AUTOFILTER: A LOW-COST BIOCOMPUTATIONAL FRAMEWORK FOR HIGH-THROUGHPUT SCREENING OF CHEMICAL DATABASES AND IDENTIFICATION OF NOVEL MALARIA INHIBITORS TARGETING PLASMODIUM FALCIPARUM

Kavin Ramadoss
Sunset High School

Kamal Singh
Bond Life Sciences Center, University of Missouri

ABSTRACT: Malaria is the third deadliest disease, with approximately 249 million cases annually, particularly in tropical regions. Caused by Plasmodium parasites transmitted through the bite of Anopheles mosquitoes, malaria remains a significant global health burden and is increasingly difficult to treat due to rising drug resistance. Drug discovery for malaria is both costly and time-consuming, typically requiring over a decade and around $3 billion before a compound gains approval. To address this challenge, AutoFilter was developed: a low-cost and novel biocomputational framework that integrates machine learning (ML) and screening tools to streamline the filtering of large chemical databases for more efficient drug discovery. AutoFilter sequentially screens compounds that violate basic chemical filters such as Lipinski's Rule of 5 (Lipinski et al., 1997), Veber's (Veber et al., 2002), and PAINS; performs molecular docking and analyzes post-docking interactions; conducts ADME filtration to identify compounds with favorable drug-like properties; employs an ML model to predict toxicity and synthetic accessibility; and finally applies molecular dynamics (MD) simulations to refine compound stability. AutoFilter was applied to screen the ChEMBL database, which contains 2.4 million bioactive compounds, to identify malaria inhibitors targeting Plasmodium falciparum apPOL. The five selected compounds demonstrated high inhibition performance and favorable drug-like properties and are currently undergoing *in vitro* trials. As the first integrated biocomputational framework for chemical database screening, AutoFilter is a transformative tool for drug discovery across diverse diseases, efficiently identifying inhibitors while reducing costs and time by 50%, with the profound potential to save lives worldwide.

1 Introduction

Malaria remains one of the most devastating infectious diseases worldwide, responsible for more than 600,000 deaths annually, particularly in tropical and subtropical regions. Caused by Plasmodium parasites and transmitted through the bite of Anopheles mosquitoes, malaria presents a significant global health burden, with Plasmodium falciparum being the deadliest species. Despite extensive efforts in drug development, malaria remains notoriously difficult to treat due to the parasite's complex life cycle, rapid evolution, and increasing resistance to frontline antimalarial therapies. One of the primary challenges in developing effective malaria drugs is Plasmodium's ability to quickly develop resistance to small-molecule inhibitors. Antimalarial drugs such as chloroquine and artemisinin have historically been effective but have seen diminishing efficacy due to widespread drug resistance. As a result, the CDC determines that all four major drugs currently targeting malaria are non-functional due to the disease being able to undergo multiple developmental stages, necessitating the need for inhibitors that can target essential, conserved pathways across its life cycle. Furthermore, the parasite's ability to sequester in deep tissues, evade immune responses, and exhibit metabolic plasticity adds to the difficulty in designing long-lasting therapeutics.

Many efforts have been made to create better drugs for malaria, but drug discovery is typically a costly and lengthy process, often requiring over a decade and approximately $3 billion before a compound reaches approval. This prolonged timeline is further complicated by the high failure rate of potential drugs at various stages of clinical trials, making the development of new antimalarials a challenging endeavor. In drug discovery, machine learning (ML) scans through biological data to identify patterns that may be missed by humans. ML models can be used to predict activity and potential side effects of hit (or lead) compounds and can suggest new scaffolds with promising properties. As such, the use of computational tools in drug discovery is not a new phenomenon. However, ML has significantly enhanced the capacity of Computer-Aided Drug Discovery (CADD). Traditionally, researchers have relied on screening drug-like compounds to identify "hit" compounds followed by conducting Quantitative Structure-Activity Relationship (QSAR) to improve upon the potency of a "hit" or a "lead" compound.

Chemical databases such as ChEMBL (Elber et al., 2011) and ZINC250 contain an abundance of information about potential drug-like compounds, yet many screening methods fail to efficiently identify viable candidates due to the lack of systematic filtering processes. Specifically, screening of hit compounds often results in predicting the compounds that may violate "empirical" rules of drug-likeness. While these empirical rules are not absolute in drug-discovery processes, their implementation can accelerate the identification of compounds that are likely to be potential drugs.

Additionally, the availability of numerous databases containing drug-like small molecules facilitates virtual screening through docking software to identify "hit" compounds.

