# KINETIC AND STRUCTURAL CHARACTERIZATION OF ISOENZYME-SELECTIVE ALDEHYDE DEHYDROGENASE 1A INHIBITORS

Mikhail Chtcherbinine

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Biochemistry and Molecular Biology,
Indiana University

August 2016

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Master of Science

_____

Thomas D. Hurley, Ph.D., Chair

Master's Thesis Committee

_____

Millie M. Georgiadis, Ph.D.

June 29, 2016

_____

Clark D. Wells, Ph.D.

**Dedication**

To my parents, for their constant support and love

**Acknowledgements**

First, I would like to thank my advisor, Dr. Tom Hurley, for all of his guidance and mentorship. While always ready to share his insight and advice, he was never the slightest bit overbearing. He continuously placed tremendous amounts of trust in his students by letting us work independently and develop our own ideas and experiments. Dr. Hurley's attitude and managerial style has been a great boon in helping me grow as a scientist. I am sincerely grateful for his support, kindness, and geniality.

Similarly, I'd like to thank my lab mates, Cameron Buchman, Kishore Mahalingan and Buyun Tang, for helping me deal with all manner of issues and predicaments in my lab work. I especially want to thank Cameron for instructing me in so many protocols and techniques, and showing incredible patience and selflessness in answering all of my questions and concerns. I don't think this work could have been done without his help.

Thank you to Dr. Cindy Morgan for all of her work to establish this project, for helping me continue it, and for her constant cheerful attitude.

Thank you to my committee members, Dr. Georgiadis and Dr. Wells. Although we did not have many committee meetings, Dr. Georgiadis and Dr. Wells were always approachable and offered great advice. Thank you to the Dr. Wells, as well as Lauren Bringman and Brandon Lane, for helping me get started with cell culture and allowing me to use their microscopy equipment. Thank you to Dr. Georgiadis, Isha Singh, and Qiujia Chen for giving me chance to send some crystals for data collection on their beam-time.

Thank you to the biochemistry office staff, Jack, Sandy, Melissa, Patty, Sheila, and Darlene, for helping out with my organizational, scheduling, and technical issues. Thank you to Dr. Harrington for allowing me to use the cell culture facility. Thank you to Dr. Goebl and Dr. Quilliuam for guidance in progressing through the Master's program and writing my thesis. Thank you to Dr. Lifan Zeng for helping me use the LC/MS and other Chemical Genomics Core equipment.

Of course, thank you to my parents, Sergey and Inna, for their unwavering support. Your love and care has carried me through some tough times. Thank you to Carter, for being a boundless source of energy, comfort, and joy.

This research was supported by the U.S. National Institute of Health grant R21CA198409 to TDH. Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

Mikhail Chtcherbinine

# KINETIC AND STRUCTURAL CHARACTERIZATION OF ISOENZYME-SELECTIVE ALDEHYDE DEHYDROGENASE 1A INHIBITORS

The human aldehyde dehydrogenase superfamily consists of 19 distinct genetic loci that play key roles in both health and disease. Aldehyde dehydrogenases are primarily involved in the metabolism of reactive aldehyde substrates; the ALDH1A subfamily, in particular, metabolizes retinaldehyde and is involved in a pathway regulating tissue differentiation, cell proliferation, and apoptosis. Recently, ALDH1 isoenzymes have been implicated as significant elements in cancer progression. ALDH1 activity has been used as a marker of cancer stem cells, a subpopulation of cancer stem cells with high drug resistance, proliferative potential, and ability to differentiate into multiple cell types. In accordance with this, ALDH1 activity and expression has been shown to correlate with lower survival, increased chemoresistance, and increased chance of relapse in multiple solid cancer types, including breast, ovarian, lung, and colorectal. Despite the clear relevance of ALDH1 enzymes in cancer, the specific roles of individual isoenzymes are unclear. Isoenzyme-selective small molecule modulators of the ALDH1A subfamily would allow the probing of the function of individual isoenzymes in healthy and disease states.

Two ALDH1A1 inhibitors, CM38 and C10, were previously identified in a high-throughput screen. In this study, CM38, an ALDH1A1-selective inhibitor, and CM10, an ALDH1A inhibitor, were characterized using kinetic assays, structural biology, and cell culture experiments. A structure-activity relationship was built for each series, and an X-ray crystallography structure was used to determine the binding mode. These approaches allowed the investigation of the ALDH1A active site and identification of structural features that can be used to design and improve selective modulators of this subfamily. CM38 and CM10 were also tested in a breast cancer cell line to determine their efficacy in a cellular environment. While the CM38 series showed warning signs of potential off-target toxicity, members of the CM10 compound series showed excellent initial characteristics as potential chemical tools. The results of this study may be useful in the design of new chemical tools to delineate the functions of individual ALDH1

isoenzymes in cancer biology, as well as in the development of drugs to selectively target cancer stem cells.

Thomas D. Hurley, PhD., Chair

**Table of Contents**

**List of Tables**

**List of Figures**

**List of Abbreviations**

4-HAP……..4-Hydroxy Acetophenone

4-HNE……..4-Hydroxynonenal

A340……..Absorbance at 340nm

ABC ……..ATP-binding cassette

ACES……N-(2-Acetamido)-2-aminoethanesulfonic acid

ADH……..Alcohol dehydrogenase

ALDH……..Aldehyde dehydrogenase

ALDH$^{bri}$……..Aldehyde dehydrogenase-bright (as determined by Aldefluor Assay)

AQUA……Automated quantitative analysis

ATRA……All-trans retinoic acid

BAA……Boron-dipyrromethene-aminoacetate

BAAA……..Boron-dipyrromethene-aminoacetaldehyde

BES……..N,N-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid

BisTris……2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol

CRABP……Cellular retinoic acid-binding protein

CSC……..Cancer stem cell

CXCL12……..Chemokine (C-X-C Motif) ligand 12

CXCR4……Chemokine (C-X-C Motif) receptor 4

CYP26……Cytochrome P450 26

DEAB……Diethylaminobenzaldehyde

DMEM……Dulbecco's Minimal Essential Medium

DMSO……..Dimethyl sulfoxide

DNA……..Deoxyribonucleic acid

DOPAC……3,4-Dihydroxyphenylacetic acid

DOPAL……3,4-Dihydroxyphenylacetaldehyde

DTT……..Dithiotheritol

EDTA……Ethylene diamine tetra–acetic acid

ERα……Estrogen receptor alpha

FBS……..Fetal bovine serum

GABA……Gamma–aminobutyric acid

GSH……Glutathione

$IC_{50}$……Half maximal inhibitory concentration

IPTG……Isopropyl–β–D–thiogalactopyranoside

$LC_{50}$……Half maximal lethal concentration

logP……Logarithm of partition coefficient

LP……Low pressure

MDA……Malondialdehyde

Min……Minute

MTT……3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide

NAD(P) ……Nicotinamide adenine dinucleotide (phosphate)

PBS……Phosphate buffered saline

PD……Parkinson's disease

PDB……Protein data bank

PDK1……Phosphoinositide-dependent kinase-1

PEG3350……Polyethylene Glycol 3350

PI3K……Phosphoinositide 3-kinase

PPAR……Peroxisome proliferator-activated receptor

qPCR……Quantitative polymerase chain reaction

RA……Retinoic acid

RAR……Retinoic acid receptor

RARE……Retinoic acid response element

RBP……Retinol-binding protein

RDH……Retinol dehydrogenase

ROS……Reactive oxygen species

Rpm……Revolutions per minute

RXR……Retinoid X receptor

SAR……Structure-activity relationship

SDS-PAGE……Sodium dodecyl sulfate polyacrylamide gel electrophoresis

STDEV……Standard deviation

STRA6……Stimulated by retinoic acid 6

TLS……Translational/libration/screw

TY……Tryptone Yeast

UV……Ultraviolet

# I. Introduction

## A. Aldehydes

Aldehydes are a commonly occurring class of organic compounds characterized by the –CHO chemical group (Figure 1A), and involved in an array of biological processes, from DNA damage to cell signalling. Aldehydes can originate from both exogenous and endogenous sources. A common exogenous source is the photochemical oxidation of hydrocarbons, which are either naturally present in the atmosphere or generated by vehicle exhaust (1). Vehicle exhaust is also responsible for the direct emission of aldehydes, such as formaldehyde, acetaldehyde, and acrolein. Non-automobile sources of aldehyde production can also include industrial emissions, forest fires, and coal-based power plants. Human exposure to aldehydes is not limited to atmospheric pollution; aldehydes are also present in food, drugs, and used substances. Acetaldehyde is commonly produced by ethanol metabolism (2). Cigarette smoke contains numerous aldehydes, including acrolein, 4-hydroxynonenal (4-HNE), formaldehyde, and acetaldehyde, that may contribute to lung disease (3). Some drugs are aldehydes or are metabolically converted to them. An example is cyclophosphamide, an anticancer drug used in the treatment of leukemia, breast, and lung cancers (4). Metabolites of cyclophosphamide include aldehydes such as aldophosphamide and acrolein, and this metabolic conversion is necessary for the drug's anticancer effect (4, 5).

Aldehydes can also originate from a multitude of biological processes. Lipid peroxidation, a process linked to oxidative stress, produces aldehydes such as 4-hydroxynonenal (4-HNE) and malondialdehyde (MDA), among other bioactive products (1, 6). Glyceraldehyde-3-phosphate is an intermediate in glycolysis, and can be further enzymatically converted to methylglyoxal (7). Retinaldehyde is a signalling intermediate that plays a key role in the retinoic acid pathway responsible for the regulation differentiation, development, and apoptosis (8).

Although aldehydes can play important roles in biological processes, the total aldehyde burden must be carefully managed due to their bioactive and potentially toxic nature. In contrast to reactive oxygen species, aldehydes are longer lived and can interact with cellular components far from their origin (9). Aldehydes can be reactive toward both nucleic acids and proteins. In the case of nucleic acids, aldehydes can form DNA adducts and cause mutations. For example, acetaldehyde can form adducts to deoxyguanosine, while MDA can form adducts

to deoxyguanosine, deoxyadenosine, and deoxycytidine (10, 11). Through direct damage of the genome, aldehydes can increase mutation rate and may contribute to cancer development. Aldehydes can also cause carbonylation of proteins. Carbonylation is a type of oxidative damage that can lead to loss of function, protein cross-linking or formation of aggregates. Furthermore, heavy carbonylation can lead to toxicity and contribute to cardiovascular or neurodegenerative diseases (1, 12).

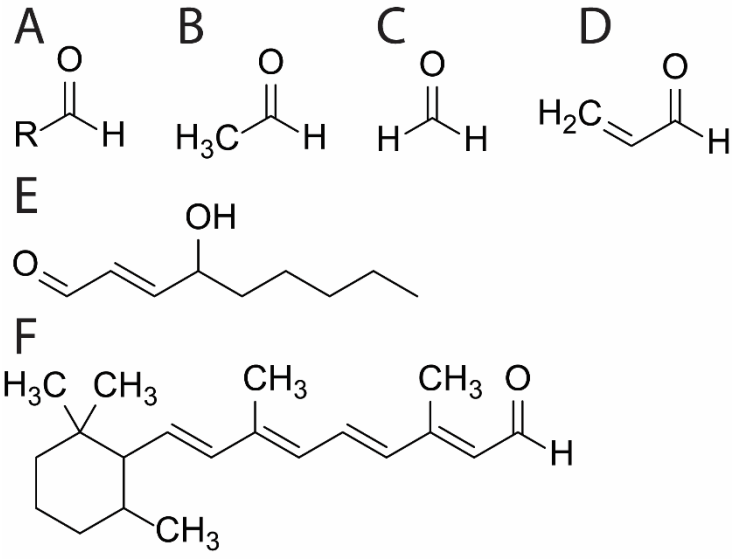



Figure 1: Chemical Structures of Aldehydes
A) General formula of aldehydes B) Acetaldehyde C) Formaldehyde D) Acrolein E) 4-hydroxynonenal F) Retinaldehyde

To control the biological levels of aldehydes, a range of enzymes are capable of metabolizing them. Broadly speaking, aldehydes can be oxidized to carboxylic acids, reduced to alcohols, or conjugated to glutathione (Figure 2) (1, 13). Oxidation is performed mainly by aldehyde dehydrogenases and cytochrome P450s. Aldehyde dehydrogenases are a superfamily of enzymes that are responsible for the irreversible, NAD(P)-dependent oxidation of aldehydes to carboxylic acids (9). Cytochrome P450s are a large family of heme thiolate enzymes responsible for the metabolism of a wide range of chemicals, including toxic ones like aldehydes (14). Aldehydes can also be reduced, which involves conversion to an alcohol that is mainly catalyzed by alpha-keto reductases and alcohol dehydrogenases. Alpha-keto reductases are a superfamily of over one hundred enzymes that generally catalyze carbonyl reduction reactions (13). Alcohol dehydrogenases are a group of enzymes with broad substrate specificity that can

facilitate the interconversion between alcohols and aldehydes; although the reactions are reversible, alcohol dehydrogenases are mainly known to catalyze the oxidation of alcohols to aldehydes as part of alcohol metabolism (15). Aldehydes can also be conjugated to glutathione by Glutathione-S-Transferases to allow excretion as mercapturic acid conjugates (13).

Aldehydes are ubiquitous organic compounds that can function as part of normal biological processes or cause toxicity via DNA and protein damage. Aldehyde dehydrogenases are one of the key groups of enzymes that modulate aldehyde levels to regulate aldehyde-associated biological pathways and reduce the toxic aldehyde burden.

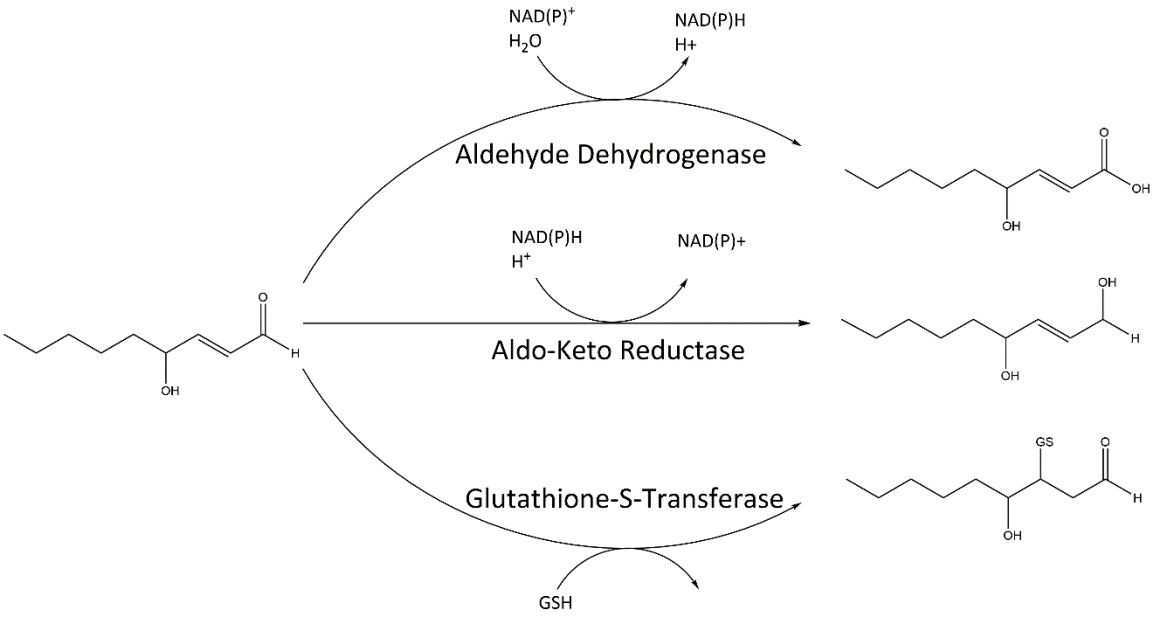



Figure 2: Primary Pathways of Aldehyde Metabolism
Potential routes of 4-HNE Metabolism are shown as an example. A) Oxidation by aldehyde dehydrogenase B) Reduction by Aldo-Keto Reductase C) Glutathione conjugation by Glutathione-S-Transferase

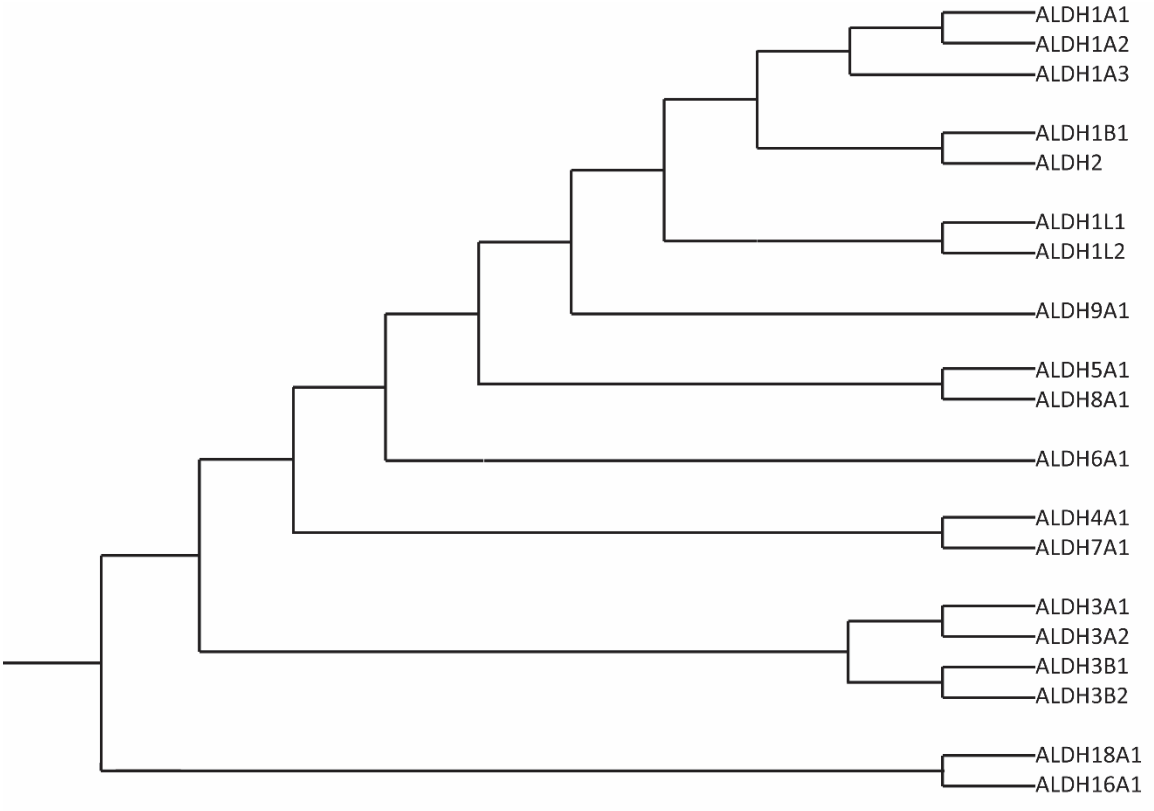



Figure 3: The Phylogenetic Tree of the 19 Currently Known Human Aldehyde Dehydrogenase Enzymes

**B. Aldehyde Dehydrogenase Superfamily**

The human Aldehyde Dehydrogenase superfamily consists of at least 19 genetic loci, the isoenzymes produced from which can vary in substrate specificity, subcellular localization, and tissue-specific expression (Figure 3) (9, 16). The current nomenclature system for the ALDH superfamily, established in 1998, states that the "ALDH" is the root symbol, the next Arabic numeral identifies the family, followed by a letter to identify the subfamily, and, lastly, another Arabic numeral to identify the gene (17). Proteins in the same family are defined as having >40% sequence identity, while members of the same subfamily exhibit >60% sequence identity. The aldehyde dehydrogenases are mainly responsible for the oxidation of both aliphatic and aromatic aldehydes to their respective carboxylic acids. The $NAD(P)^+$-dependent reaction mechanism for aldehyde oxidation is shared by all catalytic ALDH enzymes and occurs in five steps (Figure 4) (18). In the first step, the catalytic residue (Cys302 in mature ALDH2) is activated by water-mediated proton abstraction with the help of a glutamate (Glu268 in ALDH2). The

activated cysteine then performs a nucleophilic attack on the carbonyl carbon of the substrate aldehyde. A thiohemiacetal intermediate is formed, followed by hydride transfer to the $NAD(P)^+$ cofactor. The basic Glutamate activates a water molecule, allowing it to perform a nucleophilic attack on the thioester-enzyme complex. A second tetrahedral intermediate forms and then rearranges to release the carboxylic acid and regenerate the activated Cysteine. The reaction releases the carboxylic acid product and the reduced cofactor NAD(P)H.

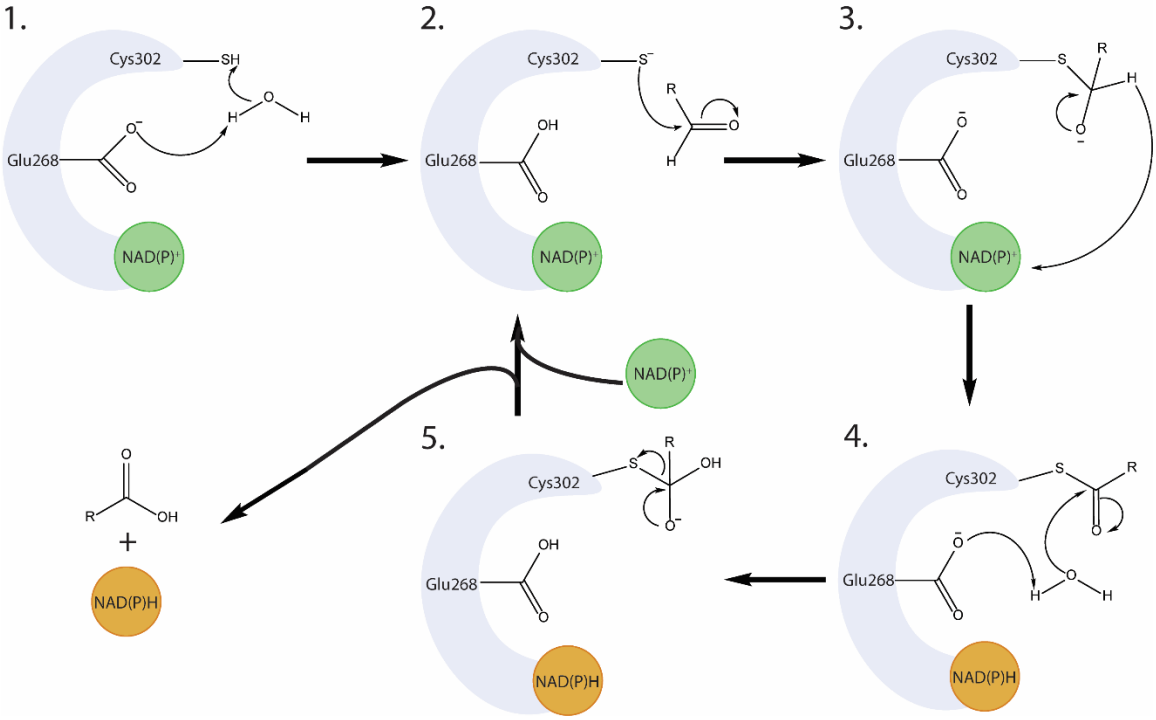



Figure 4: Catalytic Mechanism of Aldehyde Oxidation by Aldehyde Dehydrogenase Enzymes Residue labeling based on ALDH2

The shared catalytic function translates into structural similarities between the ALDH enzymes (Figure 5). Many known ALDH isoenzymes function as oligomers: either dimers or tetramers. Known ALDH structures show that the substrate and cofactor binding sites form a tunnel in order to allow both the aldehyde substrate and the $NAD(P)^+$ cofactor access to the active cysteine residue from opposing sides of the tunnel.

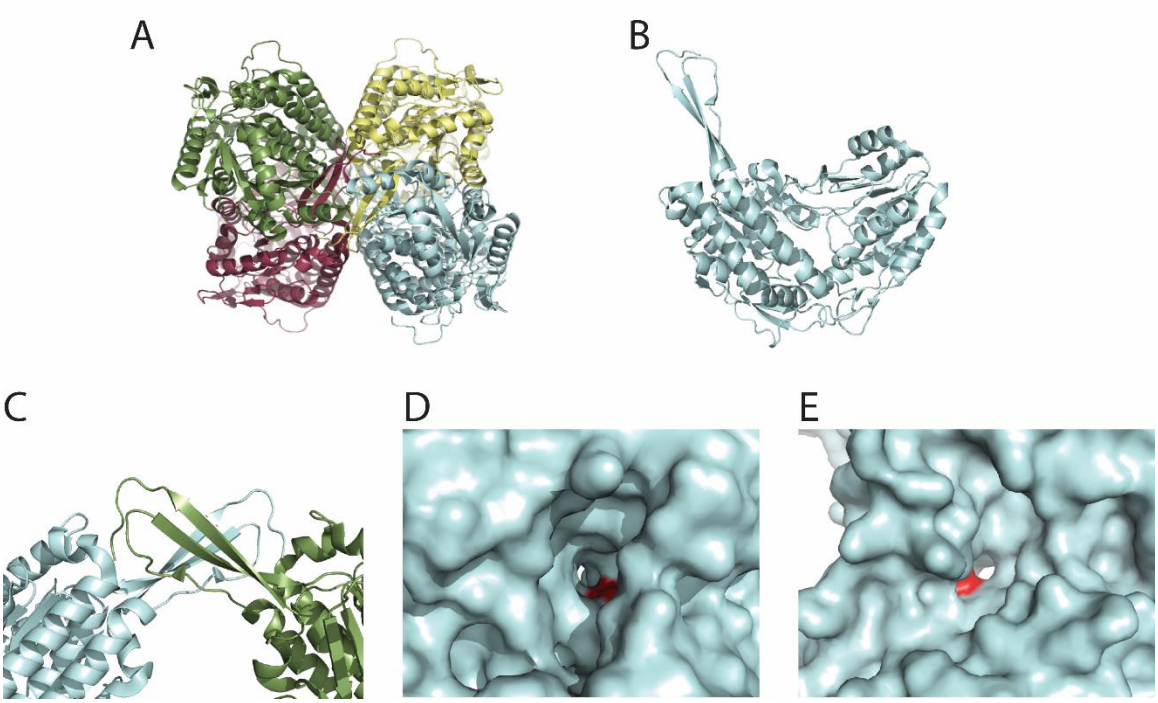



Figure 5: Structural Features of Aldehyde Dehydrogenase Enzymes
ALDH1A1 (PDB ID 4WJ9) is used to show some of the main structural features representative of most Aldehyde Dehydrogenases. A) The ALDH1A1 as a tetramer B) ALDH1A1 monomer unit C) The tetramerization domain of ALDH1A1, shown making intermolecular contacts with another subunit D) Substrate-binding pocket, with the catalytic cysteine highlighted in red E) Cofactor-binding pocket

Although their primary function is in aldehyde metabolism, aldehyde dehydrogenases can also fulfill other catalytic and non-catalytic roles. Several aldehyde dehydrogenase enzymes, namely ALDH1A1, ALDH1L1, ALDH1L2, ALDH2 and ALDH9A1, can also function as esterases and may play important roles in the bio activation of organic nitrites (19, 20). A small number of ALDH enzymes, such as ALDH1A1 and ALDH16A1, act as binding proteins for endogenous and exogenous compounds (19). Notably, ALDH16A1 lacks the catalytic cysteine, making it likely that it has no enzymatic activity and functions primarily as a binder of maspardin and/or other compounds (19, 21).

