DOE FINAL TECHNICAL REPORT
Report Number DOE-UW-0005155

Grand Identification Number: DE-SC0005155
Project Title: Protein Scaffolding for Small Molecule Catalysts
Recipient Institution: University of Washington
Principal Investigator: David Baker, PhD
Period of Grant: September 15, 2010 – September 14, 2014

Description of Project

We aim to design hybrid catalysts for energy production and storage that combine the high specificity, affinity and tunability of proteins with the potent chemical reactivities of small organometallic molecules. The widely used Rosetta and RosettaDesign methodologies will be extended to model novel protein / small molecule catalysts in which one or many small molecule active centers are supported and coordinated by protein scaffolding. The promise of such hybrid molecular systems will be demonstrated with the nickel-phosphine hydrogenase of DuBois et al. We will enhance the hydrogenase activity of the catalyst by designing protein scaffolds that incorporate proton relays and systematically modulate the local environment of the catalytic center. In collaboration with DuBois and Shaw, the designs will be experimentally synthesized and characterized.

Collaborators:

Daniel DuBois, Institute for Interfacial Catalysis, PNNL
Wendy Shaw, Fundamental and Computational Sciences Directorate, PNNL

Objectives

1. Develop methods for design of protein / small molecule hybrid catalysts

a. Incorporate non-protein chemistries and geometries into Rosetta design
b. Develop database matching approach for assembling small molecule active centers within previously determined structures
c. Develop de novo approach for constructing ideal protein scaffolding around small molecule active center
d. Develop methodology for proton relay design
e. Implement methods for computational validation of designed hybrid molecules

2. Apply methods to design protein scaffolding for nickel-phosphine hydrogenase mimic

a. Design stable protein scaffold around nickel-phosphine active center
b. Design secondary pendant base to aid stabilization of hydride
c. Engineer hydrogen bond relays to deliver protons to/from the active center
d. Systematically alter the chemical environment of the active site

3. Experimentally characterize designed hybrid hydrogenases

a. Experimentally synthesize designed hybrid catalysts
b. Characterize structure
c. Determine catalytic activity (To be carried out in DuBois and Shaw groups)

4. Based on experimental feedback from Aim 3, design and characterize a second generation of hybrid hydrogenase catalysts with improved catalytic activities, and continue to iterate between experimental characterization and computational design.

Report

The Baker lab has successfully addressed the major computational objectives related to modeling covalently linked peptide / small molecule constructs, designing peptide sequences for such constructs, and computationally validating the folding pattern of such hybrid designs.

We have successfully incorporated non-protein chemistries and geometries into Rosetta design in a number of ways (Objective 1a). The simplest method entailed modeling all or part of the linked small molecule as an amino acid with some similarity to one of the twenty natural amino acids. For example, we have modeled tris-bipyridine Ruthenium complexes, among the most studied compounds in existence, using a bipyridine non-natural amino acid which has similar behavior to Tyrosine, allowing us to take advantage of Rosetta's extensive knowledge-based force-field and kinematics for, e.g. tyrosine. A second, related method was the use of more exotic non-natural amino acids that are not analogous to natural amino acids, increasing the diversity of complexes that can be modeled. For example some of the DuBois/Shaw constructs can be modeled as exotic non-natural amino acids, but the additional flexibility comes at the cost of less accurate energetics and kinematics. The third and most general approach is to model arbitrary connectivity between a small molecule and the peptide. We are now able to model such arbitrary chemical attachments in Rosetta, however, the user must specify information about the energetics and kinematics of such constructs in the form of constraints and/or rotamer libraries.

Computational methodology for database-matching to assemble small molecule active centers within previously determined structures (Objective 1b) has been developed. For many small molecule hybrid constructs, this can be accomplished in a straightforward way with the Rosetta Matcher. However some constructs were problematic with the established methodology, particularly metal binding complexes that were of particular interest in this project. An alternative matching approach based on branch-and-bound search and inverse kinematics was developed to address limitations with respect to metal binding sites; most notably, the new method can produce designs with more interactions to the metal ligand. We have designed and synthesized a large set of de novo metal binding proteins using a single bipyridine non-natural amino acid in conjunction with several natural amino acids to form novel iron binding sites in existing protein scaffolds. Some of these designs followed the published RosettaMatch protocol, with which we produced designs coordinating the iron at four of the six octahedral binding site, and the remainder of the designs, created with the new matching protocol developed as part of this project, satisfy five or all six of the iron binding sites.

