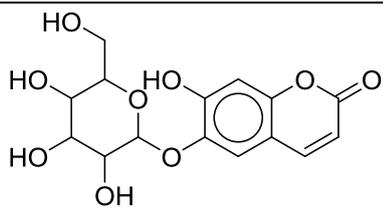
In the initial virtual screening, 55 compounds numbered from 1 to 43 and from 56 to 67 in Table 18 were used as inactives during validation of developed pharmacophores.

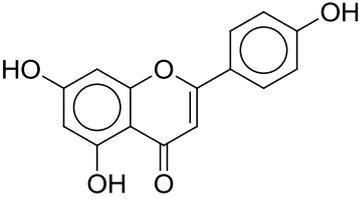
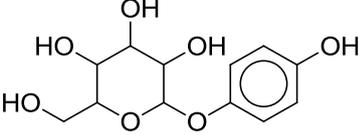
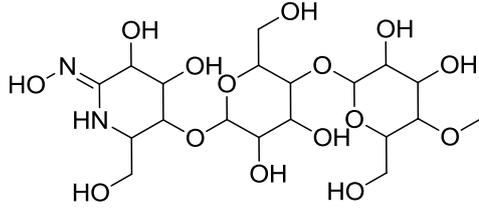
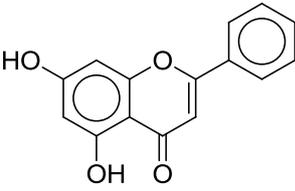
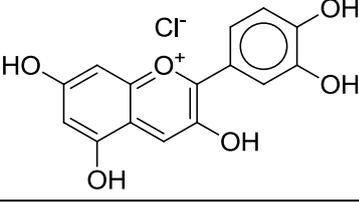
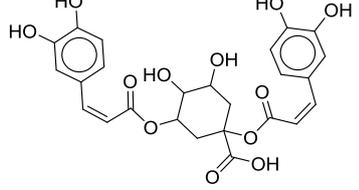
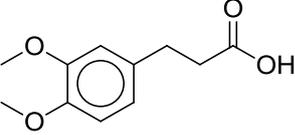
In the refined virtual screening, 60 compounds numbered from 1 to 55 in Table 18 and from 20 to 24 in Table 17 (Appendix 10.3) were used as inactives to validate the developed 3D models.

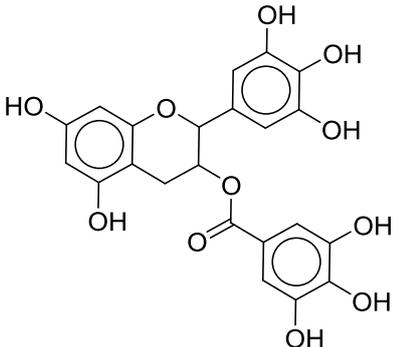
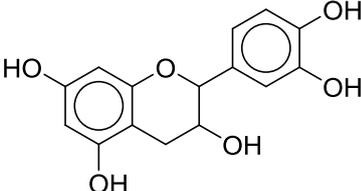
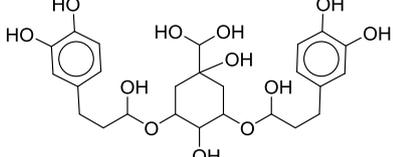
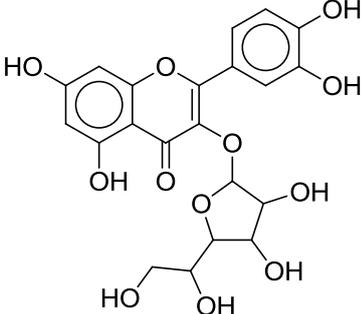
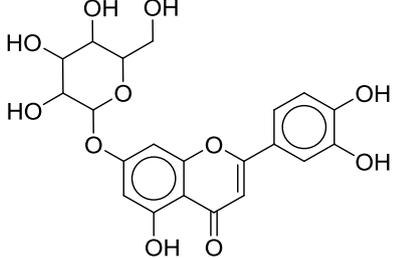
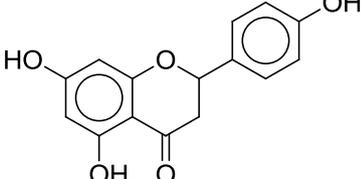
Table 18: Collected biologically known inactive α -amylase inhibitors

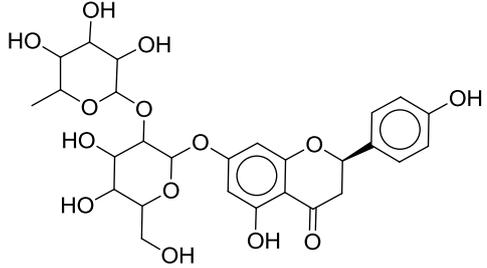
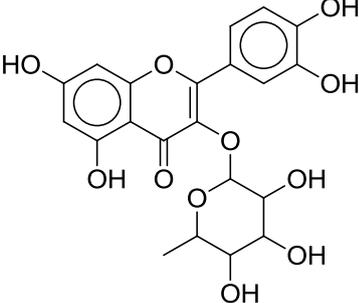
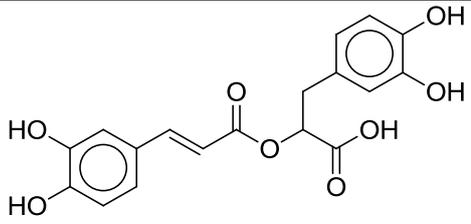
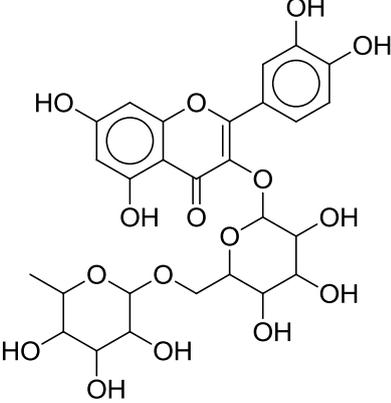
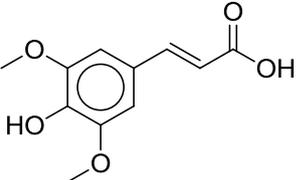
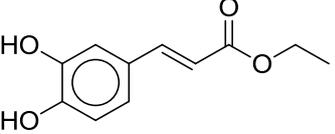
No	Name	structure	Ref.	Value	PDB
1	Kaempferol		[20]	inhibition = 34.5 %	
2	Eupafolin		[20]	IC ₅₀ = 48 μ M	
3	(+)-Catechin		[20]	inhibition = 13.1 %	
4	(-)-Catechin		[208]	inhibition = 16 %	
5	Genkwanin		[20]	inhibition = 17.5 %	