Some members of the ALDH superfamily play critical roles in the function of the eye. ALDH enzymes act as crystallins, or proteins that allow for the transparency of the lens and cornea. In most mammals, ALDH3A1 is the primary corneal crystallin, while ALDH1A1 is a lens crystallin (22). ALDH3A1 and ALDH1A1 likely play structural roles in eye; however, they also

serve a highly important function in protecting the eye against UV damage. The isoenzymes can directly absorb UV radiation and detoxify UV-induced aldehydes (22). The two can also act as antioxidants, either directly by scavenging free aldehydes, or indirectly by producing NAD(P)H.

**C. ALDH1A Subfamily and the Retinoic Acid Pathway**

The ALDH1A subfamily consists of three isoenzymes: ALDH1A1, ALDH1A2, and ALDH1A3. The three share a unique function of catalyzing the oxidation of retinaldehyde as part of the retinoic acid pathway. The retinoic acid pathway plays a crucial role in the development of many organs and tissues, including those of the central nervous, cardiovascular, and respiratory systems (8, 23). It can also regulate cellular processes, such as proliferation, differentiation, and apoptosis (8).

The retinoic acid pathway is initiated by retinol (Vitamin A) (Figure 6). Retinol cannot be synthesized by mammals, so it is instead absorbed from dietary carotenoids and retinyl esters (23). These compounds are processed into retinol, which is absorbed by mucosal cells, re-esterified and stored as retinyl esters in the liver (24, 25). From the liver, the retinyl esters are once again converted to retinol and circulated to the peripheral tissues while bound to retinol-binding protein 4 (RBP4)(23). Retinol can diffuse into cells, but can also be transported by the RBP4 membrane receptor STRA6 (24, 26). Once in the cell, retinol is bound to the cellular retinol-binding protein 1 (RBP1) (23). Retinol is then converted to retinaldehyde (retinal) by retinol dehydrogenases (RDH) and alcohol dehydrogenases (ADH) (8). Retinal can, in turn, be converted to retinoic acid by members of the ALDH1A subfamily (8, 23). Retinoic acid (RA) can bind to cellular retinoic acid binding protein (CRABP) and be shuttled to the nucleus; alternatively, it can be degraded to oxidized metabolites by CYP26 (8, 24). In the nucleus, RA can associate with retinoic acid receptors (RARs) and retinoic X receptors (RXRs)(27). RARs are activated by all-*trans* and 9-*cis* RA, while RXRs can only be activated by 9-*cis* RA. In the absence of ligands, these receptors form RAR/RXR heterodimers and recruit co-repressor proteins to RA response elements (RARE) (24). This results in the recruitment of histone deacetylases and transcriptional repression of RARE-containing genes. On the other hand, when RAR/RXR heterodimers are associated with retinoic acid ligands, they undergo a conformational change that causes them to recruit co-activating proteins (such as histone acetylates) and activate the target genes (28). Over 500 genes have been identified as potential targets of retinoic acid signalling, including genes involved in the retinoid pathway, such as RARA and CRABP2 (29). The

activation of retinoic acid transcriptional targets is generally thought to promote differentiation and apoptosis, while inhibiting cell proliferation (8). Consistent with this assumption, *all-trans* retinoic acid (ATRA) is used as a treatment in acute promyelocytic leukemia (30).

In contrast to the established role of the RA pathway as an anti-cancer pathway, the therapeutic potential of ATRA has not been validated in solid tumors. (31). In fact, retinoic acids have shown negative results as a treatment in breast and lung cancers (31-33). This discrepancy between tumor types may be attributed, at least partially, to non-classical retinoic acid pathways (Figure 6). Many of these pathways have effects that directly oppose those of their classical counterpart. Ligand-bound RAR may be able to non-genomically activate the PI3K pathway (34). ATRA may be able to directly bind and inhibit protein kinase C (35). In terms of transcriptional modulation, RARs and RXRs can form heterodimers with other receptors, such as estrogen receptor α (ERα) and peroxisome proliferator-activated receptors (PPAR) (31). In these cases, RA can actually activate the transcriptional targets of ERα and PPAR, which can include pro-survival genes (31, 36). Through this mechanism, the RA pathway can activate the transcription of PDK-1, c-MYC, and cyclin D, all of which may potentially lead to cancer progression (31, 36).

The retinoic acid pathway can promote seemingly opposing cellular responses, from differentiation and apoptosis to survival and proliferation. The exact outcome of RA signalling likely depends on the cell and tissue in which it occurs. As a result, while activation of retinoic acid signalling is an effective treatment in some cancer types, inhibition may yet prove a compelling therapeutic approach in other, distinct tumor types.

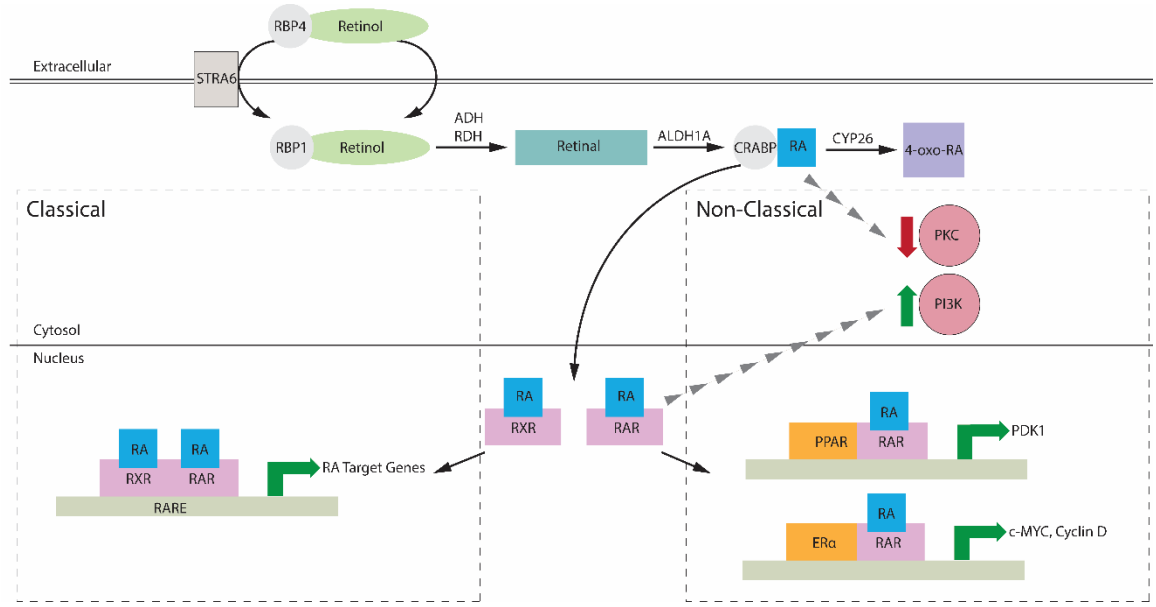



Figure 6: Retinoic Acid Signalling Pathway
Highlighted are the classical and non-classical components of retinoid signalling. Classical RA pathways promote differentiation and apoptosis, while inhibiting cell proliferation. Non-classical pathways generally promote survival and cell growth.

Table 1: Human ALDH1A, ALDH2 and ALDH1B1 Isoenzymes
Adapted from ALDH.org, Koppaka V. et al (2012) and Tomita H. et al (2016) (37-39)

| Isoenzyme | Gene Locus | Sequence Identity to ALDH1A1 | Molecular weight (kDa) | Oligomer | Tissue Expression | Subcellular localization | Known Substrate(s) | PDB Structure ID |
|---|---|---|---|---|---|---|---|---|
| ALDH1A1 | 9q21.13 | 100% | 54.9 | tetramer | Liver, kidney, red blood cells, skeletal muscle, lung, breast, lens, stomach, brain, pancreas, testis, prostate, ovary | Cytosol | Retinal, acetaldehyde, aldophosphamide, lipid-peroxidation derived aldehydes, DOPAL | 4WJ9 (human, apo-enzyme) 4WB9 (human, complex with NADH) |
| ALDH1A2 | 15q22.1 | 73.2% | 56.7 | tetramer | Testis, liver, kidney, lung, brain | Cytosol | Retinal, lipid-peroxidation derived aldehydes | 4X2Q (human, complex with NAD) |
| ALDH1A3 | 15q26.3 | 70.9% | 56.1 | tetramer | Kidney, skeletal muscle, lung, breast, stomach, salivary glands | Cytosol | Retinal, lipid-peroxidation derived aldehydes | N/A |
| ALDH1B1 | 9p11.1 | 64.7% | 57.3 | tetramer | Liver, kidney, heart, skeletal muscle, brain, prostate, lung, testes, placenta | Mitochondria | Lipid-peroxidation derived aldehydes, acetaldehyde | N/A |
| ALDH2 | 12q24.2 | 68.1% | 56.4 | tetramer | Liver, kidney, heart, skeletal muscle, lens, brain, pancreas, prostate, spleen | Mitochondria | Acetaldehyde | 1O04 (Human, complex with NAD) |

**D. ALDH1A1**

ALDH1A1 is a cytosolic homotetramer with broad tissue expression; it is near-ubiquitous in adult human organs, including brain, eyes, testes, kidneys, liver, and lungs (38). It is a well-conserved protein, showing at least 90% amino acid sequence identity in mammals. Because of its high affinity for both all-trans- and 9-cis-retinaldehyde and its high expression in adult tissues, it is considered the predominant retinaldehyde-metabolizing enzyme in adults (37, 38). Its catalytic role extends to other substrates, such as acetaldehyde, a major intermediate in ethanol metabolism. ALDH1A1 executes a role as a detoxification enzyme by eliminating reactive aldehydes such as 4-HNE and MDA; however, these substrates can also inhibit the enzyme (40). ALDH1A1, along with ALDH3A1, can confer resistance to the anti-cancer drug cyclophosphamide by metabolizing its active aldehyde intermediates (41). As mentioned previously, ALDH1A1 is very important in the lens of the eye, as it plays both a structural role as a crystallin and a protective role by minimizing UV-induced oxidative damage (22). It has an additional non-catalytic role as a compound-binding protein. ALDH1A1 can bind androgen, thyroid, and cholesterol, though the significance of this function is unknown (40).

One of the most pathology-relevant functions of ALDH1A1 is the role of the enzyme in midbrain catecholamine (dopamine) neurons. In brain dopaminergic neurons, ALDH1A1 functions as part of a pathway to synthesize and co-release GABA, a key neurotransmitter in the mammalian central nervous system (42). ALDH1A1 also catalyzes the conversion of 3,4-dihydroxyphenylacetaldehyde (DOPAL), a dopamine metabolite, to 3,4-Dihydroxyphenylacetic acid (DOPAC) (43). DOPAL is toxic and is thought to be at least partially responsible for the death of dopaminergic neurons in the striatum (43). The death of these cells is thought to be the cause of nigral depigmentation in the midbrain—one of the primary characteristics of Parkinson's disease. Thus, ALDH1A1 has a key link to the pathogenesis of one of the most clinically significant neurodegenerative diseases. Consistent with this link, the DOPAC:DOPAL ratio is lowered both in the striatum of ALDH1A1,2 knockout mice and in the putamen of PD patients (44). This evidence suggests that reduced ALDH1A1 activity (such as through inhibition by 4-HNE and MDA) may contribute to the pathogenesis of Parkinson's disease.

The role of ALDH1A1 in mammals has been studied with knockout mouse models. As expected based on the protective role of ALDH1A1 in the eye, ALDH1A1(-/-) mice develop cataracts with age (45). Notably, however, the ALDH1A1(-/-) mice are also resistant to weight

gain from a high-fat diet and have reduced adipogenesis (46). Furthermore, ALDH1A1 has also been shown to regulate fat plasticity and the balance between white adipose tissue, which is linked to obesity, and brown adipose tissue, which is linked to leanness (47). ALDH1A1 is mainly expressed in white adipose tissue, and deficiency of ALDH1A1 promotes a thermogenic program in these cells (47). Aside from altered fat metabolism, ALDH1A1(-/-) mice also have reduced hepatic gluconeogenesis, decreased fasting glucose levels, and increased fatty acid oxidation (48). In accordance with this, model diabetic rats present increased hepatic expression of ALDH1A1 (49). Together, these results indicate that the retinoid pathway and ALDH1A1 are key regulators of both fat and glucose metabolism, implicating ALDH1A1 as a potential therapeutic target in obesity and diabetes.

**E. ALDH1A2**

ALDH1A2 is a cytosolic homotetramer that is expressed in the testes, liver, kidney, lung, and brain (9). ALDH1A2 has high selectivity for *all-trans* retinaldehyde over its other substrates (50). Like ALDH1A1, its secondary substrates include lipid peroxidation products like decanal and octanal, allowing it to also serve as a detoxifying enzyme (50).

Animal models suggest that ALDH1A2 plays a key role in the early development of several organs. ALDH1A2 (-/-) mouse models die in embryogenesis due to impaired morphogenesis of the heart (51). When rescued through supplementation with RA, the mice still show defects in the development of the branchial arches, hindbrain neural patterning and the enteric nervous system (51). ALDH1A2 has also been implicated in the development of the retina, forebrain, and pancreas (52-54). As an extension of its role in development, ALDH1A2 also facilitates tissue regeneration in some species. ALDH1A2 was shown to be upregulated following fin or heart injury in zebrafish, and ALDH1A2-catalyzed production of RA was demonstrated to be necessary for regeneration (55, 56).

**F. ALDH1A3**

ALDH1A3 is a cytosolic tetramer expressed in stomach, breast, kidney, and lung tissues, among others (9). Like ALDH1A2, it is preferential for *all-trans* retinaldehyde, but can also oxidize *9-cis* retinaldehyde, octanal, and decanal (9, 57).

ALDH1A3 (-/-) mice die at birth due a blockage of the nasal passage; however, unlike ALDH1A2 knockouts, the mice show no obvious deformities or health concerns until birth (58,

59). The mice can be rescued by treatment with retinoic acid, suggesting that the primary function of ALDH1A3 in development is tissue-specific retinoic acid production (58). Knockout models have also implicated ALDH1A3 in the development of the eye, forebrain, and kidneys (60-62). In the development of the eye, ALDH1A3 and ALDH1A1 are thought to have overlapping roles, as ALDH1A1(-/-) ALDH1A3(-/-) double knockout mice demonstrated exacerbated defects compared to either single knockout (60, 63) .

**G. Cancer Stem Cells**

A growing paradigm in cancer research is the theory of cancer stem cells (CSCs), or a subpopulation of stem-like cells that fulfill a critical role in tumor development. The basic properties of cancer stem cells include cell renewal, ability to differentiate into multiple cell types, extensive proliferative potential, and resistance to classical chemotherapy and radiotherapy (64, 65). Cancer stem cells can be identified by surface cell markers, which vary with tumor type. Commonly used markers include CD24, CD44, and CD133 (65). Other properties, such as sphere-forming capacity or ALDH1 activity, can also be used as markers (65). ALDH1 enzymatic activity stands out as a universal marker in CSCs and progenitor cells across a multitude of cell types; accordingly, this property been successfully employed to isolate CSCs from head and neck squamous cell carcinomas, as well as lung, prostate, pancreas, and breast cancers (66).

Cancer stem cells are thought to be responsible for many of the properties that make cancers difficult to treat, such as drug resistance, quiescence following therapy, tumor recurrence, and metastasis (67). It is becoming increasingly apparent that many therapy options aiming to simply reduce the size of the tumor or target the most numerous cancer cell type do not effectively prevent recurrence or metastasis, as they do not affect CSCs (Figure 7) (67). Tumors are heterogeneous, and should be treated as such in order to develop effective and permanent cancer treatments. One possible approach would be to combine classical therapies with those that selectively target the cancer stem subpopulation.

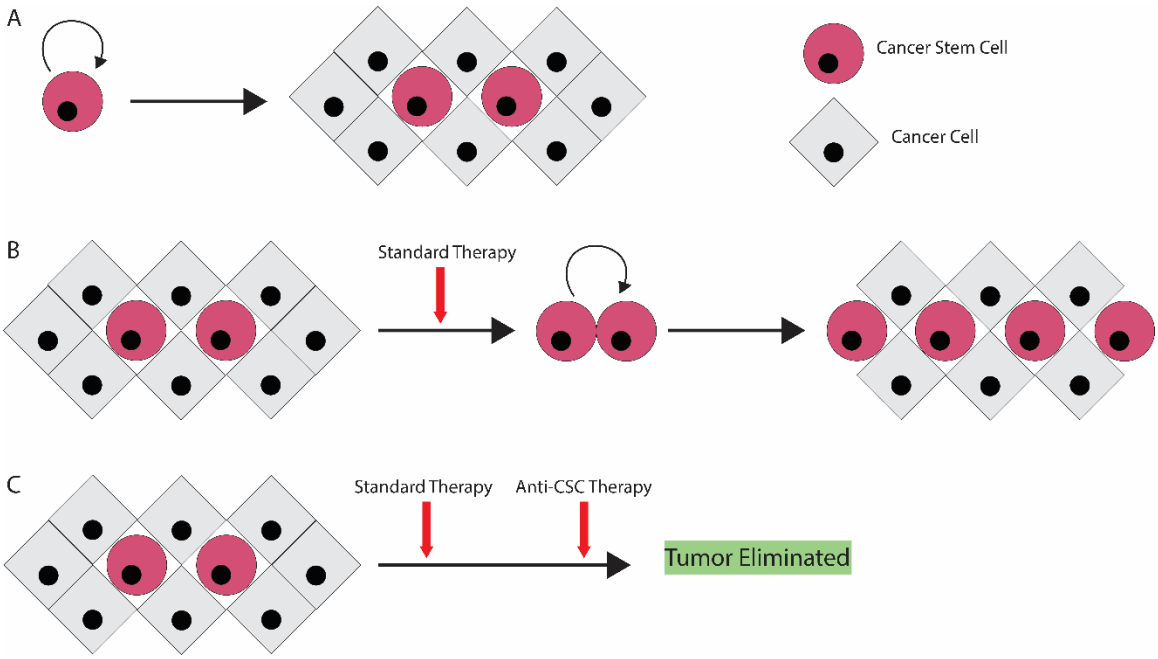



Figure 7: The Cancer Stem Cell Model
A) Asymmetric and symmetric cell division by CSCs leads to tumor growth B) CSCs are resistant to standard therapies, leading to tumor relapse C) Combinatorial therapies that target both the tumor bulk and the CSCs are a promising approach for complete elimination of the tumor

A number of strategies are being developed to target cancer stem cells. Common approaches include targeting the cell surface markers, key signalling pathways, ABC transporters, and the tumor microenvironment (65, 68). Conjugated monoclonal antibodies are a popular method of targeting CSC surface markers like CD133, CD33, CD44, and CD47 (65, 68). One such method involved anti-CD133 antibody-conjugated nanoparticles loaded with the anti-cancer agent paclitaxel. The strategy was effective in reducing the CSC population and inhibiting tumor growth (69). A number of signalling pathways have been identified as having a significant role in CSCs, including NF-κB, Notch, Hedgehog, PI3K/Akt, Wnt/b-catenin, JAK/STAT, and Hippo-YAP/TAZ (65, 67, 68). Antibodies and small-molecule compounds that inhibit these pathways are being developed and tested, including modified combinatorial strategies of targeting multiple pathways at once (65, 70). Features of the microenvironment that help support CSCs, such as secreted factors, microvasculature, and weakly acidic pH, are also potentially exploitable in treatment (68). An example of a secreted factor is CXCL12, which activates the CXCR4 receptor in CSCs and promotes close contact to the tumor stroma, proliferation, survival, and invasion (71). Plerixafor is a CXCR4 antagonist that was shown to promote mobilization of the stem cells

into blood and render them more vulnerable to drug therapy (72). ATP-binding cassette (ABC) transporters are efflux pumps that are thought to contribute to multi-drug resistance (73). A variety of ABC transporter inhibitors are currently being investigated as sensitizing agents to reduce the resistance to other drug types (65). Although ALDH1 enzymes are known to be markers of CSCs in many cancer types, they are relatively unexplored as drug targets to selectively eliminate CSCs.

**H. Aldefluor Assay**

Currently popular approaches for identifying cancer stem cells based on ALDH activity include the Aldefluor Assay and immunohistochemistry. The Aldefluor assay is a commercially available (StemCell Technologies, Vancouver, BC, Canada) assay that can measure the functional activity of ALDH in living cells (74, 75). Together with a fluorescence-activated cell sorter, the Aldefluor assay can be used to sort and isolate cells with high ALDH activity ($ALDH^{bri}$ cells) (74, 76). The assay works by providing a fluorescent cell-permeable ALDH substrate BODIPY-aminoacetaldehyde (BAAA), that can be converted to BODIPY-aminoacetate (BAA) by active ALDH enzymes (Figure 8) (75). BAA is not membrane-permeable, so it is retained by cells with active ALDH and labels them with a fluorescent signal (74, 75). DEAB is a ALDH inhibitor that is provided with the assay for use in a negative control (74). There are many advantages to the Aldefluor assay, as it is non-toxic and does not require UV light for excitation (75). However, the major disadvantage of the assay is that it is not isoenzyme-selective; neither the BAAA substrate, nor the DEAB inhibitor are specific for any one isoenzyme in the superfamily (38, 76, 77). Although the Aldefluor signal is often attributed to ALDH1A1, it could also be produced by other isoenzymes, such as ALDH1A2 and ALDH1A3 (38, 76). Isoenzyme-selective assays and inhibitors are still sorely needed in order to delineate the functions of specific ALDH1 isoenzymes in cancer biology.

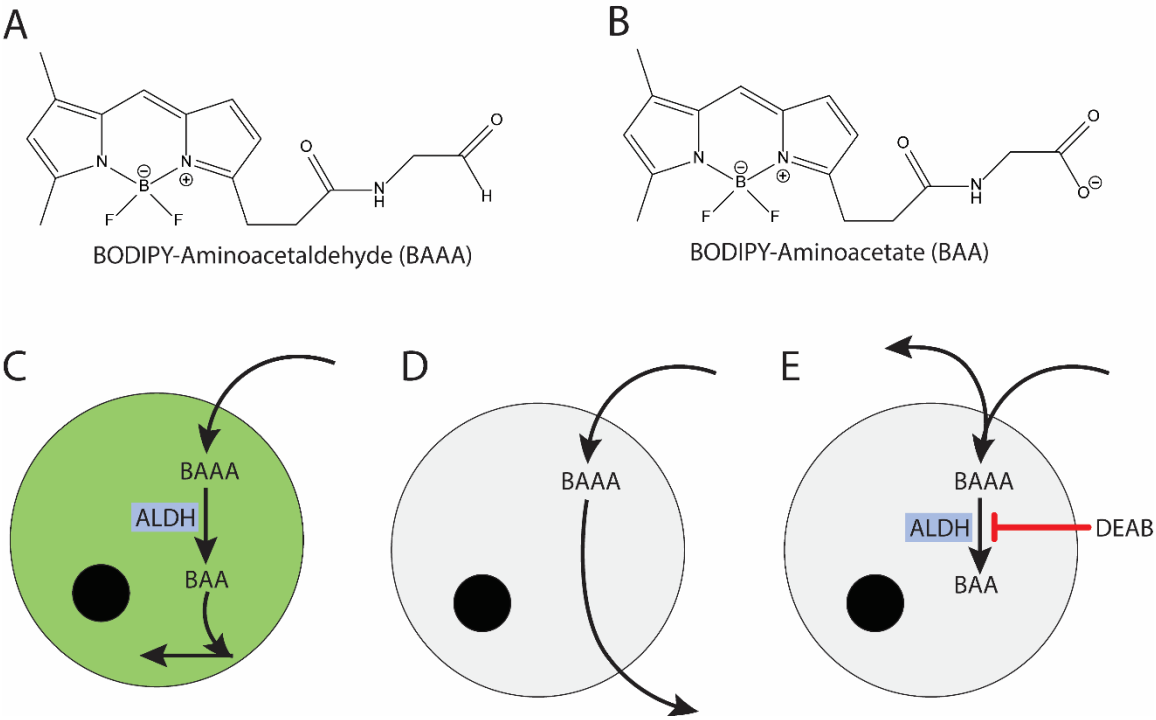



Figure 8: Aldefluor Assay
A) Structure of BODIPY-Aminoacetaldehyde (BAAA), the Aldefluor substrate B) Structure of BODIPY-Aminoacetate (BAA), the Aldefluor product C) In a cell with active ALDH, BAAA is converted to BAA, labeling the cell with a fluorescent signal D) In a cell without active ALDH, BAAA does not undergo conversion and is free to diffuse out of the cell E) In the negative control, the DEAB inhibitor prevents conversion of BAAA to BAA, thereby preventing the labeling of the cell

**I. ALDH1 Family in Breast Cancer**

The expression of the ALDH1 family has been suggested as a marker of breast cancer stem cells, and its impact on clinical prognosis has been thoroughly studied. A meta-analysis of 15 publications on breast cancer patients revealed that high expression of ALDH1A1 (as determined by immunohistochemistry) correlated with a poor prognosis, as well as unfavourable tumor characteristics like larger tumor size, greater possibility of lymph node metastasis, higher HER2 expression, and lower expression of estrogen receptor (78). A further study of 653 breast carcinoma specimens largely corroborated these results, but also suggested that ALDH1 had a significant impact on the prognosis of luminal-type cancer but not triple-negative or HER2-type (79). In the same paper, ALDH1 expression was shown to correlate to chemoresistance, even in triple-negative breast cancer (79). Furthermore, ALDH1 expression has

been suggested to be a good predictor of relapse, and the proportion of cells with high ALDH1 expression was shown to increase after relapse (80). These results are further reinforced by cell culture experiments, in which $ALDH^{bri}$ cells sorted by Aldefluor assay displayed higher invasive potential in matrigel assays, increased mammosphere-generation, and increased tumor initiation and metastasis development in mice (81, 82). All these results are consistent with the use of ALDH1 expression as a cancer stem cell marker both in cell culture and in the clinic.