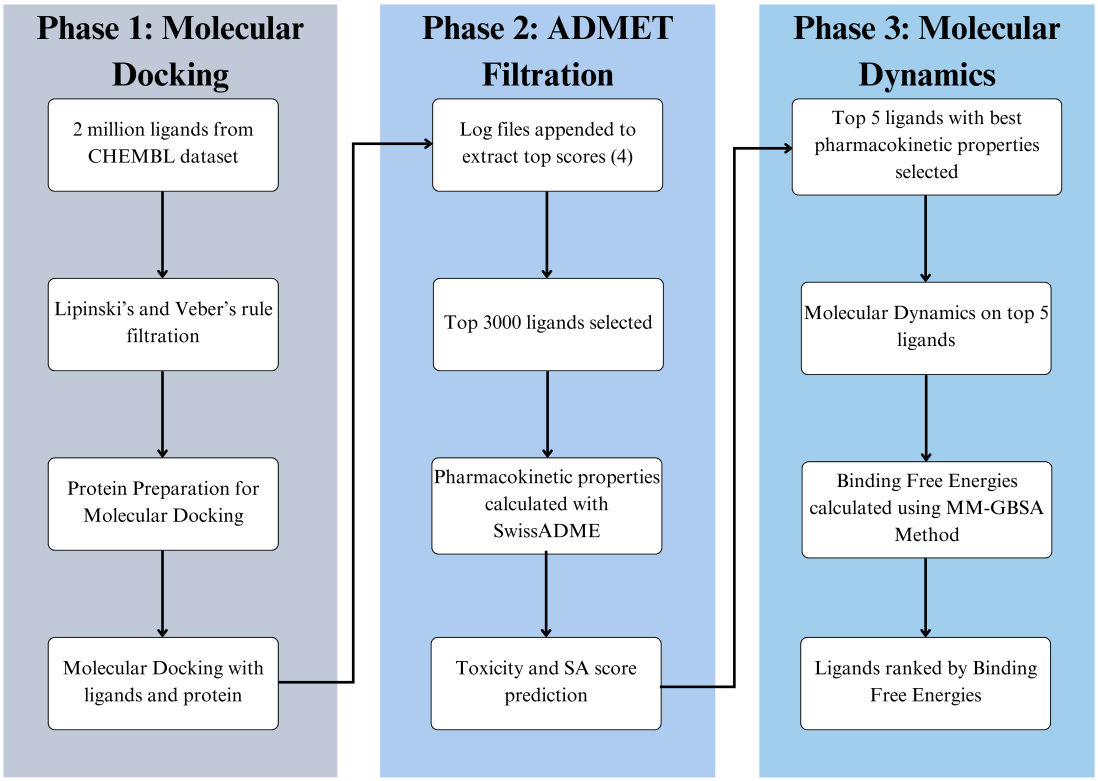

Figure 1: AutoFilter Framework

Similarly, high-throughput screening of compounds (usually in plate format) using well-developed assays has also been traditionally used to identify "hit" compounds from chemical databases and compound libraries. Regardless, neither of these methods implements the rules to exclude compounds that violate the rules of drug-likelihood.

To overcome these limitations and expedite the drug-discovery process, AutoFilter was developed–a biocomputational framework designed for efficient compound screening by excluding database entries that violate Lipinski's Rule of Five (Ro5), Veber's rule for oral bioavailability, and pan-assay interference compounds (PAINS). Figure 1 illustrates how AutoFilter docks the remaining compounds into the predefined binding pocket and prioritizes them based on docking scores for further analysis. To refine compound selection, the framework also performs molecular dynamics (MD) simulations on the top-ranked compounds. To assess its effectiveness, AutoFilter was applied to identify potential inhibitors of Plasmodium falciparum apicoplast DNA polymerase (apPOL), a critical enzyme for parasite survival.

2 Methods

2.1 Selection of Chemical Library and Initial Curation

The compounds were extracted from ChEMBL (n = 2,400,000) and were prepared for docking studies using Meeko to generate three-dimensional conformations and convert them into the PDBQT format that is required for AutoDock Vina (Trott and Olson, 2010. This preparation included the addition of hydrogen atoms and Gasteiger partial charges and the removal of water and heteroatoms. This preprocessing helped ensure that the library contained a wide range of molecular scaffolds, increasing the likelihood of identifying effective inhibitors. Any compounds that didn't pass these filters were removed, and the final processed dataset (n = 2,200,000) was then selected for further investigation.

2.2 Preparation of Target Protein

The three-dimensional structure of malaria DNA polymerase was obtained from the Protein Data Bank (PDB). The protein was prepared using AutoDockTools (Trott and Olson, 2010), which involved the addition of hydrogen atoms, the removal of water molecules and heteroatoms, and the assignment of Kollman charges to the protein. The final prepared protein was saved in PDBQT format, ensuring compatibility with the docking software. The choice of DNA polymerase as the target protein was based on its important and critical role in the parasite's life cycle and its potential as a drug target. The structural integrity of the protein was verified through visual inspection and analysis of the PDB file, confirming that no mutations or anomalies were present in the protein structure.

2.3 Allosteric Binding Site

Prior studies mentioned the correlation between drug resistance and the inhibitor binding to the active site of the protein. Therefore, using a computational approach, a novel allosteric binding site was discovered. A geometric grid-and-sphere-based method, combined with a dual-probe system, was applied to detect and characterize cavities within the protein's crystal structure. This enabled detailed analysis of the binding pocket's spatial dimensions, depth, and composition, along with its hydropathy properties. To further investigate the molecular architecture, molecular volume was calculated using van der Waals (vdW) surfaces, solvent-excluded surfaces (SES), and solvent-accessible surfaces (SAS), providing a comprehensive view of the binding site's physical and chemical environment. This analysis led to the identification of a novel, previously undiscovered allosteric site.

2.4 Molecular Docking

Molecular docking studies were conducted using AutoDock Vina . The docking protocol involved defining the 20x20x20 size grid box around the active site of the DNA polymerase. The grid parameters were carefully set to encompass the novel allosteric binding site at coordinates 10.0, -34.4, 112.4, representing the X,Y,Z coordinates used, ensuring that all potential binding interactions could be evaluated.

