**Proposal**

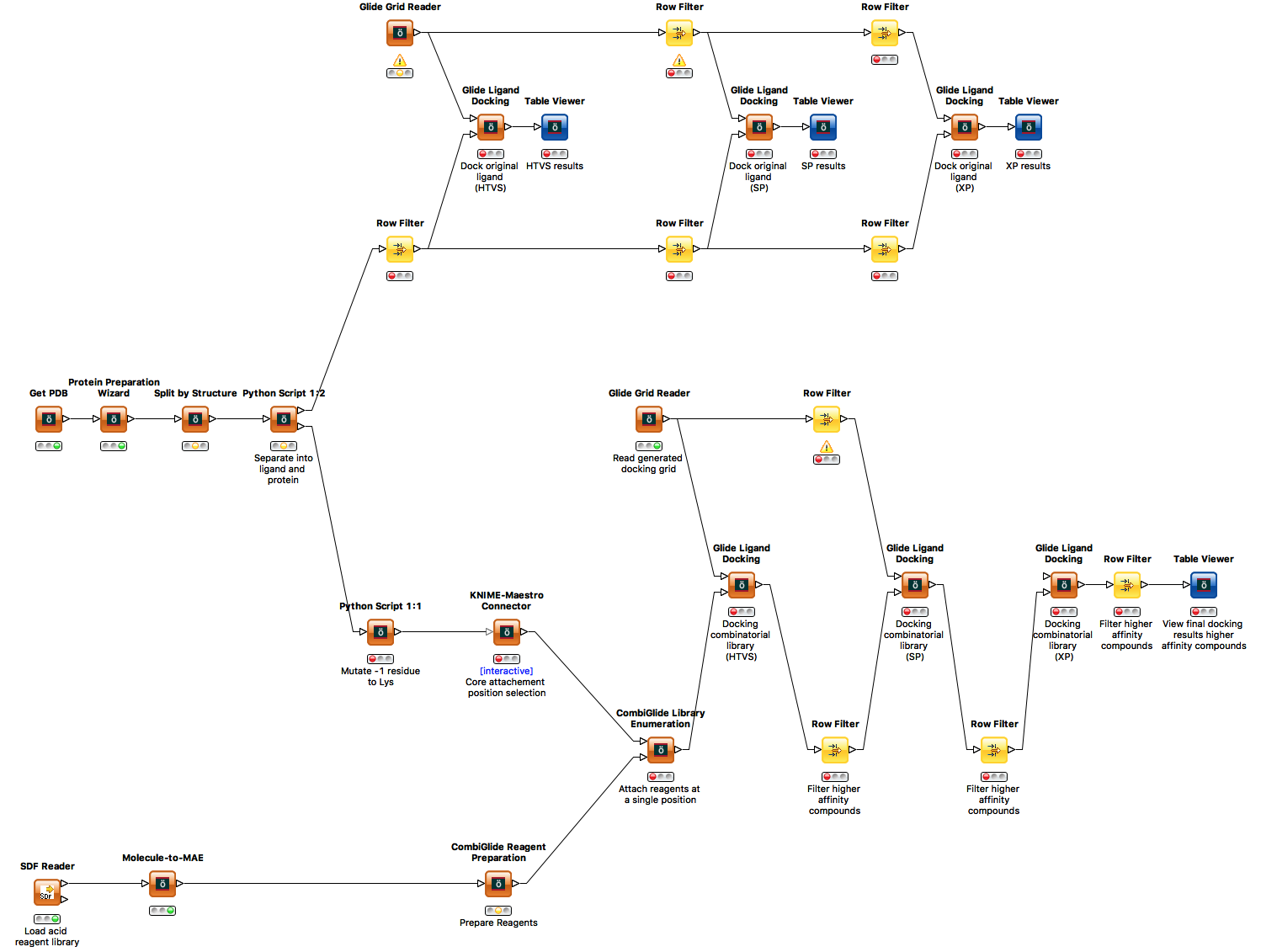
The goal of this project is to develop a novel method for identifying peptides that are potential drug targets for inhibition of PDZ domain-containing proteins. The PDZ domain is involved in interactions with other proteins and is generally responsible for co-localization and regulation. A multitude of proteins exist with variants of this domain, each with a specific target protein(s). The most promising inhibitor compounds have been small peptides which match the C-terminus tail of the particular PDZ domain’s target protein. However, two key problems have prevented any PDZ domain inhibitors from entering or succeeding in clinical trials: insufficient selectivity to the particular protein target and low binding affinity. PDZ domains are largely conserved and bind to a limited set of the C-termini target motifs; thus even leads which show positive therapeutic effects, will bind promiscuously and can cause disastrous side effects––this one of the reasons small molecules have been unsuccessful inhibitors. In addition, transient PDZ interactions lead to generally low binding affinity, typically in the micromolar range.1 Previous studies have demonstrated potential for coupling various carboxylic acids to lysine residues on the peptide ligands. This has unlocked a multitude of modified peptides one of which may have increased affinity and specificity for a particular protein target. We seek to rapidly explore this new space using virtual combinatorial screening of modified target peptides––those which are already known to bind the protein. The use of *in silico* screening to narrow down potential targets will drastically reduce the number of peptides we need to synthesize, purify, and characterize in the lab. Each synthesis procedure must be carefully planned out and complications may arise depending on the particular modification, resulting in time consuming procedures. Molecular modeling and docking software will enable us to traverse through modification combinations quickly, resulting in leads which can be further analyzed for binding energetics before picking a few to synthesize and test empirically.