While the characteristics of ALDH1-expressing cells are usually attributed to ALDH1A1, there is some evidence that ALDH1A3 also plays a significant role in this disease. It has been suggested that ALDH1A3 was primarily responsible for the $ALDH^{bri}$ signal in multiple breast cell lines, and was a better predictor of tumor progression and metastatic potential (83, 84). A study of 176 breast cancer specimens also suggested that ALDH1A3 expression had a more significant impact on survival than ALDH1A1 (84).

Although there is a strong body of evidence supporting the prognostic value of ALDH1 in breast cancer, there are also a number of studies that refute this assessment. In a study of 321 node-negative and 318 node-positive breast cancer patients, ALDH1 expression, as determined by Automated Quantitative Analysis (AQUA), was not found to significantly affect outcome (85). Another paper reported that in a study of 245 invasive cancer specimens, ALDH1 expression did not correlate with tumor size, tumor grade, lymphovascular invasion, or survival (86). In smaller cohorts of 45 and 40 triple-negative specimens, the study actually found that high stromal expression of ALDH1 correlated with increased disease-free survival (86).

There are many possible explanations for the observed discrepancies in literature, including differences in cancer type, specific tissue (stromal vs. tumor), and detection method (Immunohistochemistry, Aldefluor assay, Automated Quantitative Analysis). Given that ALDH1A1 and ALDH1A3, as well as possibly other isoenzymes, may play a role, differences in the isoenzyme selectivity of antibodies used for immunohistochemistry could seriously sway results. Although the prognostic value of ALDH1 expression is still under debate, the idea that ALDH1 expression is a marker of CSCs and an indicator of poor prognosis seems to currently have more supporting evidence (9, 38).

**J. ALDH1 Family in Ovarian Cancer**

As in the case of breast cancer, the prognostic value of ALDH1 expression in ovarian cancer in controversial. In two studies of 84 and 100 ovarian cancer specimens, high ALDH1 expression correlated with significantly worse survival (87, 88). In a study of 55 matched samples taken before and after taxane/platinum-based therapy, ALDH1 expression was increased significantly (median change of three-fold) in non-regressing tumors (89). Furthermore, the presence of these ALDH1-high cell subpopulations increased the risk of death 4.18 times, suggesting that ALDH1-high CSCs are responsible for both chemoresistance and a more aggressive phenotype (89). On the other hand, larger studies of 442 and 248 ovarian carcinoma patients found that high ALDH1 expression correlated with higher overall and disease-free survival (90, 91).

Despite the questionable role of ALDH1 in clinical cases, ALDH1 expression has consistently been shown to correlate with stemness properties in cell culture experiments. $ALDH^{bri}$ cells of multiple ovarian cancer cell lines showed higher spheroid-forming ability, invasiveness, tumorigenic potential in mice, and chemoresistance (87, 92). Another study showed that ALDH1A1 was upregulated in spheroids compared to cells grown in monolayer and that the use of a ALDH1A1-selective inhibitor reduced spheroid formation (93).

Overall, despite a strong correlation with stemness characteristic in the lab, the prognostic value of ALDH1 expression in ovarian cancer remains under question. As in breast cancer, the high variation between studies could be due to differences in cancer subtype, detection method, or the selectivity of the antibody used for immunohistochemistry.

**K. ALDH1 Family in Lung Cancer**

In lung cancer, high populations of ALDH1-expressing are generally thought to be associated with a worse prognosis. A meta-analysis of 14 publications comprising 1926 cases found that ALDH1 expression was correlated with lymph node metastasis, higher tumor stage, and decreased overall survival (94). Although the prognostic value of ALDH1 in lung cancer is generally accepted by the scientific community, it is not without challengers. For example, a study of two cohorts of 134 and 296 patients with non-small lung cancer found that the absence of ALDH1 expression (as determined by AQUA analysis) was correlated with a worse prognosis (95).

Evidence in cell culture experiments further confirms the idea that ALDH1 expression is a marker of cancer stem cells in lung cancer. $ALDH^{bri}$ cells have been found to have higher proliferation rates, colony formation ability, chemoresistance, and tumor-forming ability in nude mice (96, 97). The current consensus is that ALDH1 expression is an effective marker of cancer cell stemness in lung cancer, as supported by strong evidence *in vitro* and in the clinic.

**L. ALDH1 Family in Colorectal Cancer**

Colorectal cancer is another cancer type in which ALDH1 expression may have value as a cancer stem cell marker. A meta-analysis of 9 publications encompassing 1203 patient samples found that high ALDH1 expression was correlated with higher tumor stage and decreased disease-free and overall survival (98). On the other hand, a number of other studies have found the opposite result. A study of 1287 patient samples found no correlation between ALDH1 expression and survival time (99). Another paper looking at 659 colon cancer samples and 338 rectal cancer samples found no correlation between cytoplasmic ALDH1A1 expression and prognosis (100). However, the study did find that in a small subset of samples that showed nuclear expression of ALDH1A1, survival time was significantly reduced (100). This may hint at a yet-unexplored nuclear role of ALDH1A1 in colorectal cancer tumorigenesis.

Despite the controversy in the clinical setting, ALDH1 expression still appears to be a marker of cancer stemness in cell culture. For example, $ALDH^{bri}$ cells were shown to have significantly higher tumor-initiating ability in nude mice (101). The prognostic value of ALDH1 expression in colorectal cancer is still under debate and has no clear consensus. As before, variation in tumor subtype of experimental methods could be responsible for the high degree of disagreement observed on this topic.

**M. A Note on the Inconsistencies in Clinical Studies**

Across all the cancer types discussed, the use of ALDH1 expression as a prognostic factor in the clinic is somewhat controversial due to variability between different studies. At the same time, ALDH1 expression as measured by Aldefluor assay in cell culture experiments consistently correlates with stemness characteristics. One plausible explanation for the discrepancy is that cells in culture are inherently different from those *in vivo*. Another very likely explanation is the difference in experimental approach between the Aldefluor assay and immunohistochemistry. The Aldefluor assay is applied consistently across all studies, whereas different antibodies can

lead to a large variance between different immunohistochemistry experiments. Additionally, the Aldefluor assay measures activity rather than enzyme concentration. Furthermore, the Aldefluor assay provides a relatively non-selective substrate, thus allowing a number of isoenzymes to contribute to the signal. Since the identity of all ALDH isoenzymes that can contribute to stemness characteristics is not certain, the non-selective nature of the assay may be particularly important in some cancer types. An immunohistochemistry approach can lead to variability as different antibodies, or even antibody lots, can have different isoenzyme selectivities. Without rigorous testing, the relative contribution of any isoenzyme to an immunohistochemistry signal is both uncertain and inconsistent. More regularity in antibody use would likely improve the consistency of clinical studies. Ideally, an established protocol for antibody validation and verification of isoenzyme selectivity should be used for every antibody.

**N. Potential Functional Roles of ALDH1 Enzymes in Cancer**

Despite a strong body of evidence substantiating the role of the ALDH1 family in cancer progression, the mechanism of its involvement is largely unknown. Nevertheless, based on established properties of ALDH1 enzymes, some functions that are likely to contribute to cancer biology and stemness can be discerned.

The ability of ALDH1A isoenzyme to produce retinoic acid is one such function. As discussed in detail earlier, ALDH1A1, ALDH1A2, and ALDH1A3 regulate the retinaldehyde oxidation step in the retinoic acid pathway. The retinoic acid pathway can feed into the classical and non-classical retinoid responses. In particular, the non-classical retinoic acid pathway can lead to activation of PI3K and upregulation of PDK1, c-MYC, and Cyclin D, all of which have known links to oncogenesis (Figure 6)(31). As such, depending on the relative expression of RAR, RXR, PPAR, and ERα receptors, increased activation of retinoic acid signalling may contribute to cancer progression in some cell types.

ALDH1 enzymes have also been known to contribute to drug resistance, a key property of cancer stem cells. ALDH1A1 is known to oxidize the active metabolites of the anti-cancer drug cyclophosphamide and contribute to resistance (41). ALDH isoenzymes may also protect cells from the oxidative stress caused by radiotherapy or some anti-neoplastic agents (31). ALDH1A1 has also been proposed to contribute to drug resistance by regulating the cell cycle. An ALDH1A1 knockdown in ovarian cancer cells showed increased expression of KLF4 and

decreased expression of p21 (92). As a result, ALDH1A1-knockdown cells accumulated in the S and G2 phases of the cell cycle, which is likely to render them more vulnerable to therapy (92).

ALDH1 enzymes may also contribute to the cancer stem cell phenotype by reducing the levels of reactive oxygen species (ROS). Cancer stem cells are known to have lower levels of ROS, and maintaining this low level is necessary for the CSC phenotype (102, 103). ALDH1 enzymes are known to be able to fulfill an antioxidant role by scavenging reactive aldehydes and producing NAD(P)H (22). Therefore, high ALDH1 expression may be necessary to protect CSCs from ROS-induced DNA damage and cell death.

**O. High-Throughput Esterase Screen for ALDH1A1-Selective Inhibitors**

A high throughput screen was previously conducted to identify selective modulators of ALDH1A1 (104). The screen was successful and achieved a Z-score of 0.67. After two rounds of screening, 256 modulators of ALDH1A1 were identified. After further testing, 30 compounds were found that selectively inhibited ALDH1A1 compared to ALDH2 and ALDH3A1. The screen showed that it was possible to develop isoenzyme-selective inhibitors by making use of the structural differences between ALDH active sites. The screen also provided a number of inhibitors that could act as lead compounds for further development of isoenzyme-selective inhibitors of the ALDH1A subfamily.

**P. Hypothesis and Approach**

Members of the ALDH1 family have a clear role in cancer pathogenesis. High expression of ALDH1 has been linked to cancer stem cell characteristics like high proliferation, invasiveness, chemoresistance, and metastatic potential. Some studies have also linked ALDH1 expression in tumors to poor prognosis in breast, ovarian, lung, and colorectal cancers, among others. Despite mounting evidence, the specific contribution of each isoenzyme in the family to cancer biology remains unknown. Furthermore, the complete mechanism of ALDH1 activity in cancer is equally obscure. Isoenzyme-selective inhibitors of the ALDH1 family would allow studies to delineate the roles of select ALDH1 isoenzymes in each cancer type. In addition, inhibitors would allow researchers to verify the importance of the ALDH1 enzymatic activity and retinoic acid production while keeping any additional activities unaffected.

It was hypothesized that the ALDH1A1 inhibitors previously identified in the high-throughput screen could be used as lead compounds to develop isoenzyme-selective inhibitors

of the ALDH1A subfamily. The development of these compounds would help map out useful structural features of the ALDH1A active site and aid the structure-based drug design targeting these enzymes in the future. These compounds would also have immediate use in the identification of the roles of ALDH1A1, ALDH1A2, and ALDH1A3 in cancer biology. Beyond that, these compounds could be used in drug development in order to target in cancer stem cells in the clinic. Additionally, due to the isoenzyme's role in other diseases, ALDH1A1-selective modulators could be applied in the treatment of obesity, diabetes, and neurodegenerative disease.

The approach utilized here consisted of kinetic measurements, structural biology, and cell culture experiments. A structure-activity relationship was developed for each hit compound by analyzing the selectivity and potency of each analogue in *in vitro* kinetic assays. The binding mode of select compounds was identified by using X-ray crystallography. Cell culture experiments were used to evaluate the ability of select compounds to impair the proliferation of a breast cancer cell line.

## II. Materials and Methods

### A. Materials

All chemical reagents used for protein purification, X-ray crystallography, and enzyme kinetics were purchased from Sigma Aldrich (St. Louis, MO), unless otherwise specified. All reagents used for propagation of and experiments involving mammalian cell lines were purchased from Corning Inc. (Corning, NY), unless otherwise specified. Small molecule compounds were purchased from ChemDiv Corporation (San Diego, CA) and ChemBridge Corporation (Chicago, IL).

### B. Methods

All measurements for enzyme activity assays and cell viability assays were taken in triplicate unless stated otherwise.

#### 1. Purification of ALDH Isoenzymes

Identical protocols were used for the purification of ALDH1A1, ALDH1A2, ALDH1A3, and ALDH2. Full cDNAs of ALDH isoenzymes, previously subcloned into the pT7-7 vector, were used to transform BL21 (DE3) cells. Transformed cells were plated on Agar plates with 100 µg/mL Ampicillin and grown overnight. A single colony was grown overnight in 20 mL TY buffer containing 100 µg/mL Ampicillin (TY-amp). The overnight culture was diluted to 100 mL with TY-Amp and grown for 3 more hours. 10 mL of this starter culture was then added to each of 8 1 L cultures of TY-Amp and grown at 37 °C, with 200 rpm shaking, until an optical density of 0.6-0.8 at 600 nm was reached. Expression of the ALDH enzyme was induced by the addition of IPTG (GoldBio, St. Louis, MO) to a concentration of 100 µM. Induction was conducted overnight at 16 °C and 160 rpm. Cells were spun down at 5000 rpm, 10 min, 4 °C (Beckman Avanti J-25, JA-10 Rotor), and growth medium was decanted. Cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C. Frozen pellets were thawed slowly and suspended in cold lysis buffer (10 mM $NaH_2PO_4$, 2 mM EDTA, 1 mM benzamidine, 1 mM DTT, pH 7.0). All buffers used in enzyme purification were sparged with helium to reduce oxygen-dependent cysteine oxidation. DTT (GoldBio, St. Louis, MO) was added to buffers immediately before use. Cells were lysed by three consecutive passages through a microfluidizer (Microfluidics 110L). Lysate was pelletized by centrifugation at 35000 rpm, 45 min, 4 °C (Beckman Coulter Optima L-90K Ultracentrifuge, Ti-45

Rotor). The supernatant was dialyzed twice against 4 L of Lysis buffer at 4 °C, first for 4 hr and then overnight. The DEAE column was equilibrated with 100 mL of lysis buffer using the BioLogic LP chromatography system (BioRad, Hercules, CA). The clarified lysate was loaded and the column was washed to baseline with 300 mL of lysis buffer. Elution was performed by passing 300 mL of 0-100 % elution buffer (lysis buffer with additional 250 mM NaCl) and collecting 6-8 fractions. The presence and purity of the target enzyme in the collected fractions was examined by running samples on a 10 % SDS-PAGE gel and staining with Coomassie dye. Fractions containing the target enzyme were pooled for the next step. The 4-HAP affinity column was equilibrated with 120 mL of HAP buffer (20 mM $NaH_2PO_4$, 1 mM EDTA, 50 mM NaCl, 1 mM DTT, pH 7.5). The DEAE pool was loaded and the column washed to baseline with 200 mL of HAP buffer. Elution was performed with 100 mL of p-HAP elution buffer (10 mM p-HAP in HAP buffer). Elution was collected as one fraction and a BioRad protein assay (BioRad Laboratories, Hercules, CA), was used to observe the completion of protein elution. The HAP eluate was dialyzed three times against 4 L of storage buffer (10 mM Aces, 1 mM DTT, pH 6.6), with each dialysis being 4 hours to overnight at 4 °C. The protein was concentrated to 5-20 mg/mL using Amicon Ultra Centrifuge Devices (Millipore, Bedford, MA). Protein concentration was determined using the BioRad protein assay, and specific activity was determined by measuring the production of NADH (molar extinction coefficient of 6220 $M^{-1}$ $cm^{-1}$ ) in a solution containing 200 µM propionaldehyde and 1.5 mM $NAD^+$ on a spectrophotometer. The protein was then aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C. ALDH1B1, ALDH1L1, ALDH3A1, ALDH4A1, and ALDH5A1 were purified as described previously (105-107). Some enzyme stocks used were purified and provided by Cameron Buchman, Dr. Cynthia Morgan, Lanmin Zhai, and Dr. Bibek Parajuli.

**2. Aldehyde Dehydrogenase Activity Assays**

The activity rates of aldehyde dehydrogenase isoenzymes and their inhibition by compounds were determined by a spectrophotometric assay, in which the production of NADH was measured at 340 nm (molar extinction coefficient of 6220 $M^{-1}$ $cm^{-1}$). (Figure 9).

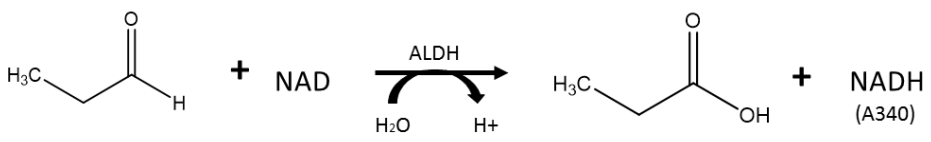


Figure 9: Reaction Used to Measure Activity Rate of Aldehyde Dehydrogenase Enzymes

For ALDH1A1, ALDH1A2, ALDH1A3, ALDH2, and ALDH1L1, 30 mM BES pH 7.5 was used as a buffer; for ALDH3A1, ALDH4A1, and ALDH5A1 100 mM $NaH_2PO_4$ pH 7.5 was used instead. Propionaldehyde was used as the substrate for all isoenzymes except ALDH3A1, for which benzaldehyde was used. Assay reagent concentrations were set up as listed in Table 2, in order to achieve saturation of $NAD^+$ and substrate for each enzyme, but still allow measurement of the reaction at initial-rate conditions.

Table 2: Aldehyde Dehydrogenase Activity Assay Reagent Concentrations

| Enzyme Used | Concentration of NAD (µM) | Concentration of substrate (µM) | Concentration of Enzyme (nM) |
|---|---|---|---|
| ALDH1A1 | 200 | 100 | 400 |
| ALDH1A2 | 200 | 100 | 200 |
| ALDH1A3 | 200 | 100 | 600 |
| ALDH2 | 200 | 100 | 100 |
| ALDH1B1 | 500 | 200 | 150 |
| ALDH3A1 | 300 | 300 | 20 |
| ALDH4A1 | 1500 | 20000 | 100 |
| ALDH5A1 | 1500 | 2000 | 100 |
| ALDH1L1 | 500 | 4000 | 200 |

Enzyme was incubated with compound for 2 minutes before substrate was added to initiate the reaction. Assays were set up in 96-well flat-bottom plates and measured on a Spectramax 340PC plate reader (Molecular Devices, Sunnyvale, CA). Readings were taken every 10 seconds for 5 minutes. Initially, all compounds were screened at 20 µM. All inhibition assays included 2 % DMSO with compound or DMSO alone as control. Activity was calculated as a ratio to control.

$$Rate\ ratio\ to\ control = \frac{Rate\ in\ the\ presence\ of\ compound}{Rate\ in\ the\ absence\ of\ compound}$$

For the CM38 series, compounds were screened with ALDH1A1 first and those that inhibited more than 50 % were screened with ALDH1A2, ALDH1A3, ALDH2, and ALDH1B1. For the CM010 series, all compounds were screened with ALDH1A1, ALDH1A2, ALDH1A3, ALDH2, and ALDH1B1. IC50s were collected in cases where a compound inhibited an isoenzyme more

than 60 % compared to DMSO control. For $IC_{50}$s, a range of concentrations composed of at least 8 points was used. Measurements were fit to a four-parameter $EC_{50}$ equation and graphed using SigmaPlot V12.3 (Systat, San Jose, CA).

**3. Steady-State Kinetics**

Steady-state kinetics were performed in order to find the mode of inhibition of the compounds. In order to determine the appropriate reaction parameters, $K_m$ values were experimentally derived for ALDH1A1 with respect to acetaldehyde and $NAD^+$. To determine the $K_m$ with respect to acetaldehyde, an activity assay was set up with 200 µM ALDH1A1 and a saturating $NAD^+$ concentration of 1 mM, while varying the concentration of acetaldehyde from 1 µM to 1 mM. Similarly, to determine the $K_m$ with respect to $NAD^+$, an assay was set up with 200 µM ALDH1A1 and a saturating acetaldehyde concentration of 500 mM, while varying the concentration of $NAD^+$ from 5 µM to 1 mM. Reactions were initiated with the addition of substrate and activity was measured at 340 nm on a Cary 50-Bio UV-Visible Spectrophotometer (Agilent, Santa Clara, CA) using 10 second intervals for 2 minutes. Results were modeled and graphed using the single-substrate substrate inhibition module of SigmaPlot V12.3. The $K_m$ values were found to be 40 ±6.4 µM with respect to acetaldehyde and 99 ±11 µM with respect to $NAD^+$ (Figure 10). Substrate inhibition was also observed, with ALDH1A1 having a $K_i$ of 2.0 ± 0.56 mM with respect to acetaldehyde and 2.9 ± 0.73 mM with respect to $NAD^+$.

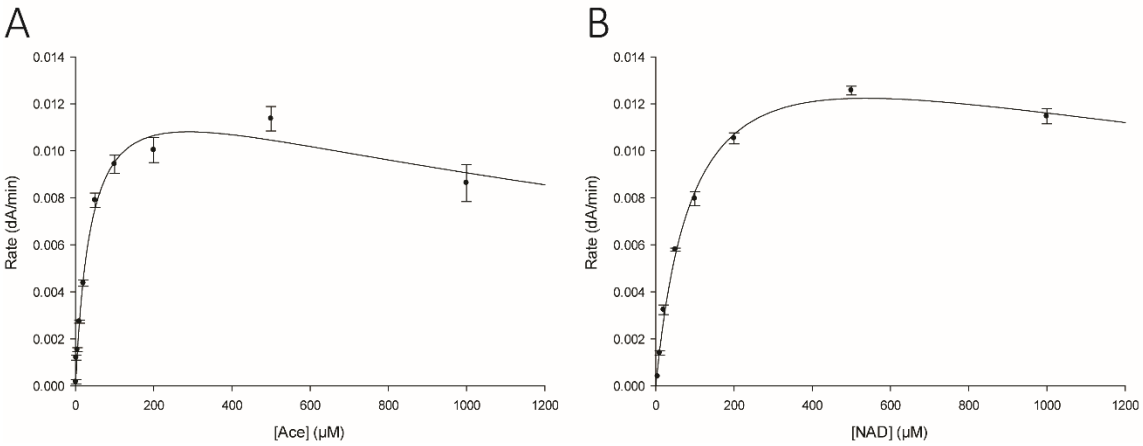



Figure 10: $K_m$ Measurement of ALDH1A1 with Respect to A) Acetaldehyde and B) $NAD^+$

To find the mode of inhibition with respect to $NAD^+$, an activity assay was set up with 150-200 nM ALDH1A1 and 1 mM propionaldehyde while varying the concentration of $NAD^+$ from 20 μM to 200 μM. Likewise, to find the mode of inhibition with respect to acetaldehyde, an activity assay was set up with 150-200 nM ALDH1A1 and 1 mM $NAD^+$ while varying the concentration of acetaldehyde from 20 μM to 200 μM. Concentration of inhibitor was varied at four concentrations in a range around the approximate $K_i$. Reactions were initiated by the addition of substrate, and activity was measured at 340 nm on a Cary 50-Bio UV-Visible Spectrophotometer. Readings were taken every 10 seconds for 2 minutes. Results were modeled and graphed using the tight-binding inhibition and single substrate—single inhibitor modules of SigmaPlot V12.3.

**4. X-ray Crystallography**

All protein used for crystallography was stored at -20 °C in 50 % (v/v) glycerol. Before use, it was dialyzed 3 times against 4 L of 10 mM Aces, 1 mM DTT, pH 6.6 buffer at 4°C. Each dialysis was at least 4 hours in length. All solutions used for setup of crystals were filtered using a 0.22 μm membrane. Crystal conditions were set up using the sitting drop method in 24-well Cryschem plates (Hampton Research, Laguna, CA), by combining 3 μL of protein stock (3-4 mg/mL) with 3 μL of crystallization solution over 500 μL of reservoir solution. Drops were mixed by pipetting. Crystallization solutions of 100 mM BisTris pH 6.1-6.4, 9-11 % PEG3350 (Hampton Research, Laguna, CA), 200 mM NaCl, and 5 mM $YbCl_3$ were used. The crystal plates were incubated at 20 °C. Crystal formation was observed within 3-20 days. The complex with CM38 was made by soaking the crystal with 500 μM compound in 2 % DMSO and 1 mM $NAD^+$ for 5 hours. The co-crystal with 3988-0485 was made by adding 250 uM of compound and 2 % DMSO to the crystallization solution. The freezing condition was composed of 20 % ethylene glycol in addition to all crystallization condition reagents and soak components at the same concentrations as in the growth conditions. Crystals were screened for diffraction on the home source Bruker X8 Prospector system. Diffracting crystals were stored in liquid nitrogen for up to 6 months. Diffraction data was collected at Beamline 19-ID of the Advanced Photon Source (Argonne National Laboratory, Chicago, IL). Data was integrated and scaled with the HKL3000 software suite (108). Rigid body, restrained refinement and TLS (translational/libration/screw) refinement were performed using RefMac5 in the CCP4 V6.5.019 Suite (109, 110). Parameters for TLS refinement were obtained using the TLS Motion Determination online server (111).

Modeling and visualization was performed using WinCoot V0.82 and PyMol V0.99 (DeLano Scientific LLC, San Francisco, CA) (112).

