Molecular Diversity
https://doi.org/10.1007/s11030-024-10960-3

ORIGINAL ARTICLE

Novel molecular inhibitor design for *Plasmodium falciparum* Lactate dehydrogenase enzyme using machine learning generated library of diverse compounds

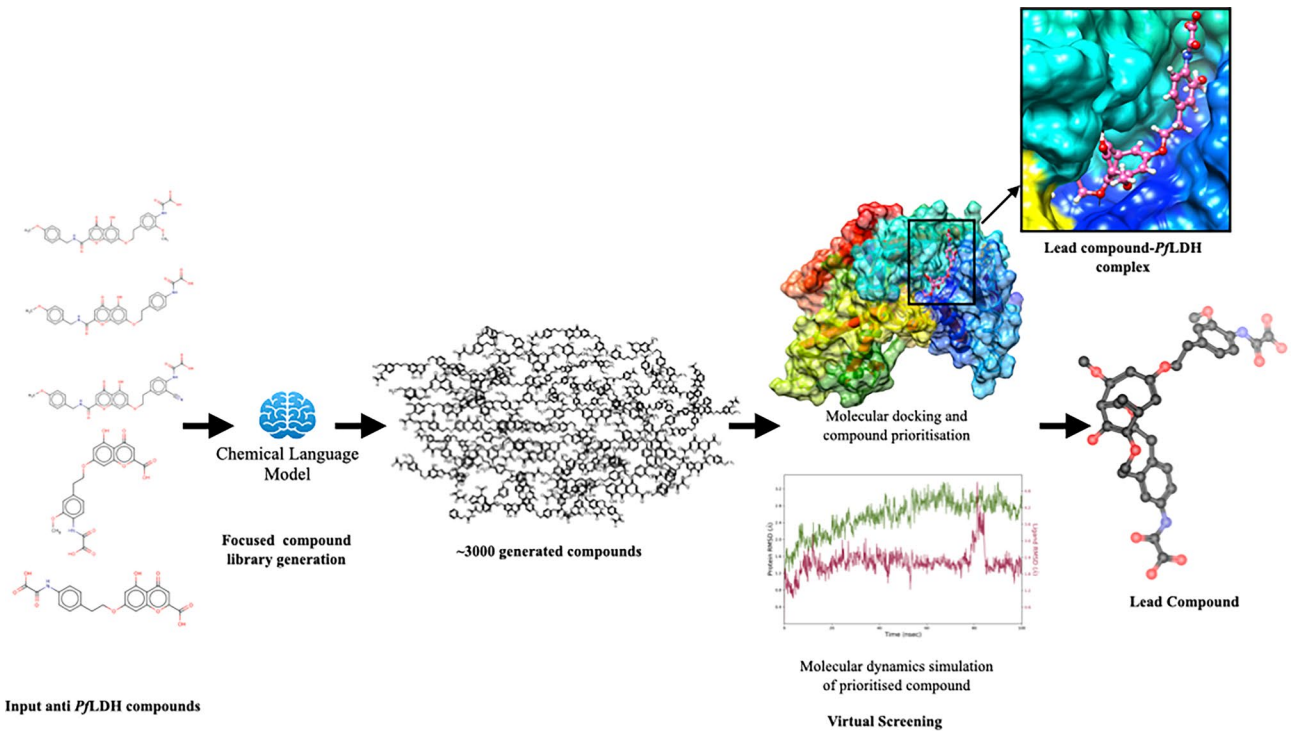
Jitendra Kuldeep[1] · **Neeraj Chaturvedi**[1] · **Dinesh Gupta**[1]

Received: 10 April 2024 / Accepted: 6 August 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Generative machine learning models offer a novel strategy for chemogenomics and de novo drug design, allowing researchers to streamline their exploration of the chemical space and concentrate on specific regions of interest. In cases with limited inhibitor data available for the target of interest, de novo drug design plays a crucial role. In this study, we utilized a package called 'mollib,' trained on ChEMBL data containing approximately 365,000 bioactive molecules. By leveraging transfer learning techniques with this package, we generated a series of compounds, starting from five initial compounds, which are potential *Plasmodium falciparum* (*Pf*) Lactate dehydrogenase inhibitors. The resulting compounds exhibit structural diversity and hold promise as potential novel *Pf* Lactate dehydrogenase inhibitors.

Graphical Abstract



Keywords Deep learning · Generative model · Transfer learning · Molecular docking · Molecular dynamics simulation

Extended author information available on the last page of the article

Published online: 20 August 2024

Introduction

Malaria is a tropical disease caused by protozoa of the genus *Plasmodium* and is often fatal due to drug resistance, leading to millions of deaths per year [1, 2]. The rise of Multiple drug resistance in *Plasmodium* has necessitated putting new efforts into drug discovery and development [3–8]. The search for new drugs which act through a novel mechanism in our bodies has received much attention. Of the many *Plasmodium falciparum* targets being worked on, lactate dehydrogenase enzyme is an attractive drug target for developing new drugs [9–11]. To meet its metabolic requirements, the parasite requires more glucose from its host cells; therefore, targeting enzymes from glycolysis can potentially impede adenosine triphosphate (ATP) production and lead to parasite mortality. The function of *Pf*LDH is the reduction of pyruvate to lactate using the cofactor NADH, has unique residues and kinetic differences from the human LDH (hLDH), which further suggests that *Pf*LDH is a suitable antimalarial drug target [12–14].

Several experimental and computational studies show candidate inhibitors' binding to *Pf*LDH [15–21]. It has been demonstrated that chloroquine binding in or near the *Pf*LDH cofactor binding site acts as a competitive inhibitor [22, 23]. Similarly, selective inhibition of the enzyme compared to the human enzyme has also been reported using gossypol derivate [24]. A series of heterocyclic, azole-based compounds that preferentially inhibit the enzyme at sub-micromolar concentrations have also been studied [25].

We obtained *Pf*LDH inhibitors from ChEMBL [26, 27] to serve as seed molecules for this study. We identified the inhibitors with the most favorable IC50 values and conducted similarity calculations on them. The resulting five compounds were chosen for transfer learning to generate application-focused libraries. Additionally, we carried out molecular docking to generate complex structures of the generated compounds with *Pf*LDH. Furthermore, we performed a similarity analysis on the top-scoring compounds compared to the input molecules. Any compounds found to be dissimilar underwent molecular dynamics simulation to evaluate their stability.

Machine learning is a rapidly evolving field with a potential to significantly impact pharmaceutical research and drug discovery, including the de novo generation of diverse novel compounds, designing and optimizing lead compounds, and predicting clinical trial outcomes [28]. In recent advancements, Stokes et al. demonstrated a successful application for antibiotic discovery leading to the identification of halicin, a structurally novel molecule with potent antibacterial activity against a wide range of pathogens [29]. Various machine learning algorithms, including support vector machines and neural networks, have been successfully employed to predict the antimalarial activity of compounds with accuracies exceeding 85% [30–33].

Structure-based drug design studies, like those conducted by Rahul et al., have explored targeting *Pf*LDH for therapeutic purposes [15]. Amod et al. have successfully identified novel *Plasmodium falciparum* inhibitors using machine learning and virtual screening methodologies [34]. However, previous studies have not integrated both machine learning and structure-based drug design specifically for the *Pf*LDH target. Building on these advancements, this study uniquely combines the power of generative AI with structure-based virtual screening to accelerate the discovery of potent and novel *Pf*LDH inhibitors. By integrating these complementary methodologies, we aim to identify promising *Pf*LDH inhibitors and advance the global fight against malaria.

Generative deep learning techniques enable the creation of novel data, like chemical structures, without explicit human-coded rules. These models, often using language modeling approaches, learn from features of large datasets of sequences, such as simplified molecular input line entry systems (SMILES). Transfer learning is crucial as it helps us apply what we've learned from bioactive molecules to address the unique needs of specific pharmacological targets [35].

Materials and methods

Compounds selection

Numerous *Pf* Lactate dehydrogenase inhibitor compounds exist in various small-molecule databases. We obtained these compounds from the ChEMBL database due to its comprehensiveness and ease of access [26, 27]. The selection of the five initial compounds was based on a two-step process (See Fig. 1 for the study pipeline). First, we retrieved compounds from the ChEMBL database that were identified as *Pf*LDH inhibitors [36]. Subsequently, we employed the Tanimoto similarity coefficient to assess the similarity between these retrieved compounds. This step allowed us to select a diverse set of five compounds that encompassed a range of bioactivity values, from strong to weak inhibitors. This diversity of compound structures ensures that the transfer learning process explores a broader chemical space, potentially generating compounds with properties similar to both highly active and less active compounds, as well as intermediate levels of activity.