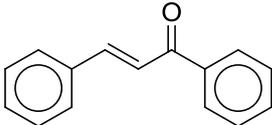
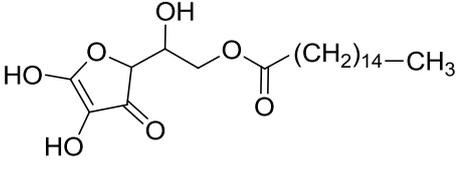
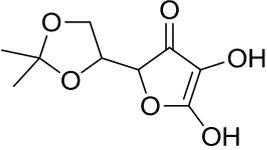
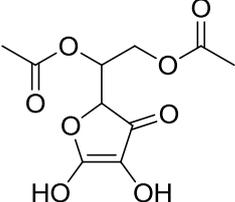
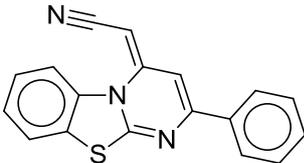
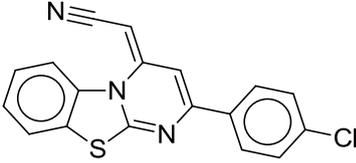
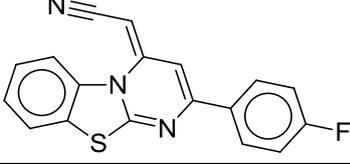
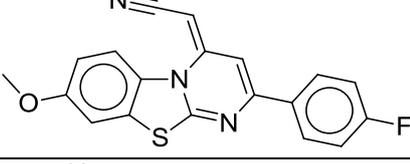
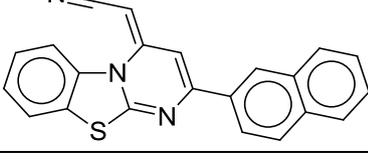
No	Name	structure	Ref.	Value	PDB
6	Acacetin		[20]	inhibition = 14.1 %	
7	Rhamnetin		[20]	inhibition = 8.1 %	
8	Isorhamnetin		[20]	inhibition = 35.4 %	
9	Hesperetin		[20]	inhibition = 39.8 %	
10	Genistein		[20]	inhibition = 25.1 %	
11	Epicatechin		[20]	inhibition = 10.3 %	
12	Daidzein		[20]	inhibition = 23.3 %	
13	Diosmetin		[20]	inhibition = 19.2 %	

No	Name	structure	Ref.	Value	PDB
14	Naringenin		[20]	inhibition = 26.8 %	
15	Myricetin		[20] [25]	IC ₅₀ = 30.2 μM	4GQR
16	Bi-fluorescence-labeled maltoside		[209]		
17	Bi-fluorescence-labeled maltohexaoside		[209]		
18	5-(2,2-Dimethyl-1,3-dioxolan-4-yl)-3-hydroxy-4-methoxyfuran-2(5H)-one		[210]	Inhibition = 0%	
19	Maltosyl-α (1,4)-d-gluconhydroximo-1,5-lactam; (G2-GHIL)		[202]	K _i = 1.8 mM	1U30
20	Aesculin		[208]	IC ₅₀ = 1.67 mM	

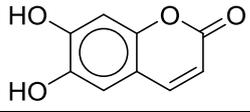
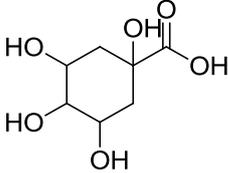
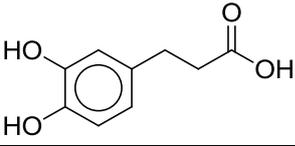
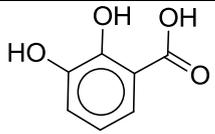
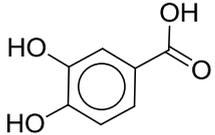
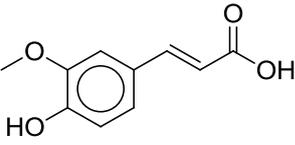
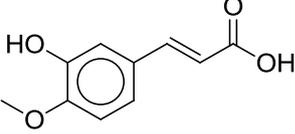
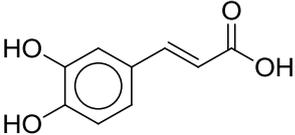
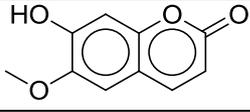
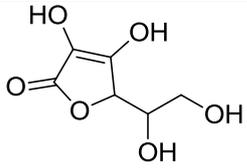
No	Name	structure	Ref.	Value	PDB
21	Apigenin		[208]	Inhibition = 12.4 %	
22	Arbutin		[208]	Inhibition = 30 %	
23	4'-O-methyl-maltosyl- α (1,4)-d-gluconhydroximo-1,5-lactam; (MeG2-GHIL)		[202]	$K_i = 25 \mu\text{M}$	1U33
24	Chrysin		[208]	inhibition = 5 %	
25	Cyanidinchloride		[208]	Inhibition = 12 %	
26	Cynarin		[208]	$\text{IC}_{50} > 2 \text{ mM}$	
27	3(3,4-Dimethoxyphenyl)-propanoic acid		[208]	$\text{IC}_{50} > 12 \text{ mM}$	

