Drug Off-Target Prediction: A Case Study with Aldehyde Dehydrogenase Inhibitor

by

Yusuf Doğuş Doğru

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Computational Sciences and Engineering

Koç University

December 2015

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Yusuf Doğuş Doğru

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Attila Gürsoy, Ph. D. (Advisor)

_____

Özlem Keskin Özkaya, Ph. D. (Advisor)

_____

Rengül Çetin Atalay, Ph. D.

_____

Nurhan Özlü, Ph. D.

_____

Halil Kavaklı, Ph. D.

Date: _____

Abstract

Aldehyde dehydrogenases (ALDH) are a large family of enzymes that maintain homeostasis through metabolizing reactive compounds. They play a pivotal role in the regulation of many important biological processes such as detoxification of alcohol and xenobiotics, amino acid salvage pathways, and retinoic acid synthesis. An attractive feature of ALDH is that it is a marker for both normal and cancer stem cells. It is thought to play a role in protection, differentiation and expansion of stem cells. DIMATE, a small molecule inhibitor of ALDH3A1, has been previously shown to display anti-proliferative effects on cancer cells.

Drug promiscuity refers to a drug having multiple targets. While multiple targets confer flexibility to a drug, it also creates unintended off-targets. Off-target detection is a crucial step of drug design to prevent unwanted side effects. Reverse docking may be used to identify off-targets, but it is very inefficient and has a high computational cost. An alternative to this is pharmacophore approach. A pharmacophore is a description of the chemical and geometrical features of a compound that is necessary for its interaction with a biological target to initiate or inhibit a biological activity. In this study, we identified the possible off-targets of DIMATE through similarity based approaches employing reverse-docking and through pharmacophore based approaches.

Özet

Aldehit dehidrogenaz (ALDH) enzimleri reaktif bileşiklerin metabolize ederek hücre içi homeostazı sağlayan büyük bir enzim ailesini oluşturur. Alkol ve ksenobiyotiklerin parçalanması, amino asitlerin geri dönüşümü ve retinoik asitin sentezlenmesi gibi önemli birçok biyolojik süreçte kilit rol oynarlar. ALDH'ın önemli bir özelliği de hem normal, hem kanser kök hücreleri için belirteç niteliğinde olmasıdır. ALDH'ın bu hücrelerde koruma, faklılaşma ve çoğalma üzerinde etkisi olduğu düşünülmektedir. ALDH3A1'in küçük molekül inhibitörü olan DIMATE'ın kanser hücreleri üzerinde büyümeyi engelleyici etkileri olduğu gösterilmiştir.

İlaç kirliliği ya da çok hedefli ilaç, bir ilacın birden fazla protein hedefleyebildiğini ifade eder. İlacın kirliliği ona esneklik sağlamakla beraber aynı zamanda istenmeyen yan etkilere sebep olabilir. Dolayısıyla ilaç tasarımı sürecinde yan etkilerin tahmini, istenmeyen yan etkileri önleyebilmek için önemli bir adımdır. Geriye dönük protein-ligand yerleştirme (docking) hedef dışı etkileri bulmak için kullanılabilir, ancak verimsizdir ve yüksek bilgisayar gücü gerektirir. Buna bir alternatif ise farmakofor yöntemleridir. Farmakoforlar bir molekülün biyolojik hedefiyle etkileşebilmesi için gereken kimyasal ve geometrik özelliklerin bütünüdür. Bu çalışmada DIMATE'ın olası istenmeyen hedeflerini benzerlik prensipleriyle geriye dönük moleküler yerleştirme ve farmakofor temelli analizler yaparak tahmin ettik.

Acknowledgements

First and foremost, I would like to offer my sincere gratitude to my two advisors, Prof. Attila Gürsoy and Prof. Özlem Keskin, for giving me the opportunity to work in their lab. They generously offered to take me in, and train a molecular biologist with an experimental background in the field of computational biology. Their support has enabled me to improve myself.

I would like to thank my thesis committee members Dr. Rengül Çetin Atalay, Dr. Nurhan Özlü and Dr. Halil Kavaklı for the critical reading and evaluation of my thesis. I am also grateful to Advanced Biodesign, especially the president and CEO of the company İsmail Ceylan, who collaborated with our lab on this project. I would also like to acknowledge Koç University for their financial support throughout my graduate studies.

I would like to thank all past and present members of the COSBI lab for making our lab a warm, energetic and fun environment. I am grateful to Engin Çukuroğlu for planning my life, Güray Kuzu for being the best mentor, Emine Güven Maiorov for our gossip time, Ece Özbabacan for being the best neighbor, Billur Engin Aras for being the best assistant. I thank MBG Bilkentiers Nilay Karahan, Derya Cavga and Emine Sıla Özdemir, and also Serena Muratçıoğlu, Farideh Halakou, Cemal Yamak, Efe Elbeyli and Bilgesu Erdoğan for a wonderful Bodrum trip. I would like to especially thank Alper Başpınar for full-time support with full-time negativity, my fellow countryman Tayfun Tümkaya, and Deniz Demircioğlu for very delightful times he enabled both intentionally and unintentionally. I thank Emel Şen Kılıç for being responsible partner, and all other lab members including Büşra Topal, Tuğçe Yıldızoğlu, Selin Karagülle.

I would especially like to thank Nilay and Derya for their invaluable contribution to the initial parts of the project. Their involvement was essential for the exploration of new approaches and development of novel ideas.

I also wish to thank Barış Çağlar, Erkut Demirhan, Ayşegül Turupçu, Süha Mutluergil, Ceren Şeref and Nazlı Ataç for their friendship, support and the best coffee times.

Most importantly I would like to thank my wife, Duygu Payzın Doğru for all she has done for me. Without her, neither this thesis, nor my life would be complete.

Finally, I thank my family: my father Musa, my mother Nurcihan, my sister Bakış and my brother-in-law Gökay and all of my other friends, whose constant love and support have enabled me to challenge myself, and continue to follow my passion to work in the biological sciences.

Table of Contents

List of Tables

List of Figures

Nomenclature

ABC	ATP-binding cassette
ALDH	Aldehyde dehydrogenase
CSC	Cancer stem cell
DIMATE	dimethylampalthiol ester
DNA	Deoxyribonucleic acid
GOLD	Genetic Optimisation for Ligand Docking
GPCR	G-protein coupled receptor
HNE	4-hydroxy-2-nonenal
MATE	morpholinoampalthiol ester
MDR	multidrug resistance
MVD	Molegro Virtual Docker
NAD	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
RNA	Ribonucleic acid
siRNA	Short interfering RNA
SMILES	Simplified molecular-input line-entry system

Chapter 1

INTRODUCTION

Aldehyde dehydrogenases are a large family of enzymes that play an active role in many important biological processes to maintain cellular homeostasis [1, 2]. Recently, they have been shown to be markers for both normal and cancer stem cells, which make them attractive candidates for targeted cancer therapy [1]. A small molecule inhibitor, DIMATE, has shown promising results in inhibiting the proliferation of cancerous, but not normal cells [3].

An important aspect to drug development is drug off-targets. Off-targets are the secondary, unintended targets of a drug. They result from drugs having multiple targets within a system, referred to as drug promiscuity [4]. The exact reason behind this promiscuity remains unknown.

In this study, multiple approaches were employed to characterize the possible interactions DIMATE within the cell. Chapter 2 provides necessary background and literature review on aldehyde dehydrogenases, drug promiscuity and pharmacophore models. Cellular functions of ALDH, previously identified inhibitors and principles of drug promiscuity are reviewed. Aldehyde dehydrogenase family of enzymes, their significance in normal physiological processes and disease phenotypes are introduced.

Chapter 3 describes the dataset and tools used in the study. The limitations of the dataset are explained. Statistical analyses used to derive principles of drug promiscuity and off-targets are introduced.

Chapter 4 presents the results of similarity based and pharmacophore based identification of possible off-targets of DIMATE. Ligand similarity, binding pocket similarity and

reverse docking approaches are used to suggest possible targets of DIMATE, other than its intended target.

Chapter 5 discusses of a significance of these findings. The limitations of the study are discussed in detail. The studies implications in drug off-target prediction and future directions are analyzed.

The thesis is concluded with a short summary of the performed study.

Chapter 2

LITERATURE REVIEW

An important consideration of drug design and development is drug off-targets. Off-targets are secondary, unintended targets of a drug [5]. In some cases, drug-off targets may turn out to be useful, as in the case of plerixafor, a drug that was initially designed against HIV to inhibit its interaction with the CXCR4 receptor, but then repurposed as a stem cell mobilizing agent [6, 7]. In other cases, off-targets cause unwanted side effects. Thalidomide is one such drug, which was initially prescribed for nausea in pregnant women, but turned out to have severe side-effects that led to birth defects in thousands of children in the 1960s [8]. Therefore, it is crucial to be aware of any potential off-targets of drugs prior to releasing them for public consumption.

Not all side-effects are easily detected prior to actual phase I trials, that is, before the use of human subjects. Pre-clinical studies in wet-laboratories are mostly conducted in cell cultures, which may not fully recapitulate the whole organism. Recently, the power of computation has been realized to screen drugs and predict their possible off-targets. In this study, we are interested in the prediction of off-targets of a small molecule inhibitor of aldehyde dehydrogenase (ALDH). ALDHs are biologically important molecules that regulate cellular homeostasis/ through the oxidation of aldehydes to their corresponding carboxylic acids [9]. These enzymes have been shown to be markers of cancer stem cells, the tumor initiating cells [10-15]. Therefore, ALDHs have recently received wide attention as possible targets for cancer therapy [16].

In this section, I will introduce the functions aldehyde dehydrogenase in the cell, describe small molecule inhibitors of ALDH, and outline work on drug promiscuity and off-target prediction.

2.1 Enzyme Inhibitors

Inhibitors of enzymes are defined as molecules that are capable of binding to an enzyme and reduce its activity. Most drugs are enzyme inhibitors that disrupt metabolic processes of the target pathogen. Inhibitors generally work to prevent an enzymes interaction with its substrate, or obstruct the enzyme's ability to catalyze the reaction. Inhibitors can be generally classified as reversible or irreversible, depending on the interactions they have with the enzyme.

2.1.1 Reversible Inhibitors

Reversible inhibitors, as the name implies, have reversible interactions with their targets. These are a result of non-covalent interactions, such as electrostatic interactions (e.g. ionic bonds and hydrogen bonding), van der Waals interactions, and hydrophobic effects, which require considerably less energy to break down than covalent interactions [17]. Therefore, the effects of reversible inhibitors can be reversed simply by diluting the inhibitor in the environment

Reversible inhibitors can be further categorized as competitive, non-competitive and uncompetitive inhibitors [17]. Competitive inhibitors work by preventing binding of the substrate to the enzyme. This could either be achieved by the inhibitor itself binding to the active site of the enzyme, to which the substrate normally binds. Non-competitive inhibitors bind to the enzyme at a site different than the active site. Thus, they do not hinder the binding of the substrate to the enzyme. Rather, they prevent enzyme turnover. Finally, uncompetitive inhibitors bind to the complex of enzyme and substrate, resulting in reduced availability of the enzyme and substrate active complex, decreasing the overall maximum velocity of the normally catalyzed reaction.

2.1.2 Irreversible Inhibitors

Irreversible inhibitors are inhibitors that permanently render the enzyme inactive [17]. In this case, covalent bonds are involved in the mechanism, which modify the functional

groups within the active site of the enzyme, which are key to the normal catalytic function. Irreversible inhibitors are classified as group-specific reagents, substrate analogs, and suicide inhibitors. Of these, suicide inhibitors are the ones with most biological application. They are essentially modified substrates, hence are still specific to the enzymes active site. When a suicide inhibitor is introduced to the environment, it is first processed by the enzyme as if it is a normal substrate. However, the intermediate generated modifies the active site such that the enzyme becomes inactive.

2.2 Aldehyde dehydrogenases

ALDH enzymes are involved in reactive compound metabolism, which make them highly attractive targets for inhibition for cancer therapy. They are specific to many substrates that allow these enzymes to oxidize many different aldehydes that are important in maintaining a number of physiological processes [18]. In the most general terms, the aldehydes that form from both endogenous and exogenous precursors may be healthy or harmful. Exogenous aldehydes are generated from drugs, such as ethanol, or from sources readily present in the environment such as cigarette smoke [1, 19, 20]. Endogenous aldehydes, on the other hand, form during normal cellular processes such as amino acid, vitamin and lipid metabolism. Different aldehydes are responsible from different functions. For example, the aldehyde retinal is necessary for vision. When it is oxidized by ALDH, it becomes retinoic acid, with is required in cellular differentiation. Another biologically important group of aldehydes are electrophiles that are generated during oxidative stress. ALDHs are able to metabolize these cytotoxic aldehydes into less reactive groups [21, 22].

The mechanism through which ALDHs oxidize aldehydes into their carboxylic acids is by using NAD or NADP as coenzymes. ALDHs are found in a number of isoforms and in almost all organisms [23-25]. Moreover, ALDH can have a variety of subcellular localizations, such as the cytoplasm, nucleus, ER and mitochondria. The expression level of ALDHs also change depending on the tissue type [19].

Interestingly, ALDHs also have enzyme activity-independent functions, such as the ability to bind to hormones and small molecules [18]. They are also important in preventing the harmful effects of UV radiation on the cornea. ALDH mutations can result in irregularities in the aldehyde metabolism, and hence lead to diseases such as type II hyperprolinemia, γ-hydroxybutyric aciduria, and Sjögren–Larsson syndrome. ALDHs have also been implicated in the developments of complex diseases, such as cancer and Alzheimer's. Excitingly, ALDH enzymes are markers for healthy and cancer stem cells. Currently there 10 known functional genes of ALDH in humans [1].

2.3 Families of ALDH proteins

In humans, the ALDH proteins are separated into 11 families and 4 subfamilies. Each has different chromosomal locations. The specific functions of each ALDH family are outlined below:

Members of the ALDH1 family are found mainly in the cytosol, and are involved in the oxidization of retinal and some other aliphatic aldehydes. They are important in the protection of the eye lens from UV absorption, which leads to the production of peroxidic aldehydes [26]. ALDH1 also has some function in the detoxification of anticancer drugs. Therefore, ALDH1s have been shown to be involved in drug resistance, a phenomenon commonly observed in cancer cells [27]. This is most likely thought to result from the activation of transcription of the ALDH1 genes [28].

ALDH2 family is made up of enzymes of the mitochondria. They have high affinity for acetaldehyde, and have functions in detoxification. It is involved in alcohol metabolism, and a deficiency of ALDH2 leads to sensitivity to alcohol [20, 29]. ALDH3 enzymes function in fatty aldehyde and peroxidic aldehyde metabolism. The members of this family are highly expressed in the cornea, liver, ER and cancer cells [30]. ALDH4 enzymes are mitochondrial, and play a role in the oxidation of γ-semialdehydes [31]. ALDH5 enzymes are succinic semialdehydes dehydrogenases. They are highly active in

the human brain [32]. Upon ALDH5 deficiency, succinic semialdehydes accumulate in the brain and result in severe neurological conditions such as mental retardation and seizures [33]. ALDH6 enzymes are CoA-dependent methylmalonate semialdehyde dehydrogenases. They are involved in the breakdown of valine and pyrimidines [34]. ALDH9 family of enzymes is responsible in the 4-aminobutyraldehyde and aminoaldehyde metabolism. ALDH9 enzymes are highly expressed in the developing brain, but poorly in the adult brain. In the adult, they are found in liver, kidney and muscle cells [31]. ALDH16 enzymes are widely distributed in the body, such as the heart, kidney and lungs. However, their functional significance remains unknown [9]. ALDH18 enzymes are found in the mitochondrial inner membrane, and are dependent on ATP and NAD(P)H. These enzymes have dual activities of γ- glutamyl kinase and γ-glutamyl phosphate reductase. ALDH18 functions in a crucial step of the proline biosynthesis pathway as well as ornithine and arginine. The enzyme is mostly found in pancreas and kidney. Deficiencies in this enzyme cause metabolic and neurologic disorders [9].

2.4 Structure of ALDH

Although ALDH enzymes have common functions and substrates, ALDH isozymes possess different specificities for their substrates [2]. These differences in their active site allow specifically targeting each isozyme for inhibition.

Many ALDH enzymes have been crystallized with their substrates, which has enabled the description of the enzyme's active site [35-39]. Through these studies it was shown that many of the isozymes possess the same dehydrogenase and esterase enzyme mechanism. One exception to this is ALDH6A1, which uses CoA instead of NADP as its cofactor [40, 41]. The great similarity between different ALDH enzymes is most likely due to the conservation of important catalytic residues within each enzyme. These are Cys302, Lys192 and Glu268 [42].

Mammalian aldehyde dehydrogenase enzymes function as homotetramers or homodimers. Each component of these complexes bears an active site. When in tetramer form, the enzyme complex may have full or half site reactivity, which is dependent on the pH [43]. The monomers are composed of 500-600 amino acids, but there are exceptions that have 800-900 residues [44, 45].

Crystallographic studies have revealed that ALDH enzymes in all of its subfamilies share some common domains, such as the NAD binding, catalytic, and oligomerization domain (Figure 2.1) [35, 36, 38, 46, 47].

Mutagenesis of key residues of ALDH enabled the study of the enzymatic catalysis. These experiments demonstrate that catalysis occurs though five distinct steps [48]. These steps are illustrated in Figure 2.2. The first is the activation of a catalytic cysteine residue (Cys302). This occurs with the proton abstraction of Glu268 residue. In step two, the thiolate group of Cys302 does a nucleophilic attack to the substrate aldehyde. In the third step, this nucleophilic attack forms a tetrahedral thiohemiacetal intermediate, and at the same time shifts a H^- anion to the cofactor NAD^+, making it NADH. In the fourth step, the thioester intermediate is hydrolyzed. Finally in the fifth step, the cofactor is released, and the enzyme is restored by NAD^+ binding [48].

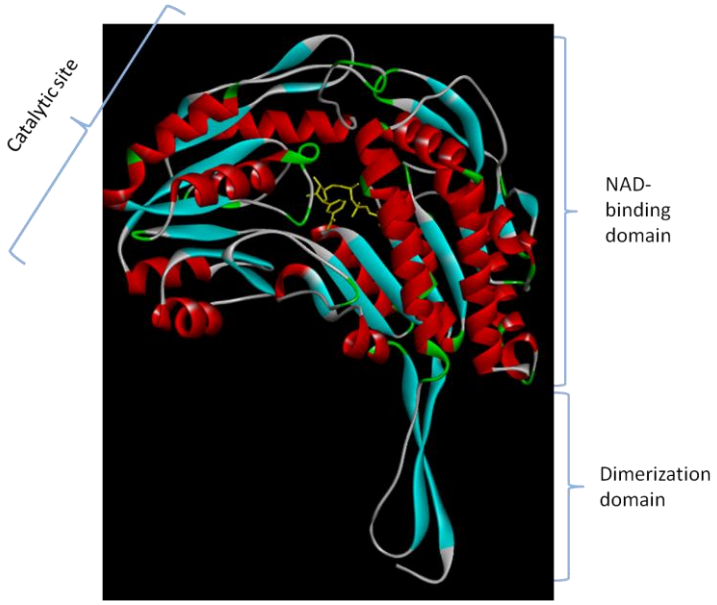


Figure 2.1 Structure of Aldehyde dehydrogenase
Figure downloaded from Protein Data Bank.

2.5 ALDH as a normal and cancer stem cell marker

Cancer stem cell theory postulates that cells within a tumor are heterogeneous, with each cell bearing a different degree of plasticity and differentiation potential [49]. Cancer stem cells possess the ability to self-renew and give rise to other specialized cells, properties attributed to normal stem cells. A striking property shared between normal and cancer stem cells is the high expression levels of the enzyme aldehyde dehydrogenase [1].

ALDH is thought to play a role in protection, differentiation and expansion of stem cells. In fact, ALDH positivity by immunohistochemistry (Aldefluor) is used to identify stem cell populations. For example, ALDH1A1 is considered to be a marker of human hematopoietic progenitor cells [50]. ALDH1A1 is also thought to be a marker or normal and cancer stem cells. ALDH1A1 and ALDH3A1 are especially important in proliferation and differentiation [51].