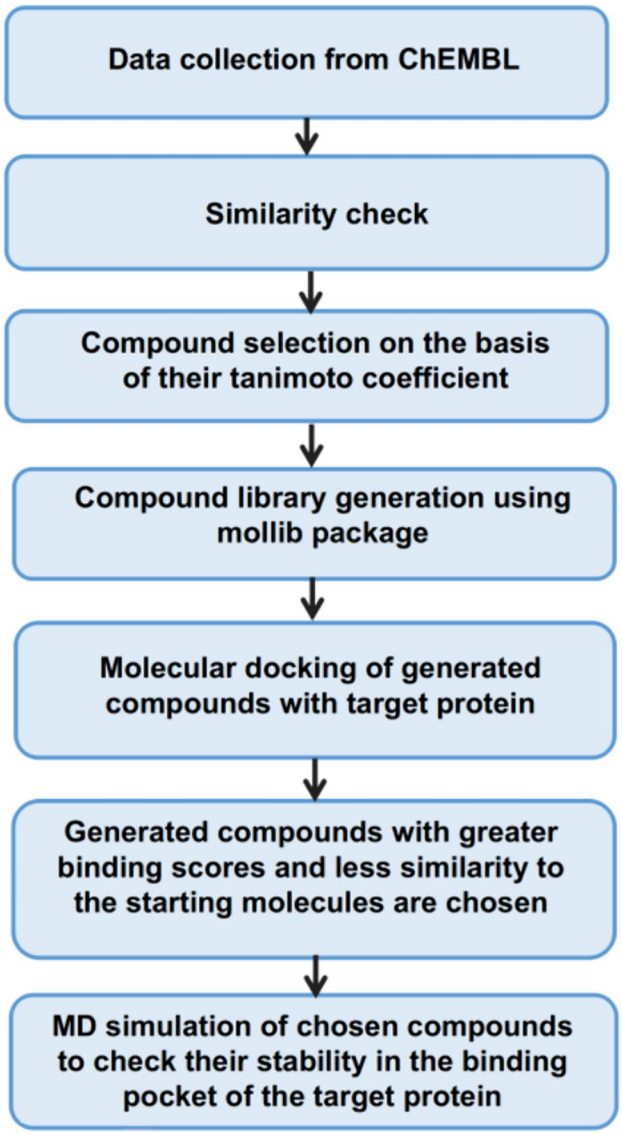


Fig. 1 Study pipeline followed in the study. Initially seed molecules were selected for generation of library of diverse compounds, followed by stringent methodology for prioritization of the best compound

Similarity calculation

We used similarity search calculations to see how many of the input compounds resembled each other. To prioritize the generated compounds, we looked for the ones that were the most different from the input compounds, indicating they were more unique. The calculation of compound similarity was carried out utilizing the online web server ChemMine tools [37]. The Tanimoto coefficient was calculated for all the retrieved compounds.

Common pharmacophores

The five seed compounds selected for this study represent a series of structurally diverse oxamic acid derivatives with potential as *Pf*LDH inhibitors. These compounds were chosen based on their unique structures and documented bioactivity against *Pf*LDH.

Fig. 2 illustrates the common pharmacophores across the selected parent compounds. The identified pharmacophores include hydrogen bond acceptors (A), which facilitate interaction with key amino acids in the active site of *Pf*LDH, and hydrogen bond donors (D), which contribute to the formation of stable hydrogen bonds with the enzyme. Additionally, hydrophobic regions (H) enhance the compounds' interaction with hydrophobic pockets within the enzyme, while positive ionic centers (P) engage in electrostatic interactions with negatively charged residues in the active site. These pharmacophores are crucial for the binding affinity and inhibitory activity of the compounds, playing a significant role in the observed bioactivity trends.

Generating application-focused compound libraries using mollib

The chemical language model (CLM) demonstrated its ability to generate a variety of molecules with a focus on specific structural features. By using transfer learning, it expanded the range of molecules it could create by learning from diverse sets of compounds. This allowed the CLM to generate molecules that were different from the ones it initially learned from, enriching the diversity of molecules it could produce [35]. The transfer learning model involves utilizing a model trained on one problem to address a related second problem. We employed the mollib package, written in Python, for the transfer learning process. The utilized CLM model employed for training is a sequential neural network with two Long Short-Term Memory (LSTM) layers. The first LSTM layer comprises 1024 units, while the second has 256 units. Batch normalization is applied before and after the LSTM layers. The output layer employs the TimeDistributed wrapper with a Dense layer having 71 units and a softmax activation function. The optimization is performed using the Adam optimizer with a learning rate of 0.0001 [35]. Our study employed this technique to generate unique and diverse virtual compound libraries to identify potential inhibitor molecules for *Pf*LDH.

Binding site studies and molecular docking

The crystal structure of the *Pf*LDH was downloaded from RCSB PDB (with PDB ID—1LDG). The 1LDG structure shows that *Pf*LDH is tetrameric, with each subunit consisting of two domains: a catalytic domain and a NADH-binding

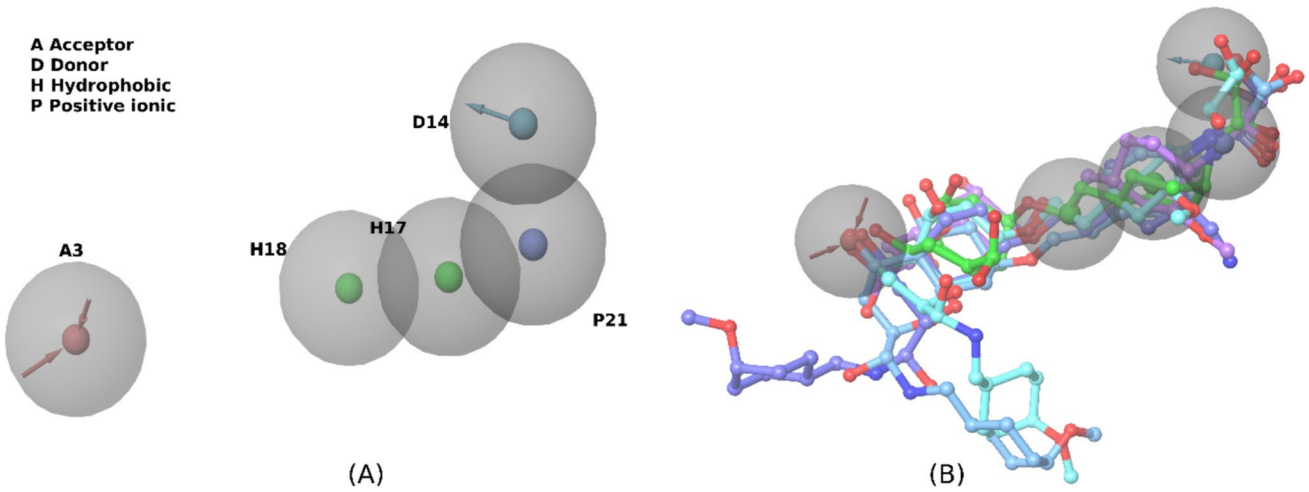


Fig. 2 **A** Common pharmacophores in parent compounds, **B** parent compound alignment on common pharmacophores

domain. The catalytic domain is similar to LDH from other organisms, but the NADH-binding domain is unique. This unique domain provides a potential target for the design of new antimalarial drugs [38, 39].

To carry out molecular docking, pre-processing of the study molecules was performed, as described below.

Protein and ligand preparation

The protein and ligands underwent preparation steps using the protein preparation module and ligand preparation module within the Cresset software (http://www.cresset-group.com/flare/). This involved the addition of hydrogens and charges to the protein and ligands. Subsequently, the structures were subjected to energy minimization.

Receptor grid generation and ligands docking

To perform ligands docking, we performed receptor grid generation which involves the creation of a three-dimensional grid around the active site of a target protein. The crucial residues for enzymatic activity, namely Gly29, Lys62, Thr97, Thr101, Lys102, His243, Als244, Ser245, and Pro246 [39], were employed to construct a grid within the protein. Once the grid was established, the docking of ligands took place using the docking module of the Cresset software (http://www.cresset-group.com/flare/).

