Plan and Schedue for 2021 - v2

James Engleback

February 27, 2021

Contents

1 Overview

1.1 Aim and context

This project has two peices of work both with the aim to engineer a P450 BM3 mutant with some sort of metabolic activity towards an important herbicide for use in herbicide-resistant crops.

Herbicide resistant crops can be engineered by introduction of a P450 that can hydroxylate the target herbicide, priming conjugation with glutathione via Glutathione-S-Transferases (GSTs) and sequestration into the vacuole.

Given the promiscuity of GSTs, there may be flexibility in the hydroxylation site of the herbicide. Therefore, the specific aim of this project is to engineer candidate P450s to hydroxylate a herbicide at any postion for further testing in plants by Syngenta.

Candidate P450s can be transformed into a model crop (e.g. tobacco) and evaluated for efficacy in glasshouse-based herbicide challenge trials.

1.2 P450 Engineering Approaches

Two approaches are being taken to engineer and then test variants of P450 BM3 with activity towards mesotrione - an important herbicide. Each approach will be written into a thesis chapter.

One approach uses a combination of classical structure prediction and molecular docking methods as a simulation envorinment for virtual directed evolution driven by a genetic algorithm.

The other uses a deep learning model built to predict binding (pK_d) for a given enzyme sequence and ligand SMILES. Model predictions of pK_a towrads mesotrione can be used as a fitness criteria in virtual directed evolution experiments at a large scale.

Both approaches will generate pools of candidate mutants that will need to be tested for efficacy in the lab. Candidate mutants will be expressed, purified and evaluated for: K_d, K_m and K_{cat} towards mesotrione, and have product formation confirmed via LCMS.

2 Structure-Based Design

2.1 Overview

This aproach uses classical protein structure prediction methods and molecular docking to simulate herbicide binding in the acive site. The simulation is used to create a virtual directed evolution experiment where a genetic algorithm uses a heuristic fitness estimate to generate candidate pools of mutants with predicted herbicide binding.

2.2 Related Work

In the source material for template-based structure prediction and molecular docking Combs et al. [1] predict the structure of lysozyme mutants based on a crystal template structure using side chain repacking and loop remodelling. They then compare the accuracy of *rosetta* and *autodock vina* docking performance based on a ligand bound reference crystal structure. They find that the structure prediction methods used are sutable for predicting the structure of near mutants and that *rosetta* docking methods perform better than *autodock vina* but incurs a significant computational cost. *autodock vina* still performed reasonably well.

2.3 *enz*

The simulation environment is contained in a python package built for this project, *enz*. The scope of *enz* is template-based structure prediction using *pyrosetta* and molecular docking using *autodock-vina* as backends.

Currently, *enz* only uses side-chain repacking for structure prediction and rigid side-chain molecular docking using *autodock vina*. *enz* works robustly and is extensible to new and modified docking and structure-prediction methods. *enz* is ready to deploy in virtual directed evolution in its current state, but modifications that may significantly improve accuracy have been identified:

- **Flexible docking** - where *vina* treats active site side chains as flexible. This is can be imlemented with *vina*, and is a viable route of investigation.
- **Flexible (loop) remodelling** - where *pyrosetta* treats unstructured regions as loops and optimizes their conformation using Cyclic Coordinate Descent (CCD). *pyrosetta* has an implementation of CCD which can be integrated into the *enz* `refold` function.

2.4 *enz* Benchmark

Combs et al. [1] benchmark template-based structure prediction and ligand docking using both *rosetta* docking and *autodock vina* against mutant lysozyme strucutres.

A performance benchmark will need to be established to evaluate efficacy of *enz* and test experimental branches. The benchmark will be based on structure prediction and docking RMSD to reference ligand bound crystal structures of BM3, of which there are X. This can be generalized to other structures for which there also exists a structure of a close mutant in future.

Discrepencies between ligand conformations in BM3 crystal structures and molecular dynamics simulations exist and may suggest that the crystal structures represent one of several stable conformations. Therefore a suitable metric will have to be chosen carefully to reflect convential knowledge whilst avoiding bias towards *enz*. At the same time it should be sensitive in the range that *enz* improvements can affect. It would be very useful to plan a suitable metric with help from someone with domain knowledge in molecular dynamics, protein strucure and molecular docking - Sam, Linus and Nathan Kidley from Syngenta come to mind.