No	Name	structure	Ref.	Value	PDB
28	Epigallocatechin gallate		[165]	IC ₅₀ = 1.4 mM	
29	(-)-Epicatechin		[208]	IC ₅₀ = 3.5 mM	
30	Isochlorogenic acid A		[208]	IC ₅₀ = 0.39 mM	
31	Isoquercitrin		[208]	inhibition = 12.2 %	
32	Luteolin-7-glucoside		[208]	IC ₅₀ = 0.28 mM	
33	Naringenin		[208]	IC ₅₀ = 1.8 mM	

No	Name	structure	Ref.	Value	PDB
34	Naringin		[208]	inhibition = 0 %	
35	Quercitrin		[208]	inhibition = 37.6 %	
36	Rosmarinic acid		[208]	IC ₅₀ = 1.4 mM	
37	Rutin		[208]	inhibition = 20.1 %	
38	Sinapinic acid		[208]	IC ₅₀ > 6.5	
39	Ethyl caffeate		[82] [25]	K _i = 1.3 mM	4GQQ

No	Name	structure	Ref.	Value	PDB
40	Trans-chalcone		[179]	$K_i = 48 \mu\text{M}$	
41	2-(4,5-Dihydroxy-3-oxo-2,3-dihydrofuran-2-yl)-2-hydroxyethyl palmitate		[210] [211]	$IC_{50} = 62.5 \mu\text{M}$	
42	5-(2,2-Dimethyl-[1,3]dioxolan-4-yl)-3,4-dihydroxy-5H-furan-2-one		[210] [211]	$IC_{50} = 28 \mu\text{M}$	
43	Acetic acid 2-acetoxy-1-(3,4-dihydroxy-5-oxo-2,5-dihydrofuran-2-yl)-ethyl ester		[210]	Inhibition = 99 %	
44	CHEMBL220794 8		[87]	$IC_{50} = 22.96 \mu\text{M}$	
45	CHEMBL220794 9		[87]	$IC_{50} = 21.85 \mu\text{M}$	
46	CHEMBL220795 0		[87]	$IC_{50} = 24.39 \mu\text{M}$	
47	CHEMBL220795 1		[87]	$IC_{50} = 20.6 \mu\text{M}$	
48	CHEMBL220795 2		[87]	$IC_{50} = 19.91 \mu\text{M}$	

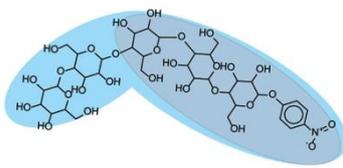
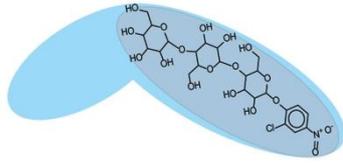
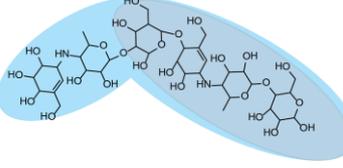
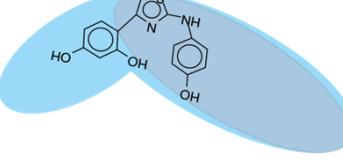
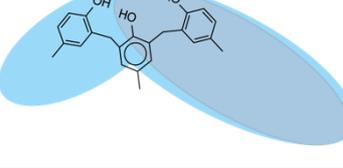
No	Name	structure	Ref.	Value	PDB
49	CHEMBL220795 3		[87]	IC ₅₀ = 18.71 μM	
50	CHEMBL220795 4		[87]	IC ₅₀ = 20.22 μM	
51	CHEMBL220795 5		[87]	IC ₅₀ = 23.24 μM	
52	CHEMBL220795 6		[87]	IC ₅₀ = 19.54 μM	
53	CHEMBL220795 7		[87]	IC ₅₀ = 22.30 μM	
54	CHEMBL220795 8		[87]	IC ₅₀ = 21.8 μM	
55	CHEMBL220795 9		[87]	IC ₅₀ = 17.43 μM	
56	Ghavamiol		[175]	K _i < 10 mM	
57	D- gluconhydroximo -1,5-lactam, (GHIL)		[202]	K _i = 18 mM	1U2Y

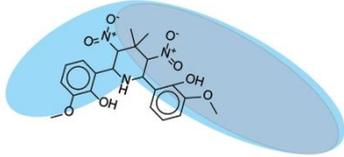
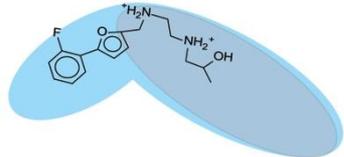
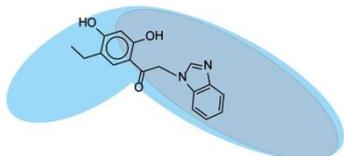
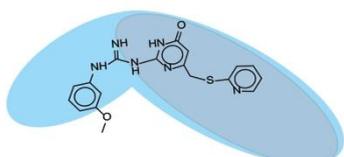
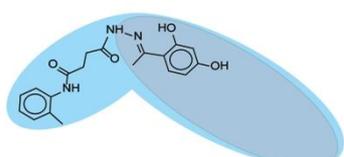
No	Name	structure	Ref.	Value	PDB
58	Aesculetin		[208]	inhibition = 17.4 %	
59	Chinic acid		[208]	IC ₅₀ = 13 mM	
60	Dihydrocaffeic acid		[208]	IC ₅₀ > 14 mM	
61	2,3-Dihydroxybenzoic acid		[208]	IC ₅₀ > 7 mM	
62	Benzoic acid		[208]	IC ₅₀ > 10 mM	
63	Ferulic acid		[208]	IC ₅₀ > 5 mM	
64	Isoferulic acid		[208]	IC ₅₀ > 7 mM	
65	Caffeic acid		[208]	IC ₅₀ = 4.8 mM	
66	Scopoletin		[208]	IC ₅₀ = 1.3 mM	
67	Ascorbic Acid		[210] [211]	IC ₅₀ = 26.5 μM	

10.5 Steric overlap of substrates with tested compounds

LigandScout 3.1 [159] was used for calculating Gaussian shape similarity score for compounds tested in both assays (first with *p*-NPG5 and second with GalG2CNP substrate) and listed in Table 19. It was performed attempting to explain the differences in inhibitory potencies (chapter 4.3.3). Cyan color represents the region of the *p*-NPG5 in both sides; silver color represents the region of GalG2CNP as appeared mainly at right side.