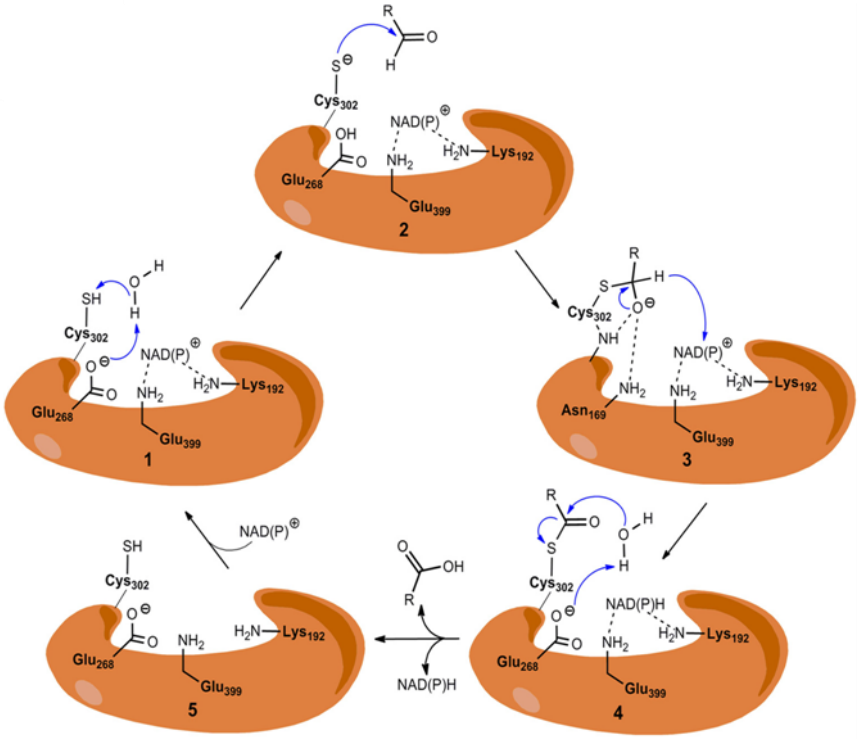


Figure 2.2 Proposed mechanism of aldehyde dehydrogenase oxidation
Figure reprinted from [42] with permission.

ALDH enzymes are much more active in hematopoietic stem cells (HSCs) than mature cells of the blood system. For example, lymphocytes only very poorly express ALDHs, whereas erythroid progenitors express ALDH in higher levels [52]. This difference makes HSCs resistant to alkylating agents such as mafosfamide or 4-hydroperoxycyclophosphamide (4-HC). If ALDH enzymes are inhibited, HSCs become sensitized to anticancer drugs that target tumor cells or drugs that are used during bone marrow transplants [53].

Aside from hematopoietic stem cells, other types of stem cells also possess high ALDH expression, such as neural [54], myogenic [55] and mammary [56] cells. High ALDH

expressing cells from mouse brain are multipotent, can self-renew and give rise to neuroepithelial stem cells. ALDH expression is higher in embryonic neural tissues compared to the adult, signifying ALDH's role in proliferation [54]. Stem cells of the muscle also express high levels of ALDH, which are resistant to oxidative stress, enabling them to be protected from cytotoxic effects of ischemic injury [55]. In mammary progenitor cells, cells highly expressing ALDH cannot self-renew but can generate luminal and myoepithelial cells [57].

2.6 ALDH3

ALDH3 family of enzymes is of particular interest because they bear important roles in regulation drug resistance and proliferation. They are responsible from the oxidization of medium-chain aliphatic and aromatic aldehydes [9]. ALDH3A1, a cytosolic ALDH, is expressed in high levels in normal tissues such as stomach and cornea, but very poorly expressed in the liver. It is also expressed in a number of human tumors, and in stem cells [58-60]. ALDH3A2 is a liver microsomal ALDH [61, 62]. It was shown to be involved in Sjögren–Larsson syndrome, an ichthyotic disease [63]. ALDH3B1 is expressed in the lungs, prostate and kidney [64, 65]. It is found in the cytosol. ALDH3B2 is expressed in the salivary gland and in the placenta [66].

Protein and activity levels of ALDH3A1 are correlated with cell proliferation rate. Inhibition or activation of ALDH3A1 show this direct correlation, and also hints at the possibility of inhibiting cell proliferation in cancer cells, and inducing cell proliferation in normal healthy stem cells in response to tissue damage [18].

2.6.1 Mechanism of ALDH3A1 regulation of cell proliferation

With the current knowledge today, how ALDH3A1 controls cell proliferation remains unknown. However, it is believed to involve the regulation of genes that are responsible for modulating cell growth, apoptosis or lipid peroxidation [18].

Many of the studies investigating the mechanism of ALDH3A1 in control of gene expression have been conducted in lung cancer cells, because ALDH3A1 was shown to be expressed in significant levels in non-small-cell lung carcinoma cells in patients. Therefore, lung cancer cells present an important model to study how ALDH contributes to homeostasis and cancer. Inhibition of ALDH3A1 has been shown to have an effect on genes such as *CCL20, GPR37, EIF1AY*, and *HMGA2*, which are known to be involved in the modulation of transcription, proliferation and apoptosis [67-69].

ALDH3A1's control over lipid peroxidation products may also be a mechanism through which ALDH3A1 regulates cell proliferation. Lipid peroxidation is a reaction associated with oxidative stress on unsaturated fatty acids, and results in the generation of reactive aldehydes. Such oxidative stress may result from xenobiotics or UV radiation [70]. ALDH3A1 is responsible from the metabolism of aldehydes that are products of lipid peroxidation, such as alkanals, which are known to be toxic [18]. For example, ALDH3A1 is highly specific to 4-hydroxy-2-nonenal (HNE), which is extremely cytotoxic. It results in glutathione depletion [71], interferes with DNA and RNA synthesis [72], and prevents mitochondrial respiration [73]. These result in serious effects and inhibit proliferation. However, ALDH3A1 activity can metabolize this toxic by-product and prevent these biological effects. It has been shown that overexpression of ALDH3A1 in normal and cancer cells confer resistance to the toxic products of lipid peroxidation [74].

2.6.2 Inhibitors of ALDH3A1

Inhibition of ALDH3A1 is achieved through the use of synthetic inhibitors or by short interfering RNA molecules (siRNAs). Moreover, indirect inhibition of ALDH3A1 is also possible through the inhibition of peroxisome proliferator-activated receptor γ (PPARγ) [18]. This is thought to be achieved through the inhibition of NF-κB or AP1 binding activity [21, 75]. Small molecule inhibitors of ALDH3A1 include tetraethylthiuram disulfide (disulfiram), an irreversible inhibitor, diethylaminobenzaldehyde, a slow

competitive inhibitor and S-methyl ester (ampalthiol ester) [18]. Two derivatives of S-methyl ester are especially useful for ALDH3A1 inhibition: dimethylampalthiol ester (DIMATE) and morpholinoampalthiol ester (MATE). These two inhibitors are shown to decrease proliferation of tumor cells significantly. This has been demonstrated in cell culture for a wide variety of cell lines, including lung cancer cells [67], glioblastoma cells [76] and hepatoma cells [77].

Inhibition of ALDH3A1 activity to same extent in different cells can result in varying results. Experiments with normal and cancerous human prostate epithelial cells demonstrated the selectivity of DIMATE and MATE. Both DIMATE and MATE were shown to be reversible inhibitors of the normal epithelial cells, but were irreversible inhibitors of cancerous epithelial cells [3]. In the cancer cells, apoptosis is induced.

2.7 Drug Promiscuity

Off-targets are the unintended targets of a drug. Until recently, a drug having multiple targets was considered an exception causing side effects, rather than being common. It is now becoming clear that a drug having multiple targets, referred to as drug promiscuity, is in fact common [4]. The exact cause of drug promiscuity remains unknown. This phenomenon is disadvantageous because it results in off-targets. However it is also advantageous because it allows drug repurposing, where new uses are discovered for readily used drugs [78]. This may be especially important in treatment of complex diseases. Current attempts to discover new targets of a drug range from genome wide association studies [79] to structural binding site comparison [80, 81].

2.8 Pharmacophore Models

Whether a small molecule interacts with a protein is determined by chemical interactions. Binding occurs when these interactions are energetically favorable. This means that only the small molecules that are of the right size, and that possess chemical properties complementary to the target protein will bind and will have a biological effect. A

protein's probability to bind to a small molecule is determined by the chemical properties of the amino acids from which it is composed, the size of the binding site, and the overall shape of the protein. This indicates that all of the small molecules that bind to the same binding site of a protein have similar chemical properties, sizes and shapes. All of these properties could be captured and represented as pharmacophore models. The idea the pharmacophore modeling relies on is that chemical properties present in similar arrangements in space result in biological activity on the same target protein [82]. These chemical properties are commonly referred to as features, and include hydrogen bond acceptors, hydrogen bond donors, hydrophobicity, ionizable groups and metal coordinating areas. The pharmacophore models do not retain information on the atoms of the small molecules and the target proteins, but only store information on the chemical features. For example, both NH_2- and OH- groups are simple hydrogen bond donors, and the pharmacophore model treats them as similar features. The aim of pharmacophore modeling is to predict the potential biological effect of the small molecule by deciding whether the small molecule matches the pharmacophore model of the target protein. The advantage of using this technique in virtual drug screening is that it allows the elimination of the compounds that are not compatible to the pharmacophore model of the target protein early on in the screen. Hence, it spares the experimenter significant resources that would be spent to investigate the biological activity of the small molecule in *in vitro* experiments [83, 84].

There are two methods through which pharmacophores are generated. The first is the ligand based method, which is used when a 3D structure of a molecule is not available, but when there is some information on the molecules that bind to the target [85]. The second method is the structure based method that is used when a 3D structure of the target protein is available. One alternative lesser known pharmacophore generation method is to predict the structure of target protein through amino acid sequences by homology modeling, and then using this structural prediction in this structure based method [86, 87].

2.9 Pharmacophore Libraries

2.9.1 Dataset generation

For both the model generation and validation steps, different kinds of reliable sources needed. The qualities of the resources are of high importance, because if the starting data is incorrect, the resulting model will be misleading. For data on the small molecules and their activities on the target molecule, there are multiple public sources available, such as ChEMBL [88] and PubChem [89]. Both of these sources contain large collections of small molecules with their activity data. For 3D structures of the target proteins, the Protein Data Bank [90] is a rich source that contains over 100,000 protein structures. These structures often include co-crystallized ligands which are important for obtaining conformations of the targets proteins and the small molecules in their biologically active states.

In the ligand based pharmacophore modeling, a training set is essential. The compounds of the training set will be the compounds on which the pharmacophore model will be based. The training set should contain a minimum of 2 compounds that are known to interact with the target protein. Some pharmacophore model programs allow the use of multiple small molecules in the training set with different degrees of biological activities on the target protein. Such an approach will allow the model to better distinguish the chemical features that are required for high and low biological activity [91].

Pharmacophore modeling also requires a test set. The test set determines whether the pharmacophore model can recognize most of the biologically active molecules and separate them from biologically inactive molecules. Such a test set allows to theoretically validate the freshly generated pharmacophore model. The training set should ideally contain molecules with great structural diversity. In addition, the use of molecules that are biologically inactive is an advantage.

As the case with any small molecule database, virtual screening can only detect a small percentage of the all the small molecules that are biologically active. This possibility should also be included in the model. Therefore, a decoy set with a small number of active molecules and a large number of inactive molecules need to be tested in the model. The decoy set could be developed from a set of random small molecules with unknown activities since the probability of these molecules to be biologically active will be quite low [92]. Directory of Useful Decoys (DUD) [93] offers readily available decoy sets for some targets which may be used in the pharmacophore model. For other targets, databases such as ChEMBL or Virtual Library [94] may be used.

2.9.2 Conformational Analysis

The structural information available for small molecule could contain 1D, 2D or 3D information. 1D information usually consists of simply the name of the molecule, its physical formula, and the chemical properties. The 2D information usually contains information on the structure of the molecule. The 3D information contains information on the geometry of the molecule. The descriptions of the molecules within the dataset that are used for the pharmacophore model are usually in the 2D representation. For pharmacophore modeling, however, 3D structures are needed. There are computational tools available that convert these 2D structures into 3D form. Examples to such tools are CONCORD [95] and CORINA [96]. These 2 tools, although very useful, represent the molecule in single conformations where the molecule is in its most relax form and hence has the lowest possible energy value. These tools assume that the active conformation of the molecule is within this state, however, this might not always be the case. There may exist other conformations of the molecule that have higher energy values that are still favorable. Therefore, a detailed conformational analysis for all of the small molecules used in the model needs to be performed [85]. The purpose of this analysis is to generate conformations that are energetically favorable among all of the possible conformations. There are a number of tools that employ different methods to generate such conformations. Examples to these tools are OMEGA [97], tools in DiscoveryStudio [98],

tools in Schrödinger Suite and in Molecular Operating Environment (MOE) [99]. All of these tools generate high quality conformations of molecules. Since these conformations complement the ligands that are available in the PDB, they are suitable for the prediction of the conformations that are biologically active.

2.9.3 Pharmacophore Model Construction

In the ligand base pharmacophore modeling method, the models are all based on the shared chemical properties of the molecules in the training set, and their representation is dependent on the algorithm. These algorithms can be categorized as RMSD based or overlay based methods [100]. RMSD based methods make use of the distance between the feature group of the molecule and the feature group in the pharmacophore model of the target protein. On the other hand, the overlay based methods consider the radius of the feature and approximate the suitability of that feature.

In structure based pharmacophore modeling, the model is generated from a ligand-protein complex [85]. Such a complex can be constructed via docking, which predicts the orientation of a small molecule in which it binds to its target. However, where available, protein structures in the protein data bank may be used instead. The analysis of the interactions between a ligand and a protein is a crucial step in these structure based pharmacophore models. This analysis can be performed manually by examining the amino acids of the binding site in close proximity to the ligand. However, there are tools that readily perform this analysis. Although these interactions could be used in the pharmacophore models as they are, some refinement is usually required.

2.10 Aim

The primary aim of this study is to generate a reproducible method that allows prediction of drug off-targets. The method proposed has been applied to predict the potential off-targets of DIMATE, a potent inhibitor of ALDH3A1 and proliferation of cancer cells. In order to achieve this, similarity based computational predictions and pharmacophore

based methods were used. In the similarity based methods, targets with binding pockets similar to that of ALDH3 are analyzed. Moreover, ligands with similar structure to DIMATE are identified, and their targets are evaluated as potential unintended targets of DIMATE. The feasibility of the predictions are discussed in detail.

Chapter 3

MATERIALS AND METHODS

3.1 Tools Used

3.1.1 BLAST

Basic Local Alignment Search Tool (BLAST) is a freely available tool that allows comparison of biological sequences at the DNA or protein level [101]. It can either search for sequence similarity of a query sequence within a specified library of sequences, or can align two or more sequences specified by the user. BLAST can be accessed at http://blast.ncbi.nlm.nih.gov/Blast.cgi. In this study, protein BLAST was used to compare sequence similarity of aldehyde dehydrogenases of different species [102].

3.1.2 SWISS-MODEL

To predict the structural similarity of the sheep aldehyde dehydrogenase with the human aldehyde dehydrogenase, homology modeling was used. SWISS-MODEL is a comparative modeling tool that allows the prediction of a protein structure based on its amino acid sequence [103-106]. The availability of a potential template, which is similar in sequence to the target, eases the process of structure prediction. The template protein with known structure is used to propose a potential crystal structure of the target. SWISS-MODEL can be accessed freely at http://swissmodel.expasy.org/.

3.1.3 Splitpocket

Splitpocket is a freely available server that allows the extraction of functional protein surfaces [107]. The idea relies on the alpha shape theory. The algorithm identifies the geometrical information of a split pocket, that is, the binding surface of a ligand., The input of the server is the simple structural data in the form of a PDB file. The output allows a dynamic visualization of the functional surface of the input protein. The server is fully automatic, and allows obtaining biological insights into the structure of input

proteins, and hence, their functions. The server can be found at http://pocket.uchicago.edu/.

3.1.4 ReliBase

ReliBase is an easy to use tool to analyze ligand-protein data [108]. It can therefore be used a s a data mining tool to obtain structural data on the particular interacting surfaces of proteins, aiding the process of drug discovery. In this study, a SMILE based search was used to search ligand substructures. The algorithm of the ReliBase computes a 2D similarity index that is based on Tanimoto coefficient [108]. The index is compared against a library of 2D similarity indices available in their database. This server can be accessed from http://relibase.ccdc.cam.ac.uk.

3.1.5 ProBis

Protein Binding Sites (ProBis) Tool is a tool that enables detection of similar binding sites in protein structures [109]. It allows local pairwise alignment of structures obtained by crystallography or NMR in the Protein Data Bank. ProBis takes PDB IDs or files as input, and compares the query structure to a non-redundant database of proteins (non-redundant PDB) through the maximum clique algorithm. The maximum clique algorithm makes use of protein structures represented as protein graphs, and returns the largest similar subgraphs among the compared graphs [110]. ProBis then calculates the degree of structure conservation for each of the amino acids in the specified protein. The web server outputs these data in a protein structure where different colors represent the degree of conservation.

ProBis has important applications in drug off-target prediction. Compared to algorithms that compare protein sequence or folding, ProBis can more effectively address the question of off-target prediction. This is because protein binding sites, rather than overall fold of the protein, determine the protein's interaction with ligands. ProBis can be accessed from http://probis.cmm.ki.si/.

3.1.6 Autodock Vina

Autodock Vina is a program for molecular docking [111]. It is an improved version jof AutoDock 4, a previously created molecular docking tool by the same group. The program uses PDBQT molecular structure files as input. In this program, specific chains can be selected to assign flexibility to a specific target. Autodock Vina can be found at http://vina.scripps.edu.

3.1.7 Gold

Genetic Optimisation for Ligand Docking (GOLD) is a molecular docking program that makes use of a genetic algorithm to predict the binding and conformational space of a ligand and its prospective target [112]. It uses pharmacophore (fitting) points to match a ligand to a potential target. It makes use of a coring function to rank probable poses of a ligand and its target. An important aspect of GOLD is that it supports flexible side-chains which allows modeling flexibility to indicated side-chains. It is also possible to perform ensemble docking with GOLD when more than one structural file is present for a protein.

3.1.8 Molegro Virtual Docker

Molegro Virtual Docker (MVD) is a molecular docking software [113]. It is used to predict the plausibility of indicated ligand-protein interactions. It uses a new heuristic search algorithm , which uses differential evolution along with a cavity prediction. MVD has a re-ranking scoring function, as well as a docking scoring function to improve accuracy. It has been demonstrated that Molegro is superior to other docking tools in terms of correct identification of binding modes, with an accuracy rate of 87% [113].

3.1.9 PharmMapper

PharmMapper is a free server that is used to search for possible targets for a set of molecules [114]. It uses a pharmacophore based modeling approach to predict binding of small molecule to its potential targets. It is a relatively fast server, allowing the

identification of possible targets from a set of candidates provided by the database in a few hours. The server can be accessed at http://59.78.96.61/pharmmapper/.

3.1.10 LigandScout

LigandScout is a software that models three dimensional 3D pharmacophore models from available structural data of a molecule with its ligand [115]. It makes use of all of the chemical feature data available for the molecules, generating a model of interaction between a ligand and its target. LigandScout allows for the overlaying and comparison of different pharmacophore models through an alignment algorithm. This way, a shared-feature pharmacophore can be generated.

3.2 Steps of Off-Target Prediction

Three separate approaches were used to predict potential off-targets of ALDH based on ligand similarity, binding pocket similarity and pharmacophore models. This strategy can be extended to virtually all drug candidates as outlined in Figure 3.1.