**5. Mammalian Cell Culture Assays**

The MDA-MB-468 triple-negative breast cancer cell line was obtained from the lab of Dr. Clark D. Wells (Indiana University Medical School, Indianapolis, IN), and was used for all cell culture assays (113). Dulbecco's Modification of Eagle's Medium (DMEM, 1 X) with 4.5 g/L glucose and L-glutamine without sodium pyruvate, supplemented with 10 % fetal bovine serum (FBS), 100 units/mL of penicillin, and 10 µg/mL of streptomycin, was used as the growth medium. Cells were subcultured at 60-80 % confluency by washing with phosphate-buffered saline (PBS) (GE Healthcare, Little Chalfont, UK), incubating with Trypsin 0.25 % 0.2 g/L EDTA ((GE Healthcare, Little Chalfont, UK) until cells detached, and transferring to fresh medium in a 1:5 ratio.

For experimental assays, cells were plated on the Matrigel Matrix basement membrane. Matrigel was thawed on ice overnight before use. 40 µL of 7 mg/mL Matrigel was plated per well on 96-well plates, and incubated at 37 °C for 30 minutes before use. 5000 MDA-MB-468 cells were then plated per well. Compound treatment was performed at 0.5 % DMSO and initiated synchronously with plating on Matrigel. Media and compound were changed every 48 hours. An MTT assay was used to measure cell proliferation and viability. The protocol for the use of the MTT assay in conjunction with Matrigel Matrix was based on previously published work (114). The wells were first washed with PBS. 100 µL of 0.5 mg/mL MTT reagent (Thermo Fisher Scientific, Waltham, MA) in PBS was added per well. Plates were incubated at 37 °C to allow reduction of the dye. 100 µL of 10 % Sodium-Dodecyl Sulfate (SDS), 0.05 N HCl was added per well. The plates were, again, incubated at 37 °C for 4 hours to promote cell lysis and Matrigel solubilisation. Absorbance readings were taken at 595 nm using the Spectramax 340PC plate reader (Molecular Devices, Sunnyvale, CA). All measurements were normalized to the Matrigel control wells, which contained no cells, compounds, or DMSO. Measurements were interpreted as a ratio to the DMSO control.

$$MTT\ signal\ ratio = \frac{A595\ after\ incuation\ in\ the\ presence\ of\ compound}{A595\ after\ incubation\ in\ the\ absence\ of\ compound}$$

## III. Results

### A. CM38 Inhibitor Properties and Structure-Activity Relationship

The CM38 compound was identified as a hit in the esterase high-throughput screen (115). The compound showed good structural characteristics as a lead compound, having a low molecular weight of 294 kDa and an approximate logP of 2.8 (Figure 11A). To further characterize the compound, it was screened for aldehyde dehydrogenase inhibition with 9 aldehyde dehydrogenase isoenzymes at 20 µM (Figure 11B). CM38 showed excellent selectivity for ALDH1A1 over the other, highly similar isoenzymes in the subfamily. CM38 also demonstrated respectable potency as an ALDH1A1 inhibitor with an $IC_{50}$ value of 440 ± 37 nM and a $K_i$ of 320 ± 14 nM. CM38 was found to be uncompetitive with respect to varied $NAD^+$, which is consistent with the idea that it does not bind the cofactor-binding site (Figure 12). Based on these characteristics, CM38 was chosen as a lead compound in structure-activity relationship development.

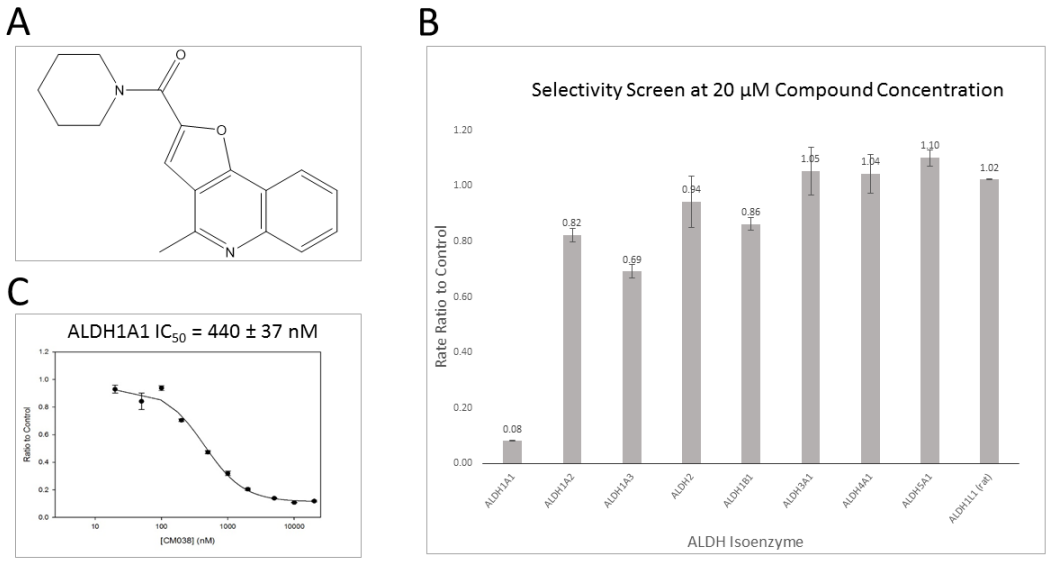



Figure 11: Characterization of CM38 as an Aldehyde Dehydrogenase Inhibitor.
A) Structure of the CM38 compound. B) Selectivity screen of CM38 with 8 human aldehyde dehydrogenase enzymes and 1 rat aldehyde dehydrogenase enzyme. C) Dose-response curve of CM38 with the ALDH1A1 isoenzyme.

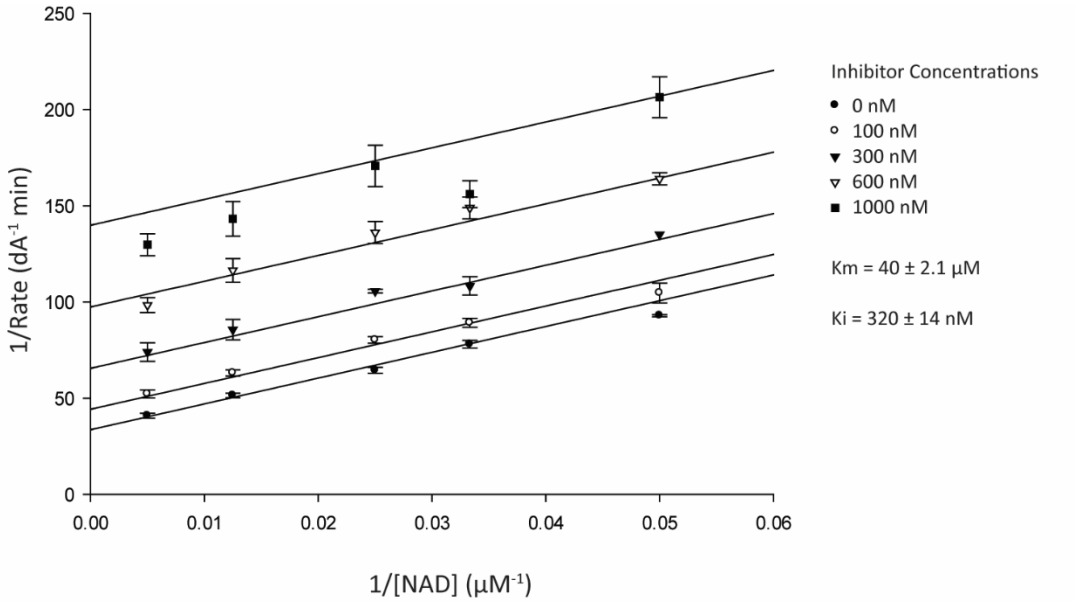
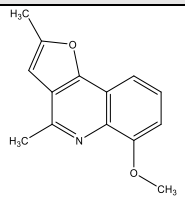
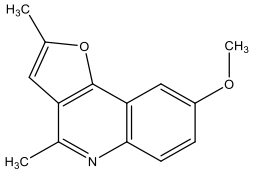
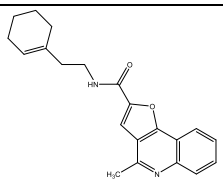
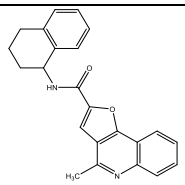
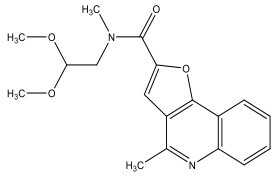
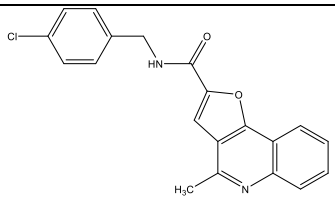
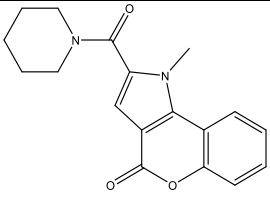



Figure 12: Lineweaver-Burk Plot of CM38 Inhibition of ALDH1A1 with Varied $NAD^+$
CM38 inhibition is uncompetitive with respect to varied $NAD^+$, and has a $K_i$ of 320 ± 14 nM

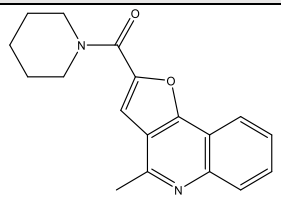
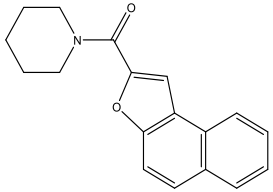
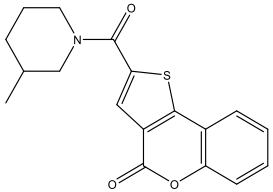
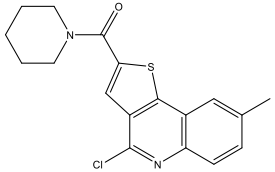
Compounds with high structural similarity to CM38 were ordered from commercial sources and evaluated as ALDH1A1 inhibitors. Compounds that showed minimal ALDH1A1 inhibition are summarized in Table 3. These compounds allowed the initial development of the CM38 structure-activity relationship by highlighting structural elements that are critical for aldehyde dehydrogenase inhibition. In particular, analogues missing the piperidine ring (57919133, 5798949 and C629-0256) showed minimal to no inhibition of ALDH1A1. Similarly, compounds where the fourth ring was separated from the main three-ring body by a linker (C629-0004, C629-0198 and C629-0250) also had a severely reduced ability to inhibit ALDH1A1.

Table 3: CM38 Analogues that are Relatively Inactive as ALDH Inhibitors

| Compound ID | Compound Structure | ALDH1A1 Rate Ratio to Control at 20 µM of Compound (± STDEV) |
|---|---|---|
| 57919133 |  | 0.66 ± 0.013 |
| 5798949 |  | 0.58 ± 0.025 |
| C629-0004 |  | 0.49 ± 0.037 |
| C629-0198 |  | 0.58 ± 0.016 |
| C629-0256 |  | 1.14 ± 0.010 |
| C629-0250 |  | 0.97 ± 0.025 |
| F210-0074 |  | 0.59 ± 0.13 |

CM38 compounds that demonstrated the ability to inhibit ALDH1A1 are summarized in Table 4. These compounds helped develop the structure-activity relationship of CM38 even further by identifying structural changes of the CM38 compound that are tolerated in the context of ALDH inhibition. Notably, all compounds have good selectivity for ALDH1A1 and do not significantly inhibit the other isoenzymes. These compounds show that the furan ring of the CM38 structure may be rotated in orientation (7998070) or converted to a pyrrole (C186-0036, C618-0097) with only a modest loss in potency. Similarly, the pyridine ring may be converted to a benzene (7998070) or a pyran ring (C186-0036). The methyl group on the pyrrole ring is not essential, and may be removed (7998070) or replaced with a carbonyl (C186-0036) or chloride (C618-0097) substitution. However, it should be noted that the removal of this methyl does appear to drastically lower the selectivity of the compound against ALDH2 (7998070). Compound 7998070 is of particular note in this compound series, as many more commercial compounds are available for it than for the original CM38 parent compound. Due to the close structural similarity and similar ALDH inhibition properties, it may be assumed that 7998070 binds and inhibits ALDH1A1 through an identical mechanism to CM38 (Figure 13). As such, any structure-activity relationship results obtained from 7998070 analogues can also be applied to the CM38 compound. Thus, 7998070 was chosen as a new parent compound to further develop the structure-activity relationship of this compound series.

Table 4: CM38 Analogues that are Active as ALDH Inhibitors

| Compound ID | Structure | ALDH1A1 IC50 | ALDH1A2 Rate ratio to Control at 20 µM | ALDH1A3 Rate ratio to Control at 20 µM | ALDHALDH2 Rate ratio to Control at 20 µM | ALDH1B1 Rate ratio to Control at 20 µM |
|---|---|---|---|---|---|---|
| CM38 |  | 440 ± 37 nM | 0.82 ± 0.024 | 0.69 ± 0.025 | 0.94 ± 0.094 | 0.86 ± 0.023 |
| 7998070 |  | 1.3 ± 0.079 µM | 0.81 ± 0.0062 | 1.02 ± 0.086 | 0.49 ± 0.024 | 0.88 ± 0.076 |
| C186-0036 |  | 500 ± 20 nM | 0.65 ± 0.026 | 1.14 ± 0.047 | 0.86 ± 0.027 | 0.77 ± 0.028 |
| C618-0097 |  | 1.3 ± 0.13 µM | 0.53 ± 0.15 | 1.04 ± 0.080 | 0.81 ± 0.015 | 0.76 ± 0.086 |

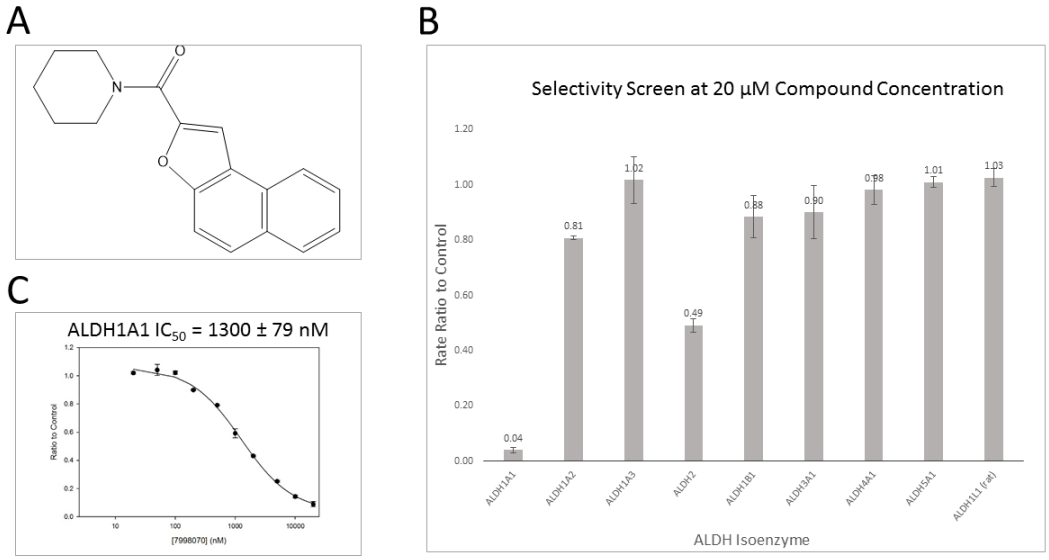
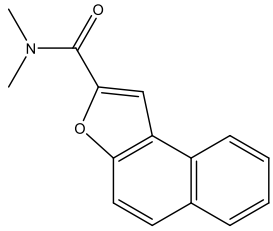
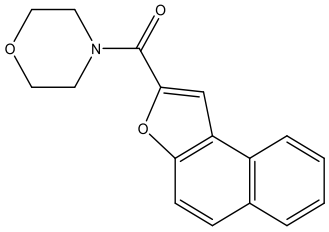
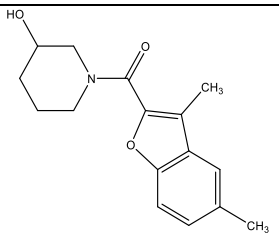
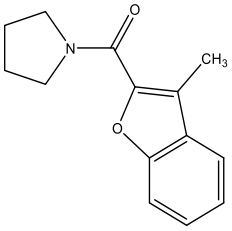
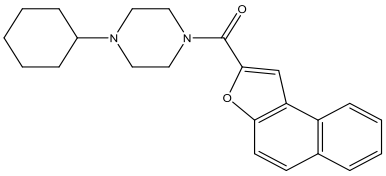



Figure 13: Characterization of 7998070 as an Aldehyde Dehydrogenase Inhibitor.
A) Structure of the 7998070 compound. B) Selectivity screen of 7998070 with 8 human aldehyde dehydrogenase enzymes and 1 rat aldehyde dehydrogenase enzyme. C) Dose-response curve of 7998070 with the ALDH1A1 isoenzyme.

Commercially available analogues of compound 7998070 that did not significantly inhibit ALDH1A1 are summarized in Table 5. As was the case with CM38, the 7998070 structure-activity relationship demonstrates that the piperidine ring is critical, and its removal has a dramatic adverse effect on ALDH inhibition (9004424). Additionally, the substitution of the piperidine ring with a morpholine ring (9018056) or a piperazine with a cyclohexane sidechain (C629-0173) is not tolerated. This likely confirms that the piperidine ring has to remain relatively non-polar in order to inhibit ALDH1A1. Additionally, removing one of the rings in the three-ring structure also minimizes inhibition, suggesting that the three-ring body of the compound is also critical.

Table 5: 7998070 Analogues that are Relatively Inactive as ALDH Inhibitors

| Compound ID | Compound Structure | ALDH1A1 Rate Ratio to Control at 20 µM of Compound |
|---|---|---|
| 9004424 |  | 0.73 ± 0.052 |
| 9018056 |  | 0.80 ± 0.018 |
| 35004881 |  | 0.76 ± 0.029 |
| 9087477 |  | 0.44 ± 0.027 |
| C629-0173 |  | 0.77 ± 0.059 |

Analogues of 7998070 that inhibited ALDH1A1 and were further characterized are summarized in Table 6. Most notably, this group of compounds demonstrates that as long as the piperidine ring is in position, it can tolerate expansion to an azepane (9040591) or addition of

methyl side chains at the meta and para positions (9129602, 9129634) with no loss of potency toward ALDH1A1. However, reducing the size of the piperidine ring to a pyrrolidine (9031547) strongly reduces ALDH inhibition. Compounds E003-0974 and C893-0789 are of particular interest in this series. The conversion of one of the aromatic rings in the three-ring body of the compound to a single-unsaturated oxane with a carbonyl side-chain (E003-0974, C893-0789) actually improves the potency of the compound toward ALDH1A1, with the $IC_{50}$ lowering from 1.3 µM for the 7998070 parent compound to 300 nM for C893-0789. Although E003-0974 and C893-0789 both feature hydrophobic side-chains on the third ring, the reduction of this sidechain from a cyclohexane to an ethyl and methyl (between E003-0974 and C893-0789) had no adverse effects on potency. Therefore, this indicates that the oxane and carbonyl groups are likely responsible for the improved potency seen in these compounds.

Table 6: 7998070 Analogues that are Active as ALDH Inhibitors

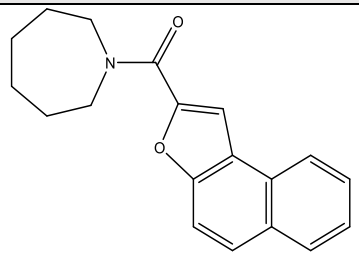
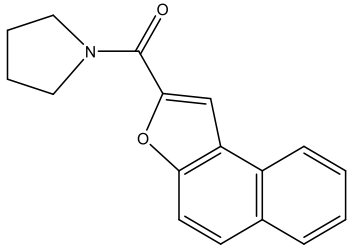
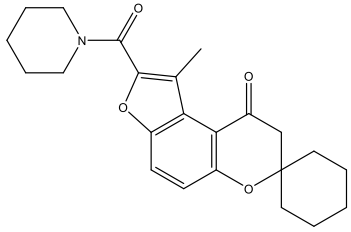
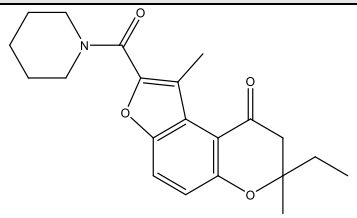
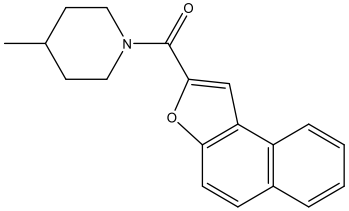
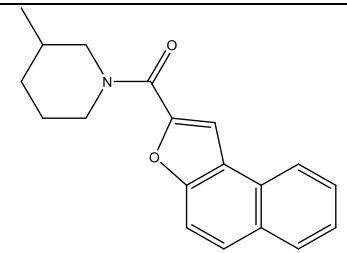
| Compound | Structure | ALDH1A1 IC50 | ALDH1A2 Rate ratio to Control at 20 µM | ALDH1A3 Rate ratio to Control at 20 µM | ALDH2 Rate ratio to Control at 20 µM | ALDH1B1 Rate ratio to Control at 20 µM |
|---|---|---|---|---|---|---|
| 9040591 |  | 1.5 ± 0.12 µM | 0.85 ± 0.013 | 1.03 ± 0.098 | 0.62 ± 0.0098 | 0.94 ± 0.0050 |
| 9031547 |  | 6.7 ± 1.3 µM | 0.97 ± 0.031 | 0.87 ± 0.032 | 0.53 ± 0.028 | 1.09 ± 0.044 |
| E003-0974 |  | 560 nM ± 110 nM | 0.63 ± 0.042 | 0.85 ± 0.040 | 0.76 ± 0.30 | 0.67 ± 0.039 |

Table 6: Continued

| Compound | Structure | 1A1 IC50 | 1A2 Rate ratio to Control at 20 µM | 1A3 Rate ratio to Control at 20 µM | ALDH2 Rate ratio to Control at 20 µM | 1B1 Rate ratio to Control at 20 µM |
|---|---|---|---|---|---|---|
| C893-0789 |  | 300 ± 14 nM | 0.58 ± 0.031 | 1.11 ± 0.032 | 0.85 ± 0.0093 | 0.60 ± 0.019 |
| 9129602 |  | 1.4± 0.18µM | 0.62 ± 0.050 | 0.61 ± 0.0065 | 0.71 ± 0.026 | 0.73 ± 0.032 |
| 9129634 |  | 1.3± 0.10µM | 0.72 ± 0.0086 | 1.03 ± 0.21 | 0.68 ± 0.028 | 0.79 ± 0.090 |

**B. CM10 Inhibitor Properties and Structure-Activity Relationship**

CM10 was identified along with CM38 in the esterase high-throughput screen for ALDH1A1 modulators. It had good basic qualities as a lead compound, featuring a size of 321 kDa and an estimated logP of 4.82. One liability of the compound was the potentially reactive alkene group (Figure 14A). The compound was screened as an ALDH inhibitor with other a range of ALDH isoenzymes (Figure 14B). Unlike CM38, CM10 was not selective for ALDH1A1 and also inhibited ALDH1A2, ALDH1A3, and ALDH1B1. However, the compound did not inhibit ALDH2, which likely means that it is selective only for the ALHD1A subfamily and ALDH1B1. Dose-response curves (Figure 14C) showed that the compound had the highest potency with ALDH1A2 and ALDH1A3. Moreover, it was a partial inhibitor of ALDH1A1, and reached only a minimum rate ratio to control of 0.30 on the $IC_{50}$ curve. Steady-state kinetics showed that ALDH1A1 inhibition was mixed partial with respect to varied $NAD^+$, and had a $K_i$ of = 1.6 ± 0.34 µM (Figure 15). Due to a very interesting selectivity profile and the potential to develop into a therapeutically applicable ALDH1A-selective inhibitor, the CM10 compound was chosen for further characterization through the assembly of a structure-activity relationship.

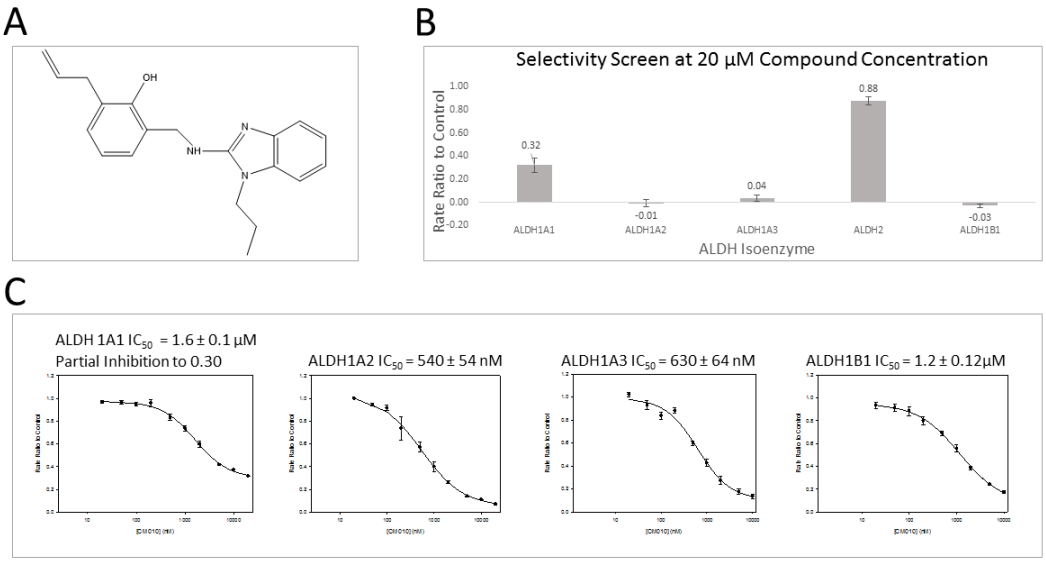



Figure 14: Characterization of CM10 as an Aldehyde Dehydrogenase Inhibitor
A) Structure of the CM10 compound. B) Selectivity screen of CM10 with 8 human aldehyde dehydrogenase enzymes and 1 rat aldehyde dehydrogenase enzyme. C) Dose-response curves of CM10 with the ALDH1A1, ALDH1A2, ALDH1A3, and ALDH1B1 isoenzymes.

