

UCLA

UCLA Electronic Theses and Dissertations

Title

Computational Studies of Organic, Organometallic, and Enzyme Catalysis

Permalink

<https://escholarship.org/uc/item/0qs475w8>

Author

Noey, Elizabeth Lynn

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Studies of Organic, Organometallic, and Enzyme Catalysis

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Chemistry

by

Elizabeth Lynn Noey

2015

© Copyright by
Elizabeth Lynn Noey
2015

ABSTRACT OF THE DISSERTATION

Computational Studies of Organic, Organometallic, and Enzyme Catalysis

by

Elizabeth Lynn Noey

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2015

Professor Kendall N. Houk, Chair

Computations are increasingly powerful tools for studying reaction mechanisms and protein catalysis. Various quantum mechanical (QM) and force field-based calculations are applied to problems in organic, organometallic, and protein chemistry. These studies span the chemistry-biology interface, progressing from theoretical studies of gold catalysis, to that of N-heterocyclic carbene (NHC) catalysis, and enzyme catalysis. The first study highlights a gold(I)-catalyzed enyne cyclization with a bifurcating potential energy surface. Several alkynylindoles undergo gold(I)-catalyzed cyclization reactions to form a single isomer in each case. This transformation involves a two-step no-intermediate mechanism with surface bifurcations leading to two or three products. The second gold study is on the mechanism of the rearrangement of acetylenic amine-N-oxides. Further work has been done on the mechanism of the Stetter reaction catalyzed by substituted NHCs. The leucine metabolic pathway was reengineered to produce biofuels, and computations showed that there is a push-pull effect between the hydrophobic effect and steric clash, which dictates the LeuA substrate scope. The redesign of a transaminase

to install the stereocenter in the blockbuster diabetes drug sitagliptin was attempted. The transaminase that was evolved for the industrial synthesis of sitagliptin was studied computationally. This study elucidates the energetic details of the transamination mechanism to form sitagliptin, and makes progress toward understanding the role of mutations in the evolution. Finally, a computational, crystallographic, and kinetic study of ketoreductases (KREDs) shows how point mutations change the enantioselectivity toward two small substrates, 3-oxa and 3-thiacyclopentanone. QM calculations of the ideal geometry for catalysis, and molecular dynamics (MD) simulations show how small changes in the size, shape, and hydrophobicity of the active site of the enzyme modulate the enantioselectivity. Here, we develop an MD method where simulations are run on the enzyme containing the cofactor-substrate transition structure for the reduction. This approach probes how well each enzyme stabilizes the transition structures and can predict the experimentally favored enantiomer. Although the subject matter varies, the underlying goal of understanding chemical reactions and catalysis from a physical organic perspective persists.

The dissertation of Elizabeth Lynn Noey is approved

Miguel Garcia-Garibay

Vidvuds Ozolins

Kendall N. Houk, Committee Chair

University of California, Los Angeles

2015

TABLE OF CONTENTS

List of Schemes	vi
List of Figures.....	vii
List of Tables	x
Acknowledgments	xi
Vita	xv
1. Selective Gold(I)-Catalyzed Formation of Tetracyclic Indolines: A Single Transition Structure and Bifurcations Lead to Multiple Products.....	1
2. Mechanism of Gold(I)-Catalyzed Rearrangements of Acetylenic Amine-N-Oxides: Computational Explanations of Regioselectivities	19
3. Quantum Mechanical Investigation of the Effect of Catalyst Fluorination in the Intermolecular Asymmetric Stetter Reaction	38
4. Catalytic Asymmetric Intermolecular Stetter Reactions of Enolizable Aldehydes with Nitrostyrenes: Computational Study Provides Insight into the Success of a New Catalyst	56
5. A Synthetic Recursive “+1” Pathway for Carbon Chain Elongation	68
6. Transaminase Redesign for the Synthesis of Sitagliptin	94
7. Toward Understanding the Activity of the Sitagliptin Transaminase	115
8. Origins of Stereoselectivity of Mutants of <i>Lactobacillus kefir</i> Ketoreductase	133

LIST OF SCHEMES

1.1 Products formed by transition metal-catalyzed enyne cyclizations	1
1.2 Proposed mechanisms for the formation of 2 from 1	2
1.3 Proposed mechanism from 9 to 10.....	4
1.4 Free energy diagram for the formation of 25 and competing paths.....	13
1.5 Free energy diagram for the formation of 25 _H and competing paths	15
2.1 Proposed mechanism for the rearrangements of acetylene amine-N-oxides to piperidinones and azapanones	21
2.2 B3LYP/6-31G(d)/CPCM(DCM) (B3LYP-D3) energetics for the conversion of the gold-coordinated N-oxide, 13, to cyclic ketones 17 and 19	23
2.3 <i>Anti</i> addition pathway	24
2.4 B3LYP-D3 free energy diagram for the lowest energy path from 13 to 17	25
2.5 B3LYP/6-31G(d)/CPCM(DCM) (B3LYP-D3) energetics for the conversion of the gold-coordinated N-oxide, 23, to piperidinone, 27 and 29	29
2.6 <i>Anti</i> addition pathway	29
2.7 B3LYP-D3 free energy diagram for the lowest energy path from 23 to 27	30
3.1 Effect of fluorination of bicyclic triazolium catalysts the Stetter reaction yield and enantiomeric ratios	39
3.2 Proline conformations reported by Raines.....	40
4.1 Derivatization of β -nitro ketone products.....	60

LIST OF FIGURES

1.1 Examples of gold-catalyzed reactions with two-step no-intermediate mechanisms	5
1.2 Optimized geometry of 16, a computational model for 11	6
1.3 Transition state 14 forms intermediates 17 and 18	7
1.4 Transition state 15 forms intermediates 20, 21 and 22	7
1.5. Exo (14 ⁺) and endo (15 ⁺) transition structures	8
1.6. Potential energy surfaces from 16 through transition structure 14 to 17 and 18	9
1.7 PES from 16 through transition structure 15 to 20, 21 and 22	10
1.8 Transformation from 17 and 17 ⁺ to the ether-bridged species.....	12
1.9 Transition structure 27 ⁺	14
2.1 Comparison of the hetero-retroene reaction with and without gold including bond distances and Mulliken charges	26
2.2 The hetero-retroene transition structure, TS14-16.....	27
2.3 Hetero-retroene reaction transition structures dictating the product distribution of 8 and 9 ...	28
2.4 The hetero-retroene transition structure, TS24-26.....	31
2.5 The hetero-retroene reaction transition structures for the formation of 11 and 12.....	32
3.1 Preferred conformations of triazolium catalysts	41
3.2 B3LYP/6-31G(d) gas phase conformations and relative enthalpies of triazolium catalysts ...	43
3.3 B3LYP/6-31G(d) acyl anion equivalents.....	44
3.4 B3LYP/6-31G(d) relative gas phase enthalpies of acyl anion equivalents	45
3.5 Enol intermediates 7-9-endo ⁺ and relative enthalpies with respect to 7-9-exo	46
3.6 Overlay of crystal structure and calculated structure of 13.	47
3.7 <i>Si</i> - and <i>Re</i> -face transition structures	48
3.8. B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger for 7	48
3.9 B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger for 8	49
3.10 B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger for 9	49
3.11 Seebach's topological rule applied to the Stetter reaction	50
3.12 Comparison of the transition states for the Stetter reaction, TS3 (<i>Si</i>), and benzoin condensation	52
4.1 Influence of catalyst structure on relative rate	60
4.2 Relative free energies of acyl anion equivalents.....	62
4.3 Transition structures TS1-TS6 and energies	64
5.1 Engineering LeuABCD “+1” biosynthetic pathway	70
5.2 Modeling of carbon-carbon bond formation and the LeuA active site	72
5.3 Comparison of substrate binding and substrate solvent accessible surface area in feedback-resistant and quintuple mutant <i>EcLeuA</i> protein-substrate complex models.....	77
5.4 Metabolic engineering of <i>E. coli</i> for long-chain alcohol production.....	82
6.1 Example of the <i>inside-out</i> protocol for the Kemp elimination	96
6.2 The transamination reaction of the pro-sitagliptin ketone, 1, to form sitagliptin, 2	97
6.3 Truncated substrate used in Codexis' directed evolution process	97

6.4 An overlay of the active sites from the aspartate aminotransferases in the pdb	99
6.5 Residues responsible for binding the PLP cofactor	100
6.6 The ketimine, 4, quinonoid, 5, and aldimine, 6, are the three covalent intermediates of the pro-sitagliptin ketone with PLP	101
6.7 The library of conformers of 5 used in the redesign	101
6.8 PLP-aspartate ketimine, 7, quinonoid, 8, and aldimine, 9, intermediates	103
6.9 Heavy atom distances monitored in the MD simulations to assess the catalytic potential of the enzyme	104
6.10 The distances monitored in the MD simulation of the WT	104
6.11 The catalytic distances sampled in the MD simulation of the WT	105
6.12 The ligand bonding contacts sampled in the MD simulation of the WT	106
6.13 The catalytic distances sampled in the MD simulation of the design K7 compared to the WT	108
6.14 The ligand binding contacts sampled in the MD simulation of the design K7 compared to the WT	109
7.1 Transaminases transform prochiral ketones into chiral amines, with a sacrificial amine donor	115
7.2 The transamination reaction of the pro-sitagliptin ketone, 1, to form sitagliptin, 2	116
7.3 Substrate scope and percent conversion for ATA-117, and the round 11 (final published round) mutant in Codexis' directed evolution.	117
7.4 The ketimine, quinonoid, and aldimine are the three covalent intermediates of the pro-sitagliptin ketone with PLP in its transformation to an amine	118
7.5 Modeled transaminase catalytic cycle	120
7.6 Optimized PLP-sitagliptin intermediates with lysine	121
7.7 Quinonoid intermediate. Distances to CA and CB are measured in the MD simulations	123
7.8 MD simulation of the commercial variant as a monomer	126
7.9 Plots of catalytic distances sampled in the MD simulations	127
7.10 Plots of the CCNC dihedral	128
7.11 Overlay of x-ray crystal structure of the commercial sitagliptin-transaminase with snapshots every 240 ns over the 2.197 μ s MD simulation	129
8.1 KRED catalytic cycle	135
8.2 Reduction and proton transfer in the enzyme	136
8.3 <i>L. kefir</i> KRED (Crystal apo WT) with acetophenone docked in	137
8.4 The alcohols formed from 1 and 2 are employed in the synthesis of fosamprenavir and sulopenem	138
8.5 <i>Theozyme</i> for 3-thiacyclopentanone reduction in KREDs	142
8.6 Log/log graph of experimental enantiomeric ratio given for oxa- and thiacyclopentanol by KRED variants	143
8.7 The electrostatic potential surfaces of 1 and 2	145
8.8 Mechanism of reduction as modeled with an abbreviated active site (<i>theozyme</i>)	147
8.9 Transition structures leading to <i>R</i> -oxacyclopentanol <i>S</i> -oxacyclopentanol, <i>R</i> -thiacyclopentanol, <i>S</i> -thiacyclopentanol	148
8.10 QM transition structure and crystal structure overlay	151
8.11 Overlay of the apo and NADP-bound WT crystal structures	152

8.12 The different H-bond contacts involving Y190, E145(S), D150, and water	153
8.13 Comparison of the apo WT and A94F crystal structures	154
8.14 Overlay of the five obtained crystal structures: apo WT, NADP-bound WT, A94F, E145S, and Sph.....	155
8.15 Comparison of the NADP-bound WT and Sph crystal structures	156
8.16 Overlay of MD snapshots showing an open and closed conformation of the substrate-binding loop	158
8.17 MD simulation of the WT	159
8.18 The catalytic distances monitored in the MD simulations with the <i>theozymes</i>	161
8.19 Log/log graph of the experimental <i>er</i> versus computed <i>er</i> from the MD simulations, where <i>er</i> is computed as $\%competentS/\%competentR$	163
8.20 MD snapshot of WT with the TS complex of <i>R</i> -thiacyclopentanone	164
8.21 WT simulation with <i>theozymes</i>	166
8.22 Snapshots from the MD simulations of A94F	167
8.23 A94F simulation with <i>theozymes</i>	169
8.24 Snapshots from the MD simulations of E145S	170
8.25 E145S simulation with <i>theozymes</i>	171
8.26 Snapshots from the MD simulations of Sph	173

LIST OF TABLES

3.1 Summary of calculated and experimental enantioselectivities	51
3.2 F-N1 distances in TS3-TS6.....	51
4.1 Catalyst and optimization studies	57
4.2 Reaction Scope.....	59
4.3 Calculated and experimental percent enantiomeric excesses	64
5.1 Quantum Mechanical Calculation Results	73
5.2 Enzyme Binding Pocket and Substrate Volumes	75
5.3 Kinetic parameters of feedback-resistant and quintuple mutant <i>EcLeuA</i>	80
5.4 Alcohol production profiles of <i>EcLeuA</i> mutants	83
6.1 Conserved residues from the WT in the enzyme design	100
6.2 Final designs on which experiments were performed.....	111
6.3 Rosetta scores for the final designs.....	112
7.1 Free energies (ΔG) and enthalpies (ΔH) for the transaminase catalytic cycle.....	123
8.1 Enantiomeric ratio (<i>S/R</i>), corresponding $\Delta\Delta G^\ddagger$ (kcal/mol), and catalytic efficiency (k_{cat}/K_m) determined experimentally for 3-oxa- and 3-thia-cyclopentanol by KRED variants	144

ACKNOWLEDGEMENTS

I would like to first express my gratitude to my loved ones, who continually support me in all my pursuits. My parents, and extended family have offered me nothing but love and encouragement throughout my life and career. I am incredibly thankful to Mike Lerch, for his continual love and support, which enable me to excel in all my endeavors. I am also grateful for my coach, Brady O'Brian, and the UCLA Triathlon Team for their passion for sport and enthusiasm.

I would also like to thank my advisor, Professor Houk, for his mentorship, and the Houk group for their helpful discussions and friendly working environment. Many group members have offered me guidance and companionship over my graduate career. Specifically, I would like to thank Professor Peng Liu, and Dr. Buck Taylor.

I am grateful to have worked with many brilliant scientists in and out of the group, so I would like to thank all of my collaborators: Professor Xiang Wang, Professor Liming Zhang, Yingdong Luo, Professor Tomislav Rovis, Dr. Daniel DiRocco, Dr. Joann Um, Professor James Liao, Dr. Ryan Mareschi, Dr. Han Li, Dr. Kechun Zhang, Dr. Sonah Kim, Dr. Asha Chaubay, Dr. Elizabeth Felnagle, Dr. Sílvia Osuna, Dr. Gonzalo Jiménez-Osés, Dr. Nidhi Tibrewal, Dr. Jiyong Park, Carly Bond, Dr. Duilio Cascio, Dr. Xiyun Zhang, Dr. Gjalt Huisman, and Professor Yi Tang.

I am grateful to be a member of the distinguished UCLA chemistry department. I appreciate all those who have made my stay here fun, and who have taken the time to teach me. Professor Garcia-Garibay has always been incredibly encouraging and friendly, in addition to mentoring me.

I am also grateful to the National Institutes of Health Chemistry-Biology Interface (CBI) Training Program for financial support, and all those involved with CBI for teaching me about topics on the frontiers of chemistry and biology. In particular, I would like to thank the current director of the CBI program at UCLA, Professor Maynard. I would like to thank the Foote family for awarding me the Foote Fellowship. I am grateful for the computing resources: UCLA IDRE Hoffman2 computing cluster, the San Diego Supercomputing Center, the Pittsburgh Supercomputing Center and the use of ANTON, and the KNH computing cluster. I am also grateful to National Science Foundation for funding.

Chapter 1 is a modified version of the publication, “Selective Gold(I)- Catalyzed Formation of Tetracyclic Indolines: A Single Transition Structure and Bifurcations Lead to Multiple Products,” Noey, E. L.; Wang, X.; Houk, K. N. *J. Org. Chem.* **2011**, 76, 3477. The Wang group at The University of Colorado, Boulder published the experimental results that were the subject of our calculations. I did the calculations and wrote the manuscript under the guidance of Professor Houk.

Chapter 2 is a modified version of the publication, “Mechanism of Gold(I)-Catalyzed Rearrangements of Acetylenic Amine-N-Oxides: Computational Explanations of Regioselectivities,” Noey, E. L.; Luo, Y.; Zhang, L.; Houk, K. N. *J. Am. Chem. Soc.* **2012**, 134, 1078. Youngding Luo did the experiments under the guidance of Professor Zhang at the University of California, Santa Barbara. I did the calculations and wrote the manuscript under the guidance of Professor Houk.

Chapter 3 is a modified version of the publication, “Quantum Mechanical Investigation of the Effect of Catalyst Fluorination in the Intermolecular Asymmetric Stetter Reaction” Um, J.

M.; DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. *J. Am. Chem. Soc.* **2011**, *133*, 11249. Joann M. Um, who was a graduate student in the Houk lab, did many of the calculations, and the initial writing of the paper. Daniel A. DiRocco did the experiments under the guidance of Professor Rovis of Colorado State University. I finished the calculations and manuscript under the guidance of Professor Houk.

Chapter 4 is a modified version of the publication, “Catalytic Asymmetric Intermolecular Stetter Reactions of Enolizable Aldehydes with Nitrostyrenes: Computational Study Provides Insight into the Success of a New Catalyst” DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. *Angew. Chem. Int. Ed.* **2012**, *51*, 2391. Daniel A. DiRocco did the experiments and wrote the experimental section under the guidance of Professor Rovis of Colorado State University. I did the calculations and wrote the theoretical section under the guidance of Professor Houk.

Chapter 5 is a modified version of the publication, “A Synthetic Recursive “+1” Pathway for Carbon Chain Elongation” Marcheschi, R. J.; Zhang, K.; Noey, E. L.; Kim, S.; Li, H.; Chauby, A.; Houk, K. N.; Liao, J. C. *ACS Chem Biol.* **2012**, *7*, 689. I performed quantum mechanical calculations, Sonah Kim and Asha Chauby performed protein-substrate modeling, Ryan J. Marcheschi, Han Li, and Kechun Zhang characterized the kinetics of proteins and quantified *n*-alcohol production from fermentations. Ryan J. Marcheschi, Kechun Zhang, and I prepared the manuscript. K. N. Houk and James C. Liao supervised the research and edited the manuscript.

Chapters 6 and 7 were in collaboration with Sílvia Osuna, Gonzalo Jiménez-Osés, Xiyun Zhang, Gjalt Huisman, Yi Tang, and K. N. Houk. Sílvia and Gonzalo worked with me on modeling, and the others helped devise the project.

Chapter 8 is a modified version of “Origins of Stereoselectivity of Mutants of *Lactobacillus kefir* Ketoreductase.” Noey, E. L.; Tibrewal, N.; Park, J.; Bond, C.; Cascio, D.; Osuna, S.; Jiménez-Osés, G.; Liang, J.; Zhang, X.; Huisman, G.; Tang, Y.; Houk, K. N. *Proc. Natl. Acad. Sci. Submitted*. K. N. Houk supervised all theoretical studies. I wrote the manuscript, did the quantum mechanical calculations and MD simulations, and analyzed data. Jiyong Park ran the MD simulations with the *theozyme*, and wrote the methods and did the statistical analysis for this section. Sílvia Osuna and Gonzalo Jiménez-Osés contributed analysis and writing. Under Prof. Yi Tang’s supervision, Nidhi Tibrewal crystallized the enzymes, and with Duilio Cascio obtained the x-ray structures. Carly Bond ran the kinetics studies, and they each wrote the methods and analysis for their sections. At Codexis Inc. Gjalt Huisman and Xiyun Zhang supervised the KRED experiments including stereoselectivity determinations and Jack Liang obtained the experimental data on the KRED library.

VITA

EDUCATION

B.S. in Chemistry, University of San Francisco, First in Class (2008)

PUBLICATIONS

- Castro, C.; Karney, W. L.; Noey, E.; Vollhardt, K. P. C. Competing Isomerizations of [12]Annulene: Diels Alder Reaction versus Electrocyclization. *Org. Lett.* **2008**, *10*, 1287.
- Noey, E.; Tam, S.; Pham, D.; Curtis, J.; Jones, E. Synthesis and Ligand-Exchange Reactions of a Tri-Tungsten Cluster with Applications in Biomedical Imaging. *J. Chem. Educ.* **2011**, *88*, 793.
- Noey, E. L.; Wang, X.; Houk, K. N. Selective Gold(I)-Catalyzed Formation of Tetracyclic Indolines: A Single Transition Structure and Bifurcations Lead to Multiple Products *J. Org. Chem.* **2011**, *76*, 3477.
- Um, J. M.; DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. Quantum Mechanical Investigation of the Effect of Catalyst Fluorination in the Intermolecular Asymmetric Stetter Reaction *J. Am. Chem. Soc.* **2011**, *133*, 11249.
- Noey, E. L.; Luo, Y.; Zhang, L.; Houk, K. N. Mechanism of Gold(I)-Catalyzed Rearrangements of Acetylenic Amine-N-Oxides: Computational Investigations Lead to a New Mechanism Confirmed by Experiment *J. Am. Chem. Soc.* **2012**, *134*, 1078.
- Felnagle, E. A. Chauby, A.; Noey, E. L.; Houk, K. N.; Liao, J. C. Engineering Synthetic Recursive Pathways *Nature Chem. Biol.* **2012**, *8*, 518.
- DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. Catalytic Asymmetric Intermolecular Stetter Reactions of Enolizable Aldehydes with Nitrostyrenes: Computational Study Provides Insight into the Success of a New Catalyst *Angew. Chem. Int. Ed.* **2012**, *51*, 2391.
- Marcheschi, R. J.; Li, H.; Zhang, K.; Noey, E. L.; Kim, S.; Chaubey, A.; Houk, K. N.; Liao, J. C. A Synthetic Recursive “+1” Pathway for Carbon Chain Elongation *ACS Chem. Biol.* **2012**, *7*, 689.
- Jiménez-Osés, G.; Osuna, S.; Noey, E. L. Houk K. N. Molecular Dynamics Explorations of Active Site Structure in Designed and Evolved Enzymes. *Acc. Chem. Res.* DOI:10.1021/ar500452q
- Jiménez-Osés, G.; Osuna, S.; Noey, E. L. Houk K. N. Molecular Dynamics Explorations of the Enhanced Activity of Evolved Enzymes for the Kemp Elimination. *J. Am. Chem. Soc.* *Submitted*
- Noey, E. L.; Tibrewal, N.; Park, J.; Bond, C.; Cascio, D.; Osuna, S.; Jiménez-Osés, G.; Liang, J.; Zhang, X.; Huisman, G.; Tang, Y.; Houk, K. N. Origins of Stereoselectivity of Mutants of *Lactobacillus kefir* Ketoreductase. *Proc. Natl. Acad. Sci.* *Submitted*.

AWARDS

Undergraduate

CRC Freshman Chemistry Award (2006)
ACS Polymer Chemistry Award (2007)
ACS Analytical Chemistry Award (2008)
Arthur Furst Award: for academics and research (2008)
Father Flynn Award: Highest GPA in graduating class (2008)
Charles Harney Award: Senior student-athlete with highest cumulative GPA (2008)
Scholar Athlete Award for Track and Field and Cross Country (2008)
Mel Gorman Award (co-recipient): outstanding senior Chemistry major (2009)
Dean's Medal Finalist: one of four seniors who exemplifies the values of USF (2009)

Graduate

NIH Chemistry and Biology Interface Training Program at UCLA (2011-13)
Christopher Foote Senior Graduate Fellowship (2013)
Spring Quarter Departmental Fellowship (2014)

TALKS

“Selective Gold(I)-Catalyzed Formation of Tetracyclic Indolines: A Single Transition Structure and Bifurcations Lead to Multiple Products.” 241st ACS National Meeting, Anaheim, CA (2011)
“Computational Study of Ketoreductase Enantioselectivity” ACS Division of Organic Chemistry Graduate Research Symposium, U. of Delaware (2013)

TEACHING EXPERIENCE

General and Organic Chemistry Lab TA (2007-08)
Instrumentation TA for Organic Chemistry Labs (2009-10)
TA for Graduate-level Physical Organic Chemistry (2010)

INTERNSHIP

Intern at Codexis Inc., Singapore (June-August, 2012)

EXTRACURRICULAR

USF Division 1 Cross Country and Track & Field, and Team Captain (2004-08)
Wildflower Triathlon Olympic Distance Champion (2012)
USA Triathlon Collegiate National Champion (2014)
UCLA Club Sports Tier 1 Athlete of the Year (2014)
California Nano Systems Institute Outreach Volunteer (2014)

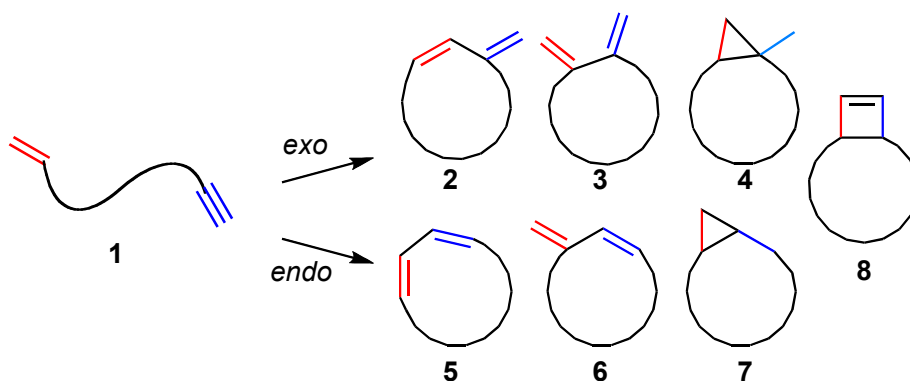
1. Selective Gold(I)- Catalyzed Formation of Tetracyclic Indolines: A Single Transition Structure and Bifurcations Lead to Multiple Products

Several alkynylindoles undergo gold(I)-catalyzed cyclization reactions to form a single isomer in each case. Density functional theory shows why this reaction is favored over the many possible regio- and stereoisomeric reaction pathways. This transformation involves a two-step no-intermediate mechanism with surface bifurcations leading to two or three products. Such bifurcations could explain reactivity in many gold(I)-catalyzed enyne cyclization reactions.¹

Gold(I) complexes are powerful catalysts for the electrophilic activation of alkynes towards a wide range of nucleophiles.² These complexes catalyze enyne cyclizations³ and are more reactive than platinum(II) and other electrophilic metal salts and complexes for this transformation.^{3e,h,4} Gold-catalyzed cyclization reactions of alkynes have been shown to take place without the change of gold oxidation state and involve complexes of one or two golds with the alkyne.⁵

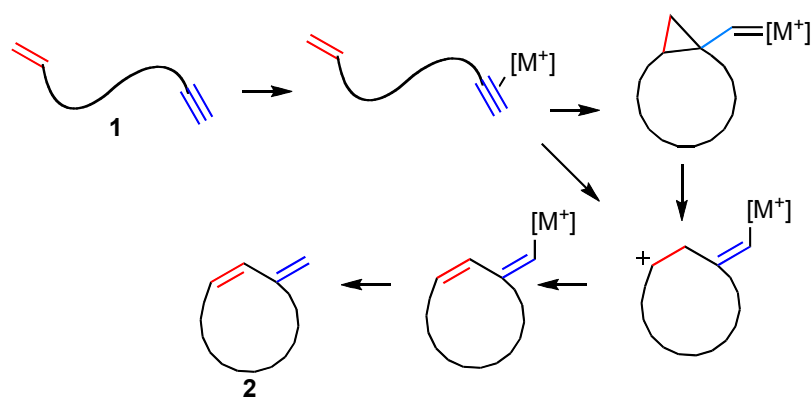
Gold(I)-catalyzed cyclizations of enynes, **1**, have been found to give a variety of products, **2-8**, depending on the nature of the substituents on the enyne³ and the ligand on the gold (Scheme 1.1).⁶ Platinum(II) catalysts give similar transformations.⁷

Scheme 1.1. Products formed by transition metal-catalyzed enyne cyclizations.



Scheme 1.2 illustrates how **2** is formed from **1**. First, the gold(I) or platinum(II) catalyst coordinates and activates the alkyne. Prior theoretical studies^{3c-f,h-i,n,7a-d,8} suggest that this intermediate can cyclize to form a cyclopropyl metal carbene. The cyclopropyl metal carbene, which has cationic character similar to that of the following structure^{3f,n,7g,9} can open to form the discrete carbocationic intermediate. Other possibilities for this intermediate are to rearrange, be deprotonated and protodemetalated, or undergo nucleophilic attack. Alternatively, the alkyne complex can undergo 6-*exo*-dig cyclization to directly form a carbocationic intermediate. Deprotonation and protodemetalation lead to the product, **2**, regenerating the catalyst.

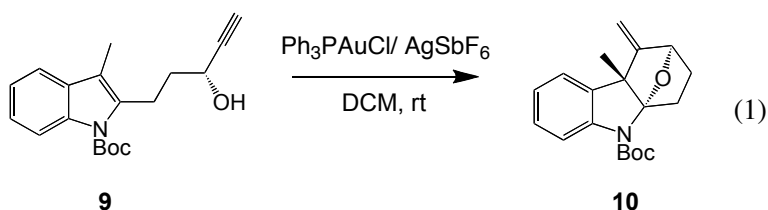
Scheme 1.2. Proposed mechanisms for the formation of **2** from **1**.



Gold catalyzed cyclizations of alkynylindoles have previously been studied, and a variety of regio- and stereoisomers are isolated.^{3f,g} We have performed a computational study of a very selective reaction of this type. Liu, Xu, and Wang discovered the gold-catalyzed cyclization of the alkynylindole, **9**, shown in equation (1).¹⁰ This reaction could have lead to many structural isomers analogous to **2-8**, but only **10** is formed. The mechanism proposed for the formation of **10** is shown in Scheme 1.3. The coordination of the alkyne by the phosphine-gold(I) catalyst (**11**) is followed by nucleophilic attack by the indole, leading to the cyclized intermediate **12**. Intramolecular attack of the alcohol on the iminium and deprotonation leads to the formation of

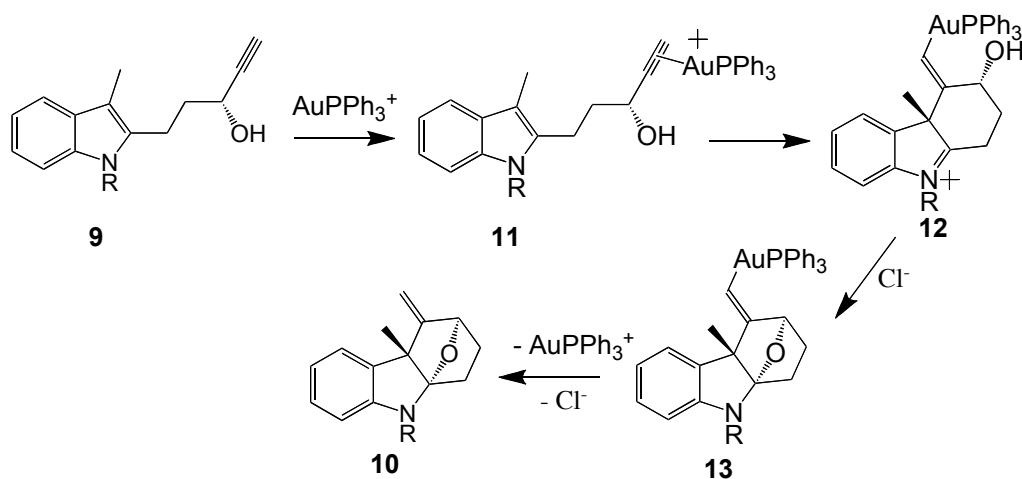
the ether bridge (**13**). Protodemetalation gives the product, **10**, and regenerates the catalyst. Prior studies involving enyne cyclizations of alkynylindoles suggest a different mechanism, in which a spiro intermediate and possibly a cyclopropyl gold carbene intermediate are formed prior to **12**.^{3f-}

^g This conclusion was based on the isolation of a spiro product.



A series of substrates and catalysts were screened for this transformation. Substitutions at various positions of the indole and of the alkyne were well tolerated. An electron-withdrawing group (EWG) on the indole nitrogen is required for the transformation from **9** to **10**. The EWGs screened include Boc, Ts, and CO₂Me. N-H and N-alkyl systems do not form products analogous to **10**, but instead form a complex mixture of products at higher temperatures.⁹ In order to elucidate the effect of the EWG, we modeled systems with and without EWGs, R=H and CN, respectively. The best catalyst screened for this reaction is PPh₃AuCl with AgSbF₆. We modeled the catalyst as AuPH₃⁺.

Scheme 1.3. Proposed mechanism from **9** to **10**



We have found that this reaction involves a bifurcating potential energy surface (PES) and is the first example of an enyne cyclization involving this type of surface. On a bifurcating PES, a single transition structure leads to two or more products, because there is a valley-ridge inflection point after the transition structure. This results in two consecutive transition structures with no minimum between them. This type of mechanism has been called a two-step no-intermediate mechanism.¹¹ Because the intermediates share a common transition structure, the shape of the PES and dynamics control the product distribution.¹² Bifurcations have been shown to be prevalent in many types of organic reactions and in two other gold(I)-catalyzed processes (Figure 1.1).¹¹

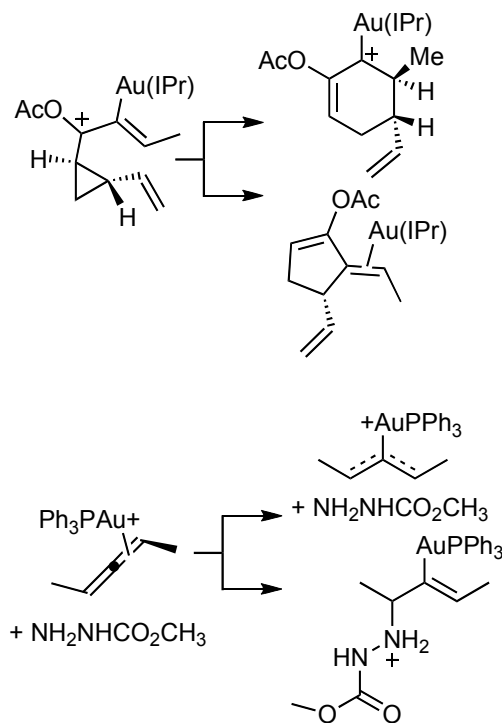


Figure 1.1. Examples of gold-catalyzed reactions with two-step no-intermediate mechanisms.^{12b-c}

All calculations were performed with Gaussian09¹³ with a hybrid density functional (B3LYP) with LANL2DZ and 6-31G(d) basis sets for gold and for all other atoms, respectively. Single point solvation energies were calculated using the CPCM model for dichloromethane. The possibility that two phosphine-gold(I) catalysts are involved was considered,^{5a} but was found to be unlikely. The results with a single phosphine-gold(I) catalyst are presented here. Transition structures **14** and **15**, along with the diastereomeric transition structures, were optimized in the CPCM solution model. Free energy surfaces from **14** and **15** were also computed in the solution model.

The origins of the regio- and stereoselectivity have been identified, and a new mechanism for enyne cyclization involving a single transition structure that can form multiple products has been discovered. We describe here our explanation of the selectivity in the transformation from **9** to **10**, and describe surface bifurcations involved in the enyne cyclization step.

Results and Discussion

Several conformations of the model for the gold(I)-alkyne complex, **11**, were optimized, and the lowest energy conformation was used as the zero of energy (**16**, Figure 1.2). The complexed C-C bond is polarized such that the terminal carbon is negatively charged and the internal carbon along with the gold and ligand are positive. The triple bond is only moderately lengthened from the uncomplexed triple bond ($\Delta=0.03$ Å). In agreement with prior gold-alkyne studies,^{4b} the gold center is “slipped” to one side of the p bond with alkynyl carbon-gold bond lengths around 2.2 to 2.4 Å.

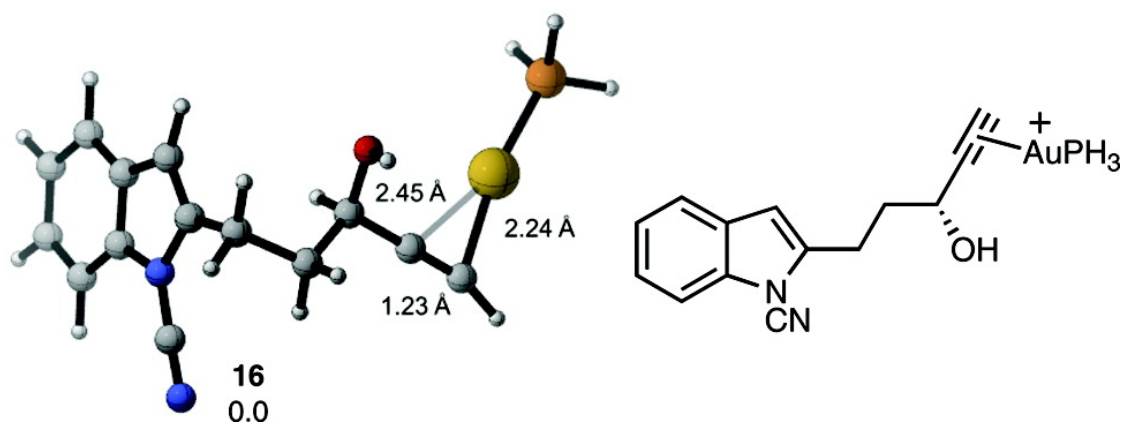


Figure 1.2. Optimized geometry of **16**, a computational model for **11**. A Chemdraw depiction of the structure is shown to the right.

The gold-alkyne complex can cyclize either by the *exo* (on the alkyne) or *endo* transition structures (**14** and **15**, Figures 1.3, 1.4) The diastereomers of **14** and **15** are **14'** and **15'** and involve attack on the opposite face of the indole (Figure 1.5). In an extensive search of the potential energy surfaces, we were surprised to find that there are far fewer saddle points than products, due to bifurcations of reaction paths. These are examples of two-step no-intermediate mechanisms.¹⁰ As shown in Figures 1.3 and 1.4, **14** forms **17** and **18** and **15** forms **20**, **21** and **22**.

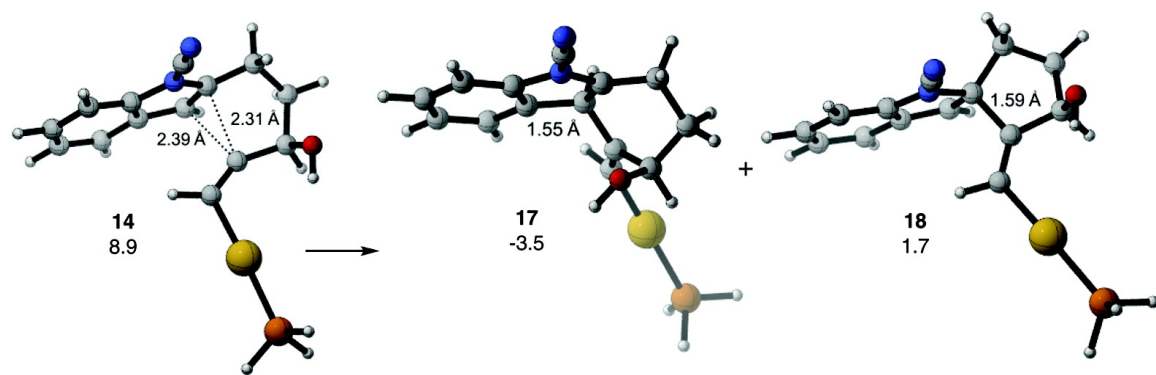


Figure 1.3. Transition state **14** forms intermediates **17** and **18**. C-C distances are in Å; energies are in kcal/mol.

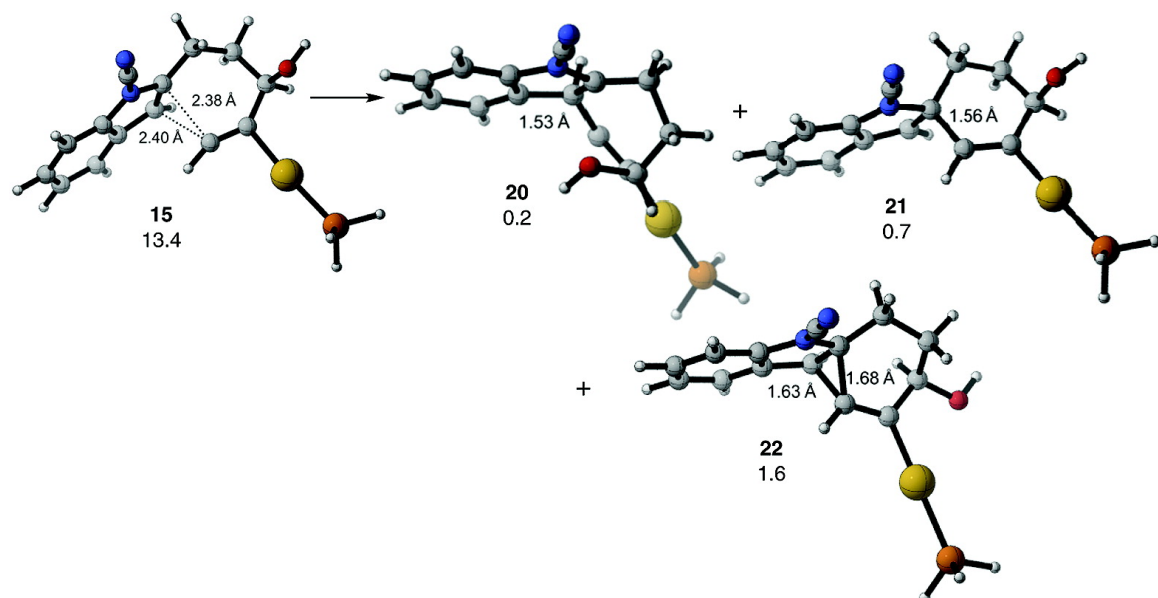


Figure 1.4. Transition state **15** forms intermediates **20**, **21** and **22**. C-C distances are in Å; energies are in kcal/mol.

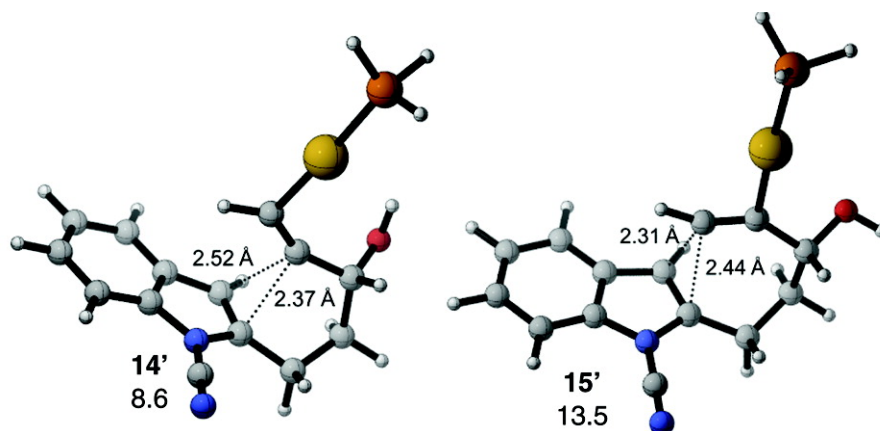


Figure 1.5. Exo (**14'**) and endo (**15'**) transition structures. C-C distances are in Å; energies are in kcal/mol.

As shown in Figure 1.6, transition structure **14** is a saddle point on a very flat surface that leads downhill to intermediates **17** and **18**, which are stabilized iminium and benzyl cations, respectively, and to transition structure **19** between them. Transition structure **19** is only slightly higher in energy than these intermediates, and involves more complete formation of both r_1 and r_2 than transition structure **14**.

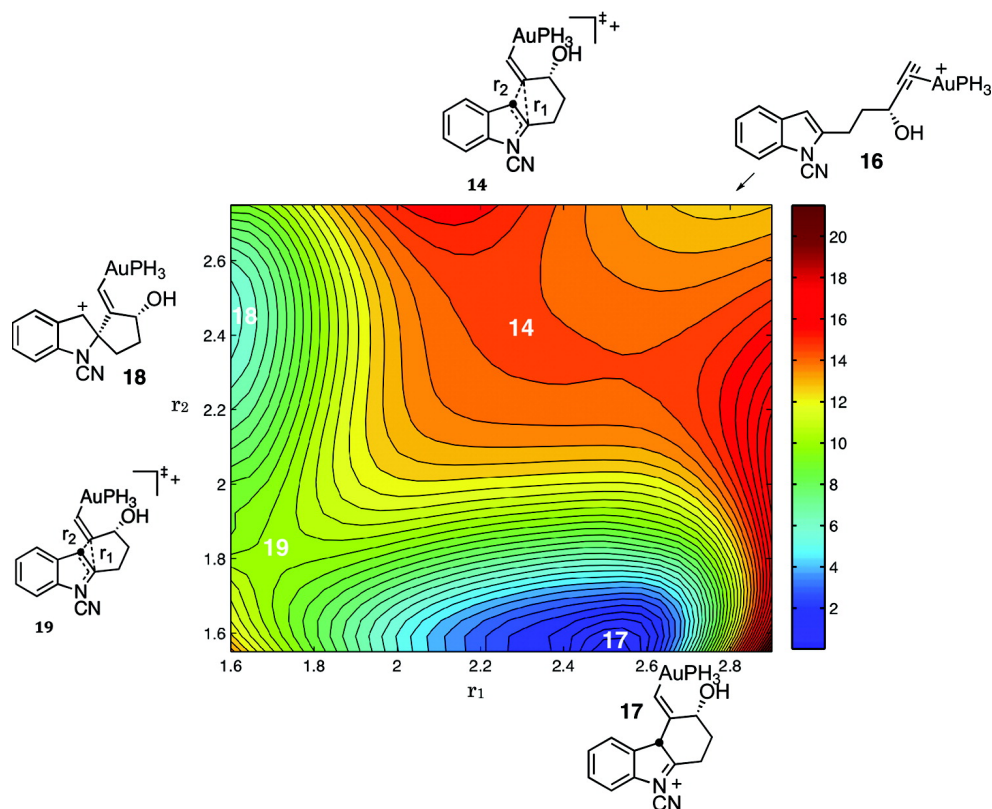


Figure 1.6. Potential energy surfaces from **16** through transition structure **14** to **17** and **18** calculated with B3LYP/6-31G(d) with the CPCM model in dichloromethane, smoothed with spline interpolation. Energies in kcal/mol.

The PES from **15** is more complex (Figure 1.7). In addition to the iminium and benzylic cations (**20** and **22**), a gold-coordinated cyclopropyl carbene intermediate, **21**, is a minimum as well. These structures are shown in Figure 1.4. This intermediate is not strictly a cyclopropyl metal carbene, because it has carbocationic and carbene character. This is consistent with previous studies on the nature of this type of intermediate.^{2f,n,6g,12}

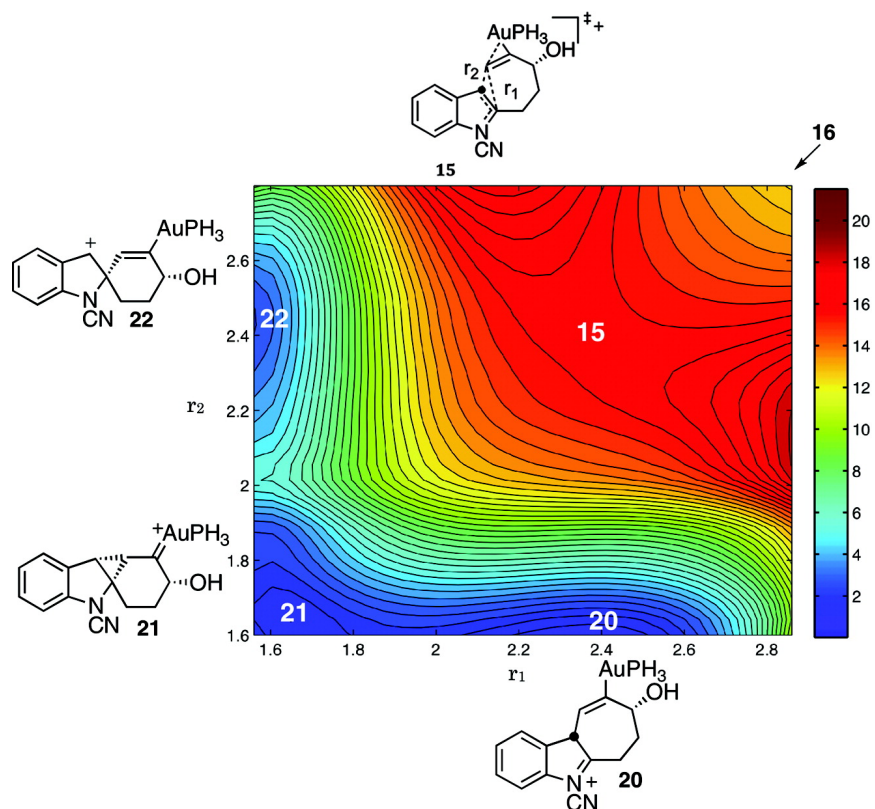


Figure 1.7. PES from **16** through transition structure **15** to **20**, **21** and **22**

Transition structure **14** and its diastereomer **14'** are 4 to 5 kcal/mol more stable than **15** and **15'**, so that **17**, **17'**, **18**, and **18'** are all formed rapidly. The formation of these intermediates is reversible, with a barrier of 13.4 kcal/mol for the conversion of **17'** back to **16**. Also, the spiro intermediates, **18**, and **18'** will convert to the more stable **17** and **17'** with a barrier of about 5 kcal/mol.

The cyclization to form the ether bridge is base-promoted. We have explored the reaction involving chloride as a base. Although there is not expected to be a high concentration of chloride in solution due to the addition of AgSbF_6 , chloride is used because the transition structure for this transformation was not found with other possible bases such as PH_3 , H_2O and OH^- . Because the alcohol must be deprotonated in order to form product, we accept chloride as a

model for the base in this transformation. In the transition structures, **23** and **23'**, chloride deprotonates the alcohol as the oxygen attacks C² of the indole ring (Figure 1.8). This transformation is highly favored for **17** but disfavored for **17'**. In structure **17**, C² and O are 2.56 Å apart, and in **17'**, they are 3.18 Å apart. Therefore, intermediate **17'** must distort more for the oxygen to be close enough to C² to form the ether bridge, so the barrier is higher, and leads to a highly strained intermediate, **24'**. However, the formation of the ether-bridged species **24** from **17** is exothermic by 18.8 kcal/mol. Because transformation from **17** to **24** is highly favored and is not reversible, only **24** is formed. Protodemetalation of **24** (also exothermic) leads to **25** and the regeneration of the gold catalyst. This agrees with experimental data, in which only **10** (which is analogous to **25**) is formed.

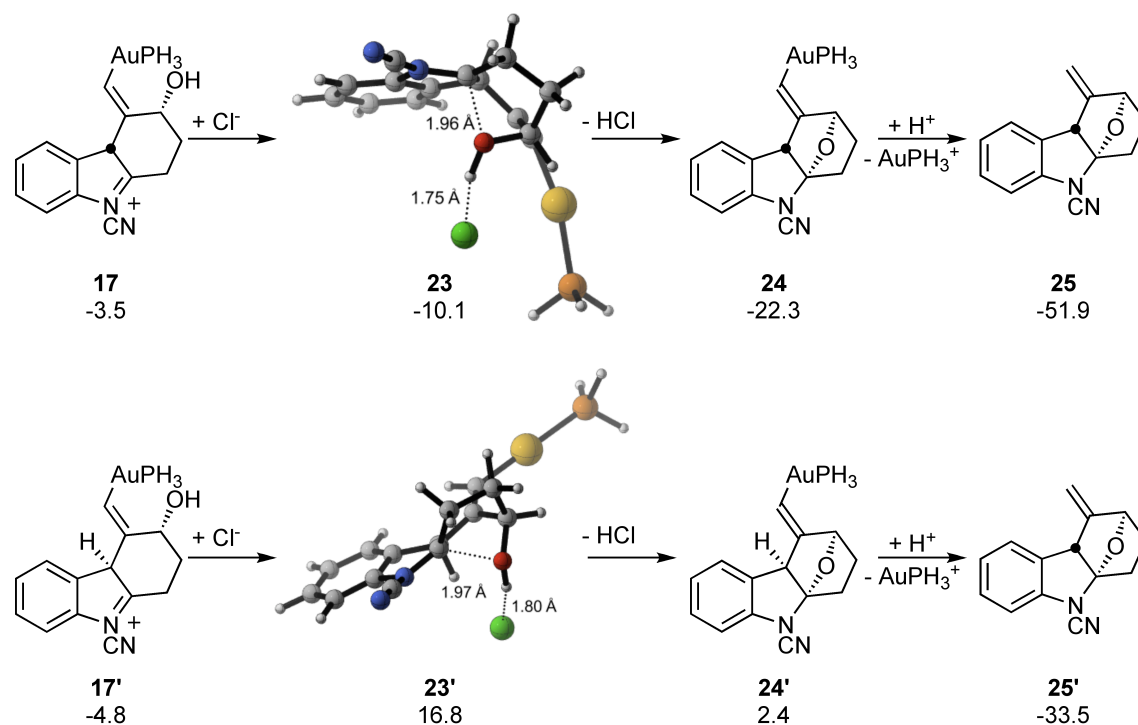


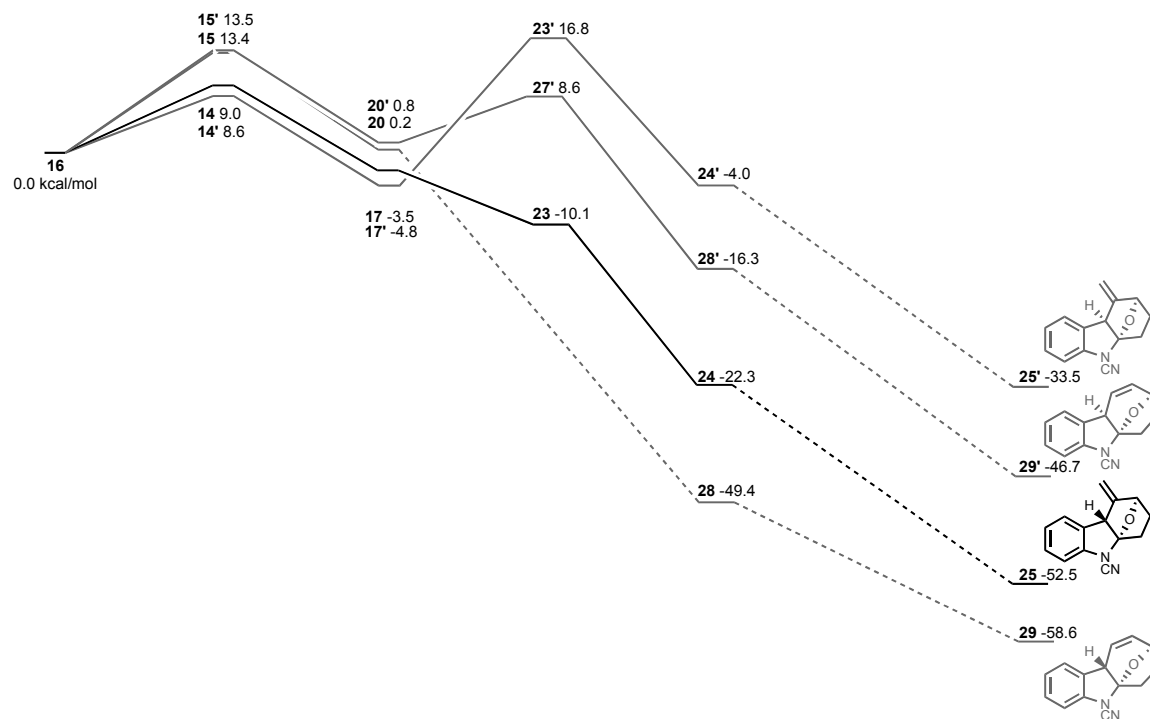
Figure 1.8. Transformation from **17** and **17'** to the ether-bridged species. Energies are in kcal/mol.

The free energy diagram (Scheme 1.4) shows the overall transformation from **16** to **25** and potential competing processes. The spiro and cyclopropyl carbene intermediates are not shown, as they do not directly lead to ether-bridged species. The diagram shows that *exo* transition structures **14** and **14'** are lower in energy than the *endo* transition structures **15** and **15'** by about 4 to 5 kcal/mol. Also, the *exo* intermediates formed, **17** and **17'**, are lower in energy than the *endo* intermediates, **20** and **20'**. Intermediate **17'** would form ether-bridged species **24'** through transition structure **23'**. The barrier from **17'** to **24'** (through **23'**) is 21.6 kcal/mol, while the barrier from **17'** back to starting material, **16**, is only 13.4 kcal/mol, so **17'** will revert back to **16**. However, **17** forms the ether-bridged species **24** with no barrier. Because intermediate **17** is cationic and transition state **23** is neutral and our solvation model (CPCM) underestimates the

solvation energy of charged species,¹⁴ we expect this barrier to be higher than indicated by our calculations. From **24**, protodemetalation is exothermic and gives the product, **25**.

As indicated above, the competing *endo* cyclizations through **15** and **15'** forming **20** and **20'** are highly disfavored compared to the *exo* cyclizations. The transition structure between **20** and **28** was not located, since **20** should not be formed. Interestingly, transition structure **27'** leads to another bifurcation point, such that in **27'** the chloride is positioned to deprotonate the alcohol, or the CH at C³ (Figure 1.9). Similar to the comparison between **24** and **24'**, intermediate **28** is lower in energy than **28'**. From **28** and **28'** protodemetalation would lead to **29** and **29'**. Ether-bridged species **29** is the lowest energy species overall but is not formed because of the inaccessibility of **15**.

Scheme 1.4. Free energy diagram for the formation of **25** and competing paths. Only the intermediates that lead to ether-bridged products are shown. Higher energy species **18**, **18'**, **21**, **21'**, **22** and **22'** are not shown for simplicity since they do not lead to cyclized product. The dashed line indicates that no transition state was located for this transformation.