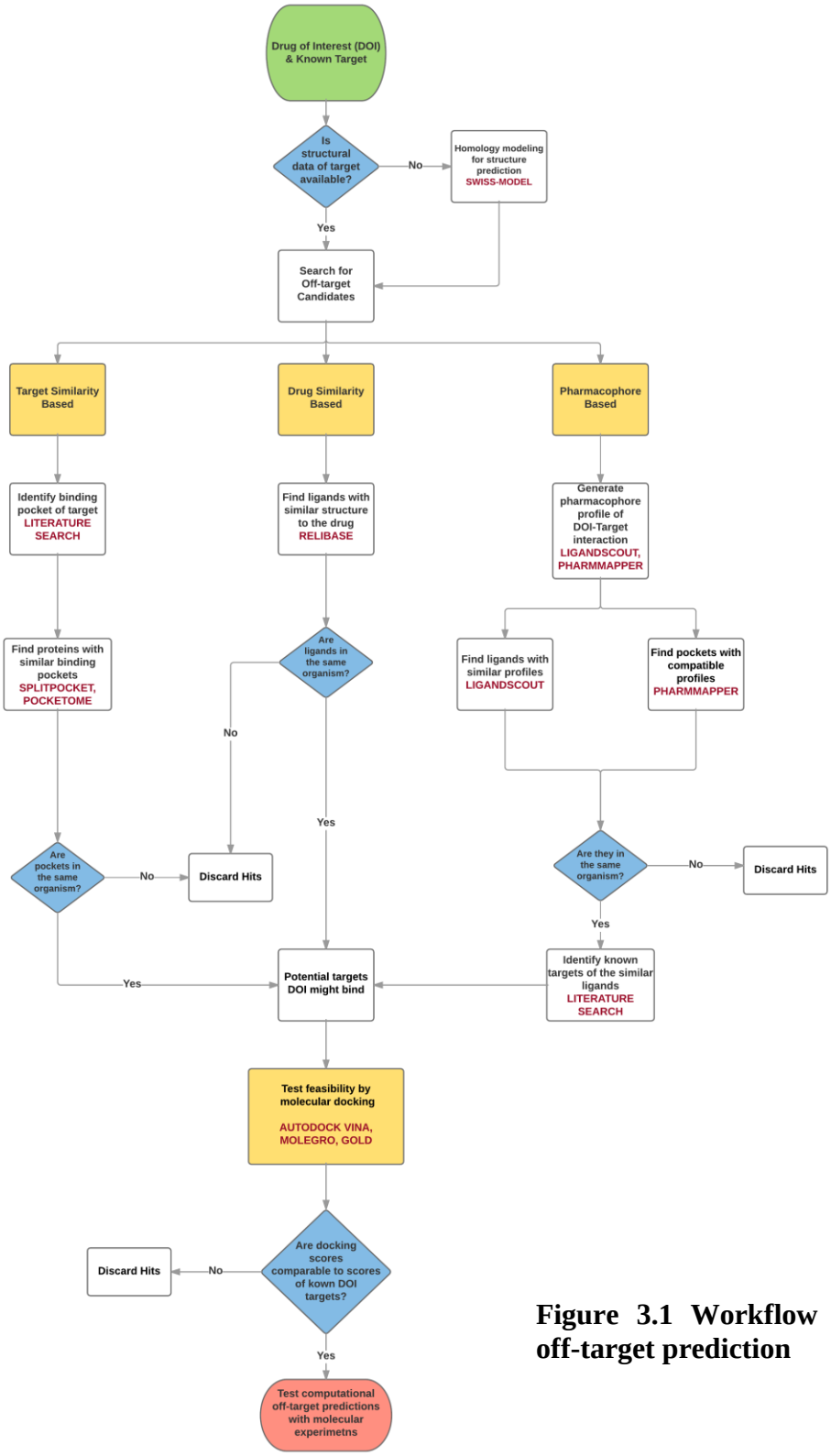


Figure 3.1 Workflow of drug off-target prediction

3.2.1 Similarity Based Approaches

In drug off-target prediction is through similarity, candidate factors were identified through two separate methods. In the first method, the binding pockets of the target were identified, and proteins that have similar binding pockets were found. To select the binding pocket of ALDH, previous data on ALDH active site was used. This region includes the critical Cys302 residue. Considering NAD is also an important cofactor in ALDH function, NAD binding region was also included in the binding pocket selection. Molegro was used to identify five potential cavities of ALDH. The top choice included the NAD binding region. In all following models, this cavity was specifically selected and centralized with a sphere radius of 16 Å. In this method, the proteins that have similar binding pockets to that of ALDH may possibly be targets of the drug of interest, since they also contain a similar binding pocket (Figure 3.2A). In the second method, ligands that are similar in structure to the drug of interest were found. The drug of interest, because it is structurally similar to the identified set of ligands, may also bind to the readily known targets of the similar ligands (Figure 3.2B). These possibilities determined by the two methods were computationally tested for their feasibility using molecular docking. The overall workflow of off-target predication using this similarity based approach is illustrated in Figure 3.3.

3.2.2 Pharmacophore Based Approaches

For pharmacophore model generation, PharmMapper and LigandScout were used to create pharmacophore models of ALDH and DIMATE interaction. These models were then used to find other interactions that are of similar nature. This method, compared to the previous similarity-based method, makes use of feature points of the pharmacophore model instead of chemical structural data.

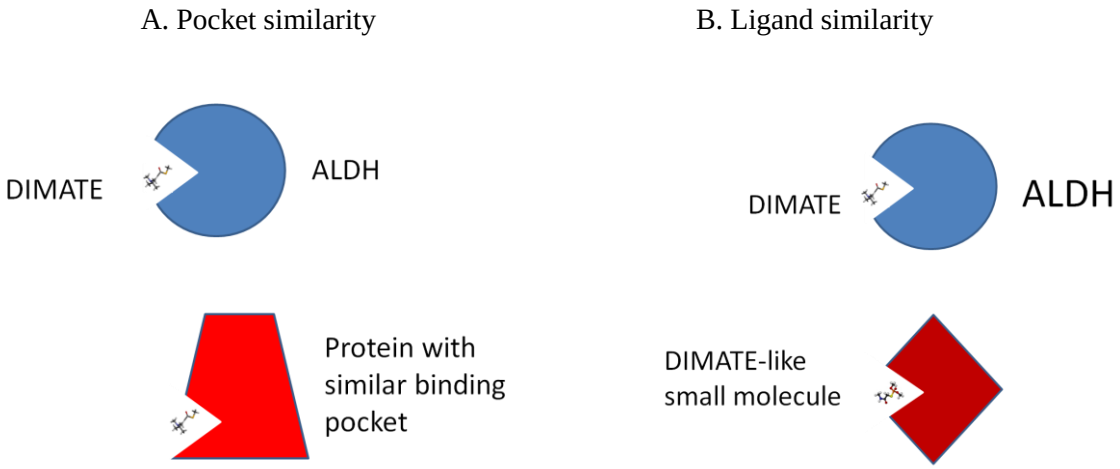


Figure 3.2 Comparison of pocket similarity and ligand similarity approach
(A) In the pocket similarity approach, the binding pocket of DIMATE in ALDH is identified, and proteins with similar binding pockets are considered as potential off-targets of DIMATE. (B) In the ligand similarity approach, ligands that are similar in structure to DIMATE are identified, and the known targets of those ligands are considered as potential off-targets of DIMATE.

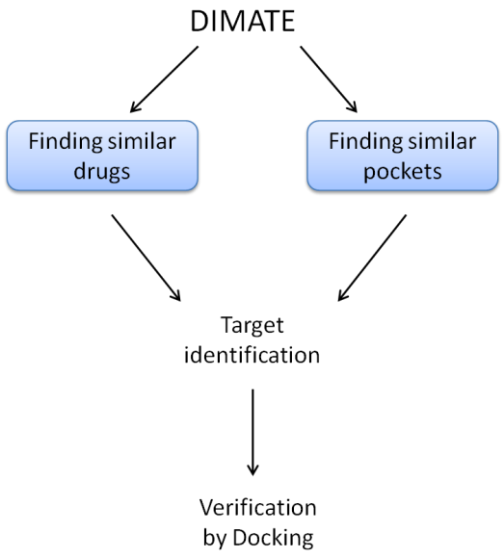


Figure 3.3 Workflow of computational off-target prediction for DIMATE by similarity based approaches
To identify possible off-targets of DIMATE, two approaches of drug similarity and pocket similarity are used. The identified targets are subsequently verified by docking.

Chapter 4

RESULTS

In this section, two different methods are employed to identify the possible off-targets of DIMATE, a novel small molecule inhibitor of ALDH3A1. In the first method, a similarity based approach is used. This method relies on the structural similarity principles to identify ligands similar to DIMATE, and targets similar to ALDH to predict off-targets. In the second method, a pharmacophore based approach is used. Pharmacophores are lists of molecular and geometrical features of a protein and its associated ligand that outline the requirements of a biological interaction.

4.1 Homology modeling to assess validity of the approach

The absence of a crystal structure of human ALDH3 from the Protein Data Bank (accessed Jan 2014) without an inhibitor limits the studies of its interactions with its partners. To investigate the potential off-targets of ALDH, we made use of rat ALDH3, whose crystal structure was readily available in the PDB. The amino acid sequence of human and rat ALDH3 are remarkably similar, sharing 82% sequence identity as determined by protein BLAST (Figure 4.1). Moreover, considering conservation of residues in terms of structure and chemical properties, 91% of amino acids are similar.

Similarity of human and rat ALDH3 proteins enabled prediction of human ALDH3 structure by homology modeling. SWISS-MODEL was used to generate a model of the protein, using the rat enzyme as a template (Figure 4.2). As expected from the high percentage of sequence similarity, the human and rat ALDH3 were predicted to be structurally similar. The quality of the model as estimated by Global Model Quality Estimation (GMQE) is considerably high, with a value of 0.97. The GMQE score is a measure of the reliability of the model for a particular template and alignment, and has a

value between 0 and 1, with higher values indicating a higher quality of the model. Moreover, estimating the quality of the model at a local level by QMEAN scoring function [116], each chain of the target was shown to be highly similar to that of the template (Figure 4.2A). With this model, the human enzyme is predicted to also function as a homo-dimer. The structures of human and rat aldehyde dehydrogenase enzymes are provided in Figure 4.2B.

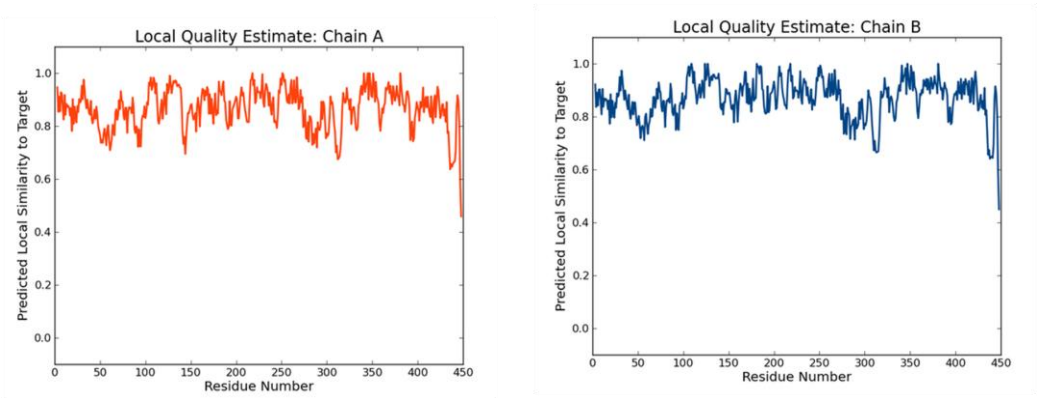
Range 1: 1 to 453 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
794 bits(2050)	0.0	Compositional matrix adjust.	370/453(82%)	416/453(91%)	0/453(0%)

```
Query  1    MSKISEAVKRARAAFSSGRTRPLQFRIQQLEALQRLIQEQEQELVGALAADLHKNEWNAY  60
            MS IS+ VKRAR AF+SG+TR LQFRIQQLEALQR+I E  + + GALA+DL KNEW +Y
Sbjct  1    MSSISDTVKRAREAFNSGKTRSLQFRIQQLEALQRMINENLKSISGALASDLGKNEWTSY  60

Query  61   YEEVVYVLEEIEYMIQKLPEWAADEPVEKTPQTQQDELYIHSEPLGVVLVIGTWNYPFNL  120
            YEEV +VLEE++  I++LP+WA DEPV KT QTQQD+LYIHSEPLGVVLVIG WNYPFNL
Sbjct  61   YEEVAHVLEELDTTIKELPDWAEDEPVAKTRQTQQDDLYIHSEPLGVVLVIGAWNYPFNL  120

Query  121  TIQPMVGAIAAGNSVVLKPSELSENMASLLATIIPQYLDKDLYPVINGGVPETTELLKER  180
            TIQPMVGA+AAGN+V+LKPSE+S +MA LLAT+IPQY+D++LY V+ GGVPETTELLKER
Sbjct  121  TIQPMVGAVAAGNAVILKPSEVSGHMADLLATLIPQYMDQNLYLVVKGGVPETTELLKER  180

Query  181  FDHILYTGSTGVGKIIMTAAAKHLTPVTLELGGKSPCYVDKNCDLDVACRRIAWGKFMNS  240
            FDHI+YTGST VGKI+M AAAKHLTPVTLELGGKSPCYVDK+CDLDVACRRIAWGKFMNS
Sbjct  181  FDHIMYTGSTAVGKIVMAAAAKHLTPVTLELGGKSPCYVDKDCDLDVACRRIAWGKFMNS  240

Query  241  GQTCVAPDYILCDPSIQNQIVEKLKKSLKEFYGEDAKKSRDYGRIISARHFQRVMGLIEG  300
            GQTCVAPDYILCDPSIQNQIVEKLKKSLK+FYGEDAK+SRDYGRII+ RHFQRV GLI+
Sbjct  241  GQTCVAPDYILCDPSIQNQIVEKLKKSLKDFYGEDAKQSRDYGRIINDRHFQRVKGLIDN  300

Query  301  QKVAYGGTGDAATRYIAPTILTDVDPQSPVMQEEIFGPVLPIVCVRSLEEAIQFINQREK  360
            QKVA+GGT D ++RYIAPTIL DVDPQSPVMQEEIFGPV+PIVCVRSLEEAIQFINQREK
Sbjct  301  QKVAHGGTWDQSSRYIAPTILVDVDPQSPVMQEEIFGPVMPIVCVRSLEEAIQFINQREK  360

Query  361  PLALYMFSSNDKVIKKMIAETSSGGVAANDVIVHITLHSLPFGGVGNSGMGSYHGKKSFE  420
            PLALY+FS+N+KVIKKMIAETSSGGV ANDVIVHIT+ +LPFGGVGNSGMG+YHGKKSFE
Sbjct  361  PLALYVFSNNEKVIKKMIAETSSGGVTANDVIVHITVPTLPFGGVGNSGMGAYHGKKSFE  420

Query  421  TFSHRRSCLVRPLMNDEGLKVRYPPSPAKMTQH  453
            TFSHRRSCLV+ L+N+E  K RYPPSPAKM +H
Sbjct  421  TFSHRRSCLVKSLLNEEAHKARYPPSPAKMPRH  453
```

Figure 4.1 Alignment of human and rat ALDH3 proteins
Protein BLAST was used to align human (UniProtKB ID: P30838) and rat (UniProtKB ID: P11883) ALDH3 enzymes.

A



B

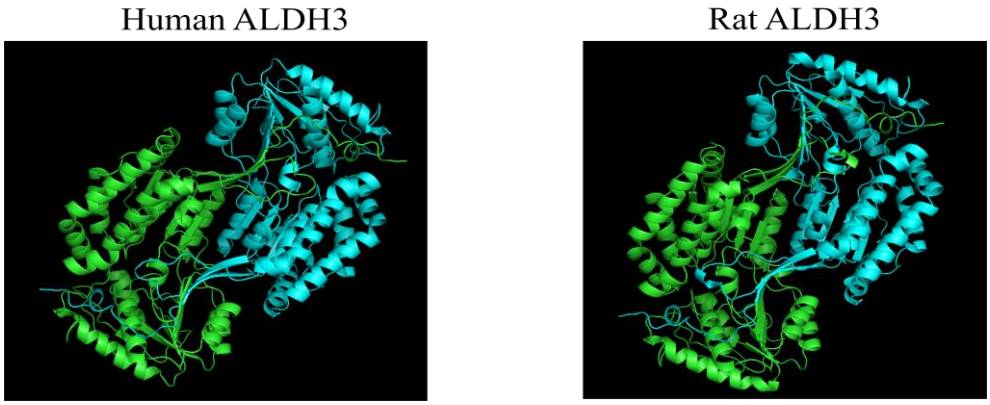


Figure 4.2 Homology modeling of human and rat ALDH3 proteins
(A) Local quality of models analyzed at the chain level. (B) Predicted structure of human ALDH3 (left) and the known crystal structure of rat ALDH3 (right).

4.2 Computational off-target prediction of DIMATE using similarity approach

4.2.1 Similar pocket approach

Using the pocket similarity approach, we identified proteins in the Protein Data Bank that have binding pockets similar to the binding pocket of ALDH3A1 using Splitpocket. The results are provided in Table 4.1. This approach did not turn out to be fruitful, because many of the proteins with similar binding pockets were found to be other aldehyde dehydrogenase proteins. The only three hits that were not from the ALDH family were Guanine nucleotide exchange factor DBS, ADP Ribosyl Cyclase and Nitrile hydratase (subunit alpha). However, the organism to which these proteins belong were not H. *sapiens*, but were the mouse *M. musculus*, the pipevine plant *A. californica* and the bacteria strain *Rodococchus sp.* Therefore, they could not be possible off-targets of DIMATE found in humans.

4.2.2 Similar ligand approach

We next adopted a ligand similarity approach to identify possible off-targets of DIMATE. We used ReliBase to find ligands that are similar in structure to DIMATE. The results of this approach are provided in Table 4.2. Interestingly, DIMATE was found to be structurally similar to carnitine and d-carnitine in many species, including humans, mouse, and bacteria. This indicates that DIMATE, because it is structurally similar to carnitine, may also bind to carnitine acetyltransferases, especially carnitine acetyltransferase isoform 2 in humans, as suggested by the results of ReliBase. Other important results for that were found in humans were 2-(butanoylsulfanyl)-N,N,N-trimethylethanaminium (Butyrlthiocholine), 4-(dimethylamino)butyl imidothiocarbamate, 2-methyl-3-(2-Aminothiazolo)propanal, tetradecanoyl-CoA, and N[4-[(R)-methylsulfinyl]butylthioformamide, making possible targets of DIMATE butyrylcholinesterase, Histamine N-methyltransferase, rennin, Acyl-CoA-binding protein, and Macrophage migration inhibitory factor, respectively (Table 4.2).

Table 4.1 Binding pocket similarity approach results for ALDH using Splitpocket

PDB ID	Protein	Gene name	Organism
1O01	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
2ONM	Aldehyde dehydrogenase, mitochondrial	ALDH2A	*H. sapiens*
1NZW	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
1NZZ	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
1ZUM	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
2ONN	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
2ONO	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
1CW3	Mitochondrial Aldehyde Dehydrogenase	ALDH2A	*H. sapiens*
1O02	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
1O05	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
2ONP	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
2VLE	Aldehyde dehydrogenase, mitochondrial	ALDH2A	*H. sapiens*
1NZX	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
1O00	Aldehyde dehydrogenase	ALDH2A	*H. sapiens*
1A4S	Betaine aldeyhde dehydrogenase from cod liver	ALDH9A1	*G. morhua*
1BXS	Sheep liver class 1 aldehyde dehydrogenase, NAD bound	ALDH1A1	*O. Aries*
2O2P	Formyltetrahydrofolate dehydrogenase	ALDH1L1 FTHFD	*R. norvegicus*
1O04	Aldehyde dehydrogenase, mitochondrial	ALDH2A	*H. sapiens*
1RJ2	Guanine nucleotide exchange factor DBS [Fragment]	Mcf2l Dbs	*M. musculus*
1LBE	ADP Ribosyl Cyclase		*A. californica*
1R15	ADP Ribosyl Cyclase		*A.californica*
1AHJ	Nitrile hydratase (subunit alpha)	nthA	*Rodococchus sp.*

Table 4.2 Ligand similarity approach results for DIMATE using ReliBase

Ligand	Ligand Name	PDB ID	Protein
BCH	2-(BUTYRYLSULFANYL)-N,N,N-TRIMETHYLETHANAMINIUM	2ha7	Acetylcholinesterase
		1p0p	butyrylcholinesterase
2MM	N,N-dimethyl-L-methionine	3cjq	Ribosomal protein L11 methyltransferase
		3cjt	Ribosomal protein L11 methyltransferase
3MM	(1R)-1-CARBOXY-N,N,N-TRIMETHYL-3-(METHYLSULFANYL)PROPAN-1-AMINIUM	3cju	Ribosomal protein L11 methyltransferase
4MM	(1S)-1-carboxy-N,N,N-trimethyl-3-(methylsulfanyl)propan- 1-aminium	3egv	Ribosomal protein L11 methyltransferase
HII	2-METHYL-3-(2-AMINOTHIAZOLO)PROPANAL	1bil	Renin
		1bim	Renin
MSN	(1R,2R,3R,4S,5R)-4-AMINO-5-(METHYLTHIO)CYCLOPENTANE- 1,2,3-TRIOL	3dx0	Alpha-mannosidase 2
		2f7o	Alpha-mannosidase 3
4DI	4-(DIMETHYLAMINO)BUTYL IMIDOTHIOCARBAMATE	2aov	Histamine N-methyltransferase
KBP	S-[2-(acetylamino)ethyl] (3R)-3-hydroxydecanethioate	4b0i	3-HYDROXYDECANOYL-[ACYL-CARRIER-PROTEIN] DEHYDRATASE
NM2	3-CARBOXY-N,N,N-TRIMETHYLPROPAN-	2wsx	CaiT carnitine transporter

	1-AMINIUM		
		3o2g	Gamma-butyrobetaine dioxygenase
152	CARNITINE	1t7q	Carnitine acetyltransferase
		2h3p	Carnitine acetyltransferase
		2h3u	Carnitine acetyltransferase
		1ndf	Carnitine acetyltransferase
		1t7o	Carnitine acetyltransferase
		1xl8	Peroxisomal carnitine O-octanoyltransferase
		1ndf	Carnitine acetyltransferase
		1s5o	carnitine acetyltransferase isoform 2

4.2.3 Reverse Docking

We next used a reverse docking approach to evaluate the affinity with which the inhibitor DIMATE binds to ALDH. Reverse docking is the docking of a ligand to the potential binding site of a target [117]. It helps uncover potential uses for a ligand other that its original intent, meaning drug repositioning. However, compared to normal molecular docking, it is more difficult. Normal docking involves docking of many ligands to one target with a known or predicted binding site. Conversely, off-target prediction involves docking one ligand to many targets.

