Università degli Studi del Piemonte Orientale
"Amedeo Avogadro"

Dipartimento di Scienze del Farmaco

Dottorato di Ricerca in Biotecnologie Farmaceutiche e Alimentari
XXVIII ciclo a.a. 2012-2015

BIOCHEMICAL AND STRUCTURAL INVESTIGATION ON HUMAN ALDH1A3

Andrea Moretti

Supervised by
Dott.ssa Silvia Garavaglia

PhD program co-ordinator Prof. Menico Rizzi

Andrea Moretti was the recipient of a Ph.D. fellowship by the European Commission under the 7th Framework Programme.

Contents

Introduction

Aldehyde dehydrogenases superfamily

Aldehydes dehydrogenases (ALDHs) are known to be $NAD(P)^+$ dependent enzymes responsible for the oxidation of a wide spectrum of aldehydes into their corresponding carboxylic acids[1]. Phylogenetic studies suggested that ALDHs derived from four ancestral genes that exist prior to the division between *Eubacteria* and *Eukaryotes* about 2 billion years ago[2]. The evolutionary history of ALDH genes started with diversifications and duplications occurred from 890 million years ago to 70 million years ago[2]. The ALDH superfamily is represented across *Archea, Eubacteria* and *Eukarya* taxa with multiple genes, implying a critical role for these enzymes in the evolutionary history of many species[3]. The actual nomenclature of human ALDH superfamily was established during the Ninth International Symposium on Enzymology and Molecular Biology of Carbonyl Metabolism, held in 1998, in Varallo Sesia, Italy[1]. The ALDH superfamily comprises 19 functional protein-coding genes located on different chromosomes. ALDH isoforms belong to families from ALDH1 to ALDH9, ALDH16 and ALDH18 (Figure 1)[4], [5]. Human ALDHs are grouped in the same family by sharing more than 40% sequence identity, meanwhile they belong to the same subfamily when they exhibit more than 60% of amino acids identity[5]. The main biological role of ALDHs is regarded as being detoxifying enzymes, due to the ability to inactivate endogenous and exogenous toxic aldehydes. Nevertheless ALDHs are essential enzymes for the biosynthesis of key molecules in different metabolic pathways.

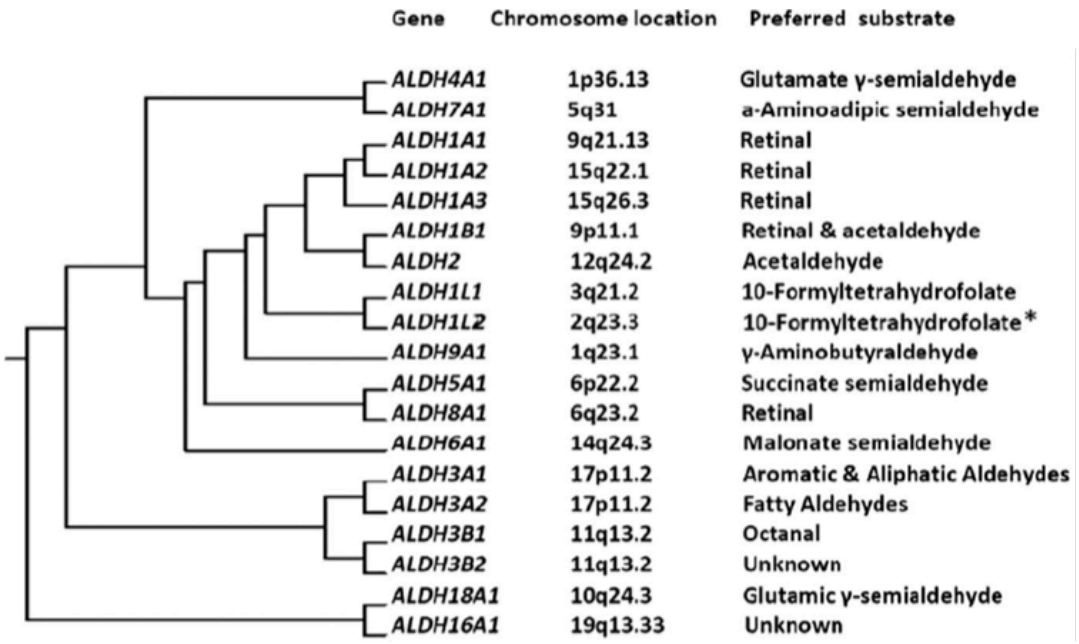


Figure 1. Evolutionary tree for human aldehyde dehydrogenases with indication of chromosomal location and preferred substrate from Koppaka et al.[6].

ALDHs catalyze the conversion of exogenous and endogenous aldehydes

Aldehydes are reactive electrophilic molecules with at least one hydrogen substituent on the carbonyl carbon at the end of the carbon chain. Aldehydes can be grouped in four large subclasses: short chain aldehydes, long chain alkanals, aromatic aldehydes and α,β-unsaturated aldehydes[7]. The anthropic activities such as combustion of hydrocarbon fuels, solvent utilization, air pollution, cigarette smoke, preservation of food and metabolism of drugs are the main causes of exogenous aldehydes formation. In addition, endogenous aldehydes are often by-products of metabolic processes such as lipid peroxidation, amino acid and

neurotransmitters catabolism, vitamins and steroids metabolism[8]. Cytotoxic endogenous and exogenous aldehydes are detoxified mainly through the oxidation catalyzed by ALDHs that prevents mutagenesis and cell death caused by DNA-adducts and cross-linking of proteins, consequence of aldehydes reactivity[9], [10]. Besides their key role as detoxifying enzymes, ALDHs are important catalysts in the xenobiotic metabolism (Figure 2). The phase 1 of drug metabolism consists in the biotransformation of the pharmaceutical compound that involves the oxidation of the drug functional group catalyzed by ALDHs.

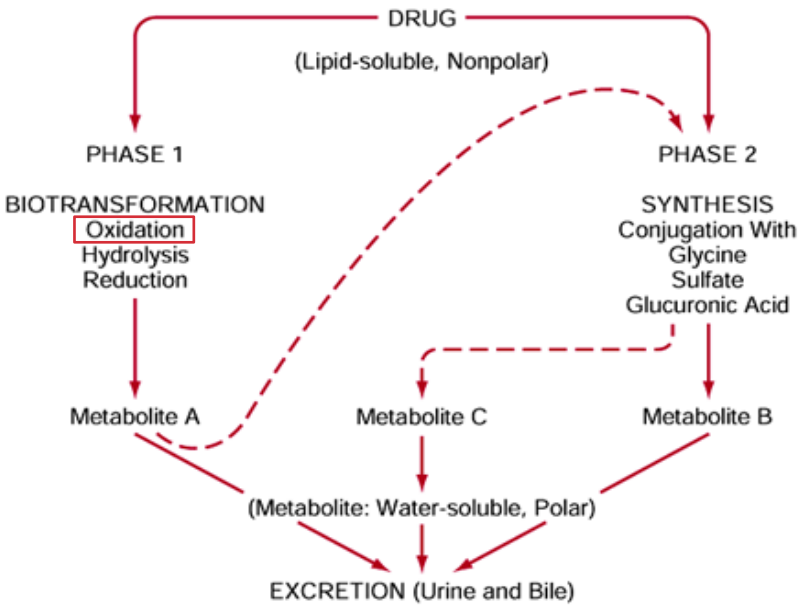


Figure 2. Schematic representation of the xenobiotic metabolism of drugs in humans.

ALDHs are implicated both in the detoxification of cells and in important metabolic and signaling processes (Figure 3). Different ALDH isoforms catalyze the biosynthesis of physiologically relevant molecules involved in embryonic development (retinoic acid), neurotransmission (γ-aminobutyric acid) and osmosis (betaine)[4].

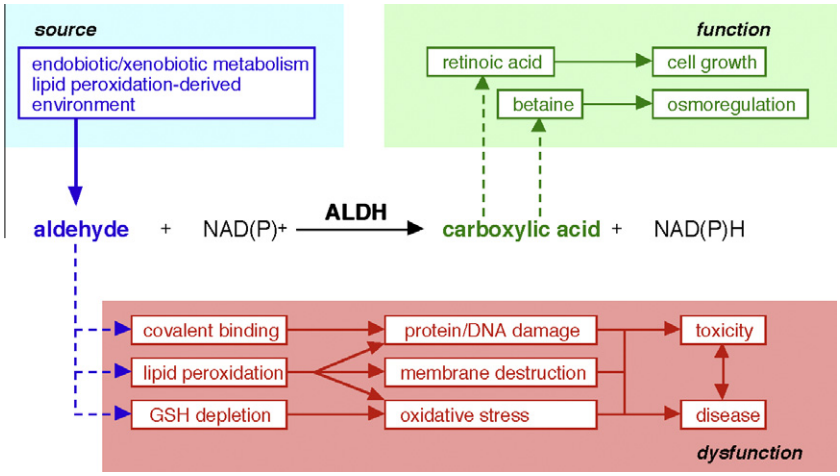


Figure 3. Schematic representation of ALDHs catalytic activity and their role in cell detoxification and metabolism, from Vasiliou et al.[5].

Representative isoforms of human ALDH superfamily and their biological roles

The 19 ALDH isoforms composing the ALDH superfamily are listed in Table 1 with a brief description of the main characteristics for each isoforms.

ALDH1A1, ALDH1A2 and ALDH1A3 are the three known isoforms of the 1A subfamily participating in the biosynthesis of retinoic acid (REA), a morphogen molecule that regulates embryonic development and cellular differentiation. By oxidizing retinaldehyde to retinoic acid the three ALDH1A isoforms are essential for the REA signaling pathway.

Table 1. List of ALDH isoforms with brief description of localization, substrate preference and other functions, from Marchitti et al.[4].

ALDH	Subcellular location	Preferred aldehyde substrate	Additional functions and characteristics
ALDH1A1	Cytosol	Retinal	Ester hydrolysis; binds androgen, cholesterol, thyroid, daunorubicin, and flavopiridol; corneal and lens crystallin; oxidizes DOPAL, acetaldehyde
ALDH1A2	Cytosol	Retinal	High affinity for LPO-derived aldehydes
ALDH1A3	Cytosol	Retinal	High affinity for LPO-derived aldehydes
ALDH1B1	Mitochondria	Acetaldehyde	May protect the cornea from UV-light
ALDH1L1	Cytosol	10-Formyltetrahydrofolate	Binds acetaminophen
ALDH1L2	Unknown	Unknown	Induced by the anti-inflammatory agent indomethacin
ALDH2	Mitochondria	Acetaldehyde	Ester hydrolysis; nitroglycerin bioactivation, oxidizes LPO-derived aldehydes; binds acetaminophen; oxidizes DOPAL and DOPEGAL
ALDH3A1	Cytosol, nucleus *	Aromatic, aliphatic aldehydes	Ester hydrolysis; scavenges ROS; UV-filter; corneal crystallin; oxidizes LPO-derived aldehydes; regulation of cell-cycle; inducted by PAHs
ALDH3A2	Microsomes, peroxisomes ‡	Fatty aldehydes	Insulin regulates gene expression
ALDH3B1	Cytosol §	Unknown	Oxidizes LPO-derived aldehydes
ALDH3B2	Unknown	Unknown	Unknown
ALDH4A1	Mitochondria	Glutamate γ-semialdehyde	Ester hydrolysis; may mitigate oxidative stress
ALDH5A1	Mitochondria	Succinate semialdehyde	May be involved in neurotransmission efficiency
ALDH6A1	Mitochondria	Malonate semialdehyde	Esterase activity; only known human CoA-dependent ALDH
ALDH7A1	Cytosol, nucleus, mitochondria ¶	α-Aminoadipic semialdehyde	Closely related to plant osmoregulatory protein; may regulate cell cycle
ALDH8A1	Cytosol	Retinal	Oxidizes LPO-derived aldehydes and acetaldehyde
ALDH9A1	Cytosol	γ-Aminobutyraldehyde	Oxidizes betaine, acetaldehyde and DOPAL; involved in carnitine biosysnthesis; esterase activity
ALDH16A1	Unknown	Unknown	Unknown
ALDH18A1	Mitochondria	Glutamic γ-semialdehyde	Unknown

ALDH1A1 is a homotetramer cytosolic enzyme distributed in retina, eye lens, lung, kidney, liver and brain. ALDH1A1 represent 2-3% of cytosolic proteins in retina and it is important to protect the eye from UV - damage generated aldehydes[11]. ALDH1A1 is mainly involved in embryonic development and in the last years has been investigated for its role in cancer stem cell biology and chemotherapy resistance[11]. Hematopoietic progenitor cells expressing high levels of ALDH1A1 showed an increased tolerance to active metabolites of cyclophosphamide anticancer drug[12]. Moreover, breast cancer cells that

overexpress ALDH1A1 demonstrated an enhanced resistance to doxorubicin and paclitaxel[13]. ALDH1A2 is the second isoform involved in REA biosynthesis. It is expressed in various embryonic tissues as a cytosolic homotetramer and it regulates early heart development in mice[4]. Furthermore human spina bifida disease is significantly associated with ALDH1A2 single nucleotide polymorphisms[14]. ALDH1A3 is the third isoform implicated in REA signaling pathway and its role has been demonstrated crucial in nasal development and more general in embryonic development[15]. ALDH1A3 is expressed in fetal nasal mucosa, kidney, breast, stomach and salivary gland and it is involved in the development of eye, olfactory bulbs, hair follicles, forebrain and cerebral cortex[4]. Mutation in human ALDH1A3 gene has been identified as the major cause of microphtalmia/anophtalmia[16]. Recently the isoenzyme is assuming a central role in the biology of different cancers in relation to its overexpression in ALDH-positive cancer stem cells.

ALDH1L1 is a cytosolic multi domain homotetramer important for the conversion of 10-formyltetrahydrofolate to tetrahydrofolate (THF). THF is an important metabolite in the dietary folate, in one-carbon metabolism and in the purine biosynthesis[17].

ALDH2 is a mitochondrial homotetramer responsible for the conversion of acetaldehyde to acetic acid. ALDH2 is the main enzyme of the ethanol metabolism and its well-known dominant allelic variant ALDH2*2 is the main cause of alcoholic liver disease and cirrhosis in Asian population [18], [19].

ALDH3A1 detoxifies aromatic and medium-chain aldehydes derived from lipid peroxidation[20]. ALDH3A2 is a microsomal homodimer that catalyzes the oxidation of fatty aldheydes and it is part of the fatty alcohol:NADoxidoreductase enzyme complex[21]. Deletions at the first five exon of the enzyme are correlated with Sjögren-Larsson syndrome [22].

ALDH4A1 participates in glutamate biosynthesis and proline degradation by promoting the conversion of pyrroline 5-carboxylate to glutamate[23]. Mutations in its coding gene are responsible of type II hyperprolinemia[24].

ALDH5A1 is a mitochondrial homotetramer, also known as succinic semialdehyde dehydrogenase (SSADH) that is implicated in the catabolism of γ-aminobutyric acid (GABA) by converting succinic semialdehyde to succinate[25]. The 4-hydroxybutyric aciduria, a neurological autosomal disorder, is a disease related to ALDH5A1 malfunction[26].

ALDH6A1 is involved in the valine and pyrimidine catabolism. The enzyme catalyzes the irreversible oxidative decarboxylation of malonate and methylmalonate semialdehydes to acetyl- and propionyl-CoA. Mutations to ALDH6A1 are responsible of dysmielination and methylmalonate aciduria[27].

ALDH7A1 is a homotetramer found both in cytosol and mitochondria. The enzyme participates to the lysine metabolism and mutations on ALDH7A1 gene cause pyridoxine dependent epilepsy[28], [29].

ALDH8A1 is the only isoform outside the ALDH1A subfamily involved in the retinoic acid metabolism. ALDH8A1 has the characteristic to prefer 9 *cis*-retinal substrate compared to all *trans*-retinal, showing a 40 times higher affinity[30].

ALDH9A1 is a key enzyme in the alternate metabolism of GABA. The cytosolic tetramer of ALDH9A1 oxidizes γ-aminobutyraldehyde to GABA[31]. ALDH16A1 is a recently discovered non-catalytic ALDH isoform. It has a preference for the $NADP^+$ cofactor and seems to interact with maspardin, a protein responsible of spastic paraplegia[32]. ALDH18A1 is the second mitochondrial ALDH involved in the proline synthesis and it is a bifunctional enzyme that utilizes ATP and NADP for γ-glutamyl kinase and γ-glutamyl phosphate reductase activities, respectively[33].

Structural and functional aspects of ALDH

ALDHs are built on three functional domains: the NAD^+ binding domain at the N-terminal, the catalytic domain and the oligomerization domain at the C-terminal (Figure 4)[34]. The global tertiary structure of ALDHs is conserved among isoforms across all families, and both the N-terminal and the C-terminal domains involve well-known and conserved amino acids in the binding of the cofactors and in the catalysis of the substrates, respectively[35].

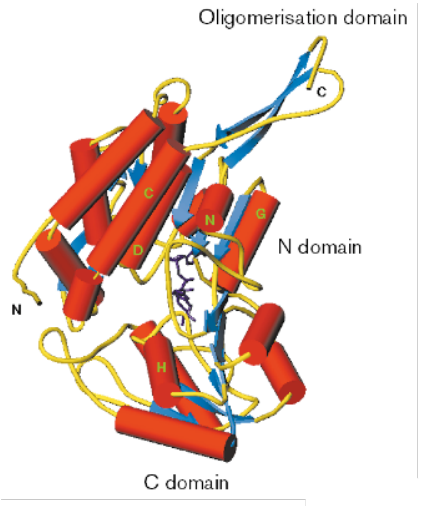


Figure 4. Structure of sheep liver cytosolic ALDH1A1 in complex with NAD^+[34]. The α-helix, the β-sheets and the connecting loop composing the N-domain, the C-domain and the oligomerization domain are respectively showed in red, blue and yellow.

ALDHs function as catalysts for the oxidation of aldehydes, although these enzymes are known to possess also an esterase activity. Beside the large differences in substrate selectivity, the ALDH superfamily shares the same catalytic mechanism. The only known exception is ALDH6A1 that uses CoA instead of $NAD(P)^+$ as cofactor for the decarboxylation of malonate[36]. The aldehyde oxidation through ALDH catalysis is well described in the Figure 5 reported from Koppaka et al.[6]. The catalytic mechanism of ALDH consists in an

ordered reaction in which the enzyme first binds the cofactor than allows aldehyde or ester to enter the catalytic site[37]. The nucleophilic attack on the substrate is operated by the active nucleophile C302 that is conserved among all the ALDHs (the numbering of amino acids is conventionally based on the mature human ALDH2). The thiol of the C302 is activated by a water-mediated proton abstraction operated by the highly conserved E268 that act as a general base[6]. The nucleophile C302 attacks the carbonyl carbon of the substrate driving through the formation of the thioemiacetal intermediate. The hydride ion transfer from the intermediate to the cofactor occurs on the C4 of the nicotinamide ring and it is stereospecific for the *pro*-R side[38]. After the hydride transfer, the E268 activates a second water molecule that is responsible of the hydrolysis of the thioester intermediate and the consequent release of the product. Simultaneously the reduced cofactor leaves the enzyme that is finally regenerated by the binding of a new $NAD(P)^+$ molecule through the conserved residues E399 and K192[6]. Studies on ALDH three-dimensional structure and its catalytic mechanism has been reported in three pivotal works on ALDHs[34], [35], [39]. Liu and collaborators focused on the different way ALDH binds NAD(P) respect to the typical binding of other enzymes sharing the Rossmann fold. Usually the pyrophosphate moiety of NAD closely interacts with the fingerprint sequence GlyXGlyXXGly that is located in the loop between the β_1-αA motifs of the $\beta\alpha\beta$-conserved Rossmann fold domain. Instead, in ALDH the nucleotide cofactor binds between the β_4-αD loop rather than in the canonical β_1-αA and the pyrophosphates does not interact strongly with any residue (Figure 6)[39]. In all the ALDH structures solved in complex with NAD, the cofactor lies in an hydrophobic pocket and it is more likely stabilized through hydrogen bonds between K192, E195, the main chain carbonyl oxygen of I166 and the adenine ring [35].

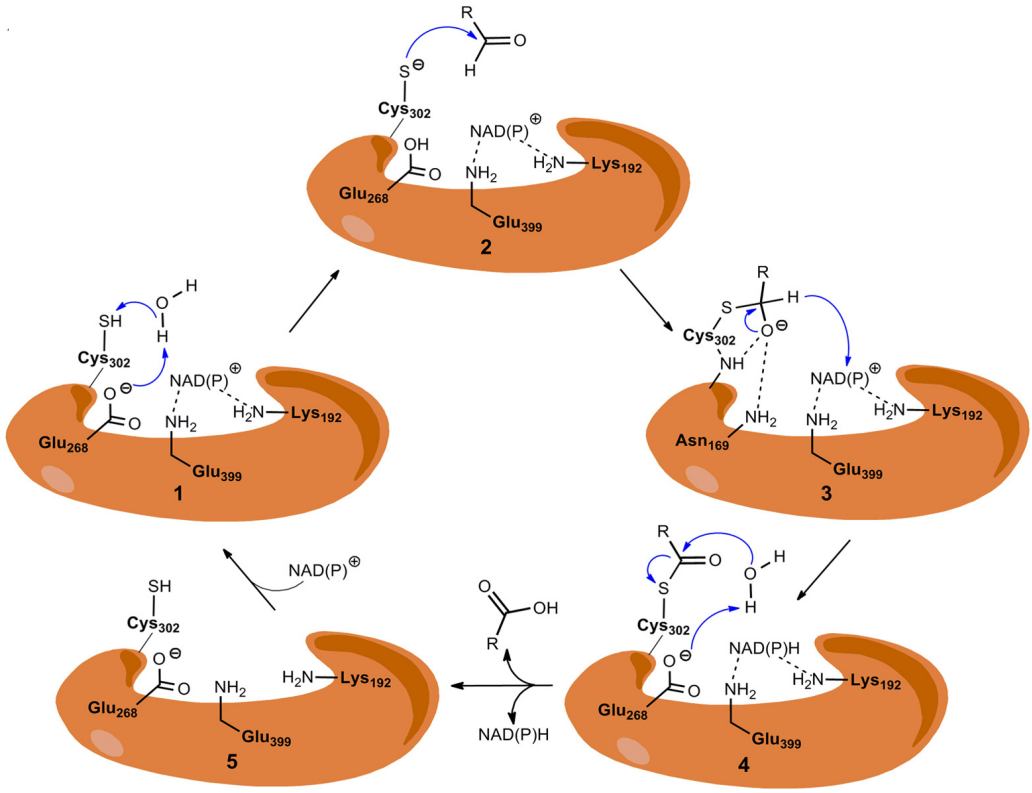


Figure 5. Description of ALDH catalytic mechanism by Koppaka et al.[6] **1:** The enzyme binds the NAD(P) cofactor and E268 activate the water molecule responsible of the thiol deprotonation. **2:** Nucleophilic attack of C302 on the carbonyl carbon of the aldehyde **3:** Thioemiacetlal intermediate formation and concomitant hydride transfer to the $NAD(P)^+$ cofactor **4:** Activation of a water molecule through E268 and hydrolysis of the thioester **5:** Release of the carboxylic acid and reduced cofactor.

In the sheep liver ALDH1A1 structure, for the first time, was described the flexibility of the nicotinamide ring in complex with ALDH. The cofactor was found in two major conformations belonging to two different catalytic steps: the hydride transfer conformation, in which NAD^+ is closer to the active site, ready to accept the hydride, and the release conformation where the reduced cofactor moved away from the catalytic site to make room for the product hydrolysis[34].