2.5 Virtual Directed Evolution Overview

Virtual directed evolution requires a sequence optimization algorithm and a sequence fitness function. In this case, a genetic algorithm is the sequence optimizer and mutant fitness is a score based on docking results of a target compound against the predicted structure.

The heuristic currently employed to estimate the desirability of each set of docking results is

$$score = \frac{1}{n} \Sigma_N \Delta G_i \times d_i \tag{1}$$

where ΔG is free energy of the interaction calculated by *autodock vina* (*kcal / mol*) and d is the distance between the heme iron and the C5 of mesotrione for n in binding poses. Where C5 is the target carbon for hydroxylation. It may be sufficient for use in the genetic algorithm but can be flexibly changed if needed.

This effectively ammounts to an average distance between the target carbon and the catalytic iron weighted by binding energy.

The genetic algorithim employed as a sequence optimizer randomly mutates input sequences at a single position randomly chosen from a set of active site residues to generate a pool of mutants. Fitness of each is queried using *enz* and the heuristic discussed above and the n fittest mutants are randomly recombined with eachother and randomly mutated again to repopulate the pool.

Mutant fitness is evaluated in parralel, enabling scale up to hardware capacity. It can run on a small virtual private server (VPS) and run 20 iterations with a batch size of 20 mutants over 2 days, though generally a large batch size results in faster convergence to good solutions.

2.6 Status, Readiness and Going Forward

An implementation of the described virtual directed evolution is ready to run in its current state. More accurate predictions enabled by changes to *enz* will result in a better pool of engineered mutants. Since there is only time to make and test one batch of mutants generated by this process in the lab, final modifications to *enz* will be made before the process is used to generate mutants to be made and tested in the lab.

3-5 days of work are required to attempt to implement the proposed changes to *enz* and set up the virtual directed evolution algorithm on a suitable peice of hardware. The small VPS can be used if necessary.

The described changes to *enz* will finish and virtual directed evolution will proceed in March. Work to construct expression plasmids for the candidate mutants can start after this. The candidate mutants will be subject to the same validation process as that used for those generated by machine earning. This process is detailed in **Mutant Validation**.

3 Machine learning-based Design

3.1 Overview

This piece of work aims to use machine learning model predictions of binding (pK_d) between an amino acid sequence and a ligand as a fitness function in virtual directed evolution experiments.

The model will be trained on in-house screening data of BM3 mutants' binding K_d with herbicides and herbicide-like compounds. A high throughput assay has been developed for this work and screening has begun.

Similar to candidate mutants genereated in the **Structure-based Design** section, candidate mutants with predicted binding towards mesotrione will be expressed, purified and assayed for K_d, K_m, K_{cat} and product formation with mesotrione.

3.2 Related Work

3.3 Screening

The screen employed here measures K_d between BM3 variants and ligands by measuring response of the P450s' UV-Vis absorbance profile to increasing ligand concentrations.

The screening assay is a 384 well plate analog of traditional cuvette-based titration. Eight concentrations per ligand is sufficient for a reasonably accurate K_d estimation, accommodating 48 compounds per plate. One plate takes 5 minutes to read in a *BMG FluoStar* plate reader, and another 10 minutes to transfer data to the dedicated computer which limits the screening rate to a maximum 4 plates per hour.

Over 2 days 12 plates (576 compounds) can feasible be set up by hand and screened against purified BM3 mutants.

During development, automation of compound dispensing using a *Labcyte Echo* was more accurate and reliable than dispensing by hand. Access to a *Labcyte Echo* in Patrick Cai's group has been arranged, so there is scope for automation of this step.

3.3.1 Assay Technical Details

The Assay takes place in clear 384 well plates and uses 30 µl of 10 µM BM3 in each well. Test compounds are added in DMSO so that the final DMSO concentration is 5 % in each well, which controls the mild binding activity between BM3 and DMSO. Assay buffer is 100 mM KPi *pH* 7 and 1 % bovine serum albumin (BSA) to mitigate BM3 precipitation. UV-Vis absorbance by the test compounds is corrected for in a control plate with only compounds in buffer. Currently, compounds are dispensed into plates using a multichannel pipette from a set of master plates that contain serial dilutions $\frac{1}{2}$ of each compound. Measurement of UV-Vis absorbance between 200 and 800 nm is a capability of most *BMG* platereaders, and is currently done on the Field group's *BMG FluoStar* on the third floor. The total time to read the plate and store the data takes 15 minutes.

