Computational Drug Design Targeting Isocitrate Lyase 1 in *Mycobacterium Tuberculosis*

Sherry Zhang[1,2], Wei-Tse Hsu[1], Linda X. Phan[1], Philip C. Biggin[1]
1 Department of Biochemistry, University of Oxford
2 Massachusetts Institute of Technology

Abstract

Despite extensive drug-development efforts targeting *Mycobacterium tuberculosis* (Mtb), tuberculosis (TB) remains a leading cause of death worldwide. One way to combat TB is to target isocitrate lyase 1 (ICL1), an enzyme involved in the glyoxylate shunt in Mtb. Although ICL1 is an attractive drug target, previous efforts to inhibit ICL1 have not yielded any viable drugs. ICL1 is known to form a dimer and tetramer, which play roles in regulating catalytic activity. Here we used a computational drug design pipeline and molecular dynamics to find new compounds to target the central cavity of the ICL1 dimer. By targeting the central cavity, we can potentially inhibit ICL1 by altering the dimer structure, which could disrupt tetramerization and allosterically reduce catalytic activity. In addition, the central cavity is fairly large and deep, making it a desirable drug binding site. Targeting this cavity, rather than the conserved catalytic site, has the potential of achieving high specificity and reducing off-target side effects. Through this approach we obtained multiple candidates that will be pursued in future studies.

Introduction

The World Health Organization estimates that around 1.2 million people die from TB yearly (https://www.who.int/news-room/fact-sheets/detail/tuberculosis). TB is currently curable but the treatment requires prolonged multidrug antibiotic regimens, and the emergence of drug-resistant Mtb strains remain major challenges[1].

One potential way to combat TB is to inhibit ICL1 and ICL2, which are necessary for the survival of TB in the host, as they jointly mediate the fatty acid catabolism. Studies have shown that inhibiting ICL1 and ICL2 together can completely eliminate TB from mice[2].

ICL1 is a protein of 428 amino acids. Two monomers associate to form a bean-shaped dimer, with the catalytic site located at the center of each monomer (Fig 1a). Previous structural studies have shown that ICL1 often forms a tetramer, made up of two dimers (Fig 1b). These observations suggest that the catalytic activity may be dependent on tetramerization, although it is not known whether the dimeric form of ICL1 is completely or partially inactive. Besides the catalytic site, preliminary work in our lab identified a cavity at the dimer interfaces, which is referred to as the central cavity (Fig 1b).

Previous efforts to inhibit ICL1 have largely focused on targeting the active site, of which, many were not successful in the long term[3]. This is because any potential drugs were non-selective, raising concerns about toxicity in humans[3]. Further, any attempts to screen for potential

compounds was very difficult because of the small size of the ICL1 active site pocket and the fact that the active site of ICL1 is polar, not ideal for high-affinity binding of drugs.

Thus, we have decided to target the central cavity of ICL1. Binding of a compound at this site can potentially disrupt the dimerization and tetramerization of ICL1. One benefit of targeting the central cavity is that it is much larger and more hydrophobic, therefore making it potentially more druggable than the active site. The central cavity, according to DogSiteScorer[4], is 629.8 $Å^3$ vs the active site which is 347.8 $Å^3$. DogSiteScorer also found that the druggability of the central pocket was 0.8, while the active site only scored 0.27. We reasoned that compounds that bind to the central cavity may inhibit ICL1 by disrupting the tetramerization and/or allosterically altering the catalytic site[5]. Compounds that bind the central cavity could be large enough to physically block tetramerization, or possess electrostatic properties that disfavor the packing of two dimers into the tetramer.

