EVOLVING ALKENE REDUCTASE ENANTIOCOMPLEMENTARITY THROUGH ITERATIVE SATURATION MUTAGENESIS

By

ADAM Z. WALTON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

© 2012 Adam Z. Walton

In memory of 1st Lieutenant Mark H. Dooley, 1978-2005.

ACKNOWLEDGMENTS

Foremost, I'd like to thank Dr. Jon Stewart for taking a chance on me, twice, and for his unrelenting support over many years. His advice and guidance have been essential in the completion this study. At the risk of failing to do so in the following text I must thank and give credit upfront to Dr. Bradford (Brad) Sullivan for his many contributions. Specifically, it was his guidance and expertise that allowed me to complete the much of the scale-up work described in Chapter 2. It was his idea to include methyl acrylate in our Baylis-Hillman adduct panel and he conducted a vast majority of substrate preparation for that work. Without Brad's diligence we may have never achieved a working mutagenesis protocol. While we did much of the work and learning together, more often than not he was the teacher and I the student. His contributions are manifest in the pBS# plasmid nomenclature referenced throughout this document. For all of these things and countless others I say cheers. Specific thanks is also due to Yuri Pompeu for solving the crystal structure of OYE 2.6 during the course of this study and numerous discussions and keen observations on structural aspects of the enzymes in this field. And finally, thanks to many unnamed Stewart group members over the past 10 years that have both influenced my education and provided precedence for and assistance with this research.

I need to thank past leadership of the Department of Chemistry and Life Science, United States Military Academy, Brigadier General David Allbee (ret.), Colonel Patricia Dooley (ret.), and Lieutenant Colonel Brad Dick (ret.) for their mentorship, guidance and faith in my abilities that gave me the opportunity to pursue this degree. Thanks is given to the current leadership of the Department, the United States Military Academy, and the U.S. Army Advanced Civilian Schooling program for financial support required to

pursue this study. I am additionally grateful to the U.S. Army for seemingly endless opportunities for growth and development over the past 23 years.

Finally, I would be remiss if I failed to acknowledge the contributions of my family. I owe thanks to my beautiful wife, Jennifer, and our four sons for their encouragement and cheerleading during this work and my career. Their unconditional support, as always, is the source of my motivation for everything that I do.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida In Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

EVOLVING ALKENE REDUCTASE ENANTIOCOMPLEMENTARITY THROUGH ITERATIVE SATURATION MUTAGENESIS

By

Adam Z. Walton

May 2012

Chair: Jon D. Stewart
Major: Chemistry

Old Yellow Enzyme (*Saccharomyces pastorianus*) was the first discovered flavo-protein and has served as a model for much of what is known about alkene reductase enzymes. These enzymes catalyze the net *trans*-addition of hydrogen to activated alkenes, namely, ketones, aldehydes, esters, and nitro alkenes. We introduce the use of two new compounds, 2-(hydroxymethyl) cyclohexenone and 2-(hydroxymethyl) cyclopentenone, to the list of known substrates. These compounds, along with hydroxymethyl methylacrylate, are reduced asymmetrically with a battery of Old Yellow Enzyme (OYE) homologs and OYE mutants with amino acid substitutions at the active site residue Trp116. The results yield a range of (*R*)- and (*S*)-reduction products. OYE 2.6 (*Pichia stipitis*) yields the best conversions and high stereoselectivity for the (*S*)-enantiomer of all three substrates. However, no variant gave a high yield and optical purity for the (*R*)-enantiomers. To obtain enantiocomplemantarity, we proposed using the strategy of Iterative Saturation Mutagenesis to evolve (*R*)-stereoselectivity for the three substrates of interest. In this approach, OYE 2.6 was subject to two rounds of saturation mutagenesis to achieve 89% enantiomeric excess of (*R*)-2-(hydroxymethyl)

cyclopentanone and 90% enantiomeric excess of (*R*)-3-hydroxy-2-methylpropanoate (Roche ester).

Also presented in this work, is a novel methodology for the generation of saturated mutant libraries. This approach includes assessment criteria for a pooled mutant plasmid library from sequencing chromatographic data. The result was an elimination of the requirement to sequence a large portion of a mutant library to assess overall library quality, dramatically streamlining the directed evolution cycle.

The applicability of alkene reductases was also demonstrated by two novel preparative-scale enzymatic approaches to useful synthetic intermediates. The synthetic menthol intermediate (*R*)-citronellal was prepared with a volumetric productivity of 2.8 g/L/hr and an enantiomeric excess of 98%. The (*S*)-enantiomer of the chiral synthetic intermediate 2-methylpentanol was prepared at a volumetric productivity of 0.7 g/L/hr and an enantiomeric excess of 99%. The combination of this scale-up methodology along with the power of directed evolution vastly enhance the utility of alkene reductase enzymes for the production of fine chemicals and other synthetic intermediates.

CHAPTER 1
OLD YELLOW ENZYME, THE ORIGINAL ALKENE REDUCTASE

Identification and Characterization

Old Yellow Enzyme was the first discovered flavo-protein and has served as a model for much of what is known about this class of enzymes.[1] In recent years the catalytic activity of this enzyme and an entire subset of related homologs have appeared in the literature as valuable catalysts with industrial applications.[2] Our interest in this enzyme is to probe the protein sequence space in order to further our understanding of these enzymes. Hopefully, this will lead to the development of biocatalysts that retain the unique activity of the enzyme while expanding its substrate scope and stereoselectivity; therefore, in theory, expanding its industrial applicability as a "green" catalyst.[3]

Old Yellow Enzyme was first isolated from the brewers' bottom yeast *Saccharomyces carlsbergensis* (so named after the Danish brewery, Carlsberg, where it was first isolated and later reclassified as *Saccharomyces pastorianus*) by Warburg and Christian in 1932.[4] Their original aim was to elucidate the nature of biological oxidations in yeast. In the process of their work they isolated a yellow protein, "gelbe ferment", that was present in appreciable quantities in yeast lysate. When investigating another enzyme that would later be identified as glucose-6-phosphate dehydrogenase, the yellow enzyme, in the presence of a thermostable extract ($NADP^+$/NADPH) and oxygen, was able to facilitate the oxidation of glucose-6-phosphate thus completing the respiratory chain (Figure 1-1). The later (1938) discovery of a new yellow enzyme homolog in yeast by Haas, led to the designation of the first reported yellow enzyme as "Old", hence the name commonly referred to today. Later, in 1956, Theorell and

Akeson reported a rigorous purification technique leading to a much more homogenous enzyme further enabling detailed study of this enzyme.[5] As a result, their work confirmed the enzyme to be a homodimer with a 1:1 ratio of FMN to protein.

In 1969, expanding on the work of Theorell, Matthews and Massey were able to improve on the purification of the Old Yellow Enzyme and set about its detailed characterization.[6] The enzyme was found to catalyze the oxidation of NADPH to $NADP^+$ where the two-electron transfer in the form of a hydride is transferred to the oxidized FMN cofactor. This now reduced flavin could then be reoxidized by molecular oxygen; however, the physiological substrate, or even putative substrates, at this point was not known. They reported that the enzyme displayed an absorption spectrum that had a maximum around 440nm that shifted to 600nm upon the addition of many phenolic compounds. This resulted in the visual color change from yellow to green and was attributed to the formation of a charge transfer complex.[7] These phenol complexes were further characterized by Raman spectroscopy and found to involve π interactions between the phenol(ate) ring and the isoalloxazine moiety of the bound flavin.[8]

Massey exploited phenol binding to develop an affinity purification technique using a phenol-functionalized resin for the sole chromatographic step.[9] This further allowed a simpler and more effective isolation of the enzyme from its native host in greater yield. In 1991 the gene for Old Yellow Enzyme (now termed OYE 1) was cloned and overexpressed in *E. coli.*[10] Comparison of the overexpressed protein with non-recombinant enzyme preparations led to the discovery that several OYE isozymes were present in brewers' yeast.[11] The cloned gene also led to a homogenous source suitable for unambiguous characterization and structural determination.

Substrate Specificity

To date, the physiological electron acceptor(s) for Old Yellow Enzyme remain unknown. The generally accepted substrate for reducing the bound flavin mononucleotide cofactor is NADPH. It has been widely noted that molecular oxygen is capable of reoxidizing the FMN; however, the rate at which it does so is thought to be physiologically insignificant.[6] Some of the first substrates proposed for OYE1 were simple quinones such as those depicted in Figure 1-2a.[11] Additionally, cyclohex-2-ene-1-one has been proposed as a simple model substrate for the reoxidization of FMN with significant activity in nearly all reported OYE homologs. This has led to the widespread investigation of α,β-unsaturated carbonyl compounds in the characterization of OYE.

Ketones and Aldehydes

One of the first and most extensive panels of substrates tested for OYE 1 was described by Vaz *et al*. (Figure 1-2b).[12] It was observed that unsaturated aldehydes and ketones, including cyclic ketones, were among the most active substrates for OYE 1. Later, while probing the kinetics of the oxidative reaction, Kohli and Massey identified several additional cinnemaldehyde derivatives that displayed modest activity in the enzyme (Figure 1-2c).[13] The enzyme was found to have little to no activity for sterols. Notably the enzyme also showed no activity for vinyl acids, esters, amides or nitriles in the study. Among the observations in this work it was observed that increasing the size of alkyl substituents at the β-position dramatically reduced the rate of hydride transfer, presumably due to steric hindrance within the active site. No characterization of the enantioselectivity of these reductions was done in this work. These workers also reported a novel dismutation reaction that occurs in the absence of the hydride donor NADPH. In this reaction OYE first oxidizes a molecule of the cyclohexenone substrate,

yielding reduced FMN and a cyclohexadienone. The $FMNH_2$ then transfers the electrons to a second molecule of the substrate producing a saturated cyclohexanone. The initial cyclohexadienone rapidly tautomerizes to the corresponding phenol, which tightly binds to the active site as previously described.

The first systematic exploration of OYE 1 enantioselectivity was conducted in our group and reported in 2006.[14] Based on results from 2- and 3-substituted cyclohexenones (Figure 1-2d), it was observed that smaller alkyl substituents at either position yielded greater conversions. This observation is consistent with those made by Vaz *et al*. Additionally, all compounds tested yielded the same enantiomer, indicating a single binding mode for the homologous substrates. This single binding mode, described hereafter as "normal" (Figure 1-3), is one where the hydride adds to the β-carbon and a proton adds to the α-carbon, from below and above the plane of the page, respectively, as drawn in Figure 1-2d.

In 2007, Muller *et al*. introduced several new compounds into the growing list of OYE substrates (Figure 1-2d).[15] These included 2-methyl pentenal and both isomers of citral (neral and geranial). The enzyme was found to have a preference for the *E*-isomer of citral (geranial) and yielded primarily the (*R*) enantiomer indicating binding analogous to that observed by Swiderska and Stewart. The *Z*-isomer, however, gave much poorer conversion and required a "flipped" binding orientation to yield predominantly the same (*R*)-enantiomer. Interestingly, the observed stereochemistry of the reduced 2-methyl pentanal indicates the same "flipped" binding mode to yield the (*S*)-enantiomer with a significantly smaller compound. In the same study the significantly larger α-methyl cinnamaldehyde conforms to the "normal" binding mode to

yield the (*R*)-enantiomer in excess. In a separate work Muller *et al*. further demonstrated the ability of OYE to also reduce an activated alkyne (Figure 1-2d).[16] In this case 4-phenyl-3-butyne-2-one was reduced to (*E*)-4-phenyl-3-butene-2-one which is also a substrate previously identified by Vaz *et al*.

In this research group, both (*R*)- and (*S*)-perillaldehyde[17] (Figure 1-2d) and (*R*)- and (*S*)-carvone[18] (Figure 1-2e) were added to the battery of substrates for OYE 1. While these compounds displayed varied binding modes when reduced with OYE homologs and mutants, the wild type enzyme displayed consistent binding and high diastereomeric excess with these compounds that did not vary the configuration of the new substrate stereo center.

In 2008, while comparing various OYE homologs, Hall *et al*. reported several new compounds for OYE1[19] (Figure 1-2e). Methyl substituted cyclopentenones, at both the 2- and 3-positions were found to be reasonable substrates for OYE. The 2-substituted compound yielded a racemic product mixture while the 3-substituted version gave predominantly (*S*)-enantiomer product, consistent with the “normal” binding mode. Additionally, 2-methyl maleimides were evaluated and found to give (*R*)-enantiomer products, also consistent with the “normal” binding mode. This binding mode is also supported by the observation that increasing the size of the *N*-substituted alkyl group had little effect on the stereoselectivity of the reaction.

Following this work, Stueckler *et al*. demonstrated an efficient reduction of additional α-methyl cinnamaldehyde derivatives (Figure 1-2f) in the preparation of fragrance compounds Lilial™ (*t*-butyl) and Helional™ (methylenedioxy) [20]. They observed high enantiomeric excess in the presence of 20% TBME co-solvent. The

predominant (*S*)-enantiomer indicates a “flipped” binding mode presumably due to the bulky nature of the *trans*-phenyl and substituted phenyl rings.

. From the same group Winkler *et al*. explored the suitability of 2-substituted alkoxy compounds (Figure 1-2f) as novel OYE substrates.[21] While overall conversion was generally poor, the 2-methoxycyclohexenone substrate was the only compound that gave the (*R*)-enantiomer corresponding to the “normal” binding mode. All other compounds indicate the alternate “flipped” binding where stereochemistry could be identified.

Esters

Hall *et al*., using dimethyl esters of citraconic, itaconic, and mesaconic acids (Figure 1-2e), presented the first report of esters as substrates for OYE 1.[19] The “normal” binding mode was observed with good stereoselectivity, except for the *E*-isomer that potentially adopts a “flipped” binding mode to avoid steric clash at the β-position. The same group also described the use of methyl 2-hydroxymethylacrylate and it’s allyl and benzyl ether derivatives (Figure 1-2e) as OYE 1 substrates to produce the synthetically useful compound 3-hydroxy-2-methylpropanoate.[22] In all cases, the observed binding mode is “flipped”, yielding the (*R*)-enantiomer presumably once again due to steric bulkiness. It appears as though considerable effort was wasted by systematically increasing the size of the ether functional group when available models were sufficiently useful to predict this outcome.

A separate effort revealed the applicability of α-halo methylcinnamate derivatives (Figure 1-2f) for use with OYE 1.[23] These compounds yield the (*S*)-enantiomer, consistent with a “flipped” binding mode that is consistent with the cinnamaldehyde derivatives discussed above.

In pursuit of synthetically useful 2,4-dimethylbutyrolactone, Korpak and Pietruszk used OYE 1 to reduce 2-methyl-4-oxopent-2-enoate (Figure 1-2f) as the first step in its preparation.[24] Assuming that the keto carbonyl is the prefered electron withdrawing group, the *E*-isomer binds in a "flipped" orientation to yield >99% (*R*)-enantiomer.

Most recently, Stueckler *et al*. introduced α,β-dehydroamino acid derivatives (Figure1-2g) as suitable substrates for OYE1.[25] In all cases, the substrates theoretically bound as depicted in Figure1-2g for OYE 1 yielding the (*S*)-enantiomers. The diester *E*-isomers each have two possible binding modes that will yield each enantiomer. Using the homolog OYE 3 (from *Saccharomyces cerevisiae*), they were able to determine that inverted binding of the di-ester compounds could be influenced by the size of the *N*-acyl substitution on these compounds.

Nitro Alkenes

In 2000, Meah and Massey first reported the reduction of α,β-unsaturated nitro compounds (Figure 1-2c) by OYE 1.[26] As would be expected, the mechanisms of these reductions were analogous to those previously described for α,β-unsaturated carbonyl compounds. The major difference was that the nitronate ion formed after hydride transfer to the β-carbon bound to the enzyme weakly and therefore could dissociate prior to enzyme-mediated protonation. This presents potential problems in directing the stereoselectivity of the enzyme at the nitro-bearing carbon.

A year later, Meah *et al*. reported that OYE 1 reduced the nitrate esters nitroglycerine and propylene dinitrate (Figure 1-2c).[27] The mechanism involves a net two-electron transfer to a nitrate ester nitrogen followed by subsequent elimination of nitrite. The enzyme preferred reaction with a terminal nitrate over secondary nitrate moieties. No further reduction was observed with the propylene mono nitrate product of

these reactions. Williams *et al*. further demonstrated the ability of OYE 1 to reduce both trinitrotoluene (TNT) and nitrobenzaldehyde (Figure 1-2c) while investigating the ability of OYE homologs to degrade these and other compounds.[28] This work further implied that hydride addition was directly to the nitro group and not to the aromatic ring.

Our research group has previously investigated the asymmetric reduction of a series of nitro acrylates compounds (Figure 1-2d) by OYE 1.[29] In all cases, the enzyme yielded a high excess of the (*R*)-enantiomer. Deuterium labeling experiments also confirmed that hydride was added to the β-carbon relative to the nitro group.

Enzyme Structure

In 1994 Fox and Karplus reported the 2.0 Å resolution crystal structure of OYE1.[30] The protein contained a single domain consisting of a parallel, eight-stranded α/β-barrel similar to that of triosephosphate isomerase (TIM). The largely hydrophobic active site is accessible to the solvent via a deep cleft. This cleft is, in part, obscured by a flexible loop that has implications in the dynamic binding of the NADPH substrate. The isoalloxazine ring of the FMN cofactor is postioned across the top of the barrel perpendicular to its axis. One crystal form contained the bound inhibitor *p*-hydroxybenzaldehyde (PHB) bound directly above the FMN prosthetic group. The aromatic rings of both ligands lay parallel to each other in close contact. The C2 of PHB lies directly above and 3.4 Å away from N5 of the flavin, which is consistent with its mimicking of C_β of a 2-cyclohexenone substrate. The π-π interactions are responsible for formation of the charge transfer complex.[8] Since the publication of this structure, several efforts have uncovered the roles of key residues within the OYE 1 active site (Figure 1-3 and 1-4).

Histidine 191 and Asparagine 194

The phenolate oxygen of PHB in the Fox and Karplus structure has the appropriate distance and orientation to be hydrogen bonded with the side chains of His191 and Asn194. Brown *et al*. constructed H191N and H191N/N194H mutants of OYE 1 and characterized their kinetic properties.[31] The enzyme maintained activity with respect to the oxidation of NADPH in both mutants, presumably because of the additional contacts the cofactor makes with the enzyme upon binding. By contrast, the mutants showed dramatically reduced activity towards the reduction of cyclohexenone and they also displayed reduced binding affinity for phenol inhibitors. These effects are likely due to altered alignment of these compounds within the active site. In order for catalysis to occur, the β-carbon of the substrate must be positioned precisely with respect to N5 of the flavin ring to facilitate efficient hydride transfer. Hydrogen bonding to His191 and Asn194 stabilizes the enolate that is formed by the addition of the hydride. As the orientation of the hydrogen bond donors are altered, the substrate is no longer optimally stabilized upon receipt of the hydride thus reducing catalytic activity.

Tyrosine 196

The same authors also probed the role of Tyr196 in the catalytic mechanism of OYE 1.[13] Tyrosine 196 is positioned 3.5 Å above C3 of the PHB ligand, opposite the flavin cofactor. This position is analogous to the α-carbon of 2-cyclohexenone. The acidity of the tyrosyl proton is potentially reduced from its normal pK_a of 10.1 by the close proximity of Asn 251 (2.7 Å). Thus Tyr196 is positioned to protonate and thereby yield a net *trans*-addition of hydrogen to the double bond. When Tyr196 was mutated to a phenylalanine, the rate of 2-cyclohexenone reduction decreased by six orders of magnitude[13] implicating its role in the reduction of α,β-unsaturated carbonyl compounds.

Threonine 37

Threonine 37 is positioned in the OYE 1 active site near the bound FMN cofactor. The side chain hydroxyl group is within hydrogen bonding distance (2.8 Å) of the C4 carbonyl oxygen of the isoalloxazine ring. By mutating the residue to an alanine, Xu *et al*. deduced that Thr 37 helped to stabilize negative charge on the flavin ring system as a result of reduction by NADPH.[32] The T37A mutant showed an order of magnitude decrease in the rate of NADPH oxidation, thus signifying a higher reduction potential for the FMN. Conversely, the mutant displayed a higher turnover (2-9 fold) for reducing all substrates investigated. This increase in rate was facilitated by the decreased stability of the reduced mutant enzyme. The overall turnover of the ping-pong mechanism was still lower with the mutant enzyme given the more pronounced effect on the NADPH oxidation.

Tryptophan 116

Given the observation that 2-cyclohexenones with large substituents at the 2- and 3-positions were reduced slowly (but with high stereoselectivity),[14] the Stewart group attempted to uncover the origins of these effects.[18] From the crystal structure of OYE 1, the side chain of Trp116 appeared to sterically crowd the active site, and computer modeling suggested that this hindrance would increase with increasing substituent size. Site saturation mutagenesis at position 116 followed by screening for increased activity with 3-methyl cyclohexenone uncovered several variants of interest. The W116I and W116F mutants were purified and screened against a battery of potential substrates. This investigation revealed an interesting result with respect to the reduction of (*R*)- and (*S*)-carvones (Figure 1-2e). With (*R*)-carvone, both W116I and W116F as well as wild type OYE1 provided (*R*)-selectivity for the newly-created sp^3 center at C_α, consistent

with previous observations and a “normal” substrate binding mode. (*S*)-carvone gave the same result for W116F and the wild type enzyme. On the other hand, the W116I mutant gave (*S*)-selective reduction at the α-carbon indicating a “flipped” binding mode. It was postulated that this alternate substrate binding orientation is possible only when the bulky tryptophan residue is replaced with an amino acid whose side chain is small enough to accommodate the isopropenyl group on C5 of the carvone substrate.

A tryptophan is highly conserved among OYE homologs at position 116; however, a BLAST search revealed a close relative of OYE 1 with an isoleucine at the analogous position (OYE 2.6 from *Pichia stipitis*[33]). We cloned and expressed this protein to provide an additional alkene reductase for our collection of biocatalysts. While its properties were similar to those of OYE 1 in some respects, it also had highly useful properties that were revealed in the course of the studies described here.

Experimental Strategy

The previous sections have described what is known about the structure, mechanism and substrate specificities of OYE 1. We had also collected a number of other OYE 1 homologs in our laboratory[18,34] as well as a complete set of amino acid replacements at position 116 in OYE 1.[35] Because most of the work to date had focused on simple model substrates, we wanted to increase the utility of our enzyme library by applying it to more complex, and therefore more useful alkenes. We set three additional criteria in our search for new targets. First, they should be well behaved with respect to enantioselectivity and lack side reactions (i.e. racemization and ketone reduction). Second, we wanted the reduction products to be useful intermediates for existing synthetic routes. Finally, the substrates needed to be commercially available or

relatively simple to prepare in-house. This search led us to Baylis-Hillman adducts as alkene reductase substrates. We chose three representative examples (Figure 1-5).

(*S*)-2-(hydroxymethyl) cyclohexanone [**4**] is a very useful chiral building block.[36,37] It has been prepared by using L-amino acids to catalyze the asymmetric aldol condensation of formaldehyde with cyclohexanone. The compound has been used to prepare the antispasmodic drug rociverine[38], the cancer therapeutic pyranicin[39], the diabetes treatment penarolide sulfate A_1[40], the plant pheromones cladospolide A-C[41], the *C. elegans* pheromone daumone[42], and the anti-viral glycolipid macroviracin D[43]. While the reported aldol route to (*S*)-**4** was successful, we reasoned that this process could be improved upon with respect to stereoselectivity and reaction time by using an appropriate enzymatic asymmetric alkene reduction. The required precursor alkene, 2-(hydroxymethyl)-cyclohex-2-enone, had been prepared previously[44] by using a base catalyzed Baylis-Hillman reaction to add an α-hydroxymethyl group to cyclohexenone in good yield (Figure 1-6).

We also chose a biocatalytic route to (*S*)-2-(hydroxymethyl) cyclopentanone[45] [**5**] since this material has been used to prepare the fragrance compound jasmine lactone[45], insect pheromones[46,47], and the signaling compound leukotriene B_5[48]. Like the cyclohexene derivative, 2-(hydroxymethyl)-cyclopent-2-enone could be prepared in a similar manner using Baylis-Hillman chemistry.[49]

Continuing our search we then focused on the Baylis-Hillman theme. (*R*)-3-hydroxy-2-methylpropanoate [(*R*)-**6**] (also known as Roche ester), is a common chiral precursor in the synthesis of vitamins, fragrance compounds, antibiotics, and natural products.[22] The Baylis-Hillman adduct precursor [**3**] to Roche ester could also be

prepared by a similar approach.[50] This compound and its derivatives had been previously tested as substrates for OYE 1 by Stueckler *et al*.[22] (Figure 1-2e). Our goal was to use enzyme – rather than substrate engineering – to identify an alkene reductase route to both enantiomers of the target compound [**6**].

The proposed enzymatic reductions of the Baylis-Hillman compounds is outlined in Figure 1-5. All three could be prepared by simple reactions from inexpensive, commercially available compounds. With the exception of [**3**], hydroxymethyl functionalized substrates represent a new class of enones not yet explored with OYE or its homologs. Additionally, all three are useful intermediates in the synthesis of a diverse range of attractive synthetic targets.

Experimental Procedures

Preparation of Baylis-Hillman Adducts

Dr. Bradford Sullivan conducted a large portion of the synthesis of these compounds in our lab. This work is summarized as follows.

