STRUCTURAL AND KINETIC STUDIES OF HUMAN ALDH7A1 AND ALDH9A1

_____

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

JESSE W. WYATT

Dr. Kent S. Gates, Dissertation Supervisor

Dr. John J. Tanner, Dissertation Supervisor

July 2021

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

STRUTURAL AND KINETIC STUDIES OF HUMAN ALDH7A1
AND ALDH9A1

Presented by Jesse W. Wyatt

a candidate for the degree of Doctor of Philosophy

and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Professor Kent Gates

_____

Professor John Tanner

_____

Professor Michael Harmata

_____

Professor Lesa Beamer

ACKNOWLEDGEMENTS

I would like to thank my family for being supportive through my tenure in graduate school. Their endless encouragement and willingness to help whenever they can have been vital to my success throughout this process. I would further extend thanks to my friends, Nate, Bryanne, and Nick, who have been steadfast in their understanding and provided guidance and support when I needed it. I also thank my partner, Ling, without whom I would not have made it through the last year during rather extraordinary changes.

Through this process, I have relied on and enjoyed the conversations with members of both my research groups. I consider them to be role models, and truly supportive friends. I am especially thankful to Kurt, Ashley, Allie, and Jay. Their willingness to talk about whatever came to mind and enjoying social outings have always left me with fond memories.

The guidance provided by Dr. Gates and Dr. Tanner, have been essential for my growth and development. Their willingness to present the hard truth and tell you how it is, truly made me more dedicated to the field and helped me grow as a person. I thank both for providing insight into my research, and for providing a positive and friendly working environment.

I extend a special thank you to my GPC family. Working with the various members of the executive board has given me not only confidence with my research, but also made me more confident in life. I was able to develop strong leadership and communication skills with their help and support.

I would also like to thank Drs. Harmata and Beamer for accepting to be on my thesis committee and provide support when needed.

Table of Contents

LIST OF FIGURES

Chapter One Figures

Chapter Two Figures

Chapter Three Figures

Chapter Three Figures

LIST OF TABLES

LIST OF SCHEMES

Chapter One Schemes

Chapter Three Schemes

STRUTURAL AND KINETIC STUDIES OF HUMAN ALDH7A1 AND ALDH9A1

Jesse W. Wyatt

Dr. Kent S. Gates, Dissertation Advisor

Dr. John J. Tanner, Dissertation Advisor

ABSTRACT

The regulation and detoxification of endogenously and exogenously derived aldehydes is paramount to cellular survival due to the highly reactive nature of aldehydes as electrophiles. Human aldehyde dehydrogenases (ALDHs) are a superfamily of oxidoreductase enzymes that have critical roles in this regulation and detoxification. Misregulation of ALDH gene expression or mutations in the genes encoding for ALDHs lead to numerous disease pathologies. While extensive work has been conducted in understanding the metabolic roles and structures of these enzymes, there remains a need to further expand the structural and kinetic understanding of members of the human ALDH superfamily. This thesis aims to utilize the tools of structural biology and enzymology to expand the understanding of the ALDH superfamily.

Chapter One: Introduction to aldehyde dehydrogenase superfamily

1.1. Introduction to NAD(P)/NAD(P)H-dependent oxidoreductases

Enzymes carry out a vast array of biological functions and are classified into seven categories with an assigned primary class number given by the Enzyme Commission [EC] based on the chemical reactions catalyzed. These classes comprise the oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5), ligases (EC 6), and translocases (EC 7). The work presented in subsequent chapters focuses on enzymes of the oxidoreductase class. Oxidoreductases comprise nearly one-third of all enzymatic activities reported in the Braunschweig Enzyme Database (BRENDA) [1], and 6.5% of protein structures reported in the Research Collaboratory for Structural Bioinformatics protein database (RCSB PDB). Enzymes of this class catalyze redox reactions in the form of $A^- + B \rightarrow A + B^-$ by oxidation of a reducing agent or electron donor and reduction of an oxidizing agent or electron acceptor. Their reactivity ranges from organic substrates (alcohols, amines, and ketones) to inorganic substrates (small anions or metals).

Of the oxidoreductases, NAD(P)/NAD(P)H-dependent enzymes comprise over 50% of the oxidoreductases reported in BRENDA and serve as the primary focus of the work presented here. NAD(P)H-dependent oxidoreductases oxidize substrates by facilitating the transfer of a hydride to the nicotinamide moiety of $NAD(P)^+$ at the C_4 position (Scheme 1.1 [2]). Typically, $NAD^+/NADH$ is utilized in catabolism whereas $NADP^+/NADPH$ is the preferred co-factor in anabolism [3,4]. Thus, NAD/NADH is more prevalent in catalysis involving smaller molecules, which include most substrates and products of catabolic pathways. These small molecules, such as aldehydes, often have central roles in redox metabolic pathways.

Scheme 1.1. Hydride transfer to the C_4 (in red) position of the nicotinamide ring of NAD^+.
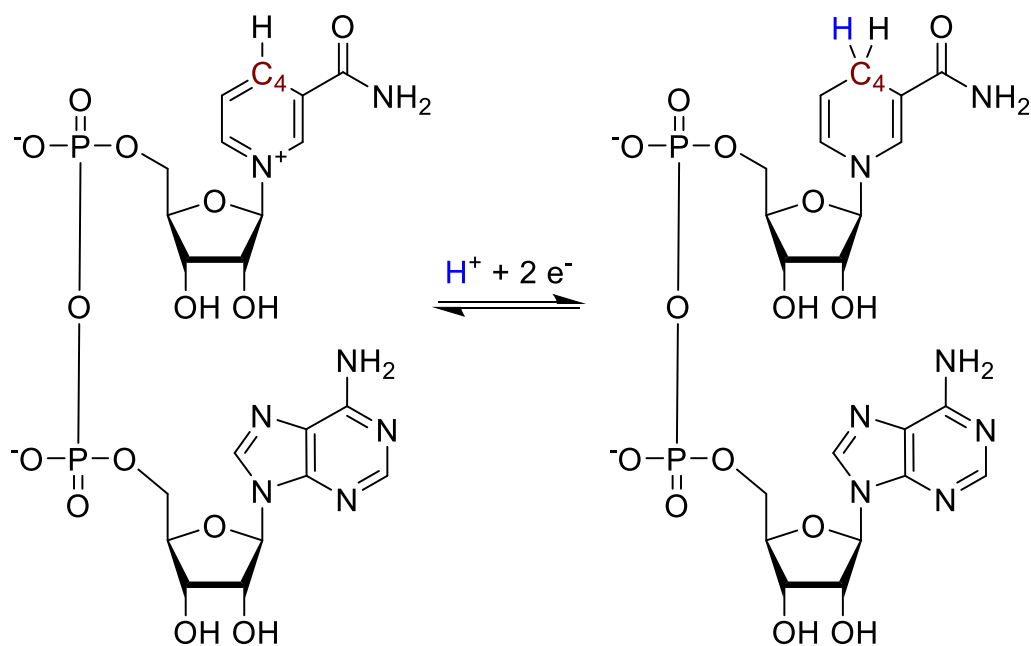
1.2. Aldehyde generation and toxicity

Aldehydes are highly reactive, strongly electrophilic chemical compounds containing a carbonyl group bonded to a hydrogen and an R-group, in the form; R-CHO. These reactive compounds can be endogenously generated from the metabolism of alcohols, amino acids, anticancer drugs, carbohydrates, neurotransmitters, and lipid peroxidation [5–7]. The autoxidation of polyunsaturated fatty acids resulting from oxidative stress is initiated by reactive oxygen species (ROS), which are formed by metabolic processes: oxidation of NADPH by NADPH oxidase, uncoupling of the mitochondrial electron transport chain, and oxidation of xanthine by xanthine oxides [8]. Physiologically abundant hydroxy or superoxide radicals act as chemical initiators to start the initiation event. Following the initiation event, propagation or termination follows [9,10]. This ultimately

can generate more than 200 aldehyde species, with 4-hydroxy-2-nonenal (4-HNE), malondialdehyde, 4-oxo-2-nonenal (4-ONE) and acrolein as the more common aldehydes generated [5]. Enzymatically derived aldehydes play important roles in metabolic pathways such as amino acid synthesis and vitamin metabolism [11]. At low concentrations aldehydes can also have roles in regulating cell proliferation [12,13] such as is the case for retinoic acid in the expression or repression of genes in embryonic development [14,15]. This important biomolecule is generated from the conversion of retinaldehyde by members of aldehyde dehydrogenase superfamily; ALDH1A1, ALDH1A2, and ALDH1A3 [16].

While aldehydes often play important roles in metabolism or cellular regulation, the overall high reactivity of aldehydes can often lead to undesirable modifications of cellular DNA and proteins, or mis-regulation of cellular processes. Modifications in the form of interstrand cross-links can be derived from the accumulation of aldehydes, such as acetaldehyde, malondialdehyde, acrolein, and 4-HNE [17,18]. Like DNA, proteins can be cross-linked either intramolecularly or intermolecularly. The aldehyde 4-ONE can facilitate these protein-protein crosslinks through first forming a Michael product with an amino acid and then formation of a Schiff base with an adjacent lysine [8,19]. Aldehydes in general have been implicated in degenerative diseases (e.g. liver cirrhosis, diabetic nephropathy), neurodegenerative processes (e.g. Alzheimer's, Parkinson's Disease), acute alcohol exposure, diabetes, and cancer [18]. Thus, cells have evolved systems for sequestering free aldehydes with glutathione for further processing or by direct catalysis of the aldehyde into other carbon sources, e.g., alcohols, carboxylic acids, acetylated molecules. This removal or transformation of aldehydes can be carried out by alcohol

dehydrogenase, P-450s, or aldehyde dehydrogenase; the later serving as the primary focus of this work [11].

1.3. Aldehyde dehydrogenase superfamily

Aldehyde dehydrogenases (ALDHs; EC 1.2.1.3) represent a superfamily of oxidoreductases that can catalyze the oxidation of endogenously or exogenously derived aliphatic and aromatic aldehydes to their corresponding carboxylic acid forms [20–22]. This enzyme family is involved in vitamin, carbohydrate, amino acid, and lipid metabolism [13,21,22]. For example, ALDH5A1 and ALDH9A1 are involved in catalyzing intermediates in the synthetic pathway for the inhibitory neurotransmitter γ-aminobutyric acid (GABA). Some aldehyde dehydrogenases also have an expanded catalytic activity and are capable of ester hydrolysis [23] or nitrate reduction [24]. Yet still, other ALDHs have structural roles such as ALDH3A1, which in addition to its ability to remove 4-HNE [25,26], forms crystallins in the cornea and lens to act as an UV-absorber [27,28]. ALDH1A1 has also been shown to detoxify 4-HNE and can additionally form crystallin structures in some mammalian species [5].

The human ALDH superfamily contains 19 known genes. These are classified based on their sequence identity [29] and help in the understanding of their evolutionary divergence (Figure 1.1). ALDH proteins containing ≥ 40% identity are assigned to a particular family designated by an Arabic numeral (e.g. ALDH**1**, ALDH**2**). The proteins having a sequence identity ≥ 60% are part of the same sub-family designated by a letter (e.g. ALDH3**A**1, ALDH3**B**1). Of the 19 human isoforms, 13 have had crystal structures solved so far. These crystal structures reveal that aldehyde dehydrogenases consist of a three-domain structure containing an NAD(P)$^+$-binding domain, catalytic domain, and an

4

oligomerization domain [30]. Some ALDHs contain additional domains flanking either the

N- or C-terminus of the three-domain core. (e.g. ALDH1L1, ALDH1L2, ALDH18A1).
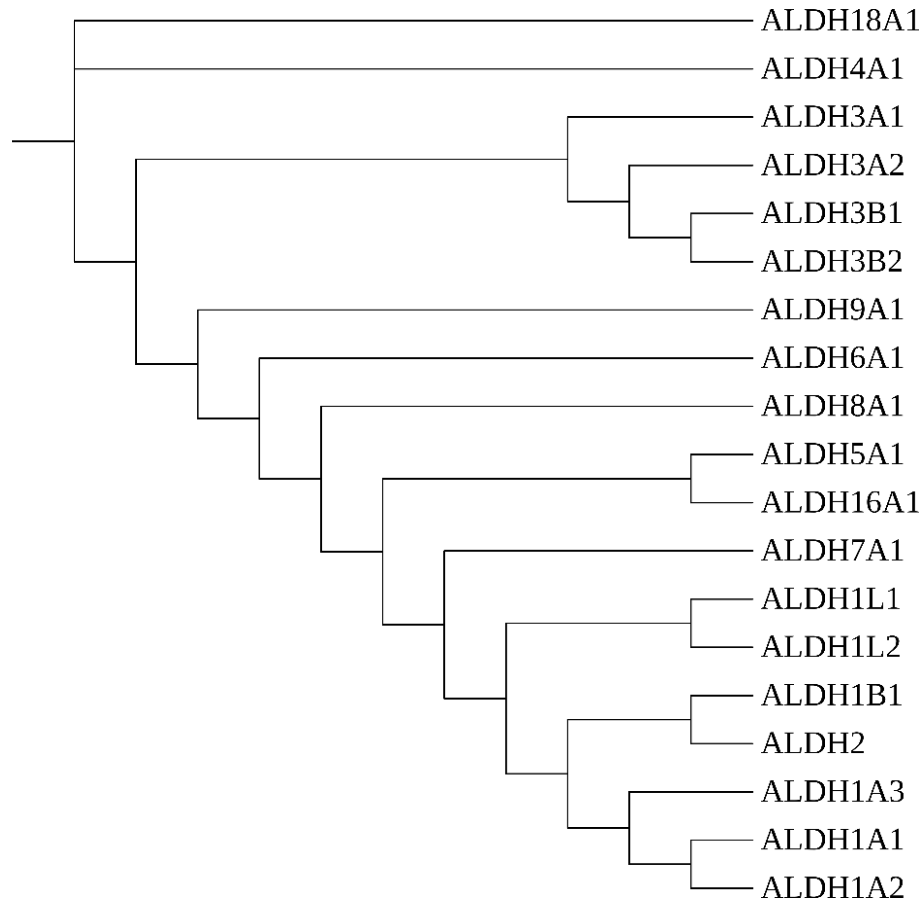
Figure 1.1. Dendrogram of the human ALDH superfamily showing similarity of the
isoforms.

NAD(P)$^+$-binding Domain

As with other NAD(P)$^+$-dependent oxidoreductases, the co-factor binding site is

essential for enzyme activity. The NAD(P)$^+$-binding domain starts at the N-terminus and

constitutes about half the residues of the monomer. The binding of NAD(P)$^+$ co-factor is

facilitated by the domain's adoption of a Rossmann fold. This fold is composed of six

parallel beta strands forming an extended beta sheet with intermittent alpha helices in a

βαβ motif. This pattern is subsequently repeated once to produce an inverted tandem repeat, with an overall arrangement of 32145 (Figure 1.2a)[31,32]. Binding of the adenine moiety is highly conserved across ALDHs, with this portion of the cofactor interacting with the αA helix [33], which provides favorable interactions to the pyrophosphate moiety (Figure 1.2b). The preference for binding of NAD^+ or $NADP^+$ is facilitated by an acidic residue of the β2 sheet. The presence of this acidic residue destabilizes the binding of $NADP^+$ within the active site, lending to the preference for NAD^+ binding. In contrast, the absence of the acidic residue eliminates repulsion effects that would occur between the residue and the $NADP^+$ phosphate [34,35]. Human ALDH3A1 and ALDH3A2 can use $NADP^+$ despite the presence of a glutamate residue on the β2 sheet.[35] In general, the NAD(P)+ co-factor is tightly bound in a snug conformation (Figure 1.2c).
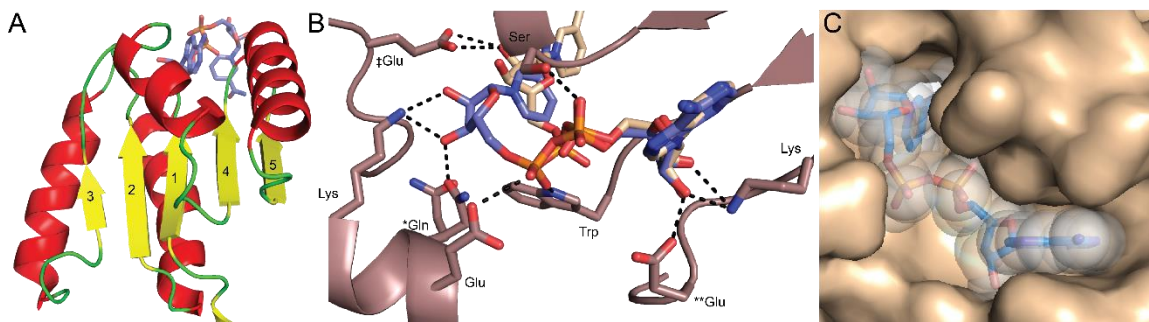
Figure 1.2. Structure of the Rossmann fold of human ALDH and binding of the NAD^+ cofactor**.** (A) ALDH1A1 complexed with NADH (PDB ID: 4WB9). Numbering of the β-sheets shows the tandem repeat pattern. (B) Bonding of the NADH (blue) by residues of the NAD+-binding domain of ALDH1A1 in the hydrolysis conformation (PDB: 4WB9). NAD+ (wheat; PDB: 6O4C) is in the hydride conformation. The * indicates different conformations for the glutamine residue. The ** indicates the glutamate residue responsible for NAD^+ or $NADP^+$ specificity. The ‡ represents a conserved glutamate residue binding to the ribose of the nicotinamide moiety (C) Space filling model of NADH shows the tight-fitting co-factor binding pocket.

Oligomerization Domain

Oligomerization of ALDHs is facilitated by the oligomerization domain, which directs the oligomerization of the monomer. Two common occurrences of oligomerization have

been observed in solved structures of ALDHs: the dimer and the tetramer (Figure 1.3a). Differences in the length of the C-terminus have been proposed to dictate the oligomeric state of the protein, whether it adopts a primarily dimeric or tetrameric state [36,37]. For dimeric ALDHs, the C-terminus is extended by approximately 17 residues, while the N-terminus is truncated by about 56 residues when compared to their tetrameric counter parts (Figure 1.3b-c) [21]. This extension of the C-terminus likely disrupts the formation of the tetramer by occupying the domain space of the second dimer. Recent evidence suggests that in some cases an extended C-terminus can stabilize a tetrameric structure, as is the case with *T. thermophilus* HB27 (TtALDH530) [38], where the extended tail wraps around the monomer of the opposing dimer. It should be noted that ALDH7A1, a tetrameric ALDH, contains both an extended C-terminus and N-terminus (Figure 1.3d). Studies of ALDH7A1 oligomerization in the absence and presence of NAD^+ showed that binding of the cofactor induced tetramer formation, possibly alluding to conformational changes of the ALDH7A1 monomer/dimer to facilitate tetramerization. Whether the conformational changes can be attributed to movement of the extended C-terminus has yet to be investigated.

A third oligomeric state, the hexamer, has been reported for some ALDHs and is comprised of three dimers on a 3-fold axis (Figure 1.3a) A hexamer, a trimer of dimers,

was reported for ScALDH4A1 [39] and TtP5CDH [40], though no human hexametric ALDHs has been reported.
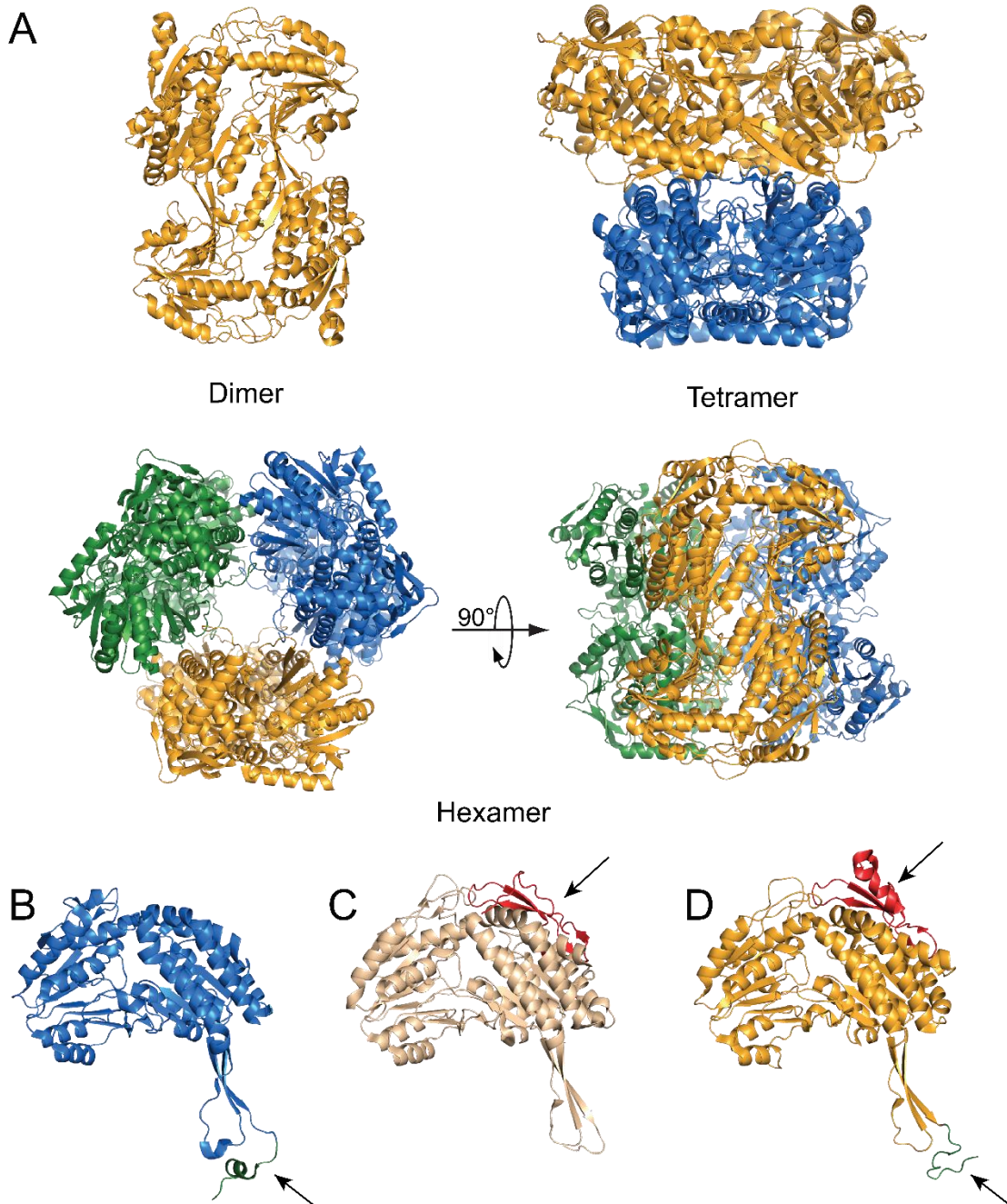
Figure 1.3. Oligomeric states of ALDHs. (A) Representation of a dimer (ALDH7A1; PDB: 4ZUK), tetramer (ALDH7A1; PDB: 4ZUK), and hexamer (P5CDh; PDB: 2BHQ). (B) Dimeric ALDH3A1 (PDB ID: 3SZA) with extended C-terminal tail (green). (C) Tetrameric ALDH1A1 (PDB ID: 4WJ9). Note the extended N-terminus (red) and truncated C-terminus. (D) ALDH7A1 (PDB ID: 4ZUK) contains both an extended N-terminus (red) and C-terminus (green).

Catalytic Domain

The catalytic domain comprises about 40 % of the enzyme and contains the catalytic

cysteine found in all human ALDHs, except for ALDH16A1, which lacks catalytic activity

and has been proposed to be a binding protein [41]. Significant conserved features of the

catalytic domain are the catalytic loop which contains the conserved catalytic cysteine, the
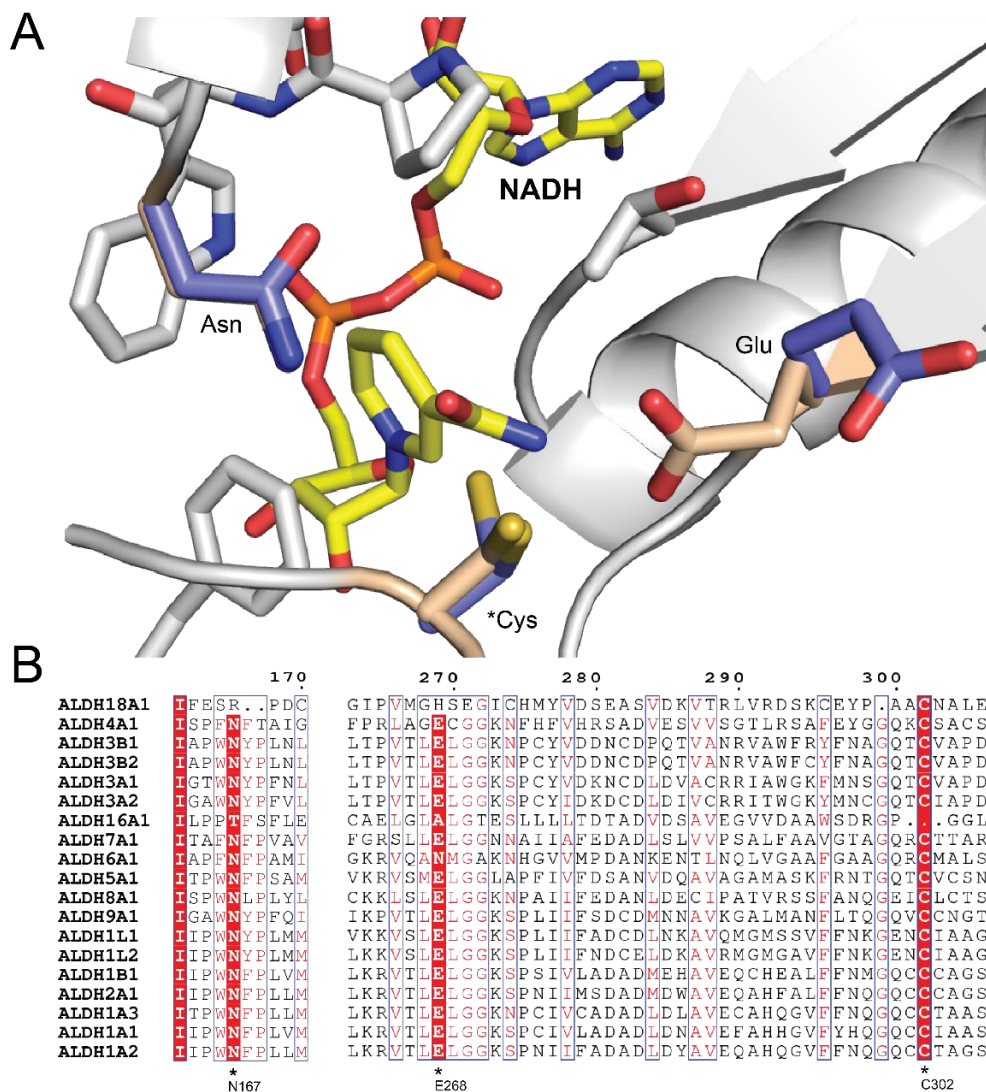
Figure 1.4. Conserved residues of the ALDH active site. (A) Representation of the conserved active site residues of human ALDHs. The * denotes the catalytic cysteine residue (PDB: 4WB9). (B) Sequence alignment of the 19 human ALDH isoforms. The active site residues are indicated by an asterisk, and numbering is from ALDH1A1.

conserved glutamate residue required for hydrolysis of the acetal intermediate (Figure 1.4),

and in some cases an anchor loop (e.g. ALDH7A1). The ability of a substrate to enter the

active site is facilitated through the substrate entry channel, whose size and shape

determine whether small or large substrates can bind. This channel and differences of the

active site residues dictates the level of broad specificity a particular ALDH has for a

substrate [15].

Detoxification of aldehyde substrates by the catalytic domain, is carried out by the

conserved cysteine residue, which acts as the nucleophile [42]. A local glutamate and

ordered water activate the catalytic cysteine through abstraction of the thiol proton [43].

Following the binding of NAD(P)$^+$ [44], the aldehyde substrate binds with the aldehyde

functional group positioned near the sulfur of the catalytic cysteine. Upon nucleophilic

attack by the thiol, a tetrahedral thiohemiacetal intermediate is formed. The collapse of the

intermediate facilitates the transfer of the aldehyde hydride to the electron acceptor

NAD(P)$^+$ generating NAD(P)H. Activation of water by the conserved active site glutamate

residue promotes hydrolysis of the thioester intermediate [6,21,45] and is followed by

subsequent release of the carboxylic acid product and the NAD(P)H cofactor (Scheme 1.2).

Scheme 1.2. Catalytic mechanism of ALDHs.
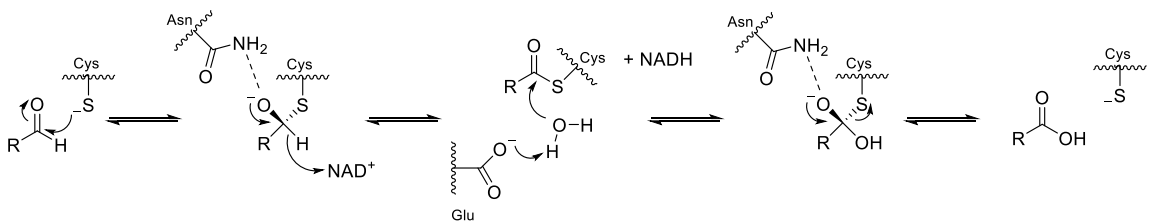
1.4. Role of ALDHs in disease

As discussed earlier, ALDHs play key roles in several biochemical pathways, in addition to several non-catalytic activities. Due to their importance in these pathways and their protective nature of removing reactive aldehydes, disruption in the proper function of theses enzymes can lead to negative or unfavorable effects on cellular systems [13]. Mutations of ALDHs have been linked to several diseases including Parkinson's disease, spina bifida, alcohol intolerance, pyridoxine-dependent epilepsy, and hyperammonemia (Table 1.1). For example, ALDH3A2 has been linked to Sjörgen-Larsson Syndrome and ALDH4A1 leads to a predisposed risk of type II hyperprolinemia. Mutations of ALDH7A1 lead to pyridoxine-dependent epilepsy, which will be explored further in chapter two.

While mutations of ALDHs can perturb the activity of the enzyme, the over expression of ALDHs has been reported extensively in cancer research, with specific interest in cancer stem cells (CSCs) [46,47]. These progenitor cells not only enhance tumorgenicity, but also engage in self-renewal and differentiation [48–50]. Further, CSCs can recapitulate the tumor, making these cells viable targets for cancer therapy. ALDHs have been reported to be a common biomarker for CSCs [51,52] with the isozyme ALDH1A1 being one of the most studied human ALDHs. Continued studies have expanded the number of ALDH enzymes as markers with isozymes ALDH1A2 and ALDH1A3 being more recently studied.