Molecular dynamics simulation

The docked complexes of the prioritized compounds were subjected to a 100 ns molecular dynamics simulation using the Desmond program *(Maestro-Desmond Interoperability Tools*, version 3.1.; Schrödinger: New York, NY, USA, 2020) with the OPLS-2005 force field parameter. Initially, the system builder module of Maestro was employed to configure the systems for the MD simulation. The protein–ligand complexes were placed within a cubic box, ensuring a solvent buffer of 2.5 Å using the SPC water model. Subsequently, the system was neutralized by adding NA + and CL- ions at a concentration of 0.15 M. Before proceeding with the final production run, the solvated system underwent a relaxation protocol implemented in Desmond, which involved a six-step procedure. Before thermalization, the protein–ligand complexes were subjected to 100 steps of steepest descent energy minimization to address any forces exceeding 100 kj/mol*nm. Following this, the system was simulated under the isothermal-isobaric (NPT) ensemble at a constant temperature of 300 K and pressure of 1 bar, using a time step of 2 fs. Subsequently, a simulation in the canonical ensemble NVT (molecules (N), volume (V), and temperature (T)) was conducted for 1 ns. Finally, the simulated conformers were equilibrated over a timescale of 100 ns.

Results and discussion

Selection of compounds

The downloaded ChEMBL compounds were chosen based on their bioactivity and their self-similarity, assessed through Tanimoto similarity. The bioactivity values for chosen parent compounds P1, P2, P3, P4, and P5 were 1750 nM, 3130 nM, 8250 nM, 191,000 nM, and 241,000 nM, respectively. The significant range of values, from high to low, signifies a diversity of activity among the initial compounds. This diversity can generate compounds exhibiting properties similar to the most active and least active compounds and

Table 1 Tanimoto similarity calculation results for the chosen parent compounds

S. No	Compounds	Tanimoto similarity
1	P1 & P2	AP Tanimoto: 0.827869 MCS Tanimoto: 0.9512 MCS Size: 39 MCS Min: 1.0000 MCS Max: 0.9512
2	P1 & P3	AP Tanimoto: 0.824249 MCS Tanimoto: 0.9070 MCS Size: 39 MCS Min: 0.9512 MCS Max: 0.9512
3	P1 & P4	AP Tanimoto: 0.557396 MCS Tanimoto: 0.7381 MCS Size: 31 MCS Min: 0.9688 MCS Max: 0.7561
4	P1 & P5	AP Tanimoto: 0.473005 MCS Tanimoto: 0.6905 MCS Size: 29 MCS Min: 0.9667 MCS Max: 0.7073
5	P2 & P3	AP Tanimoto: 0.827869 MCS Tanimoto: 0.9512 MCS Size: 39 MCS Min: 1.0000 MCS Max: 0.9512
6	P2 & P4	AP Tanimoto: 0.52716 MCS Tanimoto: 0.6905 MCS Size: 29 MCS Min: 0.9062 MCS Max: 0.7436
7	P2 & P5	AP Tanimoto: 0.533246 MCS Tanimoto: 0.7250 MCS Size: 29 MCS Min: 0.9667 MCS Max: 0.7436
8	P3 & P4	AP Tanimoto: 0.493757 MCS Tanimoto: 0.6591 MCS Size: 29 MCS Min: 0.9062 MCS Max: 0.7073
9	P3 & P5	AP Tanimoto: 0.473005 MCS Tanimoto: 0.6905 MCS Size: 29 MCS Min: 0.9667 MCS Max: 0.7073
10	P4 & P5	AP Tanimoto: 0.786948 MCS Tanimoto: 0.9375 MCS Size: 30 MCS Min: 1.0000 MCS Max: 0.9375

AP atom pairs, *MCS* maximum common substructure

those in between. We performed Tanimoto calculations to determine the similarity values between these compounds, as presented in Table 1.

Chemical language model generated compounds

In this study, we generated nearly 3000 compounds from an initial set of five compounds using the mollib package. To comprehensively evaluate the diversity, assess the physicochemical properties, and understand the relative positioning of these compounds within the chemical space, we employed a set of key metrics.

Frechet ChemNet Distance (FCD)

To assess the coverage of the chemical space during transfer learning, we utilized the feature-count distribution (FCD) metric, which measures the similarity between two sets of molecules based on their chemical structures [40]. A zero FCD value indicates complete identity between the compared molecular spaces, whereas higher values indicate greater dissimilarity. As depicted in Fig. 3, the CLM model demonstrated the ability to sample molecules from the chemical space, regardless of their similarity or dissimilarity to the source space, encompassing a wide range of bioactivities.

Furthermore, Fig. 3 reveals an initial distance between the source space (ChEMBL) and the target space right from the beginning of the epoch. This observation emphasizes the diversity of our initial five compounds, suggesting that they span distinct chemical space regions.

Fraction of sp^3-hybridized carbon atoms (FCsp^3)

To evaluate the alterations in physicochemical properties throughout transfer learning, we opted to examine the FC*sp3* parameter. FC*sp3* has been established as a predictor for a molecule's likelihood of becoming a drug and exhibits distinctions between synthetic compounds and natural products [41]. Throughout the transfer learning process, the FC*sp3* distribution approximated that of the transfer learning set distribution. Figure 4 illustrates that despite commencing with a limited assortment of similar and dissimilar compounds, the transfer learning process inherently captures pertinent physicochemical properties.

Visualization using uniform manifold approximation and projection (UMAP) plot

To depict the relative positioning of the computer-generated molecules within the chemical space, we employed the uniform manifold approximation and projection (UMAP) technique to generate UMAP plots [42]. UMAP facilitates the creation of a two-dimensional representation of high-dimensional data distributions, ensuring that the similarity relationships between data points in the original high-dimensional space are better preserved compared to