Before applying the reverse docking methodology to identify off-targets of DIMATE, the method itself should first be validated with a negative control, such as MATE that is not an inhibitor of ALDH1, a positive control such as disulfiram, an established inhibitor of ALDH2, a positive control, such as a chlorpropamide analogue

that is an inhibitor of ALDH3, and a positive control such as benzaldehyde that is a substrate of ALDH3. We therefore tested the reverse docking scores obtained by different reverse docking programs: Autodock Vina, Molegro and Gold. These results are provided in Table 4.3, Table 4.4 and Table 4.5, respectively.

Table 4.3 Reverse docking results with Autodock Vina-Molegro

Ligand – Protein	Experimental Binding	without NAD	with NAD	
		AutoDock (kcal/mol)	M Rerank S (4000 iters, 50 runs)	M MolDock S (4000 iters, 50 runs)
MATE – ALDH1	- [118]	-4,9	-56,421	-75,350
Disulfiram - ALDH2	+ [42]	-4,1	-56,427	-77,505
API-2 - ALDH3	+ [19]	-6,7	-88,728	-97,853
Benzaldehyde – ALDH3	+ [119]	-5,7	-52,364	-60,884

Table 4.4 Reverse docking results with Molegro using varying iterations and runs

Ligand - Protein	Experimental Binding	without NAD		with NAD	
		M Rerank S (4000 iters)	M MolDock S (4000 iters)	M Rerank S (4000 iters)	M MolDock S (4000 iters)
MATE – ALDH1	- [118]	-66,1653	-81,7077	-56,421	-75,350
Disulfiram - ALDH2	+ [42]	-83,9569	-102,058	-56,427	-77,505
API-2 - ALDH3	+ [19]	-84,6247	-91,601	-88,728	-97,853
Benzaldehyde – ALDH3	+ [119]	-50,6291	-58,4316	-52,364	-60,884

Table 4.5 Reverse docking results with Gold

Ligand – Protein	Experimental Binding	GOLD (PLP Score)	PLP Fitness
MATE – ALDH1	- [118]	-34,714	34,339
Disulfiram - ALDH2	+ [42]	40,384	-42,615
API-2 - ALDH3	+ [19]	-40,587	42,818
Benzaldehyde – ALDH3	+ [119]	-29,926	35,069

In order to utilize reverse docking, we made use of a collective list of proteins with similar binding pockets to that of ALDH3A1. This list was generated using pocket similarity search tools available online, such as Pocketome, ProBis, and Splitpocket. We reverse docked DIMATE to proteins that have similar binding pockets to ALDH3A1, such as other ALDH enzymes. As for controls to the reverse docking process, we also reverse docked all-trans retinal, a natural substrate of ALDH, and Aldi-1, a previously established ALDH inhibitor. Figure 4.3 provides a flowchart outlining the basic steps of reverse docking process. The results of this approach are provided in Table 4.6.

Looking at the results provided in Table 4.6, the docking scores for DIMATE should be compared to that of all-trans retinal and Aldi-1, which are both known to bind to ALDH. Negative docking scores (in kcal) indicate the feasibility of the docking process. As seen from the docking scores of DIMATE to other ALDH proteins, the binding affinity of DIMATE is in fact comparable to natural substrates of ALDH. This is indicative of the possibility that DIMATE may not only bind to ALDH3A1, but also other ALDH molecules.

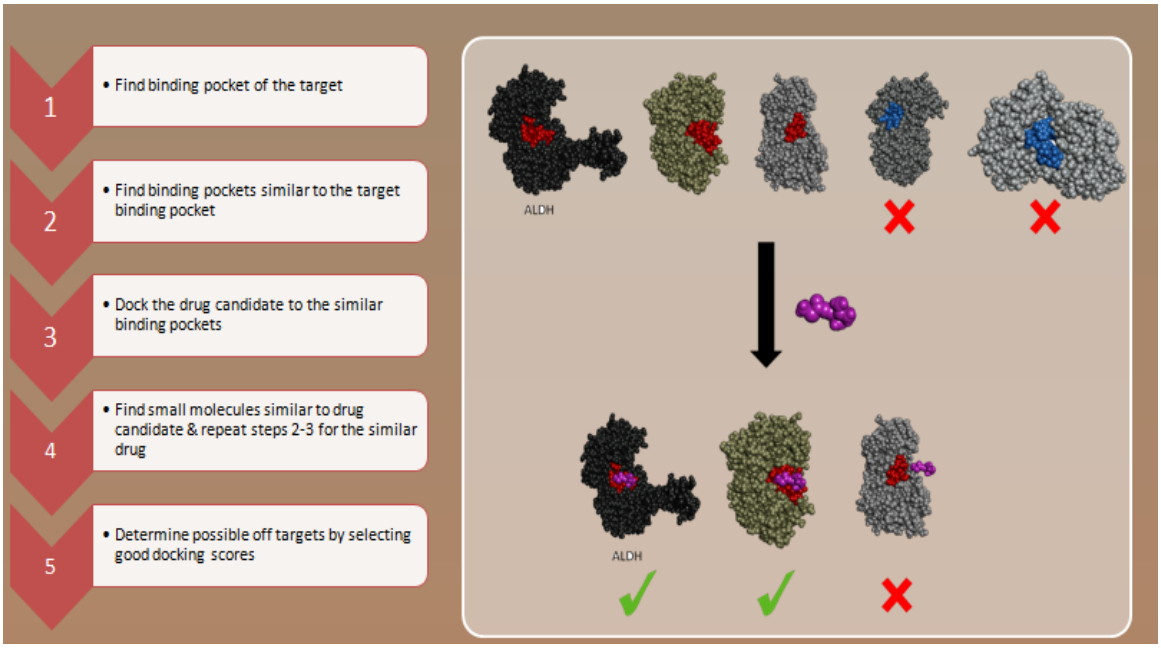


Figure 4.3 Overview of off-target prediction by reverse docking

Table 4.6 Docking scores of different ligands to ALDH structures
Docking scores are given in kcal. (All-trans retinal, a natural substrate of ALDH and Aldi-1, synthetic inhibitor of ALDH)

Gene name	PDB ID	DIMATE	ATE	All-trans retinal	ALDI-1
ALDH1A1	1BXS	-5.4	-4.6	-8.5	-5.8
ALDH1L1 FTHFD	3RHP	-5.5	-5.1	-8.0	-6.3
ALDH1L1 FTHFD	3RHQ	-5.6	-4.4	-7.9	-6.1
ALDH1L1 FTHFD	4GNZ	-5.7	-4.7	-7.9	-6.4
ALDH1L1 FTHFD	3RHJ	-5.3	-4.5	-8.2	-6.3
ALDH1L1 FTHFD	3RHL	-5.5	-4.5	-8.2	-6.4
ALDH1L1 FTHFD	2O2R	-5.6	-4.7	-8.0	-6.3
ALDH1L1 FTHFD	3RHR	-5.6	-4.6	-8.1	-6.2
ALDH1L1 FTHFD	4GO0	-5.1	-4.7	-7.3	-5.9
ALDH1L1 FTHFD	3RHM	-5.5	-5.0	-8.5	-6.0
ALDH1L1 FTHFD	4GO2	-5.0	-4.1	-8.1	-6.4
ALDH1L1 FTHFD	2O2P	-5.5	-5.0	-8.4	-6.3
ALDH1L1 FTHFD	2O2Q	-5.3	-4.1	-8.2	-6.3
ALDH2	1AG8	-5.8	-3.7	-8.1	-5.8
ALDH2	3SZ9	-5.5	-4.9	-7.6	-5.6
ALDH2	4FQF	-6.0	-4.9	-7.8	-4.7
ALDH2	4FR8	-6.1	-4.5	-7.6	-4.8
ALDH2A	2VLE	-6.2	-4.1	-8.0	-5.8
ALDH2A	1O04	-6.2	-4.8	-7.7	-6.7
ALDH2A	1O00	-5.6	-4.3	-7.7	-5.6
ALDH2A	3INJ	-6.2	-4.9	-7.8	-5.8
ALDH2A	2ONN	-5.7	-4.6	-7.7	-6.1
ALDH2A	1O05	-6.0	-4.6	-7.7	-6.0

ALDH2A	1NZZ	-6.1	-4.7	-7.5	-5.3
ALDH2A	3N82	-6.0	-4.3	-7.6	-4.9
ALDH2A	1O02	-6.1	-4.8	-8.3	-7.1
ALDH2A	2ONO	-5.5	-4.5	-7.9	-6.5
ALDH2A	1O01	-6.0	-4.8	-7.7	-5.4
ALDH2A	1ZUM	-5.7	-4.6	-7.1	-6.1
ALDH2A	3INL	-5.9	-5.0	-8.0	-6.0
ALDH2A	1NZX	-6.0	-4.9	-7.7	-6.0
ALDH2A	2ONP	-6.0	-4.7	-7.3	-6.7
ALDH2A	2ONM	-5.6	-4.9	-7.2	-5.7
ALDH2A	1NZW	-5.8	-4.9	-7.7	-5.4
ALDH2A	1CW3	-5.9	-4.8	-8.0	-5.9
ALDH3A	1AD3	-4.6	-4.6	-8.3	-6.2
ALDH3A	3SZA	-5.1	-5.0	-9.5	-6.7
ALDH3A	3SZB	-5.1	-4.5	-7.4	-6.2
ALDH3A1	4H80	-4.7	-3.9	-7.2	-6.0
ALDH3A1	4L2O	-4.8	-5.1	-7.1	-5.3
ALDH3A1	4H80	-4.7	-3.9	-7.2	-6.0
ALDH3A1	4L2O	-4.8	-5.1	-7.1	-5.3

4.3 Computational off-target prediction of DIMATE using pharmacophore approach

A different approach to identifying drug off-targets is through the use of pharmacophore models. Pharmacophore models list the chemical and geometrical features of a particular interaction between a protein and a ligand. Similar models for the possible off-targets along with the ligand of interest are constructed, and their similarity to the original pharmacophore model is compared. A high similarity indicated by z-scores shows a better probability of the ligand to bind the listed off-targets. In this section, we used PharmMapper to search for proteins that are similar pharmacophore profiles to DIMATE-ALDH interaction. The results are provided in Table 4.7. The top hits are displayed according to their z-score rankings. Using this approach, thymidylate kinase, ADP-ribosyl cyclase 2, and cytoplasmic tryptophanyl-tRNA synthetase appear to be possible off-targets of DIMATE.

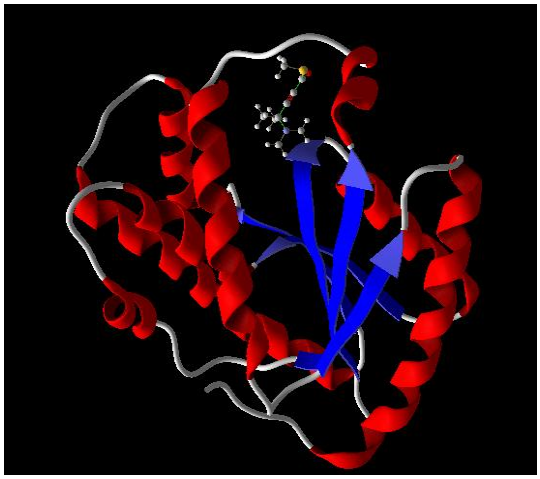
4.3.1 Analysis of Thymidylate Kinase as a Hit

To assess the validity of thymidylate kinase as a hit, DIMATE was docked to thymidylate kinase. Figure 4.4 shows the docking results of DIMATE to the ADP binding site of thymidylate kinase, which is the active site of the protein. However, lower energy scores could be obtained upon docking to FDM binding site (Figure 4.5).

Table 4.7 Proteins with similar pharmacophore profiles to DIMATE-ALDH as found by PharmMapper

Rank	Target Name	Number of Feature	Fit Score	Normalized Fit Score	z'-score
1	Thymidylate kinase	8	2.965	0.3706	104.339
2	ADP-ribosyl cyclase 2	9	2.943	0.3270	136.623
3	Tryptophanyl-tRNA synthetase, cytoplasmic	11	2.943	0.2675	0.403523
4	Glucose-6-phosphate 1-dehydrogenase	9	2.905	0.3228	0.880866
5	Signal transducer and activator of transcription 1-alpha/beta	8	2.880	0.3600	0.828366
6	Scavenger mRNA-decapping enzyme DcpS	11	2.850	0.2591	-0.0581112
7	Insulin-like growth factor IA	6	2.828	0.4714	0.531979
8	Eukaryotic translation initiation factor 4E	12	2.827	0.2356	-0.273193
9	Thymidylate kinase	9	2.801	0.3113	-0.145247
10	Thymidylate kinase	9	2.795	0.3105	-0.296493
11	Thymidylate kinase	10	2.692	0.2692	-101.914
12	Pancreatic alpha-amylase	9	2.597	0.2886	-124.964
13	Estrogen receptor beta	9	2.449	0.2721	-208.542
14	Lysozyme C	8	2.421	0.3027	-248.839

(A)



INTERACTION	POSE	MOLDOCK SCORE	RERANK SCORE
Thymidylate kinase-DIMATE	0	-74,403	-60,341

(B)

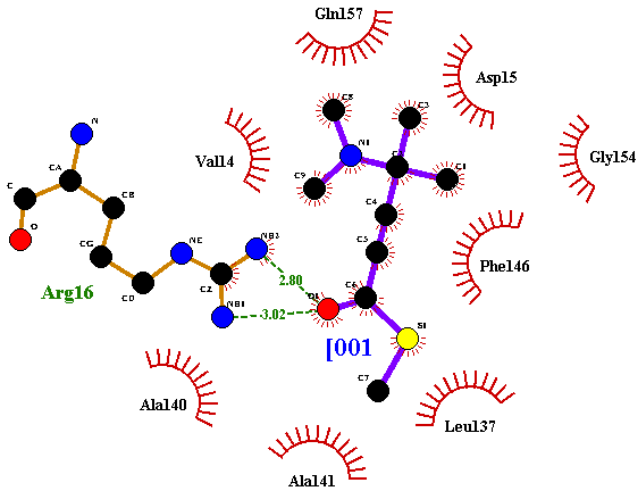
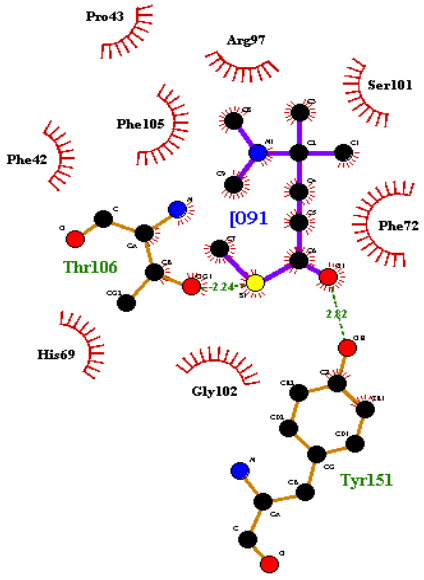


Figure 4.4 DIMATE docked to thymidylate kinase's active site
(A) DIMATE was docked to the ADP binding site of thymidylate kinase (PDB ID:1NMX) (B) Putative interactions of DIMATE with thymidylate kinase active site

(A)

INTERACTION	POSE	MOLDOCK SCORE	RERANK SCORE
Thymidylate kinase-DIMATE	9	-83,216	-73,482

(B)



1nmx

Figure 4.5 DIMATE docked to thymidylate kinase's other ligand binding site
(A) DIMATE was docked to the FDM binding site of thymidylate kinase (PDB ID:1NMX) (B) Putative interactions of DIMATE with thymidylate kinase

Chapter 5

DISCUSSION

5.1 Conclusions

In this study, our aim was to predict the potential off-targets of ALDH inhibitor DIMATE. DIMATE is a small molecule that has been shown to inhibit ALDH3A1, the aldehyde dehydrogenase enzyme that functions in drug resistance and affects cell proliferation. We employed two different approaches to predict other potential targets of DIMATE using computational approaches. In the first approach based on target binding pocket and ligand similarity, we identified carnitine acetyltransferase, histamine N-methyltransferase, rennin, Acyl-CoA-binding protein, and macrophage migration inhibitory factor as possible off-targets of DIMATE. Our reverse docking results as verification of these predictions suggested that DIMATE may also bind to other ALDH molecules to inhibit them. In the second method based on pharmacophore structures, potential off-targets of DIMATE were found to be thymidylate kinase, ADP-ribosyl cyclase 2, and cytoplasmic tryptophanyl-tRNA synthetase.

One of the more interesting potential DIMATE targets is thymidylate kinase. This enzyme is responsible for the phosphorylation of thymidine 5'-phosphate (dTMP) for the production of thymidine 5'-diphosphate [120]. It is integral to dTTP synthesis pathway, a component of pyrimidine metabolism. Interestingly, thymidylate kinase was implicated in cell cycle progression. When the human enzyme was expressed in yeast cells, it was able to functionally complement a cell cycle mutant, *cdc8* (the yeast homolog of dTMP kinase) [121]. This study suggests that the enzyme activity of human thymidylate kinase corresponds to the progression of cell cycle, with the highest gene expression levels observed in the S phase of the cell cycle. Possibly, an interaction of DIMATE with thymidylate kinase could inhibit this enzyme to result in a cell cycle arrest and stop proliferation. Such an event is plausible, because the yeast homolog is essential for DNA

replication during cell division. Therefore, the inhibition of thymidylate kinase with DIMATE could potentially be halting cell proliferation. In such an inhibition, the overall effects observed would be similar to the case where ALDH3 is targeted. However, one problem might be that although high ALDH3 expression is commonly seen in cancerous and stem cells, thymidylate kinase is expressed in virtually all cells, meaning that growth inhibition would not be specific to cancer cells. This would be a major drawback, eliminating the specificity of DIMATE treatment. Hence, it is of great importance to investigate thymidylate kinase as a potential off-target.

Another interesting potential off-target of DIMATE is tryptophanyl-tRNA synthetase. This enzyme is responsible from the aminoacylation of tRNA(trp) by tryptophan. It is induced by interferon [122]. A critical feature of the cytoplasmic form of tryptophanyl-tRNA synthetase is its implication in the negative regulation of proliferation [123]. This characteristic implies that inhibition of this enzyme by DIMATE binding may in fact increase cell proliferation, an undesirable effect for the potential use of DIMATE. However, this interaction is probably unlikely, because it was already previously shown that treatment of cancer cells with DIMATE inhibits cell proliferation [67, 76, 77].

In these types of computational approaches, it is crucial to consider the biological context at which these interactions could occur. For example, many of the possible targets of DIMATE as found by binding pocket similarity tools list proteins that are of other species. These are immediately discarded, since these two proteins would not encounter each other within the same cell. Another example could be that a computational prediction might suggest some protein to be bound by DIMATE, but these two molecules may in fact never be within the same subcellular compartment. This would prevent any interaction, but such false positives would still show up in the list. Therefore, it is important to analyze each hit one by one to more accurately predict the potential off-targets.