Table 19: Results of steric overlap and Gaussian scores for compounds tested in the first and second assays

compound	first assay_ <i>p</i> -NPG5		second assay_ GalG2CNP		Overlapping with substrates
	Gaussian score	IC ₅₀ (μM)	Gaussian score	IC ₅₀ (μM)	
<i>p</i> -NPG5	1	NA	0.39	NA	
GalG2CNP	0.396	NA	1	NA	
Bound acarbose	0.59	97.06	0.34	0.5	
2	0.24	441.5	0.24	138	
3	0.30	ND	0.19	200	

compound	first assay_ <i>p</i> -NPG5		second assay_ GalG2CNP		Overlapping with substrates
	Gaussian score	IC ₅₀ (μM)	Gaussian score	IC ₅₀ (μM)	
4	0.27	ND	0.19	300	
6	0.28	ND	0.20	1000	
7	0.30	86.62	0.37	ND	
8	0.30	97.375	0.29	ND	
9	0.35	93.58	0.22	ND	

NA means not applicable and ND means not determined.

10.6 Validation of docking experiments for α -amylase

The docking program GOLD 5.1 [184] was used first to reproduce the binding mode of the co-crystallized ligand in each used ligand-protein complex. The root mean square deviation (RMSD) was calculated in GOLD 5.1 between the heavy atoms of the original co-crystallized ligand and the docked conformation ligand. Validation of docking experiments for the PDB codes 1MFV and 3OLE for HSA and HPA, respectively, are depicted in Figure 48.

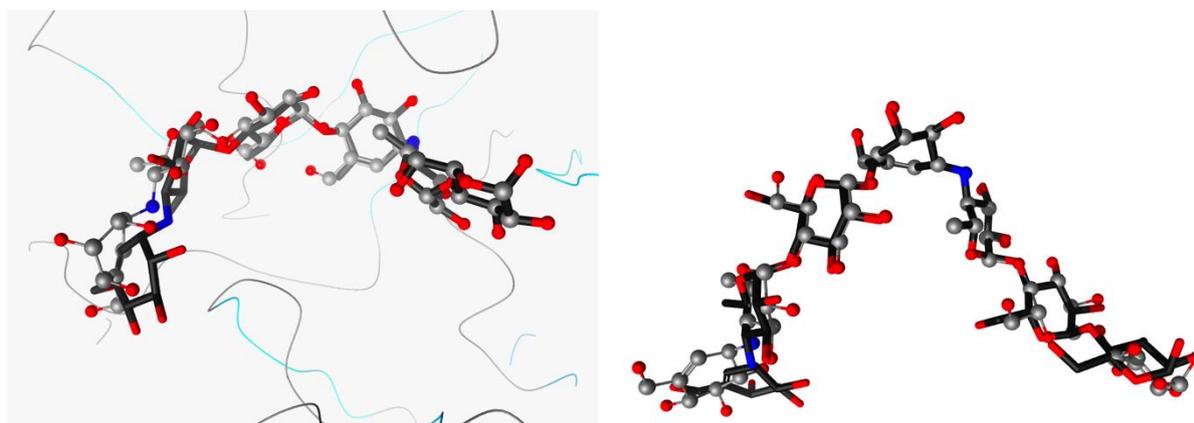


Figure 48: The software GOLD 5.1 generated binding mode (gray, balls and sticks) of pseudo-hexasaccharide bound acarbose (left, PDB: 1MFV [49]) and acarviostatin II03 (right, PDB: 3OLE [36]) compared to their experimental conformations (black sticks). RMSD between heavy atoms of the docked and co-crystallized conformations for the first ligand was 1.12 Å (left) and for the second one was 0.8 Å (right).

In HPA coded 4GQQ, three allosteric sites, located separately from each other and far away from the catalytic active site by about 20 Å, accommodate three molecules of ethyl caffeate [25]. The conformations of ethyl caffeate in the three allosteric binding sites were reproduced with GOLD 5.1. Superpositions of the original and generated conformations in sites 1, 2, and 3 are given in Figure 49.

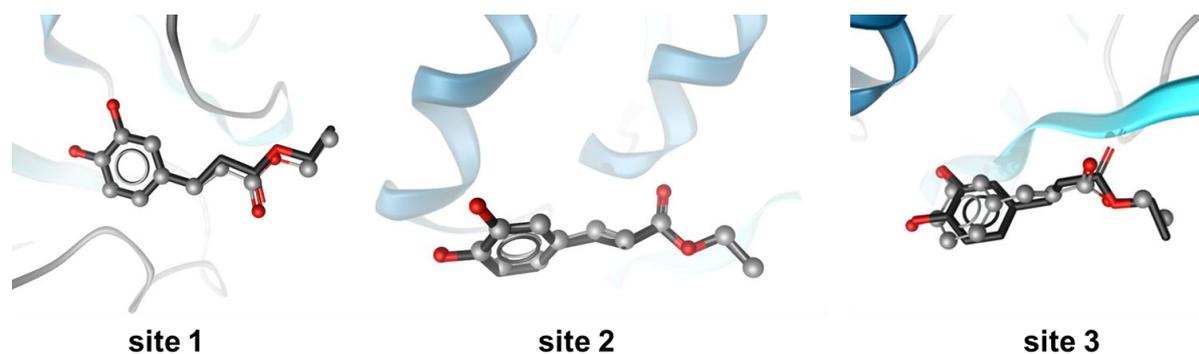


Figure 49: Docking poses of ethyl caffeate (black sticks) superposed with co-crystallized ligand (gray balls and lines) showing RMSD values of 0.5 Å, 1.07 Å and 1.46 Å in allosteric sites 1 (left), 2 (middle), and 3 (right) in the PDB code 4GQQ, respectively.

