Modeling Protein Flexibility for Structure-Based Active Site Redesign

Ryan H. Lilien[*,†] **Amy C. Anderson**[‡,§,¶] **Bruce R. Donald**[*,‡,§,¶]

Keywords: protein design, protein flexibility, ensembles, drug design

1 Introduction

A variety of microorganisms synthesize modified peptide-like products using enzymes called *Non-Ribosomal Peptide Synthetases (NRPS)*. Each NRPS protein synthesizes a specific product using multiple domains in an assembly-line fashion. These products are pharmaceutically active and include many antibiotics, antifungals, and immunosuppressants. Synthesis is initiated with the binding of an amino acid X to an adenylation (A) domain; an aminoacyl-adenylate (X-AMP) is formed and transferred down the assembly line. We consider the engineering problem of *switching* the specificity of an NRPS from its natural substrate to a new amino acid substrate using computational structural modeling techniques, in hopes of producing novel NRPS products. We hope to increase the efficiency and success rate of current protein redesign efforts. We will modify the NRPS enzyme gramicidin S synthetase A (GrsA), for which we have genetic sequence and structural data for the Phe A domain [6].

2 Methods

Our model includes the 9 residues of the Phe A domain of GrsA (1AMU) that are intimately in contact with the substrate amino acid (Fig. 1(a)) as well as 30 additional residues, near but not directly involved in the active site, which act as a steric shell. Molecular flexibility is modeled by the use of a rotamer library and soft steric potentials [2] which allow for small deviations from a rotamer conformation. We have implemented a simple *in silico* screening algorithm that applies a series of computational biophysical filters to the conformation space of rotamers. Sterically disallowed rotamer-based conformations of the free and bound active site are eliminated. The resulting conformations are scored with the nonbonded terms of the AMBER [1] energy function. The computed energies of these ensembles are then combined using a novel method for estimating binding affinity. This method strikes a balance between fidelity to the underlying physical biochemistry, and computational feasibility. Energy values are normalized by the partition function of the Boltzmann distribution in which the probability of occurrence of any state is related to the energy of that state. The use of a Boltzmann distribution scoring function was inspired by work on molecular docking to ensembles of protein models [4] and our previous work on the Core-Binding Factor complex [5].

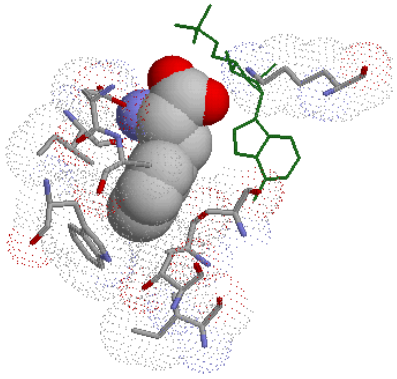
The scoring method generalizes the experimentally supported *folding constant* introduced in [3], to evaluate a target active site. It represents an approximation to the true binding constant, $K = \frac{[PS]}{[P][S]}$ by

$$K^* = \frac{\sum_{b \in B} \exp(-E_b/RT)}{\sum_{l \in L} \exp(-E_l/RT) \sum_{f \in F} \exp(-E_f/RT)} \quad (1)$$

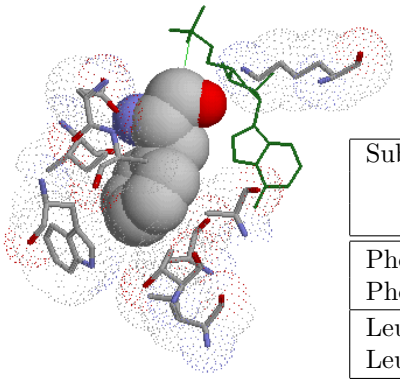
This equation[1] represents a statistical mechanics calculation of K using an ensemble-based approximation to the partition function. Intuitively, the quotient of the partition functions in Eq. (1) approximates the ratio of the probabilities of being in the bound and unbound states, that is, $K^* \propto P(\text{Bound})/P(\text{Unbound})$ — analogous to [3].

[*]Dartmouth Computer Science Department, Hanover, NH 03755, USA. [†]Dartmouth Medical School. [‡]Dartmouth Chemistry Department. [§]Dartmouth Center for Structural Biology and Computational Chemistry. [¶]Corresponding authors: brd@cs.dartmouth.edu and amy.anderson@dartmouth.edu.

[1] B is the set of bound protein states, F is the set of unbound protein states, L is the set of unbound ligand states, E_s is the energy of conformation s, R is the gas constant, and T is the temperature



(a)



(b)

Substrate	Phe A (wildtype)	Phe (T278M/ A301G)
Phe (assay†)	100	40
Phe K^*	0.50	0.23
Leu (assay†)	10	100
Leu K^*	0.21	0.40

(c)

Figure 1: (a) Crystal structure of the Phe A active site of GrsA (PDB id: 1AMU). Phe substrate is shown in spacefill, AMP in wireframe. (b) A top-scoring rotamer-based conformation, generated by *in silico* tests of our conformation algorithm, of the substrate, Phe, in the Phe A active site. (c) Modeling the Phe A active site of GrsA with substrates Phe and Leu. K^* determined with Eq. (1). †Activity assay results from [6].

3 Results

To test our model, we reproduced, *in silico*, a series of biochemistry experiments [6] in which the authors created a double mutant (T278M/A301G) of the GrsA Phe A domain which was predicted to bind leucine based on phylogenetic alignment to a leucine-binding adenylation domain. We computed K^* for the binding of Phe and Leu to both the wildtype and double mutant proteins. We predict an increase in Leu binding and a decrease in Phe binding when moving from the wildtype to the double mutant protein (Fig. 1(c)). In contrast, comparing the single conformation global energy minima across mutations did not replicate the biochemical activity assays. These results indicate that, generally, our algorithm reproduces the trends of the biochemical experimental data well but additional modeling will be necessary to reproduce the data more precisely.

Fundamental algorithmic challenges remain. We are currently examining how one would combine K^* into an efficient mutation search. We hope K^* will be useful to the structural biology community for structure-based drug design and protein redesign.

Acknowledgements. We would like to thank Anthony Yan and Lincong Wang for their assistance in the derivation of K^* and all members of the Donald Lab for helpful discussions and suggestions.

References

[1] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, and P. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[2] B. Dahiyat and S. Mayo. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci*, 94:10172–10177, 1997.

[3] V. Hilser and E. Freire. Structure-based calculation of the equilibrium folding pathway of proteins. correlation with hydrogen exchange protection factors. *J. Mol. Bio*, 262:756–772, 1996.

[4] R. Knegtel, I. Kuntz, and C. Oshiro. Molecular docking to ensembles of protein structures. *J. Mol. Biol.*, 226:424–440, 1997.

[5] R. Lilien, M. Sridharan, X. Huang, J. H. Bushweller, and B. R. Donald. Computational screening studies for core binding factor beta: Use of multiple conformations to model receptor flexibility. Poster, *The 8th Int'l Conf. on Intelligent Sys. for Mol. Bio. (ISMB-2000)*, August 2000.

[6] T. Stachelhaus, H. Mootz, and M. Marahiel. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, 6:493–505, 1999.