AI-driven drug discovery: identification and optimization of ALDH3A1 selective inhibitors with nanomolar activity

Sankalp Jain, Adam Yasgar, Anu Dalal, Marissa Davies, Aleksandra Nilova, Natalia Martinez, Anton Simeonov, Ganesha Rai*, Alexey V. Zakharov*

National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, 9800 Medical Center Drive, Rockville, MD, 20850, United States

* Address for correspondence: 9800 Medical Center Dr, Rockville, Maryland 20850, USA; Telephone: 301-480-9847; E-mails: bantukallug@mail.nih.gov and alexey.zakharov@nih.gov.

Abstract:

In the field of medicinal chemistry, the discovery of novel compounds with therapeutic potential is of utmost importance. However, the conventional approach to drug discovery, relying on high-throughput screening (HTS), encounters limitations such as reagent availability and labor costs. Although a hit-finding campaign starts with a virtual screen of millions of compounds, the hit-to-lead and lead optimization stages often require designing, synthesizing, and profiling thousands of analogs before selecting clinical candidates. Recent advances offer an innovative solution - the virtual generation of billions of synthetically feasible compounds with an impressive 80% success rate for synthesis. This breakthrough has the potential to significantly expand the chemical space available for purchase and experimental validation, bringing about a revolution in the field. Our study presents a comprehensive approach to compound discovery and optimization, harnessing quantitative high-throughput screening (qHTS), chemical databases, and reaction-based enumeration. To further enhance the synthesis process and gain deeper insights into compound formation, we utilize the Reaction Cookbook from Biosolveit, which comprises reaction SMARTs for around 300 chemical reactions. By employing this wide range of reactions, we aim to uncover the full spectrum of a chemotype's potential and customize its structure to optimize desired properties. As an illustration of this approach, our work on the ALDH3A1 project resulted in the synthesis of 50 compounds, with 21 of them exhibiting activity (negative curve class values; hit-rate ~42%). Among these active compounds, 6 displayed IC_{50} values lower than 30 μM and efficacy values less than -50%, with the most potent compound achieving an impressive potency of 447 nM. This study demonstrates the successful synergy between in-silico reaction-based analogs enumeration, molecular modeling and AI/ML-based techniques in identifying compounds with improved biological activity, offering promising prospects for the development of

ALDH3A1-targeting agents as potential cancer therapeutics. Computational workflows developed in this study can be used for similar target-based drug discovery campaigns.

Introduction:

In the field of medicinal chemistry, discovering new therapeutic agents stands as a crucial and continuous challenge. While traditional drug discovery methods, notably high-throughput screening (HTS), have been helpful in identifying potential drug candidates, they are often hindered by factors such as limited availability of specialized reagents and the significant resource demands of conducting extensive experimental screens [1–4].

The advent of computational chemistry and advancements in synthetic methodologies have brought about a paradigm shift in compound discovery and optimization [5–7]. In recent years, *in silico* reaction-based enumeration has gained prominence, offering a novel solution to overcome the limitations inherent in traditional HTS methods [8–10]. This approach enables the virtual generation of a vast array of synthetically feasible compounds, thereby broadening the scope of chemical entities available for exploration and significantly improving the rate of successful compound synthesis [11, 12].

In our study, we complemented this modern approach by integrating it with quantitative structure-activity relationship (QSAR) modeling and quantitative high-throughput screening (qHTS). This combination allowed for a more systematic and data-informed identification of potential drug compounds [12, 13]. Our methodology was further enriched by incorporating extensive chemical databases [14–16]. These databases, replete with molecular building blocks and reaction templates offer a rich resource for data-driven drug design, facilitating the exploration of a wide range of chemical structures and properties [14]. QSAR modeling has been successfully used in the past for virtual screening, where it helps predict the biological activity of compounds based on their chemical structures [13, 17–21]. This prediction capability of QSAR models aids in narrowing down the vast pool of potential compounds to those most likely to exhibit desired biological activities [22].

To illustrate the effectiveness of this approach, we evaluated its potential using ALDH3A1 as a case study. ALDH3A1 serves a critical function in metabolizing a broad spectrum of aliphatic and aromatic aldehydes, safeguarding cells from oxidative damage, and maintaining cellular balance [23, 24]. Dysregulation of this enzyme is implicated in an array of diseases, such as cancers and neurodegenerative disorders, highlighting its value as a target for therapeutic intervention across these conditions [25, 26].

