International Journal of Biological Macromolecules 242 (2023) 124749

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Raloxifene and bazedoxifene as selective ALDH1A1 inhibitors to ameliorate cyclophosphamide resistance: A drug repurposing approach

Gera Narendra [a], Baddipadige Raju [a], Himanshu Verma [a], Manoj Kumar [a], Subheet Kumar Jain [b], Gurleen Kaur Tung [c], Shubham Thakur [b], Rasdeep Kaur [d], Satwinderjeet Kaur [d], Bharti Sapra [a], Pankaj Kumar Singh [e], Om Silakari [a,*]

[a] Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab 147002, India
[b] Department of Pharmaceutical Sciences, Guru Nanak Dev University, Amritsar, India
[c] Centre for Basic and Translational Research in Health Sciences, Guru Nanak Dev University, Amritsar, India
[d] Department of Botany and Environmental Sciences, Guru Nanak Dev University, Amritsar, India
[e] Integrative Physiology and Pharmacology, Institute of Biomedicine, Faculty of Medicine, University of Turku, FI-20520 Turku, Finland.

ARTICLE INFO

Keywords:
ALDH1A1
Cyclophosphamide resistance
Drug repurposing
Raloxifene
Bazedoxifene

ABSTRACT

Cyclophosphamide (CP) is one of the most widely used anticancer drugs for various malignancies. However, its long-term use leads to ALDH1A1-mediated inactivation and subsequent resistance which necessitates the development of potential ALDH1A1 inhibitors. Currently, ALDH1A1 inhibitors from different chemical classes have been reported, but these failed to reach the market due to safety and efficacy problems. Developing a new treatment from the ground requires a huge amount of time, effort, and money, therefore it is worthwhile to improve CP efficacy by proposing better adjuvants as ALDH1A1 inhibitors. Herein, the database constituting the FDA-approved drugs with well-established safety and toxicity profiles was screened through already reported machine learning models by our research group. This model is validated for discriminating the ALDH1A1 inhibitors and non-inhibitors. Virtual screening protocol (VS) from this model identified four FDA-approved drugs, raloxifene, bazedoxifene, avanafil, and betrixaban as selective ALDH1A1 inhibitors. The molecular docking, dynamics, and water swap analysis also suggested these drugs to be promising ALDH1A1 inhibitors which were further validated for their CP resistance reversal potential by in-vitro analysis. The in-vitro enzymatic assay results indicated that raloxifene and bazedoxifene selectively inhibited the ALDH1A1 enzyme with IC_{50} values of 2.35 and 4.41 μM respectively, whereas IC_{50} values of both the drugs against ALDH2 and ALDH3A1 was >100 μM. Additional in-vitro studies with well-reported ALDH1A1 overexpressing A549 and MIA paCa-2 cell lines suggested that mafosfamide sensitivity was further ameliorated by the combination of both raloxifene and bazedoxifene. Collectively, *in-silico* and in-vitro studies indicate raloxifene and bazedoxifene act as promising adjuvants with CP that may improve the quality of treatment for cancer patients with minimal toxicities.

1. Introduction

Cyclophosphamide (CP), is a popular therapeutic option for a variety of malignancies, including chronic lymphocytic leukemia, chronic myelocytic leukemia, acute myeloid leukemia, breast cancer, and ovarian cancer [1]. However, cancer chemotherapy is often weighed down by drug resistance that hinders effective treatment [2]. Several reports indicated CP suffers from the problem of resistance via efflux, ineffective uptake, and inactivation via tumoral drug-metabolizing enzymes (DME) [3–5]. Currently, CP resistance via the inactivation mechanism is of great interest among researchers. As per the literature, DME aldehyde dehydrogenases (ALDHs) are primarily involved in CP inactivation [6].

ALDHs are phase-I oxidase enzymes that catalyze the $NAD(P)^+$-

Abbreviations: CP, Cyclophosphamide; VS, Virtual screening; DME, Drug-metabolizing enzyme; ALDH, Aldehyde dehydrogenase; 4-OH CP, 4-hydroxycyclophosphamide; PM, Phosphoramide mustard; ML, Machine learning; MD, Molecular dynamics; PDB, Protein Data Bank; RMSD, Root mean square deviation; MC, Monte Carlo; TI, Thermodynamics integration; FEF, Free energy perturbation; Maf, Mafosfamide; NCCS, National Centre for Cell Science.

* Corresponding author.
E-mail address: mmlpup73@gmail.com (O. Silakari).

https://doi.org/10.1016/j.ijbiomac.2023.124749
Received 9 December 2022; Received in revised form 25 February 2023; Accepted 1 May 2023
Available online 7 May 2023

dependent oxidation of aldehydes (R-C(=O)-H) to acids (R-C(=O)-OH) [7]. These are involved in detoxifying various endogenous and exogenous compounds such as retinaldehyde, acetaldehyde, neurotransmitters, carbohydrates, and lipids [8]. The erratic biological activity of ALDHs and their role in metabolic pathways have been linked to a range of diseases, including alcoholic liver diseases and cancer [9–11]. Several studies have manifested that overexpression of certain ALDHs, especially ALDH1A1, in a variety of tumors is associated with CP resistance [12,13]. CP is a prodrug that gets converted into 4-hydroxycyclophosphamide (4-OH CP) by a group of hepatic CYP_{450} and subsequently, bio transforms to its tautomer aldophosphamide, an active intermediate. This intermediate permeates into the cell and is converted into the active metabolite, i.e., phosphoramide mustard (PM), through spontaneous β-elimination [14]. However, the conversion of CP to PM is shunted at the aldophosphamide stage due to its conversion into an inert metabolite carboxyphosphamide via ALDH1A1 [15]. Several shreds of evidence suggest that ALDH1A1 overexpression is dependent on genetic variations, protein-protein interactions, and transcription factors [16,17]. These factors, associated with ALDH1A1 overexpression, eventually trigger CP inactivation. The detailed mechanism of ALDH1A1 overexpression and its mediated CP inactivation is displayed in Fig. 1. In many studies, it has been reported that ALDH1A1 knockdown ameliorates CP sensitivity by reducing ALDH1A1-mediated metabolism. For instance, Muramoto et al. reported that inhibition of ALDH1A1 activity in murine pluripotent hematopoietic stem cells and myeloid progenitor cells correlated well with increased in-vitro sensitivity to 4-OH CP [18].

Understanding the utility of exploring ALDH1A1 as a target, various research groups have reported ALDH1A1 inhibitors belonging to diverse chemical classes including theophyllines, indole-2,3-diones, aromatic lactones, pyrimidones, quinolines, and 3H-xanthen-3-ones [19–24]. Among them, NCT-501, a theophylline-based compound, is reported as a selective ALDH1A1 inhibitor with excellent inhibitory activity. However, the pharmacokinetic investigations revealed that NCT-501 possesses a short half-life of <1 h. Such results highlight the difficulty in the development of novel ALDH1A1 inhibitors [25]. Hence, to obtain selective ALDH1A1 inhibitors with well-established pharmacokinetic, safety, and toxicity profiles, herein drug repurposing approach has been explored [26,27]. The FDA-approved drugs in the ZINC database were virtually screened via machine learning (ML) models reported by our research group [28] to identify promising ALDH1A1 inhibitors. The obtained hits were further subjected to structure-based drug-designing approaches, including molecular docking, molecular dynamics (MD), and waterswap analysis. Finally, in-vitro enzymatic and cell line studies were carried out for the identified drugs to validate the computational results and suggest appropriate solutions to CP resistance.

2. Material and methods

2.1. FDA-approved drug collection and preparation

A total of 1600 FDA-approved drugs were retrieved from the ZINC database [29]. The structure files for the molecules were downloaded in SDF format. The collected 2D structures were converted into 3D structures using the Spark module of Cresset software. The obtained 3D molecules were further subjected to ligand preparation to add explicit hydrogens. Once the ligand preparation was performed, PaDEL software was used to generate the 1D, 2D, and 3D descriptors for prepared 3D structures of FDA-approved drugs [30]. The generated descriptor file was later used for ML-based virtual screening (VS).

2.2. Virtual screening

Previously developed and reported multiple ML models (ALDH1A1, ALDH2, and ALDH3A1) were utilized to screen FDA-approved drugs [28]. The screening process was conducted using the same protocol as discussed in our previous work [28]. Briefly, the ALDH1A1 model was utilized to screen a total of 1600 FDA-approved drugs as putative ALDH1A1 inhibitors. Later, the identified ALDH1A1 inhibitors were

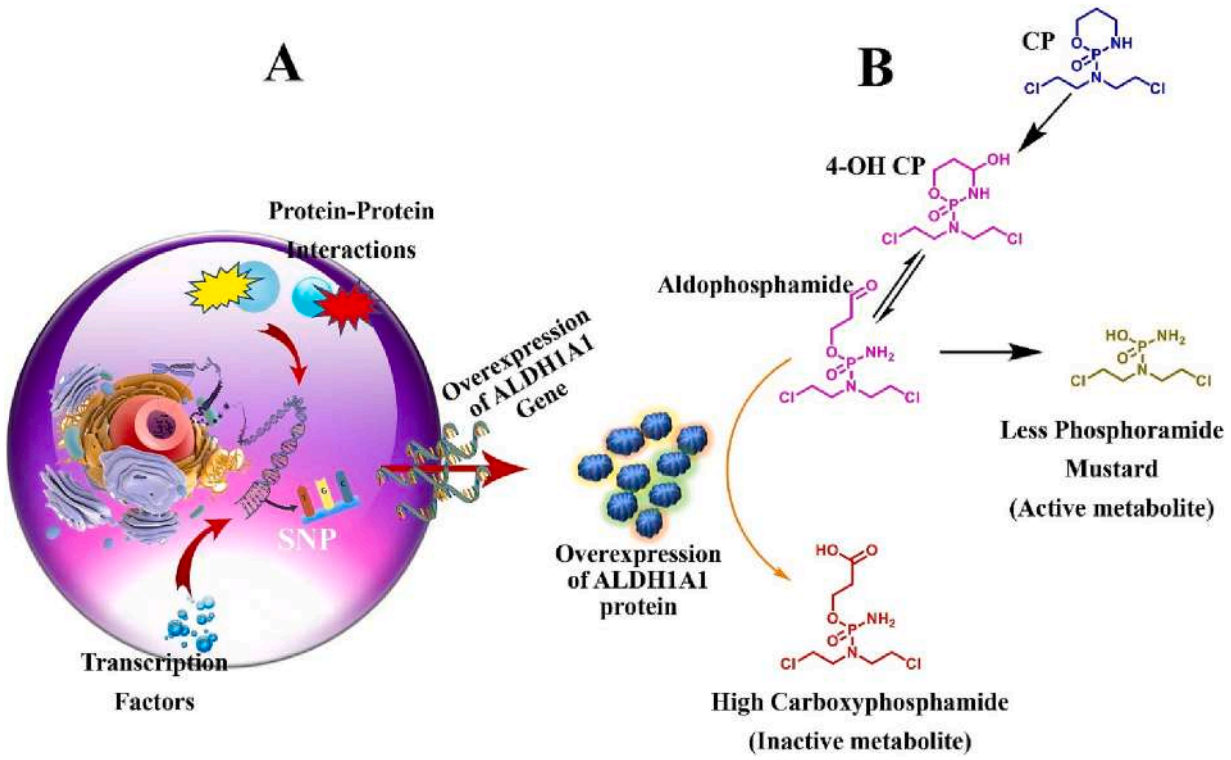

Fig. 1. Factors regulating ALDH1A1 overexpression and its mediated CP inactivation via A) protein-protein interactions, transcription factors, and SNPs, B) Conversion of CP into 4-OH CP and subsequently to its tautomer aldophosphamide, which undergoes inactivation to carboxyphosphamide via ALDH1A1 leaving scanty of active metabolite i. e. phosphoramide mustard for anti-cancer activity.

screened through validated ALDH2 and ALDH3A1 models to eliminate non-selective inhibitors. Finally, to establish their mechanism of binding with ALDH1A1 and ensure the retention of crucial interactions, molecular docking was carried out.

2.3. Molecular docking

3D crystal structure of ALDH1A1 i.e. 4X4L was retrieved from the RCSB Protein Data Bank (PDB) [31]. Our research group has selected 4X4L based on cross-docking experimentation. The cross-docking protocol has been briefly explained in our recently published paper [28]. This experimentation included five 3D crystal structures of ALDH1A1 with resolution <2 Å. All the selected crystal structures were aligned and subsequently, ligands from each crystal structure were extracted and redocked in the active site of each crystal structure. The root mean square deviation (RMSD) between the redocked crystal ligands and their native crystal conformations was computed. Based on the cross-docking analysis, the protein structure with a low average RMSD of redocked ligands i.e. 4X4L was considered for docking studies [28]. The crystal structure selected from redocking studies was prepared using the Flare module by following the insertion of missing atoms in incomplete residues, modeling the missing loops, removal of co-crystallized water, and protonation of the residues by applying a LeadFinder's force field [32]. The binding site is defined around the centroid of the co-crystallized ligand for the generation of the grid. Finally, docking was performed in the active site of the protein using BioMolTech's Lead Finder docking algorithm.

2.4. Molecular dynamics

The binding modes obtained from the molecular docking analysis only provide the least energy-stable conformation of a ligand within the receptor along with key electrostatic and hydrophobic interactions. To further ensure the ligand stability within the active site of ALDH1A1, MD study was carried out using the Flare module in Cresset Software [32]. For this, the biomolecular complex was solvated in an orthorhombic box with 10 Å dimensions using the TIP3P water molecules system. The AMBER force field in Flare was applied to the protein-ligand system. Once the system was minimized, MD simulations for the period of 100 ns were carried out under the NPT ensemble with a temperature setting of 300 $^{\circ K}$ and recording intervals of 2 ps throughout the simulation period. After the completion of MD, the binding orientations of the ligands within the active site of the protein and RMSD trajectories were further analyzed.

2.5. Energy calculations

The binding affinity of ALDH1A1 toward obtained top hits was determined using WaterSwap analysis available in the Flare module of Cresset software [33,34]. WaterSwap is a technique based on Monte Carlo (MC) simulation that is generally employed to investigate ligand-protein interactions and further calculate the absolute protein-ligand binding free energies. WaterSwap in Flare employs condensed-phase simulations using the AMBER force field. The starting conformation for WaterSwap was chosen from the final conformation obtained after MD simulations. It works by swapping the ligand bound to the protein with an equivalent shape and volume of bulk water. The binding free energy is then calculated by three different methods, such as Bennett, thermodynamics integration (TI), and free energy perturbation (FEF). A consensus of binding free energy obtained by the arithmetic mean of Bennett, TI, and FEP has been considered as total binding energy.

2.6. ALDH isoform-specific enzymatic assay

Human ALDH1A1, ALDH2, and ALDH3A1 enzymes were purchased from Abcam. The FDA-approved drugs avanafil and betrixaban were purchased from Sigma Aldrich. While raloxifene and bazadoxifene were purchased from Vivan Life Sci. Pvt. Ltd., Mumbai. The standard inhibitors NCT-501 and CB29 were purchased from Sigma Aldrich. The ALDH2 inhibitor disulfiram was received as a gift sample from Tripada Healthcare Pvt. Ltd., Ahmedabad, India. The other required chemicals including Beta-nicotinamide adenine dinucleotide and DTT were purchased from Sigma Aldrich.

The inhibitory activity of ALDH isoforms was assayed spectrophotometrically by monitoring the formation of NADH at 340 nm. Briefly, 1 μg/mL ALDH1A1 was dispatched into a 96-well plate. Later, the enzyme plate was treated with different concentrations of FDA-approved drugs, standard inhibitors, and control were incubated for 15 min at 25 °C. The reaction was initiated by the addition of a substrate mixture containing 10 mM propionaldehyde, 2 mM DTT, 100 mM KCl, 1 mM NAD^+, and 50 mM Tris pH 8.5. Then, the absorbance of NADH was recorded in kinetic mode for 5 min. NCT-501 was used as a standard in the ALDH1A1 inhibition assay. Similarly, for ALDH2, 0.5 μg/mL of the enzyme was used in the reaction with 2 mM acetaldehyde as the substrate, and 0.2 μg/mL of the enzyme was used in the ALDH3A1 inhibition assay with 1 mM 4-nitrobenzaldehyde as the substrate. Disulfiram and CB29 were used as standards in ALDH2 and ALDH3A1 reactions, respectively [22].

2.7. Mafosfamide sensitivity assay

An MTT assay was performed to identify the cytotoxicity potential of Mafosfamide (maf) alone and in the combination of test compounds among the reported ALDH1A1 overexpressing cell lines i.e. A549 and MIA PaCa [35] (Yasgar et al.,2017). Here the maf was purchased from Niomech-IIT GmbH, Germany, D-17272. Maf was chosen for this study because it is a cyclophosphamide analog that does not require cytochrome P450 to activate, making it perfect for cell-based research. Two cell lines, A549 and MIA-Paca-2, were chosen because they express high levels of ALDH1A1 [35]. A549 and MIA PaCa-2 cells were purchased from the National Centre for Cell Science (NCCS), Pune, India. The MTT assay was performed according to protocols adopted in a report published by Parajuli in 2014 [36] in BSL-3, Department of Botany and Environmental Sciences, Guru Nanak Dev University, Amritsar, India.