5.2 Future Directions

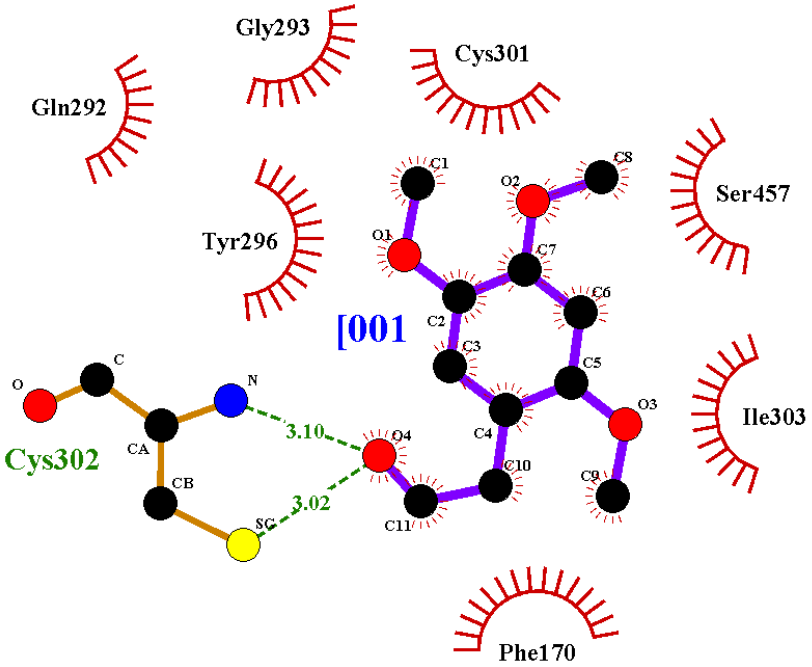
Although computational tools are invaluable to shorten the list of possible off-target interactions of a ligand, ultimately experimental approaches have to be employed to definitively determine the unintended targets. As the next step in drug-off target prediction, the suggested unintended targets could be tested for their binding to DIMATE. Direct binding experiments would be provide the most confident results. Alternatively, in an easier approach, the indirect physiological effects of DIMATE binding to its off-targets could be analyzed. For example, ligand similarity approach has suggested that a possible off-target of DIMATE could be carnitine acetyltransferase in humans. In the literature, it the function of the mouse homolog of this enzyme was previously shown to be necessary for the cells to exit from G1 phase of the cell cycle toward the S phase [124]. Therefore, a possible experiment to test whether the inhibitor also targets carnitine acetyltransferase would be to treat cells that express high levels of ALDH3A1 with DIMATE, and analyze the cell cycle stage by flow cytometry. If DIMATE actually binds to carnitine acetyltransferase, the analysis of stages of the different cell cycles by propidium iodide staining would reveal the accumulation of cells in G1 in DIMATE treated cells compared to the untreated samples. Such strategies can also be adopted to initially bypass the more laborious experiments of chemical binding.

APPENDIX

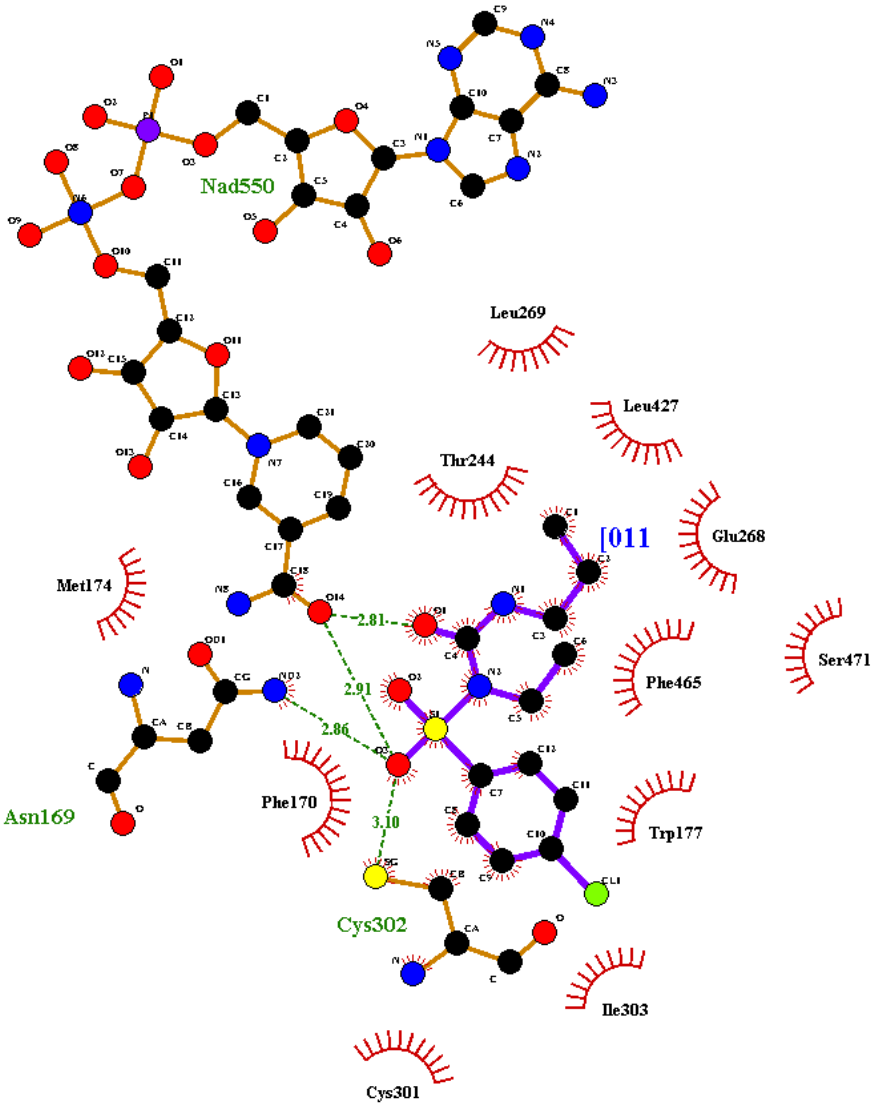
Appendix A: Reverse Docking Poses

The verification docks of ALDH1, ALDH2 and ALDH3 with acetaldehyde, API2, benzaldeyde, DIMATE, disulfiram, MATE and propanal are visualized with LigPlot in Figure 5.1.

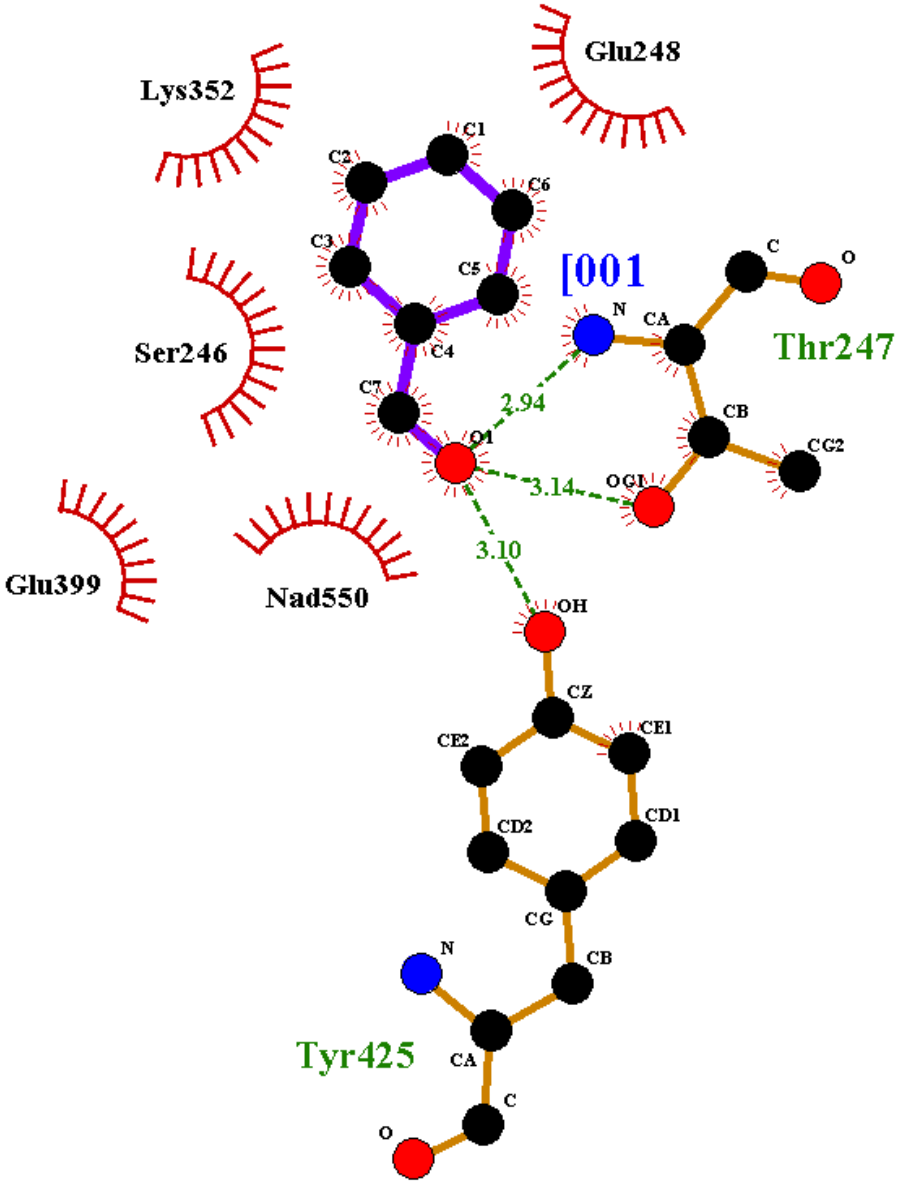
(A) ALDH1-Acetaldeyhyde



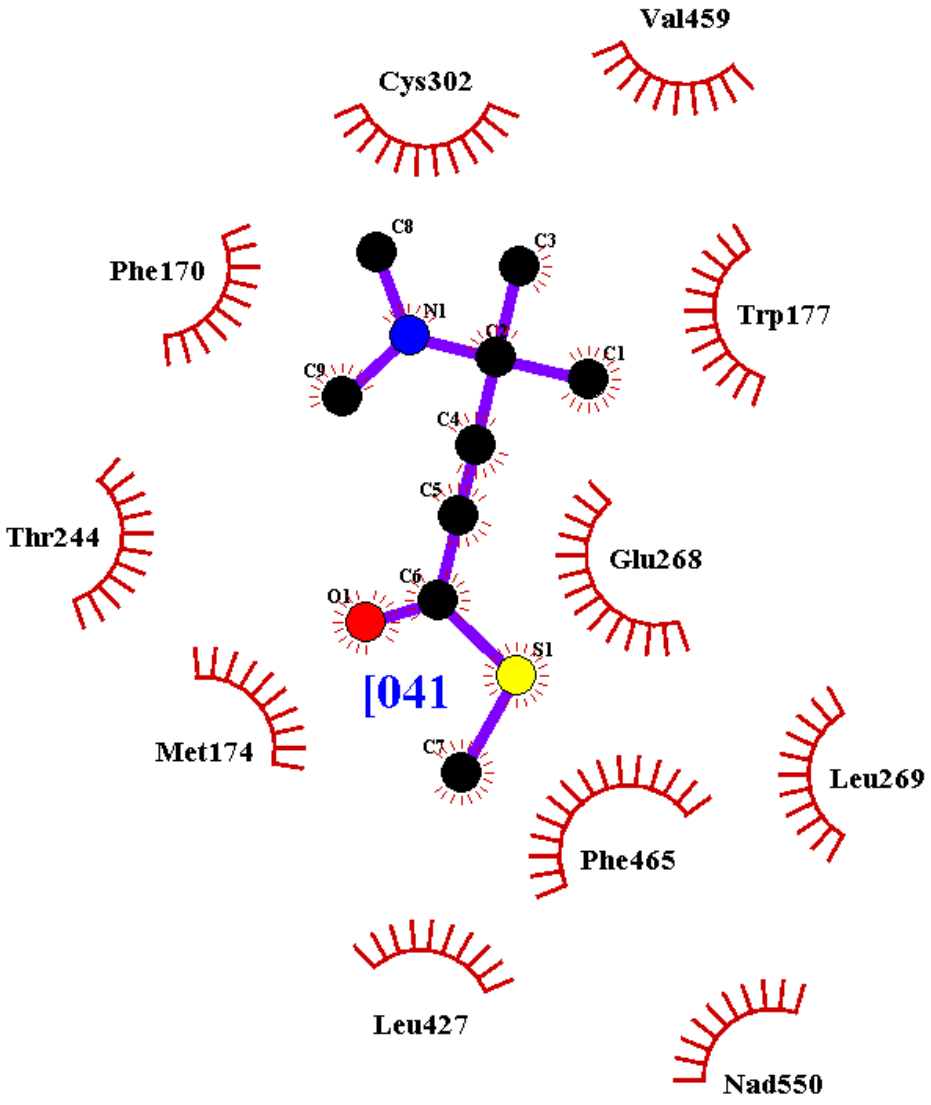
(B) ALDH1-API2


Nad550
Leu269
Leu427
Thr244
Glu268
[011
Met174
Ser471
Phe465
Asn169
Phe170
Trp177
Cys302
Ile303
Cys301
2.81
2.91
2.86
3.10