Setup of compounds can be done a day prior to the assay. Once BM3 is added to the plate, the assay remains stable at room temperature for at least an hour.

A minimum of 5 μl of compound stock (10 mM in DMSO) is used per plate plus an additional 5 μl for the control plate per batch. Therefore a batch of several mutants screened against a compound set at once is economical. One plate of 48 compounds consumes 12 ml of 10 μM BM3 stock. BM3 purified from a 6 L expression will accommodate > 10 plates.

Analysis is automated and generates a report for each test compound and assembles the K_d calculations for each into a csv file for model training. **Figure 1** shows an example report of a binding interaction between a BM3 mutant and a compound, generated in the recent pilot assay.

Currently anomalous experiments are easy to identify and discard manually however an effort is being made to automatically detect and remove anomalous data points rather than the all eight traces.

3.3.2 Automating compound dispensing with the *Labcyte Echo*

During development, a *Labcyte Echo* acoustic liquid handler was used to dispense compounds accurately and reliably. I gained extensive experience working with this machine and can adapt the assay to use it fairly easily. Going forward, an effort to use the *Echo* liquid handling robot again will be made, which greatly improves accuracy and reliability of compound dispensing - see **Figure 2** for a report of a test using the *Echo* to dispense lauric acid with BM3 wild-type. A notable difference in the data produced using the *Echo* compared to that where compounds are dispensed by hand is the consistency of compound dispensing. In the figure the Michaelis-Menten curve has not fit correctly because it is anchored by its Y intercept at 0, however this is easily handled.

Because of the improvement in pipetting accuracy, I will attempt to re-establish the *Echo* as part of the assay. This will also eliminate the risk of mis-dispensing compounds when pipetting, which is a very real risk. Additionally it increases scalability of the assay. The Cai group's *Echo* is rela-

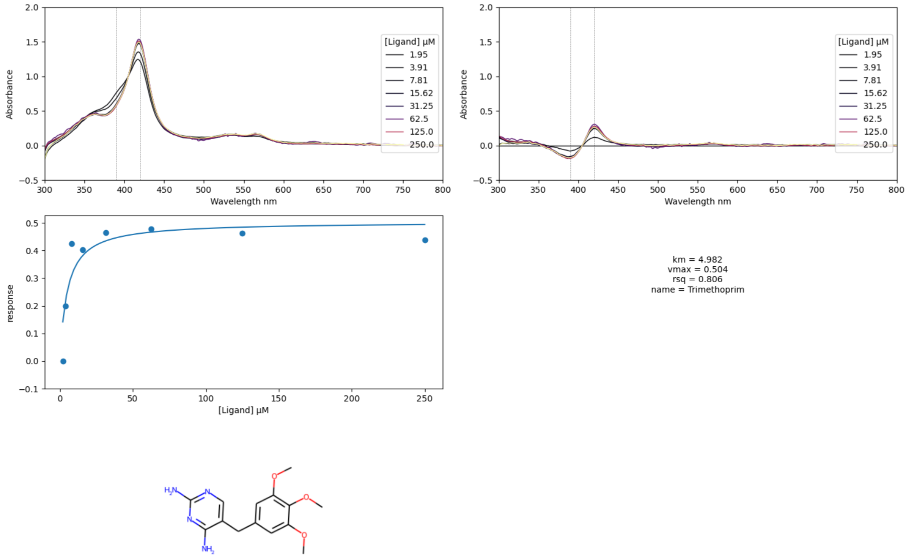


Figure 1: Example report from recent pilot assay showing response of BM3 mutants A82F to Trimethoprim

tively available compared to the FBRH's machine, so access should not be a problem.

Dispensing instructions for the *Echo* were generated with a python script which will need to be re-written. I plan on packaging it as a more flexible tool that can be applied to other assays in future.

I have been liaising with the person responsible for the *Echo* in the Cai lab and can make an attempt next week to incorporate it into my pilot screening experiment. If successful, the screen will proceed with use of the *Echo*.

