# Characterization of Two Distinct Structural Classes of Selective Aldehyde Dehydrogenase 1A1 Inhibitors

**Cynthia A. Morgan** and **Thomas D. Hurley**[*]

Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, IN 46202

## Abstract

Aldehyde dehydrogenases (ALDH) catalyze the irreversible oxidation of aldehydes to their corresponding carboxylic acid. Alterations in ALDH1A1 activity are associated with such diverse diseases as cancer, Parkinson's disease, obesity, and cataracts. Inhibitors of ALDH1A1 could aid in illuminating the role of this enzyme in disease processes. However, there are no commercially available selective inhibitors for ALDH1A1. Here we characterize two distinct chemical classes of inhibitors that are selective for human ALDH1A1 compared to eight other ALDH isoenzymes. The prototypical members of each structural class, CM026 and CM037, exhibit sub-micromolar inhibition constants, but have different mechanisms of inhibition. The crystal structures of these compounds bound to ALDH1A1 demonstrate that they bind within the aldehyde binding pocket of ALDH1A1 and exploit the presence of a unique Glycine residue to achieve their selectivity. These two novel and selective ALDH1A1 inhibitors may serve as chemical tools to better understand the contributions of ALDH1A1 to normal biology and to disease states.

## INTRODUCTION

Aldehydes are highly reactive compounds that can lead to cellular toxicity through their ability to form adducts with a variety of cellular nucleophiles found in proteins, nucleic acids, as well as small molecule metabolites. In humans, aldehyde detoxication occurs via three main enzyme systems: aldehyde oxidases, aldo-keto reductases, and aldehyde dehydrogenases. The human genome contains at least 19 functional genes for aldehyde dehydrogenases (ALDH) that catalyze the NAD(P)$^+$-dependent oxidation of endogenous and exogenous aldehydes to their corresponding carboxylic acids or CoA esters. ALDHs differ in their tissue distribution, subcellular location, structure, as well as preferred substrates and are critical enzymes that contribute to numerous biological functions as well as to the cellular defense against aldehyde toxicity[1]. They are involved in the synthesis of critical carboxylic acids including retinoic acid, a key regulator of cell growth and development[2], and the neurotransmitter, γ-aminobutyric acid[3]. A major role of the ALDH superfamily is protection from aldehyde-induced cytotoxicity[1]. Oxidative stress often results in lipid peroxidation, generating over 200 aldehydes, including 4-hydroxyhexenal, 4-hydroxynonenal, and malondialdehyde[4]. These endogenously generated compounds can carbonylate proteins and have been associated with neurodegenerative disorders[5] and

[*]*Corresponding Author*: Thomas D. Hurley, Ph.D., USA, 317-278-2008, thurley@iu.edu.

aging[6]. A variety of drugs, including ethanol and the anticancer drug cyclophosphamide, are metabolized via ALDH-dependent pathways[7]. In the environment, although some aldehydes have non-anthropogenic sources, motor vehicle exhaust, industrial applications, cigarette smoke and other human activities are the primary sources for exogenous aldehydes, including formaldehyde, acetaldehyde, and acrolein[8].

As a consequence of their critical contributions to aldehyde metabolism, loss of function mutations in ALDH genes are linked to a number of diseases. ALDH2 is the primary enzyme involved in the oxidation of acetaldehyde during ethanol metabolism[9] and a single nucleotide polymorphism (SNP) results in an enzymatically crippled protein (ALDH2*2), in which acetaldehyde, derived from ethanol oxidation, accumulates and induces alcohol toxicity[10,11]. Alterations in ALDH1A1 and ALDH2 expression or activity may play a role in Parkinson's Disease through the metabolism of the neurotransmitter dopamine, leading to increased levels of neurotoxic aldehydes, including 3,4-dihydroxyphenylacetaldehyde[12]. Modulation of ALDH2's role in dopamine metabolism has been shown to affect cocaine seeking behavior[13]. Mutations in ALDH3A2 lead to Sjögren-Larsson Syndrome, which is characterized by mental retardation, icthyosis, and spastic tetraplegia due to impaired metabolic clearance of sphingosine and plasmalogen metabolites[14,15]. Mutations in other ALDH genes have been linked to pyridoxine-dependent epilepsy (ALDH7A1)[16], type II hyperprolinemia resulting in mental retardation and seizures (ALDH4A1 and ALDH18A1)[17,18], and may possibly contribute to paranoid schizophrenia (ALDH3B1)[19]. A number of ALDHs have been associated with cancer and/or cancer stem cells, including ALDH1A1, ALDH1A2, ALDH1A3, ALDH1L1, ALDH2, ALDH3A1, ALDH4A1, and ALDH7A1[7,20]. Both ALDH1A1 and ALDH3A1 detoxify some oxazaphosphorine anticancer drugs and decrease the drug's effectiveness[21,22]. ALDH4A1 is p53- inducible and may minimize cellular damage due to oxidative stress[23]. ALDH1A2 is a possible tumor suppressor gene in prostate cancer, likely via the enzyme's role in retinoid metabolism[24]. The ALDH2*2 mutation has also been associated with a variety of cancers, possibly due to increased aldehyde-induced DNA damage[25,26]. Up-regulation of ALDH activity is also common in both normal and cancer stem cells[20]. Therefore, ALDH is considered a stem cell biomarker and the ALDEFLUOR assay (Stemcell Technologies, Vancouver, Canada) uses this ALDH activity as a means to identify cancer stem cells[27].

ALDH1A1 (retinaldehyde dehydrogenase 1, RALDH1) is a highly conserved, cytosolic homo-tetramer (~55 kDa monomers) that is widely expressed and found in numerous tissues, including brain, liver, kidneys, adipose, eye lens and retina. A key role of ALDH1A1 is the oxidation of retinaldehyde to retinoic acid (RA), forming transcriptional regulators critical for normal cell growth and differentiation[28]. Both the substrate (retinaldehyde) and product (RA) are important for normal biological processes, including vision, cellular differentiation, and immune function[7]. ALDH1A1 shares greater than 70% sequence identity to both ALDH1A2 and ALDH1A3 (RALDH2 and RALDH3, respectfully) and both also convert retinaldehyde to RA, but their roles may be more confined to embryogenesis and stem cell development[29,30]. ALDH1A1 knockout mice are viable, but ALDH1A2 and ALDH1A3 knockout mice are lethal early in development or shortly after birth[30]. In addition to its similarity with other retinaldehyde dehydrogenases, ALDH1A1 also shares nearly 70% sequence identity to the mitochondrial ALDH2 and ALDH1B1

enzymes, but shares less than 50% sequence identity with members of other ALDH families, including ALDH3A1, ALDH4A1 and ALDH5A1.

The structure of mammalian ALDHs are similar, functioning as homomultimers, and each monomer contains three structural domains: a catalytic domain, a cofactor binding domain, and an oligomerization domain[31]. The global structural similarities support identical active site residues whose contributions to aldehyde oxidation are well established. In addition to aldehyde oxidation activity, some ALDHs, including ALDH1A1, also possess an $NAD^+$-independent esterase activity[32] and this esterase activity uses the same active site residues as does the dehydrogenase activity[33]. Despite similarities in structure and function, as well as some overlap in substrate preferences, the isoenzymes of the ALDH family of proteins have evolved different aldehyde binding sites due to substitutions of the residues lining their respective substrate binding tunnels[31,34]. These differences have allowed for the development of activators and inhibitors of various isoenzymes as therapeutics. For example, disulfiram (commercial: Antabuse) is an inhibitor of both ALDH1A1 and ALDH2, and this drug can be used to treat alcoholism[31] and potentially for cocaine addiction[13]. Alda-1 is a selective activator of ALDH2 and may offer cardiac protection following ischemic events by decreasing cytotoxic aldehyde levels[35]. Since multiple cancers have increased ALDH1A1 and/or ALDH3A1 activity leading to increased resistance to chemotherapeutic agents, selective inhibitors of these two isoenzymes are also of interest[36,37].

Available inhibitors of ALDHs have been recently reviewed[31]. At this time, there are no commercially available inhibitors capable of distinguishing between ALDH1A1 and the other highly similar ALDH isoenzymes, ALDH1A2, ALDH1A3, ALDH1B1, and ALDH2. For instance, the control inhibitor used in the ALDEFLUOR assay, diethylaminobenzaldehyde (DEAB), is an effective inhibitor of at least three of these isoenzymes[38,39]. From a high-throughput screen (HTS)[40], we have identified two classes of ALDH1A1 inhibitors and have determined the structural and kinetic basis for their selectivity toward ALDH1A1.

## RESULTS

Since the esterase and dehydrogenase activity of ALDH1A1 utilize similar active site residues, modulators of esterase function as convenient surrogates for the dehydrogenase activity, as has been shown previously with the discovery of ALDH3A1 inhibitors[36]. The esterase assay was selected for the screening assay instead of aldehyde oxidation to: 1.) minimize spectral overlap with the absorbance characteristic of the compounds in the library, and 2.) eliminate compounds that compete with structurally conserved elements of the cofactor binding site. CM026 and CM037 emerged from the HTS as esterase modulators of ALDH1A1[40]. CM026 (Figure 1A) has a molecular weight of 442.5 Daltons and CM037 (Figure 2A) has a molecular weight of 431.6 Daltons. These compounds have no structural similarity to any known, commercially available aldehyde dehydrogenase modulators.

### Kinetic characterization of CM026 and CM037

CM026 is a selective inhibitor for ALDH1A1 versus eight other ALDH isoenzymes examined. At a concentration of 20 μM, CM026 had no effect on seven other human ALDH isoenzymes (ALDH1A2, ALDH1A3, ALDH1B1, ALDH2, ALDH3A1, ALDH4A1, ALDH5A1) as well as the carboxyl-terminal ALDH domain of rat ALDH1L1 (Figure 1B). At a concentration of 100 μM, CM026 modestly increased aldehyde oxidation catalyzed by ALDH1A2, ALDH1A3, and ALDH1B1. CM026 has good potency toward ALDH1A1 ($IC_{50}$ of $0.80 \pm 0.06$ μM[40], Figure 1C) for an initial hit compound. Complete inhibition of ALDH1A1 was not observed. CM026 has a noncompetitive partial mode of inhibition with respect to varied acetaldehyde, with a $K_i$ of $0.80 \pm 0.16$ μM and $\beta = 0.15 \pm 0.03$, indicating maximum inhibition at $0.15(V_{max})$ (Figure 1D). CM026 had an uncompetitive partial mode of inhibition with respect to varied $NAD^+$ and exhibited a $K_i$ of $0.72 \pm 0.03$ μM and $\beta = 0.10 \pm 0.03$ (Figure 1E).

It is possible that the size of the R-groups attached to the xanthine ring influences whether it is possible to slowly bind and release small aldehydes at the catalytic nucleophile, such substrate length dependency was also observed with the activator Alda-1 in ALDH2[41].

CM037 (Figure 2A) was also selective for ALDH1A1 at 20 μM versus eight other ALDH isoenzymes tested (Figure 2B). While 20 μM CM037 had little effect on most ALDH isoenzymes tested, ALDH1A3 was inhibited approximately 20% at this concentration. Higher concentrations were not tested due to solubility limits of CM037 under these assay conditions. CM037 exhibits an $IC_{50} = 4.6 \pm 0.8$ μM toward ALDH1A1 versus the substrate propionaldehyde (Figure 2C) and a competitive mode of inhibition with respect to varied substrate acetaldehyde and an average $K_i$ of $0.23 \pm 0.06$ μM from three independent inhibition experiments[42].