Applying this approach with the goal of discovering potent, novel, and selective inhibitors targeting ALDH3A1 through performing the comprehensive structure-activity relationships optimization resulted in the synthesis of 50 distinct compounds, demonstrating the method's ability to produce potent chemical compounds. Of these, 21 compounds showed activity, indicated by their negative curve class values. Notably, 6 of these active compounds had IC_{50} values below 30 μM and efficacy values were under -50%, suggesting their strong potency. The most potent compound identified had IC_{50} of 447 nM, showing the effectiveness of this approach in finding compounds with significant potential for drug discovery.

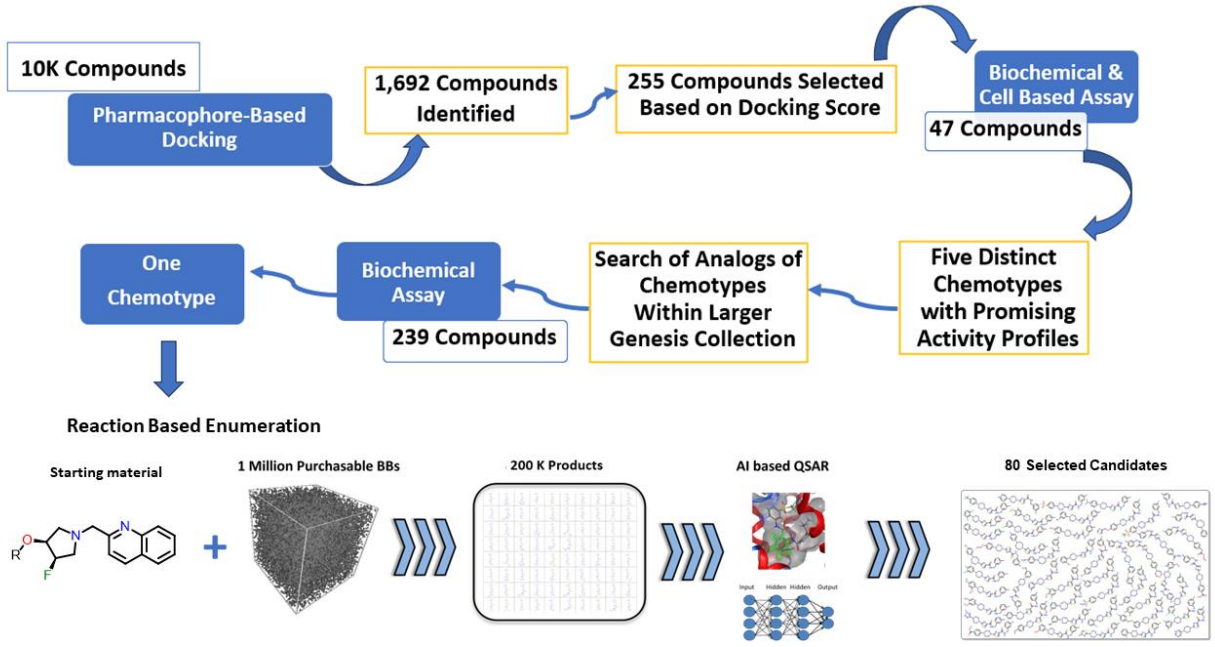


Figure 1: Flowchart of Study Approach: A flowchart that outlines the steps taken in the study: pharmacophore-based docking, reaction-based enumeration, QSAR modeling, and experimental validation.

In summary, our method, which synergistically combines data-driven compound selection, reaction-based synthesis, and comprehensive computational analysis, marks a significant advancement in the field of medicinal chemistry. It paves the way for the expedited development of novel compounds with enhanced therapeutic potential, potentially bringing substantial benefits to patient care and advancing public health initiatives.

Methods:

Pharmacophore-Based Docking and Selection

The pharmacophore-based molecular docking was performed using the Molecular Operating Environment (MOE) software [27–29]. The compound library for docking comprised a diverse set of molecular compounds (approximately 9,000 compounds), each of which was prepared using MOE's ligand preparation tools. This involves energy minimization and the generation of 3D conformations. The crystal structure of human ALDH3A1, PDB ID: 3SZA, was selected as the starting point for our pharmacophore-based molecular docking studies. Initial steps involved cleaning the protein structure, which included the removal of water molecules, irrelevant heteroatoms, and the addition of hydrogens. Protonation states were assigned, followed by energy minimization to optimize the protein's conformation [30, 31].

Active Site Identification

The active site of ALDH3A1 was identified based on the ligand-binding region observed in the crystal structure. This region was defined as the docking site for subsequent simulations.

