a. FINAL PROJECT REPORT

a. Date:
2015-07-13
b. Name:
Wanting Jiao
c. Project Title:
Develop new drugs for lung infections in patients with cystic fibrosis

Please copy the "Specific Objective(s)" statement, entered on your application form, in the space below.

1 Develop a structural model for the enzyme 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS) from *Pseudomonas aeruginosa* to provide a basis for inhibitor design (3 months)
2 Screen a virtual library of drug-like compounds to identify lead structures for drug design (4 months)
3 Evaluate lead drug compounds identified from the virtual screening (5 months).
4 Modify the structure of lead compounds to propose a series of novel drug compounds, utilizing the information obtained from the structural model for DAH7PS from *Pseudomonas aeruginosa* (5 months)
5 Synthesize lead drug compounds and test their potency in an enzyme assay and *in vivo* models (7 months)

The objectives was revised based on working progress, and the revised objectives are:

1 Develop structural models for both forms of the enzyme 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS) from *Pseudomonas aeruginosa* based on the crystal structure of type II DAH7PS from *Mycobacterium tuberculosis*.
2 Model the reaction intermediate molecule into the active site of both forms of *Pae*DAH7PS using induced fit docking protocol, in order to generate ligand-bound conformations for active site residues that are appropriate for virtual screening.
3 Set up two virtual screening calculations for each *Pae*DAH7PS enzyme, using both lead-like compound library (containing ~ 6 million compounds) and the drug-like compound library (containing ~15 million compounds).
4 Evaluate selected lead drug compounds identified from the virtual screening by experimentally testing their potencies against both forms of *Pae*DAH7PS.

Briefly describe how successful you were in achieving the stated objective(s). If the objective(s) was not achieved, explain why that is the case and describe what you did manage to achieve.

The computational objectives stated above were achieved successfully. Two forms of *Pae*DAH7PS were found (long and short forms) and homology models were generated for both forms based on crystal structure of *Mtu*DAH7PS. The reaction intermediate molecule was modelled in active site of each form of *Pae*DAH7PS and found to form very similar interactions with conserved residues. The results from these calculations were used in the virtual screening stage, to guide the search for compounds from the lead like library (containing 6 million compounds) with resembling features to reaction intermediate. Twenty compounds were identified from virtual screening that scored better than the reaction intermediate molecule, thus may be good ligands for *Pae*DAH7PS. Depending on availability some of these compounds have been ordered and will be tested in laboratory for potency soon.

Originally it was planned that another larger compound library (the drug like library which contains 15 million compounds) will also be screened, however, due to technical difficulties in the early stages of virtual screening set up, it became impractical to conduct another virtual screening calculation within the time frame of this project. However, the lead compounds identified from screening the smaller library gave good initial directions in finding potent drug structures.

Experimentally we have expressed both forms of *Pae*DAH7PS and preliminary testing showed that both enzymes are active. The long form *Pae*DAH7PS purification protocol was successfully established and suggested it adopts tetrameric quaternary structure. To study the structure of the long form *Pae*DAH7PS, initial attempts at crystallization produced crystals diffracted at 3.3 Å; crystallization conditions will be optimized further to yield better crystals. The purification protocol for the short form is still being optimized. The short form enzyme showed quite different property in the expression and purification process to the long form, this might be caused by the potentially different quaternary structures adopted by the two forms of *Pae*DAH7PS.

Briefly describe any interesting outcomes which might not have been considered in your original objectives (if any).

1. ***P. aeruginosa* have two forms of type II DAH7PS.** Initially in the original proposal, only one form of type II DAH7PS was anticipated. From sequence studies and experimental characterizations, we found that there are two forms of type II DAH7PS present in *P. aeruginosa*, which were named long and short forms based on the length of their amino acid sequences. The long form is very similar to the type II DAH7PS from *M. tuberculosis*, only missing the equivalence of the N-terminal extension in *Mtu*DAH7PS that caps the tight dimer interface. However, the short form *Pae*DAH7PS lacks the equivalence of the helices in *Mtu*DAH7PS that make up tetramer interface, which suggests that the short form *Pae*DAH7PS may adopt a different quaternary structure to the observed tetrameric form from *Mtu*DAH7PS, hence may also be regulated differently to both the long form *Pae*DAH7PS and *Mtu*DAH7PS. The reason for *P. aeruginosa* to have two forms of type II DAH7PS is unclear. To address this we are currently conducting knockout studies to determine gene product function.
2. **New combination approach for virtual screening.** Originally it was planned that the entire compound library (which contains millions of compounds) to be screened against both forms of *Pae*DAH7PS using virtual screening workflow from Schrodinger Suite. The virtual screening workflow is a structural-based method that conducts docking calculations of ligands in three stages with increasing precisions, and is considered a fast method for screening large library compounds. However, during this project, we found that screening of the smaller lead-like library, which contains 6 million compounds, took more than three months to run (and was still not finished by the end of the third months). Therefore, a new approach to virtual screening was required. The e-pharmacophore method, which was based on the fast ligand-based pharmacophore screening and also took into account contribution of surrounding residues in the enzyme active site by means of "excluded volumes", was considered to be a good option. This method was used as the first stage in screening, filtering out ligand structures that do not contain the desired functional groups at appropriate positions. Outputs from e-pharmacophore method (significantly reduced in number of compounds) were then subject to structural-based virtual screening, to identify lead structures that form the most optimized interactions with the active site of *Pae*DAH7PS. This method worked well with the system under investigation, and we successfully identified several interesting compounds to be tested against *Pae*DAH7PS as inhibitors.
3. **Novel inhibitor structures that differ from the reaction intermediate.** In previous studies for DAH7PS inhibitors, the inhibitor structures were always designed based on the structures of molecules known to bind tightly to the enzyme, i.e. the substrates, transition state or the reaction intermediate. In this project, we were able to identify novel compounds that contain different structures and functional groups. These structures would not have been designed if purely based on molecules that are known to bind. Once the lead compounds have been shown to be potent *in vitro*, we will move into direct cell studies so that we can assess the potency of these inhibitors *in vivo*.