### Structure Activity Relationship for Analogs of CM026

There were 77 compounds in the initial screening hit list that were structurally similar to CM026[40]. We have tested 17 members of this compound class and they exhibit good selectivity for ALDH1A1 compared to ALDH1A2, ALDH1A3, ALDH1B1, ALDH2, and ALDH3A1 (Table 1). At high concentrations (100 μM), some compounds activated the aldehyde oxidation activity of other ALDH1 enzymes tested (eg. CM026), but in most cases these effects were observed as the solubility limit of the compounds in the assay system were approached, precluding a full dose-response study. At 20 μM, none of the compounds demonstrated activation exceeding 10% of control activity, suggesting any activation is of modest potency and not likely to impact these enzyme activities *in vivo*. These compounds share a common xanthine core structure with theophylline and caffeine, but neither of these stimulants affected ALDH1A1 activity at concentrations up to 250 μM, indicating that substituents at the R1 and R2 positions are necessary for ALDH1A1 inhibition (Table 1). Halogens on either R-group were not tolerated. CM053 was the most potent analog examined, with an $IC_{50} = 210 \pm 40$ nM and a $K_i = 96 \pm 14$ nM with noncompetitive tight mode of inhibition compared to varied substrate acetaldehyde (Figure 3). CM028 shares the same R2 group as CM053 but the isopentyl group at R1 has been replaced with a phenylpropyl group. CM028 is less potent, with an $IC_{50} = 2.0 \pm 0.1$ μM, and exhibits a $K_i =$

240 ± 40 nM with competitive tight mode of inhibition. The lower potency suggests that the phenylpropyl group might present steric conflicts not found with the smaller phenyl and isopentyl groups (Table 1). Unlike CM026, CM028 and CM053 demonstrated complete inhibition, which confirms that the nature of the R1 group alone does not influence the final extent of inhibition, since both CM026 and CM053 share the same R1 group.

## Crystal Structure of ALDH1A1 Complexed with CM026, CM053 and CM037

To determine the mechanism that underlies the ability of these compounds to selectively inhibit ALDH1A1, we used X-ray crystallography to determine the structure of the enzyme-compound complexes. We solved the crystal structures of human ALDH1A1 in complexes with CM026, CM053, and CM037 to resolutions between 1.80 Å and 1.95 Å (Table 2). For CM026 and CM037, the naturally occurring N121S polymorphic variant of ALDH1A1 was used[40], while wtALDH1A1 was used for the structure with bound CM053. A comparison of the respective alpha carbons in the structure of the N121S·CM026 to those in the WT·CM053 generated an RMSD of 0.12Å, indicating a high degree of similarity between WT and the N121S mutant as expected since the two have very similar kinetic behavior[40]. CM026 binds near the solvent exposed exit of the substrate-binding site (Figure 4). The xanthine rings for both CM026 and CM053 are parallel to and approximately 3.6 Å from Tyr297, with which it interacts via hydrophobic pi-stacking interactions. Four residues form hydrogen bonds with CM026; the xanthine ring interacts with His293, Cys302, and Gly458, while Trp178 interacts with the ketone group on R2. The isopentyl group of R1 projects towards Cys303 and fills much of the hydrophobic space bounded by Phe171 and Phe466. CM053 differs from CM026 only in its R2 group, which can form hydrogen bonds with two residues, Trp178 and Val460 (Figure 5). CM037 binds at a similar location to the CM026 compounds, but its long axis is oriented almost orthogonal to that of CM026 and CM053 (Figure 6A). Most of the tricyclic ring of CM037 is in a hydrophobic pocket formed by Phe171, Val460, and Phe466 with a potential hydrogen bond between the ring system's carbonyl oxygen atom and the side chain of Cys302 (Figure 6B). The biggest structural adaptation to CM037 binding is the movement of Trp178 away from the substrate-binding site to accommodate the benzyl ring of CM037 (Figure 6C). This conformational movement appears to be dynamic and impacts the observed electron density for both the benzyl group of CM037 and of Trp178. Trp178 is well ordered in all other structures determined of human ALDH1A, including our CM026 and CM053 structures, but has weak density for the benzyl moiety of the indole ring in this complex (Figure 6D). We would suggest that optimization of CM037 could be achieved by altering the thiophene and benzyl ring systems to alleviate these steric conflicts. To better understand the selectivity of these compounds for ALDH1A1, we compared these structures against human ALDH2 (PDB code 1CW3[43]), ALDH3A1 (PDB code 3SZA[37]), and ALDH4A1 (PDB code 3V9G[44]) and identified a critical glycine (Gly458) that is present near the xanthine ring binding site in ALDH1A1. This glycine is replaced by larger amino acid side chains in the other three human structures examined, as well as in sheep ALDH1A1 (PDB Code 1BXS[45]). In rat ALDH1A2 (PDB code 1BI9[46]), which shares 97% sequence identity to human ALDH1A2, this location is part of a small disordered loop not observed in the crystal structure[46]. Using sequence alignments of the human genes, Gly458 in ALDH1A1 is replaced by an asparagine in ALDH1A2, ALDH1A3, and ALDH1B1, an aspartate in ALDH2, and an isoleucine in

ALDH3A1 (Figure 7A). As shown in Figure 7B, these side chains would interfere with the position of the xanthine ring, effectively eliminating the ability of these analogs to bind to any isoenzyme but ALDH1A1.

### Characterization of ALDH1A1 G458N Mutant

To confirm whether Gly458 in ALDH1A1 directly impacts the selectivity of the CM026 analogs and CM037 for ALDH1A1, we mutated the glycine at this position to asparagine, as found in ALDH1A2 and ALDH1A3. We determined the kinetic parameters for acetaldehyde oxidation for both the wild-type and G458N enzymes (Table 3). This mutation did not dramatically affect the enzyme's catalytic efficiency for aldehyde oxidation. However, when Gly458 is mutated to asparagine, CM026 no longer inhibits the enzyme at concentrations up to 100 μM and none of the CM026 analogs inhibited the mutant more than 25% (Figure 8). Similarly, 20 μM CM037 no longer inhibited the G458N enzyme. However, the non-selective inhibitors DEAB and CM302[40] both inhibit G458N, with $IC_{50}$ values of $0.52 \pm 0.10$ μM and $3.1 \pm 0.3$ μM respectively[39,40] (Table 3). These data support the hypothesis that the substrate-binding site, and in particular Gly458, determines the selectivity of both the CM026 and CM037 classes of compounds for ALDH1A1, and that bulkier side chains at this position in ALDH1A2, ALDH1A3, ALDH1B1, ALDH2, and ALDH3A1 occludes their binding to these ALDH isoenzymes.

## DISCUSION AND CONCLUSIONS

Aldehyde dehydrogenases contribute to a variety of biological processes and disease states. ALDH1A1, in particular, has been linked to such diverse diseases as cancer, Parkinson's disease, obesity, and cataracts. Therefore, selective inhibitors of ALDH1A1 would be of tremendous value to understanding the roles of this enzyme in both normal and disease processes. However, great structural and functional similarity exists, especially within the ALDH1 family, with five members sharing approximately 70% protein sequence identity or higher, plus significant overlap in substrate utilization. To date, there are no ALDH1A1-selective inhibitors commercially available. Although comparisons of available ALDH structures indicate a high degree of overlap, there exist distinct surface topographies that may enable development of selective inhibitors (Figure 9). ALDH1A1 possesses a wider opening leading to the active site, whereas ALDH2 has a much more constricted, cylindrical shaped site. ALDH3A1 possesses a wider inner vestibule near the catalytic nucleophile, with a much narrower and curved entryway. The narrower entries in ALDH2 and ALDH3A1 eliminate the binding site for the CM026 and CM037 classes of inhibitors in large part due to the side chain present at the position equivalent to Gly458. ALDH4A1 possess a serine at the position equivalent to Gly458. However the loop structure in which it resides is different enough from ALDH1A1 to prevent compound binding. Similar to daidzin, both the CM026 and CM037 classes of compounds are planar, multi-ringed structures that adopt binding modes that take advantage of the topological characteristics unique to the ALDH isoenzyme toward which they demonstrate selectivity and neither of these new compounds can be accommodated in the restricted substrate binding sites of ALDH3A1 or ALDH2. Daidzin is a strong inhibitor of ALDH2 but it also inhibits ALDH1A1[47]. Daidzin binds in a similar location to CM026, but comparison of the structure of ALDH1A1 and ALDH2 (PDB

2VLE[48]) bound to these respective compounds shows that the two compounds bind nearly perpendicular to each other in their respective binding modes. In particular, aligning the ALDH2-daidzin structure to ALDH1A1 demonstrates that daidzin binding need not engage ALDH1A1 near its unique G458 site (Figure 10). As discussed by Lowe *et al*, the tighter binding pocket of ALDH2 compared to ALDH1A1 favors more intimate interactions between the ALDH2 and daidzin, increasing specificity[48]. CM026 exploits a binding pocket that is not accessible in other ALDH isoenzymes due to their larger amino acid side chains at position 458.

In contrast to CM026, the directionality of CM037 binding within the ALDH1A1 active site resembles that of daidzin in ALDH2. However, the 4-oxo group and the branched structure of CM037 near the exit of the substrate binding site exploits the same G458 region for selectivity. Our structural studies identified Gly458 as a major contributor to the selectivity of the CM026 and CM037. Sequence alignments of the other 18 human ALDH isoenzymes indicate that the only other family member with a glycine in this position is ALDH16A1. However, the function of ALDH16A1 is unknown and has a three residue deletion in the active site loop which eliminates the conserved Cys nucleophile, suggesting this protein may have functions that are independent of aldehyde dehydrogenase activity[49]. The presence of a non-glycine residue at this position does not adversely affect catalytic activity toward small substrates or inhibition by the non-selective compounds, DEAB and CM302. Consequently, these compounds identified via an esterase-based high throughput screen successfully exploited a unique structural feature found primarily in primate ALDH1A1 enzymes, which further validates the use of the esterase activity as a screening tool for ALDH isoenzymes[40].

Selective inhibitors of ALDH1A1 are needed to understand the role of this enzyme in both normal and disease processes. As recently reviewed by Ma and Allan, a number of ALDH family members have been associated with both normal stem cells and cancer stem cells[20]. The viability of the ALDH1A1-/- mice suggests that ALDH1A1 is non-essential or can be compensated for by other family members during growth and development[50]. On the other hand, ALDH1A1 is considered a biomarker for lung[51], ovarian[52], prostate[53] and a number of other cancers. Ovarian cancer cells form spheroids, cellular aggregates that aid metastasis[54]. Recently ALDH1A1 was shown to be upregulated in ovarian cancer spheroids[42]. This study utilized our ALDH1A1- selective compound CM037 (published as A37) to disrupt spheroid formation and reduce cell viability, supporting the hypothesis that ALDH1A1 could serve as a target to improve cancer outcomes[42]. ALDH1A1 and ALDH3A1 are both involved in the metabolism of the cancer drug cyclophosphamide, metabolizing the active compound to a less active form and contributing to drug resistance. ALDH3A1 inhibitors have been shown to increase sensitivity to the cyclophosphamide analog mafosphamide when combined in cell lines with high ALDH3A1 expression[55,56]. ALDH1A1 could serve as a similar target to minimize cyclophosphamide resistance in cancers with high ALDH1A1 levels. The functional role that ALDH1A1 contributes to stem cells and cancer metastasis is not understood. However, the discovery of these two novel classes of selective ALDH1A1 inhibitors may serve as chemical tools to uncover those roles in both normal and diseased cells.