We have successfully modeled and designed novel tree-like protein topologies with central small molecule hub and protein spokes (Objective 1c). The core challenge in this work was modeling symmetrical systems in Rosetta in a way that was compatible with the novel non-protein chemistries described in the

preceding paragraph. Such symmetrical modeling is now robustly implemented in Rosetta in such a way that we can perform *de novo* protein scaffold generation, leveraging many of the same methods used in Rosetta for *ab initio* and sparse-NMR structure prediction. We are now able to build peptide scaffolding onto a small molecule hub in a way that matches the underlying symmetry of the small molecule. We have achieved experimental success with dimeric and trimeric peptide/small molecule systems and are able to model and design D2 symmetries as in the proposed Dubois / Shaw constructs, but experimental synthesis and characterization was problematic for the DuBois / Shaw constructs. We have incorporated some new technology into the fragment assembly protocol for use in protein design, most notably merging sequence information from the fragment database into the design positions at which the fragment is inserted.

We have successfully implemented several methods for *in silico* validation of designed hybrid molecules (Objective 1e). A test for stability of the folded construct has been implemented in Rosetta in which local sampling is performed starting from the designed topology. If the most energetically favorable conformations are clustered near the design, the test is considered successful. We have applied this methodology to designed symmetrical peptide assemblies and experimental characterization is underway. The most stringent computational test of a design is a forward-folding validation method similar for hybrid constructs similar to that used in other de novo design efforts. In forward-folding, *ab initio* structure prediction methodology is applied to the designed molecular topology and sequence. If the designed sequence is predicted to fold into the designed structure, the test is considered successful. We have implemented and tested this procedure with several covalently linked constructs and have preliminary evidence that designs that fold well *in silico* will also fold as desired in the test tube.

We have applied the above methodology in several ways to produce designs for a peptide and nickel-phosphine hydrogenase mimic. Though the ultimate goal is to incorporate proton relays into the peptide scaffold in order to improve catalytic performance, the main goal of the first round of design was to produce on or more stable scaffolds around the nickel-phosphine center (Objective 2a). A number of designs were provided based on peptides of known structure assembled with D2 symmetry around the catalytic center. A more ambitious set of *de novo* designs incorporating a pendant histidine amine designed to interact productively with the catalytic center were also produced (Objective 2b).

<u>Progress towards Objectives 3 and 4:</u>

By the end of our funding period we improved our capacity to model, design, synthesize and characterize hybrid peptidic structures. We verified a number of successful symmetric peptide + small molecule designs, including some incorporating catalytic units.

The course of the study involved navigating wide-ranging difficulties and technical hurdles. Some of these problems are common to protein design, but symmetric peptide / small molecule constructs presented specific concerns we were forced to address. For example, short protein sequences (peptides) are more prone to aggregation (less capacity to form hydrophobic cores, even with symmetry), and methods that prove successful for modeling proteins (a high number of mostly low-energy interactions) have a higher failure rate (very few high-energy interactions; high sensitivity to error). Despite these and other difficulties, progress toward pseudoenzymes (catalyst + peptide scaffolding), peptide nanowires, and other novel peptidic structures was realized.

"Peptide Origami": Building a Library of Symmetric Folded Peptides:

One of our accomplishments was the assembly of a library of peptide and miniprotein dimers, trimers, or other simple oligomers. Natural entries in this library are rare, especially compared to the extensive list of larger proteins. A large list of dimeric and trimeric proteins to choose from was crucial for highly successful protein (not peptide) nanocage and nanomaterial engineering efforts by this laboratory. Thus we were motivated to build our own library of peptide multimers. We chose structures that would be amenable to chemical synthesis, to allow for simpler covalent modification and eventual incorporation of catalytic centers. Part of our motivation for synthesizing and characterizing multiple symmetric peptide structures was to test our capacity to model these types of structures. In addition to building a library for higher order assembly (e.g., into nanocages or peptide crystals), they also represented one approach to catalyst-pendant structures. This approach involved working "backwards" to create a tight interface between peptide minidomains, into which a catalyst could be engineered / inserted. For example, some of these structures were designed to accommodate the Shaw Group's hydrogenase catalyst.