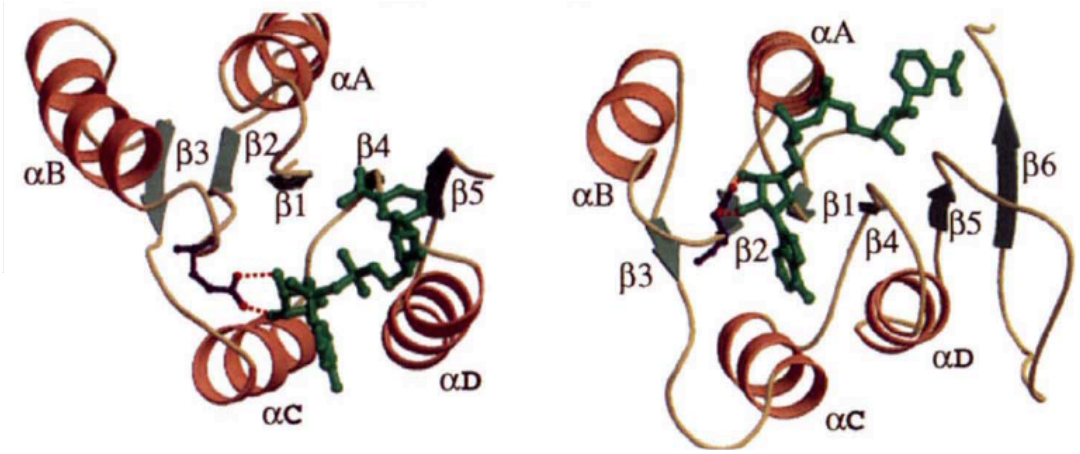


Figure 6. Comparison of NAD binding mode between ALDH3 and alcool dehydrogenase (ADH). The ALDH3 structure is reported as cartoon on the left side of the figure and it evidences how NAD binds between αD β4 and αC of the Rossmann fold. In the ADH structure, reported as cartoon on the right side of the figure, NAD exhibits the classical binding between αA and β1 of the Rossmann fold. From Liu et al., 1997[39].

ALDHs are known to process a wide spectrum of substrates, albeit each isoform displays a major affinity for its "preferred" or natural substrate. Moore et al. compared ALDH1 and ALDH2 structures to identify mechanisms involved in substrate selectivity and affinity[34]. The authors highlight the determinant role of the substrate access tunnel in the selectivity of ALDH1 towards bulky substrates such as retinal, whereas the bovine mitochondrial ALDH2 reveals a smaller and longer tunnel accessible to small substrates such as acetaldehyde[35]. The geometry of the catalytic tunnel is not the only mechanism exploited by ALDHs for substrates selectivity. Structural and biochemical studies on ALDH1A1 and ALDH1A2 suggested that the interaction of bulky substrates with a disorder loop (~ 457 - 477) located at the entrance of the substrate tunnel promotes a disorder to order transition[40], [41]. Activity assays, direct mutagenesis, Fluorescence Resonance Energy Transfer (FRET) analysis and structural studies on both hALDH1A1 and rat ALDH1A2 confirmed that the conformational adaptation of the disordered loop is a key step for the substrate recognition and catalysis[40], [41]. The methods for substrate recognition are not universal across all isoforms and

many aspects about substrate selectivity in ALDHs still need to been elucidated. Recently, MA Keller et al. (2014) published the structure of the fatty aldehyde dehydrogenase (FALDH). Their work identified a new mechanism in which FALDH uses a unique gatekeeper α-helix, at the entrance of the substrate tunnel, to regulate the access of the substrates[42].

Retinoid metabolism and retinoic acid signaling

The etymology of retinol, retinal and retinoic acid derives from the term retina that refers to the organ where they were first identified[43]. Retinoids accomplish different roles in human metabolism. They are at the basis of the visual cycle, but at the meantime, they are crucial for embryonic development and tissues differentiation. The fat-soluble retinol, also known as vitamin-A, is up taken by dietary intake and it is the main source for retinoic acid (REA) biosynthesis. The conversion of retinol to all-*trans* retinoic acid goes through two reactions: the reversible conversion of retinol to all-*trans* retinal by alcohol dehydrogenase (ADH) or short-chain dehydrogenase (SDR), and the irreversible oxidation of all-*trans* retinaldehyde to all-*trans* retinoic acid catalyzed by one of the three ALDH1A isoforms (Figure 7)[44]. The inactivation of retinoic acid is the last important step of the metabolism catalyzed by Cyp26 (also referred as P450RAI) that hydroxylate REA causing its inactivation[45]. Retinoids play an essential role in visual cycle that occurs in cone photoreceptor cell outer segments (OS) and in the retinal pigment epithelium (RPE). Photo isomerization of the 11-*cis*-retinal chromophore of rhodopsin triggers a complex set of metabolic transformations collectively termed phototransduction that ultimately lead to light perception[46].

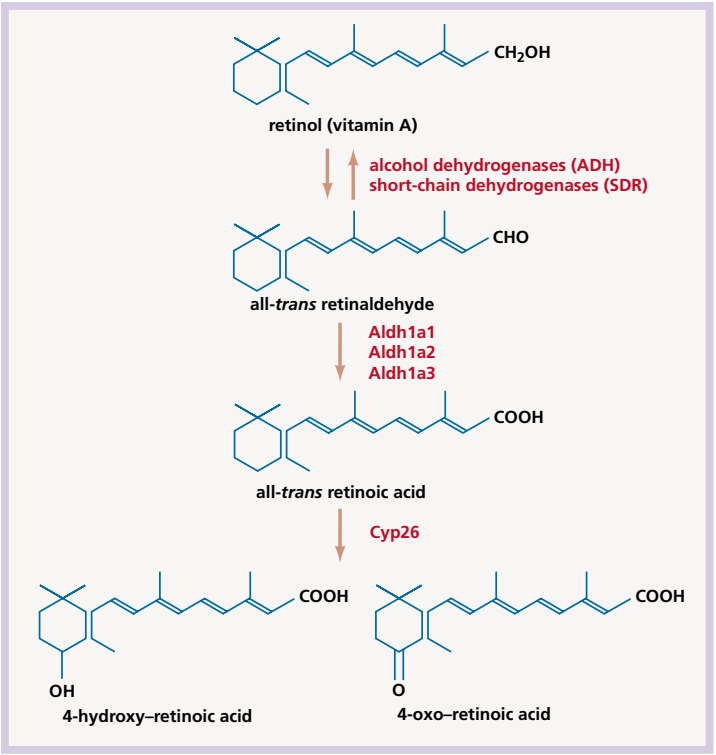


Figure 7. Overview of metabolism and catabolism of retinoids with main enzymes involved in synthesis of retinoic acid and its inactivation[44].

The signaling role of retinoic acid have to be accomplished inside target cells, since that retinol is transported from plasma into target cells through the interaction between the retinol binding protein (RBP) and its receptor stimulated by retinoic acid 6 (STRA6)[47]. Inside the cells retinol can be segregated by cellular retinol binding protein (CRBP) or can be available as substrate for ADH and SDR and enter the retinoid metabolism (Figure 7)[48]. The final product of the retinoid metabolism is retinoic acid, a well-known signaling molecule with pleiotropic functions, fundamental for the spatial-temporal development of the embryo and the tissue development in adults[49]. The morphogen REA is essential in somitogenesis and axial elongation, cardiogenesis and neurogenesis. Therefore, unbalance in REA biosynthesis is cause of severe diseases at embryonic stage and in adult.

In late 1980s were identified for the first time the Retinoic acids Receptors (RARs) that are nuclear transcription factors inducible by retinoic acid binding[50], [51]. RARs are member of the nuclear hormone receptor superfamily and are composed by three subtypes RARα, RARβ and RARγ. RARs function as heterodimers with one member of the retinoid X receptors that are in turn composed by three subtypes RXRα, RXRβ and RXRγ. RAR-RXR heterodimers exert their pleiotropic effect of retinoic acid-dependent transcriptional regulators by binding to the specific Retinoic Acid Response Element (RARE) DNA sequences found in the promoter region of retinoid target genes[52]. The repression and activation of target response genes mechanism for RXR-RAR heterodimer is illustrated in Figure 8. The RXRs are able to bind 9-*cis* retinal and other ligands termed retinoxids, but 9-*cis* retinal cannot be detected endogenously in embryos or adult tissues. It is thus likely that RAR-RXR dimers act mainly, if not solely, by ligand binding of the RAR moiety, while RXR alone is not able to determine the activation of the target gene. This theory is supported by the fact that neither 9-*cis* retinal or other RXR ligands can rescue embryonic disease related to REA defeat[53]. Once identified the molecular basis of REA signaling, it is still tricky to elucidate all the mechanisms that are relevant to the REA mediated response. If on the one hand only about 20 genes have unambiguously RARE elements, on the other hand hundreds genes are involved in REA signaling or through direct interaction or through cross talk and signal cascades activation. Reflecting the complexity of REA response, the work of Liu et al. reported a total of 169 genes, including transcriptional factors, signal transduction modulators, protein modulators and cell cycle promoters and inhibitors, to be REA-modulated and form a harmonious network in the course of the REA-induced NB_4 cell differentiation[54].

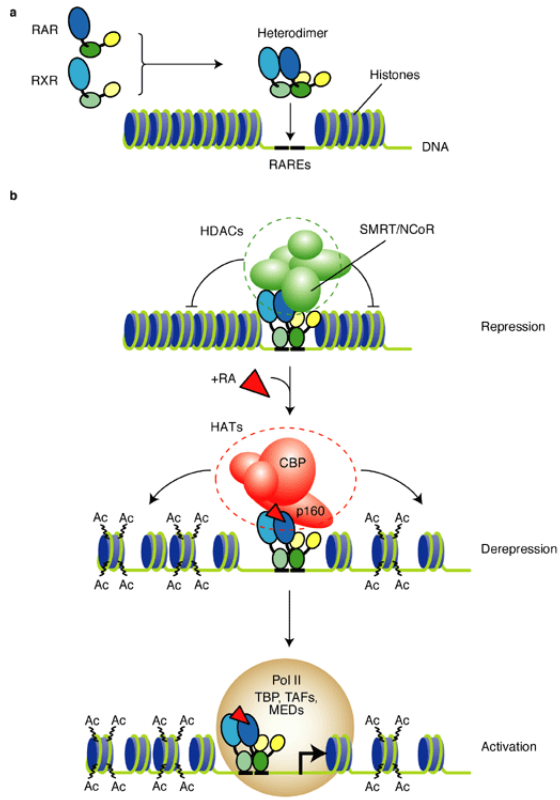


Figure 8. a) RXR-RAR heterodimers transcription factors bind to Retinoic Acid Response Element target genes. **b)** In absence of retinoic acid (REA) the system is repressed by transcriptional co-repressor (SMRT or NcoR) and by the histone deacetylase (HDACs). The binding of REA induces the release of the HDCA and the recruitment of Histone Acetyltransferase (HAT) co-activator complex that acetylates the histones for the de-repression. Activation of the target gene is consequence of the transcription started by the RNA polymerase II holoenzyme, together with the TATA-binding protein (TBP), the TBP-associated factors (TAFs) and mediator complexes (MEDs). From Clarke et al.[55].

The embryonic development and cell fate determination of organisms are orchestrated by REA signaling, since that unbalance or absence of REA in embryos is cause of defects in the growth of central nervous system (CNS), craniofacial region, limb, urogenital system, lungs and heart. REA balance is determinant also in post-embryonic development and regeneration of lung, hair, ear and CNS[49]. Although is known that REA-activated pathway plays important, evolutionarily and conserved roles in neurogenesis, many of the detailed molecular interactions required for the neural progenitors proliferation and differentiation into neurons remain obscure[56].

Retinoic acid was tested as therapeutic compound on brain cancer cell lines such as neuroblastoma and glioma. The thinking behind the use of retinoic acid against brain tumors is based on the observation that an undifferentiated cancer cell population is cause of a more aggressive cancer and poor prognosis. Since differentiation of cancer cells population is a goal in cancer treatment, the signaling role of REA in tissue differentiation should help to threat cancer. Even thought the ability of REA to differentiate neuroblastoma cell lines have been demonstrated, the clinical response to REA treatment is too variable[56]. REA-induced alterations, targeting undifferentiated glioblastoma cancer cells, lead to a differentiation of the cancer cells population, rendering them sensitive to targeted therapy. However retinoid signaling is even more complex in clinical gliomas, so that resistance and side effects were common[56], [57]. Hypotheses on inefficiency of REA treatment against brain tumors involves the possibility that REA is channeled to a pro-proliferative, oncogenic pathway depending on the relative abundance of the REA[56].

ALDHs in Cancer Stem Cells (CSCs)

First evidence of a cancer cell population with stem-like properties was reported by Bonnet and Dick in 1997 with their studies on acute myeloid leukemia[58]. In the early 2000s the Cancer Stem Cells (CSCs) theory was extended to breast cancer and glioma, two of the most studied tumors in relation to CSCs behavior[59], [60]. After almost 20 years of research, the scientific community agrees on the existence of a subpopulation of multipotent cancer cells with the ability to initiate the tumor. CSCs own the typical self-renewal/differentiation capacity of stem cells and it is quite likely that they are the engine of tumorigenisis in many cancers, not only in the formation of primary tumors but also for metastasis formation and reservoir of therapy resistant cells[61]. As clearly described by the work of Jordan et al., the origin of CSCs can't be unique. Jordan presented three different scenarios from

which CSCs can be originated (Figure 9). The first scenario explains how the cancer stem cell that generates the primary tumor can derives from aberrant mutations of a normal stem cell or progenitor cell. The second scenario describes a situation in which refractory cancer stem cells, able to form a relapsed tumor, are consequence of chemotherapy treatment. The last scenario describes a tumor cell escape of a metastatic cancer stem cell that originates metastasis.

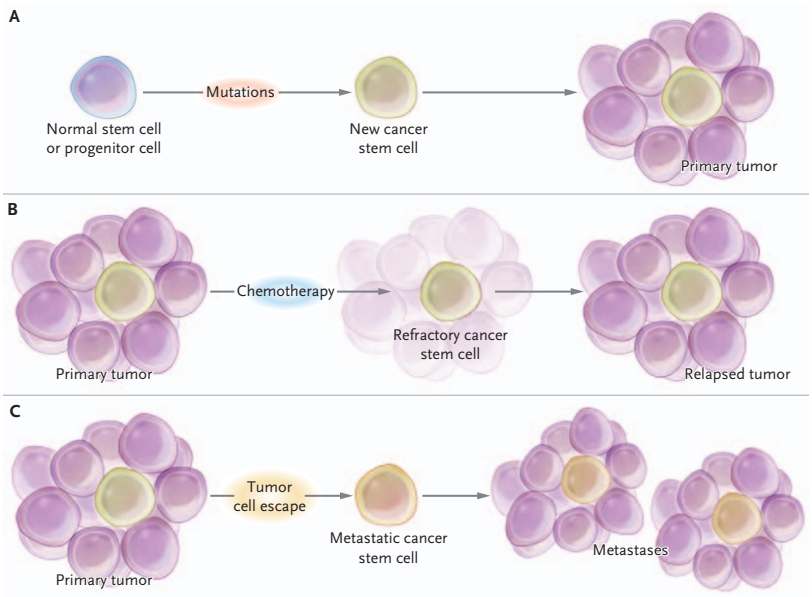


Figure 9. Three scenarios describing the origin of Cancer Stem Cells (CSCs) and their involvement in tumorigenesis, drug resistance and tumor propagation **A)** CSC generates from mutations to normal stem cell or progenitor cell. The generated CSC gave rise to the primary tumor. **B)** Chemotherapy treatment of the primary tumor cause a refractory cancer stem cell able to generates a relapsed tumor. **C)** A tumor cell escape from the primary cancer generates a metastatic cancer stem cell that originates metastasis[61].

Although the existence of a CSCs subpopulation has been proof and accepted, their identification is still controversial. Over the recognition through associated markers, CSC detection needs to be confirmed through a "gold standard" test that consists in the validation of CSCs ability to start a new tumor *ex vivo* or after xenotransplantation in animal models[60]. Up to date the relation between specific cancer markers and CSCs presence is still argument of debate. In Table 2 are listed the majority of the markers that have been somehow correlated to CSCs. Diverse

markers were scrutinized only for cell lines and does not have evidence in other cancer cell lines or *in vivo*[62]. The discovery of more specific and reliable markers among with the development of standard protocol for the evaluation of CSCs presence is a desirable goal for cancer research.

Table 2. List of molecular markers for CSC adapted from[62]. This list is not exhaustive and includes markers not exhaustively tested.

Breast	**Colon**	**Glioma**	**Liver**	**Lung**
ALDH1 CD24 CD44 CD90 CD133 Hedgehog-Gli activity α_6-integrin	ABCB5 ALDH1 β-catenin CD24 CD26 CD29 CD44 CD133 CD166 LGR5	CD15 CD90 CD133 α_6-integrin nestin	CD13 CD24 CD44 CD90 CD133	ABCG2 ALDH1 CD20 CD133 CD271
Melanoma	**Ovarian**	**Pancreatic**	**Prostate**	
ABCG2 ALDH1 CD20 CD133 CD271	CD24 CD44 CD117 CD133	ABCG2 ALDH1 CD24 CD44 CD133 c-Met CXCR4 Nestin Nodal - Activin	ALDH1 CD44 CD133 CD166 α_2-β_1-integrin α_6-integrin	

ALDHs as cancer stem cell markers

ALDHs are good candidates for the identification of CSCs and has been used as marker for many solid tumors including breast, brain, lung, liver, colon, pancreatic, ovarian, head and neck, prostate and melanoma[5]. ALDHs are a multipurpose tool in CSCs studies because they couple the detection of the CSC subpopulation with the selection of ALDH-positive cells through Fluorescence Activated Cell Sorting (FACS) with ALDEFLUOR™ assay. This technique is suitable for FACS that uses BODIPY aminoacetaldehyde (BAAA) as modified fluorescent substrate for ALDH. Cells with high ALDH activity convert BAAA in the cytoplasm into the negatively charged BODIPY aminoacetate (BAA^-), which is retained into the cells that appear as a distinct cohort of cells exhibiting green fluorescence[63]. High aldehyde dehydrogenase activity in breast, myeloid leukemia, prostate, rectal, oesophageal, lung, ovarian and gallbladder cancers is predictive of worst outcome, therefore ALDHs are being scrutiny as a potential prognostic markers[64]. The identification and correlation of specific ALDH isoforms to the presence of CSC subpopulation and a poor cancer prognosis has been reviewed from Marcato et al. [65] and Pors et al.[66] (Table 3). Initially most of the studies reported ALDH1A1 as the signature marker for CSCs in many cancers, but its correlation with CSCs was overestimated because ALDEFLUOR™ does not specifically detect only ALDH1A1 activity. The capability of ALDH to process different aldehydes doesn't guarantee the isoform specificity for the ALDEFLUOR™, therefore detection of specific isoforms in CSC is under review and it need to be supported with different experiments to investigate the specific ALDH isoform biochemical profile.

Table 3. List of ALDH isoforms detected in various cancers and their correlation with cancer prognosis. Adapted from Rodriguez-Torres et al.[64]. IHC is immunohistochemistry, HNSCC is Head and Neck Squamous cell carcinoma.

Tumor Type	**Method of ALDH detection**	**ALDH isoform**	**Clinical observation**
Breast Cancer	ALDEFLUOR, IHC, Immunoblotting, qPCR	ALDH1 ALDH1A1 ALDH1A3	Poor clinical outcome, tumor recurrence, poor response to chemoterapy
Ovarian Cancer	ALDEFLUOR, IHC, Immunoblotting, qPCR	ALDH1 ALDH1A1	Poor clinical outcome
Brain cancer	ALDEFLUOR, IHC, qPCR, DNA methylation	ALDH1 ALDH1A1 ALDH1A3	Increased metastases
Bone cancer	ALDEFLUOR	ALDH	Poor clinical outcome
Prostate Cancer	ALDEFLUOR, IHC	ALDH7A1 ALDH3A1	Poor clinical outcome
HNSCC	ALDEFLUOR, IHC	ALDH1A1	Poor clinical outcome
Colorectal cancer	ALDEFLUOR, IHC	ALDH1 ALDH1A1	Loss of ALDH expression
Lung cancer	ALDEFLUOR, IHC immunoblotting	ALDH1A1 ALDH7A1	ALDH1A1 poor clinical outcome, recurrence
Cervical cancer	ALDEFLUOR	ALDH	Not assessed
Melanoma	ALDEFLUOR, Immunoblotting	ALDH1A1 ALDH1A3	Not assessed
Endometrial cancer	ALDEFLUOR, IHC	ALDH1 ALDH1A1	Not assessed
Renal cancer	IHC, Immunoblotting	ALDH1 ALDH1A1	Poor clinical outcome
Pancreatic cancer	pPCR, Chromatin, Immunoprecipitation	ALDH1A1 ALDH1A3	Not assessed
Heptobiliary cancer	ALDEFLUOR, IHC	ALDH1A3 ALDH3A1	Poor clinical outcome
Oesophageal cancer	ALDEFLUOR, IHC	ALDH1 ALDH1A1	Poor clinical outcome

Focusing on specific ALDH isoforms that concerned cancer research, ALDH1A3 is gaining more importance in CSC biology and detection. Marcato et al. in 2011 found that the overexpression of ALDH1A3, and not ALDH1A1 was predictive of metastasis in breast cancer stem cells[67]. In 2014 Shao et al. demonstrated the importance of ALDH1A3 for the maintenance of non-small lung cancer stem cells[68], whereas Mao et al. in 2013 found overexpression of ALDH1A3 in glioblastoma stem cells[69].

ALDH1A3 in glioma

The World Health Organization (WHO) classified gliomas in four histological grades by increasing degrees of undifferentiation, anaplasia, and aggressiveness[70]. Glioblastoma or High Grade Gliomas (HGGs) are classified as WHO grade III and IV and account 82% of gliomas[71]. Glioblastoma are highly invasive, infiltrative and recurrent brain tumors with a 5 years survival < 5%[71], [72]. The presence of a cancer stem cell population in glioma and in its more aggressive declinations has been proof, thereafter glioma stem cells (GSCs) are on the apex in a cellular tumorigenic hierarchy of HGGs[73], [74]. Microarray gene expression data, immunohistochemistry and biomarkers studies propose the existence of two glioma stem cell populations: the proneural glioma stem cells (PN GSCs) and the mesenchymal glioma stem cells (Mes GSCs)[69]. In the mesenchymal phenotype of HGG it has been demonstrated that the concurrent activation of two conflicting transcriptional regulators for neurogenesis and gliogenesis operates to permanently drive the aberrant mesenchymal phenotype in the context of the genetic and epigenetic changes that accompany high-grade gliomagenesis[75].

Mao et al. found high expression of ALDH1A3 isoform in GSCs. The expression of hALDH1A3 was detected by whole transcriptome microarray analysis and was found to be about 1000 times more in Mes GSCs compared to other cancer stem cells subpopulation (Figure 10). *In vivo* studies and xenograft experiments demonstrated an enhanced tumorigenicity of HGG derived from Mes GSCs.

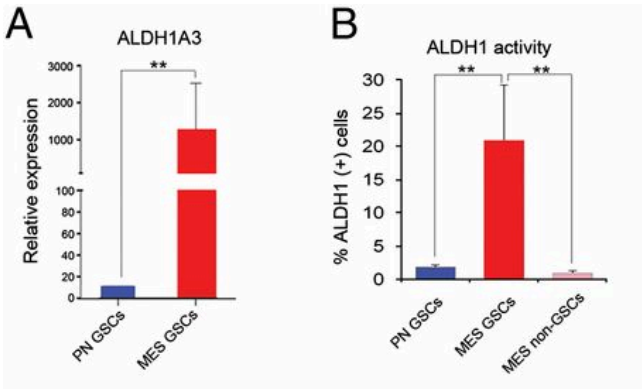


Figure 10. ALDH1A3 is a functional Mes GSC marker. (*A*) qRT-PCR analysis of ALDH1A3 expression in PN and Mes GSCs (***P* < 0.01). (*B*) FACS analysis using Aldefluor™. ALDH activities in PN GSCs (*n* = 3), Mes GSCs (*n* = 3), and non-GSCs (*n* = 3) derived from Mes GSCs (***P* < 0.01)[69].