Detection of CSCs via the ALDH biomarker is performed through the ALDEFLUOR assay. Originally designed for detection of ALDH1A1 in leukemia cells, the assay monitors the generation of BODIPY-aminoacetate (BAA) from BODIPY-aminoacetaldehyde by the overexpressed ALDH [53]. The negatively charged BAA is subsequently trapped and

accumulated in the cytoplasm and can be detected and analyzed by flow cytometry. The use of the assay has been expanded to the detection of other isozymes of the ALDH family. A study by Zhou et al., showed the overexpression and detection of 9 of the 19 ALDH isoforms in cancer cells [54]. To date, ALDHs 1A1, 1A2, 1A3, 3A1, 4A1, 5A1, 6A1, and 9A1 have been linked to cancer stem cells.

Treatment of CSCs and cancer in general can be challenging due to cellular resistance to chemotherapeutic drugs by CSCs. This is often attributed to the high activity and overexpression of some ALDHs [46,54,55]. It is important to conduct metabolic and structural work as it can shed light on the role of ALDHs and provide a basis for the development of better therapeutics. Recent work in the development of ALDH1A1 inhibitors shows promise with lead compounds ranging in the low µM to mM for inhibition [56]. It is apparent that differences even within a sub-family, having greater than 60% sequence identity, can determine the effectiveness of compounds as broad inhibitors of ALDHs.

Table 1.1. Listing of diseases associated with human ALDHs.

Gene	Diseases	Ref
Aldh1a1	Parkinson's disease; cataract formation; colorectal, ovarian, myeloma, acute myeloid leukemia, and epithelial cancer	[57–60]
Aldh1a2	Spina bifida; Disrupted embryonic development; rare cases of congenital heart disease, prostate cancer	[61–64]
Aldh1a3	Disrupted embryonic development; breast cancer	[61,65]
Aldh1b1	Ethylmalonic encephalopathy; bipolar disorder; ethanol sensitivity; hypertension	[66–69]
Aldh1l1	Cancer	[70]
Aldh2	Heart attack; alcohol intolerance; hypertension; cancers; liver cirrhosis; Parkinson's disease; late onset Alzheimer's disease; cataract formation	[66,71–73]
Aldh3a1	Cataract formation; gastric cancer	[59,74]
Aldh3a2	Sjogren-Larsson syndrome	[11,75]
Aldh4a1	Type II hyperprolinemia	[11,76]
Aldh5a1	γ-hydroxybutyric aciduria	[77]
Aldh6a1	Psychomotor delay; methylmalonic aciduria	[78–80]
Aldh7a1	Pyridoxine-dependent epilepsy; osteoporosis	[81–84]
Aldh16a1	Gout	[85,86]
Aldh18a1	Hyperammonemia	[6]

1.5. Conclusions

Given the vital role of ALDHs in metabolic pathways (e.g. amino acid synthesis), reduction in toxic aldehydes (e.g. 4-HNE), cellular regulation (e.g RA pathway) and protection (e.g. UV exposure), and their importance to cancer stem cells it remains important to continue to expand the structural and metabolic understanding of this superfamily of enzymes. This thesis aims to expand the structure and metabolic understanding of the human aldehyde dehydrogenases. Herein, the examination of the impact of the c.1512delG mutation on the catalytic activity and oligomerization of ALDH7A1 will be presented. Additionally, the fully ordered structure of ALDH9A1 and slow-tight binding inhibition by diethylamino benzaldehyde (DEAB), a component of the ALDEFLUOR assay, is reported.

References

[1]   L. Sellés Vidal, C.L. Kelly, P.M. Mordaka, J.T. Heap, Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application, Biochim. Biophys. Acta - Proteins Proteomics. 1866 (2018) 327–347.

[2]   R. Meijers, R.J. Morris, H.W. Adolph, A. Merli, V.S. Lamzin, E.S. Cedergren-Zeppezauer, On the Enzymatic Activation of NADH *, J. Biol. Chem. 276 (2001) 9316–9321.

[3]   A.K. Holm, L.M. Blank, M. Oldiges, A. Schmid, C. Solem, P.R. Jensen, G.N. Vemuri, Metabolic and Transcriptional Response to Cofactor Perturbations in Escherichia coli, J. Biol. Chem. 285 (2010) 17498–17506.

[4]   U. Sauer, F. Canonaco, S. Heri, A. Perrenoud, E. Fischer, The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli*, J. Biol. Chem. 279 (2004) 6613–6619.

[5]   S. Singh, C. Brocker, V. Koppaka, Y. Chen, B.C. Jackson, A. Matsumoto, D.C. Thompson, V. Vasiliou, Aldehyde dehydrogenases in cellular responses to oxidative/electrophilicstress, Free Radic. Biol. Med. 56 (2013) 89–101.

[6]   S.A. Marchitti, C. Brocker, D. Stagos, V. Vasiliou, Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily, Expert Opin. Drug Metab. Toxicol. 4 (2008) 697–720.

[7]   P. Zimniak, Relationship of electrophilic stress to aging, Free Radic. Biol. Med. 51 (2011) 1087–1105.

[8] P.A. Grimsrud, H. Xie, T.J. Griffin, D.A. Bernlohr, Oxidative Stress and Covalent Modification of Protein with Bioactive Aldehydes*, J. Biol. Chem. 283 (2008) 21837–21841.

[9] K.S. Fritz, D.R. Petersen, An overview of the chemistry and biology of reactive aldehydes, Free Radic. Biol. Med. 59 (2013) 85–91.

[10] K.S. Fritz, D.R. Petersen, Exploring the Biology of Lipid Peroxidation-Derived Protein Carbonylation, Chem. Res. Toxicol. 24 (2011) 1411–1419.

[11] V. Vasiliou, A. Pappa, D.R. Petersen, Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism, Chem. Biol. Interact. 129 (2000) 1–19.

[12] G. Vassalli, Aldehyde Dehydrogenases: Not Just Markers, but Functional Regulators of Stem Cells, Stem Cells Int. 2019 (2019) 3904645.

[13] B. Jackson, C. Brocker, D.C. Thompson, W. Black, K. Vasiliou, D.W. Nebert, V. Vasiliou, Update on the aldehyde dehydrogenase gene (ALDH) superfamily, Hum. Genomics. 5 (2011) 283.

[14] K. Niederreither, P. Dollé, Retinoic acid in development: towards an integrated view, Nat. Rev. Genet. 9 (2008) 541–553.

[15] T.J.P. Sobreira, F. Marlétaz, M. Simões-Costa, D. Schechtman, A.C. Pereira, F. Brunet, S. Sweeney, A. Pani, J. Aronowicz, C.J. Lowe, B. Davidson, V. Laudet, M. Bronner, P.S.L. de Oliveira, M. Schubert, J. Xavier-Neto, Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans, Proc. Natl. Acad. Sci. 108 (2011)

226 LP – 231.

[16] G. Duester, Families of retinoid dehydrogenases regulating vitamin A function, Eur. J. Biochem. 267 (2000) 4315–4324.

[17] K. Housh, J.S. Jha, T. Haldar, S.B.M. Amin, T. Islam, A. Wallace, A. Gomina, X. Guo, C. Nel, J.W. Wyatt, K.S. Gates, Formation and repair of unavoidable, endogenous interstrand cross-links in cellular DNA, DNA Repair (Amst). 98 (2021) 103029.

[18] K. Uchida, Role of reactive aldehyde in cardiovascular diseases, Free Radic. Biol. Med. 28 (2000) 1685–1696.

[19] R. Sun, L. Fu, K. Liu, C. Tian, Y. Yang, K.A. Tallman, N.A. Porter, D.C. Liebler, J. Yang, Chemoproteomics Reveals Chemical Diversity and Dynamics of 4-Oxo-2-nonenal Modifications in Cells * , Mol. Cell. Proteomics. 16 (2017) 1789–1800.

[20] J. Hempel, H. Nicholas, R. Lindahl, Aldehyde dehydrogenases: Widespread structural and functional diversity within a shared framework, Protein Sci. 2 (1993) 1890–1900.

[21] K. Shortall, A. Djeghader, E. Magner, T. Soulimane, Insights into Aldehyde Dehydrogenase Enzymes: A Structural Perspective , Front. Mol. Biosci. . 8 (2021) 410.

[22] V. Vasiliou, D.W. Nebert, Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family, Hum. Genomics. 2 (2005) 138.

[23] N.E. Sládek, Human aldehyde dehydrogenases: Potential pathological,

pharmacological, and toxicological impact, J. Biochem. Mol. Toxicol. 17 (2003) 7–23.

[24] K. Sydow, A. Daiber, M. Oelze, Z. Chen, M. August, M. Wendt, V. Ullrich, A. Mülsch, E. Schulz, J.F. Keaney Jr, J.S. Stamler, T. Münzel, Central role of mitochondrial aldehyde dehydrogenase and reactive oxygen species in nitroglycerin tolerance and cross-tolerance, J. Clin. Invest. 113 (2004) 482–489.

[25] W. Black, Y. Chen, A. Matsumoto, D.C. Thompson, N. Lassen, A. Pappa, V. Vasiliou, Molecular mechanisms of ALDH3A1-mediated cellular protection against 4-hydroxy-2-nonenal, Free Radic. Biol. Med. 52 (2012) 1937–1944.

[26] J.-H. Jang, S. Bruse, Y. Liu, V. Duffy, C. Zhang, N. Oyamada, S. Randell, A. Matsumoto, D.C. Thompson, Y. Lin, V. Vasiliou, Y. Tesfaigzi, T. Nyunoya, Aldehyde dehydrogenase 3A1 protects airway epithelial cells from cigarette smoke-induced DNA damage and cytotoxicity, Free Radic. Biol. Med. 68 (2014) 80–86.

[27] T. Estey, J. Piatigorsky, N. Lassen, V. Vasiliou, ALDH3A1: a corneal crystallin with diverse functions, Exp. Eye Res. 84 (2007) 3–12.

[28] T. Estey, M. Cantore, P.A. Weston, J.F. Carpenter, J.M. Petrash, V. Vasiliou, Mechanisms Involved in the Protection of UV-induced Protein Inactivation by the Corneal Crystallin ALDH3A1 *, J. Biol. Chem. 282 (2007) 4382–4392.

[29] V. Vasiliou, A. Bairoch, K.F. Tipton, D.W. Nebert, Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping, Pharmacogenetics. 9 (1999) 421–434.

[30] Z.-J. Liu, Y.-J. Sun, J. Rose, Y.-J. Chung, C.-D. Hsiao, W.-R. Chang, I. Kuo, J. Perozich, R. Lindahl, J. Hempel, B.-C. Wang, The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, Nat. Struct. Biol. 4 (1997) 317–326.

[31] M.G. Rossmann, D. Moras, K.W. Olsen, Chemical and biological evolution of a nucleotide-binding protein, Nature. 250 (1974) 194–199.

[32] I. Hanukoglu, Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites, Biochem. Mol. Biol. Educ. 43 (2015) 206–209.

[33] W.G.J. Hol, P.T. van Duijnen, H.J.C. Berendsen, The α-helix dipole and the properties of proteins, Nature. 273 (1978) 443–446.

[34] J. Perozich, I. Kuo, B.C. Wang, J.S. Boesch, R. Lindahl, J. Hempel, Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase, Eur. J. Biochem. 267 (2000) 6197–6203.

[35] M.A. Sharkey, T.F. Oliveira, P.C. Engel, A.R. Khan, Structure of NADP+-dependent glutamate dehydrogenase from Escherichia coli – reflections on the basis of coenzyme specificity in the family of glutamate dehydrogenases, FEBS J. 280 (2013) 4681–4692.

[36] T.D. Hurley, C.G. Steinmetz, H. Weiner, Three-Dimensional Structure of Mitochondrial Aldehyde Dehydrogenase BT - Enzymology and Molecular Biology of Carbonyl Metabolism 7, in: H. Weiner, E. Maser, D.W. Crabb, R. Lindahl (Eds.), Springer US, Boston, MA, 1999: pp. 15–25.

[37] J.S. Rodríguez-Zavala, A. Allali-Hassani, H. Weiner, Characterization of E. coli tetrameric aldehyde dehydrogenases with atypical properties compared to other aldehyde dehydrogenases, Protein Sci. 15 (2006) 1387–1396.

[38] K. Hayes, M. Noor, A. Djeghader, P. Armshaw, T. Pembroke, S. Tofail, T. Soulimane, The quaternary structure of Thermus thermophilus aldehyde dehydrogenase is stabilized by an evolutionary distinct C-terminal arm extension, Sci. Rep. 8 (2018) 13327.

[39] T.A. Pemberton, D. Srivastava, N. Sanyal, M.T. Henzl, D.F. Becker, J.J. Tanner, Structural Studies of Yeast Δ1-Pyrroline-5-carboxylate Dehydrogenase (ALDH4A1): Active Site Flexibility and Oligomeric State, Biochemistry. 53 (2014) 1350–1359.

[40] E. Inagaki, N. Ohshima, H. Takahashi, C. Kuroishi, S. Yokoyama, T. Tahirov, Crystal Structure of Thermus thermophilus Δ1-Pyrroline-5-carboxylate Dehydrogenase, J. Mol. Biol. 362 (2006) 490–501.

[41] L.-K. Liu, J.J. Tanner, Crystal Structure of Aldehyde Dehydrogenase 16 Reveals Trans-Hierarchical Structural Similarity and a New Dimer, J. Mol. Biol. 431 (2019) 524–541.

[42] J. Farres, T.T.Y. Wang, S.J. Cunningham, H. Weiner, Investigation of the Active Site Cysteine Residue of Rat Liver Mitochondrial Aldehyde Dehydrogenase by Site-Directed Mutagenesis, Biochemistry. 34 (1995) 2592–2598.

[43] X. Wang, H. Weiner, Involvement of Glutamate 268 in the Active Site of Human Liver Mitochondrial (Class 2) Aldehyde Dehydrogenase As Probed by Site-Directed

Mutagenesis, Biochemistry. 34 (1995) 237–243.

[44]   R.I. Feldman, H. Weiner, Horse Liver Aldehyde Dehydrogenase: II. KINETICS AND MECHANISTIC IMPLICATIONS OF THE DEHYDROGENASE AND ESTERASE ACTIVITY, J. Biol. Chem. 247 (1972) 267–272.

[45]   J. Hempel, I. Kuo, J. Perozich, B.-C. Wang, R. Lindahl, H. Nicholas, Aldehyde dehydrogenase, Eur. J. Biochem. 268 (2001) 722–726.

[46]   P. Marcato, C.A. Dean, C.A. Giacomantonio, P.W.K. Lee, Aldehyde dehydrogenase: Its role as a cancer stem cell marker comes down to the specific isoform, Cell Cycle. 10 (2011) 1378–1384.

[47]   V. Vasiliou, D.C. Thompson, C. Smith, M. Fujita, Y. Chen, Aldehyde dehydrogenases: From eye crystallins to metabolic disease and cancer stem cells, Chem. Biol. Interact. 202 (2013) 2–10.

[48]   M. Al-Hajj, M.S. Wicha, A. Benito-Hernandez, S.J. Morrison, M.F. Clarke, Prospective identification of tumorigenic breast cancer cells, Proc. Natl. Acad. Sci. 100 (2003) 3983 LP – 3988.

[49]   D. Bonnet, J.E. Dick, Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell, Nat. Med. 3 (1997) 730–737.

[50]   C.A. O'Brien, A. Kreso, C.H.M. Jamieson, Cancer Stem Cells and Self-renewal, Clin. Cancer Res. 16 (2010) 3113 LP – 3120.

[51]   K. Pors, J.S. Moreb, Aldehyde dehydrogenases in cancer: an opportunity for biomarker and drug development?, Drug Discov. Today. 19 (2014) 1953–1963.

[52]   J.S. Moreb, Aldehyde Dehydrogenase as a Marker for Stem Cells, Curr. Stem Cell Res. Ther. 3 (2008) 237–246.

[53]   R.W. Storms, A.P. Trujillo, J.B. Springer, L. Shah, O.M. Colvin, S.M. Ludeman, C. Smith, Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity, Proc. Natl. Acad. Sci. 96 (1999) 9118 LP – 9123.

[54]   L. Zhou, D. Sheng, D. Wang, W. Ma, Q. Deng, L. Deng, S. Liu, Identification of cancer-type specific expression patterns for active aldehyde dehydrogenase (ALDH) isoforms in ALDEFLUOR assay, Cell Biol. Toxicol. 35 (2019) 161–177.

[55]   A.K. Croker, A.L. Allan, Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44+ human breast cancer cells, Breast Cancer Res. Treat. 133 (2012) 75–87.

[56]   L. Quattrini, E.L.M. Gelardi, G. Petrarolo, G. Colombo, D.M. Ferraris, F. Picarazzi, M. Rizzi, S. Garavaglia, C. La Motta, Progress in the Field of Aldehyde Dehydrogenase Inhibitors: Novel Imidazo[1,2-<i>a</i>]pyridines against the 1A Family, ACS Med. Chem. Lett. 11 (2020) 963–970.

[57]   D.J. Pearce, D. Taussig, C. Simpson, K. Allen, A.Z. Rohatiner, T.A. Lister, D. Bonnet, Characterization of Cells with a High Aldehyde Dehydrogenase Activity from Cord Blood and Acute Myeloid Leukemia Samples, Stem Cells. 23 (2005) 752–760.

[58]   A. Matsumoto, J. Arcaroli, Y. Chen, M. Gasparetto, V. Neumeister, D.C. Thompson, S. Singh, C. Smith, W. Messersmith, V. Vasiliou, Aldehyde dehydrogenase 1B1: a novel immunohistological marker for colorectal cancer, Br.

J. Cancer. 117 (2017) 1537–1543.

[59]   N. Lassen, J.B. Bateman, T. Estey, J.R. Kuszak, D.W. Nees, J. Piatigorsky, G. Duester, B.J. Day, J. Huang, L.M. Hines, V. Vasiliou, Multiple and Additive Functions of ALDH3A1 and ALDH1A1: CATARACT PHENOTYPE AND OCULAR OXIDATIVE DAMAGE IN Aldh3a1(-/-)/Aldh1a1(-/-) KNOCK-OUT MICE*, J. Biol. Chem. 282 (2007) 25668–25676.

[60]   R. Januchowski, K. Wojtowicz, K. Sterzyĺska, P. Sosiĺska, M. Andrzejewska, P. Zawierucha, M. Nowicki, M. Zabel, Inhibition of ALDH1A1 activity decreases expression of drug transporters and reduces chemotherapy resistance in ovarian cancer cell lines, Int. J. Biochem. Cell Biol. 78 (2016) 248–259.

[61]   V. Dupé, N. Matt, J.-M. Garnier, P. Chambon, M. Mark, N.B. Ghyselinck, A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment, Proc. Natl. Acad. Sci. 100 (2003) 14036 LP – 14041.

[62]   K.L. Deak, M.E. Dickerson, E. Linney, D.S. Enterline, T.M. George, E.C. Melvin, F.L. Graham, D.G. Siegel, P. Hammock, L. Mehltretter, A.G. Bassuk, J.A. Kessler, J.R. Gilbert, M.C. Speer, N.T.D.C. Group, Analysis of ALDH1A2, CYP26A1, CYP26B1, CRABP1, and CRABP2 in human neural tube defects suggests a possible association with alleles in ALDH1A2, Birth Defects Res. Part A Clin. Mol. Teratol. 73 (2005) 868–875.

[63]   M. Pavan, V.F. Ruiz, F.A. Silva, T.J. Sobreira, R.M. Cravo, M. Vasconcelos, L.P. Marques, S.M.F. Mesquita, J.E. Krieger, A.A.B. Lopes, P.S. Oliveira, A.C. Pereira,

J. Xavier-Neto, ALDH1A2 (RALDH2) genetic variation in human congenital heart disease, BMC Med. Genet. 10 (2009) 113.

[64]   C. Wan, Y. Shi, X. Zhao, W. Tang, M. Zhang, B. Ji, H. Zhu, Y. Xu, H. Li, G. Feng, L. He, Positive association between ALDH1A2 and schizophrenia in the Chinese population, Prog. Neuro-Psychopharmacology Biol. Psychiatry. 33 (2009) 1491–1495.

[65]   P. Marcato, C.A. Dean, D. Pan, R. Araslanova, M. Gillis, M. Joshi, L. Helyer, L. Pan, A. Leidal, S. Gujar, C.A. Giacomantonio, P.W.K. Lee, Aldehyde Dehydrogenase Activity of Breast Cancer Stem Cells Is Primarily Due To Isoform ALDH1A3 and Its Expression Is Predictive of Metastasis, Stem Cells. 29 (2011) 32–45.

[66]   L.L.N. Husemoen, M. Fenger, N. Friedrich, J.S. Tolstrup, S. Beenfeldt Fredriksen, A. Linneberg, The Association of ADH and ALDH Gene Variants With Alcohol Drinking Habits and Cardiovascular Disease Risk Factors, Alcohol. Clin. Exp. Res. 32 (2008) 1984–1991.

[67]   O.M. Palo, P. Soronen, K. Silander, T. Varilo, K. Tuononen, T. Kieseppä, T. Partonen, J. Lönnqvist, T. Paunio, L. Peltonen, Identification of susceptibility loci at 7q31 and 9p13 for bipolar disorder in an isolated population, Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 153B (2010) 723–735.

[68]   J. Palmfeldt, S. Vang, V. Stenbroen, E. Pavlou, M. Baycheva, G. Buchal, A.A. Monavari, P. Augoustides-Savvopoulou, H. Mandel, N. Gregersen, Proteomics Reveals that Redox Regulation Is Disrupted in Patients with Ethylmalonic

Encephalopathy, J. Proteome Res. 10 (2011) 2389–2396.

[69]   S. Singh, Y. Chen, A. Matsumoto, D.J. Orlicky, H. Dong, D.C. Thompson, V. Vasiliou, ALDH1B1 links alcohol consumption and diabetes, Biochem. Biophys. Res. Commun. 463 (2015) 768–773.

[70]   S.A. Krupenko, N.I. Krupenko, Loss of ALDH1L1 folate enzyme confers a selective metabolic advantage for tumor progression, Chem. Biol. Interact. 302 (2019) 149–155.

[71]   Y.-C. Chao, S.-R. Liou, Y.-Y. Chung, H.-S. Tang, C.-T. Hsu, T.-K. Li, S.-J. Yin, Polymorphism of alcohol and aldehyde dehydrogenase genes and alcoholic cirrhosis in chinese patients, Hepatology. 19 (1994) 360–366.

[72]   P. Hui, T. Nakayama, A. Morita, N. Sato, M. Hishiki, K. Saito, Y. Yoshikawa, M. Tamura, I. Sato, T. Takahashi, M. Soma, Y. Izumi, Y. Ozawa, Z. Cheng, Common Single Nucleotide Polymorphisms in Japanese Patients with Essential Hypertension: Aldehyde Dehydrogenase 2 Gene as a Risk Factor Independent of Alcohol Consumption, Hypertens. Res. 30 (2007) 585–592.

[73]   S.A. Jo, E.-K. Kim, M.H. Park, C. Han, H.-Y. Park, Y. Jang, B.J. Song, I. Jo, A Glu487Lys polymorphism in the gene for mitochondrial aldehyde dehydrogenase 2 is associated with myocardial infarction in elderly Korean men, Clin. Chim. Acta. 382 (2007) 43–47.

[74]   J.-S. Lee, S.H. Kim, S. Lee, J.H. Kang, S.-H. Lee, J.-H. Cheong, S.-Y. Kim, Gastric cancer depends on aldehyde dehydrogenase 3A1 for fatty acid oxidation, Sci. Rep. 9 (2019) 16313.

[75] W.B. Rizzo, G. Carney, Sjögren-Larsson syndrome: Diversity of mutations and polymorphisms in the fatty aldehyde dehydrogenase gene (ALDH3A2), Hum. Mutat. 26 (2005) 1–10.

[76] M.T. Geraghty, D. Vaughn, A.J. Nicholson, W.-W. Lin, G. Jimenez-Sanchez, C. Obie, P. Flynn, D. Valle, C.A. Hu, Mutations in the Δ1-Pyrroline 5-Carboxylate Dehydrogenase Gene Cause Type II Hyperprolinemia, Hum. Mol. Genet. 7 (1998) 1411–1415.

[77] S.W. Sauer, S. Kölker, G.F. Hoffmann, H.J. ten Brink, C. Jakobs, K.M. Gibson, J.G. Okun, Enzymatic and metabolic evidence for a region specific mitochondrial dysfunction in brains of murine succinic semialdehyde dehydrogenase deficiency (Aldh5a1−/− mice), Neurochem. Int. 50 (2007) 653–659.

[78] J.O. Sass, M. Walter, J.P.H. Shield, A.M. Atherton, U. Garg, D. Scott, C.G. Woods, L.D. Smith, 3-Hydroxyisobutyrate aciduria and mutations in the ALDH6A1 gene coding for methylmalonate semialdehyde dehydrogenase, J. Inherit. Metab. Dis. 35 (2012) 437–442.

[79] J.L. Marcadier, A.M. Smith, D. Pohl, J. Schwartzentruber, O.Y. Al-Dirbashi, J. Majewski, S. Ferdinandusse, R.J.A. Wanders, D.E. Bulman, K.M. Boycott, P. Chakraborty, M.T. Geraghty, F.C. Consortium, Mutations in ALDH6A1encoding methylmalonate semialdehyde dehydrogenase are associated with dysmyelination and transient methylmalonic aciduria, Orphanet J. Rare Dis. 8 (2013) 98.

[80] K.L. Chambliss, R.G.F. Gray, G. Rylance, R.J. Pollitt, K.M. Gibson, Molecular characterization of methylmalonate semialdehyde dehydrogenase deficiency, J.

Inherit. Metab. Dis. 23 (2000) 497–504.

[81]   Y. Guo, L.-J. Tan, S.-F. Lei, T.-L. Yang, X.-D. Chen, F. Zhang, Y. Chen, F. Pan, H.
       Yan, X. Liu, Q. Tian, Z.-X. Zhang, Q. Zhou, C. Qiu, S.-S. Dong, X.-H. Xu, Y.-F.
       Guo, X.-Z. Zhu, S.-L. Liu, X.-L. Wang, X. Li, Y. Luo, L.-S. Zhang, M. Li, J.-T.
       Wang, T. Wen, B. Drees, J. Hamilton, C.J. Papasian, R.R. Recker, X.-P. Song, J.
       Cheng, H.-W. Deng, Genome-wide association study identifies ALDH7A1 as a
       novel susceptibility gene for osteoporosis, PLoS Genet. 6 (2010) e1000806–
       e1000806.

[82]   G. Scharer, C. Brocker, V. Vasiliou, G. Creadon-Swindell, R.C. Gallagher, E.
       Spector, J.L.K. Van Hove, The genotypic and phenotypic spectrum of pyridoxine-
       dependent epilepsy due to mutations in ALDH7A1, J. Inherit. Metab. Dis. 33 (2010)
       571–581.

[83]   P.B. Mills, E. Struys, C. Jakobs, B. Plecko, P. Baxter, M. Baumgartner, M.A.
       Willemsen, H. Omran, U. Tacke, B. Uhlenberg, B. Weschke, P.T. Clayton,
       Mutations in antiquitin in individuals with pyridoxine-dependent seizures, Nat.
       Med. 12 (2006) 307.

[84]   P.B. Mills, E.J. Footitt, K.A. Mills, K. Tuschl, S. Aylett, S. Varadkar, C.
       Hemingway, N. Marlow, J. Rennie, P. Baxter, O. Dulac, R. Nabbout, W.J. Craigen,
       B. Schmitt, F. Feillet, E. Christensen, P. De Lonlay, M.G. Pike, M.I. Hughes, E.A.
       Struys, C. Jakobs, S.M. Zuberi, P.T. Clayton, Genotypic and phenotypic spectrum
       of pyridoxine-dependent epilepsy (ALDH7A1 deficiency), Brain. 133 (2010) 2148–
       2159.

[85]   P. Sulem, D.F. Gudbjartsson, G.B. Walters, H.T. Helgadottir, A. Helgason, S.A. Gudjonsson, C. Zanon, S. Besenbacher, G. Bjornsdottir, O.T. Magnusson, G. Magnusson, E. Hjartarson, J. Saemundsdottir, A. Gylfason, A. Jonasdottir, H. Holm, A. Karason, T. Rafnar, H. Stefansson, O.A. Andreassen, J.H. Pedersen, A.I. Pack, M.C.H. de Visser, L.A. Kiemeney, A.J. Geirsson, G.I. Eyjolfsson, I. Olafsson, A. Kong, G. Masson, H. Jonsson, U. Thorsteinsdottir, I. Jonsdottir, K. Stefansson, Identification of low-frequency variants associated with gout and serum uric acid levels, Nat. Genet. 43 (2011) 1127–1130.

[86]   V. Vasiliou, M. Sandoval, D.S. Backos, B.C. Jackson, Y. Chen, P. Reigan, M.A. Lanaspa, R.J. Johnson, V. Koppaka, D.C. Thompson, 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. 202 (2013) 22–31.

Chapter Two: Importance of the C-terminus of ALDH7A1 for Oligomerization and Catalytic Activity

2.1 Introduction

The lysine catabolic pathway converts L-lysine to α-aminoadipate (AA). Aldehyde dehydrogenase 7A1 (ALDH7A1, aka antiquitin) catalyzes the final step in this pathway – the NAD^+-dependent oxidation of α-aminoadipate semialdehyde (AASAL) to α-aminoadipate (AA) (Figure 2.1A).
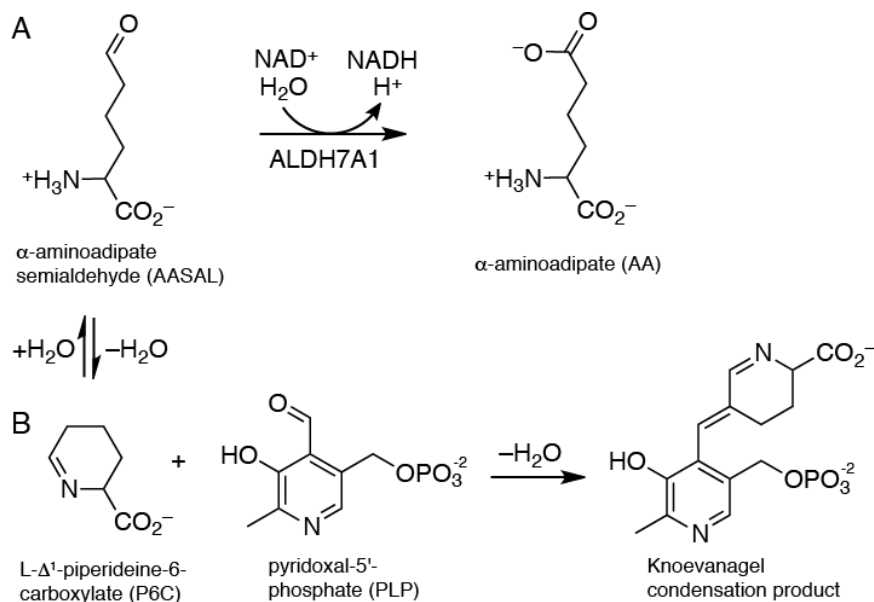
Figure 2.1. Chemical reactions relevant to ALDH7A1 and PDE. (A) The reaction catalyzed by ALDH7A1. (B) The reaction between P6C and PLP, which results in covalent inactivation of PLP. P6C forms a nonenzymatic equilibrium with AASAL.