3. Results

3.1. Virtual screening

3.1.1. ML and molecular docking-based VS

To identify selective ALDH1A1 inhibitors which could work as adjuvant drugs with CP, already reported ML models for ALDH1A1, ALDH2, and ALDH3A1 were utilized to screen a total of 1600 FDA-approved drugs. As a result of an initial screening from the ALDH1A1 model, around 528 molecules filtered out as ALDH1A1 inhibitors. To remove non-selective ALDH1A1 inhibitors, the obtained 528 molecules were screened from ML models for ALDH2 and ALDH3A1. Overall, a total of 115 molecules were found as selective ALDH1A1 inhibitors. The detailed combined VS protocol is displayed in Fig. 2. In the next step, the obtained hits were subjected to molecular docking in the active site of ALDH1A1 (PDB ID: 4X4L). The docking protocol was first validated via redocking of the co-crystallized ligand in the 4X4L structure. The poses of both redocked and co-crystallized ligands were compared and superimposed. The RMSD value of 0.49 Å was observed among the poses. It is reported that a maximum of ≤2.0 RMSD between the co-crystal ligand and redocked structure is considered realistic (Fig. 3). The same grid generated around the co-crystallized ligand was utilized for carrying out the docking of the obtained 115 hits. Among them, a total of 13 molecules were selected based on their best Lead Finder scores (LF rank score, LF dG score, and LF VSscore) compared to what was obtained with the co-crystalized ligand of ALDH1A1 i.e., -10.16, -10.05, and -09.21, respectively. Herein, the LF rank score indicates the best ligand pose, which is similar to the experimental observation. The

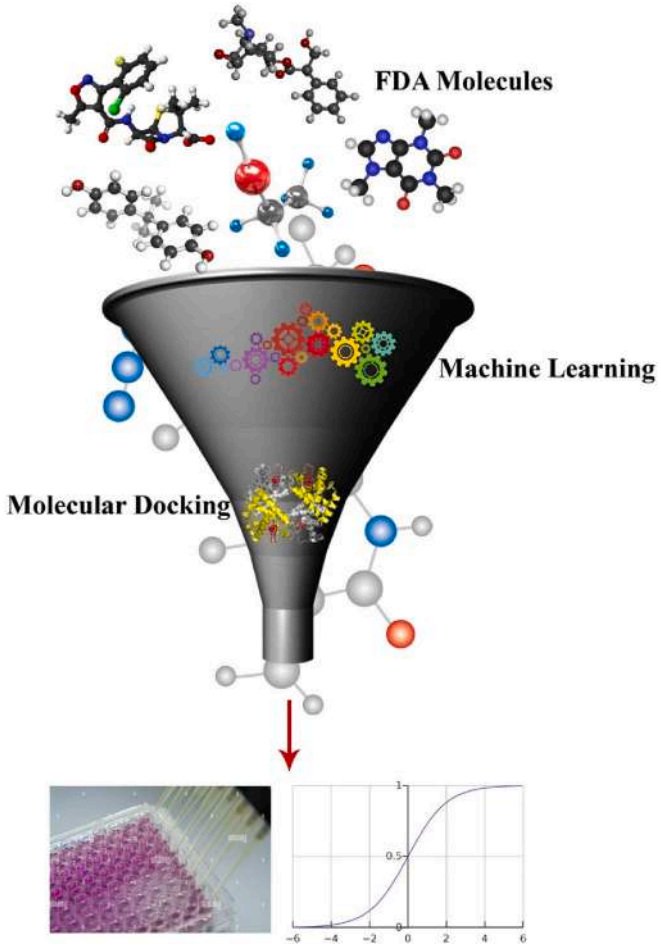

Fig. 2. Brief representation of the ML- and docking-based VS protocol.

more negative the LF rank score is, the higher the chance that the docked poses would reproduce the crystallographic pose. The LF dG scoring function estimates the protein-ligand binding free energy in the gas phase. The LF VS score represents efficacy in the VS experiment; the more negative the score is, the higher the probability of ligands being more active. The detailed LF scores for the screened hits are enlisted in Table 1 along with their ZINC codes. The top-scored four molecules (Fig. 4) were further docked against other ALDH isoforms, such as ALDH2 and ALDH3A1, to distinguish selectivity among the three isoforms and to obtain the best selective ALDH1A1 molecules. None of the molecules show better LF scores for ALDH2 and ALDH3A1 than ALDH1A1 LF scores (Supplementary Table S1 and S2). The top four drugs, including raloxifene, bazedoxifene, avanafil, and betrixaban, could not be docked in the cavity of ALDH2 and ALDH3A1 due to their narrow and constricted cavity. This distinguished front and side surface view for the cavities of ALDH2 and ALDH3A1 can be visualized in Supplementary Fig. S1 and S2. The top hits in the complex with ALDH1A1 were further subjected to MD simulations.

3.2. Molecular dynamics

The MD simulation for top-docked complexes was carried out for a 100 ns period. This study further demonstrated detailed computational insights in terms of protein-ligand stability, molecular interactions, and binding modes with the target protein. The RMSD plots displayed in Fig. 5 show that all top four complexes showed good stability throughout the simulation. During the entire simulation, all top four complexes depicted RMSD within an allowable range i.e., <1 Å, indicating all the drugs retain the stability in complex with ALDH1A1.

3.3. Interaction analysis of molecules with ALDH1A1

3.3.1. Raloxifene and bazedoxifene

The docking analysis suggested that raloxifene and bazedoxifene displayed similar interaction patterns within the active site of ALDH1A1. As observed in Fig. S3A-S3B, the phenol moiety present in raloxifene and bazedoxifene demonstrated pi-pi stacking interactions with Phe171. The other phenolic moieties in raloxifene and bazedoxifene have shown pi-pi interaction with Tyr297. In addition, the piperidine moiety in raloxifene displayed pi-cation interactions with Tyr297. All the docking interactions were found to be retained throughout the simulation except the interaction between Phe171 and the phenol moiety of raloxifene. However, Phe171 came into contact with the other phenol ring of raloxifene via pi-pi interaction during the simulation and was found to be more prominent until the end of the simulation. Similarly, the pi-pi interaction between the phenol moiety on bazedoxifene and Tyr297 was maintained throughout the simulation. The pi-pi

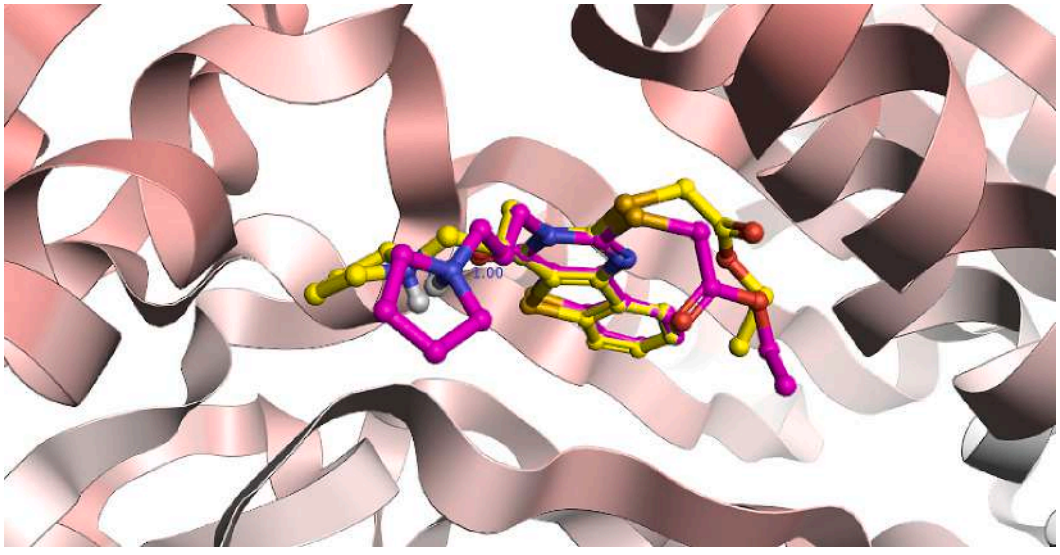

Fig. 3. Comparison between ALDH1A1 cocrystal ligand (pink) and its redocked ligand pose (yellow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Molecular docking scores of FDA approved molecules identified toward ALDH1A1.

Rank	ZINC ID	Compound Name	LF rank score[a]	LF dG score[b]	LF VSscore[c]
1	ZINC000000538275	Raloxifene	−15.55	−15.62	−16.23
2	ZINC000001895505	Bazedoxifene	−15.36	−15.44	−16.10
3	ZINC000011677857	Avanafil (Stendra)	−14.92	−15.12	−15.55
4	ZINC000030691754	Betrixaban	−13.81	−14.92	−15.24
5	ZINC000003986735	Sprycel	−12.26	−13.55	−13.28
6	ZINC000003820029	Trajenta	−11.76	−11.98	−10.36
7	ZINC000011679756	Eltrombopaq	−11.65	−11.48	−10.22
8	ZINC000011616925	Folotyn	−11.12	−10.95	−10.05
9	ZINC000084758479	Brexpiprazole	−10.84	−10-72	−10.02
10	ZINC000011617039	Pazopanib	−10.65	−10.63	−10.02
11	ZINC000008577218	Pga	−10.58	−10.42	−09.88
12	ZINC000000538658	Samsca	−10.42	−10.33	−09.72
13	ZINC000034608502	Umeclidinium	−10.21	−10.09	−09.46
14	Cocrystalized ligand	4X4L	−10.16	−10.05	−09.21

[a] Score used to rank the poses of docked compounds.
[b] Score used to rank the binding energy of docked compounds in terms of kcal/mol.
[c] Score used to rank the docked compounds in virtual screening.

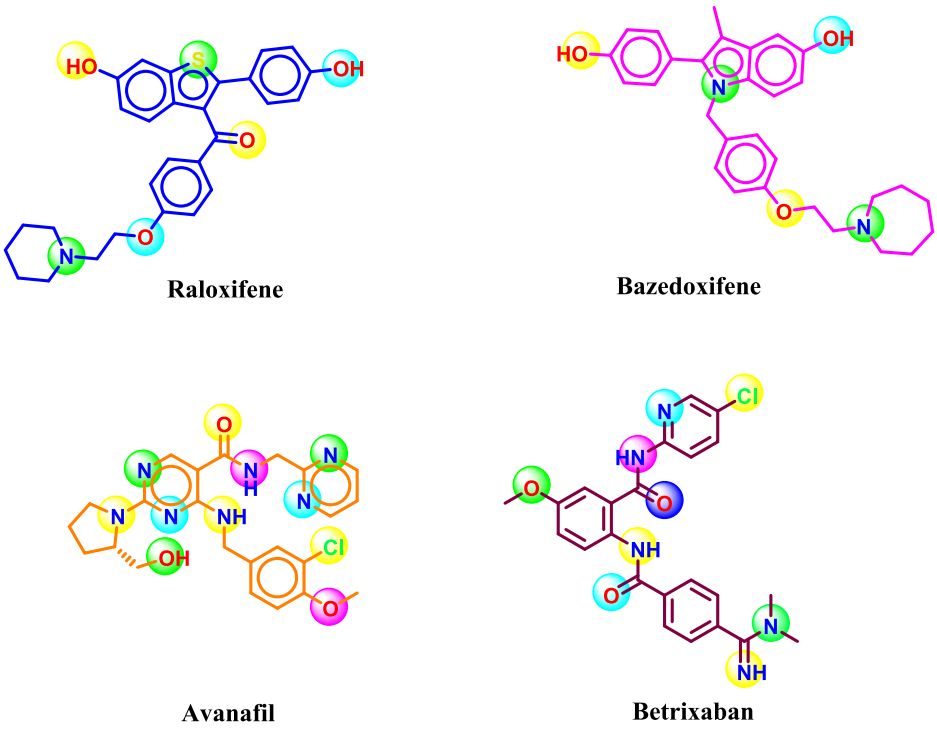

Fig. 4. The best-identified top four FDA molecules as ALDH1A1 inhibitors from Virtual Screening.

interactions with these key amino acids have already been reported to be crucial for ALDH1A1 inhibition [37,38]. The detailed pre-and post-simulation interactions of raloxifene and bazedoxifene are displayed in Fig. S3A (Supplementary) and Fig. 6A, and Fig. S3B (Supplementary) and Fig. 6B, respectively. Moreover, the initial docking studies for raloxifene revealed two weak H-bond interactions between the hydroxy group on benzothiophene and Thr129 and Ala462, respectively. A similar kind of observation was recorded between the hydroxy group on the indole moiety of bazedoxifene and Thr129. The hydroxy group on the phenol ring in both raloxifene and bazedoxifene also demonstrated a strong H-bond interaction with Cys303. However, all the H-bond interactions disappeared after 30 % of the simulation period, and observed new strong H-bond interactions between Glu255 and the hydroxy group present on the phenol ring of both raloxifene and bazedoxifene. The newly formed H-bond was considered more prominent with a % frame score value of 65 %. The benzothiophene group of raloxifene and indole group of bazedoxifene show new pi-pi interactions with the important amino acid Trp178 after MD simulation. In addition, raloxifene showed a water-mediated interaction between the hydroxy group of benzothiophene and the amine group on the side chain of Trp178 at end of the simulation.

3.3.1.1. Avanafil. The pyrimidine moiety attached to pyrrolidine mainly interacted with Tyr297 via pi-pi interactions. Similarly, the phenol moiety manifested a pi-pi interaction with Phe466 during the initial period of MD simulation. The pi-pi interaction with Phe466 was maintained throughout the simulation while the pi-pi interaction between pyrrolidine and Tyr297 was maintained for ~40 % of the total simulation period. Eventually, the pyrrolidine group came into contact with Phe466 and retained it until the end of the simulation. Two H-bonding interactions with Gly294 and Ser121 with a 50 % frame were lost at the end of the simulation. The 3D interaction poses for avanafil before and after the simulations are illustrated in Fig. S3C (Supplementary) and Fig. 6C.

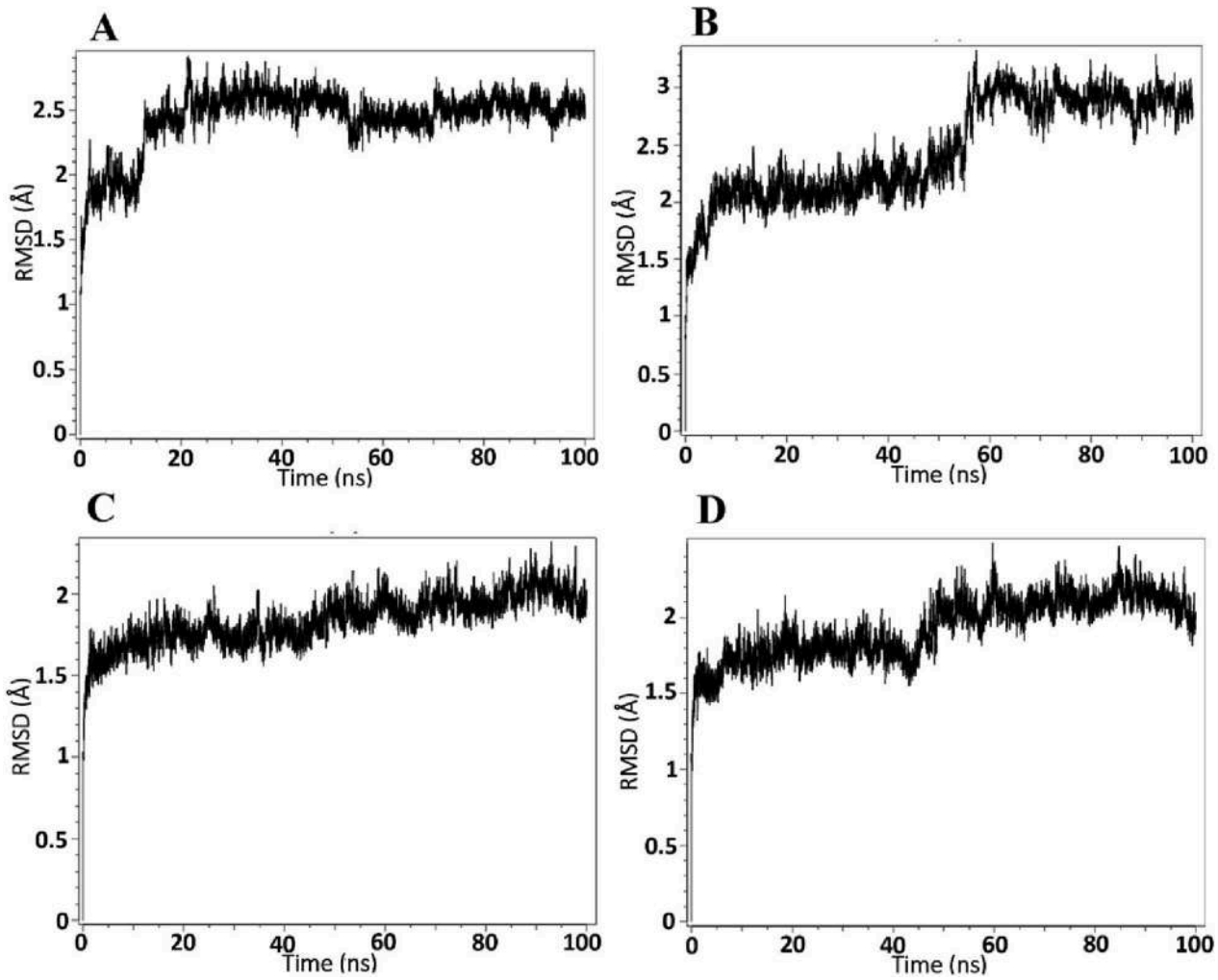

Fig. 5. RMSD analysis of ALDH1A1 in complex with A) raloxifene B) bazedoxifene C) avanafil D) betrixaban.

3.3.1.2. Betrixaban. Based on the final poses obtained from pre-and post-simulation (Fig. S3D (Supplementary) and Fig. 6D), the betrixaban interaction profiles in the ALDH1A1 active site were analyzed. During an initial period of dynamics, the benzamide moiety in betrixaban interacted with Tyr297 which was maintained for ~20 % of the simulation period, while the same in the rest of the time (80 %) displayed a new interaction with His293. The benzoyl and dimethylamino groups in this drug interact with Phe466 via pi-pi interactions, and pi-cationic interactions, respectively. The same benzoyl group displays an additional pi-cationic interaction with Trp178. The % frame contact for 100 ns trajectory suggested that the observed cationic interactions retained with 43 % while the pi-pi interaction between Phe466 and the benzoyl moiety manifested this score as 39 %. Subsequently, a new pi-pi interaction was observed between pyridine and Phe466, which was maintained throughout the simulation.