# EXPERIMENTAL SECTION

## Materials

Reagents, including acetaldehyde, propionaldehyde, para-nitrophenylacetate, $NAD^+$, and buffers were all purchased from Sigma Aldrich unless otherwise noted. Compounds were purchased from ChemDiv Corp. (San Diego, CA) and were >95% pure based on the NMR spectra provided by the vendor. Following receipt from the vendor, their chemical identities were further verified by LC/MS in the Department of Chemistry, Indiana University-Purdue University Indianapolis and used without further purification.

## ALDH expression and purification

ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, ALDH2, and ALDH3A1 were produced and purified as previously described[36,56,57]. The ALDH1A1 protein used for the ALDH1A1-CM026 crystal contained an Asn-to-Ser SNP at residue 121[40,58]. Unless where noted otherwise, ALDH1A1 WT protein was used for all aldehyde oxidation assays and for the ALDH1A1- CM053 structure. The full length cDNA for human ALDH4A1 and ALDH5A1 were generously provided by Daria Mochly-Rosen. ALDH4A1 was subcloned into the pET-28a expression plasmid and ALDH5A1 was in pTrcHis-Topo. The carboxyl terminal ALDH domain of rat ALDH1L1 was generously provided by Sergey Krupenko in the pRSET expression plasmid. ALDH1L1, ALDH4A1, and ALDH5A1 were expressed and purified as previously described for ALDH3A1[36] with the following modifications: 1) for ALDH1L1 and ALDH5A1 the medium contained 100 μg/mL ampicillin, 2) cells were lysed via 3 passages through a microfluidizer (DivTech Equipment), and 3) a single passage on a nickel-NTA column was used for purification, without the second Q-sepharose column used to purify ALDH3A1. Purified enzymes used for kinetics were stored at −80°C. ALDH1A1 used for crystallization was stored at −20°C in a 50% (v/v) solution with glycerol and dialyzed against 10 mM $Na^+$-ACES pH 6.6 and 1 mM dithiothreitol at 4°C.

## Generation of wild-type ALDH1A1 and the G458N mutant

The QuikChange site-directed mutagenesis protocol was used to make point mutations. First, wtALDH1A1 was generated from the Weiner N121S polymorphism as previously described[40]. A point mutation of wtALDH1A1 was performed to produce the G458N mutant. The ALDH1A1 G458N mutant was constructed using the forward primer 5'-GTG GGT GAA TTG CTA TAA CGT GGT AAG TGC CCAG-3' and its complement. This G458N mutant was purified in the same way as other ALDH1A1 proteins (WT, N121S). G458N was stored at 2-mg/mL and 8-mg/mL at −80°C. Kinetic experiments for G458N were performed in the same manner as WT protein.

## Aldehyde dehydrogenase activity assay

Dehydrogenase activity of purified recombinant ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1, ALDH2, and ALDH3A1 was assayed spectrophotometrically by monitoring the formation of NADH at 340 nm (molar extinction coefficient of 6220 $M^{-1}$ $cm^{-1}$) on a Beckman DU-640 or Cary 300 Bio UV-Vis spectrophotometer. Characterization of ALDH1A1 WT and ALDH1A1 G458N were performed by co-varying acetaldehyde and

$NAD^+$ concentrations in reactions containing 150 – 300 nM enzyme, 50 – 500 µM or 20 – 200 µM acetaldehyde, and 50 – 500 µM $NAD^+$ in 50 mM sodium BES, pH 7.5 at 25°C. Reactions were initiated by adding enzyme. Selectivity of compounds for dehydrogenase activity of ALDH1A1, ALDH1A2, ALDH1A3, ALDH1B1 and ALDH2 were measured in a solution containing 100 – 200 nM enzyme, 200 µM $NAD^+$, 1% DMSO, and 100 µM propionaldehyde (non-saturating for all enzymes tested, except ALDH2) in 50 mM sodium BES. ALDH3A1 activity was measured under the following conditions: 25 nM enzyme, 200 µM $NAD^+$, 1% DMSO, and 300 µM benzaldehyde (~$K_M$) in either 100 mM sodium phosphate buffer, pH 7.5 or 50 mM sodium BES, pH 7.5. All assays were performed at 25°C and were initiated by the addition of the aldehyde substrate following a 2 minute pre-incubation with compound and cofactor. Selectivity was initially tested using 20 µM compound, with ALDH1A1-selective modulators further tested using 100 µM compound. $IC_{50}$ values for propionaldehyde oxidation were calculated by varying the concentration of the compounds from 0 to 200 µM. After a 2 minute pre-incubation with compound and $NAD^+$, all reactions were initiated by the addition of propionaldehyde. Data were fit to the four parameter $EC_{50}$ equation using SigmaPlot (StatSys v12.3). The values represent the average of three independent experiments (each n = 3). The mode of inhibition was determined via steady-state kinetics by co-varying inhibitor and substrate concentrations at fixed concentration of the second substrate. All reactions contained 100–150 nM ALDH1A1 and 1% DMSO in 50 mM sodium BES, pH 7.5 at 25°C. When cofactor $NAD^+$ was varied, the reactions contained 200 µM propionaldehyde and 20 – 200 µM or 25 – 250 µM $NAD^+$ ($K_m$ = 50 µM). When acetaldehyde was varied, the reactions contained 800 or 1000 µM $NAD^+$ and 100 – 800 µM acetaldehyde ($K_m$ = 180 µM). All data were fit to competitive, noncompetitive, uncompetitive, and mixed inhibition models using both single substrate-single inhibitor or tight binding inhibition programs in SigmaPlot (StatSys v12.3). The appropriate model was selected through analysis of goodness-of-fit and the residuals of those fits. The values represent the average of three independent experiments (each n = 3) using at least two protein preps. In Sigmaplot, the data were best fit to the following equations:

$$\text{Noncompetitive partial:} \nu = \frac{V_{max}}{\left(1+\frac{K_m}{S}\right)\left(\frac{1+\frac{1}{K_i}}{1+\frac{I\beta}{K_i}}\right)}$$

$$\text{Uncompetitive partial:} \nu = V_{max}\left(\frac{1+\beta\frac{1}{K_i}}{1+\frac{1}{K_i}+\frac{K_m}{S}}\right)$$

$$\text{Noncompetitive tight:} \nu = \frac{V_0}{2E}\left(EIK_i+\sqrt{(EIK_i)^2+4EK_i}\right)$$

$$\text{Competitive tight:} \nu = \frac{V_0}{2E}\left(EIK_{ap} + \sqrt{(EIK_{ap})^2 + 4EK_{ap}}\right)$$

## Crystallization and structure determination of ALDH1A1 and its complexes with inhibitors

Crystals of ALDH1A1 at 3–5 mg/mL concentration were equilibrated against a crystallization solution of 100 mM $Na^+$ BisTris, pH 6.2–6.8, 8–12% PEG3350 (Hampton Research), 200 mM NaCl, and 5–10 mM $YbCl_3$ at 25°C. For the enzyme·CM026 and enzyme·CM037 complex, ALDH1A1·N121S was used, while wtALDH1A1 was used for the ALDH1A1·CM053 complex. CM026 and CM053 complexes were prepared by soaking crystals overnight with crystallization solution containing 500 μM inhibitor with 2% (v/v) DMSO. The CM037 complex was prepared by soaking crystals with crystallization solution containing 500 μM inhibitor, 1% (v/v) DMSO, and 1 mM $NAD^+$ for 5 hours. Cryo-protection of the crystals for flash-freezing utilized 20% (v/v) ethylene glycol in the ligand soaking solution. Diffraction data were collected at Beamline 19-ID for the ALDH1A1-CM026 complex and Beamline 19-BM for the ALDH1A1-CM053 complex, both operated by the Structural Biology Consortium at the Advanced Photon Source (APS), Argonne National Laboratory. For the ALDH1A1-CM037 complex, data were collected at Beamline 23-ID-D (GM/CA), sponsored by National Institute of General Medical Sciences and National Cancer Institute of the National Institutes of Health at APS. All diffraction data were indexed, integrated, and scaled using either the HKL2000 or HKL3000 program suites[59]. The CCP4 program suite[60] was used for molecular replacement and refinement, using human apoALDH1A1 structure (PDB code 4WJ9) as a model for both structures. The molecular graphics application Coot[61] was used for model building and TLSMD (Translation/Libration/Screw Motion Determination) server was used to determine the appropriate TLS tensors for refinement of the protein[62,63].

## ACKNOWLEDGEMENTS

## Abbreviations Used

| | |
|---|---|
| **ACES** | N-(2-Acetamido)-2-aminoethanesulfonic acid |
| **ALDH** | aldehyde dehydrogenase |
| **BES** | N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic acid |

| **BisTris** | 2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol |
|---|---|
| **RA** | retinoic acid |
| **RALDH** | retinaldehyde dehydrogenase |

## References

1. Vasiliou V, Pappa A, Estey T. Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. Drug Metab Rev. 2004; 36:279–299. [PubMed: 15237855]

2. Yoshida A, Hsu LC, Dave V. Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. Enzyme. 1992; 46:239–244. [PubMed: 1292933]

3. Kurys G, Ambroziak W, Pietruszko R. Human aldehyde dehydrogenase. Purification and characterization of a third isozyme with low Km for gamma-aminobutyraldehyde. J Biol Chem. 1989; 264:4715–4721. [PubMed: 2925663]

4. Esterbauer H, Schaur RJ, Zollner H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. Free Radic Biol Med. 1991; 11:81–128. [PubMed: 1937131]

5. Sayre LM, Smith MA, Perry G. Chemistry and biochemistry of oxidative stress in neurodegenerative disease. Curr Med Chem. 2001; 8:721–738. [PubMed: 11375746]

6. Chevion M, Berenshtein E, Stadtman ER. Human studies related to protein oxidation: protein carbonyl content as a marker of damage. Free Radic Res. 2000; 33(Suppl):S99–S108. [PubMed: 11191280]

7. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert Opin Drug Metab Toxicol. 2008; 4:697–720. [PubMed: 18611112]

8. O'Brien PJ, Siraki AG, Shangari N. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. Crit Rev Toxicol. 2005; 35:609–662. [PubMed: 16417045]

9. Klyosov AA, Rashkovetsky LG, Tahir MK, Keung WM. Possible role of liver cytosolic and mitochondrial aldehyde dehydrogenases in acetaldehyde metabolism. Biochemistry. 1996; 35:4445–4456. [PubMed: 8605194]

10. Yoshida A, Huang IY, Ikawa M. Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. Proc Natl Acad Sci U S A. 1984; 81:258–261. [PubMed: 6582480]

11. Peng GS, Chen YC, Tsao TP, Wang MF, Yin SJ. Pharmacokinetic and pharmacodynamic basis for partial protection against alcoholism in Asians, heterozygous for the variant ALDH2*2 gene allele. Pharmacogenet Genomics. 2007; 17:845–855. [PubMed: 17885622]

12. Marchitti SA, Deitrich RA, Vasiliou V. Neurotoxicity and metabolism of the catecholamine-derived 3,4-dihydroxyphenylacetaldehyde and 3,4-dihydroxyphenylglycolaldehyde: the role of aldehyde dehydrogenase. Pharmacol Rev. 2007; 59:125–150. [PubMed: 17379813]