In the same manner, GOLD 5.1 was used to reproduce the conformation of myricetin co-crystallized with HPA and coded as 4GQR. Superposition of the original orientation with the created pose is depicted in Figure 50.

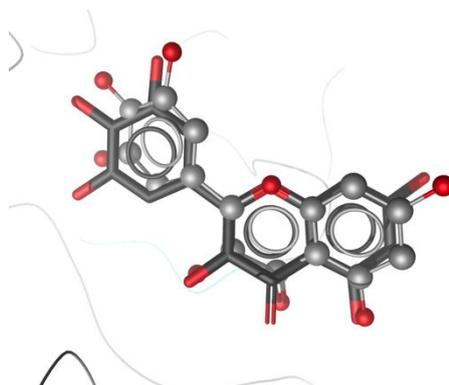


Figure 50: Generated binding mode of myricetin (gray lines and balls) superimposed with the co-crystallized ligand (black sticks) in HPA active site (PDB code: 4GQR [25]). RMSD value = 2.1 Å.

10.7 Docking of analogues of the best competitive inhibitor in the initial screening

Analogues of compound **2** were selected by similarity search from an in-house database (chapter 7.6) and biologically tested with GalG2CNP substrate and HSA as explained in chapter 4.3.10. Suggested binding modes for compounds **22**, **23**, **24**, **32**, and **33** were given in Figure 51.

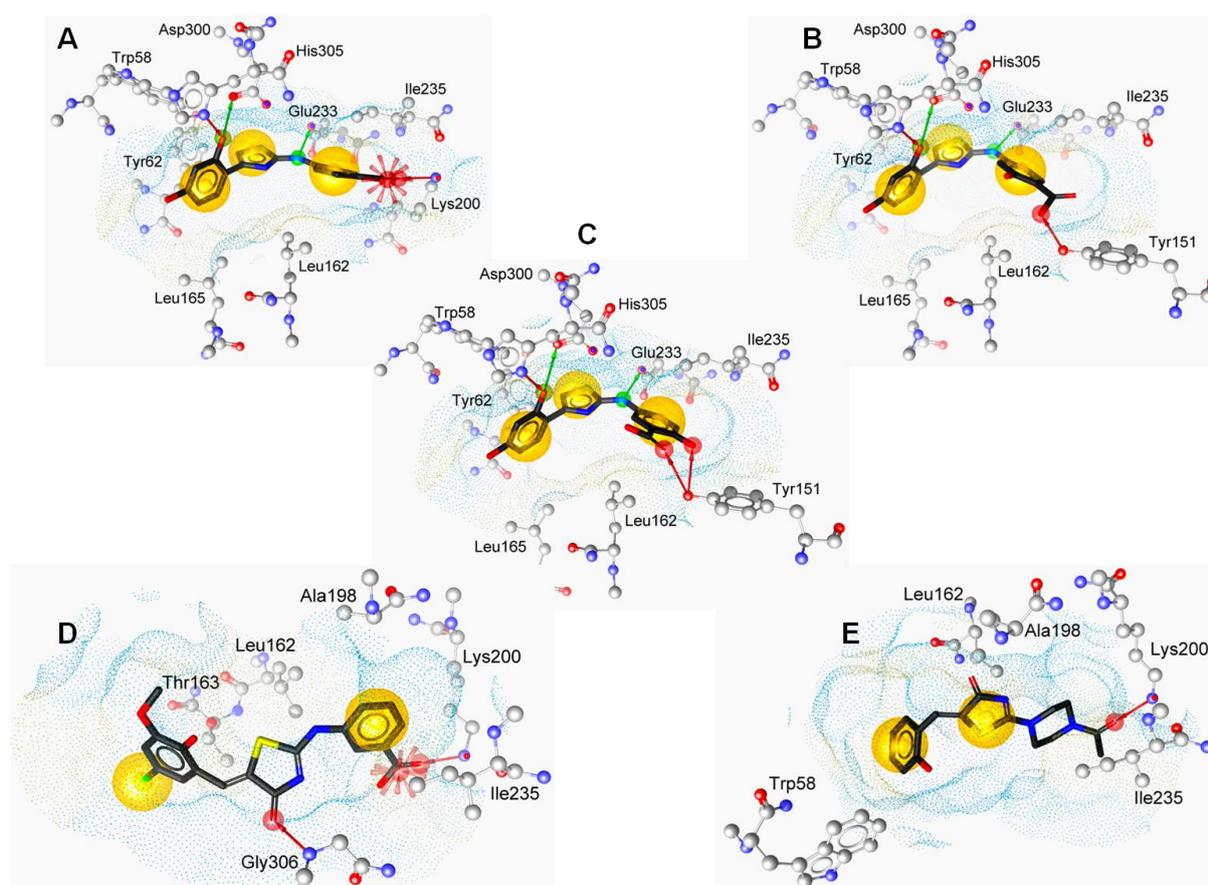


Figure 51: Docking poses of compounds 22 (A), 23 (B), 24 (C), 32 (D), and 33 (E). Ligands are shown in black stick mode and active site residues are rendered as gray balls and sticks. Yellow spheres represent hydrophobic contacts, red arrows: H-bond acceptors, green arrows: H-bond donors, red stars: Negative ionizable interactions. Blue and yellow dots represent hydrophilic and lipophilic surfaces of the receptor binding pocket.

10.8 Validation of models developed in the refined virtual screening

3D pharmacophores developed in the refined virtual screening (chapter 4.3.8) were validated using ROC curve. Dataset used for a first validation (Figure 52) were 19 biologically known active compounds (Appendix 10.3), and decoys composed of 60 inactive compounds (Appendix 10.4), and 254 created molecules used as decoys (chapter 7.2.2).