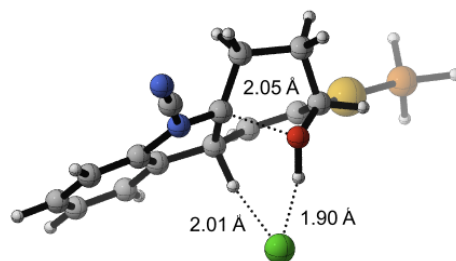
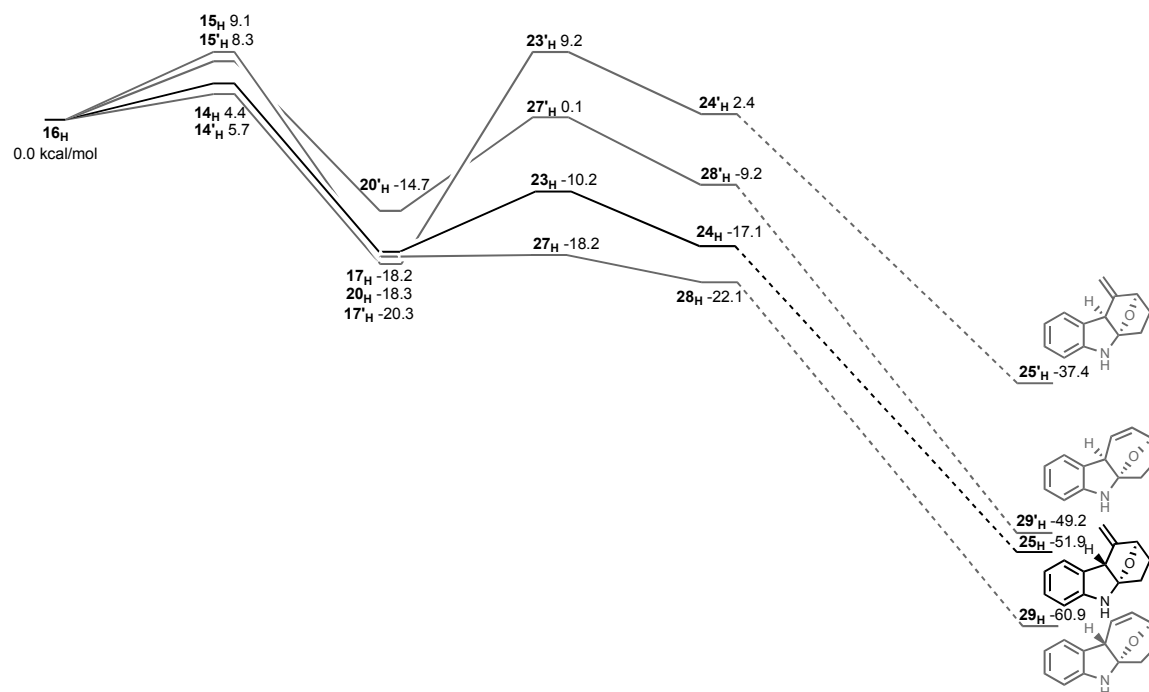


Figure 1.9. Transition structure **27**.

We also modeled the system with an unsubstituted indole nitrogen ($R=H$) in order to determine why an EWG is required for the transformation. The model system in which $R=H$ is comparable to the one discussed above, $R=CN$ (Scheme 1.5). The first cyclization step is more favorable with lower barriers (4 to 9 kcal/mol) and lower energy intermediates (-5 to -20 kcal/mol). This is because the unsubstituted indole is more nucleophilic than the cyano-substituted one and the carbocationic intermediate is more stable due to the absence of the EWG. The second step, in which the alcohol attacks C^2 , is greatly disfavored compared to that step in the cyano-substituted system. Here the barrier to cyclization is larger by 7 to 15 kcal/mol; the reactions are now endothermic in all cases except for formation of the ether bridge **28_H**. Specifically, the cyclization to form the ether bridge from **17** is barrierless ($\Delta G^\ddagger = -6.6$ kcal/mol) for the cyano-substituted system, while the analogous transformation from **17_H** is 8.0 kcal/mol ($\Delta\Delta G^\ddagger = 14.6$ kcal/mol). It is clear that an EWG is required because it makes C^2 much more electrophilic, thus promoting the second cyclization step. Without the EWG the barrier to cyclization is high and competing processes occur

Scheme 1.5. Free energy diagram for the formation of **25_H** and competing paths. Only the intermediates that lead to ether-bridged products are shown. Species **18_H**, **18'_H**, **21_H**, **21'_H**, **22_H** and **22'_H** are not shown for simplicity since they do not lead to cyclized product. The dashed line indicates that no transition state was located for this transformation.



The regio- and stereoselectivities of the gold-catalyzed indole-alkyne cyclizations of **9** and its analogs have been elucidated, and the effect of the EWG on the indole nitrogen has been determined. The first step of the reaction is selective in that it only allows for the formation of products that go through an *exo* transition structure. Of the intermediates produced, only **17** and **17'** can undergo a second cyclization step to form an ether bridge. In this step the formation of **24** from **17** is far more favorable than the formation of **24'** from **17'**. Thus, due to the trapping of **17** by the second cyclization step, **25** is the only product formed, as found experimentally. The EWG on nitrogen is essential in this reaction, because the formation of the ether bridge is disfavored without it.

For the first time a two-step no-intermediate mechanism has been identified for a transition metal-catalyzed enyne cyclization. This mechanism involves one transition structure

that can form up to three products: two different carbocationic species or the gold-coordinated cyclopropyl carbene. This is also the first time a single transition state has been found to lead to three products. Contrary to several other publications^{3d,f-j,o,7d,8} on enyne cyclizations, we find that it is not necessary to form the cyclopropyl intermediate in order to obtain the carbocationic species. This result could have mechanistic implications for a wide range of gold- and platinum-catalyzed reactions.

References

- ¹ Noey, E. L.; Wang, X.; Houk, K. N. *J. Org. Chem.* **2011**, *76*, 3477.
- ² (a) Hashmi, A. S. K. *Gold Bull.* **2003**, *36*, 3. (b) Hashmi, A. S. K.; *Gold Bull.* **2004**, *37*, 51. (c) Hoffmann-Röderand, A.; Krause, N.; *Org. Biomol. Chem.* **2005**, *3*, 387. (d) Ma, S.; Yu S.; Gu, Z. *Angew. Chem., Int. Ed.* **2005**, *44*, 200. (e) Hashmi, A. S. K.; *Angew. Chem., Int. Ed.* **2005**, *44*, 6990. (f) Li, Z.; Brouwer, C.; He, C. *Chem Rev.* **2008**, *108*, 3239. (g) Gorin, D.J.; Sjerrey, B.D.; Toste, F.D. *Chem Rev.* **2008**, *108*, 3351. (h) Hashmi, A. S. K.; Rudolph, M. *Chem. Soc. Rev.* **2008**, *37*, 1766.
- ³ Aubert, O. Buisine and M. Malacria, *Chem. Rev.* **2002**, *102*, 813. (b) G. C. Lloyd-Jones, *Org. Biomol. Chem.* **2003**, *1*, 215. (c) Echavarren A. M.; Nevado, C. *Chem. Soc. Rev.* **2004**, *33*, 431. (d) Diver, S. T.; Giessert, A. J. *Chem. Rev.* **2004**, *104*, 1317. (e) Nieto-Oberhuber, C.; Munoz, M. P.; Bunuel, E.; Nevado, C.; Cárdenas, D. J.; Echavarren, A. M. *Angew. Chem. Int. Ed.* **2004**, *43*, 2402. (f) Soriano, E.; Marco-Contelles, J. *Organometallics* **2006**, *25*, 4542. (g) Ferrer, C.; Echavarren, A. *Angew Chem.* **2006**, *118*, 1123. (h) Ferrer, C.; Amijs, C.; Echavarren, A. *Chem. Eur. J.* **2007**, *13*, 1358. (i) Amijs, C. H. M.; Lopez-Carrillo, V.; Raducan, M.; Perez-Galan, P.; Echavarren, A. M. *J. Org. Chem.* **2008**, *73*, 7721. (j) Horino, Y.; Yamamoto, T.; Ueda, K.; Kuroda, S.; Toste, F. D. *J. Am. Chem. Soc.* **2009**, *131*, 2809. (k) Shapiro, N. D.; Toste, F. D. *Synlett.* **2010**, *5*, 675. (l) Liu, Y.; Zhang, D.; Zhou, J.; Liu, C. *J. Phys. Chem. A* **2010**, *114*, 6164. (m) Sethofer, S. G.; Mayer, T.; Toste, F. D. *J. Am. Chem. Soc.* **2010**, *132*, 8276. (n) Lopez-Carrillo, V.; Echavarren, A. M. *J. Am. Chem. Soc.* **2010**, *132*, 9292.
- ⁴ (a) C. Nieto-Oberhuber, C.; Munoz, M. P.; López, S.; Jimenez-Núñez, E.; Nevado, C.; Herrero-Gomez, E.; Raducan M.; Echavarren, A. M. *Chem.–Eur. J.* **2006**, *11*, 1677. (b) Sharpio, N. D.; Toste, F. D. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2779.
- ⁵ (a) Cheong, P. H.; Morganelli, P.; Luzung, M. R.; Houk, K. N.; Toste, F. D. *J. Am. Chem. Soc.* **2008**, *130*, 4517. (b) Schmidbaur, H.; Schier, A. *Organometallics* **2010**, *29*, 2.
- ⁶ Benitez, D.; Tkatchouk, E.; Gonzalez, A. Z.; Goddard, W. A.; Toste, F. D. *Org. Lett.* **2009**, *11*, 4798.
- ⁷ (a) Chatani, N.; Kataoka, K.; Murai, S. *J. Am Chem. Soc.* **1998**, *120*, 9104. (b) Fürstner, A.; Szillat, H.; Stelzer, F. *J. Am. Chem. Soc.* **2000**, *122*, 6785. (c) Fürstner, A.; Stelzer, F.; Szillat, H. *J. Am. Chem. Soc.* **2001**, *123*, 11863. (d) Martin-Matute, B.; Nevado, C.; Cárdenas, D. J.; Echavarren, A. M. *J. Am. Chem. Soc.* **2003**, *125*, 5757. (e) Nevado, C.; Echavarren, A. M. *Chem. Eur. J.* **2005**, *11*, 3155. (f) Zhang, L.; Sun, J.; Kozmin, S. A. *Adv. Synth Catal.* **2006**, *348*, 2271. (g) Fürstner, A.; Davies, P. W. *Angew. Chem. Int. Ed.* **2007**, *46*, 3410.
- ⁸ (a) Nieto-Oberhuber, C.; Lopez, S.; Munoz, M. P.; Cardenas, D. J. Bunuel, E.; Nevado, C.; Echavarren, A. M. *Angew. Chem., Int. Ed.* **2005**, *44*, 6146. (b) Nieto-Oberhuber, C.; López, S.; Muñoz, M. P.; Jiménez-Núñez, E.; Buñuel, E.; Cárdenas, D. J.; Echavarren, A. M. *Chem. Eur. J.* **2006** *12*, 1694. (c) Jiménez-Núñez, E.; Echavarren, A. M. *Chem. Commun.* **2007**, 333. (d) Gorin, D. J.; Toste, F. D. *Nature Rev.* **2007**, *446*, 395. (e) Luzung, M. R.; Mauleon, P.; Toste, F. D. *J.*

Am. Chem. Soc. **2007**, *129*, 12402. (f) Soriano, E.; Marco-Contelles, J. *Acc. Chem. Res.* **2009**, *42*, 1026.

⁹ (a) Fürstner, A.; Morency, L. *Angew. Chem. Int. Ed.* **2008**, *47*, 5030. (b) Stephen, A.; Hashmi, K.; Benitez, D.; *Angew. Chem., Int. Ed.* **2008**, *47*, 6754. (c) Shapiro, N. D.; Tkatchouk, E.; Wang, Y.; Goddard, W. A.; Toste, F. D. *Nature Chem.* **2009**, *1*, 482. (d) Seidel, G.; Mynott, R.; Fürstner, A. *Angew. Chem. Int. Ed.* **2009**, *48*, 2510. (e) Perez-Galan, P.; Herrero-Gomez, E.; Hog, D. T.; Martin, N. J. A.; Maseras, F.; Echavarren, A. M. *Chem. Sci.* **2011**, *2*, 141.

¹⁰ Liu, Y.; Xu, W.; Wang, X. *Org. Lett.* **2010**, *12*, 1448.

¹¹ Singleton, D. A.; Hang, C.; Szymanski, M. J.; Meyer, M. P.; Leach, A. G.; Kuwata, K. T.; Chen, J. S.; Greer, A.; Foote, C. S.; Houk, K. N. *J. Am. Chem. Soc.* **2003**, *125*, 1319.

¹² (a) Ess, D. H.; Wheeler, S. E.; Iafe, R. G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K. N. *Angew. Chem., Int. Ed.* **2008**, *47*, 7592. Two investigations of Au(I)-catalyzed reactions involving a transition state bifurcation have been reported: (b) Garayalde, D.; Gomez-Bengoa, E.; Huang, X.; Goeke, A.; Nevado, C. *J. Am. Chem. Soc.* **2010**, *132*, 4720. (c) Wang, Z. J.; Benitez, D.; Tkatchouk, E.; Goddard, W. A.; Toste, F. D. *J. Am. Chem. Soc.* **2010**, *132*, 13064.

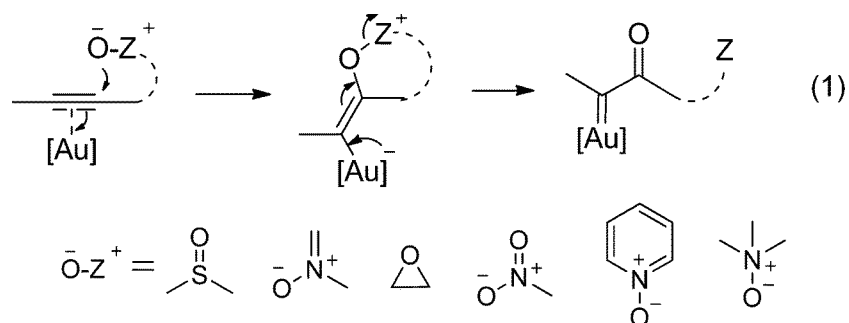
¹³ Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

¹⁴ Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70.

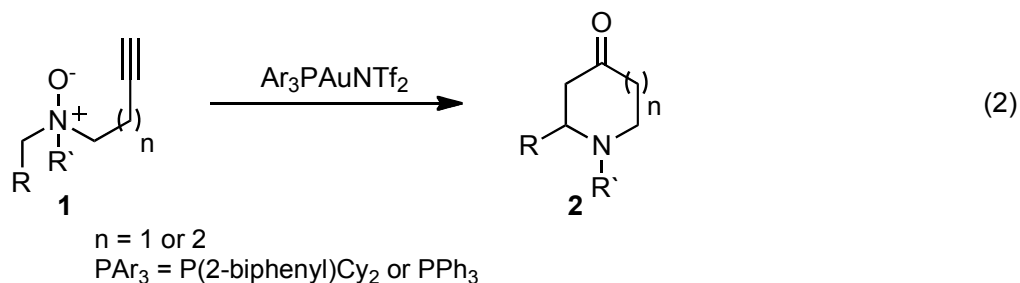
2. Mechanism of Gold(I)-Catalyzed Rearrangements of Acetylenic Amine-N-Oxides: Computational Explanations of Regioselectivities

Quantum mechanical studies of the mechanism of gold-catalyzed rearrangements of acetylene amine-N-oxides to piperidinones or azapanones are reported. A new mechanism is proposed involving a concerted hetero-retroene reaction of gold-coordinated isoxazolidiniums and oxazinaniums. Computational investigations of the hetero-retroene mechanism reproduce experimental regioselectivities and provide an explanation as to why the hydrogen shift occurs from the smaller amine substituent.¹

Cationic gold catalysts activate π bonds for reactions with a range of nucleophiles, leading to sigmatropic rearrangements, migrations, and cycloisomerizations.² A particularly important class of nucleophiles is an oxidant that possesses a nucleophilic oxygen and can formally deliver the oxygen atom during the reaction, and this type of gold-catalyzed alkyne oxidation has been proposed to generate a reactive α -oxo gold carbene intermediate, which would undergo an array of versatile transformations (eq 1)³ Early studies in this area use tethered oxidants such as sulfoxide,⁴ nitron,⁵ epoxide,⁶ nitro,⁷ and amine *N*-oxide,⁸ and recently the Zhang group used pyridine/quinoline *N*-oxides as external oxidants to achieve intermolecular alkyne oxidation.⁹ While α -oxo gold carbenes have been invoked as reactive intermediates in many of these studies, studies using external arylsulfoxides¹⁰ argue against such intermediacy. Here we present computational and experimental studies to suggest that gold carbene intermediates are not involved in two intramolecular cases; moreover, we disclose the first example of a gold-catalyzed reaction involving a heteroretroene reaction.

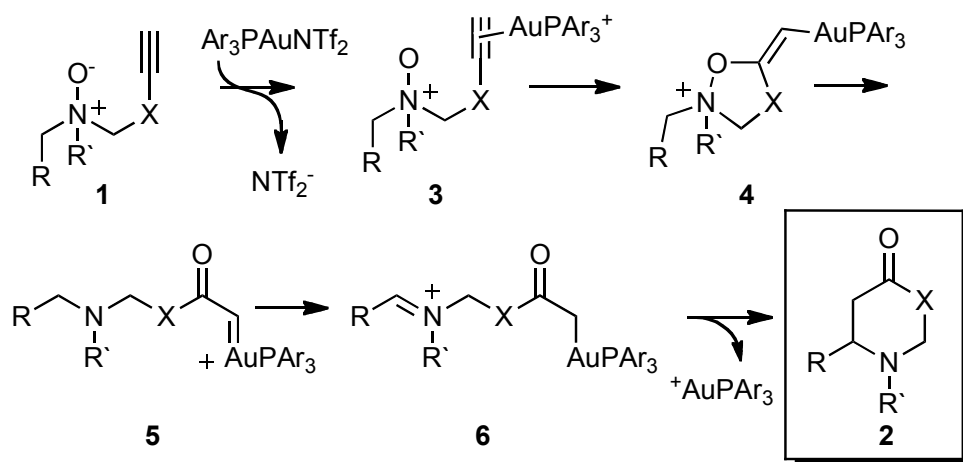


The Zhang group reported the gold-catalyzed annulation to form piperidin-4-ones or azepan-4-ones (eq 2).

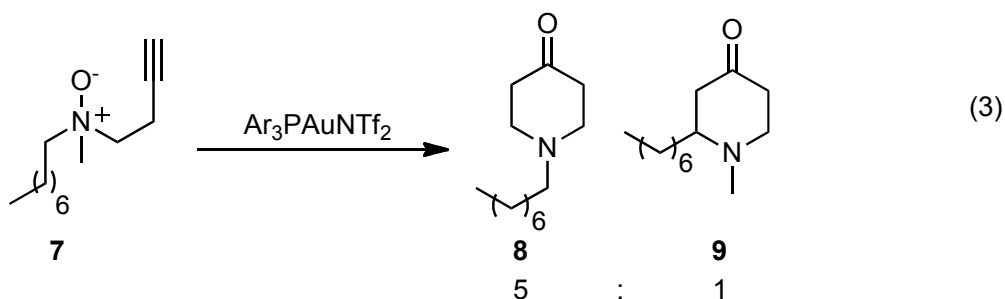


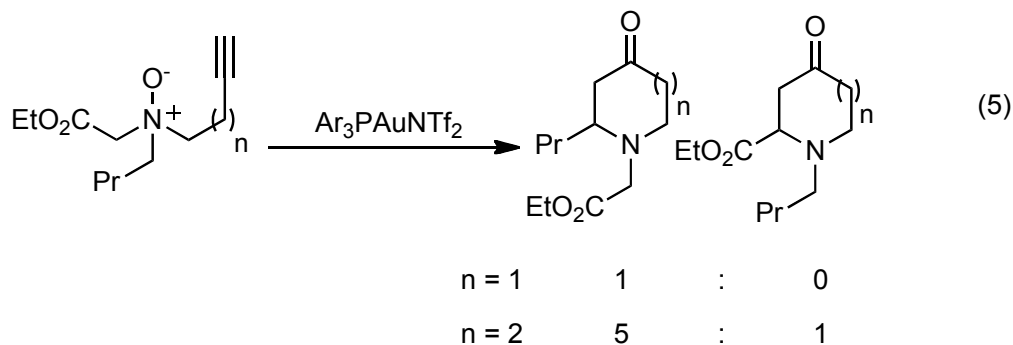
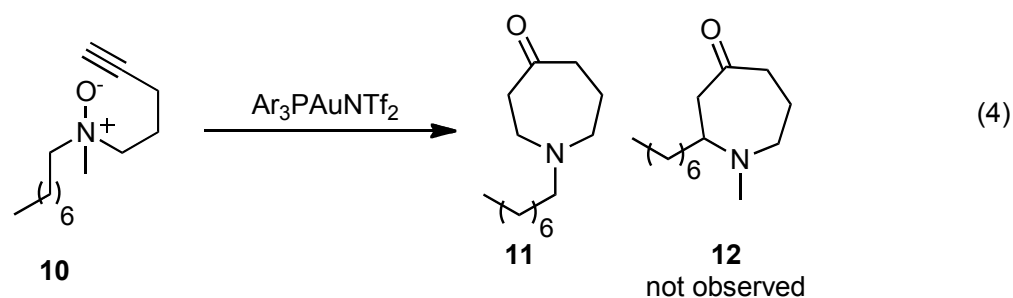
These reactions were hypothesized to proceed via the mechanism in Scheme 2.1. The amine oxides, **1**, are prepared by reactions of amines with butynyl or pentynyl triflates followed by oxidation with *m*CPBA. Coordination of **1** with gold(I) catalyst is proposed to form **3**. Nucleophilic attack on the gold-activated alkyne, **3**, forms **4**. Intermediate **4** was thought to ring-open to form a gold carbenoid intermediate, **5**. Intermediate **5** undergoes a 1,6-hydride shift or 1,7-hydride shift to form the iminium intermediate **6**, which leads to the product, **2**. While the proposed mechanism accounts for the formation of the major product, it fails to explain the selective hydride migration from one of the *N*-alkyl substituents.

Scheme 2.1. Proposed mechanism⁸ for the rearrangements of acetylene amine-N-oxides to piperidinones and azapanones. When X=CH₂ piperidinones are formed and when X= CH₂CH₂ azapanones are formed.



That is, when the amine oxide has two different substituents, the hydrogen shift occurs selectively from the smaller group. The transformation shown in eq 3^{8a} yielded a 5:1 ratio of products (**8**:**9**). The longer chain substrate, **10**, in eq 4,^{8b} yielded a single product, **11**. Acidic hydrogens α to an ester group appear to migrate either less effectively ($n = 2$) or not at all ($n = 1$) (eq 5). We now present computational evidence for a different mechanism that accounts for the selection observed in these reactions. A novel heteroretroene reaction is identified.





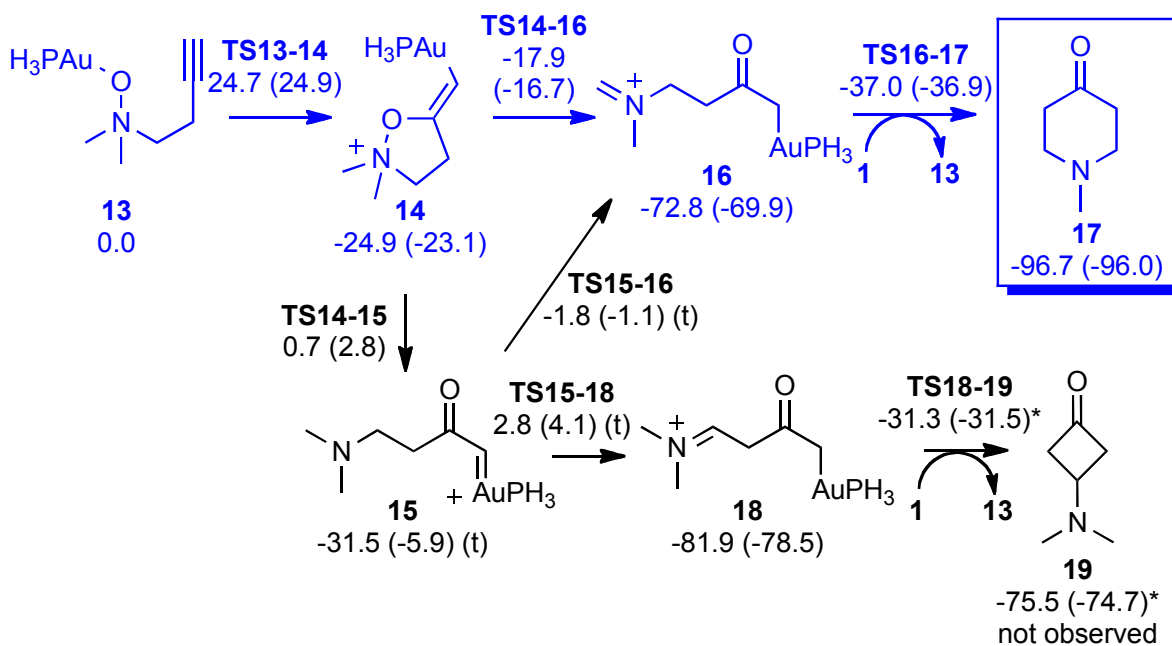
Calculations were performed with Gaussian09¹¹ with a hybrid density functional (B3LYP and B3LYP-D3¹²) with LANL2DZ and 6-31G(d) basis sets for gold and for all other atoms, respectively. Structures were optimized with the CPCM model for DCM solvent. The coordination of the gold to either the N-oxide or alkyne as well as the *syn* addition to the alkyne were additionally modeled with PPh₃ and gave comparable results to that of the PH₃ ligand. Modeling PPh₃ with PH₃ is very common¹³ because it gives good structures and approximates reaction barriers.^{13c,14} B3LYP has recently been shown to give reasonable energy and geometry predictions for homogenous gold catalysis.¹⁵

Two different mechanisms for the conversion of **13** to **17** were investigated; the original proposal is shown in black and the lowest energy path described here is shown in blue in Scheme 2.2. The cationic gold catalyst preferentially coordinates to the N-oxide to give **13**. The alkynyl-gold complex, **20**, is 15.6 kcal/mol higher in energy than **13** (Scheme 2.3). Calculations with PPh₃ as the gold ligand predict that coordination of gold to the N-oxide is favored by 15.4

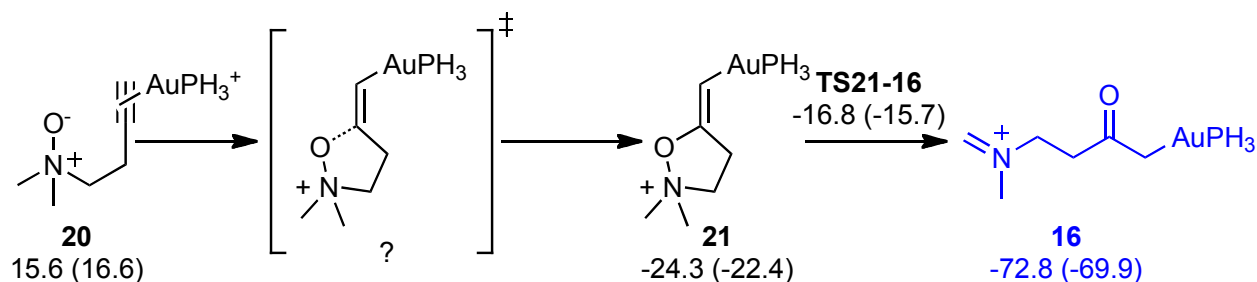
kcal/mol over coordination to the alkyne. The *syn* addition to the alkyne from **13** has a barrier of 24.7 kcal/mol (**TS13-14**). The reaction is exothermic by 24.9 kcal/mol and gives intermediate **14**. *Anti* addition has been established for a variety of nucleophilic additions to alkynes;¹⁶ however, here the N-oxide strongly coordinates to the cationic gold, and no transition structure for the *anti* addition could be located (Scheme 2.3).¹⁷

If the N-oxide were protonated by *m*-chlorobenzoic acid (*m*-CBA), the gold cation would be less likely to coordinate to the OH group. However, calculations indicate that proton transfer from *m*-CBA to the N-oxide, **1**, is endothermic by 7.3 kcal/mol in DCM.

Scheme 2.2. B3LYP/6-31G(d)/CPCM(DCM) (B3LYP-D3) energetics for the conversion of the gold-coordinated N-oxide, **13**, to cyclic ketones **17** and **19**. Compound **19** is not observed experimentally. (t) indicates the species is a ground-state triplet. *gas phase optimization with a single point solvation correction.



Scheme 2.3. *Anti* addition pathway.



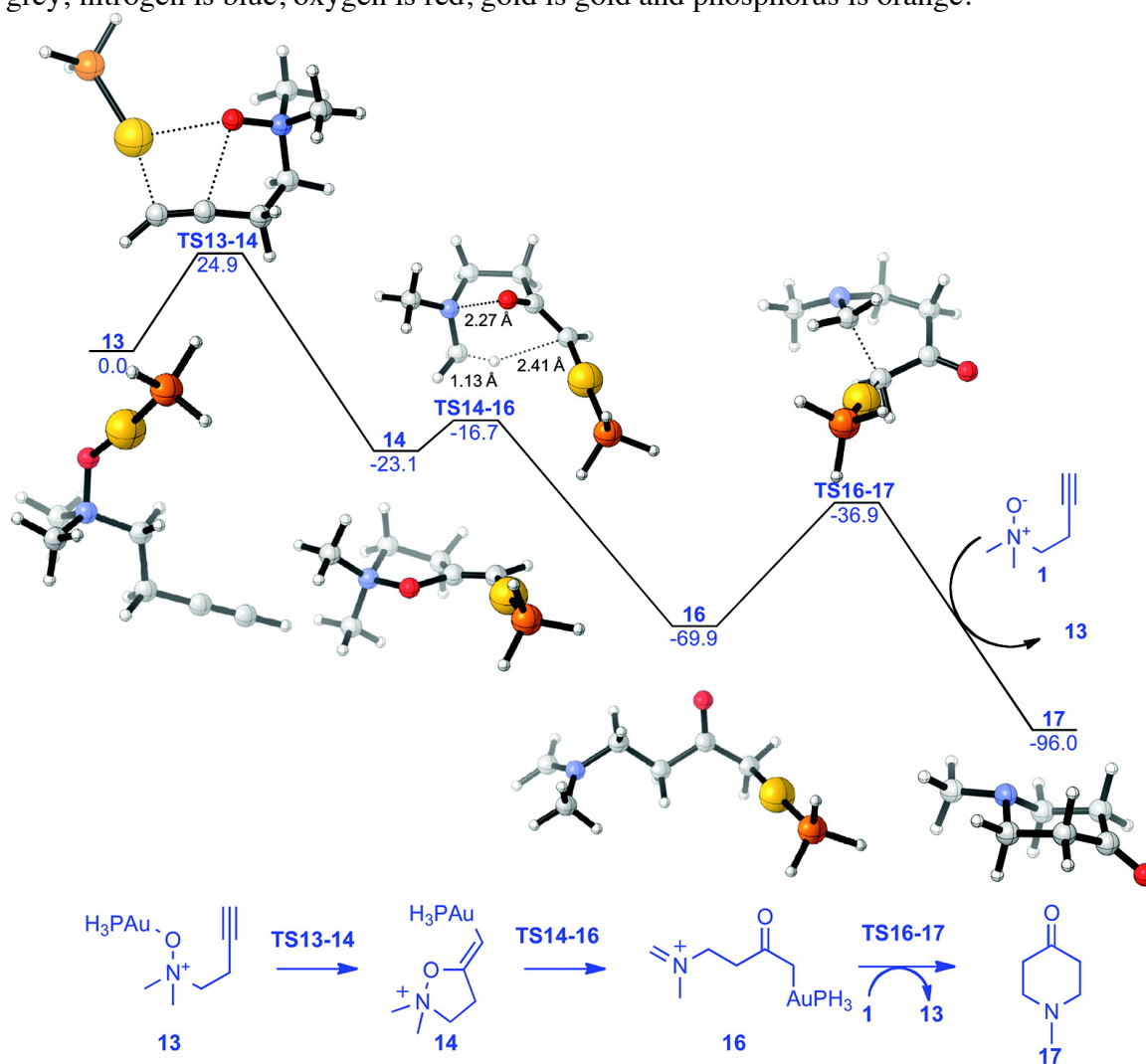
Along the proposed mechanistic path, the ring-opening of **14** has a barrier of 25.6 kcal/mol and leads to a triplet carbenoid, **15**. This triplet acyl carbenoid is stabilized by conjugation with the carbonyl, while the singlet (5.9 kcal/mol higher in energy) is less stabilized. Alpha-carbonyl carbenes are ground state triplets and adopt a planar geometry.¹⁸ Alkyl gold carbenoids have been involved in other gold(I)-catalyzed reactions and are predicted to be singlets.^{6c,19}

Intermediate **15** can undergo a 1,6-hydride shift with a barrier of 29.6 kcal/mol (**TS15-16**) to form **16**. The barrier for the 1,4-hydride shift (**TS15-18**) is 4.6 kcal/mol higher in energy than that for the 1,6-hydride shift.

An alternative to the mechanism involving the sequential ring-opening and 1,6-hydride shift was found to be energetically much more favorable. This new mechanism involves a hetero-retroene reaction and leads directly from **14** to **16**, thus avoiding the gold-carbenoid intermediate, **15**, altogether. This step is favored over ring-opening by 18.6 kcal/mol. Cyclization of **16** forms the product, **17**, and regenerates the catalyst. This final step was calculated to have a barrier of 35.8 kcal/mol. This barrier should be lowered by the presence of an NTf_2^- counterion or the coordination to a molecule of the N-oxide starting material, **1**. Also, the actual aryl phosphine ligands used experimentally are more electron-rich than our model phosphine, which should facilitate the dissociation of the gold catalyst. This step is exothermic by 23.9 kcal/mol.

Scheme 2.4 summarizes the lowest energy path from **13** to **17**. In this path, the first step is the *syn* addition of the gold-coordinated N-oxide, **13**, to form **14**. Intermediate **14** undergoes the hetero-retroene reaction, and forms **16**. The final step is the cyclization of **16** to yield the piperidine product, **17**, and regenerate the catalyst. This cyclization is calculated to be the rate-determining step. The overall transformation is exothermic by 96.7 kcal/mol.

Scheme 2.4. B3LYP-D3 free energy diagram for the lowest energy path from **13** to **17**. Carbon is grey; nitrogen is blue, oxygen is red, gold is gold and phosphorus is orange.



The role of the gold catalyst in the hetero-retroene reaction was probed by comparing the gold-catalyzed reaction to one without the AuPH₃ group. Figure 2.1 shows the transition structures for the reaction of **14** and of **22**. In **22** AuPH₃ is replaced by H. While the reactions are equally exothermic, the barrier is 9.5 kcal/mol lower in energy with the gold catalyst than without. In **14** the cationic gold-phosphine makes the adjacent carbon more negative compared to **22**. Because the carbon in **14** is more nucleophilic, the hydrogen transfer is favored, resulting in a lower energy transition structure.

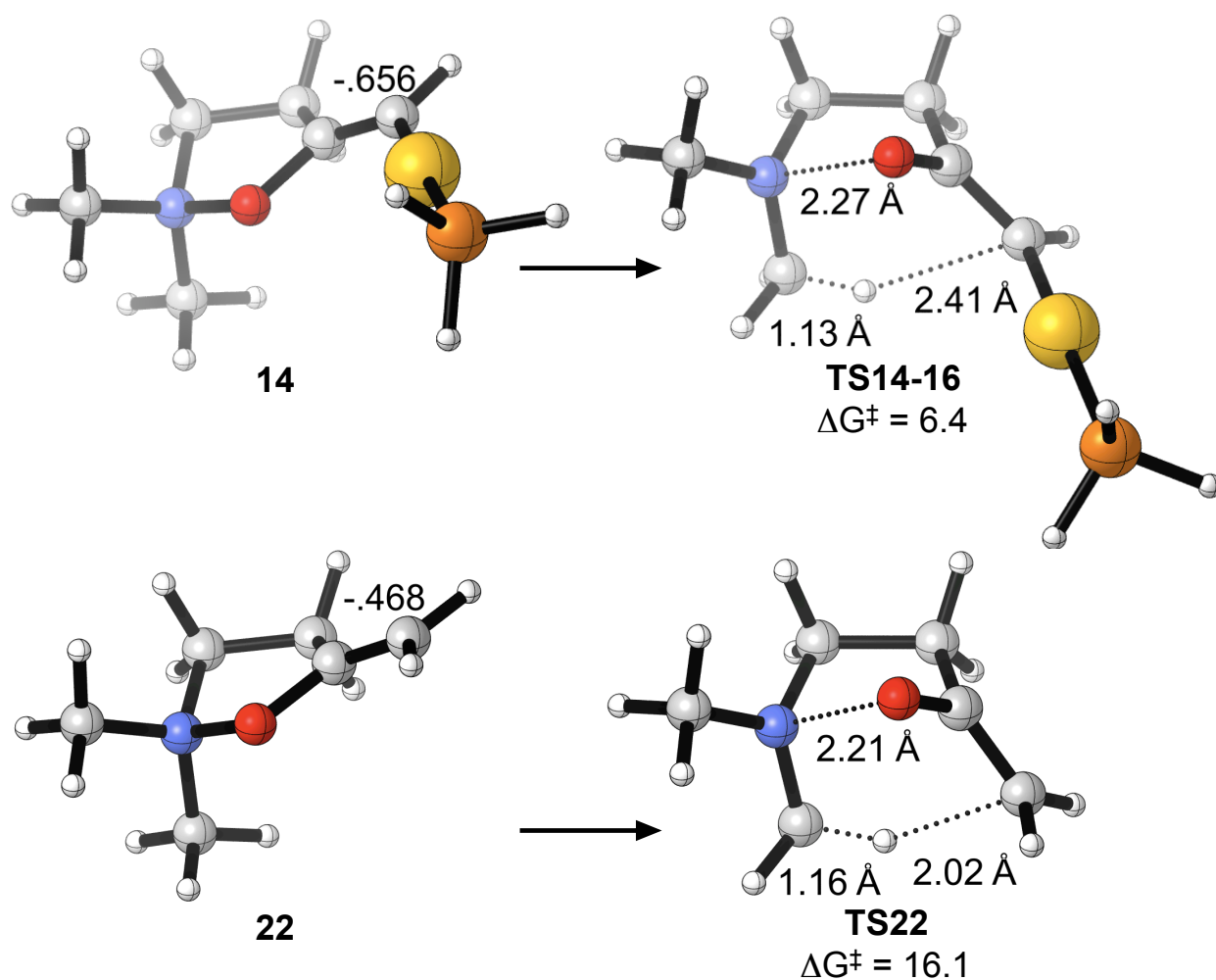


Figure 2.1. Comparison of the hetero-retroene reaction with and without gold including bond distances and Mulliken charges. The ΔG_{rxn} is -46.9 and -48.4 kcal/mol with and without gold, respectively. Free energies determined by B3LYP-D3.

As indicated earlier, experiments showed that the hydrogen is transferred from the smaller amine substituent.² This observation is consistent with the newly proposed mechanism. Figure 2.2 shows the hetero-retroene transition structure, **TS14-16**. The least sterically hindered position on the amine substituent (highlighted in green) is not involved in the hetero-retroene reaction. A quantitative study of this phenomenon was undertaken as well. The acetylenic amine-N-oxide, **7**, under experimental conditions afforded **8** and **9** in a 5:1 ratio (Equation 3). This product distribution corresponds to the hetero-retroene reaction preferentially involving hydrogen shift from the methyl group rather than the octyl substituent. Calculations on the N-methyl, N-ethyl substrate gave a 1.0 kcal/mol preference for **TS7-8**, corresponding to a 5.4:1 product ratio (Figure 2.3).

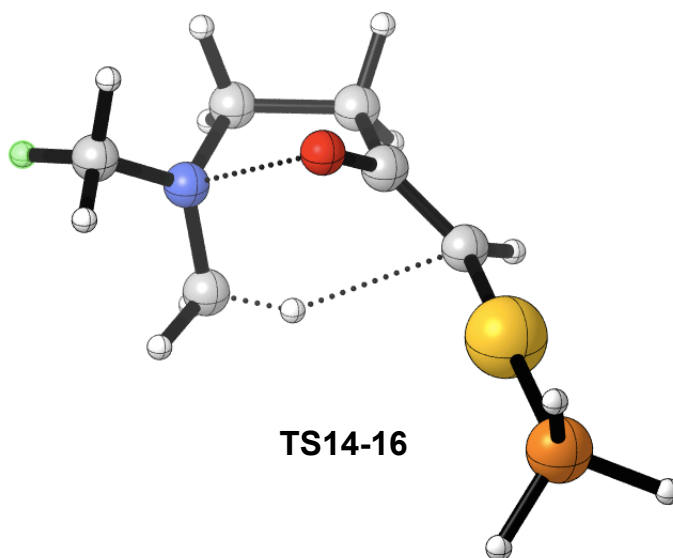


Figure 2.2. The hetero-retroene transition structure, **TS14-16**. The highlighted (green) hydrogen is the at least sterically hindered position on the N substituents.

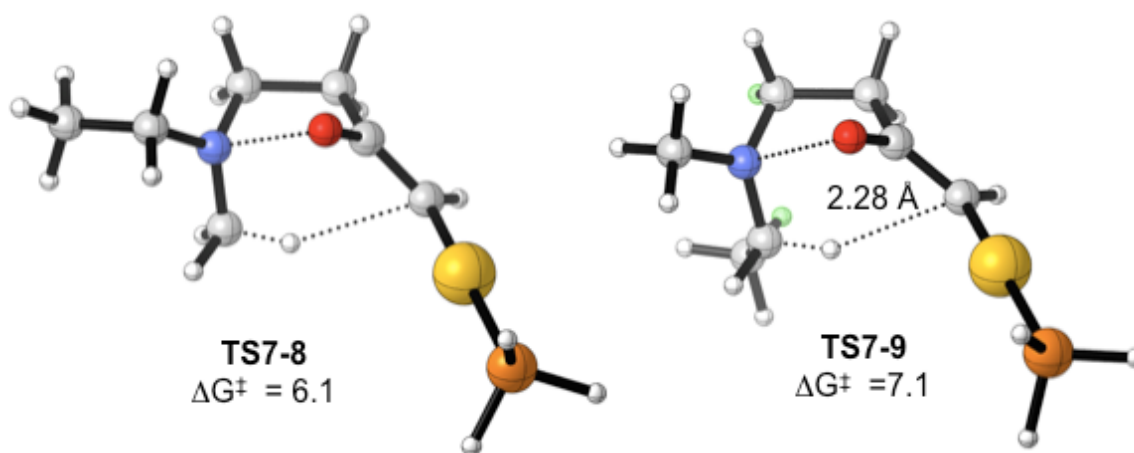
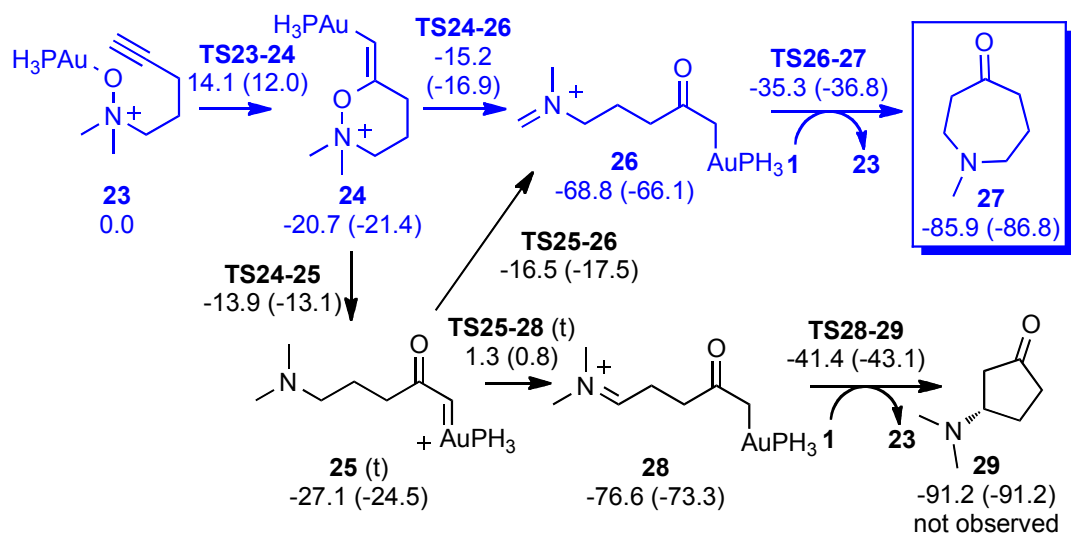


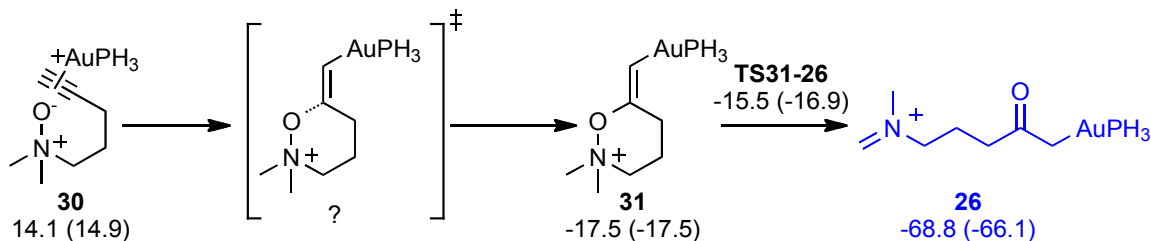
Figure 2.3. Hetero-retroene reaction transition structures dictating the product distribution of **8** and **9**. Highlighted hydrogens in **TS7-9** are 2.28 Å apart. Free energies are determined by B3LYP-D3.

Similar mechanistic and regioselectivity studies were carried out for the annulation to azepanones. The mechanism for the conversion of **23** to **27**, including competing pathways, is shown in Scheme 2.5. Again, the gold preferentially coordinates to the N-oxide rather than the alkyne, and no transition structure corresponding to the *anti* addition could be located (Scheme 2.6). When PPh₃ is used, calculations predict that coordination of gold to the N-oxide is favored by 10.0 kcal/mol and the barrier for the *syn* addition is 11.1 kcal/mol. The hetero-retroene mechanism is favored over the mechanism involving stepwise ring-opening followed by hydride shift.

Scheme 2.5. B3LYP/6-31G(d)/CPCM(DCM) (B3LYP-D3) energetics for the conversion of the gold-coordinated N-oxide, **23**, to piperidinone, **27** and **29**. Compound **29** is not observed experimentally. (t) indicates the species is a ground-state triplet.



Scheme 2.6. *Anti* addition pathway.



From the gold coordinated N-oxide intermediate, **23**, the *syn* addition to the alkyne forms **24**. Intermediate **24** can either ring-open or undergo a hetero-retroene reaction. Analogous to the annulation to piperidinones, the hetero-retroene transition structure, **TS24-26**, is favored, but only by 1.7 kcal/mol over the ring-opening. This indicates that a small amount of the gold carbenoid, **25**, could be formed, although the necessity for intersystem crossing will disfavor this pathway. Intermediate **25** undergoes a 1,7-hydride shift, rather than a 1,5-hydride shift ($\Delta\Delta G^\ddagger = 17.8$ kcal/mol). Iminium intermediate, **26**, leads to the product **27**. Again, the barrier for the cyclization is overestimated by our calculations. The transformation from **23** to **27** is exothermic

Calculations are in good agreement with the experimental regioselectivity observed in this transformation as well. The hetero-retroene transition structure, **TS24-26**, is in a chair conformation with the substituent involved in the rearrangement in an axial position (Figure 2.4). This is consistent with the experimental observation that smaller groups are involved in the hydrogen shift. The acetylenic amine-N-oxide, **10**, under experimental conditions afforded only **11** (Equation 3). As shown in Figure 2.5, the activation energy of the hetero-retroene reaction involving the methyl group is 2.2 kcal/mol lower in energy than that involving the ethyl substituent. The prediction is that a 43:1 ratio of products should occur.

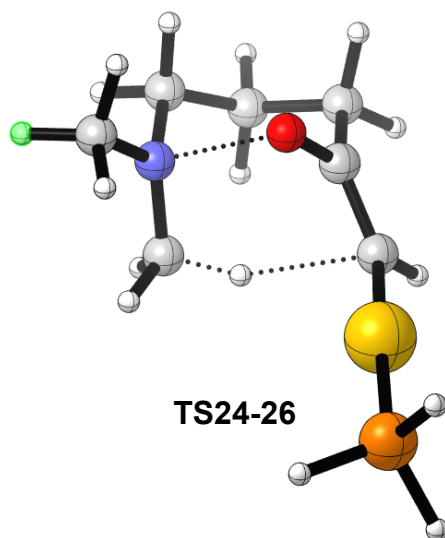


Figure 2.4. The hetero-retroene transition structure, **TS24-26**, for the transformation from **24** to **26**. The highlighted hydrogen is at the least sterically hindered position on the N substituents.

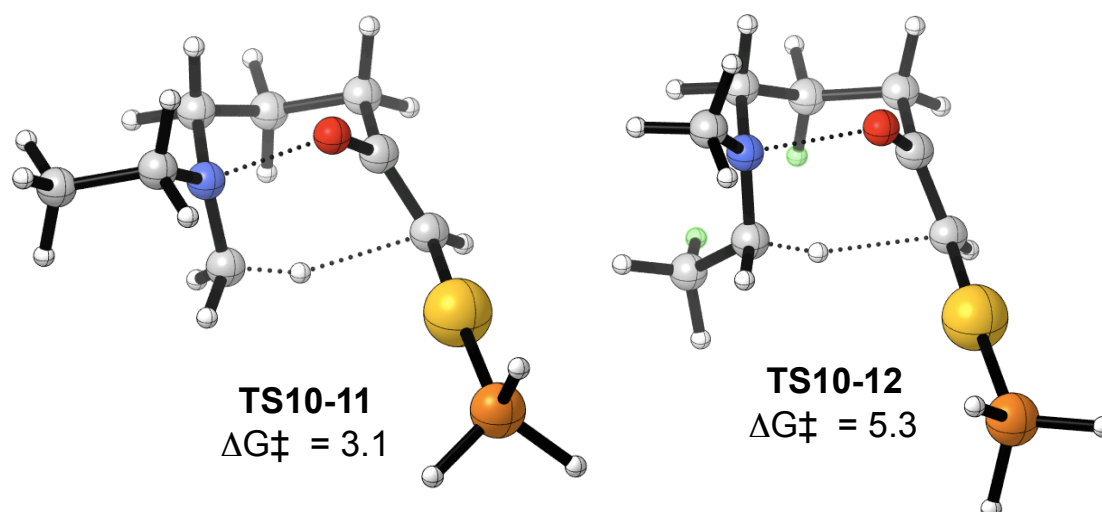
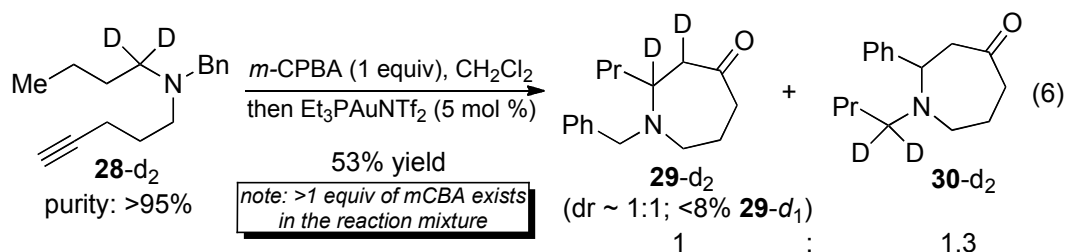


Figure 2.5. The hetero-retroene reaction transition structures for the formation of **11** and **12**. The highlighted hydrogens in **TS10-12** are 2.10 Å apart. Free energies are determined by B3LYP-D3.

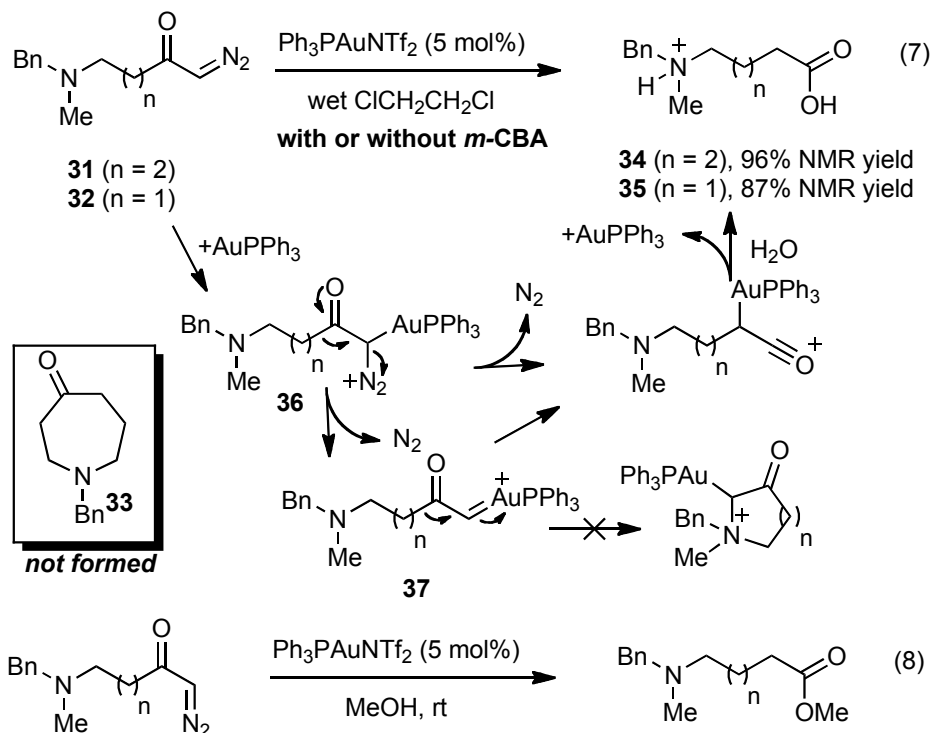
The annulation to azapanones is more regioselective than the annulation to piperidinones (Equations 2 and 3), because the axial/equatorial difference is larger in the chair-like transition structure, (**TS10-12**, Figure 2.5) than in the envelope transition structure, (**TS9-11**, Figure 2.3). The distances between the ethyl substituent and the nearest hydrogen (both highlighted in green) in **TS10-12** and in **TS9-11** are 2.10 and 2.28 Å, respectively.

Experimental Studies

These computational studies indicate that the hydrogen migration is intramolecular and the intermediacy of gold carbene intermediates of type **5** (Scheme 2.1) is unlikely. Experiments were performed to provide further support for these conclusions and thereby the calculated hetero-retroene mechanism.

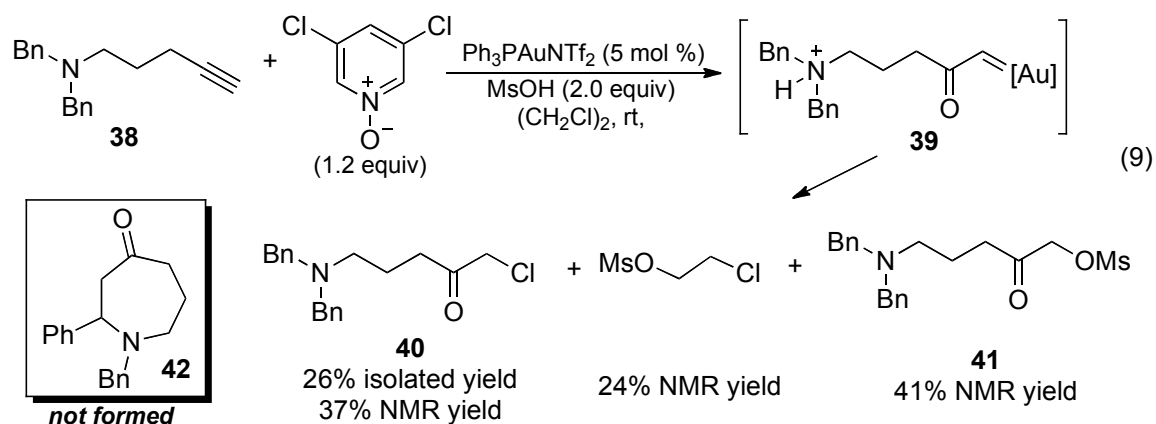


To examine the hydrogen migration, the α -methylene group in the tertiary amine **28-d₂** was fully labeled by deuterium. When compound **28-d₂** was subjected to the oxidation and gold catalysis sequence, little deuterium loss was detected in the *N*-benzylazepanone product **29-d₂** (Eq. 5), suggesting that the hydrogen/deuterium migration is intramolecular. As more than 1 equivalent of *m*-CBA was present in the reaction mixture during the gold catalysis, an intermolecular ionic process would most likely lead to substantial loss of deuterium. This result is consistent with the hetero-retroene mechanism.



To offer evidence against the intermediacy of gold carbene **5** or related carbenoid intermediates, the diazo ketone **31** was prepared²⁰ and treated with Ph₃PAuNTf₂ in 1,2-

dichloroethane. The expected azepan-4-one product **33** was not formed (Eq. 6); instead, the Wolff rearrangement product, amino acid **34**, was formed in 96% NMR yield. The addition of stoichiometric amount of *m*-CBA did not alter the reaction outcome. With **32**, similar results were observed; amino acid **34**, but not the piperidine-4-one product, was observed. When MeOH was used as solvent, the corresponding methyl ether was isolated in 70% yield (Eq. 7). While there is no previous report on gold-catalyzed Wolff rearrangements, the related Ag catalysis¹⁴ is known, but the mechanistic role of the metal is not well understood. It is commonly assumed, though, a carbene or carbenoid intermediate might be involved. In these clean Wolff rearrangements,²¹ either gold carbenoid **36** or gold carbene **37** is the mostly likely intermediate, and the 1,2-alkyl migration could either be in concert with or follow the expulsion of N₂. This migration was seemingly highly facile as the tethered tertiary amine moiety did not cyclize to the electrophilic carbene/carbenoid center. Since the gold carbene **37** would be highly electrophilic (vide infra), this unexpected preference for the Wolff rearrangement suggests that concerted 1,2-alkyl migration and nitrogen expulsion from carbenoid **36** are more likely.



An alternative approach to access gold carbene of type **37** was then pursued. One of us has recently developed a facile access to putative α-oxo gold carbenes via gold-catalyzed intermolecular oxidations of alkynes using pyridine/quinoline *N*-oxides as oxidants.²² By using

this strategy, intermediate **39**, with its gold carbene moiety similar to that of **37**, was most likely formed from the substrate **38** in the presence of MsOH (2 equiv, Eq. 8). The reaction, however, did not afford azepan-4-one **42** but instead α -chloro ketone **40** and α -methanesulfonyloxy ketone **41**. The formation of **40** is a strong indication of the formation of a highly electrophilic intermediate such as gold carbene **39** as it abstracts a chloride from the solvent, dichloroethane.²³ This chloride abstraction is supported by the observation of 2-chloroethyl mesylate, which accounts for the remaining part of the solvent molecule. When *m*-CBA (1.2 equiv) replaced MsOH, no oxidation of the C-C triple bond occurred. These studies argue against the intermediacy of gold carbene **5** (Scheme 2.1) and are fully consistent with the concerted hetero-retroene reaction predicted computationally.

In conclusion, a hetero-retroene mechanism for the rearrangements of acetylenic amine-N-oxides to piperidinones and azapanones has been identified. The experimental regioselectivities are reproduced by the computations of activation barriers for the hetero-retroene reaction. Experimental studies support the proposed mechanism. This is the first case in which a retroene reaction of a gold-coordinated intermediate has been identified. This type of mechanism may be important in other gold-catalyzed reactions.^{3a,e}

References

¹ Noey, E. L.; Luo, Y.; Zhang, L.; Houk, K. N. *J. Am. Chem. Soc.* **2012**, *134*, 1078.