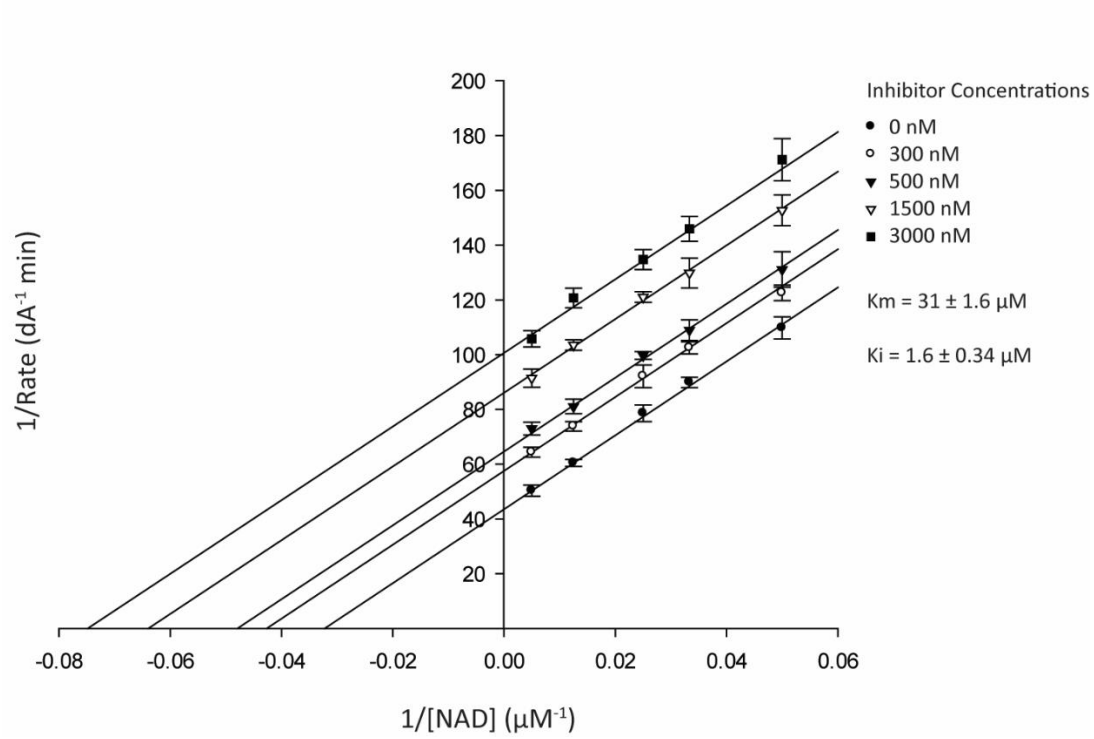



Figure 15: Lineweaver-Burk Plot of CM10 Inhibition of ALDH1A1 with Varied $NAD^+$
CM10 inhibition is mixed partial with respect to varied $NAD^+$, and has a $K_i$ of 1.6 ± 0.34 µM

Commercially available analogues of CM10 were evaluated as inhibitors of the ALDH1 family. The compounds that were not found to inhibit any ALDH1 isoenzymes are summarized in Table 7. Most of these compounds lack the hydroxyl of the phenol group (Y020-7301, D010-0120, D010-0112, 6738-0091) leading to the interpretation that this hydroxyl group is essential for ALDH inhibition. The only inactive CM10 analogue to feature the phenol ring is compound 6738-0174. In this case, the reasonable interpretation is that the compound scaffold does not tolerate a bulky adduct, such as a morpholine ring, on the alkyl side-chain.

CM10 analogues that inhibited ALDH1A enzymes were, for the most part, partial inhibitors. Since this is a property shared by the parent compound CM10, it is likely inherent to the scaffold and its binding mode. One of the early priorities in the development of the CM10 structure-activity relationship was the determination of the role of the potentially-reactive alkene group. The structure-activity relationship shows that while the alkene group does play a role in determining the inhibition characteristics of the compound, it is not absolutely crucial for inhibition. Removal of the alkene chain modestly reduced the potency of the compound with

ALDH1A1, ALDH1A2, and ALDH1B1 (3988-0434). Replacement of the side chain with an ether group at the same position strongly reduced ALDH1A2 and ALDH1B1 inhibition, but improved ALDH1A1 inhibition, thereby producing a compound selective for ALDH1A1 and ALDH1A3 (3988-0587). Removal of the alkene and addition of chloride groups at the 2- and 4- positions on the phenol produced a compound that is a highly potent ALDH1A1 and ALDH1A3 inhibitor; however, the compound loses selectivity and shows minor inhibition of ALDH2 (3988-0474). Removal of the alkene group and the addition of a benzyl ring onto the phenol improves ALDH1A1 inhibition but reduces inhibition of the other ALDH1A isoenzymes. This produces a more selective ALDH1A1 inhibitor (3988-0485). Overall, the structure-activity relationship suggests that the alkene group is suboptimal for ALDH1A1 or ALDH1A3 inhibition; however, the group does appear to be key for high potency toward ALDH1A2.

The most interesting aspect of the CM10 structure-activity relationship is the variety of ways in which the inhibitor selectivity can be modified. Branching of the linker between the phenol and the two-ring structure eliminated inhibition of ALDH1A1 and produced ALDH1A3-selective compounds (3988-0506, 3988-0580). On the other hand, branching or extending the alkyl side chain on the two ring structure eliminated ALDH1A3 inhibition, producing compounds with excellent selectivity for ALDH1A1 (5519-0624, 3988-0486).

Table 7: CM10 Analogues that are Relatively Inactive as ALDH Inhibitors

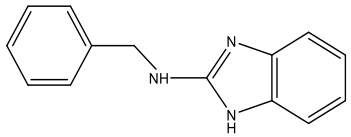
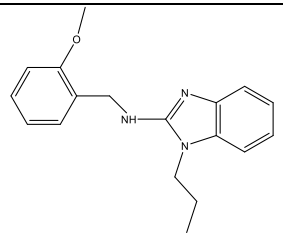
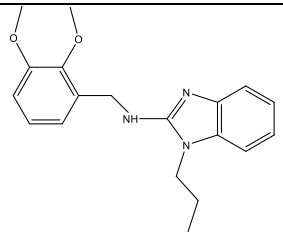
| Compound | Structure | ALDH1A1 Rate Ratio to Control at 20 µM | ALDH1A2 Rate Ratio to Control at 20 µM | ALDH1A3 Rate Ratio to Control at 20 µM | ALDH2 Rate Ratio to Control at 20 µM | ALDH1B1 Rate Ratio to Control at 20 µM |
|---|---|---|---|---|---|---|
| Y020-7301 |  | 1.0 ± 0.016 | 1.03 ± 0.0031 | 1.07 ± 0.12 | 0.92 ± 0.014 | 1.06 ± 0.015 |
| D010-0120 |  | 1.0 ± 0.0067 | 1.02 ± 0.031 | 1.22 ± 0.047 | 0.87 ± 0.020 | 1.07 ± 0.015 |
| D010-0112 |  | 1.01 ± 0.023 | 1.03 ± 0.027 | 1.10 ± 0.049 | 0.91 ± 0.0070 | 1.02 ± 0.041 |

Table 7: Continued

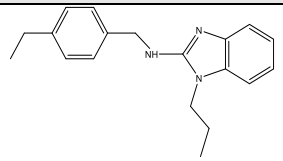
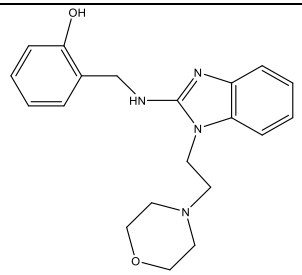
| Compound | Structure | ALDH1A1 Rate Ratio to Control at 20 µM | ALDH1A2 Rate Ratio to Control at 20 µM | ALDH1A3 Rate Ratio to Control at 20 µM | ALDH2 Rate Ratio to Control at 20 µM | ALDH1B1 Rate Ratio to Control at 20 µM |
|---|---|---|---|---|---|---|
| 6738-0091 |  | 0.99 ± 0.033 | 0.97 ± 0.012 | 1.07 ± 0.031 | 0.86 ± 0.014 | 0.98 ± 0.031 |
| 6738-0174 |  | 0.86 ± 0.034 | 0.66 ± 0.064 | 0.63 ± 0.021 | 0.91 ± 0.064 | 0.70 ± 0.039 |

Table 8: CM10 Analogues that are Active as ALDH Inhibitors

| $IC_{50}$ < 1 µM | $IC_{50}$ > 1 µM, Act. at 20 µM <0.5 | 0.4 <Act. at 20 µM <0.8 | Act. at 20 µM >0.8 |
|---|---|---|---|

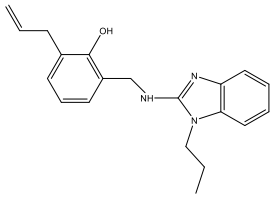
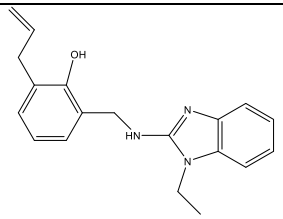
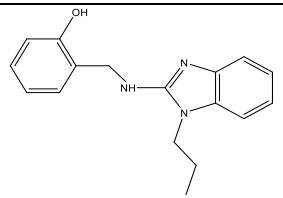
| Compound | Structure | ALDH1A1 (Min. Activity) | ALDH1A2 (Min. Activity) | ALDH1A3 (Min. Activity) | ALDH2 (Min. Activity) | ALDH1B1 (Min. Activity) |
|---|---|---|---|---|---|---|
| CM010 |  | $IC_{50}$= 1600 ± 100 nM (0.30) | $IC_{50}$= 540 ± 54 nM (0.05) | $IC_{50}$= 630 ± 64 nM (0.11) | Act. at 20 µM= 0.88 ± 0.037 | $IC_{50}$= 1200 ± 120 nM (0.073) |
| 5519-0628 |  | $IC_{50}$= 2.8 ± 0.35 µM (0.35) | $IC_{50}$= 820 ± 43 nM (0.09) | $IC_{50}$= 1.2 ± 0.13 µM (0.24) | Act. at 20 µM= 0.84 ± 0.018 | $IC_{50}$= 1.2 ± 0.081 µM (0.20) |
| 3988-0434 |  | $IC_{50}$= 3.7 ± 0.38 µM (0.26) | $IC_{50}$= 2.7 ± 0.32 µM (0.33) | $IC_{50}$= 700 ± 45 nM (0.27) | Act. at 20 µM= 0.87 ±0.027 | $IC_{50}$= 2.1 ± 0.23 µM (0.4) |

Table 8: Continued

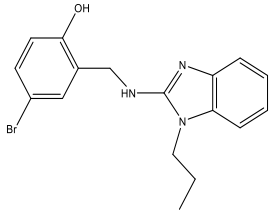
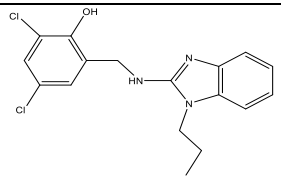
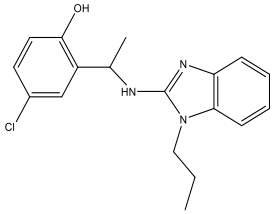
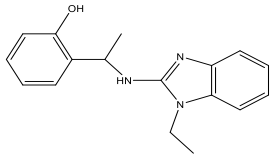
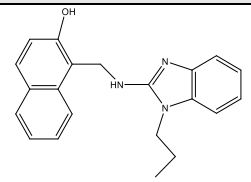
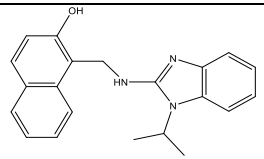
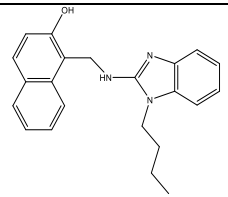
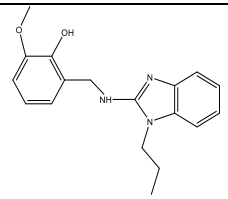
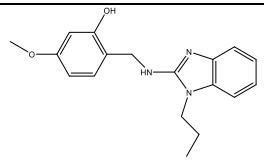
| Compound | Structure | ALDH1A1 (Min. Activity) | ALDH1A2 (Min. Activity) | ALDH1A3 (Min. Activity) | ALDH2 (Min. Activity) | ALDH1B1 (Min. Activity) |
|---|---|---|---|---|---|---|
| 3988-0451 |  | $IC_{50}$= 3.3 ± 0.44 µM (0.30) | Act. at 20 µM= 0.61 ± 0.062 | $IC_{50}$= 4.3 ± 1.6 µM | Act. at 20 µM= 0.77 ± 0.0056 | Act. at 20 µM= 0.60 ± 0.026 |
| 3988-0474 |  | $IC_{50}$= 300 ± 41 nM (0.16) | $IC_{50}$= 3.2 ± 1.2 µM | $IC_{50}$= 520 ± 39 nM | Act. at 20 µM= 0.49 ± 0.044 | $IC_{50}$= 1.9 ± 0.23 µM |
| 3988-0506 |  | Act. at 20 µM= 1.03 ± 0.011 | $IC_{50}$= 7.1 ± 2.5 µM (0.37) | $IC_{50}$= 3.1 ± 0.73 µM (0.23) | Act. at 20 µM= 0.86 ± 0.013 | Act. at 20 µM= 0.59 ± 0.0096 |
| 3988-0580 |  | Act. at 20 µM= 0.75 ± 0.023 | $IC_{50}$= 3.1 ± 0.29 µM (0.42) | $IC_{50}$= 730 ± 86 nM (0.21) | Act. at 20 µM= 0.94 ± 0.032 | Act. at 20 µM= 0.47 ± 0.018 |

Table 8: Continued

| Compound | Structure | ALDH1A1 (Min. Activity) | ALDH1A2 (Min. Activity) | ALDH1A3 (Min. Activity) | ALDH2 (Min. Activity) | ALDH1B1 (Min. Activity) |
|---|---|---|---|---|---|---|
| 3988-0485 |  | $IC_{50}$= 780 ± 78 nM (0.18) | $IC_{50}$= 2.2 ± 0.28 µM (0.49) | $IC_{50}$= 1.8 ± 0.50 µM (0.24) | $IC_{50}$= 0.65 ± 0.073 | $IC_{50}$= 4.4 ± 0.87 µM (0.58) |
| 5519-0624 |  | $IC_{50}$= 2.3 ± 0.22 µM (0.28) | Act. at 20 µM= 0.62 ± 0.086 | Act. at 20 µM= 1.01 ± 0.031 | Act. at 20 µM= 0.76 ± 0.049 | Act. at 20 µM= 0.86 ± 0.029 |
| 3988-0486 |  | $IC_{50}$= 620 ± 59 nM (0.33) | Act. at 20 µM= 0.98 ± 0.11 | Act. at 20 µM= 0.97 ± 0.060 | Act. at 20 µM= 0.85 ± 0.087 | Act. at 20 µM= 1.01 ± 0.052 |
| 3988-0587 |  | $IC_{50}$= 750 ± 74 nM (0.20) | $IC_{50}$= 9.5 ± 8.9 µM (0.18?) | $IC_{50}$= 620 ± 67 nM (0.28) | Act. at 20 µM= 0.85 ± 0.020 | $IC_{50}$= 7.6 ± 0.97 µM (0.4) |
| 3988-0586 |  | $IC_{50}$= 3.8 ± 0.60 µM (0.45) | $IC_{50}$= 8.7 ± 3.7 µM (0.26) | Act. at 20 µM= 0.99 ± 0.025 | Act. at 20 µM= 0.83 ± 0.017 | Act. at 20 µM= 0.51 ± 0.029 |

**C. X-ray Crystallography of ALDH1A1 in Complex with CM38 and CM10**

In order to characterize the binding mode of the CM38 and CM10 compound series and better explain their varied selectivity profiles, an X-ray crystallography approach was used. X-ray diffraction datasets were collected for ALDH1A1 in complex with CM38 and the CM10 analogue 3988-0485. The complex with CM38 was formed by soaking the crystal with the CM38 compound and the $NAD^+$ cofactor. The complex with 3988-0485 was formed by co-crystallization. The space groups and cell dimensions match the apo-enzyme ALDH1A1 structure previously published, indicating the presence of the same ALDH1A1 crystal isoform (Table 9) (104). The CM38 and CM10 structures were collected at resolutions of 1.8 Å and 1.65 Å, respectively. The CM38 structure was refined successfully to a $R_{work}/R_{free}$ of 0.18 / 0.23. The 3988-0485 structure, on the other hand, showed only partial compound occupancy. As such, the ALDH1A1-39880485 crystal structure should be interpreted with less confidence; however, it still holds value for forming hypotheses and making potential interpretations of the structure-activity relationship.

Table 9: Data Collection and Refinement Statistics for the structures of ALDH1A1 in Complex with CM38 and 3988-0485

| | ALDH1A1-CM38 | ALDH1A1-39880485 |
|---|---|---|
| Data Collection | | |
| Space Group | P422 | P422 |
| Cell Dimensions | | |
| a, b, c (Å) | 109, 109, 83 | 109, 109, 83 |
| A, β, γ (°) | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 50-1.8 | 50-1.65 |
| R merge | 0.09 (0.50) | 0.07 (0.70) |
| I/σ | 24.4 (3.2) | 34 (3.6) |
| Completeness (%) | 95.6 (77.6) | 100 (99.9) |
| Redundancy | 12.9 (9.9) | 12.6 (11.3) |
| Refinement Statistics | | |
| Number of Reflections | 42449 | 57482 |
| $R_{work}/R_{free}$ | 0.18 / 0.23 | 0.24 / 0.28 |
| Number of atoms | 4320 | 4085 |
| Protein | 3853 | 3852 |
| Ligand/Ion | 69 | 28 |
| Water | 398 | 205 |
| RMS Deviations | | |
| Bond Lengths (Å) | 0.011 | 0.019 |
| Bond Angles (°) | 1.6 | 1.8 |

The structure of CM38 bound to ALDH1A1 shows that CM38 binds within the substrate binding pocket of the enzyme (Figure 16). The compound binds very close to the catalytic residue (Cys303) and appears to interact with the adjacent residue (Cys302). Cys302 forms contacts with the nitrogen of the piperidine ring and the oxygen of the furan ring. The three-ring body of the compound binds in parallel to Tyr297, and forms pi-stacking interactions with its aromatic ring. Val 460 is a key residue that helps shape the hydrophobic pocket that accommodates the piperidine ring.

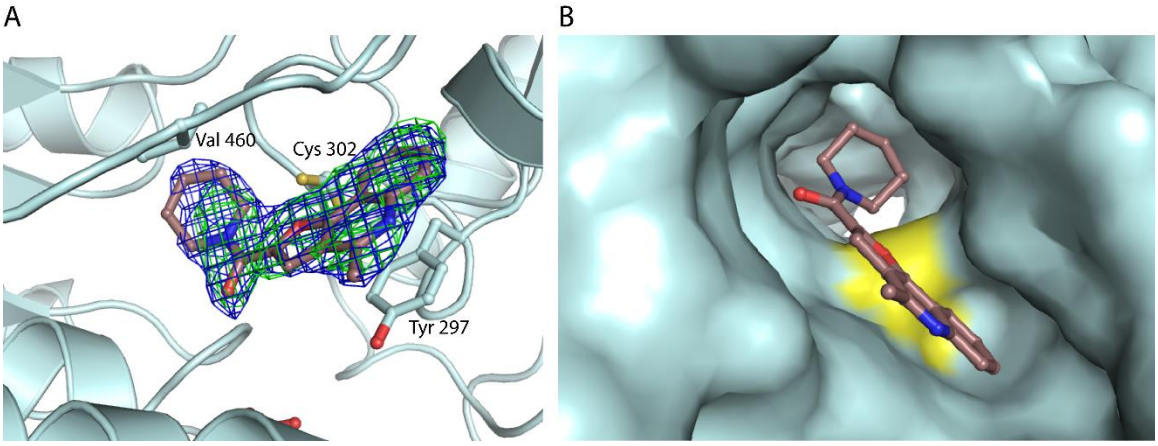



Figure 16: Structure of ALDH1A1 in Complex with CM38
A) Electron density maps of CM38 with the final refined $2F_o - F_c$ ($\sigma = 1.0$) map is shown in blue and the original $F_o - F_c$ ($\sigma = 2.5$) map is shown in green. B) The binding location of CM38 within the substrate binding pocket. Cys302 is highlighted in yellow.

3988-0485 also appears to bind within the substrate-binding pocket in its crystal structure (Figure 17). In contrast to CM38, the compound does not bind immediately adjacent to the catalytic residue, and is located a little further out of the catalytic pocket. The hydroxide of the phenol ring appears to form a hydrogen bond with Trp178. Phe171 forms pi-stacking interactions with the imidazole-benzene ring system.

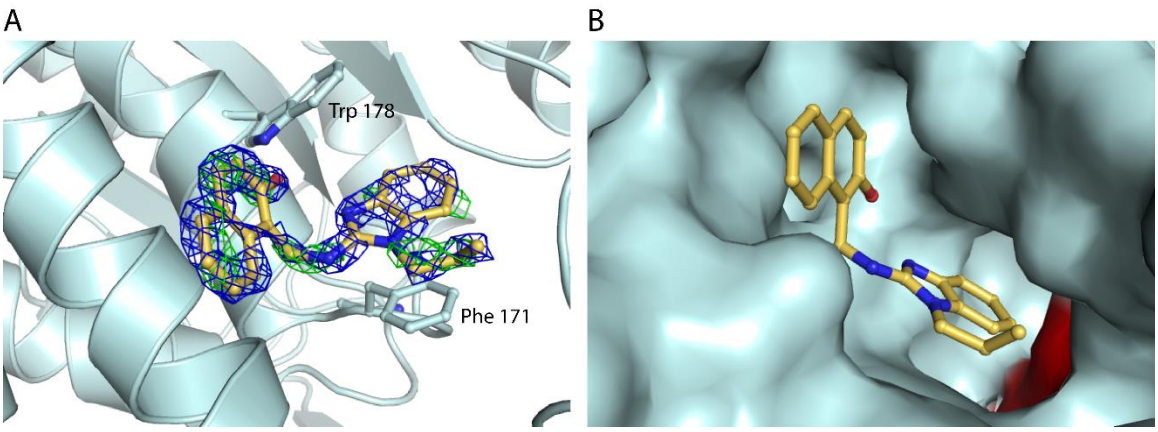



Figure 17: Structure of ALDH1A1 in Complex with CM10 Analogue 3988-0485
A) Electron density maps of 3988-0485 with the final refined $2F_o - F_c$ ($\sigma = 1.0$) map is shown in blue and the original $F_o - F_c$ ($\sigma = 2.5$) map is shown in green. B) Binding location of 3988-0485 within the substrate binding pocket. The catalytic residue Cys303 is highlighted in red.

**D. Characterization of the CM38 and CM10 Compound Series in Cell Culture**

In order to evaluate the effectiveness of the CM38 and CM10 compound series in cell culture, both a cell line and a quantitation method had to be chosen. The MDA-MB-468 human triple-negative breast cancer cell line was previously shown to contain a subpopulation of cells with high ALDH activity (116) Additionally, it has been found that this cell line is sensitive to ALDH1A1 modulators when grown on the Matrigel basement membrane (unpublished, Matthew Martien, Dr. Wells Lab, Indiana University Medical School). For these reasons, the MDA-MB-468 cell line was chosen for cell culture assays, with all inhibitor treatment performed on cells grown on the Matrigel basement membrane. A valid approach for measuring cell viability and proliferation after compound treatment also had to be established. The MTT assay is a convenient and often used cell assay to measure cell viability. A validation assay was performed to ensure that the MTT assay could be reliably used to evaluate the viability of MDA-MB-468 cells when grown on Matrigel. Cells were plated at initial densities of 1000, 4000 and 16000 cells/ well, and the MTT assay was performed after 1, 3, and 5 days. The results showed a clear distinction between all three initial densities (Figure 18). The results were also consistent with the expected growth pattern of cultured cells, having a lag phase and an exponential growth phase. According to these results, the MTT assay is a valid method of measuring cell viability and proliferation of the MDA-MB-468 cell line grown on Matrigel. As such, this methodology was used for all cell culture assays.

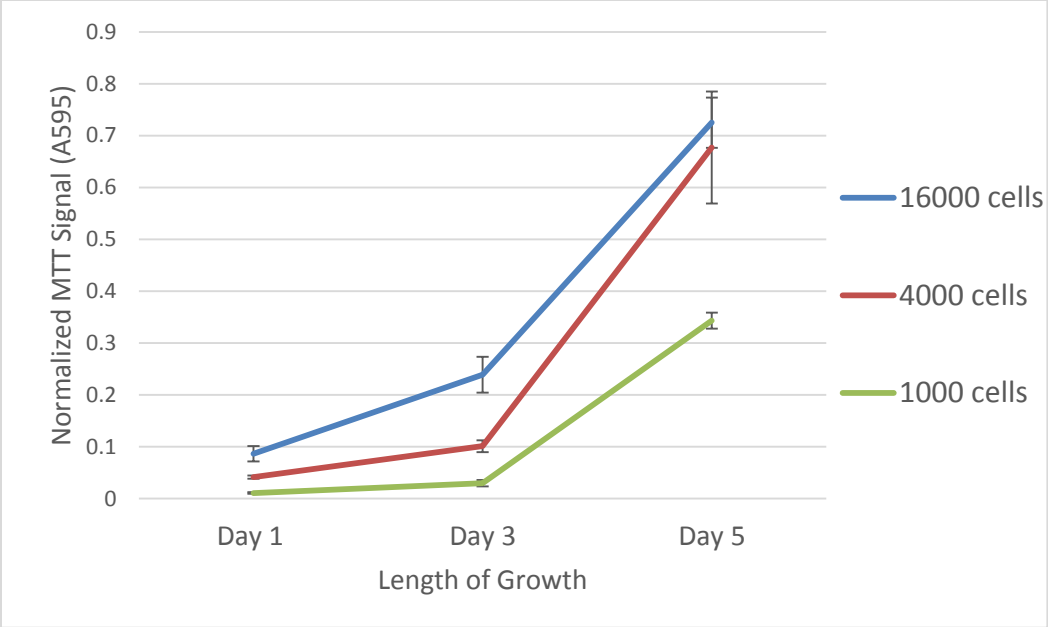
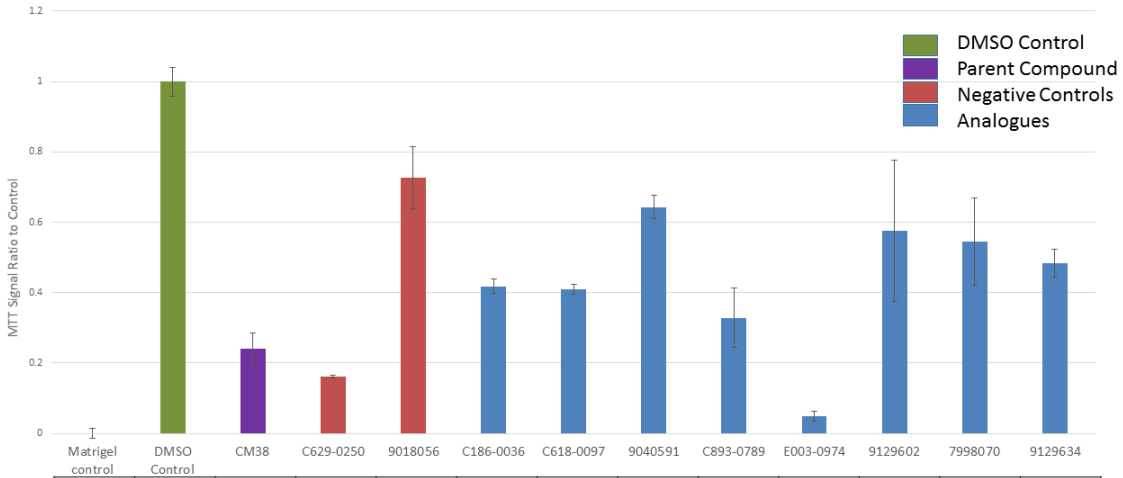



Figure 18: MTT Assay Validation
MDA-MB-468 cells were plated at three different densities on the Matrigel Basement Membrane. The MTT assay was used to measure cell viability at three time points.