Mutations in the *ALDH7A1* gene can cause pyridoxine-dependent epilepsy (PDE),[1,2] a rare autosomal recessive disorder that typically presents with seizures in the first days of life. Compromised ALDH7A1 activity leads to increased levels of AASAL and Δ^1-

piperideine-6-carboxylic acid (P6C), the cyclized form of AASAL (Figure 2.1). P6C reacts with ubiquitous cofactor, pyridoxal 5´-phosphate (PLP), and the resulting adduct is incapable as an enzyme cofactor (Figure 2.1B). PDE is evidently a consequence of the resulting inadequate levels of pyridoxal 5´-phosphate (PLP).[3] Although seizures can be controlled with pharmacologic doses of pyridoxine (vitamin B6) and a lysine-restricted diet, 75% of PDE patients nevertheless suffer intellectual developmental disability.[1] Dozens of ALDH7A1 mutations have been linked to PDE. At present, the majority of these are uncharacterized, an exception being a group of six mutations recently shown to abolish catalytic activity by altering oligomerization behavior.[4] Knowledge of the structural- and kinetic consequences of PDE-causing mutations may provide insight into the molecular basis of the disorder. An improved understanding of the disease pathology could ultimately allow for the prediction of symptom-severity and development of patient-specific medical treatments.

Although the general catalytic mechanism of ALDHs is well known,[5] recent crystallographic studies of ALDH7A1 led to the novel suggestion that the C-terminus (residues 500-511) is a mobile structural element that plays a role in catalysis.[6] Like other ALDH superfamily enzymes, ALDH7A1 has three domains: an N-terminal Rossmann fold NAD$^+$-binding domain, a catalytic domain, and an oligomerization domain (Figure 2.2A). The C-terminus is part of the oligomerization domain, which mediates domain-swapped dimerization (Figure 2.2B) and tetramerization (Figure 2.2C). The crystal structures revealed opened and closed active sites that differ in the conformation of C-terminus. In the closed state, observed when the product AA is bound, the crook of the C-terminus of one protomer of the dimer appears to stabilize the active site of the opposite protomer

(Figure 2.2D). In particular, Gln506 forms both a direct hydrogen bond with the conserved aldehyde anchor loop and a water-mediated hydrogen bond with the α-carboxylate of AA. The methyl group of Ala505 also packs tightly against the anchor loop. The open state is substantially different. In the absence of AA, the C-terminus rotates outward by 16 Å, creating a solvent-exposed active site (Figure 2.2B).[6] The interpretation of these structures is that the active site is open in the resting state of the enzyme and closed in the E-AASAL-NAD$^+$ Michaelis complex. A further inference is that AASAL binding facilitates the conformational changes that close the active site for catalysis.
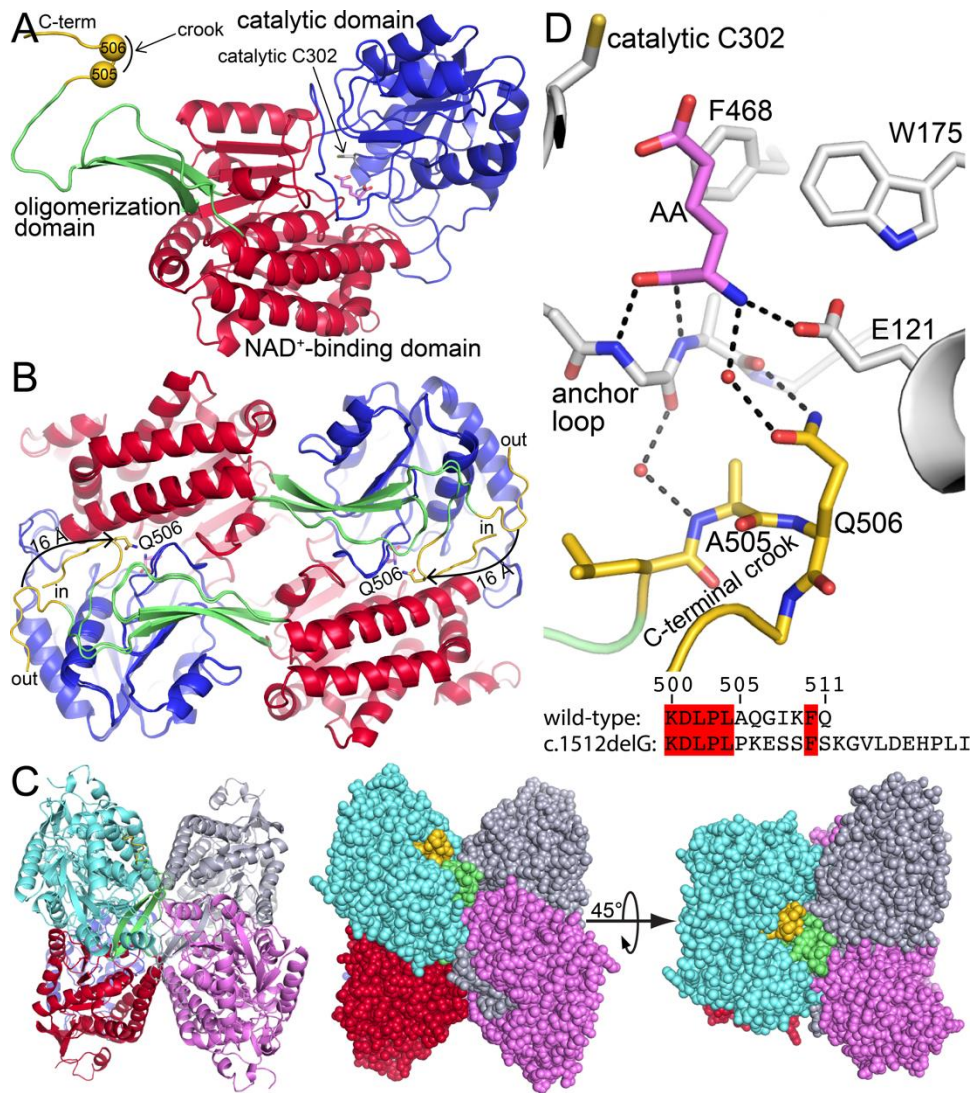
Figure 2.2. Structural and genetic contexts of the mutations studied. (A) A protomer of ALDH7A1 with the mutated residues Ala505 and Gln506 marked with spheres (PDB code 4ZUL). The three domains are colored red (NAD⁺-binding), blue (catalytic), and green (oligomerization). The residues deleted in in the truncation mutant Δ504-511 are colored gold. The product AA is in pink. (B) Superposition of the open and closed ALDH7A1 dimers (PDB codes 4ZUK (open) and 4ZUL (closed)). The domains are colored as in panel A. The arrows indicate the 16-Å motion of the C-terminus from the open (out) to the closed (in) state. (C) The dimer-of-dimers tetramer of ALDH7A1. One protomer is colored according to domains as in panel A. The other three protomers have individual colors (aquamarine, gray, and violet). (D) Close-up of the quaternary structural interactions that stabilize the aldehyde-binding site in the closed state (AA complex, PDB code 4ZUL). One protomer of the dimer is colored white with bound AA colored pink. The C-terminus of the other protomer is colored gold. Note the interactions of Ala505 and Gln506 with the active site. This panel also includes a sequence alignment of the C-termini of wild-type ALDH7A1 and the c.1512delG deletion mutant, which has been implicated in PDE.

The discovery of the C-terminus as a mobile part of the active site is potentially relevant for understanding the c.1512delG frameshift mutation [3] associated with PDE. Deletion of G1512 from exon 18 mutates six of the last seven amino acids of the wild-type polypeptide and extends the C-terminus by 10 residues (Figure 2.2D). Notably, Ala505 and Gln506 are mutated to Pro and Lys, respectively.

Herein we test the importance of the C-terminus of human ALDH7A1 using site-directed mutagenesis. Three mutations inspired by c.1512delG were generated: the single mutations A505P and Q506K, and the double mutation A505P/Q506K. The C-terminal truncation mutant Δ504-511, which lacks residues 504 – 511, was also generated. The variants were analyzed with steady-state kinetics assays and analytical ultracentrifugation. The truncation mutant was also analyzed with small-angle X-ray scattering (SAXS). The mutant enzymes exhibit a common phenotype characterized by a very high K_m for AASAL and perturbed self-association equilibrium. Overall, our results show that an intact C-terminus is essential for maximal catalytic efficiency and proper oligomerization and suggest that the c.1512delG variant is functionally compromised due to a defect in tetramer formation.

2.2 Results

Steady-State Kinetic Analysis of Wild-Type ALDH7A1 and C-terminal Variants.

Wild-type ALDH7A1 exhibited Michaelis-Menten behavior (Figure 2.3A), yielding a k_{cat} of 0.30 s^{-1}, a K_m for AASAL of 64 µM, and a catalytic efficiency (k_{cat}/K_m) of 4700 M^{-1}s^{-1} (Table 2.1). The k_{cat}/K_m is within a factor of three of values reported by other groups for human ALDH7A1. [7,8]

When assayed under the same conditions as the wild-type enzyme, the A505P variant likewise exhibited Michaelis-Menten behavior (Figure 2.3A); however, the K_m is 10 times higher, and k_{cat} is 3 times higher (Table 2.1). The resulting catalytic efficiency of A505P, 1600 $M^{-1}s^{-1}$, is approximately 3 times lower than wild-type ALDH7A1. Thus, A505P is functional, albeit with compromised efficiency.

Replacement of Gln506 with Lys has a more profound impact on catalytic activity. Unlike A505P, at experimentally accessible substrate concentrations, the reaction velocities of Q506K and the double mutant A505P/Q506K exhibit a linear dependence on AASAL concentration, indicating that the K_m is very large compared to the maximum substrate concentration used (1500 µM) (Figures 3B and 3C). For these enzymes, the catalytic efficiency was estimated from linear regression, resulting in k_{cat}/K_m ~60 $M^{-1}s^{-1}$. Clearly, the presence of Lys at position 506 of ALDH7A1 is deleterious for catalytic function.

Point mutations in the C-terminal crook of ALDH7A1 lower catalytic efficiency, suggesting that the C-terminus itself may be vital for catalysis. To examine the contribution of the C-terminus to catalytic activity, a truncated protein missing the last 8 residues of ALDH7A1 (Δ504-511) was generated. The kinetic phenotype of Δ504-511 resembles the phenotype of the point mutants in that Δ504-511 could not be saturated with AASAL (Figure 2.3D). The estimated k_{cat}/K_m, ~30 $M^{-1}s^{-1}$, is less than 1% of the wild-type value, indicating that an intact C-terminus is essential for proper catalytic activity.
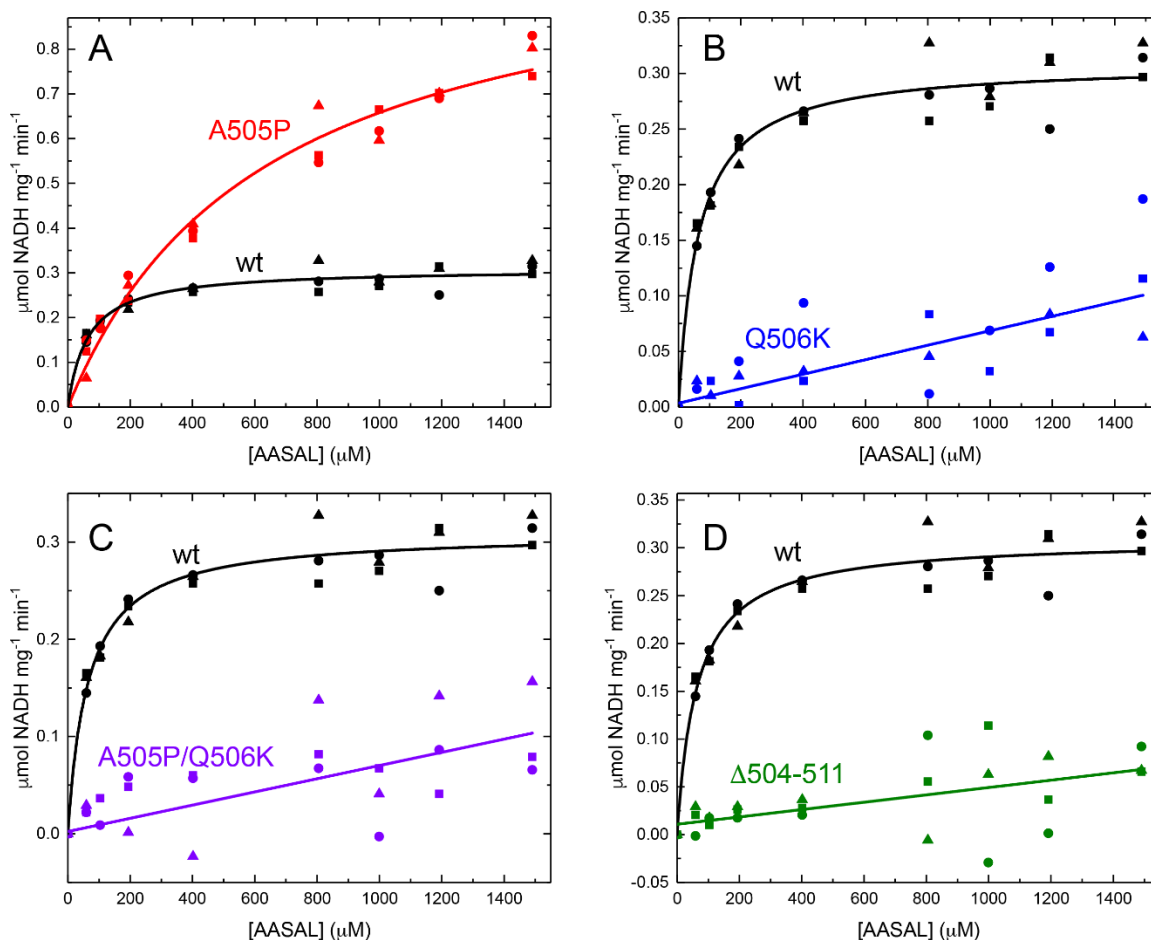
Figure 2.3. Steady-state kinetics of wild-type and mutant ALDH7A1 using AASAL as the variable substrate at a fixed NAD⁺ concentration of 2.5 mM (26°C, pH 8.0). (A) Initial velocity data for wild-type ALDH7A1 (black) and A505P (red). The curves are fits to the Michaelis-Menten equation. (B) Initial velocity data for Q506K (blue). The line was calculated from linear regression. The data for wild-type ALDH7A1 are shown for comparison. (C) Initial velocity data for A505P/Q506K (purple). The line was calculated from linear regression. The data for wild-type ALDH7A1 are shown for comparison. (D) Initial velocity data for Δ504-511 (green). The line was calculated from linear regression. The data for wild-type ALDH7A1 are shown for comparison.

Table 2.1. Kinetic Parameters for Wild-Type and Mutant ALDH7A1 Enzymes[a]

Enzyme	k_{cat} (s⁻¹)	K_m (μM)	k_{cat}/K_m (M⁻¹s⁻¹)
ALDH7A1	0.30 ± 0.01	64 ± 8	4700 ± 600
A505P	1.0 ± 0.1	630 ± 90	1600 ± 300
Q506K	N/D	N/D	60 ± 10[b]
A505P/Q506K	N/D	N/D	60 ± 20[b]
Δ504-511	N/D	N/D	35 ± 5[b]

[a]Assays were performed at 26°C and pH 8.0 with NAD⁺ fixed at 2.5 mM.
[b]Estimated from linear regression.

35

Mutations in the C-terminus of ALDH7A1 Disrupt Tetramer Formation.

Because the C-terminus contributes to the oligomerization domain, we examined the oligomeric states of the variants in solution. Sedimentation equilibrium experiments were performed on the wild-type enzyme and the four variants under identical conditions (Figure 2.4).
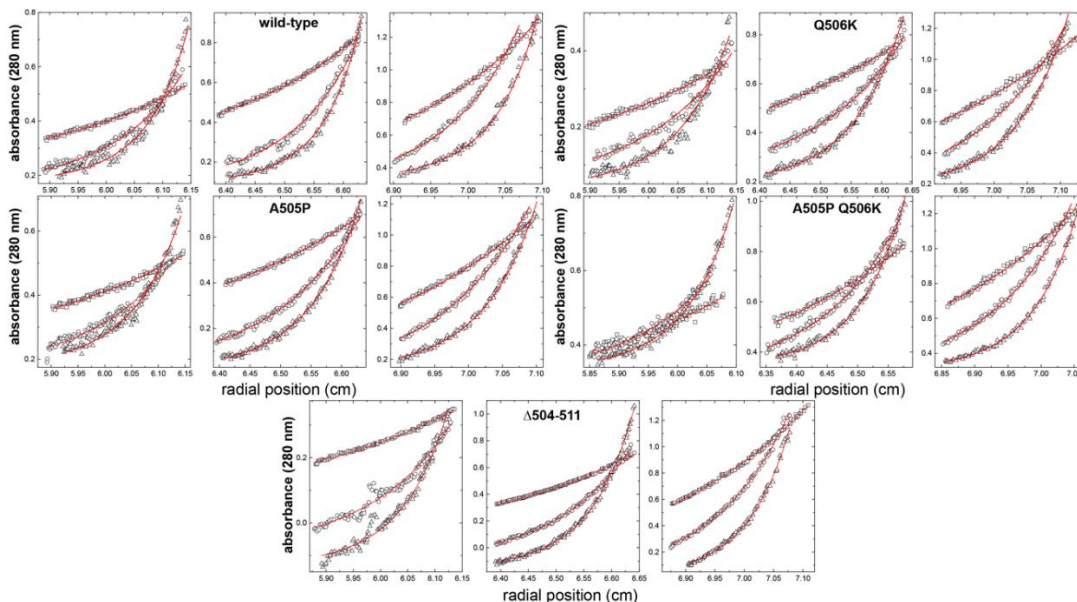
Figure 2.4. Sedimentation equilibrium analysis of wild-type and mutant ALDH7A1. For each protein, the three panels show three different protein concentrations: 0.2 mg mL^{-1} (left), 0.4 mg mL^{-1} (middle), and 0.8 mg mL^{-1} (right). Within each graph, the three data sets correspond to centrifugation speeds of 6000 (squares), 9000 (circles), and 12,000 (triangles) rpm. The red curves for wild-type ALDH7A1 represent a global fit of the data to a previously described dimer-tetramer oligomerization equilibrium model.[4] The red curves for the mutant enzymes are from global fits to a single-species model.

Sedimentation-equilibrium experiments performed at 0.2 – 0.8 mg mL^{-1} on wild-type ALDH7A1 yielded an average M_r of 141 kDa (Table 2.2). For reference, the theoretical dimer M_r is 111 kDa. The experimental M_r of 141 kDa is consistent with the presence of a dimer-tetramer equilibrium, as previously reported for ALDH7A1.[4] Introduction of a model to account for the presence of a dimer-tetramer equilibrium improved the fit to the experimental data (Figure 2.4) and revealed a K_{2-4} self-association constant of 59,000 ± 6,000 M^{-1}, corresponding to a K_d of 17 µM, in excellent agreement with our previous study

(K_d of 16 μM).[4]

Table 2.2. Oligomeric State Parameters from Sedimentation Equilibrium

	Enzyme Concentration (mg mL^{-1})	Average M_r (kDa)[a]	$K_{2\text{-}4}$ (M^{-1})[b]
Wild type	0.2 – 0.8	141 ± 2	5.9 ± 0.6 x 10^4
A505P	0.2 – 0.8	105 ± 1	N/D[c]
Q506K	0.2 – 0.8	108 ± 1	N/D[c]
A505P/Q506K	0.2 – 0.8	122 ± 1	N/D[c]
Δ504-511	0.2 – 0.8	112 ± 1	N/D[c]
Δ504-511	2 – 8	157.6 ± 0.2	2.64 ± 0.03 x 10^4

[a]Values from single-species fit of low concentration sedimentation equilibrium data.
[b]Association constant for the formation of a tetramer from two dimers.
[c]Inclusion of the parameter $K_{2\text{-}4}$ in the model did not improve the fit.
.

Comparable sedimentation-equilibrium experiments were also performed on A505P, Q506K, A505P/Q506K, and Δ504-511. Interestingly, in each case, the average M_r (ranging from 105 to 122 kDa) was significantly lower than the value of 141 kDa obtained for the wild-type protein (Table 2.2). The quality of the least-squares fits was not improved by incorporating the dimer-tetramer equilibrium into the model (Figure 2.4). Evidently, A505P, Q506K, A505P/Q506K, and Δ504-511 are predominantly dimeric under the assayed conditions, suggesting that these mutations impair tetramerization.

To determine whether the oligomeric states of the three point mutant variant enzymes were concentration-dependent, sedimentation-velocity studies were performed at high concentration (4.5 mg mL^{-1}; ~40 μM dimer). Wild-type ALDH7A1 displays a major peak at sedimentation coefficient of 6.4S (Figure 2.5A), which corresponds to M_r of 212 kDa (Figure 2.5B). For reference, the theoretical M_r of the ALDH7A1 tetramer is 222 kDa. Thus, ALDH7A1 is primarily tetrameric at this concentration, in agreement with previous

studies.[4] Importantly, each of the C-terminal crook mutant variants – A505P, Q506K, and A505P/Q506K – displays a similar distribution with a major peak near 6.4S, consistent with the M_r of a tetrameric enzyme (Figure 2.5). Overall, these data suggest that although point mutations in the C-terminal crook of ALDH7A1 weaken the dimer-dimer interaction, they do not abolish the capacity for tetramer formation.
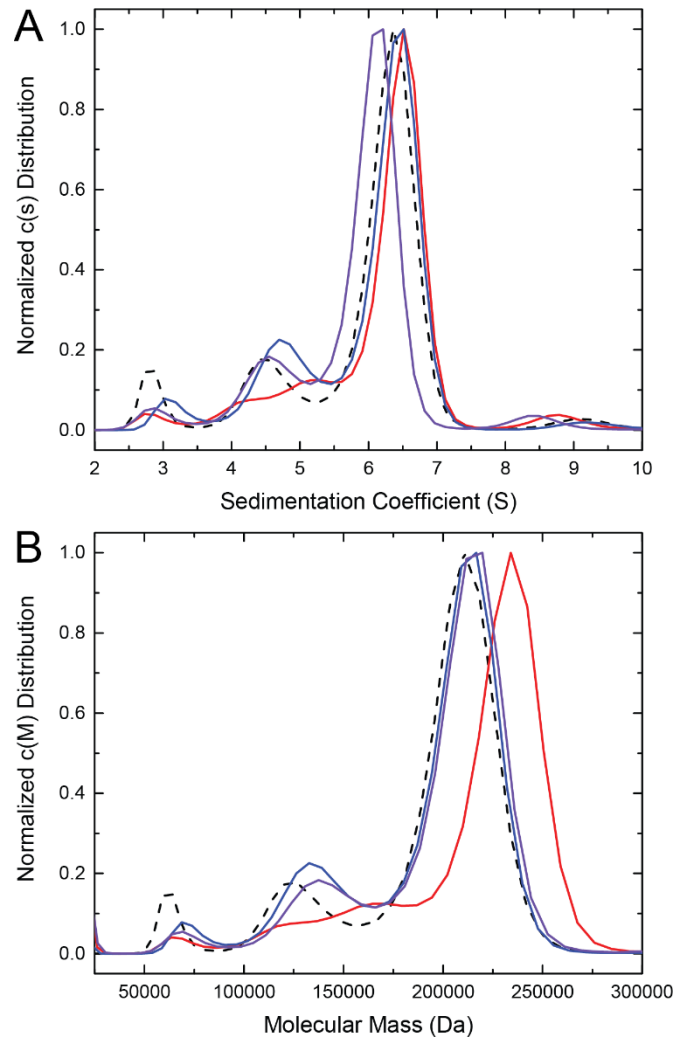
Figure 2.5. Sedimentation velocity analysis of ALDH7A1 wild-type and C-terminal point mutants. (A) The apparent sedimentation coefficient distributions for ALDH7A1 wild-type (dashed black), A505P (red), Q506K (blue), or A505P/Q506K (purple). (B) The apparent M_r distributions for ALDH7A1 wild-type (dashed black), A505P (red), Q506K (blue), or A505P/Q506K (purple). All enzymes were used at 4.5 mg mL^{-1} (40 µM using dimer M_r).

Contribution of the C-terminus of ALDH7A1 to Oligomeric State.

The Δ504-511 variant of ALDH7A1 displayed the lowest catalytic activity of the proteins studied (Table 2.1). Like the C-terminal-crook point variants, Δ504-511 is apparently predominantly dimeric at low concentrations (0.2 – 0.8 mg mL^{-1}; ~2 – 7 µM, dimer M_r) (Table 2.2). Therefore, a series of sedimentation-velocity experiments were conducted on the Δ504-511 variant to investigate any perturbations of oligomeric state at higher concentration. At the lowest concentration examined (8 µM dimer), the molecular weight distribution displayed a single major peak at ~105 kDa, consistent with a dimer (theoretical dimer M_r = 109 kDa; Figure 2.6, top left panel). With increased protein concentration, a second feature appears intermediate between the dimer- and tetramer masses. At the highest protein concentration (93 µM dimer), the M_r distribution includes two broad, overlapping peaks at a range of apparent M_r between ~100 kDa and ~200 kDa. Although no explicit feature is present at the tetramer mass, the peak at ~154-175 kDa could reflect a facile dimer-tetramer equilibrium. Alternatively, this peak may be interpreted as formation of a trimer (M_r ~164 kDa), suggesting a mixture of dimer and trimer in solution at high concentration. Importantly, there is no peak present consistent with the apparent M_r of a tetramer. These results are consistent with a dynamic mixture of multiple oligomeric species in solution and suggest that removal of the eight C-terminal residues severely impairs tetramer formation.
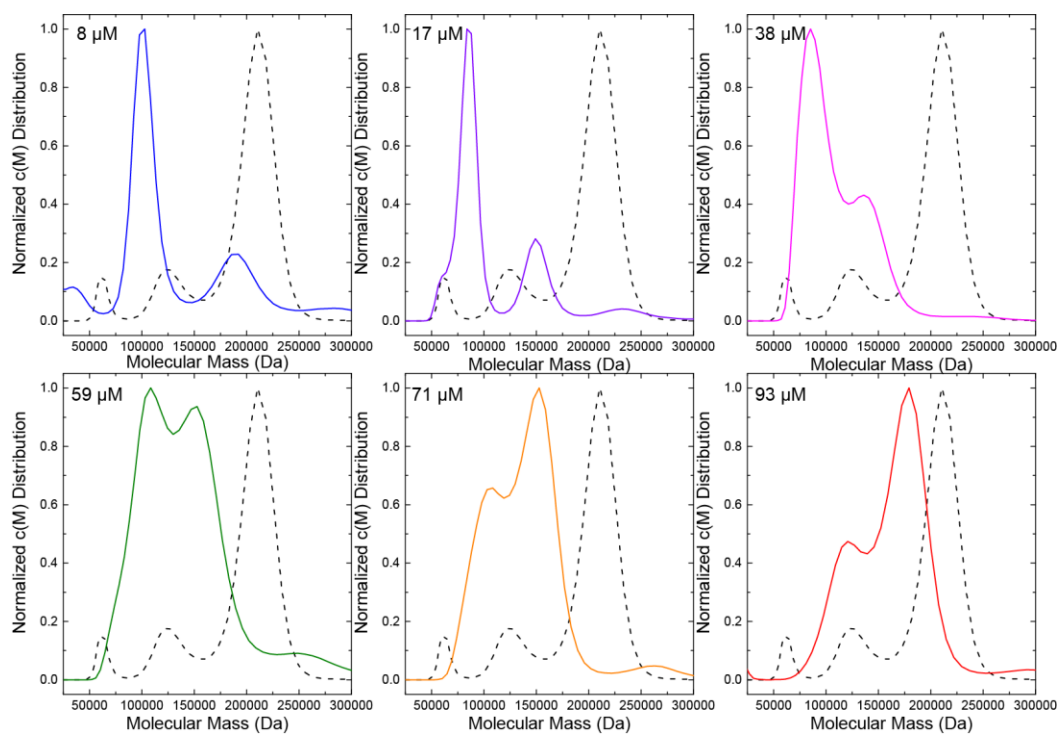
Figure 2.6. Concentration dependence of ALDH7A1 Δ504-511 self-association. Sedimentation velocity analysis was conducted on Δ504-511 at the indicated concentrations (8 – 93 μM). Concentrations were determined using the dimer M_r (109 kDa). Each panel shows the distribution of apparent molecular masses in solution at the indicated concentration. For reference and ease in identifying the peak corresponding to the ALDH7A1 tetramer, the black, dashed curve shows the distribution of molecular masses of wild-type ALDH7A1 at 40 μM.

To complement the high concentration sedimentation-velocity experiments, a sedimentation equilibrium experiment was also performed at high concentration (2 – 8 mg mL^{-1}; 18 – 73 μM dimer). The experimental data and fits to the experimental data are shown in Figure 2.7. A single-species fit to the experimental data revealed the average M_r in solution of 157.6 kDa. Because this M_r is between that of a dimer and a tetramer, a dimer-tetramer self-association model was applied to the data. Fits to the data were improved with the model, which returned a $K_{2\text{-}4}$ of 26,400 ± 300 M^{-1}, corresponding to a K_d of 38 μM. Likewise, the average in-solution M_r of 157.6 kDa is within 4% of the M_r of a Δ504-511 trimer. These results are consistent with the high-concentration sedimentation-velocity

40

experiment (Figure 2.6, top panel), which could not differentiate between a facile dimer-tetramer equilibrium or the introduction of a new oligomeric state (trimer).
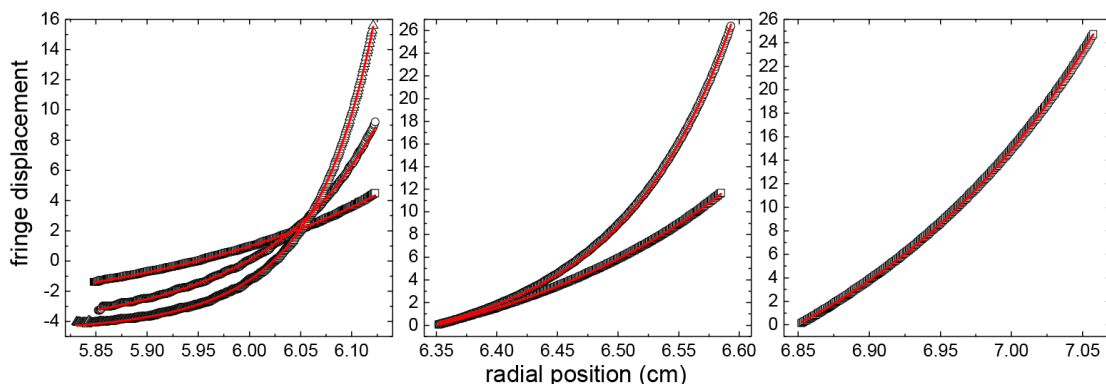
Figure 2.7. High-concentration sedimentation equilibrium analysis of ALDH7A1 Δ504-511. Sedimentation equilibrium analysis was conducted on ALDH7A1 Δ504-511 using Raleigh Interference optics at 2 mg mL^{-1} (left panel), 4 mg mL^{-1} (center panel), and 8 mg mL^{-1} (right panel) at three rotor speeds: 6000 (squares), 9000 (circles), and 12,000 (triangles) rpm. The red curves represent global fits to a dimer-tetramer equilibrium model. We note that the center and right panels are missing data from faster rotor speeds due to the limits of detection at the apparent concentrations within the sedimentation cell. Data were graphed and analyzed in Origin 2017.