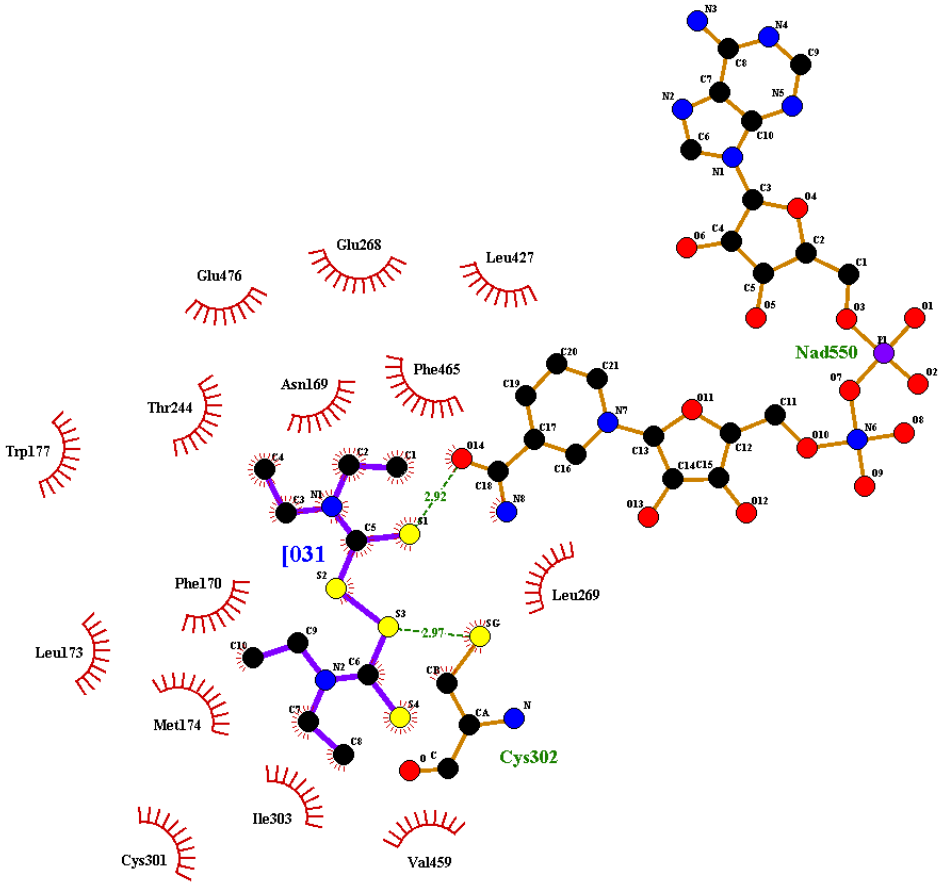
(C) ALDH1-Benzaldehyde



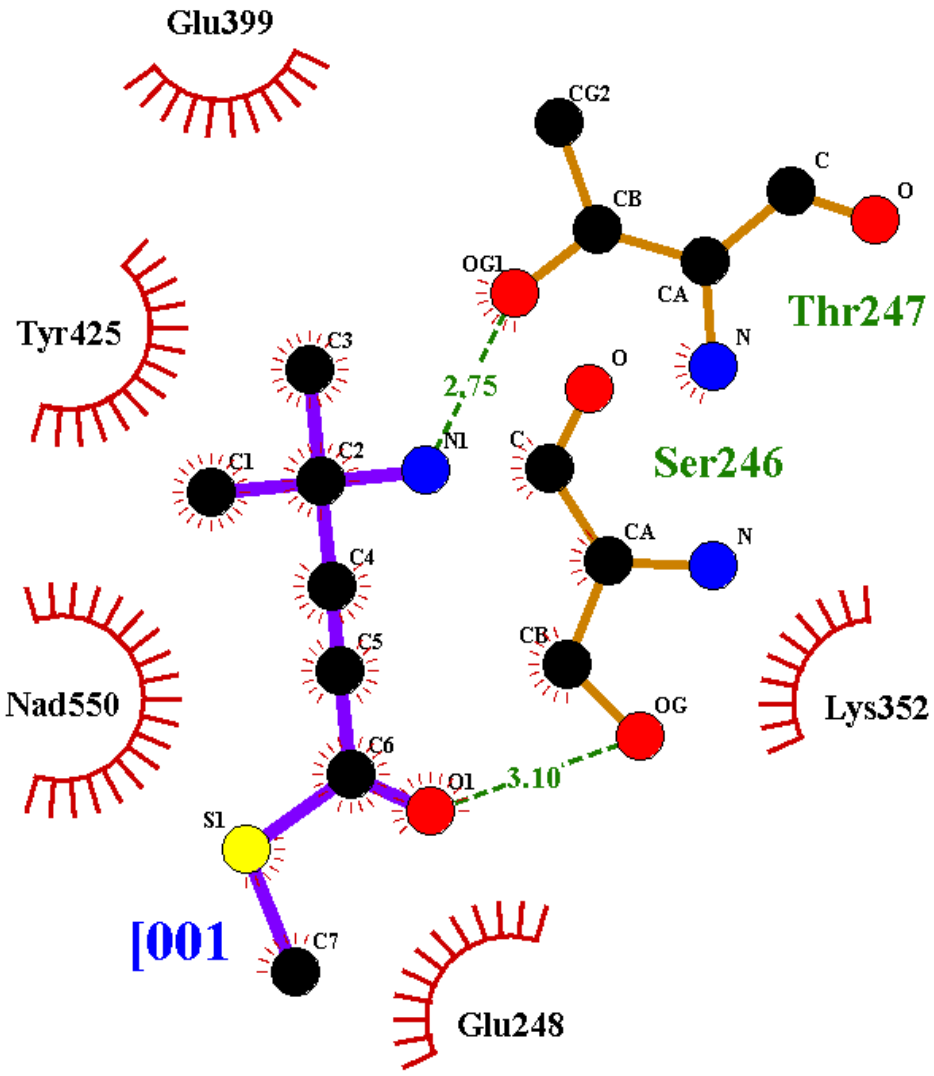
(D) ALDH1-DIMATE



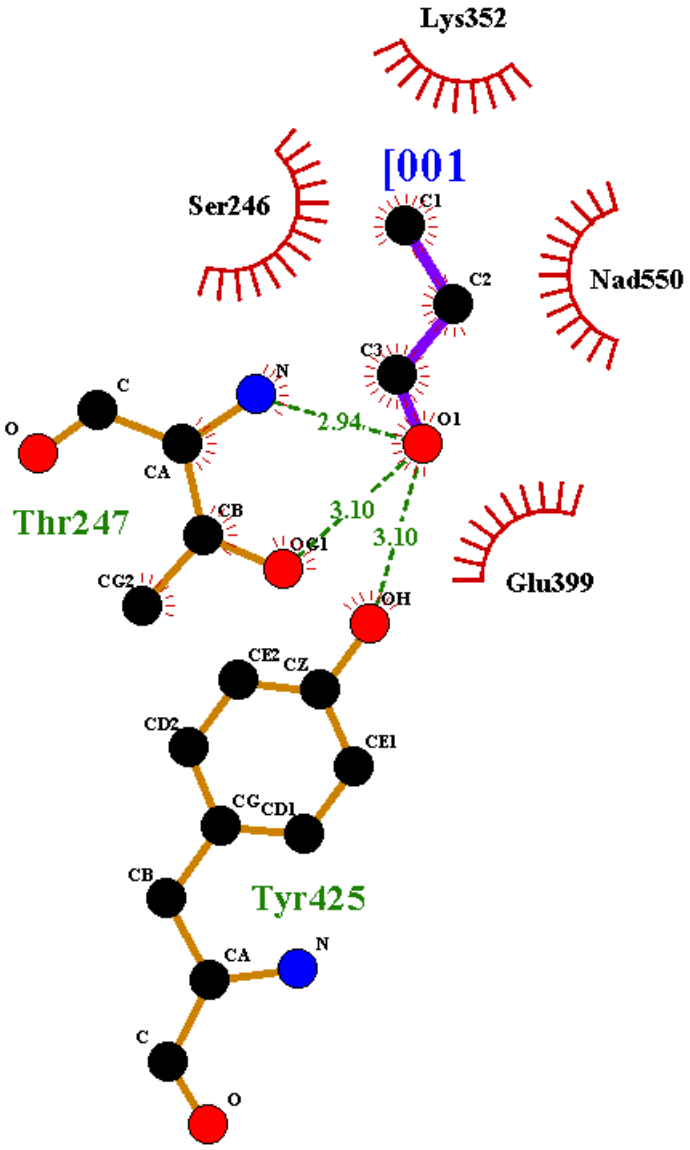
(E) ALDH1-DISULFIRAM



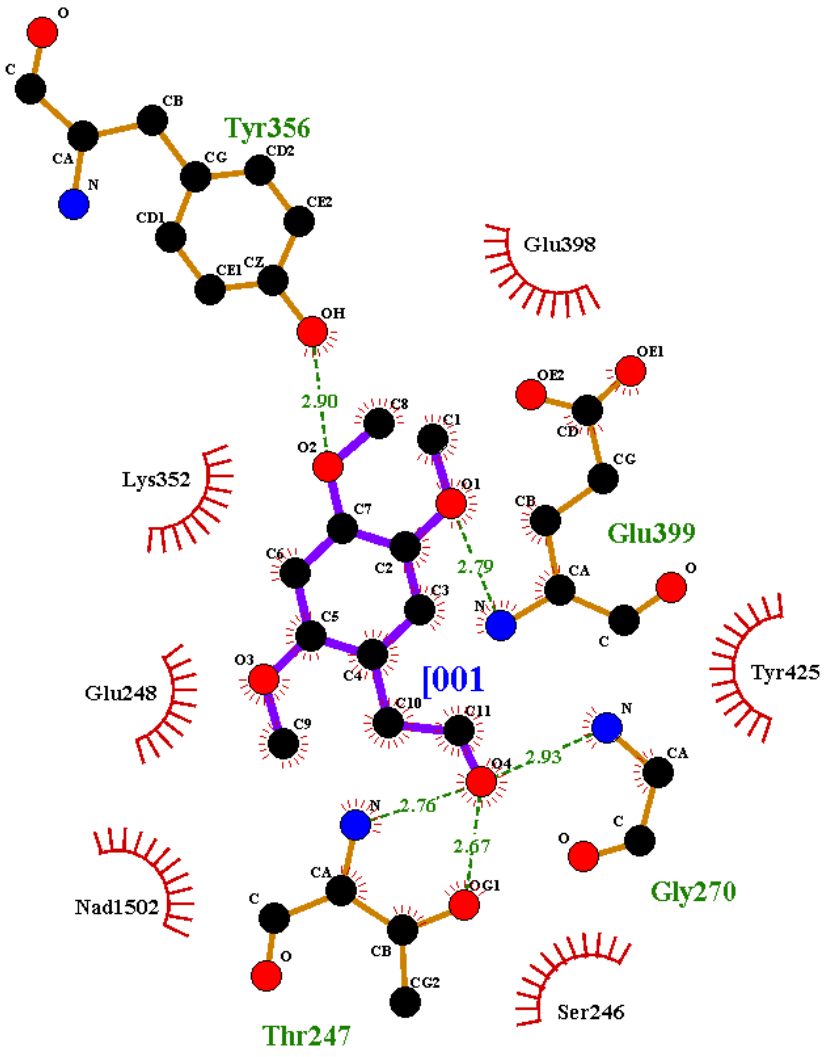
(F) ALDH1-MATE


Glu399
CG2
CB
C
O
OG1
CA
Thr247
Tyr425
C3
N
2.75
O
N1
C
C1
C2
Ser246
CA
N
C4
C5
CB
Nad550
OG
Lys352
C6
O1
3.10
S1
[001
C7
Glu248

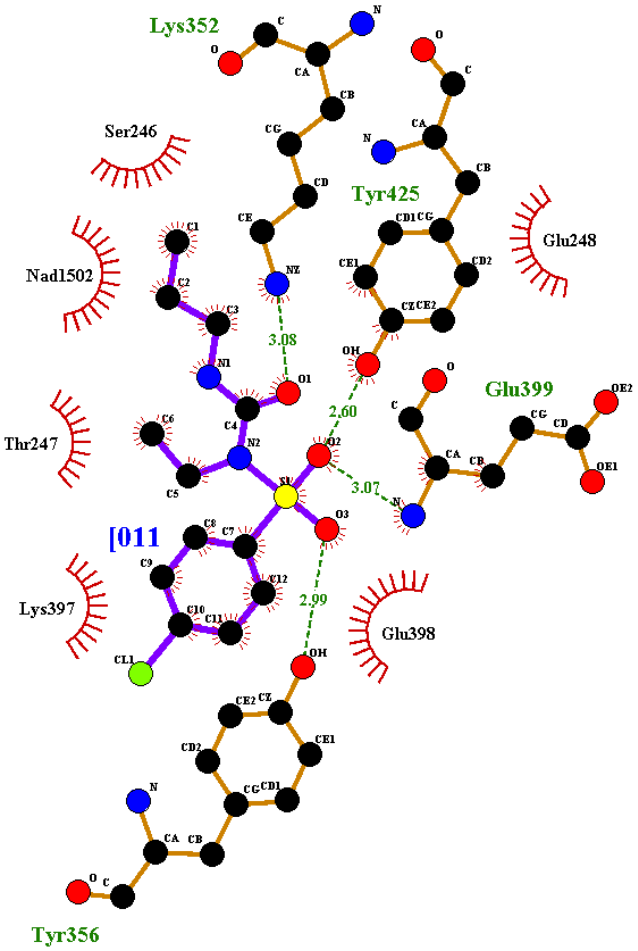
(G) ALDH1-PROPANAL



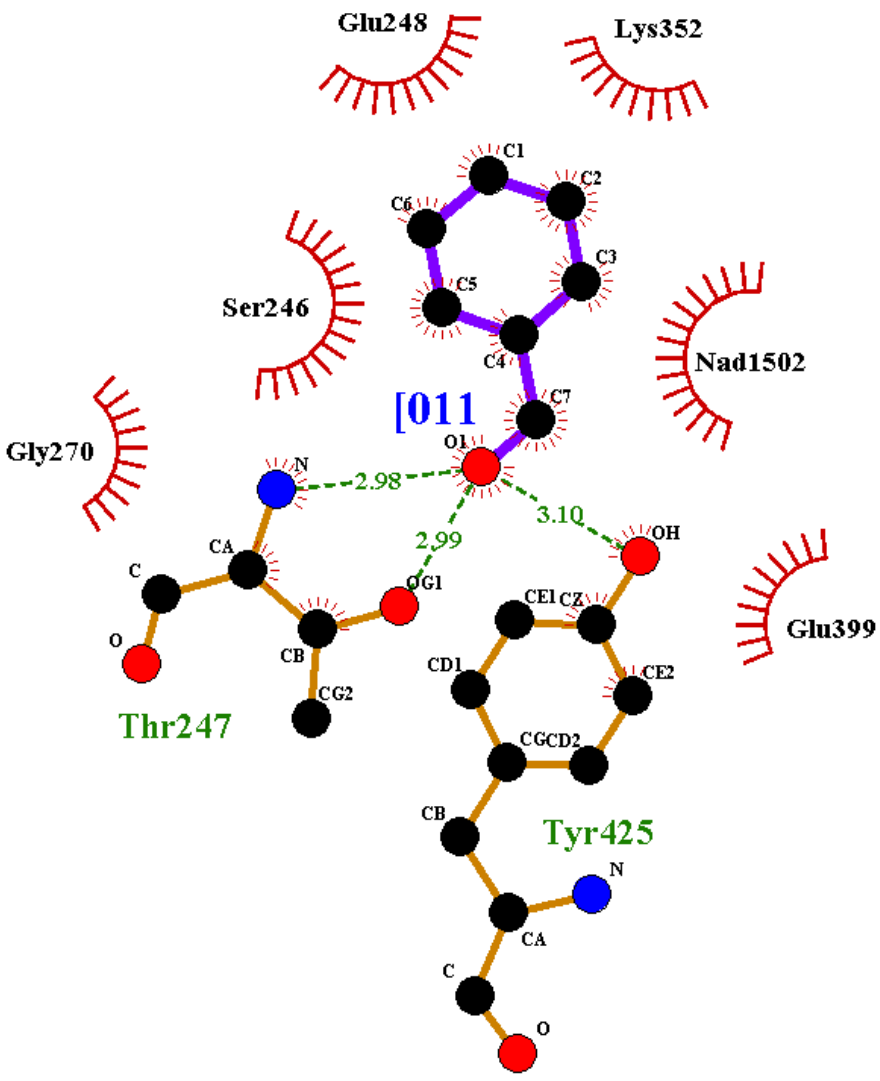
(H) ALDH2-Acetaldeyhyde



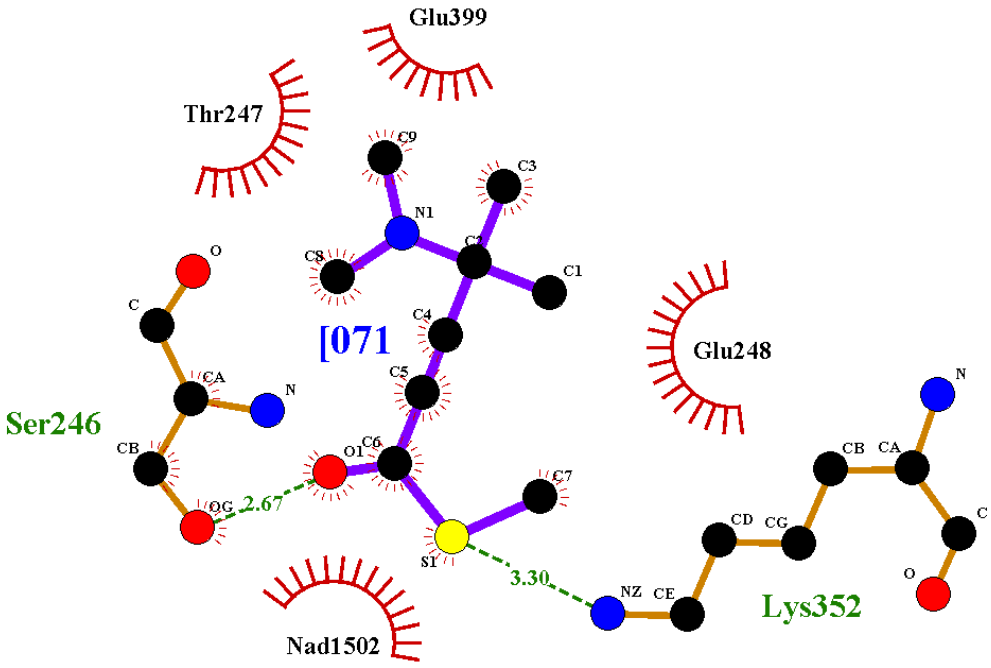
(I) ALDH2-API2


Lys352
Ser246
Tyr425
Glu248
Nad1502
Thr247
Glu399
[011
Lys397
Glu398
Tyr356
3.08
2.60
3.07
2.99