Two proteins of particular interest to our group are GIPC (GAIP interacting protein, C terminus) and CAL (CFTR-associated ligand). Our laboratory has done significant investigation into GIPC, a protein which has been shown to be unregulated in and contribute to pancreatic and breast cancer. We, and our collaborators, have developed peptides with promise as anticancer treatments; however, we do not currently have access to GIPC for binding studies, and this empirical data is required to verify the success of our computational process. We do, however, have access to CAL, and previous studies have demonstrated significant promise in utilizing CAL inhibitors to extend the half life of mutated cystic fibrosis transmembrane conductance regulator protein (ΔF508-CFTR)3. The mutation, which causes cystic fibrosis, an autosomal recessive disease, accelerates the breakdown rate of CFTR, causing an abnormal Cl- conductance across the plasma membrane of epithelial cells. However inhibition of CAL has been shown to slow down the breakdown rate and, at least to some extent, remedy the problem.5 Our preliminary target for this project is developing CAL inhibitors due to ease of verification and previous literature on modified peptide inhibitors. However, if the computational workflow demonstrates success for CAL it is relatively trivial to apply it to GIPC and other PDZ domain-containing proteins down the road.

Our screening approach is rooted in our previous study which demonstrated the feasibility of mutating residues on the target peptide for CAL to a lysine and coupling its free amine with an organic acid to create a stable amide (similar to a peptide bond). The initial study coupled the mutated lysine with the simplest carboxylic acid (acetic acid) and many of the residues tolerated this modification. Four of them even displayed a modest increase in binding affinity. This study was a proof of concept that the peptide would bind in a similar conformer with a modified lysine mutation, confirmed by crystal structures of the complex.1 Therefore, any stable acid could feasibly be coupled with the lysine residue(s), effectively creating a branch off the core peptide where novel groups can be attached. This significantly expands the chemical space and is not limited by only amino acids in standard peptides. Additionally, the branched nature of this type of modification allows the added chemical groups to interact with part of the protein that is further away from the conserved, active site. This has the potential to increase both affinity and specificity of the compound if the attached group has favorable intermolecular interactions to the particular protein. A framework has thus been established for exploring modified versions of this refined peptide; however synthesizing a large number of these peptides, especially for peptides with modifications at multiple loci, would be practically infeasible in an academic research setting. Utilizing this computational approach to narrow down the vast array of possible compounds to a few promising leads will drastically increase the likelihood of finding a viable drug target.