The high activity (detected by ALDEFLUOR™) and the overexpression (detected by mRNA microarray) of ALDH1A3 in Mes GSCs are related with the high tumorigenicity, aggressiveness and worst prognosis of Mes GSC derived HGG. This research work lays the foundation for the validation of ALDH1A3 protein as interesting target in glioma.

ALDHs and cancer stem cell therapy resistance

Chemotherapy and radio resistance in tumors are the major difficulties that cancer research is facing and represent one of the most challenging barrier in cancer treatment. The resistance to treatments can be inherent or acquired by cancer cells, and mechanism such as DNA repair, drug efflux, drug inactivation and other adaptive responses are the known defenses of cancer cells against therapies[76] (Figure 11).

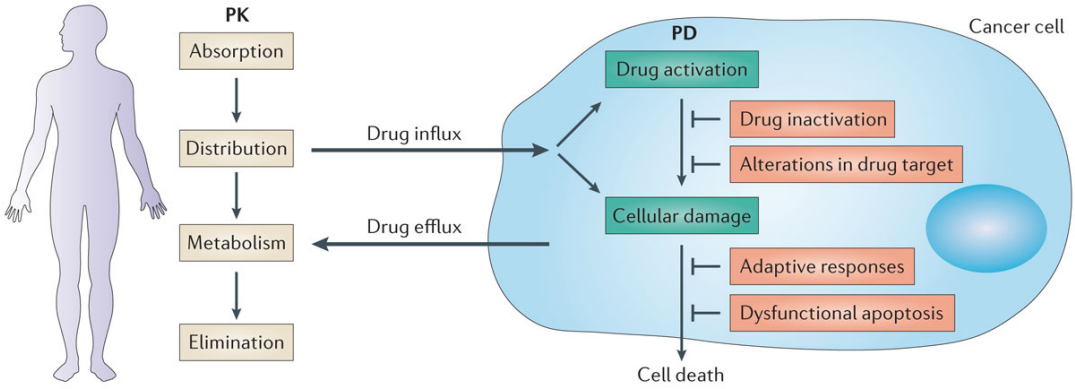


Figure 11. Overview on cancer drug resistance. Pharmacokinetics (PK) and Pharmodynamics (PD) are the two critical mechanisms where the drug can be limited on its effect in cancer treatment by drug resistance mechanism of the cell or absorption, distribution, metabolism and elimination factors. From Holohan et al.[76].

The plethora of mechanisms that tumor cells use to contrast cancer treatments includes CSCs that are often mentioned as one of the main actors in therapy resistance. Radio and chemo resistance of CSCs are in turn a consequence of a multiplicity of defense mechanisms such as DNA damage response, ABC transporters, slow cycling rate and high ALDH expression[12]. First evidence of cytotoxic drug resistance, due to ALDH high activity, was observed for cyclophosphamide treatment[12], in which ALDH1A1 has been observed to oxidize pivotal aldehyde intermediates of pro-drug cyclophosphamide and its analogues[77]. Other drugs such as doxorubicin, paclitaxel and A459-Taxol used in breast and lung cancer chemotherapy might be associated with high ALDH expression. It is still to be understood case by case if ALDHs are the cause of drug resistance or biomarkers of cancer cells that gain drug resistance[12]. The different situations in which ALDHs are involved in CSC biology can lead to different strategies: the development of selective inhibitors or the use of ALDH isoforms as biomarker and prognosis predictors. The structure determination of the ALDH isoforms, notably

in complex with its ligand and the structured-based design of small molecules are indispensable tools to investigate the ALDH biology itself or in relation to CSCs.

Bibliography

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

2 Yoshida, A., Rzhetsky, A., Hsu, L. C. and Chang, C. (1998) Human aldehyde dehydrogenase gene family. Eur. J. Biochem. FEBS **251**, 549–557.

3 Jackson, B., Brocker, C., Thompson, D. C., Black, W., Vasiliou, K., Nebert, D. W. and Vasiliou, V. (2011) Update on the aldehyde dehydrogenase gene (ALDH) superfamily. Hum. Genomics **5**, 283–303.

4 Marchitti, S. A., Brocker, C., Stagos, D. and Vasiliou, V. (2008) Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily.

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

6 Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D. and Vasiliou, V. (2012) Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. Pharmacol. Rev. **64**, 520–539.

7 LoPachin, R. M. and Gavin, T. (2014) Molecular mechanisms of aldehyde toxicity: a chemical perspective. Chem. Res. Toxicol. **27**, 1081–1091.

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

9 Brooks, P. J. and Theruvathu, J. A. (2005) DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. Alcohol Fayettev. N **35**, 187–193.

10 Burcham, P. C., Fontaine, F. R., Kaminskas, L. M., Petersen, D. R. and Pyke, S. M. (2004) Protein adduct-trapping by hydrazinophthalazine drugs: mechanisms of cytoprotection against acrolein-mediated toxicity. Mol. Pharmacol. **65**, 655–664.

11 Chen, Y., Koppaka, V., Thompson, D. C. and Vasiliou, V. (2012) Focus on molecules: ALDH1A1: from lens and corneal crystallin to stem cell marker. Exp. Eye Res. **102**, 105–106.

12 Januchowski, R., Wojtowicz, K. and Zabel, M. (2013) The role of aldehyde dehydrogenase (ALDH) in cancer drug resistance. Biomed. Pharmacother. Bioméd. Pharmacothérapie **67**, 669–680.
13 Croker, A. K. and Allan, A. L. (2012) Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44$^+$ human breast cancer cells. Breast Cancer Res. Treat. **133**, 75–87.
14 Deak, K. L., Dickerson, M. E., Linney, E., Enterline, D. S., George, T. M., Melvin, E. C., Graham, F. L., Siegel, D. G., Hammock, P., Mehltretter, L., et al. (2005) Analysis of ALDH1A2, CYP26A1, CYP26B1, CRABP1, and CRABP2 in human neural tube defects suggests a possible association with alleles in ALDH1A2. Birt. Defects Res. A. Clin. Mol. Teratol. **73**, 868–875.
15 Zhang, X., Zhang, Q.-Y., Liu, D., Su, T., Weng, Y., Ling, G., Chen, Y., Gu, J., Schilling, B. and Ding, X. (2005) Expression of cytochrome p450 and other biotransformation genes in fetal and adult human nasal mucosa. Drug Metab. Dispos. Biol. Fate Chem. **33**, 1423–1428.
16 Abouzeid, H., Favez, T., Schmid, A., Agosti, C., Youssef, M., Marzouk, I., El Shakankiry, N., Bayoumi, N., Munier, F. L. and Schorderet, D. F. (2014) Mutations in *ALDH1A3* Represent a Frequent Cause of Microphthalmia/Anophthalmia in Consanguineous Families. Hum. Mutat. **35**, 949–953.
17 Tsybovsky, Y., Donato, H., Krupenko, N. I., Davies, C. and Krupenko, S. A. (2007) Crystal structures of the carboxyl terminal domain of rat 10-formyltetrahydrofolate dehydrogenase: implications for the catalytic mechanism of aldehyde dehydrogenases. Biochemistry (Mosc.) **46**, 2917–2929.
18 Xiao, Q., Weiner, H. and Crabb, D. W. (1996) The mutation in the mitochondrial aldehyde dehydrogenase (ALDH2) gene responsible for alcohol-induced flushing increases turnover of the enzyme tetramers in a dominant fashion. J. Clin. Invest. **98**, 2027–2032.
19 Nagata, N., Hiyoshi, M., Shiozawa, H., Shiraishi, K., Watanabe, N., Tsuda, M. and Matsuzaki, S. (2002) Assessment of a difference in ALDH2 heterozygotes and alcoholic liver injury. Alcohol. Clin. Exp. Res. **26**, 11S–14S.
20 Black, W., Chen, Y., Matsumoto, A., Thompson, D. C., Lassen, N., Pappa, A. and Vasiliou, V. (2012) Molecular mechanisms of ALDH3A1-mediated cellular protection against 4-hydroxy-2-nonenal. Free Radic. Biol. Med. **52**, 1937–1944.
21 Rizzo, W. B. (2014) Fatty aldehyde and fatty alcohol metabolism: review and importance for epidermal structure and function. Biochim. Biophys. Acta **1841**, 377–389.
22 Gaboon, N. E. A., Jelani, M., Almramhi, M. M., Mohamoud, H. S. A. and Al-Aama, J. Y. (2015) Case of Sjögren-Larsson syndrome with a large deletion in the *ALDH3A2* gene confirmed by single nucleotide polymorphism array analysis. J. Dermatol. **42**, 706–709.

23 Hu, C. A., Lin, W. W. and Valle, D. (1996) Cloning, characterization, and expression of cDNAs encoding human delta 1-pyrroline-5-carboxylate dehydrogenase. J. Biol. Chem. **271**, 9795–9800.
24 Geraghty, M. T., Vaughn, D., Nicholson, A. J., Lin, W. W., Jimenez-Sanchez, G., Obie, C., Flynn, M. P., Valle, D. and Hu, C. A. (1998) Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. Hum. Mol. Genet. **7**, 1411–1415.
25 Tillakaratne, N. J., Medina-Kauwe, L. and Gibson, K. M. (1995) gamma-Aminobutyric acid (GABA) metabolism in mammalian neural and nonneural tissues. Comp. Biochem. Physiol. A Physiol. **112**, 247–263.
26 Pearl, P. L., Parviz, M., Vogel, K., Schreiber, J., Theodore, W. H. and Gibson, K. M. (2014) Inherited disorders of gamma-aminobutyric acid metabolism and advances in ALDH5A1 mutation identification. Dev. Med. Child Neurol.
27 Marcadier, J. L., Smith, A. M., Pohl, D., Schwartzentruber, J., Al-Dirbashi, O. Y., FORGE Canada Consortium, Majewski, J., Ferdinandusse, S., Wanders, R. J. A., Bulman, D. E., et al. (2013) Mutations in ALDH6A1 encoding methylmalonate semialdehyde dehydrogenase are associated with dysmyelination and transient methylmalonic aciduria. Orphanet J. Rare Dis. **8**, 98.
28 Chang, Y. F., Ghosh, P. and Rao, V. V. (1990) L-pipecolic acid metabolism in human liver: L-alpha-aminoadipate delta-semialdehyde oxidoreductase. Biochim. Biophys. Acta **1038**, 300–305.
29 Yang, Z., Yang, X., Wu, Y., Wang, J., Zhang, Y., Xiong, H., Jiang, Y. and Qin, J. (2014) Clinical diagnosis, treatment, and ALDH7A1 mutations in pyridoxine-dependent epilepsy in three Chinese infants. PloS One **9**, e92803.
30 Lin, M. and Napoli, J. L. (2000) cDNA cloning and expression of a human aldehyde dehydrogenase (ALDH) active with 9-cis-retinal and identification of a rat ortholog, ALDH12. J. Biol. Chem. **275**, 40106–40112.
31 Kikonyogo, A. and Pietruszko, R. (1996) Aldehyde dehydrogenase from adult human brain that dehydrogenates gamma-aminobutyraldehyde: purification, characterization, cloning and distribution. Biochem. J. **316 (Pt 1)**, 317–324.
32 Hanna, M. C. and Blackstone, C. (2009) Interaction of the SPG21 protein ACP33/maspardin with the aldehyde dehydrogenase ALDH16A1. Neurogenetics **10**, 217–228.
33 Hu, C. -a. A., Khalil, S., Zhaorigetu, S., Liu, Z., Tyler, M., Wan, G. and Valle, D. (2008) Human Delta1-pyrroline-5-carboxylate synthase: function and regulation. Amino Acids **35**, 665–672.
34 Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. and Baker, E. N. (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Struct. Lond. Engl. 1993 **6**, 1541–1551.

35 Steinmetz, C. G., Xie, P., Weiner, H. and Hurley, T. D. (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. Struct. Lond. Engl. 1993 **5**, 701–711.
36 Kedishvili, N. Y., Popov, K. M., Rougraff, P. M., Zhao, Y., Crabb, D. W. and Harris, R. A. (1992) CoA-dependent methylmalonate-semialdehyde dehydrogenase, a unique member of the aldehyde dehydrogenase superfamily. cDNA cloning, evolutionary relationships, and tissue distribution. J. Biol. Chem. **267**, 19724–19729.
37 Feldman, R. I. and Weiner, H. (1972) Horse liver aldehyde dehydrogenase. II. Kinetics and mechanistic implications of the dehydrogenase and esterase activity. J. Biol. Chem. **247**, 267–272.
38 Jones, K. H., Lindahl, R., Baker, D. C. and Timkovich, R. (1987) Hydride transfer stereospecificity of rat liver aldehyde dehydrogenases. J. Biol. Chem. **262**, 10911–10913.
39 Liu, Z. J., Sun, Y. J., Rose, J., Chung, Y. J., Hsiao, C. D., Chang, W. R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J., et al. (1997) The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat. Struct. Biol. **4**, 317–326.
40 Bordelon, T., Montegudo, S. K., Pakhomova, S., Oldham, M. L. and Newcomer, M. E. (2004) A disorder to order transition accompanies catalysis in retinaldehyde dehydrogenase type II. J. Biol. Chem. **279**, 43085–43091.
41 Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F. and Rahuel-Clermont, S. (2013) Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. Chem. Biol. Interact. **202**, 78–84.
42 Keller, M. A., Zander, U., Fuchs, J. E., Kreutz, C., Watschinger, K., Mueller, T., Golderer, G., Liedl, K. R., Ralser, M., Kräutler, B., et al. (2014) A gatekeeper helix determines the substrate specificity of Sjögren-Larsson Syndrome enzyme fatty aldehyde dehydrogenase. Nat. Commun. **5**, 4439.
43 Wald, G. (1968) The molecular basis of visual excitation. Nature **219**, 800–807.
44 Perlmann, T. (2002) Retinoid metabolism: a balancing act. Nat. Genet. **31**, 7–8.
45 White, J. A., Ramshaw, H., Taimi, M., Stangle, W., Zhang, A., Everingham, S., Creighton, S., Tam, S. P., Jones, G. and Petkovich, M. (2000) Identification of the human cytochrome P450, P450RAI-2, which is predominantly expressed in the adult cerebellum and is responsible for all-trans-retinoic acid metabolism. Proc. Natl. Acad. Sci. U. S. A. **97**, 6403–6408.
46 Kiser, P. D., Golczak, M., Maeda, A. and Palczewski, K. (2012) Key enzymes of the retinoid (visual) cycle in vertebrate retina. Biochim. Biophys. Acta **1821**, 137–151.
47 Das, B. C., Thapa, P., Karki, R., Das, S., Mahapatra, S., Liu, T.-C., Torregroza, I., Wallace, D. P., Kambhampati, S., Van Veldhuizen, P., et al. (2014) Retinoic

acid signaling pathways in development and diseases. Bioorg. Med. Chem. **22**, 673–683.
48 Napoli, J. L. (1996) Retinoic acid biosynthesis and metabolism. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. **10**, 993–1001.
49 Maden, M. (2000) The role of retinoic acid in embryonic and post-embryonic development. Proc. Nutr. Soc. **59**, 65–73.
50 Petkovich, M., Brand, N. J., Krust, A. and Chambon, P. (1987) A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature **330**, 444–450.
51 Giguere, V., Ong, E. S., Segui, P. and Evans, R. M. (1987) Identification of a receptor for the morphogen retinoic acid. Nature **330**, 624–629.
52 Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J. and Gronemeyer, H. (2006) International Union of Pharmacology. LX. Retinoic acid receptors. Pharmacol. Rev. **58**, 712–725.
53 Niederreither, K. and Dollé, P. (2008) Retinoic acid in development: towards an integrated view. Nat. Rev. Genet. **9**, 541–553.
54 Liu, T. X., Zhang, J. W., Tao, J., Zhang, R. B., Zhang, Q. H., Zhao, C. J., Tong, J. H., Lanotte, M., Waxman, S., Chen, S. J., et al. (2000) Gene expression networks underlying retinoic acid-induced differentiation of acute promyelocytic leukemia cells. Blood **96**, 1496–1504.
55 Clarke, N., Germain, P., Altucci, L. and Gronemeyer, H. (2004) Retinoids: potential in cancer prevention and therapy. Expert Rev. Mol. Med. **6**.
56 Janesick, A., Wu, S. C. and Blumberg, B. (2015) Retinoic acid signaling and neuronal differentiation. Cell. Mol. Life Sci. CMLS **72**, 1559–1576.
57 Karsy, M., Albert, L., Tobias, M. E., Murali, R. and Jhanwar-Uniyal, M. (2010) All-trans retinoic acid modulates cancer stem cells of glioblastoma multiforme in an MAPK-dependent manner. Anticancer Res. **30**, 4915–4920.
58 Bonnet, D. and Dick, J. E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. **3**, 730–737.
59 Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. and Clarke, M. F. (2003) Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. U. S. A. **100**, 3983–3988.
60 Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., Fiocco, R., Foroni, C., Dimeco, F. and Vescovi, A. (2004) Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. Cancer Res. **64**, 7011–7021.
61 Jordan, C. T., Guzman, M. L. and Noble, M. (2006) Cancer stem cells. N. Engl. J. Med. **355**, 1253–1261.
62 Medema, J. P. (2013) Cancer stem cells: the challenges ahead. Nat. Cell Biol. **15**, 338–344.

63 Alison, M. R., Guppy, N. J., Lim, S. M. L. and Nicholson, L. J. (2010) Finding cancer stem cells: are aldehyde dehydrogenases fit for purpose? J. Pathol. **222**, 335–344.
64 Rodriguez-Torres, M. and Allan, A. L. (2015) Aldehyde dehydrogenase as a marker and functional mediator of metastasis in solid tumors. Clin. Exp. Metastasis.
65 Marcato, P., Dean, C. A., Giacomantonio, C. A. and Lee, P. W. K. (2011) Aldehyde dehydrogenase: its role as a cancer stem cell marker comes down to the specific isoform. Cell Cycle Georget. Tex **10**, 1378–1384.
66 Pors, K. and Moreb, J. S. (2014) Aldehyde dehydrogenases in cancer: an opportunity for biomarker and drug development? Drug Discov. Today **19**, 1953–1963.
67 Marcato, P., Dean, C. A., Pan, D., Araslanova, R., Gillis, M., Joshi, M., Helyer, L., Pan, L., Leidal, A., Gujar, S., et al. (2011) Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. Stem Cells Dayt. Ohio **29**, 32–45.
68 Shao, C., Sullivan, J. P., Girard, L., Augustyn, A., Yenerall, P., Rodriguez-Canales, J., Liu, H., Behrens, C., Shay, J. W., Wistuba, I. I., et al. (2014) Essential role of aldehyde dehydrogenase 1A3 for the maintenance of non-small cell lung cancer stem cells is associated with the STAT3 pathway. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. **20**, 4154–4166.
69 Mao, P., Joshi, K., Li, J., Kim, S.-H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U. R., Benos, P. V., Smith, L., et al. (2013) Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. Proc. Natl. Acad. Sci. U. S. A. **110**, 8644–8649.
70 Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W. and Kleihues, P. (2007) The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. (Berl.) **114**, 97–109.
71 Omuro, A. and DeAngelis, L. M. (2013) Glioblastoma and other malignant gliomas: a clinical review. JAMA **310**, 1842–1850.
72 Ostrom, Q. T., Bauchet, L., Davis, F. G., Deltour, I., Fisher, J. L., Langer, C. E., Pekmezci, M., Schwartzbaum, J. A., Turner, M. C., Walsh, K. M., et al. (2014) The epidemiology of glioma in adults: a "state of the science" review. Neuro-Oncol. **16**, 896–913.
73 Heywood, R. M., Marcus, H. J., Ryan, D. J., Piccirillo, S. G. M., Al-Mayhani, T. M. F. and Watts, C. (2012) A review of the role of stem cells in the development and treatment of glioma. Acta Neurochir. (Wien) **154**, 951–969; discussion 969.
74 Cheng, L., Huang, Z., Zhou, W., Wu, Q., Donnola, S., Liu, J. K., Fang, X., Sloan, A. E., Mao, Y., Lathia, J. D., et al. (2013) Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. Cell **153**, 139–152.

75 Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., Sulman, E. P., Anne, S. L., Doetsch, F., Colman, H., et al. (2010) The transcriptional network for mesenchymal transformation of brain tumours. Nature **463**, 318–325.
76 Holohan, C., Van Schaeybroeck, S., Longley, D. B. and Johnston, P. G. (2013) Cancer drug resistance: an evolving paradigm. Nat. Rev. Cancer **13**, 714–726.
77 Sládek, N. E., Kollander, R., Sreerama, L. and Kiang, D. T. (2002) Cellular levels of aldehyde dehydrogenases (ALDH1A1 and ALDH3A1) as predictors of therapeutic responses to cyclophosphamide-based chemotherapy of breast cancer: a retrospective study. Rational individualization of oxazaphosphorine-based cancer chemotherapeutic regimens. Cancer Chemother. Pharmacol. **49**, 309–321.

Outline of the thesis

Aldehyde dehydrogenases (ALDHs) are detoxifying enzymes that uses NAD(P) as cofactor to oxidizes aldehydes into their corresponding carboxylic acids. The human ALDH superfamily consist of 19 isoforms classified in families and subfamilies by sequence identity. Human ALDHs comprise crucial enzymes involved in many metabolic and signaling processes such as the retinoid metabolism, the neurotransmitter (GABA) biosynthesis and osmosis (betaine) regulation.

ALDH1A1, ALDH1A2 and ALDH1A3 are the three cytosolic isoenzymes composing the ALDH1A subfamily. They are key regulators of the biosynthesis of retinoic acid (REA) that is an important signaling molecule at the basis of the embryonic development and tissue differentiation. ALDH1A1 and ALDH1A3 are the two isoforms that are gaining importance in cancer studies for their role in Cancer Stem Cells (CSCs) biology. CSCs are a reservoir of tumorigenic cells with self-renewal/differentiation capacity, high tolerance to radio and chemotherapy and with the ability to initiate and propagate tumors. The intrinsic CSCs high-ALDH activity is exploited for their identification and selection through ALDEFLUOR™. In summary, the high expression of ALDH isozymes in each cancer type is likely correlated to CSCs presence, worst outcome of the tumor and drug resistance. Focusing on ALDH1A3, its high expression and activity has been detected in mesenchymal Glioma Stem Cells (Mes GSCs) of High Grade Gliomas (HGGs) that are the most aggressive brain cancers, affecting people of all ages. The primary treatment for glioma is the surgical removal however, because it is an invasive and aggressive cancer, the neurosurgery can't remove the entire tumor mass and doesn't avoid the return of the glioma. The high invasive nature of HGG and its recurrence have been associated with the presence of Glioma Stem Cells (GSCs), in which ALDH1A3 seems to play a critical role.

In the present Ph.D. dissertation, we provide the biochemical characterization and the X-ray crystal structure of hALDH1A3 in complex with its ligands. Steady state kinetic constants for hALDH1A3 showed Michaelis-Menten kinetics and confirmed the preference of the enzyme for its natural substrate all-*trans* retinal. Moreover the interaction of the enzyme with its ligands was investigated by Thermofluor assay. Thermal shift data underline a high affinity of ALDH1A3 for the oxidized cofactor and evidence a certain grade of affinity also for the products of the reaction. The structure of ALDH1A3 in complex with the cofactor and the natural product was solved by means of X-ray crystallography at a resolution up to 2.9 Å. The ALDH1A3 crystal asymmetric unit comprises 8 polypeptides arranged in two distinct catalytic tetramers. Each ALDH1A3 monomer binds one molecule of NAD^+, and in addition electron density corresponding to product REA was found in complex with four chains.