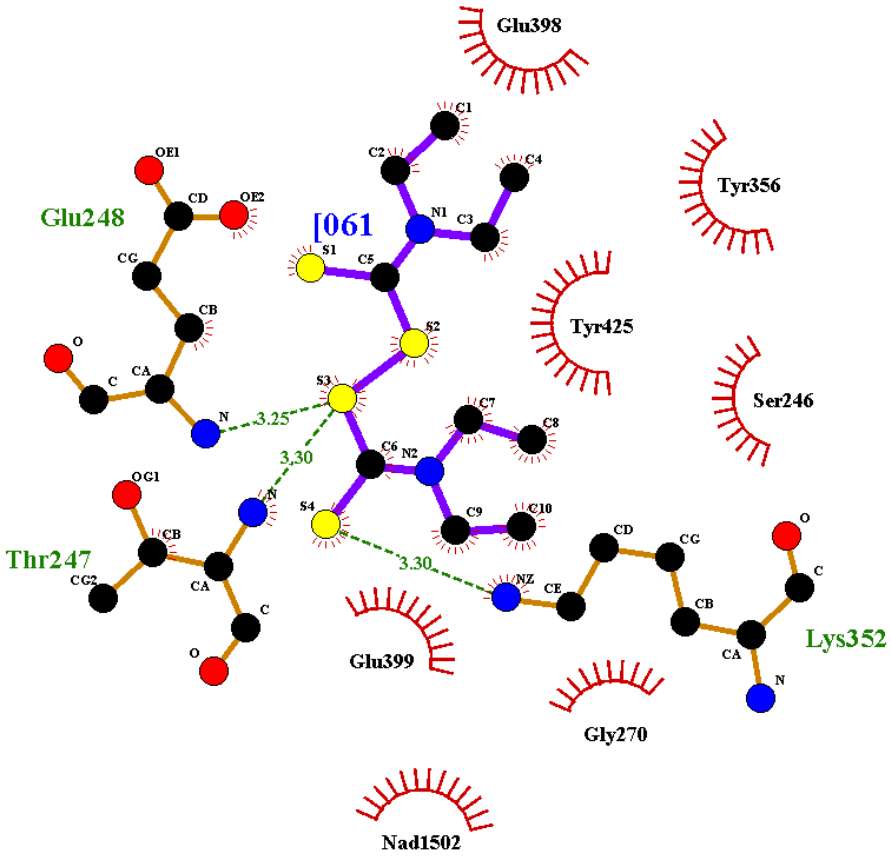
(J) ALDH2-Benzaldehyde



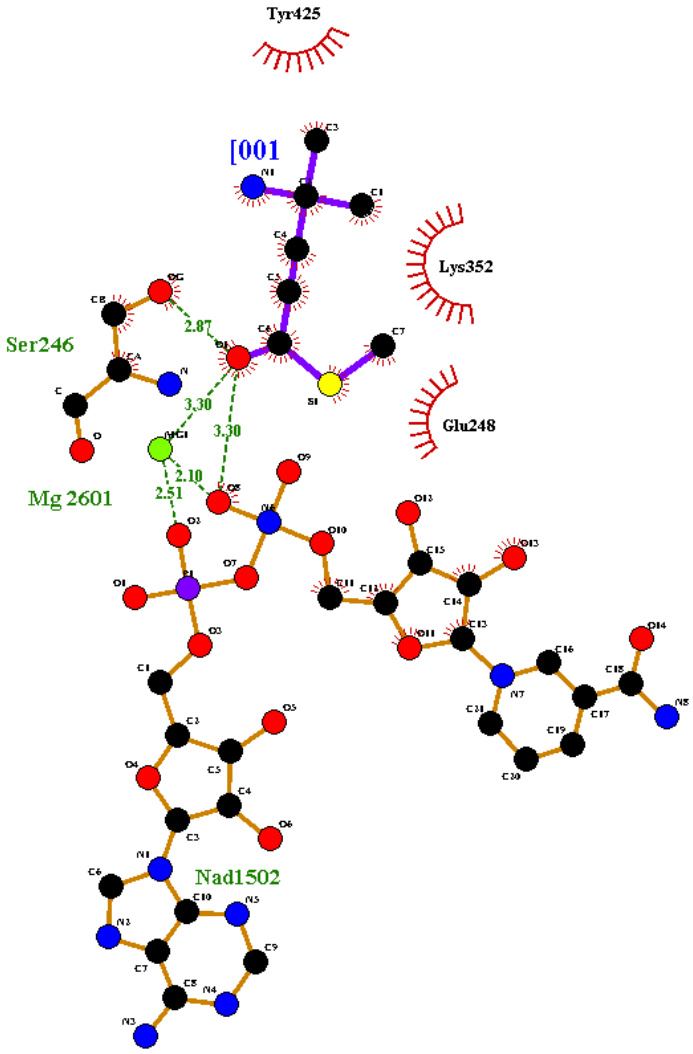
(K) ALDH2-DIMATE



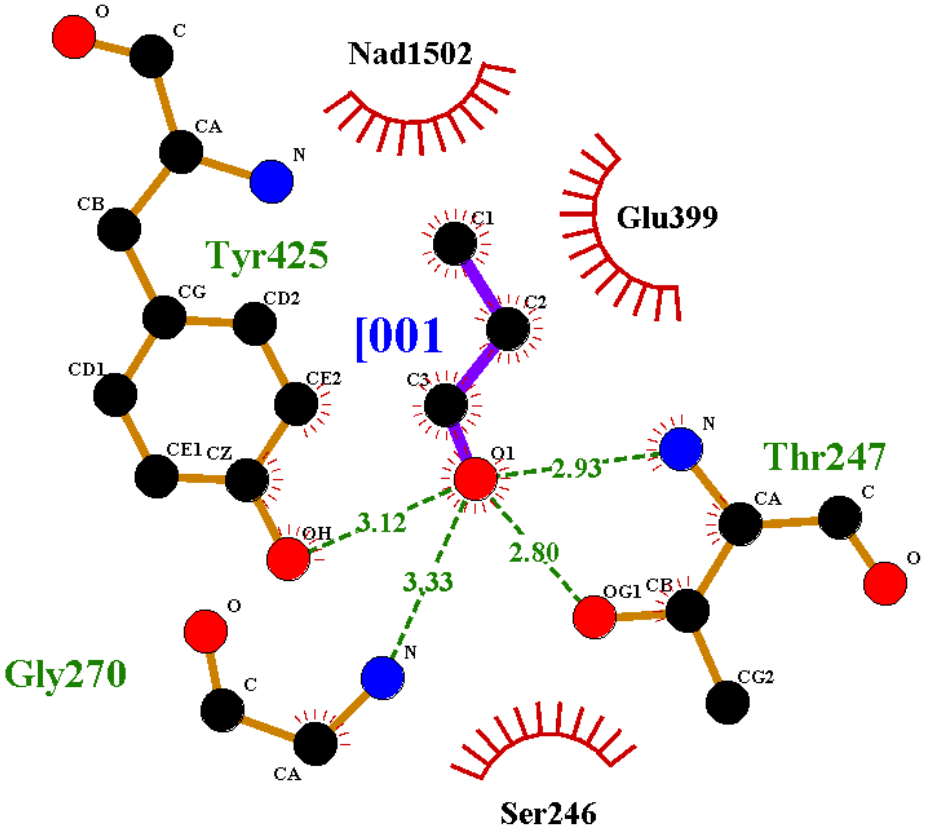
(L) ALDH2-DISULFIRAM



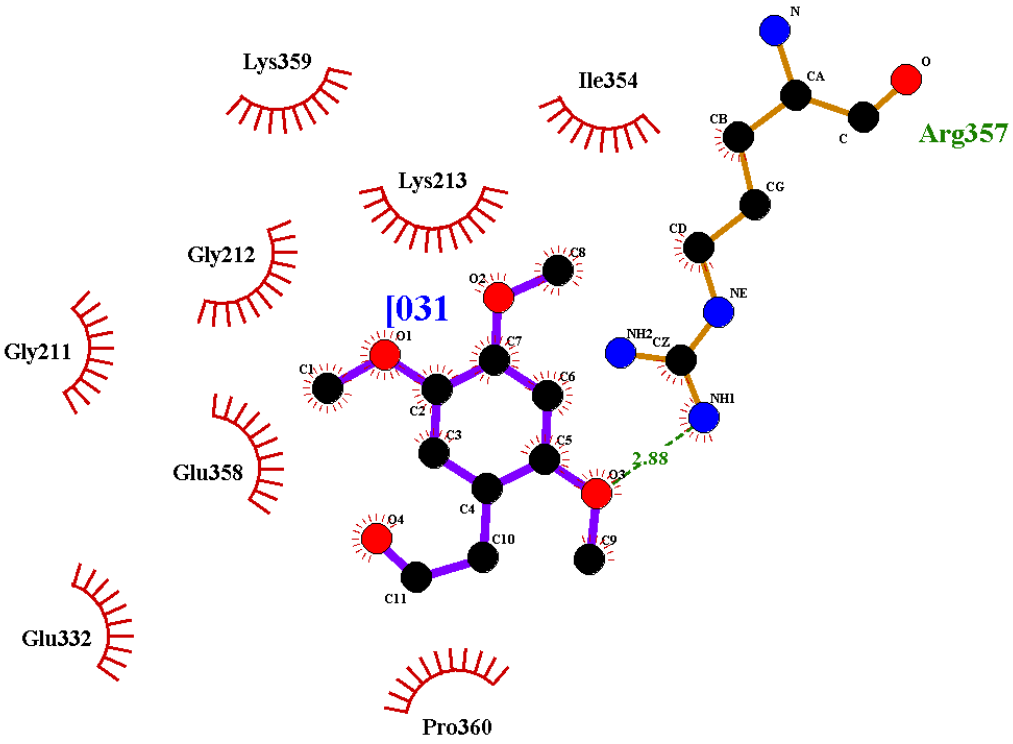
(M) ALDH2-MATE


Tyr425
[001
Lys352
Ser246
Glu248
Mg 2601
Nad1502
2.87
3.30
3.30
2.10
2.51

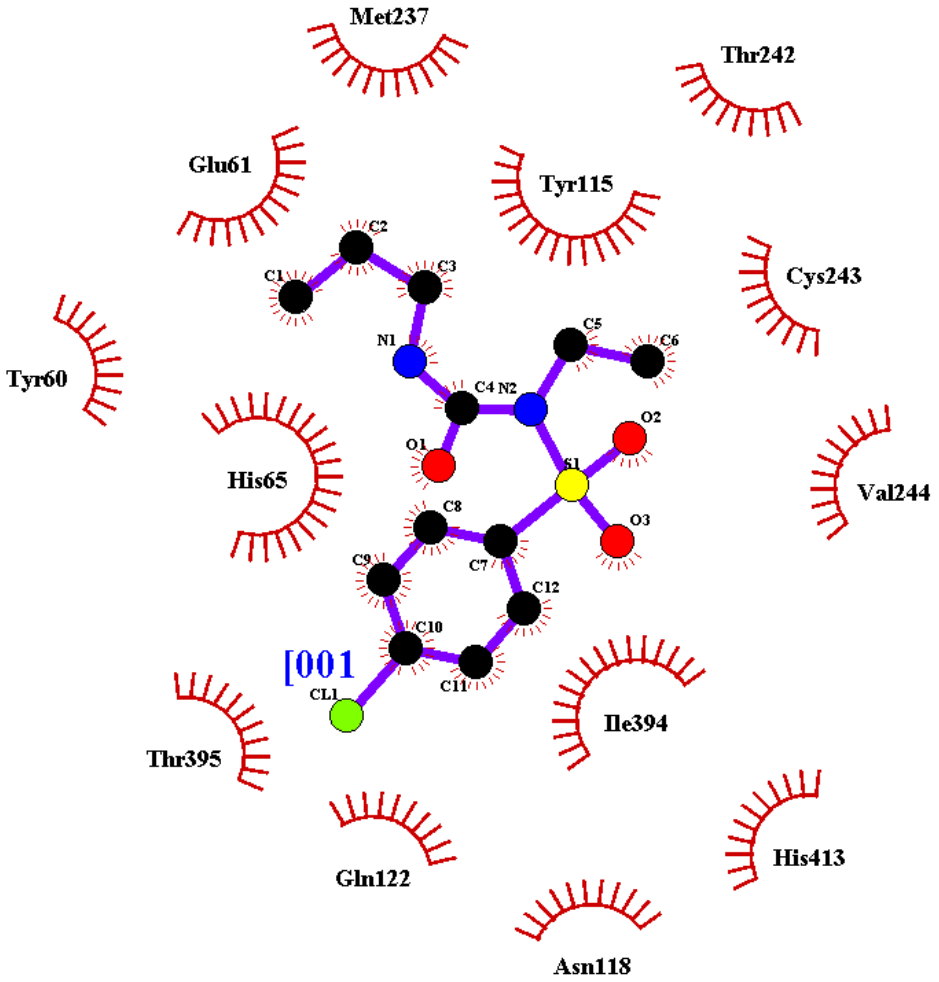
(N) ALDH2-PROPANAL



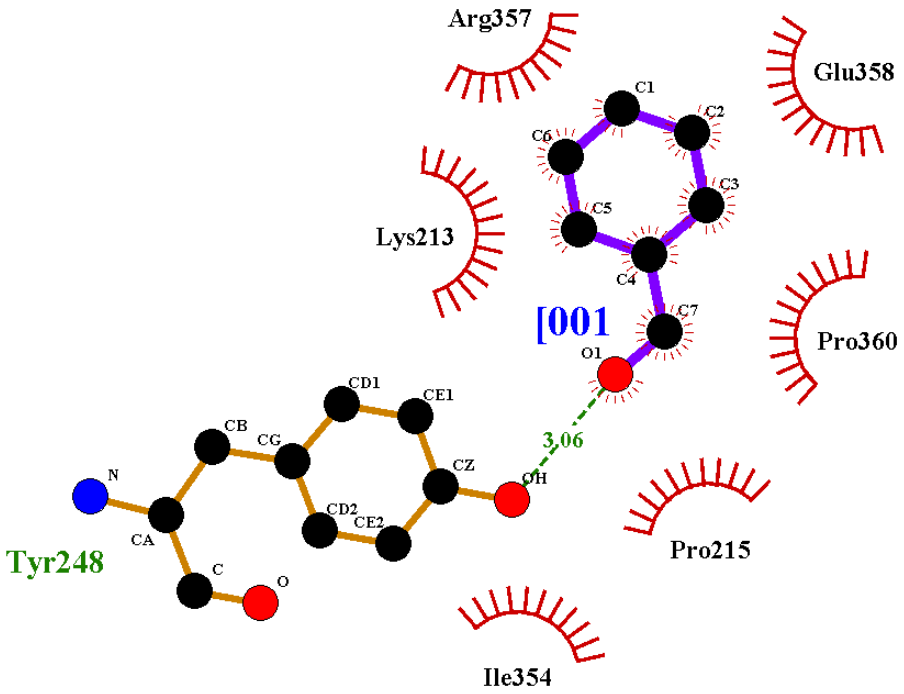
(O) ALDH3-Acetaldehyde



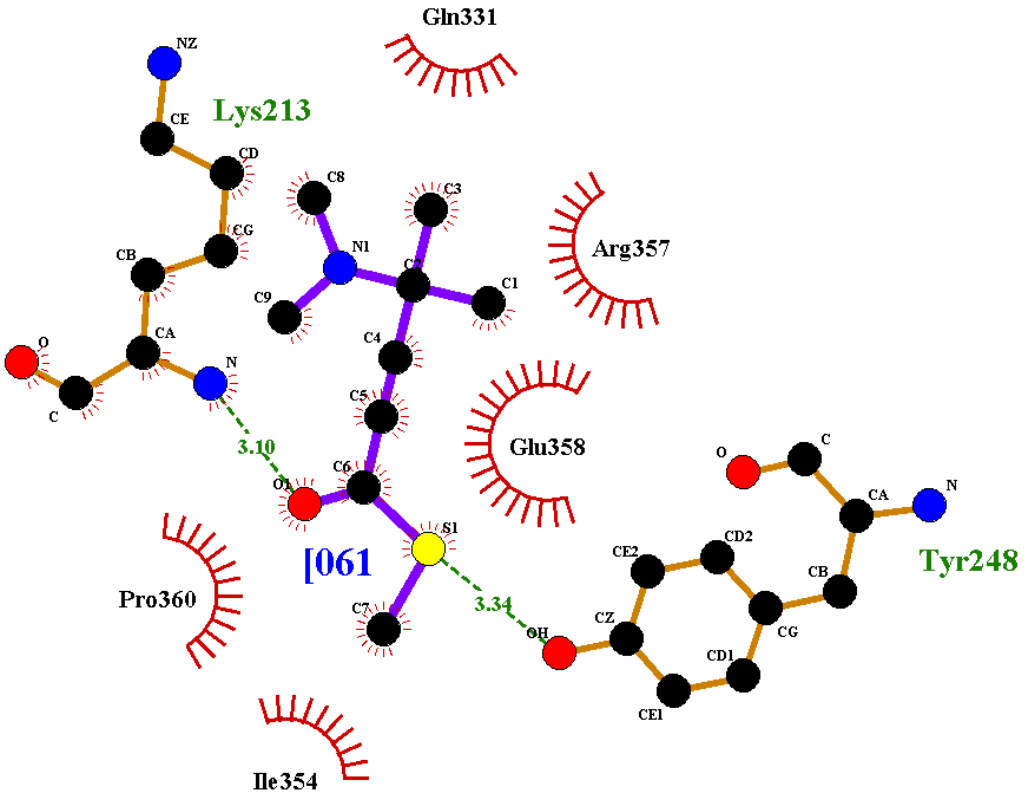
(P) ALDH3 – API2


Met237
Thr242
Glu61
Tyr115
Cys243
Tyr60
His65
Val244
Ile394
Thr395
Gln122
His413
Asn118
[001
C1
C2
C3
N1
C4
N2
C5
C6
O1
O2
S1
O3
C7
C8
C9
C10
C11
C12
CL1

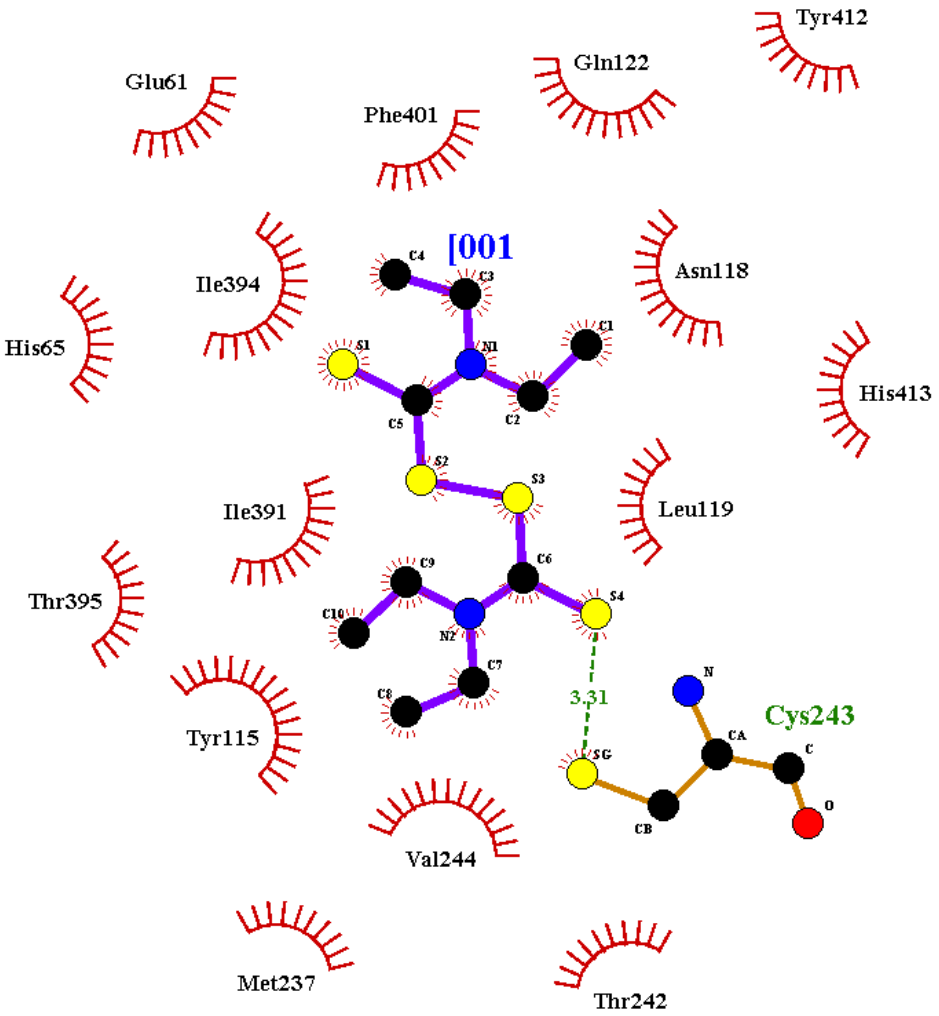
(R) ALDH3-Benzaldehyde



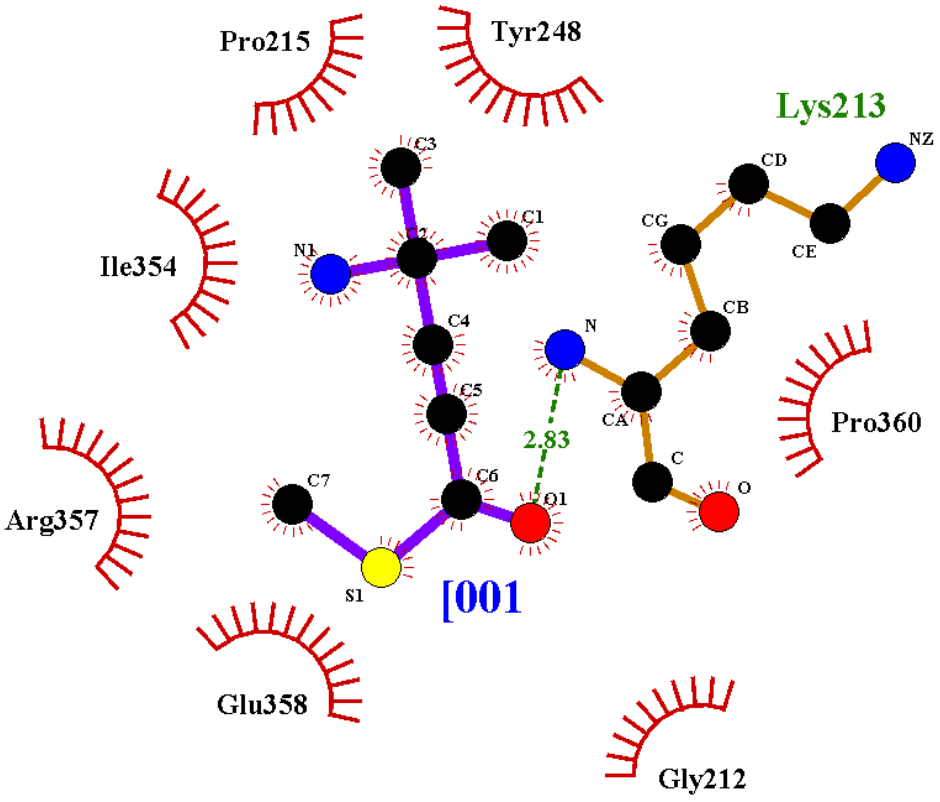
(S) ALDH3- DIMATE



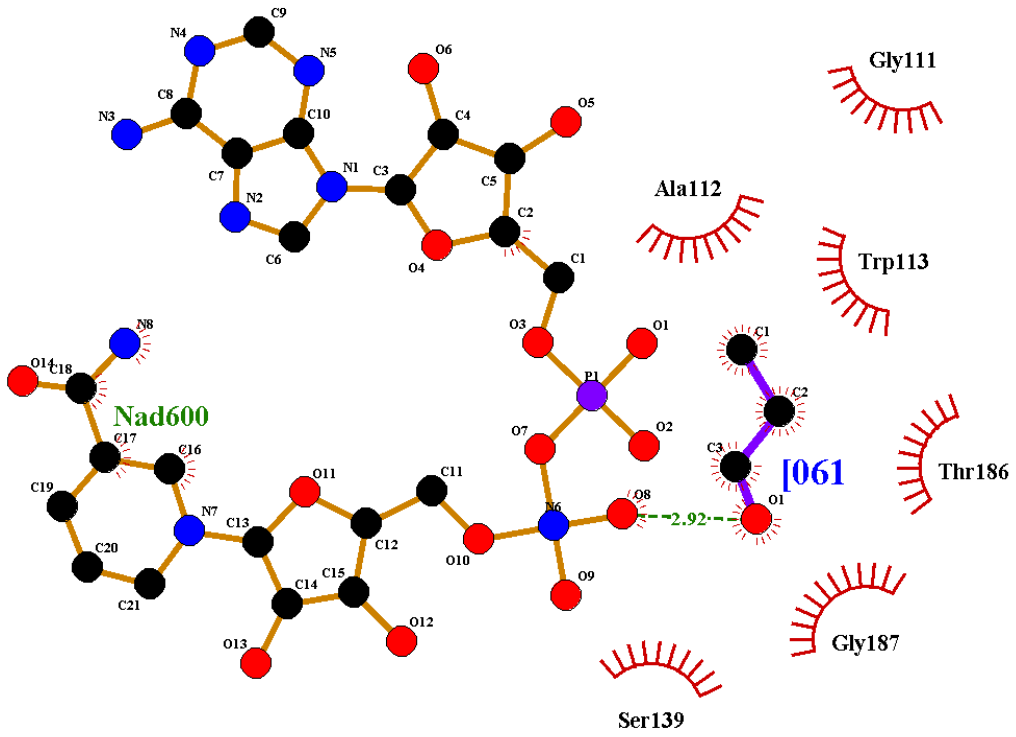
(T) ALDH3 – DISULFIRAM



(U) ALDH3 – MATE


Pro215
Tyr248
Lys213
NZ
CD
CG
CE
C3
C1
C2
N1
Ile354
CB
N
C4
C5
2.83
CA
Pro360
C
O
C7
C6
O1
Arg357
S1
[001
Glu358
Gly212

(V)ALDH3 - PROPANAL



(Y) KEY

Figure 5.1 Visualization of verification docks

Appendix B: Permissions

Yusuf Dogus Dogru
Computational Sciences and Engineering Koc University
Rumelifeneri Yolu Sariyer
Istanbul 34450 Turkey

Email: ydogru@ku.edu.tr

Dear Yusuf Dogus Dogru:

This is to grant you permission to reproduce the following figure in your thesis entitled "Prediction of Off-Targets of Aldehyde Dehydrogenase Inhibitor: An Analysis of Molecular Features of Drug Promiscuity" for Koc University:

Figure 3 from Vindhya Koppaka, David C. Thompson, Ying Chen, Manuel Ellermann, Kyriacos C. Nicolaou, Risto O. Juvonen, Dennis Petersen, Richard A. Deitrich, Thomas D. Hurley, and Vasilis Vasiliou, Aldehyde Dehydrogenase Inhibitors: a Comprehensive Review of the Pharmacology, Mechanism of Action, Substrate Specificity, and Clinical Application, *Pharmacol Rev* July 2012 64:520-539

Permission to reproduce the figure is granted for worldwide use in all languages, translations, and editions, and in any format or medium including print and electronic. The authors and the source of the materials must be cited in full, including the article title, journal title, volume, year, and page numbers.

Sincerely yours,

Richard Dodenhoff Journals Director

9650 Rockville Pike | Bethesda | MD | 20814-3995
P: (301) 634-7060 | F: (301) 634-7061 | E: info@aspet.org | www.aspet.org

Bibliography

1. Jackson, B., et al., *Update on the aldehyde dehydrogenase gene (ALDH) superfamily.* Hum Genomics, 2011. **5**(4): p. 283-303.
2. Wang, M.F., C.L. Han, and S.J. Yin, *Substrate specificity of human and yeast aldehyde dehydrogenases.* Chem Biol Interact, 2009. **178**(1-3): p. 36-9.
3. Quash, G., et al., *Aldehyde dehydrogenase inhibitors: alpha,beta-acetylenic N-substituted aminothiolesters are reversible growth inhibitors of normal epithelial but irreversible apoptogens for cancer epithelial cells from human prostate in culture.* Eur J Med Chem, 2008. **43**(5): p. 906-16.
4. Nobeli, I., A.D. Favia, and J.M. Thornton, *Protein promiscuity and its implications for biotechnology.* Nat Biotechnol, 2009. **27**(2): p. 157-67.
5. Recanatini, M., G. Bottegoni, and A. Cavalli, *In silico antitarget screening.* Drug Discov Today Technol, 2004. **1**(3): p. 209-15.
6. Flomenberg, N., et al., *The use of AMD3100 plus G-CSF for autologous hematopoietic progenitor cell mobilization is superior to G-CSF alone.* Blood, 2005. **106**(5): p. 1867-74.
7. Devine, S.M., et al., *Rapid mobilization of functional donor hematopoietic cells without G-CSF using AMD3100, an antagonist of the CXCR4/SDF-1 interaction.* Blood, 2008. **112**(4): p. 990-8.
8. Kim, J.H. and A.R. Scialli, *Thalidomide: the tragedy of birth defects and the effective treatment of disease.* Toxicol Sci, 2011. **122**(1): p. 1-6.
9. Marchitti, S.A., et al., *Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily.* Expert Opin Drug Metab Toxicol, 2008. **4**(6): p. 697-720.
10. Cheung, A.M., et al., *Aldehyde dehydrogenase activity in leukemic blasts defines a subgroup of acute myeloid leukemia with adverse prognosis and superior NOD/SCID engrafting potential.* Leukemia, 2007. **21**(7): p. 1423-30.
11. Charafe-Jauffret, E., et al., *Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature.* Cancer Res, 2009. **69**(4): p. 1302-13.
12. Carpentino, J.E., et al., *Aldehyde dehydrogenase-expressing colon stem cells contribute to tumorigenesis in the transition from colitis to cancer.* Cancer Res, 2009. **69**(20): p. 8208-15.
13. Clay, M.R., et al., *Single-marker identification of head and neck squamous cell carcinoma cancer stem cells with aldehyde dehydrogenase.* Head Neck, 2010. **32**(9): p. 1195-201.
14. Jiang, F., et al., *Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer.* Mol Cancer Res, 2009. **7**(3): p. 330-8.
15. Wang, L., et al., *Prospective identification of tumorigenic osteosarcoma cancer stem cells in OS99-1 cells based on high aldehyde dehydrogenase activity.* Int J Cancer, 2011. **128**(2): p. 294-303.
16. Alison, M.R., S. Islam, and N.A. Wright, *Stem cells in cancer: instigators and propagators?* J Cell Sci, 2010. **123**(Pt 14): p. 2357-68.
17. Berg, J., J.L. Tymoczko, and L. Stryer, *Biochemistry*. 6 ed. 2007, New York: W. H. Freeman and Company.
18. Muzio, G., et al., *Aldehyde dehydrogenases and cell proliferation.* Free Radic Biol Med, 2012. **52**(4): p. 735-46.
19. Sladek, N.E., *Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact.* J Biochem Mol Toxicol, 2003. **17**(1): p. 7-23.