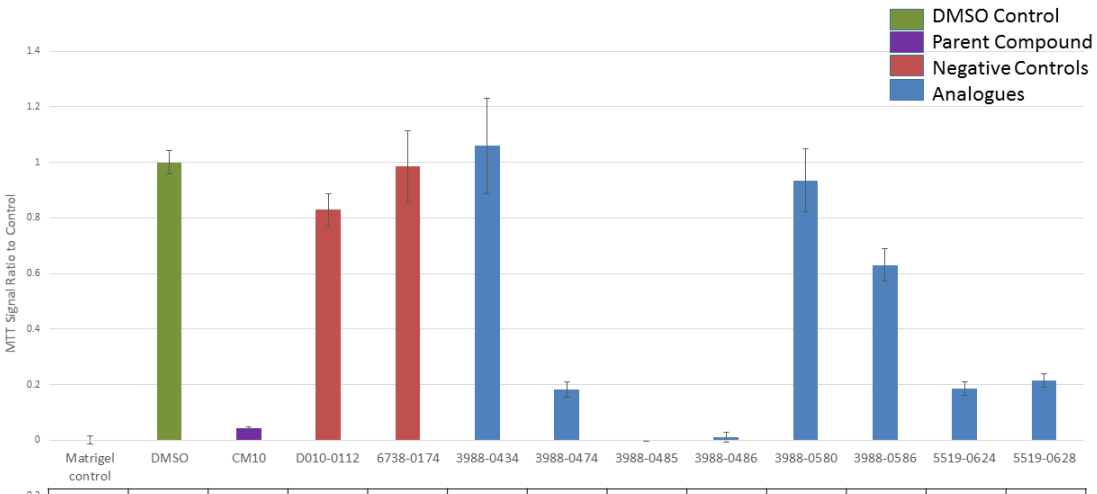
Several compounds from the CM38 compound series were screened for activity in cell culture. The tested compounds were chosen based on their physical properties and their activity as ALDH inhibitors. Two compounds (C629-0250 and 9018056), which were not effective as ALDH inhibitors in vitro, were also used in the assay as negative controls. Cells were treated with 30 µM of compound for 5 days, after which the MTT assay was performed. A number of compounds were found to successfully reduce the viability and/or proliferation of the cells in the assay (Figure 19). The parent compound had an MTT signal ratio to DMSO control of 0.24 ± 0.044; the E003-0974 compound was even more effective, with an MTT ratio to control of 0.048 ± 0.015. Both of these compounds were among the most potent ALDH1A1 inhibitors in vitro, which is a positive sign that suggests the observed effect is due, at least in part, to ALDH1A1 inhibition. However, the screen also revealed a potential concern about the CM38 compound series. The negative controls (C629-0250 and 9018056) both showed an MTT ratio to control of below 0.8, with the C629-0250 negative control showing high potency with a ratio of 0.16 ± 0.0043. This may suggest that the CM38 scaffold has some general toxicity, and that at least some of the observed effect in the MTT screen is independent of ALDH inhibition.




| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1A1 IC50** | - | 0.44uM | 0.97 at 20uM | 0.80 at 20uM | 0.5uM | 1.3uM | 1.5uM | 0.3uM | 0.56uM | 1.4uM | 1.3uM | 1.3uM |
| **LogP** | - | 2.8 | 4.35 | 2.51 | 3.76 | 4.77 | 4.09 | 4.01 | 5.07 | 4.05 | 3.53 | 4.05 |
| **MW** | - | 294.36 | 350.81 | 281 | 327.41 | 344.87 | 293 | 355.44 | 395.5 | 293 | 279 | 293 |

Figure 19: MTT Screen of CM38 Analogues at 30 µM
Screen conducted on MDA-MB-468 cells plated on Matrigel. MTT assay performed after 5-day incubation.

CM10 series compounds were screened in the same way as CM38 compounds, including two negative control compounds (Figure 20). Unlike in the CM38 series, the CM10 negative control compounds (D010-0112 and 6738-0174) showed minimal effectiveness in the assay, giving no reason to suspect off-target toxicity from this scaffold. Overall, compounds that were selective for isoenzymes other than ALDH1A1 (such as 3988-0434 and 3988-0580) performed poorly in the assay, suggesting that ALDH1A1 plays the dominant role in the MDA-MB-468 cell line. On the other hand, compounds that showed potent inhibition of ALDH1A1 in vitro (such as 3988-0485, 3988-0486, and 3988-0474) all showed good potency in the MTT screen. Compounds CM10, 3988-0485 and 3988-0486 were chosen for further characterization in cell culture based on their excellent effectiveness in the screen.




| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1A1 IC50 (MIN)** | - | 1.6uM (0.3) | NI | NI | 3.7uM (0.26) | 0.30uM (0.16) | 0.78uM (0.18) | 0.62 uM (0.33) | 0.75 at 20uM | 3.8uM (0.45) | 2.3uM (0.28) | 2.8uM (0.35) |
| **LogP** | - | 4.82 | 4.05 | 2.77 | 3.81 | 5.14 | 4.72 | 5.18 | 3.67 | 3.8 | 4.57 | 4.3 |
| **MW** | - | 321.4 | 325.41 | 352.44 | 281.36 | 350.25 | 331.42 | 345.45 | 281.36 | 311.39 | 331.42 | 307.4 |

Figure 20: MTT Screen of CM10 Analogues at 30 µM
Screen conducted on MDA-MB-468 cells plated on Matrigel. MTT assay performed after 5-day incubation.

CM10 and two of its analogues, 3988-0485 and 3988-0486, were tested for a dose dependent response in the MTT assay (Figure 21). The three inhibitors demonstrated a clear dose-dependent response and relatively high effectiveness. All three compounds showed a similar degree of potency, having $LC_{50}$ values of 18 ± 4.3 µM for CM10, 13 ± 6.1 µM for 3988-0485, and 15 ± 2.3 µM for 3988-0486. CM10, however, demonstrated a visibly lower Hill slope than its two analogues.

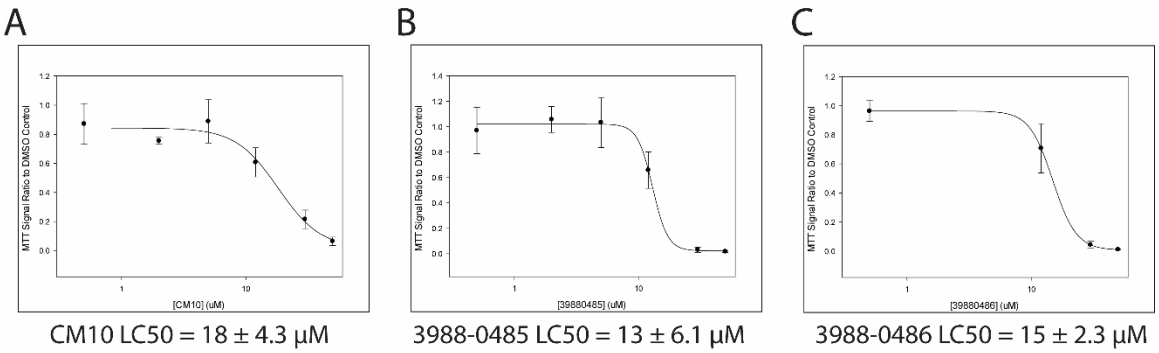



Figure 21: Dose-Response Curves of CM10 Series Compounds
Assay conducted on MDA-MB-468 cells plated on matrigel. MTT assay performed after 5-day incubation. A) Dose-response curve of CM10 B) Dose-response curve of CM10 analogue 3988-0485 C) Dose-response curve of CM10 analogue 3988-0486

## IV. Discussion

### A. Structural Features of the ALDH1 Active Site in Inhibitor Design

Despite the many known roles of aldehyde dehydrogenases, the specific roles of ALDH1 isoenzymes in healthy or pathogenic pathways remain unclear. Isoenzyme-selective inhibitors could be a key tool to help delineate the roles of each isoenzyme in biological processes. In spite of their potential importance, currently, no commercially available inhibitors show selectivity for the ALDH1A subfamily or individual ALDH1A isoenzymes (39). Available structural data suggests that the active site of ALDH1A features key differences when compared to ALDH2 and ALDH3 isoenzymes, indicating that it may be possible to develop inhibitors with the desired selectivity. As such, development of isoenzyme-selective ALDH1A inhibitors is a promising and yet-unexplored niche in aldehyde dehydrogenase research.

Currently available x-ray crystallography structures include those of ALDH1A1, ALDH2, and ALDH3A1. Comparing the active sites of the three structures, some major differences in the architecture of the active site become obvious (Figure 22). The ALDH1A1 structure features an open, conical substrate-binding pocket. The ALDH2 active site is more cylindrical, while the ALDH3A1 site is even more narrow and twisting. ALDH1A1 has the most accessible active site, meaning that modulators and substrates with large, planar structures are likely to be selective for the ALDH1 family. CM38 and CM10 compound series, which both feature planar multi-ring systems in the scaffold, are likely selective for ALDH1 enzymes because of this topological difference in shape of the substrate-binding pocket. Notably, while CM10 series compounds are selective against ALDH2, some still inhibit ALDH1B1. This suggests that while ALDH1B1 is, phylogenetically, more closely related to ALDH2, its substrate-binding site may actually resemble that of the ALDH1A subfamily. Thus, the CM10 SAR provides some insight into the structure of ALDH1B1, for which a crystal structure is not yet available.

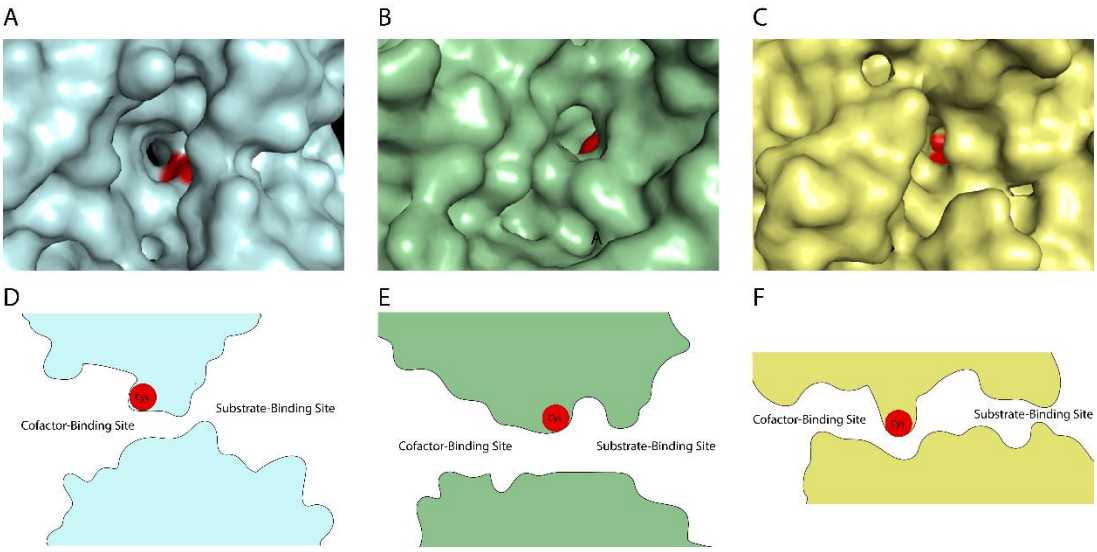



Figure 22: Architecture of the ALDH1A1, ALDH2, and ALDH3A1 Substrate-Binding Sites
Structures obtained from PDB. PDB IDs: 4WJ9, 1O04, 3SZA (104, 117, 118).
A) Active site of ALDH1A1, with the catalytic cysteine highlighted in red. B) Active site of ALDH2, with the catalytic cysteine highlighted in red. C) Active site of ALDH3A1, with the catalytic cysteine highlighted in red. D) Cross-section of the ALDH1A1 active site, demonstrating the cone-shaped, accessible substrate-binding pocket E) Cross-section of the ALDH2 active site, showing the cylindrical structure F) Cross-section of the ALDH3A1 active site, showing the narrow, winding architecture.

The binding mode of compounds discussed here can be viewed as an effort to map out useful structural features of the active site and facilitate the structure-based drug design of future compounds. The crystal structures obtained for CM38 and CM10 compounds can be used to identify key residues in the active site that can improve potency and selectivity of modulators. CM38 makes contact with Cys302 and π-stacking interactions with Tyr297 (Figure 23). Both of these interactions are well conserved in the ALDH1 and ALDH2 families, as Cys302 is perfectly retained and Tyr297 is only substituted by other aromatic residues (Table 10). In ALDH3A1, neither of these contacts are conserved, suggesting that the altered shape of the active site is not the only reason for the lack of inhibition. Furthermore, the piperidine ring of CM38 occupies a hydrophobic pocket shaped by Phe171, Trp178, Val460, and Phe466. This hydrophobic pocket is conserved in ALDH1 and ALDH2 families, but not in ALDH3A1 (Table 10). However, Val460 is substituted to the larger Leu in ALDH1A2 and ALDH1A3 and the bulkier Phe in ALDH2; this may reduce the space available for the piperidine ring and contribute to the ALDH1A1 selectivity of this compound. Despite their relevance, none of the residues mentioned so far are able to fully explain the excellent ALDH1A1 selectivity of the CM38 compound series.

In fact, the predominant residue responsible for the selectivity is likely Gly458. Gly458 has previously been found to play a key role in facilitating the binding of ALDH1A1-selective inhibitors (107). The absence of a side chain increases the room available in the ALDH1A1 active site and allows the binding of large, planar compounds. Gly458 is not conserved in other ALDH1 and ALDH2 isoenzymes, meaning that occupying the space afforded by it allows high ALDH1A1 selectivity. This is the case with CM38, as the three-ring core of the compound would be sterically hindered by the presence of a side chain at the 458 position.

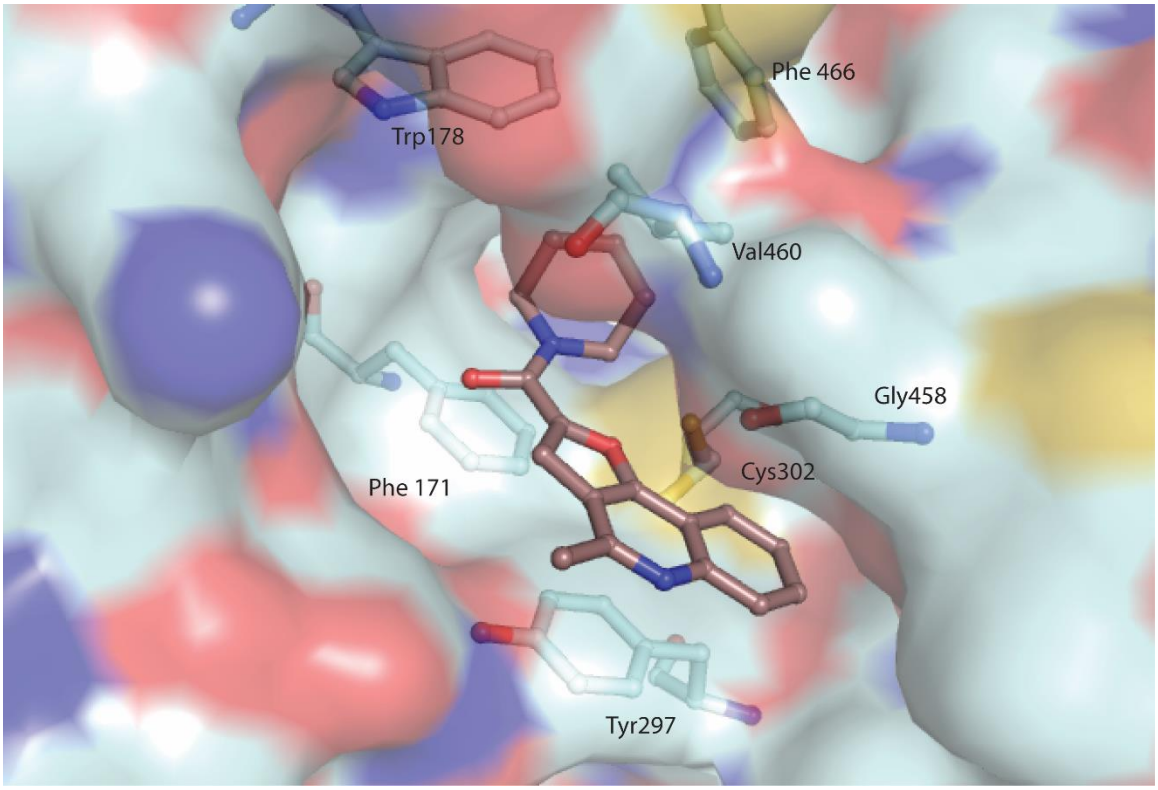



Figure 23: Key Residues in the Binding of CM38 to ALDH1A1

Table 10: Conservation of Key Residues in the ALDH1A1 Active Site in CM38 Binding
Identical residues are labeled green, similar labeled yellow, and non-conserved labeled red

| ALDH1A1 | Phe171 | Trp178 | Tyr297 | Cys302 | Gly458 | Val460 | Phe466 |
|---|---|---|---|---|---|---|---|
| ALDH1A2 | Phe | Trp | Phe | Cys | Asn | Leu | Phe |
| ALDH1A3 | Phe | Trp | Phe | Cys | Asn | Leu | Phe |
| ALDH1B1 | Phe | Trp | Phe | Cys | Asn | Val | Phe |
| ALDH2 | Phe | Trp | Phe | Cys | Asp | Phe | Phe |
| ALDH3A1 | Tyr | Gln | Met | Thr | His | Thr | Phe |

The crystal structure of the CM38 binding mode can be further integrated with its kinetic properties and the structure-activity relationship of its analogues in order to gain further insight and check for consistency. The binding mode in the crystal structure is in agreement with the results previously found in steady-state kinetic experiments, as the CM38 compound was found to be uncompetitive with respect to $NAD^+$, and, thus, was unlikely to be found in the cofactor-binding site. One of the findings of the CM38 SAR was that the piperidine ring is crucial for inhibition; loss of the ring, addition of polar elements or expansion with bulky groups compromised inhibition (Figure 24). This is in perfect agreement with the crystal structure, as the piperidine ring was found to occupy a well-defined hydrophobic pocket. The results of the SAR and the crystal structure suggest that this hydrophobic interaction is one of the primary drivers of CM38 binding. The oxygen and nitrogen atoms in the three-ring structure were not found to be critical, as removing them or altering their position had only a minor effect on potency. The crystal structure, again, is largely in agreement. While the oxygen may make polar contacts with Cys302 and provide a minor benefit to potency, the nitrogen makes no observable contribution to compound binding. The methyl group on the pyrrole was thought to contribute to selectivity against ALDH2. The crystal structure has no clear explanation for this, as any binding to ALDH2 would have an altered binding mode, due to the alternate active site architecture (Figure 22). It is likely that the absence of this methyl makes the compound narrow enough to allow suboptimal binding to the more cylindrical ALDH2 substrate-binding site. Another finding of the CM38 SAR suggested that the substitution of the terminal ring in the three-ring body of the compound with a single-unsaturated oxane and a carbonyl side-chain actually improved potency, as seen in compounds E003-0974 and C893-0789. This observation cannot be validated with the structure of the CM38-ALDH1A1 co-crystal. It is highly unlikely that E003-0974 and C893-0789 have an identical binding mode to CM38. The methyl on the furan and the carbonyl on the oxane would sterically clash with Cys302. Furthermore, there is not enough space to accommodate the additional alkyl ring seen in E003-0974. As a result, these compounds likely feature a binding mode where the main body of the compound is shifted or rotated. An additional crystal structure is needed to find the binding mode of these compounds and rationalize their improved potency.

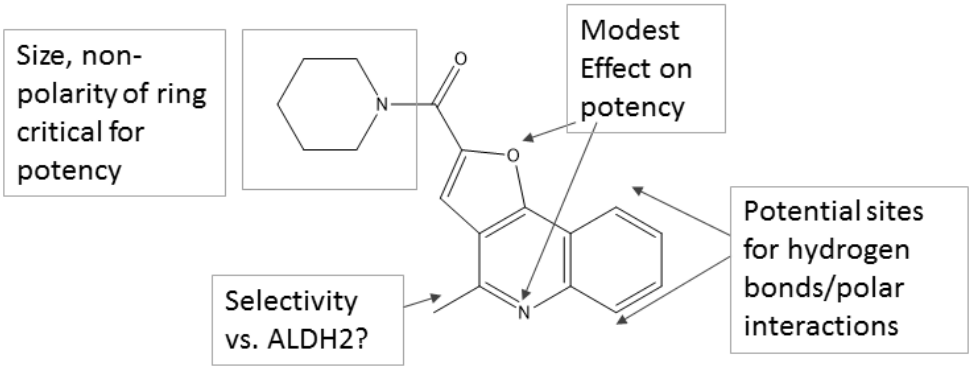



Figure 24: Summary of CM38 Structure-Activity Relationship

The CM10 series appears to bind in a slightly different location as compared to CM38 (Figure 25). Likely the primary interaction for CM10 binding is a hydrogen bond between the phenyl hydroxyl of the compound and the amine of Trp178. Trp178 is perfectly conserved in ALDH1 and ALDH2 enzymes, explaining the relatively broad selectivity profiles of CM10 analogues (Table 11). The structure also has a π-stacking interaction with Phe171, which is, again, conserved in the ALDH1 and ALDH2 families. Notably, the 3988-0485 compound in the obtained crystal structure does not bind immediately adjacent to Gly458, which is a key site for ALDH1A1-selective compounds. This likely explains why 3988-0485, like many other compounds in the CM10 series, is able to inhibit ALDH1A2 and ALDH1A3. Instead, the 3988-0485 compound occupies a space afforded by the lack of a side chain on Gly125. Unlike Gly458, Gly125 is conserved within the ALDH1A subfamily, but not by the other ALDH1 and ALDH2 isoenzymes (Table 11). Since the presence of a side chain at the 125 position would sterically hinder the binding of 3988-0485, Gly125 likely plays a predominant role in defining the ALDH1A selectivity of many CM10 series compounds.