The curated library of compounds was submitted for docking, and each compound's binding affinities were calculated. The binding affinities were calculated using:

$$\begin{aligned}\Delta G = {} & \Delta G_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \\ & + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{\text{bond}} \right) \\ & + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \\ & + \Delta G_{\text{tor}} N_{\text{tor}} \\ & + \Delta G_{\text{sol}} \sum_{i,j} S_i V_j e^{\left(\frac{-r_{ij}^2}{2\sigma^2} \right)}\end{aligned}$$

The binding interactions were further analyzed using ProLif to elucidate the nature of the interactions, including hydrogen bonds, hydrophobic constants, and π-π stacking. A custom docking was built using Python to conduct the molecular docking in a high-throughput manner, allowing for the rapid evaluation of thousands of compounds. Each docking run generated a series of conformations for each ligand, which were iteratively ranked based on their predicted binding affinities. The top-ranked compounds (n = 3,194) were selected for further analysis from the processed dataset of bioactive compounds (n = 2,200,000).

2.5 ADME-Based Filtration

After completing the molecular docking analysis, the ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties of the best-hit compounds were assessed. This filtration process is essential for reducing the likelihood of selecting compounds with adverse side effects, thereby minimizing the risk of late-stage (Phase 2 and 3) attrition due to efficacy and safety deficiencies. For this purpose, SwissADME (Daina et al., 2017) was used to analyze the pharmacokinetic profiles of the top candidates identified in the docking studies. The best-hit compounds, from molecular docking, (n = 3,194) were submitted in SMILES format to SwissADME for comprehensive ADMET

predictions. The *in silico* ADMET prediction approach has long been recognized as an important parameter in the pharmaceutical or drug discovery pipeline. It aids in the selection of potentially viable compounds from large datasets and assists in the optimization of new chemical entities. The results from SwissADME were used to filter the compounds that did not meet the desired pharmacokinetic criteria.

2.6 Machine Learning-Based Toxicity and Synthetic Accessibility Assessment

To further refine our selection of potential drug candidates, ML techniques were implemented to assess the toxicity and synthetic accessibility (SA) of the identified small molecules. For this assessment, eToxPred (Pu et al., 2019) was utilized, an ML model that predicts the toxicity of chemical compounds. The models are trained on diverse datasets that include known drugs, potentially hazardous chemicals, natural products, and synthetic bioactive compounds. The eToxPred tool uses the Extra Tree (ET) algorithm, along with other algorithms such as Gradient Boosting (GB), Balanced Random Forest (BRF), and Balanced Bagging, to provide a comprehensive analysis of the compounds' toxicity profiles.

Prior to the prediction of toxicity and synthetic accessibility scores, the developed models were validated through the calculation of receiver operating curve (ROC) and area under the curve (AUC) and accuracy metrics. This validation process ensures that the models are reliable and capable of accurately predicting the toxicity of new compounds. The toxicity values generated by eToxPred range from 0 to 1, indicating the likelihood of a compound being toxic.

Additionally, eToxPred was used to assess synthetic accessibility (SA) scores for the evaluated compounds (n = 3,194). The SA score ranges from 0 to 1, with lower scores indicating easier synthesis and higher scores suggesting more complex synthetic pathways. The SA score guided the selection of compounds that are not only biologically active but also feasible for synthesis in a laboratory setting. By integrating ML-based toxicity and synthetic accessibility assessments into our workflow, the quality of our candidate selection process was enhanced. This approach allows for a more informed decision-making process, ultimately increasing the chances of identifying viable drug candidates for further development.

After extensive ADME, toxicity, and SA filtration, compounds that didn't exhibit favorable pharmacokinetic properties were disregard to reduce the chances of potential failures in further testing stages (Phase 2 and 3). As a result, from the compounds that exhibited high binding affinity energies (n = 3,194) only the compounds that had favorable drug-like properties were chosen (n = 10).

2.7 Molecular Dynamics (MD) Simulations Analysis and MM-GBSA

To gain deeper insights into the binding interactions between the selected compounds and apPOL, MD simulations were conducted on the protein-ligand complexes. All MD simulations were performed using GROMACS (Abraham et al., 2015), as detailed below.

A total of 10 protein-ligand complexes for 100 ns all-atom classical MD simulations were selected. These 10 protein-ligand complexes have preferable molecular docking scores and have favorable ADME and drug-like properties. The MD simulations were conducted with a time step of 2 fs, under constant pressure of 1 atm and a constant temperature of 300 K. The topology of the apPOL was generated using the CHARMM36 (Z. Huang et al., 2023) all-atom force field. Each protein-ligand complex system was solvated using the TIP3P (Mark and Nilsson, 2001) water model and submerged into a cubic box before the simulation run. To neutralize the biomolecular system, appropriate numbers of Na+ and Cl ions were added.

Following the successful completion of the MD simulations, critical analytical metrics including protein backbone root-mean-square deviation (RMSD) and ligand root-mean-square fluctuation (RMSF). These metrics provided valuable insights into the stability and flexibility of the protein-ligand complexes throughout the simulation.

These MD simulations provided valuable information about protein-ligand interactions. From the 10 ligands that molecular dynamic simulations were performed on, the top 5 that exhibited preferable protein-ligand interactions. These five compounds are the final proposed potential malaria inhibitors.