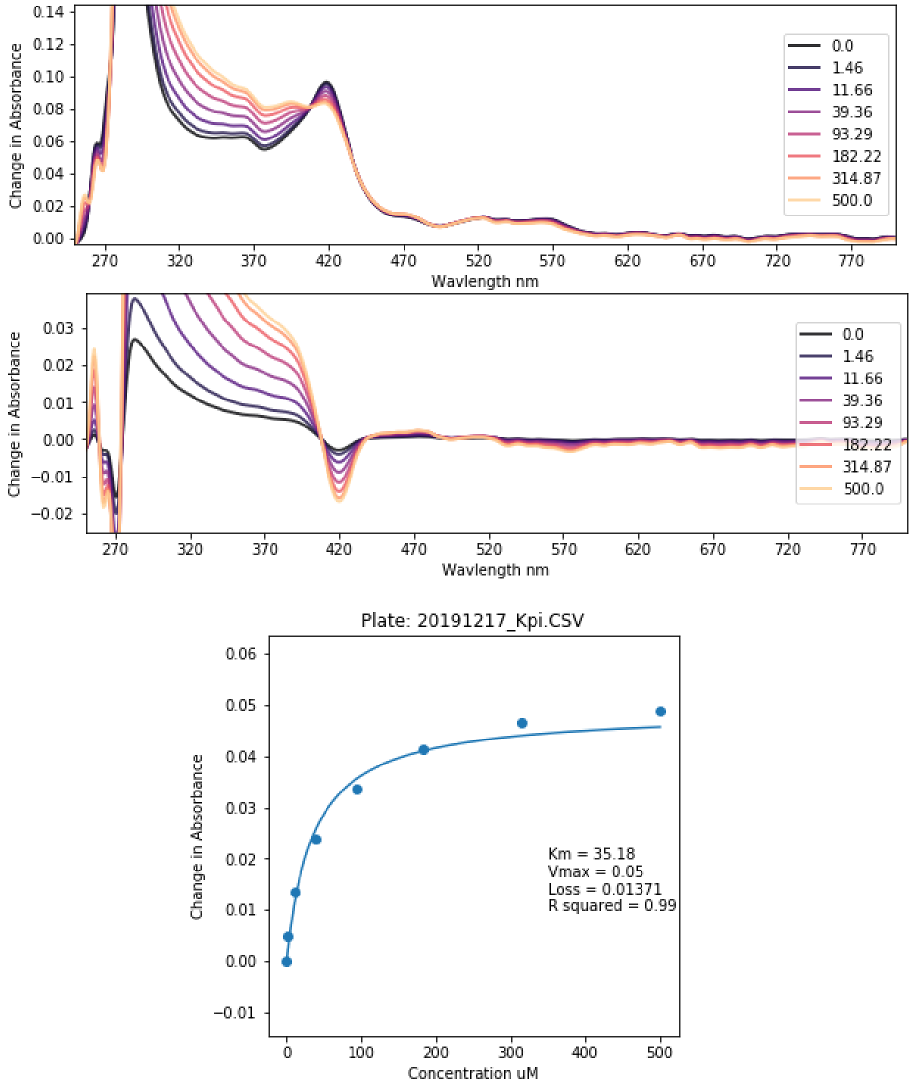


Figure 2: Report from an assay during development using the *Echo* to dispense lauric acid into wells.

3.3.3 Mutant Inventory and Purification

I currently have the following mutants purified and ready to screen:

- Wild type
- A82F
- A82F/F87V
- T268A
- T268N

I have cell pellets ready to purify for the following mutants:

- K97C
- P393A - 118g
- K97C - 88g
- A264E - 77g
- A82F - 50g
- F393W - 67g
- F393A - 36 g

Additionallly, I currently have plasmid stocks for the following mutants:

- A330P
- E267P
- F87A

Mutant expression plasmids are created by site-directed mutagenesis from existing stocks. Success rate is reasonably good and a round of mutations can be introduced in a week with some luck. To mitigate the chances of a failed set of reactions, mutagenic PCR is done in batches.

Expression is in BL21 (DE3) *E. coli* cells grown in autoinduction TB broth. 3 hours of growth at 37°C followed by 24 hours of growth at 25°C is sufficient to yield 50-120 g of cell paste, which can yield 1 μmol of pure enzyme per 6 Litre expression - enough for 10 plates (480 compounds).

Purification is one step nickel affinity chromatography and takes 2-3 days per mutant, however purifications can be batched. Based on prior batch purifications, maximum estimated purification capacity is 4 mutants in one week.

6 BM3 mutants can be purified from existing cell pellets over the course of March, during which time communal autoclaves are due to be serviced.