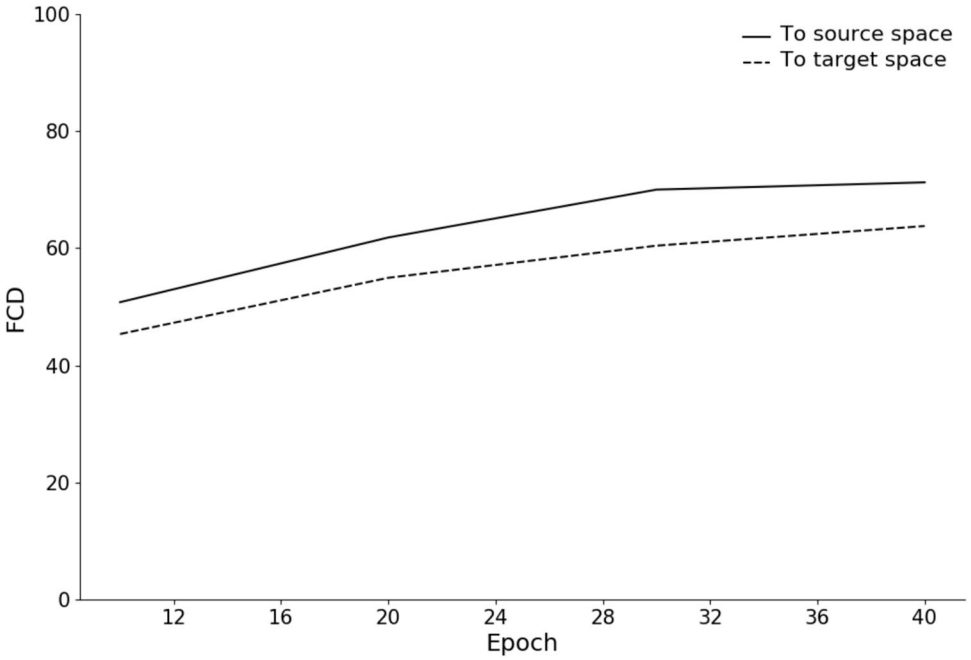


Fig. 3 FCD distance evolved continuously as a function of epochs

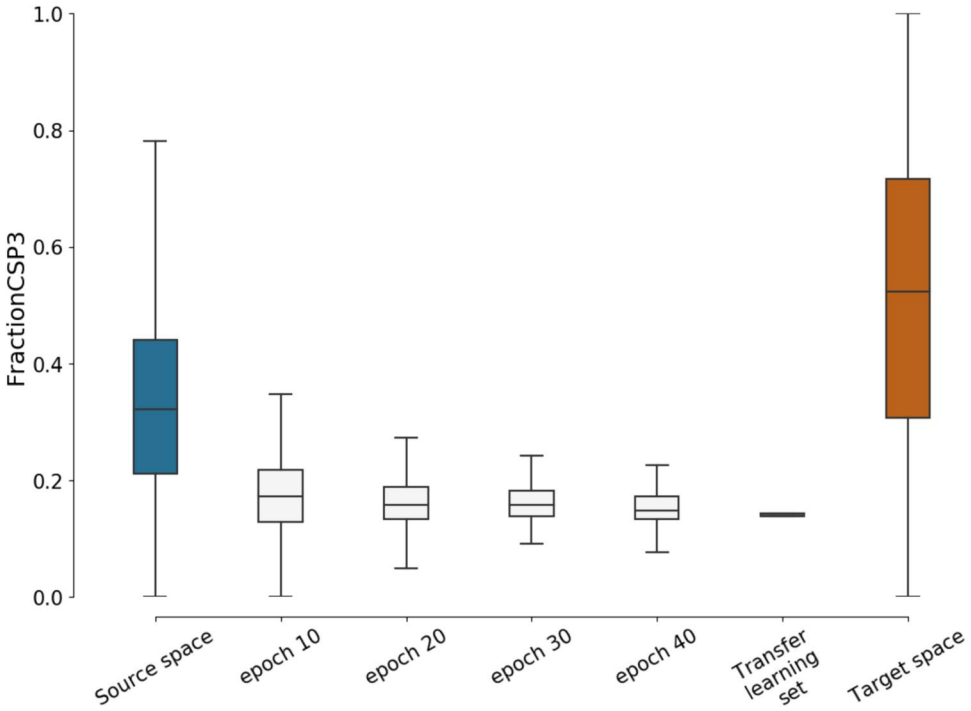


Fig. 4 The implicit capture of relevant physicochemical properties

t-distributed stochastic neighbor embedding (T-SNE). In this visualization, the generated molecules shifted toward the vicinity of the transfer learning set following the completion of transfer learning at epoch 40 (refer to Fig. 5).

Binding site and molecular docking studies

All the generated compounds at each epoch were combined and subjected to docking studies. We have used Cresset

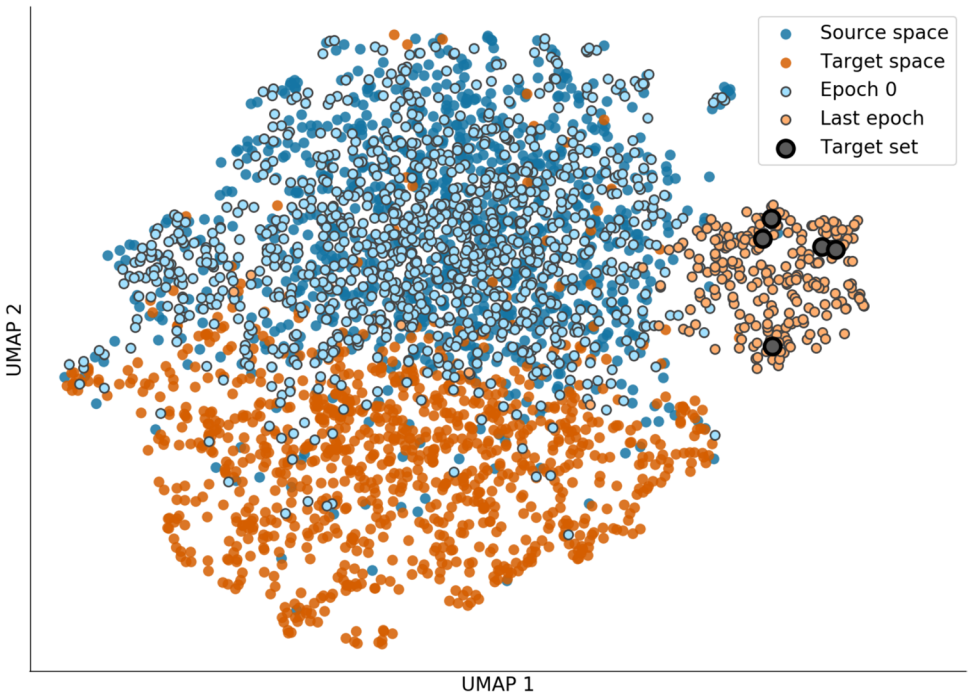


Fig. 5 The UMAP plot shows the relative placement of generated compounds in the chemical space following 40 epochs of transfer learning

software for docking. The docking revealed several compounds with higher docking scores than the control inhibitor. Supplementary Table 2S provides the docking scores and interacting residues for both the parent compounds and the top-scoring generated compound in complex with *Pf*LDH.

Parent compound 1 (P1) exhibited a docking score of -11.671 and formed hydrogen bonds with MET30, HIS195, ASN140, ARG109, and ARG171. It interacted hydrophobically with residues such as VAL26, ILE31, PHE52, ILE54, VAL55, ILE123, TYR85, ALA98, PHE100, ILE119, VAL138, LEU163, VAL142, TRP107, ALA236, PRO250, LEU167, LEU112, PRO246, and TYR247. Charged interactions with ASP53 and polar interactions with SER28, THR97, THR101, ASN116, THR139, and SER245 were also observed, along with other interactions involving GLY27, GLY29, and GLY99.

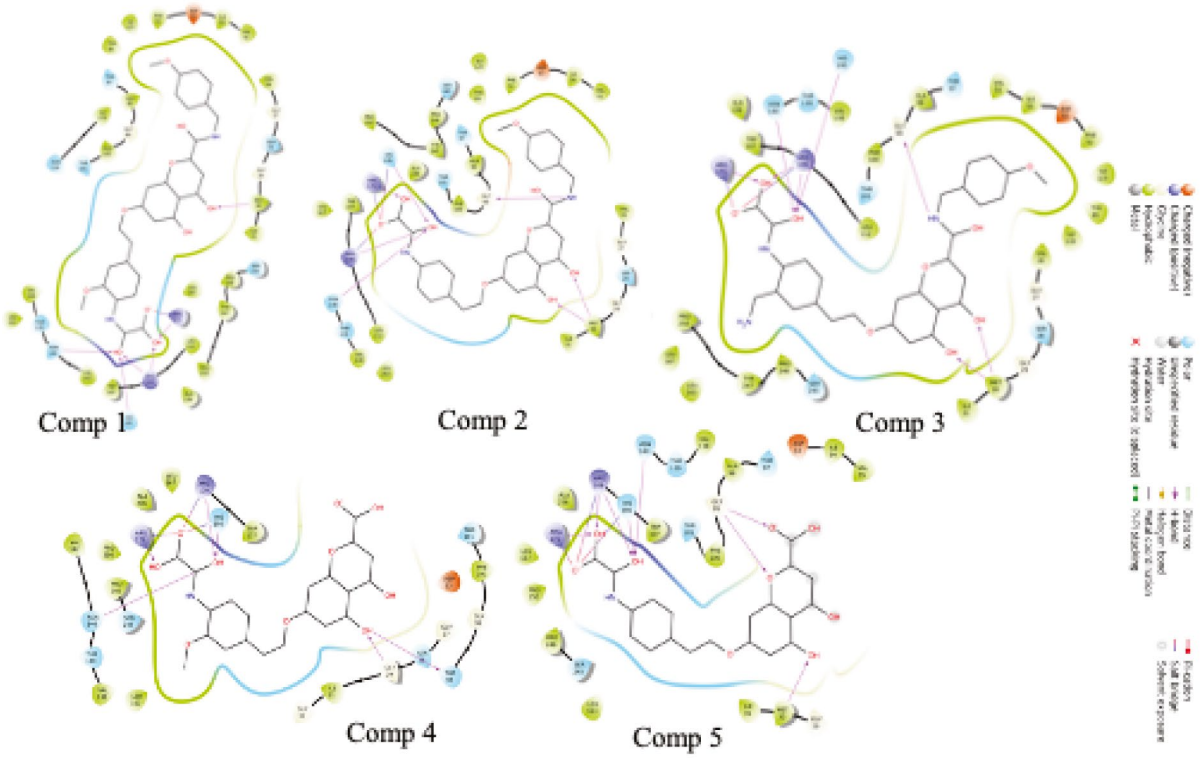
Parent compound 2 (P2), with a docking score of -11.403, formed hydrogen bonds with MET30, GLY99, HIS195, ASN140, and ARG171. Hydrophobic interactions were observed with VAL26, ILE31, PHE52, ILE54, VAL55, ILE123, TYR85, ALA98, PHE100, ILE119, VAL138, LEU163, TRP107, ALA236, PRO250, LEU167, and PRO246, TYR247. It also engaged in charged interactions with ARG109 (positive) and ASP53 (negative), polar interactions with SER28, THR97, THR101, ASN116, THR139, and SER245, and other interactions with GLY27 and GLY29.