2-(Hydroxymethyl)-cyclohex-2-enone

A solution of aqueous 37% formaldehyde (12.5 mL, 154 mmol), 2-cyclohexen-1-one (5.00 g, 52.0 mmol), and 4-dimethylaminopyridine (6.35 g, 52.0 mmol) in THF (15 mL) was stirred at room temperature for 12 h. The reaction mixture was acidified to *ca*. pH 5 with 1 M HCl and the aqueous layer was extracted with CH_2Cl_2 (5 × 15 mL). The combined organic layers were washed with H_2O (25 mL), brine (25 mL), and dried over Na_2SO_4. The crude material was subjected to flash column chromatography with a solvent gradient of hexanes and ethyl acetate (3 : 1, 1 : 1, then 1 : 3) to yield the title compound as a pale yellow oil (5.12 g, 78% yield). R_f 0.40 (1 : 1 hexanes-ethyl acetate); ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.93 (t, J = 4.1 Hz, 1H), 4.20-4.25 (m, 2H), 2.77

(s, br OH), 2.35-2.46 (m, 4H), 1.95-2.05 (m, 2H) ppm; ^{13}C NMR (75 MHz, $CDCl_3$) δ 200.8, 147.2, 138.4, 62.0, 38.4, 25.8, 22.9 ppm. Spectral data were consistent with those reported previously.[44]

2-(Hydroxymethyl)-cyclopent-2-enone

To a stirred solution of aqueous 37% formaldehyde (1.0 mL, 12 mmol) and cyclopent-2-enone (700 mg, 8.52 mmol) in THF (1 mL) was added imidazole (29 mg, 0.42 mmol). The resulting mixture was stirred at room temperature for 17 days. The reaction mixture was acidified to *ca*. pH 5 with 1 M HCl and the aqueous layer was extracted with CH_2Cl_2 (5 × 5 mL). The combined organic layers were washed with H_2O (10 mL) and brine (10 mL), then dried over Na_2SO_4. The crude material was subjected to flash column chromatography with a solvent gradient of hexanes and ethyl acetate (1 : 1, 1 : 2, then 1 : 5) to yield the title compound as a white solid (219 mg, 22% yield). R_f 0.35 (1 : 3 hexanes- ethyl acetate); m.p. 68-70°C; ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.51-7.57 (m, 1H), 4.33-4.41 (m, 2H), 2.61-2.69 (m, 2H), 2.53 (s, br OH), 2.43-2.49 (m, 2H) ppm; ^{13}C NMR (75 MHz, $CDCl_3$) δ 210.1, 159.4, 145.2, 57.5, 35.2, 27.1 ppm. Spectral data were consistent with those reported by Kar and Argade[49].

Methyl 2-(hydroxymethyl)acrylate

To a stirred solution of 37% aqueous formaldehyde (1.45 mL, 17.9 mmol) and freshly-distilled methyl acrylate (2.10 mL, 23.2 mmol) in a mixture of H_2O and 1,4-dioxane (1:1, 50 mL) was added 1,4-diazabicyclo[2.2.2]octane (1.70 g, 15.1 mmol) at room temperature. The resulting mixture was stirred for 72 h, then poured into saturated aqueous NaCl (50 mL). This was extracted with Et_2O (5 × 10 mL) and the combined organic layers were washed with water (20 mL), brine (20 mL), and dried over Na_2SO_4. The crude material was subjected to flash column chromatography with a

solvent gradient of petroleum ether and ethyl acetate (7 : 1, 4 : 1, then 1 : 1) to yield the title compound as a clear and colorless oil (1.54 g, 57% yield). R_f 0.34 (1 : 1 petroleum ether-ethyl acetate); ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.18-6.20 (m, 1H), 5.78-5.80 (m, 1H), 4.24-4.27 (m, 2H), 3.72 (s, 3H), 3.12 (s, br OH) ppm; ^{13}C NMR (75 MHz, $CDCl_3$) δ 166.9, 139.5, 125.9, 62.4, 52.1 ppm. Spectral data were consistent with those reported by Drewes et al.[50]

Screening Reactions

Screening assays were carried out in 0.3 mL volumes of 50 mM KP_i, pH 7.5, supplemented with 200 mM glucose, 25 U/mL GDH102 (Biocatalytics), 0.3 mM $NADP^+$, 10 mM substrate, and ~100 μg of GST-affinity purified alkene reductase enzyme. Reactions were allowed to proceed for 24 hours while shaken at room temperature followed by extraction with 0.5 mL of ethyl acetate and subsequent analysis by GC. Chiral-phase GC analyses were carried out with a 30 m × 0.25 mm β-Dex 225 column (Supelco) with He as the carrier gas and FID. For analysis of **1**, **2**, **4** and **5** the temperature program involved 140°C (10 min) followed by a 20°C / min increase to 200°C (3 min). Under these conditions, peaks eluted at 10.2 min ((*S*)-**4**), 10.7 min ((*R*)-**4**), and 13.1 min (**1**) or 10.3 min ((*R*)-**5**), 11.4 min ((*S*)-**5**), and 13.2 min (**2**). For analysis of **3** and **6** the temperature program involved 100°C (12 min) followed by a 20°C / min increase to 180°C (5 min). Under these conditions, peaks eluted at 10.7 min ((*S*)-**6**), 11.3 min ((*R*)-**6**), and 11.8 min (**3**).

Larger-Scale Preparation of (*S*)-4

E. coli BL21(DE3) cells overexpressing OYE 2.6 (*Pichia stipitis*) were grown on LB agar plates supplemented with 0.2 mg/mL ampicillin. A single colony was then used to prepare a 40 mL starter culture in LB media with ampicillin that was shaken overnight at

37° C. This pre-culture was used to inoculate 4 L of LB media supplemented with 80 mL of 20% glucose, 2 g/L ampicillin, and 0.5 mL AF 204 (Sigma). The cells were grown in a New Brunswick Scientific M19 fermenter at 37°C, 700 rpm, and 4 L/min airflow for 2.5 hours until the OD_{600} = 0.8. At this time the temperature was decreased to 30°C and protein expression was induced with the addition of 0.48 mL of 840 mM iPTG. Growth was continued for an additional 4 hours until the cells reached an OD_{600} = 4.98. Cells were then harvested by centrifugation at 5000 rpm to yield 28.15 g (wet cell weight) and were stored at -20°C overnight. After thawing in 30 mL of 0.1 M phosphate buffer (pH 7.5), the slurry was passed through a French Pressure cell twice at 15,000 psi in the presence of 10 µM PMSF. Insoluble material was then removed by centrifugation at 15,000 rpm for 45 min to yield 45 mL of crude cell lysate. When assayed against **1** under the conditions described for screening reactions this crude contained 75 U/mL of activity and was used as is. The enzymatic reduction of **1** was conducted in a 500 mL three-neck round bottom flask stirred continuously at 300 rpm with a magnetic bar. The reaction media consisting of 90 mL of 0.1 M phosphate buffer (pH 7.5) containing 8.0 g of solid dextrose monohydrate was degassed under reduced pressure for 1 hour prior to use. It was then transferred to the reaction vessel and kept under argon. To this 1 mg (100 U) of GDH 102 (Biocatalytics), 10 mg NADPH (12 µmol), and 450 U of freshly assayed crude OYE 2.6 from the previous step was added. This mixture was allowed to equilibrate for 15 minutes at room temperature prior to the addition of substrate. 0.63 g (5.0 mmol) of neat **1** was then added to the reaction vessel in a single bolus. Reaction pH was maintained at 7.5 by adding 1 M KOH via a pH stat. Base demand was monitored until it dropped noticeably after 1.5 hours. GC-MS

confirmed that at 1.75 hours that the reaction was approximately 100% complete. The mixture was then extracted with 100 mL of CH_2Cl_2. The aqueous layer was extracted twice more with 50 mL of CH_2Cl_2 and the organic layers were combined and filtered over Celite to remove insoluble material. The filtered organic solution was then washed with brine and dried over Na_2SO_4. The solvent was then removed under vacuum to yield 0.61 g (4.8 mmol) of (*S*)-**4**. The observed e.e. was 91% and $[\alpha]_D^{25} = +10.42$ (c = 4.03 in $CHCl_3$). Lit. $[\alpha]_D^{22}$ +11.4 (*c* 1.0, $CHCl_3$).[37]

Larger-Scale Preparation of (*S*)-5 and Assignment of Absolute Configuration

E. coli BL21(DE3) cells overexpressing OYE2.6 (*Pichia stipitis*) were grown on LB agar plates supplemented with 0.2 mg/mL ampicillin. A single colony was then used to prepare a 15 mL culture in LB media with ampicillin that was shaken overnight at 37°C. This growth was used to inoculate 500mL of LB media supplemented with 5 mL of 40% glucose, 2 g/L ampicillin. The cells were grown at 37° C while shaking at 200 rpm in a baffled flask for 2 hours until the $OD_{600} = 1.0$. The temperature was decreased to 30°C and protein expression was induced with the addition of 60 μL of 840 mM iPTG. Growth was continued for an additional 4 hours until the cells reached an $OD_{600} = 3.2$. Cells were then harvested by centrifugation at 5000 rpm to yield 3.48 g (wet cell weight). They were then resuspended in 300 mL of 50 mM phosphate buffer (pH 7.5) containing 0.1 M glucose. To this slurry 300 mg of solid **2** was then added. GC-MS confirmed that at 1.5 hours that essentially all the starting material had been consumed. The mixture was then stirred with 500 mL CH_2Cl_2 at room temperature overnight. The heavily emulsified organic layer was filtered over Celite to remove insoluble material. The filtered organic solution was then washed with brine and dried over Na_2SO_4. The

solvent was then removed under vacuum to yield 129 mg of **5** with an observed e.e. of 75%.

The absolute configuration of the OYE 2.6 reduction product was assigned after Baeyer-Villiger oxidation to the corresponding lactone. To a stirred suspension of the reduction product from **5** (65 mg, 0.57 mmol) and $NaHCO_3$ (72 mg, 0.85 mmol) in CH_2CH_2 (1.0 mL) was added 3-chloroperoxybenzoic acid (77% purity, 153 mg, 0.683 mmol) at 0°C. The reaction mixture was slowly warmed to room temperature over 3 h and then quenched by adding solid $Na_2S_2O_3$ (50 mg) and stirring for an additional 15 min. The reaction mixture was dried with Na_2SO_4 and the solids were removed via filtration through a small plug of Celite with CH_2Cl_2. The crude material was subjected to flash column chromatography with a solvent gradient of pentane and ethyl acetate (2:1, 1:1, then 1:5) to yield 59 mg (80% yield) of (*S*)-6-(hydroxymethyl)tetrahydro-2*H*-pyran-2-one as a clear and colorless oil. R_f 0.20 (1:3 hexanes-ethyl acetate); ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.41 (tdd, *J* = 11, 5.5, 3.3 Hz, 1H), 3.78 (dd, *J* = 12.3, 3.1 Hz, 1H), 3.66 (dd, *J* = 12.3, 5.4 Hz, 1H), 2.55-2.66 (m, 1H), 2.38-2.52 (m, 1H), 1.67-2.03 (m, 4H) ppm; $[\alpha]_D^{23}$ +26.5 (*c* 2.50, $CHCl_3$); Lit. $[\alpha]_D^{25}$ +33 (*c* 1.3, $CHCl_3$).[46] Spectral data were consistent with those reported by Lees and Whitesides.[51]

Results and Discussion

Alkene Reductase Homologs

Our Baylis-Hillman substrates were first screened against a collection of alkene reductase homologs previously prepared in our lab.[35] See Figure 1-7 for complete results. Remarkably, *P. stipitis* OYE 2.6 was the only clone to display both high conversion and strong enantioselectivity for the (*S*)-enantiomer of products **4**, **5** and **6** at 95%, 76% and 95% respectively. It should also be noted that **5** appears to slowly

epimerize under these reaction conditions. Shortening this product's exposure to aqueous reaction conditions can yield an enantiomeric excess of 95% (Chapter 5). Selecting the best enzyme to deliver the (*S*)-enantiomer was quite simple; however, a good solution to obtaining the (*R*)-enantiomers was not quite as obvious.

In the case of the (*R*)-product, the OYE homolog collection gave no result for **1**. For substrate **2**, similar enantiomeric excess was observed in several clones; however, conversion for these reactions was limited to 50%. Finally, for **3**, OYE 1 and OYE 2 displayed limited conversion, but also showed a strong preference for the (*R*)-product. These results are consistent with those reported by Stueckler *et al*. except in the case of OYE3, which they reported as their best enzyme for the reduction of the methyl ester (37% conversion, >99% (*R*)).[22] In summary, OYE 1 was a good enantiocomplementary catalyst for OYE 2.6 except for its poor conversion. We therefore turned to our Trp116 replacement library of OYE 1 variants to identify a solution to this problem.

Tryptophan 116 Mutations

Given its success in previous studies[18] the Baylis-Hillman adducts were screened against the set of mutants at position 116.[35] See Figure 1-8 for a summary of results.

Surprisingly, all modifications with the exception of Arg, Cys, Thr, and Pro afforded both good conversion and a strong preference for the (*S*)-enantiomer of **1**, while none of the mutants indicated any preference for the opposite enantiomer. Similarly, the same mutants yielded a dramatic improvement in conversion for substrate **2** with the same enantiopreference. Of note, only wild type OYE 1 retained any preference for the (*R*)-product.

Even more interesting were the results for the reduction of **3**. It became apparent that modifications at position 116 could yield the full spectrum of conversion and

stereoselectivity. These are graphically depicted in Figure 1-9. In pursuit of the (*R*)-product, the most suitable candidates are Val, Tyr, and to a lesser extent Phe. These mutants all give greater than 70% e.e., which is significantly less than the 95% observed with the wild type enzyme. The observed loss in e.e. for these specific mutants is offset to a certain extent by an increase from wild type conversion (19%) of 2-3 times (37-68%).

Conclusions and Future Work

At this point in our investigation, *P. stipitis* OYE 2.6 remains the clear choice to obtaining the (*S*)-products for the reductions of all three Baylis-Hillman adducts. In the case of the (*R*)-products, we found no good solution leading to **4**, a partial solution to achieve compound **5**, and multiple partial solutions to **6**. Future efforts described in this work will advance on the goal of obtaining the (*R*)-products for the reduction of compounds **1**-**3**. Logical routes to this goal are 1) to continue searching for homologs that will yield the desired result, or 2) pursue an approach using enzyme engineering via an appropriate mutagenesis strategy.

Considering the results depicted in Figures 1-7 and 1-8, the most attractive route is mutagenesis, simply due to its ability to deliver a wide variety results. However, the generation of rationally designed mutants one at a time requires a significant amount of human effort and may be no easier than cloning homologs in terms of that effort. Before embarking on our protein engineering studies, we had to determine the best starting point for mutagenesis. That is to say, is it better to attempt to evolve enantiocompementarity using a single template like OYE 2.6, where a solution to (*S*)-enantioselectivity already exists, or is it better to start with one or more of the “partial” solutions manifest in our OYE 1 mutants? Complicating this is our desire to evolve a

solution for all three Baylis-Hillman adducts in the same effort rather than conducting three separate and distinct directed evolution projects. A more streamlined approach toward enzyme engineering is certainly in order and will be discussed and exploited in future chapters.

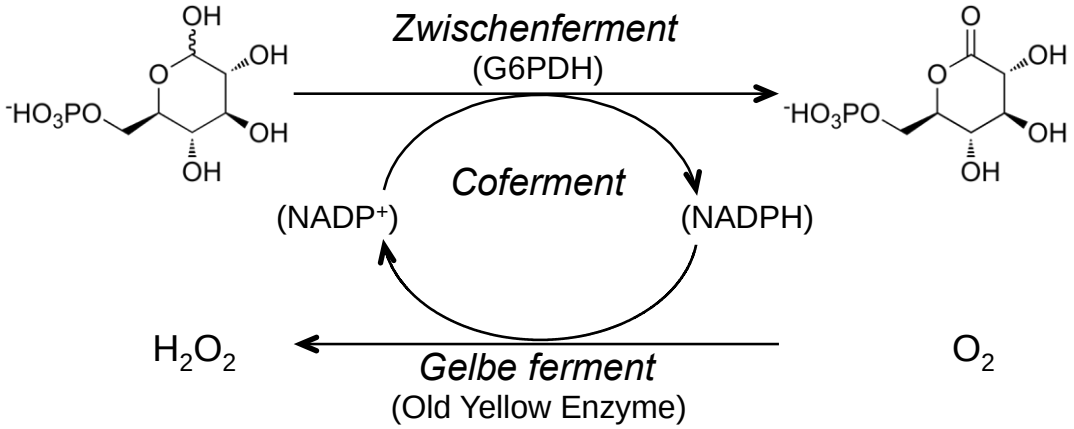
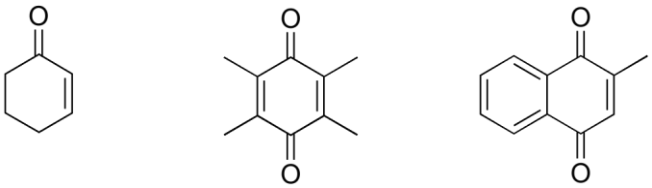


Figure 1-1. Warburg and Christian reaction system

Glucose-6-phosphate is oxidized by glucose-6-phosphate dehydrogenase (*zwischenferment*) in the presence of $NADP^+$ (*coferment*). Old Yellow Enzyme (*gelbe ferment*) serves to regenerate $NADP^+$ by facilitating a 2-electron transfer from NADPH to molecular oxygen.[1]

Stott *et al*. (1993)



(ref.) (145%) (5%)

Figure 1-2a. OYE1 substrates.

In Figures 1-2a through 2g substrates are introduced in chronological order and are presented only if identified as a new substrate for OYE 1. Reference compounds (ref.), where available, are indicated for relative comparison of activity, turnover number, or conversion within a given study. By default, unsaturated carbonyl compounds are drawn in the “normal” binding mode orientation such that the hydride hydrogen adds to the β-position from below the plane of the page and the proton hydrogen adds to the α-position from above the plane of the page to yield the stereoselectivity analogous to that observed by Swiderska and Stewart[14]. Where absolute stereochemistry was determined (Figures 1-2d through 2g) substrates are drawn in a binding mode that corresponds to the major product enantiomer. In these cases, dashed lines indicate substrates that bind in a “flipped” conformation.

Vaz *et al*. (1995)

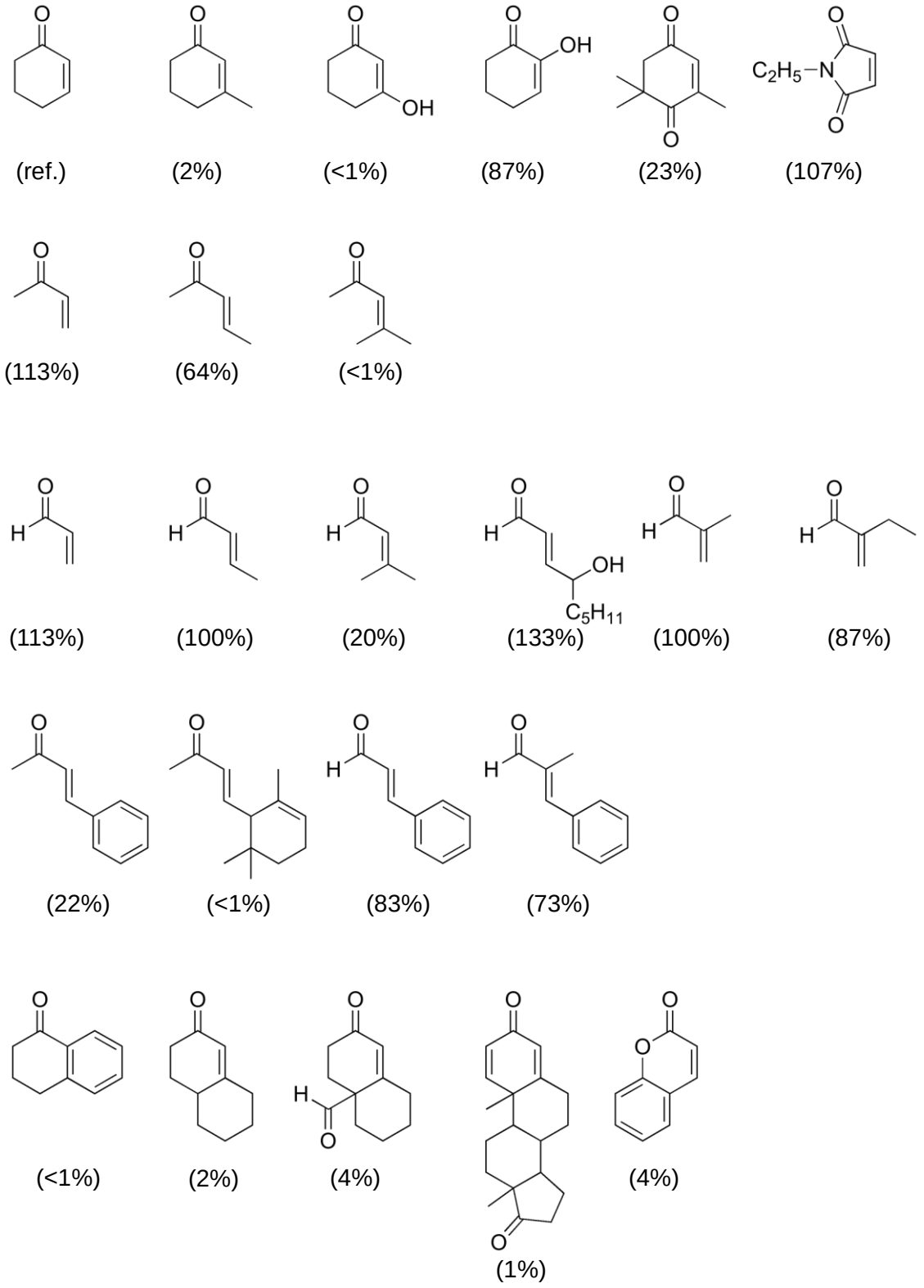
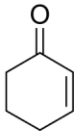
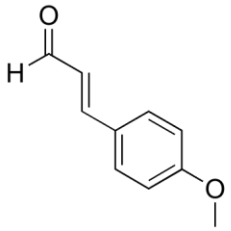


Figure 1-2b. OYE1 substrates.

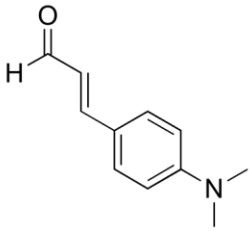
Kohli and Massey (1998)



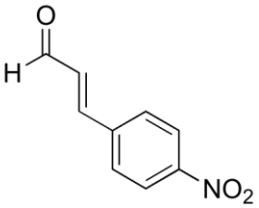
(ref.)



(260%)

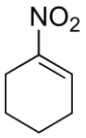


(<1%)

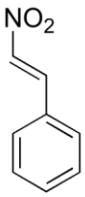


(5%)

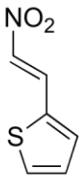
Meah and Massey (2000)



(ref.)

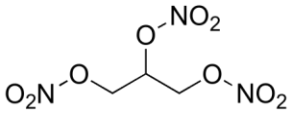


(88%)

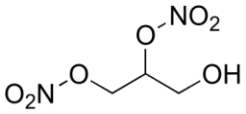


(98%)

Meah *et al*. (2001)

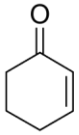


(ref.)

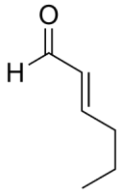


(45%)

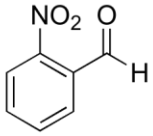
Williams *et al*. (2004)



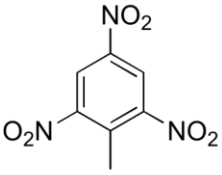
(ref.)



(82%)



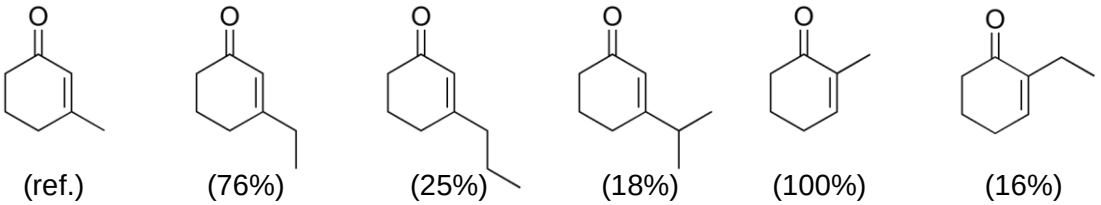
(61%)



(63%)

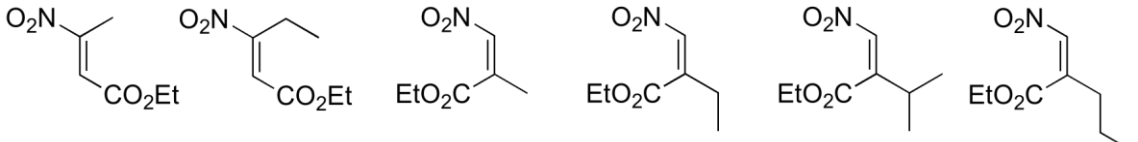
Figure 1-2c. OYE1 substrates.