In terms of design, we used a mixed approach. Most structures were envisioned and modeled using Foldit, either via in-house modeling with the stand-alone version of Foldit, or from puzzles released to the Foldit player community. Other lead structures were generated from *ab initio* forward folding and other proprietary modeling schemes (Rosetta plugins and mods) developed exclusively for this project.

Though difficulties were encountered upon proceeding to higher-order assemblies and catalyst-centered assemblies, we found considerable success with novel, covalently assembled dimers and trimers. Some examples follow:

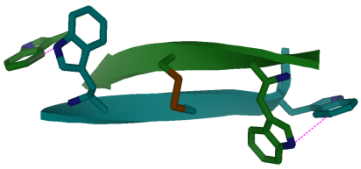

Figure 1. A simple "doubly-capped" (two sets of cross-strand tryptophan/tryptophan pairs) beta ribbon, with a disulfide center. This minimalist motif proved useful as a hub for larger, more complex designs.

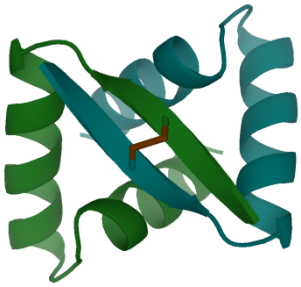

Figure 2. Beta-alpha-alpha 38, a two-helix bundle packed against the previous "beta ribbon hub".

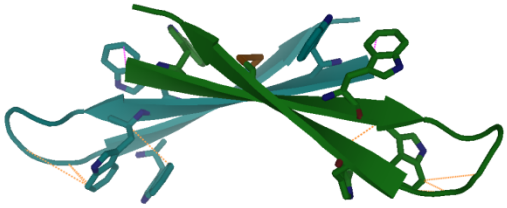

Figure 3. A four-stranded beta sheet, centered on a beta ribbon hub.

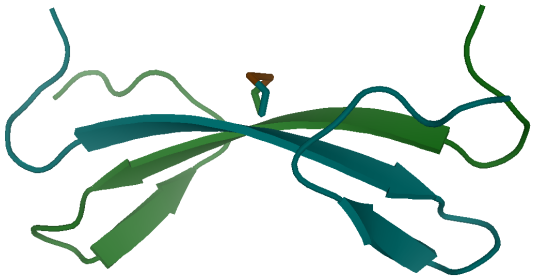

Figure 4. A strand-swapped dimer of the WW domain (the smallest and most biologically common natural miniprotein motif.)

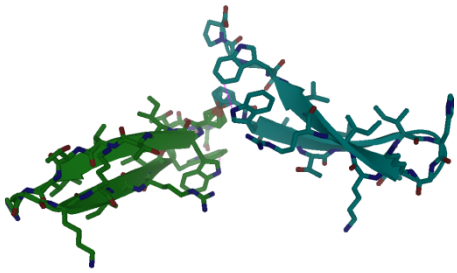

Figure 5. Independently folded beta hairpin "wings" of any length can be extended off a central hub molecule (a dicarboxylic acid) at fixed angles defined by the dicarboxylic acid hub.

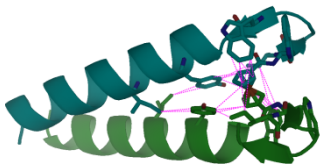

Figure 6. A Y-shaped beta-beta-alpha domain. This is effectively an alpha helical coiled-coil with two beta sheet "wings" of tunable length, stapled together with a covalent N-terminal hub. (We successfully employed a range of small molecule hubs here, including redox-relevant molecules dipicolinic acid and ferrocene 1,1'-dicarboxylic acid. The latter is pictured.)