3.4. WaterSwap analysis

WaterSwap analysis was performed to determine the binding free energy and compute the energies (ΔG, kcal/mol) for each amino acid within the active site concerning the ligand that favors the inhibition. As shown in Fig. S4 (Supplementary), the green-colored amino acids are favorable for ligand interactions, and red-colored amino acids favor only the water-mediated interactions. The analysis revealed that the amino acid and ligand interactions retained after MD simulation, i.e., Tyr297, Phe171, and Glu269, were found to favorably contribute to the binding free energy for both raloxifene and bazedoxifene. The amino acid Trp178 contributed more to the water-mediated interaction for raloxifene than to the direct ligand interaction. In addition, Val460, His293, and Ser121 have been identified as additional possible hotspot amino acids for raloxifene interaction, whereas Ala462 and Gly458 have been identified as additional amino acids for bazedoxifene interaction. Furthermore, Tyr297 and Phe171 are favorable for avanafil, whereas Phe171, Ala462, and Glu255 have been identified as favorable for the betrixaban interaction with ALDH1A1. The obtained consensus of the arithmetic mean of binding energy values is −42.00 kcal/mol, −38.03 kcal/mol, −13.82 kcal/mol, and − 11.48 kcal/mol for raloxifene, bazedoxifene, avanafil, and betrixaban, respectively. The detailed energy values calculated by various methods in WaterSwap analysis are illustrated in Table 2. The obtained results show that all the ligand complexes exhibit good binding energy. This signifies that the approved drugs might show good inhibitory potentials against ALDH1A1 in in-vitro experiments. To validate the *in-silico* results of drug repurposing, in-vitro enzymatic investigations were carried out.

3.5. In-vitro enzymatic assay

The inhibitory effects of the screened drugs, including raloxifene, bazedoxifene, avanafil, and betrixaban against recombinant human ALDH1A1, ALDH2, and ALDH3A1 were measured according to the protocol reported in previous reports [22]. Of these, both raloxifene and bazedoxifene were identified as the most promising drugs as ALDH1A1

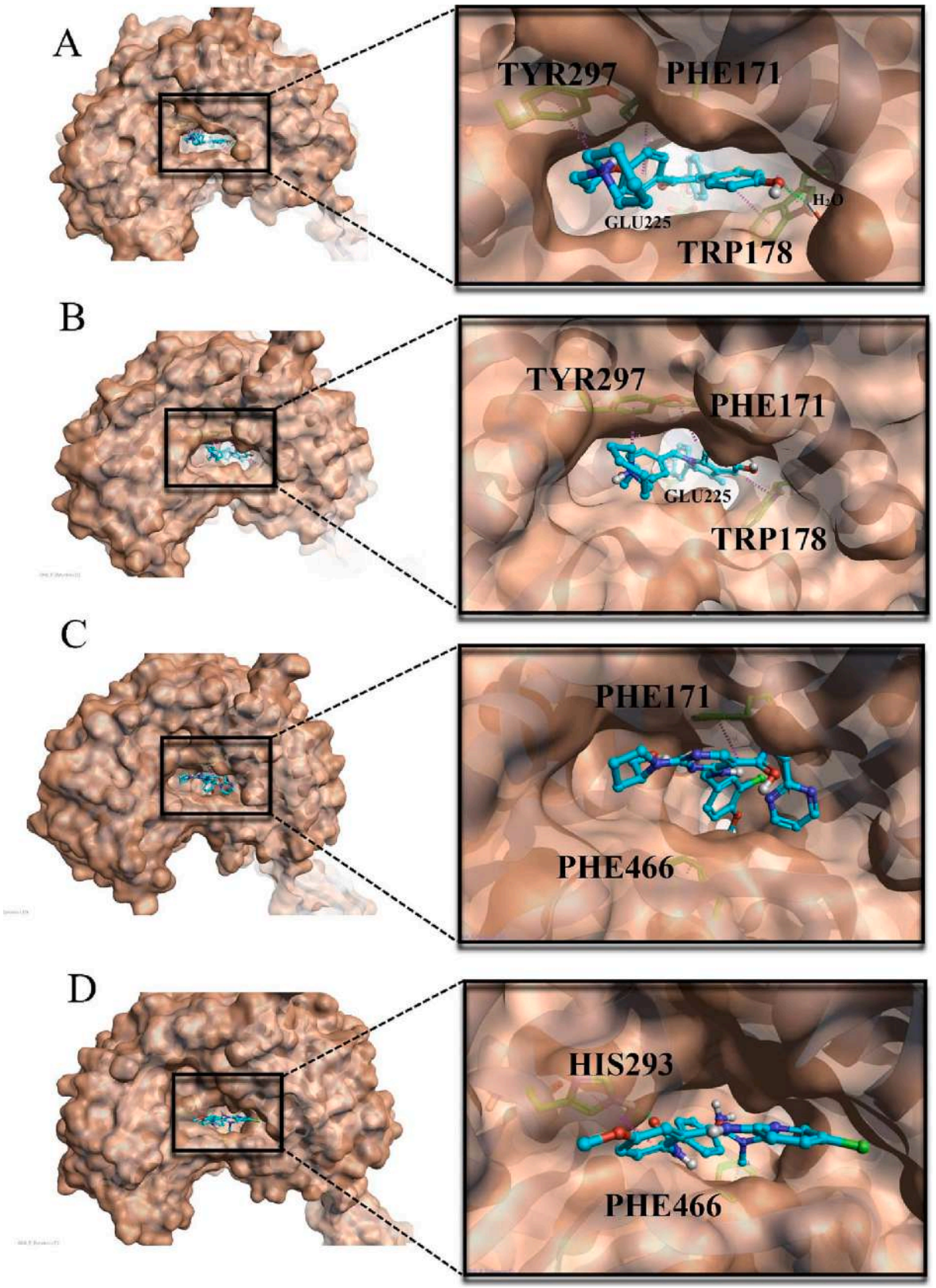

Fig. 6. The binding poses of the selected FDA molecules after MD simulation A) raloxifene, B) bazedoxifene, C) avanafil, D) betrixaban in the active site of ALDH1A1

Table 2
Binding energies calculated using WaterSwap analysis.

Drug name	Bennett	FEP[a]	TI[b]	TI Quadratic	The consensus of Arthematic mean (kcal/mol)
Raloxifene	−40.30	−41.13	−39.10	−37.57	−42.00
Bazedoxifene	−39.73	−37.26	−36.30	−38.84	−38.03
Avanafil	−11.99	−13.75	−15.24	−14.31	−13.82
Betrixaban	−10.40	−9.62	−12.36	−13.57	−11.48

[a] Free energy perturbation.
[b] Thermodynamics integration.

selective inhibitors over avanafil and betrixaban. The detailed % inhibition of all the drugs with different ALDH isoforms has been displayed in Fig. 7A, B, and C. Then the IC_{50} values were calculated using the ATT bioquest IC_{50} calculator (https://www.aatbio.com/tools/ic50-calculator). From the evaluated results it was concluded that all the tested compounds exhibited good inhibitory activity against ALDH1A1 with IC_{50} values in the range of μM concentrations (Table 3). As shown in Table 3, raloxifene and bazedoxifene were found to inhibit ALDH1A1 at 2.35 μM and 4.41 μM, respectively, while the same drugs possess very minimal inhibitory activity toward the rest of the isoforms, i.e., ALDH2 and ALDH3A1 at 100 μM. Meanwhile, avanafil and betrixaban couldn't exhibit 50 % inhibitory activity against ALDH1A1 at the given concentration (10 μM), and the drugs showed inhibitory potentials against ALDH2 and ALDH3A1 at higher concentrations. In addition, the reference NCT-501 shows 0.9 μM inhibitory potential against ALDH1A1 and minimal inhibitory activity toward ALDH2 and ALDH3A1.

Table 3
ALDH isoforms inhibition by raloxifene, bazedoxifene, avanafil, and betrixaban.

S. No.	Compound name	ALDH IC_{50} (μM) ± SD[b]		
		ALDH1A1	ALDH2	ALDH3A1
1.	Raloxifene	2.35 ± 0.45	>100(NI[a])	>100(NI[a])
2.	Bazedoxifene	4.41 ± 0.50	>100(NI[a])	>100(NI[a])
3.	Avanafil	>10	73.38 ± 3.86	64.72 ± 5.77
4.	Betrixaban	>10	79.71 ± 3.79	69.28 ± 4.81
5.	NCT-501 (Reference)	0.9 ± 0.42	>100(NI[a])	>100(NI[a])

[a] No inhibition.
[b] Standard deviation.

3.6. Mafosfamide sensitivity assays

As discussed in the methodology section, two cell lines, A549 and MIA PaCa-2 were used in the present study to identify maf sensitivity with the combination of raloxifene and bazedoxifene. These FDA-approved drugs were selected because they manifested greater inhibition and selectivity toward ALDH1A1 in the enzymatic assay studies. The obtained MTT results were expressed as % growth inhibition (Mean ± S. D.) v/s different concentrations of the FDA compounds + fixed concentration of maf ((i. e. half of the IC_{50} (138.04 for A549) and (128.08 for MIA PaCa-2) obtained individually)). As shown in Fig. 8A and B the cytotoxicity effect of maf on A549 and MIA-PaCa-2 was significantly improved with the combination of treatment with maf + raloxifene compared to maf alone. Similarly, the cytotoxicity of maf on A549 was also improved by the combination of maf + bazedoxifene than maf alone, while the same combination indicates no cytotoxicity effect on MIA PaCa-2 at the given concentrations. The graphs also clearly indicate that an increase in the concentration of raloxifene and bazedoxifene led to a reduction in cell growth viability. In addition,

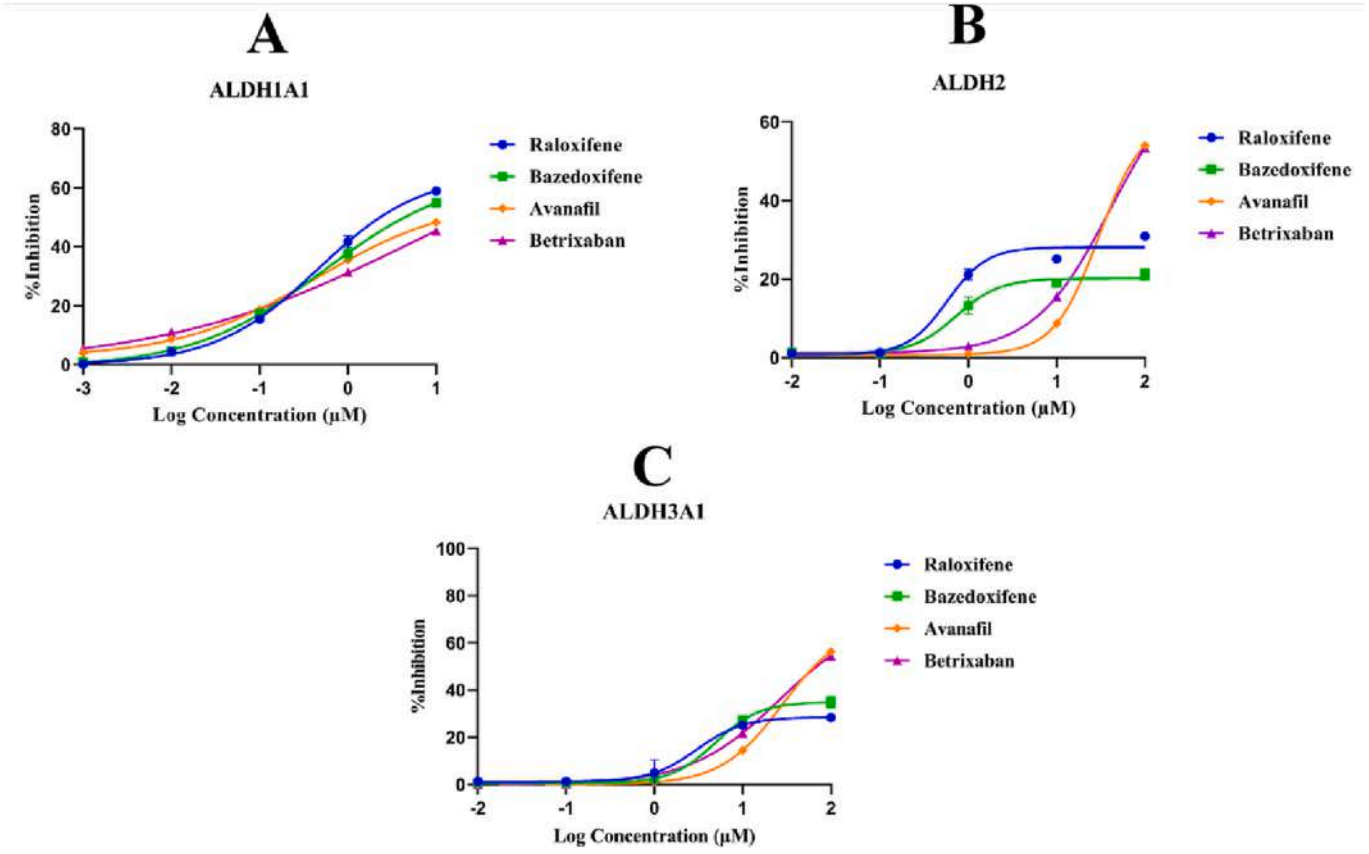

Fig. 7. Concentration response curve to identify % inhibition A) ALDH1A1, B) ALDH2, and C) ALDH3A1; D) In-vitro activity results indicating IC_{50} of FDA drug for ALDH1A1, ALDH2, and ALDH3A1.

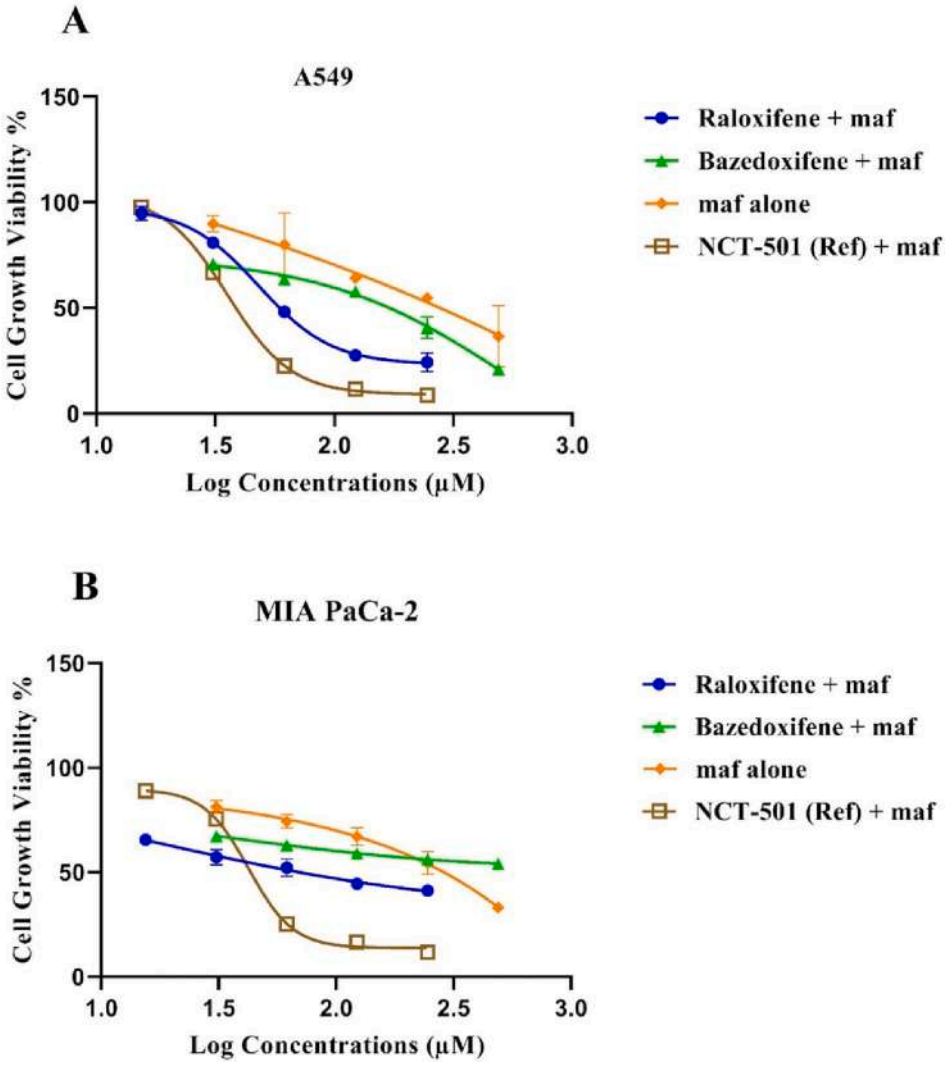

Fig. 8. Mafosfamide sensitivity assay in A) A549 and B) MIA PaCa-2 cells.

treatment of A549 and MIA PaCa-2 cells with maf in the presence of reference compounds i.e. NCT-501 showed a marked increase in maf sensitivity compared to maf alone. The antiproliferative potentials of FDA-approved drugs, such as raloxifene and bazedoxifene alone and in combination with maf were compared to reference compound NCT-501 (Table 4). The results indicated that the repurposed drugs alone indicated antiproliferative activity in both cell lines. The maf alone possesses poor antiproliferative activity toward both cell lines. However, the combination of raloxifene and bazedoxifene with maf showed the highest maf chemosensitivity than maf alone on both the cell lines. These results indicate that the use of bazedoxifene and raloxifene with maf at low concentrations could eliminate the resistance. Also, the combination enhances the maf sensitivity in ALDH1A1-overexpressing tumor types.

Table 4
In-vitro antiproliferative activity of Identified drugs alone and in combination with maf using A549 and MIA PaCa-2 cell lines.

S. No.	Drugs (Alone/ Combination)	A549 cell line (IC_{50} (μM) ± SD)	MIA PaCa-2 cell line (IC_{50} (μM) ± SD)
1.	Raloxifene alone	91.18 ± 0.21	188.32 ± 0.12
2.	Bazedoxifene alone	371.32 ± 0.11	385.23 ± 0.32
3.	NCT-501 alone	68.92 ± 0.56	179.55 ± 0.65
4.	Maf alone	276.09 ± 1.25	256.17 ± 2.10
5.	Raloxifene + Maf	71.71 ± 0.01	79.66 ± 0.05
6.	Bazedoxifene + Maf	130.60 ± 0.12	>500 ± 0.01
7.	NCT-501 + Maf	38.55 ± 0.32	60.25 ± 0.24

4. Discussion

Drug resistance is one of the most common causes of cancer therapeutic failure. Changes in drug uptake or efflux rates, target mutations, gene amplification, enhanced DNA repair, and overexpression of DMEs are a few resistance mechanisms. Among them, DME-mediated chemoresistance is one of the most overlooked causes that has become a target of interest for researchers. CP, one of the important and most prescribed chemotherapeutics, is reported to undergo inactivation to an inactive metabolite via a DME known as ALDH1A1. Moreover, its clinical utility is often limited by bone marrow toxicity especially when it is administered in higher doses in resistance conditions [39]. Thus, the development of novel ALDH1A1 inhibitors can overcome both CP resistance and toxicity problems.