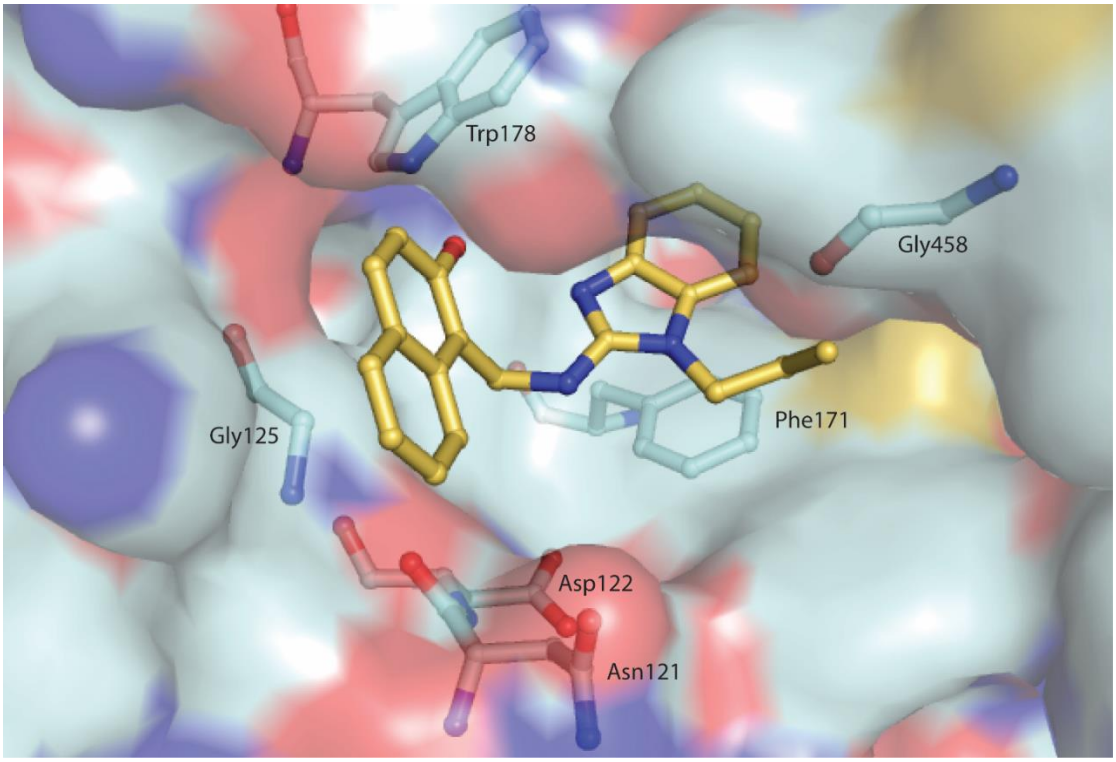



Figure 25: Key Residues in the Binding of 3988-0485 to ALDH1A1

Table 11: Conservation of Key Residues in the ALDH1A1 Active Site in 3988-0485 Binding
Identical residues are labeled green, similar labeled yellow, and non-conserved labeled red

| ALDH1A1 | Asn121 | Asp122 | Gly125 | Phe171 | Trp178 | Gly458 |
|---|---|---|---|---|---|---|
| ALDH1A2 | Val | Asp | Gly | Phe | Trp | Asn |
| ALDH1A3 | Ile | Asp | Gly | Phe | Trp | Asn |
| ALDH1B1 | Leu | Asp | Glu | Phe | Trp | Asn |
| ALDH2 | Val | Asp | Met | Phe | Trp | Asp |
| ALDH3A1 | Glu | Glu | Tyr | Tyr | Gln | His |

As was the case with CM38, the CM10 SAR is largely in agreement with the proposed binding mode of the series. The mixed partial inhibition with respect to $NAD^+$ shown in steady state kinetics experiments is consistent with this structure, as the compound is not competing with $NAD^+$ for binding. In contrast to CM38, the CM10 series features a more diverse selectivity profile, and makes use of more sites in the ALDH1A1 catalytic pocket that can be exploited to alter compound specificity. The main group necessary for binding was found to be the hydroxyl on the phenol group (Figure 26). This is explained by the crystal structure, in which this hydroxyl forms a hydrogen bond with the amine of Trp178. This interaction is likely to be the primary

determinant of compound potency of the CM10 series. Based on the SAR, it was thought that there is room to expand the compound by addition of groups onto the phenol. The compound in the structure, 3988-0485, actually features a benzene at this position. The crystal structure shows that there is sufficient room to allow these bulky adducts, as they face largely out of the substrate-binding site into solvent. The branching of the linker between the aromatic groups with aliphatic side chains was shown to almost entirely eliminate ALDH1A1 inhibition, while leaving ALDH1A2 and ALDH1A3 binding relatively intact. In the crystal structure, this branched group would come in close contact with Asn121, which, being polar, would not accommodate an aliphatic group. However, ALDH1A2 and ALDH1A3 feature Val and Ile, respectively, at this position (Table 11). These non-polar residues would be able to form hydrophobic interactions with an aliphatic group and, therefore, better accommodate the branch. It would be interesting to observe if branching with a polar group or a hydrogen-bonding partner would have the reverse effect and improve ALDH1A1 selectivity. Another structural feature important for selectivity was the alkyl side chain on the two-ring structure. It was found that lengthening or branching this group improved ALDH1A1 selectivity. The crystal structure provides a likely mechanism for this observation. While the 3988-0485 structure does not occupy the space adjacent to Gly458, the alkyl side chain extends in that direction. It is likely that an extended chain occupies this space and, therefore, faces steric clashes from side chains in other isoenzymes.

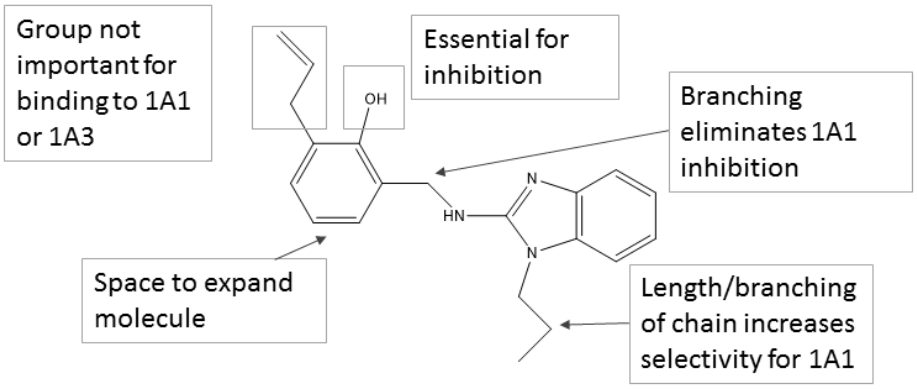



Figure 26: Summary of CM10 Structure-Activity Relationship

Overall, the crystal structures and structure-activity relationships for the CM38 and CM10 series identify a number of structural features in the ALDH1A1 active site that may be

taken advantage of in modulator design. Cys302 and Trp178 are able to make contacts with polar groups, while Phe171 and Tyr297 are available for π-stacking interactions. Slightly deeper than the catalytic cysteine in the active site, there exists a hydrophobic pocket that can accommodate six or seven-membered rings. Asn121, Gly125, and Gly458 are three residues that vary within the ALDH1 and ALDH2 families, thereby allowing the design of isoenzyme-selective inhibitors.

**B. CM38 and CM10 Analogues as Lead Compounds in Chemical Tool and Drug Development**

Cancer stem cells are a proposed subpopulation in some cancer types, which may contribute to chemotherapy and radiotherapy resistance, cancer relapse, and metastasis. ALDH1 enzymes have been successfully used as markers of cancer stem cells in multiple tumor types, but have been largely unexplored as drug targets. Furthermore, the specific functions of ALDH1 isoenzymes in cancer biology remain unclear, making it difficult to design specific therapies. While the CM38 and CM10 compound series are valuable as a means of mapping the ALDH1 active site, they can also be useful as putative lead compounds in chemical tool and drug development. The ability to use these compounds in living systems would allow the identification of specific roles of ALDH1 isoenzymes in healthy and disease pathways, as well as potentially the development of compounds to selectively target cancer stem cells. To this end, the efficacy of these compound series needs to be fully evaluated in cell and animal models.

An MTT assay was conducted on MDA-MB-468 cells grown on Matrigel in the presence of CM38 analogues. The screen showed warning signs of a potential source of general toxicity in the CM38 scaffold, as compounds that did not inhibit ALDH1 in vitro still inhibited cell proliferation. CM38 compounds feature an α,β-unsaturated carbonyl group, which is a potential source of reactivity for the compound series (Figure 27). The CM38 series should be pursued with caution for drug development until this issue is further investigated.

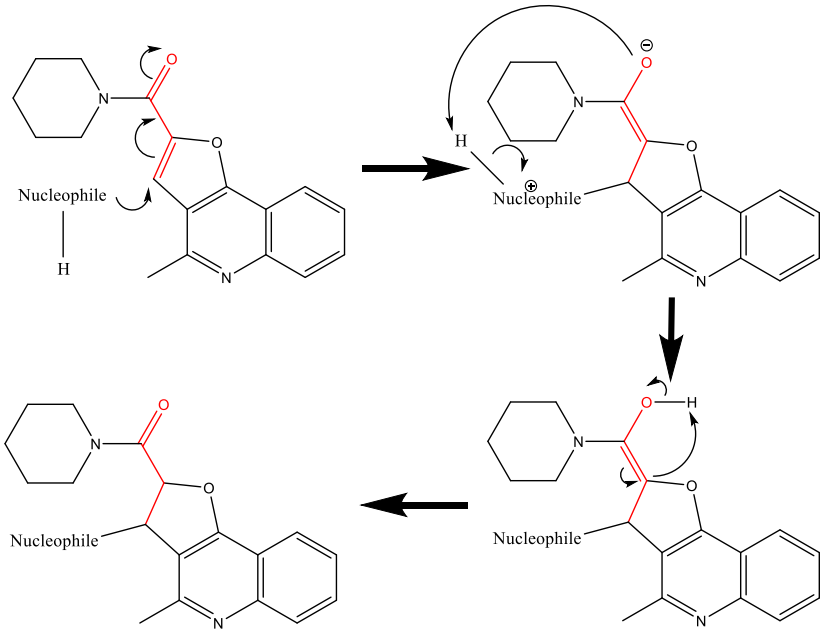
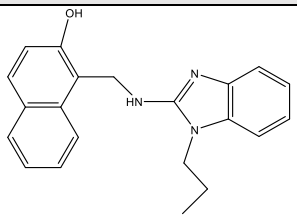
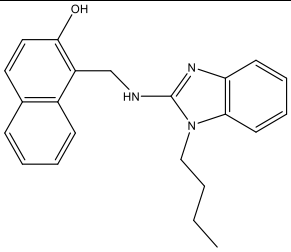



Figure 27: Potential Mechanism of the Reactivity of the CM38 Scaffold
Highlighted in red is the α,β-unsaturated carbonyl group, a potential source of reactivity and toxicity in the CM38 series

The CM10 series was also tested in an identical MTT screen. Unlike the CM38 series, the CM10 series showed no clear warning signs of potential toxicity in the screen. The negative control compounds were not effective and had MTT signal ratios to DMSO control of 0.8 and above. Furthermore, the effectiveness of the compounds in the MTT assay was generally correlated with the potency of ALDH1A1 inhibition and the logP. This is a positive sign, suggesting that the ability to inhibit ALDH1A1 is the primary reason for the effect observed in the assay. It also likely indicates that ALDH1A1 is the predominant ALDH1A enzyme in the MDA-MB-468 cell line, as compounds that were selective for ALDH1A2 or ALDH1A3 were largely ineffective. CM10, 3988-0485 and 3988-0486 were found to be the most potent in the MTT screen, having LC50 values of 18 ± 4.3 µM, 13 ± 6.1 µM, and 15 ± 2.3 µM, respectively. 3988-0485 and 3988-0486 are very promising compounds for future development as chemical tools. The two have good Lipinski properties, having 4 hydrogen bond acceptors, two donors and a molecular weight well below 500Da (Table 12). 3988-0485 has a logP of 4.69, while 3988-0486 narrowly exceeds the Lipinski guidelines at 5.11. The inhibition properties of these compounds make them very interesting as a pair. 3988-0485 inhibits ALDH1A1, ALDH1A2 and ALDH1A3, while 3988-0486 is an almost perfectly selective ALDH1A1 inhibitor. Despite the differences in

selectivity, the two display nearly identical potency toward ALDH1A1, having $IC_{50}$ values of 780 ± 78 nM for 3988-0485 and 620 ± 59 nM for 3988-0486. On top of that, the two compounds are nearly identical structurally, having only a one-methyl difference between them. Altogether, these properties make this pair of compounds extremely useful for comparing the role of ALDH1A1 with that of ALDH1A2 and ALHD1A3 in various cell lines. For example, in the MDA-MB-468 cell line used here, the two have nearly identical LC50 values, suggesting that ALDH1A2 and ALDH1A3 do not significantly contribute to the survival and stemness properties of the cell line. In a cell line where ALDH1A1 did not play a dominant role, 3988-0485 would be expected to be more effective than 3988-0486. Beside the immediate usefulness of these compounds as chemical tools, the potency, effectiveness in cells, and Lipinski properties of these compounds make them highly promising for future pharmaceutical development.

Table 12: Lipinski Properties of Select CM10 Analogues

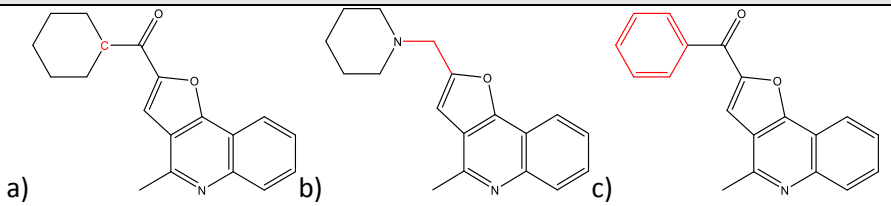
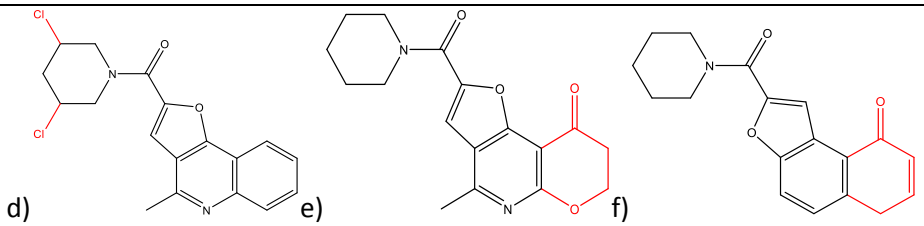
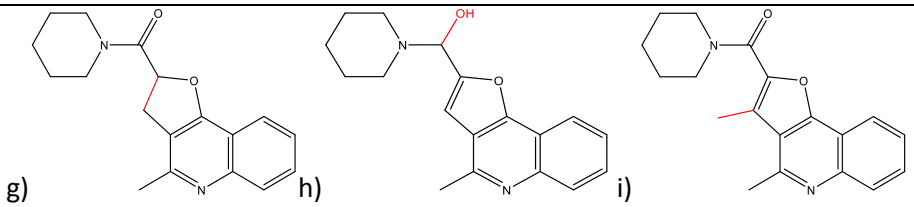
| Compound | Structure | Freely Rotatable Bonds | H-bond Acceptors | H-Bond Donors | LogP | MW (Da) |
|---|---|---|---|---|---|---|
| 3988-0485 |  | 5 | 4 | 2 | 4.69 | 331.42 |
| 3988-0486 |  | 6 | 4 | 2 | 5.11 | 345.45 |

**C. Future Directions**

Although much progress has been made to characterize and develop the CM38 and CM10 series, further experiments are still needed to complete this work. In particular, future directions should include expanding the structure-activity relationships with more compounds, crystallization of new analogues, and cell culture experiments to investigate the selectivity of compounds in cells.

The most straightforward way to continue this project is to expand the structure-activity relationships. New analogues can be used to investigate yet-unexplored structural features of the compound or to attempt to capitalize on the things already learned in order to produce superior compounds. For the CM38 series, the specific values of the nitrogen in the piperidine ring or the carbonyl were never explored in the structure-activity relationship. Replacing the nitrogen with a carbon or eliminating the carbonyl group would serve to verify the roles of these groups in ALDH1 inhibition (Table 13a,b). Furthermore, it may be possible to replace the piperidine ring with a benzene ring (Table 13c). Although it would be less flexible and perhaps a worse fit for the hydrophobic pocket, it may also form π-stacking interactions with the nearby Phe171 and improve affinity. Another prospect is to use the available SAR of CM38 and its crystal structure to guide the design of more potent ALDH1A1 inhibitors. The hydrophobic interaction with the piperidine ring plays a key role in binding; therefore, it may be possible to improve affinity by enhancing the hydrophobicity of the ring (such as with the addition of halogens like chloride) (Table 13d). Compounds E003-0974 and C893-0789 show that substitution of the terminal ring with a single-unsaturated oxane containing a carbonyl side-chain improved potency. It may be possible to produce a better inhibitor by applying this change to the more potent CM38 scaffold, rather than the 7998070 scaffold (Table 13e). Since it is unlikely that both the carbonyl and the oxane are equally responsible for the improved inhibition, one approach may be to keep only one of these groups in new analogues (Table 13f). As stated earlier, the α,β-unsaturated carbonyl group is a potential liability in the CM38 scaffold. Analogues can be used to investigate if this group can be eliminated without sacrificing inhibition. Possible approaches include reducing the α,β-unsaturated bond or replacing the carbonyl with an alcohol (Table 13g, h). It may also be possible to sterically occlude the potential nucleophilic attack with the addition of a group, such as a methyl, onto the furan ring (Table 13i).

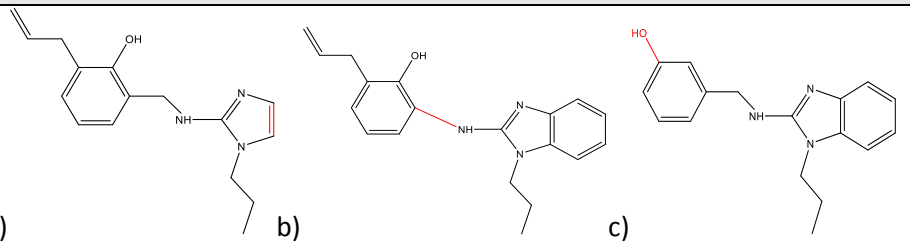
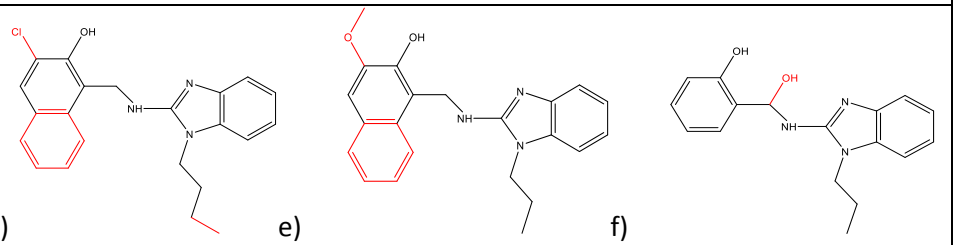
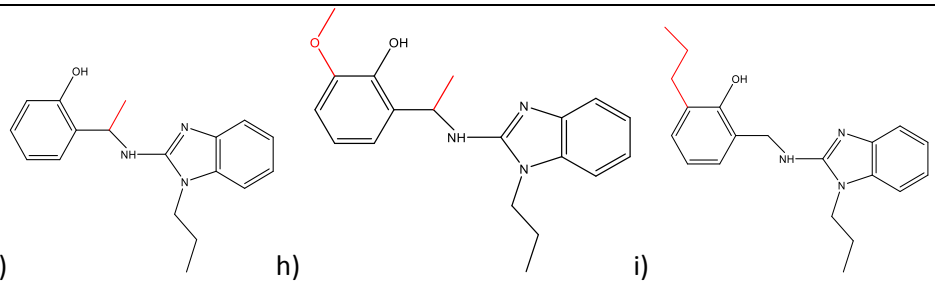
Table 13: Potential Compounds to Expand the Structure-Activity Relationship of CM38

| Objective | Possible Analogue Structures |
|---|---|
| Test new structural features |  a) b) c) |
| Improve potency |  d) e) f) |
| Eliminate α,β-unsaturated carbonyl group |  g) h) i) |

A similar approach of ordering new analogues may also be taken with the CM10 series. The importance of the benzene ring in the benzimidazole group may be tested by trying an analogue with only an imidazole (Table 14a). The length of the linker between the ring structures can also be varied (Table 14b). Although the crucial nature of the hydroxyl has been verified, its position has so far not been varied. A hydroxyl at the meta position may still be able to hydrogen bond with Trp178, and should be checked as a possible improvement (Table 14c). The potency of the CM10 series may be improved by combining features from different potent analogues. 3988-0474 was the most potent analogue, but also the least selective. By combining the chloride groups that gave it improved potency with the benzene and methyl groups that have been shown to improve ALDH1A1 selectivity, it may be possible to produce a more potent ALDH1A1 inhibitor (Table 14d). Similarly, by combining the ether of compound 3988-0587 and the benzene of compound 3988-0485, it may be possible to surpass the potency of either parent compound (Table 14e). Lastly, while a methyl branch at the linker has been shown to eliminate ALDH1A1 inhibition, other types of branches have not been attempted. A hydroxyl at this position may be able to hydrogen bond with Asn121 and improve potency of ALDH1A1

inhibition (Table 14f). While compound 3988-0486 has already demonstrated optimal ALDH1A1 inhibition, the series also shows untapped potential in developing a highly selective ALDH1A3 inhibitor. A methyl branch at the linker has been shown to increase ALDH1A3 selectivity, but has not been tested with a standard 3988-0434 scaffold (Table 14g). It may also be possible to combine features from two good ALDH1A3 inhibitors, 3988-0580 and 3988-0587, in order to produce an even more selective ALDH1A3 inhibitor (Table 14h). The original CM10 compound showed better inhibition of ALDH1A2 and ALDH1A3 than toward ALDH1A1. Although the alkene group was eliminated in order to remove a potential metabolic liability, it may be possible to recapture the potency toward ALDH1A2 and ALDH1A3 by replacing the alkene with an alkyl (Table 14i).

Table 14: Potential Compounds to Expand the Structure-Activity Relationship of CM10

| Objective | Possible Analogue Structures |
|---|---|
| Test new structural features |  a) b) c) |
| Improve potency |  d) e) f) |
| Improved ALDH1A3 selectivity |  g) h) i) |

Aside from new analogues, other aspects of this project should also be addressed in further research. New crystal structures would benefit the validation and development of the structure-activity relationships. A better-occupancy structure of any CM10 series compound

should be pursued. Since compound occupancy in the crystal is the main issue, co-crystals with a more hydrophilic compound may help inhibitor penetration of the crystal. A good candidate is compound 3988-0587, which is the most hydrophilic of the ALDH1A1 inhibitors in the current SAR. For the CM38 series, new crystal structures with compounds E003-0974 and C893-0789 would confirm whether or not these compounds assume that same binding mode as CM38.

Another gap in the data on the CM38 and CM10 compound series is the lack of information on their selectivity in cells. This is especially an issue for the CM38 series, as the initial cell culture screen suggested that the scaffold may cause off-target toxicity. One approach is to test analogues without the α,β-unsaturated carbonyl group, see if the apparent toxicity persists. A more universal approach would be to attempt to rescue the cell line following compound treatment. ALDH1A1 overexpression in the cell line should be able to rescue the cells from the effects of the CM38 and CM10 compounds, if these compounds are acting selectively on ALDH1A1. However, since the overexpressed ALDH1A1 would be able to bind and sequester the compounds even if they have an off-target effect, this approach would not provide a definitive answer. An optimal strategy would be to transfect the cell line with an ALDH1A1 mutant that is resistant to inhibition by the compounds. For the CM38 analogues, mutation of Gly458 to any residue with a bulky sidechain would likely eliminate inhibition. A G458N mutant was previously used in a similar strategy for the CM26 and CM37 compounds (107). For the CM10 series, the ALDH1A1-selective 3988-0486 would be easiest to use in this type of experiment, as it shows excellent selectivity for one isoenzyme. Compound 3988-0486 would also likely be unable to inhibit a G458 mutant. If the selectivity of other CM10 compounds needs to be evaluated, a G125 mutant may be used in a similar fashion; however, transfection with multiple enzyme mutants may need to be done if the cell line is reliant on multiple ALDH1 isoenzymes. Overall, transfection with an inhibitor-resistant enzyme mutant would be expected to rescue the cells following compound treatment if the compounds do not have significant off-target effects, thus settling the issue on in-cell compound selectivity. An entirely alternate approach would be the use of a counter cell line. A “normal” human breast cell line would not be expected to be reliant on ALDH1A1 activity, meaning that selective ALDH1A1 inhibitors should not have a toxic effect.

Other cell culture assays can also be used to evaluate the effect of these compounds on cell properties and signalling pathways. The compounds may be tested in other cell lines, such

as ovarian, lung, and colorectal to verify that they can be effective in multiple cancer cell types. Spheroid formation is a cell property indicative of cancer stemness; as such, the impact of CM10 and CM38 series compounds on a spheroid formation assay may be evaluated. An ALDH1A1 enzymatic inhibitor has previously been shown to decrease spheroid formation in an ovarian cancer cell line, validating this type of experimental approach (93). Similarly, cell invasion may be evaluated. Commercially available matrigel invasion assays (Corning Inc., Corning, NY) may be used to evaluate the effect of compounds on the invasiveness of cell lines. These types of experiments would show the effect of these compounds, and ALDH1A inhibition in general, in a greater context. While the compounds presented have been shown to be effective at reducing proliferation in a cell line, further experiments are needed to show their effect on other stemness properties. By conducting experiments to verify their effect on more clinically relevant properties, such as spheroid formation and invasiveness, a much stronger case for the use of these compounds in drug development could be made.

The effect of these inhibitors on pathways related to retinoic acid signalling should also be evaluated. qPCR can be used to measure the transcriptional expression of classical and non-classical RA-pathway target genes following compound treatment, to evaluate the ways in which these inhibitors are able to affect RA signalling. For example, ALDH1A1 inhibitors may be able to reduce the expression of non-classical targets c-MYC, Cyclin D, and PDK1, verifying the non-classical RA pathway as a drug target in cancer. Similarly, the activation of pathways that have been linked to RA signalling may be evaluated. For example, by collecting cell lysate following compound treatment and running a western using a P-AKT antibody, the relative activation of the PI3K pathway in treated and non-treated cells can be compared (119). These types of experiments would help validate the non-classical RA pathway as a drug target in cancer, and help establish CM38 and CM10 analogues as valuable chemical tools in the study of cancer stem cell signalling.

**D. Conclusion**

Enzymatic inhibition of ALDH1A enzymes, which are established cancer stem cell markers, is a promising, but yet-unexplored avenue to therapeutically target cancer stem cells in clinical treatment. The ALDH1A1-selective inhibitor, CM38 and the ALDH1A-selective inhibitor, CM10, have been characterized by establishing structure-activity relationships, finding the binding modes using X-ray crystallography, and verifying their effectiveness in a cancer cell line.

The compounds have helped establish structural features of the ALDH1A1 active site, namely the isoenzyme-specific residues Asn121, Gly125, and Gly458, that will be useful in the development of additional isoenzyme-selective inhibitors. The compound series show promise as chemical tools to help delineate the roles of ALDH1 isoenzymes in cancer biology and as lead compounds in drug development; however, further experiments are needed to verify their in-cell selectivity and their impact on cancer cell properties and pathways.