3.3.4 Compound Selection

Compound library diversity is required to minimize bias of the model when trained. Herbicide-likeness may also improve the models' accuracy over the chemical space occupied by herbicides. I have been using filtering rules based on *ref* and the *MaxMin* algorithm to pick n diverse compounds. This method can select a diverse set f n herbicide-like compounds from a given set.

This compound selection method has been applied to an in-house compounds library of 979 FDA-approved compounds for a pilot screen. Since 48 compounds pass the herbicide-likeness filtering, it is feasible to proceed with screening with compounds we already have with 9 additional in-house herbicides. Note that the Library has been stored at -80°C for 4 years and there is some concern over compound degradation and contamination.

The same compound selection method can be applied to the database of compounds available via *Molport*, a compound aggregator. Quotes (including lead time) of candidate screening libraries can be autogenerated using the *Molport* API in order to find a suitably cheap set. Compounds can arrive within 4 weeks of ordering and handling can be covered by a single COSHH form. *Molport* can deliver in many formats, including as 10 mM solutions in DMSO in a *FluidX* plate, which is most convenient to me.

Going forward, a set of 96 compounds from the in-house FDA library and the panel of 9 herbicides will be screened until it is demonstrated that additional compounds are required.

3.3.5 Pilot Screen - Initial Results

A screening set was selected from the in-house FDA library using the method described and combined with 9 herbicides from in-house stocks including mesotrione for a total of 96 compounds. Of this set, 48 were screened against three BM3 variants due to a planning oversight that can be mitigated next time.

Compounds were located in cold storage based on the manufacturer-specified positions, however they did not correspond to the serial numbers which were used to retrospectively identify each compound. This necessitates a second attempt at the pilot screen and a cataloging effort of the library, due to take place in the week starting 2 March.

Despite the mentioned issues the initial pilot assay was useful to develop an analysis package and gauge the resource requirements of the assay. As a result, it was decided that compound dispensing should happen the day before measurement to avoid the same time constraints encountered last time.

Going forward, the pilot assay will be re-attempted in the first week of March with the following differences:

- Compounds will be dispensed into plates using the *Echo*
- Compounds will be dispensed ahead of time to allow room for the *entire* screening set to be tested against the 5 purified BM3 variants.
- The correct compounds will be screened

Following this test, additional mutants will be purified and screened in March whilst additional mutants will be prepared. Screening will continue with the existing compounds until it is clear that purchasing of additional compounds is required.

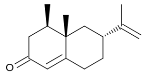
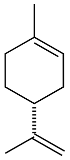
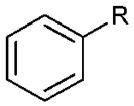
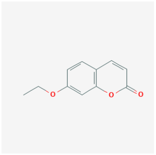
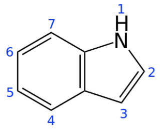
3.4 Additional Mutants for screening

Additional expression plasmids can also be assembled in March. Work is ongoing to identify BM3 mutants with known altered substrate scope that can be obtained with few point mutations relative to existing stocks.

The selection will be angled towards ring-containing substrates since most herbicides contain ring systems. The current goal is to ifentify 6 mutants for further screening. Literature is a rich source of mutants and 6 mutants with diverse substrate scope is faesible Whitehouse, Bell, and Wong [2].

Table 1 contains a selection of promising mutants found in literature so far. The *LVQ* and *GVQ* mutants show relatively broad substrate scope and can be built in two steps from existing stocks.

Table 1: Engineered BM3 Mutants

Mutant	**Substrate scope**	**Pic**
F87A/I263A/(A328I)	Nootkatone	
A328F	Limonene	
R47L/Y51F	Alkylbenzenes (no heteroatoms)	
R47L/F87V/L188Q (LVQ)	Coumarins	
A74G/F87V/L188Q (GVQ)	Indole, Lovastatinn, beta-ionine, organophosphates, polyaromatic hydrocarbons, chlorinated dioxins	

3.5 Model

The proposed model is combines two deep learning architectures that have proven efficacy in chemical data and sequence learning tasks. Graph neural networks are a class of model that are ideal for learning chemical data *todo: ref* whilst transformers are very effective on sequence learning tasks.

A model that combines transformers and graph neural networks has been constructed to estimate pK_d between a given enzyme sequence and ligand SMILES and an overview is in **Figure 3**.