Swiderska and Stewart (2006a)



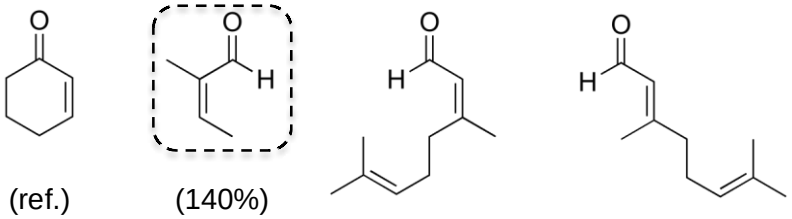
(ref.) (76%) (25%) (18%) (100%) (16%)

Swiderska and Stewart (2006b)



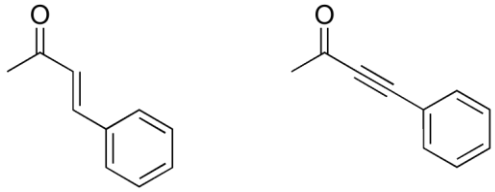
(ref.) (50%) (100%) (100%) (100%) (100%)

Mueller *et al*. (2007a)



(ref.) (140%) (10%) (110%)

Mueller *et al*. (2007b)



(ref.) (400%)

Bougioukou and Stewart (2008)

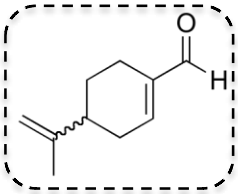
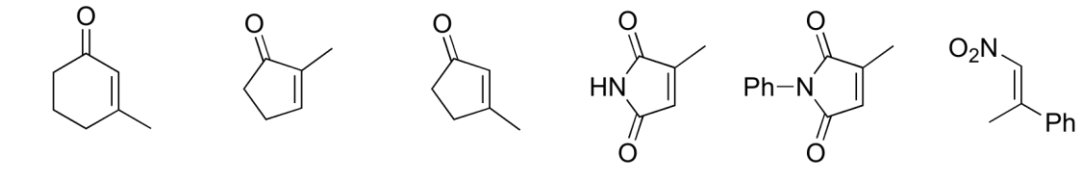
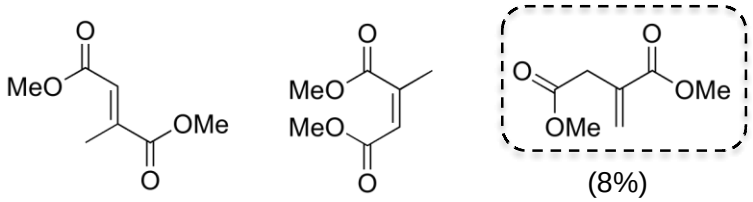


Figure 1-2d. OYE1 substrates.

Hall *et al*. (2008)

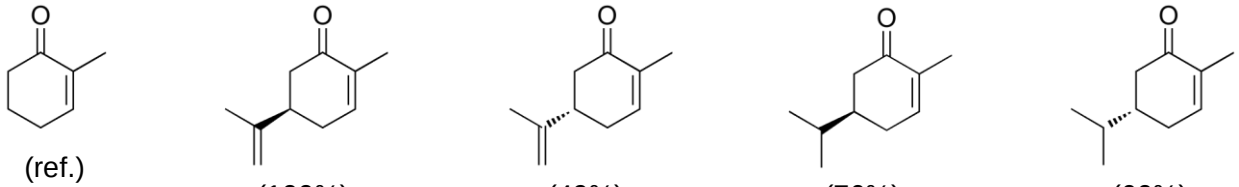


(ref.) (101%) (56%) (103%) (103%) (101%)



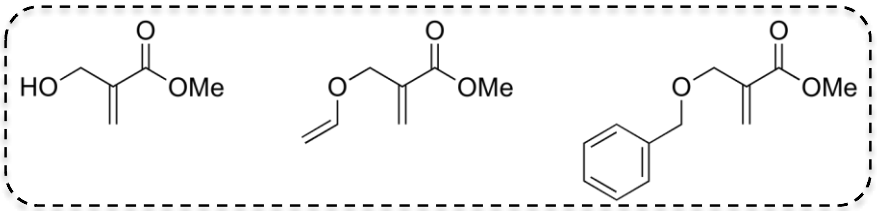
(103%) (60%) (8%)

Padhi *et al*. (2009)



(ref.) (100%) (49%) (76%) (60%)

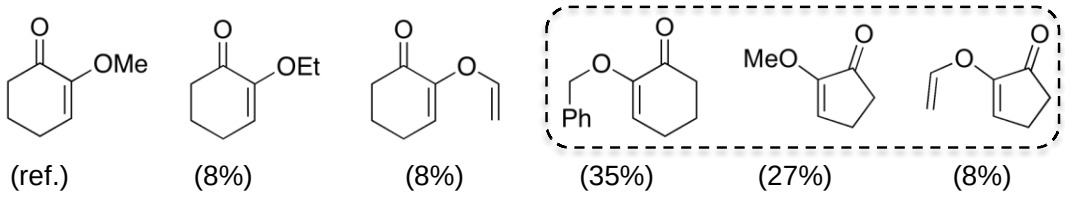
Stueckler *et al*. (2010a)



(ref.) (70%) (90%)

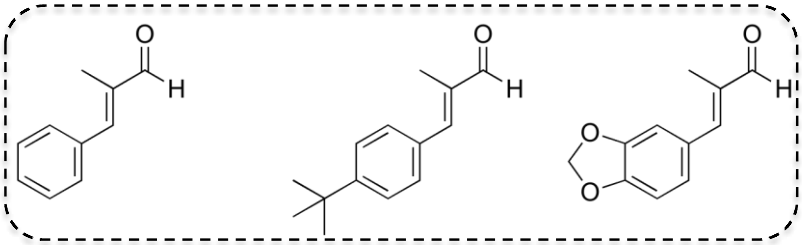
Figure 1-2e. OYE1 substrates.

Winkler *et al*. (2010)



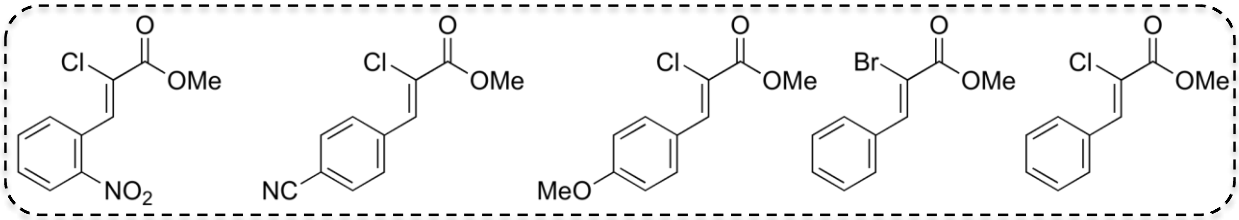
(ref.) (8%) (8%) (35%) (27%) (8%)

Stueckler *et al*. (2010b)



(ref.) Lilial™ (61%) Helional™ (59%)

Brenna *et al*. (2011)



(ref.) (74%) (37%) (26%) (10%)

Korpak and Pietruszk (2011)

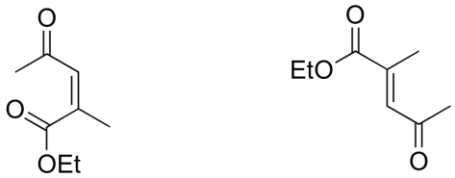


Figure 1-2f. OYE1 substrates.

Stueckler *et al*. (2011)

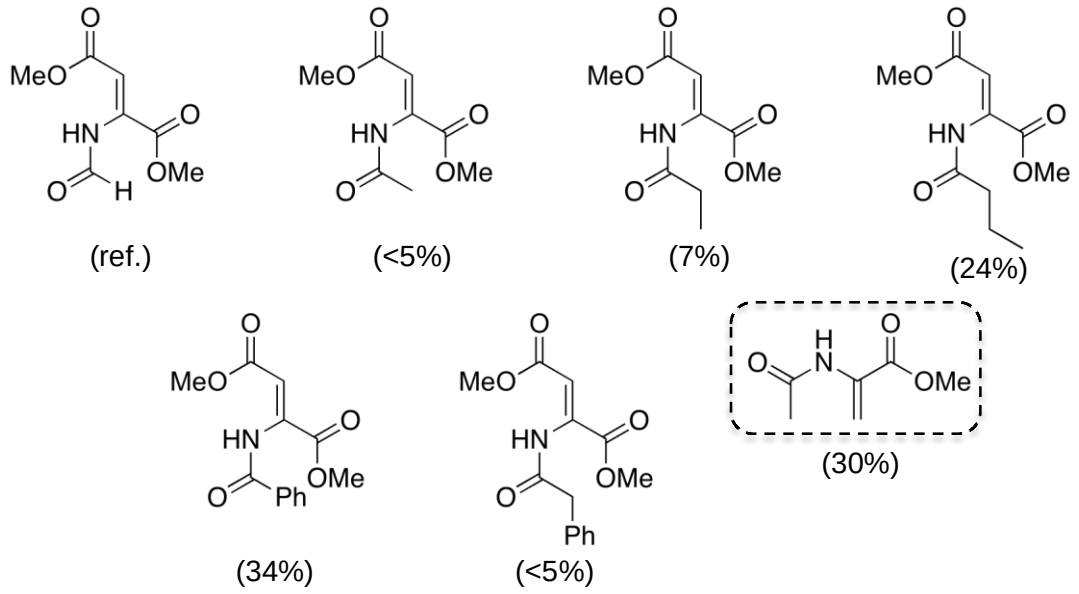


Figure 1-2g. OYE1 substrates.

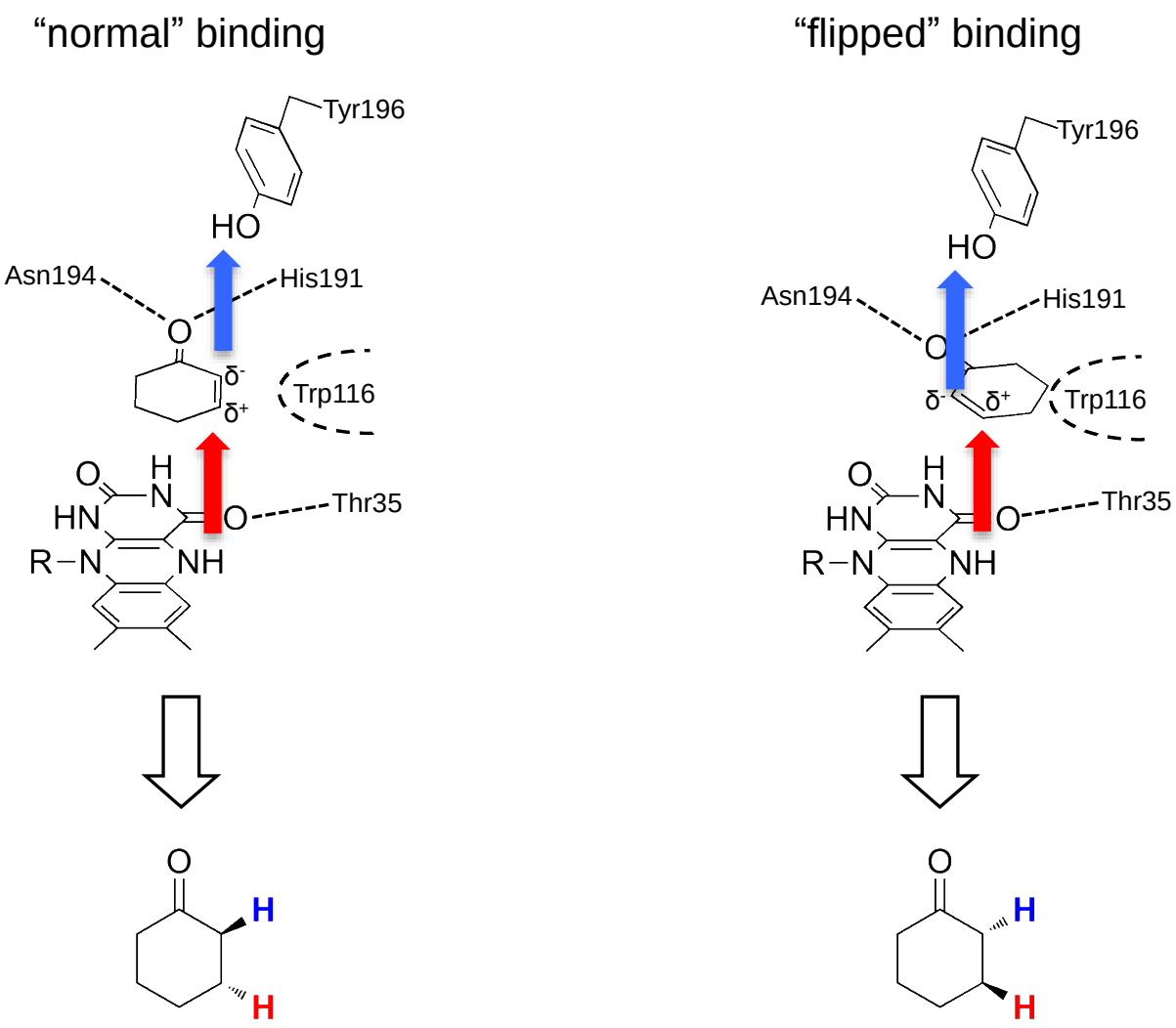


Figure 1-3. OYE substrate binding modes

In both binding modes, the model substrate (cyclohexenone) docks in the active site above and parallel to the plane of the reduced isoalloxazine ring system. The carbonyl oxygen forms hydrogen bonds with the side chains of residues Asn194 and His191. The hydride from N5 is transferred (red arrow) to the electron deficient β-carbon. Tyr196 functions as a general acid to protonate (blue arrow) the resulting enolate at the α-carbon. The “flipped” conformation theoretically requires a shift in the angle of hydride and proton transfer and is sterically crowded by the presence of Trp116. The stereochemistry of each binding mode product is indicated with the protic hydrogen (blue) and the hydride hydrogen (red) below each scheme. For a prochiral substrate the two binding modes determine product stereochemistry.

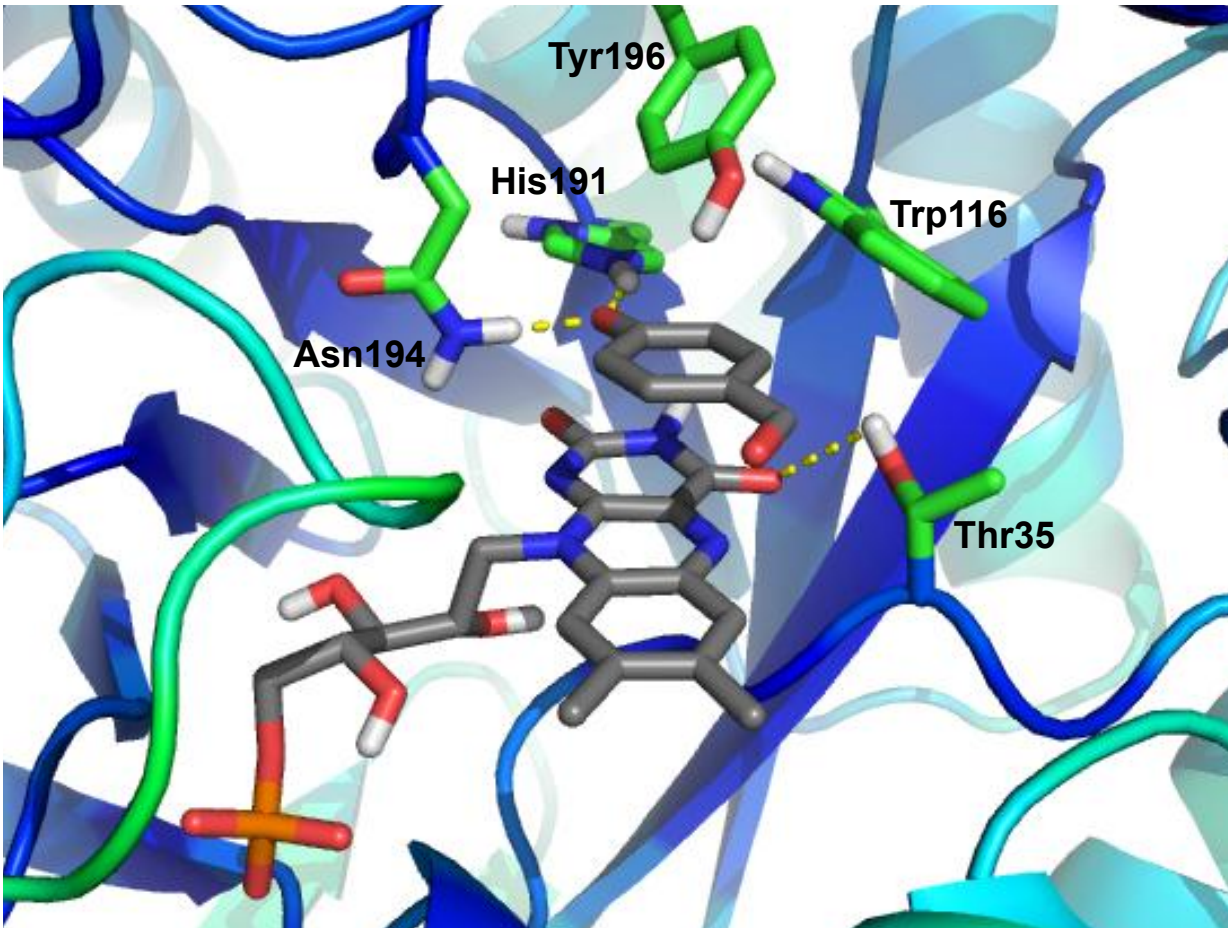


Figure 1-4. OYE 1 active site crystal structure.

Active site view of OYE 1 complexed with *p*-hydroxybenzaldehyde (PDB: 1OYB) as viewed from the entrance of the active site. Ligands are depicted in stick format with carbons in grey. Selected residue side-chains are labeled and depicted with carbons in green. Key hydrogen bond interactions are indicated by dashed yellow lines.

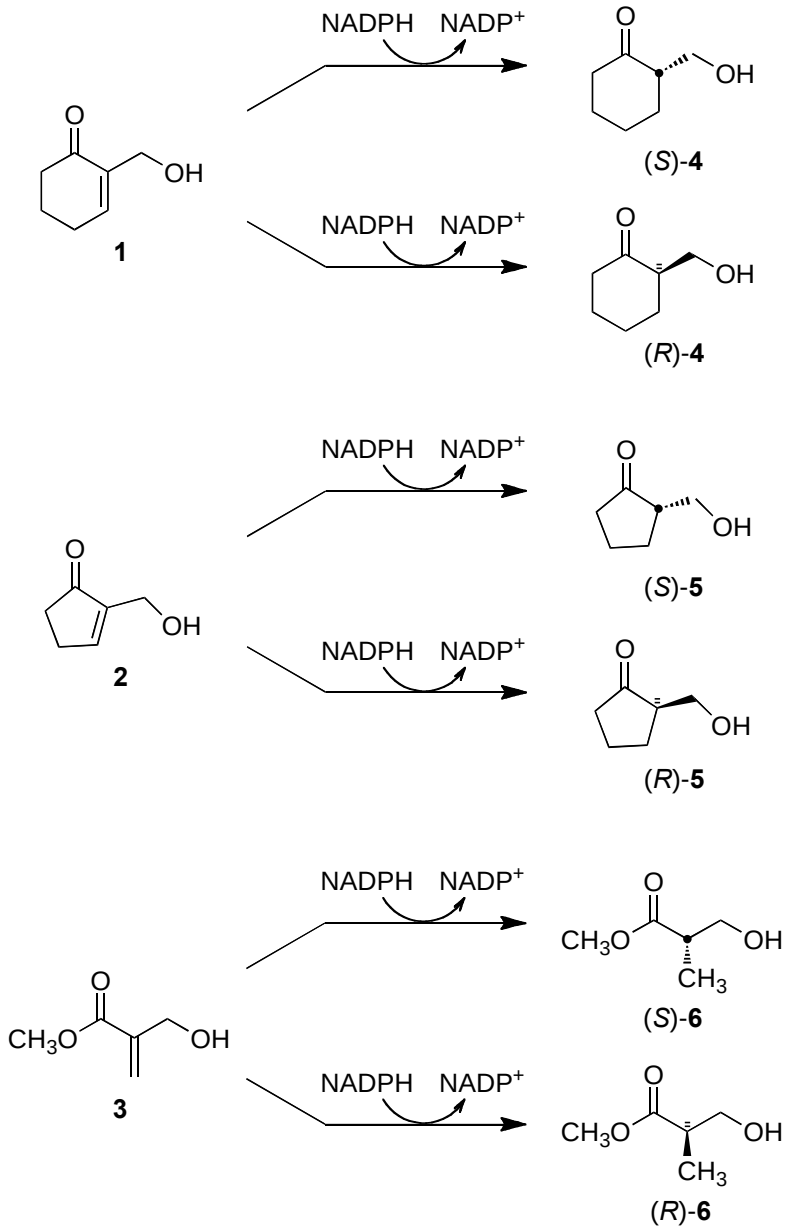


Figure 1-5. Reaction scheme for the reduction of Baylis-Hillman adducts

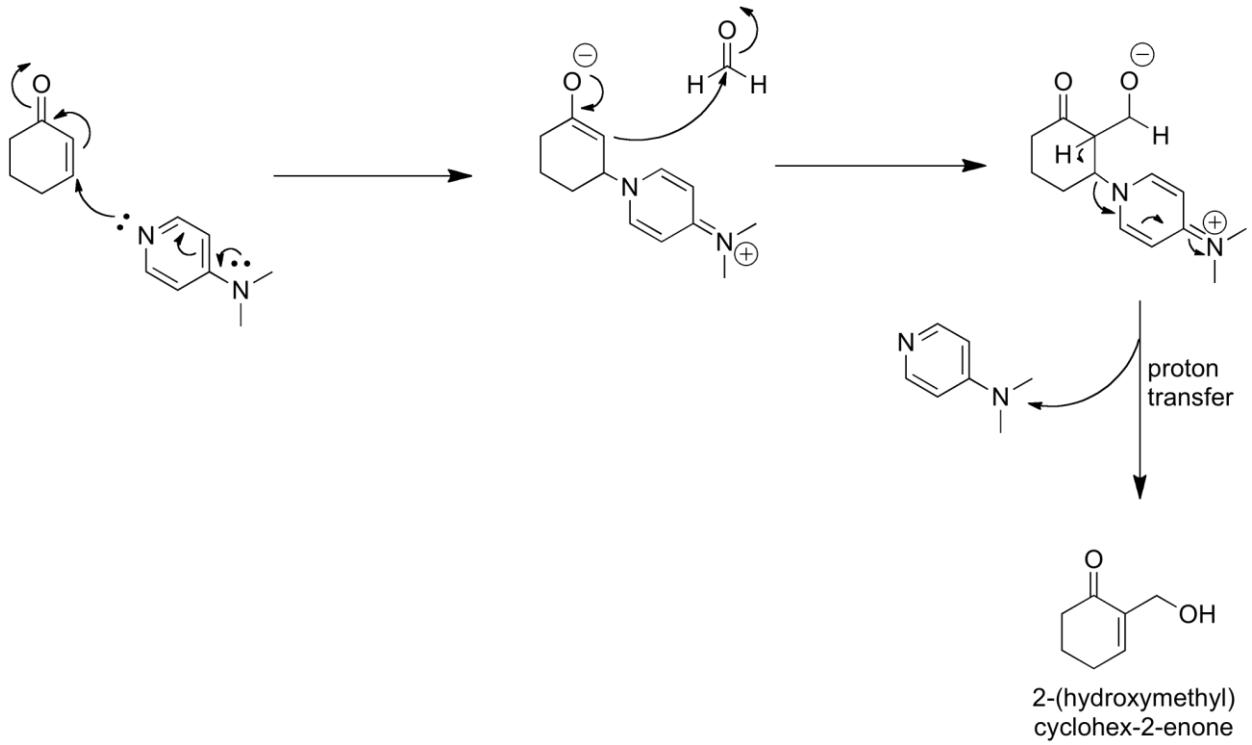
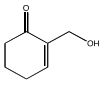
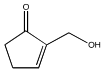
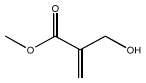
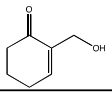
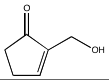
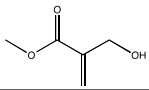


Figure 1-6. Baylis-Hillman reaction mechanism

Enzyme	% conv.	e.e.	% conv.	e.e.	% conv.	% e.e.
P. stipitis OYE 2.6	99%	**≥95% (*S*)**	99%	**76% (*S*)**	99%	**≥95% (*S*)**
S. pastorianus OYE1	4%	n.d.	51%	60% (*R*)	19%	≥95% (*R*)
S. cerevisiae OYE2	3%	n.d.	47%	71% (*R*)	11%	≥95% (*R*)
S. cerevisiae OYE3	-	-	10%	25% (*R*)	5%	n.d.
P. stipitis OYE 3.3	6%	n.d.	6%	n.d.	-	-
K. marxianus OYE	-	-	12%	59% (*R*)	2%	n.d.
E. coli NemA	5%	n.d.	46%	70% (*R*)	-	-
P. putida NemA	36%	94% (*S*)	92%	71% (*S*)	-	-
S. elongatus OYE	9%	69% (*S*)	62%	45% (*R*)	-	-
A. thaliana OPR1	1%	n.d.	1%	n.d.	3%	n.d.
A. thaliana OPR3	4%	n.d.	9%	51% (*R*)	-	-
S. lysopersicon OPR	7%	n.d.	41%	41% (*R*)	2%	n.d.

Figure 1-7. Results of OYE homolog screening with Baylis-Hillman adducts

	OYE1-W116X	conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	4%	n.d.	1%	n.d.	-	-
	K	60%	>95%	75%	76%	1%	n.d
-	D	99%	>95%	95%	77%	5%	n.d.
	E	93%	90%	96%	88%	3%	n.d.
polar	C	31%	>95%	47%	77%	-	-
	M	99%	>95%	99%	86%	15%	64%
	H	99%	>95%	99%	77%	67%	>95%
	N	99%	>95%	99%	89%	99%	41%
	Q	99%	>95%	99%	89%	78%	>95%
	S	84%	>95%	87%	>90%	13%	**46% (*R*)**
	T	28%	>95%	44%	>90%	3%	n.d.
non-polar	P	14%	>95%	16%	77%	-	-
	G	98%	>95%	99%	86%	14%	16%
	A	84%	>95%	99%	72%	9%	n.d.
	V	84%	>95%	97%	>90%	52%	**86% (*R*)**
	L	99%	>95%	99%	57%	99%	20%
	I	99%	>95%	99%	91%	50%	**9% (*R*)**
aromatic	F	99%	>95%	99%	85%	37%	**70% (*R*)**
	Y	99%	>95%	99%	87%	68%	**76% (*R*)**
	W	4%	n.d.	51%	**60% (*R*)**	19%	**>95% (*R*)**

Figure 1-8. Results of W116X OYE1 screening with Baylis-Hillman adducts

All ee% listed are for the (S)-enantiomer except where indicated in **bold**.