13. Yao L, Fan P, Arolfo M, Jiang Z, Olive MF, Zablocki J, Sun HL, Chu N, Lee J, Kim HY, Leung K, Shryock J, Blackburn B, Diamond I. Inhibition of aldehyde dehydrogenase-2 suppresses cocaine seeking by generating THP, a cocaine use-dependent inhibitor of dopamine synthesis. Nat Med. 2010; 16:1024–1028. [PubMed: 20729865]

14. Rizzo WB, Dammann AL, Craft DA. Sjogren-Larsson syndrome. Impaired fatty alcohol oxidation in cultured fibroblasts due to deficient fatty alcohol:nicotinamide adenine dinucleotide oxidoreductase activity. J Clin Invest. 1988; 81:738–744. [PubMed: 3343337]

15. Rizzo WB, Carney G. Sjogren-Larsson syndrome: diversity of mutations and polymorphisms in the fatty aldehyde dehydrogenase gene (ALDH3A2). Hum Mutat. 2005; 26:1–10. [PubMed: 15931689]

16. Mills PB, Struys E, Jakobs C, Plecko B, Baxter P, Baumgartner M, Willemsen MA, Omran H, Tacke U, Uhlenberg B, Weschke B, Clayton PT. Mutations in antiquitin in individuals with pyridoxine-dependent seizures. Nat Med. 2006; 12:307–309. [PubMed: 16491085]

17. Geraghty MT, Vaughn D, Nicholson AJ, Lin WW, Jimenez-Sanchez G, Obie C, Flynn MP, Valle D, Hu CA. Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. Hum Mol Genet. 1998; 7:1411–1415. [PubMed: 9700195]

18. Farrant RD, Walker V, Mills GA, Mellor JM, Langley GJ. Pyridoxal phosphate de-activation by pyrroline-5-carboxylic acid. Increased risk of vitamin B6 deficiency and seizures in hyperprolinemia type II. J Biol Chem. 2001; 276:15107–15116. [PubMed: 11134058]

19. Marchitti SA, Orlicky DJ, Vasiliou V. Expression and initial characterization of human ALDH3B1. Biochem Biophys Res Commun. 2007; 356:792–798. [PubMed: 17382292]

20. Ma I, Allan AL. The role of human aldehyde dehydrogenase in normal and cancer stem cells. Stem Cell Rev. 2011; 7:292–306. [PubMed: 21103958]

21. Sladek NE. Aldehyde dehydrogenase-mediated cellular relative insensitivity to the oxazaphosphorines. Curr Pharm Des. 1999; 5:607–625. [PubMed: 10469894]

22. Agarwal DP, von Eitzen U, Meier-Tackmann D, Goedde HW. Metabolism of cyclophosphamide by aldehyde dehydrogenases. Adv Exp Med Biol. 1995; 372:115–122. [PubMed: 7484368]

23. Yoon KA, Nakamura Y, Arakawa H. Identification of ALDH4 as a p53-inducible gene and its protective role in cellular stresses. J Hum Genet. 2004; 49:134–140. [PubMed: 14986171]

24. Kim H, Lapointe J, Kaygusuz G, Ong DE, Li C, van de Rijn M, Brooks JD, Pollack JR. The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. Cancer Res. 2005; 65:8118–8124. [PubMed: 16166285]

25. Matsuda T, Yabushita H, Kanaly RA, Shibutani S, Yokoyama A. Increased DNA damage in ALDH2-deficient alcoholics. Chem Res Toxicol. 2006; 19:1374–1378. [PubMed: 17040107]

26. Yokoyama A, Muramatsu T, Omori T, Yokoyama T, Matsushita S, Higuchi S, Maruyama K, Ishii H. Alcohol and aldehyde dehydrogenase gene polymorphisms and oropharyngolaryngeal, esophageal and stomach cancers in Japanese alcoholics. Carcinogenesis. 2001; 22:433–439. [PubMed: 11238183]

27. Jones RJ, Barber JP, Vala MS, Collector MI, Kaufmann SH, Ludeman SM, Colvin OM, Hilton J. Assessment of aldehyde dehydrogenase in viable cells. Blood. 1995; 85:2742–2746. [PubMed: 7742535]

28. Chambon P. A decade of molecular biology of retinoic acid receptors. FASEB J. 1996; 10:940–954. [PubMed: 8801176]

29. Kumar S, Sandell LL, Trainor PA, Koentgen F, Duester G. Alcohol and aldehyde dehydrogenases: Retinoid metabolic effects in mouse knockout models. Biochim Biophys Acta. 2012; 1821:198–205. [PubMed: 21515404]

30. Niederreither K, Fraulob V, Garnier JM, Chambon P, Dolle P. Differential expression of retinoic acid-synthesizing (RALDH) enzymes during fetal development and organ differentiation in the mouse. Mech Dev. 2002; 110:165–171. [PubMed: 11744377]

31. Koppaka V, Thompson DC, Chen Y, Ellermann M, Nicolaou KC, Juvonen RO, Petersen D, Deitrich RA, Hurley TD, Vasiliou V. Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. Pharmacol Rev. 2012; 64:520–539. [PubMed: 22544865]

32. Feldman RI, Weiner H. Horse liver aldehyde dehydrogenase. II. Kinetics and mechanistic implications of the dehydrogenase and esterase activity. J Biol Chem. 1972; 247:267–272. [PubMed: 4336042]

33. Sidhu RS, Blair AH. Human liver aldehyde dehydrogenase. Esterase activity. J Biol Chem. 1975; 250:7894–7898. [PubMed: 170275]

34. Wang MF, Han CL, Yin SJ. Substrate specificity of human and yeast aldehyde dehydrogenases. Chem Biol Interact. 2009; 178:36–39. [PubMed: 18983993]

35. Chen CH, Budas GR, Churchill EN, Disatnik MH, Hurley TD, Mochly-Rosen D. Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. Science. 2008; 321:1493–1495. [PubMed: 18787169]

36. Parajuli B, Kimble-Hill AC, Khanna M, Ivanova Y, Meroueh S, Hurley TD. Discovery of novel regulators of aldehyde dehydrogenase isoenzymes. Chem Biol Interact. 2011; 191:153–158. [PubMed: 21349255]

37. Khanna M, Chen CH, Kimble-Hill A, Parajuli B, Perez-Miller S, Baskaran S, Kim J, Dria K, Vasiliou V, Mochly-Rosen D, Hurley TD. Discovery of a novel class of covalent inhibitor for aldehyde dehydrogenases. J Biol Chem. 2011; 286:43486–43494. [PubMed: 22021038]

38. Moreb JS, Ucar D, Han S, Amory JK, Goldstein AS, Ostmark B, Chang LJ. The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. Chem Biol Interact. 2012; 195:52–60. [PubMed: 22079344]

39. Morgan CA, Parajuli B, Buchman C, Dria K, Hurley TD. N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. Chemico-Biological Interactions. 2014 in press.

40. Morgan CA, Hurley TD. Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1. Chem Biol Interact. 2014 In press.

41. Perez-Miller S, Younus H, Vanam R, Chen CH, Mochly-Rosen D, Hurley TD. Alda-1 is an agonist and chemical chaperone for the common human aldehyde dehydrogenase 2 variant. Nat Struct Mol Biol. 2010; 17:159–164. [PubMed: 20062057]

42. Condello S, Morgan CA, Nagdas S, Cao L, Turek J, Hurley TD, Matei D. beta-Cateninregulated ALDH1A1 is a target in ovarian cancer spheroids. Oncogene. 2014 In press.

43. Ni L, Zhou J, Hurley TD, Weiner H. Human liver mitochondrial aldehyde dehydrogenase: three-dimensional structure and the restoration of solubility and activity of chimeric forms. Protein Sci. 1999; 8:2784–2790. [PubMed: 10631996]

44. Srivastava S, Watowich SJ, Petrash JM, Srivastava SK, Bhatnagar A. Structural and kinetic determinants of aldehyde reduction by aldose reductase. Biochemistry. 1999; 38:42–54. [PubMed: 9890881]

45. Moore SA, Baker HM, Blythe TJ, Kitson KE, Kitson TM, Baker EN. Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Structure. 1998; 6:1541–1551. [PubMed: 9862807]

46. Lamb AL, Newcomer ME. The structure of retinal dehydrogenase type II at 2.7 A resolution: implications for retinal specificity. Biochemistry. 1999; 38:6003–6011. [PubMed: 10320326]

47. Keung WM, Vallee BL. Daidzin: a potent, selective inhibitor of human mitochondrial aldehyde dehydrogenase. Proc Natl Acad Sci U S A. 1993; 90:1247–1251. [PubMed: 8433985]

48. Lowe ED, Gao GY, Johnson LN, Keung WM. Structure of daidzin, a naturally occurring antialcohol- addiction agent, in complex with human mitochondrial aldehyde dehydrogenase. J Med Chem. 2008; 51:4482–4487. [PubMed: 18613661]

49. Vasiliou V, Sandoval M, Backos DS, Jackson BC, Chen Y, Reigan P, Lanaspa MA, Johnson RJ, Koppaka V, Thompson DC. ALDH16A1 is a novel non-catalytic enzyme that may be involved in the etiology of gout via protein-protein interactions with HPRT1. Chem Biol Interact. 2013; 202:22–31. [PubMed: 23348497]

50. Levi BP, Yilmaz OH, Duester G, Morrison SJ. Aldehyde dehydrogenase 1a1 is dispensable for stem cell function in the mouse hematopoietic and nervous systems. Blood. 2009; 113:1670–1680. [PubMed: 18971422]

51. Ucar D, Cogle CR, Zucali JR, Ostmark B, Scott EW, Zori R, Gray BA, Moreb JS. Aldehyde dehydrogenase activity as a functional marker for lung cancer. Chem Biol Interact. 2009; 178:48–55. [PubMed: 18952074]

52. Landen CN Jr, Goodman B, Katre AA, Steg AD, Nick AM, Stone RL, Miller LD, Mejia PV, Jennings NB, Gershenson DM, Bast RC Jr, Coleman RL, Lopez-Berestein G, Sood AK. Targeting aldehyde dehydrogenase cancer stem cells in ovarian cancer. Mol Cancer Ther. 2010; 9:3186–3199. [PubMed: 20889728]

53. Li T, Su Y, Mei Y, Leng Q, Leng B, Liu Z, Stass SA, Jiang F. ALDH1A1 is a marker for malignant prostate stem cells and predictor of prostate cancer patients' outcome. Lab Invest. 2010; 90:234–244. [PubMed: 20010854]

54. Sodek KL, Ringuette MJ, Brown TJ. Compact spheroid formation by ovarian cancer cells is associated with contractile behavior and an invasive phenotype. Int J Cancer. 2009; 124:2060–2070. [PubMed: 19132753]

55. Parajuli B, Fishel ML, Hurley TD. Selective ALDH3A1 inhibition by benzimidazole analogues increase mafosfamide sensitivity in cancer cells. J Med Chem. 2014; 57:449–461. [PubMed: 24387105]

56. 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. 2014; 15:701–712. [PubMed: 24677340]

57. Hammen PK, Allali-Hassani A, Hallenga K, Hurley TD, Weiner H. Multiple conformations of NAD and NADH when bound to human cytosolic and mitochondrial aldehyde dehydrogenase. Biochemistry. 2002; 41:7156–7168. [PubMed: 12033950]

58. Zheng CF, Wang TT, Weiner H. Cloning and expression of the full-length cDNAS encoding human liver class 1 and class 2 aldehyde dehydrogenase. Alcohol Clin Exp Res. 1993; 17:828–831. [PubMed: 8214422]

59. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt A. 1997; 276:307–326.

60. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:760–763. [PubMed: 15299374]

61. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]

62. Painter J, Merritt EA. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. Acta Crystallogr D Biol Crystallogr. 2005; 61:465–471. [PubMed: 15809496]

63. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr. 2006; 62:439–450. [PubMed: 16552146]

A.

8-{[4-(2-Furoyl)-1-piperazinyl]methyl}-1,3-dimethyl-7-(3-methylbutyl)-3,7-dihydro-1H-purine-2,6-dione

B.

C.

D.

E.
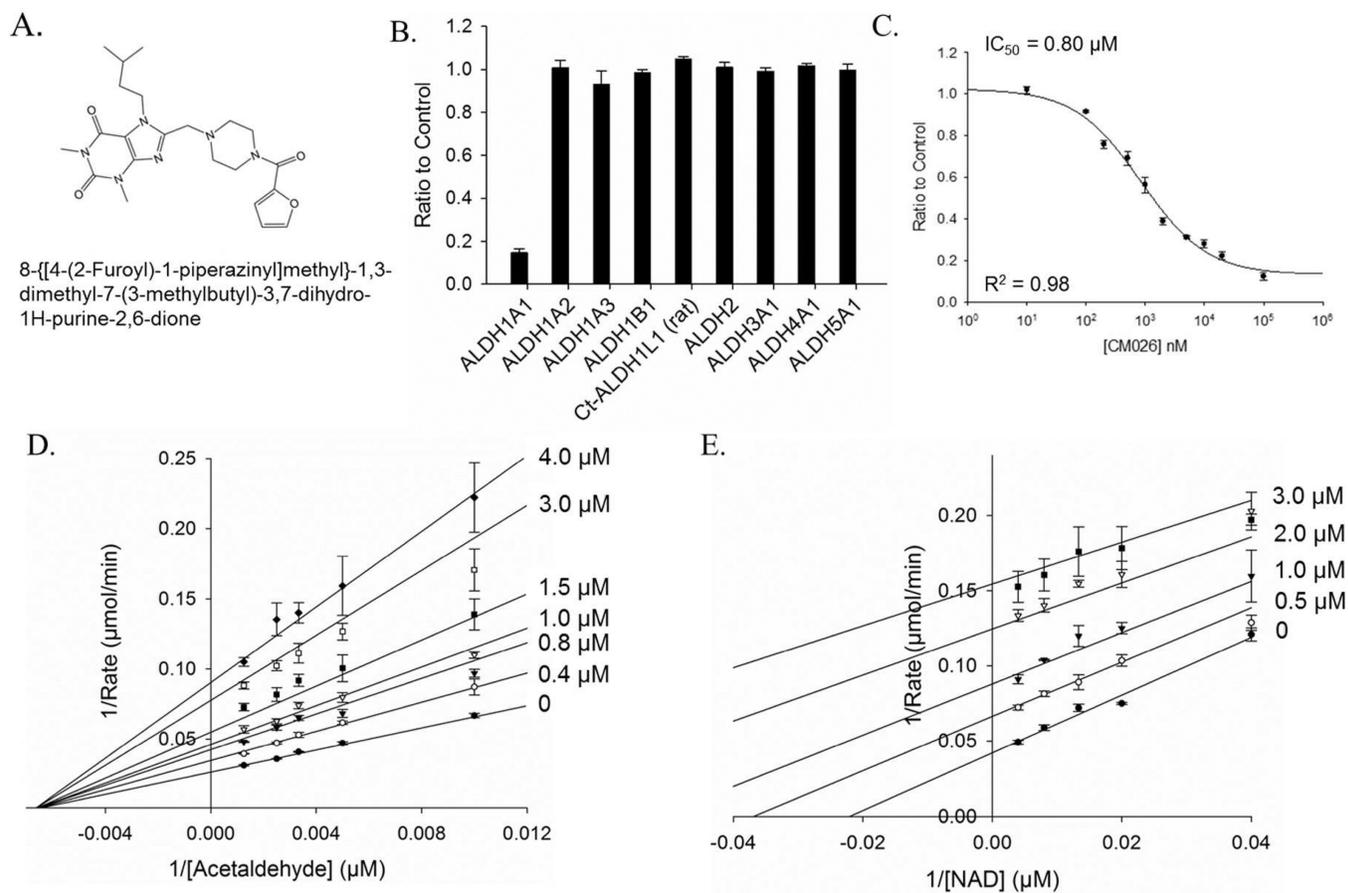
**Figure 1.**
Characterization of CM026, a selective inhibitor of ALDH1A1. (A) The structure of CM026. (B) Selectivity of 20 μM of CM026 with respect to nine ALDH isoenzymes. (C) $IC_{50}$ of CM026 with ALDH1A1[39]. (D) Lineweaver-Burk representation of noncompetitive inhibition for CM026 (0 – 4 μM) verses varied acetaldehyde (100 – 800 μM) at fixed concentration of $NAD^+$ (800 μM). (E.) Lineweaver-Burk representation of uncompetitive inhibition for CM026 (0 – 3 μM) verses varied $NAD^+$ (25 – 250 μM) at fixed concentration of propionaldehyde (200 μM). The $IC_{50}$ curves and Lineweaver-Burk plots represent one of three experiments performed for each condition, with each point the mean/SEM of three data points at each concentration.

A.

Ethyl ({4-oxo-3-[3-(1-pyrrolidinyl)
propyl]-3,4-dihydro[1]benzothieno[3,2-
d]pyrimidin-2-yl}sulfanyl)acetate

B.

C.