In the present study, a drug repurposing approach was carried out to screen some approved drugs as selective ALDH1A1 inhibitors. Herein this approach has been implemented to identify approved drugs as selective ALDH1A1 inhibitors with well-established safety, pharmacokinetics, and toxicity profiles, which is a great issue for most of the reported inhibitors of this target. Considering already published machine learning models against the targets ALDH1A1, ALDH2, and

ALDH3A1 by our research group, a total of 115 molecules were scrutinized as selective ALDH1A1 inhibitors. Among the obtained hits, four drugs, raloxifene, bazedoxifene, avanafil, and betrixaban, were considered for MD simulation based on their compromised docking scores against ALDH isoforms, i.e., ALDH2 and ALDH3A1. Raloxifene and bazedoxifene are selective estrogen receptor modulators [40]. Raloxifene is primarily used to prevent osteoporosis in postmenopausal women, and bazedoxifene is used to treat hot flushes in women during menopause [41,42]. The two drug molecules share high structural similarities with previously reported selective ALDH1A1 inhibitors by Li et al., and Yang et al., [23,37]. In addition, these drugs show good binding interactions with important amino acids responsible for ALDH1A1 inhibition along with high binding energy scores. These results encouraged to subject of these drugs for further in-vitro enzymatic assay studies. The IC50 values for raloxifene and bazedoxifene were found to be 2.35 μM and 4.41 μM by inhibition assays, respectively. However, due to surface topography dissimilarity among the ALDH isoforms, neither raloxifene nor bazedoxifene exhibited any inhibitory activity against ALDH2 and ALDH3A1. As per Morgan and Hurley, molecules showing steric clashes with residues, including Asp458 of ALDH2 and Ile458 of ALDH3A1, prohibit the compounds from inhibiting these enzymes [43]. A similar kind of observation was identified in the present study. Both raloxifene and bazedoxifene demonstrated strong clashes with the bulky amino acids in ALDH2 and ALDH3A1. Moreover, bazedoxifene show some additional steric clashes with other bulky amino acids present on the surface of ALDH3A1, the details were provided in the supplementary file. The final obtained molecules had specific features and structural shapes that fit into ALDH1A1 and possessed the ability to bind to its wider active site. The narrow entry pathways of both ALDH2 and ALDH3A1 eliminate the possibility of binding of raloxifene and bazedoxifene. These *in-silico* results were in corollary to in-vitro enzymatic inhibition studies where raloxifene and bazedoxifene showed selective inhibition against ALDH1A1 and weak inhibition against ALDH2 and ALDH3A1. Some of the already reported literature evidence indicate that raloxifene and bazedoxifene are used in cancer treatment [44–46]. In particular, raloxifene is approved for invasive breast cancer in postmenopausal women. Its combination with other anticancer drugs has also been reported to have optimal therapeutic effects. A combination of raloxifene and letrozole has been reported to possess an additive cytotoxic effect in the MCF-7-cell line [47]. Similarly, the combination of raloxifene and conjugated estrogen has been reported to reduce the severity of menopausal symptoms [48]. The effect of the combination of raloxifene and CP has not yet been reported, the drug-drug interaction information deposited in the drug bank supports the fact that the metabolism of CP decreased when combined (https://go.drugbank.com/drugs/DB00481) with raloxifene. In addition, a study by Zhong et al. confirmed that raloxifene shows the inhibition of recombinant ALDH1A1 [49]. Likewise, bazedoxifene can also be used in a variety of cancers, including breast cancer, pancreatic cancer, and gastric cancer [50–52]. Moreover, in the present investigation, raloxifene and bazedoxifene were observed to inhibit ALDH1A1 significantly, so the combination of these drugs with CP may be beneficial for the CP resistance problem.

Avanafil and betrixaban are other FDA drugs obtained from VS as ALDH1A1 inhibitors. Both drugs were found to inhibit ALDH1A1 with an IC_{50} value of >10 μM. Moreover, these drugs also showed inhibitory activity against ALDH2 and ALDH3A1 with higher IC_{50} values (>100 μM) than ALDH1A1. This is might be due to their pi-pi interactions with the ALDH2 and ALDH3A1 surface amino acids. However, the molecules still have greater IC_{50} (>10) values toward ALDH1A1. Therefore, the best-identified raloxifene and bazedoxifene were further considered for cell line studies.

The influence of ALDH1A1 inhibitors on the anticancer activity of maf was examined in A549 and MIA PaCa-2 cell lines. A significant increase in sensitivity toward maf was observed when given in the combination with raloxifene and bazedoxifene. However, the maf sensitivity in the above case is not higher than what was observed with the combination of maf and standard ALDH1A1 inhibitor (NCT-501). But, as described in the introduction NCT-501 has a lower half-life i.e., <1 h. Therefore, the obtained FDA molecules with known pharmacokinetics are more beneficial than the standard ones. Moreover, the FDA molecules show optimal maf sensitivity in both cell lines. Thus, both raloxifene and bazedoxifene could be the solution to the CP resistance.

5. Conclusion

To ameliorate the problem associated with CP resistance owing to ALDH1A1 in cancer cell lines, herein drug repurposing approach is implemented using various ligand and structure-based drug designing methods including machine learning, molecular docking, and dynamics study. Implementing the aforementioned approaches, two of the FDA-approved drugs raloxifene and bazedoxifene are predicted to be potential selective ALDH1A1 inhibitors. This type of observation was also validated by an in-vitro enzymatic assay in which both the drugs manifested IC_{50} values of 2.35 ± 0.45 μM and 4.41 ± 0.50 μM, respectively against ALDH1A1 with no inhibition against ALDH2 and ALDH3A1 at 100 μM. The cell line assay against A549 and Mia-PaCa-2 cells also demonstrated the ability of both these drugs in combination with maf (analog of CP) to reverse the CP resistance. The selectivity of these two drugs toward ALDH1A1 over ALDH2 and ALDH3A1 could be explained based on molecular docking. This computational approach revealed that both Raloxifene and bazedoxifene showed selectivity toward ALDH1A1 and no inhibition at 100 μM toward ALDH2 and ALDH3A1 due to steric clashes with residues, such as Asp458 of ALDH2 and Ile458 of ALDH3A1. These new scaffolds may be a good starting point for designing more ALDH1A1 inhibitors as adjuvant therapy for CP resistance correlated with ALDH1A1 overexpression. In addition, the drugs can take for further evaluation in other cell lines that could overexpress ALDH1A1 and also in-vivo evaluation for stronger evidence about the combinations.

Declaration of competing interest

The authors declare no potential conflicts of interest.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to thank Tripada Healthcare Pvt. Ltd., Ahmedabad, India for providing Disulfiram as a gift sample. We also thank Mr. Shubham Thakur, ICMR-SRF, Department of Pharmaceutical Sciences, Guru Nanak Dev University, Amritsar for helping us while performing in-vitro enzymatic assay and cell line studies. Mr. Gera Narendra and Mr. Baddipadige Raju are highly grateful to the Indian Council of Medical Research (ICMR) New Delhi for providing a senior research fellowship (SRF) with award no. ISRM/12(10)/2019.

Funding

This work was supported by the Indian Council of Medical Research (ICMR), New Delhi; Sanction No. ISRM/12(10)/2019. The hardware used in this research project for doing *in-silico* calculations has financed by the Department of Biotechnology (DBT), New Delhi, Sanction No: [BT/PR39876/BTIS/137/7/2021].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2023.124749.

References

[1] M. Ahlmann, G. Hempel, The effect of cyclophosphamide on the immune system: implications for clinical cancer therapy, Cancer Chemother. Pharmacol. 78 (2016) 661–671.
[2] A.P. Tran, M. Ali Al-Radhawi, I. Kareva, J. Wu, D.J. Waxman, E.D. Sontag, Delicate balances in cancer chemotherapy: modeling immune recruitment and emergence of systemic drug resistance, Front. Immunol. (2020) 1376.
[3] X. Wang, H. Zhang, X. Chen, Drug resistance and combating drug resistance in cancer, Cancer Drug Resist. 2 (2019) 141.
[4] M. Kartal-Yandim, A. Adan-Gokbulut, Y. Baran, Molecular mechanisms of drug resistance and its reversal in cancer, Crit. Rev. Biotechnol. 36 (4) (2016) 716–726.
[5] B. Raju, H. Verma, G. Narendra, B. Sapra, O. Silakari, Multiple machine learning, molecular docking, and ADMET screening approach for identification of selective inhibitors of CYP1B1, J. Biomol. Struct. Dyn. (2021) 1–16.
[6] H. Verma, M. Singh Bahia, S. Choudhary, P. Kumar Singh, O. Silakari, Drug metabolizing enzymes-associated chemo resistance and strategies to overcome it, Drug Metab. Rev. 51 (2) (2019) 196–223.
[7] B. Jackson, C. Brocker, D.C. Thompson, W. Black, K. Vasiliou, D.W. Nebert, V. Vasiliou, Update on the aldehyde dehydrogenase gene (ALDH) superfamily, Hum. Genomics 5 (4) (2011) 1–21.
[8] V. Vasiliou, D.W. Nebert, Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family, Hum. Genomics 2 (2) (2005) 1–6.
[9] S.A. Marchitti, R.A. Deitrich, V. Vasiliou, Neurotoxicity and metabolism of the catecholamine-derived 3, 4-dihydroxyphenylacetaldehyde and 3, 4-dihydroxy-phenylglycolaldehyde: the role of aldehyde dehydrogenase, Pharmacol. Rev. 59 (2) (2007) 125–150.
[10] C. Smith, M. Gasparetto, C. Jordan, D.A. Pollyea, V. Vasiliou, The effects of alcohol and aldehyde dehydrogenases on disorders of hematopoiesis, Biol. Basis Alcohol Induc. Cancer (2015) 349–359.
[11] W. Wang, C. Wang, H. Xu, Y. Gao, Aldehyde dehydrogenase, liver disease and cancer, Int. J. Biol. Sci. 16 (6) (2020) 921.
[12] M. Magni, S. Shammah, R. Schiró, W. Mellado, R. Dalla-Favera, A.M. Gianni, Induction of Cyclophosphamide-resistance by Aldehyde-dehydrogenase Gene Transfer, 1996.
[13] J. Zhang, Q. Tian, S. Yung Chan, S. Chuen Li, S. Zhou, W. Duan, Y.-Z. Zhu, Metabolism and transport of oxazaphosphorines and the clinical implications, Drug Metab. Rev. 37 (4) (2005) 611–703.
[14] B. Raju, S. Choudhary, G. Narendra, H. Verma, O. Silakari, Molecular modeling approaches to address drug-metabolizing enzymes (DMEs) mediated chemoresistance: a review, Drug Metab. Rev. 53 (1) (2021) 45–75.
[15] X. He, Y. Deng, W. Yue, Investigating critical genes and gene interaction networks that mediate cyclophosphamide sensitivity in chronic myelogenous leukemia, Mol. Med. Rep. 16 (1) (2017) 523–532.
[16] H. Verma, O. Silakari, Investigating the role of missense SNPs on ALDH 1A1 mediated pharmacokinetic resistance to cyclophosphamide, Comput. Biol. Med. 125 (2020), 103979.
[17] G. Narendra, B. Raju, H. Verma, O. Silakari, Identification of potential genes associated with ALDH1A1 overexpression and cyclophosphamide resistance in chronic myelogenous leukemia using network analysis, Med. Oncol. 38 (10) (2021) 1–10.
[18] G.G. Muramoto, J.L. Russell, R. Safi, A.B. Salter, H.A. Himburg, P. Daher, S. K. Meadows, P. Doan, R.W. Storms, N.J. Chao, Inhibition of aldehyde dehydrogenase expands hematopoietic stem cells with radioprotective capacity, Stem Cells 28 (3) (2010) 523–534.
[19] S.-M. Yang, A. Yasgar, B. Miller, M. Lal-Nag, K. Brimacombe, X. Hu, H. Sun, A. Wang, X. Xu, K. Nguyen, Discovery of NCT-501, a potent and selective theophylline-based inhibitor of aldehyde dehydrogenase 1A1 (ALDH1A1), J. Med. Chem. 58 (15) (2015) 5967–5978.
[20] A.C. Kimble-Hill, B. Parajuli, C.-H. Chen, D. Mochly-Rosen, T.D. Hurley, Development of selective inhibitors for aldehyde dehydrogenases based on substituted indole-2, 3-diones, J. Med. Chem. 57 (3) (2014) 714–722.
[21] C.D. Buchman, K.K. Mahalingan, T.D. Hurley, Discovery of a series of aromatic lactones as ALDH1/2-directed inhibitors, Chem. Biol. Interact. 234 (2015) 38–44.
[22] C.A. Morgan, T.D. Hurley, Characterization of two distinct structural classes of selective aldehyde dehydrogenase 1A1 inhibitors, J. Med. Chem. 58 (4) (2015) 1964–1975.
[23] S.-M. Yang, N.J. Martinez, A. Yasgar, C. Danchik, C. Johansson, Y. Wang, B. Baljinnyam, A.Q. Wang, X. Xu, P. Shah, Discovery of orally bioavailable, quinoline-based aldehyde dehydrogenase 1A1 (ALDH1A1) inhibitors with potent cellular activity, J. Med. Chem. 61 (11) (2018) 4883–4903.
[24] C. Anorma, J. Hedhli, T.E. Bearrood, N.W. Pino, S.H. Gardner, H. Inaba, P. Zhang, Y. Li, D. Feng, S.E. Dibrell, Surveillance of cancer stem cell plasticity using an isoform-selective fluorescent probe for aldehyde dehydrogenase 1A1, ACS Cent. Sci. 4 (8) (2018) 1045–1055.
[25] L. Sleire, H.E. Førde, I.A. Netland, L. Leiss, B.S. Skeie, P.Ø. Enger, Drug repurposing in cancer, Pharmacol. Res. 124 (2017) 74–91.
[26] N. Kumar, A. Gahlawat, R.N. Kumar, Y.P. Singh, G. Modi, P. Garg, Drug repurposing for Alzheimer's disease: in silico and in vitro investigation of FDA-approved drugs as acetylcholinesterase inhibitors, J. Biomol. Struct. Dyn. (2020) 1–15.
[27] C.G. Begley, M. Ashton, J. Baell, M. Bettess, M.P. Brown, B. Carter, W.N. Charman, C. Davis, S. Fisher, I. Frazer, Drug repurposing: misconceptions, challenges, and opportunities for academic researchers, Sci. Transl. Med. 13 (612) (2021), eabd5524.
[28] G. Narendra, B. Raju, H. Verma, B. Sapra, O. Silakari, Multiple machine learning models combined with virtual screening and molecular docking to identify selective human ALDH1A1 inhibitors, J. Mol. Graph. Model. 107 (2021), 107950.
[29] J.J. Irwin, B.K. Shoichet, ZINC– a free database of commercially available compounds for virtual screening, J. Chem. Inf. Model. 45 (1) (2005) 177–182.
[30] C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (7) (2011) 1466–1474.
[31] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (1) (2000) 235–242.
[32] G. Kiran, L. Karthik, M.S. Devi, P. Sathiyarajeswaran, K. Kanakavalli, K. Kumar, D. R. Kumar, In silico computational screening of kabasura kudineer-official siddha formulation and JACOM against SARS-CoV-2 spike protein, J. Ayurveda Integr. Med. 13 (1) (2020) 100324.
[33] C.J. Woods, M. Malaisree, S. Hannongbua, A.J. Mulholland, A water-swap reaction coordinate for the calculation of absolute protein–ligand binding free energies, J. Chem. Phys. 134 (5) (2011), 02B611.
[34] C.J. Woods, M. Malaisree, J. Michel, B. Long, S. McIntosh-Smith, A.J. Mulholland, Rapid decomposition and visualisation of protein–ligand binding free energies by residue and by water, Faraday Discuss. 169 (2014) 477–499.
[35] A. Yasgar, S.A. Titus, Y. Wang, C. Danchik, S.-M. Yang, V. Vasiliou, A. Jadhav, D. J. Maloney, A. Simeonov, N.J. Martinez, A high-content assay enables the automated screening and identification of small molecules with specific ALDH1A1-inhibitory activity, PLoS One 12 (1) (2017), e0170937.
[36] B. Parajuli, M.L. Fishel, T.D. Hurley, Selective ALDH3A1 inhibition by benzimidazole analogues increase mafosfamide sensitivity in cancer cells, J. Med. Chem. 57 (2) (2014) 449–461.
[37] B. Li, K. Yang, D. Liang, C. Jiang, Z. Ma, Discovery and development of selective aldehyde dehydrogenase 1A1 (ALDH1A1) inhibitors, Eur. J. Med. Chem. 209 (2021), 112940.
[38] H. Verma, G. Narendra, B. Raju, M. Kumar, S.K. Jain, G.K. Tung, P.K. Singh, O. Silakari, 3D-QSAR and scaffold hopping based designing of benzo [d] ox-azol-2 (3H)-one and 2-oxazolo [4, 5-b] pyridin-2 (3H)-one derivatives as selective aldehyde dehydrogenase 1A1 inhibitors: synthesis and biological evaluation, Arch. Pharm. 355 (9) (2022), e2200108.
[39] R. Brown, R. Herzig, S. Wolff, D. Frei-Lahr, L. Pineiro, B. Bolwell, J. Lowder, E. Harden, K. Hande, G. Herzig, High-dose Etoposide and Cyclophosphamide Without Bone Marrow Transplantation for Resistant Hematologic Malignancy, 1990.
[40] J.H. Pickar, B.S. Komm, Selective estrogen receptor modulators and the combination therapy conjugated estrogens/bazedoxifene: a review of effects on the breast, Post Reprod. Health 21 (3) (2015) 112–121.
[41] B. Ettinger, D.M. Black, B.H. Mitlak, R.K. Knickerbocker, T. Nickelsen, H.K. Genant, C. Christiansen, P.D. Delmas, J.R. Zanchetta, J. Stakkestad, Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial, JAMA 282 (7) (1999) 637–645.
[42] W. Utian, H. Yu, J. Bobula, S. Mirkin, S. Olivier, J.H. Pickar, Bazedoxifene/conjugated estrogens and quality of life in postmenopausal women, Maturitas 63 (4) (2009) 329–335.
[43] C.A. Morgan, T.D. Hurley, Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1, Chem. Biol. Interact. 234 (2015) 29–37.
[44] E. Zafar, M.F. Maqbool, A. Iqbal, A. Maryam, H.A. Shakir, M. Irfan, M. Khan, Y. Li, T. Ma, A comprehensive review on anticancer mechanism of bazedoxifene, Biotechnol. Appl. Biochem. 69 (2) (2022) 767–782.
[45] S. Goldstein, Selective estrogen receptor modulators and bone health, Climacteric 25 (1) (2022) 56–59.
[46] T.-H. Heo, J. Wahler, N. Suh, Potential therapeutic implications of IL-6/IL-6R/gp130-targeting agents in breast cancer, Oncotarget 7 (13) (2016) 15460.
[47] D. Vohora, A. Kalam, A. Leekha, S. Talegaonkar, A.K. Verma, Combined raloxifene and letrozole for breast cancer patients, Arch. Med. Res. 48 (6) (2017) 561–565.
[48] A.L.B. Carneiro, A.P.C. Spadella, F.A.D. Souza, K.B.F. Alves, J.T.D. Araujo-Neto, M. A. Haidar, R.D.C.D.M. Dardes, Effects of raloxifene combined with low-dose conjugated estrogen on the endometrium in menopausal women at high risk for breast cancer, Clinics 76 (2021).
[49] G. Zhong, C.J. Seaman, E.M. Paragas, H. Xi, K.-L. Herpoldt, N.P. King, J.P. Jones, N. Isoherranen, Aldehyde oxidase contributes to all-trans-retinoic acid biosynthesis in human liver, Drug Metab. Dispos. 49 (3) (2021) 202–211.
[50] C. Burkhardt, L. Bühler, M. Tihy, P. Morel, M. Forni, Bazedoxifene as a novel strategy for treatment of pancreatic and gastric adenocarcinoma, Oncotarget 10 (34) (2019) 3198.
[51] R. Zhang, T. Wang, J. Lin, Synergistic effect of bazedoxifene and PARP inhibitor in the treatment of ovarian cancer regardless of BRCA mutation, Anticancer Res. 41 (5) (2021) 2277–2286.
[52] J. Tian, X. Chen, S. Fu, R. Zhang, L. Pan, Y. Cao, X. Wu, H. Xiao, H.-J. Lin, H.-W. Lo, Bazedoxifene is a novel IL-6/GP130 inhibitor for treating triple-negative breast cancer, Breast Cancer Res. Treat. 175 (3) (2019) 553–566.