Pharmacophore Model Development

A pharmacophore model was developed (in MOE) to represent the essential features required for interaction with the crystal ligand in its protein binding site (PDB ID: 3SZA). The model included key features such as hydrogen bond donors and acceptors, hydrophobic regions, and aromatic rings, which are critical for molecular interactions.

Docking Protocol

The docking protocol involved aligning the compounds from the library to the pharmacophore model to identify potential binders [32]. MOE's docking algorithm (Affinity dG scoring function) was used to simulate the interaction between each compound and the target binding site. Post-docking, each compound was scored based on its docking pose and the degree of match with the pharmacophore model. The scoring function in MOE provided a quantitative measure of the binding affinity. Compounds were then ranked based on their scores, with higher-scoring compounds considered more likely to be active.

Experimental Testing and Assay Information

The assay has been previously published [33]. Briefly, 3 µL of enzyme (final concentration of 5 nM for ALDH3A1) or assay buffer (100 mM HEPES pH 7.5 with 0.01% Tween 20) were

dispensed into a 1,536-well solid-bottom black plate. Twenty nL of compounds (final concentration range 420 pM to 50 μM) or control (DEAB or ZM-39923 final concentration range 1.52 nM to 50 μM) were transferred via acoustic droplet ejection. Samples were incubated (room temperature, protected from light) for 15 minutes followed by an addition of 1 μL substrate mixture of $NADP^+$ and benzaldehyde (final concentrations of 1 mM and 200 μM, respectively; K_ms of 260 μM and 280 μM, respectively) [34, 35], in addition to coupling reagents Resazurin and Diaphorase (final concentrations of 100 μM and 0.7 U/mL) to right-shift the detection by monitoring the production of resorufin [36, 37]. Plates were centrifuged at 1,000 rpm (164 x g) for 15 seconds, then read (RT) in kinetic mode on a ViewLux High-throughput CCD imager equipped with standard Rhodamine optics (525 nm excitation, 598 nm emission) for 5 minutes (< 20% conversion). The change in fluorescence intensity over the 5-minute reaction period was normalized against no-inhibitor and no-enzyme controls.

Data from each assay were normalized plate-wise to corresponding intra-plate controls as noted above. The same controls were used for the calculation of the Z' factor, a measure of assay quality control. Concentration-response curves were fitted and classified as described previously [38–42]. IC_{50}'s were calculated using Prism software (version 8.1.2, GraphPad Software, Inc.), sigmoidal dose-response (variable slope). We used the Palantir Technologies (Washington, DC) data integration platform, which is configured to ingest all HTS results generated at NCATS and harmonized this data with other sources such as ChEMBL and OrthoMCL. All qHTS screening results are publicly available at PubChem**.** The chemical structures were standardized using the LyChI (Layered Chemical Identifier) program (version 20141028, https://github.com/ncats/lychi) [43].

Reaction-Based Compound Enumeration with KNIME

For the enumeration of potential compounds, we utilized the KNIME Analytics Platform [44, 45], leveraging its capabilities for integrating and automating cheminformatics processes [11]. Our approach focused on two key synthetic reactions: O-alkylation and amide coupling. We generated the reaction SMARTS for the two reactions and, using the RDKit two-component reaction node within KNIME, systematically generated a virtual library of compounds. This process was based on the chemotype we had previously identified, employing Enamine's building blocks to explore various compound permutations. The enumeration process yielded approximately 200,000 virtual compounds, thereby expanding our chemical space and facilitating the discovery of novel variants related to our target chemotype. This expansive virtual library serves as a resource for subsequent screening and experimental validation.

A QSAR model utilizing biochemical data from older protocols was then applied to these enumerated compounds to select the most promising candidates.

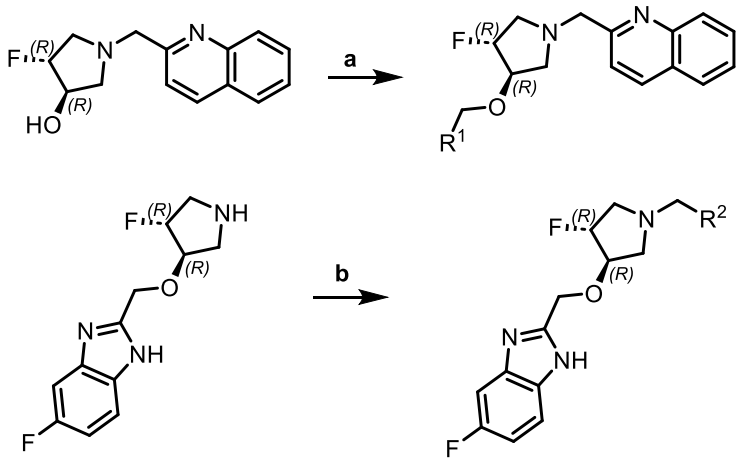
Quantitative Structure-Activity Relationship (QSAR) Modeling with KNIME