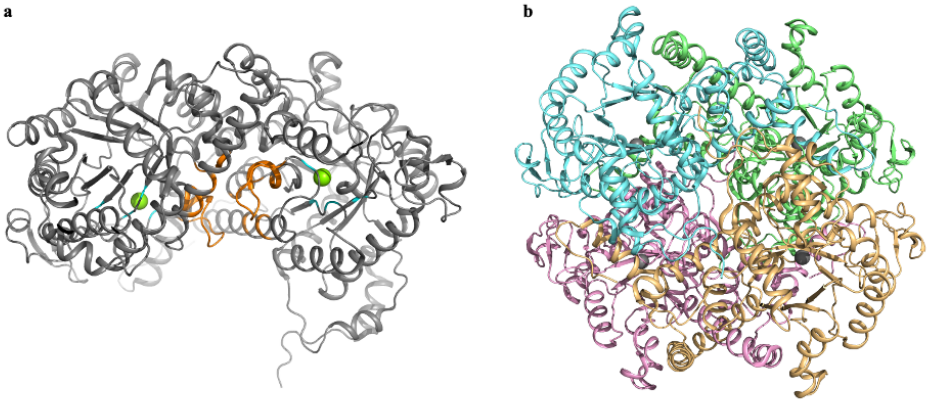
Since there were no other pockets of sufficient size to target, we decided to proceed with drug design for the central cavity under the hope that we would be able to accomplish either of these goals.

Results

System setup

We used the structure of ICL1 from (PDB ID: 1f61) for our simulations[6]. As mentioned above, the ICL1 protein exists as a tetramer formed by two identical copies of the dimer. We decided to run docking experiments targeted at the central cavity of ICL1 and molecular dynamics simulations using the dimer. We reasoned that the central cavity is located near the dimer interface and the tetramer interface, which indicates that large compounds that bind to the central cavity of the dimer will very likely block the tetramerization. To be effective at disrupting the tetramer format, the compounds must be able to bind stably to the dimer. In addition, it is easier to analyze the potential allosteric effects of the compounds on the catalytic site in the dimer.

Figure 1: ICL1 Dimer and Tetramer



a, ICL1 dimer. Active site residues in cyan, central cavity residues in orange. **b**, ICL1 tetramer. Each monomer in own color, active sites are around gray magnesium ions.

To assess druggability of the central cavity, previous work in the Biggin lab used three software packages: fpocket, DoGSiteScorer, PockDrug-Server[7,8]. Each of these packages gave different druggability scores and sizes of the pocket, but the consensus across them was that the central cavity was larger than the active site pocket and that druggability of the central cavity was, at worst, very similar to the active site, but likely much better (Fig 2). The size of the pocket and score reported by fpocket tend to deviate from those of the other packages because fpocket often splits 1 pocket into multiple smaller sub-pockets.

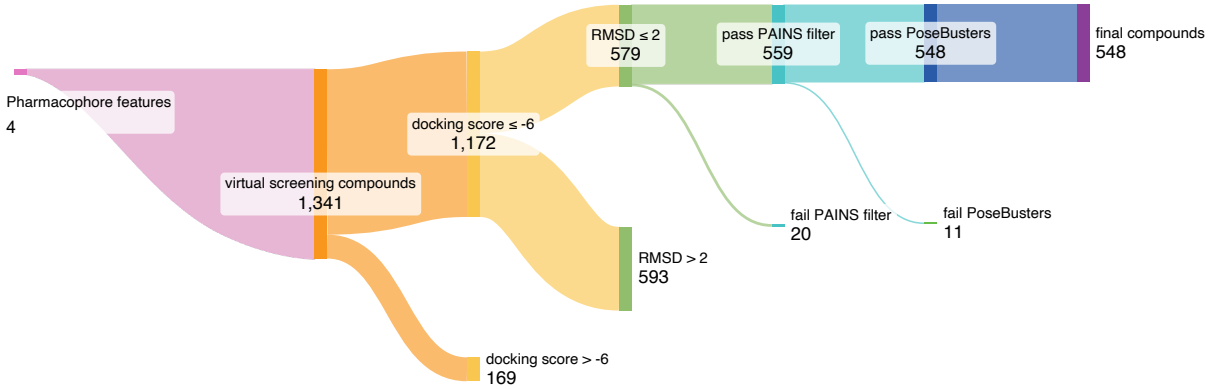
Figure 2: Druggability of Central Cavity and Active Site