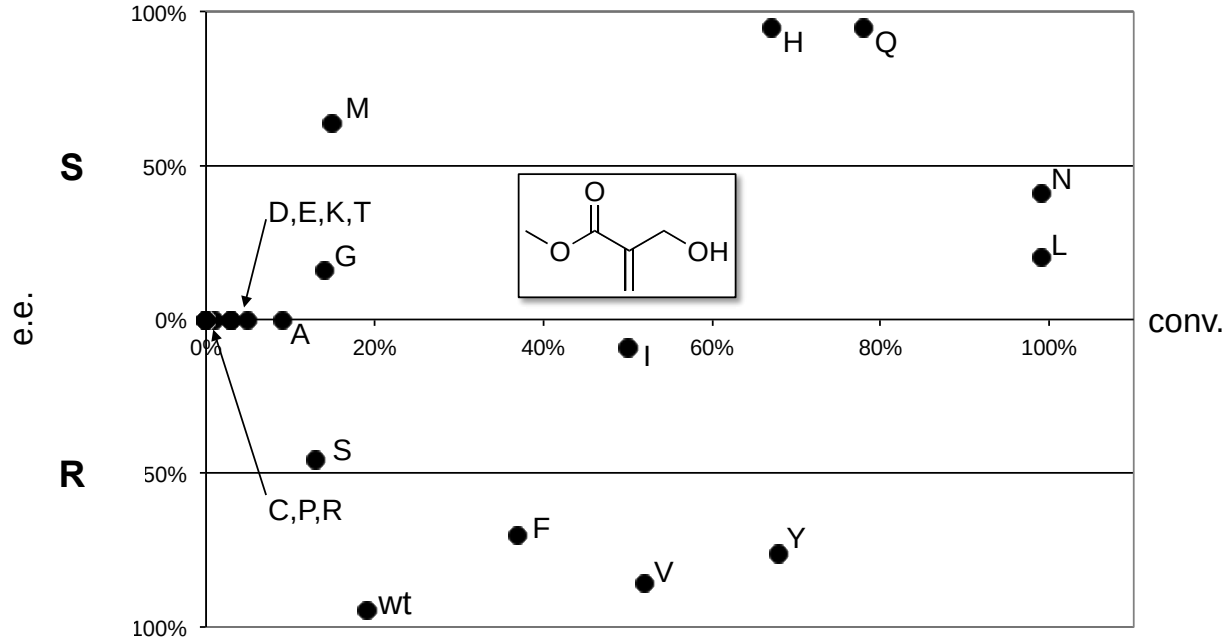


Figure 1-9. Results of W116X OYE1 screening with methyl 2-(hydroxymethyl)acrylate

Select data from Figure 1-8 with e.e.% on the vertical axis and total conversion on the horizontal axis.

CHAPTER 2
PREPARATIVE SCALE ENZYMATIC ALKENE REDUCTIONS

Background

To validate the utility of alkene reductases for industrial applications we desired to find model chiral compounds that have current relevance and production at industrial scale. A literature review revealed two interesting targets where biocatalytic asymmetric reductions might be useful. We therefore attempted to develop scalable methodologies for the production of (*R*)-citronellal and (*R*)-2-methylpentanol using alkene reductases in the key steps.

(*R*)-Citronellal

Menthol is a valuable commodity that is consumed in the production of food, pharmaceuticals, health-care products, cosmetics and a variety of synthetic applications.[52] While menthol has several isomers, the most desirable form is (-)-menthol for these applications[53] and approximately 12,000 metric tons of are consumed annually.[54] Of this amount, nearly 20% is produced by synthetic methods, whereas the remainder is isolated from plant species (*Mentha piperita*). The most common method for the production of synthetic menthol is the Takasago process (Takasago International Corporation). In this approach, myrcene is converted to (*R*)-citronellal using a chiral Rh BINAP catalyst.[55] The aldehyde intermediate is then cyclized using a heterogeneous zinc bromide catalyst followed by hydrogenation to produce pure (-)-menthol.

Previous work has shown that (*R*)-citronellal production by OYE homologs is possible; however the reported volumetric productivity was generally low (1.5 g/L/h) and reaction times longer than 1 hour were not investigated.[15] Preliminary studies by our group revealed that OYE 2.6[18], introduced in the previous chapter, yields (*R*)-citronellal

with high optical purity when provided with the α,β-unsaturated precursor, geranial. Commercially, geranial (*E*-isomer) is available as a component of citral, a mixture of geometric alkene isomers, which includes ~40% neral (*Z*-isomer) as a contaminant. Unfortunately the presence of this contaminant decreases both the rate of reaction and the optical purity of the final product when the results of geranial reduction by a series of OYE homologs was tested.[15] We proposed a reaction scheme (Figure 2-1) that commenced with the commercially available geraniol in geometrically pure form. Dehydrogeneases present in whole cells and crude cell extracts would facilitate the oxidation of geraniol to geranial. OYE 2.6 would then be used to reduce geranial to (*R*)-citronellal. In this scenario, NADPH would be recycled in a closed loop process. Unfortunately, attempts to use this scheme with whole cells and crude cell extracts overexpressing alkene reductases and an alcohol dehydrogenase were plagued by two problems. Wolken *et al*. demonstrated that isomerization can be facilitated by the presence of free amino acids in a buffer system.[56] Some isomerization of the unsaturated aldehyde was observed under our reaction conditions, resulting in a mixture of enantiomers. We also observed that the dehydrogenase present in our crude enzyme preparations had a tendency to reduce the desired product further to the unsaturated alcohol, citronellol. Combined, these two factors resulted in a complex mixture where after 6 hours a representative trial consisted of 7% geraniol starting material, 26% geranial, 19% isomerization product (neral), 48% over-reduction product ((*R*)- and (*S*)-citronellol) and *none* of the desired product ((*R*)-citronellal). This revealed that significant optimization would be needed before this approach was viable.

2-Methylpentanol

We also became interested in developing a biocatalytic route to (*R*)-2-methylpentanol ((*R*)-2MP)[57] because of its utility in the synthesis of multiple pharmaceutical compounds and in the production of liquid crystal technology. A previously described strategy was based on an evolved ketoreductase (KRED) that catalyzed the reduction of an aldehyde at the expense of the NADPH co-factor. In this case, the evolved KRED selectively reduced the (*R*)-enantiomer of racemic 2-methylvaleraldehyde. Unreacted (*S*)-2-methylvaleraldehyde was then removed by treatment with sodium bisulfite to form a water-soluble adduct that allowed the desired alcohol to be easily recovered. A volumetric productivity of 4.7 g/L/hour was reported (2 L scale).

We reasoned that an alkene reductase could provide a viable route to (*R*)-2MP and avoid the need for a kinetic resolution. In our plan 2-methylpentanal (available as the pure *trans*-isomer) would be reduced by an appropriate alkene reductase to yield enantiomerically pure methylvaleraldehyde (Figure 2-3). Reduction to the desired alcohol would take place *in situ* by the alcohol dehydrogenase activity endogenous to our crude cell extracts containing the overexpressed alkene reductase. Since both steps involve NADPH, a cofactor regeneration scheme would be required.

Experimental Strategy

Preparation of Enantiomerically Pure (*R*)- and (*S*)-Citronellal Using OYE Homologs

Our initial strategy to avoid interference between the dehydrogenase and alkene reduction involved compartmentalizing the two reactions. The idea was to generate the geranial *in situ* in one compartment ($NADP^+$ favored) but allow alkene reduction to

proceed only in a physically separated compartment (NADPH favored)(Figure 2-2). We first worked to optimize OYE 2.6-mediated alkene reduction. GST-affinity purified OYE 2.6 and the commercially available geranial source, citral, containing a mixture of isomers was used. We discovered that the enzyme displays instability in the presence of the aldehyde substrates and products. To simplify the process further we prepared geometrically pure geranial for subsequent work.

We initially attempted to compartmentalize the reaction system using a polydimethylsiloxane (PDMS) membrane as describe by Mwangi *et al*.[58] This approach was complicated by the tendency for the PDMS material to absorb both starting material and product, dramatically retarding reaction progress. Given the observation that the PDMS membrane appeared to function as a second phase we proceeded with investigation of solvent/aqueous biphasic reactions.

Our initial aim was to use a biphasic reaction system as a means to limit OYE 2.6 exposure to substrate and product. When this failed to address the issue of enzyme stability, we explored the use of cross-linked enzyme aggregates (CLEAs) since these have been shown to improve enzyme stability in both single and multi-enzyme systems.[59] General protocols for the formation of enzyme aggregates require precipitation of the enzyme (or enzyme/protein mixture) using ammonium sulfate followed by cross-linking of this material with glutaraldehyde or similar bi-functionalized reagent. Presumably, cross-linking occurs through the formation of a Schiff's base involving the exposed lysine residues, displaying a free -NH_2 group, on the protein surface and the cross-linking agent. The resulting aggregate retains enzymatic activity in its solid form and due to its nature can be regenerated by simple separation and

washing steps. Similarly and by the same logic, we pursued protection by acetylation of the soluble enzyme. The results of these isolation and protection efforts are described below.

When the above-mentioned strategies proved unsatisfactory, we adopted the most straightforward approach. This involved chemical oxidation of geraniol or nerol followed by alkene reductase-mediated conversions to yield the desired saturated aldehyde enantiomers. The alkene reductases were supplied as partially purified lysates from *E. coli* cells overexpressing OYE 2.6 (*P. stipitis*) or NemA (*E. coli*) respectively.[60] We also found that reaction progress could be monitored by the use of a pH stat where base demand is tied stoichiometrcally to the moles of gluconolactone formed by the NADPH recycling system (Figure 2-2).

Preparation of (*S*)-2-Methylpentanol Using OYE2 and YahK

Process development in this case was more straightforward. Commercially available 2-methyl-2-pentenal was sequentially reduced to (*S*)-methylvaleraldehyde by OYE 2 (*S. cerevisiae*) and then to (*S*)-2-methylpentanol by cloned alcohol dehydrogenase, *E. coli* YahK, in a single pot reaction (Figure 2-3). Glucose dehydrogenase (GDH) prepared from a donated plasmid pTgluDH3 (Biocatalyics), and processed as a crude cell lysate, was used to regenerate reducing equivalents (NADPH) for the reaction.

Experimental Procedures

GC/MS Analysis

GC/MS analyses were performed on a Chirasil-Dex CB column (25 m × 0.25 mm) using a mass-selective detector (EI, 70 eV). The temperature program for the reduction of citral reductions involved 60°C (2 min) followed by a 10°C/min increase to 180°C (10

min). Under these conditions, peaks eluted at 10.5 min (citronellal), 11.8 min (neral), 12.2 min (geranial) and 12.5 min (cintronellol).

Chiral analyses for the reduction of citral were performed on a βDex-225 column (30 m × 0.25 mm). The temperature program involved 95°C (35 min) followed by a 5°C/min increase to 160°C (2 min), then a 10°C/min increase to 200°C (5 min). Under these conditions, peaks eluted at 26.6 min ((*S*)-citronellal), 27.0 min ((*R*)-citronellal), 36.6 min ((*S*)-citronellol), 36.9 min ((*R*)-citronellol), 43.4 min (neral) and 46.2 min (geranial).

Chiral analyses for the production of 2-methylpentanol were performed on a Chirasil-Dex CB column (25 m × 0.25 mm) using a mass-selective detector (EI, 70 eV). The temperature program involved 65°C (3 min) followed by a 1°C/min increase to 75°C (2 min), then a 5°C/min increase to 120°C (2 min), and completed with a 20°C/min increase to 200°C (5 min). Under these conditions, peaks eluted at 5.6 min (2-methylvaleraldehyde), 7.8 min (2-methyl-2-pentenal), 18.2 min ((*R*)-2-methylpentanol) and 18.4 min ((*S*)- 2-methylpentanol).

PDMS Membrane Preparation

Polydimethylsiloxane (PDMS) reaction tubes were prepared using Sylgard® 184 Silicone Elastomer Kit (Dow Corning). The elastomer working solution (approximately 10 g) was prepared as described in the manufacturer's instructions. Approximately half of the solution was applied by transfer pipet to a 0.5-inch diameter brass rod while being turned on a horizontal rotisserie at room temperature. After turning for 30 minutes the rotisserie was transferred to a 65°C incubator and allowed to cure for one hour while turning. The single coat film was removed from the incubator and allowed to cool to room temperature before applying the remaining elastomer solution in the same

manner. The rod was then turned overnight at room temperature followed by final curing at 65°C for one hour. The two-layer film was removed from the rod by placing the cooled rod in a graduated cylinder containing hexane. The solvent swollen tube was then air dried to return to its original shape and trimmed to a desired length. The tubes were then sealed at one end by standing them on end in a thin layer fresh elastomer solution and cured for two hours at 65°C followed by trimming as necessary.

Small-scale Test Reactions

Reaction conditions for preparative scale citronellal production were optimized at an 8 mL working volume. These reactions were conducted in a 25 mL three-neck round bottom flask fitted with a pH probe and connected to an auto-burette delivering 1M KOH to maintain a static pH of 7.5. Reactions were conducted in phosphate buffer (100 mM) supplemented with glucose (200 mM) and $NADP^+$ (0.2 mM), which were degassed and purged with argon prior to adding enzyme. When investigating biphasic reaction protocols an 8 mL volume of hexanes was introduced to the vessel after degassing. OYE 2.6-GST (1 mg/mL) and GDH 102 (0.25 mg/mL, Codexis) were then added to the stirred reaction media and allowed to equilibrate for 15 minutes before adding substrate (50% v/v in ethanol). Alternatively, the same enzyme loading was also delivered as a CLEA (see below). Reaction progress was monitored by base addition confirmed periodically by GC-MS.

CLEA Preparation

Purified GST-OYE 2.6 (17 mg, 0.50 mL) was mixed with purified GDH-102 (6 mg) and 0.50 mL of 100 mM KP_i, pH 7.5. A 0.50 mL aliquot was transferred to a microcentrifuge tube and 0.50 mL of saturated $(NH_4)_2SO_4$ solution was added. The tube was rotated gently at 4°C for 15 min, then 384 mg of solid $(NH_4)_2SO_4$ was added the

tube was rotated gently at 4°C for an additional 15 min. Glutaraldehyde (15 µL, 50% aqueous solution) was added to a final concentration of 75 mM and the tube was rotated gently at 4°C for an additional 2 hr. The CLEA was collected by centrifugation and washed three times with cold 100 mM KP_i, pH 7.5.

Acetylation of GST-OYE 2.6.

Four microcentrifuge tubes containing 4 mg of purified GST-OYE 2.6 in a volume of 0.50 mL were prepared (1 - 4). Sodium acetate (190 mg, 50% aqueous solution) was added to tubes 2 and 4. Neat acetic anhydride (1 µL) was added to tubes 3 and 4. All tubes were gently rotated at 4°C for 1 hr, then samples 2 – 4 were dialyzed against 100 mM KPi, pH 7.5, 50 mM NaCl, 50% glycerol for 3 hr. Aliquots (250 µL) from samples 1 - 4 were transferred to microcentrifuge tubes and citronellal was added to a final concentration of 25 mM (28 µL of a 250 mM stock in EtOH) and gently rotated overnight at 4°C. Each sample, along with a control that was not treated with citronellal was diluted 1 : 10 with 100 mM KP_i, pH 7.5 and assayed for protein concentration and catalytic activity. Activity was measured by incubating 2.5 mM 2-cyclohexenone (from a 1 M EtOH stock solution) with 100 mM KP_i, pH 7.5 that contained 0.2 mM NADPH and an appropriate quantity of protein in a total volume of 1.0 mL. The change in A_{340} was measured at 25°C. The small change in A_{340} observed during the same time period in a control reaction lacking 2-cyclohexenone was subtracted from all of the determinations.

Preparation of (*R*)-Citronellal

Preparation of geranial

A slurry of 5.25 g of geraniol (98%, Sigma) (33.4 mmol), 10.85 g of activated manganese (IV) oxide (Fluka) (125 mmol), and 30 mL of methylene chloride was mixed and stirred at room temperature. After 24 hours, a 0.2 mL sample was diluted to 1.5 mL

and passed through a silica plug to remove any solid oxidant. This sample was determined to be approximately 70% oxidized when analyzed by GC-MS. An additional two equivalents of oxidant (63 mmol) was added and the reaction was allowed to proceed overnight. When analyzed at 46 hours in the same manner, all geraniol had been oxidized with the resulting product being a 96% geranial and 4% neral mixture. The reaction slurry was then passed over a silica pad and the solvent was removed under vacuum to yield 4.72 g (89% yield) of bright yellow oil. This product was stored at -20°C under argon until used.

Preparation of OYE 2.6 catalyst

E. coli BL21(DE3) cells overexpressing OYE 2.6 (*P. stipitis*) were grown on LB agar plates supplemented with 0.2 mg/mL ampicillin. A single colony was then used to prepare a 40 mL culture in LB media with ampicillin that was shaken overnight at 37°C. This pre-culture was used to inoculate 4 L of LB media supplemented with 80 mL of 20% glucose, 2 g/L ampicillin, and 0.5 mL antifoam AF 204 (Sigma). The cells were grown at 37°C, 700 rpm, and 4 L/min airflow for two hours until the OD_{600} = 0.6. At this time the temperature was decreased to 30°C and protein expression was induced with the addition of 0.48 mL of 840 mM iPTG. Growth was continued for an additional 4 hours until the cells reached an OD_{600} = 4.15. Cells were then harvested by centrifugation at 5000 rpm to yield 28.18 g wet cell weight and were stored at -20°C until the next step.

Cells from the previous growth were thawed at room temperature in 30 mL of 0.1M phosphate buffer (pH 7.5). The slurry was passed through a French Pressure cell twice at 15,000 psi in the presence of 10 µM PMSF. Insoluble material was then

removed by centrifugation at 15,000 rpm for 45 min to yield 44.5 mL of crude cell lysate. To this, 15 mL of room temperature saturated $(NH_4)_2SO_4$ was added slowly to bring the solution to 25% saturation and was then allowed to equilibrate for 10 minutes. Insoluble proteins were then removed by centrifugation at 15,000 rpm for 45 minutes to yield 55 mL of protein solution. To this, 5.0 g of solid $(NH_4)_2SO_4$ was slowly added while stirring and was allowed to equilibrate for 10 minutes. Insoluble proteins were separated by centrifugation and then redisolved in 25 mL of 0.1 M phosphate buffer and centrifuged once more to yield 29 mL of purified protein. When assayed this fraction contained 7.9 U/mL activity and approximately 47% of the total activity in the original crude preparation. Aliquots of this preparation were stored at -20°C until needed.

Enzymatic reduction of geranial to (*R*)-citronellal

The enzymatic reduction of geranial was conducted in a 500 mL three-neck round bottom flask stirred continuously at 300 rpm with a magnetic bar. The reaction media consisting of 85 mL of 0.1 M phosphate buffer (pH 7.5) containing 8.0 g of solid dextrose monohydrate was degassed for 1 hour prior to use. It was then transferred to the reaction vessel and kept under argon. To this, 1 mg (100 U) of GDH 102 (Biocatalytics), 10 mg NADPH (12 µmol), and 101 U of thawed and freshly assayed OYE 2.6 from the previous step was added. This mixture was allowed to equilibrate for 15 minutes at room temperature prior to the addition of substrate. Geranial (2.38 g) dissolved in 2.5 mL of EtOH was then divided into three portions and added to the reaction vessel upon initiation and then at 1.5 and 3 hours. Reaction pH was maintained at 7.5 by adding 1 M KOH via a pH stat. Base demand was monitored until it dropped to approximately ½ maximum and 16.5 mL of base had been added by 5.5

hours. After 5.75 hours, GC-MS analysis confirmed that the reaction was 95% complete. The mixture was acidified with 1M HCl to pH 4 and extracted with 100 mL of CH_2Cl_2 and allowed to stir overnight. The aqueous layer was extracted twice more with 50 mL of CH_2Cl_2 and the organic layers were combined and filtered over Celite to remove insoluble material. The filtered organic layer was then washed three times with brine and dried over Na_2SO_4. The resulting orange/brown solution was passed over a bed of silica resulting in a bright yellow solution. The solvent was then removed under vacuum to yield 2.12 g of yellow oil consisting of 91% citronellal when analyzed by GC-MS. This product was then loaded onto 5 g of silica deactivated with 10% H_2O and then onto a 60 g silica column equilibrated with hexanes. The product was eluted with 1:9 Et_2O/hexanes. Fractions containing citronellal as indicated by GC-MS were pooled and the solvent was removed under vacuum. The final yield was 1.59 g (10.2 mmol) of a pale yellow liquid, 99% pure by GC-MS, with an e.e. of 98%, $[\alpha]_D^{24} = +18.22$ (c = 7.30 in $CHCl_3$). Lit. $[\alpha]_D^{24} = +16.2$ (c = 1.00 in $CHCl_3$).[61]

Preparation of (*S*)-Citronellal

Preparation of neral

A slurry of 21.0 g of nerol (97%, Sigma) (136 mmol), 43.40 g of activated manganese (IV) oxide (Fluka) (500 mmol) and 150 mL of methylene chloride was prepared and stirred at room temperature. After 22 hours, a 0.2 mL sample was diluted to 1.5 mL and passed through a silica plug to remove any solid oxidant. GC-MS analysis showed that 10% of the initial starting material was still present. An additional 125 mmol of oxidant was added and the reaction was allowed to proceed for four more hours. After 27 hours, the reaction slurry was passed over a silica pad and the solvent was removed under vacuum to yield 20.30 g (98.2% yield) of yellow oil. GC-MS analysis showed the final

product mixture contained 97% neral and 3% geranial. This product was stored at -20°C under argon until used.

Preparation of NemA catalyst

E. coli BL21(DE3) cells overexpressing the OYE homolog NemA (*N*-ethylmaleimide reductase, *E. coli*) were grown on LB agar plates supplemented with 0.2 mg/mL ampicillin. A single colony was then used to prepare a 40 mL culture in LB media with ampicillin that was shaken overnight at 37°C. This growth was used to inoculate 4 L of LB media supplemented with 80 mL of 20% glucose, 2 g/L ampicillin, and 0.5 mL antifoam AF 204 (Sigma). The cells were grown at 37°C, 700 rpm, and 4 L/min airflow for 2.5 hours until the OD_{600} = 0.8. At this time the temperature was decreased to 30°C and protein expression was induced with the addition of 0.48 mL of 840 mM iPTG. Growth was continued for an additional 4 hours until the cells reached an OD_{600} = 4.98. Cells were then harvested by centrifugation at 5000 rpm to yield 28.15 g wet cell weight and were stored at -20°C overnight. After thawing at room temperature in 30 mL of 0.1M phosphate buffer (pH 7.5) the slurry was passed through a French Pressure cell twice at 15,000 psi in the presence of 10 µM PMSF. Insoluble material was then removed by centrifugation at 15,000 rpm for 45 min to yield 45 mL of crude cell lysate. When assayed this crude contained 44 U/mL activity and was used as is.

Enzymatic reduction of neral to (*S*)-citronellal

The enzymatic reduction of neral was conducted in a 500 mL three-neck round bottom flask stirred continuously at 300 rpm with a magnetic bar. The reaction media consisting of 85 mL of 0.1 M phosphate buffer (pH 7.5) containing 8.0 g of solid

dextrose monohydrate was degassed for 1 hour prior to use. It was then transferred to the reaction vessel and kept under argon. To this 1 mg (100 U) of GDH 102 (Biocatalytics), 10 mg NADPH (12 μmol), and 484 U of freshly assayed crude NemA from the previous step was added. This mixture was allowed to equilibrate for 15 minutes at room temperature prior to the addition of substrate. 2.35 g of prepared neral was mixed with 2.5 mL of EtOH was then divided and added to the reaction vessel in equal portions upon initiation and then at 1.5 and 2.5 hours. Reaction pH was maintained at 7.5 by adding 1 M KOH via a pH stat. Base demand (1 M KOH) was monitored until it dropped noticeably after 13.5 mL of base had been added by 4 hours. GC-MS confirmed that at 3.5 hours that the reaction was approximately 100% converted. The mixture was then acidified with 1M HCl to pH 4 and extracted with 100 mL of CH_2Cl_2 and allowed to stir overnight. The aqueous layer was extracted twice more with 50 mL of CH_2Cl_2 and the organic layers were combined and filtered over Celite to remove insoluble material. The filtered organic solution was then washed three times with brine and dried over Na_2SO_4. The resulting pale yellow solution was passed over a bed of silica resulting in a pale yellow solution. The solvent was then removed under vacuum to yield 1.62 g (10.3 mmol) of pale yellow oil consisting of 98% citronellal when analyzed by GC-MS. The observed e.e. was >99% and $[\alpha]_D^{24} = -13.71$ ($c = 5.55$ in $CHCl_3$). Lit. $[\alpha]_D^{24} = -16.2$ ($c = 1.00$ in $CHCl_3$).[61]

Preparation of (*S*)-2-Methylpentanol

Preparation of OYE2 and YahK catalyst

E. coli BL21(DE3) cells overexpressing OYE 2 (*S. cerevisiae*), YahK (*E. coli*), and GDH (pTgluDH3, Biocatalyics) were grown separately on LB agar plates supplemented with 0.2 mg/mL ampicillin. The same procedure was used for all three overexpression

strains. A single colony was then used to prepare a 40 mL culture in LB media with ampicillin and this was shaken overnight at 37°C. The pre-culture was used to inoculate 4 L of LB media supplemented with 80 mL of 20% glucose, 2 g/L ampicillin, and 0.5 mL antifoam AF 204 (Sigma) in a New Brunswick Scientific M19 fermenter. The cells were grown at 37°C, 700 rpm, and 4 L/min airflow for two hours until the OD_{600} ~ 0.6. At this time the temperature was decreased to 30°C and protein expression was induced with the addition of 0.48 mL of 840 mM iPTG. Growth was continued for an additional 4 hours until the cells reached an OD_{600} ~ 4. Cells were then harvested by centrifugation at 5000 rpm to yield approximately 7 g/L w.c.w. and were stored at -20°C until the next step.