To refine the selection of our chemotype-specific virtual compound library, we developed a regression-based Quantitative Structure-Activity Relationship (QSAR) model [13, 46]. This model was built using the KNIME Analytics Platform, leveraging both physicochemical descriptors and Morgan fingerprints from RDKit to characterize the molecular structures. The model was built using the tree ensemble learning node within KNIME, which is well-suited for capturing complex non-linear relationships between descriptors and biological activity.

To ensure the relevance of our QSAR model's predictions, we introduced an applicability domain framework [47, 48]. This was performed by calculating the Tanimoto similarity between the enumerated compounds and those in the training set. We set a similarity threshold of 0.7, allowing us to focus on compounds that were structurally related to the biologically active compounds in our training data. This step ensured that the model predictions remained within the chemical space for which the model was applicable and reliable.

The incorporation of the applicability domain by similarity assessment served a dual purpose: it enhanced the robustness of our predictions and allowed us to identify which novel compounds were most likely to exhibit desirable levels of biological activity. Consequently, this approach facilitated a more targeted and efficient selection of candidates for subsequent experimental validation.

Synthetic Routes Assessing Pyrrolidine-Based ALDH3A1 Inhibitors



Reagents and conditions: (a) R^1-X (X = Cl or Br, K_2CO_3, DMF, rt to 60°C, 1−24 h. (b) R^2CHO, $NaBH(OAc)_3$, DCE, rt, 1-12 h.

*General Procedure for (**1**):* To a mixture of 4-fluoro-1-(quinolin-2-ylmethyl)pyrrolidin-3-ol (10 mmol, 1 eq) and potassium carbonate (K_2CO_3, 15 mmol, 1.5 eq) was added N,N-dimethylformamide (DMF, 0.6 M). The mixture was stirred at room temperature for 5 minutes, before the addition of R^1-X (12 mmol, 1.2 eq). The reaction mixture was stirred at room temperature for 30 minutes and then heated at 60°C for 2 hours. After the reaction was complete, the mixture was filtered to remove the solids. The filtrate containing the product was purified on a preparative HPLC to obtain pure products after lyophilization.

*General Procedure for (**2**):* "To a mixture of 5-fluoro-2-((4-fluoropyrrolidin-3-yloxy)methyl)-1H-benzo[d]imidazole (10 mmol, 1 eq) and an appropriate aldehyde (12 mmol, 1.2 eq) in dichloroethane/MeOH (0.6 M) was added sodium triacetoxyborohydride (15 mmol, 1.5 eq). The reaction mixture was stirred at room temperature for 1-12 hours. After completion of the reaction, the excess solvent was removed by evaporation (instead of 'forced air'). The crude mixture was then dissolved in DMSO and purified using preparative HPLC to obtain pure products.

Results:

Identification of Compounds through Pharmacophore-Based Docking

Our pharmacophore-based docking approach, utilizing the MOE software, successfully identified 1,692 compounds that aligned with our pharmacophore hypothesis. From this set, we carefully selected 255 top-ranked compounds for further analysis. This selection was made based on criteria such as docking scores, how well each compound fit the pharmacophore model, and their chemical diversity. These compounds were then advanced to *in vitro* assays to assess their biological activity against the ALDH3A1 enzyme.

Experimental Validation and Activity Screening

Experimental validation of the 255 compounds resulted in the identification of 47 compounds with significant biological activity. This activity was characterized by their ability to inhibit the enzymatic function of the ALDH3A1 target, reinforcing the predictive power of our pharmacophore-based docking strategy.

From these active compounds, we were able to distinguish five unique chemotypes, each showing a promising profile of activity. Focusing on one of these chemotypes, we then performed a reaction-based enumeration to further explore its potential.