3.5.1 Architecture

The model constructs a chemical graph from a smiles string, which is processed by a graph convolutional network and downsampled to a fixed-size embedding vector using set2set. Meanwhile the amino acid sequence string is downsampled using three 1D-convolutional layers and processed to a short, variable-length tensor which is downsampled into a fixed-size vector using the final hidden-state from an LSTM layer. The learned embedding vectors of the sequence and conatenated into a combined representation vector. A two layer perceptron provides a point estimate of binding likelihood. A diagram of the structure of the model is in **Figure 3**

3.5.2 Transfer Learning

Transfer learning is a popular approach to enhancing model accuracy and involves pretraining the model on a large, general dataset before transferring to a new task. The effect is a reduction in the number of samples required to train the model. Pre-training datasets do no necessarily need to be closely related to the target task, for example: models initially trained on the CIFAR 100 dataset (contains cars, animals etc. has been successfully re-tasked on phenotype identification from microscopy images. Transfer learning is predicted to drive commercial success *Andrew Ng NIPS 2016 talk*.

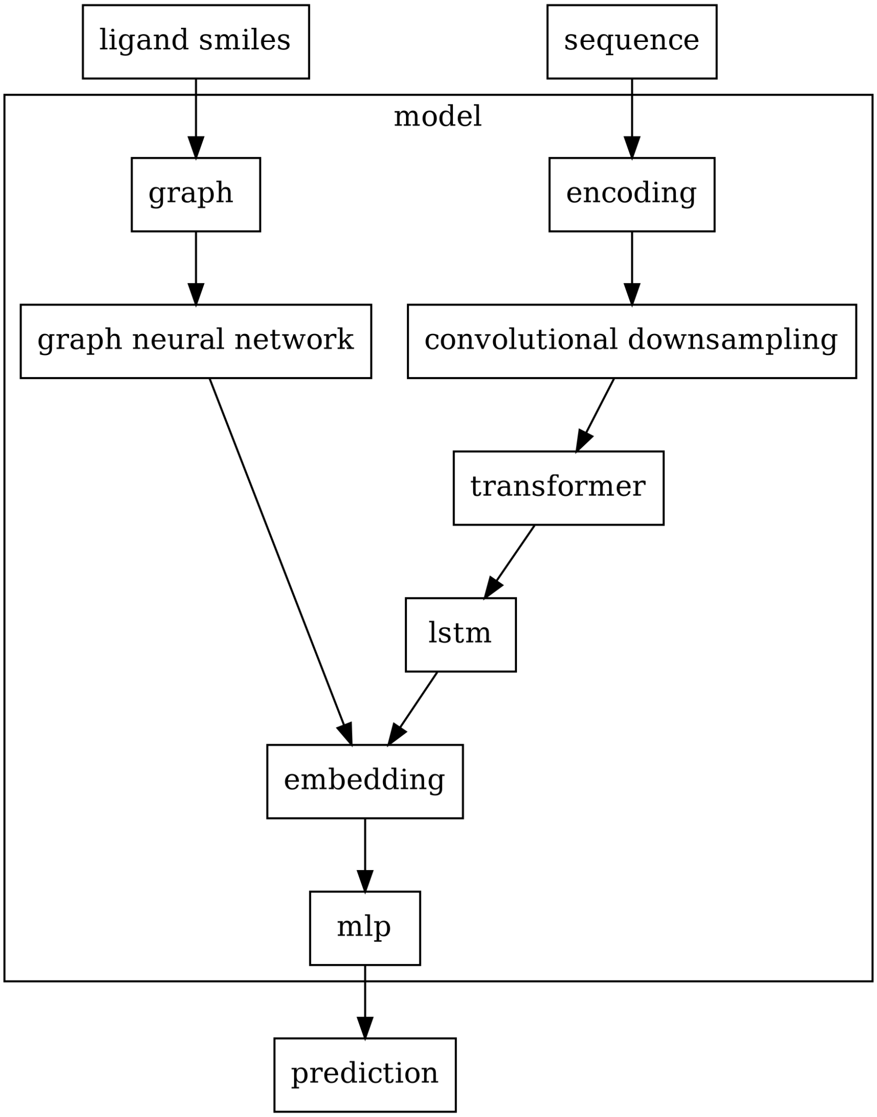
The pre-training dataset has been mined from KEGG and currently holds roughly 0.2 million data points of enzyme sequence and smiles for reactants and products. I've identified changes to the data miner that could increase the dataset size close to 10^6.

The dataset contains only positive examples of probable enzyme-substrate pairs which necessitates generative-adversarial training, a common machine learning technique.