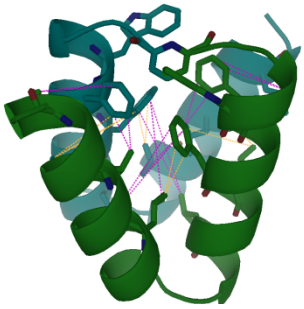

Figure 7. A two-by-two four helix bundle of unusual topology, held together by a covalent N-terminal hub (Dipicolinic acid pictured).

Other designs in our library included very long symmetric beta ribbons, various alpha helical bundles & coiled-coils (including trimers), and tris-bipyridine-centered structures. These join ranks with a very few natural peptide oligomers accessible via chemical synthesis: three ~40 residue helix bundles and Foldon, an efficient 27 residue (per chain) miniprotein with a triangular three-beta-hairpins core. We were able to pare a modified Foldon down to 22 residues (via removal of the short C-terminal helix) and preserve folding & association.

<u>Peptide scaffolds with catalytic functionality:</u>

Nickel-phosphine hydrogenase catalysts of known function were incorporated into peptide scaffolds – either as replacements of the covalent "hubs" of the structures shown above, or in de novo generated designs optimized to pack around the catalyst molecule.

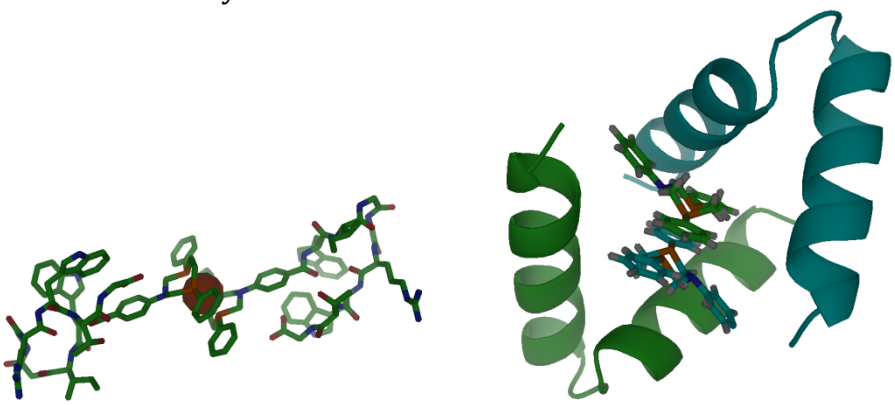

Figure 8. Left: a variant of "5" from the above list, with short beta hairpin wings as symmetric pendants attached to a central hydrogenase catalyst hub. Right: a variant of "2", with the central beta ribbon of the beta-alpha-alpha 38 dimer replaced by the hydrogenase catalyst (which is almost precisely the same length).

The structure shown on the left was shown to be a functional "mini-enzyme", with significant improvement over the free (no peptide scaffolding) catalyst. This finding was recently published:

Enzyme Design from the Bottom Up: An Active Nickel Electrocatalyst with a Structured Peptide Outer Coordination Sphere. M. L. Reback, G. W. Buchko, B. L. Kier, B. Ginovska-Pangovska, Y. Xiong, S. Lense, J. Hou, J. A. S. Roberts, C. M. Sorensen, S. Raugei, T. C. Squier and W. J. Shaw. *Chem. Eur. J.* 20, 1510-1514 (2014).

The mechanism of enhancement is thought to be via a proton shuttle (the rate limiting step for the hydrogenase catalyst is proton delivery) as well as biasing the structure of the catalyst toward its more productive conformer. In addition, the peptide pendants imparted water solubility to the otherwise strongly hydrophobic catalyst.

Higher-Order Assembly:

Despite repeated attempts, we had no success with higher-order assembly into nanocages and other complex structures. Nonspecific aggregation was common, as was lack of association. In some cases we observed a range of oligomers including sizes that matched our designs, but we had hoped to achieve monodisperse oligomerization.

Note: the dotted lines that appear in some of the images represent structure-confirming NOEs. All of these structures were either verified or explicitly determined using distances determined via NMR data.