Hello,

At my last committee meeting in march, we agreed that I would continue work on my aims, and submit a written progress report in leu of of an additional in person meeting.  Below is a summary of my current progress toward my thesis aims.

My work on RosettaLigand can be divided into 3 manuscripts, all of which are currently at different stages. Brief summaries of the plans for these manuscripts are below. My goal is to have the first manuscript in ready for submission within the next 2 months, at which point I would ask for permission to defend in late January, and the second and third manuscript would be completed in parallel with the writing of my thesis. If the progress towards the first manuscript falls substantially behind the anticipated schedule, I would instead ask that the late January defense be replaced with a committee meeting at the same time.

The first manuscript will describe the "low resolution" grid based score function.  Development and optimization of the new score function is complete, and the manuscript in preparation will describe in detail the design and implementation of the new scoring function.  Additionally, benchmarking results will be presented to illustrate the benefits to speed and sampling efficiency afforded by the new scoring function. The benchmarking experiments necessary to the completion of this manuscript are currently underway.

The second manuscript is a benchmarking study of the impact of recent additions to the high resolution Rosetta energy function on ligand docking performance. In the 6 months, several major changes have been made to the Rosetta energy function, including the implementation of energy terms that use the position of electron orbitals (unpublished work by Steven Combs).  It is hypothesized that the ability to model salt bridge, pi-pi stacking and cation-pi interactions conferred by the new energy terms will have a significant positive impact on ligand docking performance, but the impact of these terms has not yet been rigorously benchmarked.  The benchmarking experiments necessary to complete this manuscript are also underway.  Both papers will use a benchmark of 43 proteins from the CSAR dataset, selected as no waters or co-factors are present in the binding interface.

The third manuscript describes the implementation of a method to directly integrate structure based and ligand based virtual high throughput screening methods. In this method, ligands are initially docked using the new grid based scoring function I developed, and the docking poses are refined using the high resolution energy function with the addition of the new orbital based scoring terms. The resulting predicted ligand poses are used to generate fingerprint descriptors based on the pairwise scores of pairs of protein and ligand atoms. In addition to the protein-ligand interface descriptors, traditional cheminformatics descriptors are computed for each ligand describing the chemical and geometric properties of the ligand. The combined set of descriptors is used as the input to a neural network which is trained to predict –log(Ki). It is hypothesized that the combination of structure and ligand based descriptor information will counteract limitations in the ligand docking energy function, and improve the ability of Rosetta to predict the binding affinity of compounds in a vHTS scenario. The software development associated with this project is complete, but the optimization and benchmarking associated with the complete protocol continues. Specifically, I am investigating methods for correctly selecting protein-ligand binding poses for both the training of the neural network, and for scoring. This is a challenging problem, as the correct identification of the binding pose is critical for a high quality affinity prediction.

Sincerely,

Sam DeLuca