Parent compound 3 (P3), with a docking score of -12.121 in the complex, formed hydrogen bonds with MET30, GLY99, ARG109, HIS195, ASN140, and ARG171. Hydrophobic interactions were observed with VAL26, ILE31, PHE52, ILE54, VAL55, ILE123, TYR85, ALA98, PHE100, ILE119, VAL138, LEU163, TRP107, ALA236, PRO250, LEU167, LEU112, PRO246, and TYR247. It engaged in charged interactions with ASP53 and polar interactions with SER28, THR97, THR101, THR139, and SER245, along with other interactions involving GLY27 and GLY29.

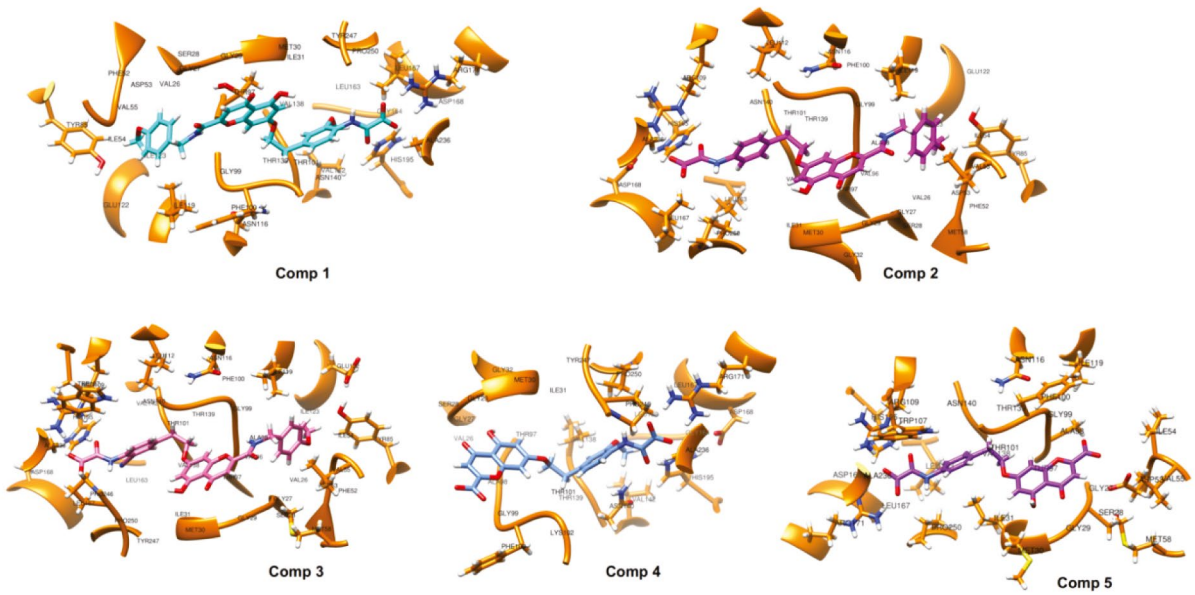
Parent compound 4 (P4) had a docking score of -9.934 and formed hydrogen bonds with GLY29, THR97, ARG109, HIS195, and ARG171. Hydrophobic interactions were observed with ILE31, PHE100, VAL138, LEU163, LEU167, TRP107, ALA236, PRO250, VAL142, and PRO246. It engaged in charged interactions with ASP53 and polar interactions with SER28, THR101, THR139, ASN140, and SER245, along with other interactions involving GLY27, GLY32, and GLY99.

Parent compound 5 (P5) exhibited a docking score of − 9.861 and formed hydrogen bonds with MET30, GLY99, ARG109, HIS195, ASN140, and ARG171. Hydrophobic interactions were observed with ILE31, ILE54, VAL55, ALA98, PHE100, VAL138, LEU163, TRP107, ALA236, PRO250, LEU167, and PRO246. It engaged in charged interactions with ASP53 and polar interactions with THR97, THR101, THR139, and SER245, along with other interactions involving GLY29.

We visually inspected the top 30 compounds and selected 10 that could form the bonds with most of the corresponding residues. Figure 6A and B depicts the 2D and 3D interaction plots of the five input molecules. Analysis of the compound



(A)



(B)

Fig. 6 Interaction plots of the five compounds used for the generation of a diverse library of compounds using mollib. **A** 2D interaction plot of the five input compounds. **B** 3D interaction plot of the five input compounds

203 revealed that it forms Hydrogen bond: ILE31, GLY99, ASN140, ARG109, ARG171 Hydrophobic: VAL26, MET30, PHE52, ILE54, VAL55, TYR85, ALA98, PHE100, LEU112, ILE119, ILE123, VAL138, TRP107, ALA236, PRO250, LEU167, PRO246, TYR247 Charged (positive): LYS118, ARG171 Charged (negative): ASP53, GLU122 Polar: SER28, THR97, THR101, THR139, SER245 also interactions with GLY27, GLY29, GLY32.

Compound 203 exhibited a docking score of -12.739 kcal/mol, which is lower than the docking scores of all the parent compounds. The parent compounds had docking scores ranging from -9.861 to -12.121 kcal/mol.

Based on the docking scores and the similarity in interacting residues, compound 203 demonstrates a superior binding affinity. This enhanced affinity may be attributed to the dioxacyclotetra group forming an additional hydrogen bond with the ILE31 residue of the target protein, a feature not observed in the parent compounds. This improved interaction is reflected in the lower docking score of -12.739 kcal/mol, suggesting that compound 203 has the potential to be a more effective inhibitor of *Pf*LDH.

Scaffolds analysis and Tanimoto coefficient calculation

On comparing the scaffold from the generated compounds with our input five compounds, we found a significant difference in diversity with a varied range of Shannon-scaled entropy (SSE) values at each epoch. The following equation is used to compute the SSE [43]:

$$SSE = \frac{-\sum_{i=1}^{n} Piln(\frac{ci}{P})}{\log_2(n)}$$

The numerator represents the Shannon entropy, where n is the number of unique scaffolds, ci is the count of compounds containing the ith scaffold, and P is the total number of compounds among the considered n scaffolds.

