JESSICA LEIGH THIBAUD

PROJECT: Targeting Plasmodium falciparum cGMP-dependent Protein Kinase: Machine learning and Medicinal Chemistry Approaches.

Background and Rationale

Malaria-related deaths are increasing annually as a result of parasite resistance. This is why continued research of the disease and development of structurally and mechanistically novel treatments is crucial.[1] *Plasmodium falciparum* cGMP-dependent protein kinase (PfPKG) has been identified as a prioritised target for drug discovery.[4] This enzyme regulates the parasite's cGMP-signalling pathway and plays a vital role in many of its key developmental stages, therefore making it a promising antimalarial drug target.[5,6] The aim of this investigation is to identify new PfPKG inhibitors using all tools at our disposal in an effort to decrease both the cost and time of the drug-discovery pipeline. Recent work in our group has led to the development of the Antiplasmodium Chemical Space (APCS) map.[2,3] Generated using principal component analysis, the map allows one to visualise a subset of antiplasmodium chemical space, and based on target class clustering, observe areas of enrichment for activity prediction. Myburgh *et al*. have demonstrated the successful use of the APCS map with respect to identifying new inhibitors of three diverse antimalarial drug targets, including PfPKG.[3] We intend to explore this further with the purpose of increasing the hit rate for discovery of new PfPKG inhibitors.

Objectives and methodology

The initial phase of this research involves a combination of machine learning and molecular modelling computational techniques to identify hit compounds with predicted activity against the malaria PfPKG.

1. The APCS map will be used as preliminary filtration step to identify a manageable subset of potential PfPKG inhibitors from large online databases that fall in regions of enrichment (Fig. 1a).
2. The inhibitor ability of these compounds as predicted by the APCS map will then be refined using a Random Forest machine learning algorithm (Fig. 1b). This model will be trained on a 2D binary molecular *fingerprin*t set and a 3D molecular *descriptor* set deemed crucial for drug-likeness.
3. As a final predictive step, molecular modelling will be used to probe receptor-ligand active site interactions, particularly those deemed most important for inhibitor binding (Figure 1c).

As with all computational investigations, model validation based on a thorough analysis of the results is essential. Therefore, the research objectives for phase two are as follows.

4. A total of 30 compounds predicted to be active and 10 compounds predicted to be inactive, will be selected and purchased from online libraries. The PfPKG enzyme inhibitory activity of these compounds will be investigated using the facilities available at the H3D laboratories.

5. Promising hit compounds obtained from biological testing will be structurally modified in an attempt to improve important drug-like characteristics such as solubility, selectivity, and potency.

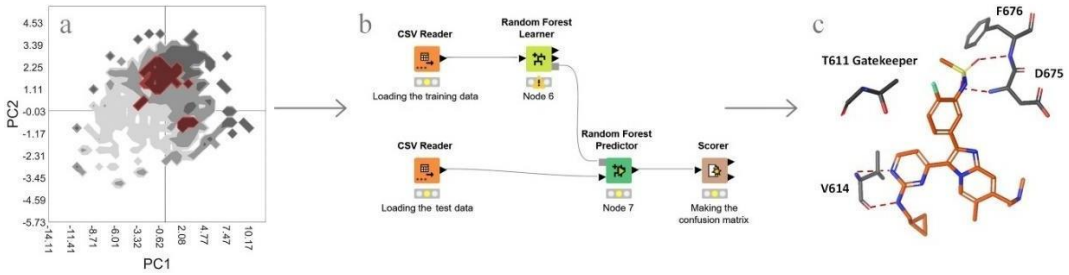

Figure 1. a) APCS map, where red indicates regions of PfPKG enrichment. b) Illustration of Random Forest machine learning. c) Molecular modelling showing hydrogen bond interactions important for inhibitors.

a. Timeline

	January - March	April - June	July - September	October - December
Year 1	Literature review and mining data for known-activity training sets.	Structure-activity analysis of training sets. This includes ligand docking.	Generation of the PfPKG percentage enrichment map (APCS map).	Plotting libraries onto APCS map and selecting regions of PfPKG enrichment.
Year 2	Develop a Random Forest machine learning model.	Use model to predict activity of compounds from APCS map.	Dock promising compounds (> 0.7) to PfPKG protein.	**Select compounds to test for model validation.**
Year 3	Functionalize hit compounds to improve potency.	Optimize model performance to improve predictions.	Drug-repurposing and dual activity study with TKI's.	Selection round 2 for improved model validation.
Year 4	Complete PhD project write-up.	Complete examiners corrections.	Finish with PhD defense.	-

b. Results or preliminary data

A total of 50 000 compounds from various online libraries have been plotted onto the PfPKG percentage enrichment map, and the largest areas of enrichment were selected (Fig. 1c). This resulted in a subset of 6255 preliminary predicted PfPKG inhibitors. The predictions were refined using both a structure-based approach in the form of molecular modelling i.e., ligand docking, and a ligand-based approach in the form of machine learning. Using the output of these computational techniques, the top 30 predicted active compounds and 10 predicted inactive compounds were selected and purchased from online libraries. The compounds were evaluated against the PfPKG enzyme to determine their inhibitory activity. As of yet two hits have been discovered, one with moderate activity and the other showing potent activity as a selective inhibitor of this enzyme (molecular structures are under embargo). Overall, these are the second set of inhibitors identified via the APCS map. Interestingly, our most potent hit shows promise as a dual-target inhibitor meaning that it is likely inhibiting two different targets in the malaria parasite. This will be the next focus of our research.

1. *World malaria report 2021*. (World Health Organisation, 2021).
2. Thibaud, J. L. *Identification and Evaluation of Novel Inhibitors to Target Plasmodium falciparum*. (Stellenbosch University, 2021).
3. Myburgh, D. *et al. Principal Component Analysis of Antiplasmodium Chemical Space Predicts Regions of Enrichment for Discovery of New Inhibitors of β-Hematin Formation, PfPKG and PfDHODH (Manuscript in Preparation)*. (2022).
4. Forte, B. *et al.* Prioritization of Molecular Targets for Antimalarial Drug Discovery. *ACS Infect. Dis.* **7**, 2764–2776 (2021).
5. Bakkouri, M. El *et al.* Structures of the cGMP-dependent protein kinase in malaria parasites reveal a unique structural relay mechanism for activation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14164–14173 (2019).

Baker, D. A. *et al.* A potent series targeting the malarial cGMP-dependent protein kinase clears infection and blocks transmission. *N*