² (a) Hashmi, A. S. K. *Gold Bull* **2003**, *36*, 3. (b) Hashmi, A. S. K; *Gold Bull* **2004**, *37*, 51. (c) Hoffmann-Roder, A.; Krause, N.; *Org. Biomol. Chem* **2005**, *3*, 387. (d) Ma, S.; Yu S.; Gu, Z. *Angew. Chem., Int. Ed* **2005**, *45*, 200. (e) Hashmi, A. S. K.; *Angew. Chem., Int. Ed* **2005**, *44*, 6990. (f) Furstner, A.; Davies, P. W. *Angew. Chem. Int. Ed.* **2007**, *46*, 3410. (g) Li, Z.; Brouwer, C.; He, C. *Chem. Rev.* **2008**, *108*, 3239. (h) Gorin, D.J.; Sjöerry, B.D.; Toste, F.D. *Chem. Rev.* **2008**, *108*, 3351. (i) Hashmi, S.K.; Rudolph, M. *Chem. Soc. Rev.* **2008**, *37*, 1766. (j) Cuenca, A.

B.; Montserrat, S.; Hossain, K. M.; Mancha, G.; Lledos, A.; Medio-Simon, M.; Ujaque, G.; Asensio, G. *Org. Lett.* **2009**, *11*, 4906. (k) Bandini, M. *Chem. Soc. Rev.* **2011**, *40*, 1358.

³ Xiao, J.; Li, X. *Angew. Chem., Int. Ed.* **2011**, *50*, 7226.

⁴ (a) Shapiro, N. D.; Toste, F. D. *J. Am. Chem. Soc.* **2007**, *129*, 4160. (b) Li, G.; Zhang, L. *Angew. Chem. Int. Ed.* **2007**, *46*, 5156.

⁵ (a) Yeom, H. S.; Lee, J. E.; Shin, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 7040. (b) Yeom, H.-S.; Lee, Y.; Jeong, J.; So, E.; Hwang, S.; Lee, J.-E.; Lee, S. S.; Shin, S. *Angew. Chem.* **2010**, *122*, 1655.

⁶ Lin, G. Y.; Li, C. W.; Hung, S. H.; Liu, R.-S. *Org. Lett.* **2008**, *10*, 5059. (b) Hashmi, A. S.; Bührle, M.; Salathé, R.; Bats, J. *Adv. Synth. Catal.* **2008**, *350*, 2059.

⁷ Jadhav, A. M.; Bhunia, S.; Liao, H.-Y.; Liu, R.-S. *J. Am. Chem. Soc.* **2011**, *133*, 1769.

⁸ (a) Cui, L.; Peng, Y.; Zhang, L. *J. Am. Chem. Soc.* **2009**, *131*, 8394. (b) Cui, L.; Ye, L.; Zhang, L. *Chem. Comm.* **2010**, *46*, 3351.

⁹ (a) Ye, L.; He, W.; Zhang, L. *Angew. Chem., Int. Ed.* **2011**, *50*, 3236. (b) He, W.; Li, C.; Zhang, L. *J. Am. Chem. Soc.* **2011**, *133*, 8482. (c) Ye, L.; He, W.; Zhang, L. *J. Am. Chem. Soc.* **2010**, *132*, 8550. (d) Ye, L.; Cui, L.; Zhang, G.; Zhang, L. *J. Am. Chem. Soc.* **2010**, *132*, 3258. (e) Lu, B.; Li, C.; Zhang, L. *J. Am. Chem. Soc.* **2010**, *132*, 14070.

¹⁰ (a) Cuenca, A. B.; Montserrat, S.; Hossain, K. M.; Mancha, G.; Lledos, A.; Medio-Simon, M.; Ujaque, G.; Asensio, G. *Org. Lett.* **2009**, *11*, 4906. (b) Li, C.-W.; Pati, K.; Lin, G.-Y.; Sohel, S. M. A.; Hung, H.-H.; Liu, R.-S. *Angew. Chem., Int. Ed.* **2010**, *49*, 9891.

¹¹ Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

¹² DFTD3 V2.0 Rev 1, Grimme, S. University Muenster, **2010** (a) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104. (b) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456.

¹³ (a) Cheong, P. H.-Y.; Morganelli, P.; Luzung, M. R.; Houk, K. N. Toste; F. D. *J. Am. Chem. Soc.* **2008**, *130*, 4517. (b) Xia, Y.; Dudnik, A. S.; Gevorgyan V.; Li, Y. *J. Am. Chem. Soc.* **2008**, *130*, 6940. (c) Benitez, D.; Tkatchouk, E.; Gonzales, A. Z.; Goddard, W. A.; Toste, F. D. *Org. Lett.* **2009**, *11*, 4798. (d) M.; Hashmi, A. S. K.; Pernpointner, M. *ChemCatChem* **2010**, *2*, 1226. (e) Touil, M.; Bechem, B.; Hashmi, A. S. K.; Engels, B.; Omary, M. A.; Rabaâ, H. *J. Mol. Struct-Theochem.* **2010**, *957*, 21. (f) Noey, E. N.; Wang, X.; Houk K. N. *J. Org. Chem.* **2011**, *76*, 3477.

¹⁴ Haeberlen, O. D.; Roesch, N. *J. Phys. Chem.* **1993**, *97*, 4970.

¹⁵ Faza, O. N.; Rodríguez, R. Á.; López, S. C. *Theor. Chem. Acc.* **2011**, *128*, 647. For the comparison of fully relativistic DHF-SCF, DFT/B3LYP and GF, see: Pernpointner M.; Hashmi, A. S. K. *J. Chem. Theory Computation* **2009**, *5*, 2717.

¹⁶ Hashmi, A. S. K. *Angew. Chem. Int. Ed.* **2010**, *49*, 5232.

¹⁷ Several optimizations were run with the forming C-O bond distances locked at distances between 1.9 and 2.9 Å. The negative force constant in these outputs was appropriate for the desired transition structure. However, full optimizations from these outputs did not give the desired transition structure. A scan of the C-O bond distance from 1.66 to 3.34 Å showed a gradual increase in energy with the maximum at 3.22 Å. A TS optimization of this point did not yield a transition structure. The energy of this structure is 19.0 kcal/mol relative to **23**.

¹⁸ (a) Bertrand, G. Singlet Carbenes. In *Reactive Intermediate Chemistry*; Moss, R. A.; Platz, M. S.; Jones, M. Jr. Ed. Wiley: New Jersey, 2004; p278. (b) Termath, V.; Tozer, D. J.; Handy, N. C. *Chem. Phys. Lett.* **1994**, *228*, 239. (c) Maier, G.; Reisenauer, H. P.; Cibulka, M. *Angew. Chem.* **1999**, *111*, 110; *Angew. Chem. Int. Ed.* **1999**, *38*, 105. (d) Scott, A. P.; Platz, M. S.; Radom, L.; *J. Am. Chem. Soc.* **2001**, *123*, 6069. (e) Kirmse, W. *Eur. J. Org. Chem.* **2002**, *14*, 2193.

¹⁹ Benitez, D.; Shapiro, N. D.; Tkatchouk, E.; Wang, Y.; Goddard, W. A.; Toste, F. D. *Nature Chem.* **2009**, *1*, 482.

²⁰ West, F. G.; Naidu, B. N.; Tester, R. W. *J. Org. Chem.* **1994**, *59*, 6892.

²¹ Kirmse, W. *Eur. J. Org. Chem.* **2002**, 2002, 2193.

²² (a) Ye, L.; He, W.; Zhang, L. *Angew. Chem., Int. Ed.* **2011**, *50*, 3236; (b) He, W.; Li, C.; Zhang, L. *J. Am. Chem. Soc.* **2011**, *133*, 8482; (c) Ye, L.; He, W.; Zhang, L. *J. Am. Chem. Soc.* **2010**, *132*, 8550; (d) Ye, L.; Cui, L.; Zhang, G.; Zhang, L. *J. Am. Chem. Soc.* **2010**, *132*, 3258; (e) Lu, B.; Li, C.; Zhang, L. *J. Am. Chem. Soc.* **2010**, *132*, 14070.

²³ This type of chloride abstraction is previously known by Rh carbenoids. For reference, see: Pirrung, M. C.; Zhang, J.; Lackey, K.; Sternbach, D. D.; Brown, F. *J. Org. Chem.* **1995**, *60*, 2112.

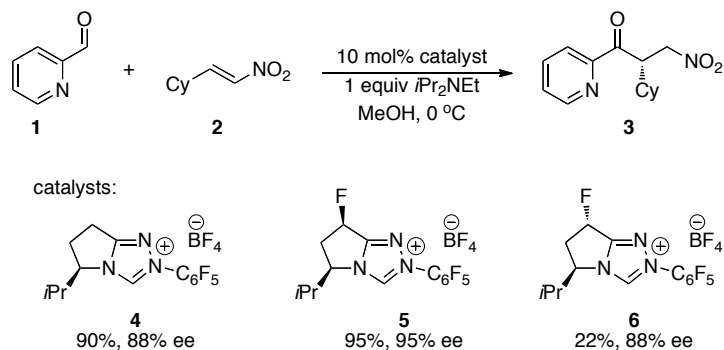
3. Quantum Mechanical Investigation of the Effect of Catalyst Fluorination in the Intermolecular Asymmetric Stetter Reaction

The asymmetric intermolecular Stetter reaction was investigated using the B3LYP and M06-2X functionals. Fluorination of a triazolium bicyclic catalyst had been found to significantly influence reaction yields and enantiomeric ratios. Computations indicate that the improved reactivity of the fluorinated catalyst is due to better electrostatic interactions between the nitroalkene and catalyst. Computational investigations of preferred conformations of the ground state catalyst and acyl anion equivalent, and the transition structures leading to both enantiomers of the products, are reported.¹

Introduction

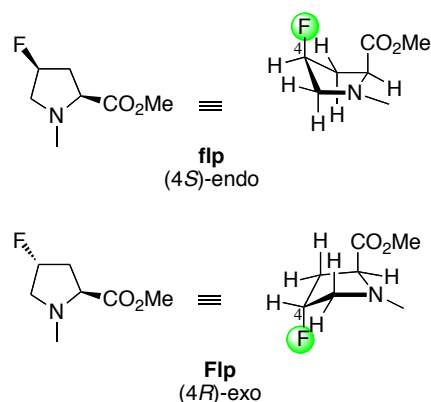
The Stetter reaction inverts the normal mode of reactivity of aldehydes by generating acyl-anion equivalents capable of reacting with Michael acceptors.² Following seminal work by Enders,³ the Rovis group and others have shown that the reaction can be organocatalyzed in excellent yields and stereoselectivities.^{4,5} In a recent study of the reaction of picolinaldehyde **1** and β -cyclohexyl nitroalkene **2**, DiRocco *et al.* found that fluorination of bicyclic triazolium catalysts has a significant effect on the reaction yield and enantiomeric ratios (Scheme 3.1).⁶

Scheme 3.1. Effect of fluorination of bicyclic triazolium catalysts the Stetter reaction yield and enantiomeric ratios.



The isopropyl-substituted triazolium catalyst **4** gave promising yields and enantioselectivity.^{2b} Introduction of a fluorine *cis* to the isopropyl group of **4** resulted in a catalyst (**5**) that gave improved yield and enantioselectivity results, while placement of a fluorine group *trans* to the isopropyl group of **4** gave **6**, which displayed a significant decrease in yield and no change in enantioselectivity.⁵ Rovis proposed that the varying reactivities were controlled by stereoelectronic effects suggested by Raines, who has shown that fluorine has a stereoelectronic effect on the preferred conformation of proline and the stabilities of fluorinated collagen models.⁷ Due to the gauche effect,⁸ fluorine-substituted prolines prefer an axial-fluoride conformation. In other words, (4*S*)-isomer **flp** (Scheme 3.2) prefers the *endo* conformation, while (4*R*) isomer **Flp** prefers the *exo* conformation.

Scheme 3.2. Proline conformations reported by Raines.



We reasoned that the same stereoelectronic effect should influence the preferred conformations of catalysts **4–6** (Figure 3.1). For catalyst **4**, the *endo* conformation (**4-endo**) is expected to be favored due to 1,3-diaxial strain between the isopropyl group and a hydrogen in the *exo* conformation (**4-exo**). Catalyst **5** is expected to prefer the *exo* conformation (**5-exo**) because of the gauche effect of fluorine. Catalyst **6** is expected to favor the *endo* conformation (**6-endo**) both because of 1,3-diaxial strain in **6-exo**, and because of the favorable gauche effect in **6-endo**. Crystal structures of **4–6** were obtained by Rovis' group; all three structures match the preferred conformations described by Rovis in Figure 3.1.⁵

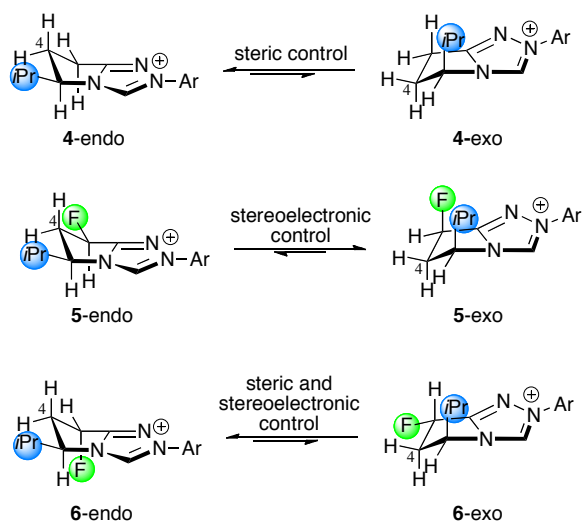


Figure 3.1. Preferred conformations of triazolium catalysts **4–6**.

Computational Methods

We undertook computational studies to explore this hypothesis and to determine the effect of conformation on selectivity.⁹ All geometries were optimized in the gas phase using B3LYP¹⁰/6-31G(d)¹¹ as implemented in the Gaussian 09¹² suite of programs. All stationary points were verified as minima or first-order saddle points by vibrational frequency analysis. Single point calculations on the transition structures were performed using M06-2X¹³/6-31+G(d)¹⁴ as implemented in Gaussian 09. Solvation free energies were calculated using the CPCM model¹⁵ (UAKS radii, methanol, $\epsilon = 32.6$). DFT calculations have provided insight into the Stetter reaction and related reactions.^{8,16} Dudding and Houk predicted the stereochemistry of asymmetric benzoin reactions although with slightly overestimated *ee*.¹⁵ While B3LYP has come under criticism for inaccuracies in reaction energy for reactions involving the transformation from π to σ bonds, a recent review validates these methods for prediction of transition state energies.¹⁷ Furthermore, M06-2X has proven to be capable of reproducing selectivities and thermochemistries when dispersion energies are important. M06-2X results are given here in addition to B3LYP.^{13b}

Results and Discussion

Figure 3.2 shows the computed *exo* and *endo* conformers of the catalysts, **4–6**. The B3LYP/6-31G(d) lowest energy conformers are the same as those found in the crystal structures, although the preference for **4**-endo is only 0.2 kcal/mol (Figure 3.2). In addition to the gauche effect, the relative stabilities of the preferred conformers of **5** and **6** may be due to the distance between the pyrrolidine fluorine and the triazole N1 (labeled in Figure 3.2).¹⁸ The pyrrolidine fluorine of **5** is farther from the triazole N1 in the *exo* (dihedral F-C3-C2-N1 = 73°; F–N1 = 3.20

Å) versus *endo* (dihedral F-C3-C2-N1 = 50°; F-N1 = 3.12 Å) conformation. Similarly, the pyrrolidine fluorine of **6** is farther from the triazole N1 in the *endo* (dihedral F-C3-C2-N1 = 77°; F-N1 = 3.21 Å) versus *exo* (dihedral F-C3-C2-N1 = 51°; F-N1 = 3.11 Å) conformation.

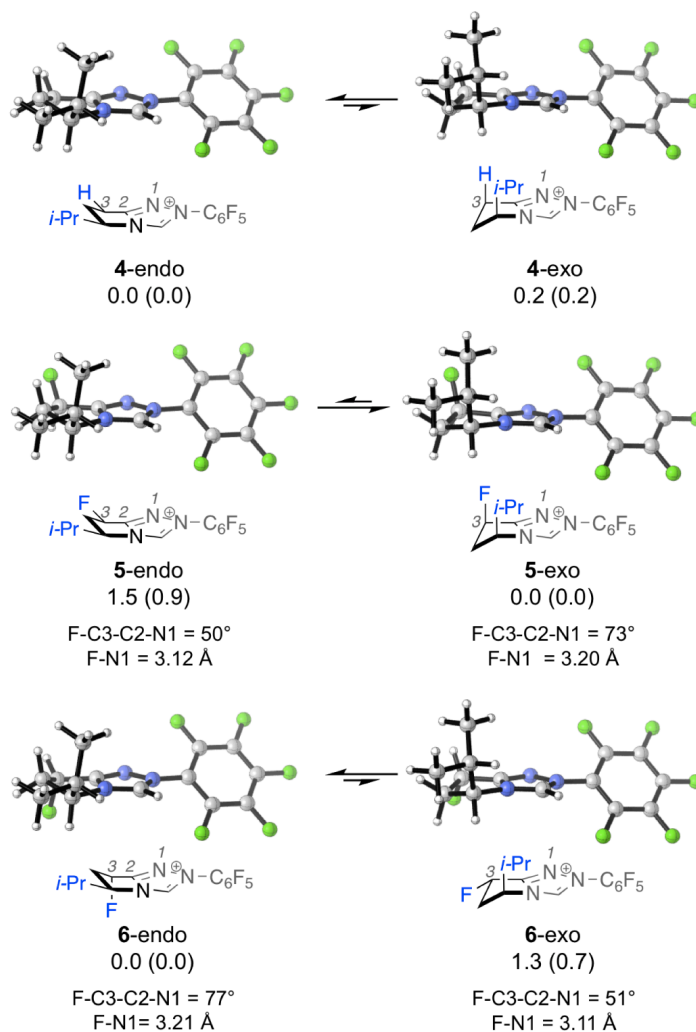


Figure 3.2 B3LYP/6-31G(d) gas phase conformations and relative enthalpies of triazolium catalysts **4–6**. B3LYP/6-31+G(d) relative enthalpies with solvation corrections (CPCM methanol) are in parentheses.

The conformations of the intermediates (**7–9**) formed by reaction of catalysts **4–6** and aldehyde **1** (Figure 3.3) were investigated next. Interestingly, no minima for the *endo* conformers could be located. Optimizations beginning from *endo* conformations consistently

converge to the *exo* conformations. This is likely due to unfavorable interactions between the isopropyl group and enol moieties in the *endo* conformations. We therefore calculated the relative stabilities of related intermediates in which the isopropyl group was replaced by hydrogen (**10–12**, Figure 3.4). The fluorine has a small effect on conformation, as reflected in the small energy difference between *exo* and *endo* in **11–12**. Recent studies support this finding, showing that different conformations about the N-C-C-F dihedral angle in *E*- α -fluoro-imines have small energy differences.¹⁹ The *endo* conformations of all three catalysts were located and are low in energy, supporting the hypothesis that an unfavorable interaction must exist between the isopropyl and enol moieties of **7–9**-endo.

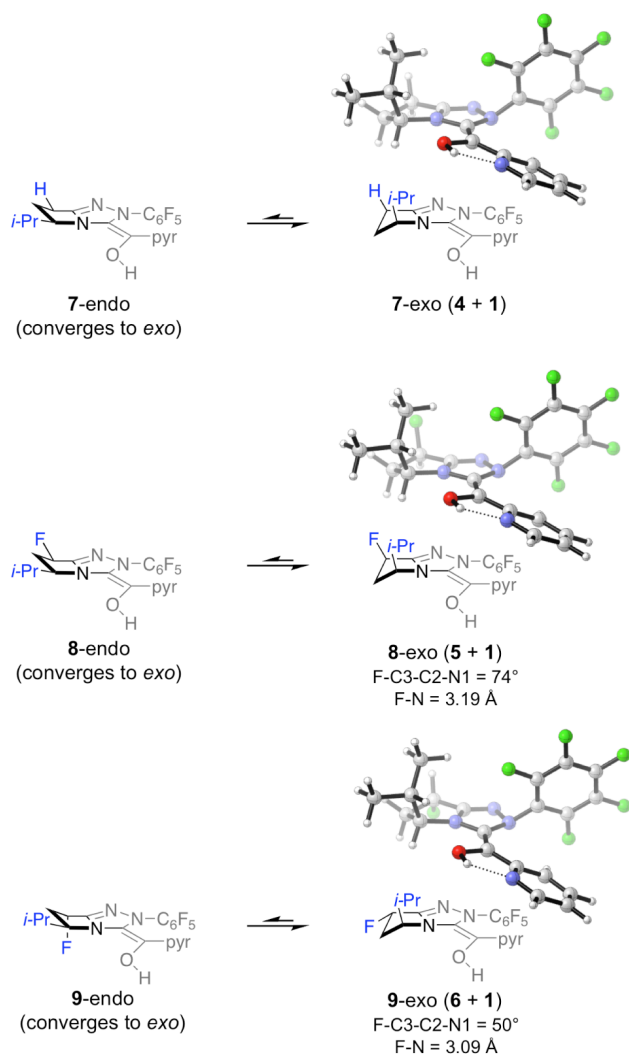


Figure 3.3. B3LYP/6-31G(d) acyl anion equivalents **7–9**.

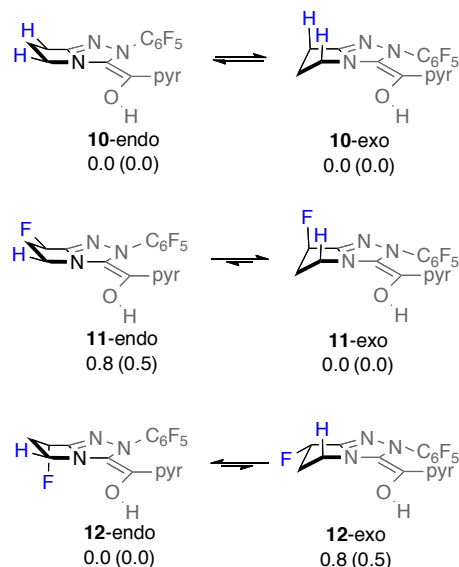


Figure 3.4. B3LYP/6-31G(d) relative gas phase enthalpies of acyl anion equivalents **10–12**. B3LYP/6-31+G(d) relative enthalpies with solvation corrections (CPCM methanol) are in parentheses.

To quantify the extent of steric strain in **7–9-endo**, isopropyl groups were placed on **10–12**, the *endo* geometries of the bicyclic triazoline were frozen, and the resulting structures were optimized (**7–9-endo'**, Figure 3.5). Intermediates **7–9-endo'** are all approximately 5–7 kcal·mol⁻¹ higher in energy than their corresponding *exo* conformers. The instability can be attributed to (1) H-H steric clash between the isopropyl group and the pyrrolidine ring, and (2) disfavored interaction between the isopropyl hydrogen and enol oxygen.²⁰ Thus, the *exo* conformations of **7–9** are strongly favored regardless of the catalyst used. *Endo* conformations **7–9-endo'** are too high in energy to be involved in the reaction. The experimentally observed stereoselectivities are attributed not to the conformation of the catalyst, but rather to (1) a steric effect, (2) a stereoelectronic effect, or (3) a combination of both, of the fluorine atom in the transition state.

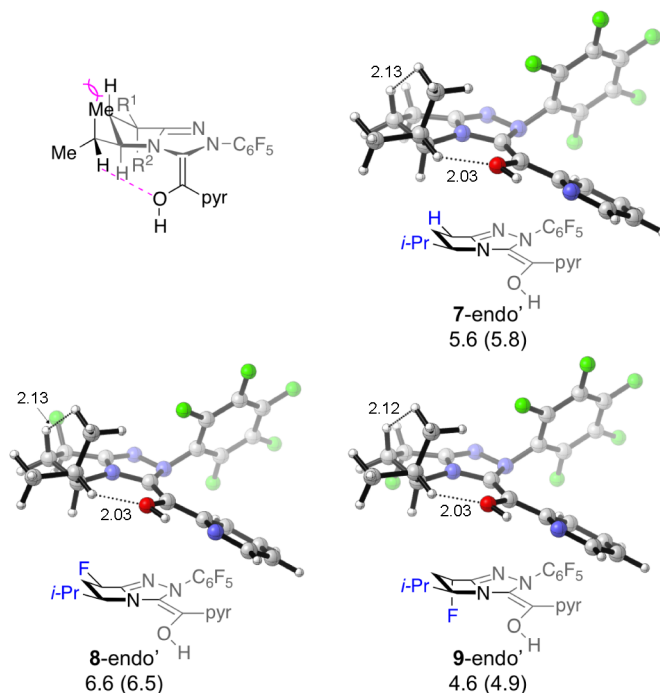


Figure 3.5. Enol intermediates **7–9-endo'** and relative enthalpies with respect to **7–9-exo**. B3LYP/6-31+G(d) relative enthalpies with solvation corrections (CPCM methanol) are in parentheses.

Compounds **7–9** are cyclic hydrazones, and the trisubstituted nitrogen in such compounds is pyramidal.⁸ The conformation of **7–9-exo** with the nitrogen pyramidalized so that C₆F₅ is in an “up” position is more stable than the “down” position by no more than 0.3 kcal/mol (Figure 3.3). The inversion barrier for the nitrogen bearing the C₆F₅ was calculated to be 5.1 kcal/mol for **8-exo**. The *Z*-isomers (not shown) are 2–3 kcal/mol higher in energy than the corresponding *E* isomers.

B3LYP gives a planar geometry of the nitrogen connected to the aromatic ring in the triazolium catalysts, **4–6**, and a tetrahedral geometry at this nitrogen in the acyl anion equivalents, **7–9**. It was brought into question whether the nitrogen in the acyl anion equivalents is truly tetrahedral or if it is analogous to the gold-coordinated triazole carbene, **13**, which is planar (Figure 3.6). In order to verify that our theoretical methods give reasonable geometries,

we compared the crystal structure of the gold-coordinated triazole carbene, **13**, to the geometry of that given by B3LYP/6-31G(d) (/LANL2DZ for gold). Figure 3.6 shows that there is excellent agreement between these two geometries. It is clear that the nitrogen connected to the aromatic ring is planar in both structures.

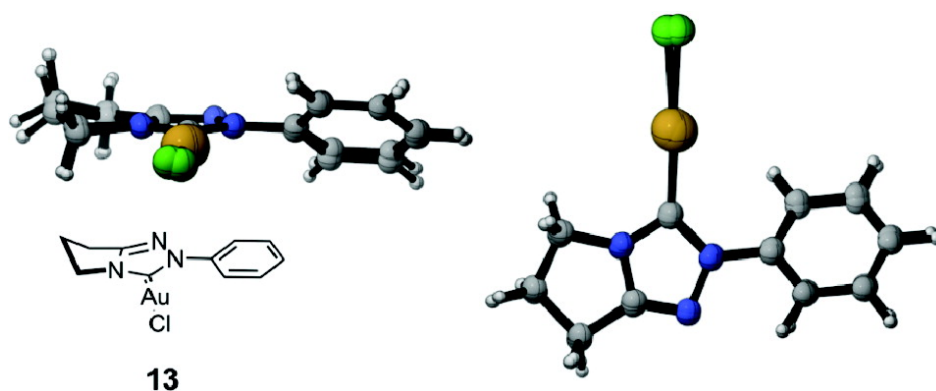


Figure 3.6. Overlay of crystal structure and calculated structure (B3LYP/6-31G(d) for C, N, Cl and H, /LANL2DZ for gold) of **13**.

Transition structures for Michael addition of enol intermediates **7–9** to β -cyclohexyl nitroalkene, **2**, were next located. The pentafluorophenyl group was modeled by a phenyl group. All lowest energy transition structures for addition to the *Si*-face of the nitroalkene (**TS1**, **TS3**, and **TS5**, favored, Figure 3.7) exhibit a stabilizing interaction between the hydrogen of the hydroxyl group and the carbon α -to the nitro group, in a transition state reminiscent of the reverse Cope elimination²¹ as previously proposed by Rovis.²² Similarly, the lowest energy transition structures for addition to the *Re*-face of the nitroalkene (disfavored, **TS2**, **TS4**, and **TS6**) show a stabilizing interaction between the hydroxy group and an oxygen of the nitro group. Transition structures for **7–9** are shown in Figures 3.8–3.10, respectively. Transition structures **TS1–TS6** are all in agreement with Seebach's topological rule, which describes a preferred

gauche arrangement of the double bonds of donors and acceptors in Michael addition transition states, including the addition of enamines to nitroalkenes (Figure 3.11).²³

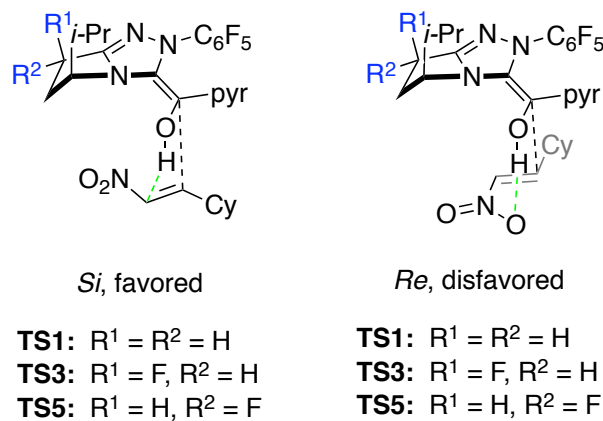


Figure 3.7. *Si*- and *Re*-face transition structures.

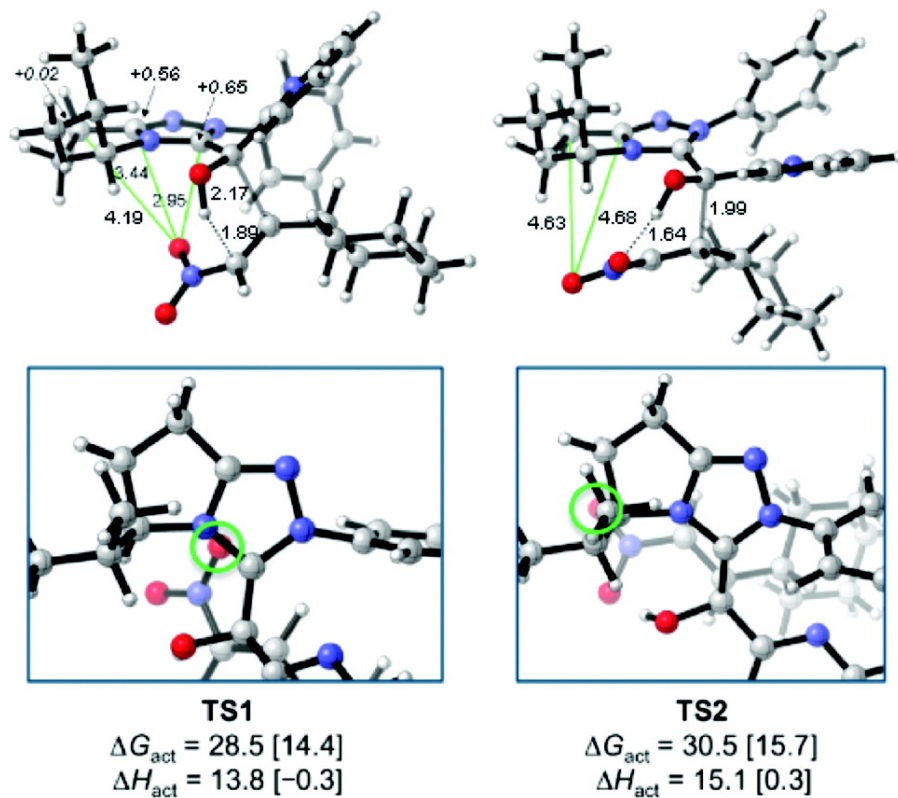


Figure 3.8. B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger for **7**. M06-2X/6-31+G(d) energies in brackets.

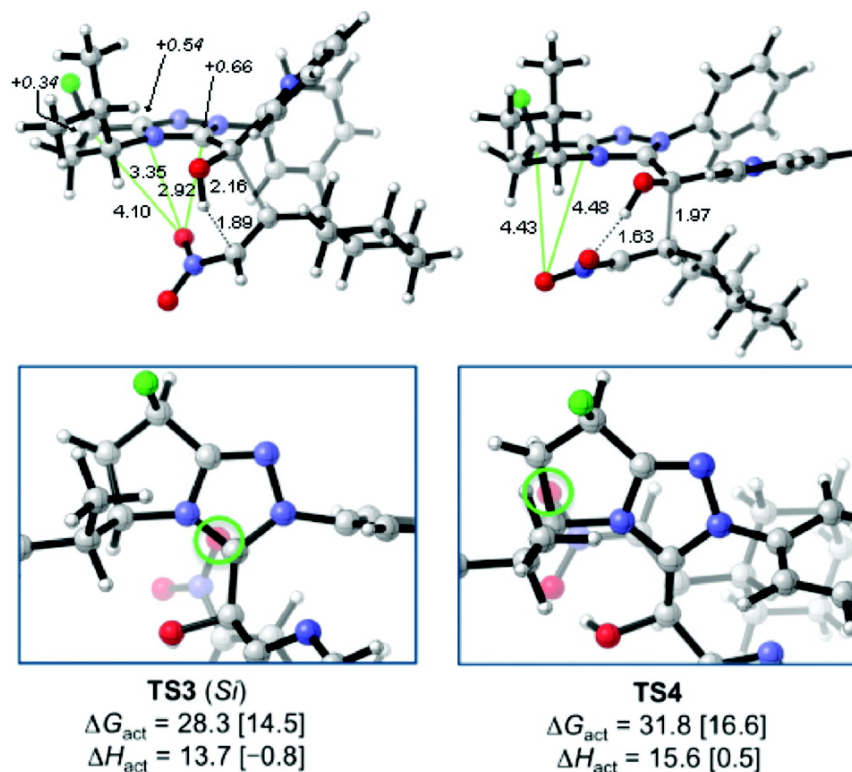


Figure 3.9. B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger for **8**. M06-2X/6-31+G(d) energies in brackets.

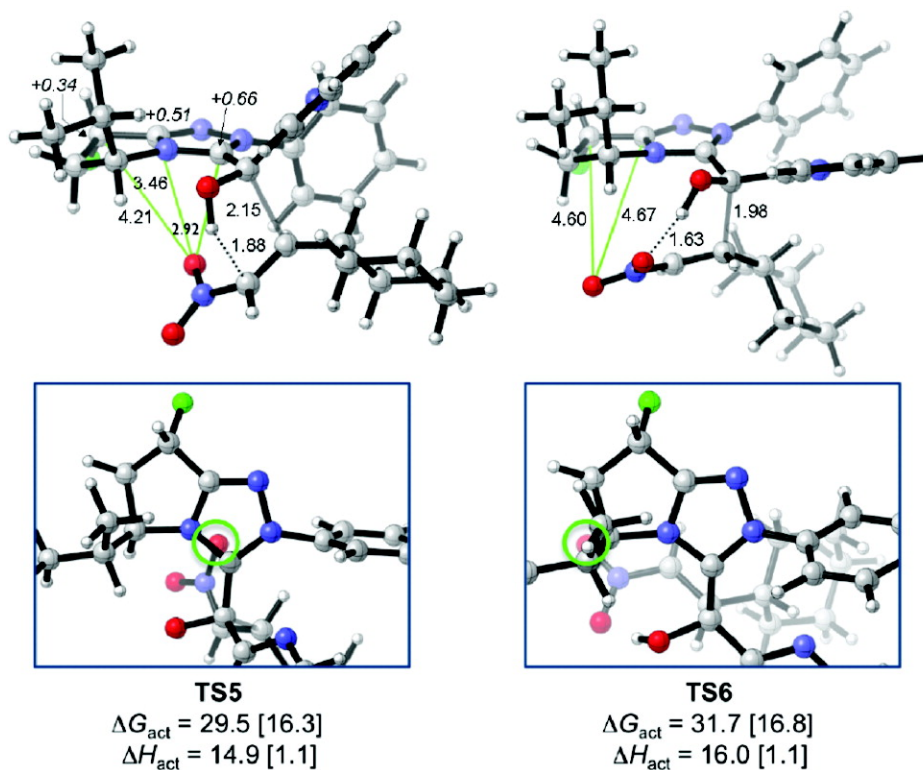


Figure 3.10. B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger for **9**. M06-2X/6-31+G(d) energies in brackets.

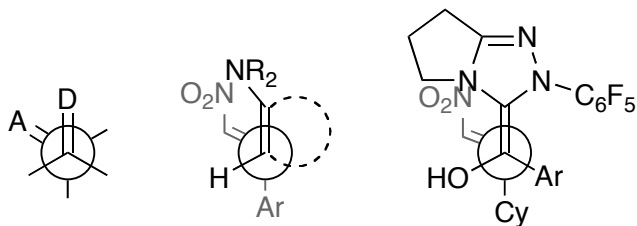


Figure 3.11. Seebach's topological rule applied to the Stetter reaction.

The experimental and calculated enantiomeric ratios are summarized in Table 3.1. The M06-2X activation barriers are approximately 15 kcal·mol⁻¹ higher in energy than the B3LYP barriers. Furthermore, the barriers for catalyst **6**, which experimentally reacted in only 22% yield, are approximately 1 kcal·mol⁻¹ higher in energy than the barriers for **4** and **5**. An examination of the transition structures led to our hypothesis for the experimentally observed enantioselectivities. We propose that **5** is more effective than **4** and **6** because the electronegative fluorine promotes a stronger attractive interaction between the developing cationic character of the catalyst bicycle and the developing anionic character of the nitroalkene in the transition state. The distances between an oxygen of the nitroalkene and the positively charged carbons of the catalyst are shorter in **TS3** (4.10 Å and 3.35 Å) compared to **TS1** (4.19 Å and 3.44 Å) and **TS5** (4.21 Å and 3.46 Å). The same holds true for the disfavored transition states. The O-C3 and O-C2 distances in **TS4** are 4.43 Å and 4.48 Å, compared to 4.63 Å and 4.68 Å in **TS2** and 4.60 Å and 4.67 Å in **TS6**. Catalyst **6** is not as effective as **5** because the fluorine projects toward the same face as the nitroalkene, and a repulsive interaction makes this arrangement less favorable. Furthermore, **9-exo** is 1.0 kcal·mol⁻¹ less stable than **8-exo** due to the disfavored F-N1 distances described earlier. This difference carries over to the transition states (Table 3.2).

Table 3.1. Summary of calculated and experimental enantioselectivities^a

entry	catalyst	experimental		B3LYP ^b gas phase		M06-2X ^c methanol		B3LYP ^c methanol	
		e.r.	$\Delta\Delta G^\ddagger$	e.r.	$\Delta\Delta G^\ddagger$	e.r.	$\Delta\Delta G^\ddagger$	e.r.	$\Delta\Delta G^\ddagger$
1	4	94:6	1.6	97:3	2.1	91:9	1.3	93:7	1.6
2	5	98:2	2.2	99:1	2.9	98:2	2.3	93:7	1.6
3	6	94:6	1.6	98:2	2.2	86:14	1.1	86:14	1.1

^aCalculated enantioselectivities include 3–5 low-energy conformations. The free energy difference here is a weighted average of the energies of these conformations. The $\Delta\Delta G_{\text{act}}$ in Figures 3.8–3.10 involve only the lowest energy transition structure. ^b6-31G(d) basis set. ^c6-31+G(d) basis set. Solvation single point energies on B3LYP/6-31G(d) geometries.

Table 3.2. F–N1 distances in **TS3**–**TS6**.

entry	TS	F–N1 (Å)
1	TS3	3.24
2	TS4	3.16
3	TS5	3.07
4	TS6	3.13

All *Re*-face transition structures are higher in energy than the *Si*-face transition structures because of poorer stabilization of the nitroalkene by the catalyst in the latter. The nitro group of the disfavored transition structures is rotated away from the catalyst, causing the distances between an oxygen of the nitro group and the partially positive carbons of the catalyst to be longer in **TS2**, **TS4**, and **TS6** compared to **TS1**, **TS3**, and **TS5** (Figures 3.8–3.10).

Triazolium catalysts, **4–6**, can competitively catalyze the benzoin condensation over the Stetter reaction. In the case reported here, a 1.5:1 ratio of nitroalkene to aldehyde was sufficient to achieve good yields.⁵ The benzoin condensation of enol intermediate, **8**, with aldehyde **1** was found to have a lower barrier than the Stetter reaction of **8** with β -cyclohexyl nitroalkene, **2** ($\Delta\Delta G^\ddagger = 2.9$ kcal/mol (B3LYP) or 0.7 kcal/mol (M06-2X)). These calculations appear to overestimate the preference for the benzoin condensation. Figure 3.12 shows the transition state for the Stetter reaction, **TS3** (*Si*), and benzoin condensation, **TS7** (*Si*).

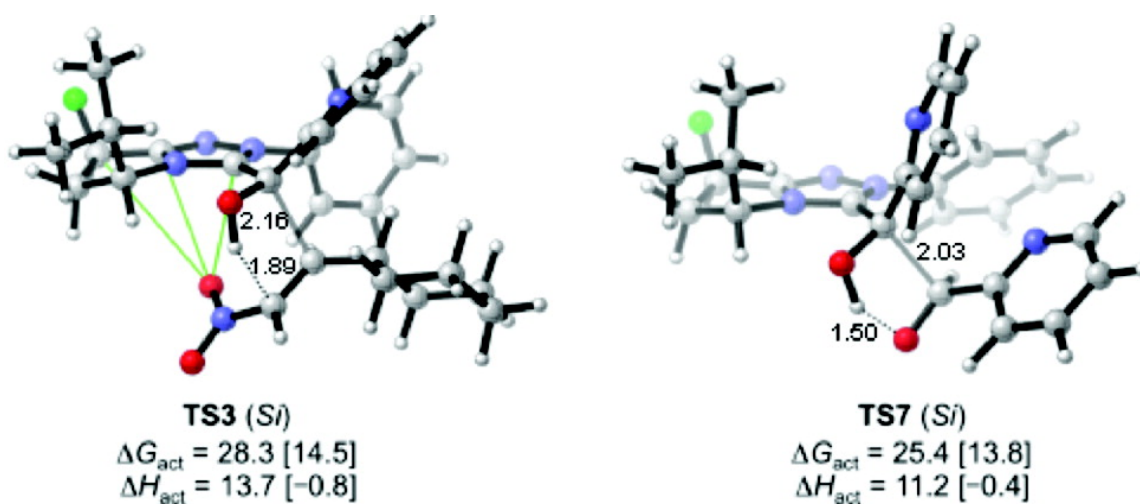


Figure 3.12. Comparison of the transition states for the Stetter reaction, **TS3** (*Si*), and benzoin condensation, **TS7**. B3LYP/6-31G(d) gas phase transition structures and ΔG^\ddagger . M06-2X/6-31+G(d) energies in brackets.

In conclusion, crystal structure and DFT analyses show that the preferred conformation of fluorinated triazolium catalysts **5–6** is controlled by stereoelectronic effects. Formation of the acyl anion equivalent results exclusively in the *exo* conformation of the pyrrolidine ring, regardless of the catalyst used. The favored transition structures are stabilized by an interaction between the developing negative charge in the nitro group of the alkene and the developing positive charge on the catalyst. The fluorinated catalyst, **5**, is more effective than **4** due to the

electronegativity of **5**. Catalyst **6** is less effective than **5** because of unfavorable electrostatic interactions between the fluorine and both the nitroalkene and a triazolium nitrogen. The synthesis of bridged bicyclic catalysts containing methylene groups and oxygen atoms in lieu of fluorine is currently underway. Testing these catalysts should give further insight into the effect of electronegative groups on reactivity and stereoselectivity.

References

¹ Um, J. M.; DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. *J. Am. Chem. Soc.* **2011**, *133*, 11249.

² (a) Stetter, H.; Schreckenberger, M. *Angew. Chem., Int. Ed. Engl.* **1973**, *12*, 81. (b) Stetter, H. *Angew. Chem., Int. Ed. Engl.* **1976**, *15*, 639. (c) Stetter, H.; Kuhlmann, H. *Org. React.* **1991**, *40*, 407.

³ (a) Enders, D.; Breuer, K.; Runsink, J.; Teles, J. H. *Helv. Chim. Acta.* **1996**, *79*, 1899. (b) Enders, D.; Breuer, K. *Comprehensive Asymmetric Catalysis*; Springer: Berlin, 1999; 1093. (c) Enders, D.; Balensiefer, T. *Acc. Chem. Res.* **2004**, *37*, 534. (d) Enders, D.; Han, J.; Henseler, A. *Chem. Commun.* **2008**, 3989. (e) Enders, D.; Han, J. *Synthesis* **2008**, 3864.

⁴ For an account, see: Read de Alaniz, J.; Rovis, T. *Synlett*, **2009**, 1189. For more recent work, see: (a) Liu, Q.; Perreault, S.; Rovis, T. *J. Am. Chem. Soc.* **2008**, *130*, 14066. (b) Liu, Q.; Rovis, T. *Org. Lett.* **2009**, *11*, 2856. (c) Jousseume, T.; Wurz, N. E.; Glorius, F. *Angew. Chem. Int. Ed.* **2011**, *50*, 1410. (d) Dresen, C.; Richter, M.; Pohl, M.; Lüdeke, S.; Müller, M. *Angew. Chem. Int. Ed.* **2010**, *49*, 6600. (e) Moore, J. L.; Silvestri, A. P.; Read de Alaniz, J.; DiRocco, D. A.; Rovis, T. *Org. Lett.* **2011**, *13*, 1742. (f) Piel, I.; Steinmetz, K.; Hirano, K.; Fröhlich, R. Grimme, S.; Glorius, F. *Angew. Chem. Int. Ed.* **2011**, *50*, 4983. (g) Hirano, K.; Biju, A. T.; Piel, I.; Glorius, F. *J. Am. Chem. Soc.* **2009**, *131*, 14190.

⁵ For reviews on carbene catalysis, see: (a) Christmann, M. *Angew. Chem. Int. Ed.* **2005**, *44*, 2632. (b) Enders, D.; Niemeier, O.; Henseler, A. *Chem. Rev.* **2007**, *107*, 5606. (c) Moore, J. L.; Rovis, T. *Top. Curr. Chem.* **2009**, *290*, 77.

⁶ DiRocco, D. A.; Oberg, K. M.; Dalton, D. M.; Rovis, T. *J. Am. Chem. Soc.* **2009**, *131*, 10872.

⁷ (a) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 777. (b) Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 9262. (c) Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2005**, *127*, 15923.

⁸ Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102.

⁹ For a recent DFT study on the mechanism of the Stetter reaction, see: Hawkes, K. J.; Yates, B. *Eur. J. Org. Chem.* **2008**, *33*, 5563.

¹⁰ (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5468. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372. (c) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *98*, 785.

¹¹ (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (c) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

¹² Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

¹³ (a) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215. (b) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

¹⁴ (a) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294. (b) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265. (c) Latajka, Z.; Scheiner, S. *Chem. Phys. Lett.* **1984**, *105*, 435.

¹⁵ (a) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995. (b) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

¹⁶ Dudding, T.; Houk, K. N. *PNAS*, **2004**, *101*, 5770.

¹⁷ Simon, L.; Goodman, J. M. *Org. Biomol. Chem.* **2011**, *9*, 689.

¹⁸ Sum of fluorine-nitrogen van der Waals radii = 3.02 Å

¹⁹ Sparr, C.; Salamanova, E.; Schweizer, W.B.; Senn, H.M.; Gilmour, R. *Chem. Eur. J.* **2011**, *In press*.

²⁰ Sum of hydrogen-hydrogen van der Waals radii = 2.40 Å; sum of oxygen-hydrogen van der Waals radii = 2.72 Å

²¹ (a) Niu, D.; Zhao, K. *J. Am. Chem. Soc.* **1999**, *121*, 2456. (b) Sibi, M. P.; Liu, M. *Org. Lett.* **2000**, *21*, 3393. (c) Sibi, M. P.; Prabakaran, N.; Ghorpade, S. G.; Jasperse, C. P. *J. Am. Chem. Soc.* **2003**, *125*, 11796.

²² The intramolecular asymmetric Stetter has been demonstrated to proceed in high diastereoselectivity when using trisubstituted Michael acceptors. See: de Alaniz, J. R.; Rovis, T. *J. Am. Chem. Soc.* **2005**, *127*, 6284.

²³ Seebach, D.; Golinski, J. *Helv. Chim. Acta* **1981**, *64*, 1413.

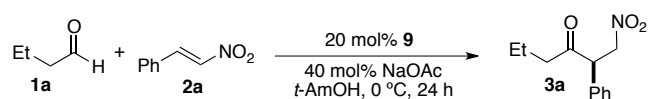
4. Catalytic Asymmetric Intermolecular Stetter Reactions of Enolizable Aldehydes with Nitrostyrenes: Computational Study Provides Insight into the Success of the Catalyst

Over the past decade, *N*-heterocyclic carbenes (NHCs) have been used as catalysts in a variety of C-C bond forming reactions.^{1,2} Our groups have been interested in the development of chiral NHC's as catalysts for the asymmetric *intra*-molecular Stetter reaction^{3,4} and more recently the *inter*-molecular variant.^{5,6} We recently reported that hetaryl aldehydes and enals react efficiently with nitroalkenes in the Stetter reaction, leading to β -nitro ketones with high enantioselectivity.^{4d} Crucial to the success of this method was the development of a fluorinated triazolium salt pre-catalyst that provides significantly enhanced enantioselectivity over *des*-fluoro analogues.⁷ Although this new catalyst system greatly expands the scope of this method, these conditions are not amenable to the use of unactivated aliphatic aldehydes. Due to their lower electrophilicity than aryl aldehydes, aliphatic aldehydes have rarely been used successfully in the asymmetric *inter*-molecular Stetter reaction.^{8,9}

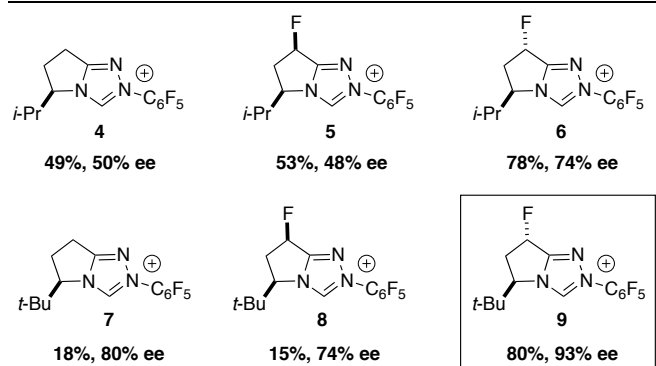
Our initial attempts at rectifying this problem began by evaluating more reactive Michael acceptors such as β -nitrostyrenes. The unstable nature of the reaction components mandated milder reaction conditions,¹⁰ and a brief screen revealed tertiary alcohol solvents and weak inorganic bases as being optimal. Under these conditions, pre-catalyst **5**, which previously demonstrated high reactivity and enantioselectivity for hetaryl aldehydes and enals, affords only modest yield (53%) with low enantioselectivity (48%) (Table 4.1). Interestingly, this reaction affords the opposite major enantiomer to that observed in our previous work using the same pre-catalyst with aliphatic nitroalkenes.^{4d} Surprisingly, *trans*-fluorinated pre-catalyst **6** provides substantial increases in both yield (78%) and enantioselectivity (74%). In order to further increase selectivity we evaluated the more sterically demanding pre-catalyst **7**, derived from *t*-

leucine. This pre-catalyst displays low reactivity compared to the valine derived pre-catalysts (**4-6**), but with greatly increased enantioselectivity (80%). Further evaluation of this scaffold shows the same trends in reactivity and selectivity to that of the valine derived series; *trans*-fluorinated pre-catalyst **9** provides drastically better selectivity (93% *ee*) than both *cis*-fluoro (74%) and des-fluoro (80%) catalysts.

Table 4.1. Catalyst and optimization studies.^a



Entry	Deviation from standard conditions	Yield (%) ^b	ee(%) ^c
1	none	80	93
2	<i>i</i> -PrOH as solvent	42	88
3	THF as solvent	0	nd
4	PhMe as solvent	0	nd
5	<i>i</i> -Pr ₂ NEt instead of NaOAc	trace	nd
6	1.0 equiv NaOAc	78	93
7	23 °C	42	90



^a Reactions conducted with 1.5 equiv **1a** and 1.0 equiv **2a**. ^b Isolated yield after chromatography. ^c Enantiomeric excess determined by HPLC analysis on a chiral stationary phase. BF_4^- counterions omitted for clarity.

Under optimized conditions the scope of this transformation was evaluated with respect to both the aldehyde and nitrostyrene derivative (Table 4.2). Using β -nitrostyrene **2a** as the Michael acceptor a variety of aliphatic aldehydes were examined. Straight-chain aliphatic substitution provides products in high yield (80-87%) and excellent enantioselectivities (92-93%) with the exception of acetaldehyde, which gives good yield (71%) but is only modestly selective (62%

ee). β -Branched aldehydes are tolerated and provide excellent enantioselectivity (95%), albeit in lower yield, while α -branched aldehydes do not participate. A variety of functional groups are well-tolerated including thio-ethers, silyl ethers, alkyl halides and terminal olefins. Substitution on the aryl ring of the nitroalkene leads to fairly invariant results. Ortho-, meta-, and para-substitution was examined providing fair to good yields (50-83%) in all cases as well as excellent enantioselectivities (91-94%). A 2.5 mmol scale experiment was performed using 10 mol% **9**, providing the product in 84% yield and 93% *ee*.

Table 4.2. Reaction Scope^a

Entry	R	Ar	yield % ^b	ee % ^c
1	<i>n</i> -Pr	Ph	80%	93%
2	Et	Ph	87%	92%
3	Me	Ph	71%	62%
4	<i>i</i> -Bu	Ph	32%	95%
5	TBSO-CH ₂ CH ₂ CH ₂ -	Ph	68%	87%
6	Me-S-CH ₂ CH ₂ CH ₂ -	Ph	67%	92%
7	Ph-CH ₂ CH ₂ CH ₂ -	Ph	76%	93%
8	Cl-CH ₂ CH ₂ CH ₂ -	Ph	83%	93%
9	CH ₂ =CH-CH ₂ CH ₂ CH ₂ -	Ph	83%	93%
10	Cy	Ph	<5%	N/A
11	<i>n</i> -Pr	2-Cl-C ₆ H ₄	70%	91%
12	<i>n</i> -Pr	2-F-C ₆ H ₄	75%	93%
13	<i>n</i> -Pr	2-MeO-C ₆ H ₄	83%	94%
14	<i>n</i> -Pr	3-MeO-C ₆ H ₄	63%	91%
15	<i>n</i> -Pr	3-Br-C ₆ H ₄	50%	91%
16	<i>n</i> -Pr	4-Cl-C ₆ H ₄	70%	92%
17	<i>n</i> -Pr	4-Me-C ₆ H ₄	81%	92%
18	<i>n</i> -Pr	4-(B(pin))-C ₆ H ₄	62%	91%

a Reactions conducted with 1.5 equiv **1** and 1.0 equiv **2** for 24–48 h. b Isolated yield after chromatography. c Enantiomeric excess determined by HPLC analysis on a chiral stationary phase.

Derivatization of product **3a** was performed to demonstrate the synthetic utility of β-nitro ketone products (Scheme 4.1). Reduction with NaBH₄ leads to the desired nitro alcohol in quantitative yield and 8:1 dr. The major diastereomer is easily isolated by chromatography providing 82% of **10** in 93% *ee*. Reduction of the nitro group is accomplished using

NiCl₂/NaBH₄ which, following a simple workup, is transformed to the corresponding benzamide in 81% yield and 98% *ee*.

We were intrigued by the large difference in both reactivity and enantioselectivity among diastereomeric catalysts **8** and **9**. To further probe this variance, a series of competition experiments were performed between catalysts that allowed us to assess the relative rates of product formation for each catalyst. By means of this assessment, *trans*-fluorinated pre-catalyst **9** is drastically (~13 fold) more active than *cis*-fluorinated pre-catalyst **8**. Even more remarkable is the difference in reactivity between pre-catalyst **9** and achiral catalyst **12**. Pre-catalyst **9** is the most sterically encumbered scaffold that we have examined to date, yet is still more reactive than achiral pre-catalyst **12** which lacks a bulky directing group.

Scheme 4.1. Derivatization of β-nitro ketone products. a Absolute and relative configuration determined by x-ray analysis.

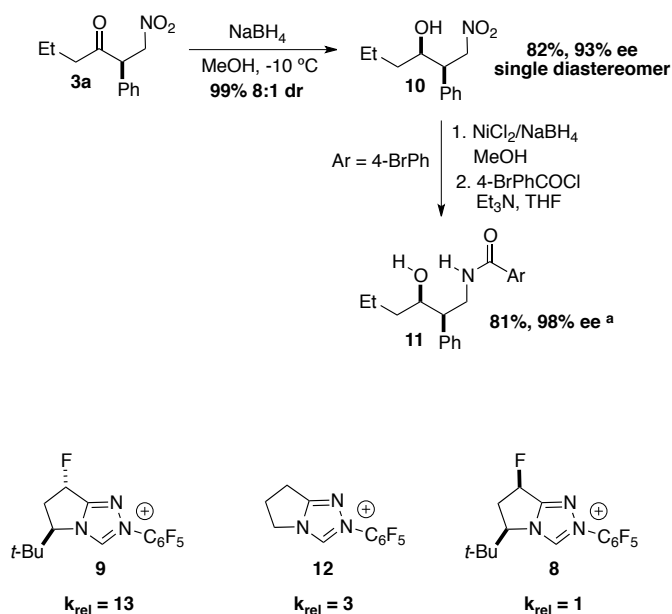


Figure 4.1. Influence of catalyst structure on relative rate. BF₄⁻ counterions omitted for clarity.

In our initial report of the asymmetric intermolecular Stetter using fluorinated triazolium salt pre-catalyst **9** (Figure 4.1), we proposed that backbone fluorination of the triazolium salt results in a conformational change due to the gauche effect.^{4c} Our recent DFT study has provided evidence that an attractive electrostatic interaction between the C-F dipole in the catalyst and the developing nitronate in the transition state is the source of increased selectivity.¹¹ Due to the divergence in both the stereochemical outcome of this reaction as well as the relative stereochemistry of the fluorinated catalyst architecture required for good selectivity, a different effect may be operative.

To further understand the effect of fluorination of the catalyst architecture in this system we undertook a DFT study. Reactions with catalysts **7-9** were quantum mechanically investigated in order to resolve the origin of selectivity. The focus of our investigation is to determine why the R enantiomer of the product is favoured, and why *trans*-fluorinated catalyst **9** is more selective than *cis*-fluorinated catalyst **8** and *des*-fluoro catalyst **7**.

Calculations were performed with Gaussian09.¹² All geometries were optimized with B3LYP/6-31G(d) with the CPCM solvation model¹³ for methanol (UAKS radii, methanol, $\epsilon = 32.6$). Single point calculations were performed on the B3LYP geometries with M06-2X/6-31+G(d,p), again using the CPCM model for methanol. Aldehyde **1a** was modelled with propionaldehyde. DFT calculations predict good transition state geometries¹⁴ and have been used to study the Stetter reaction and related reactions.^{15,11} M06-2X provides more accurate selectivities and thermochemistries and incorporates dispersion effects.¹⁶ Including the implicit solvation model was important for predicting accurate selectivities.

