An iterative approach to *de novo* computational enzyme design and the successful application to the Kemp elimination

Thesis by

Heidi Kathleen Privett

In Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended May 13, 2009)

ACKNOWLEDGEMENTS

I offer thanks to my advisor, Stephen Mayo, for his support and encouragement, for allowing me the opportunity to work on an exciting project that at times seemed impossible, and for not letting me give up when my research appeared hopeless.

I would also like to thank all of the past and present members of the Mayo Lab who contributed to a fun, congenial, and collaborative work environment and were always there to provide advice and encouragement.

I would like to thank all of my official collaborators, who have all been acknowledged individually in the relevant chapters and appendices.

In addition, I thank all of the Caltech students, postdocs, and staff who generously gave their time to teach me new techniques, trained me on equipment housed in their labs, and helped me troubleshoot problems in my projects even though my research had little bearing on their own. These generous individuals include Jonas Oxgaard (Goddard Lab), Amanda Cashin (Dougherty Lab), Ariele Hanek (Dougherty Lab), Dan Caspi (Stoltz Lab), Doug Behenna (Stoltz Lab), Robert Dirks (Pierce Lab), Jens Kaiser (Rees Lab), Jost Veilmetter (Protein Expression Facility), Rich Olson (Bjorkman Lab), Adrian Rice (Bjorkman/Rees Labs), Hernan Garcia (Phillips Lab), Brendan Mack (Davis Lab), and Jenn Stockdill (Stoltz Lab).

I also thank my many Caltech and non-Caltech friends, without whom I would have never made it, and my family for telling me that I could be anything I wanted to be when I grew up.

ABSTRACT

The development of reliable methods for the "on demand" *de novo* design of an enzymatic catalyst for an arbitrary chemical reaction has been an elusive goal of the computational protein design community. Recent successful results of *de novo* computational enzyme design have been encouraging, but the activity of the enzymes produced so far is still well below that of natural enzymes and the generalizability of these methods has yet to be established.

Presented in this thesis are methods that we have developed for the computational design of enzyme active sites as well as results from the evaluation of these methods through a test case, the Kemp elimination. Initial Kemp elimination designs were shown to be inactive. However, in the course of refining these design procedures, we carried out extensive theoretical and experimental evaluation of several of these inactive designs, which allowed us to identify the causes of the inactivity and led to adjustments of our design procedure. These modified methods were then successfully used to design four distinct enzymes for this reaction in three inert scaffolds including the scaffold that housed the previously inactive designs. In addition, we demonstrate that molecular dynamics simulations can accurately predict the activity of designed Kemp elimination enzymes and can be used as a reliable prescreening step, allowing us to focus our experimental efforts on designs that are most likely to be active.

The work presented here demonstrates that the cyclic evaluation and redesign of both active and inactive enzymes was instrumental in the identification and resolution of deficiencies in our computational methods and directly resulted in *de novo* designed enzymes with novel and increased activity.

TABLE OF CONTENTS

TABLES AND FIGURES

ABBREVIATIONS

AFU	arbitrary fluorescence units
AR	androgen receptor
CD	circular dichroism
C. fimi	*Cellullomonas fimi*
CLEARSS	combinatorial libraries emphasizing and reflecting scored sequences
CV	column volume
Da	Dalton (1 g/mol)
DEE	dead end elimination
DHT	dihydrotestosterone
DMF	dimethylformamide
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
FASTER	fast and accurate side-chain topology and energy refinement
FFH	L-2-phenyl-4-benzylphenyloxazolin-5-one hydrolase
FMEC	faster minimum energy conformation
FOX	L-2-phenyl-4-benzylphenyloxazolin-5-one
GdnHCl	guanidine hydrochloride
GMEC	global minimum energy conformation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfornic acid
HPLC	high-pressure liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infrared
k_{cat}	catalytic constant
KE	Kemp elimination
K_M	Michaelis constant
k_{uncat}	rate constant for an uncatalyzed reaction
LB	Luria-Bertani broth
LK	Lazaridis-Karplus (solvent exclusion model)

MBP	maltose binding protein
MC	Monte Carlo
MD	molecular dynamics
MES	2-(*N*-morphilino)ethanesulfonic acid
MME	monomethyl ether
MR	molecular replacement
MS	mass spectrometry
MUG	4-methylumbelliferyl-β-D-glucopyranoside
MUX	4-methylumbelliferyl-β-D-xylopyranoside
MWCO	molecular weight cut off
NBT	5-nitrobenzotriazole
NBZ	5-nitrobenzoxizole
NBX	transition state of the 5-nitrobenzoxizole Kemp elimination reaction
Ni-NTA	nickle-nitrilotriacetic
NMR	nuclear magnetic resonance
NPT	constant number of particles, pressure, and temperature
OD	optical density at a specific wavelength
ORBIT	optimization of rotamers by iterative techniques
PBS	phosphate buffered saline
PDB	protein data bank
PEG	polyethylene glycol
PNK	polynucleotide kinase
*p*NP	para-nitrophenol
PCR	polymerase chain reaction
P*i*	inorganic phosphate
PPMAL	Protein/Peptide MicroAnalytical Laboratory (Caltech)
RMSD	root mean squared deviation
S. avidinii	*Streptomyces avidinii*
S. cerevisiae	*Saccharomyces cerevisiae*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

S. solfataricus	*Sulfolobus solfataricus*
T. aurantiacus	*Thermoascus* aurantiacus
TAX	*T. aurantiacus* xylanase
TES	testosterone
T_m	midpoint of thermal denaturation curve
T. maritima	*Thermotoga maritima*
Tris	tris(hydroxymethyl)aminomethane)
TS	transition state
UV-vis	ultraviolet-visible
VDW	van der Waals