Structural investigations on hALDH1A3, together with its biochemical characterization, are useful tools to help unravel the role of this enzyme in GSCs biology. The reported structure, notably in complex with REA and NAD, was the necessary preliminary data for the first *in silico* screening of commercially available compounds with potential activity as inhibitors. Studies on inhibition and selectivity of compound hits resulting from the virtual screening are on the way. The structure-based design of small ligands is a critical step for the identification of a lead compound or for the modeling of an isoform specific fluorescent probe for prognosis and diagnostic purposes.

Crystal structure and biochemical characterization of human ALDH1A3 isoenzyme in complex with NAD^+ and retinoic acid.

Andrea Moretti[1], Jianfeng Li[2], Stefano Donini[1], Robert W Sobol[2], Menico Rizzi[1*] and Silvia Garavaglia[1*].

[1]Department of Pharmaceutical Sciences, University of Piemonte Orientale, Largo Donegani 2, 28100 Novara, Italy; [2]Department of Oncologic Sciences, Molecular & Metabolic Oncology Program, University of South Alabama Mitchell Cancer Institute, 1660 Springhill Avenue, Mobile, AL 36604, USA.

***Corresponding authors**

Silvia Garavaglia, [1]Department of Pharmaceutical Sciences, University of Piemonte Orientale, Via Bovio 6, 28100 Novara, Italy

Phone: +39 0321 375714

Fax: +39 0321 375821

Email: silvia.garavaglia@uniupo.it

Menico Rizzi, [1]Department of Pharmaceutical Sciences, University of Piemonte Orientale, Via Bovio 6, 28100 Novara, Italy

Phone: +39 0321 375712

Fax: +39 0321 375821

Email: menico.rizzi@uniupo.it

Keywords: Retinoic Acid/ Aldehyde Dehydrogenase/ ALDH1 family/ NAD cofactor/ Glioma/ Cancer Stem Cells.

ABSTRACT

The aldehyde dehydrogenase family 1 member A3 (ALDH1A3, E.C. 1.2.1.5) catalyzes the oxidation of all-*trans* retinal to retinoic acid (REA) using NAD^+ as cofactor. The human ALDH1A3 plays a key role in embryonic development and tissue differentiation by regulating the synthesis of the pleiotropic retinoic acid. ALDHs high enzymatic activity is a marker of Cancer Stem Cells (CSCs) population in some tumors and it seems to be related to cell proliferation, survival and chemoresistance of CSCs. The detection of ALDH1A3 high expression in mesenchymal Glioma Stem Cells (Mes GSCs), with respect to ALDH1A1 and ALDH1A2 isoforms, addresses potential medical interest on the isoenzyme as prognosis marker or inhibitors target.

Here we present the first X-ray structure of human recombinant ALDH1A3 co-crystallized with NAD and REA. The structure model reveals a new conformation of NAD^+ and two different conformations of REA in complex with the enzyme. In addition our work contributes to complete the functional characterization of ALDH1A3 by reporting steady state kinetic constants and melting temperatures (Tm) of the enzyme with its ligands, measured by continuous spectrophotometric assay and thermofluor experiments, respectively. In summary, our data provide a first comprehensive structural view of ALDH1A3 combined with its biochemical characterization. These results can be a valuable tool for structure-based drug design of ALDH1A3 potential inhibitors and for the detection of ALDH1A3 isoform in glioma as prognosis marker.

INTRODUCTION

Living organisms are constantly confronted with oxidative stress and with the reactive oxygen species (ROS) derived therefrom. In animals a plethora of mechanisms contribute to ROS formation and the superfamily of Aldehyde dehydrogenase enzymes (ALDHs) are known to decrease oxidative stress, particularly that caused by aldehydes. During numerous physiological processes, aldehydes are generated from a variety of precursors, including the endogenous biotransformation of most compounds such as amino acids, neurotransmitters, carbohydrates and lipids[1], [2]. Aldehydes are ubiquitous molecules present in the environment from natural sources and from pollution and cigarette smoke. Moreover, they are approved additives in various foods and are also used or generated in a wide variety of industrial processes. As some aldehydes play vital roles in normal physiological processes like vision, embryonic development, and neurotransmission, many are cytotoxic and carcinogenic[3]. ALDHs enzymes are required for the clearance of potentially toxic aldehydes, and are essential for the production of key metabolic regulators[2]. Most of the ALDHs have wide tissue distribution displaying distinct substrate specificities[2], [4]. The human ALDH superfamily consists of 19 putatively functional genes with different chromosomal locations[5]. A standardized gene nomenclature system based on divergent evolution and amino acid identity was established for the ALDH superfamily in 1998[6]. The ALDH enzymes catalyze the $NAD(P)^+$-dependent irreversible oxidation of a wide spectrum of endogenous and exogenous aldehydes and are found in all subcellular regions including cytosol, mitochondria, endoplasmic reticulum and nucleus, with several of those found in more than one compartment[7].

In particular, ALDH1A1, ALDH1A2, ALDH1A3, which are the three members of the cytosolic aldehyde dehydrogenase class 1 of higher vertebrates, have a key role in vertebrate development. Indeed, they show a high specificity for retinaldehyde,

oxidizing it to the powerful differentiation factor retinoic acid (REA)[8]. The retinoic-acid signaling pathway in vertebrates utilizes two classes of retinoid receptors, RARs and RXRs that belong to the family of nuclear hormone receptors. These proteins are ligand-regulated transcription factors that bind 9-*cis* (RXR, RAR) or all-*trans* (RAR) retinoic acid via a ligand-binding domain, and direct the transcription of target genes via a DNA-binding domain[9], [10]. Retinoic acid is derived from vitamin A (retinol) and its pleiotropic effects include spinal chord and retina development during embryogenesis, neuronal cell differentiation and maintenance of epithelial cell type in adult tissues[11]. Synthesis of REA proceeds through two steps: the oxidation of retinol by alcohol dehydrogenase producing the relative aldehyde, followed by its irreversible conversion to acid using retinal dehydrogenases[12]. The ALDH1A3 sequences shares more than 70% amino acid conservation with ALDH1A1 and ALDH1A2 and major amino acid residues found in the catalytically active sites of ALDHs are conserved. Despite processing the same substrate the expression of ALDH1A1, 1A2 and 1A3 are not entirely redundant considering their preference for specific substrates.

Indeed, ALDH1A1 shows substrate preferences for the aldehydes products of lipid peroxidation as well as retinaldehyde. Otherwise ALDH1A2 and ALDH1A3 exhibit the highest substrate specificity and catalytic efficiency for retinal oxidation to REA[13]. In addition, ALDH1A3, although capable to catalyzes the oxidation of *cis*-retinal to REA, displays a strong specificity for *all*-trans retinal with an efficiency that is 10-fold higher than that of ALDH1A1 and ALDH1A2[14]. In summary, the most efficient retinal dehydrogenase activity was reported for the cytosolic enzyme ALDH1A3 that oxidizes both all-*trans*-retinal and 9-*cis*-retinal to REA[14]. Moreover, such high specificity of ALDH1A3 for all-*trans*-retinal indicates the requirement of restricted *all*-trans configuration of the carbon side chain in the retinal substrate characterizing this isoenzyme with respect to the other two. The ALDH1A3 is differentially activated during early embryonic head and forebrain development, and it is expressed at high levels in the differentiating

keratinocytes of human and murine hair shafts. Otherwise a low to negative expression level of ALDH1A3 was reported for the interfollicular epidermis[15], [16]. Moreover, knockout of the murine ALDH1A3 gene was associated with perinatal lethality that could be rescued by maternal treatment with REA[17].

The milieu of cancerous tumors consists of heterogeneous cell population and increasing evidence indicates Cancer Stem Cells (CSCs) as the self-renewal and differentiation population of cells able to driving tumorigenesis[18]. In this contest, high aldehyde dehydrogenases activity is being scrutiny as a potential prognostic marker for cancer. Indeed, the ALDHs enzymatic activity was found modified in CSCs population and seems to be correlate with a poor outcome for many solid tumors[19]. Recently, ALDH1A1 was found to be expressed in Aldefluor™-positive CSCs from human melanoma and breast cancer. As such, it is unlikely that the ALDH enzymatic activity measured by Aldefluor™ is due only to ALDH1A1. Recent studies found that tumor aggressiveness and metastasis in human breast cancer were correlated with ALDH1A1 instead of ALDH1A3[20], [21]**.** This finding underscores the importance of defining the ALDH isoenzymes responsible for the high ALDHs activity in each cancer type and provides compelling evidence in favor of the ALDH isoenzymes functioning as key molecules governing cell proliferation, survival and chemoresistance of CSCs[22]. Given the accumulating and evolving evidence of ALDH role in CSC biology, the selective suppression of ALDH isozymes and genes may be a promising frontier for CSC-directed therapeutics in human cancers. It is hoped that this may spur the development of novel and highly specific inhibitors[23], [24].

In this contest, recently, it has been demonstrated the elevated expression of human ALDH1A3 in mesenchymal Glioma Stems Cells (Mes GSCs) with respect to human ALDH1A1 and all other ALDHs. Mao et al. investigate GSCs by transcriptome array analysis and found high expression of hALDH1A3 isoform in High Grade Gliomas (HGGs) generated form Mes GSCs[25]. The hALDH1A3 is highly expressed in clinical HGGs but not in low-grade glioma (LGGs) and in

normal brain samples. In particular, it is possible to classify two distinct populations of GSCs in HGGs: the proneuronal glioma stem cells (PN GSCs) and mesenchymal glioma stem cells (Mes GSC). The HGG derived from Mes GSCs in which hALDH1A3 is significantly up regulated displays a higher radioresistance and aggressiveness compared to HGG derived from PN GSCs. In addition, irradiation induces hALDH1A3 up-regulation in treated cancer cells transforming PN GSCs derived gliomas into tumor cells that are highly resistant to radiation treatment[24]. Taken together, this data suggest that subtypes of GSCs in clinical HGG tumor tissues are identifiable by up-regulation of ALDH1A3 activity that distinguishes Mes GSCs from PN GSCs[25].

Here we report on the first crystal structure of the human ALDH1A3 enzyme in complex with REA and NAD^+, which has been determined at a 2.9 Å resolution. Such structural analysis is the instrument in order to better understand the differences in substrate specificity between ALDH1A subfamily enzymes, and complete the overall view of the catalytic and metabolic function of this important family of enzymes. The structural determination of ALDH1A3 in complex with its ligands is the critical issue to start planning a structure based design of potential inhibitors with medical interest and to generate fluorescent probes that should be useful for early diagnosis in glioma and other tumors expressing ALDH1A3.

EXPERIMENTAL PROCEDURES

Expression and purification of human ALDH1A3

The human ALDH1A3 gene was amplified by PCR from mesenchymal Glioma Stem Cell cDNA pool with the synthetic oligonucleotides primers 5'-CACCGCCACCGCTAACGGGGCCGTG-3' and 5'-TCAGGGGTTCTTGTCGCCAAGTTTGATGGTGACAGT-3'. The full-length hALDH1A3 was cloned into the destination vector pDEST17 that provides an N-term 6xHis tag and transformed in. *E. coli* strain BL21(DE3) (Novagen). The next day, colonies were scraped and inoculated in one liter 2xTY medium supplemented with 50 μg/mL ampicillin. When OD_{600} of 0.6–0.8 was reached, the temperature was shifted to 20°C to induce the recombinant protein production. Induced cells collected by centrifugation, were resuspended in 1/25 original volume of lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 1mM β-mercaptoethanol, 10 mM imidazole at pH 7.5) supplemented with 250U of Benzonase nuclease. Afterwards, bacterial cells were fragmented by applying a high pressure through the use of the French Press. Protease inhibitor cocktail was added to the crude extract and the lysate was clarified with centrifugation at 39,000 rcf. To evaluate the purity and homogeneity of the protein, after each purification step, eluted fractions were analyzed by SDS-PAGE and the protein quantification was determined by Bradford protein assay. In the first purification step, the soluble fraction of recombinant human 6xHis-ALDH1A3 was purified with a Qiagen Ni-NTA Superflow 5 mL cartridge. The supernatant was loaded on the Ni-NTA column, previously equilibrated with 10 column volumes of lysis buffer. The Ni-NTA cartridge was washed with 50 mM Na_2HPO_4, 300 mM NaCl, 1mM β-mercaptoethanol, 50 mM imidazole pH 7.5 until the absorbance (A) at 280 nm return to the baseline (15 column volumes). The recombinant protein was eluted by a linear gradient in 20 column volumes with the buffer as follow: 50 mM Na_2HPO_4, 300 mM NaCl, 1mM β-mercaptoethanol, 250 mM imidazole pH 8.

Eluted fractions were pooled and concentrated to 5 mg/ml with Merck Millipore Amicon Ultra-15 10 kDa and loaded on a Sephacryl S200 16/60 column on AKTA FPLC system. The protein was eluted and stored in buffer containing 20 mM HEPES pH 7.5, 150 mM KCl, 1mM β-mercaptoethanol, 0.5 mM EDTA. This procedure allowed us to obtain from one liter of bacterial culture 25 mg of pure and active hALDH1A3 used for crystallization trials and kinetics analysis.

Enzyme kinetic analysis and thermal shift assay

The hALDH1A3 activity was followed by continuously measuring the NADH formation at 340 nm (molar extinction coefficient of 6220 M^{-1} cm^{-1}) for 30 minutes at 25°C with Varian Cary® 50 UV-Vis spectrophotometer. The enzymatic assays were performed in a total volume of 200 µl in 20 mM Tris HCl pH 8.0, 1 mM β-mercaptoethanol, 150 mM KCl, with 2.6 µM of pure recombinant ALDH1A3. The biochemical characterization of the enzyme was performed by measuring the enzymatic steady state kinetics at variable concentrations of all-*trans* retinal (from 0.02 µM to 60 µM), acetaldehyde (from 0.05 mM to 20 mM) and NAD from 0.2 µM to 10 mM. Michaelis-Menten kinetics[26] analysis of three independent experiments was performed using SigmaPlot (Systat Software, San Jose, CA). All the compounds were purchased from Sigma Aldrich and were dissolved in water except all-*trans* retinal that was dissolved in DMSO. Each assay was pre-incubated without enzyme for one minute, than the ALDH1A3 was added to initiate the reaction.

For the thermofluor assay, ALDH1A3 was diluted to a final concentration of 1 mg/mL using 20 mM HEPES pH = 7.5, 150 mM KCl, 1 mM β-mercaptoethanol, 0.5 mM EDTA as the buffer solution. The thermal shift assay was performed using a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad) with FAM emission wavelength. The melting temperature (Tm) were recorded in triplicates and calculated as described by Matulis et al.[27]. The Tm was calculated as the inflection point of the sigmoidal melt curve and increased Tm indicates increased global

stability. For the thermal shift analysis, ALDH1A3 was mixed with an appropriate amount of stock solutions of NAD^+, NADH, all-trans retinal and retinoic acid to a final concentration of 1 mM and with the fluorescence probe SYPRO® Orange (Sigma–Aldrich) at 1/4000 dilution. Data were harvested and analyzed using CFX Manager™ Software (Bio-Rad) and SigmaPlot.

Crystallization and Structure Determination

Crystals of hALDH1A3 were obtained by using the vapor-diffusion technique in a sitting drop and applying a spare-matrix-based strategy on a crystallization robot (Oryx4, Douglas Instruments) with commercially available crystallization kits (Qiagen, Hamapton). Best crystals grown by mixing 0.5 μl of protein solution at 5.5 mg/mL, pre-incubated with 1 mM NAD and 1 mM retinoic acid, with an equal volume of a reservoir solution containing 20% PEG 3350, 0.24 M Na_2 malonate pH 7.0, 10 mM TCEP hydrochloride and equilibrated against 50 μl of the reservoir solution, at 20 °C in about 30 days. For X-ray data collection, crystals were quickly equilibrated in a solution containing the crystallization buffer and 12.5% glycerol as cryo-protectant and flash frozen at 100 K in liquid nitrogen. A complete dataset for human ALDH1A3 was collected up to 2.9Å resolution at the ID23 EH1 beamline of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Analysis of the diffraction data set, allowed us to assign the crystal to the monoclinic P 1 2_1 1 space group with cell dimensions a = 82.66 Å b = 159.79 Å and c = 155.65 Å, containing eight molecules per asymmetric unit with a corresponding solvent content of 55.5%. Diffraction data were integrated using the program package XDS[28] and scaled with the CCP4 suite of program[29]. The structure determination of hALDH1A3 was carried out by means of the molecular replacement technique using the coordinates of *Ovis aries* ALDH1 tetramer as the search model (Protein Data Bank ID code 1BXS). The program PHASER[30] was used to automatically determine the hALDH1A3 structure and returned a unique molecular replacement solution with an LLG of 4716.971 and TFZ=11.3. The

resulting electron density map of human ALDH1A3 allowed automatic tracing by the program AUTOBUILDING[31]. The initial model was subjected to iterative cycles of crystallographic refinement with the programs REFMAC5[32] and PHENIX.REFINE[33] alternated with manual graphic session for model building using the program Coot[34]. 5% randomly chosen reflections were excluded from refinement of the structure and used for the Free R factor calculation[35]. The program ARP/wARP[36] was used for adding solvent molecules. Refinement was continued until convergence to R factor and Free R factor to 0.224 and 0.286 respectively and ideal geometry. Data collection and refinement statistics are given in Table 1. The stereochemistry of the model has been assessed with the program PROCHECK[37]. Consensus amino acids sequence for ALDH1A3 start at 1 resulting in a shift of +12 compared to the reference sequence of mature ALDH2

Illustrations

Figures were generated by using the program pymol[38] and LigPlot+[39].

Table 1. Data collection and refinement statistics for hALDH1A3 crystal.

	hALDH1A3 crystal
DATA COLLECTION	
Space group	P 1 2_1 1
Cell dimensions	
a, b, c (Å)	82.66, 159.79, 177.65
α,β,γ (°)	90, 93.69, 90
Resolution (Å)	47.51 - 2.9
R_{pim} / R_{merge}	0.109(0.490)/0.184(0.785)
Mean(I) / sd(I)	5.4 (1.6)
Completeness (%)	98.3 (97.3)
Redundancy	3.6 (3.2)
REFINEMENT	
Resolution (Å)	2.9
No. reflections	100139 (9865)
R_{work} / R_{free}	0.2244 / 0.2869
No. atoms	
Protein	29498
Water	66
Ligands	462
Mean B-factors	
Protein ($Å^2$)	40.5
Water ($Å^2$)	13.8
Ligands ($Å^2$)	45.8
R.m.s deviations	
Bond lengths (Å)	0.017
Bond angles (°)	1.74
Ramachandran outliers (%)	0.63

RESULTS AND DISCUSSION

Biochemical characterization

The kinetic characterization of hALDH1A3 was performed in presence of NAD^+ varying the concentration of the natural substrate all-*trans* retinal and acetaldehyde, a substrate accepted among different ALDH isoforms[40]. Plots of initial velocity versus either substrates or NAD^+ concentrations showed Michaelis–Menten kinetics, summarized in Table 2.

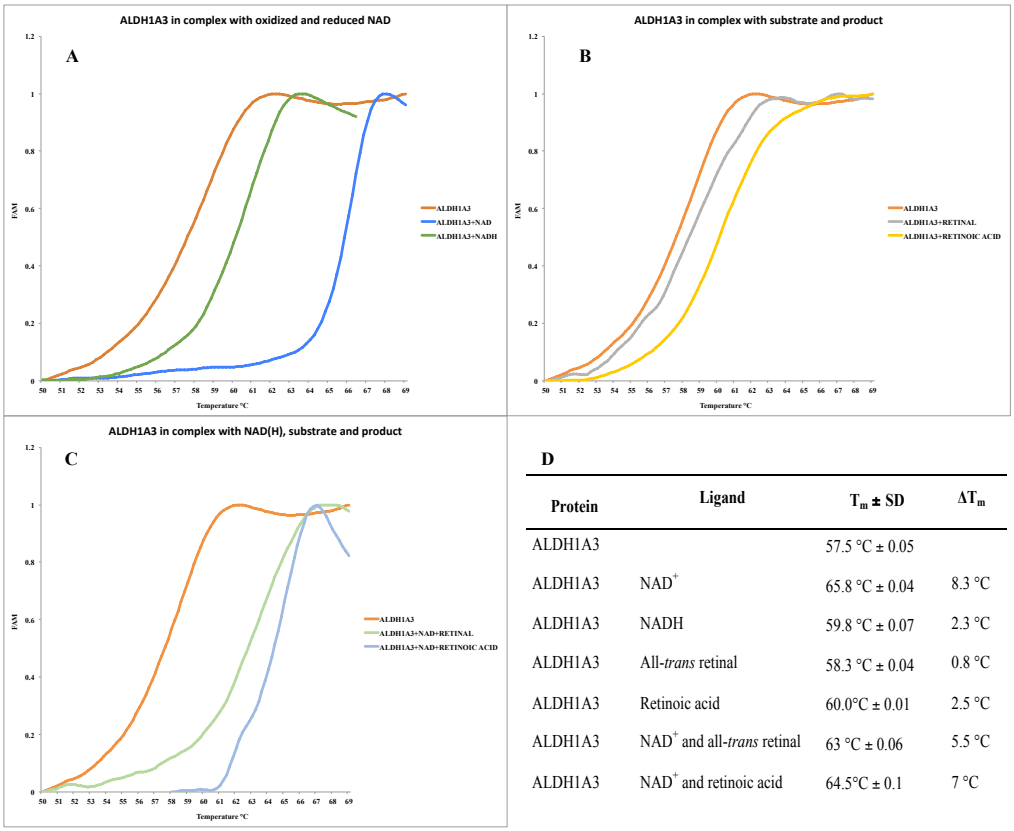
Table 2. The biochemical characterization of hALDH1A3. The steady state kinetics were measured varying the concentration of NAD^+, the natural substrate all-*trans* retinal and acetaldehyde.

Variable Substrate	Fixed Substrate	Km	Kcat	Kcat/Km
Acetaldehyde	NAD+	$2.4x10^{-3}$ M	0.1 s^{-1}	$4.1x10^{1}$ $M^{-1}s^{-1}$
NAD+	Acetaldehyde	$77.5x10^{-6}$ M	0.1 s^{-1}	$1.3x10^{3}$ $M^{-1}s^{-1}$
All-*trans* retinal	NAD+	$9.3x10^{-6}$ M	1.6 s^{-1}	$1.7x10^{5}$ $M^{-1}s^{-1}$
NAD+	All-*trans* retinal	$4.8x10^{-6}$ M	2.0 s^{-1}	$4.2x10^{5}$ $M^{-1}s^{-1}$

The steady state kinetic data are consistent with others ALDH1s described in literature[40], [41]. Indeed, the decrease in the Km values of 258-fold for all-*trans* retinal than acetaldehyde, confirms the substrate specificity of this human isoenzyme. Moreover, the measured Km for NAD^+ is almost 2 times lower than the Km for retinal and it is comparable to the constant measured for others ALDHs[42]. Noteworthy, turnover number for all-*trans* retinal oxidation (kcat=1.6 s^{-1}) is ~12 times higher compared to ALDH1A2 and ~18 times higher than ALDH1A1[43] and it confirms the preference of ALDH1A3 for all-*trans* retinal whit respect to 1A1 and 1A2.