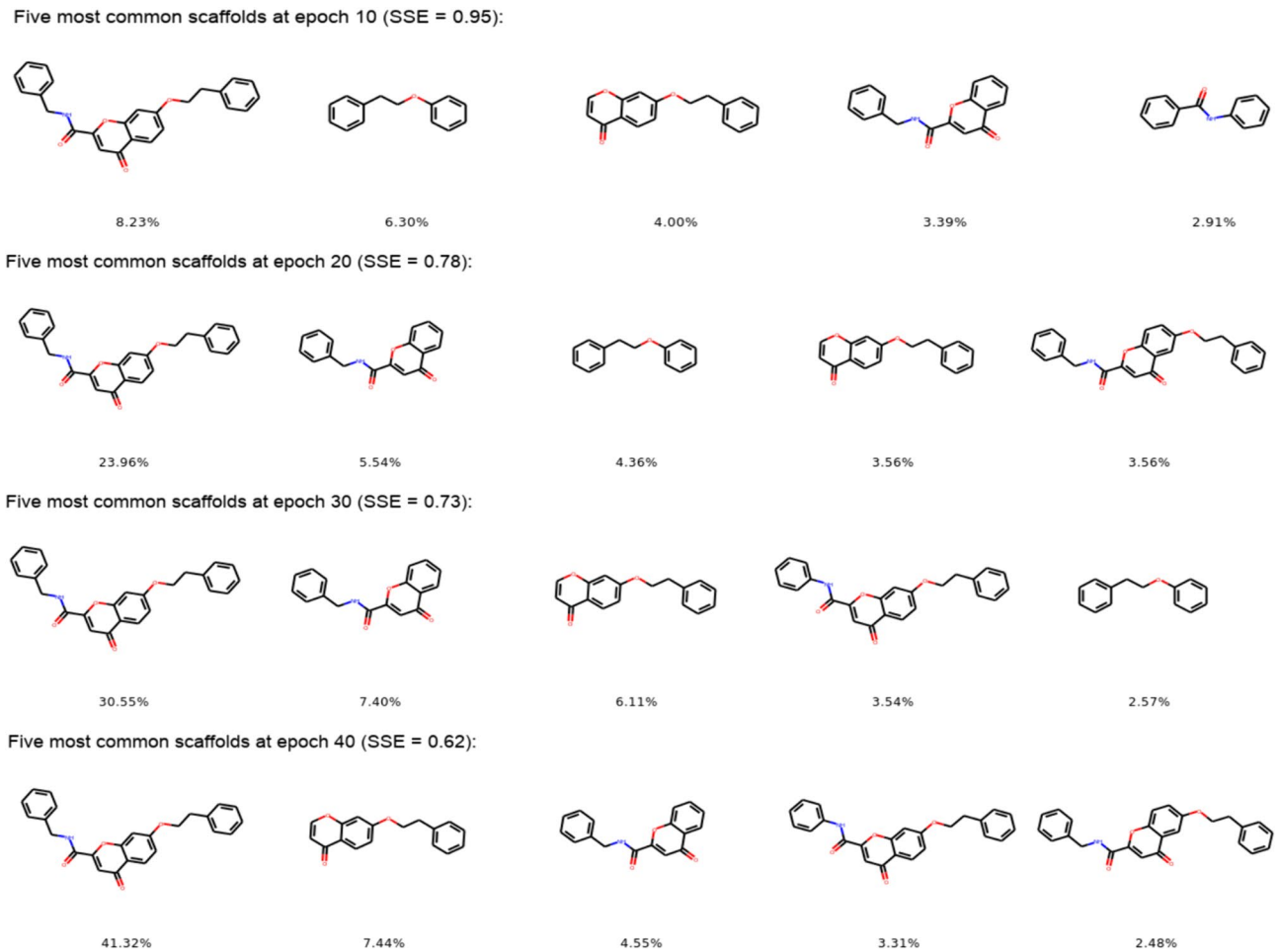


Fig. 7 The similarity of the five most common scaffolds at an interval of 10 epochs. The SSE value can be seen to start from 0.95 and drops to 0.62. A higher SSE value indicates more significant dissimilarity

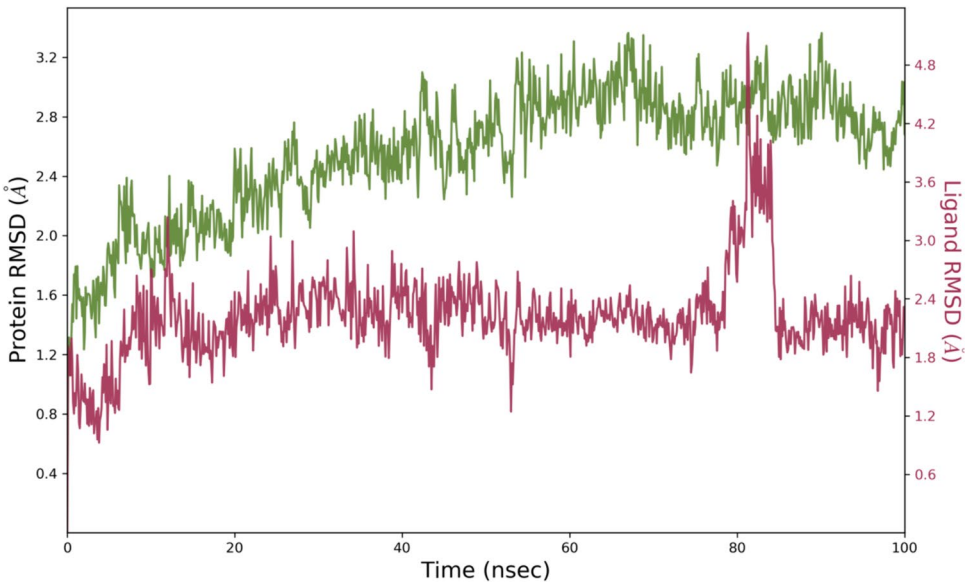


Fig. 8 The RMSD of the protein and ligand 203 during MD simulation. The lower RMSD of the ligand than that of the protein backbone indicates its stability inside the binding pocket

Figure 7 shows the significant percentage of the five most common scaffolds at each epoch. This further shows that the generated compounds are more diverse than the existing five potent compounds.

Further, to check the similarity/dissimilarity, we have calculated the Tanimoto coefficient of the top 10 generated compounds with the five input compounds as shown in Table 1S. Several initially promising compounds with novel scaffolds exhibited high RMSD values after molecular dynamics simulations. This suggests that their structural flexibility or interactions with the *Pf*LDH protein might not be optimal for long-term stability, potentially hindering their effectiveness as drug candidates. On comparing input compounds 1, 2, 3, 4, and 5 with the generated compound number 203, we obtained the Tanimoto values as 0.48, 0.44, 0.43, 0.36, and 0.31, respectively. This shows that the generated compound 203 is different from the input compounds and its RMSD values fall within acceptable ranges during the molecular dynamics simulation, indicating good potential stability. The table for the top 10 generated compounds with their Tanimoto values with each of the input compounds is shown in Supplementary Table 1S.

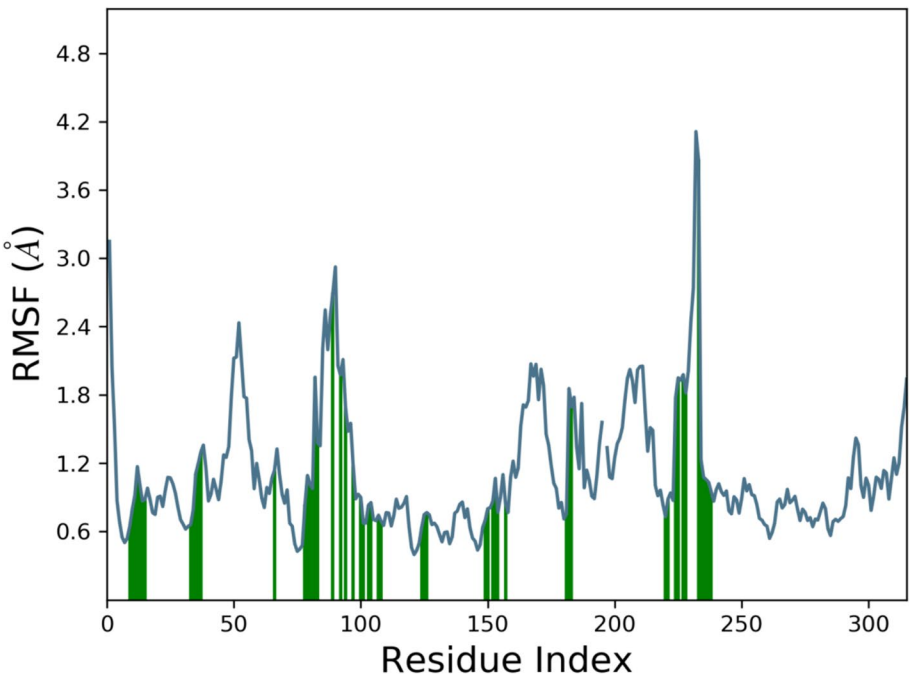


Fig. 9 RMSF of the protein residues reflects the overall stability of the protein during the MD simulation of its complex with compound 203. Protein residues that interact with the ligand are marked with green-colored vertical bars

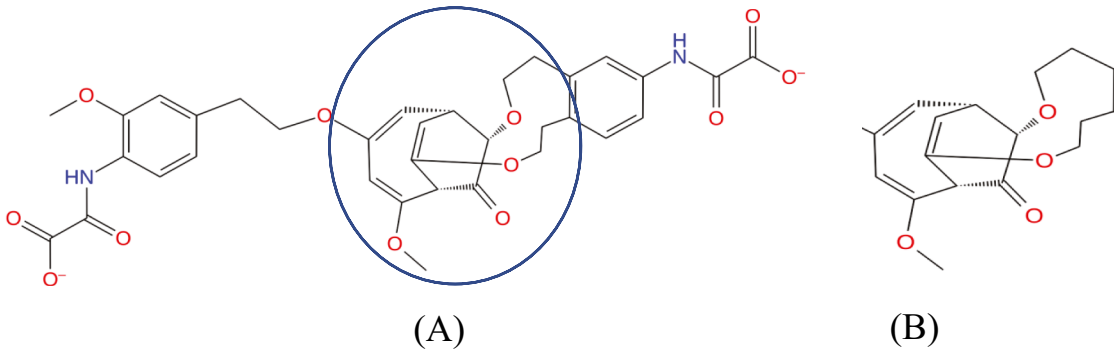


Fig. 10 **A** The generated compound 203 encircling the dioxatetracyclo group. **B** Dioxatetracyclo group