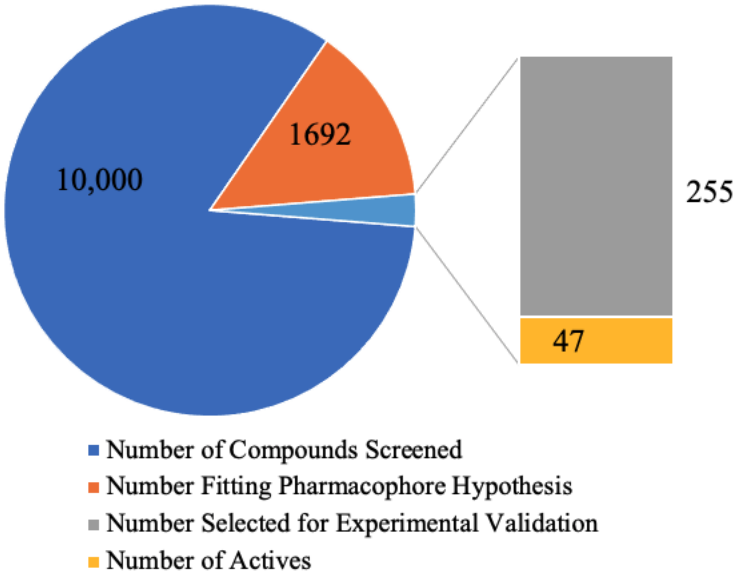


Figure 2: Docking and Activity Results: A pie chart showing the number of compounds screened, the number fitting the pharmacophore hypothesis, the number selected for experimental validation, and the number showing activity.

Compound Expansion through Reaction-Based Enumeration

We expanded our chemical exploration by employing reaction-based enumeration using one million building blocks provided by Enamine. This process included both amide coupling and O-alkylation reactions, culminating in a substantial library of 250,000 enumerated compounds. Subsequent application of a QSAR model, which was trained using biochemical data, enabled us to shortlist the most promising candidates from this extensive compound set.

QSAR Model Performance and Compound Selection

The QSAR model, built on regression analysis using RDKit descriptors and fingerprints, demonstrated an R-squared value of 0.52, indicating a moderate predictive ability. Leveraging this model, we selected the top 500 compounds for synthesis.

Synthesis and Bioactivity of Selected Compounds

From the top 500 compounds, we further selected and were able to synthesize 50. Testing of these synthesized compounds revealed that 21 exhibited measurable activity, as indicated by negative curve class values. Further scrutiny within this active group identified 6 compounds with IC_{50} values below 30 µM and efficacy (Eff) values of less than -50%. The most potent compound within this selection stood out with an IC_{50} of 447 nM, showcasing the strength of our integrated approach in discovering highly active compounds for potential therapeutic use.

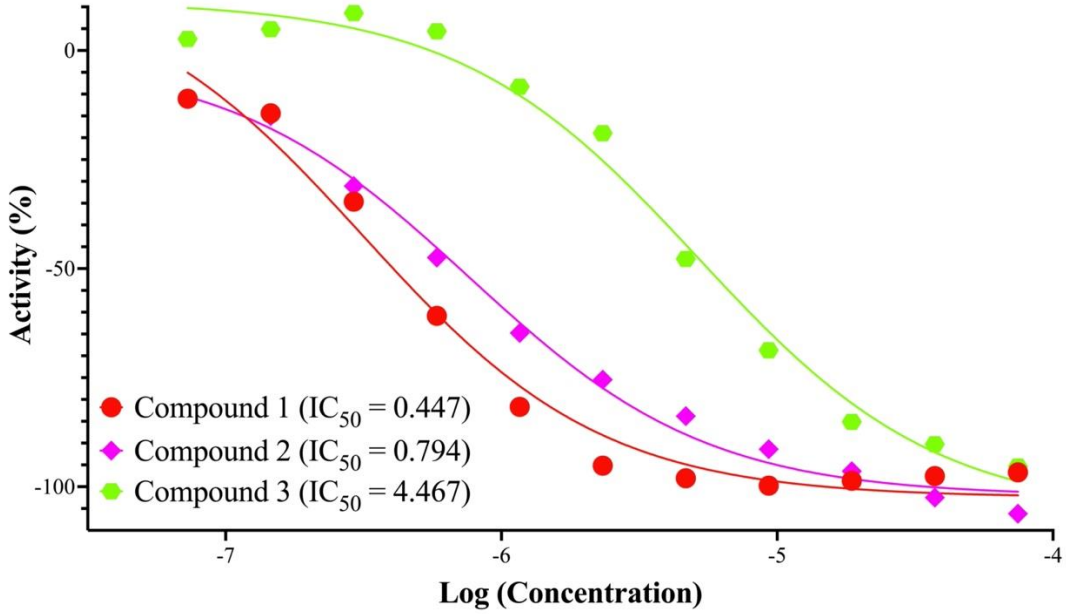


Figure 3: Lead Compound Potency: Dose-response curve for top 3 lead compounds.