$IC_{50}$ = 4.6 μM

$R^2$ = 0.99

[CM037] nM
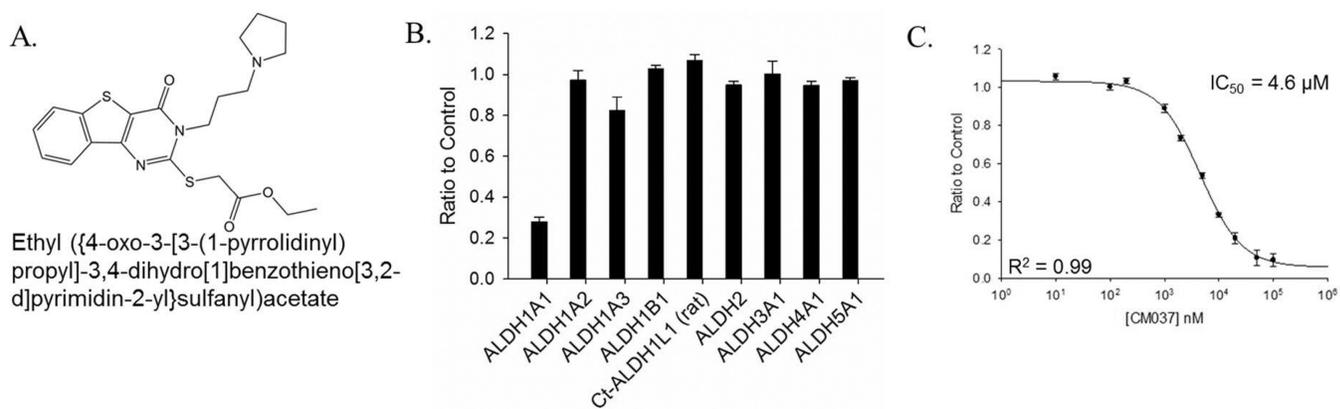
**Figure 2.**
Characterization of CM037, a selective inhibitor of ALDH1A1. (A) The structure of CM037
with a molecular weight of 431.6 Daltons. (B) Selectivity of 20 μM of CM037 with respect
to nine ALDH isoenzymes. (C) $IC_{50}$ of CM037 with ALDH1A1[42]. The $IC_{50}$ curve
represents one of three experiments performed, with each point the mean/SEM of three data
points at each concentration.
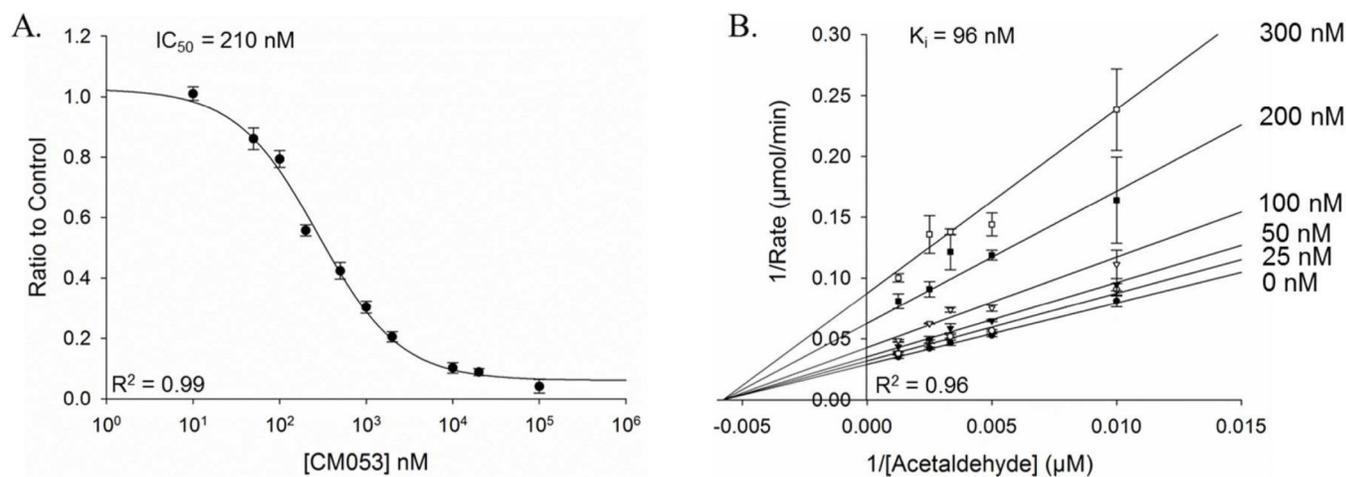
**Figure 3.**
Characterization of CM053. (A) $IC_{50}$ of CM053 with ALDH1A1. (B) Lineweaver-Burk representation of noncompetitive tight inhibition for CM053 (0 – 300 nM) verses varied acetaldehyde (100 – 800 µM) at fixed concentration of $NAD^+$ (1000 µM). Both curves represent one of three experiments performed, with each point the mean/SEM of three data points at each concentration.
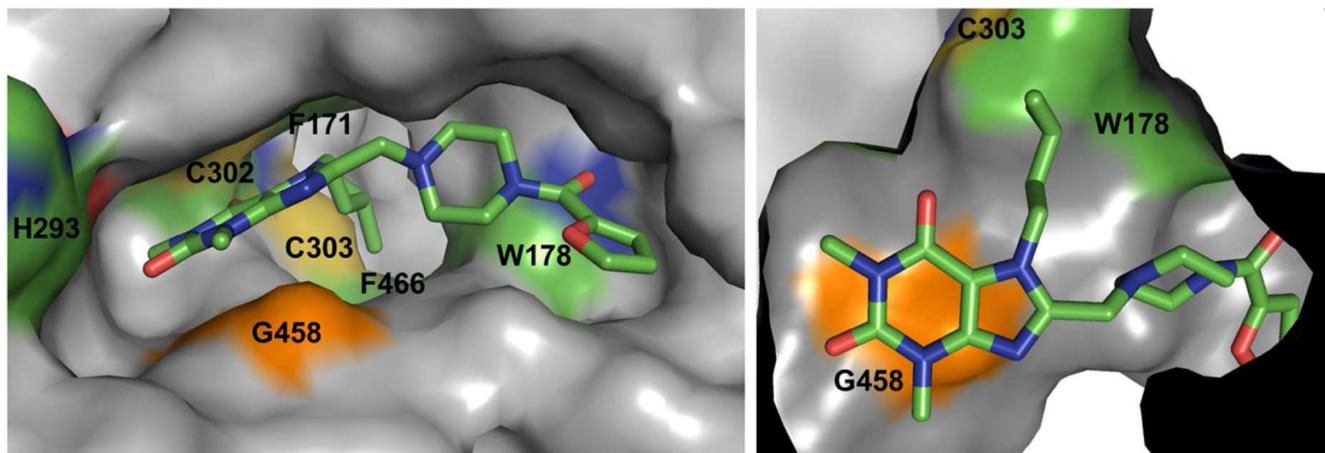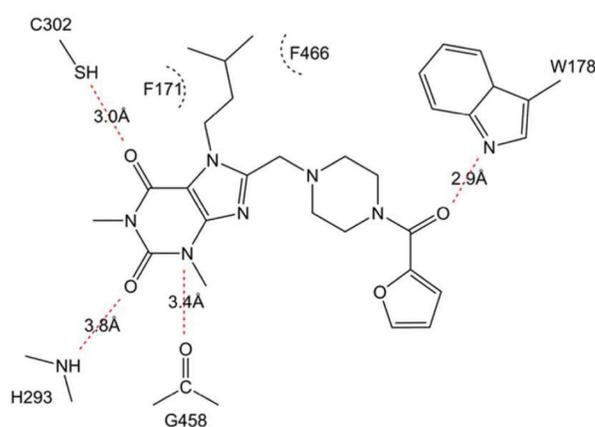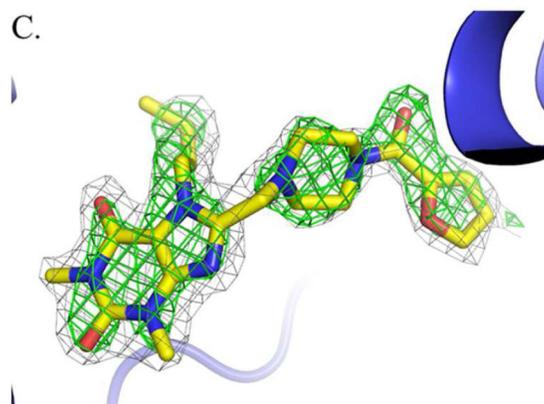
**Figure 4.**
Structure of ALDH1A1 N121S with CM026 (PDB Code 4WP7). (A) CM026 binds in the active site near Cys 303. Key residues are colored based on element, and Gly458 is shown in orange. (B) Two-dimensional representation of the key hydrogen bonds, illustrated with red dashed lines, and hydrophobic interactions, illustrated with black arcs, between ALDH1A1 and CM026. (C) The electron density maps of CM026, with the original $F_o – F_c$ map in green contoured at 2.5 standard deviations and the final $2F_o – F_c$ map in grey contoured at 1.0 standard deviations. Figures A and C generated in Pymol, while Figure B generated in ChemDraw.
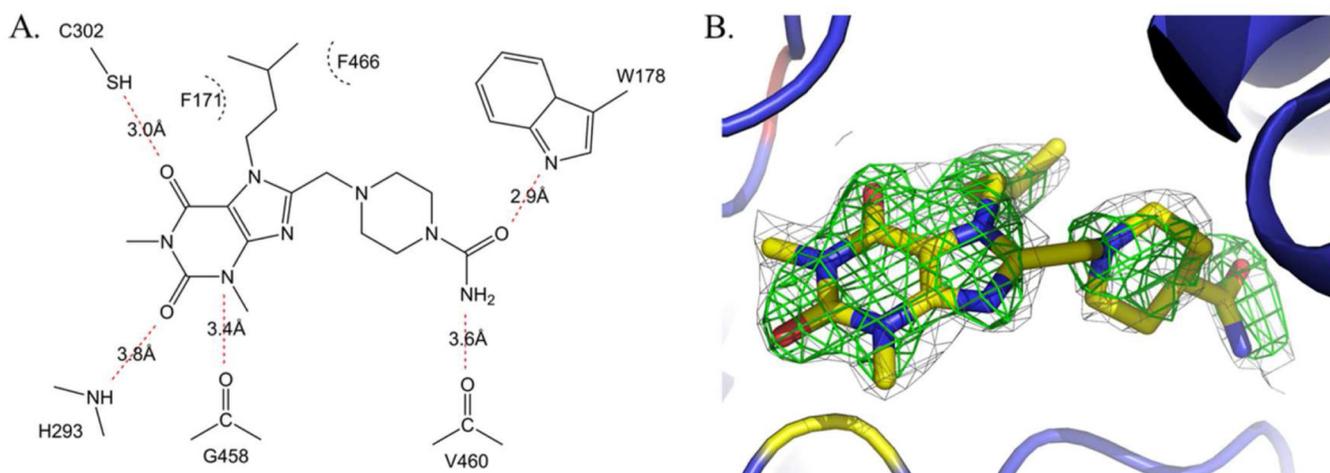
**Figure 5.**
Structure of human ALDH1A1 with CM053 (PDB Code 4WPN). (A) Two dimensional representation of the key hydrogen bonds, illustrated with red dashed lines, and hydrophobic interactions, illustrated with black arcs, between ALDH1A1 and CM053. (B) The electron density maps of CM053 with the original $F_o - F_c$ map in green contoured at 2.5 standard deviations and the final $2F_o - F_c$ map in grey contoured at 1.0 standard deviations. Figure A generated in ChemDraw and figure B generated in Pymol.
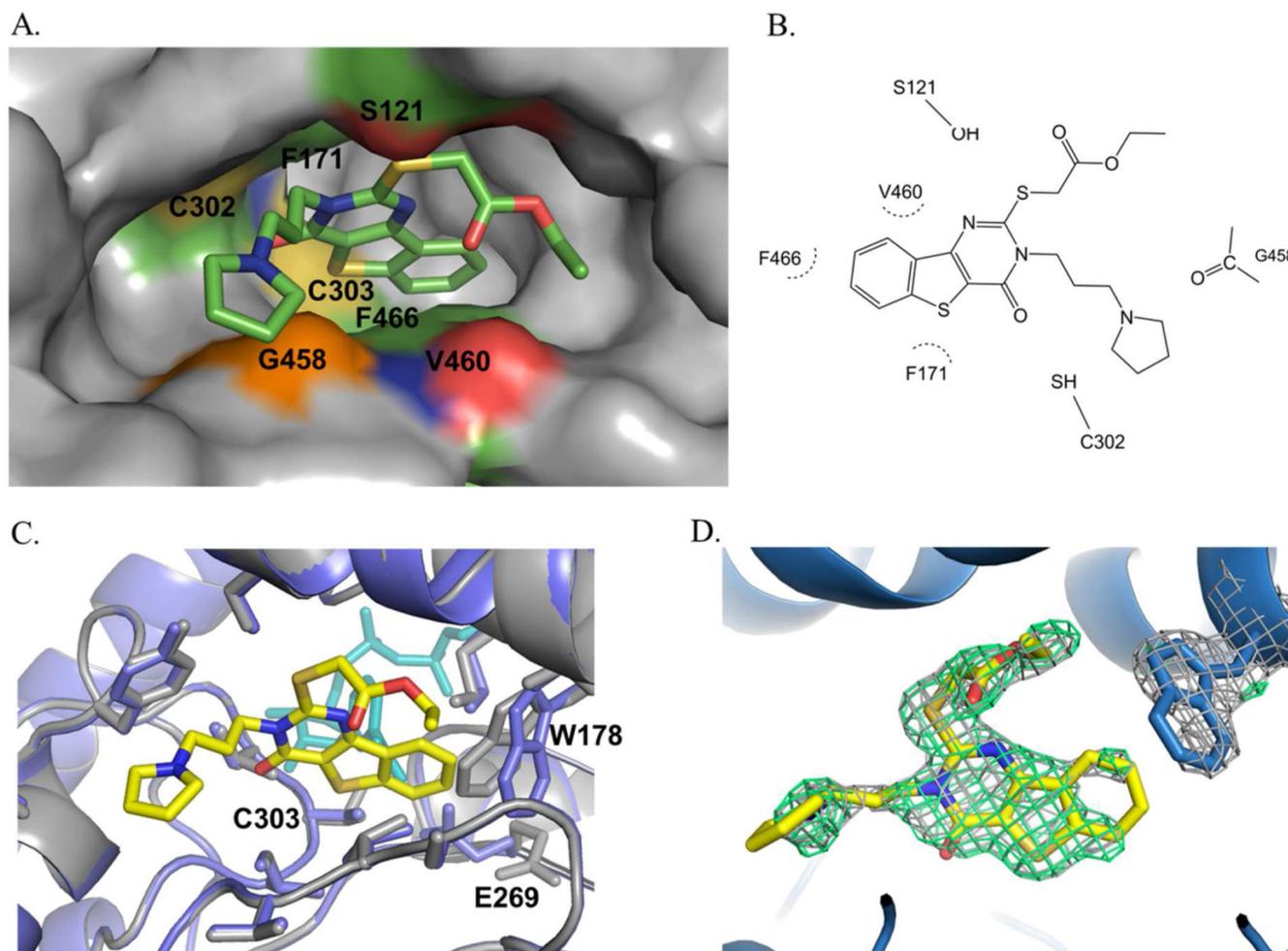
**Figure 6.**
Structure of ALDH1A1 N121S with CM037 (PDB Code 4X4L). (A) CM037 binds in the active site near Cys 303. Key residues are colored based on element, and Gly458 is shown in orange. (B) Two-dimensional representation of the hydrophobic interactions, illustrated with black arcs, between ALDH1A1 and CM037. (C) Binding of CM037 induces structural changes in ALDH1A1 (in blue) compared to apo-ALDH1A1 (in gray), particularly at W178. NADH binding (in cyan), induces conformational changes at the cofactor binding site, as seen here with E269. (D) The electron density maps of CM037, with the original $F_o - F_c$ map in green contoured at 2 standard deviations and the final $2F_o - F_c$ map in grey contoured at 1.0 standard deviations. Figures A, C and D generated in Pymol, while Figure B generated in ChemDraw.