The global stability of hALDH1A3 in complex with the oxidized and reduced cofactor and with all-*trans* retinal and retinoic acid (REA) was investigated for the first time by Thermofluor assay, as described in materials and methods [27], [44]. The

melting temperature (Tm) of hALDH1A3, calculated in absence or presence of ligands, showed a highly significant thermal shift for this enzyme coupled with the oxidized form of the cofactor (Figure 1).



D

Protein	Ligand	T_m ± SD	ΔT_m
ALDH1A3		57.5 °C ± 0.05	
ALDH1A3	NAD^+	65.8 °C ± 0.04	8.3 °C
ALDH1A3	NADH	59.8 °C ± 0.07	2.3 °C
ALDH1A3	All-*trans* retinal	58.3 °C ± 0.04	0.8 °C
ALDH1A3	Retinoic acid	60.0°C ± 0.01	2.5 °C
ALDH1A3	NAD^+ and all-*trans* retinal	63 °C ± 0.06	5.5 °C
ALDH1A3	NAD^+ and retinoic acid	64.5°C ± 0.1	7 °C

Figure 1. Thermal shift of hALDH1A3 in complex with its ligands. A) Tm of the apoALDH1A3 (orange) compared with the enzyme in complex with NAD^+ (blue) and NADH (green) **B)** Tm of the apoenzyme (orange) compared with the melting temperatures of the ALDH1A3 in complex with the substrate retinal (grey) and the product retinoic acid (light orange) **C)** Tm of the apoALDH1A3 (orange) compared with the enzyme in complex with NAD^+ and retinal (light green) and NAD and retinoic acid (light blue). **D) Table of T_m of ALDH1A3 in complex with its ligands**. Measured T_m for ALDH1A3 alone or in complex with oxidized and reduced cofactor and with all-*trans* retinal substrate and retinoic acid product. The ΔTm is expressed as difference between the Tm of ALDH1A3 in complex with ligands and the Tm of the apoenzyme.

The binds of NAD^+ increased the global stability of the enzymes (ΔTm 8.3 °C) compared to the reduced form of the cofactor (ΔTm 2.3 °C). Noteworthy was possible to identify a significant interaction of the protein with the product retinoic acid (ΔTm 2.5 °C), whit respect to the substrate all-*trans* retinal (ΔTm < 2 °C, under threshold value[44]) (Figure 1). The oxidized cofactor shows the major

capability to stabilize the complex compared to any other tested ligands. Furthermore, the measured Tm evidences a certain grade of affinity also for the products of the reaction. The NAD^+ influence on the stability of the enzyme was also tested in presence of both the substrate and product. As expected, the Tm for ALDH1A3 incubated with NAD^+/retinal and NAD^+/retinoic acid is, for both, lower than NAD^+ alone but higher respect to the single substrate or product.

Overall Quality of the model

The three-dimensional structure of human ALDH1A3 has been solved by molecular replacement using the sheep liver ALDH1A1 atomic coordinates as a search model (PDB code: 1BXS)[45] and refined at a resolution of 2.9 Å. The final human ALDH1A3 model contains eight identical chains per asymmetric unit, arranged as two independent tetramers and a total of 66 solvent molecules (Table 1). In each monomer no electron density is present for the N-terminal portion and for the last four amino acids at the C-terminal region. In addition in all eight monomers are present some disorder and no electron density is detectable in different zone of the structure (Table 3).

Table 3. Missing residues of all the eight monomers composing the asymmetric unit of human ALDH1A3 crystals.

	Monomer	Missing residues (no ρ)
	A	1-18, 508-512
Tetramer	B	1-22, 509-512
ABCD	C	1-20, 509-512
	D	1-19, 508-512
	E	1-22, 508-512
Tetramer	F	1-27, 338-343, 348-353, 386-394, 406-414 508-
EFGH	G	1-28, 383-391, 404-411, 508-512
	H	1-20, 509-512

The height monomers of the asymmetric unit share the same conformation showing an r.m.s.d. of 0.44 Å after superposition based on all 475 Cα atoms. Tetramer ABCD exhibits a global better-defined electron density than the EFGH tetramer.

Each monomer composing the asymmetric unit binds one molecule of NAD, while electron density corresponding to product retinoic acid (REA) exhibits a better defined density in chains A, B, C and D. In summary our model is the first structure of human ALDH1A3 reported in complex with NAD and REA. The stereochemistry of the model has been assessed with the program PROCHECK[37] and 99% of protein residues fall in the favored regions of the Ramachandran plot.

Overall structure of human ALDH1A3

Tetramer

The crystal structure of hALDH1A3 reveals the presences of two tetramers in the asymmetric unit with the four monomers intimately associated (Figure 2). Crystallographic data were supported by size exclusion chromatography that reveals an oligomeric state for recombinant enzyme consistent with a tetrameric quaternary structure in solution, as described for many ALDHs[46]. Indeed hALDH1A3 has a predicted molecular weight of 56 KDa and gel filtration experiment shows one defined peak corresponding to a native molecular mass of about 224 KDa (Figure 2). Moreover, the PISA analysis prediction[47] confirms that the favorite conformation of hALDH1A3 is a tetramer with an Assembled Surface Area (ASA) of 62658 $Å^2$ and a Buried Surface Area (BSA) of 17769 $Å^2$. In addition, the dimerization association, also predicted by PISA, is between chain A and B (BSA 2520 $Å^2$) and chain C and D (BSA 2456 $Å^2$). The secondary structures involved in the interaction surfaces (β5, α7, β18, β15, α13) contribute to the stability of the complex mainly with hydrogen bonds and few hydrophobic interactions. Three salt bridges between R84 - E426, R487 - E489, and R271 - G277, favor the dimerization between monomers.

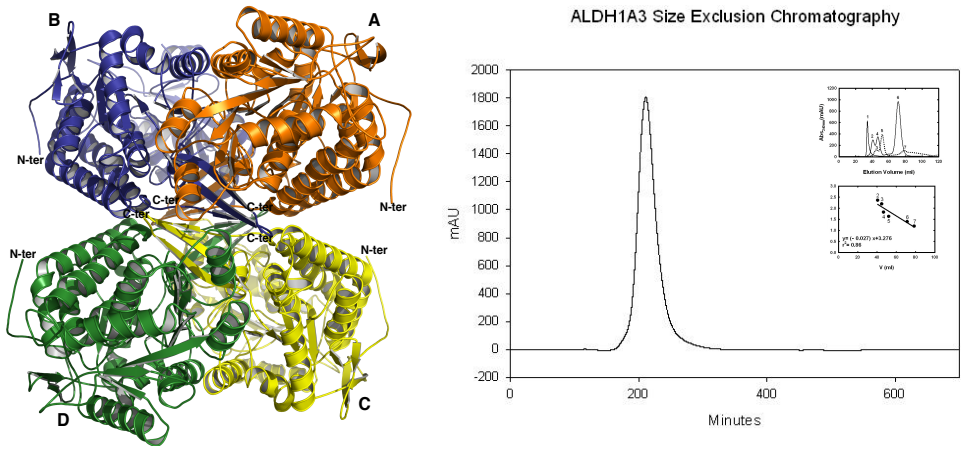


Figure 2. A) Cartoon representation of the biological tetrameric assembly of hALDH1A3 Chains A, B, C and D are colored in orange, blue, yellow and green, respectively. All the N-terminal of the monomer are flexible and exposed to the solvent, whereas the C-terminal are organized at the core of the tetramer **B) Size exclusion chromatogram** showing the elution peak of ALDH1A3 recombinant protein in solution. The elution volume corresponds to a molecular weight of 224 KDa, consistent with a tetrameric quaternary structure.

Monomer

Despite the assembly as homotetramer, each monomer of the human ALDH1A3 is independently able to convert one molecule of substrate to one molecule of product by reducing one NAD^+. The structure analysis of hALDH1A3 monomer confirmed that the protein shares high homology of the overall structure across all classes of ALDHs. Human ALDH1A3 folds into 13 α-helices, 19 β-strands and connecting loops resulting in a molecular architecture with three functional domains (Figure 3). The N-terminal NAD binding domain (L20-D149 and I171–G282), which characterizes the ALDH superfamily contains five-stranded parallel β-sheets (β9, β8, β7, β10 and β11). The C-terminal catalytic domain (G283–M482) is made by six-stranded parallel β-sheets (β15, β16, β13, β12, β17 and β18), and the oligomerization domain (K150–P170 and S483–L507) is defined by three-stranded antiparallel β-sheets (β5, β6 and β15). Both the NAD^+ binding and the catalytic domains are based on topologically related βαβ type polypeptide fold (Figure 3).

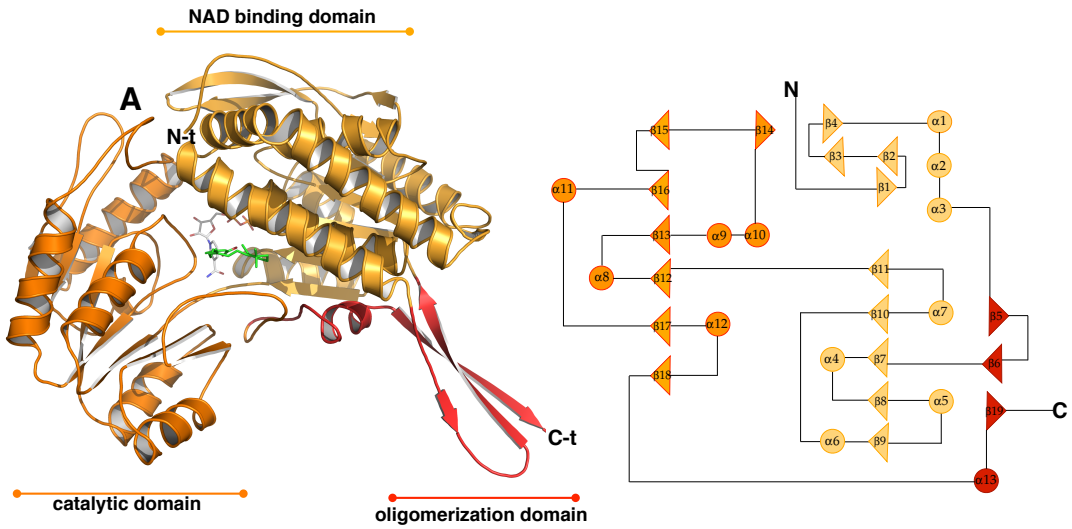


Figure 3. Crystal structure and topology of the secondary structure of hALDH1A3. The three dimensional structure is shown as cartoon representation with the N-terminal domain, the C-terminal domain, and the oligomerization domain colored in light orange, orange and red, respectively. The ligands NAD and REA are shown as white and green sticks, respectively. Beside the tertiary structure is reported the topology of the secondary assembly of the ALDH1A3 protein.

The superposition of our model with human ALDH1A1 available structure (PDB: 4WB9)[48] evidences a perfect correspondence of all secondary structures (r.m.s.d. of 0.84 Å for the 485 Cα). As well the superposition with the sheep liver ALDH1 highlight the similarity with the isoenzyme used as search model (r.m.s.d. of 0.88 Å for the entire polypeptide chain). Furthermore, ALDH1A3 shares high sequence similarity with other ALDHs with more than 70% amino acids identity with the human isozymes ALDH1A1 and ALDH1A2. The Figure 4 reports the alignment of the human ALDH1A3 with secondary structure information from PDB, with representative isoforms of the ALDH human superfamily. The sequence alignment was done with T-Coffee Expresso[49] and ESPript 3.0[50] confirming the high homology within ALDHs. The substrate selectivity of the different isozymes is ascribable to single amino acid mutations in the conserved regions and the structural analysis of these differences are the base for understanding the different enzymatic activity.

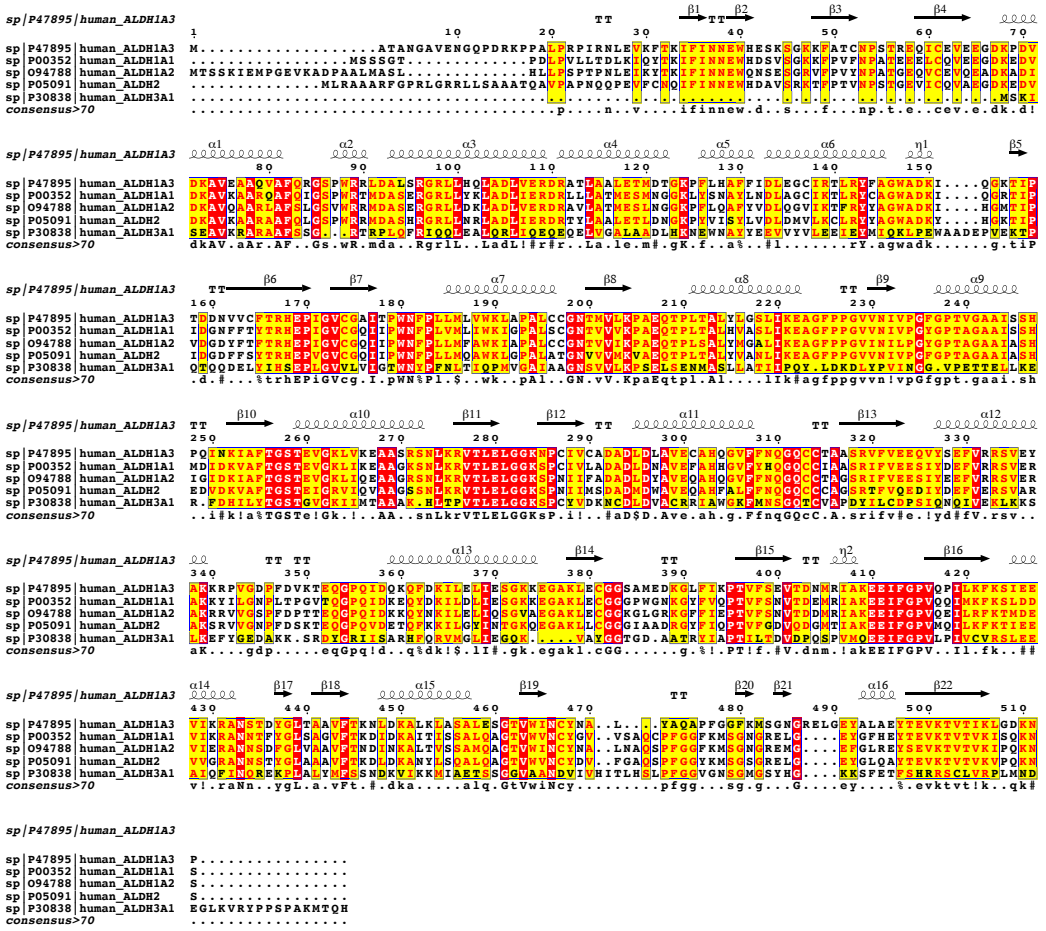


Figure 4. Alignment of the human ALDH1A3 amino acids sequence with representative human ALDHs: ALDH1A1, ALDH1A2, ALDH2 and ALDH3A1. The secondary structure indications on top of the aligned amino acids refer to the hALDH1A3 PDB. Identical and similar amino acids are boxed in red and yellow, respectively.

Retinoic acid and NAD$^+$ binding site

Our structural model highlights different conformations of REA and NAD$^+$ in complex with the ALDH1A3 isoenzyme. In particular, in monomer D, REA is mimicking the substrate with the carboxyl group close to the catalytic C314. At the same time the nicotinic acid moiety of NAD$^+$ is close to the active site like to accept the hydride from the thiohemiacetal intermediate. In monomer C, the retinoic acid is bound to the tunnel known to be the access for the substrates to the catalytic site. In parallel, NAD$^+$ nicotinic ring move closest to the surface of the

protein (Figure 5).

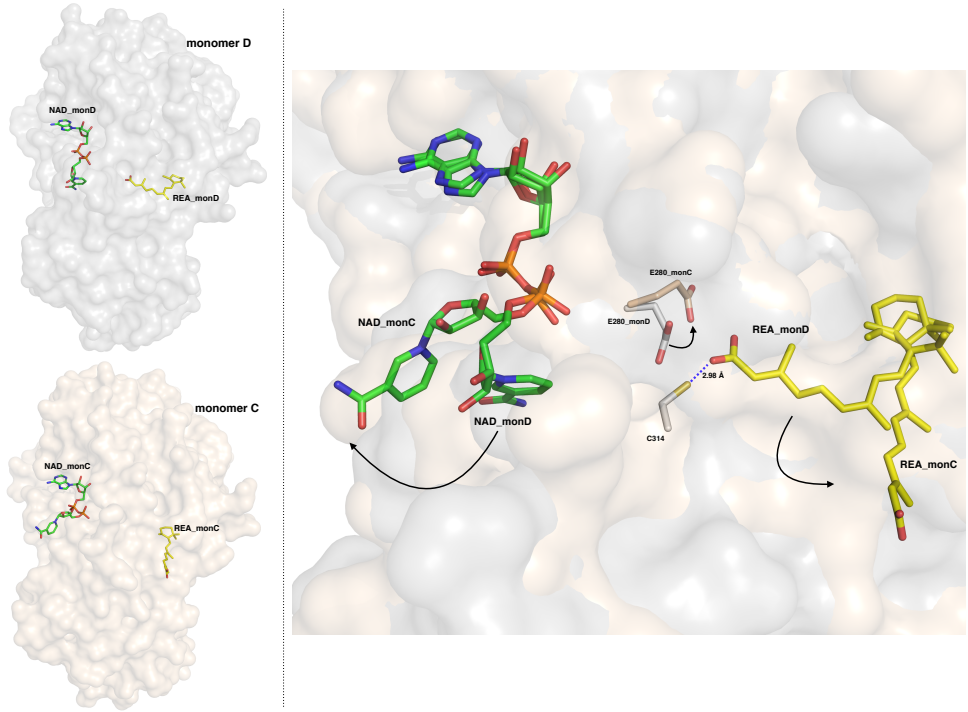


Figure 5. Overview of retinoic acid and NAD in complex with ALDH1A3 chains C and D. The figure describes the key aspect of the different conformations of the ligands co-crystallized with the protein. In monomer D, NAD binds near the canonical hydride transfer conformation, at the meantime REA is mimicking the substrate with the carboxyl group positioned close to the catalytic C314. The conserved residue E280 is closer to C314 with which establish weak interactions. In monomer C, NAD is far from the catalytic site with the nicotinamide moiety exposed to the solvent at the surface of the protein. The REA binds through hydrophobic interactions close to the opening of the tunnel that leads to the catalytic site, while E280 move outside of the catalytic site leaving room for the release of the product. We speculate that the ternary complex of monomer C can describe a putative release mechanism of the ligands after the occurred catalysis.

Architecture of the catalytic pocket in complex with REA

The most important feature of our structure is the definition of the tunnel for the retinal substrate and the interactions of the natural product with conserved residues of the catalytic pocket. Our structural data confirm the prediction analysis on substrate binding made on sheep liver and human ALDH1A1 structures[45], [48]. In human ALDH1A3 the catalytic task is made on the left-hand side by helix α3 that forms part of a three-helix bundle near the beginning of the N-terminal domain and

on the right-hand side by a surface loop (467 - 473) that precedes helix α13 near the oligomerization domain. The back of the tunnel is made up by helix α4 that is also the first α helix of the canonical Rossmann fold of the N-terminal domain. Helix α8 makes up the bottom of the substrate entrance tunnel, and originates from the catalytic domain, immediately preceding the active-site nucleophile (C314). The base of the tunnel comprises two β strands (β10, β11) (Figure 3). The first mechanism for selectivity adopted by ALDHs is the substrate access of the catalytic site, since the tunnel geometry is determinant for the specificity of ALDH isoforms. Precedent studies identified amino acids 124, 459 and 303 (based on ALDH2 sequence) as three signatures amino acids[51]. Our structural analysis confirmed that hALDH1A3 has the typical signature amino acids of ALDH1 family: G136, L471 and T315 respectively in the position corresponding to 124, 459 and 303. In particular, G136 is located at the entrance of the tunnel and its reduced steric bulk allows the entrance of large substrates. The steric occupancy of L471 positioned near to the access to the catalytic task is also determinant for ALDH1A3 selectivity. T315 lies next to reactive C314 and it is involved in the direct substrate interactions. In both the described conformations the β-ionone ring of REA doesn't just sticks out from the tunnel but establishes contacts with the protein environment. The natural product binds to the protein through hydrophobic interactions with G136, R139, W189, L185, L471, N469 and I132 conserved residues (Figure 6). In particular the β-ionone ring of REA is positioned at the entrance of the substrate access tunnel and establishes π-π stacking contacts with W189, and Van der Waals interactions with T140 and R139 conserved residues (Figure 7).

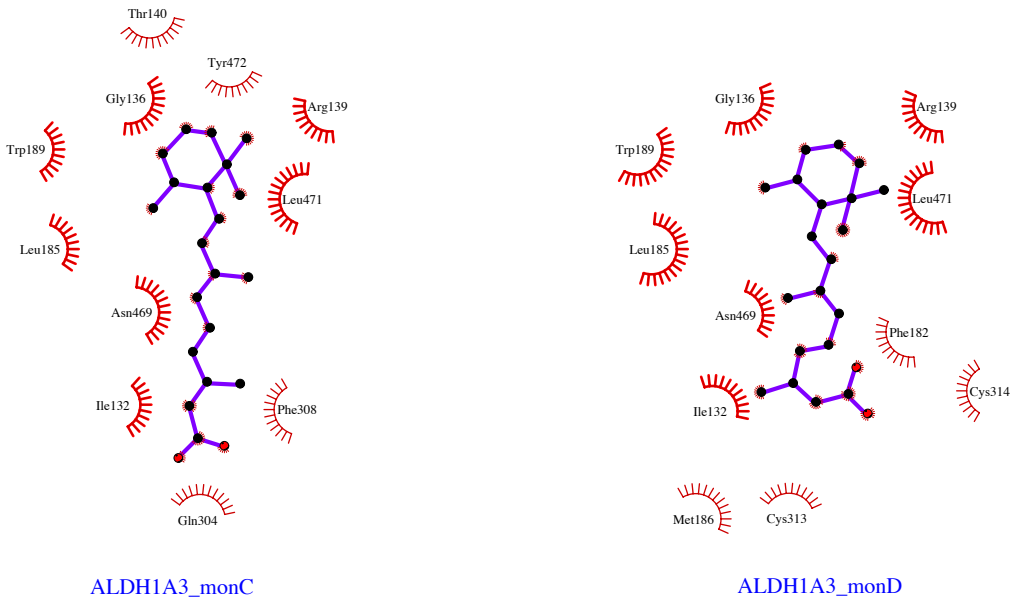


Figure 6. Hydrophobic residues involved in the binding of the product REA in complex with monomer C and monomer D. Ligplot^{+} v 1.4.5[39] representation was used to highlight the conserved interaction between the aliphatic REA and the hydrophobic residues G136, R139, W189, L185, L471, N469 and I132 in the catalytic pocket of chain C and D. Spoked arcs represent protein residues making nonbonded contacts with the ligand. The bold spoked arcs indicate protein residues that are in equivalent 3D positions when the two structural models are superposed.

In monomer D the solvent-accessible area of the substrate task corresponds to 290 Å^2 and it is compatible with a tunnel volume able to host all-*trans* retinal. At the end of the tunnel the carboxyl group of REA was stabilized by the strictly conserved M186 and C314. The catalytic SG of C314 engaged hydrogen bond with the oxygen O2 of REA carboxylic group (2.98Å) while SD of M186 stabilizes the O1 (2.93Å). In details, we observed that the methyl group (CE) of M186 undergoes a rotation of about 45° in the presence of REA with respect to chain C conformation, favoring the formation of a hydrogen bond between sulphur and oxygen (Figure 7A). In monomer C the solvent-accessible area of the substrate access tunnel was measured in 310 Å^2. The product REA is partially exposed to the solvent and the carboxyl group follows through to the exit from the tunnel by two weak interactions. In particular, OE1 of Q304 interacts with O1 (2.75Å) of the

REA while the NE2 weakly interacts with O2 of H128 (4.69Å) (Figure 7B).