Journal of Molecular Graphics and Modelling 107 (2021) 107950

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

Multiple machine learning models combined with virtual screening and molecular docking to identify selective human ALDH1A1 inhibitors

Gera Narendra, Baddipadige Raju, Himanshu Verma, Bharti Sapra, Om Silakari *

Molecular Modeling Lab (MML), Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala, Punjab, 147002, India

ARTICLE INFO

Keywords:
ALDH1A1 inhibitors
Drug resistance
Multiple machine learning models
Virtual screening
Docking
ADMET
MD simulations

ABSTRACT

Aldehyde dehydrogenases (ALDHs) are the enzymes of oxidoreductase family that are responsible for the aldehyde metabolism. The unbalanced expression of these enzymes may be associated with a variety of disease conditions including cancers. ALDH1A1 is one of the isoform of ALDHs majorly overexpressed in a variety of tumors and responsible for the anti-cancer drug resistance. This makes ALDH1A1 as a specific target to develop small molecule ALDH1A1 inhibitors for resistant cancer condition. Number of ALDH1A1 inhibitors have been developed and reported in the literature, but because of non-selectivity and inappropriate pharmacokinetic properties till now none of these have reached in the market for clinical use. Therefore, multiple machine learning models of different isoforms of ALDHs are integrated with in-*silico* techniques including virtual screening, docking, ADMET profiling, and MD simulation to identify selective ALDH1A1 inhibitors. Total ten selective ALDH1A1 inhibitors with diverse scaffolds and appropriate ADMET were identified that can be further developed as adjuvant therapy in cyclophosphamide and cisplatin resistance cancer.

1. Introduction

Aldehyde dehydrogenase (ALDH) is a superfamily of phase-I oxidases, which mainly catalyze the $NAD(P)^+$ dependent oxidation of aldehydes (R–C(=O)–H) to acids (R–C(=O)–OH) [1]. Till now 19 functional ALDH genes have been identified within the human genome [2]. These genes are involved in a wide variety of biological processes, including the metabolism of various endogenous and exogenous compounds such as acetaldehyde, retinaldehyde, amino acids, neurotransmitters, carbohydrates, and lipids [3]. The irregular biological activity of ALDHs and their contribution in metabolic pathways have been associated with a variety of diseases, including alcoholic liver diseases, cancers etc. [4–6]. Additionally, several studies have demonstrated that ALDH1A1 is overexpressed in a variety of tumors and is associated with resistance to several anti-cancer drugs, including cyclophosphamide (CPA) and cisplatin [7–9]. Initially, CPA is biotransformed into aldophosphamide followed by formation of phosphoramide mustard (PM) which is an active metabolite of the CPA. But, as described in Fig. 1, the conversion of CPA to PM is shunted at the level of aldophosphamide by converting it into carboxyphosphamide (an inactive metabolite) using ALDH1A1 [10]. Similarly, cisplatin, the standard first-line drug used for treatment of numerous cancers, has also been reported to be resistant in ALDH1A1 overexpressed cancer cell line [11]. In many studies it has been reported that ALDH1A1 knockdown protects both CPA and cisplatin from ALDH1A1 mediated metabolism and is accompanied by a remarkable reduction in chemoresistance [12–14]. Based on these reported evidence, it can be concluded that ALDH1A1 is a promising target to improve the quality of cancer treatment among the patients suffering with ALDH1A1 mediated drug resistance.

Globally, efforts are being made by researchers to develop ALDH1A1 inhibitors and address the issue of resistance. However, some of these reported inhibitors face the problem of selectivity over these isoforms of ALDH for an example *N,N*-diethylaminobenzaldehyde (DEAB), 4-dimethylamino-4-methyl-pent-2-ynthionic acid-S-methylester (DIMATE), and citral are known multi-isoform inhibitors [15–17]. As per the reported *in vivo* studies, isoform-selective ALDH1A1 inhibitors exhibit limited efficacy and poor bioavailability [18,19]. In addition to these, selective inhibitors also exhibit some degree of inhibition towards both ALDH2 and ALDH3A1. Hence there is a need to develop potent, non-toxic ALDH1A1 inhibitors with suitable drug-like properties and selectivity profiles to overcome this breach.

The main objective of the present investigation is to identify selective ALDH1A1 inhibitors with suitable pharmacokinetic (PK) profiles to overcome ALDH1A1 mediated chemoresistance and cancer progression.

* Corresponding author.
E-mail address: omsilakari@gmail.com (O. Silakari).

https://doi.org/10.1016/j.jmgm.2021.107950
Received 13 December 2020; Received in revised form 18 May 2021; Accepted 20 May 2021
Available online 28 May 2021

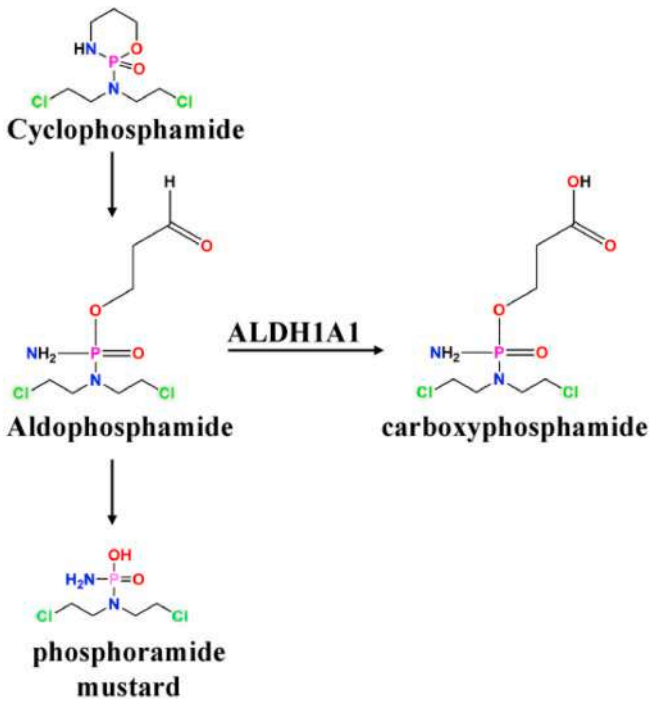

Fig. 1. Schematic representation of CPA inactivation by ALDH1A1.

To accomplish this objective, three different machine learning (ML) classification models have been generated which were used for virtual screening (VS) of three large chemical libraries to identify selective ALDH1A1 inhibitors later. These screened molecules were also analyzed for their interaction patterns with three selected isoforms using docking analysis to justify the selection of multiple machine learning (MML) based screening hits towards ALDH1A1 and further ADMET prediction with suitable bioavailability parameters. To investigate the interaction stability of the selected hits; molecular dynamic (MD) simulation studies were also performed. The potential selective ALDH1A1 inhibitors with good predicted PK profile could be anticipated from the current study.

2. Material and methods

2.1. Data preparation for ML models

To construct the three different classification models, two sets of compounds i.e. inhibitors and non-inhibitors were required. Set of inhibitors of three ALDH isoforms, including ALDH1A1, ALDH2, and ALDH3A1 and their respective biological activities in terms of IC_{50} were collected from various sources like PubChem Bioassay database, ChEMBL database, and literature [18–27]. Non-inhibitors dataset accommodates compounds which were presumed as inactives or decoys. The decoys of each target were generated using the DUD-E database [28]. The collected data was prepared by OpenBable software by eliminating duplicate entries of the compounds on the basis of SMILE consistency. Thereafter the compounds in the dataset were processed using discovery studio 4.1 to add explicit hydrogens. Moreover, the entries containing salts were removed using PaDEL and the final data was further subjected for descriptor calculations.

2.2. Descriptors calculation

During the past years, it was observed that the use of 1D, 2D, and 3D descriptors have presented an improved interpretation of predictive modeling experiments using ML based approaches. In the present study, PaDEL, an open tool was used to calculate total 1875 1D, 2D, and 3D descriptors corresponding to each compound in the datasets [29]. These datasets were subjected to data partitioning into training and test sets in the proportion of 7:3 ratio using create partition function in caret library [30]. The details of the datasets used in model development are given in Table 1. All the descriptors in the training and test sets were normalized between 0 and 1 using MinMaxScalar function. Further, the training dataset was subjected to data preprocessing to remove zero variance and highly correlated features using variance threshold and Pearson correlation coefficient cutoff (0.6), respectively. Finally, the recursive feature elimination (RFE) method was selected to remove the irrelevant features from the data [31]. The obtained features were further used for model generation. The above mentioned protocol helped to improve the model performance by reducing dimensionality of the data. The overall workflow of the dataset processing and descriptor selection are shown in Fig. 2.

2.3. Machine learning model development

ML predictive models corresponding to each of three different ALDH isoforms were developed. The classification models, including Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) were developed using caret library (version 6.0–86) [32] written for the R programming language. R version 3.5.1 is used for the analysis of descriptors and development of ML model.

2.3.1. SVM

SVM is a ML method developed by Vapnik, which has been employed in both types of regression and classification problems [33]. It is a kernel-based supervised learning algorithm which is most suited for the distinct classification of data points based on their biological data. The main aim of this algorithm is to find a maximum marginal hyperplane that best divides the data into classes. The final position and orientation of the selected hyper-plane can be defined as the subset of training vectors i.e. support vectors. This hyper-plane in the multi-dimension space is the decision boundary to classify the given data (in the present study to classify inhibitor or non-inhibitor). SVM model can be determined by various parameters like a kernel function, a kernel coefficient, and a penalty parameter. In this work, the SVM model has been developed using radial basis function (RBF) kernel function (version

Table 1
Detailed information about the data used for the ML models.

Target	Total number of inhibitors and non-inhibitors used		Data divide for ML model			
	Inhibitors	Non-inhibitors	Training set		Test set	
			Inhibitors	Non-inhibitors	Inhibitors	Non-inhibitors
ALDH1A1[a]	1340	1339	938	938	402	401
ALDH2[b]	129	128	91	90	38	38
ALDH3A1[b]	247	168	173	118	74	50

[a] Aldehyde dehydrogenase 1A1 isoform: Drug inactivating enzyme responsible for the resistance of CPA.

[b] Other Isoforms of ALDH.

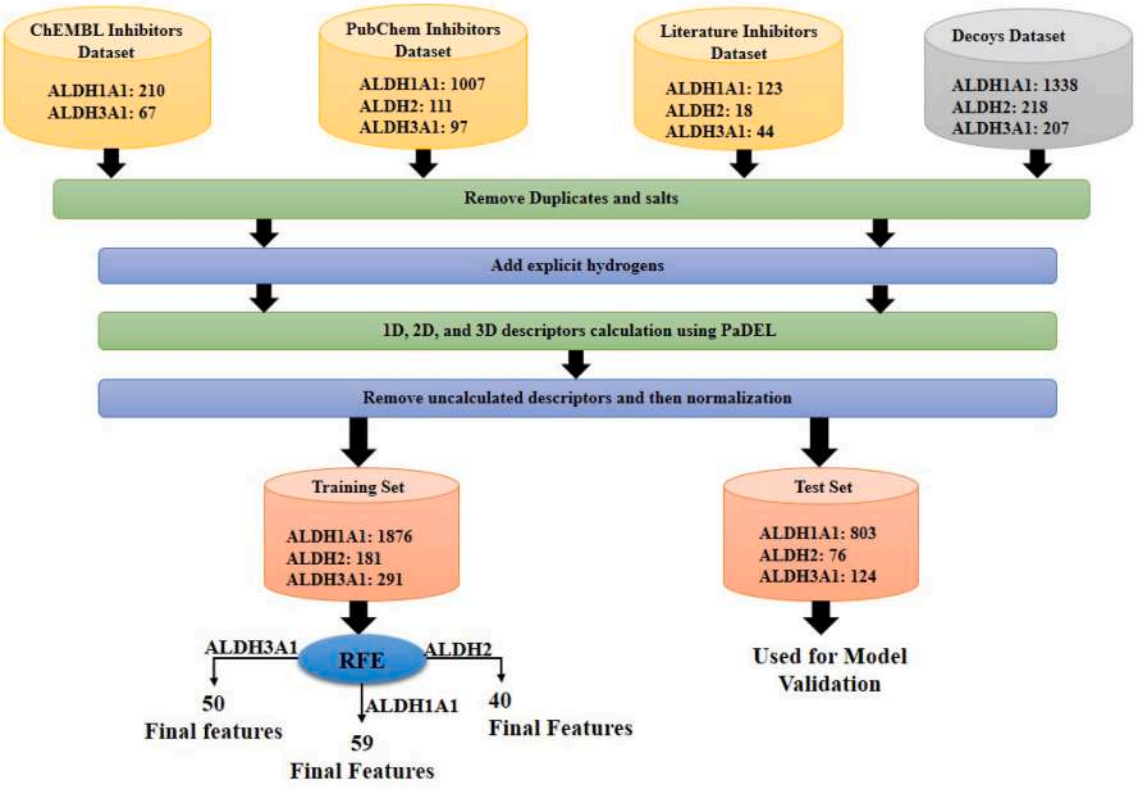

Fig. 2. Overall Workflow of dataset preprocessing and feature selection.

0.9–29) [34,35]. The 10-fold cross validation method was performed in model development and model optimization was performed using the grid program implemented in caret package with cost (C) and sigma (σ) parameters.

2.3.2. Random forest (RF)

RF is a combination of multiple decision trees. It is a rule based statistical model, suitable for cases with low variance. Decision tree based ensemble methods have proven to be an effective ML model. RF is the best and successfully applied ensemble method among many decision tree based ensemble methods with potentially low overfitting problem. RF algorithm generates decision trees on the basis of data samples and then gets the prediction from each of them and finally selects the best one [36,37]. In this study the RF models were developed using the 10-fold cross validation method. Additionally, two adjustable parameters, including number of trees and mtry were set to 500 and 4, respectively. RF version 4.6–14 [38] was used for the model development.

2.3.3. Artificial neural network (ANN)

ANN is the biological inspired technique that can be used widely to process the information [39]. An ANN has hundreds of processing units interconnected with nodes called artificial neurons. These processing units are made up of input and output units. The neural networks attempt to learn about the information taken as input to give one output. The ANN structure consists of basically 3 components, in the form of layers. Input Layer; is the initial raw information that is fed into the network. Hidden Layer; ANN architecture can have one and more hidden layers. Its main job is to process the raw information received from the input layer to the output layer. Output Layer; information received from the hidden layer is fed to the output layer and is processed to produce the desired results. In the present work, logistic activation function was implemented with hypertuning parameters, including size and decay to optimize the ANN model. 10-fold cross validation was performed in ANN model development. NNET package (version 7.3–15) [40] in R-studio was used to build neural network models [41].

2.4. Statistical analysis

All the developed models were evaluated for their performance by calculating various statistical indexes like sensitivity (SE), specificity (SP), precision (P), accuracy (Q), and Matthew's correlation coefficient (MCC). While Q is used to measure the overall classification performance, SE and SP are simply used to correctly identify the inhibitors as inhibitors and non-inhibitors as non-inhibitors. P indicates the actual actives (positives) in the predicted active molecules. Finally, MCC evaluates the model without influencing the effect of imbalanced dataset. equations (1)–(5) are used to calculate the statistical parameters i.e. SE, SP, P, Q, and MCC.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(FP + TN)(FP + TP)(FN + TN)(FN + TP)}} \quad (5)$$

where, true positive (TP) is the number of inhibitors correctly classified as inhibitors, true negative (TN) is the number of non-inhibitors correctly classified as non-inhibitors, false positive (FP) is the number of non-inhibitor classified as inhibitors, and false negative (FN) is the number of inhibitors wrongly classified as non-inhibitors. The area

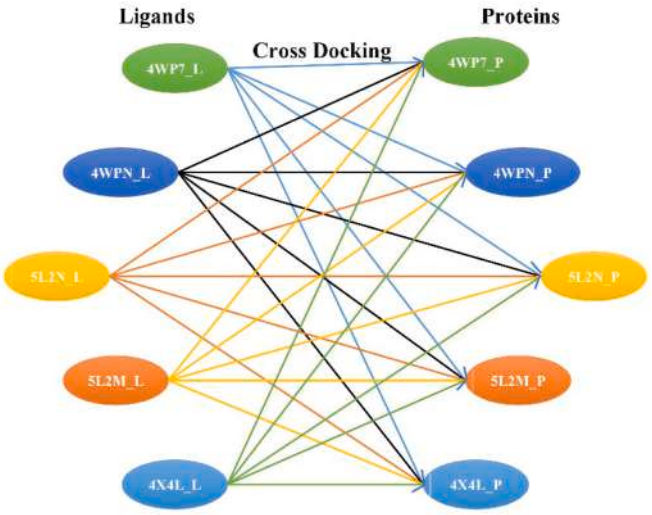

Fig. 3. Brief representation of cross docking (L_ligand, P_Protein).

under the receiver operating characteristic (ROC) curve and Y-randomization analysis were also used to evaluate the classifiers. 10-fold cross validation method was used in model development procedure.