Molecular dynamics studies

Further, we have performed the MD simulation for 100 ns using the Desmond program to check the generated compounds' stability and interaction pattern. This also helped study the conformational changes of the generated compound with respect to the active site, along with the motion of the residues of the *Pf*LDH protein. During the time period of simulation, a protein structure should converge, and the overall system's stability is measured by analyzing its RMSD. The trajectories of the start phase of the simulation were superimposed onto the production phase to analyze the progression of RMSD throughout the simulation. The RMSD of the protein backbone compared with the ligand is shown in Fig. 8. The higher RMSD of the protein backbone than the ligand RMSD shows that the ligand is stable in the protein's binding pocket. This further reflects no significant movement of the docked ligand onto the protein's active site. The protein stability was also analyzed using Root Mean Square Fluctuations (RMSF) of individual *Pf*LDH amino acid residues. Figure 9 indicates the RMSF of the protein, and the peaks indicate the most protein residues are in stable conformation during the MD simulation.

Compound 203 is a potential chemotype PfLDH inhibitor

The similarity scores between compound 203 and five known inhibitors were calculated using the Tanimoto coefficient and are shown in Table 1. The Tanimoto coefficients between compound 203 and the available inhibitors range from 0.316765 to 0.487023. These low

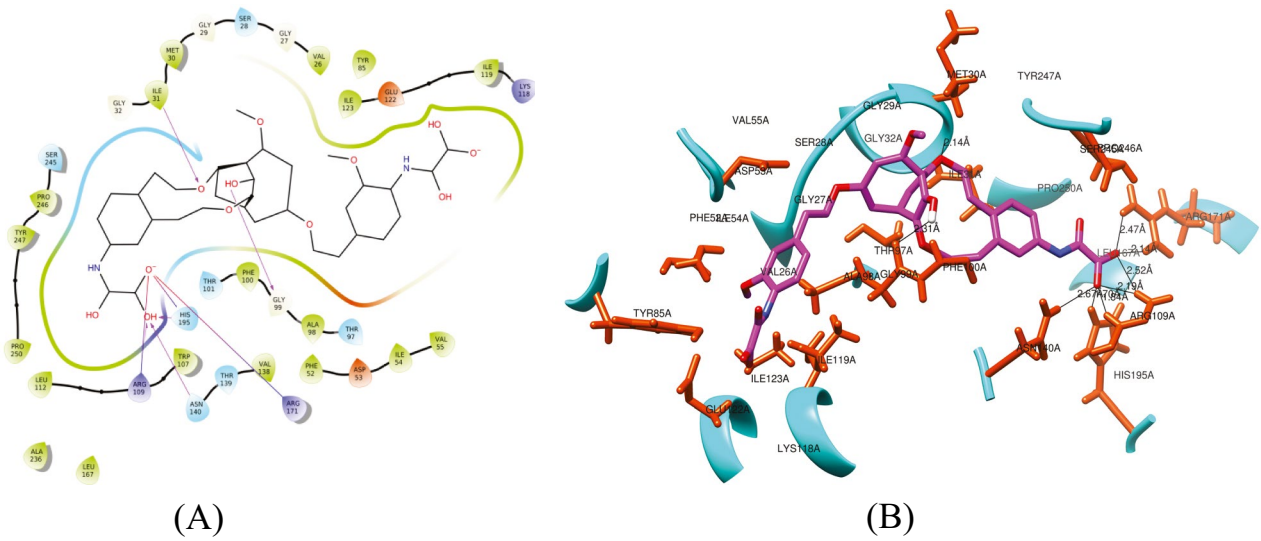


Fig. 11 **A** A 2D interaction plot of generated compound 203. **B** A 3D interaction plot of generated compound 203

Tanimoto coefficient values (< 0.5) indicate that the generated compound 203, which contains a dioxatetracyclo group, is dissimilar from the known inhibitors. As shown in Fig. 10A, compound 203 encircles the dioxatetracyclo group. Figure 10B provides a closer look at the isolated dioxatetracyclo group. Likewise, Fig. 11A depicts the 2D interaction between compound 203 and the binding pocket, and Fig. 11B zooms in 3D interaction plot of generated compound 203. Therefore, compound 203 represents a new chemotype that could be further optimized.

Conclusion

The generative deep learning technique can diversify medicinal chemistry chemical space by offering a comprehensive method of chemical transformations through structural enumeration and combinatorial sampling. This research demonstrates that deep learning models facilitate the identification of novel molecular entities tailored to a specific drug target.

In this investigation, the model based on SMILES representations effectively generated diverse molecules with the desirable characteristics, bridging the gap between synthetic compounds and natural products. Utilizing the principle of chemical similarity and *Pf*LDH inhibitors as initial reference points for drug development, this computational method successfully produced one unique and chemically diverse molecular entity that can be further refined into a potential drug-like molecule.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11030-024-10960-3.

Authors' contributions JK, NC, and DG contributed equally to this work, including conceptualization and methodology. JK and NC performed data analysis and interpretation. They also conducted the experimental procedures collaboratively. DG, JK, and NC contributed to the writing and revision of the manuscript. DG provided critical guidance and supervision throughout the study, contributing to the overall direction and quality of the research. All authors reviewed and approved the final version of the manuscript.

Funding This work was undertaken at the "Bioinformatics and Computational Biology Centre for high throughput data analysis," supported by the Department of Biotechnology, Government of India under grant No. BT/PR40151/BTIS/137/5/2021.

Data availability The dataset utilized in this study is available on the ChEMBL database (www.ebi.ac.uk/chembl).

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Krostad DJ (1996) Malaria as a reemerging disease. Epidemiol Rev 18:77–89. https://doi.org/10.1093/oxfordjournals.epirev.a017918
2. Greenwood B, Mutabingwa T (2002) Malaria in 2002. Nature 415:670–672. https://doi.org/10.1038/415670a
3. Blasco B, Leroy D, Fidock DA (2017) Antimalarial drug resistance: linking *Plasmodium falciparum* parasite biology to the clinic. Nat Med 23:917–928. https://doi.org/10.1038/nm.4381
4. Haldar K, Bhattacharjee S, Safeukui I (2018) Drug resistance in Plasmodium. Nat Rev Microbiol 16:156–170. https://doi.org/10.1038/nrmicro.2017.161
5. Tilley L, Straimer J, Gnädig NF et al (2016) Artemisinin action and resistance in *Plasmodium falciparum*. Trends Parasitol 32:682–696. https://doi.org/10.1016/j.pt.2016.05.010
6. Fairhurst RM, Dondorp AM (2016) Artemisinin-resistant *Plasmodium falciparum* Malaria. Microbiol Spectrosc. https://doi.org/10.1128/microbiolspec.EI10-0013-2016
7. Ménard S, Ben Haddou T, Ramadani AP et al (2015) Induction of multidrug tolerance in *Plasmodium falciparum* by Extended artemisinin pressure. Emerg Infect Dis 21:1733–1741. https://doi.org/10.3201/eid2110.150682
8. Hanboonkunupakarn B, White NJ (2016) The threat of artemisinin resistant malaria in Southeast Asia. Travel Med Infect Dis 14:548–550. https://doi.org/10.1016/j.tmaid.2016.11.016
9. Penna-Coutinho J, Cortopassi WA, Oliveira AA et al (2011) Antimalarial activity of potential inhibitors of *Plasmodium falciparum* lactate dehydrogenase enzyme selected by docking studies. PLoS ONE 6:e21237. https://doi.org/10.1371/journal.pone.0021237
10. Granchi C, Bertini S, Macchia M, Minutolo F (2010) Inhibitors of lactate dehydrogenase isoforms and their therapeutic potentials. Curr Med Chem 17:672–697. https://doi.org/10.2174/092986710790416263
11. Shoemark DK, Cliff MJ, Sessions RB, Clarke AR (2007) Enzymatic properties of the lactate dehydrogenase enzyme from *Plasmodium falciparum*. FEBS J 274:2738–2748. https://doi.org/10.1111/j.1742-4658.2007.05808.x
12. Wiwanitkit V (2007) Plasmodium and host lactate dehydrogenase molecular function and biological pathways: implication for antimalarial drug discovery. Chem Biol Drug Des 69:280–283. https://doi.org/10.1111/j.1747-0285.2007.00495.x
13. Laganá G, Barreca D, Calderaro A, Bellocco E (2019) Lactate dehydrogenase inhibition: biochemical relevance and therapeutical potential. Curr Med Chem 26:3242–3252. https://doi.org/10.2174/0929867324666170209103444
14. Fokou PVT, Tali BMT, Dize D et al (2022) Implementation and continued validation of the malaria *Plasmodium falciparum* lactate dehydrogenase-based colorimetric assay for use in antiplasmodial drug screening. Anal Biochem 648:114669. https://doi.org/10.1016/j.ab.2022.114669
15. Singh R, Bhardwaj V, Purohit R (2021) Identification of a novel binding mechanism of Quinoline based molecules with lactate dehydrogenase of *Plasmodium falciparum*. J Biomol Struct Dyn 39:348–356. https://doi.org/10.1080/07391102.2020.1711809
16. Joshi N, Hada R, Gupta S et al (2022) Highly potent antimalarial activity of benzopyrano(4,3-b)benzopyran derivatives and in silico interaction analysis with putative target *Plasmodium falciparum* lactate dehydrogenase. J Biomol Struct Dyn 40:5159–5174. https://doi.org/10.1080/07391102.2020.1868336
17. Tahghighi A, Mohamadi-Zarch S-M, Rahimi H et al (2020) In silico and in vivo anti-malarial investigation on 1-(heteroaryl)-2-((5-nitroheteroaryl)methylene) hydrazine derivatives. Malar J 19:231. https://doi.org/10.1186/s12936-020-03269-7