Catalysts **7-9** react with propionaldehyde to form acyl anion equivalents **13-15**, respectively (Figure 4.2). The favoured conformations of intermediates **13-15** are important for

determining the stereocenter that is formed in the product. No minima for the *endo* conformers could be located as all optimizations beginning from *endo* conformations converge to the *exo* conformations. Optimizations with the triazolium ring frozen in the *endo* conformation predict that these conformations are disfavoured by 5.2-7.5 kcal/mol and are therefore too high in energy to be involved in the reaction. As shown in our previous computational study,¹¹ the A^{1,3}-like strain involving the alkyl substituent (here *t*-Bu) is extremely large in the *endo* conformers of the enol intermediates, and these conformers are not involved in subsequent reactions steps.

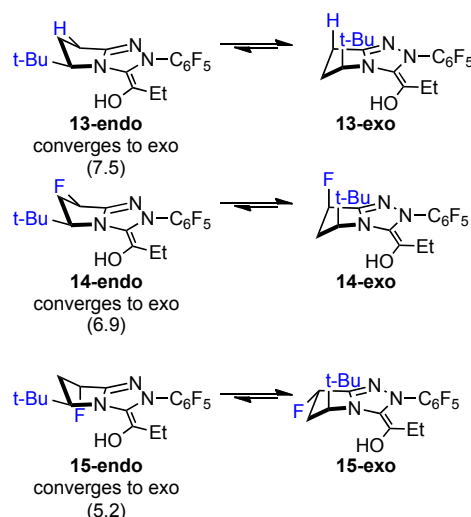


Figure 4.2. Relative free energies (M06-2X/6-31+G(d,p),CPCM(MeOH)//B3LYP/6-31G(d),CPCM(MeOH)) of acyl anion equivalents **13–15**. Endo conformations were located by freezing the triazolium ring.

Acyl anion equivalents **13–15 exo** attack either the *Si* or *Re* face of the nitrostyrene **2** leading to transition structures **TS1**, **TS3**, **TS5** or **TS2**, **TS4**, **TS6**, respectively (Figure 4.3). This step determines the stereocenter formed, but is not the rate-determining step.¹⁷ Conformations resulting from the rotation about the forming carbon-carbon bond were considered. The optimized geometries for the most favourable transition structures are shown in Figure 4.3. Unlike previous studies on similar Stetter reactions,¹¹ **TS1–TS6** do not all follow Seebach's topological rule,¹⁸ which describes a preference for a synclinal orientation of the double bonds of donors and acceptors in Michael addition transition states.

In agreement with experiment, addition to the *Re*-face of the nitrostyrene is favoured (**TS2**, **TS4**, **TS6**) and catalyst **9** is computed to be the most selective. A *gauche* – orientation of the double bonds in the *Si*-face attack (**TS1**, **TS3**, **TS5**) places the nitro group under the catalyst ring. This is electrostatically favourable for **TS1** and **TS3**. However, for **TS5** the negative electrostatic interactions between the fluorine, which points down in catalyst **9**, and the nitro group disfavors this conformation. For **TS5** the *gauche* + conformation is favoured because it is well solvated due to its large dipole moment. The *anti* conformation for the *Re*-face attack is favoured¹⁹ due to the stabilizing interaction between the hydrogen of the hydroxyl group and the carbon α to the nitro group, favourable electrostatic interactions between the alkyl group of the aldehyde and the nitro group and because it is well solvated. These combined effects are enough to favour the *Re*-face attack over the *Si*-face attack. Additionally, **TS6** is especially favourable because the negative p cloud of the phenyl ring is near the electropositive catalyst ring and the positive part of the phenyl ring is near the electronegative fluorine. Therefore, **TS6** has the lowest barrier, making catalyst **9** the most selective.

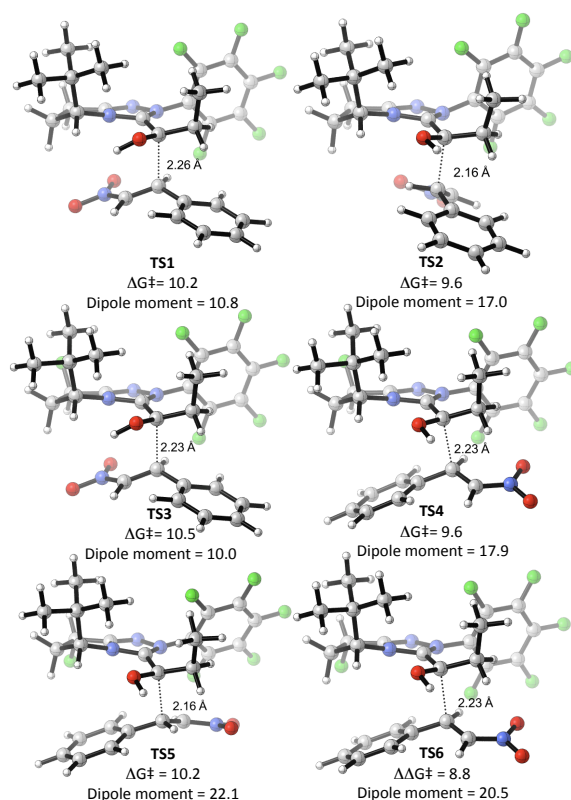


Figure 4.3. Transition structures **TS1-TS6** and energies (M06-2X/6-31+G(d,p),CPCM(MeOH)//B3LYP/6-31G(d),CPCM(MeOH)). The dipole moments are based on Mulliken charges and are given in Debye.

The experimental and calculated % ee are given in Table 4.3. The calculated % ee is determined by a Boltzman distribution of all optimized transition structures. These values correlate well with the experiment but are slightly underestimated.

Table 4.3. Calculated and experimental percent enantiomeric excesses.

Catalyst	Calculated ^[a]	Experimental
	%ee	%ee
7	62	80
8	51	74
9	80	93

In conclusion, we have identified a novel catalyst, one capable of inducing high levels of enantio-induction in the intermolecular Stetter reaction of aliphatic aldehydes and nitrostyrenes.

Trans fluorination of the catalyst architecture promotes unparalleled reactivity and enantioselectivity in this transformation, as compared to previously known scaffolds. Computations show that catalyst 9 is the most stereoselective because the Re-face attack (**TS6**) is stabilized by favorable electrostatic interactions between the phenyl group and the fluorine on the catalyst backbone.

References

-
- ¹ DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. *Angew. Chem. Int. Ed.* **2012**, *51*, 2391.
- ² For reviews see: (a) Marion, N.; Díez-González, S.; Nolan, S. P. *Angew. Chem. Int. Ed.* **2007**, *46*, 2988. (b) Enders, D.; Niemeier, O.; Henseler, A. *Chem. Rev.* **2007**, *107*, 5606-5655. (c) Moore, J.; Rovis, T. *Top. Curr. Chem.* **2009**, *290*, 77. (d) Vora, H. U.; Rovis, T. *Aldrichimica Acta.* **2011**, *44*, 3.
- ³ (a) Stetter, H.; Schreckenberger, M. *Angew. Chem., Int. Ed. Engl.* **1973**, *12*, 81. (b) Stetter, H. *Angew. Chem., Int. Ed. Engl.* **1976**, *15*, 639. (c) Stetter, H.; Kuhlmann, H. *Org. React.* **1991**, *40*, 407. (d) Christmann, M. *Angew. Chem., Int. Ed.* **2005**, *44*, 2632. (e) Rovis, T. *Chem. Lett.* **2008**, *37*, 2.
- ⁴ For a review of the asymmetric intramolecular Stetter reaction, see: de Alaniz, J. R.; Rovis, T. *Synlett.* **2009**, *8*, 1189.
- ⁵ (a) Liu, Q.; Perreault, S.; Rovis, T. *J. Am. Chem. Soc.* **2008**, *130*, 14066. (b) Liu, Q.; Rovis, T. *Org. Lett.* **2009**, *11*, 2856. (c) DiRocco, D. A.; Oberg, K. M.; Dalton, D. M.; Rovis, T. *J. Am. Chem. Soc.* **2009**, *131*, 10872. (d) DiRocco, D. A.; Rovis, T. *J. Am. Chem. Soc.* **2011**, *133*, 10402.
- ⁶ For other contributions to the asymmetric intermolecular Stetter reaction, see: (a) Enders, D.; Han, J.; Henseler, A. *Chem. Commun.* **2008**, 3989. (b) Enders, D.; Han, J. *Synthesis* **2008**, 3864. (c) Jousseume, T.; Wurz, N. E.; Glorius, F.; *Angew. Chem. Int. Ed.* **2011**, *50*, 1410. (d) Sanchez-Larios, E.; Thai, K.; Bilodeau, F.; Gravel, M.; *Org. Lett.* **2011**, *13*, 4942. (e) Fang, X.; Chen, X.; Lv, H.; Chi, Y. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 11782.
- ⁷ For reviews of fluorine's effect on molecular conformation see: (a) L. Hunter, *Beilstein J. Org. Chem.* **2010**, *6*, 38. (b) Zimmer, L. E.; Sparr, C.; Gilmour, R. *Angew. Chem., Int. Ed.* **2011**, *50*, 11860.

⁸ Enders has reported the asymmetric Stetter of butanal and chalcone with two different chiral thiazolium catalysts (4% yield, 39% ee and 29% yield, 30% ee); see: (a) Enders, D.; Bockstiegel, B.; Dyker, H.; Jegelka, U.; Kipphardt, H.; Kownatka, D.; Kuhlmann, H.; Mannes, D.; Tiebes, J.; Papadopoulos, K. *Wege zu neuen Verfahren und Produkten der Biotechnologie, DECHEMA-Monographie* **1993**, 129, 209. (b) Enders, D. *Stereoselective Synthesis*, Ottow, E.; Schöllkopf, K.; Schulz, B.- G. (Eds), Springer-Verlag, Berlin, **1994**, 63. (c) Enders, D.; Breuer, K. *Comprehensive Asymmetric Catalysis*; Springer: Berlin, **1999**; 1093. (d) Enders, D.; Balensiefer, T. *Acc. Chem. Res.* **2004**, 37, 534.

⁹ (a) Glorius et al reported a single example of an asymmetric intermolecular Stetter reaction of an aliphatic aldehyde adding to a β -unsubstituted Michael acceptor; see ref. 5c. (b) An enzyme catalyzed asymmetric Stetter reaction has recently been published providing Stetter products of acetaldehyde: : Dresen, C.; Richter, M.; Pohl, M.; Lüdeke, S.; Müller, M. *Angew. Chem. Int. Ed.* **2010**, 49, 6600.

¹⁰ The use of stronger bases and/or more polar solvents leads to base-induced decomposition products.

¹¹ Um, J. M.; DiRocco, D. A.; Noey, E. L.; Rovis, T.; Houk, K. N. *J. Am. Chem. Soc.* **2011**, 133, 11249.

¹² Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

¹³ (a) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, 102, 1995. (b) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, 24, 669.

¹⁴ Simon, L.; Goodman, J. M. *Org. Biomol. Chem.* **2011**, 9, 689.

¹⁵ (a) Dudding, T.; Houk, K. N. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 5770. (b) Hawkes, K. J.; Yates, B. F. *Eur. J. Org. Chem.* **2008**, 33, 5563.

¹⁶ (a) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215. (b) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, 41, 157.

¹⁷ Moore, J. L.; Silvestri, A. P.; de Alaniz, J. R.; DiRocco, D. A.; Rovis, T. *Org. Lett.* **2011**, *13*, 1742.

¹⁸ Seebach, D.; Golinski, J. *Helv. Chim. Acta* **1981**, *64*, 1413.

¹⁹ The *anti* conformation is favoured for **TS4** and **TS6**. The *gauche* + conformation is only slightly favored (0.5 kcal) over the *anti* conformation for **TS2**.

5. A Synthetic Recursive “+1” Pathway for Carbon Chain Elongation

Using a combination of quantum mechanical (QM) modeling, protein-substrate modeling, and protein and metabolic engineering, we have engineered the enzymes involved in leucine biosynthesis for use as a synthetic “+1” recursive metabolic pathway to extend the carbon chain of 2-ketoacids. This modified pathway preferentially selects longer chain substrates for catalysis, as compared to the non-recursive natural pathway, and can recursively catalyze 5 elongation cycles to synthesize bulk chemicals such as 1-heptanol, 1-octanol and phenylpropanol directly from glucose. The “+1” chemistry is a valuable metabolic tool in addition to the “+5” chemistry and “+2” chemistry for biosynthesis of isoprenoids, fatty acids or polyketides.¹

Introduction

Engineering microorganisms to synthesize bio-derived molecules from renewable carbon sources has drawn increasing attention in recent years. Biological synthesis is an important step towards sustainable production of fuels² or chemicals³ to replace petroleum-based approaches. However, even though nature has a diverse set of enzymes and metabolic pathways to generate primary or secondary metabolites,⁴ many useful chemicals are not accessible through natural biosynthetic pathways. For example, while more than 1 billion pounds of 1-pentanol, 1-hexanol, 1-heptanol and 1-octanol are produced annually by chemical synthesis,⁵ no metabolic pathways have been identified to make these alcohols. To broaden the applications of biosynthesis, it is necessary to expand the metabolite repertoire of living systems to cover more industrial chemicals. Liao and coworkers previously engineered the isoleucine biosynthetic pathway to elongate the carbon chain of 2-keto-3-methylvalerate and produced a branched six-carbon alcohol.⁶ In the current study, quantum mechanical modeling aided in the expansion of the

substrate range of this “+1” pathway from branched-chain ketoacids to linear-chain or even aromatic-chain ketoacids.

The natural role of our target “+1” pathway is for the biosynthesis of ketoleucine.⁷ The catalytic cycle (Figure 5.1a) begins with the condensation of acetyl-CoA to ketovaline catalyzed by 2-isopropylmalate synthase (LeuA).⁸ Then isopropylmalate isomerase complex (LeuCD) catalyzes the conversion of (2*S*)-2-isopropylmalate (**1**) to (2*R*,3*S*)-3-isopropylmalate (**3**) by dehydration and trans-rehydration.⁹ Finally, isopropylmalate dehydrogenase (LeuB) catalyzes the oxidation of (**3**) to (2*S*)-2-isopropyl-3-oxosuccinate (**4**) and decarboxylation of (**4**) to ketoleucine.¹⁰ The net effect of one catalytic cycle is the addition of one carbon unit to the input substrate. Altering the selectivity of the “+1” pathway to accept other substrates, makes it a useful tool for biosynthesis of nonnatural ketoacids with elongated carbon chain, and nonnatural alcohols upon enzymatic decarboxylation and reduction Figures 5.21b and c outline how this biosynthetic pathway can be used to make alcohols for biofuels. Since LeuA is the first committed enzyme of the elongation pathway, and mutating LeuA can change the pathway selectivity on branched-chain substrates,⁶ we investigated the effect of modulating the substrate specificity of LeuA on linear-chain and aromatic-chain substrates.

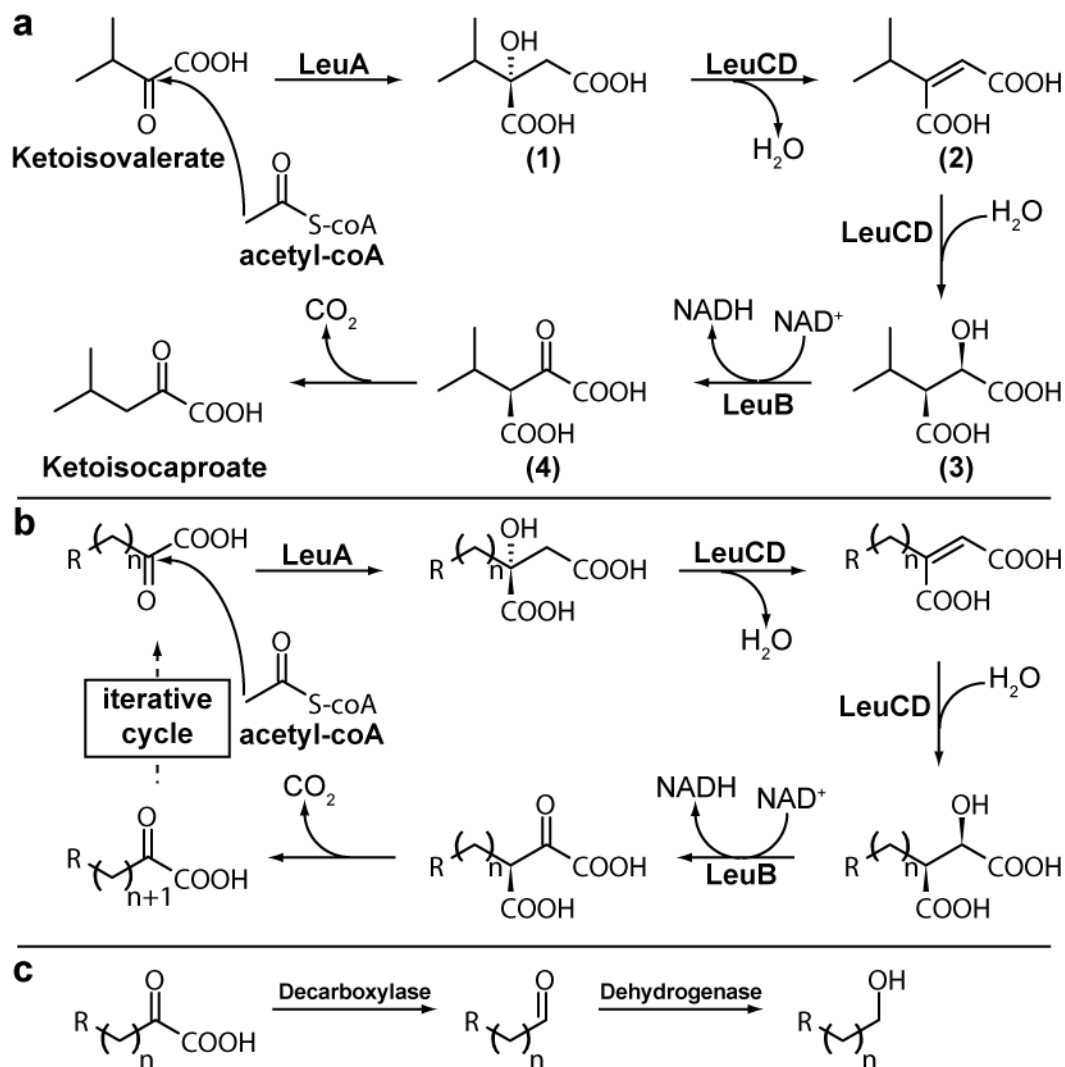


Figure 5.1. Engineering LeuABCD “+1” biosynthetic pathway. **(a)** Reaction mechanism of LeuABCD pathway. Natural substrate ketovaline is converted to ketoisocaproate after five enzyme-catalyzed chemical steps. **(b)** Theoretical “+1” iterative elongation of 2-ketoacid substrates containing different R-groups. This pathway enables the biosynthesis of nonnatural ketoacids. **(c)** Alcohols are produced from 2-ketoacids by enzymatic decarboxylation and reduction.

Results and Discussion

To address the feasibility of using LeuA to catalyze recursive elongation of long chain substrates, we performed quantum mechanical calculations of the transition state of the carbon-carbon bond formation step catalyzed by LeuA, modeled as shown in Figure 5.2. These calculations determined the best conformation and geometry for the transition state for this step,

and showed how the free energy changes as a function of the R group on the 2-ketoacid. Because this model is simplified compared to the reaction in the protein, the absolute energy barrier is not highly accurate, but gives a good indication of how the barrier changes as a function of the R group on the 2-ketoacid.

The conformation of the transition state follows the Felkin-Anh model for nucleophilic attack on a carbonyl.¹¹ The largest substituent on the carbon adjacent to the carbonyl is pointed away from the incoming nucleophile. This conformation is nearly identical to that of ketoisovalerate in the *Mycobacterium tuberculosis* LeuA crystal structure¹² (Figure 5.2b) and fits well into the active site of LeuA (Figure 5.2c). The forming carbon-carbon bond distance is about 2.04 Å for alkyl chain substrates and lengthens as the R group becomes more electrophilic, corresponding to an earlier transition state (Table 5.1).

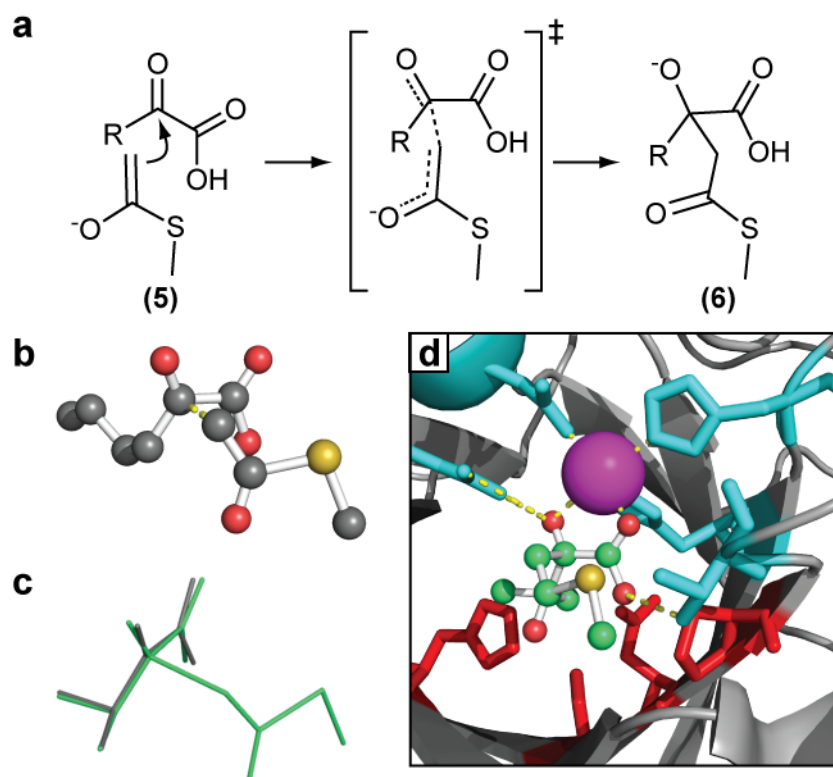


Figure 5.2. Modeling of carbon-carbon bond formation and the LeuA active site. (a) Quantum mechanical model of the carbon-carbon bond formation step catalyzed by LeuA. A 2-ketoacid is attacked by 1-(methylthio)ethenolate **5** (a model for acetyl-CoA), resulting in the formation of a 2-ketoacid enolate **6**. (b) Stereo view of ketoisovalerate from PDB 1SR9 (gray) and the ketoisovalerate enolate calculated by QM (green) aligned over 8 shared heavy atoms (r.m.s.d. = 0.15 Å). (c) LeuA active site modeled with ketoisovalerate enolate. Residues involved in coordinating the Zn^{2+} ion (magenta sphere) and the substrate (green, red, and yellow ball-and-stick) are shown in cyan (Arg12, Asp14, Thr171, His202, His204 in *E. coli* LeuA; corresponding to Arg80, Asp81, Thr254, His285, His287 in *Mycobacterium tuberculosis* LeuA¹²). Coordination interactions are shown as dashed yellow lines. Residues within 4.5 Å of the γ -methyl of ketoisovalerate (other than those involved in metal ion and substrate coordination) are shown in red. The protein backbone is shown as a gray cartoon.

As shown in Table 5.1, all alkyl chain substrates have activation energies of about 9 kcal/mol, indicating that increasing chain length or the branched nature of the R group does not change the activation free energy of the reaction. Therefore, there is no inherent barrier to the use of long alkyl chain substrates by LeuA. In addition to alkyl chain substrates, we also performed calculations for the aromatic substrates phenylpyruvate and homophenylpyruvate to examine the

potential for their elongation by LeuA. We found that in these cases, as compared to alkyl chains, the activation energy (ΔG^\ddagger) is lowered by 2.7 kcal/mol for phenylpyruvate and by 1.6 kcal/mol for homophenylpyruvate, indicating that there is also no barrier to the use of these substrates by LeuA. This decreased activation energy is due to the electron-withdrawing nature of the phenyl ring, which makes the alpha-keto carbon of the substrate that is the target of nucleophilic attack by acetyl-coA more electrophilic. This observation was confirmed using calculations of *p*-chloro-phenylpyruvate, which had a ΔG^\ddagger of 3.5 kcal/mol, and a product energy of 3.1 kcal/mol.

Table 5.1. Quantum Mechanical Calculation Results. B3LYP/6-31-G(d) CPCM(ethylphenylether, $\epsilon=4.1797$) Transition state distance is the length of the forming carbon-carbon bond.

Substrate	ΔG^\ddagger (kcal/mol)	$\Delta G_{\text{reaction}}$ (kcal/mol)	Transition state distance (Å)
Ketoisovalerate	9.1	9.0	2.03
Pyruvate	8.6	6.4	2.01
2-Ketobutryate	9.3	7.4	2.05
2-Ketovalerate	9.2	8.6	2.04
2-Ketocaproate	9.4	8.8	2.04
2-Ketoheptanoate	9.2	8.2	2.04
2-Ketooctanoate	9.0	8.6	2.04
Phenylpyruvate	6.4	6.4	2.12
Homophenylpyruvate	7.5	7.5	1.98

These calculations show that the sterics of the R group do not greatly affect the barrier for this reaction, but the electronics of the substituent are important. Combined, these data suggest that there is no barrier to the recursive elongation of linear chain substrates or even aromatic substrates as long as they can fit into the enzyme active site and substrate binding pocket and are oriented appropriately with respect to the residues involved in coordination of the metal ion and the substrate (Figure 5.2c).

Structure-based protein engineering and modeling of LeuA-substrate molecular complexes.

Residues in the substrate-binding pocket of *Ec*LeuA (His97, Ser139, Asn167, Pro169) were identified based on the crystal structure of *Mt*LeuA (PDB 1SR9, corresponding residues are His167, Ser216, Asn250 and Pro252).¹² These residues are located within a radius of 4.5 Å of the γ -methyl group of the bound ketoisovalerate in the structure and likely influence substrate binding (Figure 5.2d). To increase the size of the substrate-binding pocket, these residues were systematically mutated to smaller amino acids such as alanine or glycine to enlarge the binding pocket of *Ec*LeuA and allow the accommodation of bigger substrates. A series of *Ec*LeuA mutants based on the leucine-feedback insensitive *G462D Ec*LeuA mutant¹³ were made and their binding pocket volumes were analyzed by CASTp¹⁴ (<http://sts.bioengr.uic.edu/castp/index.php>) (Table 5.2). A comparison of the increased active site volume in the largest of these mutants (50.0 Å³, 83% increase over the *G462D* mutant) *versus* the increase in molecular van der Waals volume of the substrates relative to 2-ketobutyrate predicts that the *H97A/S139G/N167G/P169A/G462D Ec*LeuA mutant should be accessible to the range of substrates examined by QM modeling (Table 5.2).

Table 5.2 .Enzyme Binding Pocket and Substrate Volumes

Enzyme Mutation(s)	Increase in Solvent Accessible Volume (\AA^3) ^b	Substrate	Increase in Molecular Volume (\AA^3) ^c
G462D ^a	0	2-Ketobutyrate	0
S139G, G462D	5.1	2-Ketovalerate	17.0
S139G, N167A, G462D	9.2	Ketoisovalerate	17.2
H97A, S139G, G462D	33.4	2-Ketocaproate	34.0
H97A, S139G, N167A, G462D	42.6	2-Ketoheptanoate	50.9
H97A, S139G, N167G, G462D	47.0	Phenylpyruvate	53.7
H97A, S139G, N167G, P169A, G462D	50.0	2-Ketoocatanoate	67.8
		Homophenylpyruvate	70.6

^a The feedback-resistant *EcLeuA* mutant. ^b Relative to a solvent accessible volume for *G462D EcLeuA* of 60.0 \AA^3 , as calculated by CASTp (17) (<http://sts.bioengr.uic.edu/castp/index.php>). ^c Relative to a molecular volume for 2-Ketobutyrate of 92.1 \AA^3 , as calculated by chemicalize.org (<http://www.chemicalize.org>).

Molecular modeling of the *G462D EcLeuA* and the *H97A/S139G/N167G/P169A EcLeuA* mutants with RosettaDesign¹⁵ was performed to examine this prediction in greater detail. The quantum mechanically optimized conformations of each substrate were used as an input along with either the *G462D* mutant enzyme based on PDB 1SR9 or a mutant enzyme based on 1SR9 containing the *H97A/S139G/N167G/P169A/G462D* mutations. Overall, all models exhibited similar packing scores, indicating that the conformation of the active site is well folded and internal to the protein structure, an observation that is confirmed upon analysis of the energy-minimized output structures in comparison to their respective input structures. To prevent substrates from leaving the active site during refinement, a high force constant was applied to keep them in the vicinity of the catalytic residues (His202, His204 in *EcLeuA*). Modeling results indicate that the KIV, 2-ketobutyrate (C4), and 2-ketovalerate (C5) substrates are well-positioned in the active site of the *G462D* enzyme and have ligand binding scores of 2.4, 3.3, and 4.1,

respectively (Figure 5.3a). The 2-ketocaproate (C6) substrate shows a potential steric clash with the interior of the active site and has a ligand binding score of 18.2, while the 2-ketoheptanoate (C7), 2-ketooctanoate (C8) and aromatic substrates have significant steric clashes with the protein, indicated by their high (*i.e.* >40) ligand binding scores (Figure 5.3a). These observations predict that the *G462D* enzyme will use the KIV, C4, C5 and possibly the C6 substrate and will not use the C7-C8 or aromatic substrates due to their inability to fit well into the active site. Modeling results for the *H97A/S139G/N167G/P169A/G462D* mutant enzyme revealed that the KIV, C4-C7, and aromatic substrates all fit well into the active site and have ligand binding scores that are all less than 3.5 (Figure 5.3a), predicting that this mutant enzyme will be able to use all these substrates. The C8 substrate has a higher ligand binding score of 27.6, but shows only limited potential steric clashes with the active site (Figure 5.3a), so its use by the *H97A/S139G/N167G/P169A/G462D* mutant enzyme is possible but not predicted definitively.

In addition to these predictions of enzyme binding of substrates, the solvent accessible surface area (SASA) was examined for each substrate in the protein-substrate complexes (Figure 5.3b) to predict which substrate(s) would be most preferred by the *G462D* and *H97A/S139G/N167G/P169A/G462D* mutant enzymes. The basis for this predicted substrate selectivity is the hydrophobic effect, where the substrate able to exclude the most water from the active site (thus having the least SASA) will be the preferred substrate for the enzyme. For the *G462D* enzyme model, the SASA is smallest for the C4 substrate, while for the *H97A/S139G/N167G/P169A/G462D* mutant enzyme the SASA is smallest for the HPPy aromatic substrate, although the C6 substrate has a similarly small SASA (Figure 5.3b).

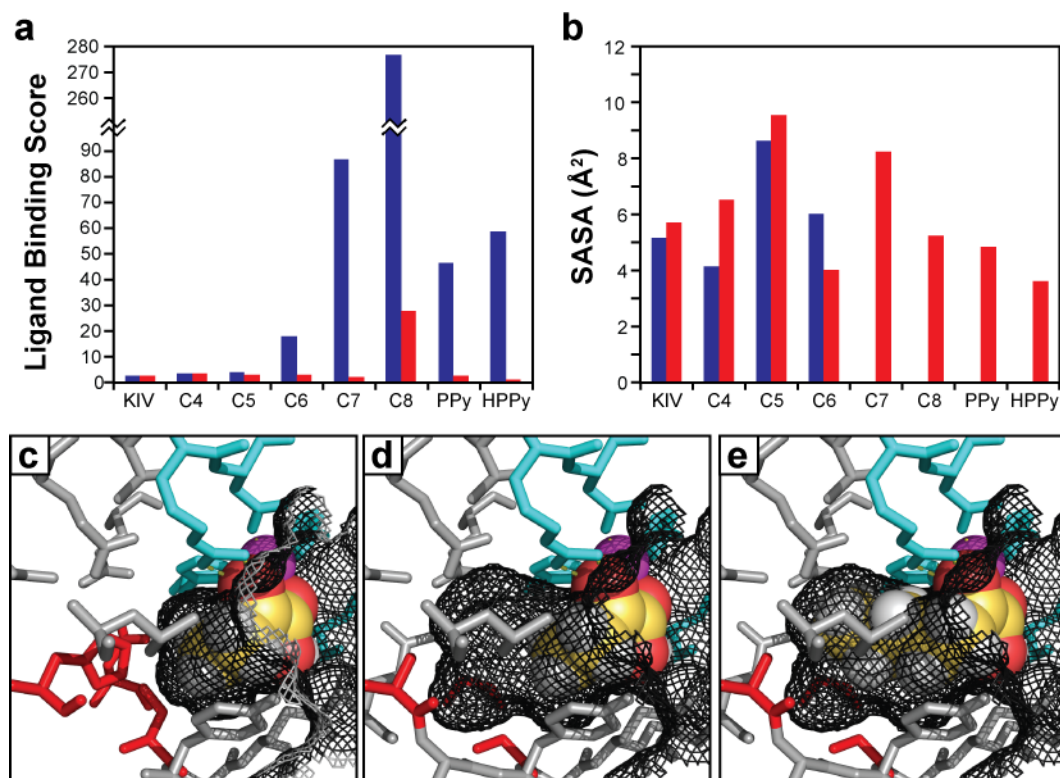


Figure 5.3. Comparison of substrate binding and substrate solvent accessible surface area in feedback-resistant and quintuple mutant *EcLeuA* protein-substrate complex models. (a) Ligand binding score of substrates, as calculated by Rosetta. Smaller scores indicate lower calculated energy parameters and better substrate binding. Scores for substrates in complex with the *G462D* and *H97A/S139G/N167G/P169A/G462D* mutant proteins are shown in blue and red, respectively. (b) Solvent accessible surface area (SASA) of substrates in the binding pockets of *G462D* (blue) and *H97A/S139G/N167G/P169A/G462D* mutant (red) proteins, as calculated by RosettaDesign. SASA values for C7, C8, PPy, and HPPy in the *G462D* protein binding pocket have no meaning due to the steric clash observed between these substrates and the enzyme and are therefore not shown. (c-e) Protein-substrate models with binding pocket are shown as wire mesh and substrates shown as spheres. Protein residues, metal ion, and coordination interactions are shown and colored as indicated (Figure 5.2). (c) *G462D* protein with C4 substrate. (d) *H97A/S139G/N167G/P169A/G462D* mutant protein with C4 substrate. (e) *H97A/S139G/N167G/P169A/G462D* mutant protein with C7 substrate.

Additionally, in the protein-substrate complex models, the C4 substrate fits well into the *G462D* enzyme binding pocket and occupies most of the available volume (Figure 5.3c), as compared to the *H97A/S139G/N167G/P169A/G462D* mutant enzyme binding pocket where the substrate occupies only a portion of the available volume (Figure 5.3d). These observations indicate that

the C4 substrate may be the preferred substrate for the *G462D* enzyme, while either the C6 or the HPPy substrate (or both) will be the preferred substrate(s) for the *H97A/S139G/N167G/P169A/G462D* mutant enzyme. Interestingly, the C6 substrate is approximately 34 Å³ larger than the C4 substrate (Table 5.2), correlating well with the gain in solvent-accessible volume in the *H97A/S139G/N167G/P169A/G462D* mutant protein active site relative to the *G462D* protein active site (50.0 Å³, Table 5.2) and strengthening the prediction that the C6 substrate will be preferred by the *H97A/S139G/N167G/P169A/G462D* mutant enzyme. Also of note, the C7 substrate is approximately 50 Å³ larger than the C4 substrate (Table 5.2), an identical volume increase as that of the active site, and it fits well into the *H97A/S139G/N167G/P169A/G462D* mutant enzyme binding pocket (Figure 5.3e). This volumetric correlation and protein-substrate complex model predict that the C7 substrate may be used preferentially to the C4, C5 and C8 substrates (although not the C6 substrate) in the *H97A/S139G/N167G/P169A/G462D* mutant enzyme.

Enzyme kinetics of feedback-resistant and mutant EcLeuA.

Enzymatic assays were performed on *G462D* EcLeuA (feedback-resistant) and the quintuple mutant *H97A/S139G/ N167G/P169A/G462D* EcLeuA to test the QM and molecular modeling predictions and examine substrate range and specificity. *G462D* EcLeuA is active on the C4, C5, and C6 substrates (Table 5.3). The enzyme has the highest activity for 2-ketobutyrate, as measured by its catalytic rate constant (k_{cat}). Activity decreases 3-fold towards 2-ketovalerate and 100 fold towards 2-ketocaproate. The most preferred substrate, as measured by catalytic efficiency (k_{cat}/K_m), is 2-ketobutyrate (Table 5.3). Catalytic efficiency is 1.3-fold lower for 2-ketovalerate and 28-fold lower for 2-ketocaproate as compared to 2-ketobutyrate.

There is no detectable activity for *G462D EcLeuA* towards 2-ketoheptanoate, 2-ketooctanoate, or aromatic substrates (Table 5.3).

Several differences in specificity were observed for the *H97A/S139G/N167G/P169A/G462D* mutant *EcLeuA*. This mutant *EcLeuA* has a broader substrate range than *G462D EcLeuA*, and activity is seen for all substrates, as predicted by the QM calculations. Despite having similar activities towards the alkyl chain C4-C8 substrates with k_{cat} values in the range of 3.27 to 6.51 s⁻¹, the most preferred substrates for the enzyme are 2-ketocaproate and 2-ketoheptanoate (Table 5.3). The catalytic efficiencies of these substrates are approximately 2-fold higher than that of 2-ketooctanoate and approximately 3-fold higher than for 2-ketobutyrate and 2-ketovalerate. The enzyme has less activity for the aromatic substrates than for the linear substrates, with k_{cat} values of 0.091 s⁻¹ and 0.38 s⁻¹ for phenylpyruvate and homophenylpyruvate respectively, and has similar catalytic efficiencies for both aromatic substrates (0.77 and 0.66 respectively). The lower activity of the quintuple mutant *EcLeuA* for the aromatic substrates may be due to their deviation from the ideal transition state distance for all other substrates (2.04 ± 0.01 Å *versus* 2.12 Å and 1.98 Å for PPy and HPPy, respectively, Table 5.1) as predicted by QM calculations. In the context of the enzyme active site, these deviations may result in the observed decrease in activity for the aromatic substrates.

These kinetic data agree well with the predictions of the QM and RosettaDesign modeling, particularly for the alkyl chain substrates. As predicted, the preferred substrate for the *G462D* enzyme is C4 and the preferred substrate(s) for the *H97A/S139G/N167G/P169A/G462D* mutant enzyme are C6/C7. However, the prediction that the aromatic substrate homophenylpyruvate would be the most preferred substrate for the *H97A/S139G/N167G/P169A/G462D* mutant enzyme was not supported, indicating that other

factors, such as deviation from the ideal transition state distance, are involved in the use of the aromatic substrates by this mutant *EcLeuA* in addition to solvent exclusion by the substrate and the electronic state of its alpha-keto carbon. Nonetheless, the *H97A/S139G/N167G/P169A/G462D* mutant *EcLeuA* was capable of using the aromatic substrates, as predicted by the QM calculations, enabling the recursive elongation of aromatic 2-ketoacids.

Table 5.3. Kinetic parameters of feedback-resistant and quintuple mutant *EcLeuA*.

	G462D <i>EcLeuA</i>^a			<i>EcLeuA</i> H97A, S139G, N167G, P169A, G462D mutant		
Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1} s^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1} s^{-1}$)
Ketoisovalerate (KIV)	206±21 ^b	1.15±0.03	5.58±0.59	583±68	0.96±0.04	1.65±0.20
2-Ketobutyrate (C4)	345±21	2.43±0.04	7.04±0.44	657±31	3.27±0.05	4.98±0.25
2-Ketovalerate (C5)	135±10	0.73±0.012	5.41±0.41	815±63	4.52±0.13	5.54±0.46
2-Ketocaproate (C6)	105±10	0.026±0.001	0.25±0.03	456±86	6.51±0.39	14.28±2.83
2-Ketoheptanoate (C7)	0	0	0	349±80	5.21±0.35	14.93±3.57
2-Ketooctanoate (C8)	0	0	0	455±69	3.49±0.17	7.67±1.22
Phenylpyruvate	0	0	0	118±26	0.09±0.01	0.77±0.18
Homophenylpyruvate	0	0	0	575±101	0.38±0.02	0.66±0.12

^a The feedback-resistant *EcLeuA* mutant. ^b All values represent the average of three independent experiments, with standard deviation indicated.

Metabolic engineering of *EcLeuA*-catalyzed "+1" recursive elongation pathway in *E. coli*.

Finally, we examined the activity of the *EcLeuA* mutants (Table 5.2) *in vivo*. The series of mutants were individually constructed into an operon composed of six genes in the

transcriptional order *leuA** *leuB* *leuC* *leuD* *kivd* *adh6* (Figure 4a) under the control of the *PLlacO1* promoter on a high copy plasmid (pZE_LeuABCDKA6). Kivd is a mutant (*F381L/V461A*) form of 2-ketoacid decarboxylase from *Lactococcus lactis*,⁶ and Adh6 is the alcohol dehydrogenase VI from *Saccharomyces cerevisiae*.¹⁶ Introduction of Kivd and Adh6 enables the conversion of 2-ketoacid products into alcohols. The production host is a modified threonine-hyperproduction strain (ATCC98082 Δ *rhtA*) with the deletion of the threonine transporter gene *rhtA*.¹⁷ Overexpression of *thrABC*⁶ in a low copy plasmid helps drive the carbon flux towards threonine. Another operon in the transcriptional order *ilvA leuA* was also built on a medium copy plasmid (Figure 4a). IlvA is the *Bacillus subtilis* threonine dehydratase that is active in converting threonine into 2-ketobutyrate.¹⁷ The additional overexpression of *G462D EcLeuA* in the medium copy plasmid is to provide basal elongation activity for small substrates,¹⁸ which is required for more robust production of alcohols longer than 1-butanol (Table 5.4). The final production strain has a total of 11 genes overexpressed.

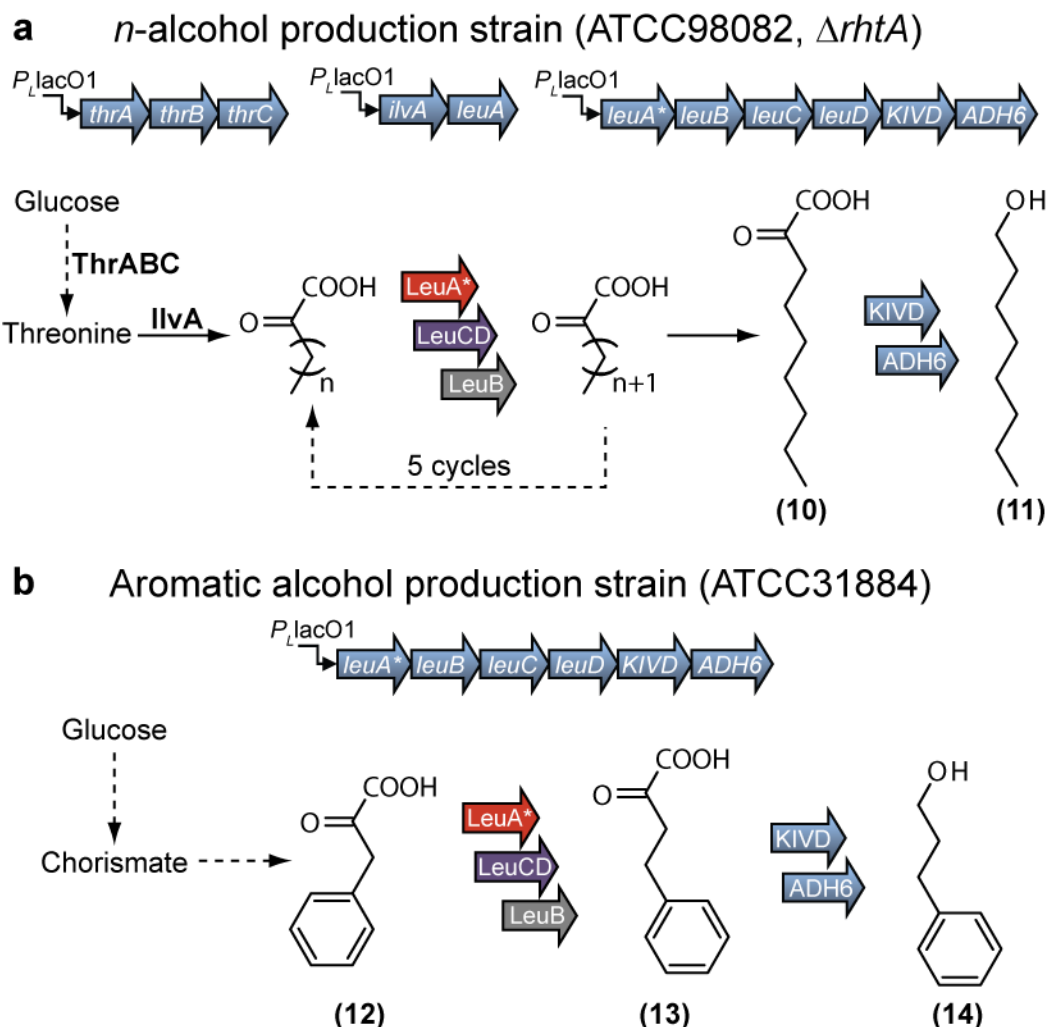


Figure 5.4. Metabolic engineering of *E. coli* for long-chain alcohol production. (a) Synthetic operons for gene expression in threonine overproducer to produce linear chain alcohols. Overexpression of ThrABC and IlvA drives the carbon flux towards 2-ketobutyrate. Engineered protein *H97A/S139G/N167G/P169A/G462D EcLeuA* (LeuA*) enables *EcLeuA*BCD* to catalyze recursive elongation of natural metabolite 2-ketobutyrate ($n=1$) to 2-ketononanoate **10**. Decarboxylation and reduction of 2-ketononanoate by Kivd and Adh6 results in 1-octanol **11**. (b) Synthetic operons for gene expression in phenylalanine overproducer. LeuA* enables the elongation of phenylpyruvate **12** into homophenylpyruvate **13**. Decarboxylation and reduction of 2-ketononanoate by Kivd and Adh6 results in phenylpropanol **14**.

Table 5.4. Alcohol production profiles of *EcLeuA* mutants.

Enzyme Mutation(s)	Additional <i>EcLeuA</i> G462D ^a	<i>n</i> -Alcohols produced (mg l ⁻¹)					
		1-propanol	1-butanol	1-pentanol	1-hexanol	1-heptanol	1-octanol
G462D ^a	+	35±10 ^b	58±1	953±39	31±3	ND ^c	ND
S139G, G462D	+	316±78	993±181	2220±142	219±11	4.8±1.4	ND
S139G, N167A, G462D	+	303±145	843±223	1380±69	204±13	8.9±1.7	ND
H97A, S139G, G462D	+	285±34	872±71	1751±120	302±33	24±3.9	ND
H97A, S139G, N167A, G462D	+	329±38	897±81	957±42	148±6.5	77±9.8	ND
H97A, S139G, N167G, G462D	+	368±15	932±79	1131±107	104±11	41±9.9	ND
H97A, S139G, N167G, P169A, G462D	+	105±20	406±47	656±50	146±10	80±11	2.0±0.26
H97A, S139G, N167G, P169A, G462D	-	187±37	374±38	148±8	109±12	23±4	ND

^a The feedback-resistant *EcLeuA* mutant. ^b All values represent the average of three independent experiments, with standard deviation indicated. ^c ND = not detected.

Alcohol products produced during fermentative growth were identified by GC-MS and quantified by GC-FID (Table 5.4). The *G462D EcLeuA* produced primarily 1-pentanol (955 mg l⁻¹), with small amounts of 1-butanol and 1-hexanol also produced (Table 5.4). This reflects the *in vitro* enzyme activity, where the C4 and C5 substrates are efficiently used, converting the initial 2-ketobutyrate supplied through the threonine biosynthesis pathway and the action of *IlvA* into 2-ketovalerate and 2-ketocaproate recursively, with the 2-ketocaproate subsequently decarboxylated and dehydrogenated by *Kivd* and *Adh6*, resulting in 1-pentanol formation. The *EcLeuA* mutants with expanded active site volumes yielded additional long-chain alcohols. The *S139G/G462D EcLeuA* mutant produced 4.8 mg l⁻¹ 1-heptanol, an alcohol never before produced directly from glucose, in addition to significant amounts of 1-butanol and 1-pentanol (Table 5.4). As the volume of the active site was increased by further mutation, the amount of 1-heptanol also increased up to 77 mg l⁻¹ for the *S139G/H97A/N167A/G462D EcLeuA* mutant (Table 5.4). This is not a perfect correlation, however, as the *S139G/H97A/N167G/G462D EcLeuA* mutant only produced 41 mg l⁻¹ 1-heptanol (Table 5.4). The mutant with the largest active site volume, *H97A/S139G/N167G/P169A/G462D EcLeuA*, recaptured the trend of increasing long chain alcohol production and produced 80.2 mg l⁻¹ 1-heptanol and even 2.0 mg l⁻¹ 1-octanol (Table 5.4), another alcohol never before produced directly from glucose. None of the other mutants were capable of producing 1-octanol. These results indicate that our engineered *EcLeuABCD* module can catalyze the recursive elongation of the natural metabolite 2-ketobutyrate over 5 cycles to 2-ketononanoate **10** (Figure 4a). This C9 ketoacid and the other smaller ketoacids produced by recursive elongation were then converted into 1-butanol, 1-pentanol, 1-hexanol, 1-heptanol and 1-octanol *in vivo* by *Kivd* and *Adh6*.¹⁹

To investigate the applicability of the “+1” pathway to aromatic substrates *in vivo*, we transformed the *leuA*-leuB-leuC-leuD-kivd-adh6* operon encoding *S139G/H97A/N167G/P169A/G462D EcLeuA* into a phenylalanine-overproducing *E. coli* strain ATCC31884 (Figure 4b). Remarkably, besides 664.4 mg l⁻¹ phenylethanol, 4.1 mg l⁻¹ phenylpropanol was identified in the final fermentation mixture. Without the quintuple mutant *EcLeuA*, phenylpropanol was not produced (data not shown). These results demonstrate that the *EcLeuABCD* “+1” pathway could even elongate aromatic substrate phenylpyruvate **12** into homophenylpyruvate **13** (Figure 4b) *in vivo*.

Conclusion

We have demonstrated the utility of using a combination of QM modeling, protein-substrate complex modeling, structure-based protein engineering, and metabolic engineering to design, analyze and produce bacterial strains capable of producing 1-heptanol, 1-octanol, and phenylpropanol. Prior to this study, these long chain alcohols had never before been produced directly from glucose, a process that represents 5 iterations of the engineered *EcLeuABCD* pathway. The other enzymes involved in the alcohol production pathway (LeuCD, LeuB, Kivd and Adh6) have demonstrated remarkable plasticity in enabling production of long chain and aromatic alcohols despite not being selected and modified for such activity. Indeed, the enzymatic reactions catalyzed by these enzymes may be limiting steps for the continued iteration of carbon-chain elongation in the *EcLeuABCD* pathway, as *in vitro EcLeuA* enzyme assays indicate that robust catalytic efficiency is retained even for the C8 substrate, while only a small quantity of 1-octanol is produced *in vivo*. In particular, LeuCD may not be capable of using much larger substrates relative to the native isopropylmalate substrate. This enzyme complex

has been studied in *M. jannaschii*²⁰ and is approximately 35% identical to the *E. coli* LeuCD, as determined by BLAST²¹ (www.ncbi.nlm.nih.gov/BLAST). The *M. jannaschii* LeuCD actually had higher activity for smaller substrate analogs of isopropylmalate than for the native substrate itself.²⁰ Although no larger substrates were analyzed, if this trend continues, substrates such as butylmalate, pentylmalate and hexylmalate, may not be usable by the enzyme, hence limiting the throughput of the "+1" elongation pathway.

LeuB has been studied in both *E. coli*²² and *T. thermophilus*,²³ with differing substrate ranges reported for each organism-specific enzyme, despite being 51% identical.²² *E. coli* LeuB had similar catalytic activities of 70 s⁻¹ and 60 s⁻¹ for ethylmalate and isopropylmalate, respectively, but no activity was observed for isobutylmalate and no larger substrates were tested,²¹ while *T. thermophilus* LeuB had catalytic activities from approximately 30-49 s⁻¹ for the substrates ethylmalate, isopropylmalate, isobutylmalate and isoamylmalate.²³ These observations suggest that LeuB may be substrate-limited in *E. coli* but LeuB enzymes from other organisms may be able to circumvent this potential limitation. We are currently investigating this hypothesis.

Kivd, while not limiting to the process of "+1" chain elongation, may possibly be limiting for alcohol production at the 2-ketoacid decarboxylation step of the pathway (Figure 5.1). This enzyme has reduced activity on substrates larger than the native ketoisovalerate, with about 23%, 17%, 7%, and 9% relative activities for ketoisocaproate, ketomethylvalerate, ketomethylthiobutyrate and phenylpyruvate, respectively.²⁴ We have also previously shown that Kivd has about 28% relative activity on ketomethylhexanoate as compared to ketoisovalerate.⁶ These data indicate that larger substrates are not used as efficiently as the native substrate. While our mutant Kivd has an enlarged active site that allows for about 110% relative activity on

ketomethylhexanoate as compared to ketoisovalerate,⁶ its properties for other substrates remain to be determined and are the subject of our current studies.

Adh6, like Kivd, is not limiting to "+1" chain elongation. Further, this enzyme is active on a broad range of substrates with similar activities for aldehydes from pentanal to octanal,²⁵ so it is not expected to be limiting to alkyl-chain or aromatic alcohol production.

QM modeling of the reaction catalyzed by *EcLeuA*, and molecular modeling of the protein-substrate complex have given insight into the role of steric effects, SASA and hydrophobicity of the ligand on binding. These models show the strengths and weaknesses of *EcLeuA* for metabolic engineering and will be expanded upon to reengineer *EcLeuA* to either attain greater substrate specificity for long chain 2-ketoacids or accept even more diverse substrates to allow further recursive cycles. Engineering of these enzymes for increased activity on long chain and aromatic substrates using the methods described here may further advance long chain alcohol production, as well as the production of an expanded range of nonnatural metabolites, which are useful for a variety of purposes.

In summary, by engineering *EcLeuA* to be active on aromatic or long chain aliphatic ketoacids, we turned *EcLeuABCD* into a flexible biosynthetic pathway for recursive "+1" elongation of the carbon chain of branched-chain,⁶ linear-chain and even aromatic-chain ketoacids. This synthetic elongation pathway is a nice complement to the "+5" and "+2" chemistries existing in nature to produce isoprenoids,²⁶ polyketides and fatty acids²⁷ for the production of industrially and pharmaceutically relevant metabolites. The combination of QM calculations, molecular modeling, and metabolic engineering described in this work has been successful in identifying and modeling enzymes and reactions that can be modified to perform in a recursive manner and may be applied to the investigation and modification of other

biosynthetic pathways, such as branched-chain amino acid synthesis and isobutanol production using acetolactate synthase,^{2a} to produce additional nonnatural metabolites and long-chain biofuels.

Methods

Vector construction. Plasmids pZS_thrO, containing the *thrA^{fbr}BC* operon under control of the P_{LlacO1} promoter, and pZE_LeuABCDKA6, containing the *leuA^{fbr}BCD; kinvd; adh6* operon under control of the P_{LlacO1} promoter, were previously constructed.⁶ Plasmid pZAlac_ilvA_{BS}leuA, containing the *ilvA* gene from *Bacillus subtilis* and the *leuA^{fbr}* gene from *Escherichia coli* under control of the P_{LlacO1} promoter, was constructed by amplifying the threonine dehydratase gene (*ilvA*) from the genomic DNA of *B. subtilis* using the primer pair IlvAbsaccfwd/IlvAbssalrev.¹⁷ This fragment was digested with Acc65I and SalI. The *leuA^{fbr}* gene was amplified from *E. coli* genomic DNA using the primer pair LeuAsalfwd/LeuAxbarev (sequences available in Supplemental Information). This fragment was digested with SalI and XbaI. The fragments were ligated and inserted into pZElac¹⁷ to create plasmid pZElac_ilvA_{BS}leuA. The replication origin was changed from colE1 to p15A by digesting pZA31-luc²⁸ with AvrII and SpeI. The shorter fragment was gel purified and cloned into pZElac_ilvA_{BS}leuA cut with the same enzymes, creating pZAlac_ilvA_{BS}leuA.

Fermentation procedure. For n-alcohol production, the threonine-hyperproduction strain ATCC98082 with *rhtA* knockout¹⁷ was transformed with pZS_thrO, pZAlac_ilvA_{BS}leuA, and pZE_LeuABCDKA6 containing various *EcLeuA* mutations. For aromatic alcohol production, the phenylalanine-hyperproduction strain ATCC31884 was transformed with pZE_LeuABCDKA6 containing the *H97A/S139G/N167G/P169A/G462D EcLeuA* mutant.

Fermentations for alcohol production were carried out as previously described,⁶ with the total incubation time being 48 hours.

GC-MS and GC-FID Analysis. Alcohols were subjected to mass analysis and quantification as previously described.⁶ For detection of phenylpropanol and 1-octanol, alcohols were extracted by 0.5 mL n-hexane from 5 mL of fermentation medium. GC oven temperature was initially placed at 85 °C for 2 min. Temperature was increased with a gradient of 45 °C min⁻¹ until 118 °C. Then temperature was increased with a gradient of 0.5 °C min⁻¹ until 119 °C, followed by a gradient of 65 °C min⁻¹ until 182 °C, which was held for 4 min. Finally, temperature was increased with a gradient of 65 °C min⁻¹ until 235 °C and held for 3 min.

Protein Expression and Purification. Genes encoding *H97A/S139G/N167G/P169A/G462D* LeuA were amplified from plasmid pZE_LeuABCDKA6 and cloned into pQE9 (Qiagen) as previously described⁶ to create pQE_hisleuA_GSAGA. The plasmid pQE_hisleuA_GD, containing the *leuA^{br}* gene was previously created.⁶ Transformation, growth, induction and purification conditions were performed as previously described.⁶

Enzymatic Assay. Compounds 2-ketoisovalerate, 2-ketobutyrate, 2-ketovalerate, 2-ketocaproate, 2-ketooctanoate, phenylpyruvate, 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) and acetyl-CoA were purchased from Sigma-Aldrich. Homophenylpyruvate was purchased from Chembridge Inc. and 2-ketoheptanoate was custom synthesized by AsisChem, Inc. Protein concentration was determined by measuring UV absorbance at 280 nm using an extinction coefficient of 30020 cm⁻¹ M⁻¹. Enzyme was incubated at 30 °C with various ketoacids in 100 mM KCl, 2 mM MgCl₂, 100 mM Tris buffer pH 8.0 in a total volume of 50 or 100 μ l. Acetyl-CoA was added to 1 mM to start the reactions. After 20 min, reactions were stopped by adding 3 assay volumes of ethanol. Thereafter, 2 assay volumes of 1 mM DTNB in 100 mM Tris buffer,

pH 8.0 was added and the resulting yellow color was measured at 412 nm. All values were baseline-corrected using blank lacking the respective ketoacids in each assay. A molar extinction coefficient of $13600\text{ cm}^{-1}\text{ M}^{-1}$ was used in these calculations.