2.8 Proof of Concept (POC)

To validate the proposed framework, a comprehensive proof of concept (POC) was conducted using DNA polymerase theta (Pol Theta) as the target protein. Pol Theta is a key enzyme involved in DNA repair pathways. The goal of this validation was to determine whether AutoFilter could accurately identify known inhibitors of Pol Theta from ChEMBL. AutoFilter went through all six stages of its workflow, including compound preprocessing, molecular docking, feature extraction, and ranking. The binding site was defined at coordinates X = -93.3, Y = -104.8, and Z = 54.7, a region chosen because previously characterized inhibitors are known to interact at this site. Within the ChEMBL database, seven compounds were experimentally validated inhibitors of Pol Theta based on past research studies. AutoFilter successfully identified all seven of these inhibitors during the screening process. This result demonstrates that the framework not only performs consistent and accurate compound prioritization but also effectively distinguishes true inhibitors from a vast background of non-binding molecules. The ability to recover all known inhibitors from such a large dataset provides strong evidence for AutoFilter's robustness, reliability, and potential utility in large-scale virtual screening and drug discovery applications.

2.9 *in vitro* Validation

To experimentally validate the efficacy of the final selected compounds (n = 5), *in vitro* validation was performed and is currently ongoing. DNA polymerization was the chosen process to study because it is a fundamental process by which nucleotides (dNTPs) are sequentially added to a growing DNA strand, guided by a template strand and initiated by a primer with a free 3'-hydroxyl (3'-OH) group. In malaria-causing Plasmodium parasites, this mechanism is essential for the replication and repair of their genomic DNA, enabling the production of new infected cells that contribute to the rapid spread of the disease within the host. Because this process is vital for parasite survival and proliferation, it serves as a key target for antimalarial drug development. DNA polymerization assays are commonly employed to evaluate a compound's ability to inhibit the synthesis of parasite DNA, thereby blocking the replication of malaria-infected cells. By interfering with this critical biological function, candidate drugs can disrupt the parasite's life cycle and offer a potential therapeutic strategy against malaria.

3 Results

AutoFilter, when tested against apPOL, successfully identified promising inhibitor candidates (n = 5) from the curated ChEMBL library. The bioactive compounds (n = 2,400,000) in the ChEMBL databases were processed and then molecular docking was performed on them given certain molecular docking conditions. Following the steps in AutoFilter (Figure 1), 5 compounds emerged as the most favorable candidates based on their binding affinities and interaction profiles. These compounds were subjected to further evaluation, including their physicochemical properties, ADMET profiles, toxicity predictions, and synthetic accessibility scores.

3.1 Molecular Docking Results

The molecular docking studies revealed that the top inhibitor candidates bind to the active site of malaria DNA polymerase, forming key interactions with critical residues. Compound L1 exhibited the strongest binding affinity with a docking score of *-10.830 kcal/mol*, followed by Compound L2 *(-10.317 kcal/mol)*, Compound L3 *(-10.204 kcal/mol)*, Compound L4 *(-10.108 kcal/mol)*, and Compound L5 *(-9.828 kcal/mol)*. These docking scores, all below *-9.8 kcal/mol*, indicate strong and energetically favorable binding for each compound. A score lower than *-9.0 kcal/mol* is generally considered indicative of high-affinity ligand binding, suggesting that each of these compounds has significant potential as an effective inhibitor. Compound L1, with the most negative score, appears particularly promising, while the others closely follow, demonstrating only slight variations in predicted binding strength.

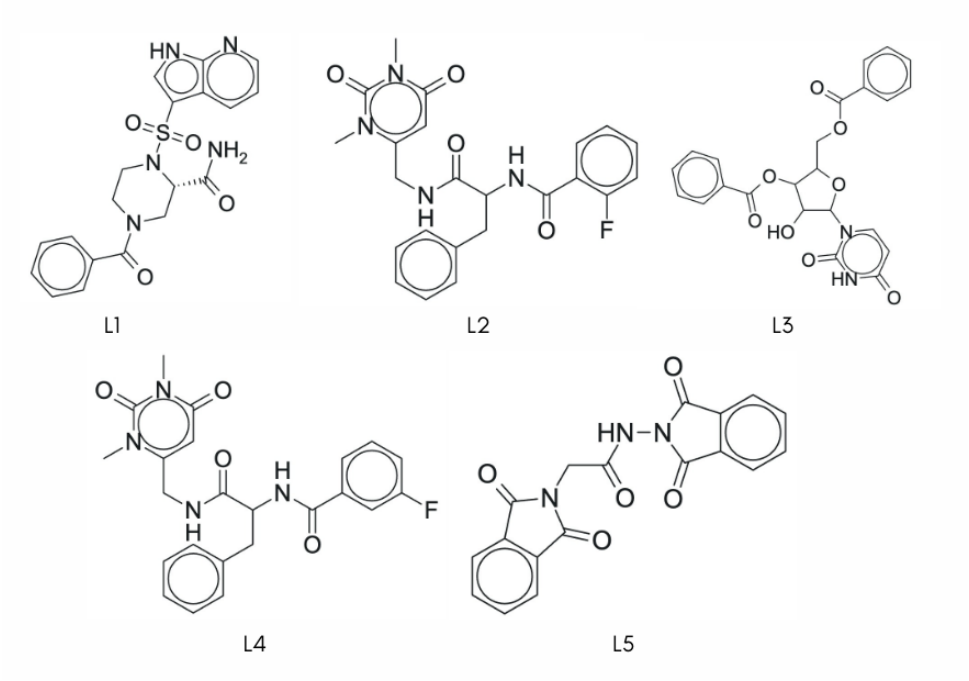

Figure 2: Final Selected Ligands (2D)

