

By folding into structures with pockets, proteins enable small-molecule binding events associated with molecular recognition, metabolic pathways or cell signaling. Likewise, computational design of ligand-binding proteins and enzymes relies on finding existing proteins with pockets satisfying the ligand binding site requirements while tolerating the incorporated mutations. Despite the successes of this approach for repurposing native proteins and designing new enzyme catalysts, this dependence on existing protein structures can be a limitation for certain applications – the geometry of the ideal active site may not be realized in a native structure or the active site placement may lead to unexpected changes in structure, stability or expression. Instead, building proteins *de novo* with custom-made binding sites should be more effective, but remains an outstanding challenge. *De novo* design methods have succeeded in designing thermostable proteins in a variety of folds, but with pockets that are too small to create binding sites. Yet, unlike most native proteins, current examples of *de novo* proteins lack non-ideal features that are critical to shape binding pockets like kinked helices, curved β -sheets or long loops. Among them, β -sheet curvature is ubiquitous in native ligand-binding protein structures, like the β -barrel, NTF2-like and jelly-roll fold domains; and the control of its geometry by design could open up new possibilities for customizing ligand-binding sites. *The aim of this project is to extend the current de novo computational protein design methodology towards the design of protein folds with cavities of variable size and incorporate binding and/or catalytic functions.*

I started this project by identifying β -sheet design principles from the analysis of natural protein structures and implemented these principles in a computational protocol to design proteins with curved β -sheets using the Rosetta software. One fundamental principle is that β -sheet bending can be controlled by combining β -bulges and register shifts in key positions of the β -sheet. With this approach six protein folds with cone-like shapes inspired by the naturally occurring cystatin and NTF2-like superfamilies were computationally designed. Designed proteins representative of each fold were experimentally characterized. Those designs with soluble expression in *E. coli* that were monomeric, thermostable and cooperatively folded were pursued for structure determination. The structures of eight proteins from five different fold variants were solved by X-ray crystallography or NMR spectroscopy, and the structures were very similar to the design models ($C\alpha$ -RMSD in the 1.0-2.1 Å range).

We reasoned that when designing function into these *de novo* scaffolds the proximity between the active site and the protein core could compromise protein stability, and explored two stabilization strategies preserving pocket accessibility: disulfide bonds and homodimer interfaces. Both strategies increased the stability of the designs and the structures of the stabilizing components were validated by X-ray crystallography. We then tested the ability of these proteins to support internal cavities, and designed mutations at the pocket entrance of these proteins. In general, most of the proteins tolerated the incorporated mutations with some detriments in stability as expected. The experimental structures of two designs showed internal cavities providing further support on their potential for designing active sites. This work provides a framework for designing new protein structures well suited for customizing ligand-binding and enzyme active sites.

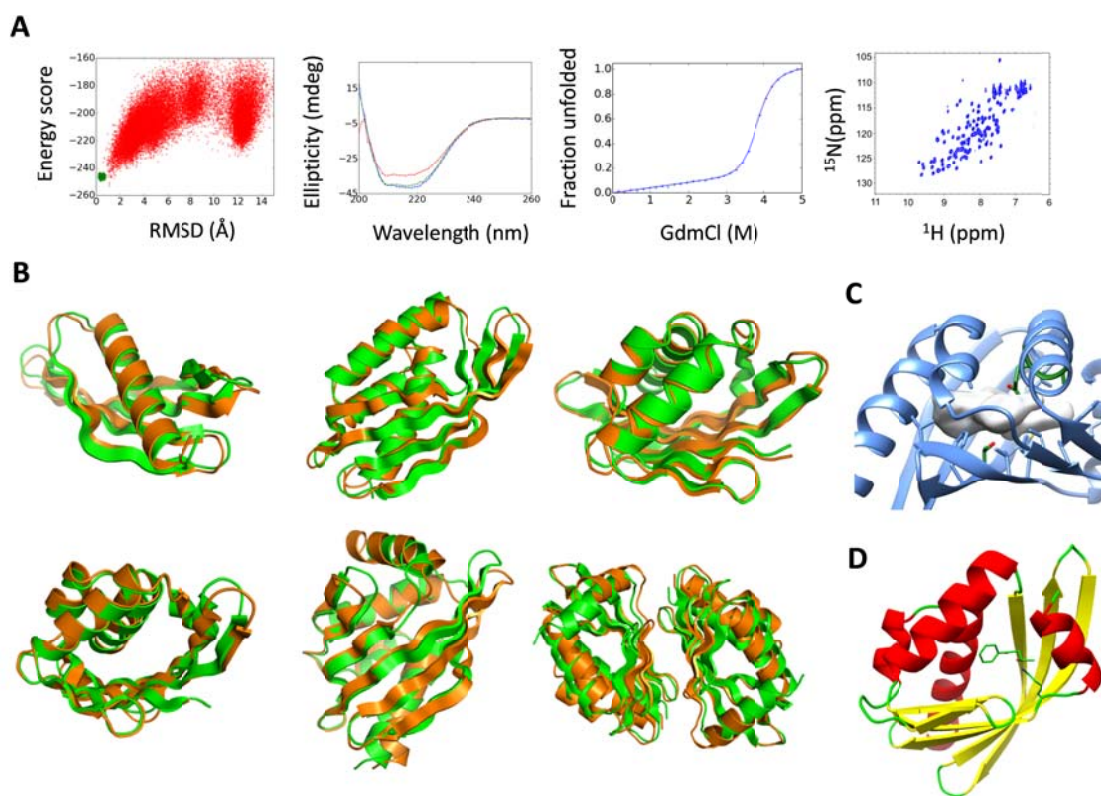


Figure 1. Computationally *de novo* designed proteins. (A) Designed proteins with funnel-shaped energy landscapes simulated by ab initio structure prediction (left panel) are experimentally characterized with circular dichroism and NMR HSQC spectra (other panels). (B) Design models (green) and experimentally determined structures (orange) are superimposed. (C) Example of a designed internal cavity. (D) Design model of most active variant with a nucleophilic lysine.

As a first application of these *de novo* proteins, we have designed active sites with a nucleophilic lysine. The activation of protein lysines is required to catalyze aldol and retro-aldol reactions, what are frequently used reactions for the synthesis of new chemicals. In addition, covalent binding of fluorescent molecules *via* these active lysines is also a powerful approach to label proteins being monitored in live cells with fluorescence microscopy. I computationally designed and experimentally tested active lysine designs in one of the protein variants that were previously characterized. With mass spectrometry four designed proteins with labeling activity were identified, and the most active variant has been evolved for labeling function with yeast surface display to higher levels. Interestingly, the computational designs did not have retro-aldol activity, but this evolved variant do shows low levels of enzymatic activity.

With this new approach it should now be possible the holistic design of protein structure and binding site, and therefore overcome some of the limitations of the early efforts of enzyme design with naturally occurring protein structures. This has already opened a research field on the customization of protein structure and active site with applications in biocatalysis, biosensing and bioscavenging.