A.

```
ALDH2      TYGLAAAVFTKDLDKANYLSQAL--Q--AGTVWVNCYDVFGAQS--PFGGY   485
ALDH1A1    FYGLSAGVFTKDIDKAITISSAL--Q--AGTVWVNCYGVVSAQC--PFGGF   469
ALDH1A2    DFGLVAAVFTNDINKALTVSSAM--Q--AGTVWINCYNALNAQS--PFGGF   486
ALDH1A3    DYGLTAAVFTKNLDKALKLASAL--E--SGTVWINCYNALYAQA--PFGGF   480
ALDH1B1    RYGLAAAVFTRDLDKAMYFTQAL--Q--AGTVWVNTYNIVTCHT--PFGGF   485
ALDH1L1    EFGLASGVFTRDINKALYVSDKL--Q--AGTVFVNTYNKTDVAA--PFGGF   885
ALDH1L2    EYGLASGVFTRDINKAMYVSEKL--E--AGTVFINTYNKTDVAA--PFGGV   896
ALDH3A1    EKPLALYMFSSNDKVIKKMIAET--S--SGGVAANDVIVHITLHslPFGGV   405
ALDH3A2    EKPLALYVFSHNHKLIKRMIDET--S--SGGVTGNDVIMHFTLNsfPFGGV   402
ALDH3B1    EKPLALYAFSNSSQVVKRVLTQT--S--SGGFCGNDGFMHMTLAslPFGGV   405
ALDH3B2    EKPLALYAFSNSSQVVNQMLERT--S--SGSFGGNEGFTYISLLsvPFGGV   324
ALDH4A1    SYGLTGAVFSQDKDVVQEATKVL--RnaAGNFYINDKSTGSIVGqqPFGGA   523
ALDH5A1    DVGLAGYFYSQDPAQIWRVAEQL--E--VGMVGVNEGLISSVEC--PFGGV   520
ALDH6A1    PYGNGTAIFTTNGATARKYAHLV--D--VGQVGVNVPIPVPLPMf-SFTGS   487
ALDH7A1    KQGLSSSIFTKDLGRIFRWLGPKgsD--CGIVNVNIPTSGAEIGg-AFGGE   499
ALDH8A1    KYGLAATVWSSNVGRVHRVAKKL--Q--SGLVWTNCWLIRELNL--PFGGM   460
ALDH9A1    TFGLAAGVFTRDIQRAHRVVAEL--Q--AGTCFINNYNVSPVEL--PFGGY   484
ALDH16A1   PRGGSASVWSERLGQALELGYGL--Q--VGTVWINAHGLRDPSV--PTGGC   467
ALDH18A1   PVGLEGLLTTKWLLRGKDHVVSDfsEh--GSLKYLHENLPIPQRntN----   795
```

B.
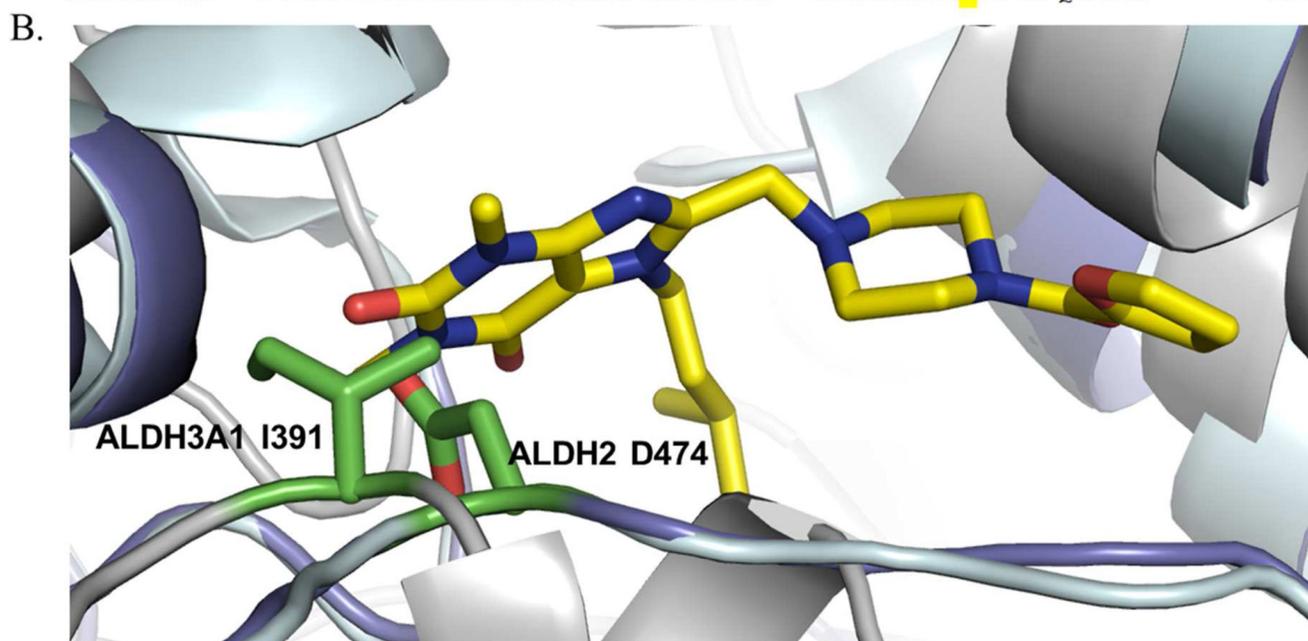


**Figure 7.**
Structural basis of selectivity of CM026 for ALDH1A1. (A) Multiple sequence alignment in the region of ALDH1A1 Gly458 to the mature form of ALDH2. (B) Structure of ALDH1A1 with bound CM026 (blue) compared to ALDH2 (cyan) and ALDH3A1 (grey) indicating that a bulky amino acid such as the Asp of ALDH2 and Ile of ALDH3A1 would clash with CM026 and prevent the compound from inhibiting the enzyme. Sequence alignment was performed using NCBI delta-BLAST while structural alignment was performed using least

square fit (LSQ) in Coot. Figure B generated in Pymol. PDB Codes: ALDH1A1 (4WP7), ALDH2 (2VLE), and ALDH3A1 (3SZA)
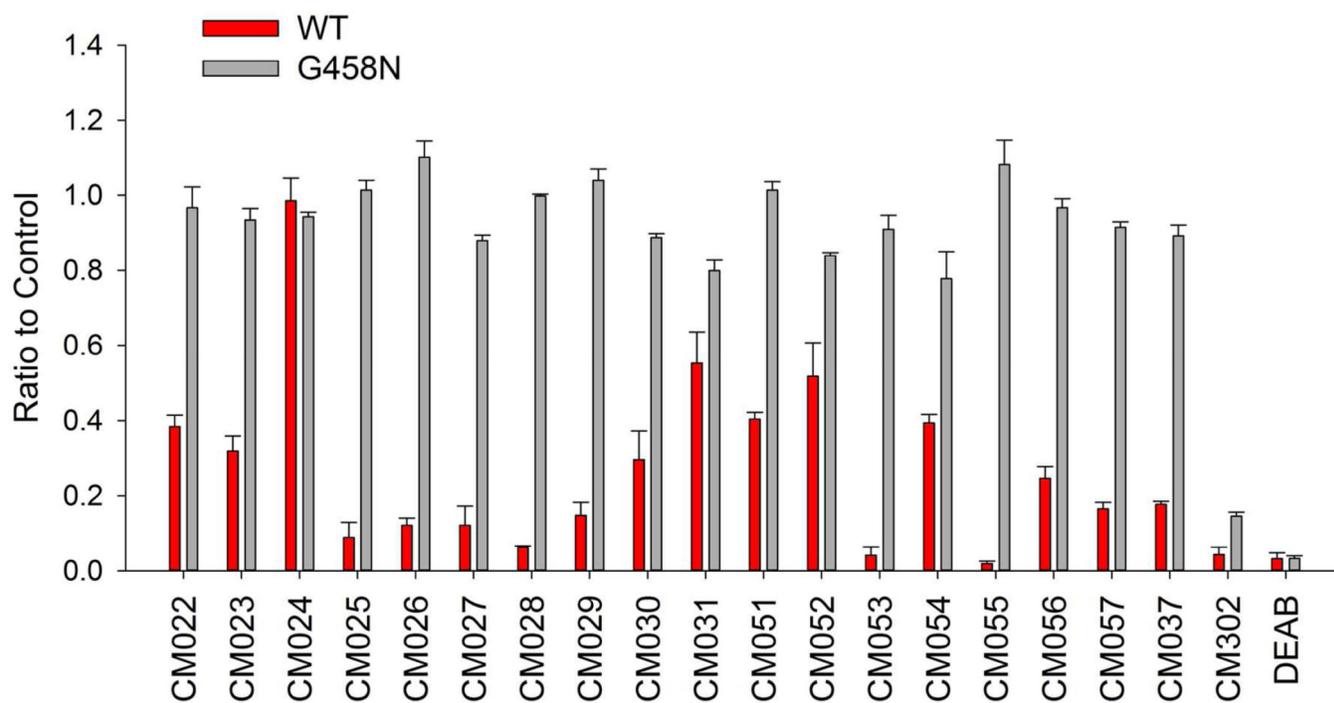
**Figure 8.**
Selectivity of compounds for WT vs G458N mutant. For CM026 and its analogs, 100 μM of compound was used. For CM037 and CM302, 20 μM of compound was used. Each value is mean/SEM (n=3).
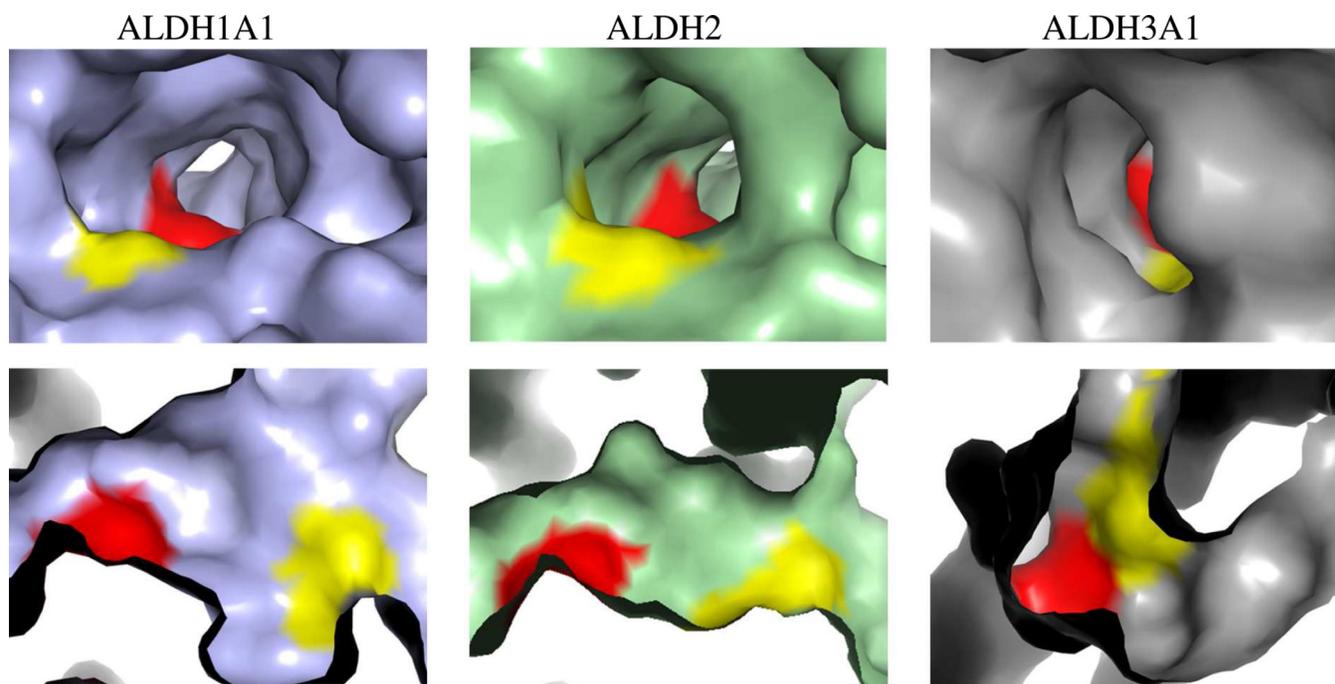
**Figure 9.**
Comparison of the active site topography of human ALDH1A1, ALDH2, and ALDH3A1.
The three isoenzymes were aligned using LSQ in Coot and the surface figures generated via
Pymol. The active site cysteine is shown in red for all three isoenzymes. G458 in ALDH1A1
(PDB Code 4WJ9) and its equivalent residues D457 in ALDH2 (PDB Code ICW3) and I391
in ALDH3A1 (PDB Code 3SZA) are shown in yellow.
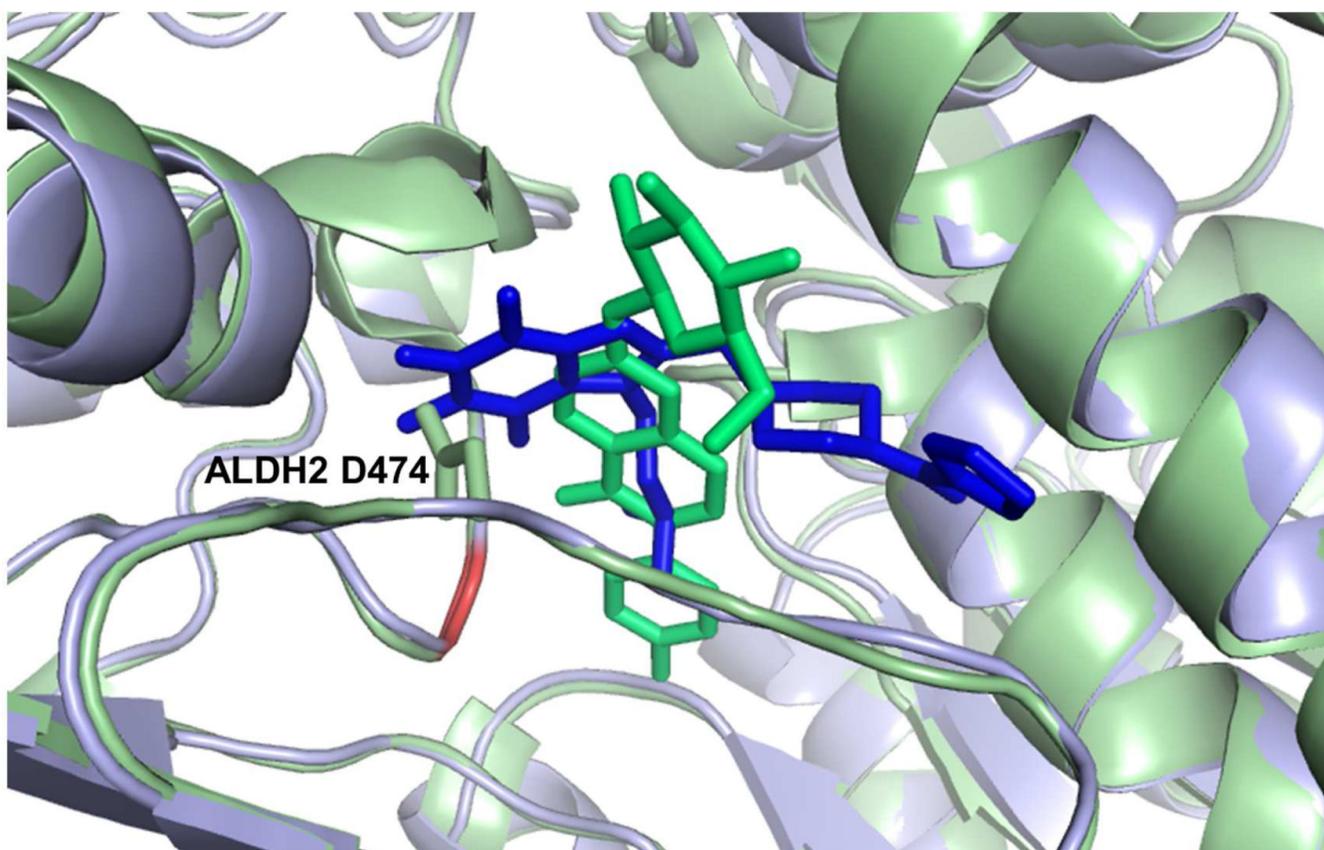
**Figure 10.**
Comparison of the binding of CM026 to ALDH1A1 (PDB Code 4WP7) with the binding of
daidzin to ALDH2 (PDB Code 2VLE). ALDH1A1 is shown in light blue with CM026 in
dark blue. ALDH2 is shown in light green with daidzin in dark green. The active site
cysteine is shown in red for both isoenzymes. Although both compounds bind in the same
location, they bind in a different orientation enabling CM026 to inhibit ALDH1A1 but not
ALDH2 due to steric hindrance of D474 (D457 in the mature sequence, in orange), while
daidzin inhibits both isoenzymes. Figure generated in Pymol.