Quantum Mechanical Calculations. All calculations were performed with Gaussian09²⁹ with a hybrid density functional (B3LYP) with 6-31G(d) basis set. Optimization and frequency calculations were done in solution using the CPCM model for ethylphenylether, $\epsilon=4.1797$. This solvent was chosen for its dielectric constant, which is close to that of the inside of a protein.³⁰ A model for the C-C bond formation step in LeuA was used to determine the activation free energy for this step with a variety of 2-ketoacids. This model includes 1-(methylthio)ethenolate³¹ as a model for acetyl-CoA attacking the 2-ketoacids (Figure 5.2).

Protein Modeling Using RosettaDesign. Initial protein-substrate complex models were made by superimposing geometry-optimized (resembling the transition state for carbon-carbon bond formation) substrates containing various R groups (alkyl or aromatic) onto the ketoisovalerate substrate present in the *Mt*LeuA crystal structure (PDB 1SR9). Amino acid side chains in the vicinity of these superimposed substrates were then re-oriented using low energy rotamers to minimize unfavorable steric interactions with the substrates. Full optimization of protein packing and active site residue conformations was then carried out to generate one energy-minimized protein-substrate model for each substrate examined. The substrate coordinates were held fixed for this first step. Next, mutations were made in the initial models (H167A, S216G, N150G, P252A) to generate a mutant protein-substrate complex model for each substrate. Fifty independent energy minimized structures of the *G462D Ec*LeuA and *H97A/S139G/N167G/P169A/G462D Ec*LeuA mutant protein-substrate complexes were then calculated. Ligand translational movement and protein side chain torsional movement were

allowed during this phase. The top 20 models with the lowest computed ligand energies were used as seeds for final optimization of protein packing and stability with protein backbone coordinates kept fixed. Four final models were calculated from each seed model with force constants (300 and 100 for distance and angle restraints, respectively) applied to the substrates to maintain their position with respect to the residues involved in coordination of the metal ion and the substrate. The best model was selected based on the calculated ligand binding score and energy score, with a more negative energy score and smaller ligand binding score denoting a more stable complex. Protein-substrate models are presented in Supplemental Information. Solvent accessible surface area (SASA) values for each atom of each substrate were calculated by RosettaDesign. These atomic SASA values were then added together to determine the SASA of each substrate.

Enzyme Active Site and Substrate Volume Calculations. Models of the *EcLeuA* mutants were generated based on the 1SR9 PDB structure, which was defined as *G462D EcLeuA*. PyMol was used to mutate the active site residues (His167, Ser216, Asn150, Pro252) to their backbone-dependent rotamers. Only a single rotamer was available for each mutated residue, removing any ambiguity in side chain orientation. *G462D* and additional mutant models were analyzed using the Computed Atlas of Surface Topography of proteins (CASTp) server (<http://sts.bioengr.uic.edu/castp/index.php>) to determine the approximate solvent accessible volume of the active site in each model. A probe radius of 1.4 Å was used for these calculations. Substrate van der Waals molecular volumes were calculated using chemicalize.org (<http://www.chemicalize.org>, accessed August 2011) by ChemAxon (<http://www.chemaxon.com>).

References

-
- ¹ Marcheschi, R. J.; Zhang, K.; Noey, E. L.; Kim, S.; Li, H.; Chauby, A.; Houk, K. N. Liao, J. C. *ACS Chem. Biol.* **2012**, *7*, 689.
- ² (a) Atsumi, S.; Hanai, T.; Liao, J. C. *Nature* **2008**, *451*, 86. (b) Steen, E. J. et al. *Nature* **2010**, *463*, 559. (c) Schirmer, A.; Rude, M. A.; Li, X.; Popova, E.; del Cardayre, S. B. *Science* **2010**, *329*, 559. (d) Stephanopoulos, G. *Science* **2007**, *315*, 801.
- ³ (a) Causey, T. B.; Zhou, S.; Shanmugam, K. T.; Ingram, L. O. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 825. (b) Lin, H.; Bennett, G. N.; San, K. Y. *Metab. Eng.* **2005**, *7*, 116. (c) Zha, W.; Shao, Z.; Frost, J. W.; Zhao, H. *J. Am. Chem. Soc.* **2004**, *126*, 4534.
- ⁴ Cane, D. E.; Walsh, C. T. & Khosla, C. *Science* **1998**, *282*, 63.
- ⁵ McCoy, M.; Reisch, M. S.; Tullo, A. H.; Tremblay, J. F.; Voith, M. *Chemical & Engineering News* **2010**, *88*, 54.
- ⁶ Zhang, K.; Sawaya, M. R.; Eisenberg, D. S.; Liao, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20653.
- ⁷ Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*. 5th edn, (W.H. Freeman, **2002**).
- ⁸ Stieglitz, B. I.; Calvo, J. M. *J. Bacteriol.* **1974**, *118*, 935.
- ⁹ Cole, F. E.; Kalyanpur, M. G.; Stevens, C. M. *Biochemistry* **1973**, *12*, 3346.
- ¹⁰ Burns, R. O.; Umbarger, H. E.; Gross, S. R. *Biochemistry* **1963**, *2*, 1053.
- ¹¹ Chérest, M.; Felkin, H.; Prudent, N. *Tetrahedron Lett.* **1968**, *9*, 2199. (b) Anh, N. T.; Eisenstein, O.; Lefour, J. M.; Dau M. E. T. H. *J. Am. Chem. Soc.*, **1973**, *95*, 6146.
- ¹² Koon, N.; Squire, C. J.; Baker, E. N. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8295.
- ¹³ Gusyatiner, M. M.; Lunts, M. G.; Kozlov, Y. I.; Ivanovskaya, L. V.; Voroshilova, E. B. **2002** (Office, U. S. P., Ed.), Anjinomoto Co., Inc., Tokyo (JP), U.S.
- ¹⁴ Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. *Nucleic Acids Res.* **2006**, *34*, W116.
- ¹⁵ Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10383.
- ¹⁶ Larroy, C.; Rosario Fernandez, M.; Gonzalez, E.; Pares, X.; Biosca, J. A. *Chem. Biol. Interact.* **2003**, *143*, 229.

-
- ¹⁷ Zhang, K.; Li, H.; Cho, K. M.; Liao, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6234.
- ¹⁸ Shen, C. R.; Liao, J. C. *Metab. Eng.* **2008**, *10*, 312.
- ¹⁹ Ehrlich, F. *Ber Dtsch Chem Ges* **1907**, *40*, 1027.
- ²⁰ Drevland, R. M.; Waheed, A.; Graham, D. E. *J. Bacteriol.* **2007**, *189*, 4391.
- ²¹ Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389.
- ²² Wallon, G.; Kryger, G.; Lovett, S. T.; Oshima, T.; Ringe, D.; Petsko, G. A. *J. Mol. Biol.* **1997**, *266*, 1016.
- ²³ Miyazaki, K.; Kakinuma, K.; Terasawa, H.; Oshima, T. *FEBS Lett.* **1993**, *332*, 35.
- ²⁴ de la Plaza, M.; Fernandez de Palencia, P.; Pelaez, C.; Requena, T. *FEMS Microbiol. Lett.* **2004**, *238*, 367.
- ²⁵ Larroy, C.; Fernandez, M. R.; Gonzalez, E.; Pares, X.; Biosca, J. A. *Biochem. J.* **2002**, *361*, 163.
- ²⁶ Kellogg, B. A.; Poulter, C. D. *Curr. Opin. Chem. Biol.* **1997**, *1*, 570.
- ²⁷ Carreras, C. W.; Pieper, R.; Khosla, C. *Top. Curr. Chem.* **1997**, *188*, 85.
- ²⁸ Lutz, R.; Bujard, H. *Nucleic Acids Res.* **1997**, *25*, 1203.
- ²⁹ Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.
- ³⁰ (a) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400. (b) Simonson, T.; Brooks, C. L. *J. Am. Chem. Soc.* **1996**, *118*, 8452.
- ³¹ van der Kamp, M. W.; Perruccio, F.; Mulholland, A. J. *J. Mol. Graph. Model* **2007**, *26*, 676.

6. Transaminase Redesign for the Synthesis of Sitagliptin

Introduction

Enzymes are powerful catalysts. If they could be reliably and rationally designed to accelerate any given reaction, synthetic and materials chemistry would be revolutionized.¹ The chemical space available to enzymes is immense, such that it is nearly impossible to randomly achieve the design of an active enzyme. The probability of generating a protein with a specific function is about $1/10^{77}$ and only 1 in 10^{24} mutants of chorismate mutase is active.^{2,3}

In recent years, there have been landmark designed enzymes from the research groups of Baker, Mayo, Tawfik, Hilvert, DeGrado, our group and others.⁴ Enzymes have been designed to perform the Kemp elimination, retro-aldol, and Diels-Alder reactions. These designs were generated with the *inside-out* methodology, shown schematically in Figure 6.1, and described below. Further enzyme design projects based on this protocol have accomplished different levels of success.^{5,6} Despite the great advances, this highly multidisciplinary area of research is not yet robust, and successful designs are infrequent.

Directed evolution is a strategy used to improve the activity of an enzyme.⁷ This strategy is able to greatly improve an enzyme for a given process, and can be used to change the environment (i.e. solvent conditions, pH, temperature) in which the enzyme operates. However, directed evolution requires an enzyme with some initial activity for the reaction of interest. This strategy, combined with enzyme design, is a successful approach for generating proficient enzymes,⁵ and offers the potential of creating more highly proficient catalysts in the future.

A collaborative effort between the Baker lab and our group led to the development of the *inside-out* approach for designing artificial enzymes, which is shown in Figure 6.1. Central to this design process is the calculation of a *theozyme*, short for theoretical enzyme,⁸ which is the

ideal arrangement of catalytic side chains around the transition state (TS) of a reaction as calculated by quantum mechanics (QM). Calculating the *theozyme* is the first step in this protocol and has the highest precision of any of the steps in the design process. The *theozyme* concept was developed by Houk *et al.* in order to provide a detailed description of *de novo* designed catalytic active sites.⁸ *Theozymes* serve as models for active site structures and for the prediction of activation barriers for the theoretical active site relative to the uncatalyzed reaction in aqueous solution. These calculations provide an understanding of the chemical reaction of interest at the atomic level, by determining what functional groups provide optimal catalysis, and in which precise geometry they should be arranged to achieve it.

Pauling⁹ established the central idea of how enzymes catalyze reactions more than fifty ago, by recognizing that enzymes selectively stabilize the rate-limiting transition state of the catalyzed reaction relative to the ground state. As Hilvert and Mayo have recently reported, achieving enough precision in the designed active site with respect to the ideal one (*i.e.* the calculated *theozyme*), is crucial to attaining high levels of catalysis with artificial enzymes.^{5,10} For example, directed evolution of a computationally designed Kemp eliminase improved the active shape complementarity for productive substrate binding, and placed catalytic groups precisely to stabilize the transition state of the reaction. The evolved enzyme showed $\sim 10^9$ fold acceleration over the background reaction.⁵ This result illustrates the importance of achieving the ideal geometry for catalysis. To do this with *in silico* methods one must calculate and characterize optimal geometries for transition states and the catalytic groups, and incorporate this arrangement into a protein scaffold.

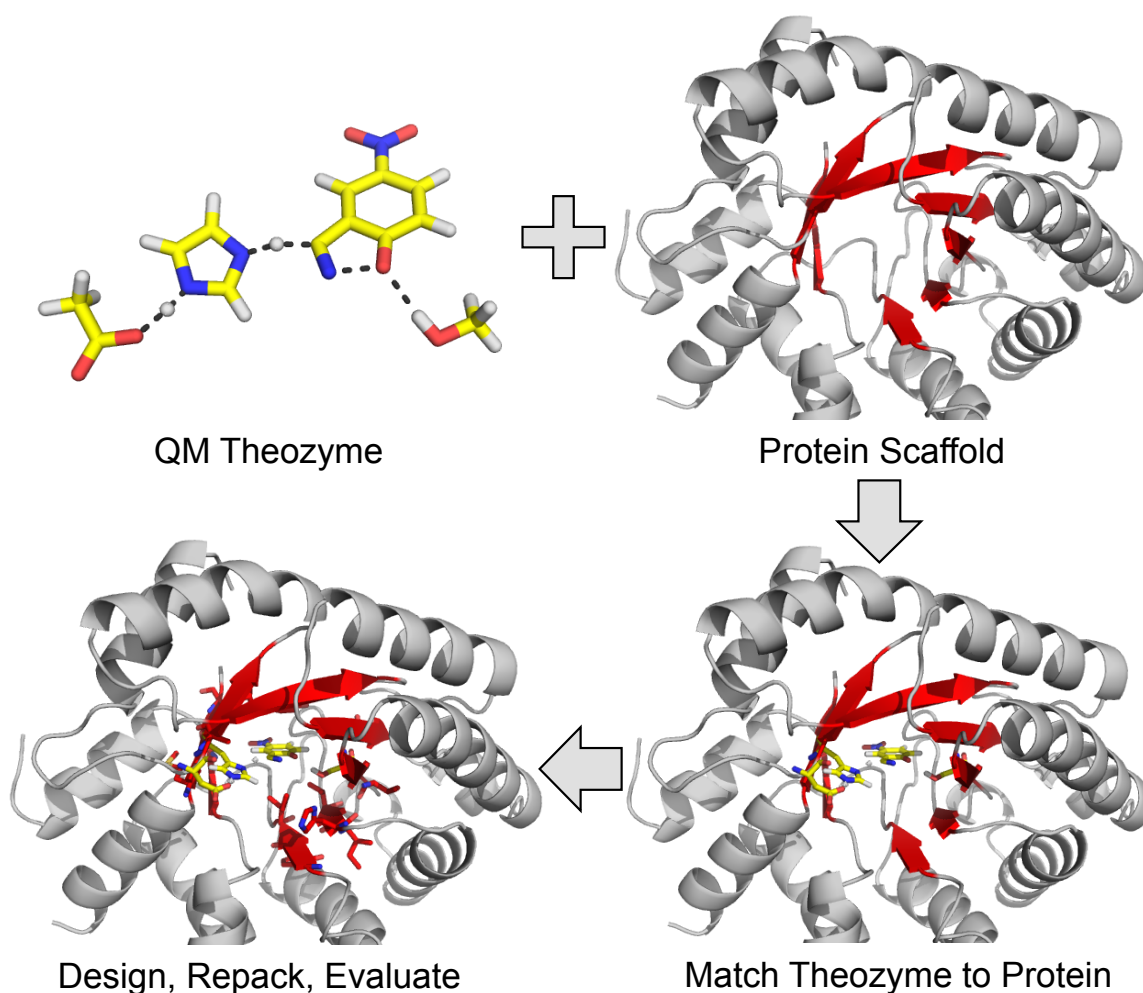


Figure 6.1. Example of the *inside-out* protocol for the Kemp elimination.

The next steps in the protocol involve incorporating the *theozyme* into protein scaffolds and stabilizing the protein with the newly incorporated active site. *Rosetta*,¹¹ a suite of programs developed by Baker to design enzymes, is one of the most reliable tools. *Rosetta* has been used in a significant and increasing number of applications.⁴ *Rosetta*'s Matcher or *SABER* (Selection of Active/Binding sites for Enzyme Redesign), which was developed in our group¹² are used to incorporate the *theozyme* into the protein. How well the match resembled the *theozyme* is then assessed and designs are ranked according to the summation of the deviations of selected distances, angles and dihedrals. In the next step of the *de novo* computational protocol, amino

acid residues surrounding the theozyme are mutated to optimize packing and stability, using *RosettaDesign*.¹³ The designed enzymes are then screened using Molecular Dynamics simulations.

Coworkers, Drs. Sílvia Osuna, and Gonzalo Jiménez-Osés, and I took on the challenge of designing a transaminase for the synthesis of sitagliptin. The transamination reaction from the pro-sitagliptin ketone, **1**, to sitagliptin, **2**, is shown in Figure 6.2. Codexis accomplished this feat by directed evolution, starting from a truncated substrate, **3**, and resulted in an enzyme that formed sitagliptin in 99.9 % *ee*.¹⁴ The enzyme they generated is used industrially. In 2010 this process won a Presidential Green Chemistry Challenge award.¹⁵ Our goal in the design was to create an enzyme that could accomplish the transamination on the pro-sitagliptin ketone, **1**.

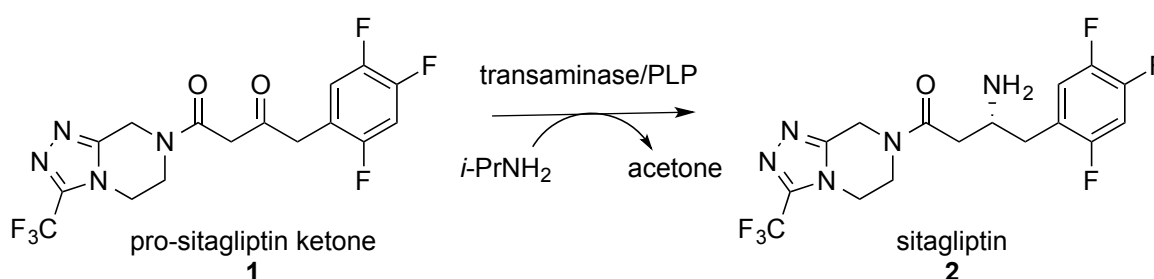


Figure 6.2. The transamination reaction of the pro-sitagliptin ketone, **1**, to form sitagliptin, **2**.

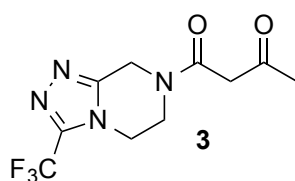


Figure 6.3. Truncated substrate used in Codexis' directed evolution process.

Since Nature already does transamination chemistry, we decided to redesign a natural transaminase to accept the much larger than natural, pro-sitagliptin ketone. This redesign project is seemingly simpler than the *de novo* enzyme design projects discussed earlier and several steps of the *inside-out* protocol can be skipped. Here the *theozyme* is still calculated, but does not need

to be incorporated into a protein scaffold. We saw the major challenges as fitting the substrate into the active site without destroying the transaminase catalytic machinery.

Codexis' evolution involved a transaminase from fold-type IV of pyridoxal phosphate (PLP)-dependent enzymes,¹² and Drs. Sílvia Osuna, and Gonzalo Jiménez-Osés, redesigned two backbones from this fold type. Here the redesign of a fold-type I transaminase is described. This backbone was chosen in order to produce a novel design, and because it has the most literature precedent of the fold-types. Fold-type I contains transaminases, which are not present in all fold-types, and has the most structures in the protein data bank (pdb).¹⁶ At the time of the design there were 223 fold-type I transaminase structures deposited in the pdb, which provided various primary sequences, backbone conformations, and bound ligands. The starting structure for the design was an *E. coli* aspartate aminotransferase crystal structure, 1ARS.¹⁷ An overlay of the active sites from the aspartate aminotransferases in the pdb is shown in Figure 6.4. This crystal structure, 1ARS, is a dimer and contains the PLP covalently bound to Lys258 in the active site, which is the resting state for this enzyme.

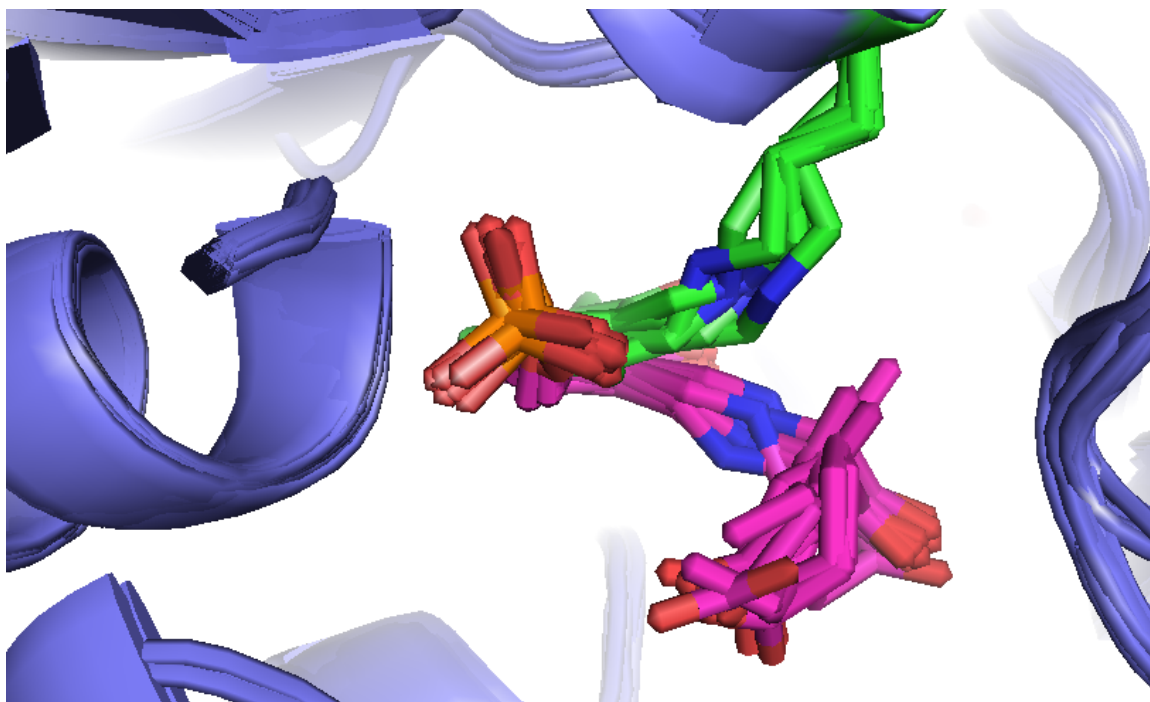


Figure 6.4. An overlay of the active sites from the aspartate aminotransferases in the pdb. The covalent PLP-Lys intermediate is shown in green, and other PLP derivative ligands are shown in pink.

The amino acids involved in catalysis and in binding the PLP cofactor were identified and conserved through the design process, and are shown in Table 6.1. The residues responsible for binding are shown in Figure 6.5. Residues within 5 Å from the cofactor, excluding the ones responsible for binding PLP, were selected for mutation in the design process. Twelve residues met these criteria (15, 17, 18, 39, 142, 360, 382, 386, 69, 292, 296, and 194; shown in Table 6.2).

by overlaying the pyridoxal ring in the conformer with the one in the crystal structure. The conformers that clashed with the backbone were eliminated. The library of conformers of **5** is shown in Figure 6.7. Figure 6.7b shows these conformers in the active site of the enzyme.

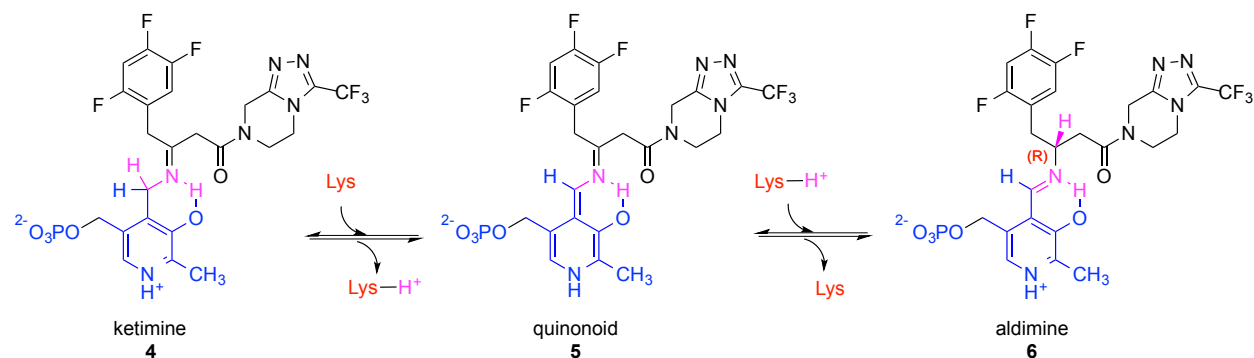


Figure 6.6. The ketimine, **4**, quinonoid, **5**, and aldimine, **6**, are the three covalent intermediates of the pro-sitagliptin ketone with PLP.

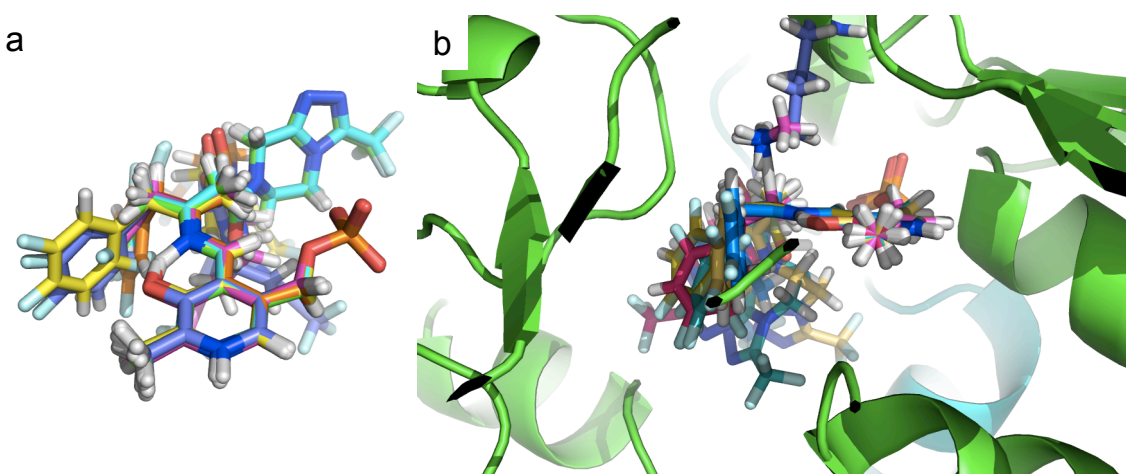


Figure 6.7. The library of conformers of **5** used in the redesign (a) overlaid. (b) docked into the transaminase active site.

RosettaDesign was then used to generate a mutant library. The input for *RosettaDesign* was the protein and a library of the selected conformations of the intermediates. *RosettaDesign* was allowed to mutate the selected residues in the active site (Table 6.2). Each design was ranked and filtered based on energetic scores generated by the *RosettaDesign* program, and compared to that of the wild type (WT) with the natural substrate. These rankings are based on a score for the total energy of the enzyme (total score), and the energy of ligand binding (ligand

binding score), where negative numbers are more stable. The cutoff for the total score is -127, which is within 61 % (80 units) of the stability of the WT. No designs were predicted to be more stable than the WT with its natural substrate. The total charge of the enzyme of the enzyme was also considered, where only enzymes with a charge within 4 units of the WT were kept. In order to prevent elimination of good designs that were not sufficiently optimized, eight conformations of each design were generated with Rosetta, each conformation was scored, and the best score was used to rank and filter the designs. All designs that scored worse than the cutoff in all eight Rosetta optimizations were eliminated.

The best designs were then subjected to all-atom AMBER²⁰ molecular dynamic simulations with explicit water for 20 ns. The simulations of the designs were compared to the simulations of the WT with intermediates of the natural substrate. The intermediates of the natural substrate are shown in Figure 6.8. The catalytic and binding distances that were monitored are shown in Figure 6.9. In some simulations the dimeric enzyme separated into two monomers, which indicated that the protein was unstable as a dimer, and would likely not be a functioning enzyme. Designs were eliminated if the catalytic and binding distances exceeded 1 Å of the distances sampled in the WT simulation, or if the dimer separated.

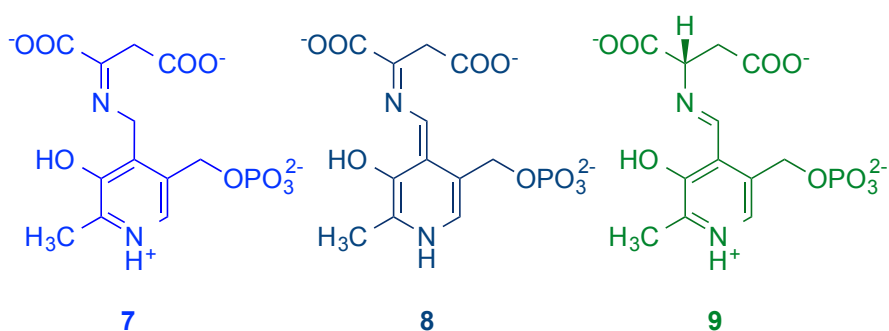


Figure 6.8 PLP-aspartate ketimine, **7**, quinonoid, **8**, and aldimine, **9**, intermediates color-coded to match the plots in Figure 6.11.

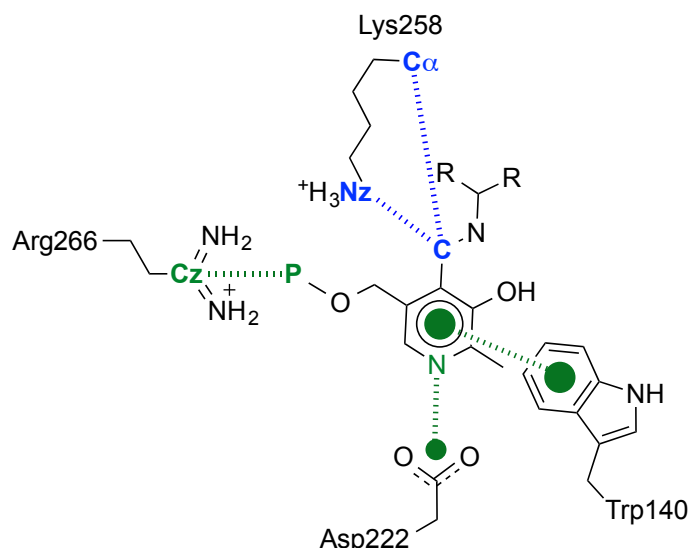


Figure 6.9. Heavy atom distances monitored in the MD simulations to assess the catalytic potential of the enzyme. The catalytic distances (Lys258 Nz – substrate C, Lys258 Ca - substrate C) are shown in blue, and the binding contacts (Asp222 average O – substrate N, Arg266 Cz - substrate P, and Trp140 π – substrate π) are shown in green.

The PLP-aspartate ketimine, **7**, quinonoid, **8**, and aldimine, **9**, intermediates are shown in Figure 6.9, and the distances sampled in the simulation of the WT with these intermediates are shown in Figures 6.11 and 6.12. The simulations with the WT served as reference to guide the assessment of the designs.

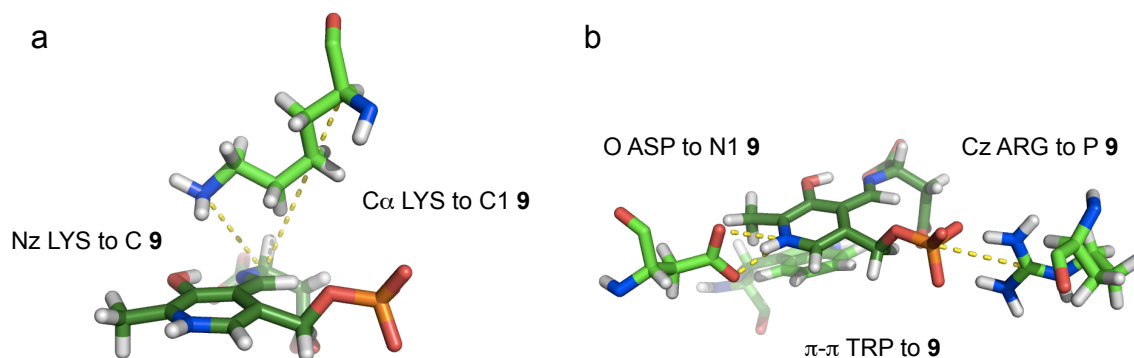


Figure 6.10. The distances monitored in the MD simulation of the WT (a) The catalytic contacts, C α and N of the Lys258 to C of **9**. (b) The binding contacts Asp222, Arg266, and Trp140 to **9**.

These simulations show that the lysine catalytic distance is not close to that in the QM TS (2.7 Å). Distances in the simulations are between 3 and 7 Å, with peaks at 3.6, 4.9, and 5.9 Å

(Figure 6.11b). It is not expected that Lys258 would sample distances as close as the TS, because this is an activated complex, not a minimum on the potential energy surface. However, distances over 4 Å were unexpected, and not close to what is needed for catalysis. These populations were a result of Lys258 interacting with the phosphate of the substrate.

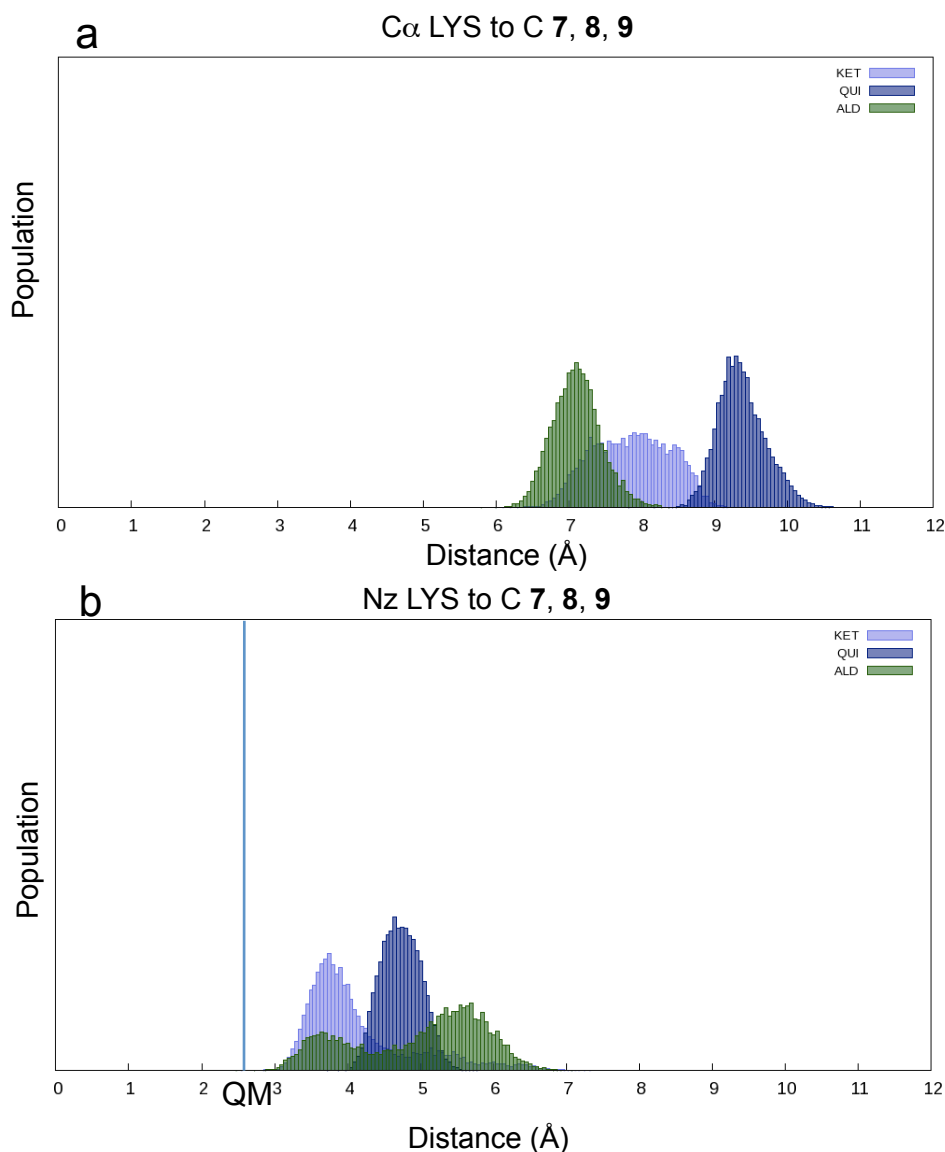


Figure 6.11. The catalytic distances sampled in the MD simulation of the WT. (a) Lys258 Cα to C of **7** (light blue), **8**, (dark blue) and **9** (green). (b) Lys258 Nz to C of **7** (light blue), **8**, (dark blue) and **9** (green). The TS distance calculated by QM is labeled.

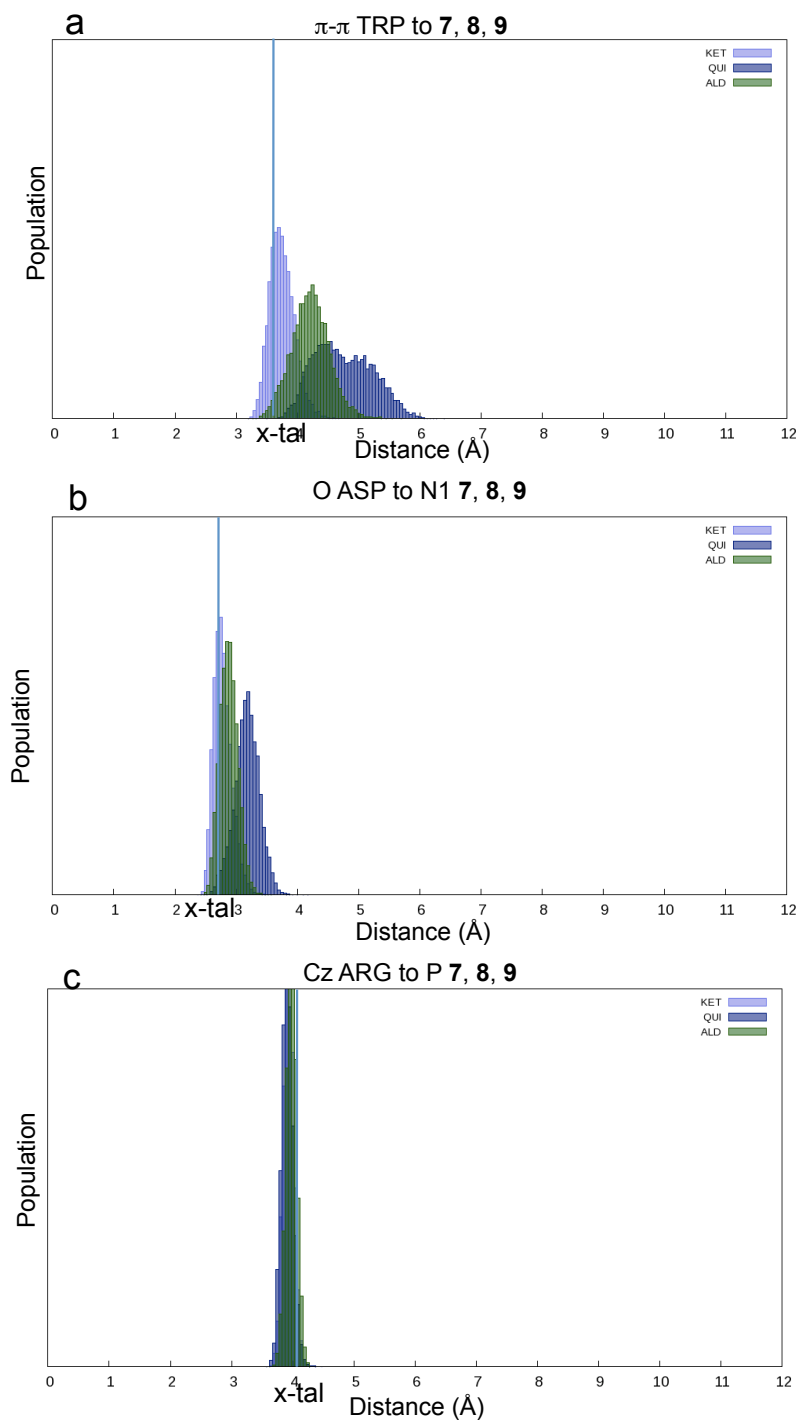


Figure 6.12. The ligand bonding contacts sampled in the MD simulation of the WT. The distance found in the crystal structure is labeled (a) π - π distance between Trp140 to **7** (light blue), **8**, (dark blue) and **9** (green). (b) Asp222 to **7** (light blue), **8**, (dark blue) and **9** (green). (c) Arg266 to phosphate of **7** (light blue), **8**, (dark blue) and **9** (green).

The results from MD simulations of the design K7 are shown in Figures 6.13 and 6.14. Figure 6.13 shows a comparison of the catalytic distances sampled in the MD simulation of the design K7 with **4** to the WT with **7**. The ligand bonding contacts sampled in the MD simulation of the design K7 are compared to that of the WT in Figure 6.14. Both the catalytic and ligand binding contacts in the simulations of the design K7 match quite well with that of the WT, as can be seen by the nice overlay of peaks in the histograms. These simulations are representative of what was observed in the MD simulations of the designs, as long as the dimer was maintained. Specifically, the catalytic and binding distances were well maintained in almost all of the designs that remained dimeric.

The designs that remained a dimer and maintained catalytic distances within 1 Å of the distances sampled in the WT simulation for the majority of the time (> 50%) were expressed experimentally. The final designs selected for experiment are shown in Table 6.2, and their Rosetta Scores are shown in Table 6.3. The total and ligand binding scores are given, along with the total charge, the root mean squared (rms) deviation from the WT, and the mutations per chain. The rms ranges from 0.13 to 0.74 and the number of mutations ranges from 3 to 10.

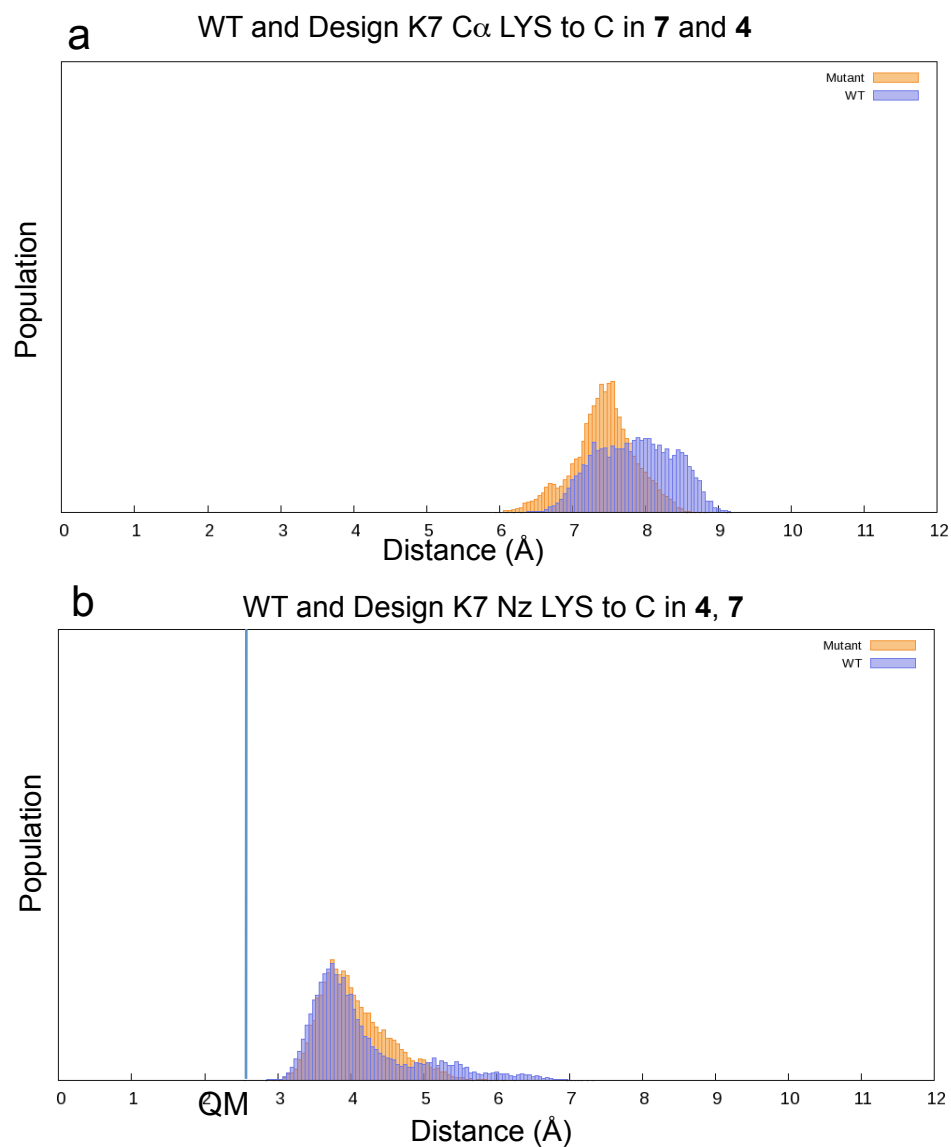


Figure 6.13. The catalytic distances sampled in the MD simulation of the design K7 (orange) compared to the WT (light blue). (a) Lys258 C α to C of the ketimine intermediates **4** in K7, and **7** in the WT. (b) Lys258 Nz to **4** in K7, and **7** in the WT.

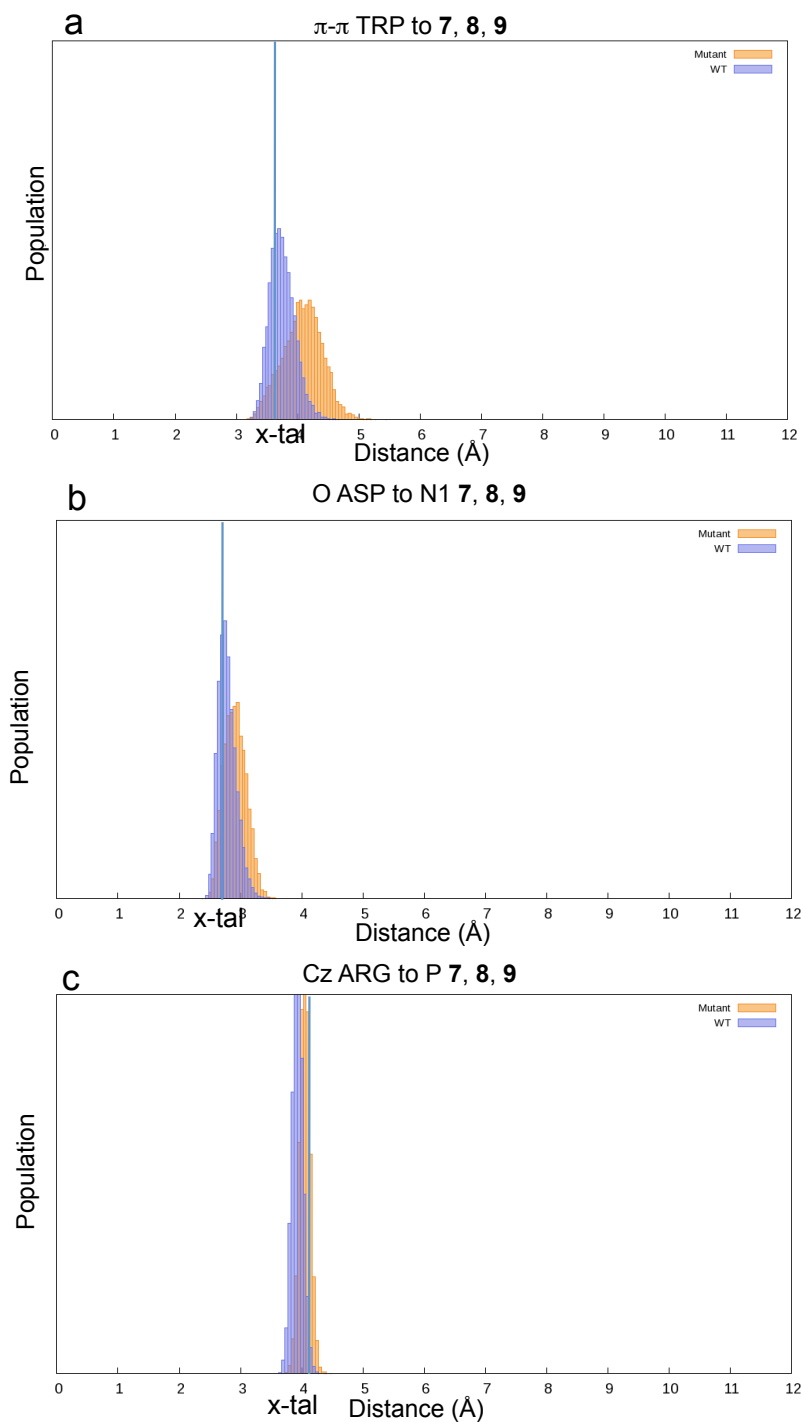


Figure 6.14. The ligand binding contacts sampled in the MD simulation of the design K7 (orange) compared to the WT (light blue). (a) π - π distance between Trp140 to **4** in K7, and **7** in the WT. (b) Asp222 to **4** in K7, and **7** in the WT. (c) Arg266 to phosphate of **4** in K7, and **7** in the WT.

Table 6.2. Final designs on which experiments were performed. The design name is given on the left, where K, Q and TS indicate the ligand used in the design process (**1**, **2**, or the TS that connects **1** and **2**, respectively). The mutations at each residue are color-coded based on the properties of the amino acid: hydrophobic (grey), negatively charged (red), positively charged (blue), polar (light blue), containing hydroxyl (orange), containing amide (gold), containing sulfur (yellow), tyrosine (pink). Residues conserved from the WT are in bold.

Design	Residue											
	15	17	18	39	142	360	382	386	69	292	296	194
WT	D	I	L	V	N	F	V	R	N	R	S	N
K2	D	A	A	S	N	A	V	A	N	R	S	S
K7	H	A	E	T	N	L	H	A	N	A	S	S
Q3	D	I	R	T	N	I	A	A	A	A	R	N
Q4	D	A	R	T	N	I	A	A	A	A	R	N
Q5	H	L	R	T	N	A	A	V	N	A	G	N
Q7	H	L	R	V	N	A	A	V	N	A	G	N
Q8	H	L	R	T	N	A	A	V	N	A	S	N
Q9	H	L	R	T	N	A	A	V	N	A	S	S
Q11	H	H	N	D	N	T	N	V	N	A	G	S
Q12	H	I	N	T	N	T	L	V	N	A	G	S
Q14	H	H	R	T	N	V	H	A	S	A	G	N
Q16	H	T	R	T	N	T	H	V	S	A	G	N
Q17	D	T	L	V	N	A	V	V	N	A	G	N
Q18	H	T	L	V	N	T	V	V	N	A	G	N
Q21	H	L	L	V	N	A	V	V	N	A	G	N
Q22	H	V	L	V	N	A	V	V	N	A	G	N
Q23	H	T	L	V	N	A	V	V	N	A	G	N
Q24	D	T	L	V	N	T	V	V	N	A	G	N
Q30	H	Y	R	V	N	M	V	A	N	A	A	S
Q32	H	I	R	V	N	A	H	V	N	A	S	S
Q33	D	H	L	V	N	V	H	A	N	K	S	S
TS112	D	I	E	V	R	H	V	V	N	A	G	N
TS116	H	I	R	V	A	L	V	V	N	A	G	N
TS12	H	I	R	V	A	H	V	V	N	A	G	N
TS122	H	I	R	V	A	A	V	V	N	A	G	N
TS124	H	I	R	V	A	T	V	V	N	A	G	N
TS127	H	I	E	V	A	L	V	V	N	A	G	N
TS129	D	I	N	V	R	Y	V	V	N	A	G	N
TS132	H	I	L	V	A	H	V	V	N	A	G	N
TS15	H	I	R	V	A	M	V	V	N	A	G	N
TS23	H	Y	L	V	A	H	V	V	N	A	G	N
TS26	H	I	L	V	A	L	V	V	N	A	G	N
TS27	H	Y	L	V	H	T	V	V	N	A	G	N
TS229	H	R	L	V	A	H	V	V	N	A	G	N
TS230	D	Y	L	V	K	C	V	V	N	A	G	N
TS232	H	Y	L	V	A	R	V	V	N	A	G	N
TS319	D	I	L	V	N	R	V	V	N	R	G	N

TS331	D	I	L	V	N	H	V	V	N	R	G	N
TS56	H	I	L	V	N	L	V	V	N	A	G	N
TS513	H	I	L	V	N	C	V	V	N	A	G	N

Table 6.2 continued

Design	Residue											
	15	17	18	39	142	360	382	386	69	292	296	194
TS526	H	I	L	V	N	T	V	V	N	A	G	N
TS528	D	I	L	V	N	V	V	A	N	A	G	N
TS535	H	I	L	V	N	A	V	V	N	A	G	N

Table 6.3. Rosetta scores for the final designs.

Design	Rosetta Scores				
	Total Score	Total Charge	Ligand Binding	RMS	Mutations
WT	-207	-14	-7.8	0	0
K2	-133	-16	-3.3	0.58	6
K7	-156	-18	-3.1	0.47	8
Q3	-152	-14	0.0	0.35	8
Q4	-128	-14	-9.0	0.28	9
Q5	-153	-14	-2.9	0.34	9
Q7	-182	-14	-6.6	0.33	8
Q8	-155	-14	-5.6	0.36	8
Q9	-156	-14	-6.7	0.50	9
Q11	-168	-18	-5.2	0.17	10
Q12	-176	-16	-3.0	0.25	9
Q14	-149	-14	-6.3	0.13	10
Q16	-162	-14	-5.1	0.24	10
Q17	-159	-18	-0.6	0.31	5
Q18	-157	-16	-1.4	0.26	6
Q21	-127	-16	-3.4	0.31	6
Q22	-160	-16	-4.1	0.32	6
Q23	-127	-16	-3.8	0.36	6
Q24	-151	-18	-0.3	0.17	5
Q30	-145	-14	-4.8	0.42	8
Q32	-161	-14	-7.0	0.73	7
Q33	-167	-16	-1.5	0.54	7
TS112	-159	-18	-2.2	0.38	6
TS116	-165	-14	-10.1	0.48	7
TS12	-166	-14	-6.5	0.18	7
TS122	-163	-14	-1.8	0.30	7
TS124	-154	-14	-9.4	0.36	7
TS127	-167	-18	-4.8	0.44	7
TS129	-146	-16	-3.4	0.22	6
TS132	-168	-16	-4.7	0.22	6
TS15	-194	-14	-1.0	0.48	7
TS23	-158	-16	-2.6	0.40	7

TS26	-164	-16	-0.9	0.31	6
TS27	-197	-16	-4.2	0.74	7
TS229	-196	-14	-4.7	0.42	7
TS230	-156	-16	-0.7	0.51	6
TS232	-149	-14	-4.4	0.28	7
TS319	-193	-14	-3.7	0.53	3

Table 6.3 continued

Design	Rosetta Scores				
	Total Score	Total Charge	Ligand Binding	RMS	Mutations
TS331	-193	-16	-3.8	0.54	3
TS56	-174	-16	-3.4	0.39	5
TS513	-160	-16	-6.2	0.33	5
TS526	-163	-16	-3.6	0.59	5
TS528	-167	-18	-2.6	0.26	4
TS535	-205	-16	-1.9	0.35	5

Codexis scientists then expressed the final designs and tested them for transaminase activity toward **1**. Additionally, a library of mutants that were generated by the random combination of mutations from the designs was tested for activity. The genes were cloned into expression vector pCK110900²¹ under the control of a lac promoter. The vector was transformed into *E. coli* W3110, where the plasmid DNA was prepared using standard methods. The activity experiments were run under two conditions. A transaminase with established activity toward acetophenone was included as a positive control. The first set of conditions included 100 mL lysate, 2 g/L pro-sitagliptin ketone, **1**, 5 % methanol, 500 mM isopropylamine in 100 mM TEA buffer at pH 7.5. The reaction was shaken for 20 h at room temperature. High-pressure liquid chromatography (HPLC) was used to analyze the samples. The second set of conditions included 100 mL lysate, 2 g/L acetophenone, 5 % DMSO, 1 g/L PLP, 500 mM isopropylamine in 100 mM TEA buffer pH 7.5. The reaction was shaken for 20 h at room temperature. Gas chromatography (GC) was used to analyze the samples.

None of these designs were found to be active under Codexis' library screening process. It is unclear why none of the proteins were active toward **1**. However, this results suggests that

there are factors beyond the active site size impacting the substrate scope of transaminases. It could likely be due to protein instability, lack of ligand binding, or that the enzyme was unable to catalyze any of the steps in the reaction. This failed design process has led our group to seek methods to make the design process more robust. To that end we have undertaken projects to understand successful design and directed evolution processes.^{10,22} The study of Codexis' directed evolution process is described in the next chapter.

References

¹ Nanda, V.; Koder, R. L. *Nat. Chem.* **2010**, *2*, 15.

² Axe D. D. *J. Mol. Biol.* **2004**, *341*, 1295.

³ Taylor, S. V.; Walter, K. U.; Kast, P.; Hilvert, D. *Proc. Natl. Acad. Sci.* **2001**, *98*, 10596.

⁴ (a) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190. (b) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387. (c) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St. Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* **2010**, *329*, 309. (d) Bjelic, S.; Nivón, L. G.; Çelebi-Ölçüm, N.; Kiss, G.; Rosewall, C. F.; Lovick, H. M.; Ingalls, E. L.; Gallaher, J. L.; Seetharaman, J.; Lew, S.; Montelione, G. T.; Hunt, J. F.; Michael, F. E.; Houk, K. N.; Baker, D. *ACS Chem. Biol.* **2013**, *8*, 749. (e) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovnik, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8013. (f) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. *Nature* **2013**, *503*, 418. (g) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. *Nat. Chem. Biol.* **2013**, *9*, 494. (h) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3790. (i) Khersonsky, O.; Kiss, G.; Rothlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10358. (j) Khersonsky, O.; Rothlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss, G.; Houk, K. N.; Baker, D.; Tawfik, D. S. *J. Mol. Biol.* **2011**, *407*, 391-412. (k) Khersonsky, O.; Rothlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. *J. Mol. Biol.* **2010**, *396*, 1025. (l) Kaplan, J.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11566. (m) Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; DeGrado, W. F. *Nat Chem* **2012**, *4*, 1050. (n) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew Chem, Int Ed.* **2013**, *52*, 5700.