2.5. Virtual screening

To identify new hits as selective ALDH1A1 inhibitors, the best selected ML models (each for ALDH1A1, ALDH2 and ALDH3A1) were used to screen four small molecule databases, including ChemBridge [42], Maybridge [43], Natural Product, and Natural Product-like Compound Library from Life Chemicals database. Initially, the screening was performed using the best selected ALDH1A1 model. Later, the obtained inhibitors were screened through ALDH2 and ALDH3A1 models respectively. Finally, the molecules with the selective ALDH1A1 features were considered for docking and ADMET based screening.

2.6. Molecular docking

2.6.1. Protein preparation

3D structures of ALDH1A1 were collected from the protein data bank (PDB). Total ten 3D structures of this enzyme are available in the protein data bank (PDB; https://www.rcsb.org/) with PDB IDs: 4WJ9, 4WB9, 4WP7, 4WPN, 5L2N, 5L2M, 5L2O, 5AC2, 5TEI, and 6DUM [22–24,44,45]. Among all the 3D structures, structures having resolution more than 2Å (4WB9, 5L2O, 5TEI, and 6DUM), apo form of structure (4WJ9), and the structure co-crystallized with covalent inhibitors (5AC2) are excluded. The remaining five 3D structures (4WP7, 4WPN, 4X4L, 5L2N, and 5L2M) were further considered for cross-docking to select the most appropriate structure of ALDH1A1 for final docking analysis.

2.6.1.1. Cross docking. As the dataset consists of ligands with diverse scaffolds, appropriate structure of protein must be selected so that it can accommodate all molecules of the dataset in cross-docking analysis [46]. Five 3D structures of ALDH1A1 selected in previous step were prepared using Discovery studio 4.1 (DS 4.1) [47] by inserting the missing atoms in the incomplete residues, modeling of missing loops, removing non-active water molecules, and protonation of titratable residues using predicted pKa at a pH value of 7.4. The CHARMm force field was used for preparing these structures. All the prepared crystal structures were aligned and subsequently ligands from each crystal structure were extracted out. Extracted ligands and proteins were saved as general codes as mentioned in Fig. 3 (for protein PDB-ID_P e.g. 4WP7_P and for ligands PDB-ID_L e.g. 4WP7_L). Later, the active site in each protein was specified using define and edit binding site tool. The grid of 10Å radius was ultimately created to redock the ligands. The root mean square deviation (RMSD) between the redocked crystal ligands and their native crystal conformations were computed.

On the basis of cross-docking analysis, the protein structure with low average RMSD of redocked ligands i.e. 4X4L was considered for docking of screened molecules. The number of possible conformations for each ligand was set to 5 and the top scored conformations are selected and analyzed based on -CDOCKER interaction energy.

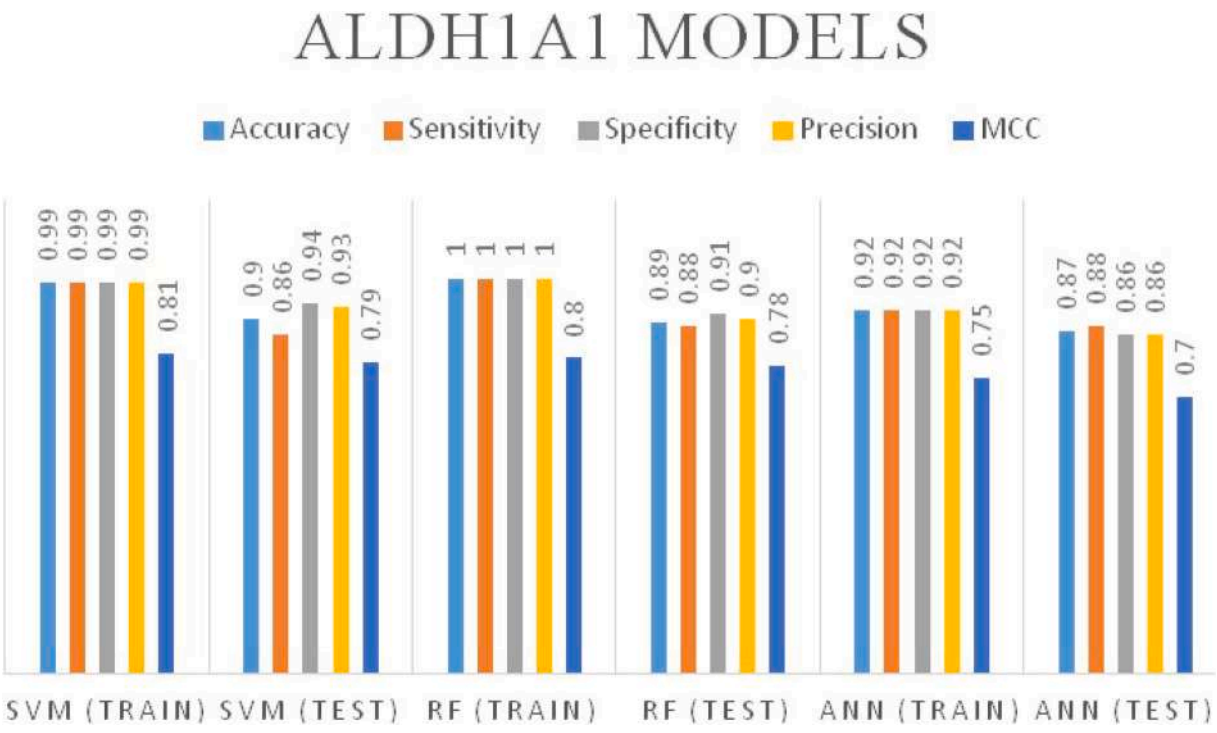

Fig. 4. Performance of ALDH1A1 machine learning models.

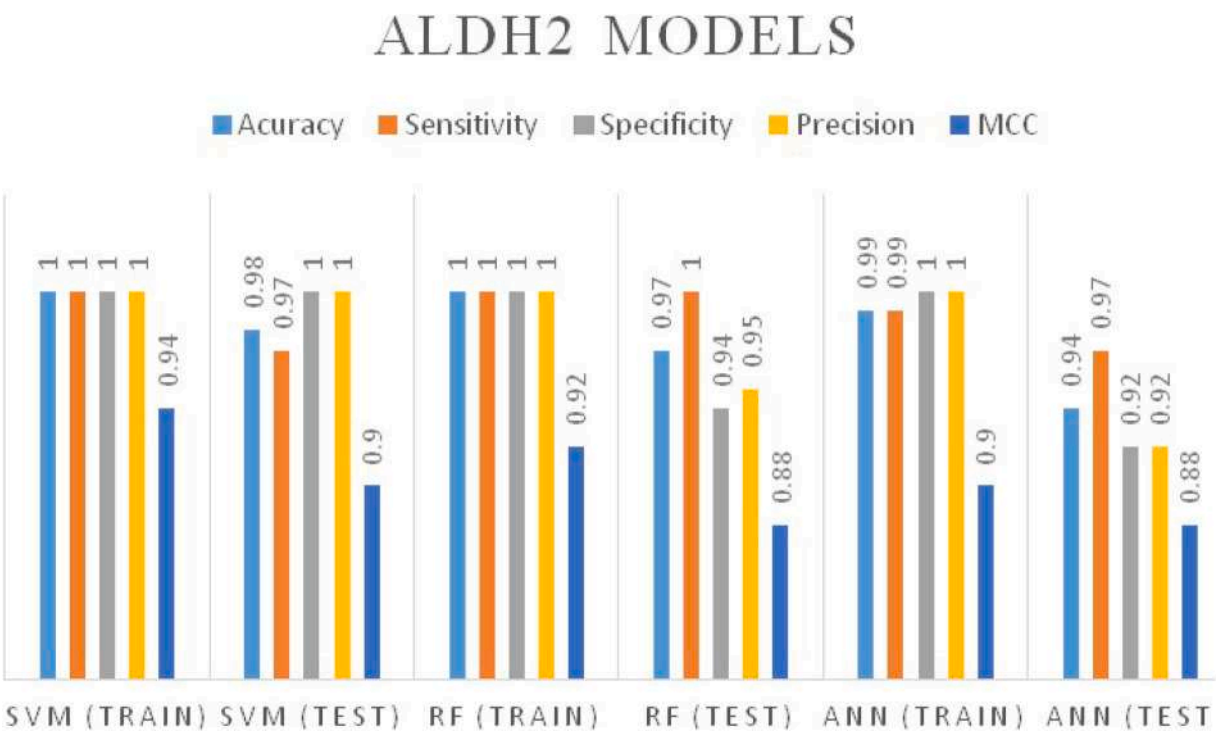

Fig. 5. Performance of ALDH2 machine learning models.

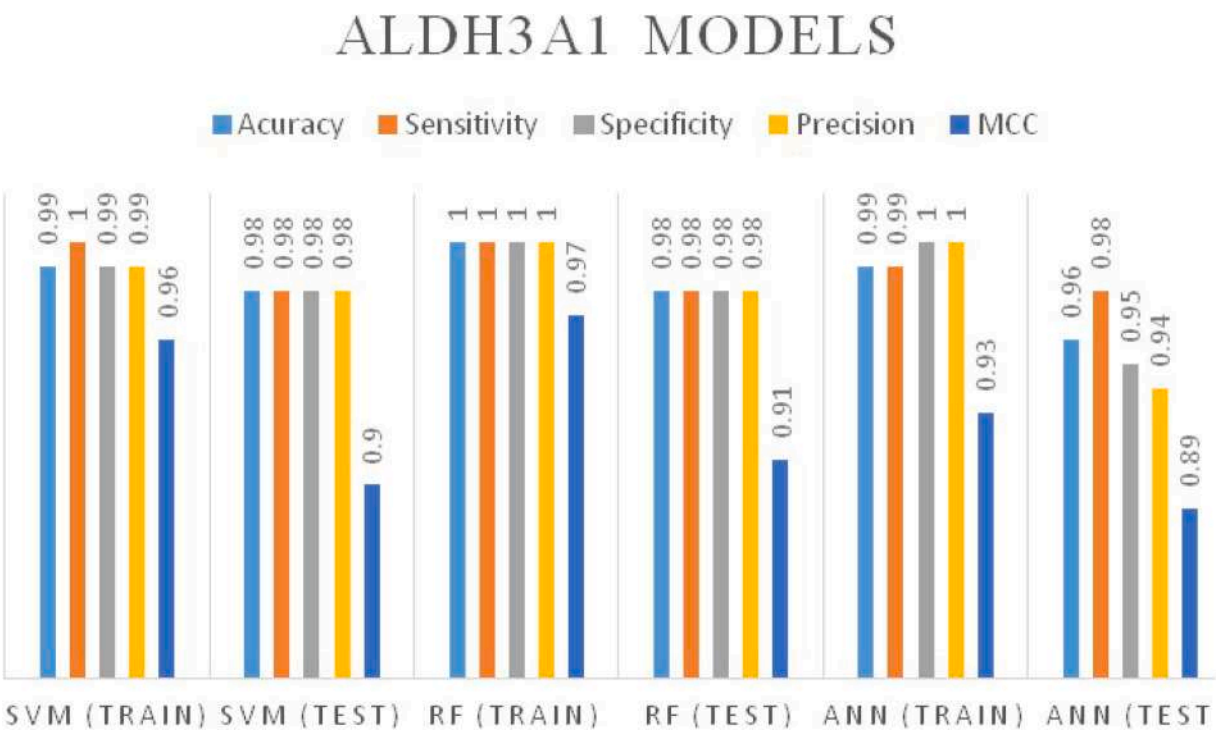

Fig. 6. Performance of ALDH3A1 machine learning models.

2.7. ADME and toxicity prediction

To estimate the drug-likeness of compounds, *in-silico* ADME and toxicity predictions were carried out. SwissADME [48] is the freely available online server used to investigate the various Lipinski rule parameters, including molecular weight (MW), LogP, topological polar surface area (TPSA), number of hydrogen bond acceptors (HBA), and number of hydrogen bond donors (HBD). In addition to these, LogS was calculated to determine the solubility. Besides, PAINS and Brenk alert filtrations were applied to remove compounds with reactive and toxic fragments. In early drug discovery process, toxicity prediction would be of more importance. So, the compounds filtered through ADME were further subjected to toxicity screening with the help of online tool admetSAR [49]. The calculated parameters were compared with standard ranges to check if the values lie within the limit or not.

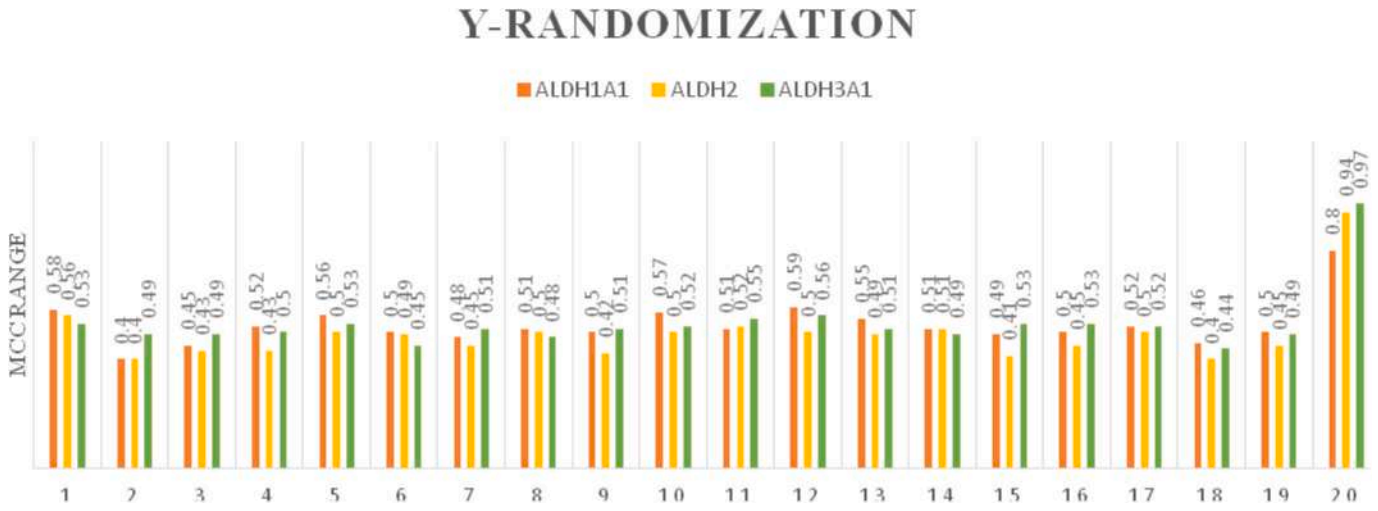

Fig. 7. ALDH models Robustness indicating Plot (1–19 is scrambled data, 20 is original data).

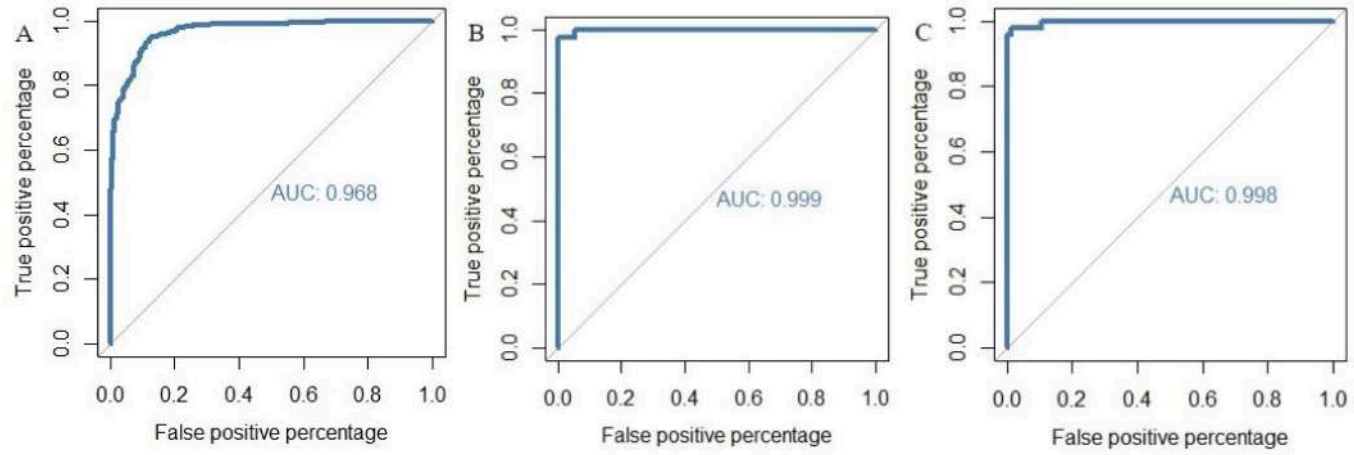

Fig. 8. (A) ROC curve of 10-fold cross validated best selected ALDH1A1 model, (B) ROC curve of 10-fold cross validated best selected ALDH2 model, (C) ROC curve of 10-fold cross validated best selected ALDH3A1 model.

2.8. Molecular dynamics

In order to check the stability of the predicted hits in the active site of ALDH1A1, the docked complexes were subjected to MD simulations for a time period of 30 ns using Desmond program which is based on OPLS-2005 force-field [50]. Initially, the system builder module implemented in Maestro was used to build the system, including protein and inhibitor complex for carrying MD simulation. The protein-ligand complexes were placed in an orthorhombic water box with dimensions $10 \times 10 \times 10$ Å^3 using the TIP3P water model as a solvent system [51]. Following this, at physiological pH, complexes were charged accordingly, the system was neutralized by adding of Na^+ and Cl^- ions at a concentration of 0.15 M. Before proceeding toward the final run, the energy was minimized, the isothermal-isobaric (NPT) ensemble at a constant temperature of 310 K and 1.01325 bar pressure over 30 ns with recording intervals of 1ps, throughout the simulation. Then, the RMSD was calculated to monitor the stability of the protein ligand complex in its native state. Finally, RMSD plots and ligand interaction diagrams were sketched with the help of the Desmond simulation interaction diagram (SID) tool of Maestro.