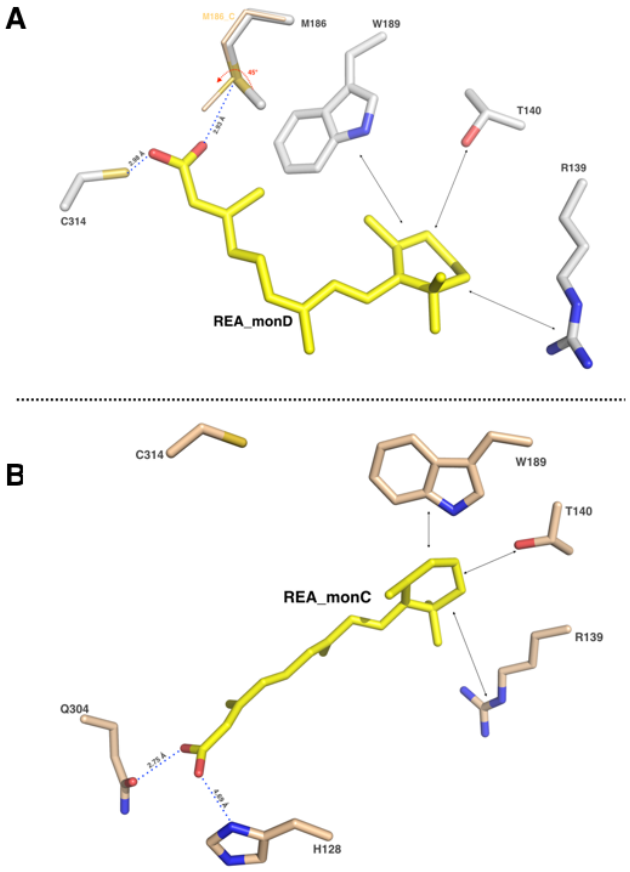


Figure 7. Representation of the main interactions within the two conformations of REA and the respective monomer. A) REA in complex with monomer D (white). The β-ionone ring establishes hydrophobic interactions with W189, T140 and R189, while the carboxyl group makes hydrogen bonds with C314, M186. **B)** REA in complex with monomer C (wheat). The β-ionone ring maintains the same hydrophobic interactions with W189, T140 and R189. The carboxyl group it is within interacting distance for hydrogen bonds with Q304 and H126.

In addition we observed in our structure a double conformation of the strictly conserved E280 that acts like a general base and it is responsible of the activation of two water molecules essential for the reduction of the catalytic C314 and for the hydrolysis of the thioester intermediate. In the hALDH1A3 chain D, the E280 is in the "inside" position close to the C314 (3.69 Å), meanwhile in monomer C is in the "outside" position referring to the description by Muñoz-Clarez et al.[52] (Figure 5).

NAD^+ binding at the N-terminal domain of hALDH1A3

The N-terminal NAD^+-binding domain of human ALDH1A3 enzyme, exhibits the typical ALDH cofactor-binding motif (β, α, β) that differs from the classical Rossmann fold generally involved in the NAD^+ recognition[53]. Many ALDHs structures evidence the ability of the enzyme to bind a population of NAD conformers[54]. In our structure all the NAD^+ molecules in complex with ABCD tetramer have traceable electron density for the adenine moiety and the pyrophosphates, whereas the nicotinamide ring of the cofactor is mobile. In the reported hALDH1A3 crystal structure, the adenine lies in a conserved hydrophobic pocket between helix α4 and strand β8, as described for the majority of human ALDHs in complex with NAD (PDB code: 4WB9, 1O00, 1BXS). In particular, the adenine moiety was taken in place through π-π stacking interaction with P238 and hydrophobic interaction with V261 (Figure 8). In all the NAD conformations in complex with the ALDH1A3 the NZ of K204 engaged an hydrogen bond with the oxygen O2B of the adenine ribose (2.05 Å in chain C), and the first phosphate group of NAD is stabilized by another hydrogen bond between the oxygen O1A and OG of S258 (2.80 Å in chain C) (Figure 8). The network of interactions described for the adenine and the pyrophosphates refers to all the molecules of NAD in complex with the hALDH1A3 protein, whereas for what concern the nicotinamide moiety we observed in our structure two different conformations in complex with monomers C and D. In the monomer D, the NAD^+ nicotinamide group was oriented near the catalytic site at about 10 Å from the SG of catalytic cysteine C314. In this conformation the O2D of nicotinamide ribose is stabilized through an hydrogen bond with N7 of K364 (2.31 Å) and the O3D engaged hydrogen bonds with OE1 (2.80 Å) and NE2 (2.37 Å) of Q361. The carbamide group is stabilized through the binding between its N7N and OG1 of T259 (3.55 Å), and between O7N and OE1 of E260 (2.97Å) (Figure 8A).

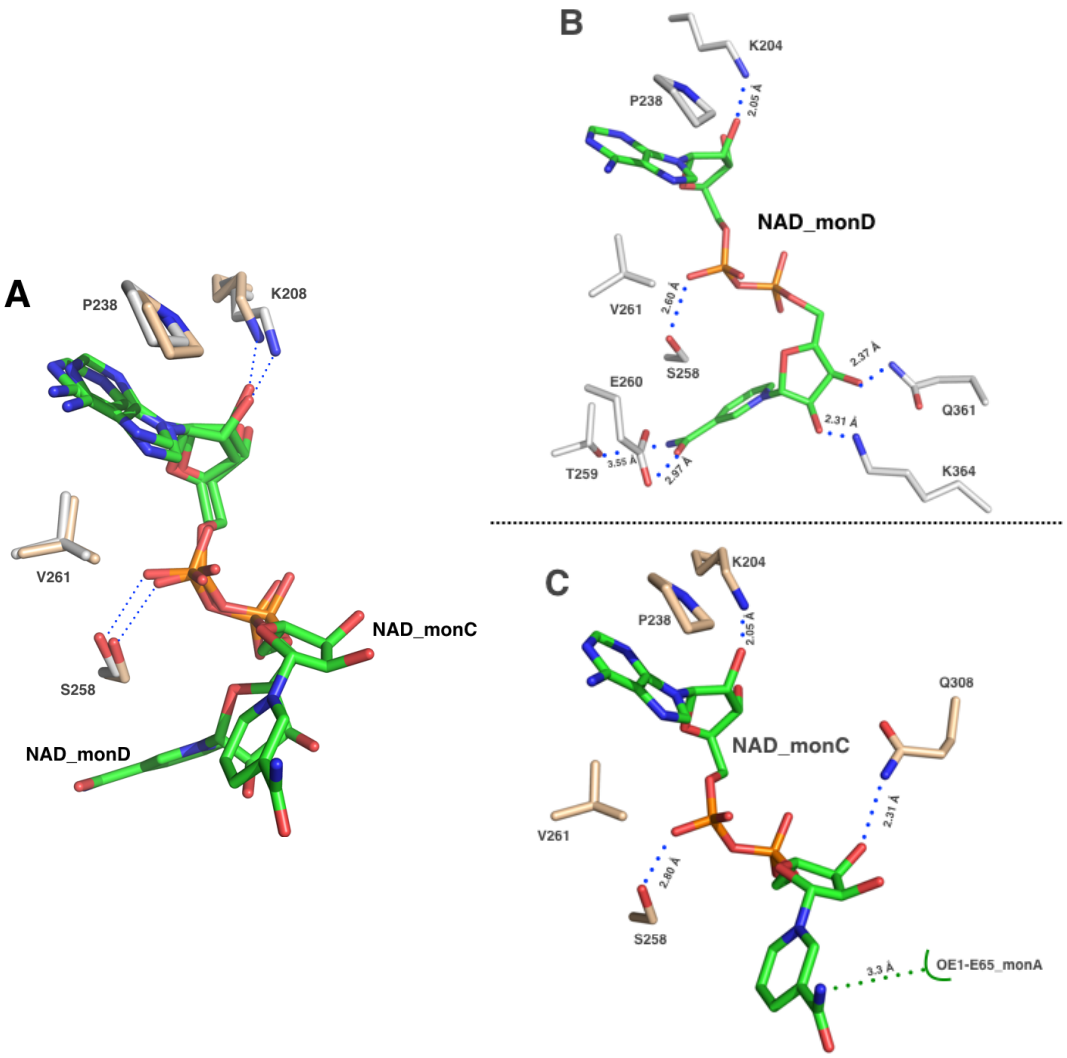


Figure 8. Superposition on NAD conformations in complex with chain C and D. Residues from the two monomers are represented as sticks and colored in wheat for monomer C and in white for monomer D. **A)** The figure illustrates the superposition of the two conformations and the conserved residues that are involved in the stabilization and binding of the adenine moiety and the pyrophosphates. P238 and V261 stabilize through π-π stacking and hydrophobic interactions the adenine. K204 and S258 are involved in hydrogen bonds with adenine ribose and the pyrophosphate closer to the adenine. **B)** Conformation of NAD in complex with chain D. The K204 is within interacting distance for hydrogen bonds with the oxygen O2B of the adenine ribose (2.05 Å). The S258 make a hydrogen bond with the O1A of the first phosphate group of NAD (2.80 Å). The O2D and O3D of the nicotinamide ribose make hydrogen bonds with N7 of K364 (2.31 Å) and NE2 (2.37 Å) of Q361. C) The NAD conformation of chain C shares all the binding of adenine moiety described for chain D. While the O3D of the nicotinamide ribose contacts the NE2 of Q208 (2.31 Å) and the N7N of the carboxyamide engaged hydrogen bonds with OE1 (3.3 Å) of E65 of the symmetry mate of monomer A.

In monomer C, by moving close to the surface of the protein, the nicotinamide moiety of NAD^+ loose the network of interactions described before[55]. The O3D of nicotinamide ribose contacts the NE2 of Q208 (2.31 Å), while the N7N of the

carboxyamide in our crystal assembly is within hydrogen binding distance with OE1 (3.3 Å) of E65 of the symmetry mate of monomer A (Figure 8B, Figure 9).

Otherwise in chain C the position of the C4 of nicotinamide group is located 16.1 Å far from the SG of catalytic cysteine C314, which is more than double the distance compared to the reported NAD conformations in complex with other ALDHs (1BXS, 4WB9) (Figure 9).

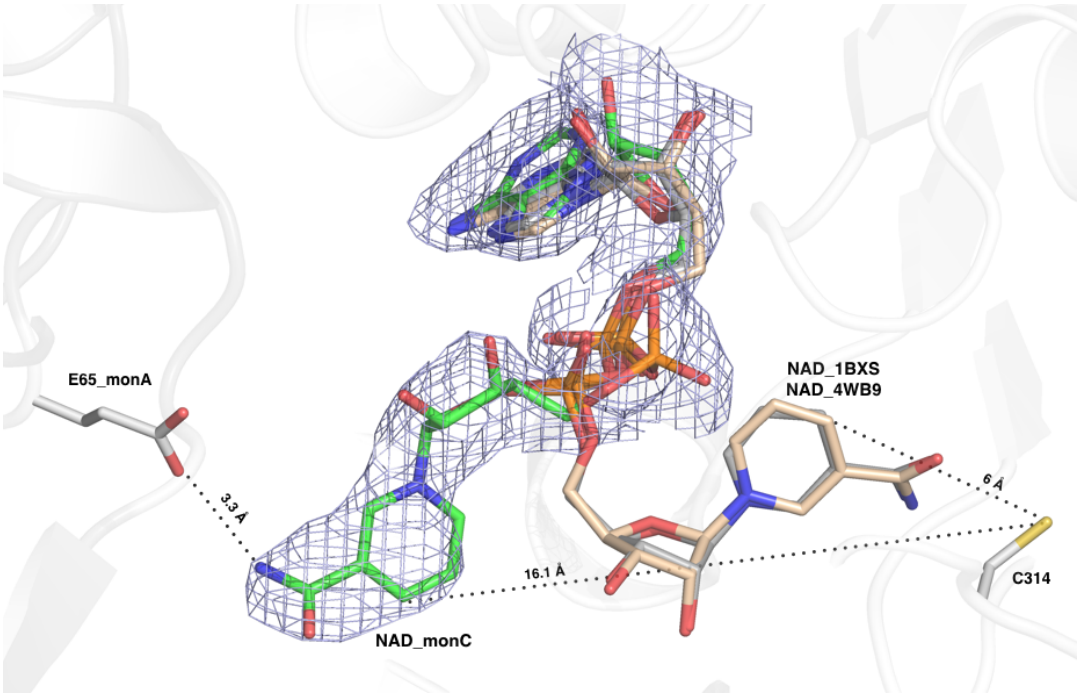


Figure 9. Superposition of NAD in complex with chain C (green) with NAD from 4WB9 (grey) and 1BXS (wheat). The overlap of the three cofactors highlights the movement of the nicotinamide of NAD in the direction of the surface of monomer C. The C4 of the nicotinamide ring is 16.1 Å far from C314, while the nicotinamide ring of NAD from human or sheep ALDH1A1 is 6 Å from the same catalytic residue. The nicotinamide is stabilized through hydrogen bond with the OE1 (3.3 Å) of E65 belonging to the symmetry mate of monomer A. The experimental electron density map for NAD in complex with chain C is represented in blue.

CONCLUSIONS

The human ALDH1A3 is a member of the cytosolic aldehyde dehydrogenase class 1 family with a key role in vertebrate embryonic development. Indeed, the enzymes of the ALDH1A subfamily show high specificity for retinaldehyde, oxidizing it to the powerful differentiation factor retinoic acid (REA)[8]. The ALDH1A3 has the most efficient dehydrogenase activity towards all-*trans*-retinal with respect to the others two isoenzymes.

In the present study we provide the first structural model of human recombinant ALDH1A3 in complex with NAD^+ and REA. Moreover the authors introduce an exhaustive biochemical characterization of the human isozyme, including the first reported Thermofluor experiment for ALDHs. Our biochemical data highlights the specificity of ALDH1A3 for all-*trans* retinal and demonstrate that human ALDH1A3 has greater affinity for NAD^+ supporting the assumption that in ALDHs the cofactor is the first molecule to bind to the enzyme[56]. In addition, our thermal shifts analysis shown accordance with precedent studies that identify the release of the oxidized product and the reduced cofactor as the rate-limiting steps in ALDH1s catalysis[57]. Our crystallographic study on human ALDH1A3 provides the first structure that describes an isoform of ALDH1A subfamily co-crystallized with the natural product. Retinoic acid has been described in two different conformations. In chain C, REA was detected at the exit of the tunnel leading to the catalytic site, picturing a ternary complex in which NAD^+ is far from the catalytic site and REA is almost on the protein surface. In chain D, REA is at the end of the tunnel with the carboxyl group close to the catalytic C314, in a conformation that mimic the substrate. The comparison of the two conformations of REA highlights a rotation of the aliphatic chain around carbon C10 that brings the functional carboxyl group or close to the active site or in the exit tunnel near the enzyme surface. Previous studies on ALDH1A2 suggested that the interaction of long aliphatic substrates, such as all-*trans* retinal, with a disorder loop is a key mechanism for the disorder to

order transition that accompanies the retinal catalysis[58]. In ALDH1A3 structure in complex with REA, the electron density of the described disorder loop (N469-L489) is well defined, confirming that the recognition of the substrate is linked to disorder/ordered transition of the loop. We can speculate that the different conformation of REA influences the NAD^+ nicotinamide moiety movement. Indeed, the superposition between chain C and D of ALDH1A3 confirmed the "swing" of the nicotinamide ring. In particular, NAD^+ in complex with monomer D was oriented near the catalytic site, 10 Å far from the SG of catalytic C314. Otherwise in monomer C, in which REA binds the enzyme close to the exit tunnel, the nicotinamide moiety moved away from the catalytic site in the direction of the protein surface, 16.4 Å far from C314. This conformation describes a new orientation of NAD^+ that differs from the known conformers of NAD^+ reported in ALDH structures.

In summary our data provide a first comprehensive structural view of different conformations of NAD and REA bound to ALDH1A3. The structural analysis of the human enzyme in complex with its ligand, together with its biochemical characterization can be a valuable starting point for the rational base drug design of molecules potentially active as inhibitors. This data takes on even more importance in light of the latest findings that demonstrated an elevated expression of human ALDH1A3 in mesenchymal Glioma Stem Cells (Mes GSCs) with respect to other ALDHs, including human ALDH1A1. This findings underscores the importance of defining the ALDH isoenzymes responsible for the high ALDHs activity in each cancer type and provide compelling evidence in favor of the ALDH isoenzymes functioning as key molecules governing cell proliferation, survival and chemoresistance of CSCs. In this contest, the complete structural determination of hALDH1A3 is a critical issue for rational design of small molecules to be used as prognostic marker and diagnostic tools.

BIBLIOGRAPHY

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

2 Vasiliou, V., Pappa, A. and Estey, T. (2004) Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. Drug Metab. Rev. **36**, 279–299.

3 Yokoyama, A., Muramatsu, T., Omori, T., Yokoyama, T., Matsushita, S., Higuchi, S., Maruyama, K. and Ishii, H. (2001) Alcohol and aldehyde dehydrogenase gene polymorphisms and oropharyngolaryngeal, esophageal and stomach cancers in Japanese alcoholics. Carcinogenesis **22**, 433–439.

4 Sládek, N. E. (2003) Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact. J. Biochem. Mol. Toxicol. **17**, 7–23.

5 Vasiliou, V. and Nebert, D. W. (2005) Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. Hum. Genomics **2**, 138–143.

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

7 Braun, T., Bober, E., Singh, S., Agarwal, D. P. and Goedde, H. W. (1987) Evidence for a signal peptide at the amino-terminal end of human mitochondrial aldehyde dehydrogenase. FEBS Lett. **215**, 233–236.

8 Zhao, D., McCaffery, P., Ivins, K. J., Neve, R. L., Hogan, P., Chin, W. W. and Dräger, U. C. (1996) Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase. Eur. J. Biochem. FEBS **240**, 15–22.

9 Kastner, P., Mark, M. and Chambon, P. (1995) Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life? Cell **83**, 859–869.

10 Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., et al. (1995) The nuclear receptor superfamily: the second decade. Cell **83**, 835–839.

11 Hofmann, C. and Eichele, G. Retinoids in development. Retin. Biol. Chem. Med. 2nd Ed Raven Press, New York, 387–441.

12 Duester, G. (2001) Genetic dissection of retinoid dehydrogenases. Chem. Biol.

Interact. **130-132**, 469–480.

13 Maly, I. P., Crotet, V. and Toranelli, M. (2003) The so-called "testis-specific aldehyde dehydrogenase" corresponds to type 2 retinaldehyde dehydrogenase in the mouse. Histochem. Cell Biol. **119**, 169–174.

14 Sima, A., Parisotto, M., Mader, S. and Bhat, P. V. (2009) Kinetic characterization of recombinant mouse retinal dehydrogenase types 3 and 4 for retinal substrates. Biochim. Biophys. Acta **1790**, 1660–1664.

15 Everts, H. B., King, L. E., Sundberg, J. P. and Ong, D. E. (2004) Hair cycle-specific immunolocalization of retinoic acid synthesizing enzymes Aldh1a2 and Aldh1a3 indicate complex regulation. J. Invest. Dermatol. **123**, 258–263.

16 Everts, H. B., Sundberg, J. P., King, L. E. and Ong, D. E. (2007) Immunolocalization of enzymes, binding proteins, and receptors sufficient for retinoic acid synthesis and signaling during the hair cycle. J. Invest. Dermatol. **127**, 1593–1604.

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

18 Reya, T., Morrison, S. J., Clarke, M. F. and Weissman, I. L. (2001) Stem cells, cancer, and cancer stem cells. Nature **414**, 105–111.

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

20 Luo, Y., Dallaglio, K., Chen, Y., Robinson, W. A., Robinson, S. E., McCarter, M. D., Wang, J., Gonzalez, R., Thompson, D. C., Norris, D. A., et al. (2012) ALDH1A isozymes are markers of human melanoma stem cells and potential therapeutic targets. Stem Cells Dayt. Ohio **30**, 2100–2113.

21 Marcato, P., Dean, C. A., Pan, D., Araslanova, R., Gillis, M., Joshi, M., Helyer, L., Pan, L., Leidal, A., Gujar, S., et al. (2011) Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. Stem Cells Dayt. Ohio **29**, 32–45.

22 Jordan, C. T., Guzman, M. L. and Noble, M. (2006) Cancer stem cells. N. Engl. J. Med. **355**, 1253–1261.

23 Khanna, M., Chen, C.-H., Kimble-Hill, A., Parajuli, B., Perez-Miller, S., Baskaran, S., Kim, J., Dria, K., Vasiliou, V., Mochly-Rosen, D., et al. (2011) Discovery of a novel class of covalent inhibitor for aldehyde dehydrogenases. J. Biol. Chem. **286**, 43486–43494.

24 Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D. and Vasiliou, V. (2012) Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. Pharmacol. Rev. **64**, 520–539.

25 Mao, P., Joshi, K., Li, J., Kim, S.-H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U. R., Benos, P. V., Smith, L., et al. (2013) Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. Proc. Natl. Acad. Sci. U. S. A. **110**, 8644–8649.

26 Michaelis, L., Menten, M. L., Johnson, K. A. and Goody, R. S. (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. Biochemistry (Mosc.) **50**, 8264–8269.

27 Matulis, D., Kranz, J. K., Salemme, F. R. and Todd, M. J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. Biochemistry (Mosc.) **44**, 5258–5266.

28 Kabsch, W. (2010) XDS. Acta Crystallogr. D Biol. Crystallogr. **66**, 125–132.

29 Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. **50**, 760–763.

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

31 Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J. and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr. D Biol. Crystallogr. **64**, 61–69.

32 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53**, 240–255.

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

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

35 Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature **355**, 472–475.

36 Perrakis, A., Harkiolaki, M., Wilson, K. S. and Lamzin, V. S. (2001) ARP/wARP and molecular replacement. Acta Crystallogr. D Biol. Crystallogr. **57**, 1445–1450.
37 Laskowski, R. A., Moss, D. S. and Thornton, J. M. (1993) Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol. **231**, 1049–1067.
38 Schrodinger. (2010, August) The PyMOL Molecular Graphics System, Version 1.3r1.
39 Laskowski, R. A. and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. **51**, 2778–2786.
40 Klyosov, A. A., Rashkovetsky, L. G., Tahir, M. K. and Keung, W. M. (1996) Possible role of liver cytosolic and mitochondrial aldehyde dehydrogenases in acetaldehyde metabolism. Biochemistry (Mosc.) **35**, 4445–4456.
41 Labrecque, J., Dumas, F., Lacroix, A. and Bhat, P. V. (1995) A novel isoenzyme of aldehyde dehydrogenase specifically involved in the biosynthesis of 9-cis and all-trans retinoic acid. Biochem. J. **305 (Pt 2)**, 681–684.
42 Ho, K. K., Hurley, T. D. and Weiner, H. (2006) Selective alteration of the rate-limiting step in cytosolic aldehyde dehydrogenase through random mutagenesis. Biochemistry (Mosc.) **45**, 9445–9453.
43 Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F. and Rahuel-Clermont, S. (2013) Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. Chem. Biol. Interact. **202**, 78–84.
44 Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. **357**, 289–298.
45 Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. and Baker, E. N. (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Struct. Lond. Engl. 1993 **6**, 1541–1551.
46 Yoshida, A., Rzhetsky, A., Hsu, L. C. and Chang, C. (1998) Human aldehyde dehydrogenase gene family. Eur. J. Biochem. FEBS **251**, 549–557.
47 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. **372**, 774–797.
48 Morgan, C. A. and Hurley, T. D. (2015) Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1. Chem. Biol. Interact. **234**, 29–37.
49 Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B.,

Keduas, V. and Notredame, C. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res. **34**, W604–608.