Discussion

The results obtained from our integrated drug discovery approach illustrate the effectiveness of combining pharmacophore-based docking with reaction-based enumeration and QSAR modeling. The successful identification of 47 active compounds from an initial pool of 1,692, derived via pharmacophore-based docking, validates our hypothesis-driven selection criteria. Notably, the reduction to 255 compounds for experimental validation underscores the importance of a multi-parameter selection process that balances docking scores with pharmacophore fit and chemical diversity.

Our reaction-based enumeration strategy, leveraging the extensive Enamine building block repository, demonstrates the utility of combinatorial chemistry in modern drug discovery. By generating a vast library of 250,000 compounds through targeted reactions, we effectively expanded the chemical space under investigation. The subsequent application of a QSAR model,

with a moderate R-squared value of 0.52, was instrumental in prioritizing 500 compounds with the highest potential for biological activity.

The synthesis and experimental validation of these compounds led to a substantial hit rate of 42%, with 21 active compounds emerging from a synthesized set of 50. This high success rate demonstrates the robustness of our reaction enumeration and QSAR predictions and the chemical intuition underlying the selection of these molecules for synthesis. The further delineation of 6 compounds with IC_{50} values lower than 30 μM and efficacy values less than -50% reinforces the potential of these compounds as potent inhibitors of ALDH3A1. Finding a lead compound with an IC_{50} value of 447 nM is a noteworthy step forward in our study. It exemplifies the transformative potential of a methodical and data-driven approach to drug discovery. This compound, alongside the other active compounds identified, provides a promising foundation for the development of new therapeutic agents.

In conclusion, our study demonstrates the power of an integrated approach that combines computational and experimental methodologies. This hybrid strategy maximizes the efficiency of the drug discovery process and paves the way for the rapid development of novel therapeutic compounds. Future work will focus on further optimization of these active compounds and validation of their efficacy in relevant biological models, with the ultimate goal of contributing meaningful advances to the field of medicinal chemistry and drug development.

Acknowledgments

This research was supported in part by the Intramural/Extramural research program of the NCATS, NIH.

References:

1. Major J (1998) Challenges and Opportunities in High Throughput Screening: Implications for New Technologies. J Biomol Screen 3:13–17. https://doi.org/10.1177/108705719800300102

2. Attene-Ramos MS, Austin CP, Xia M (2014) High Throughput Screening. In: Wexler P (ed) Encyclopedia of Toxicology (Third Edition). Academic Press, Oxford, pp 916–917

3. Maas RGC, Beekink T, Chirico N, et al (2023) Generation, High-Throughput Screening, and Biobanking of Human-Induced Pluripotent Stem Cell-Derived Cardiac Spheroids. J Vis Exp JoVE. https://doi.org/10.3791/64365

4. Liu B, Li S, Hu J (2004) Technological advances in high-throughput screening. Am J Pharmacogenomics Genomics-Relat Res Drug Dev Clin Pract 4:263–276. https://doi.org/10.2165/00129785-200404040-00006

5. Cai JH, Zhu XZ, Guo PY, et al (2023) Recent updates in click and computational chemistry for drug discovery and development. Front Chem 11:1114970. https://doi.org/10.3389/fchem.2023.1114970

6. Azad I, Khan T, Ahmad N, et al Updates on drug designing approach through computational strategies: a review. Future Sci OA 9:FSO862. https://doi.org/10.2144/fsoa-2022-0085

7. Gioiello A, Piccinno A, Lozza AM, Cerra B (2020) The Medicinal Chemistry in the Era of Machines and Automation: Recent Advances in Continuous Flow Technology. J Med Chem 63:6624–6647. https://doi.org/10.1021/acs.jmedchem.9b01956

8. Sadybekov AV, Katritch V (2023) Computational approaches streamlining drug discovery. Nature 616:673–685. https://doi.org/10.1038/s41586-023-05905-z

9. Konze KD, Bos PH, Dahlgren MK, et al (2019) Reaction-Based Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-Dependent Kinase 2 Inhibitors. J Chem Inf Model 59:3782–3793. https://doi.org/10.1021/acs.jcim.9b00367