		Druggability	Volume (Å^3)
Central Cavity	fpocket	0.007, 0.011	957.6
	DGSS	0.8	629.8
	PDS	0.58	1806.3
Active Site	fpocket	0.005, 0.020	504.7
	DGSS	0.27	347.8
	PDS	0.97	539.2

Compound selection

We started with 1,341 compounds from virtual screening and narrowed down to 548 final compounds after filtering described in the methods section below (Fig 3). From these compounds we selected based upon AEV-PLIG, a machine-learning based scoring function for binding affinity prediction, and the pose of the compounds[9]. For some compounds, in order to maximize their docking score, they were bent into unnatural shapes during the docking process, which were unlikely to occur naturally since we were quite lenient with the number of rotational bonds allowed. Thus, these compounds were discarded.

Figure 3: Compound Filtering Procedure



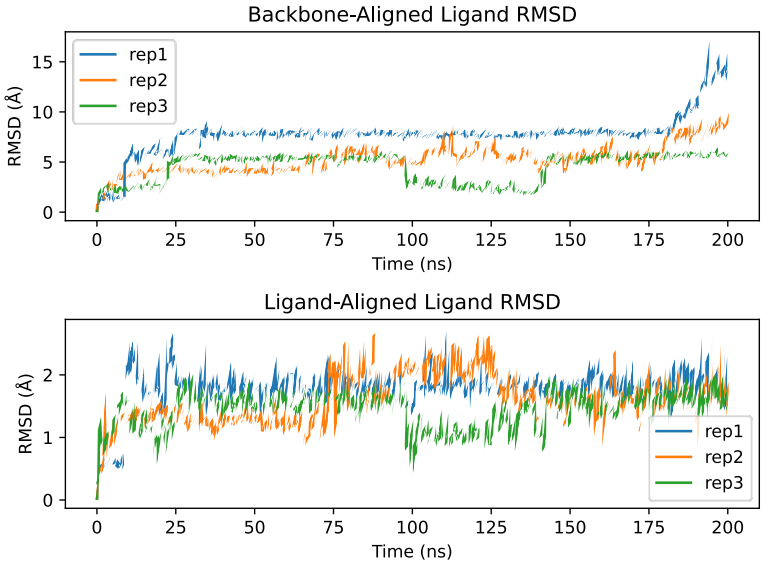
After selecting a final compound, we proceeded to run three replica of 200 ns molecular dynamics simulations to visualize and analyze the movement of the compound in the pocket. Initially, the compound has around 3 hydrogen bonds with the protein, which is a sufficient but improvable amount for stability.

In replicate 1, the compound seems to escape the pocket at around 180 ns, suggesting that the binding may not be very stable. But before that, the compound displayed some promising features. For a majority of the simulation, the small tail of the compound remained in the pocket and the rest of the compound stuck out of the pocket which could potentially physically prevent tetramerization.

In replicates 2 and 3, the compound stays in the pocket during the entire trajectory, as indicated by the low RMSD value shown in the figure below (Fig 4). In replicate 2, approximately 56% of the surface area of the compound was buried in the central cavity from frame 817 until the end of the simulation (around 59% of the simulation), which is a positive sign in terms of blocking tetramerization. On the other hand, for replicate 3, only a small portion of the compound stuck out of the compound. During the entire simulation, over 73% of the surface area of the compound was buried in the protein.

One positive aspect is that the ligand's internal RMSD values were quite low. The RMSD values for all 3 reproductions fluctuated around 2 Å which shows that the ligand conformation remains stable throughout each of the trajectories.

