

**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)**

© 2009

**Heidi K. Privett**

**All Rights Reserved**

## 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 Veilmutter (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**

<b>Acknowledgements</b>		iii
<b>Abstract</b>		iv
<b>Table of Contents</b>		v
<b>Tables and Figures</b>		vi
<b>Abbreviations</b>		xi
<hr/>		
<b>Chapters</b>		
Chapter I	<i>Introduction</i>	1
Chapter II	<i>Combinatorial methods for small molecule placement in computational enzyme design</i>	17
Chapter III	<i>Towards the computational design of a Kemp elimination enzyme and crystallographic conformation of the active site configuration of an inactive design</i>	46
Chapter IV	<i>Completing the protein design cycle: the computational design and molecular dynamics simulation analysis of five Kemp elimination enzymes</i>	87
<hr/>		
<b>Appendices</b>		
Appendix A	<i>Toward the computational design of a novel enantioselective hydrolase</i>	131
Appendix B	<i>Using computational library design to alter the specificity of a xylanase</i>	167
Appendix C	<i>Altering the specificity of an androgen receptor</i>	189
Appendix D	<i>Recombinant expression and purification of a thermophilic xylanase</i>	204

**TABLES AND FIGURES**

Figure 1-1.	The protein design cycle	15
Figure 1-2.	The Kemp elimination of 5-nitrobenzisoizole	16
Figure 1-3.	Kemp elimination catalytic antibody 34E4	16
<hr/>		
Table 2-1.	RMSD and number of wild-type contacts as a function of rotational step size and rotamer library	39
Table 2-2.	RMSD and number of wild-type contacts as a function of rotational and translational step sizes	40
Table 2-3.	Results from targeted placement procedure as a function of rotamer library	41
Figure 2-1.	Contact geometries specified in small molecule pruning step	42
Figure 2-2.	Sample results from test calculations presented in Table 2-1	43
Figure 2-3.	Effect of rotational and translational step sizes	44
Figure 2-4.	Targeted placement procedure	44
Figure 2-5.	The three clustering moves are illustrated by showing the state of a sample system before and after the move is performed	45
<hr/>		
Table 3-1.	Partial atomic charges for 5-NBX	69
Table 3-2.	Variation of contact geometry for targeted ligand placement	70
Table 3-3.	Geometric constraints between the active site residues and the transition state in the active site search	71
Table 3-4.	Geometric constraints for additional base contact	72

Table 3-5.	Thermocycler temperature programs for gene construction and mutagenesis reactions	73
Table 3-6.	Crystallographic statistics for HG-1	74
Figure 3-1.	5-NBX transition state	75
Figure 3-2.	Kemp elimination ideal active site	76
Figure 3-3.	Additional hydrogen bond contact	76
Figure 3-4.	Predicted active site structure of HG-1	77
Figure 3-5.	First-order rate constants for KE reaction	78
Figure 3-6.	CD analysis of HG-1	79
Figure 3-7.	Crystal structure of HG-1	80
Figure 3-8.	Electron density of HG-1 mutation sites	81
Figure 3-9.	Overlay of HG-1 crystal structure active site with design model	82
Figure 3-10.	Ordered water molecules near E237 in the active site	83
Figure 3-11.	MD analysis of HG-1	84
Figure 3-12.	Active site of HG-1h	85
Figure 3-13.	MD analysis of HG-1h	86
<hr/>		
Table 4-1.	Summary of design calculations for Kemp elimination enzymes	111
Table 4-2.	Physical characteristics of protein variants	112
Table 4-3.	Experimental characterization of designed Kemp elimination enzymes	113
Table 4-4.	Design summary of Kemp elimination enzymes	114
Figure 4-1.	Active site locations of first- and second-generation designs in TAX	115