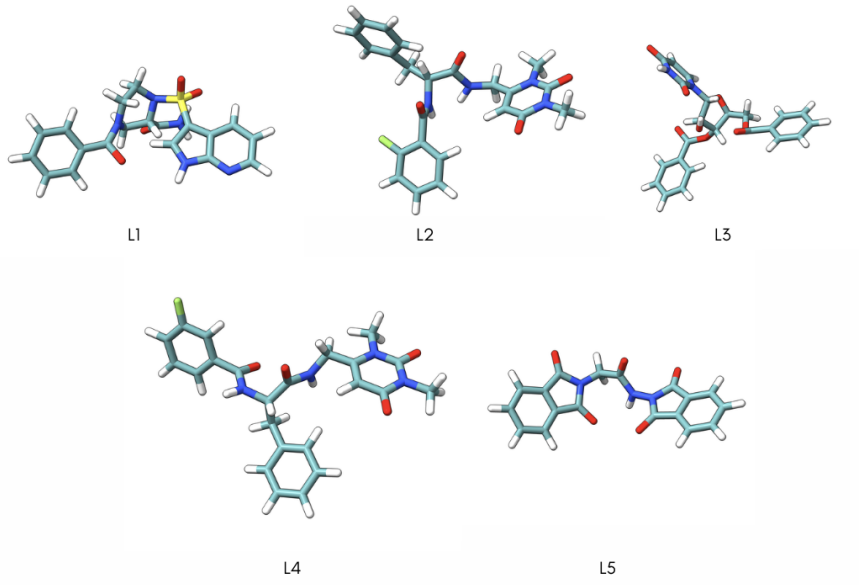

Figure 3: Final Selected Ligands (3D)

The binding modes of the compounds were further analyzed using ProLIF in both 2D and 3D formats, which provides detailed insights into the nature of the interactions between the ligands and the protein. As seen in Figure 4 (link), Compound L1 formed Van Der Waals (VdW) bonds with residues Met490, Ala482, Ile446, Tyr506, and more, which are crucial for stabilizing the ligand within the active site. Additionally, hydrophobic interactions with Leu478, Met428, and Phe498