50 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. **42**, W320–W324.

51 Sobreira, T. J. P., Marlétaz, F., Simões-Costa, M., Schechtman, D., Pereira, A. C., Brunet, F., Sweeney, S., Pani, A., Aronowicz, J., Lowe, C. J., et al. (2011) Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. Proc. Natl. Acad. Sci. U. S. A. **108**, 226–231.

52 Muñoz-Clares, R. A., González-Segura, L. and Díaz-Sánchez, A. G. (2011) Crystallographic evidence for active-site dynamics in the hydrolytic aldehyde dehydrogenases. Implications for the deacylation step of the catalyzed reaction. Chem. Biol. Interact. **191**, 137–146.

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

54 Hammen, P. K., Allali-Hassani, A., Hallenga, K., Hurley, T. D. and Weiner, H. (2002) Multiple conformations of NAD and NADH when bound to human cytosolic and mitochondrial aldehyde dehydrogenase. Biochemistry (Mosc.) **41**, 7156–7168.

55 Perez-Miller, S. J. and Hurley, T. D. (2003) Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. Biochemistry (Mosc.) **42**, 7100–7109.

56 Feldman, R. I. and Weiner, H. (1972) Horse liver aldehyde dehydrogenase. II. Kinetics and mechanistic implications of the dehydrogenase and esterase activity. J. Biol. Chem. **247**, 267–272.

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

58 Bordelon, T., Montegudo, S. K., Pakhomova, S., Oldham, M. L. and Newcomer, M. E. (2004) A disorder to order transition accompanies catalysis in retinaldehyde dehydrogenase type II. J. Biol. Chem. **279**, 43085–43091.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Nadia Raffaelli (Università Politecnica delle Marche, Italy) for critical reading of the manuscript and helpful discussion.

Contributions

A.M., S.D., J.L. and S.G. performed experiments, collected and analyzed the data;

S.G. designed the research;

S.G. and A.M. wrote the paper.

Competing financial interests

The authors declare no competing financial interests

In silico virtual screening and development of high throughput screening for identification of potential inhibitors of human ALDH1A3

Andrea Moretti[1], Zhaofeng Ye[2], Carlos J Camacho[2], Robert W Sobol[3], Menico Rizzi[1*] and Silvia Garavaglia[1*].

[1]Department of Pharmaceutical Sciences, University of Piemonte Orientale, Largo Donegani 2, 28100 Novara, Italy; [2]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15260, USA; [3]Department of Oncologic Sciences, Molecular & Metabolic Oncology Program, University of South Alabama Mitchell Cancer Institute, 1660 Springhill Avenue, Mobile, AL 36604, USA.

***Corresponding authors**

Silvia Garavaglia, [1]Department of Pharmaceutical Sciences, University of Piemonte Orientale, Via Bovio 6, 28100 Novara, Italy

Phone: +39 0321 375714

Fax: +39 0321 375821

Email: silvia.garavaglia@uniupo.it

Menico Rizzi, [1]Department of Pharmaceutical Sciences, University of Piemonte Orientale, Via Bovio 6, 28100 Novara, Italy

Phone: +39 0321 375712

Fax: +39 0321 375821

Email: menico.rizzi@uniupo.it

Keywords: Aldehyde Dehydrogenase/ ALDH1A3/ *In silico* screening/ Structure-based drug design/ Glioma/ Cancer Stem Cells.

INTRODUCTION

Aldehydes dehydrogenases (ALDHs) are $NAD(P)^+$ dependent enzymes that functions as catalysts for the dehydrogenase oxidation of a wide spectrum of aldehydes into their corresponding carboxylic acids[1]. The human ALDH superfamily consists of 19 putatively functional genes located on different chromosomes and grouped by sequence identity[2]. ALDHs sharing more than 40% sequence identity belong to the same family, meanwhile isoforms exhibiting more than 60% amino acids identity, reside in the same subfamily[2]. ALDHs are ubiquitous and conserved enzymes, essential for the clearance of potentially toxic aldehydes[3]. The importance of ALDHs in human metabolism is not only defined to their action as detoxifying enzymes but includes their role as key metabolic regulators[4]. ALDHs catalyze the biosyntheses of physiologically relevant molecules such as retinoic acid, γ-aminobutyric acid and betaine which are involved in embryonic development and tissue differentiation (retinoic acid), neurotransmission (γ-aminobutyric acid) and osmosis (betaine)[1]. The three isozymes belonging to the ALDH1A subfamily (ALDH1A1, ALDH1A2 and ALDH1A3) are all involved in the metabolism of retinoic acid (REA), a morphogen molecule that regulates embryonic development and cellular proliferation and differentiation. The aldehyde dehydrogenase family 1 member A3 (ALDH1A3, E.C. 1.2.1.5) catalyzes the oxidation of all-*trans* retinal to REA using NAD^+ as cofactor. ALDH1A3 is an essential enzyme for nasal development and more general for the embryonic development[5], and it is expressed in fetal nasal mucosa, kidney, breast, stomach and salivary gland. ALDH1A3 through REA signaling pathway, is implicated in the development of eye, olfactory bulbs, hair follicles, forebrain and cerebral cortex[1]. The three isozymes display differences in substrate selectivity, although they all share the ability to convert retinal isomers to REA. ALDH1A1 has a substrate preference for aldehydes deriving from lipid peroxidation as well as retinaldehyde. ALDH1A2 shows the highest catalytic

efficiency for the conversion of 9-*cis* retinal to REA[6], whereas ALDH1A3, despite it is capable to catalyzes the oxidation of *cis*-retinal to REA, displays a higher specificity for *all*-trans retinal with 10-fold the efficiency of ALDH1A1 and ALDH1A2[7]. We have been demonstrated that ALDH1A3 assemblies as a homotetramer in solution, moreover the enzyme shares high homology of three-dimensional global structure across all classes of ALDHs. ALDH1A3 folds in three functional domains namely, NAD^+ binding domain, catalytic domain and oligomerization domain, and its structure has been solved in complex with NAD and REA. The catalytic mechanism adopted by ALDH involves highly conserved amino acids and proceed through five critical steps. After the protein binds to NAD^+ cofactor[8], the active C302 (numbering based on mature ALDH2) is activated through a water-mediated proton abstraction by E268, than the thiolate of C302 performs a nucleophilic attack on the electrophilic aldehyde. The formation of the thiohemiacetal intermediate is concomitant with the hydride transfer to the pyrimidine ring of the cofactor. Finally the hydrolysis of the thioester allows he release of the carboxylic acid and the reduced cofactor[9]. Despite similarity in the global structure, high homology in secondary sequence and conserved mechanism of reaction, ALDHs process a wide spectrum of substrates, exploiting for their selectivity different molecular architectures of the catalytic pocket[10]. These differences in substrate selectivity have been exploited to develop inhibitors of ALDHs that have been reviewed recently[9]. Up to date none of the known ALDHs inhibitors have been demonstrated to be isoform selective. For example disulfiram (also known with the commercial name Antabuse) is known to inhibit more potently the ALDH1A1 instead of the putative target ALDH2[11], while DEAB has been demonstrated to be at the same time a classic substrate (ALDH3A1), a covalent inhibitor, (ALDH2 and ALDH1A2), and an intermediate between substrate and inhibitor (ALDH1A1, ALDH1A3, ALDH1B1, and ALDH5A1)[12]. One of the first successful research on the development of isoform specific inhibitors is described by Morgan et al. that identify two compounds selective

towards human ALDH1A1[10] The increasing studies on Cancer Stem Cells (CSCs) are drawing the attention of the scientific community on ALDH role in cancers. High ALDH1 expression and/or enzymatic activity are consolidate markers for CSCs detection[13]. At the meantime high ALDH1 activity is a useful molecular tool for selection of ALDH-positive CSCs through FACS analysis with ALDEFLUOR™ assay. Noteworthy, high ALDH activity in CSCs was observed during cyclophosphamide treatment and correlated with cytotoxic drug resistance of CSC population[15]. The critical role of ALDH in cancer metabolism is coming down to specific isoforms that are being scrutiny as potential prognostic marker for many cancers and are often predictive of worst outcome of the tumors[14]. Notably, ALDH1A3 is gaining more importance in CSCs biology. Marcato et al. found that was the overexpression of ALDH1A3, and not ALDH1A1, that was predictive of metastasis in breast cancer stem cells[16]. In 2014 Shao et al. demonstrated the importance of ALDH1A3 for the maintenance of non-small lung cancer stem cells[17], whereas Mao et al. in 2013 found that ALDH1A3 is overexpressed in mesenchymal Glioma Stem Cells (Mes GSCs)[18].

The structure determination of hALDH1A3 in complex with NAD and REA provides the necessary tool for the rational design of small ligands. In this work we provide a first *in silico* screening of compounds potentially active as inhibitors against the human ALDH1A3 and the first high throughput screening to test the inhibiting efficiency of the selected small molecules. Overall the structure-based drug design, beside its role in discovering potential inhibitors for ALDH1A3, is an interesting tool for the development of fluorescent ligands with prognosis and diagnostic purposes.

MATERIALS AND METHODS

Cloning, Expression and purification of human ALDH1A1 and ALDH1A3

The full-length human ALDH1A1 and ALDH1A3 genes were PCR amplified and cloned into the destination vector pDEST17 (providing a N-term 6xHis tag) using pENTR/D-TOPO Invitrogen Gateway® recombinant technology according to manufacture's protocol. The constructed plasmids were then transformed into *E. coli* BL21(DE3) (Novagen), and expressed in a 2xTY medium supplemented with Ampicillin (50 μg/ml). One liter of medium was inoculated with cells directly collected from overgrown 2xTY agar plates (Ampicillin 50 μg/ml) and incubated at 37 °C to OD_{600} = 0.6, then cells were grown at 20ºC overnight. After cultures were harvested by centrifugation, cells were resuspended in 40 mL of lysis Buffer A (50 mM Na_2HPO_4 pH 7.5, 300 mM NaCl, 1mM β-mercaptoethanol, 10 mM imidazole) supplemented with 250U of Benzonase nuclease, and then lysed by sonication on ice. After addition of protease inhibitor cocktail (100 μL), the lysate was clarified by centrifugation. Soluble fraction was loaded onto a Ni-NTA Superflow 5mL cartridge (Qiagen) equilibrated with 10 column volumes (CV) of Buffer A and washed with Buffer A supplemented with 50 mM imidazole until the absorbance at 280 nm return to the baseline (15 CV). Then the recombinant protein was eluted with Buffer A supplemented with 250 mM imidazole by a linear gradient in 20 CV. Fractions containing the expected enzyme were pooled and concentrated to 5 mg/ml with Merck Millipore Amicon Ultra-15 10 kDa. The pool was further purified by a gel filtration step on a HiPrep 16/60 Sephacryl 200 High Resolution column equilibrated in the same buffer containing 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM β-mercaptoethanol, 0.5 mM EDTA. The resulting procedure allowed us to obtain pure and active hALDH1A1 and hALDH1A3. All the purification steps described above were performed using a BioRad BioLogic DuoFlow FPLC-

systems. Concentrations were determined by Bradford assay, and sample purity was assessed by SDS-PAGE.

In silico analysis for the virtual screening of a ALDH1A3 pharmacophore

The amino acids describing the catalytic pocket of ALDH1A3 have been identified as target for the design of small ligands potentially active as inhibitors. The pharmacophore features were identified with the online tool ZINCpharmer[19] directly form structural information of ALDH1A3 in complex with the ligand retinoic acid. ZINCpharmer recognizes and ranks those clusters of residues that define a binding site and provide an initial target for rational small-molecule design. ZINCpharmer identifies hydrophobic, hydrogen bond donor/acceptor, positive/negative ions and aromatic features that are critical for the interaction between the ligand and the receptor. The resulting pharmacophoric structural features, including the search radius and the steric limits determined by the geometry of the catalytic site, defined the pharmacophore models that were used to screen the purchasable compounds from the ZINC library (176 million conformers of 18.3 million compounds)[20]. The screening of the ZINC library returned a list of commercially available molecules that should satisfy the predicted contacts and receptor recognition preferences.

The compounds of the first *in silico* screening were analyzed for their affinity for ALDH1A3 with smina[21], a fork of the AutoDock[22] program. A small, highly enriched set of sixteen compounds was selected by applying the expertise of the authors. Hot spots residues for selectivity of the docked compounds were confirmed by superposition[23] of the ALDH1A3 (unpublished) structures with human ALDH2 (PDB: 3N80), ALDH3A1 (PDB: 3SZA[24]) and ALDH1A1 (PDB: 4WJ9[25]) that are representative ALDH isoforms and putative off-targets of the small ligands.

High-throughput screening for the identification of ALDH1A3 hits

The previously reported continuous spectrometric assay for ALDH1A3 [to be published] was adapted to support a high throughput screening, and optimized to suit Corning ® 96-well format plate. For ALDH1A3 inhibition test, 200 µl reaction mixture containing 20 mM Tris HCl pH 8.0, 1 mM β-mercaptoethanol, 150 mM KCl, 100 µM NAD^+ and 20 µM all-*trans* retinal, was set up in a Corning® 96 Well Clear Flat Bottom UV-Transparent Microplate.

Reactions were started by the addition of 2.6 µM pure recombinant ALDH1A3. Change in absorbance at 340 nm (ε_{NADH}= 6220 M^{-1} cm^{-1}) was monitored for 30 min in a BioTek ® Synergy 2 Multi-Mode Reader at 25 °C.

The ALDH1A1 spectrometric assay to test selectivity of chosen compounds was identical to hALDH1A3 assay with the only difference of 200 µM acetaldehyde used as substrate.

Compounds chosen from virtual screen ZINC library collection based on their predicted potency on ALDH1A3 were dissolved in DMSO and tested at 100 µM in triplicates. The compounds with more than 50% inhibition were chosen for further confirmation using six points dose response concentration ranging from 500 µM down to 0.5 µM. The concentration of inhibitor (IC_{50}) required to reduce the fractional enzyme activity to half of its initial value in the absence of inhibitor by the selected compound was calculated plotting the enzyme fractional activity against the logarithm of inhibitor concentration, and fitting the curves to a dose response curve (Eq.1):

$$y=min+ ((max-min)/(1+10^{(LogIC50-x)})) \quad (Eq. 1)$$

In which y is the fractional activity of the enzyme in the presence of inhibitor at concentration [I], max is the maximum value of y observed at [I]= 0, and min is the minimum limiting value of y at high inhibitor concentrations. Results of the high throughput screening were analyzed with Sigmaplot (Systat Software, San Jose, CA) and GraphPad Prism 6.0e.

RESULTS AND DISCUSSION

In silico virtual screening of small molecules

Up to date there are no available ALDH inhibitors that are specific for ALDH1 subfamily isoforms. The development of a selective compound able to target the human ALDH1A3 is a difficult challenge that needs the integrated approach of structural studies and *in silico* drug design. The resulting pharmacophore has to distinguish between a high number of ALDH isoforms with high similarity in secondary and tertiary structures. We thought that the affinity of most the ALDHs for different substrates could be exploited for the design of a selective molecule. The geometry of the substrate access tunnel has a central role in ALDHs specificity[26], since that the catalytic pocket is the ideal target for the *in silico* screening of small ligand with potential inhibiting action. The rational design of an ALDH1A3 pharmacophore is necessary to identify a ligand that can discern among the same subfamily enzymes. All the ALDH1A isozymes are able to catalyze the conversion of retinal to retinoic acid, but they show preferences for different isomer of retinal, implying some diversity in the catalytic site (i.e. ALDH1A3 has major affinity for all-*trans* retinal)[7].

Identification of pharmacophoric features in ALDH1A3 structure

The three isoforms belonging to ALDH1A subfamily share more than 60% sequence identity, in details ALDH1A3 have 71% amino acids identity with ALDH1A1 and 72% identity with ALDH1A2[27]. The three sequences were aligned in T-Coffee Expresso[28], [29] with information on the topology of the secondary structure referred to the ALDH1A3 structure edited according to ESPript3[30]. The resulting alignment (Figure 1) underlines the high homology between the isozymes and the difficulties to identify critical amino acids target for the inhibitors. It is of our interest to find a compound that can discern in particular between ALDH1A1 and ALDH1A3, considering the high expression of the two

isoforms in many CSCs, whereas high expression of the 1A2 isoform has not been related to solid tumor or CSCs[31].

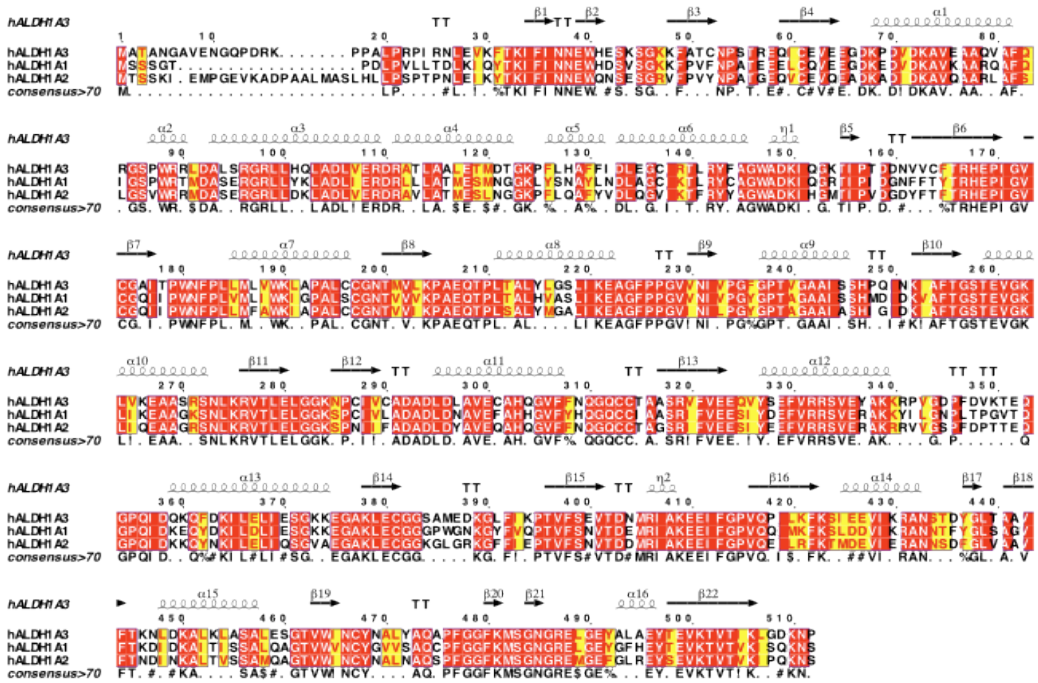


Figure 1. Sequence alignment of the three isoforms (UniProt ID) of ALDH1A subfamily. Aligned sequences of ALDH1A3 (P47895), ALDH1A1 (P00352), ALDH1A2 (O94788) (4). Red boxes highlight the residues conserved for all the isoforms; yellow boxes indicate residues conserved between two isozymes. The topology of the secondary structure refers to the unpublished ALDH1A3 PDB. The sequences alignment was performed using T-Coffee Expresso online tool[28], [29] and sequence information are edited according to the Espript3[30] notation conservation.

The pharmacophore features that were satisfied for the *in silico* screening of small molecules with inhibiting potential are described in Figure 2. The superposition of ALDH1A3 three-dimensional structure (not published) to ALDH1A1 (PDB: 4WJ9), ALDH2 (PDB: 3N80) and ALDH3A1 (PDB: 3SZA) structures allowed us to confirm the hot spots and the key residues critical for selectivity (Figure 3). The overlap of ALDH1A3 with ALDH2 and ALDH3A1 draw attention to the different architecture of the substrate access tunnels. ALDH2 and ALDH3A1 have two bulky amino acids located at the entrance and in the middle of the pocket, respectively the M124 and F292 for ALDH2, and the Y65 and the W233 for

ALDH3A1, which build a narrow tunnel accessible by small substrates such as acetaldehyde (Figure 3A).

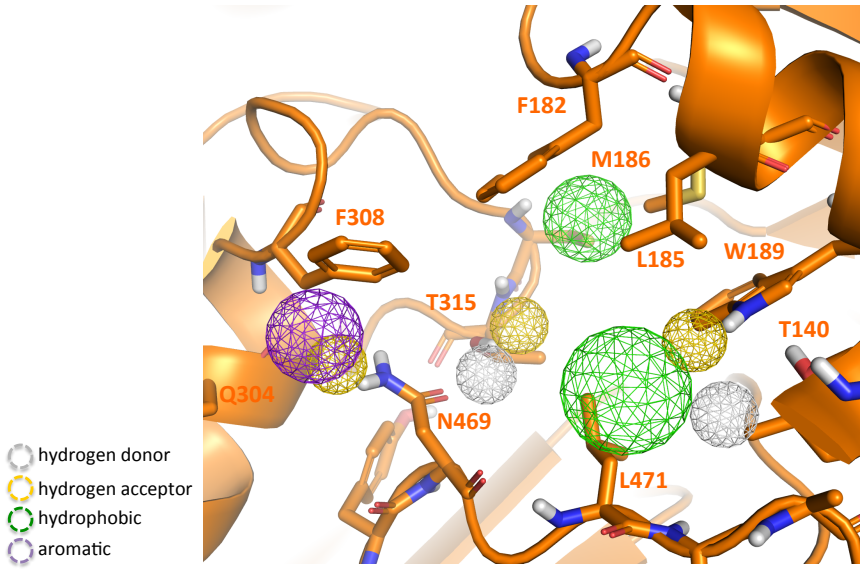


Figure 2. ZINCPharmer identification of pharmacophoric features for ligand-receptor interaction. The pharmacophore model needs to satisfy the selective interacting forces relevant for a putative pharmacological activity. The figure describes the binding site with the pharmacophoric points that are critical for the interaction within the small ligand and the receptor. The ALDH1A3 structure is reported in orange and the interacting forces representation follows the legend reported on the figure.

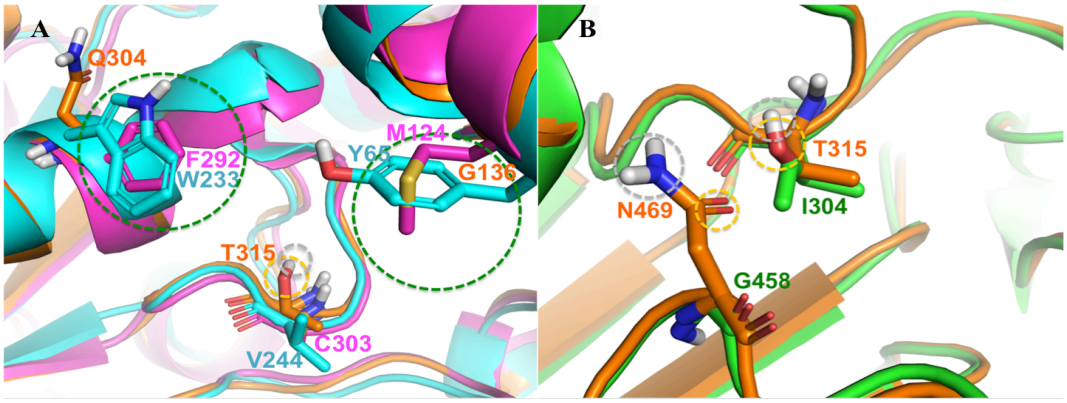


Figure 3. A) Superposition of ALDH1A3 structure (orange) with ALDH2 (magenta, PDB: 3N80) and ALDH3A1 (cyan, PDB: 3SZA). The green dot lines circle the bulky residues that in ALDH2 and ALDH3A1 occupy the access tunnel for the substrate. The yellow dot line and the grey dot line evidence the oxygen and the hydrogen of T315 that are possible target for hydrogen bond interaction with the inhibitor. B) **Overlap of ALDH1A3 and ALDH1A1 (green PDB: W4J9).** The superposition of the two ALDHs evidences the residues target for selectivity between the two isozymes. N469 and T315 are the two amino acids targeted as potential hydrogen donor (yellow dot line) or acceptor (grey dot line) for a possible hydrogen bond with the selective compound.