SAXS analysis was also performed to further investigate the in-solution oligomeric structure of ALDH7A1 Δ504-511. SAXS data collected at four protein concentrations (1 – 6 mg mL^{-1}; 9 – 55 μM using dimer M_r) are shown in Figure 2.8. Qualitative changes in the SAXS curve are evident with increasing protein concentration. In particular, a shoulder appears in the region of $q = 0.10 – 0.15$ Å^{-1} with increasing protein concentration. Also, the radius of gyration (R_g) from Guinier analysis increases from 34.5 Å at the lowest concentration to 37.2 Å at the highest concentration (Figures 2.8A,B). Likewise, the real space R_g calculated from the distance distribution function increases from 34.5 to 37.2 Å (Figure 2.8C). For reference, the calculated R_g of the ALDH7A1 domain-swapped dimer is 30.4 Å, and the previously reported experimental SAXS R_g for the ALDH7A1 tetramer is 37.7 Å. [6] Further, calculation of the in-solution M_r from the experimental SAXS curves using the MoW2 server [9] revealed that the apparent M_r is within 10% of a dimer at the lowest concentration, and very close to the trimer at the two highest concentrations (Table

2.3). All of these results are consistent with concentration-dependent self-association, in agreement with the sedimentation analysis.

Table 2.3. Calculated Molecular Masses from Experimental SAXS Data[a]

Concentration (mg mL^{-1})	M_r	Predicted Oligomeric State	Discrepancy from Theoretical Mass (%)[b]
1	120	dimer	9.7
2	146	trimer	11.0
4	158	trimer	3.7
6	163	trimer	0.7

[a]Calculated using the MoW2 Server [9]
[b]The theoretical M_r of the Δ504-511 monomer is 54.7 kDa.

The experimental SAXS curves were compared to theoretical curves calculated from models of Δ504-511 using FoXS and MultiFoXS. [10,11] The models were made by deleting residues 504-511 from the wild-type structure (Figure 2.2), and no attempt was made to model any structural perturbations caused by the truncation. The FoXS goodness-of-fit parameter (χ) for single-body and two-body ensemble fits are listed in Table S2.3 of the Supporting Information.

We first analyzed the Δ504-511 SAXS data using a dimer-tetramer ensemble, as we had done previously for wild-type ALDH7A1.[6] Neither the dimer nor the tetramer alone provided a satisfactory fit to the Δ504-511 SAXS data, as evidenced by χ values of 2 – 16 (Table S2.3). Using a dimer-tetramer equilibrium model improved the fits ($\chi = 0.83 – 5.0$); however, the experimental and theoretical SAXS curves deviated significantly at $q > 0.05$ Å^{-1} (Figure 2.8A). The discrepancy is particularly noticeable for the two higher concentration SAXS samples. These results suggest that the standard dimer-tetramer equilibrium model may be insufficient to describe the in-solution behavior of Δ504-511.

Because the dimer-tetramer model was unsatisfactory and the sedimentation velocity data suggested the possibility of dimers and trimers in solution at high concentration (Figure 2.6), we incorporated a trimer into the SAXS analysis. The trimer model was generated by removing one protomer from the crystallographic tetramer and deleting residues 504-511. Interestingly, for the two highest concentration samples, the trimer alone gave the best fit to the SAXS data (Figure 2.8B, Table S2.3); this result is consistent with the SAXS M_r calculations, which indicated an average M_r very close to that of a trimer (Table 2.3). For the two lowest concentration samples, the trimer also yielded the best one-body fit; however, the use of a dimer-trimer ensemble improved the fit (Figure 2.8B, Table S2.3). Likewise, this result is consistent with the SAXS-derived estimate of M_r, which indicated an average M_r between those of the dimer and trimer (Table 2.3).
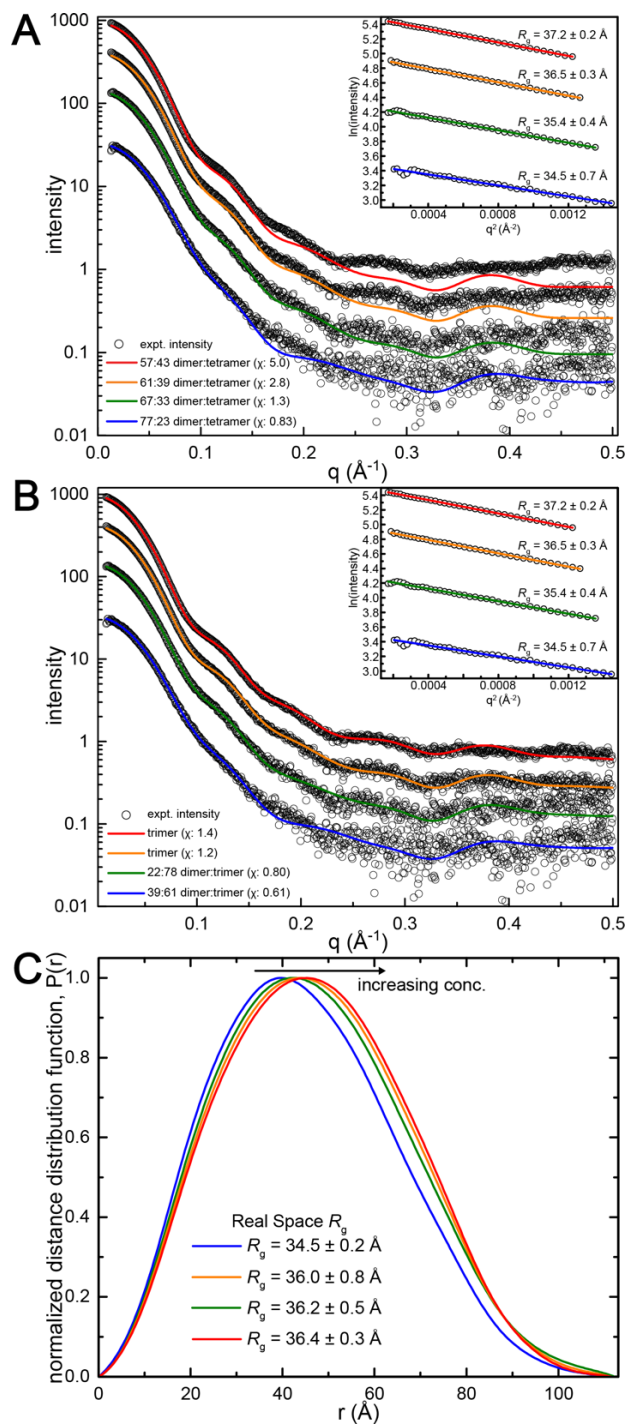
Figure 2.8. SAXS analysis of ALDH7A1 Δ504-511. (A) Experimental SAXS curves for ALDH7A1 Δ504-511 (open circles) at four concentrations (from bottom to top): 1, 2, 4, and 6 mg mL^{-1} (9 – 55 µM using dimer M_r). The theoretical SAXS curves were calculated using MultiFoXS from a two-body ensemble consisting of the domain-swapped dimer and the tetramer of ALDH7A1 (PDB: 4ZUK) lacking the terminal eight residues. The χ values from MultiFoXS are listed. An arbitrary scale factor has been applied for curve separation. The inset shows Guinier plots. (B) Analysis of the SAXS data from panel A using a model of the trimer of Δ504-511. The blue and green curves show the two-body fit obtained with a dimer:trimer ensemble. The orange and red curves show the single-body fit obtained with a trimer. (C) Experimental distance distribution functions calculated with GNOM via PRIMUS.

44

2.3 Discussion

This study was motivated by crystallographic data indicating that the C-terminus contributes to the active site of ALDH7A1 (Figure 2.2C). We chose to focus on Ala505 and Gln506 because they are identically conserved in family 7 of the ALDH superfamily, implying functional importance. Moreover, in the c.1512delG deletion mutation present in a subset of PDE patients, Ala505 and Gln506 are replaced by Pro and Lys, respectively (Figure 2.2D). Therefore, instead of substituting alanine for the wild-type residue, we elected to make mutations that incorporate the changes caused by c.1512delG, anticipating that the observed behavior of the variants might provide insight into the molecular basis of this form of PDE.

Our results suggest that Ala505 and Gln506 are important for AASAL binding. Mutation of Ala505 to Pro increases the K_m for AASAL by a factor of 10. Mutation of Gln506 to Lys – either alone or in combination with A505P – caused a profound catalytic defect characterized by a lack of saturation kinetics with AASAL. These results are consistent with Ala505 and Gln506 contributing to the binding of AASAL, a conclusion also suggested by the crystal structure showing that these residues stabilize the aldehyde anchor loop (Figure 2.2D).

Because the C-terminus contributes to the oligomerization domain, we hypothesized that the compromised catalytic activity of the variants might reflect an alteration of oligomeric structure. The oligomerization domain of ALDH7A1 is a β-substructure, consisting of a β-hairpin and the C-terminus of the polypeptide chain. It is involved in both dimerization and tetramerization. In the domain-swapped dimer, the oligomerization domain of one protomer contacts the catalytic domain of the other (Figure 2.2B). Via this

interaction, the C-terminus contributes to the aldehyde-binding site (Figure 2.2D). The oligomerization flap also contributes to tetramerization, mediating association of the two dimers (Figure 2.2C).

Although it has long been regarded as a tetrameric enzyme, wild-type ALDH7A1, in fact, exhibits interesting self-association behavior. At high-protein concentrations, it is undoubtedly tetrameric. It invariably crystallizes as a dimer-of-dimers tetramer. [6,7,12,13] And the presence of the tetramer in solution has been confirmed by SAXS. [6] Recently, however, we have learned that, at lower concentrations, wild-type ALDH7A1 exists as a dimer-tetramer equilibrium in solution, [4] a result confirmed in this work (Figure 2.4, Table 2.2). Thus, regarding ALDH7A1 strictly as a tetrameric enzyme, as it is often referred to in the literature (including by the authors), neglects a subtle aspect of the enzyme, an aspect that we suggest is related to catalytic function.

Related to this complex self-association behavior, our thermodynamic data suggest wild-type ALDH7A1 is dimeric at the concentrations used in enzyme activity assays. This observation raises the question: what is the true active form of the enzyme? One possibility is that the dimer, itself, is active. Another intriguing idea is that the binding of substrate or product enhances affinity for tetramerization, even at enzyme concentrations at which our data suggest tetramerization is improbable. Further study of this interesting phenomenon must be conducted to differentiate between these two possibilities.

All of the variant enzymes characterized in this study are defective in tetramer formation. At protein concentrations where the wild-type enzyme exhibits dimer-tetramer interconversion, the mutant enzymes exist as lower-order oligomeric assemblies. These data are consistent with perturbation of the dimer-tetramer equilibrium toward the dimer.

Another interpretation is that the mutants – particularly the Δ504-511 truncation variant – form an apparent trimer, a species not evident in studies of the wild-type enzyme. The formation of alternative oligomeric assemblies via perturbation of a fragile self-association equilibrium is characteristic of morpheein enzymes.[14–16] Interestingly, previous studies of porphobilinogen synthase (PBGS) revealed that ALAD porphyria-related mutations in PBGS favor formation of the less-active hexameric form over the more-active octamer.[15] Similarly, both previously[4] and in the current study we present PDE-related mutations that shift the self-association equilibrium toward less active oligomers. As observed in the PBGS mutations that cause ALAD porphyria,[15] the PDE-related mutations here result in high substrate apparent K_m values. In this case, we possibly identified a trimeric form of ALDH7A1 that is a less active than wild-type. Likewise, other PDE-related mutations apparently cause formation of inactive dimers.[4] Together, these observations suggest PDE-related mutations in ALDH7A1 that perturb the self-association equilibrium may be added to a list of morpheein-based conformational diseases.

Finally, our results provide insight into the molecular basis of PDE caused by the c.1512delG frameshift mutation. The enzyme encoded by *ALDH7A1* harboring c.1512delG includes the double mutation A505P/Q506K plus 4 other mutations and 10 extra residues in the C-terminus (Figure 2.2D). Because the double mutant A505P/Q506K has a severe kinetic defect, it is likely that the enzyme encoded by *ALDH7A1* c.1512delG is even more catalytically disabled. Indeed, Mills et al. found no detectable AASAL dehydrogenase activity in CHO cells transfected with the *ALDH7A1* gene encoding c.1512delG.[3] Presuming the frameshift mutant enzyme is soluble, we suggest that disruption of tetramerization may be the structural basis for their observation.

2.4 Materials and Methods

Site-Directed Mutagenesis and Protein Production.

The site-directed mutants of ALDH7A1 A505P, Q506K, and A505P/Q506K were generated using the QuikChange method by former graduate student Min Luo. Pfu Turbo DNA Polymerase was purchased from Stratagene-Agilent. The primers used in mutagenesis (from Integrated DNA Technologies) are listed in Table S2.1 of the Supporting Information. The mutations were confirmed by DNA sequencing. The C-terminal truncation mutant Δ504-511 was generated by introduction of a premature stop codon by site-directed mutagenesis using the QuikChange Lightning site-directed mutagenesis kit (Agilent) by former postdoctoral associate Dr. David Korasick. Primers are listed in Table S2.1.

Expression and purification of wild-type and mutant enzymes followed the protocols described previously for the wild-type enzyme.[12] Briefly, cell cultures were grown at 37°C at 250 rpm and induced with 0.2 mM IPTG when the OD_{600} reached 0.8. Induction was performed for 18 hours at 18°C at 200 rpm. Cells were harvested and resuspended in a buffer containing 50 mM Hepes pH 8.0, 200 mM NaCl, 10 mM imidazole, and 5% (v/v) glycerol. Following purification by immobilized Ni^{2+}-affinity chromatography, the His tag was removed with tobacco etch virus protease, and the protein was dialyzed into 50 mM Hepes, 100 mM NaCl, and 5% (v/v) glycerol at pH 7.8. Size exclusion chromatography was then performed using a Superdex 200 column. The pooled fractions were concentrated and stored at -80°C in 50 mM Hepes pH 7.8, 100 mM NaCl, 5% (v/v) glycerol, and 0.5 mM dithiothreitol.

Mass Spectrometry,

The sequences of the ALDH7A1 variants were confirmed by mass spectrometry. Nano LC-Nanospray QTOF MS analysis was performed on the ALDH7A1 variants at the University of Missouri Charles W. Gehrke Proteomics Center. Samples were prepared to 5 mg/mL in 50 mM HEPES at pH 7.8, 100 mM NaCl, 5 mM dithiothreitol, and 5% glycerol. Each sample was diluted to 2 pmole/µL in 30/970/1 (v/v/v) acetonitrile/water/99% formic acid. A 0.1 µL aliquot was injected for fractionation and mass analysis by Nano LC on an Agilent Protein Chip column interfaced with a Nanospray QTOF MS (Agilent 6520A MS). Prior to the sample analysis, the system was primed with an injection of cap pump solvent (2 µL of 30/970/1 (v/v/v) acetonitrile/water/99% formic acid).

From selected retention time windows of the total ion chromatograms, multiple-charge ion spectra were obtained. These spectra were deconvoluted with the maximum entropy algorithm, and its peaks were labeled with the peak height and calculated neutral average masses. Protparam [17] was used to predict theoretical masses for the ALDH7A1 variants. The experimental masses were within 1 Da of the theoretical values, validating that site-directed mutagenesis was accurate (Table S2.2).

Synthesis of P6C/AASAL.

P6C/AASAL was prepared using a modified version of the previously reported procedure.[3,18] A 50 mg sample of L-allysine ethylene acetal (Sigma-Aldrich) was stirred with 800 mg of Amberlyst-15 (dry) ion exchange resin (Sigma-Aldrich) in water (5 mL) for 30 minutes at 45°C. The reaction mixture was filtered through a sintered funnel, and the resin was washed with water (5 mL). P6C/AASAL was eluted from the resin with 4

mL of 28% ammonia solution. The filtrate was evaporated to dryness under a stream of nitrogen, producing a pale yellow solid.

P6C/AASAL was quantitated by reaction with o-aminobenzaldehyde (oAB). The synthesized P6C/AASAL sample was dissolved in 1 mL of 100 mM sodium pyrophosphate buffer (pH 8.0). A dilution series of P6C/AASAL was incubated with 200µM oAB in a total volume of 1 mL for 1 hour at 37°C. The concentration of the P6C-oAB reaction product was determined from absorbance at 465 nm using an extinction coefficient of $\varepsilon = 2.8$ mM^{-1} cm^{-1}. [19]

Steady-state Kinetics Assays.

ALDH activity was measured at 26°C in an Epoch 2 microplate reader (Biotek) by continuously monitoring NADH formation at 340 nm for 20 minutes using AASAL as the variable substrate and NAD$^+$ at a fixed, saturating concentration. The final assay mixture (200 µL) contained 90 nM enzyme, 2.5 mM NAD$^+$, and 60-1500 µM AASAL in 0.1 M sodium pyrophosphate buffer (pH 8.0). The enzyme concentration was determined using the bicinchoninic acid method (Pierce). The pH of the aldehyde stock solution was adjusted to 8.0 prior to addition to the reaction mixture. Kinetic constants for wild-type ALDH7A1 and A505P were estimated by fitting the initial-rate data (2 min) to the Michaelis-Menten equation using Origin 2017. The initial-rate data for Q506K, A505P/Q506K, and Δ504-511 were linear, rather than hyperbolic, presumably because the Michaelis constant for AASAL vastly exceeded the accessible substrate concentrations. In these cases, the slope of the best-fit line was interpreted as an approximation of k_{cat}/K_m.

Analytical Ultracentrifugation.

Analytical ultracentrifugation experiments were conducted by Dr. David Korasick. Sedimentation experiments were performed at 20°C using an An50Ti rotor in a Beckman XL-I analytical ultracentrifuge. For sedimentation-velocity analysis, a sedimentation-velocity cell, bearing a two-sector charcoal-Epon centerpiece, was loaded with 430 µL of reference buffer (size-exclusion flow through) and 400 µL of protein solution. The sample was allowed to equilibrate at temperature for two hours prior to initiating the analysis. Following equilibration, samples were centrifuged at 35,000 rpm for 300 scans. Scans were acquired using Rayleigh interference optics and were collected at two-minute intervals. Every other scan from 10-300 (145 total scans) was analyzed using Sedfit. [20]

For sedimentation-equilibrium analysis, data were collected at three rotor speeds – 6,000 rpm, 9,000 rpm, and 12,000 rpm – at three different protein concentrations. Samples and reference buffer were loaded into a six-sector charcoal-Epon centerpiece and allowed to equilibrate for 16 h at 6,000 rpm prior to collection of the first scan. Prior to the first scan at 9,000 and 12,000 rpm, the system was allowed to equilibrate for 8 h. At each rotor speed, six scans were collected at hourly intervals after equilibration. The final scan at each rotor speed was used for data analysis. Data were analyzed using the previously described models [4] in Origin 2017.

Small-Angle X-ray Scattering (SAXS).

The Δ504-511 variant was analyzed using SAXS. Shutterless SAXS data collection was performed with a Pilatus detector at beamline 12.3.1 of the Advanced Light Source through the mail-in program. [21,22] Scattering intensities were measured at protein concentrations of 1, 2, 4, and 6 mg mL^{-1} (9 – 55 µM by dimer M_r). For each sample, 33 images were recorded evenly during a total exposure time of 10.2 s. Background

subtraction data were collected similarly on the dialysis buffer. For each protein sample, the low q region of the first 3 images (0.9 s of exposure) were averaged and merged with the high q region of the average of the first 12 images (3.6 s of exposure) using PRIMUS. [23] Guinier analysis was performed with PRIMUS using a maximum qR_g of 1.3. Distance distribution functions were calculated with GNOM [24] through PRIMUS. Single- and multiple-species model fitting were carried out using FoXS and MultiFoXS,[10] respectively. Results were graphed in Origin 2017.

Supporting Information

Table S2.1. Primers Used for Site-Directed Mutagenesis

A505P fwd	5'– GACCTTCCTCTGCCTCAAGGAATCAAGTTTCAG –3'
A505P rev	5'– CTGAAACTTGATTCCTTGAGGCAGAGGAAGGTC –3'
Q506K fwd	5'– GACCTTCCTCTGGCCAAGGGAATCAAGTTTCAG –3'
Q506K rev	5'– CTGAAACTTGATTCCCTTGGCCAGAGGAAGGTC –3'
A505P/Q506K fwd	5'– GACCTTCCTCTGCCTAAGGGAATCAAGTTTCAG –3'
A505P/Q506K rev	5'– CTGAAACTTGATTCCCTTAGGCAGAGGAAGGTC –3'
Δ504-511 fwd	5'– CAGTAAAGACCTTCCTTAGGCCCAAGGAATCAAG –3'
Δ504-511 rev	5'– CTTGATTCCTTGGGCCTAAGGAAGGTCTTTACTG –3'

Table S2.2. Mass Spectrometry Results for Wild-Type ALDH7A1 and Variants

	Theoretical Mass (Da)	Experimental Mass (Da)
ALDH7A1	55560.61	55561.69 ± 1.15
A505P	57883.08[a]	57884.17 ± 1.15
Q506K	55560.66	55561.91 ± 1.15
A505P/Q506K	55586.69	55587.61 ± 1.15
Δ504-511	54674.55	54675.64 ± 1.15

[a]Theoretical molecular mass includes the His-tag.

Table S2.3. FoXS/MultiFoXS χ Values from Fitting Models to the SAXS Curves[a]

Model	1 mg mL^{-1}	2 mg mL^{-1}	4 mg mL^{-1}	6 mg mL^{-1}
	Single-body fits			
Monomer	5.2	11	20.	33
Dimer	1.8	4.6	9.4	16.
Trimer	0.84	0.98	**1.2**	**1.4**
Tetramer	2.9	4.2	6.8	9.9
	Two-body fits			
Dimer:Trimer	**0.61 (39:61)**	**0.80 (22:78)**	N/D[b]	N/D[b]
Dimer:Tetramer	0.83 (77:23)	1.3 (67:33)	2.8 (61:39)	5.0 (57:43)

[a]Bold red font indicates the lowest χ for each sample.
[b]N/D denotes that MultiFoXS did not produce a two-body fit, indicating that the inclusion of a second body did not significantly decrease χ.

References

[1] C.D. van Karnebeek, S.A. Tiebout, J. Niermeijer, B.T. Poll-The, A. Ghani, C.R. Coughlin, J.L. Van Hove, J.W. Richter, H.J. Christen, R. Gallagher, H. Hartmann, S. Stockler-Ipsiroglu, Pyridoxine-dependent epilepsy: An expanding clinical spectrum, Pediatr Neurol. 59 (2016) 6.

[2] S. Stockler, B. Plecko, S.M. Gospe, M. Coulter-Mackie, M. Connolly, C. van Karnebeek, S. Mercimek-Mahmutoglu, H. Hartmann, G. Scharer, E. Struijs, I. Tein, C. Jakobs, P. Clayton, J.L. Van Hove, Pyridoxine dependent epilepsy and antiquitin deficiency: clinical and molecular characteristics and recommendations for diagnosis, treatment and follow-up, Mol. Genet. Metab. 104 (2011) 48.

[3] P.B. Mills, E. Struys, C. Jakobs, B. Plecko, P. Baxter, M. Baumgartner, M.A. Willemsen, H. Omran, U. Tacke, B. Uhlenberg, B. Weschke, P.T. Clayton, Mutations in antiquitin in individuals with pyridoxine-dependent seizures, Nat. Med. 12 (2006) 307.

[4] D.A. Korasick, J.J. Tanner, M.T. Henzl, Impact of disease-Linked mutations targeting the oligomerization interfaces of aldehyde dehydrogenase 7A1, Chem.-Biol. Interact. (2017).

[5] V. Koppaka, D.C. Thompson, Y. Chen, M. Ellermann, K.C. Nicolaou, R.O. Juvonen, D. Petersen, R.A. Deitrich, T.D. Hurley, V. Vasiliou, Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application, Pharmacol. Rev. 64 (2012) 520.

[6] M. Luo, J.J. Tanner, Structural basis of substrate recognition by aldehyde

dehydrogenase 7A1, Biochemistry. 54 (2015) 5513.

[7]   C. Brocker, N. Lassen, T. Estey, A. Pappa, M. Cantore, V. V Orlova, T. Chavakis, K.L. Kavanagh, U. Oppermann, V. Vasiliou, Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress, J. Biol. Chem. 285 (2010) 18452.

[8]   R. Koncitikova, A. Vigouroux, M. Kopecna, T. Andree, J. Bartos, M. Sebela, S. Morera, D. Kopecny, Role and structural characterization of plant aldehyde dehydrogenases from family 2 and family 7, Biochem. J. 468 (2015) 109.

[9]   H. Fischer, M. de Oliveira Neto, H.B. Napolitano, I. Polikarpov, A.F. Craievich, Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale, J. Appl. Crystallogr. 43 (2010) 101.

[10] D. Schneidman-Duhovny, M. Hammel, J.A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles, Nucleic Acids Res. 44 (2016) W424.

[11] D. Schneidman-Duhovny, M. Hammel, A. Sali, FoXS: a web server for rapid computation and fitting of SAXS profiles, Nucleic Acids Res. 38 (2010) W540.

[12] M. Luo, K.S. Gates, M.T. Henzl, J.J. Tanner, Diethylaminobenzaldehyde is a covalent, irreversible inactivator of ALDH7A1, ACS Chem. Biol. 10 (2015) 693.

[13] W.K. Tang, K.B. Wong, Y.M. Lam, S.S. Cha, C.H. Cheng, W.P. Fong, The crystal structure of seabream antiquitin reveals the structural basis of its substrate specificity, FEBS Lett. 582 (2008) 3090.

[14] S.H. Lawrence, T. Selwood, E.K. Jaffe, Environmental contaminants perturb fragile

protein assemblies and inhibit normal protein function, Curr. Chem. Biol. 7 (2013) 196.

[15] E.K. Jaffe, L. Stith, ALAD porphyria is a conformational disease, Am. J. Hum. Genet. 80 (2007) 329.

[16] E.K. Jaffe, Morpheeins–a new structural paradigm for allosteric regulation, Trends Biochem. Sci. 30 (2005) 490.

[17] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server BT - The Proteomics Protocols Handbook, in: J.M. Walker (Ed.), Humana Press, Totowa, NJ, 2005: pp. 571–607.

[18] A. Rumbero, J. Fco. Martin, M. Angeles Lumbreras, P. Liras, C. Esmahan, Chemical synthesis of allysine ethylene acetal and conversion in situ into 1-piperideine-6-carboxylic acid: key intermediate of the alpha-aminoadipic acid for beta-lactam antibiotics biosynthesis, Bioorg. Med. Chem. 3 (1995) 1237.

[19] M.L. Wu, J.H. Chen, C.T. Ho, T.C. Huang, Synthesis of 1-piperideine-6-carboxylic acid produced by L-lysine-epsilon-aminotransferase from the Streptomyces clavuligerus gene expressed in Escherichia coli, J. Agric. Food Chem. 55 (2007) 1767.

[20] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, Biophys. J. 78 (2000) 1606.

[21] G.L. Hura, A.L. Menon, M. Hammel, R.P. Rambo, F.L. Poole, S.E. Tsutakawa, F.E. Jenney, S. Classen, K.A. Frankel, R.C. Hopkins, S.J. Yang, J.W. Scott, B.D. Dillard, M.W. Adams, J.A. Tainer, Robust, high-throughput solution structural analyses by

small angle X-ray scattering (SAXS), Nat. Methods. 6 (2009) 606.

[22]   S. Classen, G.L. Hura, J.M. Holton, R.P. Rambo, I. Rodic, P.J. McGuire, K. Dyer, M. Hammel, G. Meigs, K.A. Frankel, J.A. Tainer, Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source, J. Appl. Crystallogr. 46 (2013) 1.

[23]   P. V Konarev, V. V Volkov, A. V Sokolova, M.H.J. Koch, D.I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis, J. Appl. Crystallogr. 36 (2003) 1277.

[24]   D. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495.

Chapter Three: Inhibition, crystal structures, and in-solution oligomeric structure of aldehyde dehydrogenase 9A1

3.1. Introduction

The aldehyde dehydrogenase (ALDH) structural superfamily is a large group of enzymes that catalyze the NAD^+-dependent oxidation of aldehydes to carboxylic acids [1,2]. The superfamily comprises hundreds of distinct genes, including 19 ALDHs expressed in humans. ALDHs share a common protein fold and catalytic mechanism, but subtle differences in their active sites result in different preferences for the aldehyde substrate.

ALDH9A1 is a human enzyme that catalyzes the NAD^+-dependent oxidation of a variety of aldehydes including the carnitine precursor 4-trimethylaminobutyraldehyde (TMBAL, Scheme 3.1) [3], the GABA precursor aminobutyraldehyde [4], the dopamine metabolite 3,4-dihydroxyphenylacetaldehyde [5], and betaine aldehyde [6]. A recent comprehensive study of substrate specificity showed that ALDH9A1 has a strong (> 10-fold) preference for TMBAL [7], confirming the assertion of Vaz, *et al.* [3] that the major *in vivo* function of this enzyme is to catalyze the penultimate step of carnitine biosynthesis (E.C. 1.2.1.47), the oxidation of TMBAL to 4-N-trimethylaminobutyrate. Carnitine functions in the transport of long-chain fatty acids from the cytosol to the mitochondrial matrix for the synthesis of acyl-CoAs for β-oxidation. Because the mitochondrial membrane is impermeable to acyl-CoAs, fatty acids are conjugated to carnitine to enter mitochondria. Thus, ALDH9A1 functions indirectly in β-oxidation.
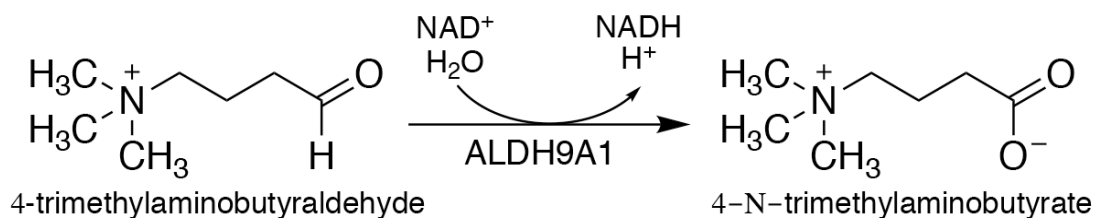
Scheme 3.1. Reaction catalyzed by ALDH9A1.

Recently the first crystal structure of human ALDH9A1 was reported [7] (PDB ID: 6QAP). Although ALDH9A1 exhibits the basic ALDH superfamily fold, the structure revealed two remarkable features. First, the final α-helix and β-strand of the Rossmann dinucleotide-binding fold are disordered. Referred to as αE-βE in the closely-related betaine ALDH [8], these secondary structural elements form extensive interactions with NAD$^+$ in other ALDHs. Further, αE forms a major part of the dimer interface in other ALDHs. The second atypical aspect of the apo ALDH9A1 structure concerns the ~25-residue inter-domain linker, which connects the catalytic domain to the oligomerization domain. In other ALDH structures, the linker folds into a long β-hairpin, wherein the strands of the β-hairpin form the floor of the aldehyde substrate tunnel, while the tip helps stabilize the NAD$^+$ binding site. In contrast, the linker in the apo ALDH9A1 structure adopts a novel S-shaped conformation, which blocks access to the aldehyde substrate binding pocket. The apo ALDH9A1 structures appear to represent an inactive enzyme, raising the question of how αE-βE and the inter-domain linker isomerize into their active conformations.