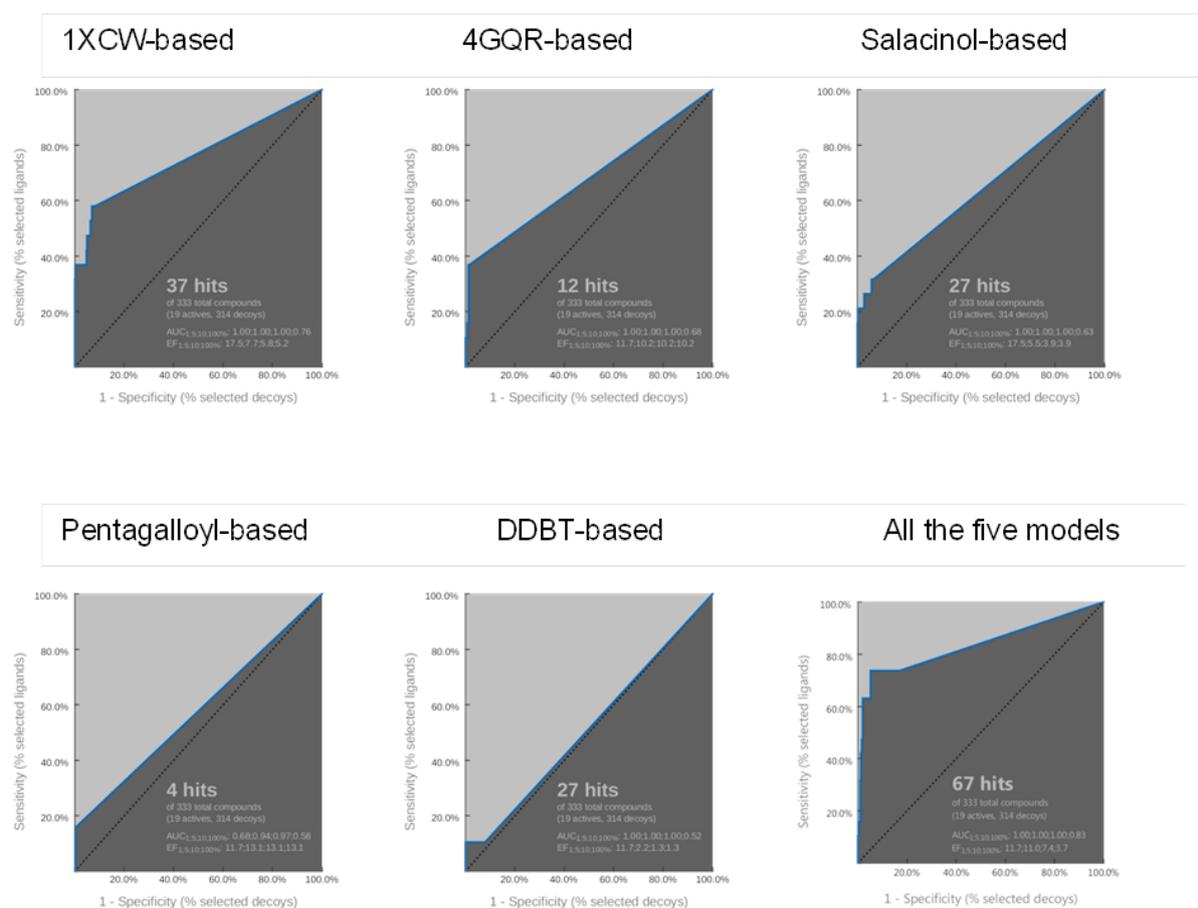


Figure 52: ROC curves for the first validation of 1XCW-based model, 4GQR-based model, salacinol-based model, pentagalloyl-based model, and DDBT-based model.

Second validation for the 3D models developed in the refined virtual screening (Figure 53) was performed using the previously mentioned dataset, in addition to 64,255 compounds from WDI2005 utilized as decoys.

The goal of this validation was to check the selectivity of the pharmacophores to retrieve the active compounds from large screening database. Validation results show that these models can retrieve the known active compounds and rank them within the first 500 hits based on the pharmacophore fit scores using LigandScout 3.1. The $EF_{1\%}$ as well as the $AUC_{100\%}$ reveal the quality of these pharmacophores and show promising reliability to be used for the refined virtual screening.