Cell pellets were thawed in 0.1M phosphate buffer (pH 7.5). The slurries were passed through a French Pressure cell twice at 15,000 psi in the presence of 10 µM PMSF. Insoluble material was them removed by centrifugation at 15,000 rpm for 45 min to yield crude cell lysates which were then divided into aliquots and either used immediately or stored at -20°C until needed. Prior to use, activity was estimated by measuring the decrease in A_{340} at 25°C in pH 7.5 phosphate buffer (100 mM) containing NADPH (0.2 mM) and 10 mM of the appropriate substrate (2-methyl-2-pentenal, 2-methylvaleraldehyde, or glucose).

Enzymatic reduction of 2-methyl-2-pentenal to (*S*)-2-methylpentanol

The enzymatic reduction of 2-methyl-2-pentenal was conducted in a 500 mL three-neck round bottom flask stirred continuously at 300 rpm with a magnetic bar. The reaction media consisting of 90 mL of 0.1 M phosphate buffer (pH 7.5) containing 8.0 g of solid dextrose monohydrate was degassed for 1 hour prior to use. It was then

transferred to the reaction vessel and kept under argon. To this, 1 mL (300 U) of GDH lysate, 10 mg $NADP^+$ (13 µmol), and 325 U of freshly assayed OYE 2 lysate from the previous step was added. This mixture was allowed to equilibrate for 30 minutes at room temperature prior to the addition of substrate. 2-methyl-2-pentenal (0.30 g, Sigma) was then added to the reaction vessel. Reaction pH was maintained at 7.5 by adding 1 M KOH via a pH stat. Base demand indicated that approximately half of the substrate had been reduced after 1 hour. At that time 1.5 mL YahK (122 U) was added to the reaction to initiate reduction to the final product. A second 0.30 g aliquot of neat 2-methyl-2-pentenal was added at 2 hours when base demand indicated that approximately 90% of the initial bolus had been consumed. Base demand was monitored until it dropped to approximately half the maximal value and a total of 15.5 mL of base had been added by 5.5 hours. The reaction was then acidified with 1M HCl to pH 4 and stirred overnight with 100 mL CH_2Cl_2. The aqueous layer was extracted twice more with 100 mL of CH_2Cl_2 and the organic layers were combined and filtered over Celite to remove insoluble material. The filtered organic layer was then washed once with 1M $NaHCO_3$ and twice with brine and dried over Na_2SO_4. The solvent was then removed under vacuum to yield 0.37 g (59% yield) of a pale yellow oil consisting of ~98% 2-methylpentanol when analyzed by GC. Analysis of the final crude product indicates an e.e. of 99.6%, $[\alpha]_D^{24} = -11.10$ (c = 4.30 in $CHCl_3$). Lit. $[\alpha]_D^{20} = -8.2$ (c = 1.48 in $CHCl_3$).[62]

Results and Discussion

(*R*)- and (*S*)-Citronellal

Attempts to use a biphasic system led to the conclusion that substrate partitioning severely limited the overall reaction rates by effectively reducing the concentration of

substrate available to the enzyme, which remained in the aqueous phase. Figure 2-4 shows the reaction time course for the reduction of citral (3 : 2, geranial : neral) in a two-phase system. The data indicated that OYE 2.6 preferentially reduces geranial over neral. Because it was not clear whether the decrease in neral level was due to isomerization to geranial or by direct reduction of neral to citronellal, geometrically pure geranial was substituted for citral.

Figure 2-5 shows the relative reaction progress of the biphasic reaction compared to a single-phase aqueous reaction. Both reactions appear to stop after approximately 12 hours. Surprisingly, the single-phase aqueous reaction yielded approximately 2-fold higher product titer than the 2-phase strategy. We therefore carried out all subsequent studies under single-phase conditions.

Given that the enzyme tended to inactivate after exposure to the aldehyde, we pursued preparation of CLEAs. The assumption was that the immobilized aggregate would prove more stable in the presence of our aldehyde substrates. This was based on the fact that the cross-linking agents were also aldehydes and might therefore pre-condition the enzyme prior to substrate addition. We found that the reaction rate for aggregate enzymes was dramatically reduced for our system (Figure 2-5) and displayed approximately a 4-fold decrease in activity from that of the free enzyme experiments in both single- and bi-phase approaches.

Acetylation was investigated as an approach to enhance enzyme stability without the decreased reaction rates associated with CLEA immobilization. We used the acetylation protocol described by Riordan and Vallee[63] and discovered that the diminished activity for the enzyme yielded no apparent protection from aldehyde

exposure. Table 2-1 shows that after the acetylation protocol, both the untreated and acetylated enzyme (NaOAc, AcO_2) preparations saw a ~16% reduction in specific activity after 24 hour incubation with citronellal. Therefore, we concluded no benefit was obtained by acetylation and further attempts to functionalize the enzyme were abandoned.

Based on these preliminary results, the final optimized processes for producing (*R*)- and (*S*)-citronellal utilized unmodified alkene reductases and aqueous reaction mixtures in a single compartment. Our final reaction schemes (Figure 2-6 and 2-7) deviated from our original concept (Figure 2-2) of a compartmentalized reaction scenario. In this approach, alcohol oxidation and alkene reductions were separated both spatially and temporally. The productivity of these schemes are listed in the Table 2-2 and are consistent with general requirements for economically feasible industrial biocatalysis.[3] Under our reaction conditions we observed very little geometric isomerization of the substrate, geranial or neral. This eliminated the previously assumed requirement to produce geranial or neral *in situ*. Ultimately, our solution to the carbonyl over-reduction problem was to eliminate dehydrogenase activity by using simple ammonium sulfate fractionation of the crude cell lysate prior to alkene reduction. While the volumetric productivities listed in Table 2-2 certainly are capable of improvement through process design, they are further limited by enzyme exposure to a total aldehyde concentration of ~ 200 mM.

(*S*)-2-Methylpentanol

Screening against our available alkene reductase library revealed that no enzyme could afford the (*R*)-enantionmer of the desired target. On the other hand, OYE 2 from *Saccharomyces cerevisiae* yielded high enantiomeric excess for the (*S*)-enantiomer. To

accomplish the second carbonyl reduction, we introduced a crude preparation of a previously cloned alcohol dehydrogenase, YahK (*E. coli*). The YahK enzyme displayed high activity toward the primary aldehydes in this study and could be used to enhance the rate of secondary reduction as needed in the reaction scheme (Figure 2-3).

This process was complicated by two phenomena. First, the intermediate methylvaleraldhyde produced in the reaction demonstrated the tendency to racemize under reaction conditions. Secondly, the YahK dehydrogenase had a 3 : 1 preference for reducing the saturated aldehyde compared to the carbonyl of the unsaturated starting material; however, at lower valeraldehyde concentrations the enzyme would alternatively reduce the pentenal starting material to produce quantities of the undesired allylic alcohol. As described in the experimental section, the addition of the YahK enzyme was delayed until approximately 50% of the initial substrate bolus had reacted. This served to avoid both intermediate reacemization and staring material reduction. As with the citronellal example, the total aldehyde exposure proved problematic and thus limited substrate additions to ~60 mM. The productivity of this reaction scheme was determined to be 0.67 g/L/h, significantly less than the approach published described by Gooding *et al*.[57], but was capable of yielding pure product with high enantiomeric excess (Table 2-2).

Conclusion

Using simple reaction schemes, we were able to demonstrate the applicability of alkene reductases in preparation of valuable synthetic intermediates. With the exception of the 2-methylpentanol target, the reaction schemes developed in this study show sufficiently high volumetric productivity to be economically feasible.[3] Aldehyde targets continue to be challenging substrates due to their apparent toxicity toward alkene

reductase enzymes. Future applications of alkene reductases in the production of enantiomerically pure primary aldehydes and alcohols could be greatly enhanced with the development of more substrate tolerant enzyme variants.

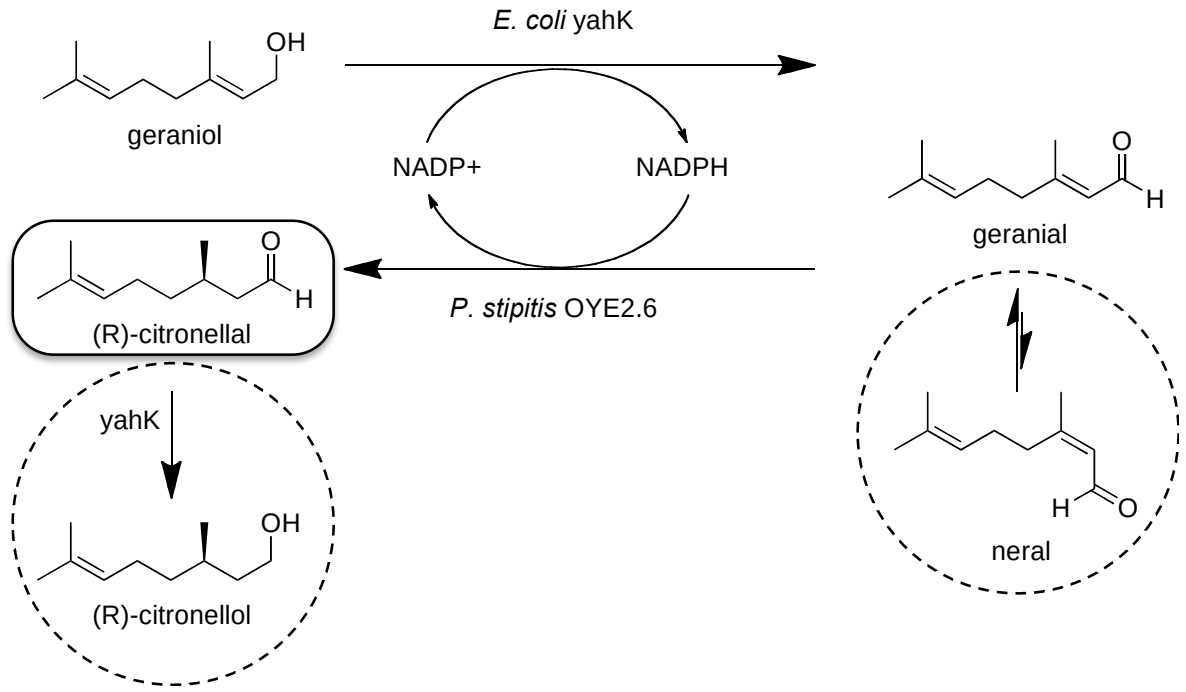


Figure 2-1. Initial proposed scheme for the preparation of (*R*)-citronellal

Geraniol starting material is available from Sigma Aldrich at >97% purity and $0.06 per gram. Enzymes are used as crude cell lysates after protein expression in *E. coli* BL21 (DE3). Exogenous $NADP^+$ cofactor is added to a concentration of 0.2 mM. Dashed line circles indicate observed undesired side reactions.

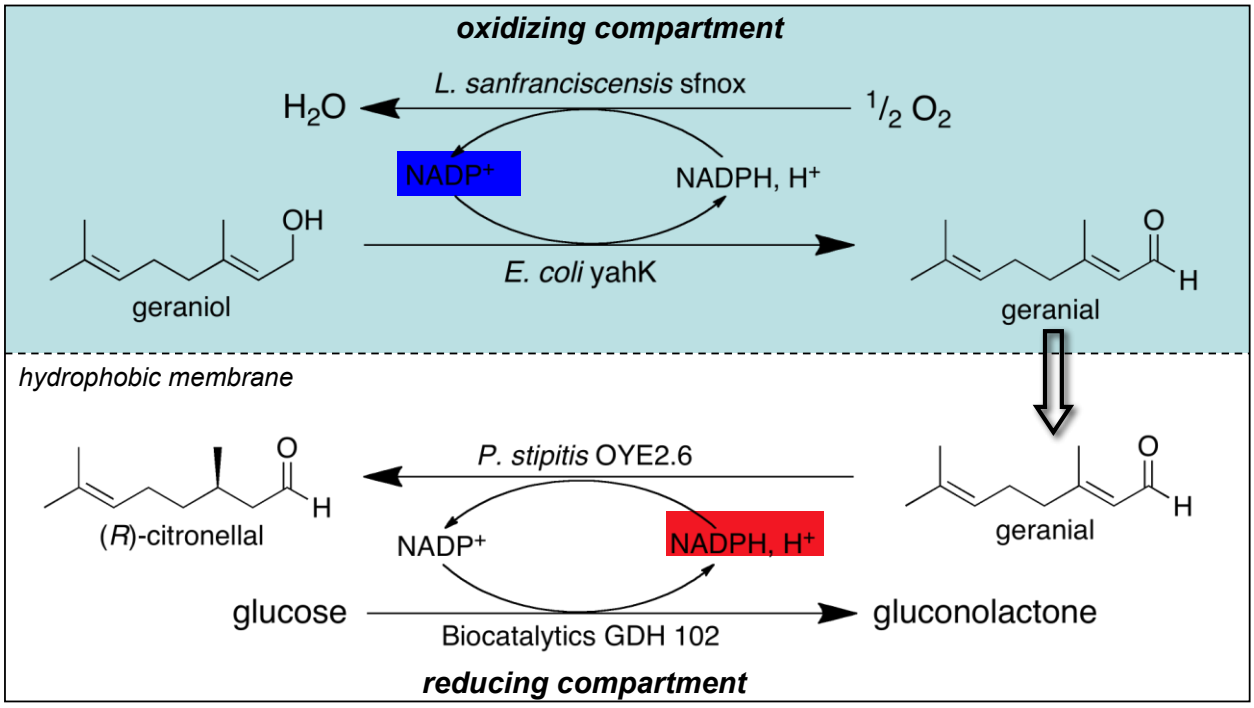


Figure 2-2. Proposed 2-compartment scheme for avoiding undesired side reactions

NADPH oxidase from *L. sanfranciscensis* is used to regenerate oxidized cofactor without production of peroxide.[64] All organic species are able to migrate across the indicated membrane (dashed line). In the reducing compartment, reduction of the unsaturated aldehyde is driven by the irreversible formation of gluconolactone.

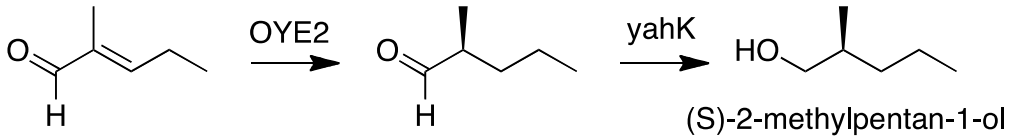


Figure 2-3. General scheme for the preparation of (*S*)-2-Methylpentanol

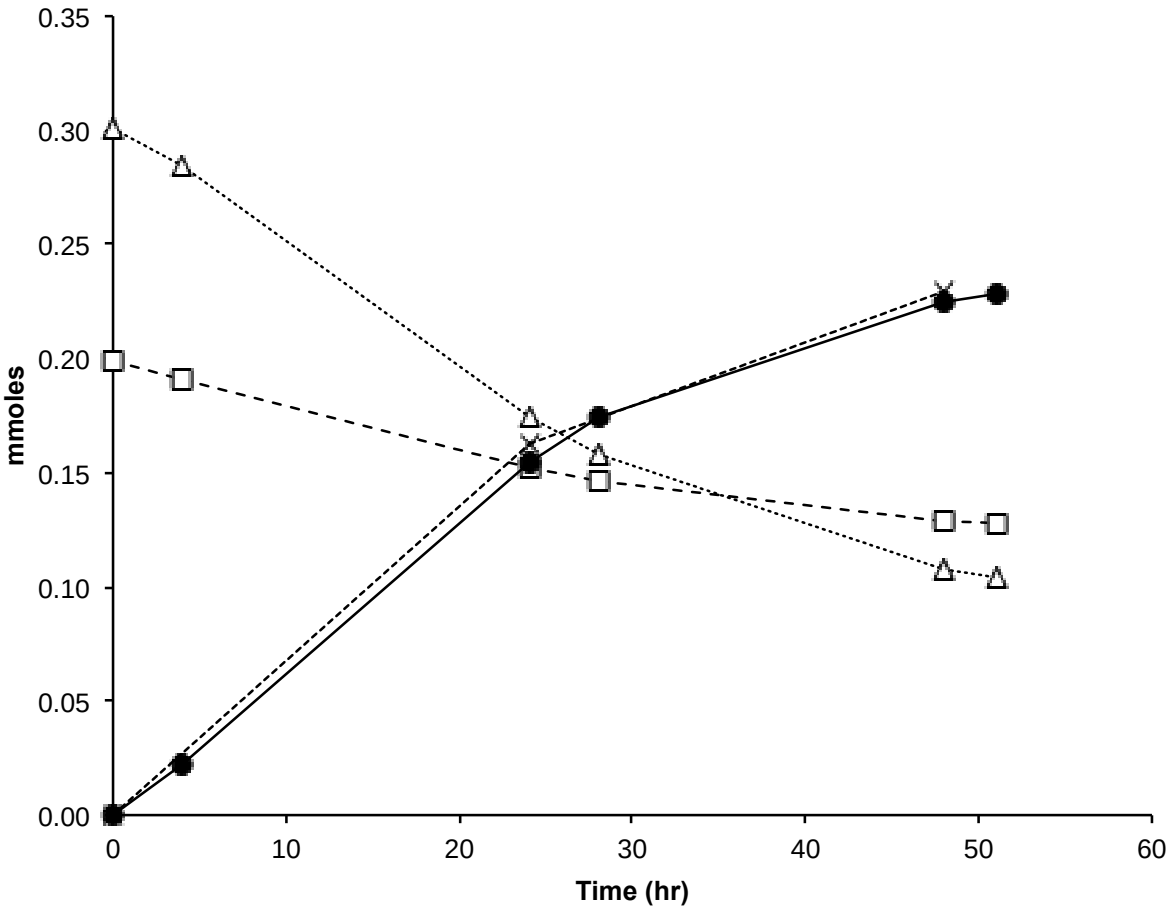


Figure 2-4. Time course for OYE 2.6 reduction of citral under biphasic conditions

Quantities determined by GC are plotted as mmoles present (geranial, ∆; neral, □; citronellal, •) or mmoles consumed (glucose, ×). The reaction was carried out in a 1 : 1 mixture of hexanes and 100 mM KP_i, pH 7.5 using purified GST-OYE 2.6 and GDH at room temperature.

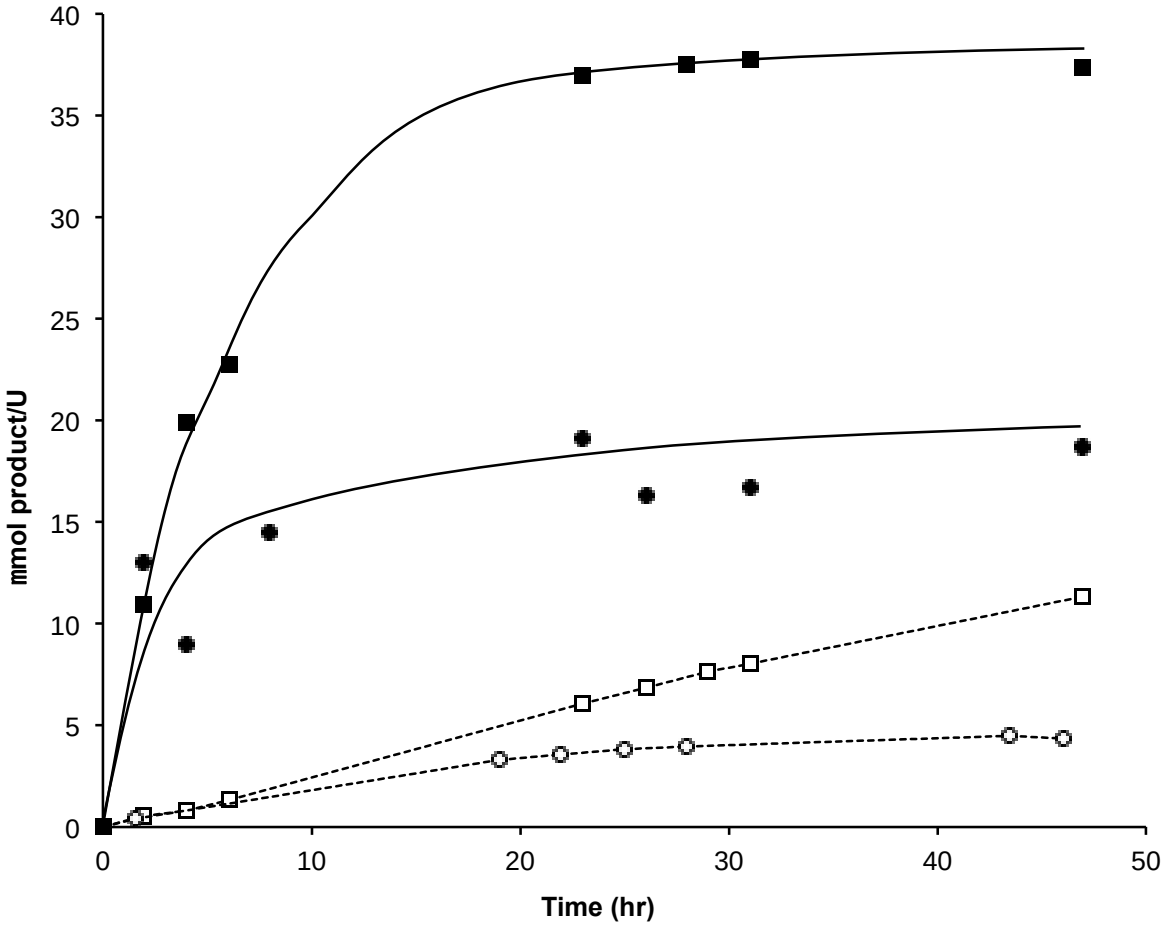


Figure 2-5. Comparison of strategies for OYE 2.6-mediated reduction of geranial

Quantities determined by GC are plotted as mmoles of citronellal formed per unit of GST-OYE 2.6 activity. The enzyme was employed as free protein in buffer (■) or 1 : 1 mixture of hexanes and buffer (•) or as a CLEA in buffer (□) or a 1 : 1 mixture of hexanes and buffer (**O**).

Table 2-1. Acetylation of OYE 2.6

	untreated	NaOAc	Ac_2O	NaOAc, Ac_2O
not incubated	1.75 ± 0.17	1.43 ± 0.08	0.12 ± 0.01	1.02 ± 0.04
incubated (25mM citronellal)	1.47 ± 0.09	1.46 ± 0.14	0.26 ± 0.02	0.84 ± 0.04

Specific activity of enzyme preparations are listed in μmoles of cyclohexene-2-one reduced per minute per mg of protein as measured by the decrease in the A_{340} peak of NADPH cofactor. The acetylation protocol used requires the addition acetic anhydride to the protein in the presence sodium acetate. Samples treated with only sodium acetate or acetic anhydride were run as controls. Following treatment, samples were incubated with 25 mM aldehyde product, where indicated, for 24 hours prior to specific activity assays.

Table 2-2. Productivity of alkene reductase biotransformations

Product	e.e. %	Productivity g product/L/h	Productivity μmol product/min/g d.c.w.
(*R*)-citronellal	98	2.8	17
(*S*)-citronellal	>99	4.1	25
(*S*)-2-methylpentanol	>99	0.67	6

Dry cell weight (d.c.w.) of alkene reductase/dehydrogenase catalyst is calculated using 0.23 conversion factor from wet weight.[65]

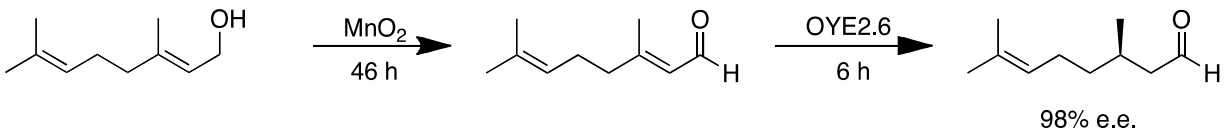


Figure 2-6. General scheme for the preparation of (*R)*-citronellal

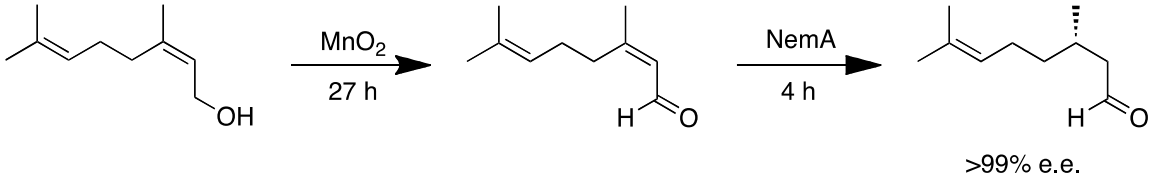


Figure 2-7. General scheme for the preparation of (*S)*-citronellal

CHAPTER 3
PROTEIN ENGINEERING TO EVOLVE ENANTIOSELECTIVITY

Introduction

Our interest in advanced protein engineering techniques is fundamentally a proof of principle endeavor. While approaches to engineering vary from rational to non-rational we have observed the power and potential of a semi-rational approach described in Chapter 1 and elsewhere.[18] The next logical step in complexity is that of directed evolution which by definition potentially requires successive combinations of beneficial mutations. When one considers the sheer magnitude of effort typically expended on these iterative cycles one can easily be dissuaded. We did not attempt to continue evolutionary studies until ”ideal” variants were obtained. Instead, our goal was to use directed evolution as a tool to explore structure-function relationships in alkene reductases.