20. Ducci, F. and D. Goldman, *Genetic approaches to addiction: genes and alcohol.* Addiction, 2008. **103**(9): p. 1414-28.
21. Muzio, G., et al., *Arachidonic acid suppresses growth of human lung tumor A549 cells through down-regulation of ALDH3A1 expression.* Free Radic Biol Med, 2006. **40**(11): p. 1929-38.
22. Pappa, A., et al., *Aldh3a1 protects human corneal epithelial cells from ultraviolet- and 4-hydroxy-2-nonenal-induced oxidative damage.* Free Radic Biol Med, 2003. **34**(9): p. 1178-89.
23. Feldman, R.I. and H. Weiner, *Horse liver aldehyde dehydrogenase. II. Kinetics and mechanistic implications of the dehydrogenase and esterase activity.* J Biol Chem, 1972. **247**(1): p. 267-72.
24. Greenfield, N.J. and R. Pietruszko, *Two aldehyde dehydrogenases from human liver. Isolation via affinity chromatography and characterization of the isozymes.* Biochim Biophys Acta, 1977. **483**(1): p. 35-45.
25. Hempel, J., H. von Bahr-Lindstrom, and H. Jornvall, *Aldehyde dehydrogenase from human liver. Primary structure of the cytoplasmic isoenzyme.* Eur J Biochem, 1984. **141**(1): p. 21-35.
26. King, G. and R. Holmes, *Human corneal and lens aldehyde dehydrogenases. Purification and properties of human lens ALDH1 and differential expression as major soluble proteins in human lens (ALDH1) and cornea (ALDH3).* Adv Exp Med Biol, 1997. **414**: p. 19-27.
27. Hilton, J., *Role of aldehyde dehydrogenase in cyclophosphamide-resistant L1210 leukemia.* Cancer Res, 1984. **44**(11): p. 5156-60.
28. Yoshida, A., et al., *Enhanced transcription of the cytosolic ALDH gene in cyclophosphamide resistant human carcinoma cells.* Adv Exp Med Biol, 1993. **328**: p. 63-72.
29. Yoshida, A., *Genetic polymorphisms of alcohol metabolizing enzymes related to alcohol sensitivity and alcoholic diseases.* Alcohol Alcohol, 1994. **29**(6): p. 693-6.
30. Holmes, R.S. and J. Hempel, *Comparative studies of vertebrate aldehyde dehydrogenase 3: sequences, structures, phylogeny and evolution. Evidence for a mammalian origin for the ALDH3A1 gene.* Chem Biol Interact, 2011. **191**(1-3): p. 113-21.
31. Yoshida, A., et al., *Human aldehyde dehydrogenase gene family.* Eur J Biochem, 1998. **251**(3): p. 549-57.
32. Trettel, F., et al., *Human succinic semialdehyde dehydrogenase. Molecular cloning and chromosomal localization.* Adv Exp Med Biol, 1997. **414**: p. 253-60.
33. Jakobs, C., J. Jaeken, and K.M. Gibson, *Inherited disorders of GABA metabolism.* J Inherit Metab Dis, 1993. **16**(4): p. 704-15.
34. Gibson, K.M., et al., *Combined malonic, methylmalonic and ethylmalonic acid semialdehyde dehydrogenase deficiencies: an inborn error of beta-alanine, L-valine and L-alloisoleucine metabolism?* J Inherit Metab Dis, 1993. **16**(3): p. 563-7.
35. Liu, Z.J., et al., *Crystal structure of a class 3 aldehyde dehydrogenase at 2.6 A resolution.* Adv Exp Med Biol, 1997. **414**: p. 1-7.
36. Cobessi, D., et al., *Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from Streptococcus mutans.* J Mol Biol, 2000. **300**(1): p. 141-52.
37. Larson, H.N., H. Weiner, and T.D. Hurley, *Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase "Asian" variant.* J Biol Chem, 2005. **280**(34): p. 30550-6.

38. D'Ambrosio, K., et al., *The first crystal structure of a thioacylenzyme intermediate in the ALDH family: new coenzyme conformation and relevance to catalysis.* Biochemistry, 2006. **45**(9): p. 2978-86.
39. Lowe, E.D., et al., *Structure of daidzin, a naturally occurring anti-alcohol-addiction agent, in complex with human mitochondrial aldehyde dehydrogenase.* J Med Chem, 2008. **51**(15): p. 4482-7.
40. Kazmi, S.M., et al., *Role of aldehyde dehydrogenase in the biological activity of spermine dialdehyde, a novel immunosuppressive/purging agent.* Pharmacol Res, 1992. **25**(4): p. 383-92.
41. Xiao, T., et al., *Molecular cloning and oxidative modification of human lens ALDH1A1: implication in impaired detoxification of lipid aldehydes.* J Toxicol Environ Health A, 2009. **72**(9): p. 577-84.
42. Koppaka, V., et al., *Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application.* Pharmacol Rev, 2012. **64**(3): p. 520-39.
43. Takahashi, K., H. Weiner, and D.L. Filmer, *Effects of pH on horse liver aldehyde dehydrogenase: alterations in metal ion activation, number of functioning active sites, and hydrolysis of the acyl intermediate.* Biochemistry, 1981. **20**(21): p. 6225-30.
44. Min, H., B. Shane, and E.L. Stokstad, *Identification of 10-formyltetrahydrofolate dehydrogenase-hydrolase as a major folate binding protein in liver cytosol.* Biochim Biophys Acta, 1988. **967**(3): p. 348-53.
45. Strickland, K.C., et al., *Enzymatic properties of ALDH1L2, a mitochondrial 10-formyltetrahydrofolate dehydrogenase.* Chem Biol Interact, 2011. **191**(1-3): p. 129-36.
46. Steinmetz, C.G., et al., *Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion.* Structure, 1997. **5**(5): p. 701-11.
47. Colby, T.D., et al., *Active site modifications in a double mutant of liver alcohol dehydrogenase: structural studies of two enzyme-ligand complexes.* Biochemistry, 1998. **37**(26): p. 9295-304.
48. Hempel, J., et al., *Aldehyde dehydrogenase catalytic mechanism. A proposal.* Adv Exp Med Biol, 1999. **463**: p. 53-9.
49. Gupta, P.B., C.L. Chaffer, and R.A. Weinberg, *Cancer stem cells: mirage or reality?* Nat Med, 2009. **15**(9): p. 1010-2.
50. Christ, O., et al., *Improved purification of hematopoietic stem cells based on their elevated aldehyde dehydrogenase activity.* Haematologica, 2007. **92**(9): p. 1165-72.
51. Ma, I. and A.L. Allan, *The role of human aldehyde dehydrogenase in normal and cancer stem cells.* Stem Cell Rev, 2011. **7**(2): p. 292-306.
52. Kastan, M.B., et al., *Direct demonstration of elevated aldehyde dehydrogenase in human hematopoietic progenitor cells.* Blood, 1990. **75**(10): p. 1947-50.
53. Moreb, J., et al., *Role of aldehyde dehydrogenase in the protection of hematopoietic progenitor cells from 4-hydroperoxycyclophosphamide by interleukin 1 beta and tumor necrosis factor.* Cancer Res, 1992. **52**(7): p. 1770-4.
54. Obermair, F.J., et al., *A novel classification of quiescent and transit amplifying adult neural stem cells by surface and metabolic markers permits a defined simultaneous isolation.* Stem Cell Res, 2010. **5**(2): p. 131-43.
55. Jean, E., et al., *Aldehyde dehydrogenase activity promotes survival of human muscle precursor cells.* J Cell Mol Med, 2011. **15**(1): p. 119-33.

56. Martignani, E., et al., *Human milk protein production in xenografts of genetically engineered bovine mammary epithelial stem cells.* PLoS One, 2010. **5**(10): p. e13372.
57. Ginestier, C., et al., *ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome.* Cell Stem Cell, 2007. **1**(5): p. 555-67.
58. Jones, D.E., Jr., et al., *Cloning and complete nucleotide sequence of a full-length cDNA encoding a catalytically functional tumor-associated aldehyde dehydrogenase.* Proc Natl Acad Sci U S A, 1988. **85**(6): p. 1782-6.
59. Holmes, R.S., R.A. Popp, and J.L. VandeBerg, *Genetics of ocular NAD+-dependent alcohol dehydrogenase and aldehyde dehydrogenase in the mouse: evidence for genetic identity with stomach isozymes and localization of Ahd-4 on chromosome 11 near trembler.* Biochem Genet, 1988. **26**(3-4): p. 191-205.
60. Boesch, J.S., C. Lee, and R.G. Lindahl, *Constitutive expression of class 3 aldehyde dehydrogenase in cultured rat corneal epithelium.* J Biol Chem, 1996. **271**(9): p. 5150-7.
61. Miyauchi, K., et al., *Molecular cloning, sequencing, and expression of cDNA for rat liver microsomal aldehyde dehydrogenase.* J Biol Chem, 1991. **266**(29): p. 19536-42.
62. Rogers, G.R., et al., *Genomic organization and expression of the human fatty aldehyde dehydrogenase gene (FALDH).* Genomics, 1997. **39**(2): p. 127-35.
63. De Laurenzi, V., et al., *Sjogren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene.* Nat Genet, 1996. **12**(1): p. 52-7.
64. Hsu, L.C., et al., *Molecular cloning, genomic organization, and chromosomal localization of an additional human aldehyde dehydrogenase gene, ALDH6.* Genomics, 1994. **24**(2): p. 333-41.
65. Hsu, L.C. and W.C. Chang, *Sequencing and expression of the human ALDH8 encoding a new member of the aldehyde dehydrogenase family.* Gene, 1996. **174**(2): p. 319-22.
66. Hsu, L.C., W.C. Chang, and A. Yoshida, *Human aldehyde dehydrogenase genes, ALDH7 and ALDH8: genomic organization and gene structure comparison.* Gene, 1997. **189**(1): p. 89-94.
67. Moreb, J.S., et al., *ALDH isozymes downregulation affects cell growth, cell motility and gene expression in lung cancer cells.* Mol Cancer, 2008. **7**: p. 87.
68. Ruzinova, M.B. and R. Benezra, *Id proteins in development, cell cycle and cancer.* Trends Cell Biol, 2003. **13**(8): p. 410-8.
69. Sarhadi, V.K., et al., *Increased expression of high mobility group A proteins in lung cancer.* J Pathol, 2006. **209**(2): p. 206-12.
70. Marchitti, S.A., et al., *Ultraviolet radiation: cellular antioxidant response and the role of ocular aldehyde dehydrogenase enzymes.* Eye Contact Lens, 2011. **37**(4): p. 206-13.
71. Cadenas, E., et al., *Effects of 4-hydroxynonenal on isolated hepatocytes. Studies on chemiluminescence response, alkane production and glutathione status.* Biochem J, 1983. **214**(2): p. 479-87.
72. Poot, M., et al., *Reversible inhibition of DNA and protein synthesis by cumene hydroperoxide and 4-hydroxy-nonenal.* Mech Ageing Dev, 1988. **43**(1): p. 1-9.
73. Canuto, R.A., et al., *The effect of various aldehydes on the respiration of rat liver and hepatoma AH-130 cells.* Cell Biochem Funct, 1985. **3**(1): p. 3-8.
74. Canuto, R.A., et al., *Role of aldehyde metabolizing enzymes in mediating effects of aldehyde products of lipid peroxidation in liver cells.* Carcinogenesis, 1994. **15**(7): p. 1359-64.
75. Trombetta, A., et al., *Arachidonic and docosahexaenoic acids reduce the growth of A549 human lung-tumor cells increasing lipid peroxidation and PPARs.* Chem Biol Interact, 2007. **165**(3): p. 239-50.

76. Quemener, V., et al., *Aldehyde dehydrogenase activity in xenografted human brain tumor in nude mice. Preliminary results in human glioma biopsies.* J Neurooncol, 1990. **9**(2): p. 115-23.
77. Muzio, G., et al., *Inhibition of cytosolic class 3 aldehyde dehydrogenase by antisense oligonucleotides in rat hepatoma cells.* Chem Biol Interact, 2001. **130-132**(1-3): p. 219-25.
78. O'Connor, K.A. and B.L. Roth, *Finding new tricks for old drugs: an efficient route for public-sector drug discovery.* Nat Rev Drug Discov, 2005. **4**(12): p. 1005-14.
79. Sanseau, P., et al., *Use of genome-wide association studies for drug repositioning.* Nat Biotechnol, 2012. **30**(4): p. 317-20.
80. Konc, J., et al., *ProBiS-database: precalculated binding site similarities and local pairwise alignments of PDB structures.* J Chem Inf Model, 2012. **52**(2): p. 604-12.
81. Xie, L., L. Xie, and P.E. Bourne, *A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery.* Bioinformatics, 2009. **25**(12): p. i305-12.
82. Guner, O.F. and J.P. Bowen, *Setting the record straight: the origin of the pharmacophore concept.* J Chem Inf Model, 2014. **54**(5): p. 1269-83.
83. Doman, T.N., et al., *Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B.* Journal of Medicinal Chemistry, 2002. **45**(11): p. 2213-2221.
84. Hein, M., D. Zilian, and C.A. Sotriffer, *Docking compared to 3D-pharmacophores: the scoring function challenge.* Drug Discovery Today: Technologies, 2010. **7**(4): p. e229-e236.
85. Vuorinen, A. and D. Schuster, *Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling.* Methods, 2015. **71**: p. 113-34.
86. Klabunde, T., C. Giegerich, and A. Evers, *Sequence-Derived Three-Dimensional Pharmacophore Models for G-Protein-Coupled Receptors and Their Application in Virtual Screening.* Journal of Medicinal Chemistry, 2009. **52**(9): p. 2923-2932.
87. Sanders, M.P.A., et al., *Snooker: A Structure-Based Pharmacophore Generation Tool Applied to Class A GPCRs.* Journal of Chemical Information and Modeling, 2011. **51**(9): p. 2277-2292.
88. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic Acids Research, 2012. **40**(D1): p. D1100-D1107.
89. Bolton E, W.Y., Thiessen PA, Bryant SH., *PubChem: Integrated Platform of Small Molecules and Biological Activities.* Annual Reports in Computational Chemistry, 2008. **4**.
90. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
91. Leach, A.R., et al., *Three-dimensional pharmacophore methods in drug discovery.* J Med Chem, 2010. **53**(2): p. 539-58.
92. Fox, S., et al., *High-throughput screening: update on practices and success.* J Biomol Screen, 2006. **11**(7): p. 864-9.
93. Huang, N., B.K. Shoichet, and J.J. Irwin, *Benchmarking Sets for Molecular Docking.* Journal of Medicinal Chemistry, 2006. **49**(23): p. 6789-6801.
94. Virshup, A.M., et al., *Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds.* Journal of the American Chemical Society, 2013. **135**(19): p. 7296-7303.
95. Wei, Y., J. Thompson, and C.A. Floudas, *CONCORD: a consensus method for protein secondary structure prediction via mixed integer linear optimization*. Vol. 468. 2012. 831-850.

96. Sadowski, J., J. Gasteiger, and G. Klebe, *Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures.* Journal of Chemical Information and Computer Sciences, 1994. **34**(4): p. 1000-1008.
97. 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.* Journal of Chemical Information and Modeling, 2006. **46**(4): p. 1848-1861.
98. Ref. Dassault Systèmes BIOVIA, D.S.M.E., Release 4.5, San Diego: Dassault Systèmes, 2015.
99. Molecular Operating Environment (MOE), C.C.G.I., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2015. .
100. Sanders, M.P.A., et al., *Comparative Analysis of Pharmacophore Screening Tools.* Journal of Chemical Information and Modeling, 2012. **52**(6): p. 1607-1620.
101. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
102. Gish, W. and D.J. States, *Identification of protein coding regions by database similarity search.* Nat Genet, 1993. **3**(3): p. 266-72.
103. Biasini, M., et al., *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W252-8.
104. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.
105. Kiefer, F., et al., *The SWISS-MODEL Repository and associated resources.* Nucleic Acids Res, 2009. **37**(Database issue): p. D387-92.
106. Guex, N., M.C. Peitsch, and T. Schwede, *Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective.* Electrophoresis, 2009. **30 Suppl 1**: p. S162-73.
107. Tseng, Y.Y., et al., *SplitPocket: identification of protein functional surfaces and characterization of their spatial patterns.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W384-9.
108. Hendlich, M., et al., *Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions.* J Mol Biol, 2003. **326**(2): p. 607-20.
109. Konc, J. and D. Janezic, *ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W214-21.
110. Konc, J. and D. Janezic, *An improved branch and bound algorithm for the maximum clique problem.* proteins, 2007. **4**: p. 5.
111. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-61.
112. Jones, G., et al., *Development and validation of a genetic algorithm for flexible docking.* J Mol Biol, 1997. **267**(3): p. 727-48.
113. Thomsen, R. and M.H. Christensen, *MolDock: a new technique for high-accuracy molecular docking.* J Med Chem, 2006. **49**(11): p. 3315-21.
114. Liu, X., et al., *PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W609-14.
115. 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.

116. Benkert, P., M. Kunzli, and T. Schwede, *QMEAN server for protein model quality estimation.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W510-4.
117. Kharkar, P.S., S. Warrier, and R.S. Gaud, *Reverse docking: a powerful tool for drug repositioning and drug rescue.* Future Med Chem, 2014. **6**(3): p. 333-42.
118. Quash, G., et al., *Novel competitive irreversible inhibitors of aldehyde dehydrogenase (ALDH1): restoration of chemosensitivity of L1210 cells overexpressing ALDH1 and induction of apoptosis in BAF(3) cells overexpressing bcl(2).* Biochem Pharmacol, 2002. **64**(8): p. 1279-92.
119. Morgan, C.A., et al., *N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes.* Chem Biol Interact, 2015. **234**: p. 18-28.
120. Liang, P., et al., *Molecular characterization of the murine thymidylate kinase gene.* Cell Growth Differ, 1995. **6**(10): p. 1333-8.
121. Huang, S.H., et al., *Human dTMP kinase: gene expression and enzymatic activity coinciding with cell cycle progression and cell growth.* DNA Cell Biol, 1994. **13**(5): p. 461-71.
122. Bange, F.C., et al., *An interferon-induced protein with release factor activity is a tryptophanyl-tRNA synthetase.* FEBS Lett, 1992. **300**(2): p. 162-6.
123. Rubin, B.Y., et al., *Interferon induces tryptophanyl-tRNA synthetase expression in human fibroblasts.* J Biol Chem, 1991. **266**(36): p. 24245-8.
124. Brunner, S., et al., *Cloning and characterization of murine carnitine acetyltransferase: evidence for a requirement during cell cycle progression.* Biochem J, 1997. **322 (Pt 2)**: p. 403-10.

Vita

Yusuf Doğuş Doğru was born in İçel, Turkey on May 5, 1987. He received his Bachelor of Science Degree from Bilkent University, Ankara in 2012. From September 2012 to August 2015, he worked in the Department of Computational Sciences and Engineering in Koç University, Turkey. His thesis is on the prediction of off-targets of DIMATE, a small molecule inhibitor of aldehyde dehydrogenase, and on drug off-target prediction.