Herein we describe a structural and biochemical study of ALDH9A1 designed to provide understanding of the catalytic function of this enzyme, in light of the atypical structural features observed previously for the apo enzyme. We demonstrate that

diethylaminobenzaldehyde (DEAB) is a reversible, time-dependent inhibitor of ALDH9A1. We report two crystal structures of ALDH9A1 complexed with NAD^+. In one of the structures, which was obtained from enzyme treated with DEAB and NAD^+, αE-βE and the inter-domain linker are ordered and adopt the canonical secondary, tertiary, and quaternary structures observed in other ALDHs. In contrast, the other NAD^+ complex structure exhibits the atypical features of the apo enzyme, showing that NAD^+ binding alone is not sufficient to promote the active conformation of the enzyme. The oligomeric structure of ALDH9A1 was determined using analytical ultracentrifugation, small-angle X-ray scattering, and negative stain electron microscopy. These studies show that ALDH9A1 forms the classic ALDH superfamily dimer-of-dimers tetramer in solution. Our results provide insight into the mechanism by which ALDH9A1 is activated for catalysis.

3.2. Results

Time-dependent Inhibition of ALDH9A1 by DEAB

DEAB (Figure 3.1A) is variously a substrate, competitive tight-binding inhibitor, or irreversible covalent inactivator of ALDHs [9,10]. We therefore performed activity assays to understand how DEAB interacts with ALDH9A1.

We first tested whether DEAB is a substrate using the approach of Morgan *et al.* [9] by incubating ALDH9A1 (0.5 μM) with DEAB (30 μM) and NAD^+ (1.5 mM), while monitoring the absorbance at 360 nm and 300 nm. In this assay, a decrease at 360 nm indicates consumption of DEAB, while an increase at 300 nm indicates production of the carboxylic acid product, diethylaminobenzoic acid. The spectrum of DEAB alone shows a broad peak centered at 360 nm, as expected (Figure 3.1A). Spectra acquired for the mixture of ALDH9A1, DEAB, and NAD^+ show little change over 4 hours (Figure 3.1A). These

results suggest that DEAB is not an efficient substrate for ALDH9A1.

Because we observed no significant activity with DEAB as the aldehyde substrate, we next performed time-dependent inhibition assays. A progress curve approach was used in which the enzyme (0.5 μM) was mixed with the aldehyde substrate hexanal (2 mM), NAD$^+$ (2.5 mM), and DEAB (0 – 5.0 μM) (Figure 3.1B). The initial increase in absorbance at 340 nm indicates the production of NADH associated with the catalytic turnover of hexanal to hexanoic acid. The flattening of the curves at longer time is characteristic of covalent inactivation, similar to what has been observed with ALDH2 [9] and ALDH7A1 [10]. Alternatively, time-dependent loss of activity is associated with slow, tight-binding inhibition [11–13]. A control experiment confirmed that flattening of the curves was not due to depletion of substrate or cofactor (inset, Figure 3.1B). Global fitting of the data to combined Equations 1a and 1b was performed to estimate kinetic parameters for enzyme inactivation, assuming $k_6 << k_5$ (see Material and Methods). These calculations yielded k_5 of 0.184 ± 0.052 min^{-1}, K_I of 0.054 ± 0.018 μM, and an apparent second-order rate constant for enzyme inactivation by DEAB (k_5/K_I) of 57000 ± 25000 M^{-1}s^{-1}. This value is similar to that of ALDH2 (86000 M^{-1}s^{-1}) [9].
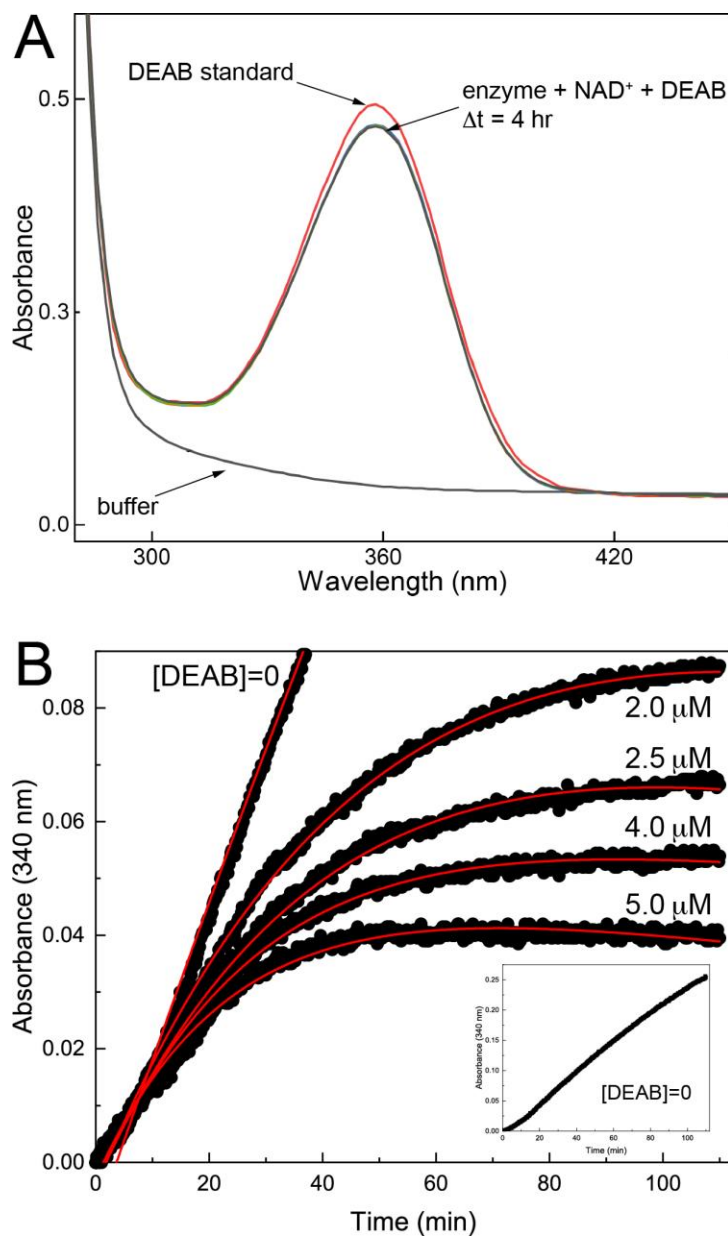
Figure 3.1. Inhibition of ALDH9A1 by DEAB. (A) Evidence that DEAB is not a substrate of ALDH9A1. ALDH9A1 (0.5 µM) was incubated with 30 µM DEAB and 1.5 mM NAD$^+$, and spectra were collected every 10 minutes for a total of 4 hours. For reference, the red curve is the spectrum of DEAB. (B) Time-dependent inhibition of ALDH9A1 by DEAB. The black circles are experimental progress curves monitoring NADH production from ALDH9A1 in the presence of 2 mM of the aldehyde substrate hexanal, 2.5 mM NAD$^+$, and various concentrations of DEAB. The red curves represent the global fit to combined Equations 1a and 1b. The parameters from global fitting are $k_5 = 0.184 \pm 0.052$ min^{-1}, $K_I = 0.054 \pm 0.018$ µM. For ease of viewing, a vertical offset was applied to each data set to account for the absorbance due to the different DEAB concentrations. The inset shows the progress curve in the absence of DEAB. The unshifted progress curves with error bars are shown in Figure S3.2.

The reversibility of inhibition was assessed by monitoring the return of enzyme activity upon dilution. ALDH9A1 (5 μM) was incubated with 10 μM DEAB and 1 mM NAD$^+$ at 4°C for 60 minutes, and then assayed for activity by diluting 10-fold into assay buffer containing 0.5 mM hexanal and 1 mM NAD$^+$. The progress curve for the DEAB-treated sample shows a ~5-minute lag in NADH production, followed by an increase in activity (Figure 3.2A). The steady-state velocity of the DEAB-treated sample was ~70% compared to a control that was treated identically but lacked DEAB. These results show that inhibition of ALDH9A1 by DEAB is reversible.

The reversibility of inhibition was confirmed by measuring the return of enzyme activity after removal of DEAB. ALDH9A1 (5 μM) was incubated with 10 μM DEAB and 1 mM NAD$^+$ for 60 minutes, and then the excess DEAB was removed by spin ultrafiltration. The activity was then measured in an assay containing 0.5 mM hexanal and 1 mM NAD$^+$. The progress curves for the DEAB-treated and control samples are very similar (Figure 3.2B). Note the progress curve for the DEAB-treated sample does not exhibit a lag. These results are consistent with a covalent reversible mechanism of inhibition.
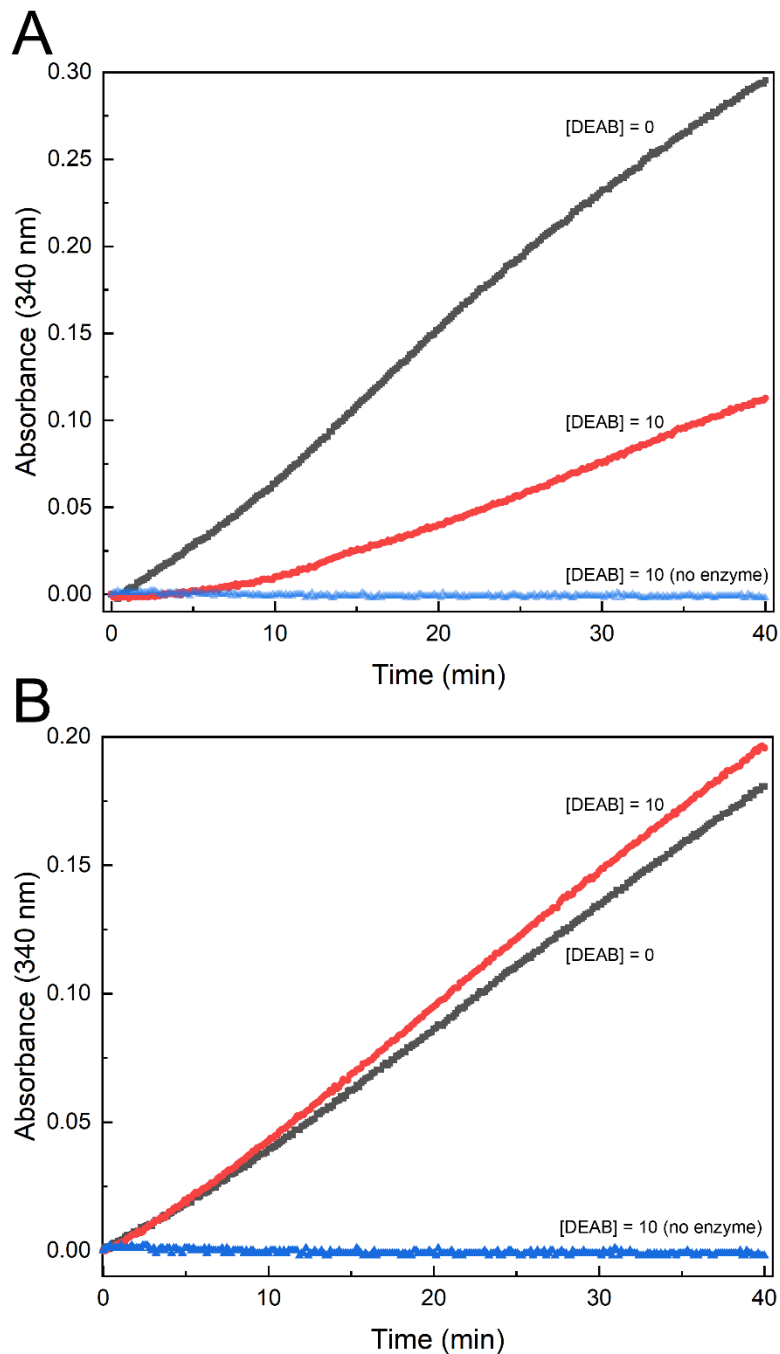
Figure 3.2. Inhibition of ALDH9A1 by DEAB is reversible. (A) Progress curves from a dilution assay. ALDH9A1 (5 μM) was incubated with 10 μM DEAB and 1 mM NAD$^+$ at 4°C for 60 minutes, and then assayed for activity by diluting 10-fold into assay buffer containing 0.5 mM hexanal and 1 mM NAD$^+$. (B) Residual activity after removing excess DEAB. ALDH9A1 (5 μM) was incubated with 10 μM DEAB and 1 mM NAD$^+$ for 60 minutes, and then the excess DEAB was removed by spin ultrafiltration using Amicon Ultra-0.5 mL centrifugal filter units. The activity was then measured in an assay containing 0.5 mM hexanal and 1 mM NAD$^+$. Prior to the enzyme assays, the enzyme concentration was measured and used to adjust the progress curves.

The Fold of ALDH9A1 Complexed with NAD$^+$

Two structures of ALDH9A1 complexed with NAD$^+$ were determined. A 2.5 Å resolution structure with space group *P*1 was obtained using enzyme that had been incubated with both NAD$^+$ and DEAB (Table 3.1). Also, a 2.64 Å structure in space group *C*222 was determined using enzyme incubated with NAD$^+$, without DEAB. We note these crystal forms are different from those used for structure determination of apo ALDH9A1 [7].

As expected, ALDH9A1 complexed with NAD$^+$ exhibits the 3-domain architecture of apo ALDH9A1, as well as other ALDHs (Figure 3.3A). The NAD$^+$-binding domain features a 5-stranded Rossmann-dinucleotide binding fold at its core. The catalytic domain has an α/β fold and contains the catalytic cysteine residue (Cys288). The oligomerization domain adopts the expected β- substructure protruding from the NAD$^+$-binding domain.

The structures of ALDH9A1-NAD$^+$ were compared to each other and with the contents of the PDB using PDBeFold [14] to identify the closest structural neighbors. Table 3.2 lists the pairwise RMSD values between the chains of the various structures used in the analysis. Comparison of our *P*1 and *C*222 NAD$^+$ structures yields RMSD values of 1.0 – 1.3 Å. This range is 3-5 times larger than the chain-chain RMSDs within either structure, suggesting that the two structures represent distinct conformations. This idea was supported by comparing our structures to the PDB. We found that the *P*1 ALDH9A1-NAD$^+$ structure is most similar to betaine ALDH from cod liver (RMSD = 0.6 – 0.7), whereas the *C*222 ALDH9A1-NAD$^+$ structure most resembles apo ALDH9A1 (RMSD = 0.4 – 0.6 Å).

Table 3.1.

X-ray diffraction data collection and refinement statistics.

Space group	*P*1	*C*222
Unit cell parameters (Å,°)	$a = 82.79$, $b = 90.30$, $c = 145.24$, $\alpha = 89.37$, $\beta = 84.04$, $\gamma = 73.87$	$a = 158.15$, $b = 163.47$, $c = 84.34$
Wavelength (Å)	0.97918	1.00000
Resolution (Å)	79.09 – 2.54 (2.54 – 2.50)	47.13 – 2.64 (2.77 – 2.64)
Observations[a]	177440 (9039)	229337 (30035)
Unique reflections[a]	106567 (5667)	32497 (4262)
$R_{merge}(I)$[a]	0.090 (0.500)	0.151 (1.070)
$R_{meas}(I)$[a]	0.127 (0.708)	0.163 (1.156)
$R_{pim}(I)$[a]	0.090 (0.500)	0.061 (0.434)
Mean I/σ[a]	4.5 (1.0)	11.6 (1.7)
$CC_{1/2}$	0.991 (0.610)	0.996 (0.803)
Completeness (%)[a]	76.7 (82.8)	100.0 (100.0)
Multiplicity[a]	1.7 (1.6)	7.1 (7.0)
No. of protein chains	8	2
No. of protein residues	3939	900
No. of atoms		
Protein	29077	6729
NAD^+	233	46
Water	N/A	31
R_{cryst}[a]	0.213 (0.2996)	0.218 (0.288)
R_{free}[a,b]	0.269 (0.352)	0.273 (0.332)
rmsd bonds (Å)	0.002	0.002
rmsd angles (°)	0.591	0.511
Ramachandran plot[c]		
Favored (%)	97.30	98.19
Outliers (%)	0.03	0.00
Clashscore (PR)[c]	5.73 (99)	3.24 (100)
MolProbity score (PR)[c]	1.96 (96)	1.41 (100)
Average B (Å^2)		
Protein	38.7	50.8
NAD^+	39.5	58.6
Water	N/A	30.9
Coord. error (Å)[d]	0.38	0.41
PDB code	6VR6	6VWF

[a]Values for the outer resolution shell of data are given in parenthesis. [b]2% test set for 6VR6; 5% test set for 6VWF. [c]From MolProbity. The percentile ranks (PR) for Clashscore and MolProbity score are given in parentheses. [d]Maximum likelihood-based coordinate error estimate from PHENIX.

The RMSD data suggest that the current ALDH9A1 structures can be classified into two groups, one corresponding to the $P1$ NAD$^+$ complex structure, and the other consisting of the $C222$ NAD$^+$ complex structure and the apo structures. Additional analysis shows that the two groups differ by a domain rotation. For example, when the NAD$^+$-binding domains of the $P1$ NAD$^+$ complex and apo ALDH9A1 are superimposed, the catalytic domains differ by an apparent rigid body rotation of ~5° (Figure 3.3B). Apparently, interactions of the enzyme with DEAB and NAD$^+$ caused the NAD$^+$-binding and catalytic domains to rotate closer together.

Table 3.2.
Pairwise RMSDs between ALDH structures.[a]

	ALDH9A1-NAD$^+$ ($P1$)	ALDH9A1-NAD$^+$ ($C222$)	ALDH9A1 apo (6QAP)	Betaine ALDH apo (1A4S)	Betaine ALDH-NAD$^+$ (1BPW)
ALDH9A1-NAD$^+$ ($P1$)	0.00 – 0.24	1.01 – 1.26	1.35 – 1.51	0.65 – 0.69	0.62 – 0.65
ALDH9A1-NAD$^+$ ($C222$)		0.00 – 0.28	0.44 – 0.58	1.27 – 1.36	1.21 – 1.30
ALDH9A1 apo (6QAP)			0.00 – 0.24	1.46 – 1.56	1.41 – 1.51
Betaine ALDH apo (1A4S)				0.00 – 0.09	0.31 – 0.32
Betaine ALDH-NAD$^+$ (1BPW)					0.00 – 0.00

[a]Calculated with PDBeFold. The range was calculated from all pairwise chain-chain comparisons.
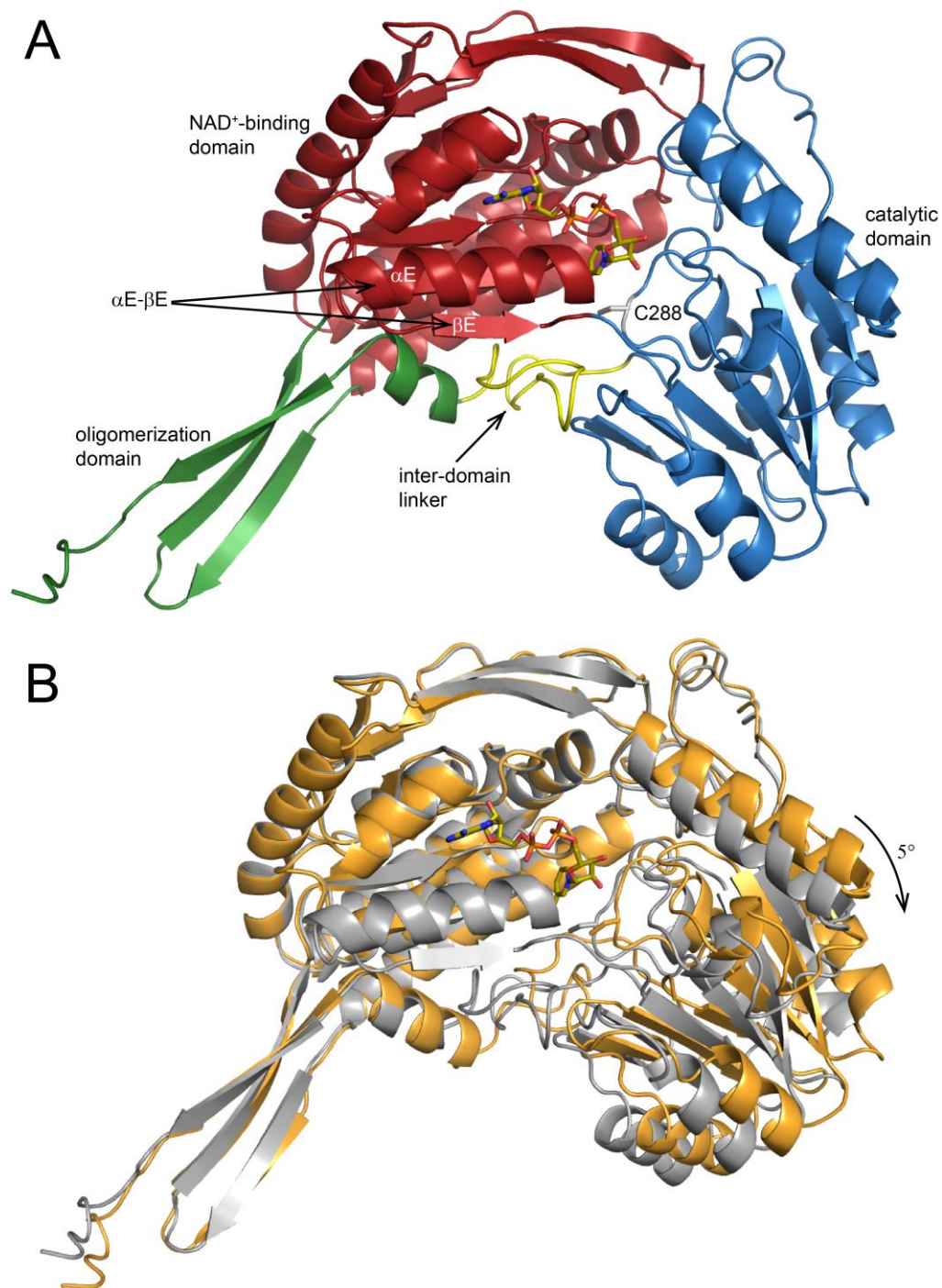
Figure 3.3. The crystal structure of *P*1 ALDH9A1 complexed with NAD$^+$. (A) One protomer (chain A) of ALDH9A1 complexed with NAD$^+$ in space group *P*1. The NAD$^+$-binding, catalytic, and oligomerization domains have different colors. Arrows indicate αE-βE of the Rossmann fold domain and the β-hairpin inter-domain linker (yellow). NAD$^+$ is shown in sticks. (B) Superposition of NAD$^+$-bound (gray, PDB ID: 6VR6) and apo ALDH9A1 (gold, PDB ID: 6QAP). The superposition calculation was based on the NAD$^+$-binding domains to accentuate the difference in the orientations of the catalytic domains. The arrow shows the direction of domain closure associated with NAD$^+$ binding.

Conformation and Interactions of NAD⁺ Bound to ALDH9A1

Electron density for NAD$^+$ was observed in all eight chains of the *P*1 asymmetric unit. In chain A, electron density was present for the entire cofactor (Figure 3.4A), whereas in the other chains only the ADP portion could be modeled with confidence (Figure S3.3). Similarly, only the AMP fragment could be modeled in the C222 structure (Figure S3.4). We note that weak density for the nicotinamide riboside of NAD$^+$ is common in ALDHs [15–19]. Electron density extending from catalytic Cys288 was present in all eight chains of the *P*1 structure, possibly indicating covalent modification by DEAB; however, we were not able to model DEAB covalently bound to Cys288 satisfactorily into these features. The lack of convincing density for the covalent adduct is perhaps consistent with the apparent reversibility of inactivation (Figure 3.2).

NAD$^+$ binds in the expected site at the C-termini of the β-strands of the Rossmann fold. NAD$^+$ forms several electrostatic interactions with the protein (Figure 3.4A). The adenine ribose hydrogen bonds with Lys180. The pyrophosphate interacts with Trp156, Ser233, and Thr236. The nicotinamide ribose of the one complete NAD$^+$ forms a hydrogen bond with Glu391, a residue identically conserved in the ALDH superfamily [20].
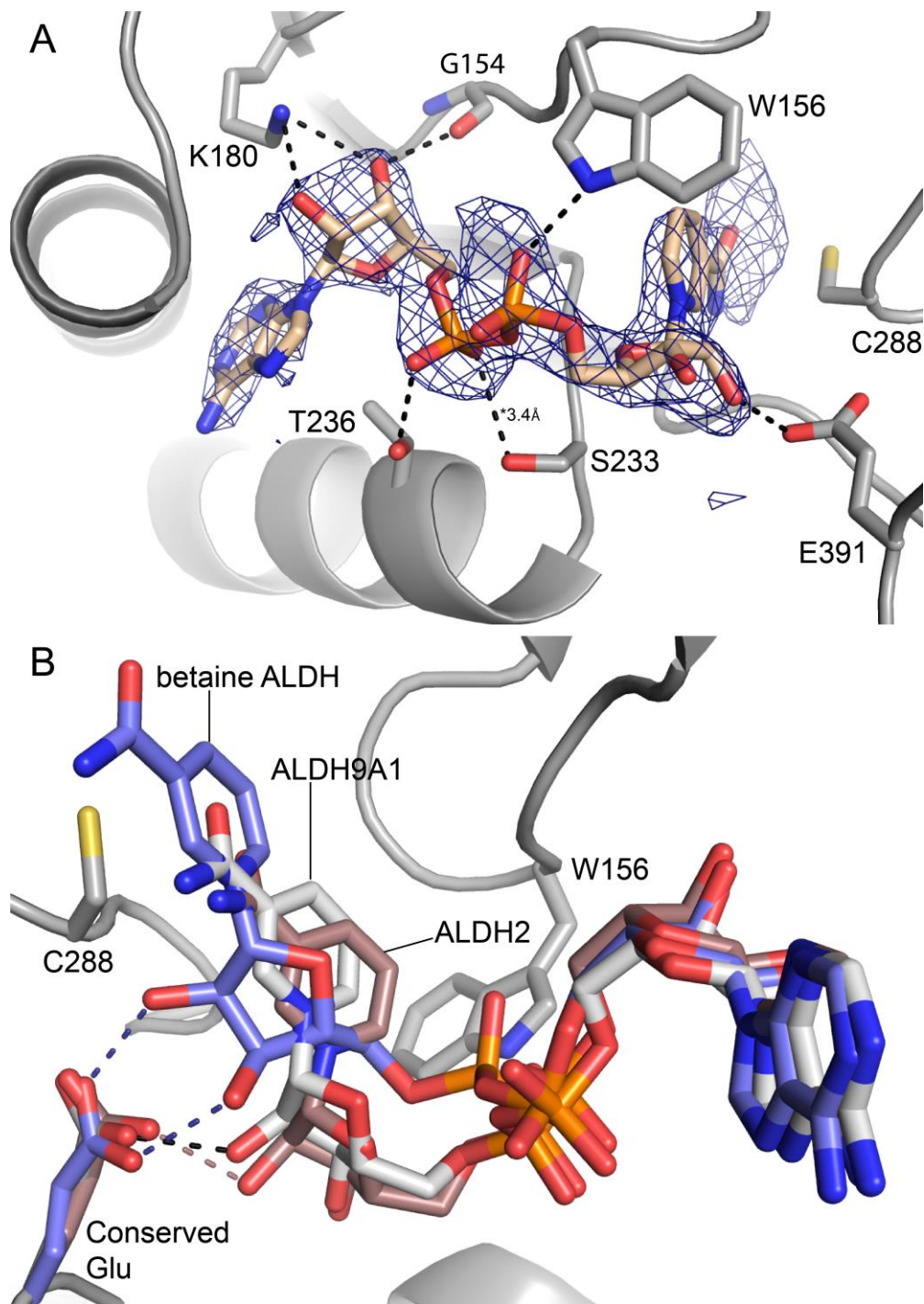
Figure 3.4. Conformation of NAD$^+$ bound to chain A of the P1 ALDH9A1 (PDB ID: 6VR6). (A) Electron density and interactions for NAD$^+$. The cage represents a polder omit map (3.5σ). The dashed lines indicate interactions less than 3.2 Å, except where noted. (B) Superposition of the cofactors of P1 ALDH9A1 (gray), betaine ALDH (blue, PDB ID: 1BPW), and ALDH2 (brown, PDB ID: 1O02). The betaine ALDH and ALDH2 structures show the "hydride transfer" and "hydrolysis" conformations, respectively. Hydrogen bonds with the conserved glutamate residue are colored as follows: black, ALDH9A1; blue, betaine ALDH; brown, ALDH2.

The conformation of the complete NAD$^+$ (chain A of P1 structure) is unusual in that the nicotinamide riboside group is shifted by 3 – 4 Å out of the active site compared to the normal (a.k.a. hydride transfer) conformation (Figure 3.4B). For example, a structure of betaine ALDH shows the hydride transfer conformation (Figure 3.4B). The retracted pose in our structure is similar to the hydrolysis conformation described by the Hurley group in their studies of ALDH2 [15] (Figure 3.4B). A key difference between the hydride transfer and hydrolysis conformations is that the former makes two hydrogen bonds to a conserved glutamate residue, whereas the latter makes only one (Figure 3.4B). Note the NAD$^+$ in our structure exhibits this feature of the hydrolysis conformation.