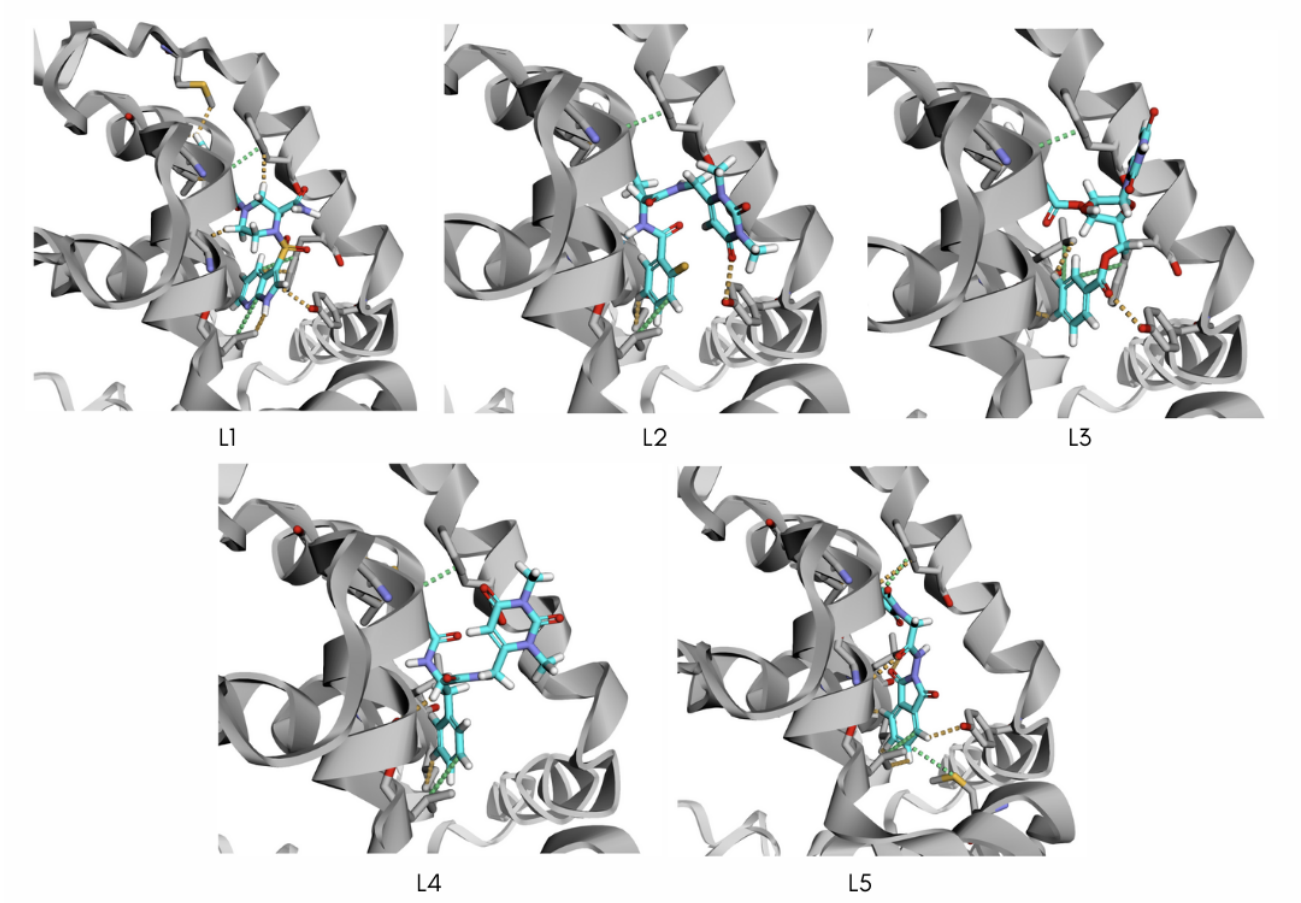

Figure 4: 3D Protein-Ligand Interactions Post-Docking

further enhance the binding affinity of Compound L1, indicating a strong and stable interaction with the enzyme.

Compound L2, while exhibiting a slightly lower docking score, also showed significant binding interactions. It formed VdW bonds with Ile470, Asn466, Ile446, and Tyr506. The presence of hydrophobic contacts with Phe498, Leu478, and Ile438 suggests that Compound L2 has a binding profile similar to that of Compound L1, although slightly less favorable due to its higher docking score.

Compound L3 established VdW bonds with Tyr506, Leu469, Leu478, Asn466, Phe502, Met490, and Met428. Compound L3 also had hydrophobic interactions with Phe498, Leu438, and Leu478.

Compound L4 exhibited a binding profile comparable to the other compounds, with notable interactions within the active site of malaria DNA polymerase. It formed VdW bonds with Ala482, Ile446, Leu469, Tyr506, His505, and Ser501, similar to Compound L1, indicating its ability to

engage with crucial residues for ligand stabilization. Hydrophobic interactions were observed with Val344, Phe583, and Tyr612, further enhancing L4's binding affinity.

Finally, Compound L5 demonstrated a binding mode that aligned well with the other compounds in the series. It formed VdW bonds with Tyr506, Ala482, Thr442, and more. The presence of hydrophobic contacts with Phe498, Met428, and Leu428 was also noted, consistent with the binding patterns of the other compounds. The similarity in its interaction profile to the other compounds suggests that it could be a viable candidate for inhibition of malaria DNA polymerase, warranting further investigation into its potential as a therapeutic agent.

3.2 ADMET Filtration Results

Table 1: Pharmacokinetic properties for final selected molecules

	MW	NRB	TPSA	LogP	PAINS	Veber	Lipinski	GIA	SP	BBBP
L1	413.45	5	137.84 Å²	1.40	0 Alerts	Yes	Yes	High	-8.56	No
L2	438.45	9	102.20 Å²	2.42	0 Alerts	Yes	Yes	High	-7.83	No
L3	452.41	8	136.92 Å²	2.32	0 Alerts	Yes	Yes	High	-7.69	No
L4	438.45	9	102.20 Å²	3.00	0 Alerts	Yes	Yes	High	-7.83	No
L5	349.30	4	103.86 Å²	1.30	0 Alerts	Yes	Yes	High	-7.64	No

[a] Molecular Weight;

[b] Number of Rotatable Bonds;

[c] Topological Polar Surface Area;

[d] Lipophilicity;

[e] Pan-assay interference compound (PAINS);

[f] Gastrointestinal Absorption;

[g] Skin Permeation;

[h] Blood-Brain Barrier Permeation

The physicochemical properties of the top inhibitor candidates (Table 1) indicate that the selected compounds possess favorable drug-like characteristics. Compound L5 has the lowest molecular weight compared to all the others, with L5 having a molecular weight of 349.30 g/mol. Compound L3 had the largest molecular weight at 452.50 g/mol. The number of rotatable bonds is higher for

L2 and L4 (8 each) and L3 (8) compared to L1 (5) and L5 (4), indicating their higher flexibility and binding affinity. A higher number of rotatable bonds can increase the conformational flexibility of a molecule, potentially enhancing its ability to adapt to the binding site and form favorable interactions with the target protein (also known as 'Wiggling and Giggling'). However, excessive flexibility can lead to entropic penalties upon binding.

The polar surface area (PSA) values suggest that L2, L4, and L5 have lower polarity compared to L1, and L3. The lipophilicity values (iLOGP) for the compounds range from 1.40 to 3.00, indicating a balance between lipophilicity and hydrophilicity. All compounds pass Lipinski's rule of five, indicating that they have properties similar to known orally active drugs. The bioavailability scores for the compounds range from 0.55 to 0.56, indicating moderate to high potential for oral absorption.

Table 2: Toxicity and Synthetic Accessibility Score for final selected molecules

	TS	SAS
L1	0.26715	0.15
L2	0.28116	0.18
L3	0.36736	0.09
L4	0.28439	0.18
L5	0.35509	0.09

[a] Toxicity score;

[b] Synthetic accessibility score

The SA scores for the five compounds range between 0.09 and 0.18 (on a scale of 0 to 1, where 1 indicates that it is hard to synthesize), suggesting that these drugs are relatively easy to synthesize. The toxicity scores for the top inhibitor candidates are presented in Table 2. The toxicity scores, predicted using ML models, range from 0.26 to 0.36, with lower values indicating a lower likelihood of toxicity. Compound L1 has the lowest toxicity score, suggesting a favorable safety profile.