**Table 1**

SAR for CM026 Analogs.

| | R1 | R2 | IC$_{50}$ in µM (SE) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ALDH 1A1 | ALDH 1A2 | ALDH 1A3 | ALDH2 | ALDH 1B1 | ALDH 3A1 |
| CM022 | | | 5.2* (0.8) | >100 | NI(A) | NI | NI | NI |
| CM024 | | | >20 | NI | >100 | NI | >100 | NI |
| CM031 | | | >20 | >100 | >20 | >100 | >100 | NI(A) |
| CM053 | | | 0.21 (0.04) | NI | NI(A) | NI | NI | NI(A) |
| CM026 | | | 0.80 (0.06) | NI(A) | NI(A) | NI | NI(A) | NI(A) |
| CM057 | | | 0.92 (0.2) | NI | >100 | >100 | >100 | NI |
| CM054 | | | 3.4* (0.7) | >100 | NI(A) | >100 | >100 | NI |
| CM056 | | | 5.8 (1.2) | NI | NI | >100 | >100 | NI |

| | R1 | R2 | IC$_{50}$ in µM (SE) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ALDH 1A1 | ALDH 1A2 | ALDH 1A3 | ALDH2 | ALDH 1B1 | ALDH 3A1 |
| CM030 | | | >20 | NI | NI | NI | NI | NI |
| CM055 | | | 0.24 (0.04) | NI(A) | NI(A) | NI | NI | NI |
| CM029 | | | 8.4 (1.0) | NI | NI(A) | NI | NI | NI |
| CM025 | | | 2.1 (0.7) | NI | NI(A) | NI | >100 | NI(A) |
| CM023 | | | 14* (2) | NI | NI(A) | NI | NI | >100 |
| CM051 | | | >20 | NI(A) | NI(A) | NI | NI | NI |
| CM052 | | | >20 | >100 | NI(A) | NI | NI(A) | NI(A) |
| CM027 | | | 6.1 (1.1) | NI | NI | NI | NI | NI |
| CM028 | | | 2.0 (0.1) | NI | NI | >100 | >100 | NI |

| | | | IC$_{50}$ in µM (SE) | | | | | |
|---|---|---|---|---|---|---|---|---|
| R1 | R2 | ALDH 1A1 | ALDH 1A2 | ALDH 1A3 | ALDH2 | ALDH 1B1 | ALDH 3A1 |
| Theophylline H | H | NI | NI | NI | NI | NI | NI |
| Caffeinea CH$_3$ | H | NI | NI | NI | NI | NI | NI |
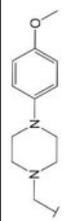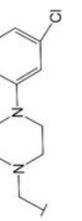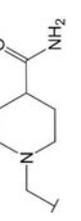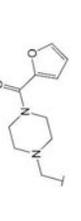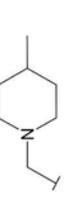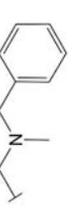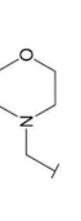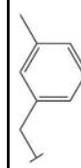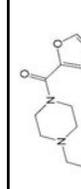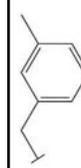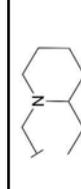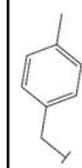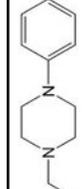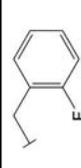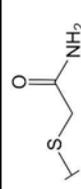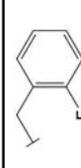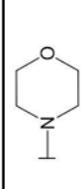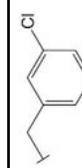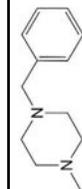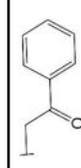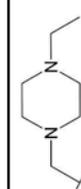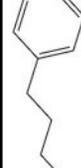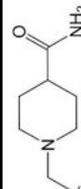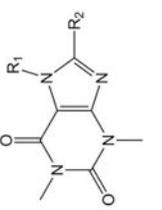
At 100 µM compound, NI stands for no inhibition and NI(A) indicates no inhibition but activation

*
indicates < 70% maximum inhibition).

**Table 2**

Data collection and refinement statistics.

| Data Collection | ALDH1A1-CM026 PDB 4WP7 | ALDH1A1-CM053 PDB 4WPN | ALDH1A1-CM037 PDB 4X4L |
|---|---|---|---|
| Space Group | P422 | P422 | P422 |
| Cell Dimensions | | | |
| a, b, c (Å) | 109, 109, 83 | 109, 109, 83 | 109, 109, 83 |
| A,β, γ (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 50 – 1.80 | 50 – 1.95 | 50 – 1.85 |
| $R_{merge}$ | 0.082 (0.59) | 0.11 (0.66) | 0.058 (0.70) |
| $I/\sigma_i$ | 22.7 (4.7) | 18.3 (3.7) | 27.4 (3.3) |
| Completeness (%) | 99 (100) | 99 (100) | 99 (100) |
| Redundancy | 11.7 (11.7) | 9.3 (6.9) | 8.5 (8.8) |
| **Refinement** | | | |
| No. of Reflections | 44544 | 35048 | 40517 |
| $R_{work}/R_{free}$ | 0.19 / 0.22 | 0.19 / 0.24 | 0.19 / 0.22 |
| No. of Atoms | 4109 | 4080 | 4066 |
| Protein | 3833 | 3858 | 3806 |
| Ligand/Ion | 35 | 33 | 80 |
| Water | 241 | 189 | 180 |
| R.M.S. Deviations | | | |
| Bond Lengths (Å) | 0.008 | 0.009 | 0.010 |
| Bond Angles (°) | 1.28 | 1.30 | 0.135 |

Numbers in parenthesis represent values of highest resolution shell.

**Table 3**

Kinetic parameters of ALDH1A1 WT and mutant G458N.

| | $K_M^{Acetaldehyde}$ (µM) | $k_{cat}/K_M$ (min$^{-1}$ µM$^{-1}$) | $K_i^{CM026}$ (µM) | $K_i^{CM037}$ (µM) | $IC_{50}^{CM026}$ (µM) | $IC_{50}^{CM037}$ (µM) | $IC_{50}^{CM302}$ (µM) | $IC_{50}^{DEAB}$ (µM) |
|---|---|---|---|---|---|---|---|---|
| WT | 177±19 | 0.18±0.02 | 0.80±0.16 | 0.23±0.06[42] | 0.80±0.06 | 4.6±0.8[42] | 1.0±0.1[40] | 0.057±0.005[39] |
| G458N | 85.8±1.6 | 0.21±0.02 | | | NI | NI | 3.1±0.3 | 0.52±0.10 |