Figure 4: RMSD Plots



Analysis of the allosteric network between the central cavity and catalytic site

We initially hypothesized that the binding of compounds to the central cavity of ICL1 may inhibit the catalytic activity through an allosteric effect based on the short distance between the two sites. To more rigorously analyze this possibility, we carried out allosteric network analysis in the ICL1 dimer using two software packages, TEntropy and MDPath[10,11].

TEntroPy uses transfer entropy from molecular dynamics trajectories to calculate information flow between residues to indentify communication paths between functional sites. MDPath uses

correlations in residue φ-ψ angle movements to identify allosteric communication pathways through the protein.

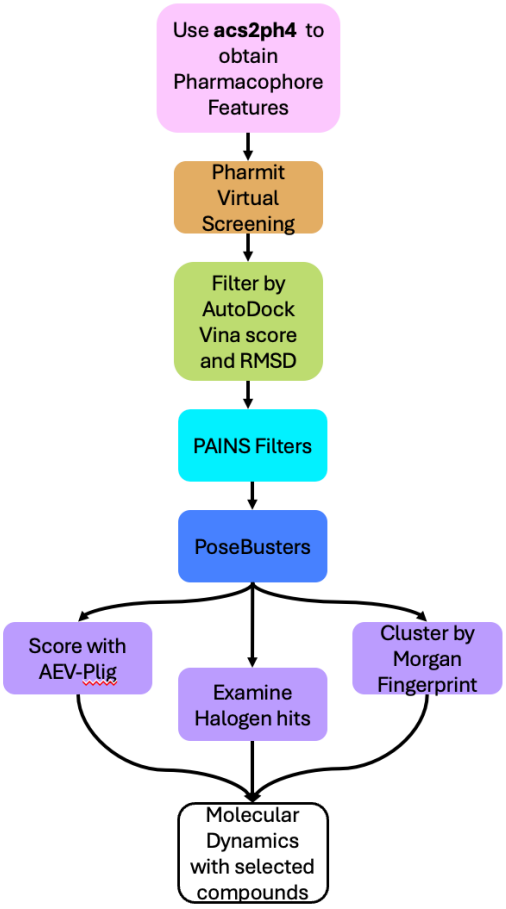
Both showed mixed results. For MDPath, one path from the central cavity to the active site had a score of 5.06, which was on the fairly lower side of all paths, ranking 378 out of the 500 that MDPath saves. For TEntropy, we found that there was transfer entropy between the central cavity and active site. However over multiple replicates, the hot spot residues were not conserved, suggesting that the indicated relationship may not be very robust.

Methods

Compound selection

We first used acs2ph4, a custom package written in our lab, to convert hot spots from E-FTMap solvent mapping analysis into pharmacophore features that are compatible with Pharmit[12,13] (Fig 5). We selected three to five of the 38 pharmacophore features generated by acs2ph4 to build each pharmacophore models, which was then used to screen against ZINC using Pharmit[14]. We set an exclusive shape tolerance of 1.0 on the receptor. Under the "Hit Reduction" section we limited maximum hits per conformation to 10, maximum hits per molecule to 1, and no cap on maximum total hits. For hit screening, we limited the molecular weight to less than or equal to 500 Da. We initially used less than or equal to 10 rotational bonds but later lowered it to 5, LogP less than or equal to 5, maximum polar surface area to less than or equal to 140, less than or equal to 4 aromatics, less than or equal to 10 hydrogen bond acceptors, and less than or equal to 5 hydrogen bond donors.After compounds from Pharmit were obtained, we used Smina (https://github.com/quantori/scip-smina) to minimize the energy of the compounds relative to the central cavity. Smina outputs a sdf file with each compound and includes an RMSD value for the minimized vs unminimized position. Using a Jupyter notebook, we removed compounds with a docking score (generated by Smina, based on AutoDock Vina[15]) that were greater than -6 kcal/mol and any compounds with an RMSD greater than 2 Å. From there, we ran PAINS filters as a part of RDKit (https://github.com/rdkit/rdkit), which screen for compounds that are likely to give false positives in biological assays[16]. To further filter, we used PoseBusters to eliminate compounds which are unlikely to occur physically due to geometric inconsistencies like incorrect bond angles and internal ligand

Figure 5: Drug Design Workflow



clashes[17]. This is our final filter, from which we then manually select compounds to proceed with.