-
- ⁵ Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. *Nature* **2013**, *503*, 418.
- ⁶ Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J. T.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.; Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D. *J Am Chem Soc* **2012**, *134*, 16197.
- ⁷ Bornsheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature*, **2012**, *485*, 185.
- ⁸ (a) Tantillo, D. J.; Chen, J.; Houk, K. N. *Curr Opin Chem Biol* **1998**, *2*, 743. (b)Ujaque G.; Tantillo, D. J.; Hu Y. F.; Houk, K.N.; Hotta, K.; Hilvert, D. *J Comput Chem* **2003**, *24*, 98. (c) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Rothlisberger, D.; Baker, D. *Protein Sci* **2006**, *15*, 2785.
- ⁹ Pauling, L. *Nature* **1948**, *161*, 707.
- ¹⁰ Jiménez-Osés, G.; Osuna, S.; Noey, E. L. Houk K. N. *J Am Chem Soc.* *Submitted*.
- ¹¹ (a) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Rothlisberger, D.; Baker, D. *Protein Sci.* **2006**, *15*, 2785.
- ¹² Nosrati, G. R.; Houk K. N. *Protein Sci.* **2012**, *21*, 697.
- ¹³ Das, R.; Baker D. *Annu Rev Biochem* **2008**, *77*, 363.
- ¹⁴ Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. *Science* **2010**, *329*, 305.
- ¹⁵ Ritter, S. K. 2010 Green Chemistry Awards. *Chemical and Engineering News*, **2010**, 88, 9. (June 23, 2010)
- ¹⁶ (a) Jansonius, J. N. *Curr. Opin. Struct. Biol.*, **1998**, *8*, 759. (b) Soda, K.; Yoshimura, T.; Esaki, N. *Chem. Rec.* **2001**, *1*, 373.
- ¹⁷ Okamoto, A., Higuchi, T., Hirotsu, K., Kuramitsu, S., Kagamiyama, H. *J. Biochem.* **1994**, *116*, 95.
- ¹⁸ Schrodinger Release 2013-1: MacroModel, version 10.0, Schrodinger, LLC, New York, **2013**.
- ¹⁹ Gaussian 09, Revision **D.01**, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;

Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

²⁰ Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12, University of California, San Francisco*. **2012**.

²¹ Davis, S. C.; Grate, J. H.; Gray, D. R.; Gruber, J. M.; Huisman, G. W.; Ma, S. K.; Newman, L. M.; Sheldon, R.; Wang, L. A. US patent 7,132,267, **2006**.

²² (a) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk K. N. *Nature Chem Biol*, **2014**, *10*, 431. (b) Noey, E. L.; Tibrewal, N.; Park, J.; Bond, C.; Cascio, D.; Osuna, S.; Jiménez-Osés, G.; Liang, J.; Zhang, X.; Huisman, G.; Tang, Y. Houk, K. N. *Proc. Natl. Acad. Sci. Submitted*.

7. Toward Understanding the Activity of the Sitagliptin Transaminase

Transaminases are pyridoxal-5'-phosphate (PLP) dependent enzymes that transaminate prochiral ketones stereoselectively as shown in Figure 7.1. Natural transaminases are involved in many pathways including the metabolism of amino acids and the biosynthesis of antibiotic compounds.¹ This transformation proceeds through covalent intermediates between the substrate and cofactor and requires a sacrificial amine donor.

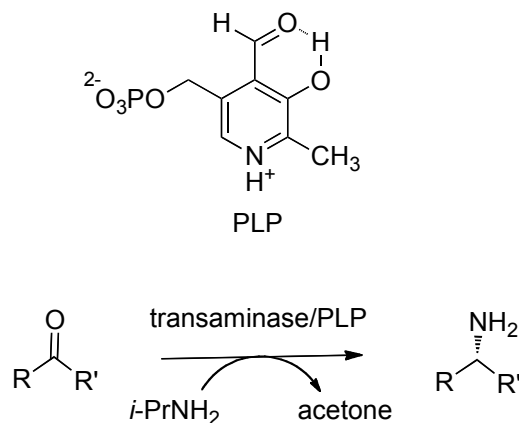


Figure 7.1. Transaminases transform prochiral ketones into chiral amines, with a sacrificial amine donor.

Transaminases have been evolved for industrial-scale synthesis of chiral amines for pharmaceuticals,² such as (S)-rivastigmine,³ (S)-repaglinide,⁴ (R)-levocetirizine,⁵ sitagliptin⁶ and suvorexant.⁷ The latter two were synthesized with the enzyme that Codexis evolved for sitagliptin.⁶ Sitagliptin (Januvia) is a diabetes drug, and is one of Merck & Co. best-selling drugs.⁸ The evolved enzyme replaced an asymmetric hydrogenation step⁹ in the industrial synthesis of sitagliptin, which required rhodium and hydrogen under high pressure. This effectively lowered the cost of production by eliminating heavy metal waste, and avoiding the necessity of carrying out the reaction with specialized high-pressure equipment. This process

earned a Presidential Green Chemistry Challenge Award in 2010.¹⁰ The transamination of the pro-sitagliptin ketone, **1**, to sitagliptin, **2**, is shown in Figure 7.2.

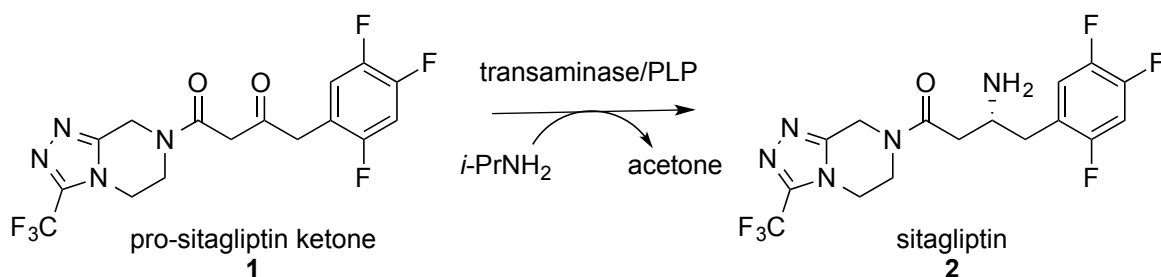


Figure 7.2. The transamination reaction of the pro-sitagliptin ketone, **1**, to form sitagliptin, **2**.

The transaminase for sitagliptin was evolved from ATA-117, which is a non-natural homolog of a transaminase from *Arthrobacter* sp. KNK168¹¹ and is homologous to D-amino acid aminotransferases and branched-chain amino acid transferases, which are in fold-class IV of PLP-dependent enzymes.¹² ATA-117 and the evolved enzymes are homodimers in solution, with the active site located at the protein-protein interface.⁶ The cofactor binds with the *re* face toward the catalytic lysine, accounting for its stereospecificity.¹³ Directed evolution improved activity and enables the enzyme to operate under more challenging conditions, which are necessary for industrial synthesis (i.e. increased temperatures, pH, concentration of methanol or dimethyl sulfoxide (DMSO) cosolvents, isopropylamine (IPA), and substrate).⁶ Here, IPA is the sacrificial amine source. A substrate scope has been assessed for ATA-117^{14,6} and for the variant from the final round of the published evolution,⁶ which is shown in Figure 7.3. The commercial variant has three additional mutations in addition to those in the final published variant. To our knowledge, the commercial variant's substrate scope has not been assessed. ATA-117 is unable to convert any substrates with the small ketone substituent larger than a methyl, unless it is part

of a ring, as shown in Figure 7.3. In line with its observed substrate scope, ATA-117 is inactive towards the pro-sitagliptin ketone, **1**. Codexis used the truncated ketone substrate that forms **3**. The smaller substituent, **3**, lacks the trifluorophenyl substituent.

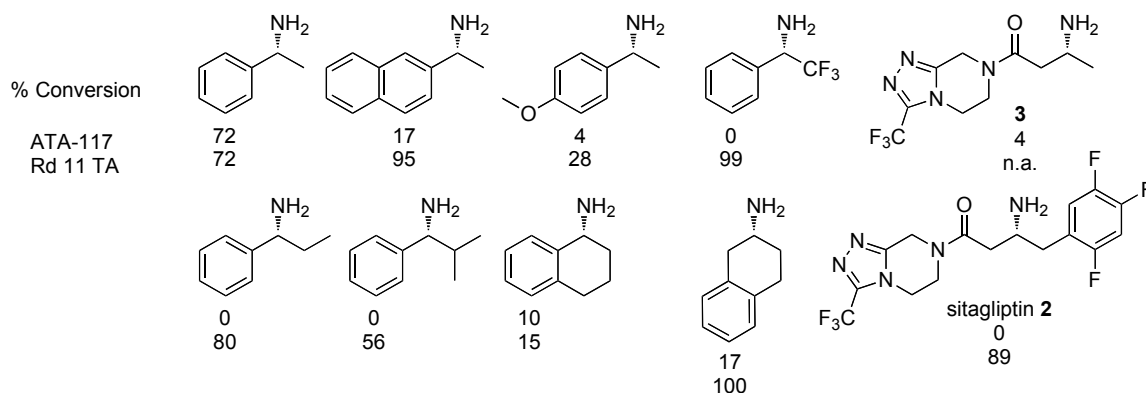


Figure 7.3. Substrate scope and percent conversion for ATA-117, and the round 11 (final published round) mutant in Codexis' directed evolution.

Our study began with an attempt to computationally redesign an enzyme for the formation of sitagliptin. In this study, Drs. Sílvia Osuna, and Gonzalo Jiménez-Osés and I redesigned three separate backbones from two different fold-classes of transaminases. This redesign project had the apparent task of making space in the active site for the large pro-sitagliptin ketone substrate, **1**, so the redesign focused exclusively on the active site of the enzyme. However, all the designs failed to lead to an active catalyst, as detailed in Chapter 6. This suggests that there are factors beyond the active site size impacting the substrate scope of transaminases.

A grand goal in our group is to reliably design enzymes, and our objective here is to advance enzyme design strategies. The transaminase redesign and other projects illuminate the fact that there is much we still do not know about enzyme catalysis. Here we computationally study Codexis' successful generation of a transaminase used to synthesize sitagliptin on the industrial scale with the idea that understanding how activity was created in this enzyme will

inform us on how to design enzymes. Specifically, we study the role of the mutations in the directed evolution, and probe why the dimer and not the monomer is active.

The multi-step transamination reaction involves three intermediates of the covalently bound pro-sitagliptin ketone, **1**, to PLP. These intermediates are the ketimine, **4**, quinonoid, **5**, and aldimine, **6**, as shown in Figure 7.4.

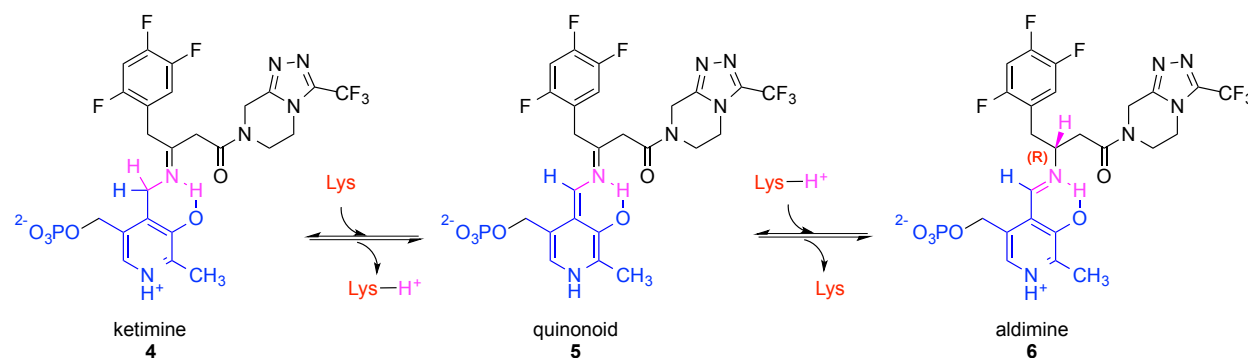


Figure 7.4. The ketimine, quinonoid, and aldimine are the three covalent intermediates of the pro-sitagliptin ketone with PLP in its transformation to an amine.

Methods

The catalytic cycle for the transamination¹⁵ of the pro-sitagliptin ketone was modeled quantum mechanically with Gaussian09.¹⁶ Lysine was modeled with methylamine/methylammonium, and the phosphate was removed from the cofactor to avoid complications in geometry optimization due to the densely charged nature of this moiety. The stationary points were optimized with B3LYP/6-31G(d), and single point energies on these geometries were calculated with M06-2X/6-311+G(d,p). The single point energies were calculated using a polarized continuum model (CPCM) with water (a dielectric constant, $\epsilon = 78$) and with $\epsilon = 4$,¹⁷ which gave similar energies.

All-atom, molecular dynamics (MD) simulations with explicit water with ATA-117 and Codexis' commercial variant were run on ANTON.¹⁸ The ATA-117 structure is a homology model built from the crystal structure of the commercial variant. The systems were prepared using AMBER,¹⁹ where the substrate-bound enzymes were immersed in a cuboidal box of TIP3P²⁰ explicit waters. The parameters for the substrate and cofactor were generated from optimizations with HF/6-31G(d) and processed by the antechamber module in Amber12 using the general Amber force field (GAFF), to provide atomic partial charges. Explicit counter ions (Na⁺, Cl⁻) were added to neutralize the protein. The protein and surrounding water environment was minimized and equilibrated, each in two stages. In the first minimization stage only the positions of solvent molecules and ions are minimized, and in the second of all the atoms in the simulation are minimized. In the first equilibration the systems were heated by increasing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions over 1 ns with a 2 fs time step. Harmonic restraints of 10 kcal/mol were applied to the substrate and cofactor. The second equilibration ran for 10 ns at a constant volume and temperature of 300 K. All simulations were run on ANTON for between 2 - 3 μ s. The commercial variant was simulated as a monomer and as a dimer with the sitagliptin quinonoid, **5**, in the active site. ATA-117 was simulated as a dimer with **5**, and with the quinonoid formed from acetophenone, **16**. All dimers have the internal aldimine in the second active site.

Results

The quantum mechanically calculated thermal and free energies calculated with the two different solvation models are shown in Table 7.1. The catalytic cycle showing the calculated free energies with the dielectric equal to 4 are shown in Figure 7.5. In Figure 7.5 the ketimine,

quinonoid, aldimine, and pyridoxamine intermediates (**9-15**) were modeled with lysine (methylamine/methylammonium) coordinated to the structure. Figure 7.6 shows the optimized ketimine, **13**, quinonoid, **14**, aldimine, **15**, and the transition structures that connect them, **TS13-14**, and **TS14-15**. Energies of the intermediates with and without lysine coordinated are included in Table 7.1.

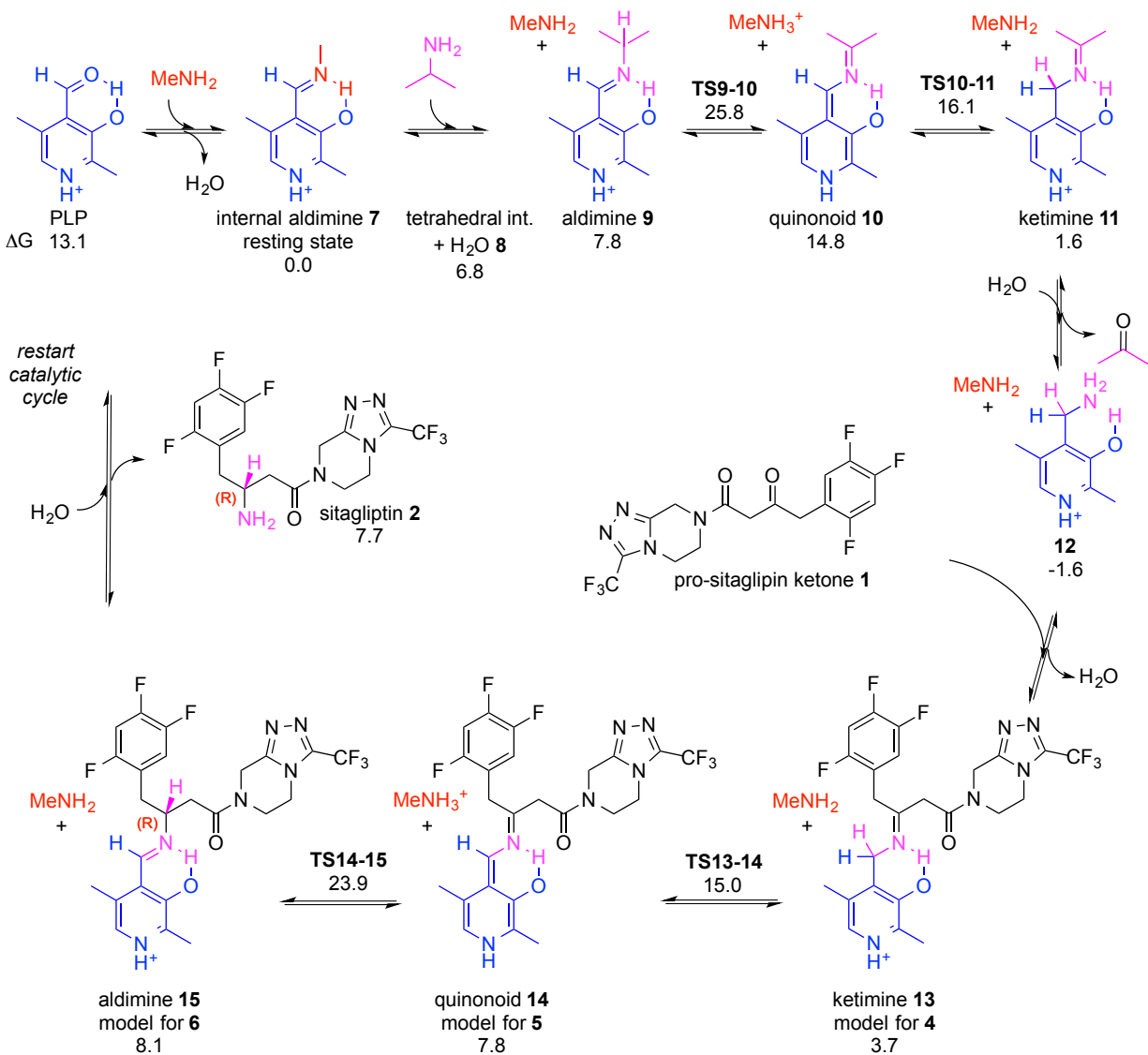


Figure 7.5. Modeled transaminase catalytic cycle. Free energies are given in kcal/mol (M06-2X/6-311+G(d,p) CPCM $\epsilon=4$ /B3LYP/6-31G(d)).

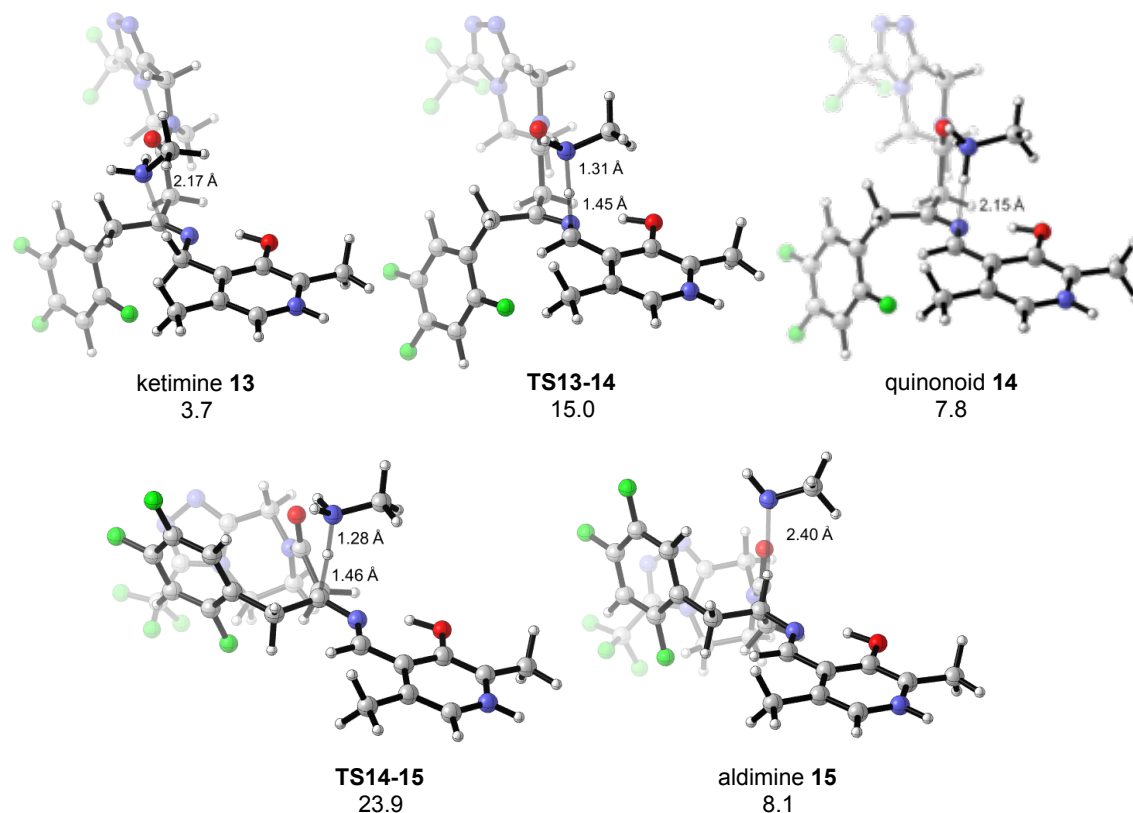


Figure 7.6. Optimized PLP-sitagliptin intermediates with lysine (methylamine/methylammonium). Free energies are given in kcal/mol (M06-2X/6-311+G(d,p) CPCM $\epsilon=4$ /B3LYP/6-31G(d)).

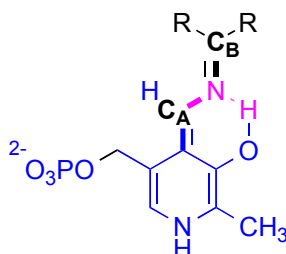
As shown by the catalytic cycle in Figure 7.5, the internal aldimine, **7**, in which PLP is covalently bound to a lysine in the active site, is the resting state in the natural reaction.¹⁰ The energy of **7** was set to 0.0 kcal/mol, to which all energies are compared. The formation of the tetrahedral intermediate, **8**, from the attack of IPA on the internal aldimine, is endothermic by 6.8 kcal/mol. Release of the lysine from the tetrahedral intermediate forms the aldimine, **9**. The aldimine, **9**, is higher in energy than the internal aldimine due to the entropic cost of coordinating **9** to methylamine. Deprotonation of the aldimine is facilitated by conjugation with the pyridine ring,^{13d} and is calculated to be the rate-determining step with a barrier of 25.8 kcal/mol from the resting state. The barrier is overestimated due the difficulty of calculating entropy inside the

enzyme; the thermal energy of this transition state, **TS9-10**, is 15.1 kcal/mol. This step forms the quinonoid, **10**, which is 14.8 kcal/mol, which is less stable than **9** and **11** due to the loss of aromaticity. The subsequent protonation has a very low barrier, 1.3 kcal/mol from **10**, and is exothermic, forming the ketimine intermediate, **11**. Like the deprotonation before, the protonation step is stabilized by conjugation, and requires a planar geometry about the CCNC bond.¹⁵ The CCNC bond is bolded in Figure 7.7. Hydrolysis of the ketimine releases acetone, and forms the PLP-free amine, **12**. Condensation of the pro-sitagliptin ketone, **1**, onto **12**, forms the ketimine, **13**, with an energy of 3.7 kcal/mol. Analogous to the previous steps, protonation and then deprotonation to form the quinonoid, **14**, and then the aldimine, **15**, occur. The ketimine, **13**, is the most stable of the covalent PLP-sitagliptin intermediates at 3.7 kcal/mol, which is 4.1 and 4.4 kcal/mol more stable than **14** and **15**, respectively. The deprotonation and protonation are endothermic. The deprotonation of the ketimine requires rotation so that the C-H bond is perpendicular to the pyridine ring, which costs 1.1 kcal/mol (compare of **13** to **13r**, Table 7.1). In **TS13-14**, the breaking C-H and forming N-H bonds are 1.45 and 1.31 Å, respectively. The protonation to form the aldimine, **TS14-15**, is the stereoselectivity-determining step and occurs with a barrier of 16.1 kcal/mol (23.9 kcal/mol from the resting state). This transition state is early; the forming C-H and breaking N-H bonds are 1.46 and 1.28 Å, respectively. The C-N distance in the transition states for deprotonation and protonation (**TS13-14** and **TS14-15**) are 2.72 and 2.73 Å, respectively. Hydrolysis forms the product and regenerates the PLP.

Table 7.1. Free energies (ΔG) and enthalpies (ΔH) for the transaminase catalytic cycle.

Solvation	$\epsilon = 4$		water	
	ΔG	ΔH	ΔG	ΔH
PLP	13.1	14.5	12.1	13.5
7	0.0	0.0	0.0	0.0
8	6.8	-15.2	8.8	-13.2
9a	1.5	1.4	1.8	1.6
9	7.8	0.7	9.2	2.1
TS9-10	25.8	15.1	26.5	15.8
10c	13.5	3.5	14.6	4.5
10a	14.4	15.2	8.4	9.2
10	14.8	5.1	15.1	5.4
TS10-11	16.1	5.4	16.9	6.2
11	1.6	-5.4	3.5	-3.5
11a	-3.6	-2.6	-3.1	-2.2
12	-1.6	-7.1	-0.1	-5.6
12a	-4.5	-1.7	-4.8	-1.9
13a	-0.6	1.4	-1.7	0.3
13	3.7	-3.7	4.3	-3.0
13r	4.8	-0.3	5.1	0
TS13-14	15.0	4.2	16.0	5.2
14	7.8	-1.9	8.0	-1.7
14a	13.7	15.8	7.1	9.2
14c	8.7	-2.7	9.8	-1.5
TS14-15	23.9	12.8	24.6	13.5
15	8.1	-0.7	7.6	-1.1
15a	3.3	2.8	1.4	0.8
2	7.7	10.8	6.1	9.2

Energies are calculated with M06-2X/6-311+G(d,p)//B3LYP/6-31G(d). The single point energies were calculated with CPCM water or $\epsilon = 4$. The bold and italicized free energies are the ones reported in Figure 7.5. **9a**, **10a**, **11a**, **12a**, **13a**, **14a**, and **15a** are calculated with lysine and intermediate separated. **13r** is a rotated conformation of **13**. In **10c** and **14c** the lysine is coordinated to the hydroxyl on pyridoxal ring.

**Figure 7.7.** Quinonoid intermediate. Distances to C_A and C_B are measured in the MD simulations.

In the three covalent PLP-sitagliptin intermediates, shown in Figure 7.6, the lysine has a hydrogen bond to the ketone of sitagliptin. Conformations with the ketone down, away from the lysine, are less favorable. The coordination of the lysine to the PLP-quinonoid, **10**, is exergonic by 11.7 kcal/mol compared to the uncoordinated lysine. The coordination to the sitagliptin-quinonoid, **14**, is exergonic by 19.5 kcal/mol and exothermic by 4.9 kcal/mol. However, this does not consider possible hydrogen bond contacts between the lysine and the protein backbone or side chains.

All atom MD simulations were run on with ATA-117 and Codexis' commercial variant on the ANTON supercomputer. In the MD simulations catalytic distances were measured, and the conformation of **5** was analyzed geometrically. The distances between the lysine and the substrate over the simulation are shown in Figures 7.8 (monomer) and 7.9 (dimers). These plots are distances from the catalytic lysine alpha carbon ($C\alpha$) or the nitrogen (Nz) to C_A and C_B of the substrate (Figure 7.7), where C_B is the site of protonation for the forward reaction and protonation at C_A would revert the intermediate back to the ketimine, **15**.

Simulations of the monomer of the commercial variant show an unstable backbone and active site. The distance between the lysine and the substrate and representative backbone conformations are shown in Figure 7.7. The $C\alpha$ - C_B plot shows two populations of backbone conformations, near 10 and 6 Å. The backbone's 4 Å shift relative to the substrate, hinders the ability of the lysine to protonate the intermediate as necessary in the reaction. The catalytic lysine is on average 6.5 Å from the C_B .

In the simulation of the dimer of the final variant with the quinonoid, Nz is significantly closer to C_B (4.4 Å) than in the monomer (Figure 7.9). The Nz - C_A distance averages 3.3 Å. Figure 7.11 shows an overlay of the backbone conformations of the dimeric enzyme compared to

the crystal structure. These structures show that the dimeric enzyme is less mobile than the monomer. This is also demonstrated by the $C\alpha$ to C_A and C_B distances, which deviate by less than 3 Å over the simulation, compared to 6.6 Å in the monomer. This suggests that the dimer is structurally important for the enzyme to function.

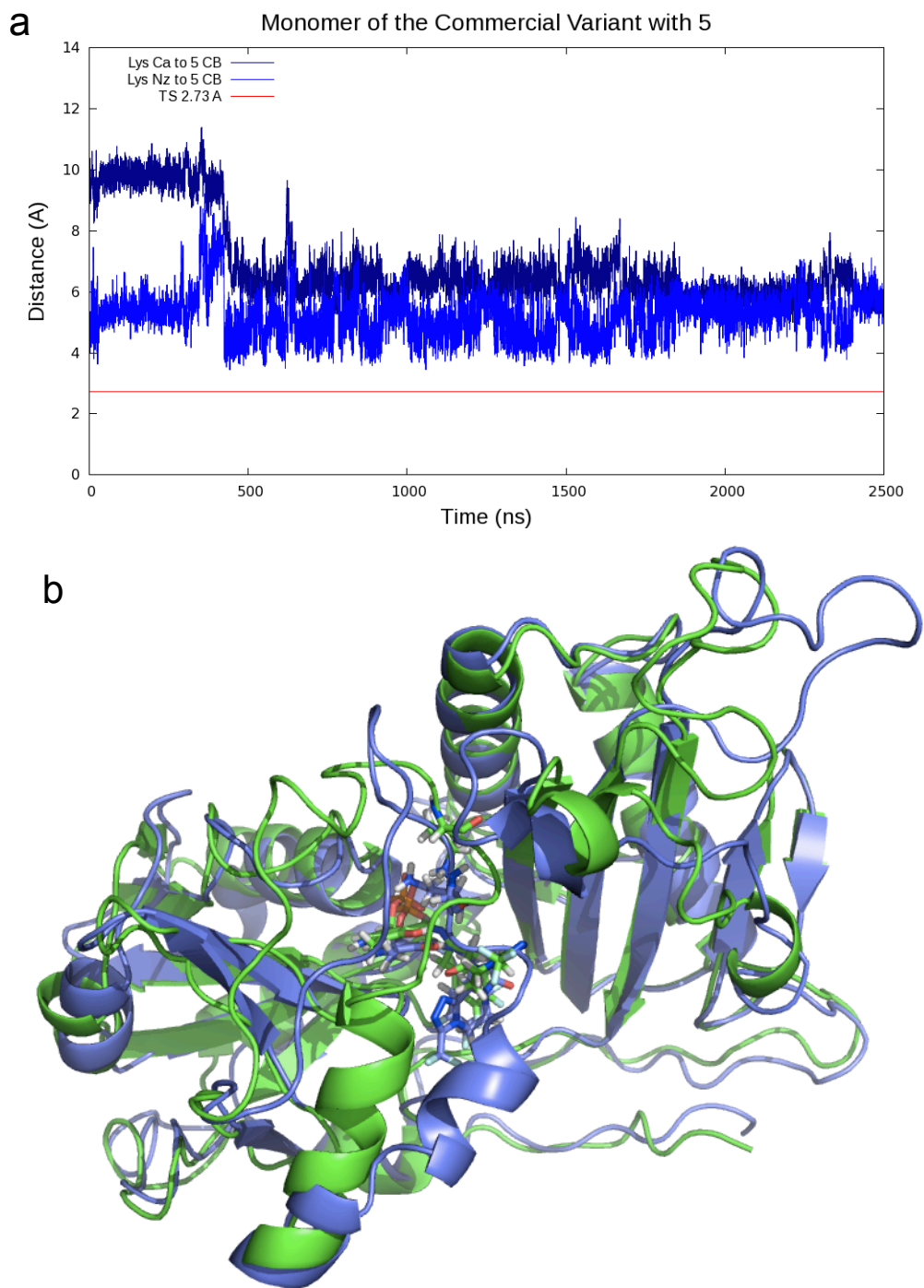


Figure 7.8. MD simulation of the commercial variant as a monomer. (a) The distance in Å from the lysine C α and N z to the substrate C $_B$ over the simulation (time is in ns). C α to C $_B$ is dark blue and N z to C $_B$ is light blue. The quantum mechanically calculated N z -C $_B$ distance for protonation in the transition state 2.73 Å is shown for reference as a red line. (b) The two representative backbone conformations are shown as ribbons.

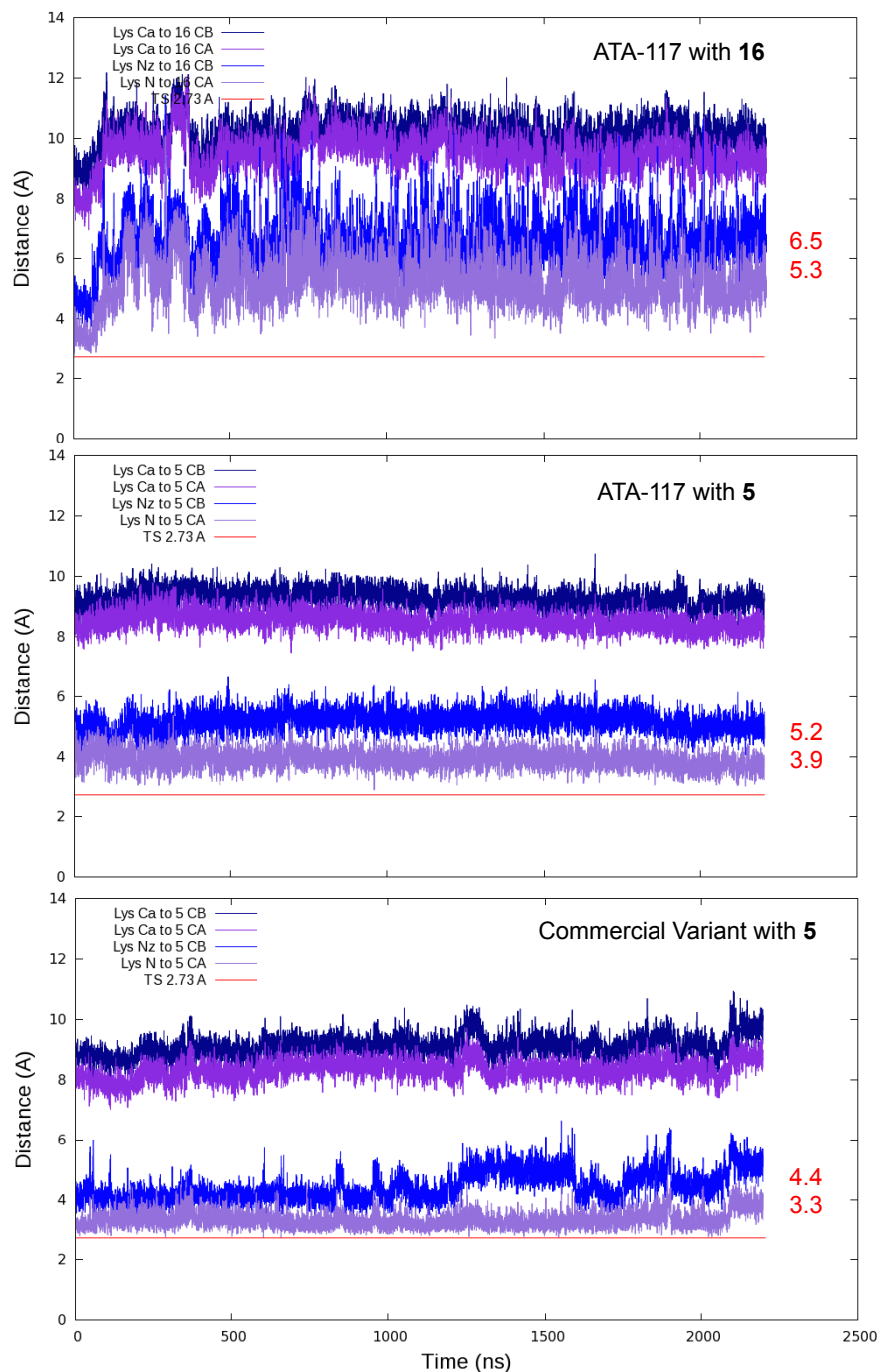


Figure 7.9. Plots of catalytic distances sampled in the MD simulations. The distances shown are of the lysine Nz and C α to C_B and C_A, where C α to C_B is dark blue, Nz to C_B is light blue, C α to C_A is dark purple, Nz to C_A is light purple and the Nz-C_B transition state distance for protonation, 2.73 Å, is shown as a red line. The average Nz-C_A and Nz-C_B distances are printed in red.

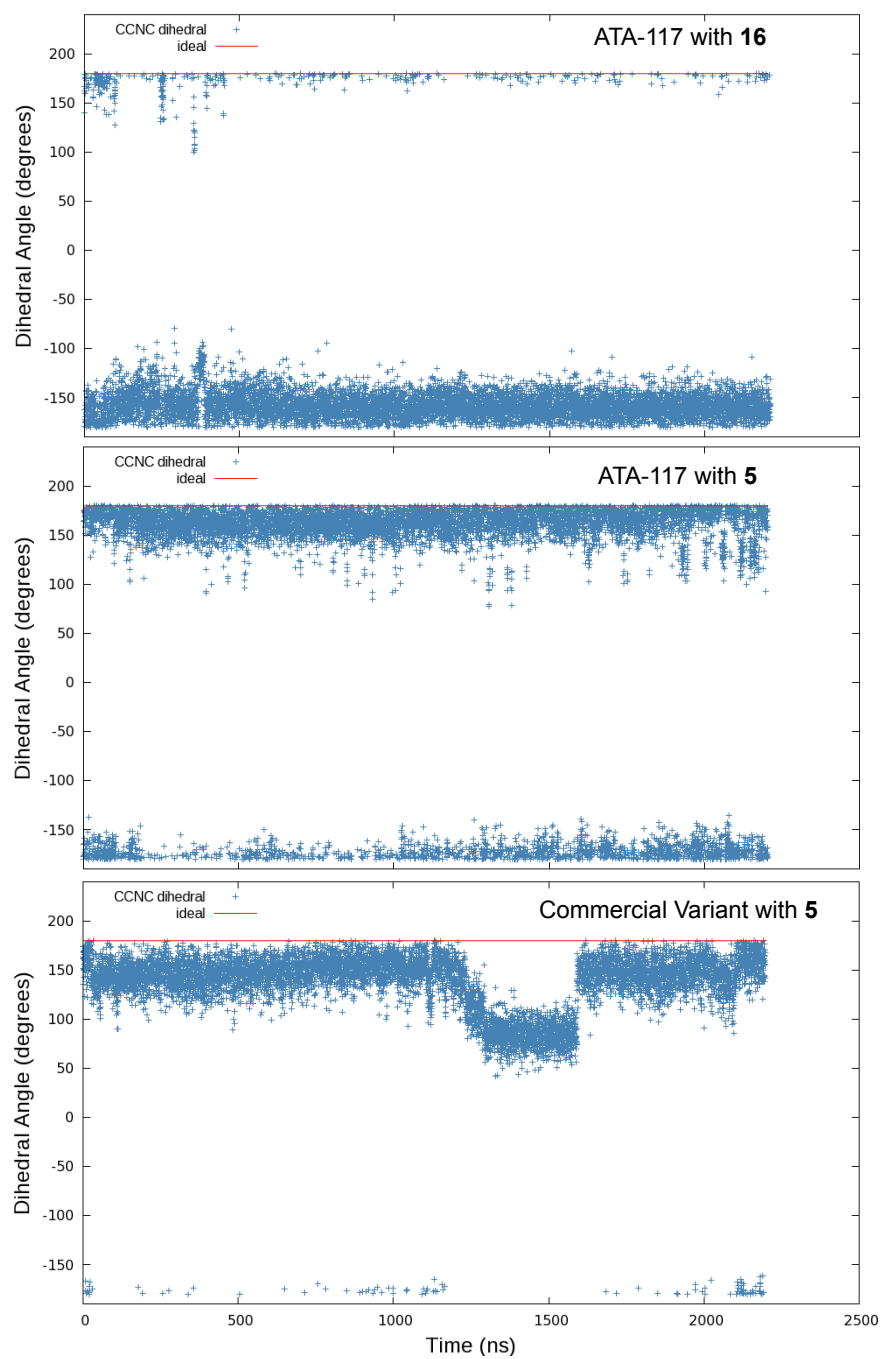


Figure 7.10. Plots of the CCNC dihedral. A red line is drawn at 180° .

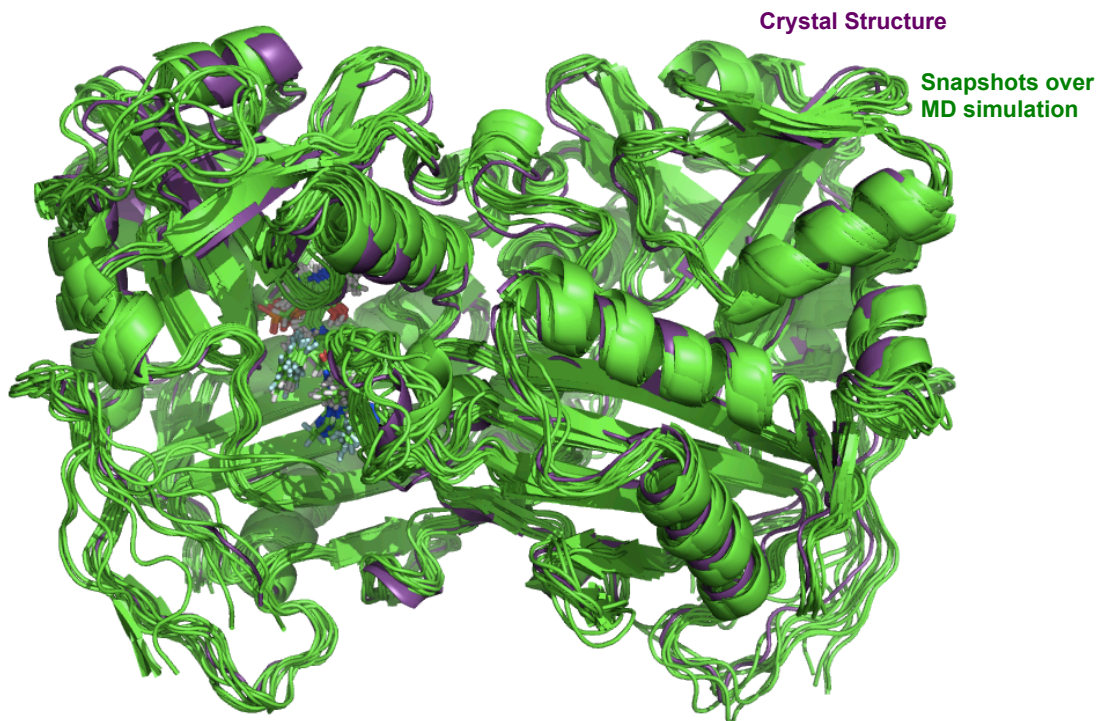


Figure 7.11. Overlay of x-ray crystal structure of the commercial sitagliptin-transaminase (purple) with snapshots every 240 ns over the 2.197 μ s MD simulation (green).

ATA-117 is not active toward the pro-sitagliptin ketone. Unexpectedly, several of the catalytic distances are maintained in the MD simulation. The contacts with the quinonoid in ATA-117 average 5.2 and 3.9 Å (Nz to C_B and C_A). This is slightly longer than those in the *active* commercial variant.

ATA-117 is active toward acetophenone, and was simulated with the acetophenone quinonoid, **16**. This offers a reference for how ATA-117 behaves with a substrate for which it is active. As shown in Figure 7.8, this substrate displays worse catalytic distances than the simulation of ATA-117 with the sitagliptin quinonoid, **5**. For **16** the average Nz-C_A and Nz-C_B distances are 6.5, and 5.3 Å, respectively, which is over an angstrom longer than the simulation with **5**. In these simulations catalytic distances are unable to predict the activity of the enzyme with a given substrate.

Because protonation of the quinonoid is facilitated by conjugation, the planarity of the quinonoid in the active site was assessed. Figure 7.10 shows the dihedral about the CCNC bond for the three simulations of the dimeric proteins. In ATA-117, the acetophenone-quinonoid, **16**, stays nearly planar (around -160°) for the entire simulation. In ATA-117, the sitagliptin-quinonoid, **5**, is even closer to planar for the entire simulation. In the final variant **5** is nearly planar (around 150°) most of the time, and deviates by about 100° for 15 % of the time (Figure 7.10). The simulation in which the substrate maintained the most planar geometry was ATA-117 with the sitagliptin-quinonoid, but ATA-117 is not active toward the pro-sitagliptin ketone, **1**. This result was unexpected, but suggests that the flexibility of the substrate inside the active site is more important than maintaining the ideal planar geometry of this intermediate. This enzyme must accommodate several different intermediates throughout the complex the multistep reaction, so a very rigid binding of the substrate may be unfavorable for some steps.

In conclusion, the quantum mechanical study of the mechanism of transamination of the pro-sitagliptin ketone gives valuable energies and geometries of the intermediates and transition states involved in this multi-step reaction. This study gives a better understanding of what needs to be achieved by the enzyme for transamination to occur. The MD simulations show that there are slightly better catalytic contacts with the quinonoid, **5**, in the *active* commercial variant than in the *inactive* ATA-117. The quinonoid intermediate for acetophenone, **16**, has poor catalytic contacts in ATA-117, even though ATA-117 is active for acetophenone. The quinonoid conformations are lower in energy in the *inactive* ATA-117 than in the *active* commercial variant. While it is still unclear as to how the evolution of ATA-117 was able to create an active transaminase toward the pro-sitagliptin ketone, some hypotheses have been generated. It is

possible that sitagliptin is unable to enter the active site of the ATA-117 or that some steps of the mechanism are disallowed in the ATA-117 and allowed in the commercial variant.

References

- ¹ (a) Krozowski, Z. *J. Steroid Biochem. Mol. Biol.* **1994**, *51*, 125. (b) Jörnvall, H.; Persson, B.; Krook, M.; Atrian, S.; Gonzalez-Duarte, R.; Jeffery, J.; Ghosh, D. *Biochemistry* **1995**, *34*, 6003.
- ² (a) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature*, **2012**, *485*, 185. (b) Hill, R. A.; Sutherland, A. *Nat. Prod. Rep.* **2014**, *31*, 706.
- ³ (a) Emre, M. et al. *New Engl. J. Med.* **2004**, *351*, 2509. (b) Fuchs, M.; Koszelewski, D.; Tauber, K.; Kroutil, W.; Faber, K. *Chem. Commun.* **2010**, *46*, 5500.
- ⁴ Plosker, G. L.; Figgitt, D. P. *PharmacoEconomics*, **2004**, *22*, 389.
- ⁵ Chen, C. *Curr. Med.* **2008**, *15*, 2173.
- ⁶ Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. *Science* **2010**, *329*, 305.
- ⁷ Mangion, I. K.; Sherry, B. D.; Yin, J.; Fleitz, F. J. *Org. Lett.* **2012**, *14*, 3458.
- ⁸ <http://www.statista.com/statistics/272367/revenues-of-merck-and-co-top-selling-drugs>
- ⁹ Hansen, K. B.; Hsiao, Y.; Xu, F.; Rivera, N.; Clausen, A.; Kubryk, M.; Krska, S.; Rosner, T.; Simmons, B.; Balsells, J.; Ikemoto, N.; Sun, Y.; Spindler, F.; Malan, C.; Grabowski, E. J. J.; Armstrong, J. D. *J. Am. Chem. Soc.* **2009**, *131*, 8798.
- ¹⁰ Ritter, S. K. 2010 Green Chemistry Awards. *Chemical and Engineering News*, **2010**, *88*, 9. (June 23, 2010)
- ¹¹ Iwasaki, A.; Yamada, Y.; Kizaki, N.; Ikenaka, Y. *Appl. Microbiol. Biotechnol.* **2006**, *69*, 499.
- ¹² (a) Jansonius, J. N. *Curr. Opin. Struct. Biol.* **1998**, *8*, 759. (b) Soda, K.; Yoshimura, T.; Esaki, N. *Chem. Rec.* **2001**, *1*, 373.
- ¹³ Schneider, G.; Kack, H.; Lindqvist, Y. *Structure* **2000**, *8*, R1.
- ¹⁴ (a) Truppo, M. D.; Turner, N. J.; Rozzell, J. D. *Chem. Commun.* **2009**, *45*, 2127. (b) Koszelewski, D.; Clay, D.; Rozzell, D.; Kroutil, W. *Eur. J. Org. Chem.* **2009**, 2289. (c) Hopwood, J.; Truppo, M. D.; Turner, N. J.; Liody, R. *Chem. Commun.* **2011**, *47*, 773.

¹⁵ (a) Dunathan, H. C. *Proc. Nat. Acad. Sci. U.S.A.* **1966**, 55, 712. (b) Martell, A. E. *Acc. Chem. Res.* **1989**, 22, 115. (c) Toney, M. D.; Kirsch, J. F. *Biochem.* **1993**, 32, 1471. (d) Nero, T. L.; Iskander, M. N.; Wong, M. G. *J. Am. Chem. Soc. Perkin Trans.* **1993**, 2, 431. (e) Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, 73, 383. (f) Griswold, W. R.; Castro, J. N.; Fisher, A. J.; Toney, M. D. *J. Am. Chem. Soc.* **2012**, 134, 8436. (g) Major, D. T.; Gao, J. *J. Am. Chem. Soc.* **2006**, 128, 16345.

¹⁶ Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

¹⁷ (a) Li, L.; Li, C.; Zhang, Z.; Alexov, E. *Theory Comput.* **2013**, 9, 2126. (b) Wattana-amorn P.; Williams, C.; Płoskoń, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P. *Biochemistry* **2010**, 49, 2186.

¹⁸ Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, Douglas J.; Kolossváry, I.; Klepeis, John L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang S. C. *Commun. ACM* **2008**, 51, 91.

¹⁹ Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12, University of California, San Francisco.* **2012**.

²⁰ Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, 79, 926.

8. Origins of Stereoselectivity of Mutants of *Lactobacillus kefir* Ketoreductase

Mutants of *Lactobacillus kefir* short chain alcohol dehydrogenase, used here as ketoreductases (KREDs), enantioselectively reduce the pharmaceutically relevant substrates 3-thiacyclopentanone and 3-oxacyclopentanone. These substrates differ by only the sulfur or oxygen atom in the ring, but the KRED mutants reduce them with different enantioselectivities. The polar and steric factors that control these selectivities are analyzed. X-ray crystal structures of apo and NADP-bound *L. kefir* KRED *wild type*, apo E145S, A94F, and the evolved enzyme for the industrial production of the antibiotic sulopenem, Sph, show that the substrate-binding loop conformational preferences are modified by these mutations. This is correlated with observed selectivities. Kinetic studies show that these enzymes are more efficient with thiacyclopentanone than with oxacyclopentanone. Quantum mechanical calculations and molecular dynamics (MD) simulations are used to understand the mechanism of reduction by the enzyme. We have developed a new method in which MD simulations are run on the enzyme into which we have incorporated the transition state complexes, or *theozymes*, for the reductions to form the *R*- or the *S*- alcohol. This method probes the stability of the catalytic arrangement with the *theozyme* and shows a correlation between the relative fractions of catalytically competent poses for the enantiomeric reductions, and the experimental enantiomeric ratio (*er*). The mutation from alanine to phenylalanine in the large binding pocket (A94F) increases the hydrophobicity of this pocket, and this favors the orientation of the methylene side of both substrates toward the large binding pocket, and provides a stabilizing CH- π interaction, increasing the *S*-selectivity of the enzyme. In the E145S mutant, and in Sph, there is a change of the conformation of tyrosine 190, and this is related to the enhanced *R* selectivity with thiacyclopentanone.

Significance

Ketoreductases (KREDs) are the most commonly used enzymes in industrial pharmaceutical synthesis. We have explored the enzymes in a library of KREDs that reduce 3-oxa- and 3-thia-cyclopentanone, which are pharmaceutically relevant, and are difficult to reduce enantioselectively by other means. Our experimental and theoretical study reveals how mutations modulate the stereoselectivity of the reduction. We present the crystallographic structure and efficiencies of select variants. We have developed a molecular dynamics method, in which simulations are run on the enzyme containing the *theo*zyme for the reduction. This approach probes how well each enzyme stabilizes these transition structures and can predict the experimentally favored enantiomer. These studies show that enantioselectivity can be influenced by creating new hydrophobic contacts, or by blocking or allowing water access to the active site.

Introduction

Ketoreductases (KREDs) reduce ketones to alcohols with high stereoselectivity. They are the most commonly used enzymes in industrial pharmaceutical synthesis.¹ Biocatalysis is a commonly used method for stereoselective ketone reduction.² The biocatalytic process often replaces multistep syntheses, and uses renewable, biodegradable, and nontoxic reagents.³ These reactions are run at ambient temperature and atmospheric pressure. KREDs can be used to synthesize alcohols of a wide range of sizes, can be chemo- and diastereoselective, and have been engineered to synthesize alcohols for the production of atorvastatin (Lipitor), montelukast (Singulair), and atazanavir (Reyetaz).⁴

The hydride reductant in KREDs is nicotinamide adenine dinucleotide phosphate (NADPH). To regenerate the NADPH and complete the catalytic cycle, isopropyl alcohol (IPA)

is oxidized to acetone (Figure 8.1). In this study, the KREDs used are mutants of *L. kefir* short chain alcohol dehydrogenase. This enzyme is in the short chain dehydrogenase/reductase (SDR) family.^{5,6}

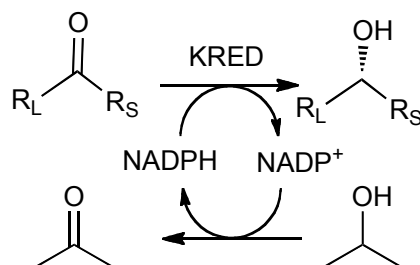


Figure 8.1. KRED catalytic cycle.

Figure 8.2 shows the proposed reduction step and accompanying proton transfer.⁷ In *L. kefir* numbering, Ser143 and Tyr156 stabilize the alkoxide formed from the reduction.^{8,9} The alkoxide is protonated by a series of proton transfers involving this tyrosine, the cofactor, Lys160, and a backbone carbonyl, leading to bulk water. When the substrate's substituents differ in size, the smaller substituent binds to the small binding pocket and the larger substituent to the large binding pocket.¹⁰

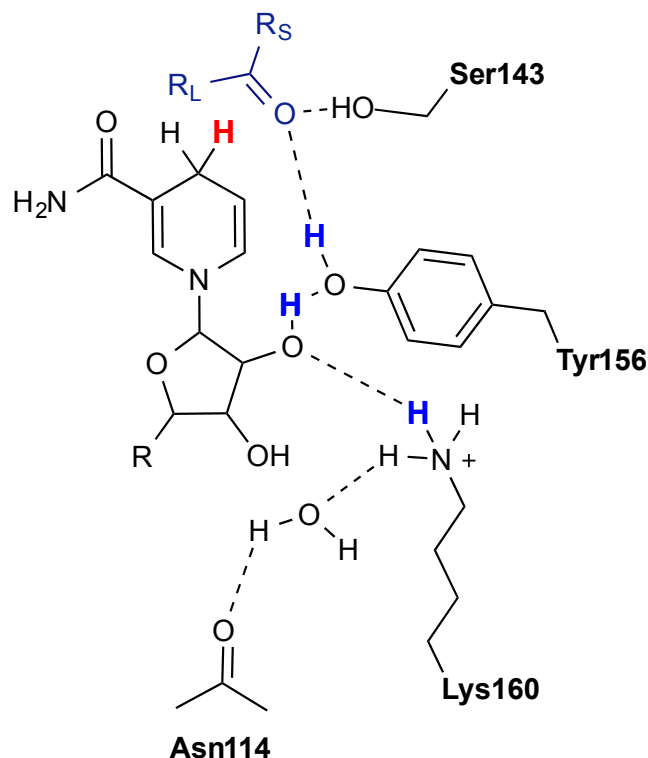


Figure 8.2. Reduction and proton transfer in the enzyme. The NADPH cofactor is abbreviated. The substrate is blue, the cofactor, and catalytic residues, and water are black, the hydride is highlighted in red, and the protons in the proton relay are highlighted with boldface blue.

Figure 8.3 shows the small and large binding pockets in the x-ray structure of the apo *wild type* (WT) enzyme with acetophenone docked in. Tyr190 prevents binding of the large group in the small binding pocket.¹¹ An important feature present in all of these enzymes is the substrate-binding loop, which flanks one side of the active site, and closes around the bound cofactor and substrate. This loop spans roughly residues 190-210 (Figure 8.4).¹¹ This loop has been implicated as being important in determining selectivity of KREDs¹² because it forms the active site pocket and it is the most variable portion of SDRs.

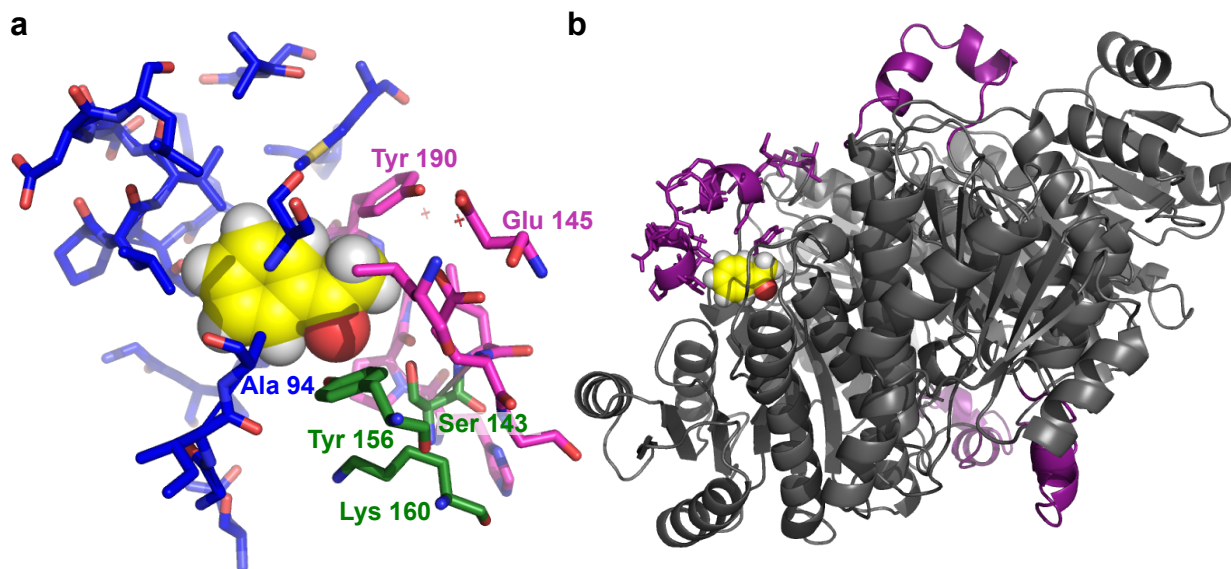


Figure 8.3. *L. kefir* KRED (Crystal apo WT) with acetophenone (yellow spheres representation) docked in. (a) The residues in the active site are shown as sticks. The catalytic triad, Ser143, Tyr156, and Lys160, are green; the residues along the small binding pocket (144-146, 150, 153, 187-190, 212) are pink, and the residues along the large binding pocket (18-19, 90, 92-94, 96, 114, 191-199, 202-203, 206, 217) are blue. (b) The 4 substrate-binding loops in this tetrameric protein are highlighted in purple.