Directed Evolution

Evolution in nature has really one ultimate strategy: survival. However in directed or laboratory evolution of proteins, the researcher controls the process and strategies leading to the desired outcome. *In vitro* protein engineering has become increasingly common over the past 20 years (Figure 3-1). Typically, the objectives of these efforts are thermo-stability or solvent tolerance, however, enantioselectivity or stereoselectivity have become increasingly common goals in recent years.

Protein engineering in the laboratory requires a sound strategy for success. Manipulating proteins requires consideration of the following facts: 1) Protein sequence space is large: with 20 amino acids, there are 20^{300} different ways to assemble a 300 residue enzyme. Add to that the three dimensional nature of protein structure and the

possible number of combinations approach the infinite. 2) Protein space is mostly empty of function. 3) The prospect of finding combinations of beneficial mutations are extremely rare.[66] The two basic approaches to protein engineering are rational design and directed evolution (Figure 3-2). In cases where much is known about the protein or enzyme structure and function, rational design can yield significant results.[67] Rational design in this case is defined as the introduction of a deliberate mutation for the purpose of changing the property of a given enzyme. In cases where little is known about the enzyme or how to evolve the desired property, directed evolution is often the preferred solution.

Directed evolution can be summarized as successive rounds of mutational pressure on a desired gene followed by subsequent screening that allows for selection of an improved gene product. The two most critical phases of a directed evolution effort are the generation of sufficient diversity so that a beneficial change can occur and a screening or selection strategy that uncovers these beneficial mutations. The following sections will outline existing approaches to mutagenesis and screening as applied to directed evolution.

Generating Diversity

After having identified the target protein of interest, the first step in directed evolution is creating the diversity. Numerous approaches have been applied to this task.[68] In general, they can be classified into three broad categories: recombinative (random), non-recombinative (random) and semi-rational approaches.

Recombinative methods

Recombinative approaches, often termed sexual evolution because of their attempt to mimic combination of two parent genes as in nature[66], involve various *in vitro* techniques based on the principle of the original method of DNA shuffling.

DNA shuffling generates large libraries of potentially beneficial mutations. The original approach by Stemmer involved random fragmentation of a double stranded parent gene using DNaseI[69,70] followed by PCR amplification of resulting fragments both with and without primers for the gene. The result was a reassembled gene with a relatively high frequency of point mutations due to various conditions of the PCR reaction.[71] When a single gene is subject to DNA shuffling, the diversity generated is the result of these point mutations. It may also be desired to shuffle homologous genes in an attempt to combine properties of multiple parent genes. In this approach, point mutations may mask or complicate the effort, however, optimization to reduce these unintended mutations is possible.[71]

The staggered extension process (StEP) is an alternate method for random combinations of multiple closely related parent genes.[72] In this approach, a single primer is introduced into a pool of homologous genes and subject to annealing and a significantly abbreviated extension step. The partially extended primer is then re-annealed to one of the parent templates in the pool and the process repeated. The result is a fully extended gene that is a random mixture of all the parents in the pool.

Similar to these approaches is a technique called random chimeragenesis on transient templates (RACHITT).[73] In RACHITT a pool of single strand homologous genes is partially digested by DNaseI. The fragments are then annealed to a homologous single strand template not included in the original digestion pool. Non-

hybridized ends (5' and 3') are removed by a combination of endonuclease activity of *Taq* and *Pfu* DNA polymerases as well as gaps filled between the annealed fragments. The strand is ligated and then cloned into an appropriate vector. The crossover is the junction point in a reassembled sequence where a template switch takes place from one parent sequence to another.[74] Crossovers are influenced by the homology and the size of the fragment annealing to the template. The RACHITT strategy has the benefit of much higher number of crossovers per gene as compared to previous methods as well as a significant increase in crossovers in regions with less than ten base pairs of sequence identity.

The efficacy of these approaches has been furthered by the advent of methodology termed assembly of designed oligonucleotides (ADO).[75,76] In these approaches, synthetic oligonucleotides containing degeneracy at targeted positions are shuffled and reassembled using their inherent sequence overlap in a PCR based assembly process. This allows the researcher to insert a variety of point mutations in combination throughout the gene as well as predefine the number of crossover events in the mutagenic product.

The previously described approaches to recombinative mutations all require a certain degree of sequence homology between the members of the recombinative population theoretically limiting the scope of the directed evolution effort. Several methods have been developed to harness the utility of recombinative methods with a lack of sequence homology among the theoretical parents.

Ostermeier *et al*. developed several approaches based on an incremental truncation approach.[77] In these approaches, two genes are designated for

recombination at a single crossover point. One gene is digested by controlled exonuclease activity from the 3'-end the other gene is digested in a similar manner from the 5'-end and the reactions are quenched at set time intervals. The result of these digestions is gene fragments of all possible lengths for each gene. The resulting fragments are ligated and fusion products corresponding to an approximate size are isolated by gel-electrophoresis. While there is a chance that the fusion products are ligated out of frame and therefore useless the overall fusion construct was recombined at random from two potentially non-homologous genes. This method and variations of it are given the title incremental truncation for the creation of hybrid enzymes (ITCHY). Traditional DNA shuffling of ITCHY libraries (termed SCRATCHY) can then be used to generate further diversity given that the ITCHY library now contains a degree of homology required.

Sieber *et al*. further developed an approach to fuse non-homologous genes with a single crossover point while keeping the construct in frame.[78] This approach is termed sequence homology–independent protein recombination (SHIPREC). In this approach genes are joined by a linker sequence containing useful restriction sites. This dimer is subject to fragmentation by DNaseI, treated to produce blunt ends and subsequently ligated to form circular constructs. Circular fragments that correspond to the size of a single gene are separated by gel electrophoresis. The linker is then removed by restriction digest revealing genes with crossover points distributed throughout the new gene construct.

Non-recombinative methods

One of the oldest and potentially the most frequently used technique for diversity generation in directed evolution is error prone PCR (epPCR). This approach exploits

the potentially low fidelity of DNA polymerases under certain conditions to insert random base substitutions in a given gene sequence during the PCR cycle.[79] The frequency of these mutations can be controlled by optimization of reaction conditions to yield anywhere from 1-20 nucleotide substitutions per 1 kb.[80] The most common methods for altering the fidelity of *Taq* DNA polymerase are the addition of Mn^{2+}, unbalanced nucleotide concentrations, and the use of nucleotide analogs. Alternatively, a lower fidelity mutant polymerase may also be selected.[81] The method has the advantage of being able to insert mutations at nearly every position in the gene of interest. However, not all nucleotide substitutions lead to changes in amino acid incorporation. Base substitutions that occur in the third position often lead to no change in the amino acid at that position. Two sequential base substitutions would have a high probability of altering the amino acid but is statistically disfavored in the controlled epPCR reaction scenario. Statistical analysis of gene sequences indicates that only 45% of all nucleotide substitutions will yield an amino acid substitution. Of those substitutions, only 4-7 of the 20 possible amino acids will be represented at a given position with a disproportionate bias towards glycine or proline substitution.[82]

Another approach designed to overcome the inherent biases of epPCR is the technique of Sequence Saturation Mutagenesis (SeSaM). Key to this approach is the construction of single strand sequences of the gene of interest of all lengths using a bead immobilization technique. The population of varied length nucleotides is then elongated with deoxyinosine using a terminal transferase and then fully elongated by an additional PCR reaction. The inosine containing genes are then subject to PCR where the inosine is replaced in a theoretically unbiased manner with the four standard

nucleotides. The result is a gene that contains mutations throughout its length that are independent of the mutational biases of DNA polymerase.

Another random PCR based approach is the technique of random insertion/deletion (RID).[83] With this method a cyclic single strand copy of a gene is subject to random cleavage by a Ce(IV)–EDTA complex. Double strand anchor sequences containing the insertion sequence and unique restriction sites are then ligated to the randomly positioned ends of the cleaved gene. PCR fills in the new double stranded construct that is subsequently cleaved by restriction enzyme to remove the anchor sequence and the specified number of bases targeted for deletion. This blunt ended construct is ligated and digested again by unique restriction enzyme to reveal a full-length gene with the inserted sequence substituted randomly throughout the gene. To add diversity to the insertion library, anchors constructed with degenerate bases can be used.

Semi-rational methods

Semi-rational approaches are so termed because they require some insight into the overall structure and function of the target enzyme. That is to say that, we have some idea of where to look for improved function; therefore, that area of the protein space is targeted in a random manner. The earliest example of this approach is combinatorial cassette mutagenesis (CCM).[84] In this method, an oligonucleotide mimicking the target site on the gene is synthesized with degenerate codons. The resulting degenerate cassette containing randomized nucleotides at desired positions is then ligated into the gene to create a targeted mutant library for screening. A variation of this method has been described that reduces the redundancy of the genetic code by creating a cassette randomized with a single codon for each of the 20 amino acids.[85] In

this approach a set of 20 primers coding for each amino acid (MAX) is constructed for each position in a given cassette. These primers containing MAX randomization are hybridized with a synthetic template that is randomized with NNN at the same positions. The construct is then ligated, amplified by PCR, and restriction digested to isolate a cassette suitable for cloning. The resulting cassette therefore lacks the degeneracy of the genetic code theoretically simplifying subsequent screening efforts.

One drawback to the use of cassette-based strategies is the requirement for the parent gene to contain restriction sites suitable for ligation of the synthetic cassette. A solution to this is to introduce degeneracy via mutagenic primers using a variation[86] of the sequence overlap extension (SOE)[87] method. This approach requires the use of complementary mutagenic primers that are first used in separate reactions to amplify the forward and reverse halves of the gene when paired with appropriate non-mutagenic primers. The two new overlapping halves are combined and extended via PCR to generate a double stranded construct suitable for cloning.

Another approach to saturation mutagenesis is the use of a whole plasmid strategy based on the popular QuikChange® mutagenesis method[88] marketed by Stratagene. When applied with degenerate mutagenic primers, the method was patented as Gene Site Saturation Mutagenesis (GSSM)[89] by Diversa Corporation, and is also referenced in the literature as mutagenic plasmid amplification (MPA).[86] The replication of the whole plasmid using degenerate mutagenic primers eliminates the need for cloning as with the previous methods. The method also employs the use of template plasmid from a *dam*+ *E. coli* strain such that it is susceptible to digestion by

DpnI endonuclease. This therefore eliminates bias toward the wild type enzyme in the degenerate library.

The introduction of multiple mutagenic sites using variations of these whole plasmid approaches has also been described. The QuikChange® Multi[90] kit as marketed by Stratagene requires the use of a single degenerate mutagenic primer for each site of interest in the gene. All mutagenic primers are designed in the same direction. After annealing, the plasmid is amplified and ligated in a single reaction using a polymerase and ligase blend. The parent template is then digested to leave a single strand circular multi-site mutated fragment that is then transformed in to competent cells. Each cell then contains a single copy of the multi-site mutant suitable for isolation and screening. An alternate approach to this has also been described[91,92] and recently optimized[93] based on the generation of mega-primers. In these methods sense and anti-sense primers located at distal sites and containing degenerate mutations are used to amplify a large fragment of a target plasmid. In the later stages of the PCR process, the large fragment functions as a mega-primer for the amplification of the whole plasmid followed by subsequent DpnI digestion.

Building on site saturation mutagenesis techniques, Reetz and Carballeira described an approach that mimicked the directed evolution schemes involving multiple random mutagenesis cycles termed Iterative Saturation Mutagenesis (ISM).[94] Figure 3-3 outlines ISM methodology. In the simplest version of their approach they use MPA to create degenerate mutagenic plasmids that are transformed into an expression strain and screened for desired activity. Positive results are then isolated, identified, and used as template for subsequent rounds of MPA at other beneficial sites. Key to their

approach is the semi-rational selection of sites subjected to saturation mutagenesis. When enhanced thermostability was the objective residues, were selected based upon their crystallographic B-values. These values are a numeric reflection of smearing of atomic electron densities with respect to their equilibrium positions as a result of thermal motion and positional disorder.[94] Residues that display the highest degree of disorder are then subject to ISM. Conversely, when changes in substrate specificity or enantioselectivity were the objective, they proposed the use of a combinatorial active-site saturation test (CAST).[95] In this approach, residues in the active site are grouped based on their side chain relative proximity to each other. Co-randomization then in theory leads to synergistic effects not predictable by substitution of only a single amino acid side chain. When one considers that when two sites are randomized simultaneously with NNN degeneracy the statistical oversampling required to observe all possible combinations exceeds 10^4 (for three sites this number approaches 10^6, see Table 3-3) it is clear that the CAST approach requires a degree of modification to be practical.

The numbers problem

To make the numbers manageable, Reetz proposed the use of restricted codon libraries. Using NNK degeneracy (Table 3-1 for IUBMB abbreviations) this reduces the number of possible codons for 64 to 32 while still coding for all 20 amino acids. It also has the beneficial effect of reducing the number of stop codons from three to one thereby reducing the theoretical number of inactive mutants in a given population. NNK codon usage is listed in Table 3-4. Further restricted codon usage can be used to target subsets of amino acids further reducing the total screening effort. Examples of a few of these are listed in Table 3-2. For example, using an NDT degenerate codon will

reduce the number of codons to 12 each coding for a single amino acid. These 12 amino acids are a balanced mix of polar, nonpolar, aliphatic, aromatic, negative and positive charged residues that exclude structurally similar amino acid side chains.[96] Using restricted codon libraries also facilitates combinatorial degeneracy by reducing the total number of possible combination requiring screening. Specifically NNN degeneracy at two sites would create 4,096 possible variants (64×64) whereas with NNK the number is reduced to 1,024 possible combinations with a 67% reduction in the number of stop codons.

Using Poisson statistics, Reetz *et al*. proposed the use of an oversampling factor, O_f, based on the percent probability of selecting all possible unbiased combinations of degenerate mutations.[96] In general terms, they determined that three-fold oversampling was required to achieve 95% coverage of all possible mutant combinations. These oversampling requirements are forecasted for a few degenerate profiles in Table 3-3. Using this as a guideline, a researcher can tailor the library generation to fit the required screening effort.

Screening

The second phase of directed evolution is the screening the generated diversity for beneficial mutations (Figure 3-2). Screening large libraries is often the bottleneck of a directed evolution project.[97] Typically, the reaction is chosen so that the enzyme yields a product that is both easily visible and non-transient in nature. In some scenarios, a pre-screen can be adapted to eliminate non-beneficial or deleterious mutations. The now reduced population can then be screened through more time consuming

approaches. Examples of some of these screening methods are described in the following sections.

The relative difficulty of a screening effort is more often than not a matter of perspective. That is to say, if high throughput technology is available for a desired screen then the generation of an unusually large degenerate library may not seem like an insurmountable task. In our case we are interested in the stereoselectivity of an enzyme where the product is most simply assayed by chiral gas chromatography. Depending on the product of interest the time to analyze 100 samples is measured in days (typically 1-3). Therefore, the construction of a 10,000-member library in a few days time could lead to a screening effort lasting a few months. Clearly, the size of the library generated must be proportional to the capabilities of the screening system and the patience of the researcher. The diversity generation methods using the ISM approach described previously are the most straightforward way to generate manageable libraries for use in a directed evolution effort given our screening parameters. While it appears that this semi-rational approach to directed evolution is the most manageable, its relative effectiveness compared to earlier traditional directed evolution methods needs to be evaluated.

Comparing Directed Evolution Approaches

Given the myriad of diversity generation techniques used in directed evolution it would be useful to compare successful approaches and the resulting outcomes. Many reviews have outlined directed evolution successes.[67,98-108] However, one significant problem arises in review of these works is that typically only successful directed evolution efforts are published. Paramesvaran *et al*. statistically compared the results of successful epPCR and site saturation approaches to determine the relative success of

each in obtaining the “optimal” directed evolution solution.[109] Their finding was that site saturation was more likely to identify the “optimal” solution when mutations in the active site affecting substrate specificity were concerned. But it is very rare to find successful directed evolution efforts that use the same starting point enzyme, screening conditions and substrates required to truly evaluate each approach side by side. Three particular examples serve to compare random diversity generation methods with semi-rational methods in their efficacy. These examples include those of *Escherichia coli* β-galactosidase, *Aspergillus niger* epoxide hydrolase, and *Pseudomonas aeruginosa* lipase.

β-galactosidase

β-galactosidase from *Escherichia coli* (BGAL) is an enzyme that catalyzes the hydrolysis of β-galactosides into their respective monosaccharides. The enzyme is also known to cleave the glycosidic bond in 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-Gal) into galactose and the respective indole that further dimerizes into an insoluble and intensely blue compound useful for visual screening. The enzyme is also known to catalyze the hydrolysis of glycosidic linkages in *p*-nitrophenyl β-D-galactopyranoside (PNPG), and *o*-nitrophenyl β-D-galactopyranoside (ONPG). The resulting nitrophenols are both soluble and yellow in color (420 nm) making them useful for measuring kinetic activity of β-galactosidase.

Using these known substrate specificities of β-galactosidase, Zhang *et al*. set about *in vitro* directed evolution of the enzyme using successive rounds of DNA shuffling.[110] The specific aim of this work was to evolve β-fucosidase activity in the native enzyme. They introduced low frequency point mutations into the β-galactosidase gene by DNA shuffling.[70] The resulting mutant plasmids were transformed into an

expression strain and plated on selection media containing 5-bromo-4-chloro-indolyl-β-D-fucopyranoside (X-Fuc). The resulting library of ~10,000 transformants contained 2-5% that stained blue by visual inspection. 20-40 of the bluest colonies were selected and their DNA pooled for seven successive rounds of DNA shuffling by the same technique. The most blue colony at the end of the seventh round was selected, sequenced and kinetically characterized. When characterized against *p*-nitrophenyl β-D-fucopyranoside (PNPF) and o-nitrophenyl β-D-fucopyranoside (ONPF) the resulting mutant showed a 300-fold increase in substrate specificity for PNPF and a 1000-fold increase in the specificity for ONPF over that of the wild type enzyme. The evolved gene contained 11 nucleotide substitutions that are further translated into 6 amino acid changes in the peptide sequence. Three substitutions (P511S, Q573R, and N604S) were thought to directly influence the active site and a fourth (D908N) may potentially influence the active site. The remaining amino acid changes lie on the surface of the protein and are thought to not be catalytically significant.

Nearly a decade later, Parikh and Matsumura used a semi-rational approach to evolving β-fucosidase activity.[111] In a direct comparison to the work described by Zhang *et al*. they focused their attention of three residues (Asp201, His540, and Asn604) known to coordinate a sodium ion in the enzyme active site. This analysis is similar to the CAST approach described earlier. Each of these three residues was randomized with NNK degeneracy simultaneously. After transforming the multi-site-saturated plasmid into an expression strain the library was screened as previously described. Approximately 10,000 mutants (25% coverage) were screened, a number significantly smaller than the 10^5 that required for 95% coverage. The eight transformants that

displayed the highest β-fucosidase activity remained unchanged at position 201. Remarkably they also contained the same mutations at the other two positions (H540V and N604T). By comparing the mutants containing each of the single substitutions with the evolved double mutant they were able to observe that the greatest contributor to fucosidase activity was the H540V mutation. Most impressive in this effort was a 700,000-fold selectivity increase for PNPF compared to that of the wild type enzyme.

This direct comparison of random and semi-rational directed evolution approaches shows that the multi-site saturation protocol is the clear winner with over 2000 times the enhancement in selectivity for β-D-fucopyranosides. However, the semi-rational approach required detailed information of the structure and mechanism of the galactosidase enzyme. It is important to note the magnitudes of the screening effort in both approaches. The random approach required visual screening of ~10,000 transformants for each of the seven rounds followed by detailed characterization of the "best" mutant in the final round. The semi-rational approach required a single round of visual characterization followed by activity assays of the 30 best mutants. The top eight performers in this round were sequenced to reveal identical mutation profiles. Detailed characterization was done on the two single and the evolved double mutant to finalize the effort. Given the relative speed of the visual screen, the major advantage with the semi-rational approach is in the time required to conduct mutagenesis. Combining the time advantage with greater enhancement in substrate selectivity, the semi-rational site saturation approach is clearly the superior technique in this direct comparison.

Epoxide hydrolase

Epoxide hydrolase from *Aspergillus niger* (ANEH) is an enzyme that catalyzes the hydrolysis of the model compound glycidyl phenyl ether to its respective diol with a

slight preference for the (*S*) product (56% ee, 33% conversion in 45 minutes, *E* = 4.6).[112] Enantioselectivity, *E*, has been described previously[113] as the ratio of specificity constants, *V*/*K*, for each of the two competing enantiomers in a reaction and can be calculated as a function of conversion and enantiomeric excess.

$$E = \frac{V_A/K_A}{V_B/K_B} = \frac{\ln[1 - c(1 + ee(P))]}{\ln[1 - c(1 - ee(P))]}$$

The presence of this epoxide can be quantified by simple assays. One such assay is the formation of a blue dye when in the presence 4-*p*-nitrobenzyl pyridine that can be quantified by absorbance at 560 nm.[114] Another is by simply using the inherent toxicity of the epoxide to select for bacterial transformants that display the ability to hydrolyze the epoxide to its less toxic diol product.[112]

Using these properties, Reetz *et al*. set about *in vitro* directed evolution of the enzyme using a single round of epPCR.[112] Their aim was to increase the selectivity factor for the glycidyl phenyl ether substrate mutagenesis followed by subsequent screening. The mutagenic ANEH bearing plasmids were transformed into an *E. coli* host. Following expression, 20,000 clones were incubated with racemic epoxide followed by staining with *p*-nitrobenzyl pyridine. The quick screen revealed approximately 400 mutants with increased hydrolase activity. Those mutants were further subject to ESI-MS to determine absolute stereoselectivity of the library population.[115] This assay requires the use of *pseudo*-enantiomers in a mixture of 1:1 of (S)-glycidyl phenyl ether and (*R*)-D_5-glycidyl phenyl ether. The ratio of the mixed diol product can then be determined by the 5 mass unit difference between the two possible products. Of the 10 most improved mutants, the best was a triple mutant (A217V,

K332E, A390E) that displayed a more than 2-fold increase in enantioselectivity (74% ee, 39% conversion in 2 minutes, *E* = 10.8).

Seeking to improve on these results, Reetz *et al*. pursued a new approach to directed evolution of ANEH.[116] This involved inspecting the ANEH active site and designating six regions using the CAST approach. Three 3-site libraries and three 2-site libraries were generated and designated A through F (193/195/196 (A), 215/217/219 (B), 329/330 (C), 349/350 (D), 317/318 (E), and 244/245/249 (F)). Bacterial growth was used as the initial quick screen followed by ESI-MS as before. After screening libraries A, B and C, they discovered that only library B contained mutants with improved enantioselectivity. Two of these mutants yielded an *E* = 14, three times greater than the wild type enzyme and slightly better than the epPCR directed evolution result (*E* = 10.8). The library B variant was further randomized iteratively against the best variant of each subsequent round in sequence B□ □, D, F, E. The best resulting mutant from this series of iterative mutations contained nine total amino acid substitutions (L215F, A217N, R219S, L249Y, T317W, T318V, M329P, L330Y, and C350V) and gave an *E* = 115±10 (95% ee and 48% conversion in 1 hour). This result is 25 times greater than the wild type enzyme and a 10-fold improvement over the previous mutagenesis effort. This result is even more remarkable considering the total screening effort was identical in size (20,000 mutants screened) to the epPCR effort. This translates into less than 10% coverage for the three position libraries and less than 70% coverage for most two position libraries. However, the resulting quality of the targeted CAST libraries is significantly greater than the random approach library,

which then translates into greater enantioselectivity improvements for the same screening effort.

In a separate and later effort, Reetz and Zheng revisited directed evolution of ANEH.[117] The focus of this approach was to evolve an ANEH variant with improved expression, a problem that had plagued previous efforts to make ANEH a commercially viable catalyst. In this effort, the ANEH gene was randomized by low frequency mutation epPCR and inserted in a plasmid connected by a linker region upstream from a gene coding for the β-galactosidase α fragment. When transformed into *E. coli* DH5α and plated onto agar supplemented with X-Gal colonies stained blue in proportion to the relative expression of the fused α fragment. Fourteen blue colonies were selected visually from a 15,000-member library. Expression was further quantified by whole cell ONPG assays and the mutants were sequenced. The best performing mutant was determined to be a single substitution variant, P221S, and was found to have retained wild type hydrolase activity (E = 5). This mutant was further cloned into a pET22b vector under the control of the T7 promoter and co-expressed with a chaperone bearing plasmid to yield a 50-fold increase in soluble protein expression. This plasmid construct was used as the starting point for site-saturated mutagenesis study to evolve enantioselectivity.

This second attempt at directed evolution of the ANEH enzyme was significantly different than the first approach. This later analysis of the ANEH active site yielded four regions containing two positions each using the CAST approach (215/219 (A), 349/350 (B), 317/318 (C), and 244/249 (D)). This selection contains six of the nine positions substituted in the previously evolved variant. To simplify screening effort, codons at

these positions were randomized with NDT degeneracy, which reduced the sampling effort for each library to 480 (five 96-well plates) in order to achieve 95% coverage. The P221S high expression variant was randomized iteratively against the best variant of each subsequent round in sequence D, A, C, B. The best resulting mutant from this series of iterative mutations contained nine total amino acid substitutions (F244C, L249F, L215F*, T317F, T318V*, and L349V) (*Mutation found in previous ISM best variant.) and gave an *E* = 160 (97% ee and 45% conversion in 1 hour). Interestingly, they also observed a 13-mer peptide insert after the library C randomization that had no negative effect on enantioselectivity and was therefore left in place for the subsequent saturation mutagenesis cycle. This insert is due to a double insert of the mutagenic primer used to construct the C library. The extent of this insertion within the library was not determined; however, it can be assumed that an insert of this size would typically have a significant negative impact on protein structure and folding, particularly if the primer insertion throws translation out of frame. Despite the relative success of this attempt, the overall library quality is questionable in this case. It is highly probable that successive primer sequence inserts in fact deactivated a large portion of the library. Given the limited screening employed (3,000 mutants) and the flawed library this mutagenesis effort was remarkably successful with respect to enantioselectivity.