On contrary the ALDH1A3 exhibits at the corresponding residue positions a G136 and a Q303 that are amino acids with smaller side chains, building an open catalytic pocket accessible to larger substrate such as retinal. The structural analysis of ALDH isoforms at issue identifies the not conserved T315 as an interesting target for the design of a selective inhibitor. The T315 is located just next to the highly conserved catalytic C314 (C302 in the ALDH2 numbering) and it should be selective not only against ALDH2 and ALDH3A1 but also against the isozyme ALDH1A1. The geometry of the access tunnel cannot be a discriminant for selectivity against ALDH1A1. The docked compound needs to target the few amino acids that are not conserved and not similar between the catalytic sites of the two isoforms. The first target residue is exactly the hydrophilic T315 that have opposite properties compared to hydrophobic I304 in ALDH1A1. The second target residue is N469 of ALDH1A3 that corresponds to a G458 in the ALDH1A1 (Figure 3B). Both the identified residues are at the meantime hydrogen donor and acceptor, being available for hydrogen bonds with the suitable pharmacophore. Primary *in silico* virtual screening gave 16 resulting hits of commercially available molecules that are potential inhibitors of human ALDH1A3.

High throughput screening of potential inhibitors

All the hits were tested for inhibition on pure recombinant ALDH1A3 as described in materials and methods. Out of the 16 predicted hits, compounds 1, 3, 9 and 11 were responsible of 50% or more inhibition towards ALDH1A3 activity and were taken in account for further investigation (Figure 4). Out of the selected compounds only the molecules 3 and 9 gave reproducible and reliable data for half maximal inhibitory concentration (IC_{50}) calculation. Compounds 1 and 11 evidence solubility problems or interference with the 340 nm wavelength emissions of NADH used to calculate the ALDH1A3 enzymatic activity, making difficult to obtain reliable data on a dose-response curve by spectrometric assay. Compounds 3

and 9 were tested at different concentrations in presence of natural substrate all-*trans* retinal and NAD^+ cofactor.

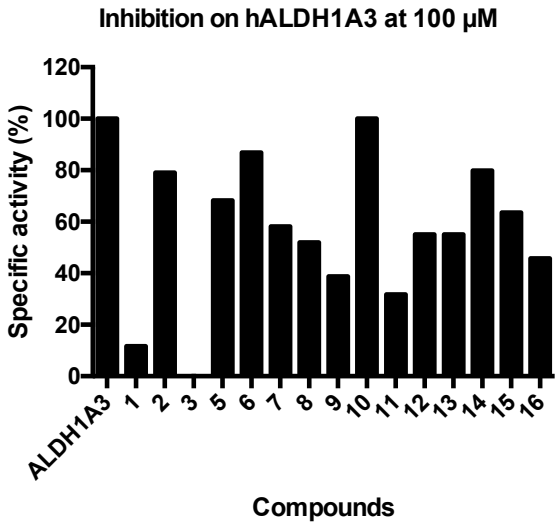


Figure 4. Screening of the 16 hits resulting from the ZINC library virtual screen. The commercially available compounds were tested for inhibition against ALDH1A3 specific activity. Data for compound number 4 are not available due to solubility problems of the molecule. The inhibition was initially estimated as the percentage of residual fractional activity of the enzyme with 100 µM inhibitor concentration and compared with the ALDH1A3 activity. Compounds 1, 3, 9, 11 exhibited more than 50% inhibition and were considered for further investigations.

The dose-response curve for antagonist number 3 and 9 versus human ALDH1A3 allowed us to calculate an IC_{50} of 21 ± 2 µM and 52 ± 1.4 µM, respectively (Figure 5). The compounds were tested also as inhibitors against ALDH1A1 activity. Preliminary data were collected by measuring the ALDH1A1 specific activity in presence of 50 µM of molecule 3 and 9 with fixed concentration of NAD^+ and acetaldehyde as substrate. The compounds showed almost no inhibition towards ALDH1A1 activity (Figure 6).

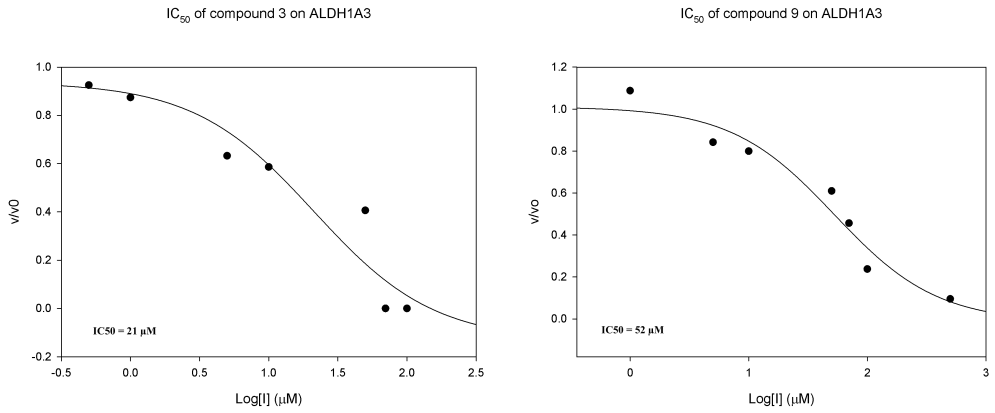


Figure 5. A) IC_{50} calculation for compounds 3 and 9 on pure recombinant human ALDH1A3. The plot of the logarithm of inhibitor concentration versus the ratio of the velocity to initial velocity was fitted with four-parameter sigmoidal curve equation (Eq.1). The IC_{50} for compound 3 was calculated as 21 ± 2 µM with an R= 0.96. The IC_{50} for compound 9 was calculated as 52 ± 1.4 µM with an R= 0.97

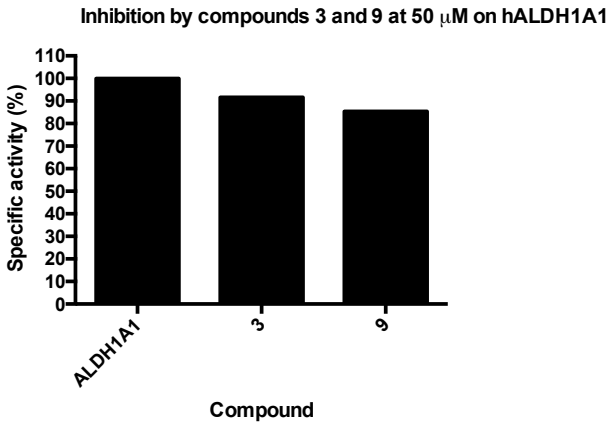


Figure 6. Inhibition of compounds 3 and 9 at 50 µM concentration on ALDH1A1 specific activity. The histogram is the result of the mean of three independent measurement reported as percentage of specific activity. ALDH1A1 maintain 92% of specific activity after inhibiting action of compound 3 and 85% of specific activity after inhibiting action of compound 9.

CONCLUSIONS

The structure-based drug design is an essential tool for the identification and the optimization of a lead compound, which is the first critical issue in the drug discovery process.

The rational design of pharmacophores needs the structure determination of the biological target to identify hot spots of the protein critical for its catalytic mechanism or for the interaction with its ligands. The analysis of the X-ray structure of human ALDH1A3 in complex with its natural ligands and the comparison with other ALDHs, potential off-target of the compound, was essential for the successful results of the first *in silico* virtual screening. By screening only sixteen commercially available hits resulting from the virtual screening, we were able to obtain interesting preliminary data on selective compounds with inhibiting action. Although we are still far from the development of a potential inhibitor with pharmaceutical activity and therapeutic prospective, the selectivity of the molecules 3 and 9 for ALDH1A3 towards ALDH1A1 lays good basis for the identification of a pharmacophore model. The identified compounds represent a good starting scaffold for the design and the chemical synthesis of further molecules with the objective for the identification of a lead compound with specificity for ALDH1A3 isoform.

BIBLIOGRAPHY

1 Marchitti, S. A., Brocker, C., Stagos, D. and Vasiliou, V. (2008) Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily.

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

3 Vasiliou, V., Pappa, A. and Estey, T. (2004) Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. Drug Metab. Rev. **36**, 279–299.

4 Sládek, N. E. (2003) Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact. J. Biochem. Mol. Toxicol. **17**, 7–23.

5 Zhang, X., Zhang, Q.-Y., Liu, D., Su, T., Weng, Y., Ling, G., Chen, Y., Gu, J., Schilling, B. and Ding, X. (2005) Expression of cytochrome p450 and other biotransformation genes in fetal and adult human nasal mucosa. Drug Metab. Dispos. Biol. Fate Chem. **33**, 1423–1428.

6 Maly, I. P., Crotet, V. and Toranelli, M. (2003) The so-called "testis-specific aldehyde dehydrogenase" corresponds to type 2 retinaldehyde dehydrogenase in the mouse. Histochem. Cell Biol. **119**, 169–174.

7 Sima, A., Parisotto, M., Mader, S. and Bhat, P. V. (2009) Kinetic characterization of recombinant mouse retinal dehydrogenase types 3 and 4 for retinal substrates. Biochim. Biophys. Acta **1790**, 1660–1664.

8 Feldman, R. I. and Weiner, H. (1972) Horse liver aldehyde dehydrogenase. II. Kinetics and mechanistic implications of the dehydrogenase and esterase activity. J. Biol. Chem. **247**, 267–272.

9 Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D. and Vasiliou, V. (2012) Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. Pharmacol. Rev. **64**, 520–539.

10 Morgan, C. A. and Hurley, T. D. (2015) Characterization of Two Distinct Structural Classes of Selective Aldehyde Dehydrogenase 1A1 Inhibitors. J. Med. Chem. **58**, 1964–1975.

11 Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. and Baker, E. N. (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Struct. Lond. Engl. 1993 **6**, 1541–1551.

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

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

14 Rodriguez-Torres, M. and Allan, A. L. (2015) Aldehyde dehydrogenase as a marker and functional mediator of metastasis in solid tumors. Clin. Exp. Metastasis.

15 Januchowski, R., Wojtowicz, K. and Zabel, M. (2013) The role of aldehyde dehydrogenase (ALDH) in cancer drug resistance. Biomed. Pharmacother. Biomed. Pharmacothérapie **67**, 669–680.

16 Marcato, P., Dean, C. A., Pan, D., Araslanova, R., Gillis, M., Joshi, M., Helyer, L., Pan, L., Leidal, A., Gujar, S., et al. (2011) Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. Stem Cells Dayt.

Ohio **29**, 32–45.

17 Shao, C., Sullivan, J. P., Girard, L., Augustyn, A., Yenerall, P., Rodriguez-Canales, J., Liu, H., Behrens, C., Shay, J. W., Wistuba, I. I., et al. (2014) Essential role of aldehyde dehydrogenase 1A3 for the maintenance of non-small cell lung cancer stem cells is associated with the STAT3 pathway. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. **20**, 4154–4166.

18 Mao, P., Joshi, K., Li, J., Kim, S.-H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U. R., Benos, P. V., Smith, L., et al. (2013) Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. Proc. Natl. Acad. Sci. U. S. A. **110**, 8644–8649.

19 Koes, D. R. and Camacho, C. J. (2012) ZINCPharmer: pharmacophore search of the ZINC database. Nucleic Acids Res. **40**, W409–W414.

20 Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. and Coleman, R. G. (2012) ZINC: A Free Tool to Discover Chemistry for Biology. J. Chem. Inf. Model. **52**, 1757–1768.

21 Koes, D. R., Baumgartner, M. P. and Camacho, C. J. (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J. Chem. Inf. Model. **53**, 1893–1904.

22 Trott, O. and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. **31**, 455–461.

23 Schrodinger. (2010, August) The PyMOL Molecular Graphics System, Version 1.3r1.

24 Khanna, M., Chen, C.-H., Kimble-Hill, A., Parajuli, B., Perez-Miller, S., Baskaran, S., Kim, J., Dria, K., Vasiliou, V., Mochly-Rosen, D., et al. (2011) Discovery of a novel class of covalent inhibitor for aldehyde dehydrogenases. J. Biol. Chem. **286**, 43486–43494.

25 Morgan, C. A. and Hurley, T. D. (2015) Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1. Chem. Biol. Interact. **234**, 29–37.

26 Sobreira, T. J. P., Marlétaz, F., Simões-Costa, M., Schechtman, D., Pereira, A. C., Brunet, F., Sweeney, S., Pani, A., Aronowicz, J., Lowe, C. J., et al. (2011) Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. Proc. Natl. Acad. Sci. U. S. A. **108**, 226–231.

27 Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., et al. (2013) BLAST: a more efficient report with usability improvements. Nucleic Acids

Res. **41**, W29–W33.

28 Notredame, C., Higgins, D. G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302**, 205–217.

29 O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G. and Notredame, C. (2004) 3DCoffee: combining protein sequences and structures within multiple sequence alignments. J. Mol. Biol. **340**, 385–395.

30 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. **42**, W320–W324.

31 Rodriguez-Torres, M. and Allan, A. L. (2015) Aldehyde dehydrogenase as a marker and functional mediator of metastasis in solid tumors. Clin. Exp. Metastasis.

Conclusions

The human ALDH1A3 is a key enzyme of the retinoid metabolism in which it catalyzes the oxidation of retinaldehyde to retinoic acid (REA), a key signaling molecule for embryonic development and tissue differentiation[1]. High expression and activity of ALDH1A3 has been identified in different Cancer Stem Cells including Glioma Stem Cells[2], [3], [4]. In particular, overexpression of human ALDH1A3 was detected in mesenchymal GSCs of High Grade Gliomas (HGGs)[4], which are highly infiltrative and aggressive brain cancers characterized by poor prognosis (5 years survival < 5%), resistance to chemotherapy and recurrence after neurosurgery. This dramatic picture for grade III and grade IV gliomas highlight the necessity of a complete understanding of glioblastoma molecular bases that cannot discern from the understanding of Glioma Stem Cells biology.

In the present study we provide the first structural model of human recombinant ALDH1A3 in complex with NAD^+ and REA and an exhaustive biochemical characterization of the human isozyme, including the first reported thermofluor experiment.

The analysis of the biochemical and biophysical observations on human ALDH1A3 supports the assumption of an ordered reaction mechanism on which the cofactor is the first molecule to bind the enzyme [5], as evidenced by the high affinity of the enzyme for NAD^+. The increased Tm of ALDH1A3 in complex with the reduced cofactor NADH and the product REA is the result of an enhancement in the global stability of the enzyme. These thermal shift data are in accordance with the work of Rodríguez-Zavala et al. that identifies the release of the oxidized product and the reduced cofactor as the rate-limiting step in ALDH1 catalysis[6].

The human ALDH1A3 is the first structure that describes an isoform of the ALDH1A subfamily co-crystallized with the natural product. REA has been co-crystallized in complex with ALDH1A3 in two different conformations. In chain C, the ligand was fitted in the electron density detected at the “exit” of the tunnel

leading to the catalytic site. The ternary complex of the protein in presence of REA and NAD^+ shows a global conformation in which the product is close to the protein surface, at the meantime the nicotinamide ring of NAD is exposed to the solvent, far from the catalytic C314. Instead in ALDH1A3 chain D, the carboxyl group of REA is located at the end of the access tunnel in a conformation that mimic the substrate, within interacting distance with the catalytic C314.

Previous studies on rat ALDH1A2 suggest that the interaction of long aliphatic substrates with a disorder loop (~ aa 457 - aa 477, based on ALDH2), part of the substrate access tunnel, is a key mechanism for the disorder to order transition that accompanies the retinal catalysis[7]. In the reported ALDH1A3 structure in complex with REA, the electron density of the loop involved in substrate recognition (N469-L489) is well defined. Interestingly in both REA conformations the β-ionone ring is within interacting distances with L469, A470 and Q471 of the loop, supporting the hypothesis that the interaction of this loop with the substrate, or in our case with the product that mimic the substrate, is suitable with a disorder/ordered transition, necessary for catalysis. The geometry of the substrate access tunnel is also critical for substrate selectivity[8]. ALDH1A3 has the typical signature amino acids of ALDH1 subfamily catalytic pocket. The architecture of the catalytic site of ALDH1A3 is determined by a G136 located at the opening of the tunnel that allows the entrance of large substrates, a L471 positioned at the funnel of the access to the catalytic site that can accommodate large substrates and a T315 at the bottom of the active site, next to reactive C314, that can be involved in direct interactions with the substrate.

NAD^+ appears to bind to ALDHs with a high degree of conformational heterogeneity, due to the rotation of nicotinamide about the glycosidic bond[9], [10]. In the reported ALDH1A3 X-ray crystal structure, we described a new NAD^+ isomer that differs from the known conformers of the cofactor in complex with ALDH (PDB: 1BXS, 4WB9). The superposition of NAD from ALDH1A3 chain C with known conformation of the cofactor from other ALDH structures (1BXS,

4WB9), highlight the "swing" of the nicotinamide ring close to the surface of the protein. The novel isomer of NAD maintains the network of interactions that stabilizes the adenine and the pyrophosphates, while the nicotinamide ring is stabilized by the E65 of the symmetry mate of monomer A, that is next to monomer C in our crystal assembly. Notably, when REA mimics the substrate in chain D, the nicotinamide is likely to be located close to the catalytic site. Whereas, when REA binds close to the entrance of the substrate access tunnel that in this case exploit the function of "exit tunnel", the nicotinamide of the cofactor showed a defined electron density next to the surface of the protein, leaving room at the catalytic site as consequence of the occurred hydrolysis. Overall the structural investigation on hALDH1A3 in complex with NAD and REA, together with its biochemical characterization, provides the experimental critical issue to develop novel potential selective inhibitors that can be used as chemical tools to investigate the role of the ALDH1A3 isoform in normal cells and cancer stem cells.

A first structure-based *in silico* screening of small molecules with potential activity as inhibitors on ALDH1A3 was performed in the present research. The analysis of the pharmacophoric features that need to be satisfied for the building of a pharmacophore model were determined through ZINCPharmer and confirmed by structural analysis of the ALDH1A3 in complex with its ligand. The superposition of ALDH1A3 with representative ALDHs (ALDH3A1, ALDH2) and with the human isozyme ALDH1A1 confirmed the identification of N469 and T315 as hot spots for the rational design of small molecule selective mainly towards ALDH1A1. In addition, the geometry of the catalytic site is exploited for the selectivity against ALDH3A1, ALDH2 and all the ALDH isoforms that process medium and small aldehydes. Sixteen compounds, resulting as the best hits from the virtual screening, were tested for inhibition on ALDH1A3 with a spectrometric high throughput screening. Out of sixteen hits, four compounds showed more than 50% inhibiting activity at 100 μM. Among the selected molecules compounds 3 and 9 showed a promising IC_{50} of 21 and 52 μM, respectively. Considering that the

two hits are the result of a preliminary screening between only sixteen commercially available molecules, the compounds 3 and 9 are interesting starting points for the development of more potent inhibitors. Moreover the two ligands seem to be selective on ALDH1A3 towards ALDH1A1, which are the two critical ALDH isoenzymes overexpressed in CSCs. The structure determination of hALDH1A3 in complex with NAD and REA among with the rational design of small molecules are potent tools for the identification of lead compounds and for the development of an isoform specific probe that would be of interest for prognosis or diagnostic purposes.

Bibliography

1 Zhao, D., McCaffery, P., Ivins, K. J., Neve, R. L., Hogan, P., Chin, W. W. and Dräger, U. C. (1996) Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase. Eur. J. Biochem. FEBS **240**, 15–22.

2 Marcato, P., Dean, C. A., Pan, D., Araslanova, R., Gillis, M., Joshi, M., Helyer, L., Pan, L., Leidal, A., Gujar, S., et al. (2011) Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. Stem Cells Dayt. Ohio **29**, 32–45.

3 Shao, C., Sullivan, J. P., Girard, L., Augustyn, A., Yenerall, P., Rodriguez-Canales, J., Liu, H., Behrens, C., Shay, J. W., Wistuba, I. I., et al. (2014) Essential role of aldehyde dehydrogenase 1A3 for the maintenance of non-small cell lung cancer stem cells is associated with the STAT3 pathway. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. **20**, 4154–4166.

4 Mao, P., Joshi, K., Li, J., Kim, S.-H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U. R., Benos, P. V., Smith, L., et al. (2013) Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. Proc. Natl. Acad. Sci. U. S. A. **110**, 8644–8649.

5 Feldman, R. I. and Weiner, H. (1972) Horse liver aldehyde dehydrogenase. II. Kinetics and mechanistic implications of the dehydrogenase and esterase activity. J. Biol. Chem. **247**, 267–272.

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

7 Bordelon, T., Montegudo, S. K., Pakhomova, S., Oldham, M. L. and Newcomer, M. E. (2004) A disorder to order transition accompanies catalysis in retinaldehyde dehydrogenase type II. J. Biol. Chem. **279**, 43085–43091.

8 Sobreira, T. J. P., Marlétaz, F., Simões-Costa, M., Schechtman, D., Pereira, A. C., Brunet, F., Sweeney, S., Pani, A., Aronowicz, J., Lowe, C. J., et al. (2011) Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. Proc. Natl. Acad. Sci. U. S. A. **108**, 226–231.

9 Hammen, P. K., Allali-Hassani, A., Hallenga, K., Hurley, T. D. and Weiner, H. (2002) Multiple conformations of NAD and NADH when bound to human cytosolic and mitochondrial aldehyde dehydrogenase. Biochemistry (Mosc.) **41**, 7156–7168

10 Perez-Miller, S. J. and Hurley, T. D. (2003) Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. Biochemistry (Mosc.) **42**, 7100–7109.

Acknowledgments

Vorrei ringraziare il professor Menico Rizzi per avermi dato la possibilità di crescere dal punto di vista sia scientifico sia umano tramite il percorso di dottorato presso il laboratorio di Biochimica e Biologia Strutturale.

Ringrazio la dott.ssa Silvia Garavaglia per la dedizione che ha mostrato nel insegnarmi l'arte dei cristalli di proteina e per avermi guidato durante gli studi di dottorato.

Ringrazio la dott.ssa Franca Rossi per la sua disponibilità e la sua conoscenza sempre a disposizione di tutti in laboratorio.

Ringrazio i compagni di laboratorio Stefano, Davide, Riccardo e Samar per avermi trasmesso la loro passione per la scienza, per avere sempre reso leggere le giornate in laboratorio e per avermi consigliato durante questi tre anni.

Ringrazio il dott. Robert W Sobol e tutti i colleghi del Mitchell Cancer Institute per avermi accolto nel loro gruppo di ricerca e per avermi fatto sentire a casa, anche in Alabama.

Ringrazio Alessandra che ormai da anni mi sopporta e che ha scelto di sopportarmi ancora a lungo.

Ringrazio la mia famiglia e la famiglia di Alessandra che sono sempre presenti e che sanno sempre consigliarmi per il meglio.

Ringrazio tutti i miei amici, se sono quel che sono è anche colpa loro.

"Ma nulla si conosce interamente finché non vi si è girato tutt'attorno per arrivare al medesimo punto provenendo dalla parte opposta"
A. Schopenhauer