The chiral alcohols produced from oxacyclopentanone, **1**, and thiacyclopentanone, **2**, are precursors for the synthesis of pharmaceuticals and can be incorporated into fosamprenavir and sulopenem, respectively (Figure 8.4). These substrates are difficult to reduce asymmetrically by other means,^{13,14} and biocatalysts offer a method to reduce these ketones with high enantioselectivity. Sph was evolved and used for the industrial reduction of **2** for the production of sulopenem.^{15,16}

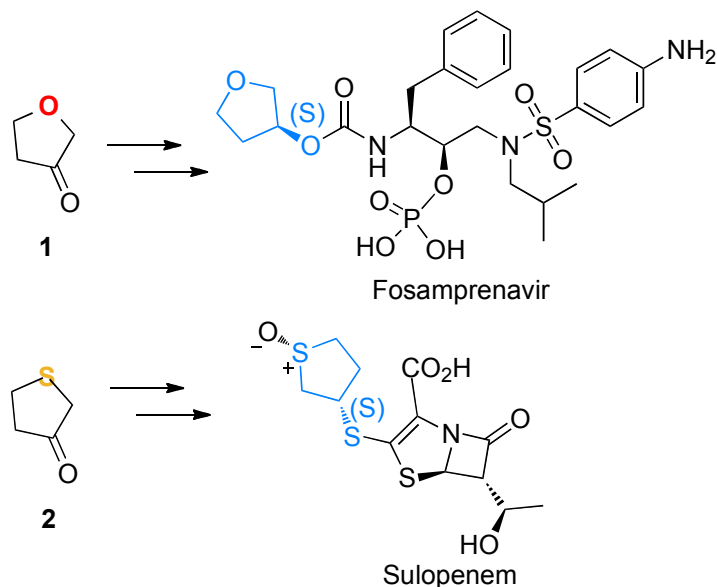


Figure 8.4. The alcohols formed from **1** and **2** are employed in the synthesis of fosamprenavir and sulopenem.

Materials and Methods

Assessment of the enantiomeric ratio given by variants in the mutant library. Experiments were run following Codexis' standard KRED panel screening protocol.¹⁷ A total of 175 KRED variants, with 0 to 16 mutations from *L. kefir* WT KRED, were mixed with each substrate, **1** and **2**, in 96-well plates. Each well contained one substrate, cell lysate, which includes one KRED variant, NADP sodium salt, and triethanolamine in isopropanol. Reactions were allowed to proceed at room temperature for 24 hours. Gas chromatography was used to determine the conversion and the *er* of the resulting alcohol.

Crystallization of the WT KRED and the mutants. Crystals of the WT KRED and variants were grown by a hanging-drop vapor diffusion method. Crystals of the WT KRED, both apo and NADP bound, were grown at room temperature. Crystals of A94F and E145S were grown at 4°C.

Data collection and structure determination. X-ray diffraction data for the WT KRED, A94F, and E145S were collected at the Advanced Photon Source (Argonne National Laboratory), beamline NECAT 24ID-C, using a DECTRIS PILATUS-6M pixel detector. Data reduction and scaling were performed using XDS for all crystals. Diffraction to 2.2 Å resolution was observed for the WT KRED. Diffraction to 1.6 Å resolution was observed for A94F. Diffraction to 1.7 Å resolution was observed for E145S. X-ray diffraction data for the NADP bound WT KRED crystal was collected at the University of California Los Angeles (UCLA), using a RIGAKU FRE+ rotating anode and a RIGAKU HTC image plate detector. Diffraction to 2.1 Å resolution was observed. All of the crystal structures were determined by the molecular replacement method using the program PHASER and using the KRED from *Lactobacillus brevis* as the search model (PDB entry 1NXQ).

Kinetics studies. To measure the efficiency (k_{cat}/K_m), the concentration of enzyme was fixed at 0.05 mM, while the concentration of **1** was varied from 10 mM to 160 mM or the concentration of **2** was varied from 0.2 to 1 mM, since catalysis of its reduction is less efficient. DMSO was added to a final concentration of 2% to facilitate the solubilization of the substrates. The reactions took place at 25°C in 150 mM phosphate buffer, pH 7.0. The final reaction solution was 7% glycerol and contained 2 mM MgSO₄ and 1 mM NADPH. The rate of NADPH depletion was measured using a Tecan Infinite M200 Pro plate reader by recording the absorbance of the reaction solutions at 340 nm relative to an NADPH standard.

Quantum mechanical calculations. A series of quantum mechanical and force field-based methods were used to understand how these enzymes distinguish between **1** and **2** and how they achieve their enantioselectivity. Quantum mechanics was used to study the properties of the substrates and to determine approximate reaction barriers by calculating the stationary points in

the reduction with an unconstrained *theozyme* (short for ‘theoretical enzyme’, the optimal transition structure stabilized by the catalytic amino acid side chains¹⁸). These calculations provide the ideal geometry for catalysis of the reduction and show the role of the cofactor and amino acids in stabilizing the transition state. These calculations were performed with the Gaussian 09¹⁹ suite of programs. The geometries were optimized with the TPSS functional and 6-31G(d)²⁰ basis set with the Conductor-like Polarizable Continuum Model²¹ (CPCM) for water. Single-point energy calculations were performed on these geometries using the M06-2X method²² and 6-311+G(d,p) basis set, and with CPCM water. The complexations of individual water molecules to substrates **1** and **2**, were calculated with MP2/6-311+G(d,p).²³

Standard MD simulations. All-atom, AMBER²⁴ MD simulations were run on selected variants with the substrate and cofactor bound in explicit water. The starting structures for the MD simulations were homology models, or subsequently obtained crystal structures, where the enzyme is tetrameric.¹¹ The NADPH cofactor plus substrate is in one active site and only the NADP cofactor is in the others. The protein is surrounded by an octahedral box of TIP3P²⁵ explicit waters. The parameters for the substrate and cofactor were generated from optimizations with HF/6-31G(d) and processed by the antechamber module in Amber12 to provide atomic partial charges. Counter ions (Na⁺, Cl⁻) were added to neutralize the protein. The protein and surrounding water environment was minimized and equilibrated, each in two stages. In the first minimization stage only the positions of solvent molecules and ions are minimized, and in the second of all the atoms in the simulation are minimized. In the first equilibration the systems were heated by increasing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions over 5 ns with a 2 fs time step. Harmonic restraints of 10 kcal/mol were applied to the substrate and cofactor. The second equilibration ran for 4 ns at a

constant volume and temperature of 300 K. The production MD simulations were then run for additional 100 ns under the same simulation conditions.

MD simulations with the *theozyme*. In order to assess the stability of the catalytic arrangement near the TS-complex, MD simulations of the tetrameric enzymes were performed with the *theozyme* and the NADPH cofactor in one active site and only the NADP cofactor in the others. The geometry and charge distribution of the *theozyme* was used to build the substrate-cofactor transition state complexes. Throughout the MD simulation, the TS geometry was restrained, because it would otherwise be changed by the force field to resemble a minimum, rather than a TS structure. An example of substrate-cofactor transition state complexes and the QM charges applied to this complex are shown in Figure 8.5. These complexes were docked into the active site of the protein, and the protein with the substrate-cofactor transition state complexes served as the starting structures for the MD simulations. We monitored four hydrogen bond distances that are critical for the catalysis driven by the enzyme.

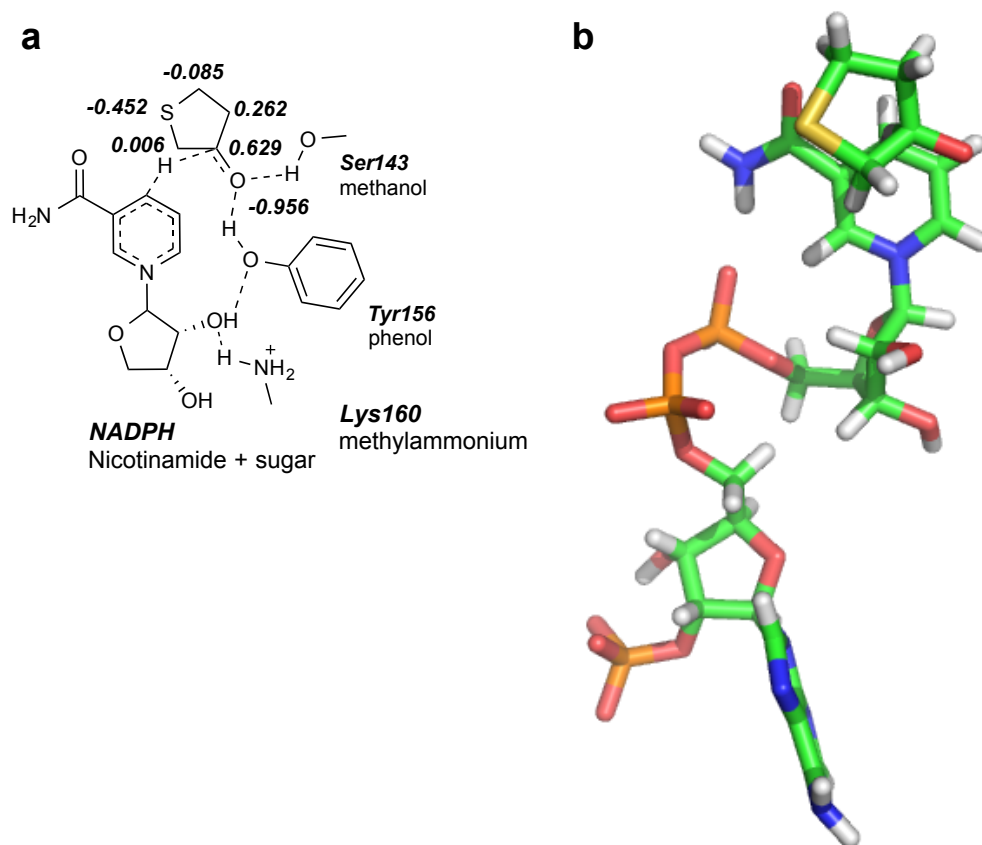


Figure 8.5. *Theozyme* for 3-thiacyclopentanone reduction in KREDs. (a) The atomic partial charges of the substrate form the *theozyme*. (b) The complete pro-*R* transition structure used in the MD simulations; the geometry of this activation complex was partially restrained to match that of the QM-calculated *theozyme*.

Results and Discussion

Experimental Selectivities. A plot of the enantioselectivities given by 175 KRED variants for the reduction of **1** and **2** is shown in Figure 8.6. These enzymes can be highly enantioselective, meaning they distinguish between a sulfur and a methylene group or an oxygen and a methylene. The variants do not give the same enantiomeric ratio (*er*) for the two substrates.

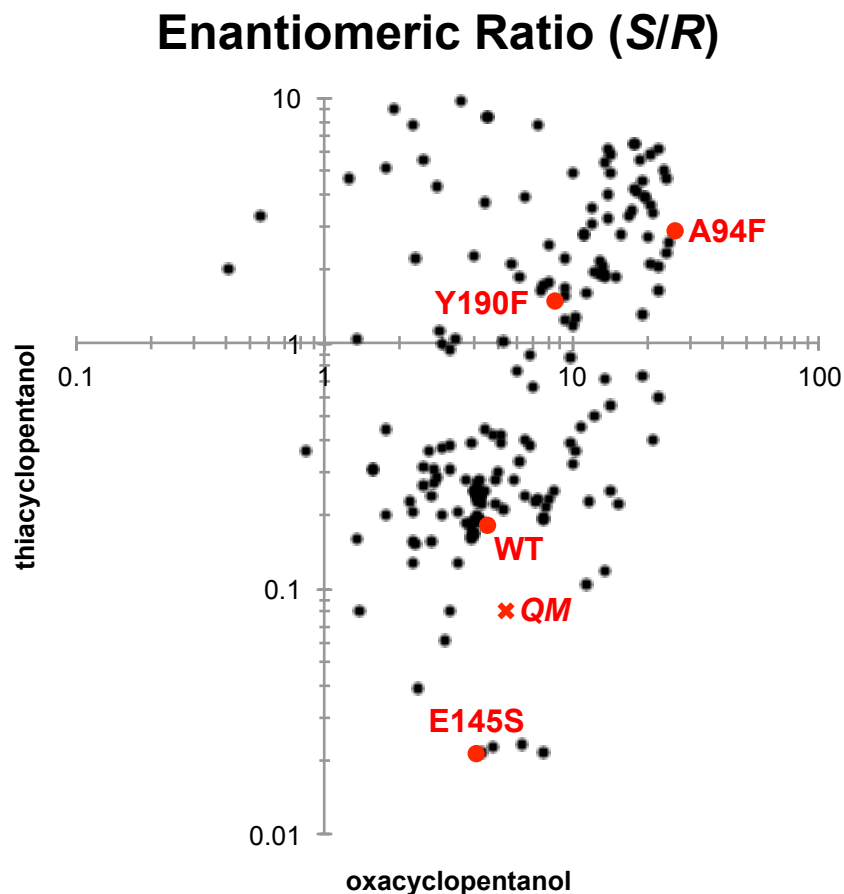


Figure 8.6. Log/log graph of experimental enantiomeric ratio (*er*) given for oxa- and thiacyclopentanol by KRED variants. Each point is a KRED variant plotted according to the *er* it gives for the two alcohols.²⁶

Nearly all of the mutants, including the WT, give (*S*)-3-oxacyclopentanol from **1**. This is in accord with the selectivity expected based on size: this selectivity arises if the small oxygen binds in the small binding pocket and the larger methylene binds in the large binding pocket. However, the selectivity of reduction of **2** ranges from 0.02 to 9.7 *er* (100 % *ee R* to 80 % *ee S*), and clearly size alone cannot explain this variation. The sizes and polarities of the two halves of the substrates are analyzed in detail later in this chapter.

We selected a set of single and double mutants from these variants for further computational, crystallographic, and kinetics studies. The selected variants show the greatest changes caused by a single mutation, enabling us explore the origin of selectivity. A table of the

selected variants, and the selectivities of the reduction of both substrates is given in Table 8.1. The values of $\Delta\Delta G^\ddagger$ (*S* vs *R*) given are calculated from the *er* values. Over the entire library screened, the reduction of **1** ranges from -1.95 to 0.52 kcal/mol, while that of thiacyclopentanone ranges from -1.34 to 3.33 kcal/mol.

Table 8.1 Enantiomeric ratio (*S/R*), corresponding $\Delta\Delta G^\ddagger$ (kcal/mol), and catalytic efficiency (k_{cat}/K_m) determined experimentally for 3-oxa- and 3-thia- cyclopentanol (**1** and **2**, respectively) by KRED variants

Variant	1			2		
	<i>er</i>	$\Delta\Delta G^\ddagger$ kcal/mol	k_{cat}/K_m [mM ⁻¹ min ⁻¹]	<i>er</i>	$\Delta\Delta G^\ddagger$ kcal/mol	k_{cat}/K_m [mM ⁻¹ min ⁻¹]
WT	4.19	-0.85	0.0251	0.19	0.98	1.46
A94F	27.5	-1.95	0.00462	2.56	-0.55	0.0164
Y190F	9.27	-1.31	n.a.	1.56	-0.26	n.a.
E145S	4.18 (5.06 ^a)	-0.84	0.00278	0.02	2.34	0.772
Sph	1.28 ^{a,b}	-0.14	0.00428	0.004 ^a	3.33	0.703

a. Experimental conditions¹⁵ b. Experimental *er* for Sph with oxacyclopentanone was not measured, but these data were obtained for a related variant with 9 of the 10 mutations in Sph (G7S/A94T/S96P/R108H/G117S/E145S/N157T/M206Q/I223V, lacking P194N), which gave 1.28 *er*.

Kinetics studies. The efficiencies of the WT, A94F, E145S, and Sph were measured and are shown in Table 8.1. The WT is the most efficient of these enzymes. Each enzyme is a better catalyst for reduction of **2** than of **1**. Their efficiency range is 0.0164 - 1.46 min⁻¹mM⁻¹ and 0.00278 - 0.0251 min⁻¹mM⁻¹, respectively. The WT is about 58 times faster with **2** than with **1**. E145S and Sph are two orders of magnitude faster with **2** than **1**. A94F, which is the slowest enzyme toward **2** is still 3 times faster with **2** than **1**. The decreased rate for reduction of **1** with A94F, which favors the formation of the *S*-enantiomer, implies that the rate of formation of the normally preferred *R*-enantiomer is reduced significantly by the A94F mutation.

Analysis of substrates. The electrostatic potential surfaces for **1** and **2** are shown in Figure 8.7. The substrates differ in size and polarity of the groups at position 3 and 4. Oxygen is smaller than sulfur with van der Waals radii of 1.55 and 1.80 Å, respectively. The C-O bond (1.43 Å) is shorter than the C-S bond (1.85 Å). Molar volume calculations reveal that **1**, is 13 % smaller than **2** . Furthermore, the electron density on the oxygen of **1** is more localized than that on the sulfur of **2**, as shown by electrostatic potentials in Figure 8.7. The molar volume of the O-CH₂ fragment is 25 % smaller than the CH₂-CH₂ fragment in **1**, while the S-CH₂ fragment is 6 % larger than the CH₂-CH₂ in **2**. Therefore, in **1** the oxygen comprises the small, polar side, and the methylene the large, non-polar side. In **2**, sulfur is polar, although to a lesser degree than oxygen, and the non-polar methylene is slightly smaller than sulfur. The enzyme recognizes these differences to give the enantioselectivity in the reduction, and reduces these substrates with different enantioselectivities.

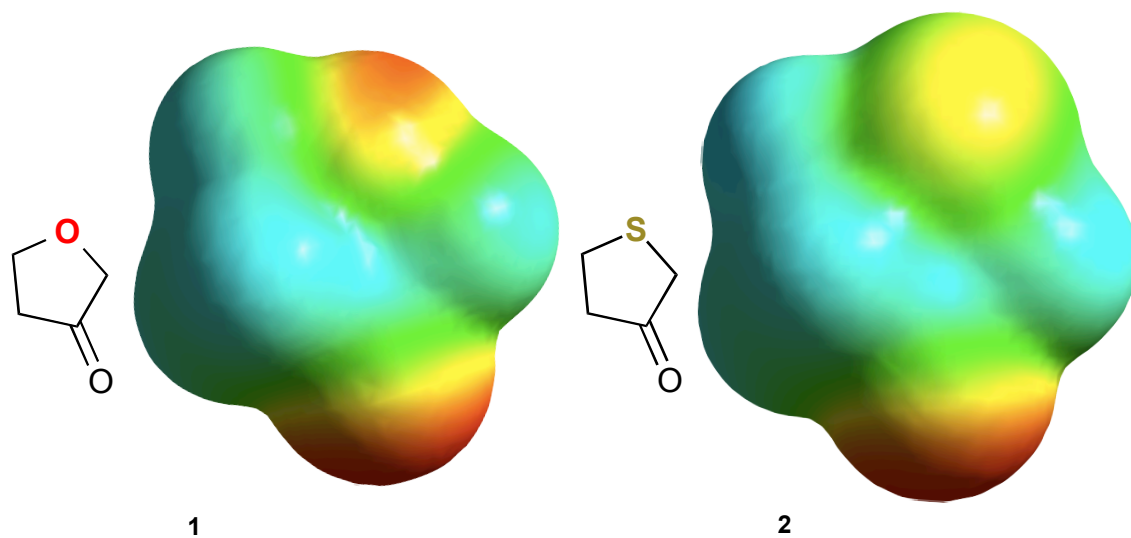


Figure 8.7. The electrostatic potential surfaces of **1** and **2**, where red is negative and blue is positive. The surfaces are electron density isovalues of 0.02, and the scale is from -27.6 (red) to +27.6 (blue) kcal/mol. Calculated with B3LYP/3-31G(d).

Oxygen and sulfur hydrogen bond (H-bond) with different energies and geometries.^{27,28,29}

The complexation of one and two waters with **1** is more favorable than with **2** by 1.4 and 2.3 kcal/mol, respectively. The lengths of the H-bonds also differ: the O-H distance is 1.91 – 1.94 Å, and the O-S distance is 2.47 – 2.65 Å, so the complex of **2** with water is considerably larger than with **1**. All of these differences contribute to how the enzyme can distinguish the two substrates, and why the enzymes have very different selectivities for the two.

Quantum mechanical models of transition states. The enzymatic reduction was modeled with DFT (TPSS functional and 6-31G(d) basis set), using an abbreviated model of the active site. The NADPH cofactor was truncated to the nicotinamide and sugar moiety. The catalytic serine, tyrosine and lysine were abbreviated to methanol, phenol, and methylammonium, respectively. The calculated stationary points for the reduction of the ketone to alcohol are shown in Figure 8.8, also with computed energies. The transition structures are shown in Figure 8.9.

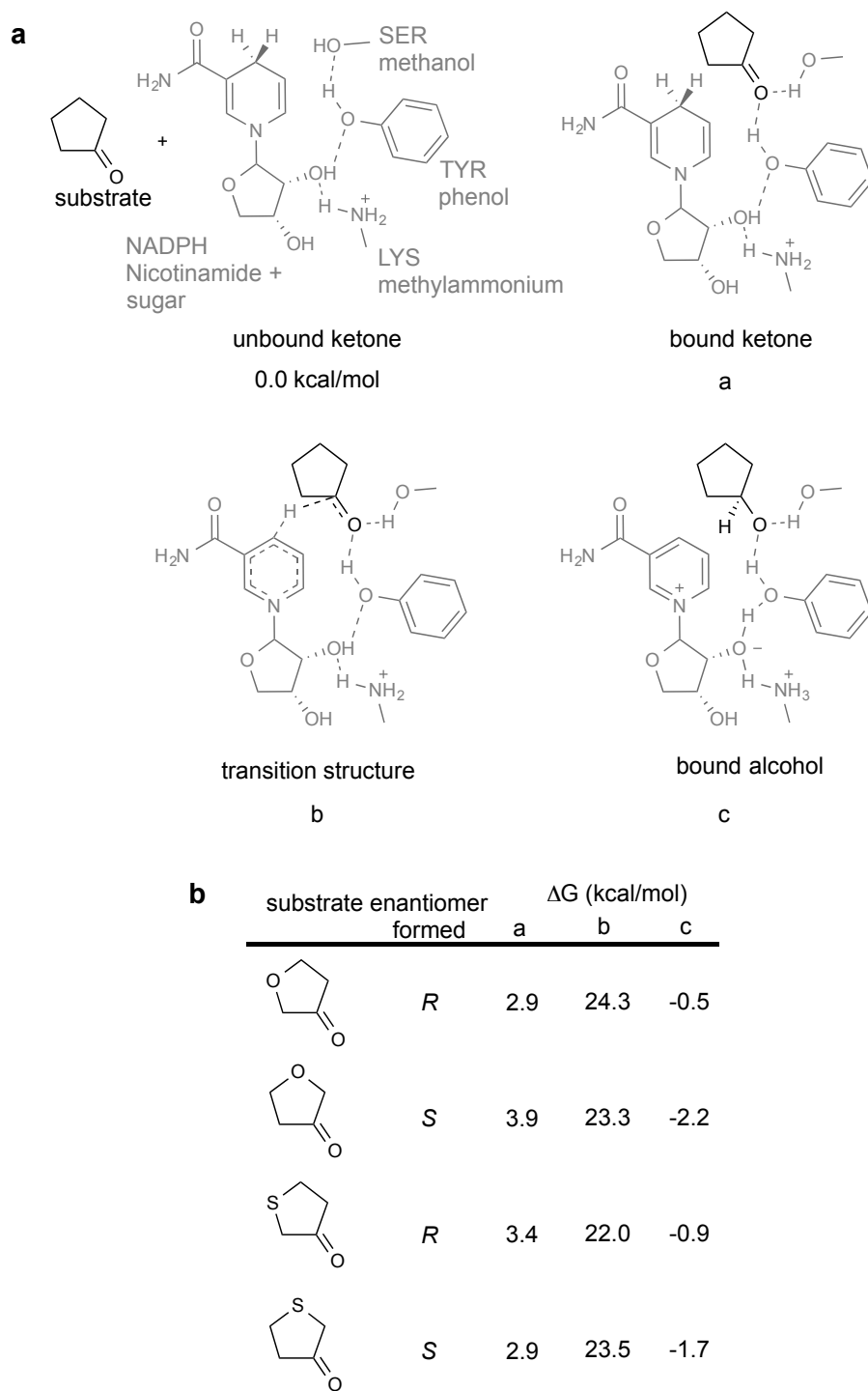


Figure 8.8. Mechanism of reduction as modeled with an abbreviated active site (*theozyme*). (a) ChemDraw depiction of each stationary point calculated. (b) The relative free energies calculated with M062X/6-311+G(d,p) with CPCM water single point energies on TPSS/6-31G(d) with CPCM water geometries.

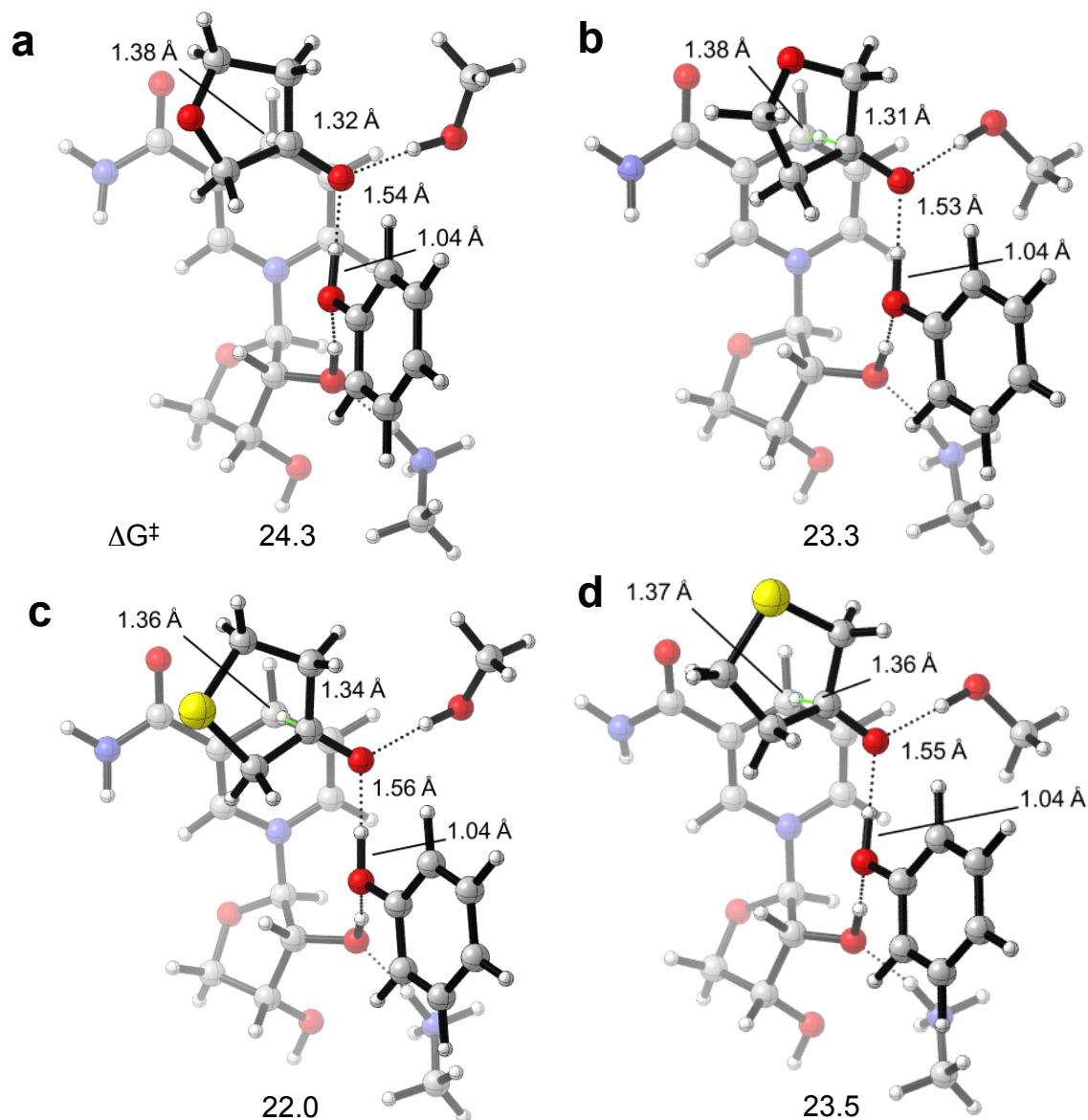


Figure 8.9. Transition structures leading to (a) *R*-oxacyclopentanol (b) *S*-oxacyclopentanol, (c) *R*-thiacyclopentanol, (d) *S*-thiacyclopentanol calculated with M062X/6-311+G(d,p)//TPSS/6-31G(d), CPCM water (same as in Figure 8.8 and 8.10). ΔG^\ddagger is given in kcal/mol.

The complex of the substrates with the amino acid and cofactor model is disfavored by 3 kcal/mol. The barriers for the reduction range from 22 to 25 kcal/mol, and **2** is reduced more easily than **1** by about 2 kcal/mol. This is consistent with the measured efficiencies of the enzymes, where **2** is reduced more readily than **1** for all enzymes tested. The differences in the

activation free energies between the *R* and *S* reductions of **1** and **2** are 1.0 and 1.5 kcal/mol, respectively. The enantioselectivity comes from the asymmetry of the cofactor and catalytic residues. This model predicts 5.4 *er* for **1** and 0.08 *er* for **2** (69 % *ee* *S*-oxacyclopentanol and 86 % *ee* *R*-thiacyclopentanol). Hence intrinsically, *S*-selectivity is preferred for **1** and *R*-selectivity is preferred for **2**. This is consistent with the selectivity seen in the WT enzyme. Nearly all-175 variants favor the formation of *S*-oxacyclopentanol, suggesting the enzyme does not overcome the intrinsic selectivity towards **1** very easily.

The transition structures and distances are shown in Figure 8.8. The donation of the hydride is concerted with the proton transfer from tyrosine to the substrate leading to the alcohol. In these transition states, the hydride is slightly closer to the substrate (1.31-1.36 Å) than the cofactor (1.36-1.38 Å). In the *S*-thia transition state the two C-H bond distances are 1.36 and 1.37 Å, respectively. The proton transfer is very early; the proton that is transferred from the catalytic tyrosine is still mostly bonded to tyrosine, reflected by the 1.04 Å O-H bond distance in all transition states. In the complex with the product alcohol, the tyrosine and sugar share a proton in a strong H-bond; for *R*-oxa, *S*-oxa, and *S*-thiacyclopentanol the tyrosine has a shorter bond to the proton than does the sugar (1.06 and 1.52 Å, respectively). In the complex with *S*-thiacyclopentanol, the opposite is true, with bond distances of 1.37 and 1.10 Å, respectively. In all cases the lysine remains partially protonated, where the N-H bond is 1.1 Å.³⁰

The higher reactivity of the thiaketone would not be expected if only inductive effects were important. A distortion/interaction analysis³¹ shows that it is harder to distort the oxaketone to the transition state geometry than it is to distort the thiaketone to its geometry, presumably due to the longer C-S bond lengths and consequent greater flexibility of the **2** geometry. The

interaction energies indeed favor the reaction of the **1**, although the differences in distortion energies dominate.

Figure 8.10 is an overlay of the optimized transition states with the NADP-bound WT crystal structure. The arrangement of the catalytic amino acids in the optimized *theozyme* closely matches the enzyme crystal structure. In this *theozyme* the four heteroatoms involved in the H-bond network differ from the NADP-bound WT crystal by only about 0.5 Å; that is, the enzyme is pre-organized to stabilize the transition state of this reduction. Several conformations of the H-bond network were tested. Lysine forms one H-bond to the sugar of the cofactor; the cofactor also has a H-bond to the tyrosine, the tyrosine is H-bonded to the partially formed alkoxide, and serine directly stabilizes the alkoxide through an additional H-bond. This arrangement is in good agreement with previously proposed mechanisms for this reaction.⁷⁻⁹ Less favorable arrangements include serine H-bonding with tyrosine, or lysine having two H-bonds to the sugar, or lysine H-bonding to tyrosine.

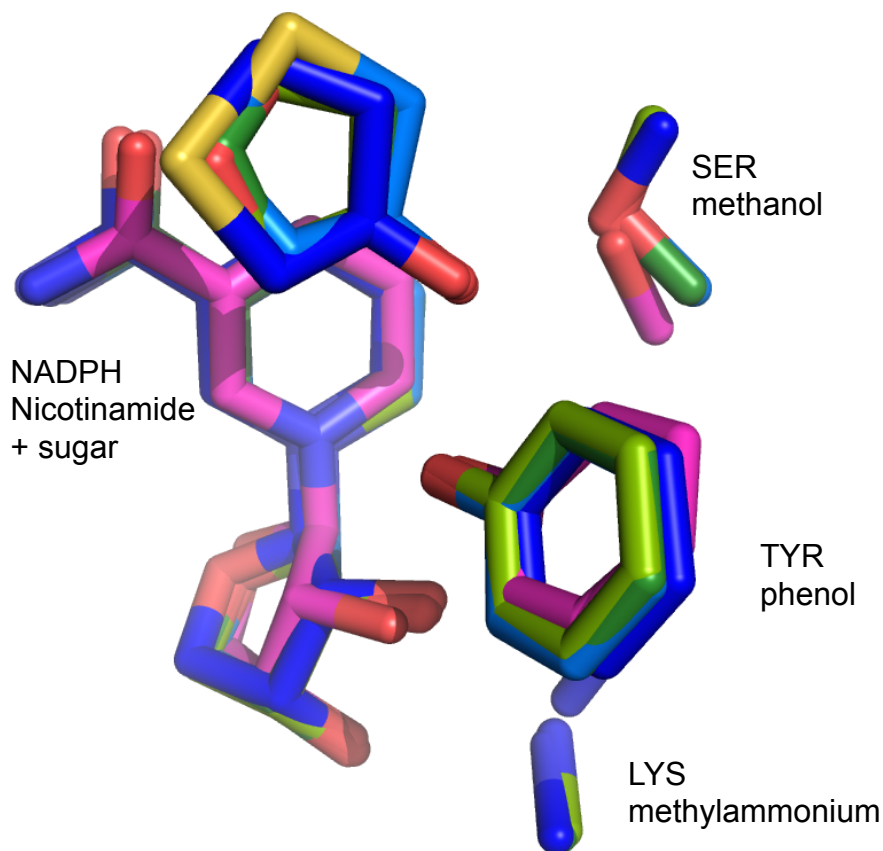


Figure 8.10. QM transition structure and crystal structure overlay .The four calculated transition structures (oxa- (green) and thia- (blue) transition structures in the pro-*R* and pro-*S* conformations) overlaid with the NADP-bound crystal structure (pink). Transition structures and energies were calculated with M062X/6-311+G(d,p)//TPSS/6-31G(d), CPCM water.

Analysis of x-ray crystallographic structures. Five crystal structures of *L. kefir* KRED variants have been obtained. We report the apo and NADP-bound *L. kefir* KRED WT, in addition to apo E145S, A94F, and Sph. This set of structures provides comparisons of a cofactor-bound and unbound structure and of single mutant variants and the best variant for the transformation involved in the production of sulopenem.

Figure 8.11 shows an overlay of the apo and NADP-bound WT crystal structures. The WT apo and NADP-bound structures vary mainly in the conformation of the substrate-binding loop, which is open in the apo form and closed when NADP(H) is bound. The deviation in conformation stems from key H-bond contacts between the loop and the cofactor and to other

residues in the protein. The cofactor makes six H-bond contacts to three residues, 191, 193, and 195, in the loop. The two phosphates that link the sugars in the cofactor make three of these contacts, which are to residues 193 and 195. The amide makes the other 3 contacts to residues 191-193. The contact between residues 145 and 190 help hold the loop closed and keep water out of the small binding pocket.

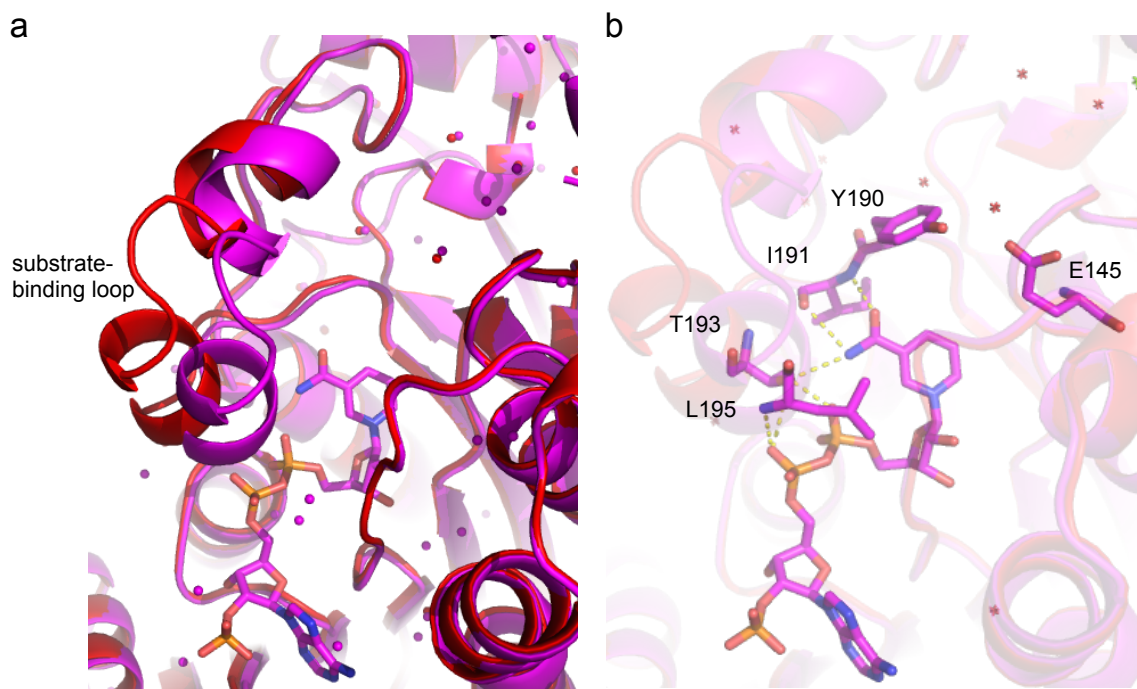


Figure 8.11. Overlay of the apo (red) and NADP-bound (pink) WT *L. kefir* KRED crystal structures. (a) The cartoon representation of the backbone shows that the substrate-binding loop has different conformations when NADP is bound or not bound. (b) The 6 H-bond contacts between the cofactor and this loop at residues 191, 193, and 195. The residues are drawn with sticks and the contacts are shown in yellow dashed lines. Tyr190 and Glu145 are also shown in sticks.

Figure 8.12 shows the contact between residue 145 and Tyr190 in the crystal structures of the apo and NADP-bound WT, apo A94F, E145S, and Sph crystal structures. The contact varies between the apo and NADP-bound structures (Figure 8.12a and b). In the apo structure a water molecule links these residues. When NADP is bound, these two residues form a direct H-bond, while still maintaining their contact with a similarly positioned water molecule.

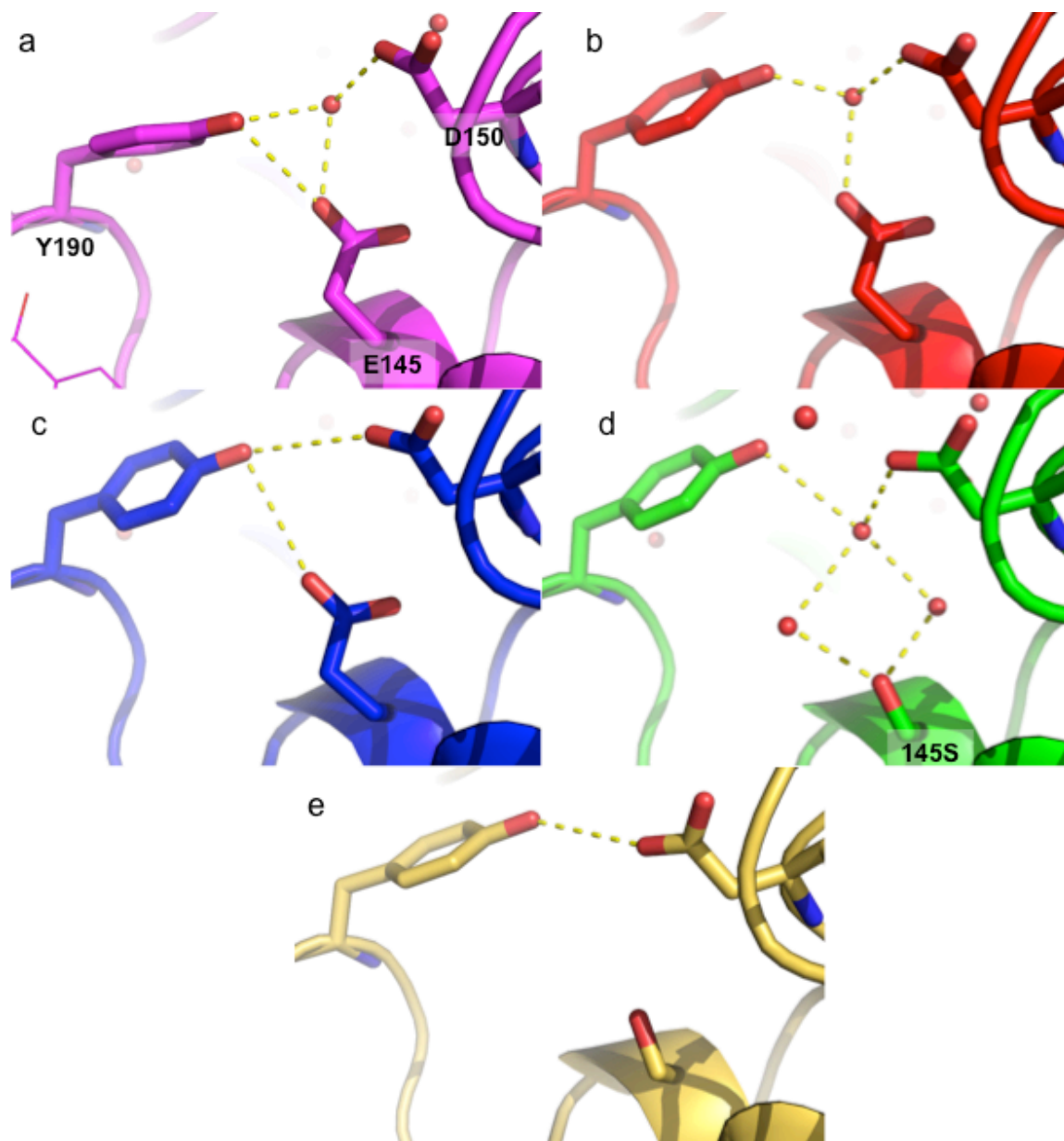


Figure 8.12. The different H-bond contacts involving Y190, E145(S), D150, and water. (a) NADP-bound WT (pink), where the top of NADP is shown in lines, (b) apo WT (red), (c) A94F (blue), (d) E145S (green). (e) Sph (no waters were resolved in this structure).

Figure 8.13 shows the apo WT and apo A94F crystal structures. The A94F mutant has a less solvent-exposed active site, and Phe94 blocks the large binding pocket from the bulk water. In the crystal structure, Phe94 makes a van der Waals contact with leucine 195 on the substrate-

binding loop. A closed loop in the apo structure is especially notable because it is lacking the NADP(H) cofactor, which, in the WT, is needed to hold the loop closed. In the A94F crystal structure, the Tyr190-Glu145 phenolic OH to carboxylate contact shown in Figure 8.12 is lacking the joining water that is present in the apo WT structure. The O-O distance is also slightly longer, at 4.1 Å, compared to 3.9 Å in the apo WT structure.

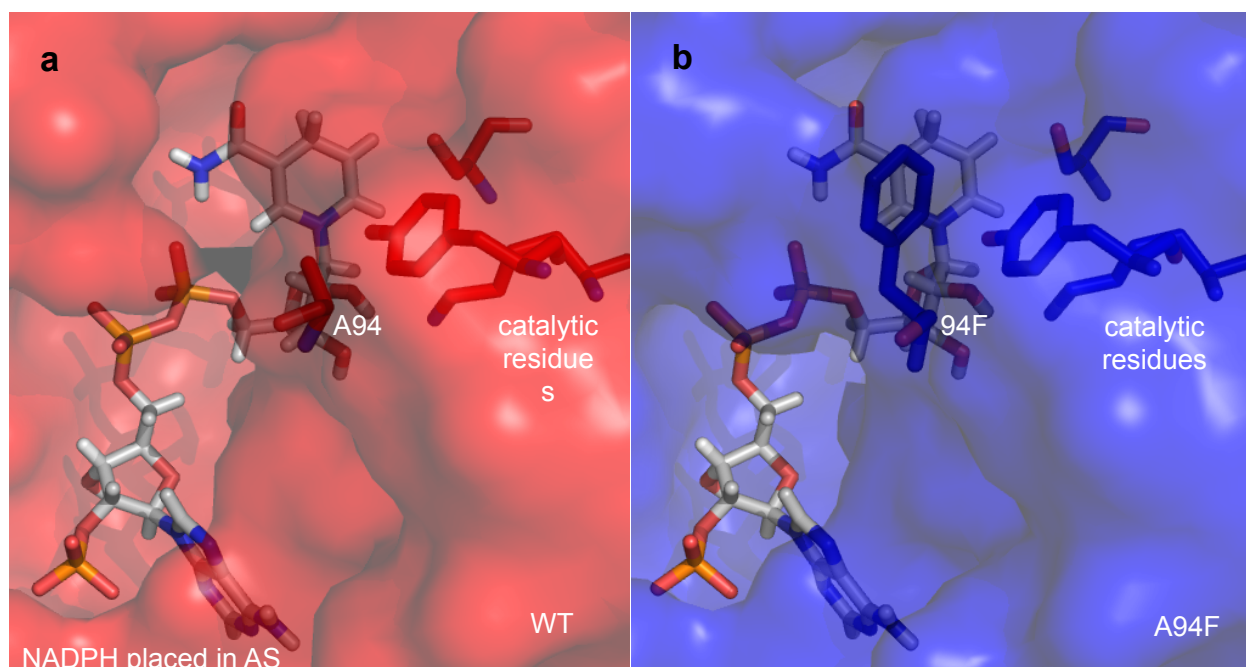


Figure 8.13. Comparison of the (a) apo WT and (b) A94F crystal structures. NADPH is placed in both active sites for clarity. A94F blocks the large binding pocket.

Sph has ten mutations. There are two mutations in the small binding pocket (E145S, N157T), two in the large binding pocket (A94T, S96P), and two in the substrate-binding loop (P194N, M206Q). There are also two mutations in α -helix E¹⁰(R108H, G117S), one in α -helix G (I223V) and one more in a loop (G7S). The conformations and H-bonds of these residues are often different in the Sph and WT crystal structures.

An overlay of the five crystal structures obtained in this work is shown in Figure 8.14. The NADP-bound WT and apo Sph structures overlay well, with a C α RMSD of 0.35 Å. This is less deviation than between the apo WT and apo Sph structures, which have an RMSD of 0.38 Å. As can be seen by the overlay of the crystal structures in Figure 8.14a, the difference in RMSD stems from the fact that Sph more closely resembles the closed loop in the cofactor-bound WT than the open loop in the apo WT.

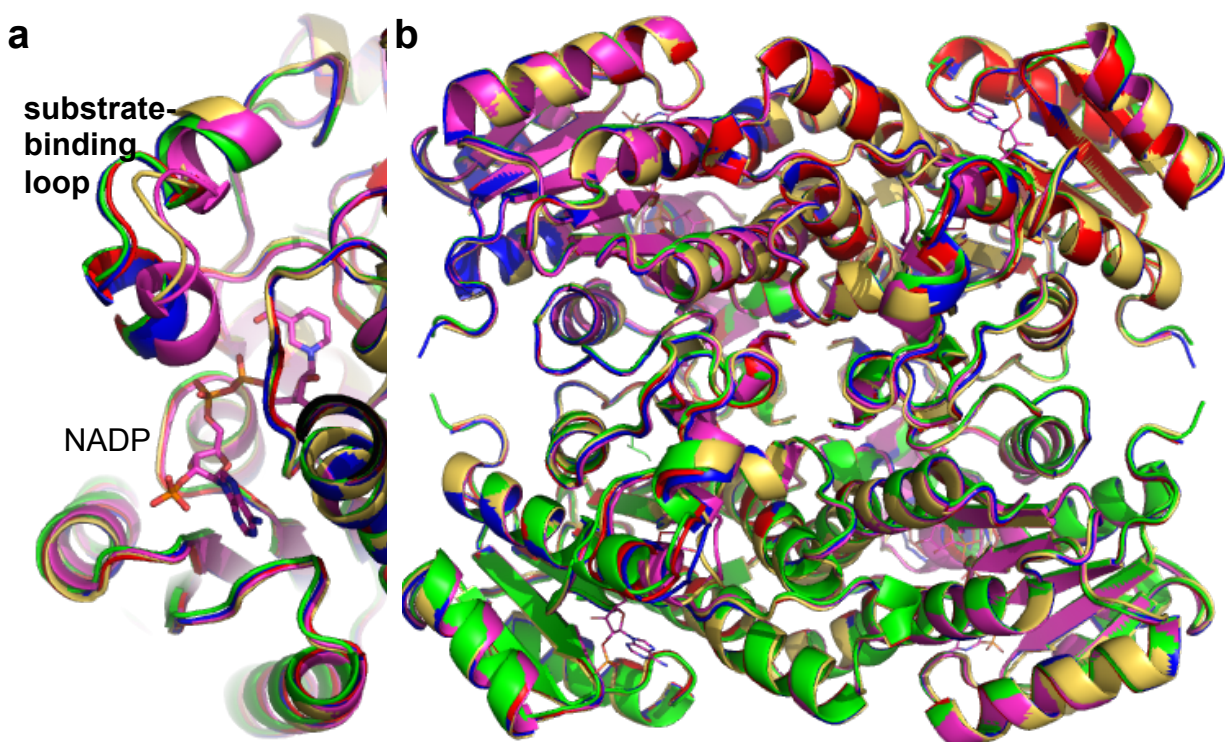


Figure 8.14. Overlay of the five obtained crystal structures: apo WT (red), NADP-bound WT (magenta), A94F (blue), E145S (green), and Sph (yellow). (a) The ribbon representation of the backbone shows the conformations of the substrate-binding loop (b) Overlay of the backbone of the tetrameric crystal structures.

As discussed above, E145S has a serine hydroxyl and two waters where the carboxylate resided in the WT. This region of Sph is shown in Figure 8.12e. The Sph structure has an

Asp150-Tyr190 H-bond, and likely has the same H-bond network involving waters between Ser145 and Tyr190 as observed in the E145S structure (Figure 8.14d).

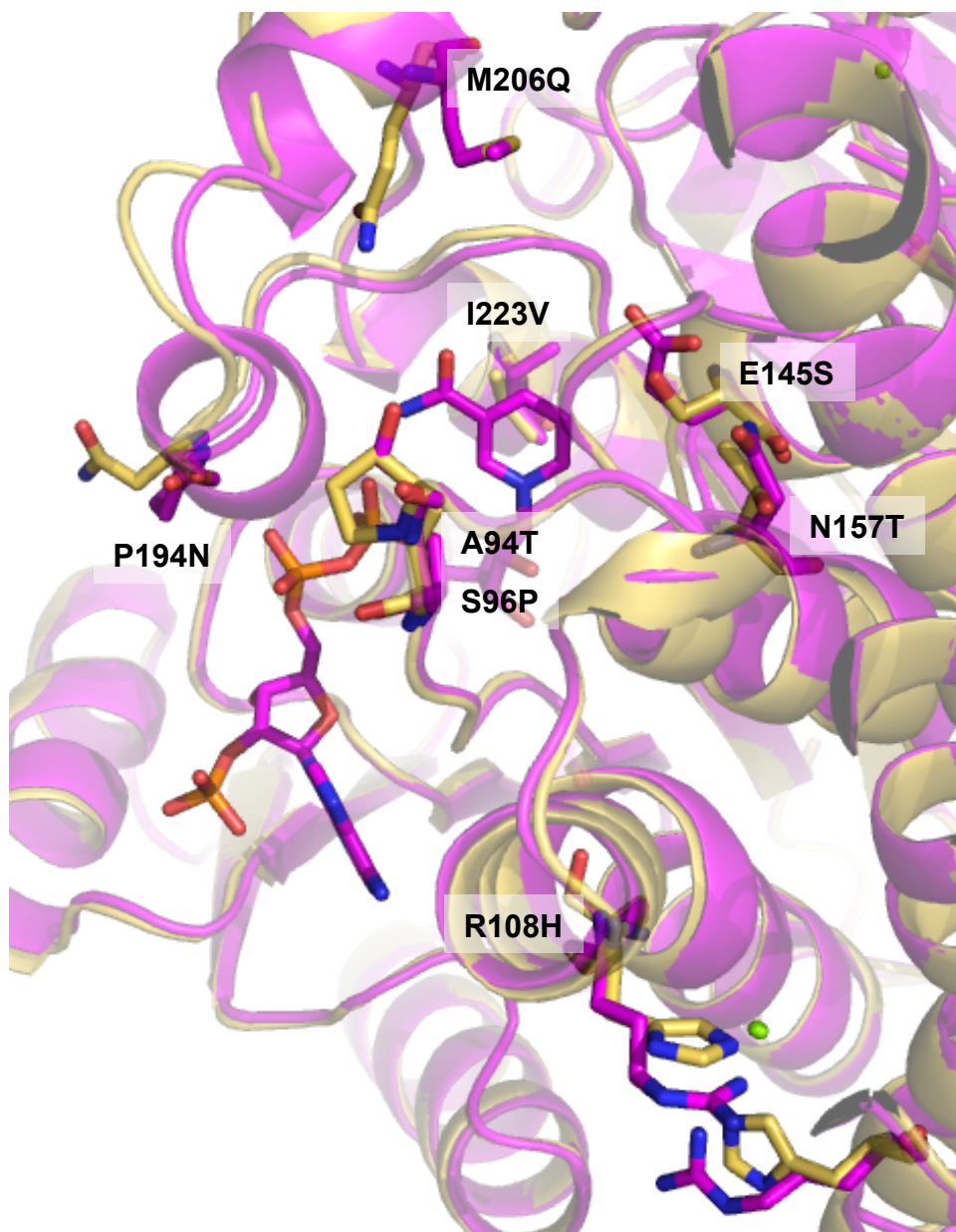


Figure 8.15. Comparison of the NADP-bound WT (pink) and Sph (yellow) crystal structures. The mutations near the active site are shown in stick representations.

Figure 8.15 shows an overlay of the NADP-bound WT and Sph crystal structures, at the active site. The mutations are shown in stick representations. The A94T mutation adds a hydrophilic hydroxyl group into the large binding pocket, whereas the A94F mutation, discussed

earlier, places a hydrophobic group in the large binding pocket. The S96P mutation does not change any H-bond patterns, but proline often increases stability.³²

In the crystal structure of Sph, the P194N mutation lengthens the distance from residue 194 to residue Asp198, and results in a less ordered helix turn in the substrate-binding loop (Figure 8.15). Note the pink helix including residue 194 in the WT becomes a less-ordered loop in Sph (yellow). Also, Asp 198 is resolved in the WT, while the side chain is unresolved in Sph. In Sph the conformation of Gln206 places the amide side chain at the top of the large binding pocket, this unlike in the apo WT, where Met206 sits above Tyr190, outside of either binding pocket. In the NADP-bound WT structure Met206 has a conformation like that of Gln206,.

Standard MD simulations. MD simulations were performed on selected variants with the substrate and cofactor in explicit water. These simulations predict the shape and water accessibility of the active site, including the conformational states of the enzyme. These conformations can be classified according to whether the loop is open, closed or in an intermediate conformation.. The most dynamic region is the substrate-binding loop. Figure 8.16 shows conformations of the open and closed substrate-binding loop. Residues Leu199 and Leu195 in the substrate-binding loop make hydrophobic contacts with Ile93, Ala94, and Ser96 when the loop is closed, with C α distances near 7 Å. In the WT, when the loop is open, the C α distances from Leu195 to Ala94, Leu195 to Ile93, and Leu199 to Ser96 are around 12, 12.5, and 16 Å, respectively, and can reach distances over 26 Å.

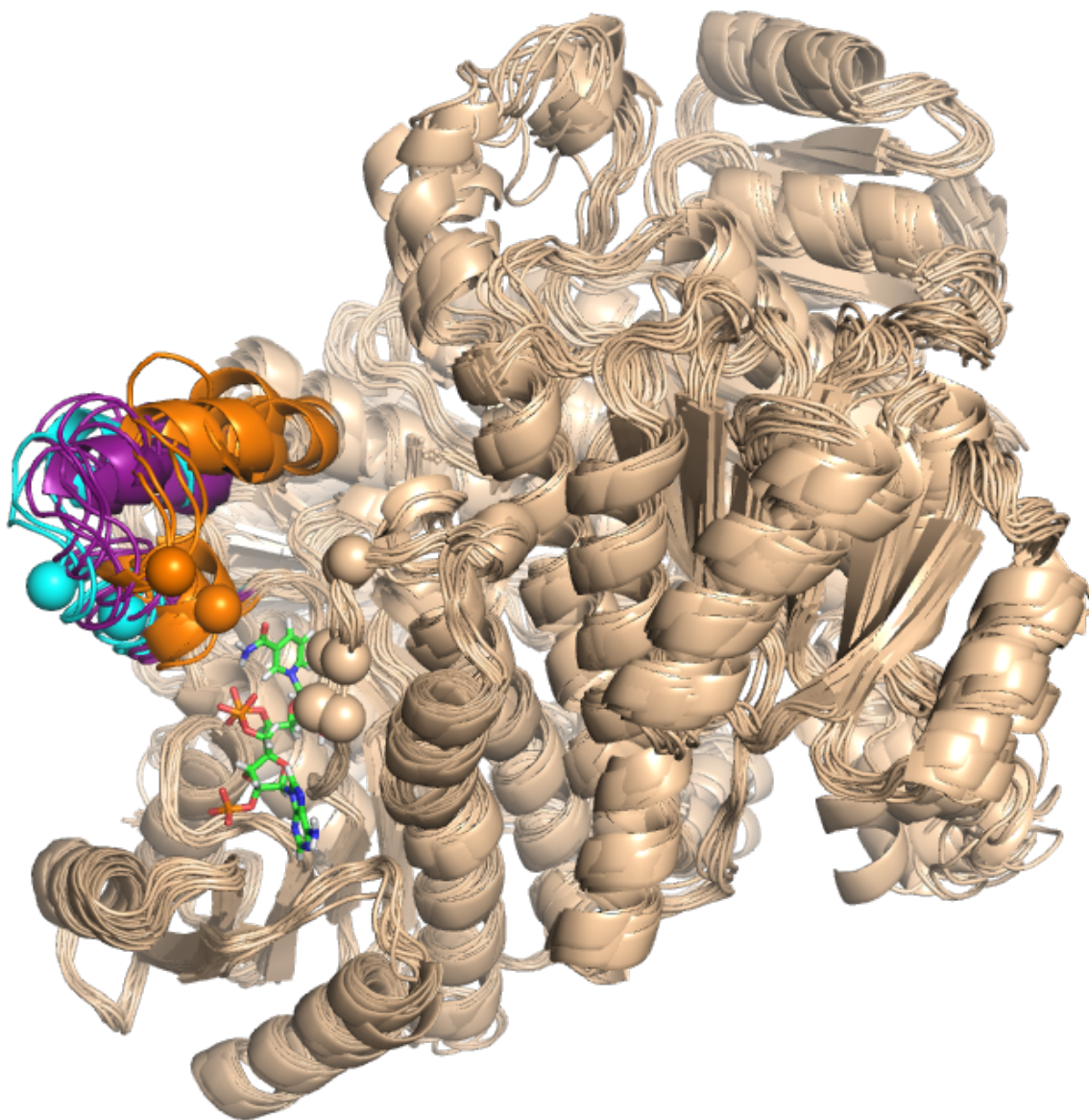


Figure 8.16. Overlay of MD snapshots showing an open and closed conformation of the substrate-binding loop. The loop (residues 190-210) is highlighted in orange (closed), cyan (open), and intermediate (purple) conformations and the cofactor is green. The C α of residues Ile93, Ala94, Ser96, Leu195 and Leu 199 are shown as spheres.