3. Results and discussion

3.1. Machine learning models

Different ML classification models were developed using SVM, RF, and ANN algorithm. Initially, 1875 molecular descriptors were calculated and normalized using MinMaxScalar method. Further, zero variance and highly correlated features were removed. Ultimately a total number of 179, 125, and 149 features were obtained for ALDH1A1, ALDH2, and ALDH3A1 datasets, respectively. Later, the preprocessed

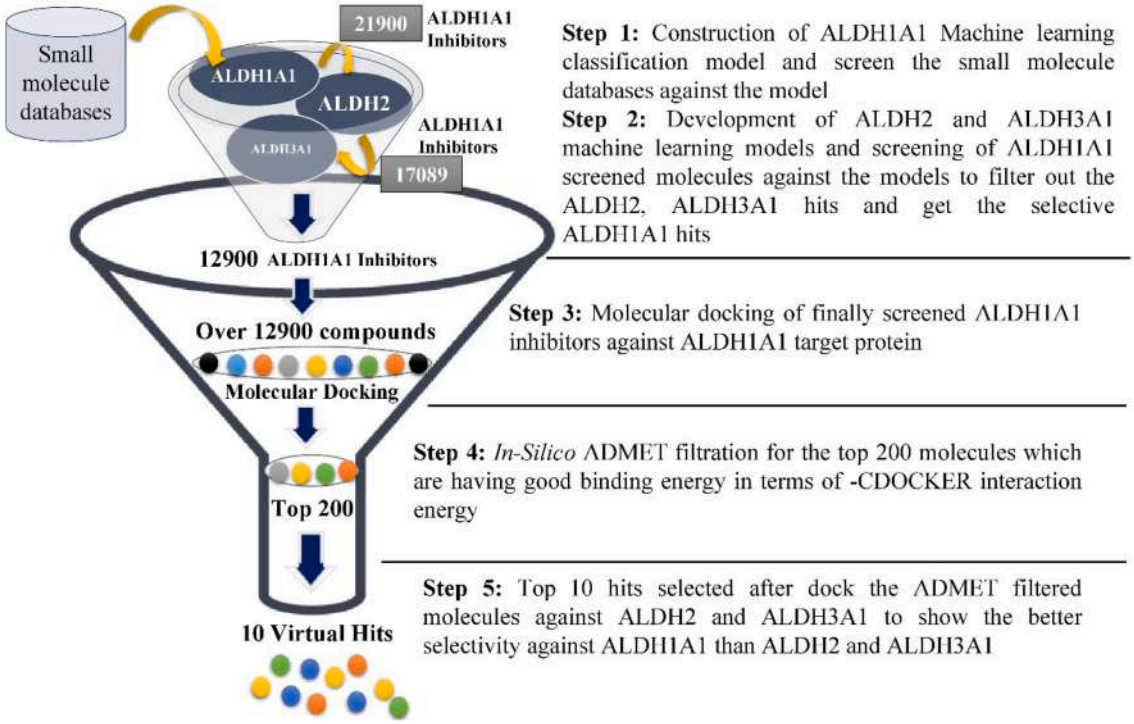

Fig. 9. Flowchart of Virtual screening adopted.

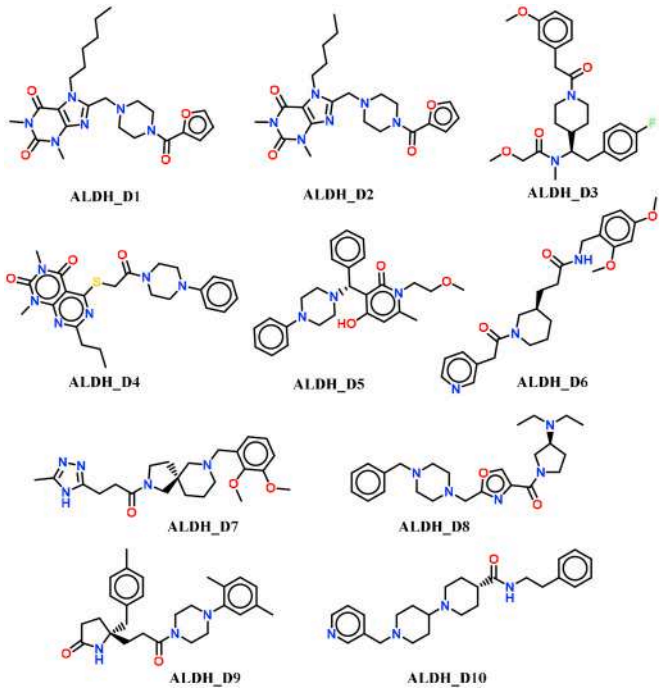

Fig. 10. Top 10 hits identified through MML and in-silico screening.

data were subjected to the RFE feature selection method to identify relevant features. Finally a subset of 59, 40, and 50 features were identified for ALDH1A1, ALDH2, and ALDH3A1, respectively, and the RFE selected features have been displayed in Table 2 (SI). The selected features were further employed in the model development procedure. Different ML models were developed and the performance of the developed models was later displayed in the form of bar charts Figs. 4–6. To evaluate the performance of the developed models, different

Table 3
Cross-docking RMSD analysis results.

PDB ID → ↓	4WP7–P[a]	4WPN-P[a]	4X4L-P[a]	5L2M-P[a]	5L2N–P[a]
4WP7-L[b]	0.45	0.75	0.57	1.84	0.72
4WPN-L[b]	1.17	0.98	1.02	2.07	1.37
4X4L-L[b]	0.98	2.38	0.49	1.14	1.77
5L2M-L[b]	1.06	1.21	0.90	0.63	1.69
5L2N-L[b]	0.77	1.42	0.92	1.11	1.08
Average RMSD	0.88	1.34	**0.78**	1.35	1.32

[a] ALDH1A1 3D-crystal structures used for Cross docking.

[b] Co-crystallized ligands used for cross docking.

statistical parameters were considered, including Q, SE, SP, P, and MCC. All the developed models showed good predictive Q (>0.9) for both training and test sets. Similarly, it was identified that all models show high SE, SP, P scores (>0.9) for both training and test sets. Among the developed models, the significance of the best model was judged by the value of MCC which brings out more reliable and informative score in evaluating binary classification models than accuracy. After analyzing the MCC results obtained for all the models, it was found that the SVM algorithm performs better on dataset of ALDH1A1, ALDH2, whereas, RF algorithms perform better on a dataset of ALDH3A1 which were evaluated by 10-fold cross-validation. The sigma (σ) 0.03 and Cost (C) 2.1 parameters for ALDH1A1 and sigma (σ) 0.03 and Cost (C) 1.3 parameter for ALDH2 were utilized in the best selected SVM model. In addition to this, Y-randomization and ROC tests were performed on best-selected models to identify the predictive robustness and screening efficiency, respectively. In Y- randomization test as described in Fig. 7, each training dataset of best model was scrambled 19 times and models were regenerated by random spreadsheets. None of the random spreadsheet showed better MCC than the original best models. This clearly indicates that each best-developed model is robust in nature and not produced by chance. Furthermore, the curve (ROC) corresponding to the best model

Table 4
ADME properties of top 10 compounds.

Ligands	MW (g/mol)[a]	LogP[b]	LogS[c]	HBA[d]	HBD[e]	TPSA (Å^2)[f]	PAINS[g]	Brenk[h]
ALDH-D1	456.54	1.89	−3.94	6	0	98.51	0	0
ALDH-D2	442.51	1.54	−3.51	6	0	98.51	0	0
ALDH-D3	456.55	3.56	−5.20	5	0	59.08	0	0
ALDH-D4	468.57	2.28	−4.77	5	0	118.63	0	0
ALDH-D5	433.54	3.05	−5.01	4	1	57.94	0	0
ALDH-D6	425.52	2.64	−4.48	5	1	80.76	0	0
ALDH-D7	427.54	2.33	−4.17	6	1	83.58	0	0
ALDH-D8	425.57	2.23	−4.00	6	0	56.06	0	0
ALDH-D9	433.59	3.83	−5.82	2	1	52.65	0	0
ALDH-D10	406.56	2.95	−4.79	4	1	48.47	0	0

[g,h]Number of alerts for undesirable substructures.
[a] Molecular weight.
[b] Lipid-water partition coefficient.
[c] Aqueous solubility.
[d] Hydrogen bond acceptor.
[e] Hydrogen bond donor.
[f] Topological polar surface area.

Table 5
Predicted toxicity parameters of top 10 compounds.

Molecule number	AM[a]	A-O-T[b]	AT[c]	CT[d]	HT[e]
ALDH-D1	–	III	–	–	–
ALDH-D2	–	III	–	–	–
ALDH-D3	–	III	–	–	–
ALDH-D4	–	III	–	–	–
ALDH-D5	–	III	–	–	–
ALDH-D6	–	III	–	–	–
ALDH-D7	–	III	–	–	–
ALDH-D8	–	III	–	–	–
ALDH-D9	–	III	–	–	–
ALDH-D10	–	III	–	–	–

[a] Ames mutagenesis.
[b] Acute oral toxicity (III indicates slightly toxic and slightly irritating).
[c] Avian toxicity.
[d] Cardiotoxicity.
[e] Hepatotoxicity.

of each target was plotted. The area under the curve (AUC) in the ROC was determined and displayed in Fig. 8. All the selected best models showing the AUC values range from 96% to 100% indicates that all the developed models are capable enough to discriminate the inhibitors from non-inhibitors. The developed models were further implemented for VS against four large databases i. e. ChemBridge, Maybridge, Natural Product, and Natural Product-like Compound Library to scrutinize selective ALDH1A1 inhibitors.

3.2. Virtual screening

MML based VS protocol in combination with docking and ADMET filters to identify selective ALDH1A1 inhibitors with drug-like property and minimum toxicity is displayed in Fig. 9.

In order to check the performance of the selected SVM and RF models, three commercially available databases have been screened. A total number of 21900 hits were identified as ALDH1A1inhibitors from SVM model of ALDH1A1 using VS. These hits were then further subjected to screening by imposing filter of ALDH2 and ALDH3A1 models to filter out inhibitors of these two isoforms. Finally, 12900 selective inhibitors of ALDH1A1 were identified. These identified selective ALDH1A1 hits were then further considered for docking and ADMET studies.

3.3. Selection and preparation of protein structures for docking analysis

The most appropriate structure was selected on the basis of cross-docking analysis. In this cross docking experiment, average RMSD of all redocked co-crystallized ligands in each protein was determined to assess suitability of each crystal protein. The protein structure, i.e. 4X4L having lowest average RMSD of 0.78 Å (bold value in Table 3) is the most appropriate protein structure that could accommodate structurally diverse ligands in its active site and therefore, it was selected for further docking analysis. The cross-docking results for all the 3D structures were mentioned in Table 3.

3.4. Docking and ADMET analysis

Total 12900 ALDH1A1 selective inhibitors, which are retrieved from MML, were docked into the active site of ALDH1A1 protein (4X4L). Among these, 200 molecules, which show -CDOCKER interaction energy scores better than the co-crystallized ligand, were further selected for ADMET analysis.

Top ranked ten drug-like molecules displayed in Fig. 10 with good -CDOCKER interaction energy were filtered out on the basis of ADMET parameters. These 10 molecules satisfy all the Lipinski parameters, including MW < 500, LogP <5, HBA <10, HBD <5, and TPSA 20–120 Å^2 and therefore, would be considered probably oral bioavailable. LogS, a common parameter for measuring aqueous solubility was estimated by the FILTER-IT program in SwissADME. LogS of all ten compounds were found to be within the reference range i.e. <6. Besides, all the molecules possess no PAINS and Brenk violations. PAINS and Brenks are filters for the structural fragments in compounds to be putatively toxic, metabolically unstable, and chemically reactive to interfere biological assays. By applying physicochemical filtration, the compounds with PAINS and Brenk violations were removed. In toxicity prediction, all the compounds were found to be free from Ames mutagenesis, acute oral

Table 6
Molecular docking scores of ADMET screened molecules against ALDH1A1, ALDH2, and ALDH3A1 isoforms.

Molecule Number	-CDOCKER interaction energy[a]	-CDOCKER interaction energy[b]	-CDOCKER interaction energy[c]
ALDH-D1	53.85	−12.39	30.93
ALDH-D2	53.04	14.24	–
ALDH-D3	53.22	–	38.89
ALDH-D4	52.00	35.09	–
ALDH-D5	51.74	–	−35.73
ALDH-D6	50.67	−45.98	–
ALDH-D7	50.29	25.21	–
ALDH-D8	50.00	31.24	–
ALDH-D9	49.77	37.95	–
ALDH-D10	49.52	31.86	21.55

[a] CDOCKER interaction energy of ALDH1A1 in terms of kcal/mol.

[b] CDOCKER interaction energy of ALDH2 in terms of kcal/mol.

[c] CDOCKER interaction energy of ALDH3A1 in terms of kcal/mol.

toxicity, avian toxicity, cardiotoxicity, and hepatotoxicity as these compounds showed negative values for these toxicities. Both ADME and toxicity results for the top ten compounds are given in Tables 4 and **5**, respectively. These ten molecules obtained through ADMET filters were also docked against ALDH2 and ALDH3A1 to determine selectivity towards ALDH1A1 over other two isoforms in terms of -CDOCKER interaction energy. Comparative docking results of all three ALDH isoforms have been displayed in Table 6. All the ten molecules showed good docking scores towards ALDH1A1 while displayed compromised docking scores towards ALDH2 and ALDH3A1. Docking results indicate that MML model based prediction was accurate and reliable. The binding conformations of all justified hits in the active site of ALDH1A1 crystal structure have been displayed in Fig. 11. As in Figs. 11A–J the molecules displayed important interactions with active site amino acids, including Y297, F171, W178, and C302. In addition to these, some of the molecules also show interactions with S121 and K128 amino acids. In docking analysis top two theophylline based hits were obtained as selective ALDH1A1 inhibitors. Similar kind of observations was recorded in the study by Yang et al. in which they designed, synthesized and evaluated some selective theophylline derivatives having greater inhibitory activity towards ALDH1A1 than ALDH2 and ALDH3A1. This experimental study supports the observed interactions in our finally screened molecules that contain features important for selectivity toward ALDH1A1 inhibition.

3.5. Molecular dynamics

The MD simulations were performed for top four hits; ALDH-D1, ALDH-D3, ALDH-D4, and ALDH-D5 with the highest predictive -CDOCKER interaction energy of 53.85, 53.22, 52.00, and 51.74 kcal/mol, respectively. RMSD plots of all docked complexes were plotted for 30 ns, and were analyzed in order to check the stability and newly formed interactions. From the RMSD plots as displayed in Fig.12 A-1, B-1, C-1 and D-1, it was observed that each complex is stable as minimum RMSD fluctuations are being observed during the entire simulation period. Although the complexes were stable, changes in ligand RMSD (1.5 Å, 2.8Å, 1.5 Å and 2.2 Å for ALDH-D1, ALDH-D3, ALDH-D4, ALDH-D5 respectively) are more prominent than the changes observed for protein RMSD (1.3 Å, 1.4 Å, 1.0 Å, and 1.8 Å for ALDH1A1 with compounds ALDH-D1, ALDH-D3, ALDH-D4, and ALDH-D5 respectively) during MD simulation of each complexes. This clearly indicates that conformational changes in ligand are more than conformational changes in protein and some new interactions between ligand proteins were also observed after MD simulation.

Among these, the RMSD plot of ALDH-D1 protein complex (Fig. 12 A-1) shows high fluctuation of 2 Å in protein RMSD without deviating ligand RMSD during the time interval of 13ns–23 ns of trajectory. However, fluctuation in protein again regain its stability to the acceptable range i.e. 1 Å after 23 ns which remain consistent till the end of the simulation. An interesting observation was also recorded in this simulation period that two new interactions i.e. water mediated interaction with Y457 and a H-bond interaction with C302 (2.3 Å distance) are appeared along with initial interactions with Y297, and W178. After 23 ns, interaction with Y457 disappeared whereas, H-bond interaction with C302 was maintained during the entire simulation (Fig. 12 A-2). The RMSD plot in Fig. 12 B-1 showed simulation of complex ALDH-D3-ALDH1A1 and indicated that there is high fluctuation in conformation of ligand (RMSD value vary within 2.8 Å) which is highest among all four ligands. The high degree of fluctuation in ligand with respect to small fluctuation in protein results into entirely new H-bond interaction between N170 and oxygen of terminal OCH_3 of ALDH-D3 in active site at 2.2 Å distance (Fig. 12 B-2). In case of ALDH-D4 and ALDH-D5, ligand RMSDs fluctuated within 2 Å and their initial contacts were maintained throughout the simulation. In case of ligand RMSD of compound ALDH-D4, notable divergence in RMSD of ligand at the end of the simulation was observed (Fig. 12 C-1). This divergence resulted in the formation of two important hydrogen bond interactions i.e. one between key residues V460 and oxygen of carbonyl of linker in ligand at a distance of 1.9 Å and second between Y457 and oxygen of one of the two carbonyl of the ring at a distance of 2.3 Å (Fig. 12 C-2). Compound ALDH-D5 also showed important interactions with W178 and F466 (Fig. 12 D-2). In this entire analysis, the protein fluctuation lies between the permissible range of 1Å -3Å and ligand fluctuation are also not having larger deviations as compared to protein RMSD.

4. Conclusion

ALDH1A1 is a potential target for combating resistance to CPA and its analogues. Inspite of the presence of several crystal structures of the protein, none of the selective ALDH1A1 could be effectively designed for seeking the approval for reversing resistance. In order to design selective ALDH1A1 inhibitors over other isoforms i.e. ALDH2 and ALDH3A1, a ligand based ML classification models were developed. These models were validated using standard procedures, including Y-randomization and ROC curve analysis. The screened selective ALDH1A1 inhibitors were also analyzed for retention of key interactions with ALDH1A1 and also for compromised docking scores with other isoforms. Since PK and toxicity are major issues with some of the existing ALDH1A1 inhibitors, the ADMET parameters were also calculated to ensure drug likeness of screened molecules, their orally bioavailability and toxicity. All these *in-silico* findings suggested that the identified compounds would serve as novel scaffolds of ALDH1A1 inhibitors. These new scaffolds may be the best starting point for designing more ALDH1A1 inhibitors as adjuvant therapy for CPA resistance and treatment of other diseases correlated with ALDH1A1 overexpression.