10. Ghiandoni GM, Bodkin MJ, Chen B, et al (2019) Development and Application of a Data-Driven Reaction Classification Model: Comparison of an Electronic Lab Notebook and Medicinal Chemistry Literature. J Chem Inf Model 59:4167–4187. https://doi.org/10.1021/acs.jcim.9b00537

11. Saldívar-González FI, Huerta-García CS, Medina-Franco JL (2020) Chemoinformatics-based enumeration of chemical libraries: a tutorial. J Cheminformatics 12:64. https://doi.org/10.1186/s13321-020-00466-z

12. Mouchlis VD, Afantitis A, Serra A, et al (2021) Advances in De Novo Drug Design: From Conventional to Machine Learning Methods. Int J Mol Sci 22:1676. https://doi.org/10.3390/ijms22041676

13. Jain S, Talley DC, Baljinnyam B, et al (2021) Hybrid In Silico Approach Reveals Novel Inhibitors of Multiple SARS-CoV-2 Variants. ACS Pharmacol Transl Sci 4:1675–1688. https://doi.org/10.1021/acsptsci.1c00176

14. Grygorenko OO, Radchenko DS, Dziuba I, et al (2020) Generating Multibillion Chemical Space of Readily Accessible Screening Compounds. iScience 23:101681. https://doi.org/10.1016/j.isci.2020.101681

15. Meier K, Bühlmann S, Arús-Pous J, Reymond J-L (2020) The Generated Databases (GDBs) as a Source of 3D-shaped Building Blocks for Use in Medicinal Chemistry and Drug Discovery. Chimia 74:241–246. https://doi.org/10.2533/chimia.2020.241

16. Wang J, Hou T (2010) Drug and Drug Candidate Building Block Analysis. J Chem Inf Model 50:55–67. https://doi.org/10.1021/ci900398f

17. Neves BJ, Braga RC, Melo-Filho CC, et al (2018) QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. Front Pharmacol 9:1275. https://doi.org/10.3389/fphar.2018.01275

18. Hochuli JE, Jain S, Melo-Filho C, et al (2022) Allosteric Binders of ACE2 Are Promising Anti-SARS-CoV-2 Agents. ACS Pharmacol Transl Sci 5:468–478. https://doi.org/10.1021/acsptsci.2c00049

19. Fedorova EV, Buryakina AV, Zakharov AV, et al (2014) Design, Synthesis and Pharmacological Evaluation of Novel Vanadium-Containing Complexes as Antidiabetic Agents. PLOS ONE 9:e100386. https://doi.org/10.1371/journal.pone.0100386

20. Abrams RPM, Yasgar A, Teramoto T, et al (2020) Therapeutic candidates for the Zika virus identified by a high-throughput screen for Zika protease inhibitors. Proc Natl Acad Sci U S A 117:31365–31375. https://doi.org/10.1073/pnas.2005463117

21. Zhao Y, Ung PM-U, Zahoránszky-Kőhalmi G, et al (2020) Identification of a G-Protein-Independent Activator of GIRK Channels. Cell Rep 31:107770. https://doi.org/10.1016/j.celrep.2020.107770

22. Zakharov AV, Peach ML, Sitzmann M, et al (2012) Computational tools and resources for metabolism-related property predictions. 2. Application to prediction of half-life time in human liver microsomes. Future Med Chem 4:1933–1944. https://doi.org/10.4155/fmc.12.152

23. Estey T, Piatigorsky J, Lassen N, Vasiliou V (2007) ALDH3A1: a corneal crystallin with diverse functions. Exp Eye Res 84:3–12. https://doi.org/10.1016/j.exer.2006.04.010

24. Voulgaridou G-P, Theologidis V, Venetikidou M, et al (2023) Investigating the Functional Roles of Aldehyde Dehydrogenase 3A1 in Human Corneal Epithelial Cells. Int J Mol Sci 24:5845. https://doi.org/10.3390/ijms24065845

25. Xia J, Li S, Liu S, Zhang L (2023) Aldehyde dehydrogenase in solid tumors and other diseases: Potential biomarkers and therapeutic targets. MedComm 4:e195. https://doi.org/10.1002/mco2.195

26. Xanthis V, Mantso T, Dimtsi A, et al (2023) Human Aldehyde Dehydrogenases: A Superfamily of Similar Yet Different Proteins Highly Related to Cancer. Cancers 15:4419. https://doi.org/10.3390/cancers15174419