Conformations of the αE-βE Region of the Rossmann Fold and the Inter-domain Linker

The αE-βE region of the Rossmann fold (residues 232 – 258) exhibits a well-defined conformation in the P1 NAD$^+$ complex, in contrast to the C222 NAD$^+$-complex and apo enzyme structures in which these residues are disordered. The electron density is strong and continuous for the main chain, and most of the side chains also have clear electron density indicating their conformations (Figure 3.5A). Residues 234 – 247 form an α-helix, and residues 251 – 254 form a β-strand (Figure 3.5A); these are the final two secondary structural elements of the Rossmann dinucleotide binding fold. Comparison to betaine ALDH shows that the αE-βE region of P1 ALDH9A1-NAD$^+$ adopts the canonical secondary and tertiary structure of the ALDH superfamily (Figure 3.5B). Furthermore, the αE helices of the domain-swapped dimer interact across a molecular two-fold axis, as in other ALDHs (Figure 3.5C).
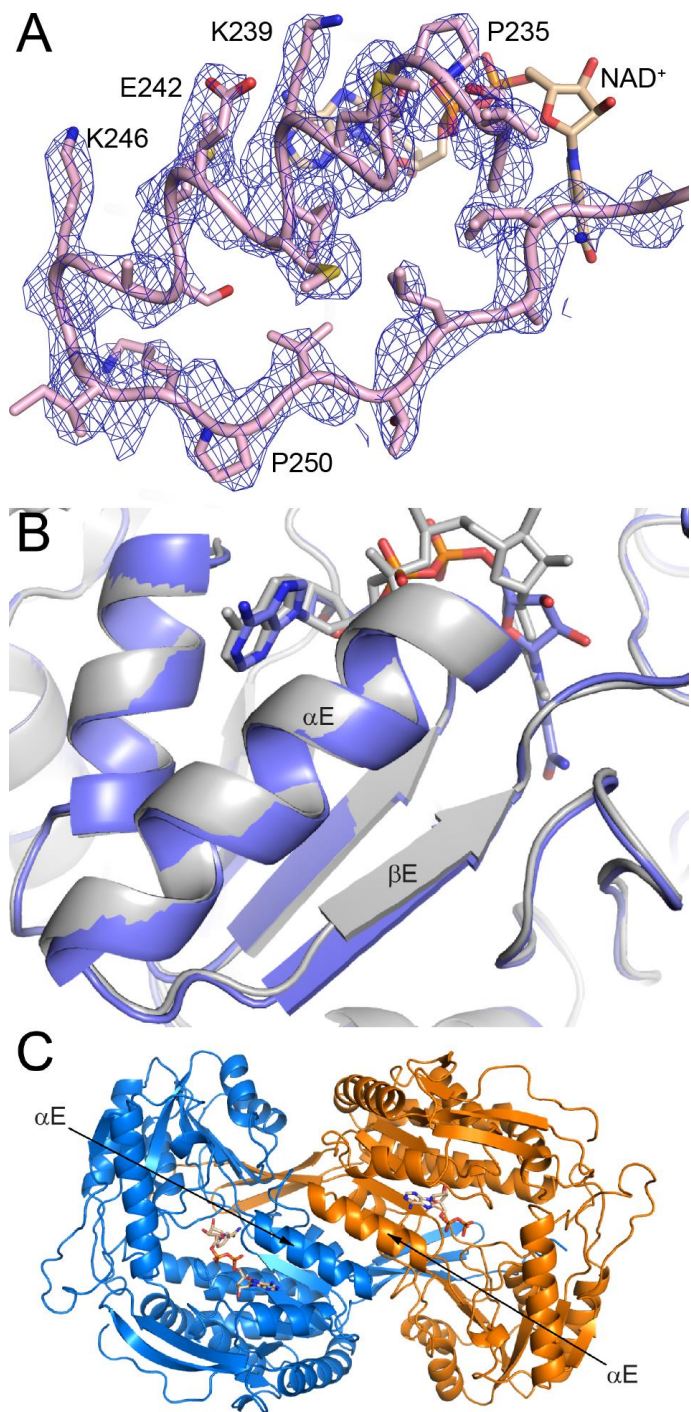
Figure 3.5 Conformation of αE-βE in the *P*1 ALDH9A1-NAD⁺ structure. (A) Electron density for αE-βE. The cage represents a refined 2F_o-F_c map (1σ). (B) Comparison of the αE-βE regions of *P*1 ALDH9A1-NAD⁺ (gray, PDB ID: 6VR6) and betaine ALDH (blue, PDB: 1BPW). (C) The domain-swapped dimer of ALDH9A1, viewed down the 2-fold axis that passes between the αE helices of adjacent protomers. The two protomers have different colors.

The electron density maps also clearly defined the conformation of the inter-domain linker in the $P1$ NAD$^+$ complex (residues 451 – 470) (Figure 3.6A). This section of the polypeptide chain connects the catalytic and oligomerization domains, and forms part of the active site. The linker in our $P1$ structure adopts the β-hairpin structure typical of the ALDH superfamily, as shown by the comparison with betaine ALDH (Figure 3.6B). Several interactions within the β-hairpin stabilize its conformation, including three main-chain hydrogen bonds and charged-hydrogen bonds between Arg467 and the carbonyls of Gly466 and Gly459. These interactions are also present in betaine ALDH. The canonical β-hairpin is very different from the S-shaped conformation in apo ALDH9A1 (Figure 3.6C). We note that the electron density for the inter-domain linker in the $C222$ NAD$^+$ complex is weak and discontinuous, but nevertheless consistent with the S-conformation.

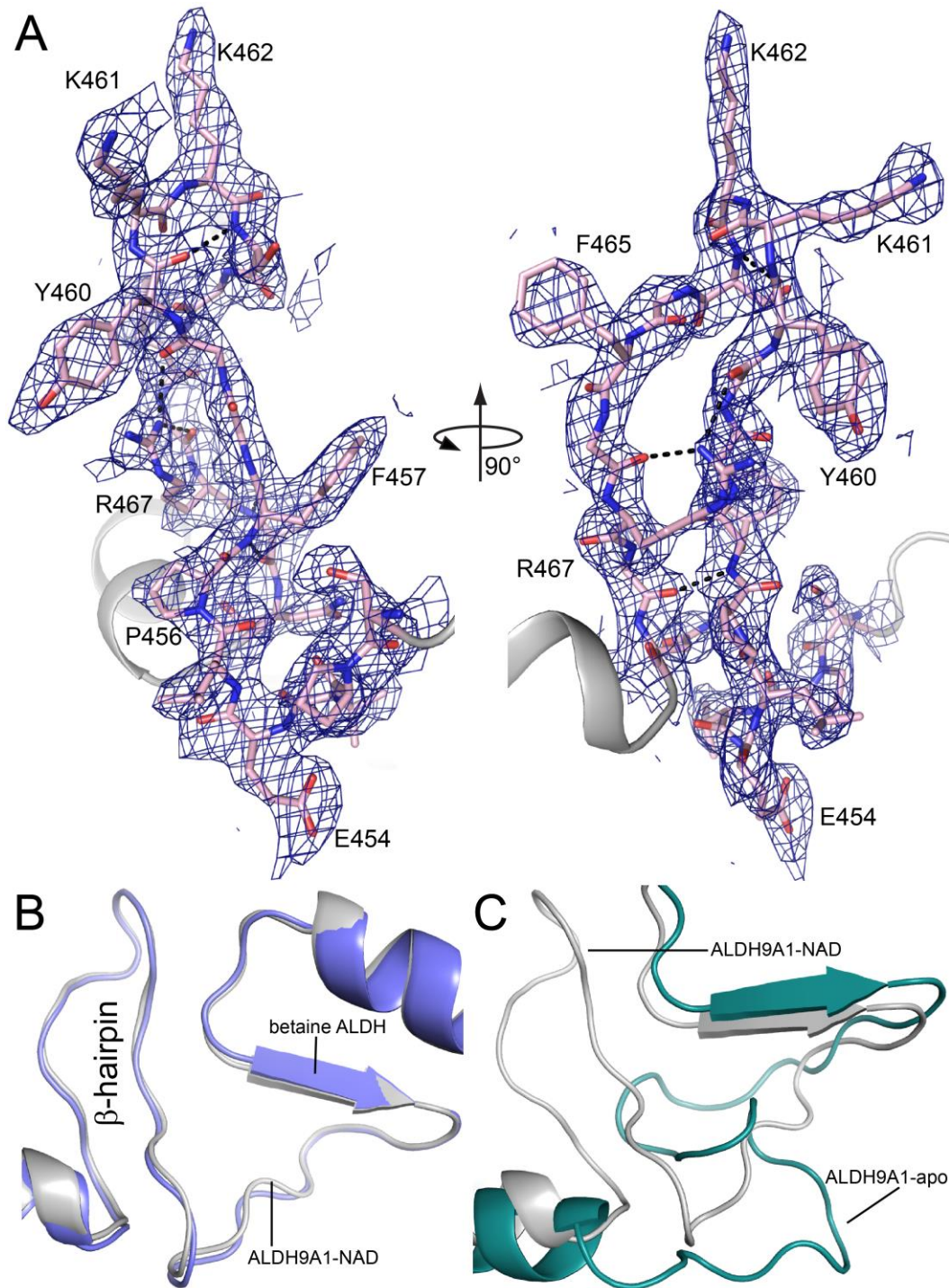**Figure 3.6.** Conformation of the inter-domain linker in the *P*1 ALDH9A1-NAD$^+$ structure. (A) Two views of electron density for the linker (refined $2F_o$-F_c map at 1σ). (B) Comparison of the inter-domain linkers of *P*1 ALDH9A1-NAD$^+$ (gray, PDB ID: 6VR6) and betaine ALDH (blue, PDB: 1BPW). (C) Comparison of the inter-domain linkers of *P*1 ALDH9A1-NAD$^+$ (gray, PDB ID: 6VR6) and apo ALDH9A1 (green, PDB ID: 6QAP).

The ordering of αE-βE and the remodeling of the inter-domain linker result in the formation of new tertiary and quaternary structural interactions (Figure 3.7). For example, βE makes three main chain hydrogen bonds with the inter-domain linker to form a two-stranded anti-parallel β-sheet structure (Figure 3.7A). And the tip of the inter-domain linker forms several hydrogen bonds with the catalytic domain. The αE helices of adjacent protomers pack together to form a new dimer interface (Figure 3.7B). This interface involves the nonpolar face of αE, which includes Met238, Met241, Ala245, and Ile248. Also, Phe465 of the linker intercalates between the βE strands of two adjacent protomers, forming intermolecular nonpolar contacts with Ile248 and Pro250 (Figure 3.7B). In summary, the isomerization of the αE-βE region and the inter-domain linker into their canonical conformations results in the formation of many noncovalent interactions that stabilize the active form of ALDH9A1.
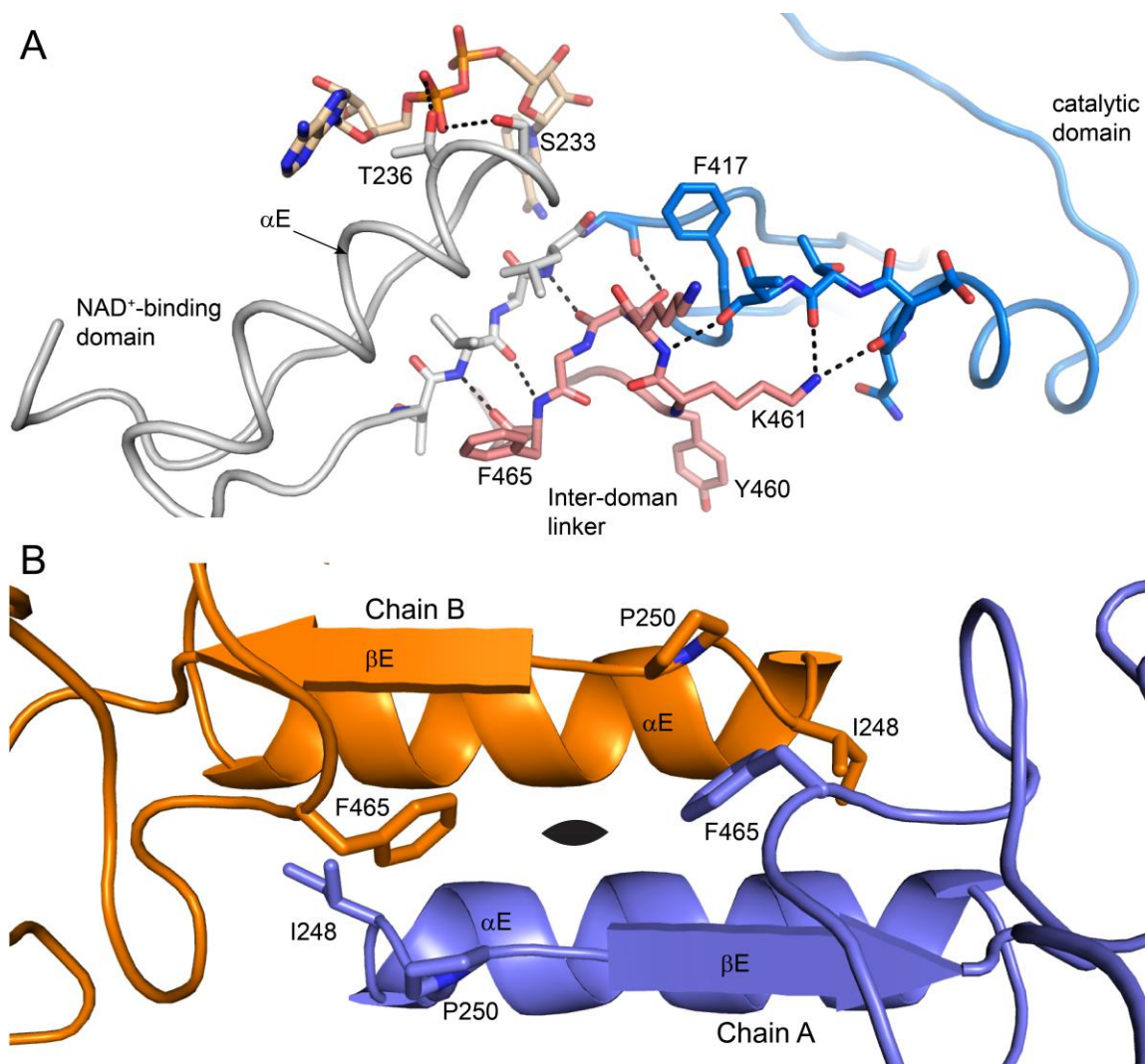
Figure 3.7. Interactions involving the αE-βE region and the inter-domain linker in the *P*1 ALDH9A1 structure. (A) Interaction of the inter-domain linker (pink) with αE-βE (gray) and the catalytic domain (blue). (B) The dimer interface formed by the αE-βE regions of adjacent protomers. The black oval denotes the 2-fold axis of the dimer.

Oligomeric Structure of ALDH9A1

Sedimentation velocity analytical ultracentrifugation was used to determine the oligomeric state of ALDH9A1 in solution. The distribution of apparent sedimentation coefficients, c(s), of apo (NAD$^+$-free) His-ALDH9A1 at 3 mg mL^{-1} revealed a single major peak at 6.8 S (Figure 3.8). Likewise, the distribution of molecular masses, c(M), revealed a single major peak corresponding to a molecular mass of 223 kDa, which is within 1% of

the molecular mass of the His-ALDH9A1 tetramer (theoretical molecular mass: 224.5 kDa). Thus, sedimentation velocity suggests apo His-ALDH9A1 is tetrameric in solution at 3 mg mL^{-1}. This result is consistent with the observation of the classic dimer-of-dimers ALDH tetramer in both of our structures, as well as the crystal forms reported by Koncitikova *et al.* [7].
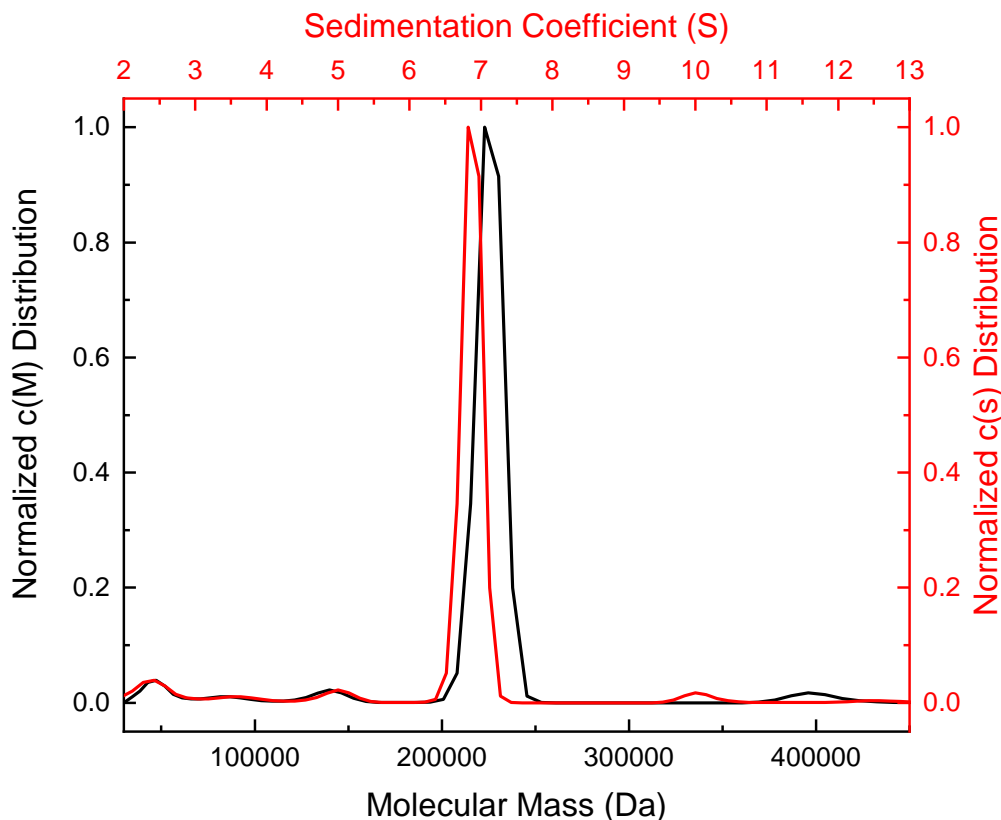
Figure 3.8. Sedimentation velocity analysis of apo ALDH9A1. The graph shows the distribution of molecular masses, c(M), (black curve), and the distribution of sedimentation coefficients, c(s) (red curve).

The in-solution quaternary structure of ALDH9A1 was determined using SAXS. Static SAXS data were collected at three nominal protein concentrations in the range of 1 – 5 mg mL^{-1}, in the presence of 1 mM NAD$^+$ (Figure 3.9A) The average R_g from Guinier analysis of 37.9 ± 0.3 Å (Table 3.3) is in agreement with the SAXS R_g of 38 Å for the ALDH7A1 tetramer [18,21] and the R_g of 35.5 Å calculated from the ALDH9A1 crystallographic

tetramer. For each concentration, the theoretical SAXS curve generated from the crystallographic tetramer provided a reasonably good fit to the experimental data (Figure 3.9A). Imposition of a multi-body fit to simulate a mixture of dimer and tetramer did not statistically improve the fits to the experimental data. Finally, the *ab initio* shape reconstruction shows good agreement with the tetramer (Figure 3.9B). Overall, these data suggest ALDH9A1 is primarily tetrameric under the conditions used, consistent with the sedimentation velocity data.

Table 3.3.
ALDH9A1 solution structural parameters from SAXS

Concentration (mg mL^{-1})	1.25	2.5	5
Guinier analysis			
Points used	1-47	1-48	1-48
qR_g range	0.31-1.28	0.31-1.30	0.30-1.29
R_g (Å)	38.1 ± 0.8	37.9 ± 0.2	37.6 ± 0.1
$I(0)$	8.5 ± 0.1	16.68 ± 0.04	33.39 ± 0.06
$P(r)$ analysis			
D_{max} (Å)	109	108	105
R_g (Å)	37.0 ± 0.5	37.3 ± 0.9	37.0 ± 0.6
Porod volume (Å^3)	252,000	237,000	238,000
Mass from SAXSMoW (kDa)	203	182	167
FoXS χ^2	0.24	1.0	2.1
SASBDB Code	SASDHV5	SASDHW5	SASDHX5
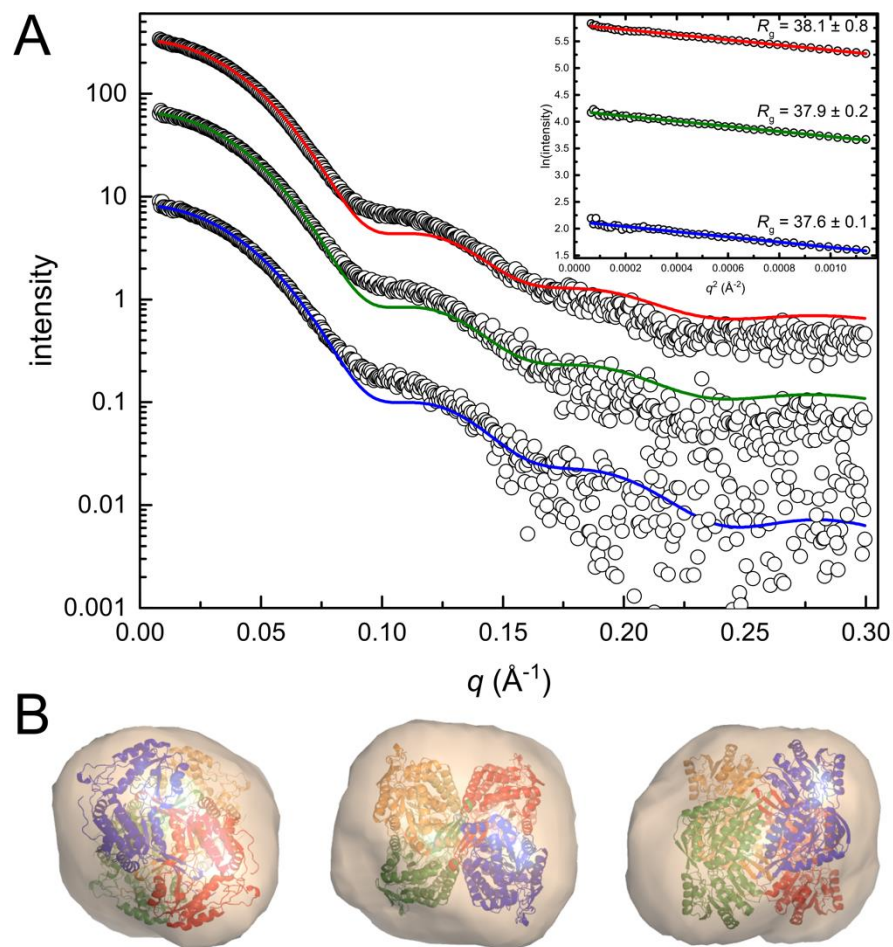
Figure 3.9. SAXS analysis of ALDH9A1 in the presence of 1 mM NAD[+]. (A) SAXS experimental data (open circles) and the corresponding FoXS single-body fit to the ALDH9A1 crystallographic tetramer (blue, green, and red curves) at three nominal protein concentrations. (B) Three views of the electron density *ab initio* shape reconstruction generated using DENSSWeb. The envelope is contoured to 1.0σ with the crystallographic tetramer of ALDH9A1 fit into the map using the "fit in map" utility of Chimera [22,23].

The quaternary structure of apo ALDH9A1 in solution was confirmed with negative stain electron microscopy (Figure 3.10). The micrographs exhibited good contrast and particle separation (Figure 3.10A). The 2D class averages calculated from 7430 particles are shown in Figure 3.10B, and for comparison, views of the ALDH9A1 tetramer and dimer are shown in Figure 3.10C. Most of the 2D classes appear to resemble the tetramer viewed down one of its three mutually orthogonal 2-fold axes (classes 1, 4, 6-10). In these classes, which account for ~80% of the particles, the four subunits of the tetramer and a

two-fold symmetry axis perpendicular to the page are apparent. Classes 3 and 5 (16% of the particles) appear to represent the tetramer rotated by 45° around a 2-fold axis. Class 2 may represent a dimer viewed down its 2-fold axis. Note the dimer represents only 4% of the particles. Thus, apo ALDH9A1 is predominantly tetrameric under the conditions used for negative stain EM. In summary, the conclusion from multiple biophysical measurements – X-ray crystallography, sedimentation velocity, SAXS, and electron microscopy – is that ALDH9A1 exists in solution primarily as the classic ALDH superfamily dimer-of-dimers tetramer under the conditions used.
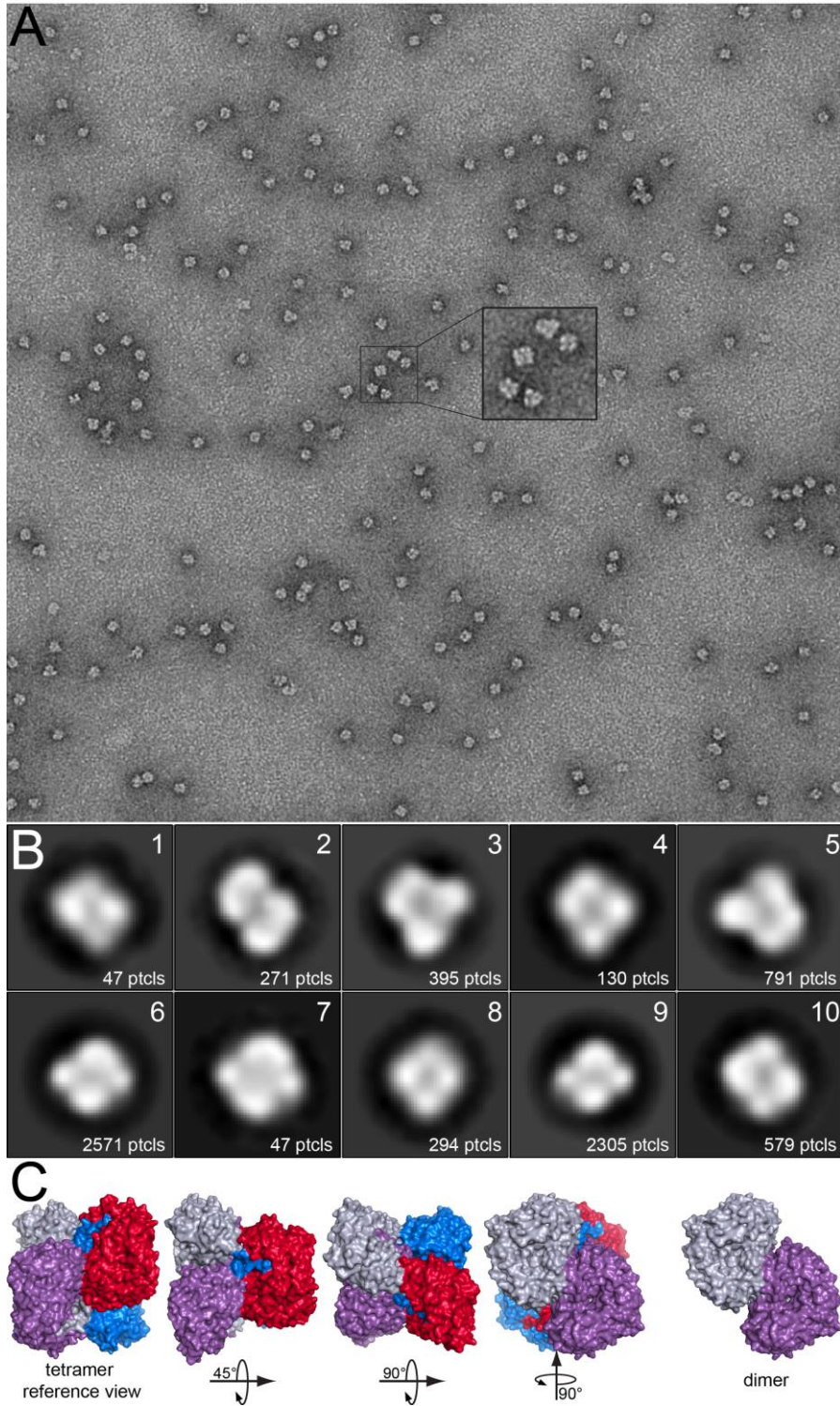
Figure 3.10. Negative-stain EM analysis of apo ALDH9A1. (A) Representative micrograph. (B) 2D class averages calculated from 7430 particles. (C) Surface representations of the ALDH9A1 tetramer viewed in various orientations relative to a reference orientation, and a dimer viewed down its 2-fold axis.

3.3. Discussion

The new structures described here provide insight into the mechanism by which ALDH9A1 is activated for catalysis. The inactive conformation first observed in the apo enzyme is characterized by disorder of αE-βE and a novel S-shaped inter-domain linker [7]. Because these atypical features are also present in our C222 NAD$^+$-complex structure, it appears that NAD$^+$-binding is not sufficient to activate the enzyme. Instead, we found that treatment of ALDH9A1 with DEAB and the cofactor NAD$^+$ apparently induced αE-βE and the inter-domain linker to assume their canonical conformations, as observed in the P1 structure. We conclude that the presence of an aldehyde substrate and NAD$^+$ promotes isomerization of the enzyme into the active conformation.

The disorder-order transitions discussed here are not unique to ALDH9A1. Disorder in the interdomain linker has also been observed in ALDH1A2 (PDB ID: 1BI9) [24], ALDH4A1 (PDB IDs: 4OE5/6) [17], and ALDH7A1 (PDB IDs: 4ZVX/Y) [18]. In ALDH1A2, ordering of the linker is thought to be related to aldehyde binding, similar to ALDH9A1 [25]. In ALDH4A1 and ALDH7A1, no clear connection between ligand binding and disorder-order transitions has been established. Disorder in αE-βE is rarer. The best example is a naturally-occurring disease variant of ALDH2 (ALDH2*2) in which Glu487 is mutated to Lys (PDB ID: 1ZUM) [26]. Glu487 in wild-type ALDH2 forms intersubunit ion pairs with two arginine residues, one located on βE (Arg264), and another in the interdomain linker (Arg475). The introduction of Lys487 obviously disrupts these stabilizing interactions, resulting in substantial disorder in both αE-βE and the interdomain linker. ALDH9A1 does not have an analogous ionic interaction network, due to the substitution of a glutamine for Glu487, and perhaps more importantly, a proline in place

of Arg264. The absence of these interactions may explain the propensity for disorder in ALDH9A1. On the other hand, the close ALDH9A1 homolog, betaine ALDH, also lacks the cross-subunit ion pairs, yet it exhibits canonical αE-βE and interdomain linker conformations, even in the absence of bound ligands (PDB ID: 1A4S). Although the mechanisms of structural ordering in ALDHs remain to be determined, the observation of large swaths of disorder in the active sites of multiple ALDHs suggests disorder-order transitions may be an inherent feature of the ALDH superfamily fold and important for catalytic function.

The NAD^+ in our P1 structure unexpectedly has a retracted pose (Figure 3.4). Two conformations of the cofactor bound to ALDHs have been described: "hydride transfer" and "hydrolysis". These designations refer to steps in the catalytic mechanism. NAD^+ tends to bind in the hydride transfer conformation, which positions the hydride acceptor of NAD^+ (C4 atom) ~3 Å from the S-atom of the catalytic cysteine, ready to accept a hydride ion from the thiohemiacetal intermediate. Following hydride transfer and formation of the acyl-enzyme intermediate, the newly formed NADH retracts from the active site into the hydrolysis conformation. This movement allows room for a water molecule to enter the active site for hydrolysis of the acyl-enzyme intermediate. NAD^+ in crystal structures typically adopts the hydride transfer pose; however, NAD^+ has also been observed in the hydrolysis pose in ALDH1 (PDB ID: 1BXS) [27] and ALDH2 (PDB ID: 1O00) [15]. Also, mutation of the conserved glutamate that hydrogen bonds to the nicotinamide ribose in ALDH7A1 causes NAD^+ to switch to a hydrolysis-like pose (PDB ID: 6O4L) [28]. NADH bound to ALDH2 adopts the mechanistically-expected hydrolysis pose (PDB ID: 1O02) [15]. On the other hand, a structure of ALDH7A1 shows NADH in the hydride transfer

pose (PDB ID: 2J6L) [29]. Altogether, the crystal structures of ALDHs provide evidence of the flexibility of the bound cofactor, which is essential for catalysis. Why NAD^+ in our structure is retracted into a hydrolysis-like conformation is unclear. It is possible that covalent modification of the catalytic cysteine by DEAB, which is suggested by the electron density but could not be modeled with certainty, prevents NAD^+ from moving into the hydride transfer conformation.

We showed that ALDH9A1 forms the classic ALDH superfamily dimer-of-dimers tetramer in solution. This is not surprising considering that the closest sequence homologs of ALDH9A1 in the PDB also form the tetramer (at least *in crystallo*), including betaine ALDHs (40 – 70% identical [8,30]) and ALDH2 (43% identical).