3.5.3 Generative-Adversarial Networks (GANs)

In generative-adversarial training loops, the model competes with a generative-adversarial network (GAN) to discriminate between real data points and

Figure 3: Graph illustrating the layers of the predictive model used in this work with layer types indicated.



those generated by the GAN. Competitive training results in a GAN that generates realistic samples from a similar distribution to the training data whilst the predictive model predicts the likelihood of a given sample in the training distribution. This approach is utilised for sample generators (e.g. deepfake) and predictive models that have learned a data distribution (e.g. for anomaly detection).

There is scope to monitor training by assessing the quality of the generated sequences and molecular graphs, indicating performance of the entire system. Candidate models will be saved for later retraining on screening data.

3.5.4 Model-Based Mutant Design

To design new mutants with predicted activity towards mesotrione (or any other herbicide) the model's predicted pK_d between a candidate amino acid sequence and mesotrione is used as a fitness function. A genetic algorithm can be used to evolve a pool of new mutants for lab testing.

The model can evaluate a large number of binding predictions in parallel very quickly which enables a large scale virtual directed evolution experiment. In future work, alternative sequence search algorithms can replace the genetic algorithm for improved performance but genetic algorithms are the simplest case and should be sufficient for this use case.

Prediction uncertainty estimates are a result of *ref dropout bayes* and can be used to inform decisions based on model predictions.

3.6 Status and Going Forward

The pilot assay will transition into the screen proper in the first week of March using 9 herbicides including mesotrione and 48 herbicide-like compounds from the FDA library against 5 purified BM3 variants. An effort to use the *Echo* will be made.

Pre-training and re-training will be attempted by the second week of March. A set of candidate models will be saved and training parameters will be swept.

Models will be evaluated on performance on the screening set. If no suitable model can be produced, a fall-back option is to investigate simpler models constrained to a select few amino acid positions.

The next 6 mutants for screening will be purified in March, batching 3 purifications to a lab week. Screening all 6 purified mutants will be batched to a week.

Plasmid stocks of an extra set of mutants with known altered substrate scope will be designed and made in March.

Expression and purification of these mutants may last from April-May. After these mutants are screened, then the project will transition into a mutant-generating and testing phase.

A pool of mutants will be engineered using a virtual directed evolution experiment using model predictions of binding activity towards mesotrione as a fitness function.

The pool of candidate mutants will be made in the lab and tested for K_d, K_M and K_{cat} and have an indication of product formation.

4 Mutant Validation

1. Expression and purification using nickel affinity chromatography
2. K_d measurement with mesotrione using titration and UV-Vis spectroscopy
3. K_{cat} measurement with mesotrione by monitoring reaction NADPH consumption by UV-Vis spectroscopy
4. Product analysis with LCMS where a hydroxylation at any position is considered a hit

5 Schedule

5.1 Feb - work done and work outstanding

5.1.1 Pilot Screen

A pilot screen of 48 compounds against 3 BM3 mutants has been done. Lessons learned have helped inform the practicalities of the assay. I'm unsatisfied with the accuracy of dispensing compounds with a multichannel pipette so an **outstanding** task is to reimplement the *Echo* as part of the assay.

5.1.2 Compound Selection

48 compounds in the FDA Library are herbicide like, and will be used for screening alongside 9 real herbicides for now. The *Molport* compound selection program needs some modifications to access the *Molport* API to auto generate quotes and lead times, allowing me to choose the cheapest possible screening set.

5.1.3 Improvements to *enz*

Still outstanding. Work done is refactoring the *enz* `refold` function for extensibility to alternative folding methods. 3 Python Club sessions on *enz* were taught.

5.1.4 Virtual Directed Evolution with *enz*

Still outstanding. Work done is prepare a script that can run the experiment.

5.2 March

5.2.1 Spillover from Feb

- *enz* improvements
- run virtual directed evolution
- finish pilot with *Echo*
- set up lcms pipeline

5.2.2 Structure-based

In March I will prepare DNA stocks for mutants predicted using *enz*. The maximum number of rounds of site-directed mutagenesis will be three, so this work can be fit into one month. Any available slots for incubator and centrifuge time will be booked for next month. I will create benchmarks for mesotrione product detection by LCMS using BM3 A82F/F87V I have on hand. This establishes a pipeline for upcoming product detection experiments.