We have developed a workflow using KNIME, an open-source data analytics and integration tool2, in concert with Schrödinger, a commercial molecular modeling package.4 An outline of the workflow can be seen below (figure 1). The workflow automates protein and ligand preparation, combinatorial enumeration, docking and scoring. Crystal structures of CAL complexed with modified peptides from our previous study will be utilized as inputs. Protein preparation is a standard process of adding missing hydrogens and heavy chain atoms, removing invalid water molecules, assigning bond orders, etc. Next, the ligand structure is extracted and one, or more, of the residues is modified to a lysine which will be attachment points for acids. We acquired acid libraries from chemical manufacturers, pre-filtered by pharmacological properties. The workflow further filters the library with general rules which prevent particular acids from being feasible in peptide synthesis. Each 2D structure in the library is then converted into multiple 3D conformers based on energetic favorability. Reagent attachment positions on the ligand peptide are defined and combinatorially enumerated with the acid library using CombiGlide.4 Once enumerated, the workflow docks each modified compound to the prepared protein with a set of predefined constraints on the active site, such as required hydrogen bonds and spacial occupation constraints. Glide docking can be performed at three increasing levels of precision, each with longer computing times: HTVS (high throughput virtual screening), SP (standard precision), and XP (extra precision). The workflow funnels the results by first screening all compounds using HTVS and comparing the resulting binding score to the unmodified peptide. Those, which exceed the original in affinity, pass on to SP and then XP docking protocols which yields our final assortment of leads.

We cannot currently process our full acid libraries due to insufficient computing power, but plan to execute the software on the Dartmouth Discovery computing cluster. KNIME allows asynchronous execution of the workflow and distribution of resources across a cluster and would enable us to rapidly get lead compounds while the process is running. The overall goal of this project is take leads from the computational software and test their empirical binding affinities. If the projects succeeds (for some or all compounds) we can validate the approach and identify points of error. However, this is not simply a proof of concept, it will hopefully bring us closer to developing a therapy for a variety of currently untreatable ailments.

**References**

**Figure 1: KNIME Data Flow Representation**

**References**

1 Amacher JF, Zhao R, Spaller MR, Madden DR (2014) Chemically Modified Peptide Scaffolds Target the CFTR-Associated Ligand PDZ Domain. PLoS ONE 9(8): e103650. doi:10.1371/journal.pone.0103650

2 "About KNIME." KNIME | About KNIME. N.p., n.d. Web. 11 Jan. 2017. <https://www.knime.org/about>.

3 Cushing, Patrick R., Lars Vouilleme, Maria Pellegrini, Prisca Boisguerin, and Dean R. Madden. "A Stabilizing Influence: CAL PDZ Inhibition Extends the Half-Life of ΔF508-CFTR." Angewandte Chemie 122.51 (2010): 10103-0107. Web.

4 "Overview | Schrödinger." Overview | Schrödinger. N.p., n.d. Web. 11 Jan. 2017. <https://www.schrodinger.com/about>.

5 Piserchio, Andrea, Abigail Fellows, Dean R. Madden, and Dale F. Mierke. "Association of the Cystic Fibrosis Transmembrane Regulator with CAL:  Structural Features and Molecular Dynamics†." Biochemistry 44.49 (2005): 16158-6166. Web.

**Budget**

|  | Name | Description | Cost |
| --- | --- | --- | --- |
| Software | **Schrödinger Small Molecule Discovery Suite** | Gives access to AutoQSAR, Canvas, ConfGen, Core Hopping, Desmond, Epik, Field-based QSAR, Glide, Jaguar, Jaguar pKa, LigPrep, MacroModel, Phase, Shape Screening, Prime, PrimeX, QikProp, QSite, SiteMap, Strike, and XP Visualize. Additionally it gives us the ability to run multiple docking jobs at a time to parallelize the process of virtual screening. | $‎7,500.00 |
| Hardware | **Dartmouth Discover Computing Cluster Node** | Gives access to a 16 core, 64G Ram, 1TB hard drive unix node on the computing cluster. Assuming available resources, we can use up to four times the computing space purchased. This computing power should be able to handle the enormous amount of computing time required to dock these molecules. | $‎4,900.00 |
| Personell | **One Undergraduate Biochemistry Research Assistant** | Paid $15/hr X 20 hours per week X 40 weeks for the year (4 Dartmouth terms). Responsible for all computational development, execution, and monitoring. Lead compounds will be synthesized by others. | $‎12,000.00 |
| Total |  |  | $‎24,400.00 |