ALDH9A1 may be added to the list of ALDHs that are inhibited by DEAB. This information is pertinent to studies that use the ALDEFLUOR flow cytometry assay to detect cancer stem cells based on ALDH activity (ALDH-bright cells) [31–35]. DEAB is used in the control arm of the ALDEFLUOR assay, and any ALDHs that are reversibly inhibited or covalently inactivated by DEAB will contribute to a cell being labeled as ALDH-bright and thus classified as a cancer stem cell. DEAB was once thought to be specific for ALDH1A1, which led researchers to assume that the ALDEFLUOR assay was specific for the detection of ALDH1A1. Recent work has proven this assumption incorrect, as DEAB inhibits or inactivates several ALDHs including ALDH1A1, ALDH1A3, ALDH1B1, ALDH5A1, ALDH1A2, ALDH2, and ALDH7A1 [9,10,36].

Hurley's group proposed a general mechanism for the inhibition of ALDHs by DEAB. They proposed that DEAB functions as a substrate through the hydride transfer step, but that the acyl-enzyme intermediate is protected against hydrolysis by resonance stabilization

[9]. The extent of resonance stabilization, and therefore the potency of inhibition, depends on the noncovalent interactions between the acyl-enzyme and the active site, which varies among ALDH isozymes. For example, the active sites of ALDH1A2, ALDH2 and ALDH7A1 apparently provide exceptional stabilization, since DEAB is a covalent irreversible inhibitor of these enzymes [9,10]. For ALDH1A1, ALDH1A3, ALDH1B1, and ALDH5A1, DEAB inhibits by virtue of being a slow substrate; the active sites of these enzymes apparently provide less efficient stabilization of the acyl-enzyme than in ALDH1A2, ALDH2 and ALDH7A1.

We found that the inhibition of ALDH9A1 by DEAB is time-dependent, consistent with covalent inactivation. In contrast to ALDH2 and ALDH7A1, the inhibition was diminished by jump-dilution and abrogated by removal of excess DEAB, consistent with a covalent, reversible mechanism. The irreversible inactivation of ALDH2 and ALDH7A1 by DEAB in the presence of NAD^+ is due to formation of the acyl-enzyme [9,10]. In principle, covalent reversible inhibition of an ALDH can arise by slow hydrolysis of the acyl-enzyme intermediate to generate active enzyme and diethylaminobenzoic acid. This is not likely to be the case for ALDH9A1, given the rapid return of enzyme activity upon removal of DEAB coupled with the fact that we find no evidence that DEAB serves as a substrate for this enzyme (see Figure 3.1A). A more plausible mechanism is formation of a reversible thiohemiacetal covalent intermediate (Scheme 3.2A) that fails to progress through the hydride transfer step of the normal catalytic mechanism that would generate the acyl-enzyme intermediate. The covalent, reversible inhibition due to thiohemiacetal formation with an active site cysteine residue proposed here finds a direct mechanistic analogy in the inhibition cysteine proteases by aldehydes [37–39] and is mechanistically

related to the inhibition of viral proteases by ketoamide compounds that form a reversible, covalent tetrahedral species with the catalytic serine [40–42]. This mechanism (Scheme 3.2A) could also explain the covalent, reversible inhibition of ALDH1A2 by DEAB in the absence of NAD^+ [9].

The progress curves in Figure 1B and the reversibility of inhibition are also consistent with slow, tight-binding inhibition (Scheme 3.2B) [11]. In this mechanism, inhibitor binding could involve the rapid formation of an initial collision complex (E·I) that subsequently undergoes a slow isomerization reaction to E·I* (with or without formation of the thiohemiacetal covalent intermediate). Presumably the initial encounter would involve the apo, inactive enzyme, followed by an isomerization step resulting in an enzyme conformation similar to the P1 structure described here. Isomerization between these two conformations likely would be quite slow (seconds to minutes), as it involves very large structural changes, including folding of the αE-βE region and reorganization of the interdomain linker.
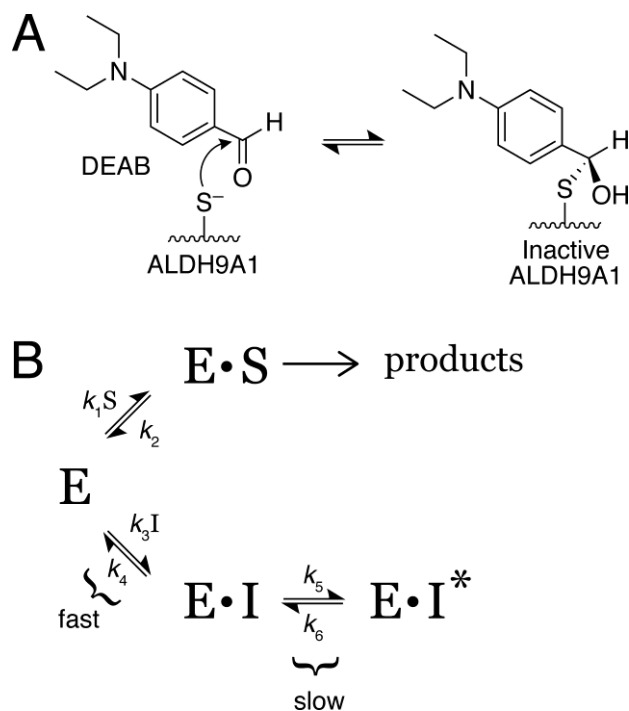
Scheme 3.2. Possible mechanisms of inactivation/inhibition of ALDH9A1 by DEAB.

3.4. Materials and methods

Expression Plasmids

A synthetic gene encoding human ALDH9A1 with codons optimized for expression in *Escherichia coli* was purchased from Genscript. The gene was subcloned by Genscript into plasmid pET-24b(+) between NdeI/XhoI restriction sites. The expressed protein contains an N-terminal hexahistidine tag and tobacco etch virus protease cleavage site.

Removal of the His-tag from the protein expressed from the aforementioned pET-24b(+) construct was problematic, so another expression plasmid encoding ALDH9A1 fused to SUMO was created. The coding sequence of human ALDH9A1 was PCR amplified from the pET24b(+) construct, gel-purified, and cloned into pET-SUMO using *BamHI* and *XhoI* restriction sites by visiting professor, Dr. Insaf Qureshi. Positive clones were identified by colony PCR followed by DNA sequencing using T7 forward and reverse primers.

Purification of SUMO-ALDH9A1 and His-ALDH9A1

To purify SUMO-tagged ALDH9A1, the fusion protein was over-expressed in *E. coli* BL21 (DE3) with induction by 0.25 mM IPTG at 18°C for 20 hrs. Cells were collected by centrifugation and resuspended in buffer A (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole and 5% glycerol (v/v) containing 0.1% Triton X-100. Cells were lysed by sonication and then centrifuged at 20,000 rpm in a SS-34 rotor for 30 min at 4ºC to remove the remaining insoluble material. The supernatant was subjected to immobilized metal-affinity chromatography (IMAC) on a HisTrap HP column (GE Healthcare) charged with nickel ion and equilibrated with buffer A and unbound proteins were washed sequentially with ten column volumes each of buffer A with imidazole concentrations of 40, 100 and 150 mM. The His-tagged proteins were eluted with buffer A containing 200 mM imidazole. To remove the N-terminal SUMO-tag, His-tagged SUMO protease (purified separately by IMAC chromatography) was added to IMAC-purified His-SUMO-ALDH9A1 in a 1:100 ratio and subjected to dialysis for 12 – 16 h at 4°C in 25 mM Tris-HCl pH 8.0, 100 mM NaCl, 5% (v/v) glycerol and 0.5% Tris(2-carboxyethyl)phosphine (TCEP). The digested protein sample was then passed through the IMAC column, which retained the cleaved SUMO tag, SUMO protease, and uncleaved fusion protein. The pure, untagged ALDH9A1 protein was collected in the flow through and then protein quality was analyzed with SDS-PAGE. Subsequently, ALDH9A1 was concentrated using a 50 kDa MilliporeSigma Amicon Ultra centrifugal concentrator and purified further using size-exclusion chromatography (SEC) on a HiLoad 16/60 Superdex 200 column with a buffer containing 50 mM Tris-HCl pH 8.0, 600 mM NaCl, 5% (v/v) glycerol and 0.5% TCEP. The purity of the recombinant ALDH9A1 was verified by SDS-PAGE. Protein concentration was

determined by absorbance using an estimated extinction coefficient calculated by ProtParam [43].

His-tagged ALDH9A1 was over-expressed from the pET-24b(+) vector in *E. coli* BL21 (DE3) at OD ~ 0.6 with 0.25 mM IPTG at 18°C for 18 – 21 hrs. Cells were harvested by centrifugation and resuspended in buffer A (50 mM HEPES pH 8.0, 300 mM NaCl, 10 mM imidazole, and 5% glycerol) containing 0.1% Triton X-100 and 0.4 mM PMSF. Cells were lysed by sonication and centrifuged at 16,500 rpm in a SS-34 rotor for 1 h at 4°C. The supernatant was subjected to IMAC using a HisTrap HP column (GE Healthcare) charged with nickel ion and equilibrated in buffer A. After washing the column with 100 mM imidazole, His-ALDH9A1 was eluted with 200 mM imidazole. The enzyme was concentrated with a 50 kDa MilliporeSigma Amicon Ultra centrifugal concentrator and further purified by SEC on a Superdex 200 column in the presence of 50 mM HEPES pH 8.0, 100 mM NaCl, 5% (v/v) glycerol, and 0.5 mM dithiothreitol (DTT). The purity of His-ALDH9A1 was verified with SDS-PAGE and protein concentration was determined by A_{280} using a NanoDrop 2000c (ThermoFisher) and confirmed via bicinchoninic acid assay (Pierce).

Kinetics of Enzyme Activity and Inhibition

All kinetic measurements were performed using His-tag-free ALDH9A1. The steady state kinetic parameters of ALDH9A1 for the aldehyde substrate hexanal at saturating NAD^+ concentration were determined. These assays monitored NADH production (340 nm) at 27°C in an Epoch 2 plate reader (BioTek, Winooski, VT, USA). The assay buffer contained 100 mM sodium pyrophosphate pH 8.0. The aldehyde substrate concentration

was varied in the range of 2.5 – 150 µM, while the NAD$^+$ concentration was fixed at 1.5 mM. The untagged enzyme was used at 500 nM. Each measurement was performed in triplicate. Initial rates were estimated by linear regression of the first five minutes of the assay. The absorbance at 340 nm was converted to NADH concentration using an extinction coefficient of 6220 M^{-1}cm^{-1}. Kinetic parameters were obtained by fitting the initial rate data to the Michaelis-Menten equation using Origin version 2019 (Figure S3.1).

DEAB is a substrate for some ALDHs, so we first tested this possibility for ALDH9A1 using the strategy of Hurley's group [9]. In this experiment, ALDH9A1 is combined with DEAB and NAD$^+$, and then the absorbance at 360 nm and 300 nm are monitored. A decrease at 360 nm indicates the consumption of DEAB, while an increase at 300 nm indicates the production of the carboxylic acid product, diethylaminobenzoic acid. For these experiments, the final assay mixture (total volume of 1 mL) contained 0.5 µM ALDH9A1, 30 µM DEAB, and 1.5 mM NAD$^+$ in a buffer containing 100 mM sodium pyrophosphate pH 8.0 and 0.1% DMSO. The reaction was initiated by the addition of enzyme. The absorbance was measured for 240 minutes using the cuvette port of a NanoDrop 2000c (ThermoFisher). Scans covering the wavelength range of 270 – 460 nm were recorded every 10 minutes.

The time-dependent inhibition of ALDH9A1 by DEAB was studied using a progress curve approach [44,45]. Progress curve experiments were performed in a BioTek Epoch 2 plate reader in duplicate at 27°C with an assay mixture consisting of 0.5 µM ALDH9A1, 2 mM hexanal, 2.5 mM NAD$^+$, 2 – 5.0 µM DEAB, 0.04% (v/v) DMSO, and 100 mM sodium pyrophosphate pH 8.0. The total reaction assay volume was 200 µL. The reaction was initiated by the addition of an aliquot of an enzyme- NAD$^+$ solution and then monitored

by observing the production of NADH at 340 nm. The progress curves measured at different DEAB concentrations were analyzed using global fitting to combined Equations 1a and 1b [11].

$$Absorbance = v_s t + \frac{v_i - v_s}{k}[1 - e^{-kt}] + a_0 \quad (1a)$$

$$k = k_6 + k_5 \frac{[I]/K_I}{\left(1 + \frac{[S]}{K_m} + \frac{[I]}{K_I}\right)} \quad (1b)$$

Equations 1a and 1b describe the inhibition model in Scheme 3.2B (see equations 6 and 9 of Morrison and Walsh [11]). The parameters v_i and v_s are the initial and steady-state velocities, t is time, a_o is the initial absorbance offset parameter, and k the apparent first-order rate constant for the establishment of the equilibrium between E·I and E·I*. In Equation 1b, k_5 and k_6 are the forward and reverse rate constants for the isomerization of E·I to E·I*, respectively; K_I is dissociation constant for the E·I complex; [I] is the inhibitor concentration; [S] is the hexanal substrate concentration; and K_m is the Michaelis constant for hexanal, which was estimated to be 6.3 μM in a separate experiment performed in the absence of DEAB (Figure S3.1). Global fitting was done with Origin 2019, with k_5, k_6, and K_I considered to be global parameters (shared by all data sets), while K_m, [S], and [I] were fixed at their known values. During fitting, k_6 refined to a very small value (2×10^{-5} min^{-1}) compared to k_5 (0.2 min^{-1}); however, the uncertainty of k_6 was large, which prevented convergence. Therefore, k_6 was fixed at 2×10^{-5} min^{-1} to aid convergence of the fitting algorithm.

Crystallization of ALDH9A1-NAD$^+$ in Space Group P1

Prior to crystallization, DEAB and NAD$^+$ were added to the protein stock solution (6

mg mL^{-1} tag-free ALDH9A1; 50 mM Tris-HCL pH 8.0, 600 mM NaCl, 5% glycerol, and 0.5% TCEP) to final concentrations of 5 mM and 10 mM, respectively. Initial crystal screening was performed in 96-well Swissci 2 drop MRC Crystallization plates with Hampton Index at 8°C. Promising crystals were obtained in a condition containing 0.2 M NaCl, 0.1 M Bis-Tris pH 6.5, and 25% (w/v) PEG 3350. Further improvement was obtained using an additive screen approach. In this approach, the reservoir consisted of equal volumes of the aforementioned base condition (0.2 M NaCl, 0.1 M Bis-Tris pH 6.5, and 25% (w/v) PEG 3350) and an additional crystal screen condition. Improvement in the crystal quality was observed when using Hampton Index condition G8 as the additive (0.2 M ammonium acetate, 0.1 M HEPES pH 9.5, and 25% (w/v) PEG 3350). Thus, the final reservoir solution contained 0.1 M NaCl, 0.05 M Bis-Tris 6.5, 0.1 M ammonium acetate, 0.05 M HEPES pH 9.5, and 25% (w/v) PEG 3350. The crystals were cryoprotected with 0.1 M NaCl, 0.05 M Bis-Tris pH 6.5, 0.1 M ammonium acetate, 0.05 M HEPES pH 9.5, 25% (w/v) PEG 3350, 20% (v/v) hexylene glycol, 2.5 mM DEAB and 5 mM NAD$^+$. Although the presence of DEAB in the active site was not obvious from the electron density, its inclusion in the crystallization protocol apparently facilitated growth of a new crystal form, which revealed the active conformation of the enzyme.

Crystallization of ALDH9A1-NAD$^+$ in Space Group C222

Prior to crystallization, NAD$^+$ was added to the protein stock solution (10 mg mL^{-1} His-tagged ALDH9A1; 50 mM HEPES pH 8.0, 100 mM NaCl, 5% (v/v) glycerol, and 0.5 mM DTT) to a final concentration of 30 mM. A crystal screen was performed in 96-well Swissci 2 drop MRC Crystallization plates at 8°C using Hampton Index with a drop ratio of 0.6/0.4

(enzyme/reservoir). Microseed matrix screening [46,47] was used with a seed stock prepared from previously grown *C*222 crystals. The crystal used for data collection was harvested directly from Hampton Index condition H5 (0.1 M succinic acid pH 7.0, 15% (w/v) PEG 3350) without optimization and cryoprotected with the reservoir supplemented with 20% (v/v) hexylene glycol.

Crystal Structure Determination

X-ray diffraction data from a *P*1 crystal were collected at NECAT beamline ID-E of the Advanced Photon Source using an EIGER 16M detector in shutterless mode. The dataset used for refinement consisted of 700 frames spanning a rotation range of 140°. Diffraction data from a *C*222 crystal were collected at beamline 4.2.2 of the Advanced Photon Source (900 frames, 180° rotation). The data from both beamlines were integrated and scaled with XDS [48], and intensities were merged and converted to amplitudes with AIMLESS [49]. The unit cell dimensions of both crystal forms are listed in Table 3.1. The method of Matthews [50] predicts eight chains in the *P*1 asymmetric unit with 49% solvent, and 2 chains in the *C*222 asymmetric unit with 49% solvent. Data processing statistics are listed in Table 3.1.

Initial phases for the *P*1 structure were calculated using molecular replacement as implemented MOLREP [51]. The search model was derived from a protomer of cod liver betaine ALDH ([8], PDB ID: 1A4S, 70% identity to ALDH9A1). The side chains of the search model were trimmed to the C-γ atom with CHAINSAW [52]. MOLREP returned a solution consisting of eight protein chains arranged as two tetramers. We note that the molecular replacement solutions were not as clear when search models derived from the

structure of apo ALDH9A1 (PDB ID: 6QAP) were used, indicating conformational differences between our structure and apo ALDH9A1.

Initial phases for the *C222* structure were calculated using molecular replacement with Phaser [53]. The search model was derived from a partial structure obtained by refinement against an earlier *C222* data set, which exhibited pseudo-merohedral twinning. Initial phases for the twinned data set were calculated using Phaser with a search model derived from cod liver betaine ALDH (PDB ID: 1A4S).

PHENIX [54] was used for refinement, and Coot [55] was used for model building. The *B*-factor model consisted of an isotropic *B*-factor for each non-hydrogen atoms and one TLS group per chain. Because of the modest resolution, non-crystallographic symmetry restraints were enforced during refinement of both structures. The occupancies of the NAD^+ molecules in both structures were set to 1.0. The structures were validated with MolProbity [56] and the wwPDB validation service [57]. Refinement statistics are listed in Table 3.1.

Analytical Ultracentrifugation

Sedimentation velocity experiments were performed in a Beckman XL-I analytical ultracentrifuge using an An50Ti rotor at 20°C by former postdoctoral associate, Dr. David Korasick. For this analysis, His-tagged ALDH9A1 at 3 mg mL^{-1} protein sample was dialyzed overnight against a buffer containing 50 mM HEPES pH 7.8, 100 mM NaCl, 5% (v/v) glycerol, and 0.5 mM DTT. A two-sector charcoal-Epon sedimentation velocity centerpiece was loaded with reference buffer and protein samples. After an equilibration period of 2 h, the sample was centrifuged at 35,000 rpm for a total of 300 radial scans

spaced at 2 min intervals. Data were acquired using Rayleigh interference optics, which allowed measurement of the actual experimental protein concentration, assuming 1 mg mL^{-1} is equivalent to 3.33 fringes [58]. Scans 10 – 300 were analyzed and the distribution of apparent sedimentation coefficients, c(s), and distribution of apparent molecular masses, c(M), were determined using SEDFIT [59].

Small-angle X-ray Scattering

Shutterless SAXS data collection was performed at beamline 12.3.1 of the Advanced Light Source through the SIBYLS Mail-in High Throughput SAXS program [60]. Prior to SAXS analysis, purified His-tag-free ALDH9A1 was passed over a Superdex 200 10-30 size-exclusion chromatography column in the presence of a buffer containing 50 mM HEPES pH 8.0, 600 mM NaCl, 2% (v/v) glycerol, 1 mM DTT, and 1 mM NAD$^+$. Samples were then supplemented with 1 mM NAD$^+$ and dialyzed overnight at 4°C against a buffer containing 50 mM HEPES pH 8.0, 600 mM NaCl, 2% (v/v) glycerol, 1 mM DTT, and 1 mM NAD$^+$.

SAXS data were collected on a Pilatus detector operating in shutterless mode, writing frames every 0.3 s. Buffer subtracted SAXS curves were averaged using SAXS FrameSlice. PRIMUS [61] was used to inspect the merged data and to derive SAXS parameters. The maximum particle dimension was estimated from calculations of the pair distribution function using GNOM [62] via PRIMUS. The molecular mass was estimated using the SAXSMoW [63]. Theoretical SAXS curves were calculated using FoXS [64]. DENSSWeb [65] was used for *ab initio* electron density determination from solution scattering data.

Negative Stain Electron Microscopy

Negative stain electron microscopy experiments were carried out by former graduate student Dr. Ashley Campbell and former postdoctoral associate Dr. David Korasick. Preparation of negative-stained grids was carried out adapted from the previously described protocol [66]. Briefly, carbon-coated copper grids (Electron Microscopy Sciences, Hatfield, PA) were glow discharged, and a 5 µL drop of His-tagged ALDH9A1 (~0.08 mg mL^{-1}) was applied to the grid, followed by a 2 min incubation at room temperature. Excess protein was removed by blotting with Whatman P4 filter paper. Blotting was immediately followed by two quick washes in water and a quick wash in 0.75% (w/v) uranyl formate (UF; Ted Pella, Reading, CA), blotting with filter paper in between wash steps. A final incubation, by floating the grid for 2 min on a 0.75% UF (w/v) drop, was performed before excess UF was removed from the grid with filter paper. The grid was air-dried. Protein particles were observed using a JEOL JEM 1400 transmission electron microscope (Peabody, MA) at 80 kV.

The data were analyzed with RELION-3 [67]. Contrast transfer function correction was performed using CTFFind 4.1 using a spherical aberration of 3.2 mm, amplitude contrast of 0.5, and magnified pixel size of 3.51 Å. The resolution range used for fitting was 5–30 Å. The defocus range was 5000–50000 Å with a defocus step of 500 Å and an astigmatism of 200 Å. An initial particle set was obtained using template-free auto-picking (Laplacian-of-Gaussian-based) with a minimum diameter of 50 Å, maximum diameter of 120 Å, and threshold of 3.5. This calculation generated 10373 particles. Low quality particles (e.g., random noise, overlapped protein molecules) were eliminated during several rounds of

iterative 2D class averaging and selection, resulting in a set of 7430 particles that was used

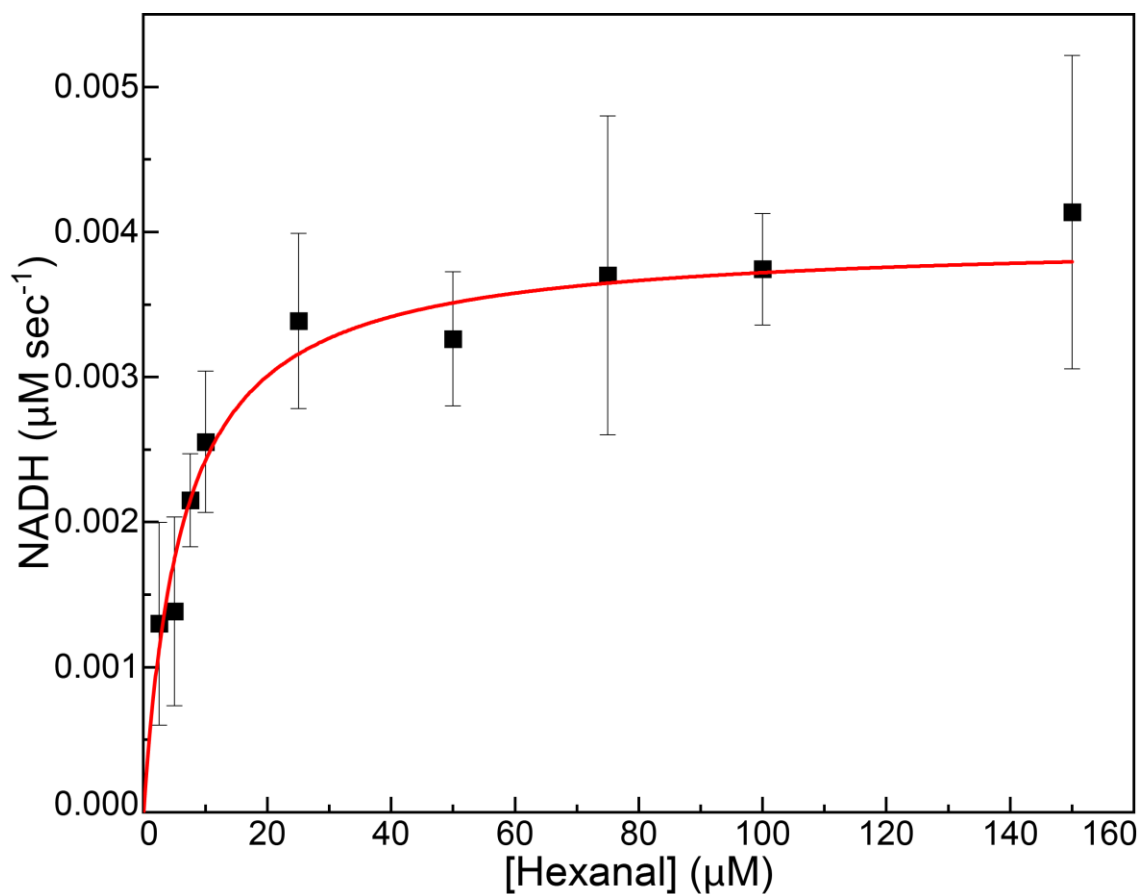for the final 2D class average calculation presented here.