In order to select which compounds to perform molecular dynamics on, we considered a variety of factors. First, whether or not the compounds matched with any halogen hot spots. Since Pharmit does not consider halogen pharmacophore models, of which we had already found from apo-site pharmacophore modeling using PharmacoNet, we looked at the compounds that we had that would match with the halogen hotspots found[18]. We also clustered all the compounds by their Morgan Fingerprint (radius of 2, fpSize of 2048) so we could proceed with compounds that represented a large portion of our options[19]. Finally, we considered predicted binding affinity scores from AEV-PLIG and aimed to choose the compounds with the highest scores[9,20,21]. The AEV-PLIG binding affinity scores had a very narrow score range (4 to 6), which might be because our compounds are unlikely to dock well or because the training data was not very representative. Because of this, we were unsure about how reliable the AEV-PLIG scores were but decided to proceed with the top scorers.

Molecular dynamics simulations

From there, we completed 200 ns molecular dynamics simulations using GROMACS[22]. We used AMBER forcefield 99SB-ILDN for the protein and the microMg forcefield designed by Grotz et al. for the magnesium ions in the active site (which AMBER cannot handle), and used H++ to generate forcefields for both ICL1 and for the compound of choice[23-25]. From there we combined the topology files. We first solvated the protein in water and added Na+ and Cl- ions to neutralize the system. After initial energy minimization, we performed an NVT equilibration, where the system was heated to 300K. Next we completed an NPT equilibration. The production run was performed under NPT conditions using the Bussi-Donadio-Parrinello thermostat with a time step of 2 fs. We used periodic boundary conditions and long-range electrostatics were calculated using particle-mesh Ewald summations. We also used a van der Waals cut-off of 1nm.

Discussion

Near the end of the simulation, the compound seemed to be coming out of the pocket. This could be for a number of reasons which all would prevent the drug from being effective. One possible reason is the hydrophobicity of the central cavity. When one compound we chose was docked in the central cavity, a vast majority of its polar components were near the hydrophobic area of the pocket. This hydrophobicity mismatch could have pushed the compound out of the pocket.

Another reason could be that the compound is too large. One potential problem with our method of inhibition (specifically blocking tetramerization) is that a fine balance needs to be achieved – the compound must be big enough to physically block another dimer from binding, but small enough that it still fits into the pocket.

To solve either of these problems, more stringent filters on virtual screening might need to be placed. If we picked hydrophobic features that aligned more with the hydrophobicity of the pocket, it might yield better results. If we lowered the number of rotational bonds or molecular weight, it might give us compounds that are smaller and potentially stay in the pocket larger.

After we do further virtual screening and molecular dynamics simulations, we plan to send the most promising compounds for experimental validation.

Acknowledgements

This was research was funded in part by the Laidlaw Foundation.

References

1 Suvvari, T. K. The persistent threat of tuberculosis - Why ending TB remains elusive? *J Clin Tuberc Other Mycobact Dis* **38**, 100510 (2025). https://doi.org/10.1016/j.jctube.2025.100510

2 Munoz-Elias, E. J. & McKinney, J. D. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* **11**, 638–644 (2005). https://doi.org/10.1038/nm1252

3 Bhusal, R. P., Bashiri, G., Kwai, B. X. C., Sperry, J. & Leung, I. K. H. Targeting isocitrate lyase for the treatment of latent tuberculosis. *Drug Discov Today* **22**, 1008–1016 (2017). https://doi.org/10.1016/j.drudis.2017.04.012

4 Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**, 2074–2075 (2012). https://doi.org/10.1093/bioinformatics/bts310