Figure 4-2.	Active site of HG-2	115
Figure 4-3.	Active site structures of HG-2 during the MD simulation	116
Figure 4-4.	Distance distributions of HG-2	117
Figure 4-5.	Expression and purification of HG-2	118
Figure 4-6.	CD analysis of wild-type TAX and HG-2	119
Figure 4-7.	Kinetic characterization of second-generation enzymes	120
Figure 4-8.	Effect of pH on the activity and structure of HG-2	121
Figure 4-9.	Kinetic characterization of third-generation enzymes	122
Figure 4-10.	Active sites of designs in scaffold 1A53	123
Figure 4-11.	Kinetic characterization of designs in scaffolds 1A53 and 1THF	124
Figure 4-12.	Active sites of designs in scaffold 1THF	125
Figure 4-13.	MD analysis of 1A53-1	126
Figure 4-14.	MD analysis of 1A53-2	127
Figure 4-15.	MD analysis of 1A53-3	128
Figure 4-16.	MD analysis of 1THF-1	129
Figure 4-17.	MD analysis of 1THF-2	129
<hr/>		
Table A-1.	Apparent pseudo-first-order rate constants for F-FOX hydrolysis	154
Figure A-1.	Enantioselective hydrolysis of S-2-benzyl-4-phenyl-oxazolone-5-one ((S)-F-FOX)	155
Figure A-2.	Maltose binding protein structure	155
Figure A-3.	FOX transition state structure	156
Figure A-4.	Ideal active site contacts	156



Figure A-5.	Geometric constraints for the contacts between the catalytic residues and the (S)-F-FOX transition state (TS)	157
Figure A-6.	Arginine-(S)-F-FOX geometric constraints	158
Figure A-7.	UV-vis spectra of F-FOX and N-benzoyl-phenyl-alanine	159
Figure A-8.	Fluorescence of F-FOX	160
Figure A-9.	Active site structure	161
Figure A-10.	Repacked active site	161
Figure A-11.	1ANF-FFH CD analysis	162
Figure A-12.	F-FOX hydrolysis rate constants determined by UV-vis kinetics assays	163
Figure A-13.	Additional potential beneficial mutations to 1ANF-FFH	164
Figure A-14.	1ANF-FFH mutant apparent rate constants	164
Figure A-15.	1ANF-FFH mutant apparent rate constants	166
<hr/>		
Table B-1.	Specific activity of TAX on 2.5 mM pNP-glycosides	181
Table B-2.	Mutagenesis primers for site saturation mutagenesis libraries	181
Table B-3.	Designed TAX libraries	182
Table B-4.	Thermocycler temperature programs for mutagenesis reactions	182
Table B-5.	Kinetic constants for TAX variants with MUX or MUG	183
Figure B-1.	Mechanism of retaining glycosidases	184
Figure B-2.	Xylanase activity assays	185
Figure B-3.	Predicted clashes of mannose in the TAX active site	186

Figure B-4.	Wild-type hydrogen bonds preserved in TAX calculations	187
Figure B-5.	Site-saturation mutagenesis positions in TAX	188
<hr/>		
Table C-1.	AR-19PT design summary	197
Figure C-1.	Activation mechanism of AR	198
Figure C-2.	Chemical structures of androgens of interest	198
Figure C-3.	Rotamers of 19PT	199
Figure C-4.	Design positions in the active site of AR	200
Figure C-5.	Wild-type hydrogen bonds to 19PT	200
Figure C-6.	ORBIT designs for AR binding of 19PT	201
<hr/>		
Table D-1.	Assembly oligonucleotides for the construction of the TAX-His <sub>6</sub> gene	213
Figure D-1.	Protein and DNA sequences for TAX-His <sub>6</sub>	214
Figure D-2.	SDS-PAGE analysis of TAX-His <sub>6</sub> expression and purification	215
Figure D-3.	Mass spectrometry analysis of TAX-His <sub>6</sub>	215
Figure D-4.	CD analysis of TAX-His <sub>6</sub>	216

**ABBREVIATIONS**

AFU	arbitrary fluorescence units
AR	androgen receptor
CD	circular dichroism
<i>C. fimi</i>	<i>Cellulomonas fimi</i>
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
<i>E. coli</i>	<i>Escherichia coli</i>
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-piperazineethanesulfonic acid
HPLC	high-pressure liquid chromatography
IPTG	isopropyl $\beta$ -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-( <i>N</i> -morpholino)ethanesulfonic acid
MME	monomethyl ether
MR	molecular replacement
MS	mass spectrometry
MUG	4-methylumbelliferyl- $\beta$ -D-glucoopyranoside
MUX	4-methylumbelliferyl- $\beta$ -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
<i>p</i> NP	para-nitrophenol
PCR	polymerase chain reaction
<i>P</i> <sub><i>i</i></sub>	inorganic phosphate
PPMAL	Protein/Peptide MicroAnalytical Laboratory (Caltech)
RMSD	root mean squared deviation
<i>S. avidinii</i>	<i>Streptomyces avidinii</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

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