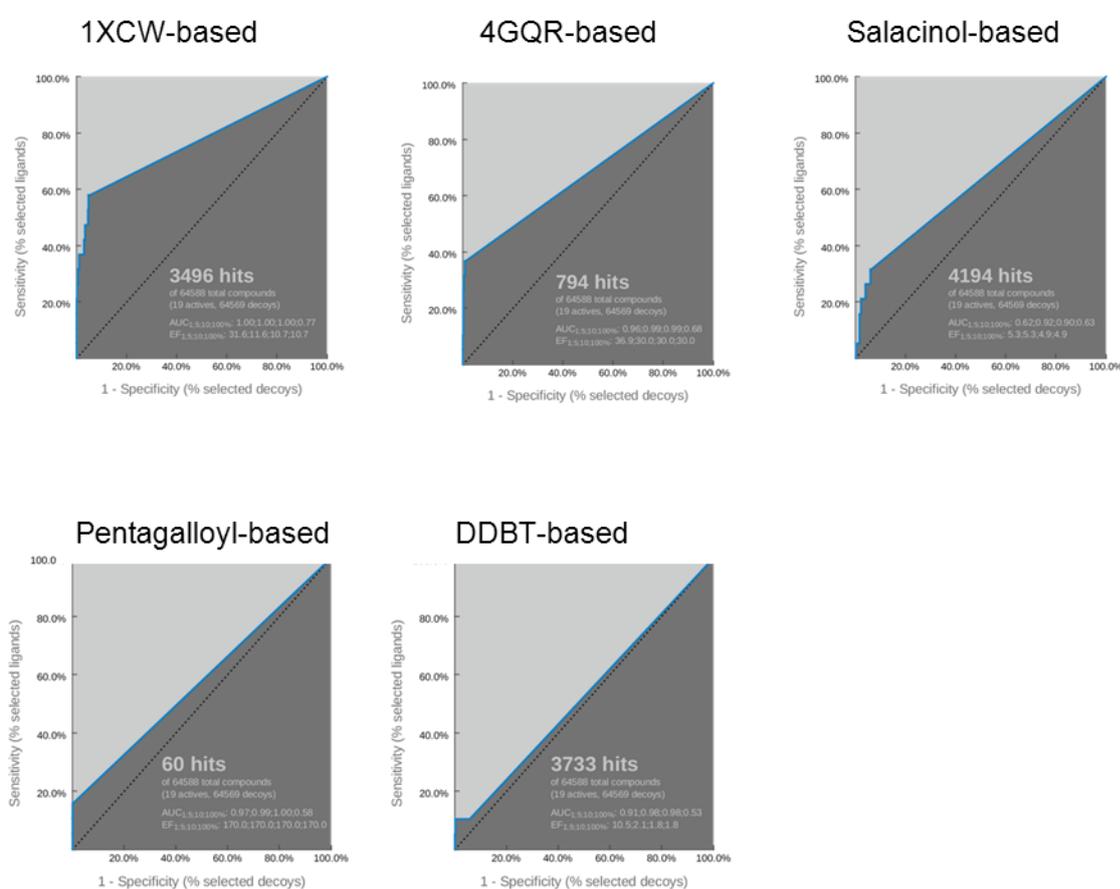


Figure 53: ROC curves for the second validation of 3D pharmacophores used in the refined virtual screening.

10.9 Docking of the best competitive inhibitor in the refined virtual screening

First, compound **20** was docked into PDB code 3OLE but resulted poses does not show any interactions with the catalytic residues. Therefore, another binding mode was suggested by docking experiments into the PDB code 4GQR (Figure 54). The most plausible pose occupies subsites -3 to -1 and shows five H-bonds at subsite -1: two HBDs between *meta*- and *para*-hydroxyl groups with the catalytic residue Asp197, two HBDs between *ortho*- and *meta*-hydroxyl groups with His101, and one HBA between *para*-hydroxyl group and Arg195. Compound **20** adequately displays stability by forming hydrophobic contacts with the lipophilic residues located in subsites -2 and -3 in the active site. These residues are Trp58, Trp59, Tyr62, Thr163, and Leu165.

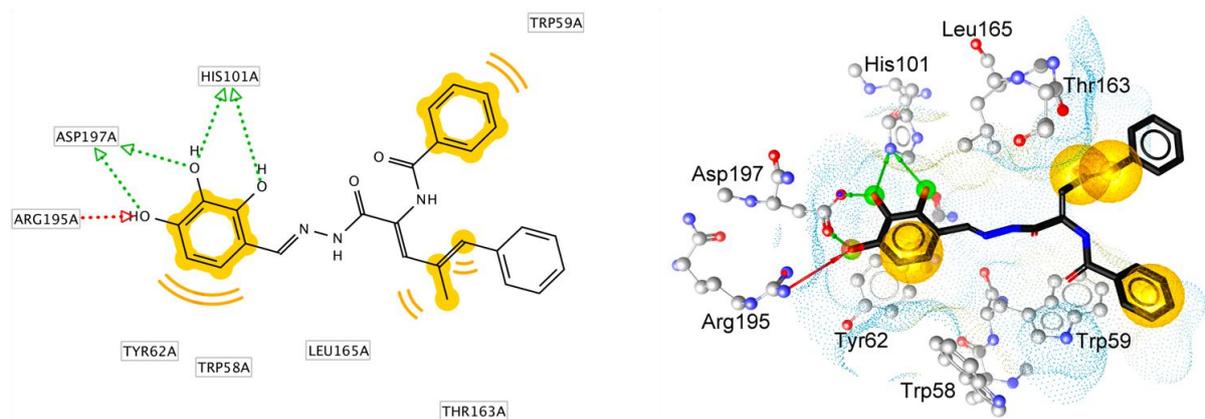


Figure 54: Suggested binding mode of compound 20 docked into active site of α -amylase (PDB: 4GQR). Yellow spheres represent hydrophobic contacts, green and red arrows symbolize H-bond donors and acceptors, respectively.

10.10 Quality control performed by vendors for selected compounds

Compounds selected in the initial and refined virtual screening campaigns and biologically tested against α -amylase were basically obtained from commercial vendors. Quality control for these selected hits was performed by commercial supplier.

Key Organics compounds

All compounds are subjected to rigorous quality control procedures with purity of more than 90 %. Purity is checked by ^1H NMR and ^{13}C , elemental analysis and mass spectroscopy.

<http://www.keyorganics.net>

Asinex

All compounds have a minimum purity of 90%. They have been characterized by either NMR and/or HPLC/MS methods. Complementary analytical data are available free of charge under the Customer's request.

<http://asinex.com/Libraries.html>

ChemBridge

All compounds are tested by LC-MS and/or NMR to confirm sample identity and minimum purity of 90 %. Electronic copies of NMR or LC-MS data are available upon request.

http://www.chembridge.com/screening_libraries/quality_control/index.php

Vitas-M

All compounds have a minimum purity level of 90%. All compounds are tested by NMR. NMR spectra can be supplied upon request.

<http://www.vitasmlab.com/compound-libraries-quality-control-2.htm>

Princeton Biomolecular Research

All compounds are required minimum purity level of 90% and are tested by NMR and LC/MS.

