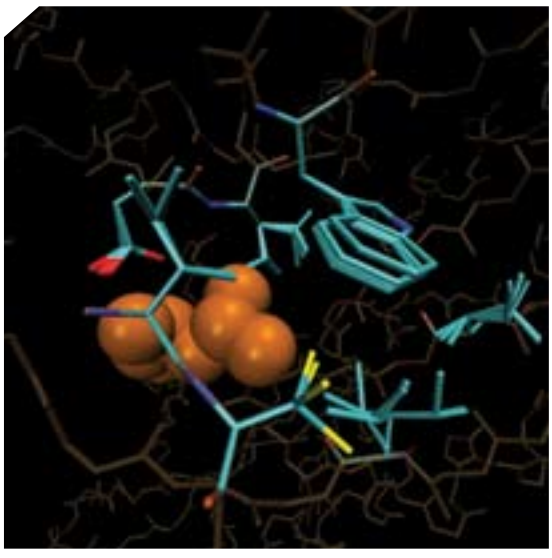

Enzyme redesign provides a good test of our understanding of proteins. We are developing algorithms to plan structure-based site-directed mutations to a protein's active site in order to modify its function. Natural NRPS enzymes make natural antibiotics, antifungals, antivirals, immunosuppressants, and antineoplastics. Redesigned NRPS enzymes should enable combinatorial biosynthesis of novel pharmacologically-active compounds, yielding new leads for drug design.

COMPUTATIONAL REDESIGN OF NATURE'S MOLECULAR ASSEMBLY LINES

ENZYME REDESIGN provides a good test of our understanding of proteins. The Donald Laboratory develops new algorithms to plan structure-based site-directed mutations to a protein in order to modify its function. They develop general planning software that can reprogram the specificity of many proteins, including "NRPS domains," whose products include natural antibiotics, antifungals, antivirals, immunosuppressants, and antineoplastics. These engineered enzymes should enable combinatorial biosynthesis of novel pharmacologically-active compounds, yielding new leads for drug design.

Redesign of NRPS enzymes offers the opportunity to reengineer biosynthetic pathways, greatly increasing the number and types of NRPS products, specifically, to develop new libraries of antibiotics. Bruce Donald, William and Sue Gross Professor of Computer Science and Professor of Biochemistry, leads a laboratory of students and postdocs in reprogramming NRPS enzymes using *K** ("K-star"), an ensemble-based protein redesign algorithm. Structure-based protein redesign algorithms, such as *K**, exploit the high-resolution protein structures as well as biophysical modeling. When Donald's lab applies *K** to an NRPS, or, more generally, to other proteins, they modify the active sites to switch the specificity of the amino acid-accepting domains from their natural substrates to different amino acids. The modifications are planned and analyzed *in silico*, using novel geometric algorithms. The Donald Lab's "enzyme reprogramming" could allow the modified NRPS to synthesize different modified peptides. Exploration of the combinatorial space of new NRPS "programs" will generate a large number of new compounds, which could then be screened for pharmaceutical activity.

How does *K work?**

Realization of novel molecular function requires the ability to alter molecular complex formation. Enzymatic function can be altered by changing enzyme-substrate interactions via modification of an enzyme's active site. *K** searches over possible active site mutations and combines a statistical mechanics-derived ensemble-based approach to computing the binding constant with the speed and completeness of a branch-and-bound pruning algorithm. Two graduate students in Donald's lab, Ivelin Georgiev and Ryan Lilien (now an Assistant Professor of Computer Science and Medicine at the University of Toronto), developed an efficient deterministic approximation algorithm, which approximates the binding constant to arbitrary precision. To test their predictions, Georgiev worked with two other graduate students in Donald's laboratory, Cheng-Yu Chen (Department of Biochemistry) and John MacMaster (CS), to redesign the phenylalanine-specific adenylation domain of the NRPS Gramicidin Synthetase A (GrsA-PheA). Using predictions made by Georgiev and Lilien, Chen and MacMaster create the mutant constructs by cloning, expressing, and purifying the mutant proteins from the PheA gene. Having obtained the purified proteins, they then subject them to biochemical activity assays and high-field solution-state nuclear magnetic resonance (NMR).

A major challenge has been the development of accurate ensemble-based redesign algorithms that efficiently prune mutations and conformations. *K** flexibly models both protein and ligand using rotamer-based partition functions for application in enzyme redesign, the prediction of protein-ligand binding, and computer-aided drug design. The *K** ϵ-approximation algorithms can prune the vast majority of conformations from more computationally expensive consideration, thereby reducing execution time and making a mutation search that considers both ligand and protein flexibility computationally feasible.

While the correctness of *K** is clear from statistical mechanics over the ensemble of protein and ligand conformations, before this algorithm it has been difficult and expensive to compute partition functions accurately for large systems. One key contribution of Donald's lab is a proof that a large number of conformations and mutations may be pruned while still approximating the partition function to within a multiplicative factor of $(1-\epsilon)$, where the error bound ϵ may be chosen by the user of the software.

*K** is biologically accurate too. Ensemble scoring, using a rotameric approximation to the partition functions of the bound and unbound states for GrsA-PheA, was used to predict binding of the wild-type protein to switch the enzyme specificity toward two novel substrates using several novel active site sequences computationally predicted by searching through the space of possible active site mutations. The top-scoring *in silico* mutants were created *in vitro* and binding constants and catalytic activity were experimentally determined. Most of the tested mutations exhibited the desired change in binding specificity or catalytic specificity.

Donald and his students are developing a general planner that can reprogram the specificity of NRPS domains, whose products include natural antibiotics, antifungals, antivirals, immunosuppressants, and antineoplastics. Fundamental to this algorithm is a novel extension of Dead-end elimination (DEE) (called "MinDEE") to include provable bounds on pruning correctness when the final rotameric conformations are energy-minimized.

For more information, see www.cs.duke.edu/donaldlab/