3.3 Molecular Dynamics Results

After conducting MD simulations, root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyzes were performed. RMSD was calculated by measuring the average deviation in the atomic positions of the protein backbone atoms relative to the initial structure. RMSF was computed for each residue by averaging the fluctuations of its atomic positions across the trajectory, offering residue-level information on flexibility.

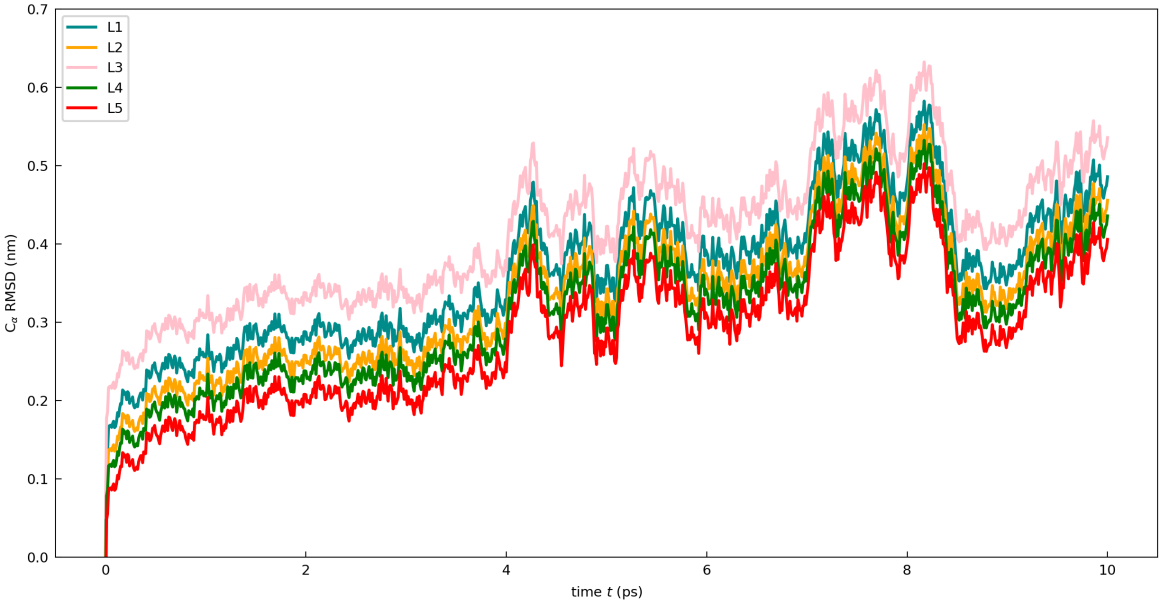

Figure 5: RMSD over time (picosecond)

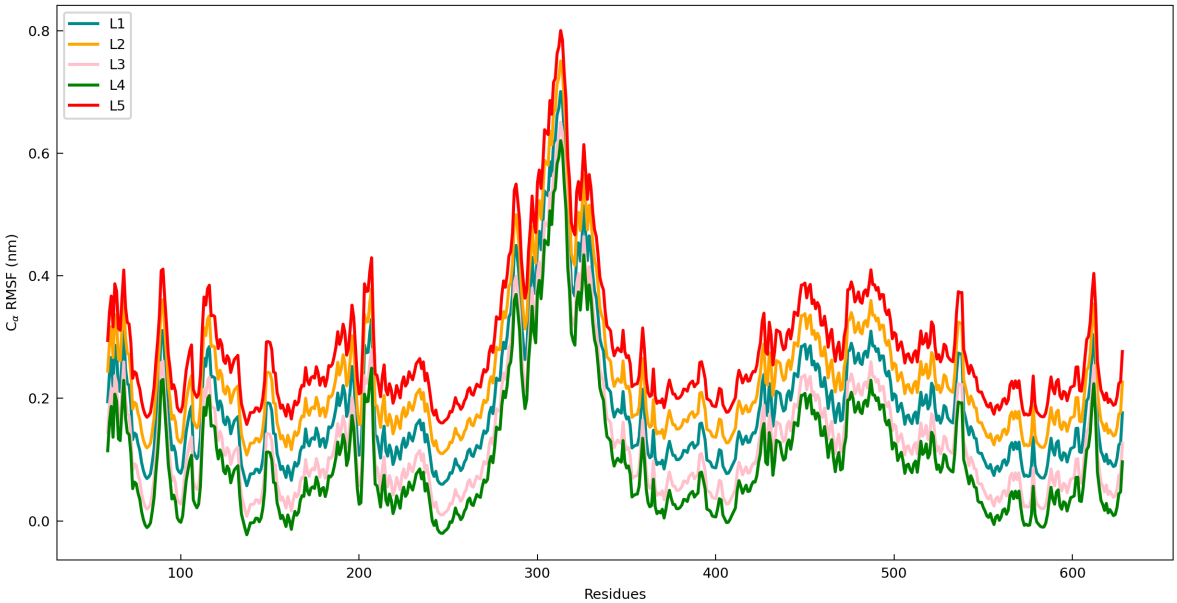

Figure 6: RMSF over residues

3.4 in vitro Results

The five final selected compounds are currently undergoing *in vitro* assays to test their efficacy performance against malaria. Preliminary results show that Compound L1 is able to inhibit by at least 50%. Further trials will be conducted to provide more definitive results, and other important metrics such as IC50 will be calculated to quantify its efficacy.