5.2.3 Machine learning-based

I will proceed with screening the 48 herbicide-like compounds from the FDA-approved library and the 9 herbicides I have with the 5 purified mutants. I will also purify and screen the cell pellets I have (5 unscreened mutants) and screen them too. I will decide on whether additional compounds should be acquired from molport and order if necessary. Lead time is 4 weeks so I will be able to screen these compounds with remaining mutants stocks in April. Early in the month I will decide on a final batch of mutants to screen and do the required DNA work over the course of the month. I will make sure that I have shakers booked for expressing these mutants as soon as they and the autoclave are available.

5.2.4 Validation prep

I will establish an LCMS method for product detection this month, ready for validating my mutants.

5.2.5 Targets

- March 7th - order primers for both sets of mutants
- March 30th - finish DNA work
- March 30th - Benchmark LCMS experiment of mesotrione and BM3 A82F/F87V
- March 30th - Write computational methods for *enz* and *rio*

5.3 April

April will be dedicated to expression and purification of mutants from both projects. By batching expressions and purifications I can work on several mutants concurrently. *rio* mutants have higher priority.

5.4 Writing

Write all lab methods.

5.4.1 Targets

- April 30th - Finish mutant expression
- April 30th - Have purified most *rio* mutants
- April 30th - have started all purifications
- April 30th - finish writing all lab methods

5.5 May

5.5.1 Structure-based

enz mutants finish purification in May. By the end of the month all *enz* mutants should be purified and ready for testing. Testing of these mutants takes place in batches of 3-4, to be completed by end of June.

5.5.2 Machine learning-based

All *rio* mutants are to be purified in the first weeks of May with higher priority than the *enz* mutants. The screen will happen in May. The screen can be done in 1 week, but I'm allowing 1 month for up to 8 mutants and 96 compounds for leeway. It'll be done in one batch as far as posible. Model training can begin as soon as the data is available. Some parameter searches will take place here, lasting one week. The trained model can be deployed with an existing genetic algorithm to generate fit mutants immediately with indicated model certainty for each. Mutant generation will be constrained to within 3 mutations of existing DNA templates. Primers for site-directed mutagenesis to be ordered immediately.

5.5.3 Writing

Make any necessary retrospective changes to methods. Start preparing results section based on screen.

5.5.4 Targets

- May 30th - Finish screening
- May 30th - Train model on screening data
- May 30th - Predict new *rio* mutants and order primers
- May 30th - Finish purification of all *enz* mutants, start testing

5.6 June

5.6.1 Structure-based

Test K_d, K_{cat} and indication of metabolite production for all *enz* mutants by end of month. Finish *enz* lab work.

5.6.2 Machine learning-based

Express all *rio* mutants by end of month. Begin purifications for all *rio* mutants. Begin testing K_d, K_{cat} and indication of metabolite production for all mutants by end of month.

5.6.3 Writing

Write results for *enz* mutants. Write results so far for *rio*.

5.6.4 Targets

- End of month - Finish all lab testing for *enz*
- End of month - Have begun all purifications for *rio* mutants
- End of month - Finish writing results for *enz*

5.7 July

5.7.1 Machine learning-based

Purification and testing of *rio* mutants must proceed as fast as possible in July, after which all lab work will finish.

5.7.2 Writing

I will write a discussion for *enz* and results and discussion for *rio*. I will also start writing introductions for each peice of work. I aim to reach a minumum viable product (MVP) for each introduction by the end of the month.

5.7.3 Targets

- Finish lab testing of *rio* mutants
- End of month - End all lab work
- End of month - Finish *enz* results and discussion to MVP
- End of month - Finish *rio* results
- End of month - MVP for both introductions

5.8 August

5.8.1 Writing

At this point, I should have finished writing methods and results sections for both projects and the discussion for *enz*. I will write a discussion for *rio*. I will finish introductions for each section and iteratively apply suggestions from readers. By the end of the month I will need a peice of work ready to submit.

5.8.2 Targets

- Finish *rio* discussion
- Finish *rio* introduction

- Finish *enz* introduction
- Ready to submit

References

[1] Steven A Combs et al. “Small-molecule ligand docking into comparative models with Rosetta”. In: *Nature protocols* 8.7 (2013), pp. 1277–1298.

[2] Christopher JC Whitehouse, Stephen G Bell, and Luet-Lok Wong. “P450 BM3 (CYP102A1): connecting the dots”. In: *Chemical Society Reviews* 41.3 (2012), pp. 1218–1260.