18. Saxena S, Durgam L, Guruprasad L (2019) Multiple e-pharmacophore modelling pooled with high-throughput virtual screening, docking and molecular dynamics simulations to discover potential inhibitors of *Plasmodium falciparum* lactate dehydrogenase (PfLDH). J Biomol Struct Dyn 37:1783–1799. https://doi.org/10.1080/07391102.2018.1471417
19. Shamsuddin MA, Ali AH, Zakaria NH et al (2021) Synthesis, molecular docking, and antimalarial activity of hybrid 4-aminoquinoline-pyrano[2,3-c]pyrazole derivatives. Pharmaceuticals (Basel). https://doi.org/10.3390/ph14111174
20. Rajesh K, Lavanya P, Iniyavan P et al (2015) Regioselective synthesis of 2-chloroquinoline based ethyl 4-(3-hydroxyphenyl)-2,7,7-trimethyl-5-oxo-1,4,5,6,7,8-hexahydroquinoline-3- carboxylates and their in-silico evaluation against *P. falciparum* lactate dehydrogenase. Med Chem 11:789–797. https://doi.org/10.2174/1573406411666150430132055
21. Ramli AH, Swain P, Mohd Fahmi MSA et al (2024) Preliminary insight on diarylpentanoids as potential antimalarials: in silico, in vitro pLDH and in vivo zebrafish toxicity assessment. Heliyon 10:e27462. https://doi.org/10.1016/j.heliyon.2024.e27462
22. Read JA, Wilkinson KW, Tranter R et al (1999) Chloroquine binds in the cofactor binding site of *Plasmodium falciparum* lactate dehydrogenase. J Biol Chem 274:10213–10218
23. Surolia I (2000) Chloroquine binds in the cofactor binding site of *Plasmodium falciparum* lactate dehydrogenase—a response. Parasitol Today 16:133. https://doi.org/10.1016/s0169-4758(99)01552-5
24. Gomez MS, Piper RC, Hunsaker LA et al (1997) Substrate and cofactor specificity and selective inhibition of lactate dehydrogenase from the malarial parasite *P. falciparum*. Mol Biochem Parasitol 90:235–246. https://doi.org/10.1016/s0166-6851(97)00140-0
25. Cameron A, Read J, Tranter R et al (2004) Identification and activity of a series of azole-based compounds with lactate dehydrogenase-directed anti-malarial activity. J Biol Chem 279:31429–31439. https://doi.org/10.1074/jbc.M402433200
26. Mendez D, Gaulton A, Bento AP et al (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47:D930–D940. https://doi.org/10.1093/nar/gky1075
27. Davies M, Nowotka M, Papadatos G et al (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res 43:W612–W620. https://doi.org/10.1093/nar/gkv352
28. Subramaniam S, Mehrotra M, Gupta D (2011) Support vector machine based classification model for screening *Plasmodium falciparum* proliferation inhibitors and non-inhibitors. Biomed Eng Comput Biol 3:BECB.S7503. https://doi.org/10.4137/becb.s7503
29. Stokes JM, Yang K, Swanson K et al (2020) A deep learning approach to antibiotic discovery. Cell 181:475–483. https://doi.org/10.1016/j.cell.2020.04.001
30. Mswahili ME, Martin GL, Woo J et al (2021) Antimalarial drug predictions using molecular descriptors and machine learning against plasmodium falciparum. Biomolecules 11:1–15. https://doi.org/10.3390/biom11121750
31. Egieyeh S, Syce J, Malan SF, Christoffels A (2018) Predictive classifier models built from natural products with antimalarial bioactivity using machine learning approach. PLoS ONE 13:1–15. https://doi.org/10.1371/journal.pone.0204644
32. Liu Q, Deng J, Liu M (2020) Classification models for predicting the antimalarial activity against *Plasmodium falciparum*. SAR QSAR Environ Res 31:313–324. https://doi.org/10.1080/1062936X.2020.1740890
33. Danishuddin MG, Malik MZ, Subbarao N (2019) Development and rigorous validation of antimalarial predictive models using machine learning approaches. SAR QSAR Environ Res 30:543–560. https://doi.org/10.1080/1062936X.2019.1635526
34. Amod L, Mohunlal R, Teixeira N et al (2023) Identifying inhibitors of β-haematin formation with activity against chloroquine-resistant *Plasmodium falciparum* malaria parasites via virtual screening approaches. Sci Rep 13:2648. https://doi.org/10.1038/s41598-023-29273-w
35. Moret M, Friedrich L, Grisoni F et al (2020) Generative molecular design in low data regimes. Nat Mach Intell 2:171–180. https://doi.org/10.1038/s42256-020-0160-y
36. Choi SR, Pradhan A, Hammond NL et al (2007) Design, synthesis, and biological evaluation of *Plasmodium falciparum* lactate dehydrogenase inhibitors. J Med Chem 50:3841–3850. https://doi.org/10.1021/jm070336k
37. Backman TWH, Cao Y, Girke T (2011) ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Res 39:W486–W491. https://doi.org/10.1093/nar/gkr320
38. Dunn CR, Banfield MJ, Barker JJ et al (1996) The structure of lactate dehydrogenase from *Plasmodium falciparum* reveals a new target for anti-malarial design. Nat Struct Biol 3:912–915. https://doi.org/10.1038/nsb1196-912
39. Cortopassi WA, Oliveira AA, Guimarães AP et al (2011) Docking studies on the binding of quinoline derivatives and hematin to Plasmodium falciparum lactate dehydrogenase. J Biomol Struct Dyn 29:207–218. https://doi.org/10.1080/07391102.2011.10507383
40. Preuer K, Renz P, Unterthiner T et al (2018) Fréchet ChemNet distance: a metric for generative models for molecules in drug discovery. J Chem Inf Model 58:1736–1741. https://doi.org/10.1021/acs.jcim.8b00234
41. Stratton CF, Newman DJ, Tan DS (2015) Cheminformatic comparison of approved drugs from natural product versus synthetic origins. Bioorg Med Chem Lett 25:4802–4807. https://doi.org/10.1016/j.bmcl.2015.07.014
42. McInnes L, Healy J, Melville J (2018) UMAP: uniform manifold approximation and projection for dimension reduction
43. Medina-Franco JL, Martínez-Mayorga K, Bender A, Scior T (2009) Scaffold diversity analysis of compound data sets using an entropy-based measure. QSAR Comb Sci 28:1551–1560. https://doi.org/10.1002/qsar.200960069

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Jitendra Kuldeep[1] · Neeraj Chaturvedi[1] · Dinesh Gupta[1]

✉ Dinesh Gupta
dinesh@icgeb.res.in

[1] Translational Bioinformatics Group, International Center for Genetic Engineering and Biotechnology (ICGEB), New Delhi 110067, India