4 Conclusion

Through this study, two main objectives were accomplished: (i) create the first biocomputational framework for the screening of large chemical databases, and (ii) develop five potential drugs that inhibit malaria and are currently undergoing *in vitro* assays to test their inhibition performance, one of which has been shown to inhibit the disease by at least 50%. AutoFilter has established itself as the first biocomputational framework available for the screening of large chemical databases, revolutionizing the process of drug discovery. While this research specifically applied AutoFilter to malaria drug discovery, it is a universal framework and can be adapted for any disease, given a target protein. For example, AutoFilter can be leveraged to develop potential inhibitors for HIV, tuberculosis, and even neurodegenerative disorders by identifying inhibitors that interact with their respective disease-related proteins. This universality makes AutoFilter a tool that extends beyond malaria research, offering the potential to accelerate drug discovery across a wide range of diseases.

Beyond its ability to scan and identify inhibitors, AutoFilter significantly enhances the efficiency of drug discovery by reducing both time and cost. Traditional drug discovery is a lengthy and expensive process, often taking up to a decade and billions of dollars to bring a single drug to market. AutoFilter mitigates these challenges by reducing the time for the initial screening phase, rapidly narrowing down vast chemical libraries to a select few high-potential compounds. Specifically, it reduces the current cost of drug discovery by around 75%. Furthermore, the speed of drug discovery is significantly increased by around 75% as well. This not only reduces the time for early stages of drug development but also minimizes wasted resources on compounds that are unlikely to succeed. Given its universal applicability, AutoFilter has the potential to transform drug discovery by accelerating the discovery of new treatments for a wide range of diseases.

Since 1 of the five compounds has demonstrated efficacy after preliminary *in vitro* assays, it could benefit millions worldwide but also validate AutoFilter as a groundbreaking biocomputational framework. In fact, AutoFilter and the drugs it identifies have the ability to contribute to global healthcare advancements.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Bryant, P., Kelkar, A., Guljas, A., Clementi, C., & Noé, F. (2024). Structure prediction of protein-ligand complexes from sequence information with umol. *Nat Commun*, *15*(1), 4536. https://doi.org/10.1038/s41467-024-48837-6

Daina, A., Michielin, O., & Zoete, V. (2017). Swissadme: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*, *7*(1), 42717. https://doi.org/10.1038/srep42717

Elber, R., Ruymgaart, A. P., & Hess, B. (2011). Shake parallelization. *Eur. Phys. J. Spec. Top.*, *200*(1), 211–223. https://doi.org/10.1140/epjst/e2011-01525-9

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics*, *103*(19), 8577–8593. https://doi.org/10.1063/1.470117

Fikadu, M., & Ashenafi, E. (2023). Malaria: An overview. *IDR*, *16*, 3339–3347. https://doi.org/10.2147/IDR.S405668

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). Chembl: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, *40*(D1), D1100–D1107. https://doi.org/10.1093/nar/gkr777

Genheden, S., & Ryde, U. (2015). The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461. https://doi.org/10.1517/17460441.2015.1032936

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2017). CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods*, *14*(1), 71–73. https://doi.org/10.1038/nmeth.4067

Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T. J., & Zou, J. (2023). A Visual–Language Foundation Model for Pathology Image Analysis Using Medical Twitter. *Nat Med*, *29*(9), 2307–2316. https://doi.org/10.1038/s41591-023-02504-3

Lee, K., Jang, J., Seo, S., Lim, J., & Kim, W. Y. (2022). Drug-likeness scoring based on unsupervised learning. *Chem. Sci.*, *13*(2), 554–565. https://doi.org/10.1039/D1SC05248A

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development

settings. *Advanced Drug Delivery Reviews*, *23*(1–3), 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1

Mark, P., & Nilsson, L. (2001). Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *J. Phys. Chem. A*, *105*(43), 9954–9960. https://doi.org/10.1021/jp003020w

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open babel: An open chemical toolbox. *J Cheminform*, *3*(1), 33. https://doi.org/10.1186/1758-2946-3-33

Pu, L., Naderi, M., Liu, T., Wu, H.-C., Mukhopadhyay, S., & Brylinski, M. (2019). Etoxpred: A machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol Toxicol*, *20*(1), 2. https://doi.org/10.1186/s40360-018-0282-6

Santos, L. H. S., Ferreira, R. S., & Caffarena, E. R. (2019). Integrating molecular docking and molecular dynamics simulations. In W. F. De Azevedo (Ed.), *Docking screens for drug discovery* (pp. 13–34, Vol. 2053). Springer New York. https://doi.org/10.1007/978-1-4939-9752-7_2

Thomas, C. M., Stauffer, W. M., & Alpern, J. D. (2023). Food and drug administration approval of artesunate for severe malaria: Enough to achieve best practice? *Clinical Infectious Diseases*, *76*(3), e864–e866. https://doi.org/10.1093/cid/ciac728

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Veber, D. F., Johnson, S. R., Cheng, H.-Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, *45*(12), 2615–2623. https://doi.org/10.1021/jm020017n

Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011). Swissparam: A fast force field generation tool for small organic molecules. *J Comput Chem*, *32*(11), 2359–2368. https://doi.org/10.1002/jcc.21816

All graphics were created by student researcher using Canva, ProLif, and Matplotlib, 2025