Figure S3.1. Steady-state kinetics of ALDH9A1 with hexanal as the variable substrate and NAD$^+$ fixed at 1.5 mM. The squares represent experimental data. The red curve is the fit to the Michaelis-Menten equation. The kinetic parameters estimated from fitting are k_{cat}= 0.0080 ± 0.0003 s^{-1} and K_m = 6.3 ± 0.8 µM.
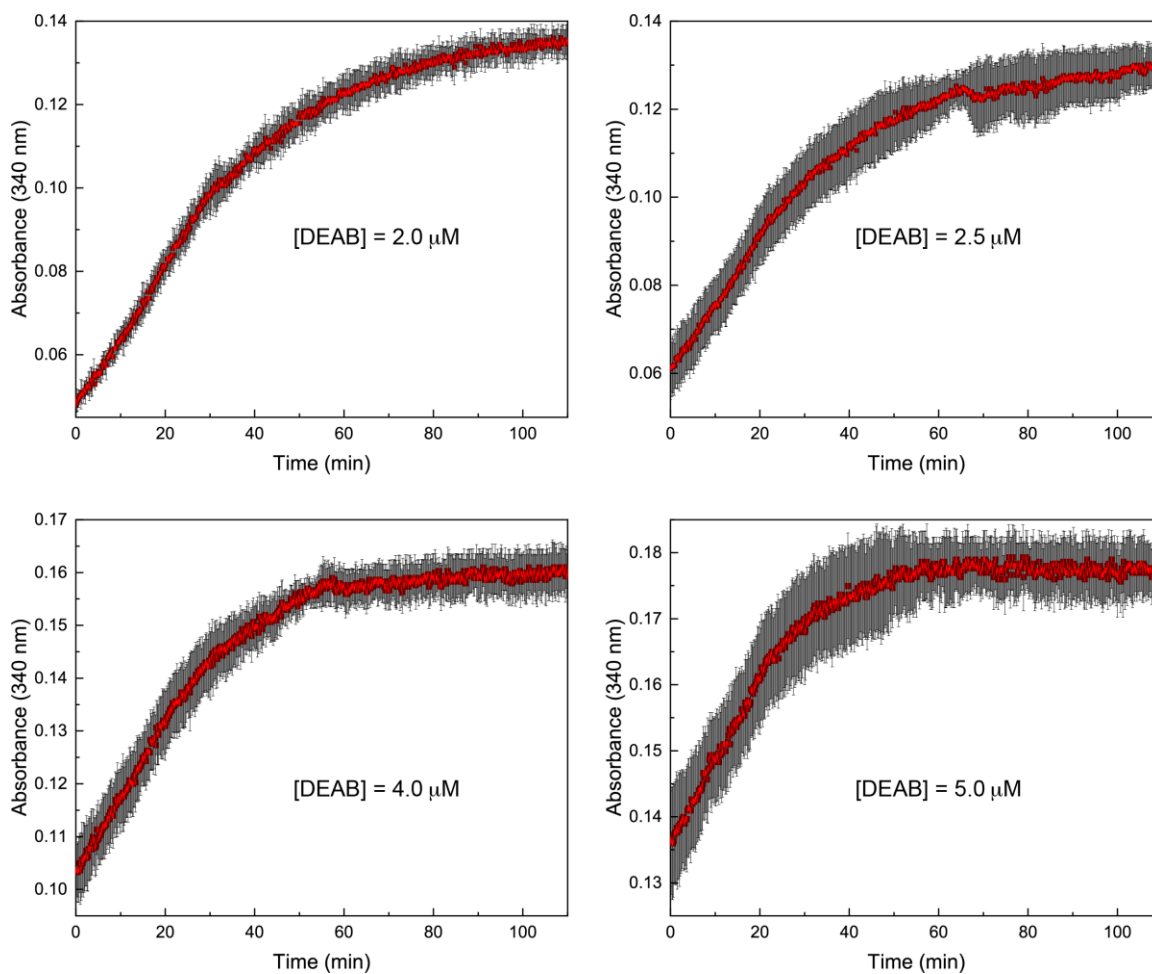
Figure S3.2. Progress curves for the inhibition of ALDH9A1 by DEAB, with error bars. The red symbols are the experimental progress curves obtained by monitoring NADH production from a mixture of ALDH9A1 (0.5 μM), the aldehyde substrate hexanal (2 mM), NAD^+ (2.5 mM), and various concentrations of DEAB. Each symbol represents the average of two trials.
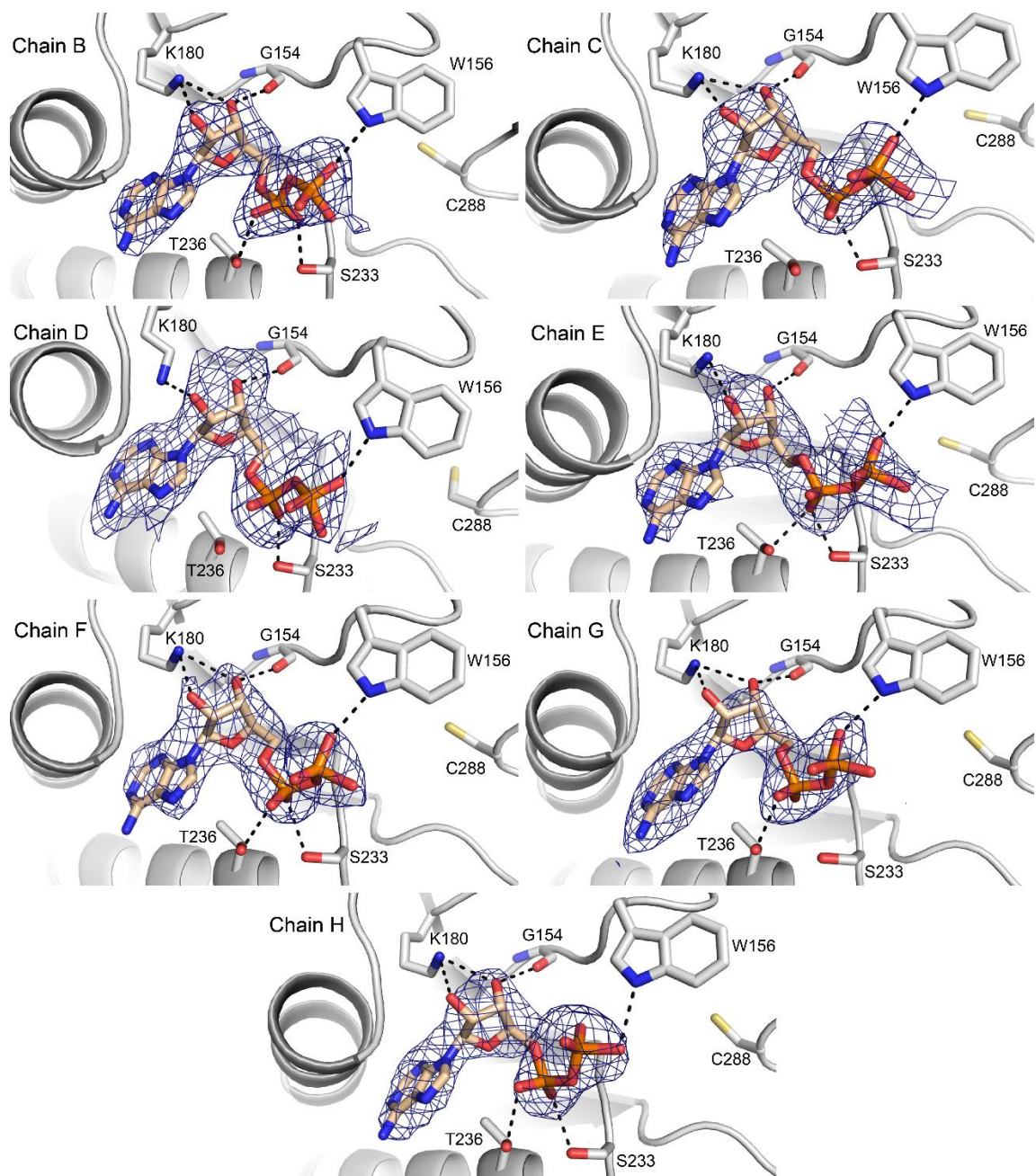
Figure S3.3. Electron density and interactions for NAD$^+$ bound to ALDH9A1 in space group *P*1 (chains B – H). The cage represents a polder omit map (3.5σ).
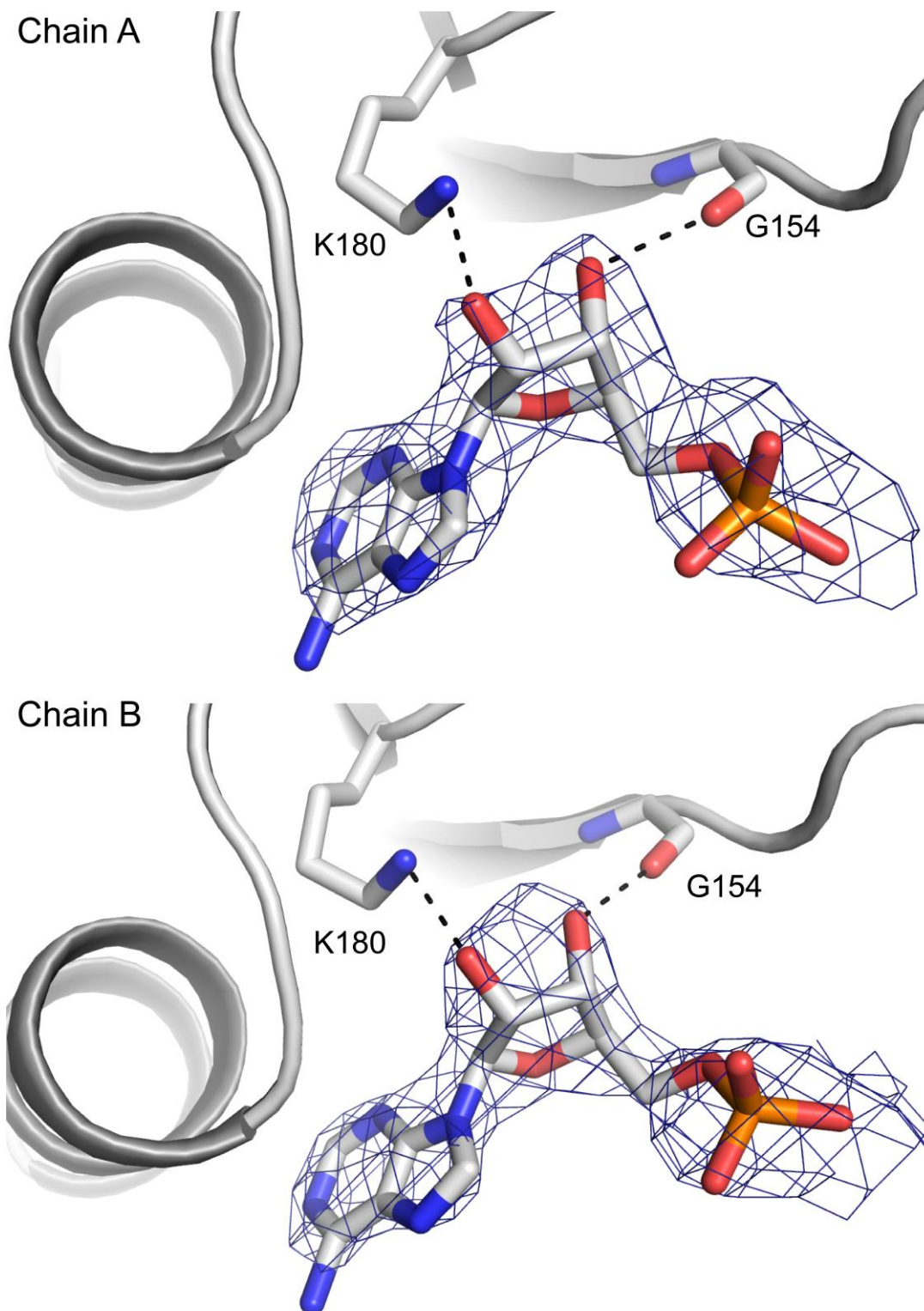
Figure S3.4. Electron density and interactions for NAD$^+$ bound to ALDH9A1 in space group *C*222. The cage represents a polder omit map (3.0σ).

References

[1]   N.A. Sophos, V. Vasiliou, Aldehyde dehydrogenase gene superfamily: the 2002 update, Chem. Biol. Interact. 143–144 (2003) 5–22.

[2]   V. Koppaka, D.C. Thompson, Y. Chen, M. Ellermann, K.C. Nicolaou, R.O. Juvonen, D. Petersen, R.A. Deitrich, T.D. Hurley, V. Vasiliou, Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application, Pharmacol. Rev. 64 (2012) 520.

[3]   F.M. Vaz, S.W. Fouchier, R. Ofman, M. Sommer, R.J.A. Wanders, Molecular and Biochemical Characterization of Rat γ-Trimethylaminobutyraldehyde Dehydrogenase and Evidence for the Involvement of Human Aldehyde Dehydrogenase 9 in Carnitine Biosynthesis*, J. Biol. Chem. 275 (2000) 7390–7394.

[4]   G. Kurys, W. Ambroziak, R. Pietruszko, Human aldehyde dehydrogenase: Purification and characterization of a third isozyme with low Km for γ-aminobutyraldehyde., J. Biol. Chem. 264 (1989) 4715–4721.

[5]   W. Ambroziak, R. Pietruszko, Human aldehyde dehydrogenase. Activity with aldehyde metabolites of monoamines, diamines, and polyamines, J. Biol. Chem. 266 (1991) 13011–13018.

[6]   M.K. Chern, R. Pietruszko, Human Aldehyde Dehydrogenase E3 Isozyme Is a Betaine Aldehyde Dehydrogenase, Biochem. Biophys. Res. Commun. 213 (1995) 561–568.

[7]   R. Končitíková, A. Vigouroux, M. Kopečná, M. Šebela, S. Moréra, D. Kopečný,

Kinetic and structural analysis of human ALDH9A1, Biosci. Rep. 39 (2019).

[8]   K. Johansson, S. Ramaswamy, H. Eklund, M. El-Ahmad, L. Hjelmqvist, H. Jörnvall, Structure of betaine aldehyde dehydrogenase at 2.1 Å resolution, Protein Sci. 7 (1998) 2106–2117.

[9]   C.A. Morgan, B. Parajuli, C.D. Buchman, K. Dria, T.D. Hurley, N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes, Chem. Biol. Interact. 234 (2015) 18–28.

[10]   M. Luo, K.S. Gates, M.T. Henzl, J.J. Tanner, Diethylaminobenzaldehyde is a covalent, irreversible inactivator of ALDH7A1, ACS Chem. Biol. 10 (2015) 693.

[11]   J.F. MORRISON, C.T. WALSH, The behavior and significance of slow-binding enzyme inhibitors, Adv. Enzymol. 61 (1988) 201–301.

[12]   J. Pandhare, C. Dash, M. Rao, V. Deshpande, Slow Tight Binding Inhibition of Proteinase K by a Proteinaceous Inhibitor: CONFORMATIONAL ALTERATIONS RESPONSIBLE FOR CONFERRING IRREVERSIBILITY TO THE ENZYME-INHIBITOR COMPLEX*, J. Biol. Chem. 278 (2003) 48735–48744.

[13]   C. Dash, V. Vathipadiekal, S.P. George, M. Rao, Slow-Tight Binding Inhibition of Xylanase by an Aspartic Protease Inhibitor: KINETIC PARAMETERS AND CONFORMATIONAL CHANGES THAT DETERMINE THE AFFINITY AND SELECTIVITY OF THE BIFUNCTIONAL NATURE OF THE INHIBITOR*, J. Biol. Chem. 277 (2002) 17978–17986.

[14]   E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions., Acta Crystallogr. D. Biol.

Crystallogr. 60 (2004) 2256–2268.

[15] S.J. Perez-Miller, T.D. Hurley, Coenzyme Isomerization Is Integral to Catalysis in Aldehyde Dehydrogenase, Biochemistry. 42 (2003) 7100–7109.

[16] D. Srivastava, R.K. Singh, M.A. Moxley, M.T. Henzl, D.F. Becker, J.J. Tanner, The Three-Dimensional Structural Basis of Type II Hyperprolinemia, J. Mol. Biol. 420 (2012) 176–189.

[17] T.A. Pemberton, D. Srivastava, N. Sanyal, M.T. Henzl, D.F. Becker, J.J. Tanner, Structural Studies of Yeast Δ1-Pyrroline-5-carboxylate Dehydrogenase (ALDH4A1): Active Site Flexibility and Oligomeric State, Biochemistry. 53 (2014) 1350–1359.

[18] M. Luo, J.J. Tanner, Structural basis of substrate recognition by aldehyde dehydrogenase 7A1, Biochemistry. 54 (2015) 5513.

[19] D.A. Korasick, R. Končitíková, M. Kopečná, E. Hájková, A. Vigouroux, S. Moréra, D.F. Becker, M. Šebela, J.J. Tanner, D. Kopečný, Structural and Biochemical Characterization of Aldehyde Dehydrogenase 12, the Last Enzyme of Proline Catabolism in Plants, J. Mol. Biol. 431 (2019) 576–592.

[20] J. Perozich, H. Nicholas, B.-C. Wang, R. Lindahl, J. Hempel, Relationships within the aldehyde dehydrogenase extended family, Protein Sci. 8 (1999) 137–146.

[21] D.A. Korasick, T.A. White, S. Chakravarthy, J.J. Tanner, NAD+ promotes assembly of the active tetramer of aldehyde dehydrogenase 7A1, FEBS Lett. 592 (2018) 3229–3238.

[22] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A visualization system for exploratory

research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[23] T.D. Goddard, C.C. Huang, T.E. Ferrin, Visualizing density maps with UCSF Chimera, J. Struct. Biol. 157 (2007) 281–287.

[24] A.L. Lamb, M.E. Newcomer, The Structure of Retinal Dehydrogenase Type II at 2.7 Å Resolution: Implications for Retinal Specificity, Biochemistry. 38 (1999) 6003–6011.

[25] T. Bordelon, S.K. Montegudo, S. Pakhomova, M.L. Oldham, M.E. Newcomer, A Disorder to Order Transition Accompanies Catalysis in Retinaldehyde Dehydrogenase Type II*, J. Biol. Chem. 279 (2004) 43085–43091.

[26] H.N. Larson, H. Weiner, T.D. Hurley, Disruption of the Coenzyme Binding Site and Dimer Interface Revealed in the Crystal Structure of Mitochondrial Aldehyde Dehydrogenase "Asian" Variant*, J. Biol. Chem. 280 (2005) 30550–30556.

[27] S.A. Moore, H.M. Baker, T.J. Blythe, K.E. Kitson, T.M. Kitson, E.N. Baker, Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases, Structure. 6 (1998) 1541–1551.

[28] A.R. Laciak, D.A. Korasick, K.S. Gates, J.J. Tanner, Structural analysis of pathogenic mutations targeting Glu427 of ALDH7A1, the hot spot residue of pyridoxine-dependent epilepsy, J. Inherit. Metab. Dis. 43 (2020) 635–644.

[29] C. Brocker, N. Lassen, T. Estey, A. Pappa, M. Cantore, V. V Orlova, T. Chavakis, K.L. Kavanagh, U. Oppermann, V. Vasiliou, Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress, J. Biol. Chem. 285 (2010) 18452.

[30]   L. González-Segura, E. Rudiño-Piñera, R.A. Muñoz-Clares, E. Horjales, The
       Crystal Structure of A Ternary Complex of Betaine Aldehyde Dehydrogenase
       from Pseudomonas aeruginosa Provides New Insight into the Reaction Mechanism
       and Shows A Novel Binding Mode of the 2′-Phosphate of NADP+ and A Novel
       Cation Binding Site, J. Mol. Biol. 385 (2009) 542–557.

[31]   R.W. Storms, A.P. Trujillo, J.B. Springer, L. Shah, O.M. Colvin, S.M. Ludeman,
       C. Smith, Isolation of primitive human hematopoietic progenitors on the basis of
       aldehyde dehydrogenase activity, Proc. Natl. Acad. Sci. 96 (1999) 9118 LP –
       9123.

[32]   L. Zhou, D. Sheng, D. Wang, W. Ma, Q. Deng, L. Deng, S. Liu, Identification of
       cancer-type specific expression patterns for active aldehyde dehydrogenase
       (ALDH) isoforms in ALDEFLUOR assay, Cell Biol. Toxicol. 35 (2019) 161–177.

[33]   L. Mele, D. Liccardo, V. Tirino, Evaluation and Isolation of Cancer Stem Cells
       Using ALDH Activity Assay BT - Cancer Stem Cells: Methods and Protocols, in:
       G. Papaccio, V. Desiderio (Eds.), Springer New York, New York, NY, 2018: pp.
       43–48.

[34]   P. Marcato, C.A. Dean, C.A. Giacomantonio, P.W.K. Lee, Aldehyde
       dehydrogenase: Its role as a cancer stem cell marker comes down to the specific
       isoform, Cell Cycle. 10 (2011) 1378–1384.

[35]   K. Pors, J.S. Moreb, Aldehyde dehydrogenases in cancer: an opportunity for
       biomarker and drug development?, Drug Discov. Today. 19 (2014) 1953–1963.

[36]   J.S. Moreb, D. Ucar, S. Han, J.K. Amory, A.S. Goldstein, B. Ostmark, L.-J.
       Chang, 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. 195 (2012) 52–60.

[37] C.A. Lewis, R. Wolfenden, Thiohemiacetal formation by inhibitory aldehydes at the active site of papain, Biochemistry. 16 (1977) 4890–4895.

[38] H.-H. Otto, T. Schirmeister, Cysteine Proteases and Their Inhibitors, Chem. Rev. 97 (1997) 133–172.

[39] E. Dufour, A.C. Storer, R. Menard, Peptide Aldehydes and Nitriles as Transition State Analog Inhibitors of Cysteine Proteases, Biochemistry. 34 (1995) 9136–9143.

[40] R.B. Perni, S.J. Almquist, R.A. Byrn, G. Chandorkar, P.R. Chaturvedi, L.F. Courtney, C.J. Decker, K. Dinehart, C.A. Gates, S.L. Harbeson, A. Heiser, G. Kalkeri, E. Kolaczkowski, K. Lin, Y.-P. Luong, B.G. Rao, W.P. Taylor, J.A. Thomson, R.D. Tung, Y. Wei, A.D. Kwong, C. Lin, Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease, Antimicrob. Agents Chemother. 50 (2006) 899–909.

[41] B.A. Malcolm, R. Liu, F. Lahser, S. Agrawal, B. Belanger, N. Butkiewicz, R. Chase, F. Gheyas, A. Hart, D. Hesk, P. Ingravallo, C. Jiang, R. Kong, J. Lu, J. Pichardo, A. Prongay, A. Skelton, X. Tong, S. Venkatraman, E. Xia, V. Girijavallabhan, F.G. Njoroge, SCH 503034, a mechanism-based inhibitor of hepatitis C virus NS3 protease, suppresses polyprotein maturation and enhances the antiviral activity of alpha interferon in replicon cells, Antimicrob. Agents Chemother. 50 (2006) 1013–1020.

[42]   K.P. Romano, A. Ali, C. Aydin, D. Soumana, A. Ozen, L.M. Deveau, C. Silver, H. Cao, A. Newton, C.J. Petropoulos, W. Huang, C.A. Schiffer, The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors, PLoS Pathog. 8 (2012) e1002832–e1002832.

[43]   E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on ExPASy server, 2005.

[44]   J.I. Juncosa, K. Takaya, H. V Le, M.J. Moschitto, P.M. Weerawarna, R. Mascarenhas, D. Liu, S.L. Dewey, R.B. Silverman, Design and Mechanism of (S)-3-Amino-4-(difluoromethylenyl)cyclopent-1-ene-1-carboxylic Acid, a Highly Potent γ-Aminobutyric Acid Aminotransferase Inactivator for the Treatment of Addiction, J. Am. Chem. Soc. 140 (2018) 2151–2164.

[45]   K.A. Salminen, J. Leppänen, J.I. Venäläinen, M. Pasanen, S. Auriola, R.O. Juvonen, H. Raunio, Simple, Direct, and Informative Method for the Assessment of CYP2C19 Enzyme Inactivation Kinetics, Drug Metab. Dispos. 39 (2011) 412 LP – 418.

[46]   A. D'Arcy, T. Bergfors, S.W. Cowan-Jacob, M. Marsh, Microseed matrix screening for optimization in protein crystallization: what have we learned?, Acta Crystallogr. Sect. F, Struct. Biol. Commun. 70 (2014) 1117–1126.

[47]   M. Till, A. Robson, M.J. Byrne, A. V Nair, S.A. Kolek, P.D. Shaw Stewart, P.R. Race, Improving the success rate of protein crystallization by random microseed matrix screening, J. Vis. Exp. (2013) 50548.

[48]   W. Kabsch, XDS, Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 125–132.

[49]   P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution?,

Acta Crystallogr. D. Biol. Crystallogr. 69 (2013) 1204–1214.

[50] B.W. Matthews, Solvent content of protein crystals, J. Mol. Biol. 33 (1968) 491–497.

[51] A. Vagin, A. Teplyakov, MOLREP: an Automated Program for Molecular Replacement, J. Appl. Crystallogr. 30 (1997) 1022–1025.

[52] N. Stein, CHAINSAW: a program for mutating pdb files used as templates in molecular replacement, J. Appl. Crystallogr. 41 (2008) 641–643.

[53] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software., J. Appl. Crystallogr. 40 (2007) 658–674.

[54] P. V Afonine, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, P.D. Adams, Towards automated crystallographic structure refinement with phenix.refine, Acta Crystallogr. D. Biol. Crystallogr. 68 (2012) 352–367.

[55] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 486–501.

[56] V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 12–21.

[57] S. Gore, E. Sanz García, P.M.S. Hendrickx, A. Gutmanas, J.D. Westbrook, H. Yang, Z. Feng, K. Baskaran, J.M. Berrisford, B.P. Hudson, Y. Ikegawa, N. Kobayashi, C.L. Lawson, S. Mading, L. Mak, A. Mukhopadhyay, T.J. Oldfield, A. Patwardhan, E. Peisach, G. Sahni, M.R. Sekharan, S. Sen, C. Shao, O.S. Smart,

E.L. Ulrich, R. Yamashita, M. Quesada, J.Y. Young, H. Nakamura, J.L. Markley, H.M. Berman, S.K. Burley, S. Velankar, G.J. Kleywegt, Validation of Structures in the Protein Data Bank, Structure. 25 (2017) 1916–1927.

[58] J.L. Cole, J.W. Lary, T. P. Moody, T.M.B.T.-M. in C.B. Laue, Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium, in: Biophys. Tools Biol. Vol. One Vitr. Tech., Academic Press, 2008: pp. 143–179.

[59] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, Biophys. J. 78 (2000) 1606.

[60] K.N. Dyer, M. Hammel, R.P. Rambo, S.E. Tsutakawa, I. Rodic, S. Classen, J.A. Tainer, G.L. Hura, High-throughput SAXS for the characterization of biomolecules in solution: a practical approach, Methods Mol. Biol. 1091 (2014) 245–258.

[61] P. V Konarev, V. V Volkov, A. V Sokolova, M.H.J. Koch, D.I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis, J. Appl. Crystallogr. 36 (2003) 1277.

[62] D. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495.

[63] V. Piiadov, E. Ares de Araújo, M. Oliveira Neto, A.F. Craievich, I. Polikarpov, SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale, Protein Sci. 28 (2019) 454–463.

[64] D. Schneidman-Duhovny, M. Hammel, J.A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their

complexes based on SAXS profiles, Nucleic Acids Res. 44 (2016) W424.

[65]   T.D. Grant, Ab initio electron density determination directly from solution

scattering data, Nat. Methods. 15 (2018) 191–193.

[66]   M. AU - Rames, Y. AU - Yu, G. AU - Ren, Optimized Negative Staining: a

High-throughput Protocol for Examining Small and Asymmetric Protein Structure

by Electron Microscopy, JoVE. (2014) e51087.

[67]   J. Zivanov, T. Nakane, B.O. Forsberg, D. Kimanius, W.J. Hagen, E. Lindahl, S.H.

Scheres, New tools for automated high-resolution cryo-EM structure determination

in RELION-3, Elife. 7 (2018) e42166.

Appendix – Development of proline analogs as PRODH inhibitors

A1.1. Synthesis of benzyl-proline analogs

L-proline was functionalized with benzaldehyde derivatives through reductive amination. Reaction mixtures contained 1 eq of L-proline and 1.2 eq of benzaldehyde derivative in 3 mL of 95 % ethanol. Following incubation at 40 °C for 18 hrs, 3 eq of sodium triacetoxyborohydride was added and the mixture allowed to incubate for a further 12 hrs. Product formation was verified by TLC, and column chromatography of the product on silica gel was eluted with ethyl acetate and methanol (3:1).

A1.2. Kinetic Analysis of analogs as inhibitors for PRODH

All kinetic experiments were performed with *E. coli* proline utilization A (PutA) containing residues 86 – 630. Purified compounds were tested against the proline dehydrogenase (PRODH) domain. Steady state kinetic parameters for the PRODH substrate L-proline and inhibition constants for the proline analogs were determined by a dye-coupled assay using dichlorophenolindophenol (DCPIP) and phenazine methosulfate (PMS) to monitor reduction of FAD following oxidation of proline to (P5C). The reduction of DCPIP was monitored by the decrease in absorption at 600 nm at 25 °C in an Epoch 2 plate reader (BioTek, Winooski, VT, USA). The assay buffer contained 50 mM Tris pH 8.0, 10 % glycerol, 75 µM DCPIP, 1.5 mM PMS, 5 % ethanol, and 2 % DMSO. The proline concentration was varied in the range of 0 – 500 mM at fixed benzyl-proline analog concentrations of 0.0, 0.1, 0.25, 0.5, 0.75 mM. The enzyme was used at 100 nM. Each measurement was performed in duplicate. Initial rates were estimated by linear regression to the first three minutes of the assay. Kinetic parameters were obtained by fitting the initial rate data to the mixed inhibition equation (Equation A1; [1]) using Origin version 2020.

Where [S] indicates the substrate concentration, [I] represents the inhibitor concentration, V_{max} is the maximum velocity, K_i is the inhibitor constant, K_m is the Michaelis constant, and α defines the effect of the inhibitor binding on the affinity of the enzyme for the substrate.

$$v = \frac{\frac{V_{max}}{\left(1+\frac{[I]}{\alpha K_i}\right)}[S]}{[S]+K_m\left(\frac{\left(1+\frac{[I]}{K_i}\right)}{1+\left(\frac{[I]}{\alpha K_i}\right)}\right)} \quad (A1)$$

A1.3. Results

Synthesized benzyl-proline analogs were tested for inhibitory properties against the PRODH domain of *Ec*PutA. These compounds were designed to retain proline as an active site binding moiety, with the functionalized aromatic ring extending outwards from the active site. The (4-(dimethylamino)benzyl)-*L*-proline (DMLP) showed promise with a K_i of 0.25 mM. For comparison, the best known PRODH inhibitor, *L*-tetrahydro-2-furoic acid (THFA), has a *Ki* of 0.2 mM [2,3] To determine how a larger functionalized group on the aromatic ring would affect inhibition, a pyrrolidine was substituted for the dimethylamine group to yield (4-(pyrrolidine-1-yl)benzyl)-*L*-proline. This substitution negatively affected inhibition, increasing the inhibition constant by about a factor of two; $K_i = 0.43$ mM.

A1.4. Conclusion and further studies

The generation of proline analogs containing functionalized aromatic rings as inhibitors is promising. While a limited number of compounds has currently been tested, the inactivation by larger molecular weight and branched compounds when compared to THFA and other reported PRODH inhibitors allows for future development of more drug-

like inhibitors. Further investigation is needed to assess the range of chemical space that can be utilized in generating these analogs. Fitting of the kinetic curves to a mixed inhibition model suggests multiple binding sites, thus the mode of binding should be determined through protein crystallography and binding experiments.
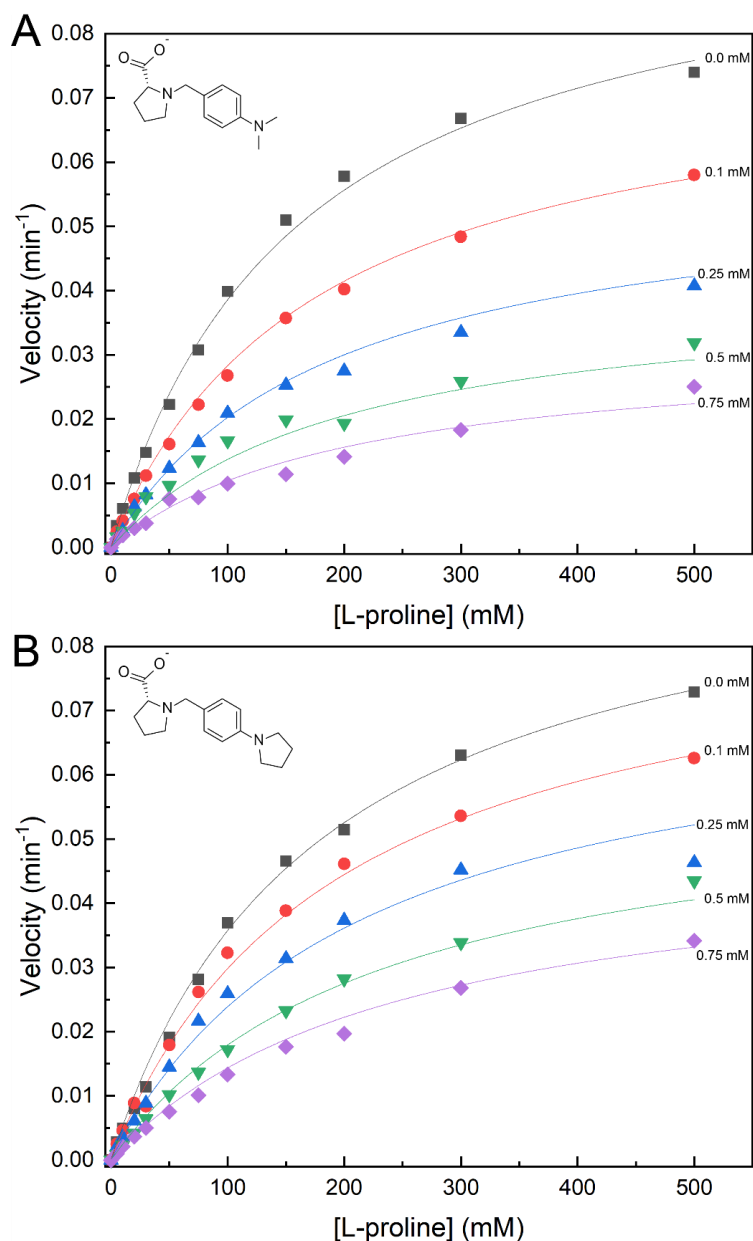

Figure A1. Mixed inhibition of *Ec*PutA by benzyl-proline analogs. Proline concentration was varied from 0 – 500 mM at fixed inhibitor concentrations of 0.0, 0.1, 0.25, 0.5, and 0.75 mM. (A) (4-(dimethylamino)benzyl)-*L*-proline (DMLP), (B) (4-(pyrrolidine-1-yl)benzyl)-*L*-proline (PBLP).

[1]   Reversible Modes of Inhibitor Interactions with Enzymes, Eval. Enzym. Inhib. Drug Discov. (2013) 57–121.

[2]   W. Zhu, Y. Gincherman, P. Docherty, C.D. Spilling, D.F. Becker, Effects of proline analog binding on the spectroscopic and redox properties of PutA, Arch. Biochem. Biophys. 408 (2002) 131–136.

[3]   T.A. White, N. Krishnan, D.F. Becker, J.J. Tanner, Structure and Kinetics of Monofunctional Proline Dehydrogenase from Thermus thermophilus*, J. Biol. Chem. 282 (2007) 14316–14327.

VITA

Jesse W. Wyatt received his bachelor's degree in biochemistry from Blackburn College, where he gained a fascination with protein function. Later, he obtained a master's in molecular biology from the Southern Illinois University Edwardsville. Here, he found a passion for the study of protein folding and enzymology. He joined the lab of Dr. Gates and Dr. Tanner as a doctoral student at the University of Missouri in 2014. As a doctoral student, he gained knowledge and an appreciation for bio-analytical techniques, including protein chromatography, crystallography, and various enzyme kinetic techniques applying them to the study of protein structure and enzymatic function of the human aldehyde dehydrogenase superfamily. He obtained a Ph.D. in 2021, after which he took a position as a Staff Scientist I for Eurofins in Columbia, MO.