Comparing these three directed evolution experiments for ANEH gives great insight into directed evolution theory. Table 3-5 summarizes these efforts and results. It can be seen that these methods can dramatically reduce screening effort while increasing coverage of all possible mutational combinations. Presupposing that the most appropriate sites are selected in a semi-rational approach it can be postulated that

this increased coverage is directly related to the magnitude of the desired and improved result.

Lipase

The lipase from *Pseudomonas aeroginosa* (PAL) is the most highly studied directed evolution target with respect to substrate acceptance and stereoselectivity.[118] PAL catalyzes the enantioselective hydrolysis of the model compound racemic *p*-nitrophenyl 2-methyldecanoate to the corresponding acid with a slight preference for the (*S*) substrate (2% ee, *E* = 1.1). Hydrolysis of the model compound can be quickly screened by the formation of *p*-nitrophenol (410 nm). Detailed study of the hydrolysis reaction can be monitored by traditional chiral GC methods after an efficient pre-screening.

In 1997 Reetz *et al*. conducted the first reported directed evolution of an enzyme to enhance its stereoselectivity.[119] In this work, the gene encoding PAL was subject to four successive rounds of low frequency mutational epPCR. In each round, the best improvement in stereoselectivity was used as the template for the following round. The result was a substantial increase in the (*S*)-selectivity of the enzyme (81% ee, 25% conversion, *E* = 11.3). The total screening effort for all four rounds was 7,600 mutants with only 29 displaying sufficient activity to justify detailed characterization.

In a continuation of this effort, Liebeton *et al*. constructed six sequential low error rate epPCR generations on the PAL enzyme[120]. Each round involved the screening of 1000-7000 clones. After the sixth generation they had successfully evolved from wild type (*E* = 1.1) to a variant with an *E* = 13.5 containing nine amino acid substitutions (S149G, S155L, V47G, F259L, L110R, Y8H, N21D, S158T, S248C). This result was very similar to the previous directed evolution effort (*E* = 11.3). The selected best

mutants from each of the first five rounds each harbored a single unique amino acid substitution. To explore further possibilities with their mutants they turned to saturation mutagenesis. NNN mutagenic primers were used to saturate each of the best variants from the first four generations at positions corresponding to their respective amino acid substitutions (in order 149, 155, 47, 259). 800 mutants were screened in each of the four degenerate libraries. Interestingly, only one mutant from the second-generation variant (S149G, S155L) showed any improvement in enantioselectivity. In this case, the leucine to phenylalanine substitution at position 155 gave a small improvement from E = 4.4 to 5.7. Following this observation the third generation variant (S149G, S155L, V47G) was randomized at position 155 by the same approach. The best mutant isolated from this library contained, in fact, the S155F mutation and exhibited an E = 20.5. With the S155F mutation now considered critical to evolve the PAL enzyme, the mutation was introduced into the wild type enzyme, and the fourth and fifth generation variants by site directed mutagenesis. Each showed an increase in E but not greater than the value of 20.5. This best variant (S149G, S155F, V47G) was subject to yet another round of epPCR where yet another improvement was observed with two new unique substitutions at positions 55 and 164 which gave a further improvement to E = 25.3.

The work described above was done without the aid of a PAL crystal structure; thus, epPCR served as a probe of enzyme protein sequence in search of influential positions. It also demonstrated that while epPCR may be efficient at finding those residues, it may not be best suited for determining the optimal amino acid substitution as evident by the position 155 saturation mutagenesis.

Reetz *et al*. further pursued an improvement on the best variant from the Liebeton *et al*. effort.[121] This involved DNA shuffling of this variant with two others (E = 3 and E = 6.5) obtained from high error rate epPCR, each containing three random amino acid substitutions. The resulting best mutant gave an *E* = 32, a new high, and contained all three mutations from the Liebeton *et al*. variant plus two new substitutions (S199G, T234S).

Continuing further, the Reetz group also pursued a strategy of combinatorial multiple-cassette mutagenesis (CMCM). In this approach, a 69 bp cassette with NNN degeneracy at positions corresponding to residues 155 and 162 was shuffled with the wild type PAL enzyme. The result was a new variant with an *E* = 34. When this cassette was shuffled with the two previous high error rate epPCR, variants the result was a new enzyme with high enantioselectivity (95% ee, 24% conversion *E* = 51) containing six amino acid substitutions (D20N, S53P, S155M, L162G, T180I, and T234S). Notably, this variant contained a methionine residue at position 155 and not the previously assumed critical phenylalanine. The overall effort for this work involved the screening of approximately 40,000 mutants. The new mutagenesis strategy resulting from this work can be described in three steps as 1) using high error epPCR to "probe" the protein sequence, 2) using CMCM to validate the perceived "hot-spots" against the wild type enzyme, and 3) extending the CMCM modification to the early "probe" results. For reference, this approach yielded a 2-fold increase[120] over the single-site saturated technique for validating the random "probe" and a 5-fold improvement[119] over the "probe" alone.

Similar to this work, Zha *et al*. set about evolving (*R*) enantioselectivity from the PAL enzyme.[122] In this approach, the wild type enzyme was subjected to three rounds of epPCR. This was followed by two successive rounds of DNA shuffling of the epPCR variants, then another round of epPCR. The result was a mutant with (*R*) enantioselectivity (*E* = 30) similar to that of the (*S*) evolved variants. The mutant contained eleven amino acid substitutions and was the product of a net 45,000 mutant screening effort. Clearly, the advent of an efficient UV-vis pre-screen is essential in these approaches to directed evolution.

In the most recent and effective approach to evolving the PAL enzyme, Reetz *et al*. applied the previously discussed CAST technique in an ISM approach.[123] They opted to designate three libraries of two positions each (A (Met16/Leu17), B (Leu159/Leu162), and C (Leu231/Val232)) and introduced NNK randomization via degenerate mutagenic primers. This level of degeneracy required screening 3000 mutants per library in order to obtain 95% coverage. After screening, libraries A and C yielded no results while library B yielded several, the best being a single mutant (L162N) which gave an *E* = 8 for (*S*) *p*-nitrophenyl 2-methyldecanoate.

This mutant was used as the starting point for a second round of mutagenesis. In this round, the poor results for the A and C libraries combined with a desire to reduce the screening effort prompted the use of DNT degeneracy for the two-position randomization of libraries A and C. This amino acid subset was chosen because leucine is notably absent from the 11 possible amino acids (Ala, Asn, Asp, Cys, Gly, Ile, Phe, Ser, Thr, Tyr, and Val) and the screening effort could be reduced to 430 mutants per library to obtain 95% coverage. When randomized in this manner against the A

grouping, the best result obtained was an unprecedented value of *E* = 594 for the (*S)* substrate. This variant contained two new mutations (M16A, L17F). Given that randomization for the C grouping gave no improvements, no further iterative cycles were performed. The total screening effort for this remarkable result was 10,000 mutants. The substrate scope of PAL has also been explored in detail the CAST technique for site selection[95] and ISM[118] to achieve dramatic improvements for a wide range of *p*-nitrophenyl esters.

Based on all of these precedents, the ISM technique is the most straightforward and concise approach to altering the stereoselectivity of our target enzyme, OYE 2.6. The focus of our study involves the uncovering of an alternative, flipped binding mode of a *pro*-chiral alkene in order to produce the opposite enantiomer. Precedence for using ISM for this purpose has been set with the alkene reductase YqjM.

ISM of an Alkene Reductase, YqjM

The alkene reductase and Old Yellow Enzyme homolog YqjM from *Bacillus subtilis* catalyzes the asymmetric reduction of prochiral α,β-unsaturated ketones in a reaction mechanism analogous to that of OYE 1 discussed in Chapter 1. The enzyme poorly reduces 3-methyl 2-cyclohexen-1-one (3% conversion) with a slight preference for the (*R*)-product (79% ee). Bougioukou *et al*. used the ISM approach to improve both activity (conversion) and steroselectivity of this 3-substituted model compound.[124] Given that the only feasible method to observe this reaction was through chiral GC analysis, they established a modified screening protocol. Mutants arrayed on a 96-well plate were pooled by column and screened for conversion by a short, achiral GC method. Pools that formed product quantities above the wild type threshold were subsequently deconvoluted to identify individual mutants of interest. Positive results at

this stage were then further subject to chiral GC analysis to assess enantioselectivity. To further reduce screening effort, individual amino acids were selected for randomization by the CAST approach. By not grouping CAST selections, individual sites could be randomized and sampled to 95% confidence with only 94 mutants per site. Using a crystal structure with a bound inhibitor (PDB 1Z42), twenty residues were selected with varying potentials to affect catalysis based on their distance from the β-carbon of the substrate analog. By randomizing each site individually, the first round screening effort (not including the pooling strategy or additional oversampling) was estimated at 2000 mutants or approximately 7 days of continuous non-chiral GC operation for identification of active variants. Conversely, construction of ten 2-site randomized libraries would theoretically increase the first round screening effort to 30,000 mutants in order to achieve similar coverage. The assumption was then made that positive effects would be apparent at least to a small degree with single site randomization and that combinatorial effects could then be exploited in subsequent rounds of mutagenesis.

After first round screening, 35 variants were identified as having increased (*R*) or (*S*) enantiopreference and increased conversion. Each was given an impact score derived from the product of the % conversion and % ee. In this manner, each property was given equal weight in ranking the variants. Several variants were chosen as starting points for the second round of single site randomization ((*R*) selective I69T and C26W; (*S*) selective C26D, C26G, T70H and A60C). Non-systematic second round randomizations at these positions led to higher overall impact scores. The net result was several mutants displaying greater than 95% ee and greater than 50% conversion

for both (*R*) and (*S*) reductions of the model compound ((*R*) selective C26W(A104Y and A104F), C26D(I69T), C26D(A104Y, A104F, and A104W); (*S*) selective C26G(A60C, A60I and A60V). Interestingly, the C26D mutant, that gave (*S*)-selectivity in round one, yielded the best variants for (*R*)-selectivity in round two. This observation further emphasizes the importance of combinatorial effects and the non-linearity of a directed evolution effort.

The evolved variants were further screened against 3-substituted ethyl, isopropyl, *n*-butyl, and methyl carboxylate derivatives of cyclohexenone as well as 3-substituted methyl and methyl carboxylate cyclopentenone. Several evolved variants displayed specificity for the additional substrates that was not present in the wild type enzyme. In this case, the pool of evolved variants provided an effective source library for screening a family of related substrates.

Conclusions

Our goal is to evolve OYE 2.6 to induce the opposite binding mode (pro-*R*) for our Baylis-Hillman derived substrates discussed in Chapter 1. Precedence for successful evolution of an Old Yellow Enzyme homolog is demonstrated by the efforts of Bougioukou *et al*. Key to that work was the selection of a directed evolution strategy that could be conducted with an abbreviated screening effort. The use of ISM with restricted codon usage enabled the construction of a focused high-quality library that was manageable in size for a GC based assay system. Several examples show that combinatorial site saturated libraries as used in ISM methodology provide superior directed evolution results when compared side by side with more random mutagenesis approaches (e.g. DNA shuffling, epPCR). Key to using a site saturated approach such as ISM is the generation of high quality mutant libraries. In this case quality depends

selection of the right sites and completeness of randomization while eliminating redundant and erroneous gene constructs. Next chapter describes our efforts for constructing mutant libraries for the Iterative Saturation Mutagenesis of OYE 2.6.

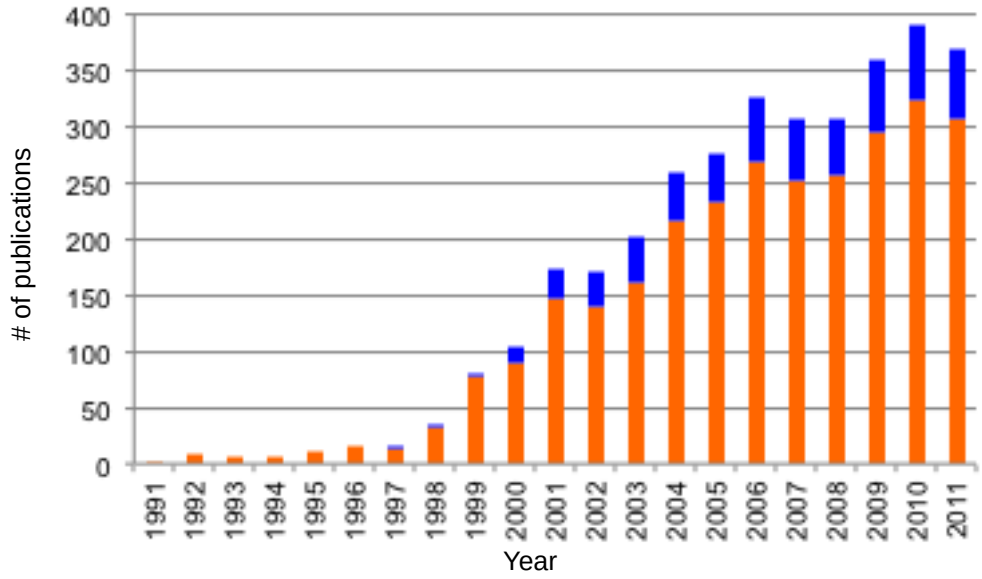


Figure 3-1. Numbers of publications dealing with *in vitro* directed evolution

The number of publications per year in the field of directed evolution (orange) and fraction of those dealing with enantioselectivity improvement (blue) were retrieved from 'ISI web of knowledge' (http://apps.webofknowledge.com). The search term "directed evolution" was used to retrieve the number of publications on directed evolution, and the additional search criteria "enantio*" or "stereo*" were used to identify publications dealing with enantioselectivity and stereoselectivity.

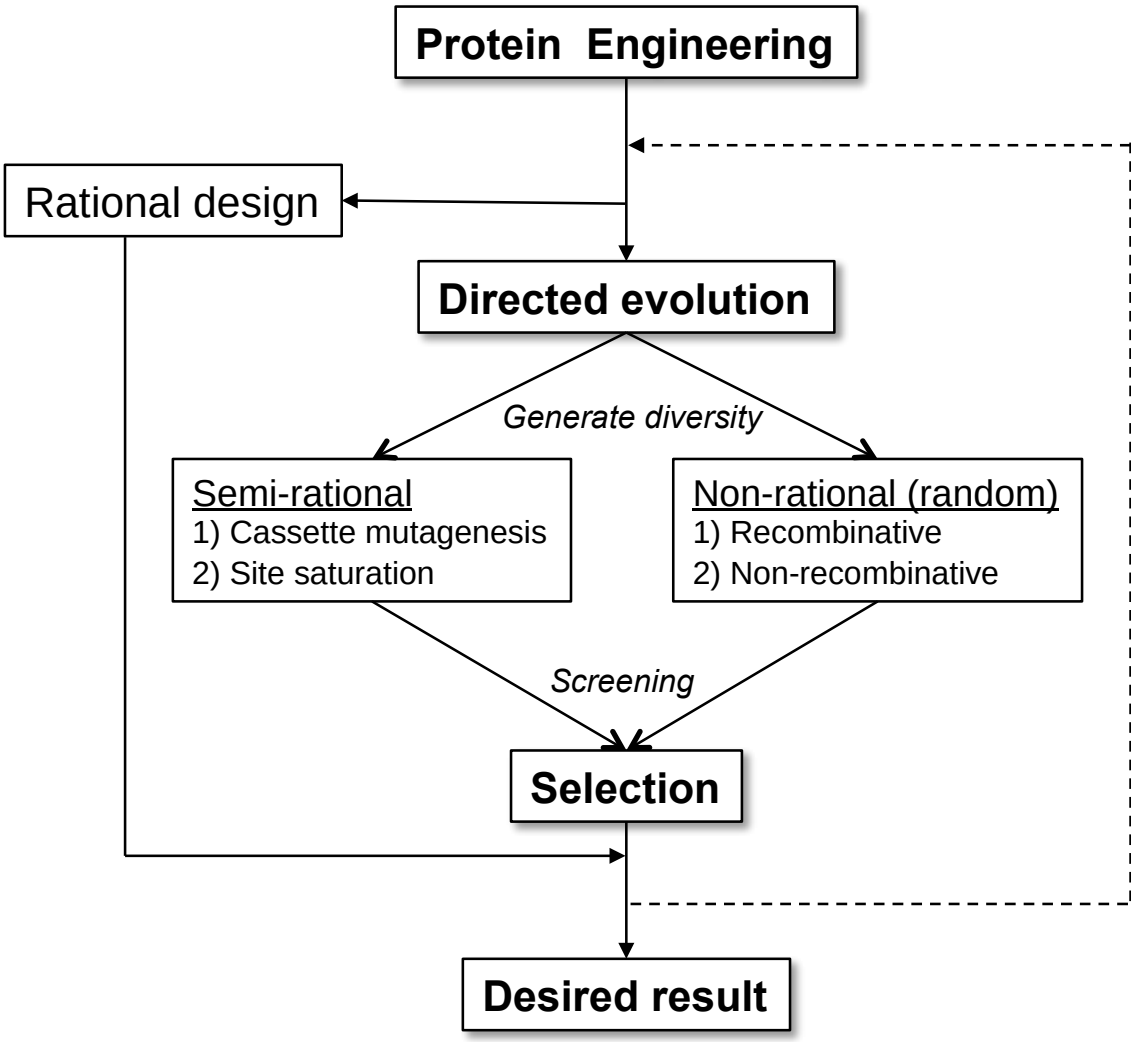


Figure 3-2. The *in vitro* protein engineering process

Table 3-1. International Union of Biochemistry and Molecular Biology (IUBMB) nucleotide nomenclature

Letter code	Name	Bases	Complement
A	Adenine	A	T
C	Cytosine	C	G
G	Guanine	G	C
T	Thymine	T	A
R	puRine	A G	Y
Y	pYrimidine	C T	R
S	Strong (3 H-bonds)	G C	S
W	Weak (2 H-bonds)	A T	W
K	Keto	T G	M
M	aMino	A C	K
B	not A	C G T	V
D	not C	A G T	H
H	not G	A C T	D
V	not T	A C G	B
N	Unknown	A C G T	N

Table 3-2. Coding description of selected degenerate codons

Degeneracy	# of codons	# of amino acids	Distribution	Properties
NNN	64	20	all + 3 stop	
NNK	32	20	all + 1 stop	
NDT	12	12	CDFGHILNRSVY	mixed
DYK	12	8	AFILMSTV	no charged
RRK	8	7	DEGKNRS	all polar

Codon usage determined using CASTER v. 2.0 software available at for download at http://www.kofo.mpg.de/en/research/organic-synthesis.

Table 3-3. Oversampling for 95% coverage as a function of degeneracy and number of simultaneously randomized sites.

# of codons	NDT	NNK	NNN
1	34	94	190
2	430	3,066	12,269
3	5,175	98,163	78,5312
4	62,118	3,141,251	50,260,046

Oversampling calculated using CASTER v. 2.0 software available at for download at http://www.kofo.mpg.de/en/research/organic-synthesis.

Table 3-4. NNK degenerate codon usage.

Amino Acid	Letter code	Codons	# Codons	% Occurrence
Alanine	A	GCG, GCT	2	6%
Cysteine	C	TGT	1	3%
Aspartate	D	GAT	1	3%
Glutamate	E	GAG	1	3%
Phenylalanine	F	TTT	1	3%
Glycine	G	GGG, GGT	2	6%
Histidine	H	CAT	1	3%
Isoleucine	I	ATT	1	3%
Lysine	K	AAG	1	3%
Leucine	L	CTG, CTT, TTG	3	9%
Methionine	M	ATG	1	3%
Asparagine	N	AAT	1	3%
Proline	P	CCG, CCT	2	6%
Glutamine	Q	CAG	1	3%
Arginine	R	AGG, CGG, CGT	3	9%
Serine	S	AGT, TCG, TCT	3	9%
Threonine	T	ACG, ACT	2	6%
Valine	V	GTG, GTT	2	6%
Tryptophan	W	TGG	1	3%
Tyrosine	Y	TAT	1	3%
Stop	none	TAG	1	3%

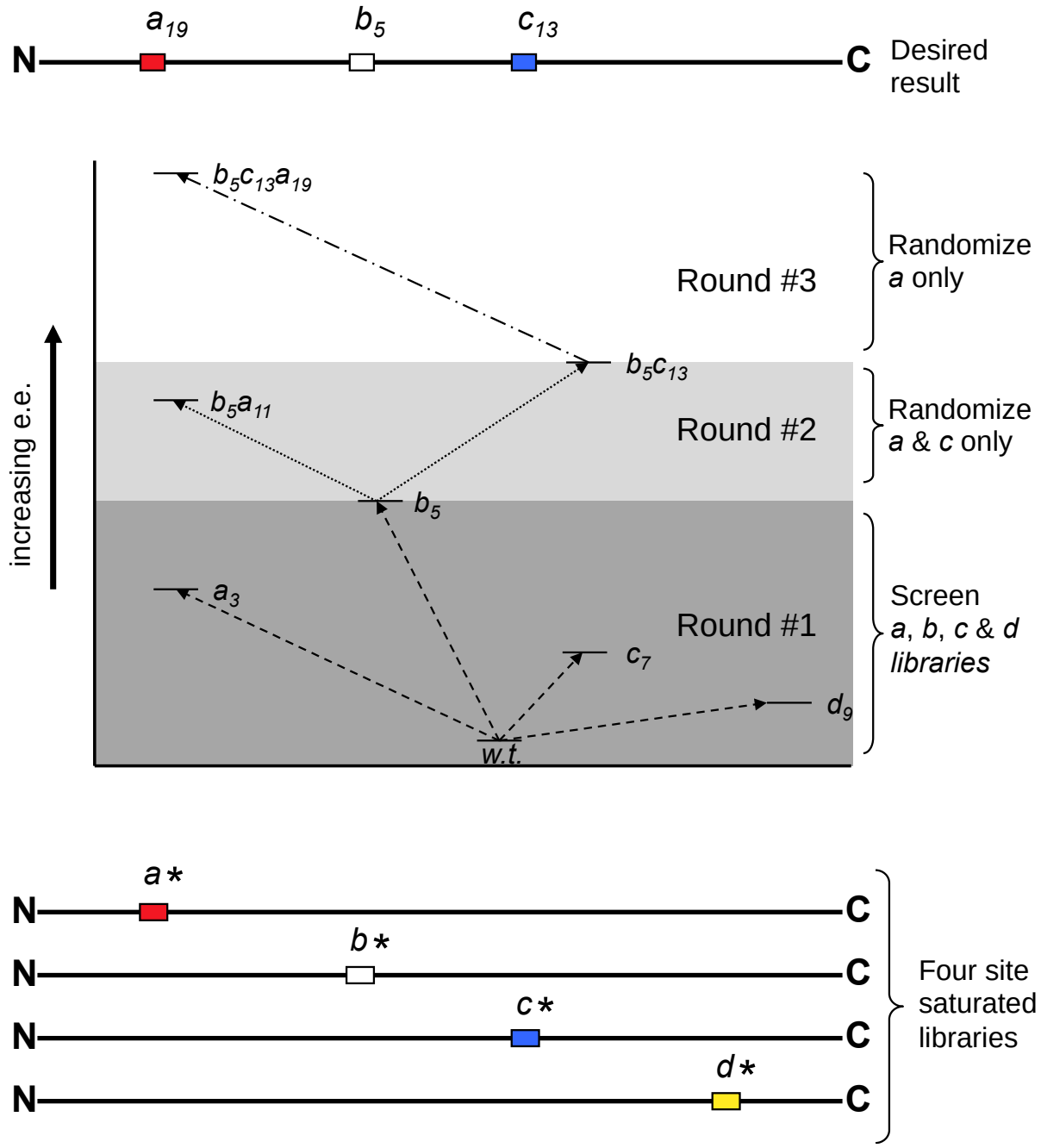


Figure 3-3. Iterative Saturation Mutagenesis methodology for the evolution of increased stereoselectivity.

In this example libraries designated by *a*, *b*, *c*, and *d* correspond to selected positions in the peptide sequence. In practice each library may consist of multiple sites randomized simultaneously by the CAST approach. Asterisk (*) indicates saturation with desired degeneracy (i.e. NNK). Subscript numbers correspond specific amino acid substitutions (1-19 possible). The best mutant in each round serves as the starting template for the following round of mutations.

Table 3-5. Summary of ANEH directed evolution efforts.

Methodology		Screening effort	Coverage	Result (*E*)
wild type		-	-	4.6
epPCR	Single round	20,000	-	10.8
ISM	NNK 2-3 a.a./library	20,000	5-70%	115
ISM	NDT 2 a.a./library	3,000	95%	160

CHAPTER 4
CREATING SITE SPECIFIC SATURATED LIBRARIES

Introduction

This chapter describes the studies of how best to create degenerate libraries of OYE 2.6 variants used in directed evolution experiments. The objectives and efforts described in this chapter are summarized in Figure 4-1. The screening of these libraries is a separate and distinct effort described in detail in the following chapter.

Experimental Strategy

Site Selection

At the onset of this work there was no solved crystal structure for the OYE 2.6 enzyme. We therefore chose to establish candidate residues for OYE 2.6 CASTing by selecting active site residues in OYE 1, then identifying the OYE 2.6 counterparts by comparing amino acid sequences. Protein Data Bank accession number 1OYB is a 2.0 Å resolution crystal structure for OYE1 with a bound inhibitor, *p*-hydroxybenzaldehyde.[30] The phenol moiety in the bound ligand is oriented toward and most likely engaged in hydrogen bonding with the side chains of His191 and Asn194. This binding mode is analogous to that of our carbonyl-containing substrates.[14] For the purposes of CASTing, we assume that bound *p*-hydroxybenzaldehyde is a good approximation of the size and orientation of our proposed substrates **1**, **2**, and **3**.