Funding

This work was supported by the Indian Council of Medical Research (ICMR), New Delhi; Sanction No. ISRM/12(10)/2019.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2021.107950.

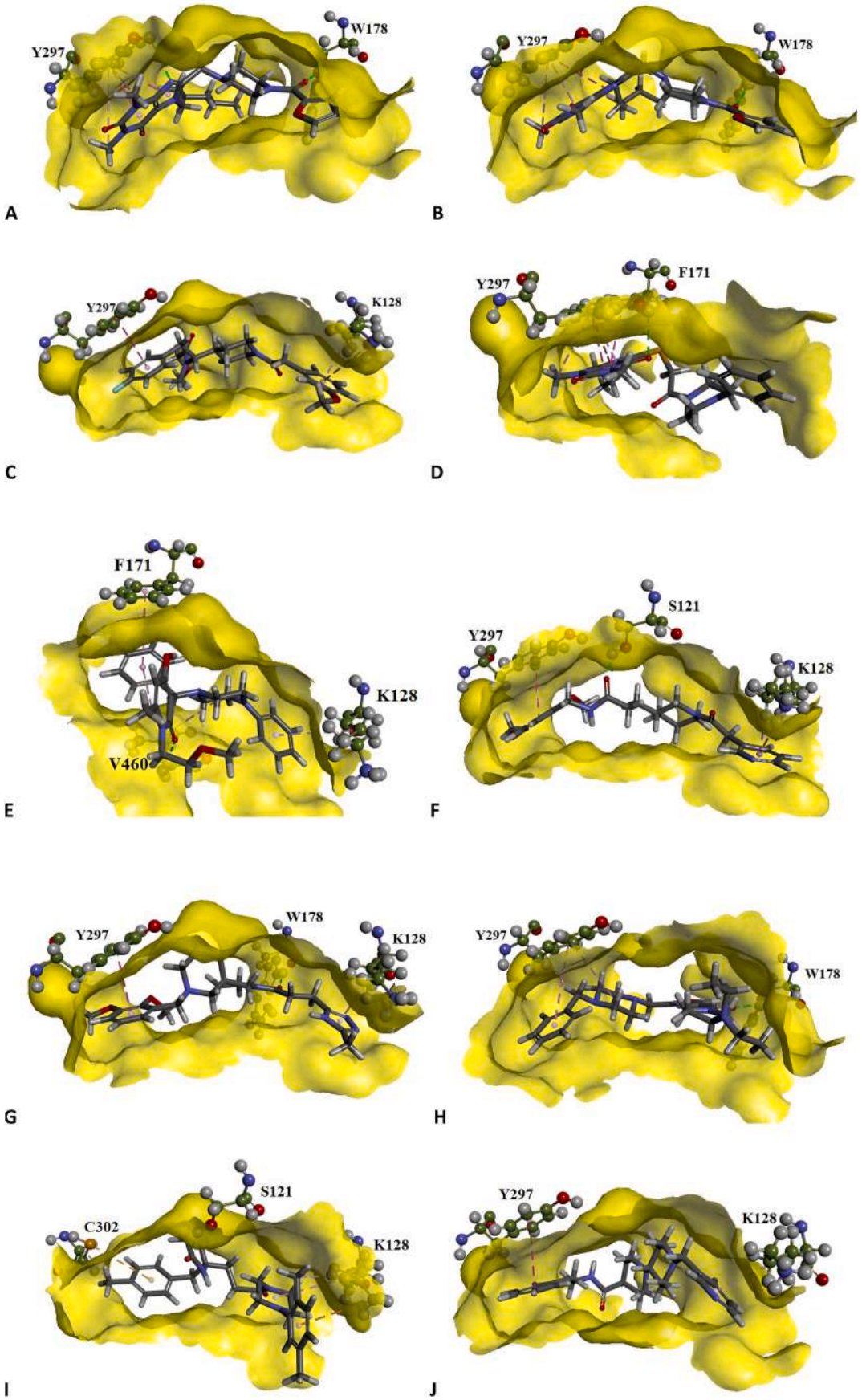

Fig. 11. Docking orientation and interactions of **A)** ALDH-D1 **B)** ALDH-D2 **C)** ALDH-D3 **D)** ALDH-D4 **E)** ALDH-D5 **F)** ALDH-D6 **G)** ALDH-D7 **H)** ALDH-D8 **I)** ALDH-D9 **J)** ALDH-D10 within the ALDH1A1 active site.

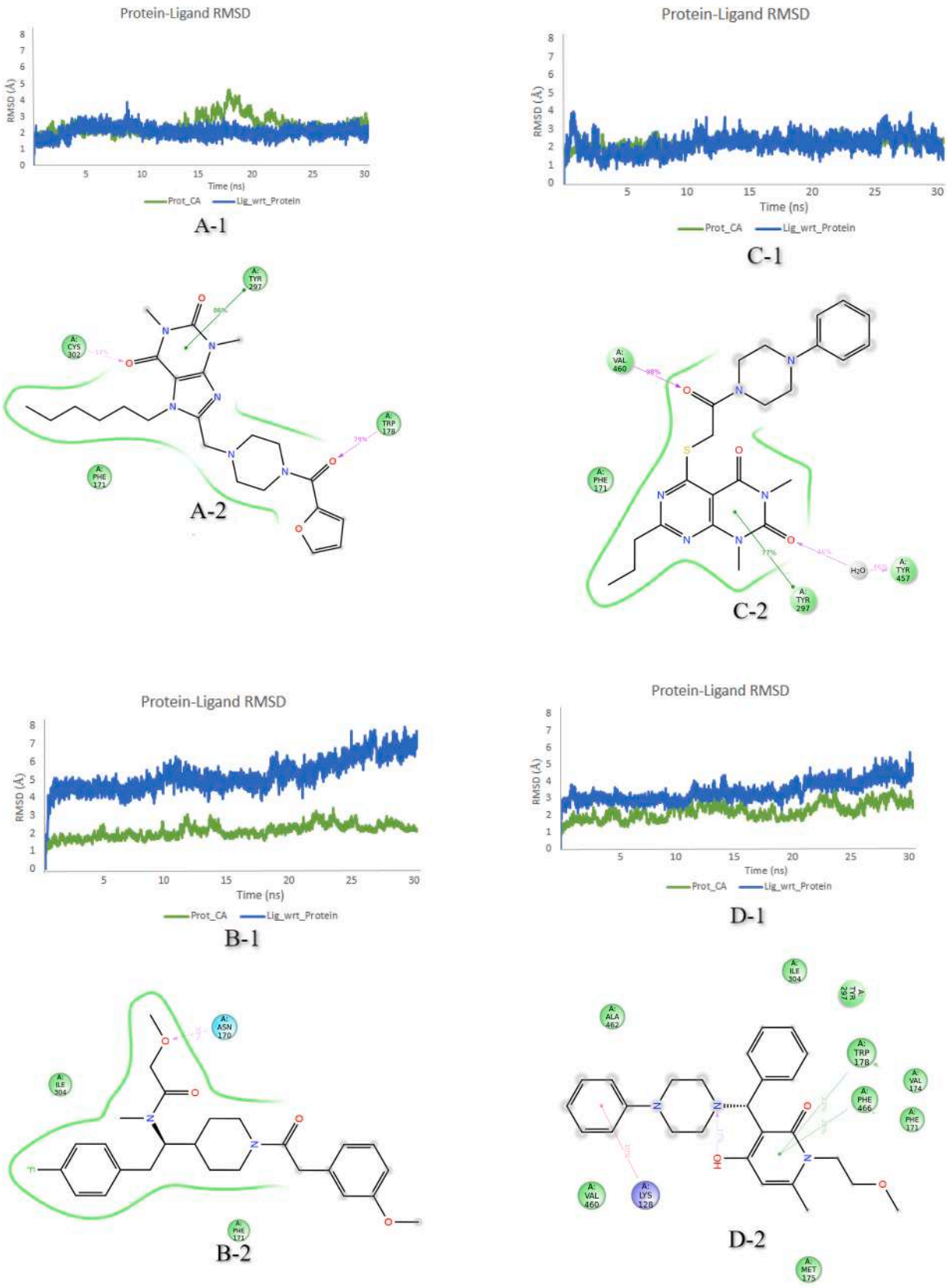

Fig. 12. RMSD analysis of MD simulation trajectory and analysis of molecular interactions of contact with ALDH1A1 after MD Simulation. **A-1)** RMSD of ALDH-D1 **A-2)** Interactions of ALDH-D1 **B-1)** RMSD OF ALDH-D3 **B-2)** Interactions of ALDH-D3 **C-1)** RMSD OF ALDH-D4 **C-2)** Interactions of ALDH-D4 **D-1)** RMSD OF ALDH-D5 **D-2)** Interactions of ALDH-D5.

References

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

[2] S. Gui, X. Xie, W.Q. O'Neill, K. Chatfield-Reed, J.-G. Yu, T.N. Teknos, et al., p53 functional states are associated with distinct aldehyde dehydrogenase transcriptomic signatures, Sci. Rep. (2020) 10.

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

[4] C. Smith, M. Gasparetto, C. Jordan, D.A. Pollyea, V. Vasiliou, The effects of alcohol and aldehyde dehydrogenases on disorders of hematopoiesis, in: Biological Basis of Alcohol-Induced Cancer, Springer, 2015, pp. 349–359.

[5] S.A. Marchitti, R.A. Deitrich, V. Vasiliou, Neurotoxicity and metabolism of the catecholamine-derived 3, 4-dihydroxyphenylacetaldehyde and 3, 4-dihydroxy-phenylglycolaldehyde: the role of aldehyde dehydrogenase, Pharmacol. Rev. 59 (2007) 125–150.

[6] W. Wang, C. Wang, H. Xu, Y. Gao, Aldehyde dehydrogenase, liver disease and cancer, Int. J. Biol. Sci. 16 (2020) 921.

[7] T. Khoury, F.O. Ademuyiwa, R. Chandraseekhar, M. Jabbour, A. DeLeo, S. Ferrone, et al., Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to neoadjuvant chemotherapy, Mod. Pathol. 25 (2012) 388–397.

[8] M. Magni, S. Shammah, R. Schiro, W. Mellado, R. Dalla-Favera, A.M. Gianni, Induction of Cyclophosphamide-Resistance by Aldehyde-Dehydrogenase Gene Transfer, 1996.

[9] E. Meng, A. Mitra, K. Tripathi, M.A. Finan, J. Scalici, S. McClellan, et al., ALDH1A1 maintains ovarian cancer stem cell-like properties by altered regulation of cell cycle checkpoint and DNA repair network signaling, PloS One 9 (2014), e107142.

[10] H. Verma, M. Singh Bahia, S. Choudhary, P. Kumar Singh, O. Silakari, Drug metabolizing enzymes-associated chemo resistance and strategies to overcome it, Drug Metab. Rev. 51 (2019) 196–223.

[11] M.H. Uddin, B. Kim, U. Cho, A.S. Azmi, Y.S. Song, Association of ALDH1A1-NEK-2 axis in cisplatin resistance in ovarian cancer cells, Heliyon 6 (2020), e05442.

[12] Y. Wei, S. Wu, W. Xu, Y. Liang, Y. Li, W. Zhao, et al., Depleted aldehyde dehydrogenase 1 A 1 (ALDH1A1) reverses cisplatin resistance of human lung adenocarcinoma cell A 549/DDP, Thorac. cancer. 8 (2017) 26–32.

[13] L.M. van der Waals, I.H. Borel Rinkes, O. Kranenburg, ALDH1A1 expression is associated with poor differentiation,'right-sidedness' and poor survival in human colorectal cancer, PloS One 13 (2018), e0205536.

[14] J.S. Moreb, A. Gabr, G.R. Vartikar, S. Gowda, J.R. Zucali, D. Mohuczy, Retinoic acid down-regulates aldehyde dehydrogenase and increases cytotoxicity of 4-hydroperoxycyclophosphamide and acetaldehyde, J. Pharmacol. Exp. Therapeut. 312 (2005) 339–345.

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

[16] N. Matsunaga, T. Ogino, Y. Hara, T. Tanaka, S. Koyanagi, S. Ohdo, Optimized dosing schedule based on circadian dynamics of mouse breast cancer stem cells improves the antitumor effects of aldehyde dehydrogenase inhibitor, Cancer res 78 (2018) 3698–3708.

[17] S. Zeng, A. Kapur, M.S. Patankar, M.P. Xiong, Formulation, characterization, and antitumor properties of trans-and cis-citral in the 4T1 breast cancer xenograft mouse model, Pharm. Res. (N. Y.) 32 (2015) 2548–2558.

[18] S.-M. Yang, N.J. Martinez, A. Yasgar, C. Danchik, C. Johansson, Y. Wang, et al., Discovery of orally bioavailable, quinoline-based aldehyde dehydrogenase 1A1 (ALDH1A1) inhibitors with potent cellular activity, J. Med. Chem. 61 (2018) 4883–4903.

[19] S.-M. Yang, A. Yasgar, B. Miller, M. Lal-Nag, K. Brimacombe, X. Hu, et al., Discovery of NCT-501, a potent and selective theophylline-based inhibitor of aldehyde dehydrogenase 1A1 (ALDH1A1), J. Med. Chem. 58 (2015) 5967–5978.

[20] S. Kim, P. Thiessen, E. Bolton, J. Chen, G. Fu, A. Gindulyte, et al., BS the PubChem project, Nucleic Acids Res. 44 (2016) D1202–D1213.

[21] D. Mendez, A. Gaulton, A.P. Bento, J. Chambers, M. De Veij, E. Félix, et al., ChEMBL: towards direct deposition of bioassay data, Nucleic Acids Res. 47 (2019) D930–D940.

[22] C.A. Morgan, T.D. Hurley, Characterization of two distinct structural classes of selective aldehyde dehydrogenase 1A1 inhibitors, J. Med. Chem. 58 (2015) 1964–1975.

[23] B.C. Huddle, E. Grimley, C.D. Buchman, M. Chtcherbinine, B. Debnath, P. Mehta, et al., Structure-based optimization of a novel class of aldehyde dehydrogenase 1A (ALDH1A) subfamily-selective inhibitors as potential adjuncts to ovarian cancer chemotherapy, J. Med. Chem. 61 (2018) 8754–8773.

[24] C.D. Buchman, T.D. Hurley, Inhibition of the aldehyde dehydrogenase 1/2 family by psoralen and coumarin derivatives, J. Med. Chem. 60 (2017) 2439–2455.

[25] B. Parajuli, M.L. Fishel, T.D. Hurley, Selective ALDH3A1 inhibition by benzimidazole analogues increase mafosfamide sensitivity in cancer cells, J. Med. Chem. 57 (2014) 449–461.

[26] D. Liang, Y. Fan, Z. Yang, Z. Zhang, M. Liu, L. Liu, et al., Discovery of coumarin-based selective aldehyde dehydrogenase 1A1 inhibitors with glucose metabolism improving activity, Eur. J. Med. Chem. 187 (2020) 111923.

[27] A.C. Kimble-Hill, B. Parajuli, C.-H. Chen, D. Mochly-Rosen, T.D. Hurley, Development of selective inhibitors for aldehyde dehydrogenases based on substituted indole-2, 3-diones, J. Med. Chem. 57 (2014) 714–722.

[28] M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking, J. Med. Chem. 55 (2012) 6582–6594.

[29] C.W. PaDEL-descriptor Yap, An open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (2011) 1466–1474.

[30] M. Kubo, K. Yamamoto, T. Itoh, Design and synthesis of selective CYP1B1 inhibitor via dearomatization of α-naphthoflavone, Bioorg. Med. Chem. 27 (2019) 285–304.

[31] H. Lin, L. Han, C. Yap, Y. Xue, X. Liu, F. Zhu, et al., Prediction of factor Xa inhibitors by machine learning methods, J. Mol. Graph. Model. 26 (2007) 505–518.

[32] M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, et al., Package 'caret'. R J. (2020) 223.

[33] V. Vapnik, The Nature of Statistical Learning Theory, Springer science & business media, 2013.

[34] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 1–27.

[35] A. Karatzoglou, A. Smola, K. Hornik, M.A. Karatzoglou, Package 'kernlab', CRAN R Project, 2019.

[36] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[37] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (2003) 1947–1958.

[38] S. RColorBrewer, M.A. Liaw, Package 'randomForest', University of California, Berkeley: Berkeley, CA, USA, 2018.

[39] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133.

[40] B. Ripley, W. Venables, M.B. Ripley, Package 'nnet'. R package version. 7 (2016) 3–12.

[41] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, in: IEEE International Conference on Neural Networks, IEEE, 1993, pp. 586–591.

[42] L. Pérez-Regidor, M. Zarioh, L. Ortega, S. Martín-Santamaría, Virtual screening approaches towards the discovery of toll-like receptor modulators, Int. J. Mol. Sci. 17 (2016) 1508.

[43] D.V. Green, Virtual screening of chemical libraries for drug discovery, Expet Opin. Drug Discov. 3 (2008) 1011–1026.

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

[45] M.F. Koch, S. Harteis, I.D. Blank, G. Pestel, L.F. Tietze, C. Ochsenfeld, et al., Structural, biochemical, and computational studies reveal the mechanism of selective aldehyde dehydrogenase 1A1 inhibition by cytotoxic duocarmycin analogues, Angew. Chem. Int. Ed. 54 (2015) 13550–13554.

[46] J.J. Sutherland, R.K. Nandigam, J.A. Erickson, M. Vieth, Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy, J. Chem. Inf. Model. 47 (2007) 2293–2302.

[47] D.S. Biovia, Discovery Studio Modeling Environment, Release 4.5, Dassault Systèmes, San Diego, 2015.

[48] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 42717.

[49] F. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, et al., admetSAR: a Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties, ACS Publications, 2012.

[50] S. Release, 1: Desmond Molecular Dynamics System, Version 3.7, DE Shaw Research, New York, NY, 2014. Maestro-Desmond Interoperability Tools, version. 2014, 3.

[51] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.