Accelerating the Interplay Between Theory and Experiment in Protein Design

Thesis by

Alex Nisthal

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2012

(Defended January 31, 2012)

Acknowledgements

I'd first like to thank Steve Mayo for all of the support and advice he has given me over the years, including the opportunity to work on large expensive equipment with lots of bells and whistles.

I would also like to thank the rest of my thesis committee (Frances, Doug, and Dave) and Jost Vielmetter at the Protein Expression Center for providing direction when I needed it the most.

As the bridge between two distinct eras of the Mayo Lab, I've witnessed a complete overhaul of students and postdocs. From my time as a learner, I would especially like to thank Oscar Alvizo, Christina Vizcarra, Marie Ary, Rhonda Digiusto, Corey Wilson, Roberto Chica, Heidi Privett, Ben Allen, and Jennifer Keeffe for their time and geniality in teaching me the ways of the lab. Now, as the master, I'd like to thank Matt Moore, Kurt Mou, Alexandria Berry, Toni Lee, Tim Wannier, Bernardo Sosa Padilla Araujo, Mohsen Chitsaz, Gene Kym, Ernest Lee, Grace Lee, Samy Hamdouche, Emzo de los Santos, and Seth Lieblich for making the lab a fun, collaborative, and politically incorrect place to work.

Lastly, I would like to thank my friends for keeping me sane and my family for letting me be whatever I wanted to be.

Abstract

Protein engineering techniques such as directed evolution and structure-based design aim to improve the properties of natural proteins. The next step, the *de novo* insertion of function into previously inert protein scaffolds, is the lofty promise of computational protein design. In order to achieve this goal reliably and efficiently, computational methods can be iteratively improved by cycling between theory and experiment.

Efforts to both accelerate the rate and broaden the information exchanged within protein design cycles form the core of this thesis. Improvements in the throughput of experimental stability determination allowed the thorough assessment of new multi-state and library design tools. Intending to alleviate the fixed backbone, single native state design approximation, the study found constrained molecular dynamics ensembles useful for core repacking applications. The subsequent development of automated liquid handling protocols for common molecular biology techniques brings design experiments to new levels of sample throughput. This technology facilitated the creation of a stability database encompassing every single mutant in a small protein domain. Although constructed to facilitate future computational training efforts, we answer a multitude of questions pertaining to mutational outcomes, distributions, positional sensitivity, tolerance, and additivity in the context of a protein domain.

By expanding the constraints of experimental molecular biology, this work opens up new possibilities in the efforts to train and assay new computational methodologies for protein engineering applications.

TABLE OF CONTENTS

Figures and Tables