<http://www.princetonbio.com/pages4.html>

Enamine

All compounds have minimum purity level of > 90%. All compounds are tested by NMR spectroscopy, LC/MS, GC, FT/IR and quantitative elemental analysis.

http://www.enamine.net/index.php?option=com_content&task=view&id=22

Analyticon Discovery GmbH

Purity of compounds is more than 90% on average. Identity is confirmed by MS/NMR.

http://www.ac-discovery.com/content/Downloads/Product_Information.php

ChemDiv

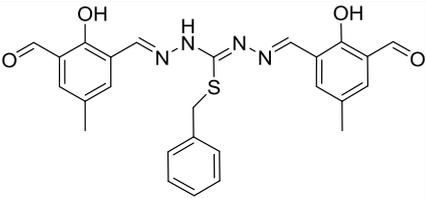
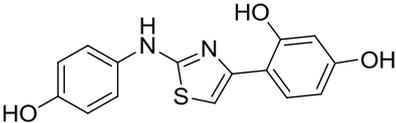
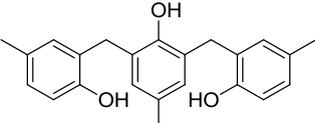
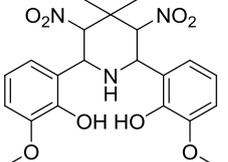
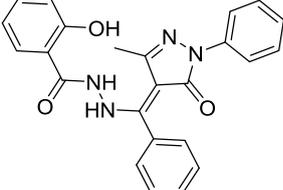
ChemDiv supports comprehensive QC of all compounds; by NMR, HPLC/MS/UV and other options. ChemDiv provides 100% quality control for all compounds and guarantee more than 90% purity (+/- 5% accuracy). The purity accuracy is confirmed by ¹H NMR and/or LC (UV)/ MS spectra in electronic format (MS TIF files) for all stock available compounds.

<http://www.chemdiv.com/additional-chemistry-services/>

10.11 Commercial compounds selected by virtual screening approaches

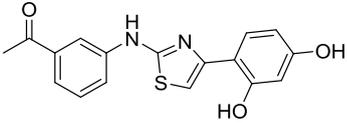
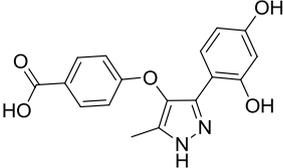
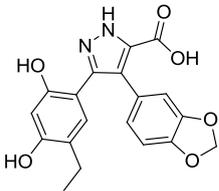
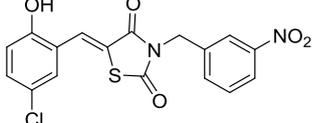
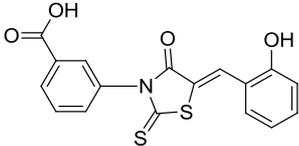
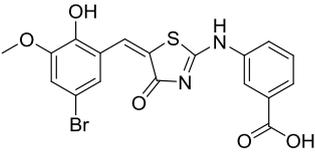
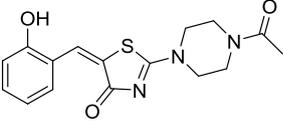
Virtual hits selected from the initial (compounds **1** - **14**) and the refined virtual screening campaigns (compounds **15** - **21**) were biologically evaluated. Compound **1** was excluded due to solubility issues. Compounds **22** - **33** were chosen from an in-house database by 2D searching for analogues similar to compound **2** (Table 20). Known biological activities of these compounds against targets other than α -amylase were also searched using CAS SciFinder (www.cas.org/products/scifinder) and listed below.

Table 20: Commercial chemical compounds chosen by both initial and refined virtual screening approaches and biologically tested against α -amylase. Known biological activities with targets other than α -amylase were also included below

Assay code	Structure	Vendor	Vendor's code	Known activity in literature
1		Asinex	BAS 00115253	-
2		Asinex	BAS 04880564	Fructose 1,6-bisphosphatase (FBPase) inhibitor, IC ₅₀ = 11 μ M [169]
3		Asinex	BAS 00411559	Food additive [212]
4		Asinex	BAS 01579757	-
5		Asinex	BAS 00341615	-

Assay code	Structure	Vendor	Vendor's code	Known activity in literature
6		Asinex	BAS 13152985	-
7		Asinex	BAS 01057772	Tumor Hsp90 Inhibitor, IC ₅₀ > 50 μM [164]
8		Chembridge	7973051	-
9		Chembridge	5468077	-
10		Key Organics	4F-062	-
11		Key Organics	10L-950	-
12		Analyticon	NP-003952	-
13		Analyticon	NP-007087	-
14		Analyticon	NAT19- 353800	-
15		Specs	AG- 219/09467025	-
16		Specs	AH- 262/36680018	-

Assay code	Structure	Vendor	Vendor's code	Known activity in literature
17		Specs	AK-968/41922716	-
18		Specs	AN-329/40366541	-
19		Specs	AG-205/04837062	Anti-microbial [213]
20		Chembridge	5668374	-
21		Specs	AG-690/09194027	-
22		ChemDiv	5186-0425	DNA gyrase B inhibitor, IC ₅₀ = 56 μM [214]
23		Vitas-M	STK017577	DNA gyrase B inhibitor, IC ₅₀ = 39 μM [214]
24		Vitas-M	STK051250	MurC ligase inhibitor, IC ₅₀ = 347 μM [215]. DNA gyrase B inhibitor, IC ₅₀ = 36 μM [214]
25		Vitas-M	STK087186	MurC ligase inhibitor, IC ₅₀ = 531 μM [215]. FBPase inhibitor, IC ₅₀ = 55 μM [169]. DNA gyrase B inhibitor, IC ₅₀ = 25 μM [214]
26		Vitas-M	STK031735	MurC ligase inhibitor, residual activity = 90% at 100 μM [215]

Assay code	Structure	Vendor	Vendor's code	Known activity in literature
27		Vitas-M	STK947495	MurC ligase inhibitor, residual activity = 83% at 100 μ M [215]
28		Princeton	OSSK_04528 7	DNA gyrase B inhibitor, $IC_{50} > 500$ μ M [214]
29		Princeton	OSSK_93743 8	DNA gyrase B inhibitor, $IC_{50} > 500$ μ M [214]. Heat-shock protein 90 (Hsp90) inhibitor, $K_d = 680$ nM [216]
30		Vitas-M	STK091393	DNA gyrase B inhibitor, $IC_{50} > 1000$ μ M [185]
31		Enamine	T0520-5342	DNA gyrase B inhibitor, $IC_{50} > 1000$ μ M [185]
32		Chembridge	6341222	DNA gyrase B inhibitor, residual activity = 86 % at 125 μ M/L [185]
33		Vitas-M	STK864500	DNA gyrase B inhibitor, residual activity = 97 % at 500 μ M/L [185]