Figure 8.17 shows the distances between the C α of the residue pairs that were sampled in the MD simulation of the WT. The variation in these distances illustrates the dynamics of the loop. The substrate binds nonspecifically, and only rarely samples a productive conformation.

Additionally, the quantum mechanical *theozyme* calculations described earlier predict that binding the substrate is disfavored by 3-4 kcal/mol.

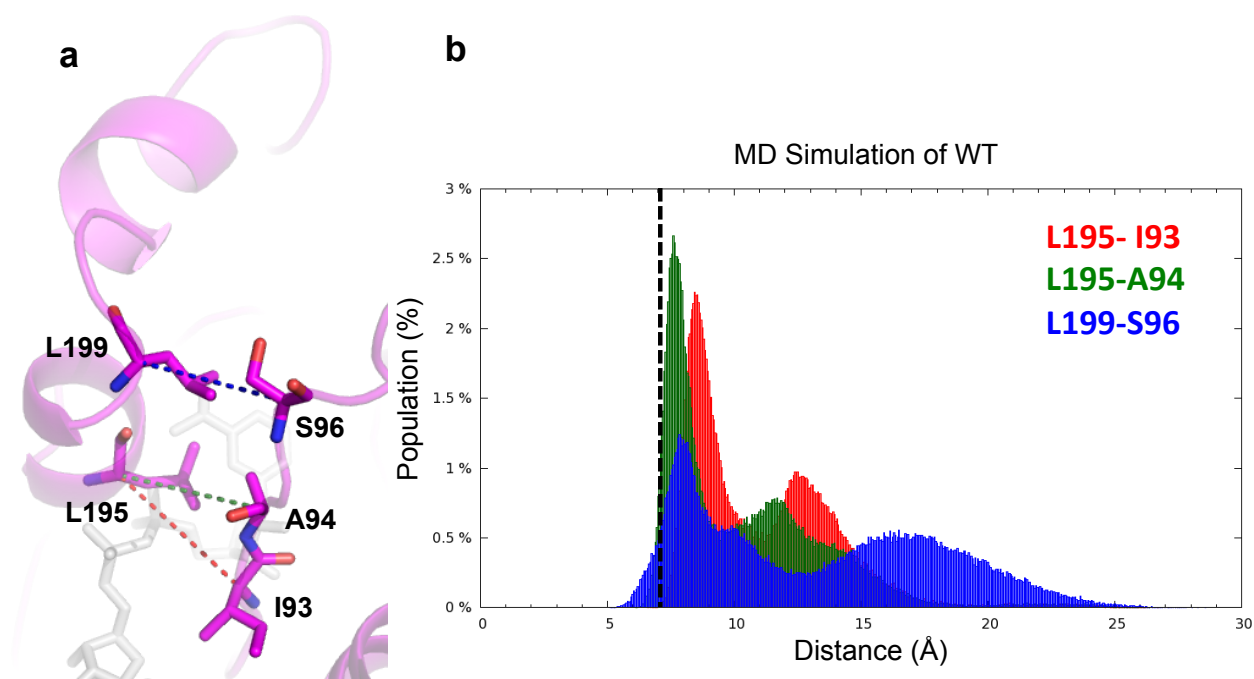


Figure 8.17. MD simulation of the WT. (a) The contacts made with the substrate-binding loop at the opening of the small binding pocket in the WT simulation. The hydrophobic contacts are shown by dashed lines between Leu199 and Ser96 (blue), Leu195 and Ala94 (green), and Leu195 and Ile93 (red). (b) The histograms show the distances between C α of interacting residues: Leu199 with Ser96 (blue), Leu195 with Ala94 (green), and Leu195 with Ile93 (red) in the 2900 ns MD simulations of the WT. The distances in the NADP-bound WT x-ray crystal structure are 7 Å, indicated by the dashed line.

The loop dynamics will occur relatively slowly but will maximize the interactions of the catalytic groups, including the cofactor, and the substrates. Mutations affect the dynamics of the loops, and consequently the catalysis. In A94F, the loop is found to remain closed significantly more often, although the C α distances for the closed conformation are longer than that of the WT due to the introduction of the large phenyl group. In A94F, the plots of distances sampled from Leu195 to Ile93 and to Phe94 have one peak each at 9 and 11 Å. These distances are longer than is seen in the crystal structure (8.4 and 9.4 Å, respectively). On the other hand, E145S and Sph

sample open conformations to a greater degree than the WT. Here the loop appears to have roughly 3 conformations. Although the crystal structure of Sph has a more closed conformation of this loop, the MD simulations indicate that a closed loop conformation is not highly stable.

MD simulations with the *theozyme*. MD simulations of the tetrameric enzymes were performed with the *theozyme* and the NADPH cofactor in one active site and only the NADP cofactor in the others. Pauling proposed, and others have since elaborated, that enzymes are complementary in structure to the transition state of the reaction that they catalyze.³³ These simulations are intended to provide a more quantitative assessment of the stability of catalytic arrangement near the TS complex.^{34,35} Figure 8.18 shows the catalytic contact distances observed from the MD simulations. We simulated two diastereomeric *theozymes* (for the formation of *R*- and *S*-alcohols) and observed the amount of time during the MD simulation in which the active site residues are in an arrangement suitable for catalysis. Four of the catalytic distances were monitored (see Figure 8.18) in order to assess relative stability of the catalytic arrangement with regard to two of the TS structures: the distance from (1) the oxygen of Tyr156 to the developing alkoxide oxygen in the substrate, (2) the oxygen of Ser143 to the developing alkoxide oxygen of the substrate, (3) the oxygen of the NADPH sugar to the oxygen of Tyr156, and (4) the oxygen of the NADPH sugar to the nitrogen of Lys160. Here the catalytically competent conformation is defined as: all four catalytic distances defined above are shorter than 3.4 Å.

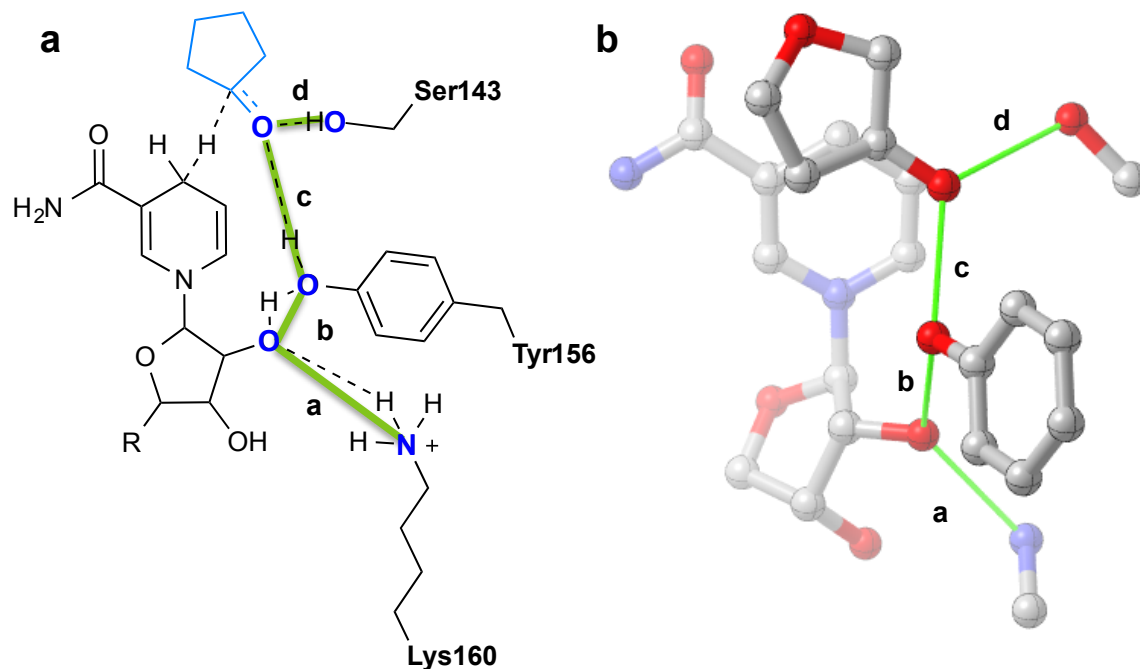


Figure 8.18. The catalytic distances monitored in the MD simulations with the *theozymes*. The lettered (a, b, c, d) green lines indicate the distances monitored. (a) ChemDraw representation (b) CYLView representation with the *S*-oxa *theozyme* as the example. The distances from the QM calculation are a. 2.8, b. 2.7, c. 2.6, and d. 2.8 Å.

Figure 8.19 shows a plot of the experimental *er* for several mutants and the difference between populations of catalytically competent *theozyme* conformations from the MD simulations. The probability that the enzyme maintains the catalytic contacts to each diastereomeric transition state within the 3.4 Å heavy atom cutoff distance was found to roughly correlate with the experimental enantiomeric ratio. In what follows we provide detailed analysis on how individual mutations alter energetics of the active-site and modulate enantioselectivity of the enzyme.

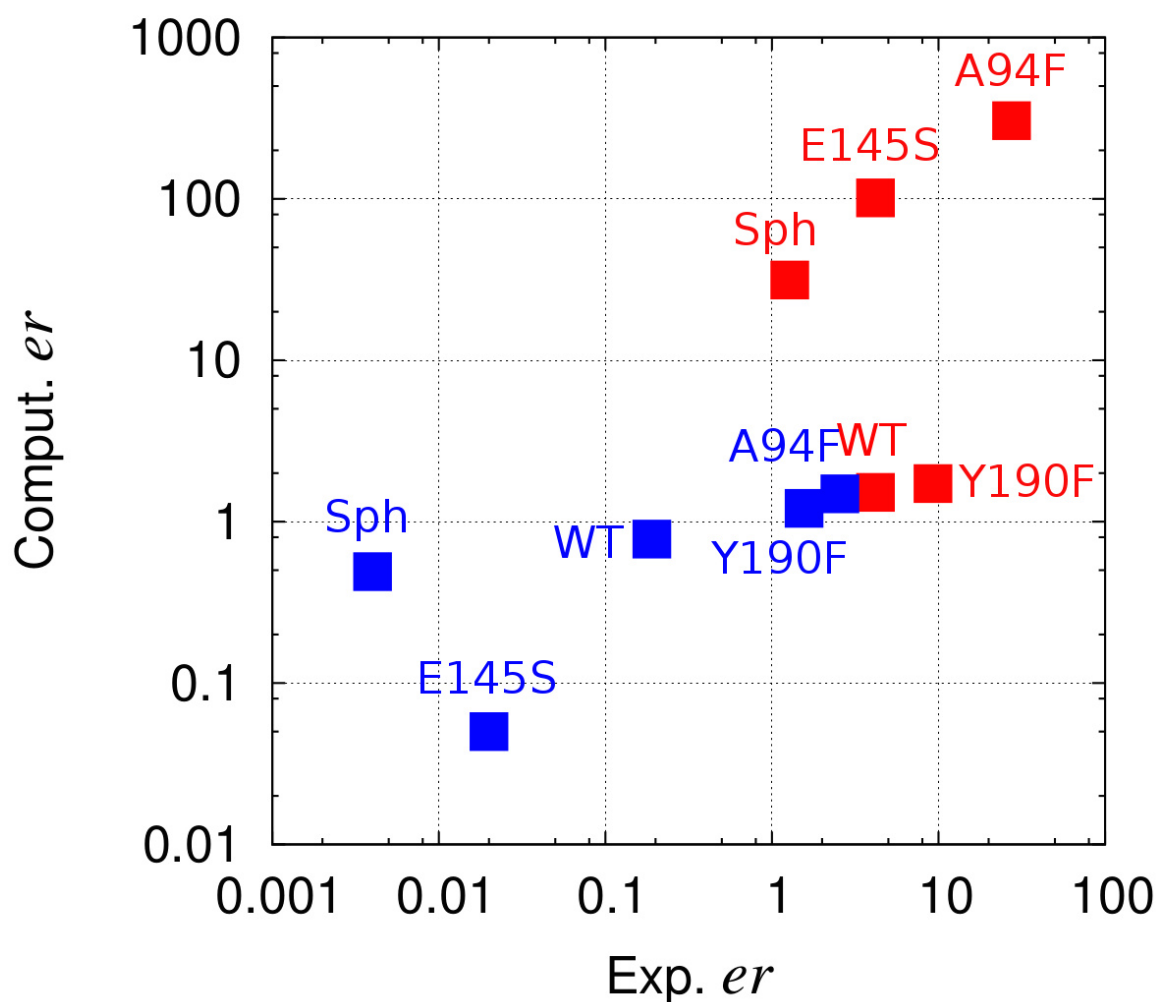


Figure 8.19. Log/log graph of the experimental *er* versus computed *er* from the MD simulations, where *er* is computed as $\%competentS/\%competentR$. Oxa is shown by red squares and thia is shown by blue squares.

The enzymes each have a small and a large binding pocket, which accommodate the two substituents of the ketone, or in these cyclic cases the CH₂ or the S or O. The MD snapshot in Figure 8.20 shows the active site of the WT enzyme with the cofactor, substrate, catalytic residues, Ala94, Tyr190, and Glu145 shown in sticks. In the WT protein, the small binding pocket is sterically hindered by Tyr190 and disallows water. The cavity in the large binding pocket that is filled with the phenyl group when acetophenone is bound fills with water when **1**

or **2** is bound, since the smaller substrates do not completely fill this space. Shown in Figure 8.20, several waters are present in the large binding pocket near the *R*-thia *theozyme*. Thus, the small binding pocket is relatively non-polar, and the large binding pocket is polar. The match of both the size and polarity of the substrates to those of the enzyme active site determines the enantioselectivities for these substrates, since the size difference is rather small for both substrates. Nevertheless, in the WT, for **1**, the size is the determining factor. In the favored *S*-orientation, the small oxygen binds to the small binding pocket, and the larger methylene group binds to the large binding pocket, even though the polarity is mismatched. This mismatch gives rise to a modest 4.19 *er* (61 % *ee S*) of the alcohol from **1**. The size and polarity of **2** both match that of the WT enzyme when **2** is reduced to the *R* alcohol, such that the large polar sulfur is in the large polar binding pocket, and the small nonpolar methylene group binds to the small nonpolar binding pocket as pictured in Figure 8.20. This results in 0.19 *er* (68 % *ee R*) of the alcohol formed from **2**. Interestingly, the quantum mechanical calculation on the reduction (Figure 8.7) is consistent with the selectivity of the WT, predicting 5.4 *er* (69 % *ee S*) for **1** and 0.08 *er* (86 % *ee R*) for **2**.

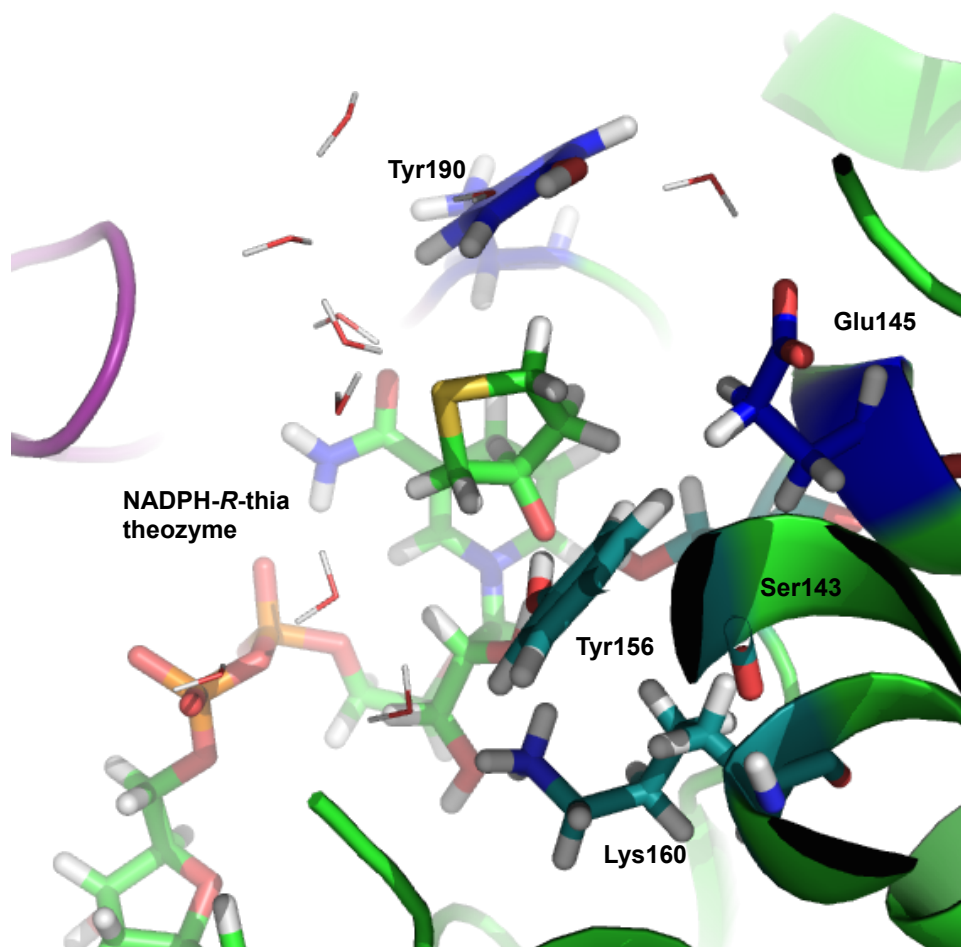


Figure 8.20. MD snapshot of WT with the TS complex of *R*-thiacyclopentanone. The cofactor and substrate (light green), Tyr190, Glu145, (blue), and catalytic residues (green) are shown as sticks, substrate binding loop (purple) and rest of the enzyme (light green) are shown as cartoons. The waters in the active site are shown as lines.

The plots of the four catalytic distances monitored in the simulations of the WT with each of the *theozymes* are shown in Figure 8.21. These simulations predict 1.5 *er* (21 % *ee S*-oxacyclopentanol), compared to the experimental value of 4.2 *er* (61 % *ee S*) and 0.25 *er* (12 % *ee R*-thiacyclopentanol), while the experimental *er* is 0.8 (68 % *ee R*).

WT

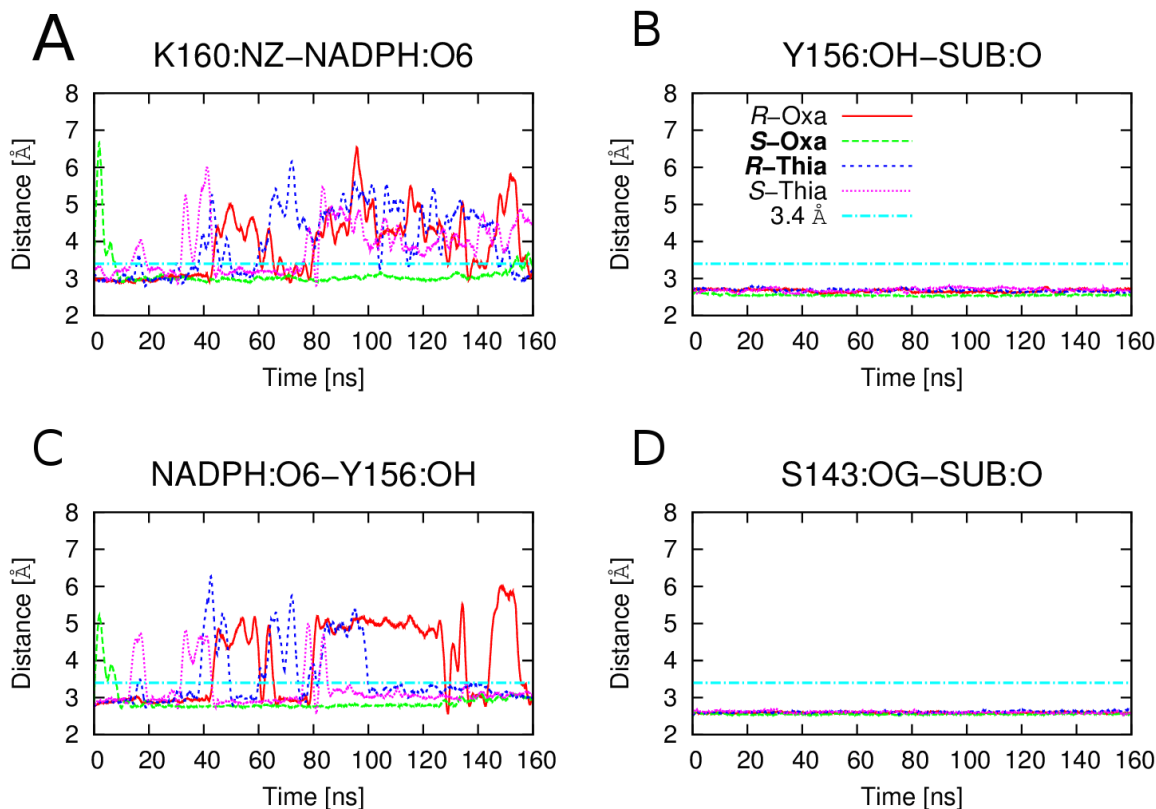


Figure 8.21. WT simulation with *theozymes*. Plots of the four catalytic distances monitored in the simulations of the WT with the *R*-oxa (red), *S*-oxa (green), *R*-thia (blue), *S*-thia (pink) *theozymes*. The cutoff distance (3.4 Å) is shown as a light blue line and the experimentally favored enantiomer is marked in boldface. (a) The side chain N of Lys160 to the sugar O. (b) Tyr156 O to substrate O. (c) NADPH sugar O to Tyr156 O. (d) Ser143 O to substrate O. The noise is eliminated from the distance plots with a smoothing algorithm.^{36,37}

When a single residue in the active site, alanine, is mutated to a phenylalanine (A94F), the selectivity is enhanced for **1** and is reversed for **2** (Table 8.1). The A94F mutant is particularly interesting, because of all the single-mutation variants, A94F gives the largest change in selectivity for thiacyclopentanol. The A94F mutant gives 27.5 and 2.6 *er* from **1** and **2**, respectively (100 % ee *S*-oxacyclopentanol, and 44 % ee *S*-thiacyclopentanol). The A94F mutation also lowers the efficiency of the enzyme for **2** by two orders of magnitude, from 1.46 to 0.0164 min⁻¹mM⁻¹, by blocking the normally favored *R*-pathway. In the *S*-orientation, which is

favorable for both substrates, the non-polar methylene of the substrate binds to the large binding pocket.

Figure 8.22 shows snapshots from the MD simulations of A94F with the *S*- (favored) and *R*- (disfavored) oxa transition structures, while Figure 8.23 shows the distance plots during the simulations with *theozymes*. In the simulations with the oxa transition structures, there are favorable hydrophobic interactions between the methylene of the *S*- and *R*-oxa transition structure and Phe94, causing the methylene of the oxa transition state to orient toward the hydrophobic phenylalanine. In the simulation of the favored *S*-oxa transition structure, shown in Figure 8.22a, the catalytic arrangement is maintained. In this simulation, the average distance between C4 of the substrate and the center of the Phe94 phenyl ring is 4.3 ± 0.7 Å, which can be stabilizing on the order of 1 kcal/mol.³⁸ At 5 Å, the interaction energy between a methylene and benzene is about 0.3 kcal/mol.³⁷ The C4-Phe94 distance is less than 5 Å for 92 % of the simulation time. When the *S*-oxa *theozyme* is in a catalytically competent arrangement, the number of waters within the first solvation shell (3.4 Å)³⁹ of C4 is decreased compared to that of the WT from 1.7 ± 0.9 to 1.1 ± 0.9 . Fewer waters in this pocket corresponds to a decrease in polarity of the pocket.

The snapshot in Figure 8.22b shows the disfavored *R*-oxa transition structure. Here the Phe94–methylene interaction disrupts the catalytic arrangement, and several critical H-bond contacts are lost in the disfavored transition state. When the *R*-oxa *theozyme* is in contact with the catalytic amino acids, the oxygen in the ring experiences a more hydrophobic environment as the number of waters near O is decreased from 2.2 ± 0.7 in the WT to 1.0 ± 0.7 .

These simulations predict 305.0 *er* (100 % *ee* for oxacyclopentanol), which overestimates the experimental value of 27.5 *er* (99 % *ee*). *S*-oxa is predicted to be highly

favored even though in the *S*-oxa simulation the cofactor to Lys160 and Tyr156 distances are not maintained well (green line in Figure 8.22a and c), because in the *R*-oxa simulation the Ser143-substrate distance (red line in Figure 8.22d) is always long.

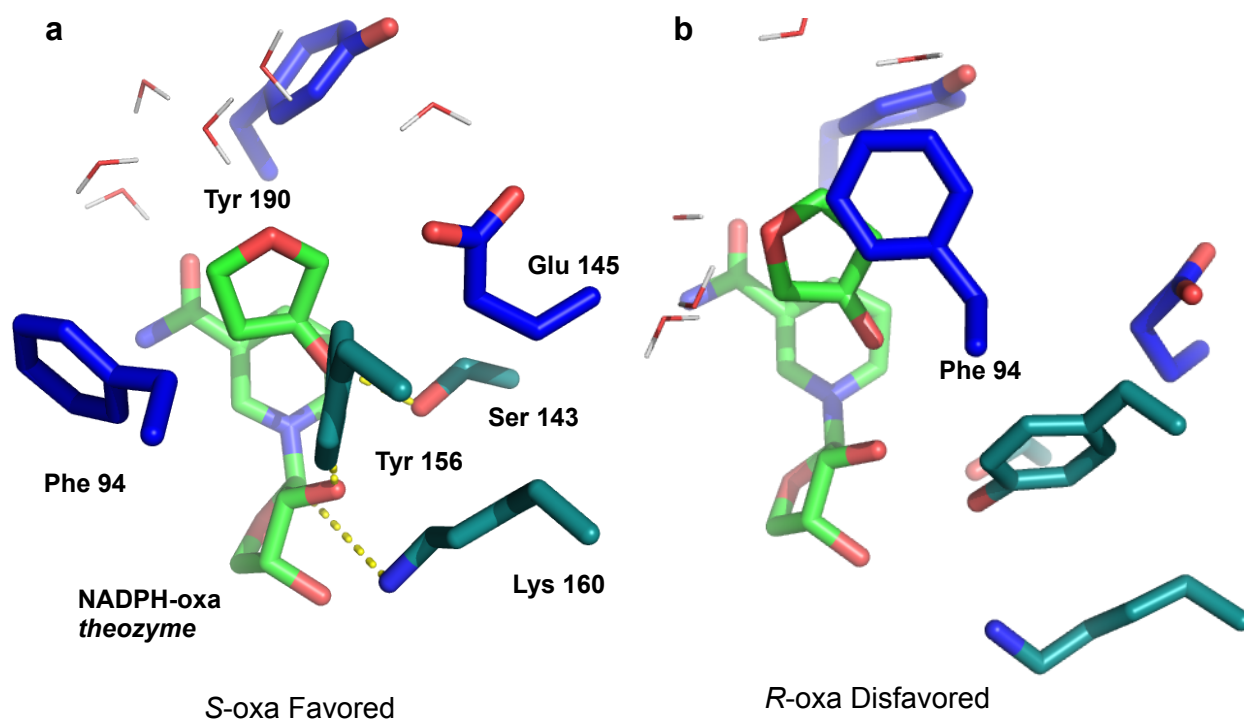


Figure 8.22. Snapshots from the MD simulations of A94F with the (a) favored *S*- and (b) disfavored *R*-oxa transition structures. The cofactor and substrate (light green), Tyr190 and Glu145 (blue), and catalytic residues (green) are shown as sticks. The water in the active site are shown as lines.

The same phenomena are seen in the thia case, which now favors the *S* product. In the favored *S*-thia simulation, the average distance between C4 of the substrate and the center of the Phe94 phenyl ring is 5.0 ± 0.6 Å, and this distance is under 5 Å for 48 % of the simulation time. When the *S*-thia theozyme is in a catalytically competent arrangement, the number of waters near C4 decrease from 1.7 ± 0.9 in the WT to 1.1 ± 0.9 . These simulations predict 1.5 *er* (20 % *ee*) for thiacyclopentanol; experimentally this enzyme gives 2.6 *er* (44 % *ee*).

A94F

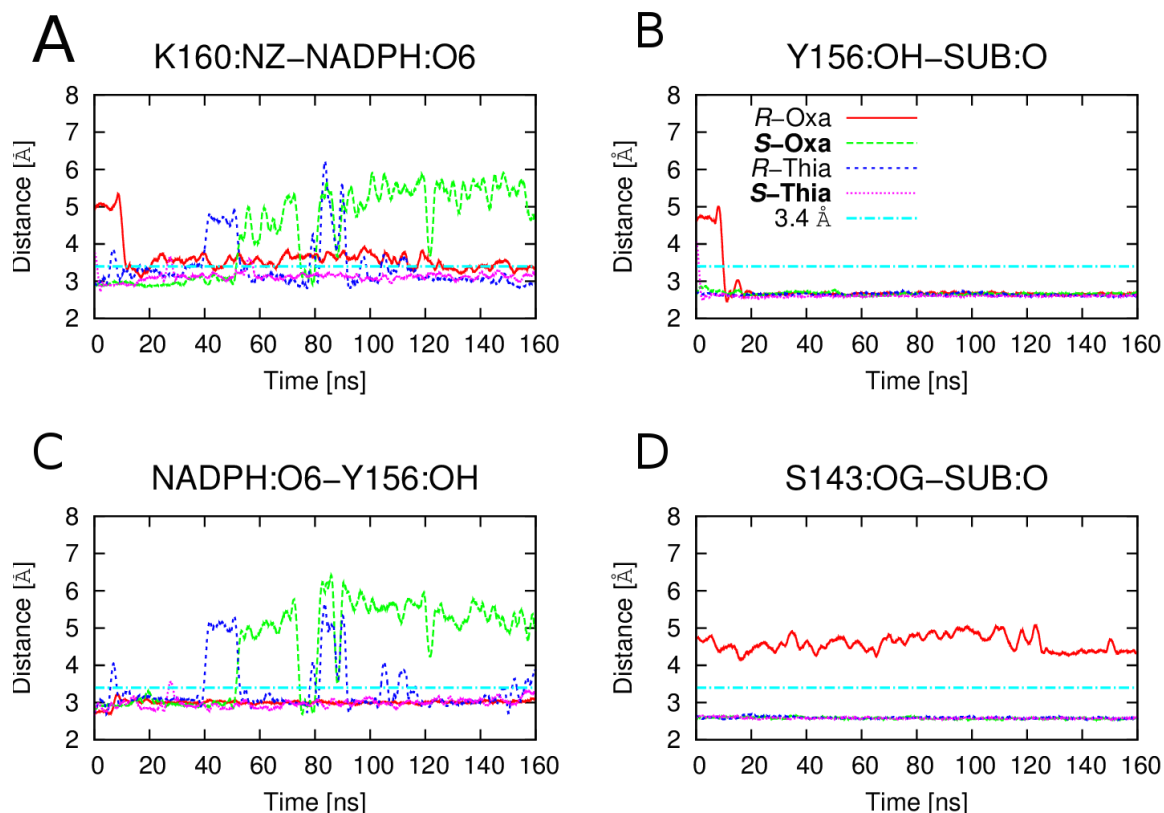


Figure 8.23. A94F simulation with *theozymes*. The plots of the four catalytic distances monitored in the simulations of the A94F mutant with the *R*-oxa (red), *S*-oxa (green), *R*-thia (blue), *S*-thia (pink) *theozymes*. (a) The side chain N of Lys160 to the sugar O. (b) Tyr156 O to substrate O. (c) NADPH sugar O to Tyr156 O. (d) Ser143 O to substrate O.

There is a moderate correlation between hydrophobicity⁴⁰ of the amino acid at residue 94 and the enantioselectivity. The *S*-selectivity is further increased by a second mutation near the large binding pocket, S96Y. The triple mutant (A94F, S96Y, F147L) gives 24.7 and 6.8 *er* (100 and 72 % *ee*) for **1** and **2**, respectively. A similar correlation is seen between the hydrophobicity of the amino acid at residue 96 and the enantioselectivity.

MD snapshots of E145S with *S*- and *R*-thia transition structures are shown in Figure 8.24, and the plots of the catalytic distances monitored in the simulations of the E145S mutant are shown in Figure 8.25. Mutating 145 from glutamate to serine, as in E145S and in Sph, eliminates the H-bond contact between Tyr190 and Glu145. Instead, Tyr190 H-bonds to Asp150 or

surrounding water molecules, which brings Tyr190 closer to the transition structure. On average, Tyr190 is 1.3 Å closer to C4 of the *R*-thia transition structure in the simulation with E145S than in the WT (3.9 ± 0.4 vs. 5.2 ± 0.9 Å, respectively). The close proximity of Tyr190 to the transition structure ensures that 2 is reduced to *R*-thiacyclopentanol, as it places the smaller methylene under Tyr190, thus improving the hydrophobic contact between Tyr190 and the methylene group of the substrate in the *R*-thia transition structure (Figure 8.24a). The *R*-thia transition structure also has the sulfur in a polar region of the active site, the large binding pocket. In the simulation of the E145S mutant with the favored *R*-thia *theozyme*, all four contacts are quite stable, and when they deviate the distances never become extremely long (all of the distances stay below 6 Å).

In the disfavored, *S*-thia simulation (Figure 8.24a), the unfavorable contact between Tyr190 and sulfur forces the transition structure to move out from under Tyr190 and away from the catalytic residues, before eventually completely leaving the active site. These simulations predict 0.05 *er* (90 % *ee R*-thiacyclopentanol), which compares well to the experimental value of 0.02 *er* (100 % *ee R*).

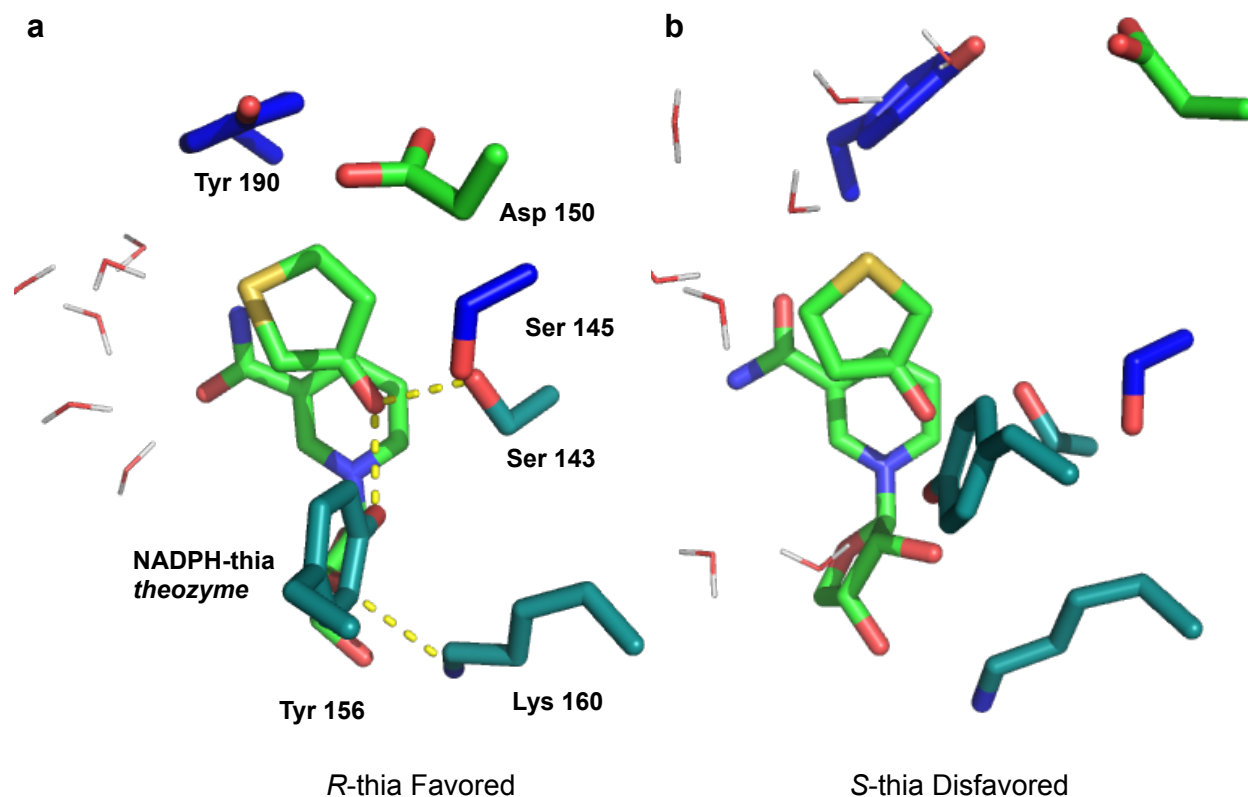


Figure 8.24. Snapshots from the MD simulations of E145S with the (a) favored *R*-thia and (b) disfavored *S*-thia transition structures. The catalytic residues, Ser143, Tyr156, and Lys160 are shown in dark green. Asp150 and the transition state complex are shown in light green, and Tyr190, and Ser145 are shown in blue. The yellow dashes indicate H-bonds.

The E145S mutation does not change the enantioselectivity of **1**. Here the size model still holds; that is small oxygen binds to the small binding pocket, and the larger methylene bind to the large binding pocket. In the simulation with the favored *S*-oxa *theozyme*, all of the contacts are better maintained than in the *R*-oxa simulation, but none of them are maintained the entire time. These simulations predict 101.4 *er* (98 % *ee* for *S*-oxacyclopentanol), the experimental *er* is 4.2 (61 *ee* %). The simulations overestimate the *er* because in the disfavored *R*-Oxa simulation, the Lys160-cofactor contact (red line in Figure 8.24a) is lost for almost the entire simulation.

E145S

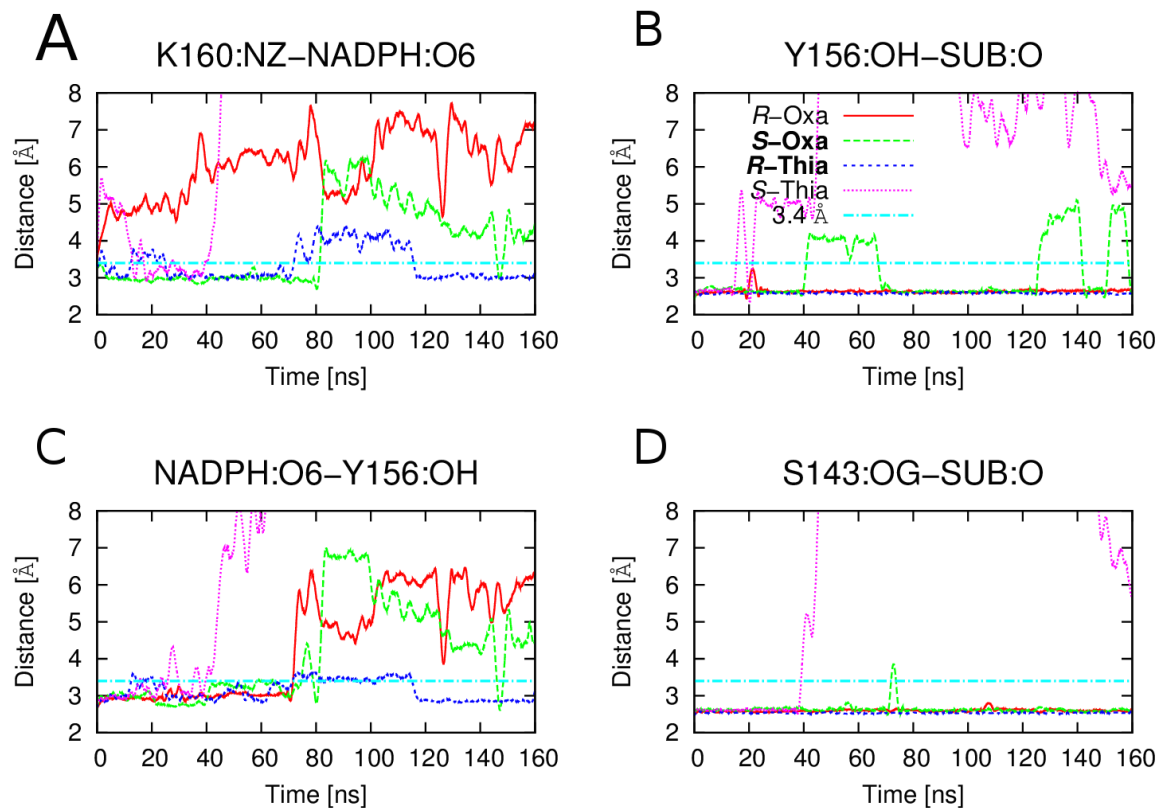


Figure 8.25. E145S simulation with *theozymes*. The plots of the four catalytic distances monitored in the simulations of the E145S mutant with the *R*-oxa (red), *S*-oxa (green), *R*-thia (blue), *S*-thia (pink) *theozymes*. (a) The side chain N of Lys160 to the sugar O. (b) Tyr156 O to substrate O. (c) NADPH sugar O to Tyr156 O. (d) Ser143 O to substrate O.

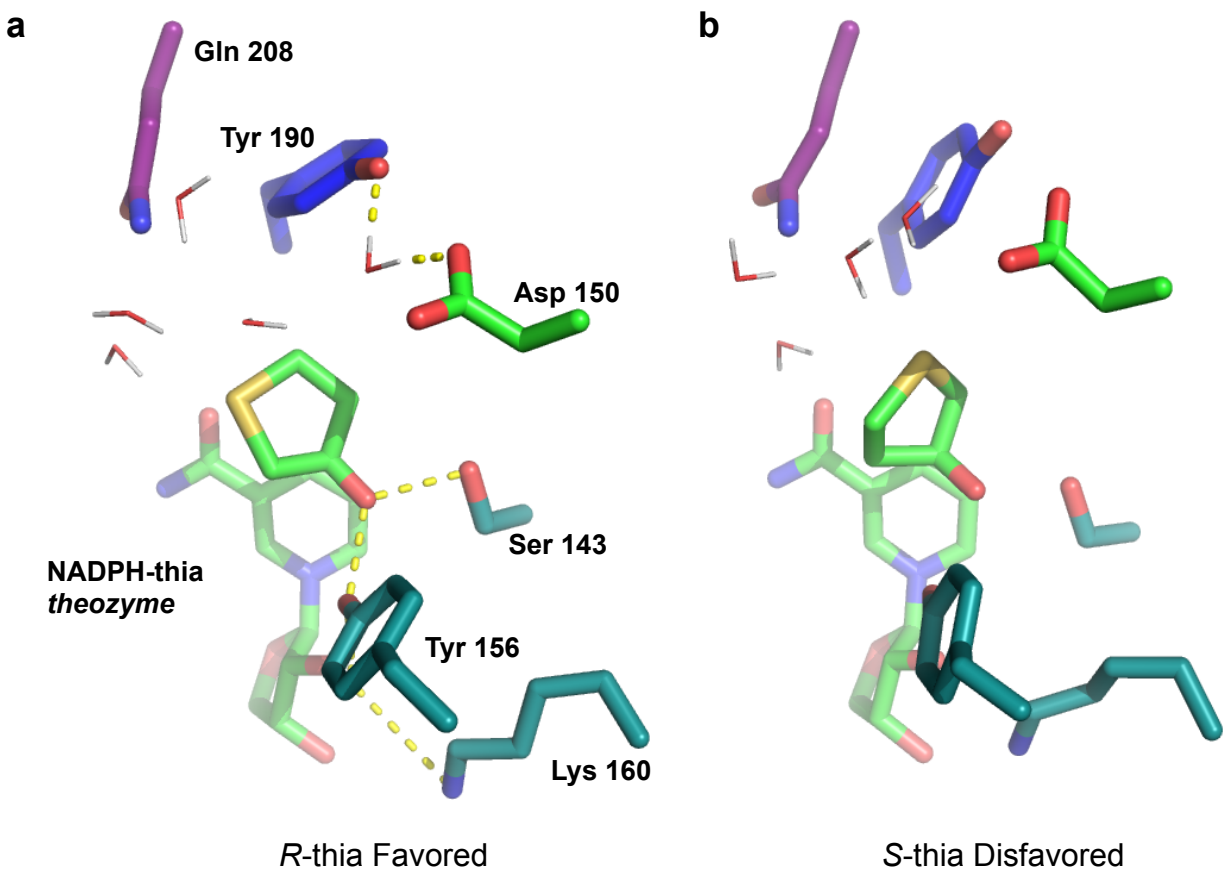


Figure 8.26. Snapshots from the MD simulations of Sph with the (a) *R*- and (b) *S*-thia transition structures. The catalytic residues, Ser143, Tyr156, and Lys160 are shown in dark green. Asp150 and the transition state complex are shown in light green, Tyr190 is shown in blue, and Gln206 is shown in purple. The yellow dashes indicate H-bonds.

In Sph, the E145S mutation gave the largest increase in *R*-selectivity in the evolution process, and the same explanation of enantioselectivity holds. That is the selectivity for *R*-thiacyclopentanol is improved because the conformational shift of Tyr190 makes the small binding pocket smaller and improves the hydrophobic contact between Tyr190 and the methylene group at C4 of the substrate in the *R*-thia transition structure. In the simulation with Sph, the average substrate-methylene to Tyr190 distance is 4.7 ± 1.0 Å. Figure 8.26 shows the active site of Sph with the *R*- and *S*-thia transition structures and shows how the environment of the active site better matches the *R*-thia transition structure (Figure 8.26a). In addition to Tyr190

H-bonding to Asp150, the M206Q mutation introduces the amine of the glutamine side chain to the large binding pocket able to stabilize the sulfur in the *R*-thia transition structure.

In the simulation of the disfavored *S*-thia transition structure sulfur, which is large and polar, is placed under Tyr190 (Figure 8.26b). This orientation has the size and polarity of the substrate mismatched with that of the active site.

While the selectivity is dictated by the same phenomena caused by the E145S mutation, the Sph mutant maintains the catalytic contacts to the *theozymes* much more often. Plots of the catalytic distances monitored in the simulations of the Sph mutant are shown in Figure 8.27. Both of the thia simulations are stable and the four catalytic contacts are maintained for 72 % and 35 % of the simulation for *R*- and *S*-thia, respectively. In the disfavored *S*-thia simulation, the cofactor-Tyr156 distance hovers near the 3.4 Å cutoff. These simulations predict 0.49 *er* (34 % *ee* *R*-thiacyclopentanol), which is an underestimate of the experimental value of 0.004 *er* (99.3 % *ee* *R*¹⁵). The simulations underestimate the *er* because both *S*- and *R*-thia simulations are stable.

The simulation with the favored *S*-oxa *theozyme* is quite stable as well, where all four contacts are maintained for 139 ns of the entire simulation time (160 ns). The simulation with the disfavored *R*-oxa *theozyme* is the least stable, maintaining all four contacts for only 4.5 ns of total 160 ns simulation, where the cofactor-Lys160 and substrate-Tyr156 (red line in Figure 8.27a and b, respectively) contacts are mostly lost. This results in a predicted overestimated *er* of 31.4 (of 94 % *ee*), compared to the experimental estimated *er* of 1.3 (12 % *ee*⁴¹).

Sph

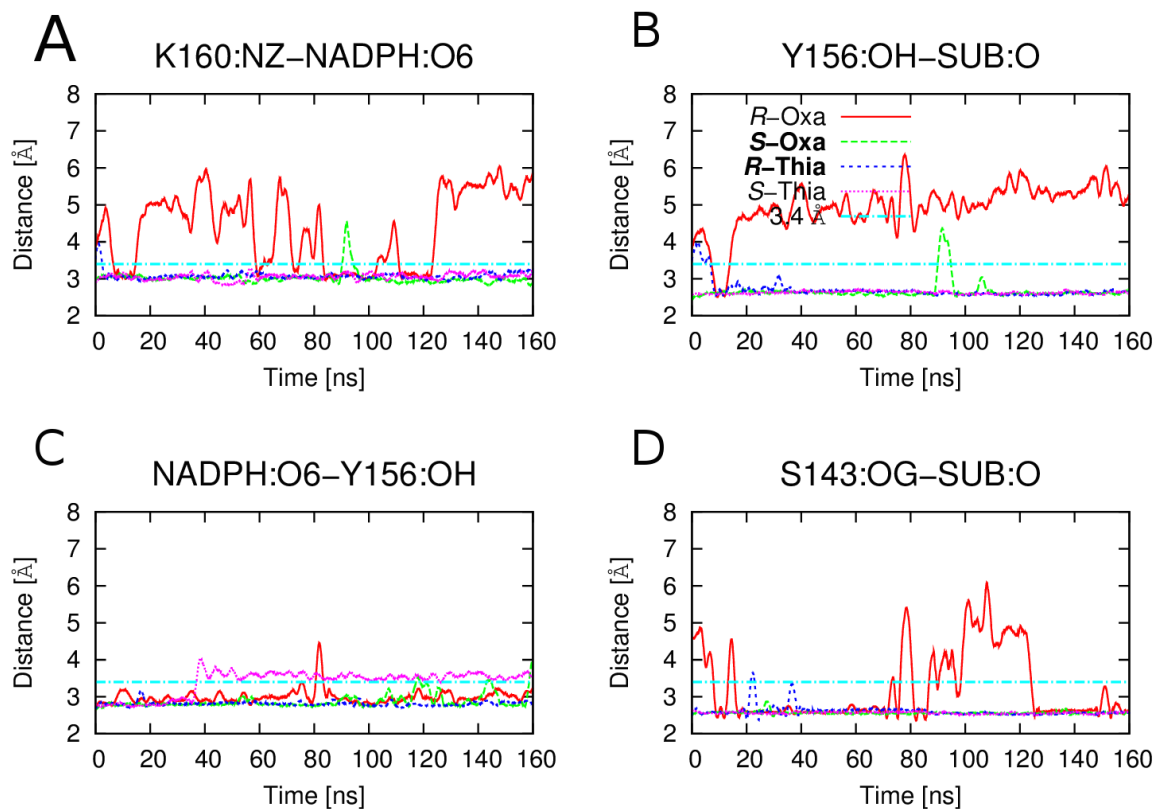


Figure 8.27. Sph simulation with *theozymes*. Plots of the four catalytic distances monitored in the simulations of the Sph mutant with the *R*-oxa (red), *S*-oxa (green), *R*-thia (blue), *S*-thia (pink) *theozymes*. (a) The side chain N of Lys160 to the sugar O. (b) Tyr156 O to substrate O. (c) NADPH sugar O to Tyr156 O. (d) Ser143 O to substrate O.

Conclusion

The mutations to *L. kefir* KRED change the size and water-accessibility of the large and small binding pockets, modulating the stereoselectivity of these enzymes. We show that point mutations change the size and polarity of the two binding pocket regions of the active site by allowing water to enter, or introducing new hydrophobic or hydrophilic contacts. The mutation from alanine to phenylalanine in the large binding pocket at residue 94 displaces water and creates a hydrophobic contact for the substrate. This orients the non-polar groups toward the large binding pocket, effectively enhancing the *S*-selectivity toward both substrates. Mutating

Tyr190, in the small binding pocket, to phenylalanine increases the polarity of the small binding pocket, by allowing water to enter, and also enhances the enzyme *S*-selectivity. The E145S mutation changes the conformation of the Tyr190, which makes a hydrophobic contact with the substrate in the small binding pocket. Additional mutations in Sph slightly enhance the *S*-selectivity toward **2**, and allow the enzyme to maintain better catalytic contacts to the transition structure. We believe that KRED enzymes can be further improved by modulating the size and polarity of the active site to control enantioselectivity.

References.

- ¹ Huisman G. W.; Liang J.; Krebber A. *Curr. Opin. Chem. Bio.* **2010**, *14*, 122.
- ² Moore, J. C.; Pollard, D. J.; Kosjek, B.; Devine, P. N. *Acc. Chem. Res.* **2007**, *40*, 1412.
- ³ Strohmeier, G. A.; Pichler, H.; May, O.; Gruber-Khadjawi, M. *Chem. Rev.* **2011**, *111*, 4141.
- ⁴ Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature* **2012**, *485*, 185.
- ⁵ Oppermann, U.; Filling, C.; Hult, M.; Shafqat, N.; Wu, X.; Lindh, M.; Shafqat, J.; Nordling, E.; Kallberg, Y.; Persson, B.; Jörnvall, H. *Chem. Biol. Interact.* **2003**, *143*, 247.
- ⁶ Weckbecker, A.; Hummel, W. *Biocatal. Biotransfor.* **2006**, *24*, 380.
- ⁷ Filling, C.; Berndt, K. D.; Benach, J.; Knapp, S.; Prozorovskin, T.; Nordling, E.; Ladenstein, R.; Jörnvall, H.; Oppermann, U. *J. Biol. Chem.* **2002**, *277*, 25677.
- ⁸ Blankenfeldt, W.; Kerr, I. D.; Giraud, M. F.; McMiken, H. J.; Leonard, G.; Whitfield, C.; Messner, P.; Gtanager, M.; Naismith, J. H. *Structure* **2002**, *10*, 773.
- ⁹ Thoden.; J. B.; Wohlers.; T. M.; Fridovich-Keil J. L.; Holden.; H. M. *Biochemistry* **2000**, *39*, 5691.
- ¹⁰ Schliebena, N. H.; Niefinda, K.; Müllera, J.; Riebelb, B.; Hummelb, W.; Schomburg, D. *J Mol. Biol.* **2005**, *349*, 801.

-
- ¹¹ Niefind, K.; Muller, J.; Riebel, B.; Werner, H.; Schomburg, D. *J. Mol. Biol.* **2003**, *327*, 317.
- ¹² Benach, J.; Atrian, S.; Gonzales-Duarte, R.; Ladenstein, R. *J. Mol. Biol.* **1999**, *289*, 335.
- ¹³ Quallich, G. J.; Woodall T. R. *Tetrahedron Lett.* **1993**, *34*, 785.
- ¹⁴ Zhang, X.; Taketomi, T.; Yoshizumi, T.; Kumobayashi, H.; Akutagawa, S.; Mashima, K.; Takaya, H. *J. Am. Chem. Soc.* **1993**, *115*, 3318.
- ¹⁵ Liang, J.; Mundroff, E.; Voladri, R.; Jenne, S.; Gilson, L.; Conway, A.; Krebber, A.; Wong, J.; Huisman, G.; Truesdell, S.; Lalonde, J. *Org. Process Res. Dev.* **2010**, *14*, 188.
- ¹⁶ G7S/A94T/S96P/R108H/G117S/E145S/N157T/P194N/M206Q/I223V
- ¹⁷ <http://www.codexis.com/documents/Codexis%20Codex%20KRED%20Panel%20-%20Screening%20Protocol%20-%2020100728%20-%20UPDATEDv1.pdf>
- ¹⁸ Tantillo, D. J.; Chen J.; Houk, K. N. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743.
- ¹⁹ Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.
- ²⁰ Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294.
- ²¹ Takano, Y.; Houk, K. N. *J. Chem. Theory Comput* **2005**, *1*, 70.
- ²² Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- ²³ Wennmohs, F.; Staemmler, V.; Schindler, M. *J. Chem. Phys.* **2003**, *119*, 3208.
- ²⁴ Case, D. A.; Darden, T. A.; T. E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.;

Swails, J.; Goetz, A. W.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; *University of California*; *San Francisco*. **2012**.

²⁵ Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

²⁶ The lowest detected peak area is 10.0, so when the peak was reported to be 0, which would lead to an *er* of 0 or infinity, 0 was set to 9 because it is incrementally lower than the detection limit. The measured peak values range up to 475.2.

²⁷ Zhou, P.; Tian, F.; Lv, F.; Shang, Z. *Proteins* **2009**, *76*, 151.

²⁸ Kaur, D.; Aulakh.; D.; Sharma, R.; Singh, H. *J. Sulfur Chem.* **2013**, *34*, 512.

²⁹ Platts, J. A.; Howard, S. T.; Bracke, B. R. F. *J. Am. Chem. Soc.* **1996**, *118*, 2726.

³⁰ Wu, Y. D.; Houk K. N. *J. Am. Chem. Soc.* **1987**, *109*, 906.

³¹ For applications of the distortion/interaction model.; see Medina J. M.; Mackey J. L.; Garg N. K.; Houk K. N. *J. Am. Chem. Soc.* **2014**, *136*, 15798 and references therein.

³² For an example of how a mutation to proline increase the thermostability of an alcohol dehydrogenase see, Goihberg E.; Dym, O.; Tel-Or, S.; Levin, I.; Peretz, M.; Burstein, Y. *Proteins, Struct. Funct. Bioinf.* **2007**, *66*, 196.

³³ Pauling L. *Nature* **1948**, *161*, 707.

³⁴ Smith A. J.; Muller R.; Toscano, M. D.; Kast P.; Hellinga H.; Hilvert D.; Houk K. N. *J. Am. Chem. Soc.* **2008**, *130*, 15361.

³⁵ Jiménez-Osés, G.; Osuna, S.; Gao X.; Sawaya, M. R.; Gilson L.; Collier, S. J.; Huisman, G. W.; Yeates T. O.; Tang, Y.; Houk, K. N. *Nature Chem. Biol.* **2014**, *10*, 431.

³⁶ Savitzky A.; Golay M. J. E. *Analytical Chemistry* **1964**, *36*, 1627.

³⁷ Numerical Recipes 3rd Edition, The Art of Scientific Computing W.H. Press.; S.A. Teukolsky.; WT Vetterling.; BP Flannery Cambridge University Press ISBN-13, 9780521880688

³⁸ Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122*, 3746.

³⁹ The AMBER MD default value for the first solvation shell is 3.4 Å.

⁴⁰ Monera, O. D.; Sereda, T. J.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1995**, *1*, 319.

⁴¹ Experimental *ee* for Sph with oxacyclopentanone was not measured.; but these data were obtained for a related variant with 9 of the 10 mutations in Sph. G7S/A94T/S96P/R108H/G117S/E145S/N157T/M206Q/I223V.; lacking P194N). This mutant gave 12.1 % *ee* *S*-oxacyclopentanol.