5 Kwon, S., Chun, H. L., Ha, H. J., Lee, S. Y. & Park, H. H. Heterogeneous multimeric structure of isocitrate lyase in complex with succinate and itaconate provides novel insights into its inhibitory mechanism. *PLoS One* **16**, e0251067 (2021). https://doi.org/10.1371/journal.pone.0251067

6 Sharma, V. *et al.* Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis. *Nat Struct Biol* **7**, 663–668 (2000). https://doi.org/10.1038/77964

7 Hussein, H. A. *et al.* PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Res* **43**, W436–442 (2015). https://doi.org/10.1093/nar/gkv462

8 Schmidtke, P., Le Guilloux, V., Maupetit, J. & Tuffery, P. fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res* **38**, W582–589 (2010). https://doi.org/10.1093/nar/gkq383

9 Valsson, I. *et al.* Narrowing the gap between machine learning scoring functions and free energy perturbation using augmented data. *Commun Chem* **8**, 41 (2025). https://doi.org/10.1038/s42004-025-01428-y

10 Yovanno, R. A. & Lau, A. Y. Directional information flow as a tool for analyzing protein allostery. *bioRxiv* (2026). https://doi.org/10.64898/2026.06.22.733418

11 Doering, N. P., Taterra, M., Bermudez, M. & Wolber, G. MDPath: Unraveling Allosteric Communication Paths of Drug Targets through Molecular Dynamics Simulations. *J Chem Inf Model* **65**, 11123–11135 (2025). https://doi.org/10.1021/acs.jcim.5c01590

12 Sunseri, J. & Koes, D. R. Pharmit: interactive exploration of chemical space. *Nucleic Acids Res* **44**, W442–448 (2016). https://doi.org/10.1093/nar/gkw287

13 Khan, O. *et al.* E-FTMap: A Protein Structure Based Pharmacophore Identification Server for Guiding Fragment Expansion. *J Mol Biol* **437**, 168956 (2025). https://doi.org/10.1016/j.jmb.2025.168956

14 Tingle, B. I. *et al.* ZINC-22 horizontal line A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *J Chem Inf Model* **63**, 1166–1176 (2023). https://doi.org/10.1021/acs.jcim.2c01253

15 Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461 (2010). https://doi.org/10.1002/jcc.21334

16 Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* **53**, 2719–2740 (2010). https://doi.org/10.1021/jm901137j

17 Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem Sci* **15**, 3130–3139 (2024). https://doi.org/10.1039/d3sc04185a

18 Seo, S. & Kim, W. Y. PharmacoNet: deep learning-guided pharmacophore modeling for ultra-large-scale virtual screening. *Chem Sci* **15**, 19473–19487 (2024). https://doi.org/10.1039/d4sc04854g

19 Morgan, H. L. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of chemical documentation* **5**, 107–113 (1965).

20 Li, X. *et al.* A High-Quality Data Set of Protein-Ligand Binding Interactions Via Comparative Complex Structure Modeling. *J Chem Inf Model* **64**, 2454–2466 (2024). https://doi.org/10.1021/acs.jcim.3c01170

21 Wang, Y. *et al.* A workflow to create a high-quality protein-ligand binding dataset for training, validation, and prediction tasks. *Digit Discov* **4**, 1209–1220 (2025). https://doi.org/10.1039/d4dd00357h

22 Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013). https://doi.org/10.1093/bioinformatics/btt055

23 Grotz, K. K., Cruz-Leon, S. & Schwierz, N. Optimized Magnesium Force Field Parameters for Biomolecular Simulations with Accurate Solvation, Ion-Binding, and Water-Exchange Properties. *J Chem Theory Comput* **17**, 2530–2540 (2021). https://doi.org/10.1021/acs.jctc.0c01281

24 Gordon, J. C. *et al.* H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**, W368–371 (2005). https://doi.org/10.1093/nar/gki464

25 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–1174 (2004). https://doi.org/10.1002/jcc.20035