## References


1. O'Brien PJ, Siraki AG, Shangari N. Aldehyde Sources, Metabolism, Molecular Toxicity Mechanisms, and Possible Effects on Human Health. Critical Reviews in Toxicology. 2005;35(7):609-62.
2. Ceni E, Mello T, Galli A. Pathogenesis of alcoholic liver disease: Role of oxidative metabolism. World Journal of Gastroenterology : WJG. 2014;20(47):17756-72.
3. Sapkota M, Wyatt TA. Alcohol, Aldehydes, Adducts and Airways. Biomolecules. 2015;5(4):2987-3008.
4. Jing Z, Quan T, Sui Yung C, Shu Chuen L, Shufeng Z, Wei D, et al. Metabolism and Transport of Oxazaphosphorines and the Clinical Implications. Drug Metabolism Reviews. 2005;37(4):611-703.
5. Ludeman SM. The chemistry of the metabolites of cyclophosphamide. Current pharmaceutical design. 1999;5(8):627-43.
6. Schaur RJ, Siems W, Bresgen N, Eckl PM. 4-Hydroxy-nonenal—A Bioactive Lipid Peroxidation Product. Biomolecules. 2015;5(4):2247-337.
7. Thornalley PJ. Pharmacology of methylglyoxal: formation, modification of proteins and nucleic acids, and enzymatic detoxification--a role in pathogenesis and antiproliferative chemotherapy. General pharmacology. 1996;27(4):565-73.
8. Das BC, Thapa P, Karki R, Das S, Mahapatra S, Liu T-C, et al. Retinoic acid signaling pathways in development and diseases. Bioorganic & Medicinal Chemistry. 2014;22(2):673-83.
9. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert opinion on drug metabolism & toxicology. 2008;4(6):697-720.
10. Marnett LJ. Oxy radicals, lipid peroxidation and DNA damage. Toxicology. 2002;181–182:219-22.
11. Brooks PJ, Zakhari S. Acetaldehyde and the genome: Beyond nuclear DNA adducts and carcinogenesis. Environmental and Molecular Mutagenesis. 2014;55(2):77-91.
12. Dalle-Donne I, Aldini G, Carini M, Colombo R, Rossi R, Milzani A. Protein carbonylation, cellular dysfunction, and disease progression. Journal of cellular and molecular medicine. 2006;10(2):389-406.
13. Conklin D, Prough R, Bhatanagar A. Aldehyde metabolism in the cardiovascular system. Molecular BioSystems. 2007;3(2):136-50.
14. Mittal B, Tulsyan S, Kumar S, Mittal RD, Agarwal G. Chapter Four - Cytochrome P450 in Cancer Susceptibility and Treatment. In: Gregory SM, editor. Advances in Clinical Chemistry. Volume 71: Elsevier; 2015. p. 77-139.
15. Cederbaum AI. ALCOHOL METABOLISM. Clinics in liver disease. 2012;16(4):667-85.
16. Muzio G, Maggiora M, Paiuzzi E, Oraldi M, Canuto RA. Aldehyde dehydrogenases and cell proliferation. Free Radical Biology and Medicine. 2012;52(4):735-46.
17. Vasiliou V, Bairoch A, Tipton KF, Nebert DW. Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. Pharmacogenetics. 1999;9(4):421-34.
18. Hempel J, Perozich J, Chapman T, Rose J, Boesch JS, Liu ZJ, et al. Aldehyde dehydrogenase catalytic mechanism. A proposal. Advances in experimental medicine and biology. 1999;463:53-9.
19. Vasiliou V, Thompson DC, Smith C, Fujita M, Chen Y. Aldehyde dehydrogenases: From eye crystallins to metabolic disease and cancer stem cells. Chemico-biological interactions. 2013;202(0):2-10.

20. Blatter EE, Abriola DP, Pietruszko R. Aldehyde dehydrogenase. Covalent intermediate in aldehyde dehydrogenation and ester hydrolysis. The Biochemical journal. 1992;282 ( Pt 2):353-60.
21. Hanna MC, Blackstone C. Interaction of the SPG21 protein ACP33/maspardin with the aldehyde dehydrogenase ALDH16A1. neurogenetics. 2009;10(3):217-28.
22. Chen Y, Thompson DC, Koppaka V, Jester JV, Vasiliou V. Ocular Aldehyde Dehydrogenases: Protection against Ultraviolet Damage and Maintenance of Transparency for Vision. Progress in retinal and eye research. 2013;33:28-39.
23. Maden M. Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat Rev Neurosci. 2007;8(10):755-65.
24. Applegate CC, Lane MA. Role of retinoids in the prevention and treatment of colorectal cancer. World Journal of Gastrointestinal Oncology. 2015;7(10):184-203.
25. Harrison EH. Mechanisms involved in the intestinal absorption of dietary vitamin A and provitamin A carotenoids. Biochimica et biophysica acta. 2012;1821(1):70-7.
26. Kawaguchi R, Yu J, Honda J, Hu J, Whitelegge J, Ping P, et al. A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. Science. 2007;315(5813):820-5.
27. Amann PM, Eichmuller SB, Schmidt J, Bazhin AV. Regulation of gene expression by retinoids. Current medicinal chemistry. 2011;18(9):1405-12.
28. le Maire A, Bourguet W. Retinoic acid receptors: structural basis for coregulator interaction and exchange. Sub-cellular biochemistry. 2014;70:37-54.
29. Balmer JE, Blomhoff R. Gene expression regulation by retinoic acid. Journal of Lipid Research. 2002;43(11):1773-808.
30. Lo-Coco F, Cicconi L, Breccia M. Current standard treatment of adult acute promyelocytic leukaemia. British Journal of Haematology. 2016;172(6):841-54.
31. Xu X, Chai S, Wang P, Zhang C, Yang Y, Yang Y, et al. Aldehyde dehydrogenases and cancer stem cells. Cancer Letters.369(1):50-7.
32. Cassidy J, Lippman M, Lacroix A, Peck G. Phase II trial of 13-cis-retinoic acid in metastatic breast cancer. European journal of cancer & clinical oncology. 1982;18(10):925-8.
33. Budd GT, Adamson PC, Gupta M, Homayoun P, Sandstrom SK, Murphy RF, et al. Phase I/II trial of all-trans retinoic acid and tamoxifen in patients with advanced breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 1998;4(3):635-42.
34. Masiá S, Alvarez S, de Lera AR, Barettino D. Rapid, Nongenomic Actions of Retinoic Acid on Phosphatidylinositol-3-Kinase Signaling Pathway Mediated by the Retinoic Acid Receptor. Molecular Endocrinology. 2007;21(10):2391-402.
35. Radominska-Pandya A, Chen G, Czernik PJ, Little JM, Samokyszyn VM, Carter CA, et al. Direct interaction of all-trans-retinoic acid with protein kinase C (PKC). Implications for PKC signaling and cancer therapy. J Biol Chem. 2000;275(29):22324-30.
36. Schug TT, Berry DC, Shaw NS, Travis SN, Noy N. Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. Cell. 2007;129(4):723-33.
37. Black W, Vasiliou V. The Aldehyde Dehydrogenase Gene Superfamily Resource Center. Human Genomics. 2009;4(2):136-42.
38. Tomita H, Tanaka K, Tanaka T, Hara A. Aldehyde dehydrogenase 1A1 in stem cells and cancer2016.
39. Koppaka V, Thompson DC, Chen Y, Ellermann M, Nicolaou KC, Juvonen RO, et al. Aldehyde Dehydrogenase Inhibitors: a Comprehensive Review of the Pharmacology, Mechanism

of Action, Substrate Specificity, and Clinical Application. Pharmacological Reviews. 2012;64(3):520-39.
40. Chen Y, Koppaka V, Thompson DC, Vasiliou V. Focus on molecules: ALDH1A1: from lens and corneal crystallin to stem cell marker. Experimental eye research. 2012;102:105-6.
41. Januchowski R, Wojtowicz K, Zabel M. The role of aldehyde dehydrogenase (ALDH) in cancer drug resistance. Biomedicine & Pharmacotherapy. 2013;67(7):669-80.
42. Kim J-I, Ganesan S, Luo SX, Wu Y-W, Park E, Huang EJ, et al. Aldehyde dehydrogenase 1a1 mediates a GABA synthesis pathway in midbrain dopaminergic neurons. Science. 2015;350(6256):102-6.
43. Goldstein DS, Kopin IJ, Sharabi Y. Catecholamine autotoxicity. Implications for pharmacology and therapeutics of Parkinson disease and related disorders(). Pharmacology & therapeutics. 2014;144(3):268-82.
44. Goldstein DS, Sullivan P, Holmes C, Miller GW, Alter S, Strong R, et al. Determinants of buildup of the toxic dopamine metabolite DOPAL in Parkinson's disease. Journal of neurochemistry. 2013;126(5):591-603.
45. Lassen N, Bateman JB, Estey T, Kuszak JR, Nees DW, Piatigorsky J, et al. Multiple and Additive Functions of ALDH3A1 and ALDH1A1: CATARACT PHENOTYPE AND OCULAR OXIDATIVE DAMAGE IN Aldh3a1(−/−)/Aldh1a1(−/−) KNOCK-OUT MICE. The Journal of biological chemistry. 2007;282(35):25668-76.
46. Ziouzenkova O, Orasanu G, Sharlach M, Akiyama TE, Berger JP, Viereck J, et al. Retinaldehyde represses adipogenesis and diet-induced obesity. Nature medicine. 2007;13(6):695-702.
47. Kiefer FW, Vernochet C, O'Brien P, Spoerl S, Brown JD, Nallamshetty S, et al. Retinaldehyde dehydrogenase 1 regulates a thermogenic program in white adipose tissue. Nature medicine. 2012;18(6):918-25.
48. Kiefer FW, Orasanu G, Nallamshetty S, Brown JD, Wang H, Luger P, et al. Retinaldehyde Dehydrogenase 1 Coordinates Hepatic Gluconeogenesis and Lipid Metabolism. Endocrinology. 2012;153(7):3089-99.
49. Takitani K, Inoue K, Koh M, Miyazaki H, Inoue A, Kishi K, et al. Altered retinol status and expression of retinol-related proteins in streptozotocin-induced type 1 diabetic model rats. Journal of Clinical Biochemistry and Nutrition. 2015;56(3):195-200.
50. Wang X, Penzes P, Napoli JL. Cloning of a cDNA Encoding an Aldehyde Dehydrogenase and Its Expression in Escherichia coli: RECOGNITION OF RETINAL AS SUBSTRATE. Journal of Biological Chemistry. 1996;271(27):16288-93.
51. Niederreither K, Vermot J, Roux IL, Schuhbaur B, Chambon P, Dollé P. The regional pattern of retinoic acid synthesis by RALDH2 is essential for the development of posterior pharyngeal arches and the enteric nervous system. Development. 2003;130(11):2525-34.
52. Martin M, Gallego-Llamas J, Ribes V, Kedinger M, Niederreither K, Chambon P, et al. Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. Developmental biology. 2005;284(2):399-411.
53. Mic FA, Molotkov A, Molotkova N, Duester G. Raldh2 expression in optic vesicle generates a retinoic acid signal needed for invagination of retina during optic cup formation. Developmental dynamics : an official publication of the American Association of Anatomists. 2004;231(2):270-7.
54. Ribes V, Wang Z, Dolle P, Niederreither K. Retinaldehyde dehydrogenase 2 (RALDH2)-mediated retinoic acid synthesis regulates early mouse embryonic forebrain development by controlling FGF and sonic hedgehog signaling. Development. 2006;133(2):351-61.

55. Mathew LK, Sengupta S, Franzosa JA, Perry J, La Du J, Andreasen EA, et al. Comparative Expression Profiling Reveals an Essential Role for Raldh2 in Epimorphic Regeneration. The Journal of Biological Chemistry. 2009;284(48):33642-53.
56. Kikuchi K, Holdway JE, Major RJ, Blum N, Dahn RD, Begemann G, et al. Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. Developmental cell. 2011;20(3):397-404.
57. Graham Caroline E, Brocklehurst K, Pickersgill Richard W, Warren Martin J. Characterization of retinaldehyde dehydrogenase 3. Biochemical Journal. 2006;394(Pt 1):67-75.
58. Dupe V, Matt N, Garnier JM, Chambon P, Mark M, Ghyselinck NB. A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(24):14036-41.
59. Kumar S, Sandell LL, Trainor PA, Koentgen F, Duester G. Alcohol and Aldehyde Dehydrogenases: Retinoid Metabolic Effects in Mouse Knockout Models. Biochimica et biophysica acta. 2012;1821(1):198-205.
60. Matt N, Dupe V, Garnier JM, Dennefeld C, Chambon P, Mark M, et al. Retinoic acid-dependent eye morphogenesis is orchestrated by neural crest cells. Development. 2005;132(21):4789-800.
61. Molotkova N, Molotkov A, Duester G. Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone. Developmental biology. 2007;303(2):601-10.
62. Rosselot C, Spraggon L, Chia I, Batourina E, Riccio P, Lu B, et al. Non-cell-autonomous retinoid signaling is crucial for renal development. Development. 2010;137(2):283-92.
63. Molotkov A, Molotkova N, Duester G. Retinoic acid guides eye morphogenetic movements via paracrine signaling but is unnecessary for retinal dorsoventral patterning. Development (Cambridge, England). 2006;133(10):1901-10.
64. Jordan CT, Guzman ML, Noble M. Cancer Stem Cells. New England Journal of Medicine. 2006;355(12):1253-61.
65. Dragu DL, Necula LG, Bleotu C, Diaconu CC, Chivu-Economescu M. Therapies targeting cancer stem cells: Current trends and future challenges. World Journal of Stem Cells. 2015;7(9):1185-201.
66. Douville J, Beaulieu R, Balicki D. ALDH1 as a functional marker of cancer stem and progenitor cells. Stem cells and development. 2009;18(1):17-25.
67. Ajani JA, Song S, Hochster HS, Steinberg IB. Cancer stem cells: the promise and the potential. Seminars in oncology. 2015;42 Suppl 1:S3-17.
68. Chen K, Huang Y-h, Chen J-l. Understanding and targeting cancer stem cells: therapeutic implications and challenges. Acta Pharmacologica Sinica. 2013;34(6):732-40.
69. Swaminathan SK, Roger E, Toti U, Niu L, Ohlfest JR, Panyam J. CD133-targeted paclitaxel delivery inhibits local tumor recurrence in a mouse model of breast cancer. Journal of Controlled Release. 2013;171(3):280-7.
70. Brechbiel J, Miller-Moslin K, Adjei AA. Crosstalk between hedgehog and other signaling pathways as a basis for combination therapies in cancer. Cancer treatment reviews. 2014;40(6):750-9.
71. Sun Y, Mao X, Fan C, Liu C, Guo A, Guan S, et al. CXCL12-CXCR4 axis promotes the natural selection of breast cancer cell metastasis. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2014;35(8):7765-73.
72. Welschinger R, Liedtke F, Basnett J, Dela Pena A, Juarez JG, Bradstock KF, et al. Plerixafor (AMD3100) induces prolonged mobilization of acute lymphoblastic leukemia cells and increases

the proportion of cycling cells in the blood in mice. Experimental hematology. 2013;41(3):293-302.e1.
73. Moitra K. Overcoming Multidrug Resistance in Cancer Stem Cells. BioMed research international. 2015;2015:635745.
74. Alison MR, Guppy NJ, Lim SML, Nicholson LJ. Finding cancer stem cells: are aldehyde dehydrogenases fit for purpose? The Journal of Pathology. 2010;222(4):335-44.
75. Storms RW, Trujillo AP, Springer JB, Shah L, Colvin OM, Ludeman SM, et al. Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(16):9118-23.
76. Dollé L, Boulter L, Leclercq IA, van Grunsven LA. Next generation of ALDH substrates and their potential to study maturational lineage biology in stem and progenitor cells. American Journal of Physiology - Gastrointestinal and Liver Physiology. 2015;308(7):G573-G8.
77. Morgan CA, Parajuli B, Buchman CD, Dria K, Hurley TD. N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. Chem Biol Interact. 2015;234:18-28.
78. Liu Y, Lv D-l, Duan J-j, Xu S-l, Zhang J-f, Yang X-j, et al. ALDH1A1 expression correlates with clinicopathologic features and poor prognosis of breast cancer patients: a systematic review and meta-analysis. BMC Cancer. 2014;14:444-.
79. Kida K, Ishikawa T, Yamada A, Shimada K, Narui K, Sugae S, et al. Effect of ALDH1 on prognosis and chemoresistance by breast cancer subtype. Breast cancer research and treatment. 2016;156(2):261-9.
80. Zhong Y, Shen S, Zhou Y, Mao F, Guan J, Lin Y, et al. ALDH1 is a better clinical indicator for relapse of invasive ductal breast cancer than the CD44+/CD24- phenotype. Medical oncology (Northwood, London, England). 2014;31(3):864.
81. Charafe-Jauffret E, Ginestier C, Iovino F F, Tarpin C, Diebel M, Esterni B, et al. ALDH1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16(1):45-55.
82. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, 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):555-67.
83. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, et al. Aldehyde Dehydrogenase Activity of Breast Cancer Stem Cells Is Primarily Due To Isoform ALDH1A3 and Its Expression Is Predictive of Metastasis. STEM CELLS. 2011;29(1):32-45.
84. Marcato P, Dean CA, Liu R-Z, Coyle KM, Bydoun M, Wallace M, et al. Aldehyde dehydrogenase 1A3 influences breast cancer progression via differential retinoic acid signaling. Molecular Oncology. 2015;9(1):17-31.
85. Neumeister V, Agarwal S, Bordeaux J, Camp RL, Rimm DL. In Situ Identification of Putative Cancer Stem Cells by Multiplexing ALDH1, CD44, and Cytokeratin Identifies Breast Cancer Patients with Poor Prognosis. The American Journal of Pathology. 2010;176(5):2131-8.
86. Resetkova E, Reis-Filho JS, Jain RK, Mehta R, Thorat MA, Nakshatri H, et al. Prognostic impact of ALDH1 in breast cancer: a story of stem cells and tumor microenvironment. Breast cancer research and treatment. 2010;123:97+.
87. Wang Y-C, Yo Y-T, Lee H-Y, Liao Y-P, Chao T-K, Su P-H, et al. ALDH1-Bright Epithelial Ovarian Cancer Cells Are Associated with CD44 Expression, Drug Resistance, and Poor Clinical Outcome. The American Journal of Pathology. 2012;180(3):1159-69.

88. Sun Y, Jia X, Wu X. High Expressions of Lgr5 and ALDH1 in Primary Epithelial Ovarian Cancer Correlate with Advanced Tumor Stage and Grade as well as Poor Prognosis of the Patients. Gynecologic and Obstetric Investigation. 2016;81(2):162-8.
89. Ayub TH, Keyver-Paik M-D, Debald M, Rostamzadeh B, Thiesler T, Schröder L, et al. Accumulation of ALDH1-positive cells after neoadjuvant chemotherapy predicts treatment resistance and prognosticates poor outcome in ovarian cancer2015.
90. Chang B, Liu G, Xue F, Rosen DG, Xiao L, Wang X, et al. ALDH1 expression correlates with favorable prognosis in ovarian cancers. Mod Pathol. 2009;22(6):817-23.
91. Huang R, Li X, Holm R, Trope CG, Nesland JM, Suo Z. The expression of aldehyde dehydrogenase 1 (ALDH1) in ovarian carcinomas and its clinicopathological associations: a retrospective study. BMC Cancer. 2015;15:502.
92. Meng E, Mitra A, Tripathi K, Finan MA, Scalici J, McClellan S, et al. ALDH1A1 Maintains Ovarian Cancer Stem Cell-Like Properties by Altered Regulation of Cell Cycle Checkpoint and DNA Repair Network Signaling. PLoS ONE. 2014;9(9):e107142.
93. Condello S, Morgan CA, Nagdas S, Cao L, Turek J, Hurley TD, et al. [beta]-Catenin-regulated ALDH1A1 is a target in ovarian cancer spheroids. Oncogene. 2015;34(18):2297-308.
94. Wei D, Peng JJ, Gao H, Zhang T, Tan Y, Hu YH. ALDH1 Expression and the Prognosis of Lung Cancer: A Systematic Review and Meta-Analysis. Heart, lung & circulation. 2015;24(8):780-8.
95. Dimou A, Neumeister V, Agarwal S, Anagnostou V, Syrigos K, Rimm DL. Measurement of Aldehyde Dehydrogenase 1 Expression Defines a Group with Better Prognosis in Patients with Non-Small Cell Lung Cancer. The American Journal of Pathology. 2012;181(4):1436-42.
96. Jiang F, Qiu Q, Khanna A, Todd NW, Deepak J, Xing L, et al. Aldehyde Dehydrogenase 1 Is a Tumor Stem Cell-Associated Marker in Lung Cancer. Molecular cancer research : MCR. 2009;7(3):330-8.
97. Liang D, Shi Y. Aldehyde dehydrogenase-1 is a specific marker for stem cells in human lung adenocarcinoma. Medical oncology (Northwood, London, England). 2012;29(2):633-9.
98. Chen J, Xia Q, Jiang B, Chang W, Yuan W, Ma Z, et al. Prognostic Value of Cancer Stem Cell Marker ALDH1 Expression in Colorectal Cancer: A Systematic Review and Meta-Analysis. PLoS ONE. 2015;10(12):e0145164.
99. Lugli A, Iezzi G, Hostettler I, Muraro MG, Mele V, Tornillo L, et al. Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer. British Journal of Cancer. 2010;103(3):382-90.
100. Kahlert C, Gaitzsch E, Steinert G, Mogler C, Herpel E, Hoffmeister M, et al. Expression Analysis of Aldehyde Dehydrogenase 1A1 (ALDH1A1) in Colon and Rectal Cancer in Association with Prognosis and Response to Chemotherapy. Annals of Surgical Oncology. 2012;19(13):4193-201.
101. Huang EH, Hynes MJ, Zhang T, Ginestier C, Dontu G, Appelman H, et al. Aldehyde Dehydrogenase 1 Is a Marker for Normal and Malignant Human Colonic Stem Cells (SC) and Tracks SC Overpopulation during Colon Tumorigenesis. Cancer research. 2009;69(8):3382-9.
102. Ye XQ, Li Q, Wang GH, Sun FF, Huang GJ, Bian XW, et al. Mitochondrial and energy metabolism-related properties as novel indicators of lung cancer stem cells. International journal of cancer. 2011;129(4):820-31.
103. Diehn M, Cho RW, Lobo NA, Kalisky T, Dorie MJ, Kulp AN, et al. Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature. 2009;458(7239):780-3.
104. Morgan CA, Hurley TD. Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1. Chemico-Biological Interactions. 2015;234:29-37.

105. Parajuli B, Kimble-Hill AC, Khanna M, Ivanova Y, Meroueh S, Hurley TD. Discovery of Novel Regulators of Aldehyde Dehydrogenase Isoenzymes. Chemico-biological interactions. 2011;191(1-3):153-8.
106. Parajuli B, Georgiadis TM, Fishel ML, Hurley TD. Development of Selective Inhibitors for Human Aldehyde Dehydrogenase 3A1 (ALDH3A1) for the Enhancement of Cyclophosphamide Cytotoxicity. Chembiochem : a European journal of chemical biology. 2014;15(5):701-12.
107. Morgan CA, Hurley TD. Characterization of Two Distinct Structural Classes of Selective Aldehyde Dehydrogenase 1A1 Inhibitors. Journal of Medicinal Chemistry. 2015;58(4):1964-75.
108. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta crystallographica Section D, Biological crystallography. 2006;62(Pt 8):859-66.
109. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D. 2011;67(4):235-42.
110. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallographica Section D. 2004;60(12 Part 1):2184-95.
111. Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. Journal of Applied Crystallography. 2006;39(1):109-11.
112. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallographica Section D. 2010;66(4):486-501.
113. Cailleau R, Olive M, Cruciger QV. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. In vitro. 1978;14(11):911-5.
114. Sasaki CY, Passaniti A. Identification of anti-invasive but noncytotoxic chemotherapeutic agents using the tetrazolium dye MTT to quantitate viable cells in Matrigel. BioTechniques. 1998;24(6):1038-43.
115. Morgan CA, Hurley TD. Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1. Chem Biol Interact. 2015;234:29-37.
116. Croker AK, Goodale D, Chu J, Postenka C, Hedley BD, Hess DA, et al. High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. Journal of cellular and molecular medicine. 2009;13(8b):2236-52.
117. Perez-Miller SJ, Hurley TD. Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. Biochemistry. 2003;42(23):7100-9.
118. Khanna M, Chen CH, Kimble-Hill A, Parajuli B, Perez-Miller S, Baskaran S, et al. Discovery of a novel class of covalent inhibitor for aldehyde dehydrogenases. J Biol Chem. 2011;286(50):43486-94.
119. Toulany M, Iida M, Keinath S, Iyi FF, Mueck K, Fehrenbacher B, et al. Dual targeting of PI3K and MEK enhances the radiation response of K-RAS mutated non-small cell lung cancer2016.

**Curriculum Vitae**

**Mikhail Chtcherbinine**

## Education

Bachelor of Science, McGill University 2010-2014

Honours Biochemistry

Cumulative GPA of 3.93

Master of Science, Indiana University 2015-2016

Biochemistry and Molecular Biology

Cumulative GPA of 4.00

## Academic Awards

First Class Honours, 2014

Dean's Honour List, 2014

Emily Ross Crawford Scholarship, 2012

Faculty of Science Scholarship, 2012

Richard and Mary Shaw Scholarship, 2011

J. W. McConnel Scholarship, 2010

## Certification

CITI Biomedical Responsible Conduct of Research, 2016

Biological Safety Training, 2015

Bloodborne Pathogens Training, 2016

IUPUI Laboratory Safety Training, 2015

**Research Experience**

McGill Department of Biochemistry May 2012 - Aug. 2012

Dr. Gehring Lab

*Summer Undergraduate Student*

·Assisted with a structural biology project

·Gained experience working with crystallography, gel electrophoresis, protein expression, protein purification by chromatography, DNA purification and PCR methods

McGill Department of Biochemistry May 2013 - Aug. 2013

Dr. Pause Lab

*Undergraduate Honours Project Student*

· Completed Independent research project on cell signalling

· Utilized cell culture, Western blots, cell fractionation techniques and qPCR analysis in order to characterize a signalling pathway

McGill Department of Physiology Jan. 2014 - Jul. 2014

Dr. Lukacs Lab

*Summer Undergraduate Student, Undergraduate Research Course Student*

· Independent project to characterize interaction between ubiquitin chains and a ubiquitin-binding protein

· Utilized tissue culture, site-directed mutagenesis, molecular cloning and pull-down assays

IUSM Department of Biochemistry and Molecular Biology Oct. 2015 - present

Dr. Hurley Lab

*Master's Student*

· Master's Thesis project to characterize two series of isoenzyme-selective ALDH1 inhibitors

· Used in-vitro enzymatic assays, X-ray crystallography, and human cell culture assays

**Presentations**

**Chtcherbinine M** and Hurley TD. Characterization of Two Novel Series of ALDH1 Family Inhibitors. Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, March 2016