27. Chemical Computing Group - Citing MOE. https://www.chemcomp.com/Research-Citing_MOE.htm. Accessed 1 Aug 2016

28. Molecular Operating Environment (MOE), 2019.01

29. Giordano D, Biancaniello C, Argenio MA, Facchiano A (2022) Drug Design by Pharmacophore and Virtual Screening Approach. Pharmaceuticals 15:646. https://doi.org/10.3390/ph15050646

30. Jain S, Grandits M, Ecker GF (2018) Interspecies comparison of putative ligand binding sites of human, rat and mouse P-glycoprotein. Eur J Pharm Sci 122:134–143. https://doi.org/10.1016/j.ejps.2018.06.022

31. Klepsch F, Vasanthanathan P, Ecker GF (2014) Ligand and Structure-Based Classification Models for Prediction of P-Glycoprotein Inhibitors. J Chem Inf Model 54:218–229. https://doi.org/10.1021/ci400289j

32. Jain S, Grandits M, Richter L, Ecker GF (2017) Structure based classification for bile salt export pump (BSEP) inhibitors using comparative structural modeling of human BSEP. J Comput Aided Mol Des 31:507–521. https://doi.org/10.1007/s10822-017-0021-x

33. Yang S-M, Martinez NJ, Yasgar A, et al (2018) Discovery of Orally Bioavailable, Quinoline-Based Aldehyde Dehydrogenase 1A1 (ALDH1A1) Inhibitors with Potent Cellular Activity. J Med Chem 61:4883–4903. https://doi.org/10.1021/acs.jmedchem.8b00270

34. Parajuli B, Fishel ML, Hurley TD (2014) Selective ALDH3A1 Inhibition by Benzimidazole Analogues Increase Mafosfamide Sensitivity in Cancer Cells. J Med Chem 57:449–461. https://doi.org/10.1021/jm401508p

35. Morgan CA, Parajuli B, Buchman CD, et al (2015) N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. Chem Biol Interact 234:18–28. https://doi.org/10.1016/j.cbi.2014.12.008

36. Davis MI, Shen M, Simeonov A, Hall MD (2016) Diaphorase Coupling Protocols for Red-Shifting Dehydrogenase Assays. Assay Drug Dev Technol 14:207–212. https://doi.org/10.1089/adt.2016.706

37. Rohde JM, Brimacombe KR, Liu L, et al (2018) Discovery and optimization of piperazine-1-thiourea-based human phosphoglycerate dehydrogenase inhibitors. Bioorg Med Chem 26:1727–1739. https://doi.org/10.1016/j.bmc.2018.02.016

38. Auld D, Thorne N, Boxer M, et al (2010) Understanding Enzymes as Reporters or Targets in Assays Using Quantitative High-throughput Screening (qHTS)

39. Inglese J, Auld DS, Jadhav A, et al (2006) Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci 103:11473–11478. https://doi.org/10.1073/pnas.0604348103

40. Wang Y, Jadhav A, Southal N, et al (2010) A Grid Algorithm for High Throughput Fitting of Dose-Response Curve Data. Curr Chem Genomics 4:57–66. https://doi.org/10.2174/1875397301004010057

41. Huang R (2016) A Quantitative High-Throughput Screening Data Analysis Pipeline for Activity Profiling. Methods Mol Biol Clifton NJ 1473:111–122. https://doi.org/10.1007/978-1-4939-6346-1_12

42. Seethala R, Zhang L (2009) Handbook of drug screening, 2nd ed. Informa Healthcare, New York

43. Stefaniak F (2015) Prediction of Compounds Activity in Nuclear Receptor Signaling and Stress Pathway Assays Using Machine Learning Algorithms and Low-Dimensional Molecular Descriptors. Front Environ Sci 3:

44. KNIME | Open for Innovation. https://www.knime.org/. Accessed 1 Nov 2016

45. Berthold MR, Cebron N, Dill F, et al (2009) KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. SIGKDD Explor Newsl 11:26–31. https://doi.org/10.1145/1656274.1656280

46. Achary PGR (2020) Applications of Quantitative Structure-Activity Relationships (QSAR) based Virtual Screening in Drug Design: A Review. Mini Rev Med Chem 20:1375–1388. https://doi.org/10.2174/1389557520666200429102334

47. Zakharov AV, Zhao T, Nguyen D-T, et al (2019) Novel Consensus Architecture To Improve Performance of Large-Scale Multitask Deep Learning QSAR Models. J Chem Inf Model 59:4613–4624. https://doi.org/10.1021/acs.jcim.9b00526

48. Tropsha A, Golbraikh A (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Curr Pharm Des 13:3494–3504. https://doi.org/10.2174/138161207782794257