Active site residues in OYE 1 were selected if their side chains met the following three criteria:

- Located above the plane of the catalytic FMN.
- Oriented toward the center of the active site pocket.
- Capable of interacting directly with the substrate by both position and distance.

Once residues were selected by visual inspection, the distances between the β-carbon of the residue side chain and the nearest carbon of the bound inhibitor were measured (Figure 4-2) and recorded (Table 4-1). Those residues with distances of 5 Å or less were determined to have a high probability of interacting with our substrates while between 5 and 10 Å were designated as possibly interacting with our target substrates. Figure 4-3 shows a schematic array of these residues in the OYE 1 active site. These residues were matched in an alignment with OYE 2.6 (Figure 4-4). From this alignment, we determined that the best candidate residues for CASTing in OYE 2.6 were Thr35, Phe37, Ile113, His188, His191, Tyr193, Phe247, Asn293, Val294, Phe373, and Tyr 374. Residues Leu115 and Gln248 were not selected based on their partial masking by adjacent active site residues. Figure 4-6 shows a schematic array of the OYE 2.6 active site residues.

During the construction of these randomized libraries, parallel efforts in our group eventually yielded crystal structures of OYE 2.6.[125] One structure contained a bound inhibitor, *p*-chlorophenol, that binds above the FMN in an analogous manner to the phenolic ligand in the 1OYB structure. Figure 4-5 depicts a ribbon diagram of OYE 1 (green) and OYE 2.6 (blue) overlaid with the FMN as the focal point of the alignment. From this depiction we can see that there is a large degree of structural similarity between the secondary and tertiary structures of the two enzymes. Additionally, the phenolic ligands bind with very close alignment. One region that did show significant deviation, however, is the loop region centered on Pro295 of OYE 1. Sequence alignment revealed that OYE 2.6 lacks a corresponding proline residue and the loop

itself is several amino acids shorter in length. Based on these observations, Gly292 of OYE 2.6 was added to our list of CASTing libraries.

Finally, we also observed that OYE 2.6 Ala68 met the criteria for selection. This residue had been overlooked in our original analysis based on OYE 1 because the corresponding residue in OYE 1 is a glycine that lacks a β-carbon and was interpreted as backbone in our earlier measurements. We had also overlooked Tyr78 in the OYE 2.6 active site due to its orientation away from the bound ligand in the OYE1 crystal structure. This residue displays a significant shift in position toward the bound ligand in OYE 2.6. Both of these residues were added to our list of libraries needed for complete coverage of the OYE 2.6 active site.

Library Creation

Our goal was to construct libraries using the protocol described by Bougioukou *et al.* This approach was based on the QuikChange® (Stratagene) protocol for site directed mutagenesis using degenerate primers. Modifications to this methodology are discussed in the results section of this chapter. The template DNA containing the OYE2.6-GST fusion protein was constructed in our lab previously[18] and further mutated to remove a second NdeI restriction site that would interfere with future cloning steps. The resulting plasmid, pBS2 (Figure 4-7), was further purified from a *dam*$^+$ strain so as to facilitate its digestion by DpnI endonuclease. Mutagenic primers were designed with NNK degeneracy to provide the potential for 32 codons encoding all 20 amino acids. As previously noted, this library size required screening 94 samples to obtain a 95% probability of identifying at least one example of each codon. This modest effort fits within our constraints for minimized downstream screening.

Assessing Degeneracy

To the best of our knowledge, there is no published method for determining the quality of a degenerate library from pooled sequencing data. This must be established prior to library screening to ensure that all possible clones actually are present within the collection. Kille developed an approach to estimate the relative amount of parent template in a pooled sample and used it to determine the degree of oversampling required to achieve an adequate sample size.[126] To analyze the quality of our attempts in a quantitative manner we propose a method for calculating a value for the quality of degeneracy, Q. This value is determined by calculating the relative percentages of peak amplitudes for each base at each degenerate position from capillary sequencing data. Then at each position in the degenerate codon the absolute value of the deviation from the "perfect" value (25% for N and 50% for K) is subtracted from the "perfect" value for each base and is then totaled at each position for all bases. The result of this treatment yields a Q value of 1.0 for "perfect" degeneracy at that position. The weighted average of these values for all three positions in the codon can be used to calculate the quality of degeneracy for the codon, Q. When calculated from fluorescence sequencing chromatography (chromat) data from pooled plasmid DNA, the value is referred to as Q_{pooled}, which is an estimate of degeneracy. Figure 4-8 shows a sample calculation of Q_{pooled}. When determined from a sample population of individually sequenced mutants the value is referred to as Q_{codon} and is an actual measure of degeneracy. The equations for determining Q are as follows:

$$Q_{pooled} = (0.4 \times Q_{N1}) + (0.4 \times Q_{N2}) + (0.2 \times Q_K)$$

where

$$Q_N = \Delta_A + \Delta_C + \Delta_G + \Delta_T$$

and

$Q_K = \Delta_G + \Delta_T$.

At any N degenerate position for any base, B,

$\Delta_B = 0.25 - |0.25 - X_B|$

and at any K degenerate position for any base, B,

$\Delta_B = 0.50 - |0.50 - X_B|$

where X is a base's fractional peak amplitude compared to the total of peak amplitudes at that position.

This approach places a quantitative value on what was originally a qualitative observation when viewing sequencing chromatograph data, i.e. the visual assessment that all four peaks for an "N" base are identical (or not) in height. It assumes that *all* peaks and *only* those peaks expected (A, C, G, and T for N; G and T for K) are present in the sequencing chromat. Attempts that display what are obviously wild type or other disproportionate bias should be qualitatively assessed and not subject to this mathematical assessment. Using the Q value facilitates discrimination between experiments that display varying degrees of degeneracy.

Experimental Procedures

Construction of Pooled Degenerate Plasmid

Saturation mutagenesis libraries were prepared by a modification of the methods reported by Zheng et al.[127] Each PCR reaction (total volume 100 µL) containing 5 × Phusion® HF Buffer (20 µL), 1 ng of pBS2 (Figure 4-7 and Figure B-1), forward and reverse NNK degenerate primers (0.5 µM each) (Table C-1), dNTPs (200 µM each), and Phusion® Hot Start II High-Fidelity DNA Polymerase (1 U) was subjected to an initial denaturation step of 98°C (30 s) followed by 25 cycles of 98°C (10 s) followed by a range of extension temperatures from 62 to 72°C (4 min) and completed by a final

incubation at 72°C (7 min). Amplicons were purified by DNA spin columns, digested with DpnI at 37°C (10 U for 4 h followed by an additional 10 U for 4 h) to remove adenine methylated template DNA and then purified by an additional DNA spin column. Aliquots (5 µL) were used to transform ElectoTen Blue (Stratagene) electocompetent cells (75 µL) by electroporation. SOC medium was added (600 µL), then samples were incubated for 1 h at 37°C prior to selection on LB media plates supplemented with 200 mg/L ampicillin. After overnight incubation, reactions yielding ≥300 colonies (3 plates combined) were pooled with the aid of LB media and pelleted. Plasmid DNA was then purified (spin columns) and analyzed by DNA sequencing to identify samples with the highest degree of pooled degeneracy at the desired position.

Degenerate Library Creation

After confirming adequate degeneracy of each pooled library, 1 ng of each was used to transform electrocompetent *E. coli* BL21 Gold (DE3) (Stratagene, 40 µL). SOC medium was added (600 µL), then samples were incubated for 1 h at 37°C prior to selection on LB medium supplemented with ampicillin. After incubation at 37°C for 16 hours, individual colonies (95) were selected and used to inoculate wells containing 600 µL LB medium supplemented with ampicillin. The 96^{th} position in the plate was inoculated with a colony containing a plasmid encoding the wild-type protein. The resulting 2 mL deep-well plate was then shaken (250 rpm) at 37°C for 6 hours. Sterile glycerol was then added to bring each culture to a final concentration of 15% wt/vol. These plates were then replicated into multiple 150 µL volume copies and stored at -80°C until needed for screening (Chapter 5). Eleven of these libraries were selected for

whole plate sequencing. Copies of these plates were frozen at -20°C and sent for RCA protocol sequencing (UF ICBR) using appropriate primers.

Results and Discussion

Developing a Mutagenesis Method

A significant amount of time and effort was put into developing an adequate saturation mutagenesis method. Our primary focus at the onset of this project was to develop a library that was truly degenerate so that we had a 95% chance of obtaining all 32 possible codons in 95 randomly selected transformants. Practically speaking, an adequate library is one that yields all 20 amino acids. Eventually, we discovered that with our self imposed sample size restrictions and the limits of PCR based methodologies our definition of adequate would have to be further modified. Issues requiring attention during method development included:

- Transformation efficiency of PCR products
- Extraneous PCR derived errors
- Frequency of wild type appearance
- Distribution of degeneracy

Table 4-2 and Figure 4-9 outline five distinct approaches that we took in the search to resolve these issues. It is worth noting that methodology changes were attempted in a semi-sequential manner with the goal of finding a workable method. Parallel assessment and optimization of the methods was therefore not performed and the discussion here does not serve to compare the methods to each other.

Transformation efficiency

One bottleneck shared by all library construction approaches is the transformation efficiency of mutagenized PCR products. We accepted the requirement that we obtain 300+ colonies from our PCR product transformants as a sample large enough to assess

degeneracy when pooled and sequenced.[124] During our investigations we frequently failed to achieve this number of transformants, which caused us to discard the attempt due to presumed inadequate degeneracy and actual insufficient quantity of DNA for sequencing.

Several factors affect transformation efficiency. First is the concentration of mutagenic plasmid. Approaches described by Bougioukou[124] and Edelheit[128] are linear amplification methods where the product of the preceding PCR cycle does not function as a template for subsequent cycles. This results in a non-exponential amplification and provides lower DNA yields from the PCR process. We increased the volumes of the PCR reactions to 100 µL in an effort to achieve a greater number of transformants with these methods. However, we ultimately moved on to the other methods listed in Table 4-2 which result in an exponential amplification, higher DNA yields and, in theory, more mutagenic transformants.

The quality of the electrocompetent cells used to transform PCR products also played a significant role in yield. In general, we found that electrocompetent cells prepared fresh, not frozen, and used the same day for electroporation resulted in the greatest transformation efficiencies. This observation held true for both *E. coli* DH5α and JM109 cell lines.

The most significant improvement in transformation efficiencies came with the use of *E. coli* ElectroTen Blue (Stratagene) cells. ElectroTen Blue cells were developed and are marketed as a high electroporation efficiency (*Hee*) phenotype. According to the manufacturer these cells better survive the electroporation event that in turn results in a larger mutant library. Using them afforded us up to two orders of magnitude greater

transformants. The resulting ≥200 colonies per plate allowed us to achieve a larger sample size with fewer plates. Practically speaking, we were able to reduce the number electroporations per library from two or three to one and the resulting number of plates pooled from six or nine to three.

Primer concatemers

The next issue was the elimination of PCR-derived errors. Specifically, these were manifest in the form of multiple repeating inserts of our mutagenic primers into the plasmid sequence.

We originally set out to reproduce the mutagenesis method described by Bougioukou.[124] After dealing with transformation efficiencies described in the previous section we obtained a pooled plasmid library for OYE 2.6 I113X that appeared significantly degenerate in the sequencing chromatogram. Using the previously described assessment, the degeneracy estimate for that library scored a $Q_{pooled} = 0.84$. The library was then transformed into an *E. coli* BL21 (DE3) expression strain and 95 random colonies, along with one wild type control, were arrayed in a 96-well plate format and sequenced by RCA. While the quality of sequencing results was less than desired (66 of 96 wells sequenced) the results were surprising.

The degeneracy of this library was poor with only 15 of 20 amino acids present in the sample. The presence of wild type enzyme was slightly higher than expected in a truly degenerate population with 9% isoleucine present as opposed to 3% in a model population. At first glance, we attributed the poor degeneracy to the smaller sequencing sample and the slightly higher parent template carry over. Upon closer inspection, we observed that 23% of the sequenced population contained multiple repeating primer

inserts. The number of inserts ranged from 2 up to 9 (!). More troubling was the fact that these multiple inserts were not apparent in DNA sequencing data from the pooled population. To determine whether the problem was site-specific or a more general flaw in the methodology, we analyzed two additional libraries. Pooled plasmid DNA sequencing data from both the F37X and V294X libraries revealed multiple primer inserts, leading us to conclude that the problem was inherent in the PCR method.

How multiple primers are inserted is unknown. Our initial approach, to solve this problem was to adapt the Single-Primer Reactions IN Parallel (SPRINP) protocol to our methodology.[128] In this approach the PCR reaction is divided in two and each portion is combined with either the forward or reverse mutagenic primers. During the PCR program only one strand is amplified in each reaction. Upon completion the two half reactions are combined and subject to a slow annealing step where, in theory, the two complementary degenerate products combine. The objective of this modification is to bypass the multiple insertion mechanism. This method did eliminate primer concatemers; however, the resulting pooled libraries possessed poor degeneracy according to the pooled DNA sequencing data. Moreover, this protocol yielded a high degree of parent template carry-through. We therefore expanded our search for an adequate mutagenesis method.

Obtaining degeneracy

Since library construction was hindered by low DNA quantities obtained from linear amplification based protocols, we examined the viability of other methods involving exponential amplification. Three approaches were evaluated: Phusion® Site-directed Mutagenesis Kit protocol, Liu and Naismith[129], and Zheng *et al*.[127] The results of these trials are summarized in Table 4-2.

The Phusion® Site-directed Mutagenesis Kit protocol is a significant deviation from previous attempts. In this strategy, primers for both template strands are designed to be completely offset from one another but directly adjacent. Melting temperature values were calculated according to the manufacture's protocol. In addition, the primers were phosphorylated at their 5' termini. Following the PCR amplification and DpnI digestion steps, the mutagenized DNA was treated with ligase prior to transformation. The resulting circular DNA has a higher transformation efficiency that was also desirable in our case. Unfortunately, results using this approach failed to yield qualitatively acceptable degeneracy. In general, pooled sequencing results appeared to be a combination of parent template and a "GGG" codon (Figure 4-10). This was observed consistently even after a degree of annealing temperature optimization therefore the investigation was abandoned in search of a more appropriate strategy.

The methods described by Liu and Naismith and Zheng *et al*. use partially overlapping primers in which the degenerate codon is positioned in the center of the complementary segment. The methods vary in the length of both the overlapping and non-overlapping segments. Ultimately, the method described by Zheng *et al*. gave us the best qualitative assessment of degeneracy, lacked an obvious favoring of parent template or other bias (i.e. GGG), and gave no indication of primer concatemers (in either pooler or individual sequencing experiments). Additionally, their approach was designed and validated for use in the generation of site-saturated libraries. Given this we continued to construct all first round libraries using this method, which is described in detail in the experimental section of this Chapter.

It is interesting to note that when we conducted the Zheng *et al*. approach with 5'-phosphorylated primers followed by ligation, as done in the Phusion® kit, the resulting pooled library overwhelmingly displayed a double primer insert. This implies that the duplex formed by the amplification product is further extended by the polymerase to form blunt-ended product. Blunt-ended product will transform at a much lower efficiency than the sticky-ended (and potentially circular) product we see in our normal protocol. Ligation, however, shows that a significant portion of the population is in fact blunt-ended. This leads us to the conclusion that "over-activity" of the polymerase is most likely responsible for the PCR derived errors we have observed.

Assessing Degeneracy

In order to go forward into the screening phase of this project with a sample size of approximately 100 mutants, we needed to ensure that the degree of degeneracy was as high as possible while the occurrence of wild type was minimal. The pooled plasmid sequencing data rarely indicated perfect degeneracy in our library plasmid mixtures. Degeneracy was therefore assessed at two stages: in the initial population of pooled plasmids and at the individual clone level following transformation into the overexpression strain.

After transforming the pooled degenerate plasmid library into our expression strain, 95 randomly selected colonies were arrayed and cultured on a 96-well plate along with a wild type control. These bacterial cultures were sequenced in plate format using RCA methodology. From these sequencing results a value for Q_{codon} can be calculated based on the appearance of each base in the NNK codon using the same treatment previously described using chromatograph peak amplitudes. We were able to use these values in combination to assess the quality of our degeneracy and the

efficacy of the Q value. The Q value now serves as both an estimate and a threshold for the construction of a pooled degenerate library.

Library Quality

We sequenced 95 mutants along with a wild-type control from eleven first round libraries in order to determine the effectiveness of our mutagenesis methodology. The results of these sequencing attempts and the Q values associated with each library are summarized in Table 4-4. Among multiple attempts for each position library, those with the highest Q_{pooled} values were selected for transformation and sequencing of arrayed mutants.

As we had hoped, there was a good correlation between Q_{pooled} (pooled plasmids from the initial transformation) and Q_{codon} from sequencing the resulting individual clones derived from the pooled plasmids. Averaging the results of all libraries yields a Q value of 0.71 for both Q_{pooled} and Q_{codon}. However, for each library average deviation between Q_{pooled} and Q_{codon} is 0.07.

On average libraries with a $Q_{pooled} = 0.71 \pm 0.07$ yielded 27.4 out of 32 possible codons and 17.5 out of 20 possible amino acids. The highest number of codons observed across our sample was 31 and the lowest was 22. Only once did we obtain all 20 possible amino acids in a library. Our lowest level was 15. The actual amino acid distribution of these first eleven libraries is depicted in Figures 4-11 thru 4-13. The actual amino acid saturation profile obtained generally correlates but is not an absolute function of the Q_{pooled} values.

With NNK degenerate codon usage, we expected that amino acids with 1, 2 and 3 codons would represent 3%, 6%, and 9%, respectively, of the population sampled. Table 4-3 lists the actual amino acid bias we observe using our methodology. The most

overrepresented amino acids are Gly and Pro, each with only 2 codons but over 10% occurrence across our libraries. Since glycine uses G-G-G/T codons and proline uses C-C-G/T it is likely that the overabundance of these amino acids is sequence specific and related to hydrogen bonding between base pairs. The most underrepresented amino acids are Cys, Ile, and Phe, all of which only have one codon in NNK degenerate libraries.

There are very few published examples with which to compare our analysis of library quality. We originally carried out this analysis to assess the actual content of libraries created using our mutagenesis methodology; ultimately, however we would also like to know whether these results are also typical of similar efforts. Unfortunately, few researchers have published the chromatographic sequence data required for comparison.

Bougioukou *et al*. published pie charts showing the base compositions of degenerate libraries similar to our presentation in Appendix A.[124] From this, we can qualitatively conclude that our degeneracy is similar, or in most cases, better than that obtained by their methodology. We originally employed this methodology. As described above, the degeneracy of pooled plasmid DNA appeared adequate; unfortunately, sequencing individual library members revealed that a large fraction conatained primer concatemers. Based on our results, we suspect that some portion of the libraries reported earlier may also contain similar concatemer events that were undetected or underestimated during attempts at library validation.

Zheng *et al*. reported chromatographic sequencing data from their library methodology[127] for both the PCR product and DNA from transformed and pooled

libraries. By our analysis, their PCR product gives a Q_{pooled} value of 0.78, which is acceptable by our analysis. However, they report pooled sequencing data with a Q_{pooled} = 0.48 that is also visibly high in parent template sequence. They attribute this to inadequate DpnI digestion. This is a plausible explanation given the high transformation efficiency of the circular template plasmid and the fact that they conducted only a single DpnI digestion as opposed to the two sequential digestions that we employed.

We also purified our PCR products prior to DpnI digestion so as to conduct a buffer exchange for the digestion as prescribed by Zheng *et al*. Sanchis *et al*. performed similar analyses in developing an improved saturation mutagenesis method[93] however, they imply that they did not conduct purification at this step. Even though they conducted direct comparisons to the Zheng *et al*. approach and others[91,92] the data they report is significantly high in parent template. Q_{pooled} for their data is consistently less than 0.50. While they do conduct two sequential DpnI digestions it appears that their digestions are poorly effective in removing parent template.

Conclusions

In summary, we have successfully used OYE 1 as a model for identifying active site residues in OYE 2.6. After obtaining structural data for OYE 2.6, only minor changes were made to our original group of targeted residues. These sites were then further exploited using saturation mutagenesis.

We developed a successful approach for constructing site-saturated mutagenesis libraries. For optimal success our approach requires degenerate primer design that facilitates exponential amplification of template plasmid, robust elimination of template plasmid by DpnI digestion, and subsequent transformation at high efficiency. These

measures ensure elimination of PCR derived errors and minimize the presence of parent template while maximizing degeneracy of the target codon.

Finally, we have also developed a simple measure of degeneracy based on data provided from standard fluorescence-based DNA sequencing data. On average, our methodology yields an estimate of degeneracy quality, $Q_{pooled} = 0.71 \pm 0.07$. For a sample library of 95 members this translates into an average of 17.5 of 20 possible amino acids. This estimate serves as a guide in selecting libraries with sufficient degeneracy for further screening endeavors. In order to achieve all possible outcomes the sample library size would need to be increased to compensate for the carry over of wild type from the parent template and any inherent sequence specific bias as a result of the PCR methodology.

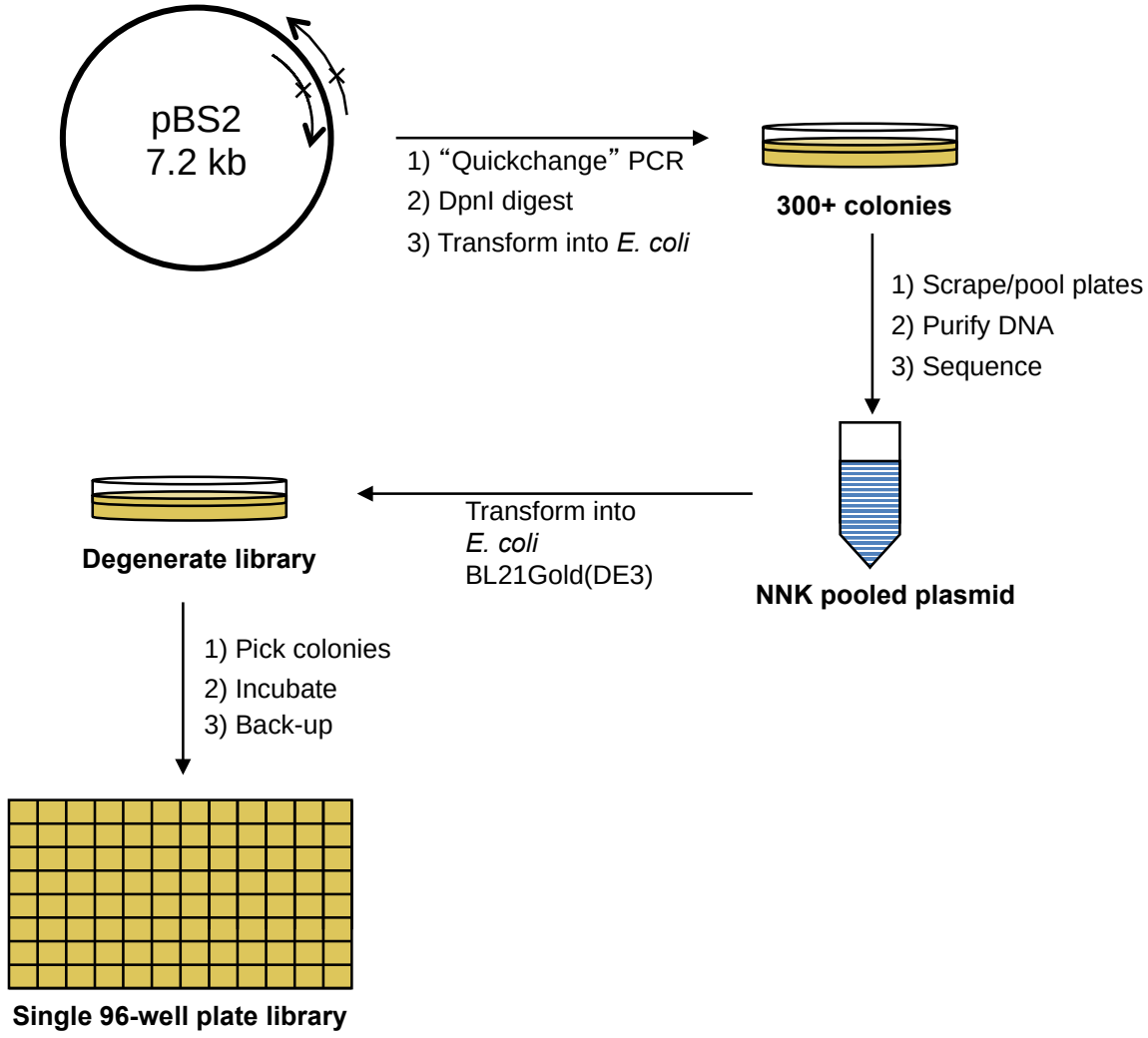


Figure 4-1. Summary of Chapter 4 experimental objectives and efforts

Table 4-1. OYE active site measurements.

OYE1 Residue	Å	Ref. carbon #	Å	OYE2.6 Residue
N194	4.1	1	4.4	H191
Y196	6.7	1	7.0	Y193
G72*	6.4	2	4.8	A68
W116	7.0	2	6.8	I113
L118	8.9	2	9.1	L115
H191	7.3	2	7.5	H188
T37	4.0	3	3.9	T35
M39	9.4	3	9.7	F37
Y82	9.0	3	8.6	Y78
F374	9.1	4	8.9	F373
Y375	9.1	4	8.6	Y374
		5	4.5	G292
N294**	9.4	5	4.5	N293
P295**	5.1	5	9.0	V294
F249	6.8	6	6.6	F247
N250	8.3	6	8.7	Q248

Distances are measured in angstroms from the β-carbon of the indicated residue to the nearest ring carbon of the bound ligand. Phenolic carbon is designated as 1 and proceeds clockwise as viewed from above the bound FMN cofactor.
*measurement taken from α-carbon
**region of poor sequence alignment with OYE2.6

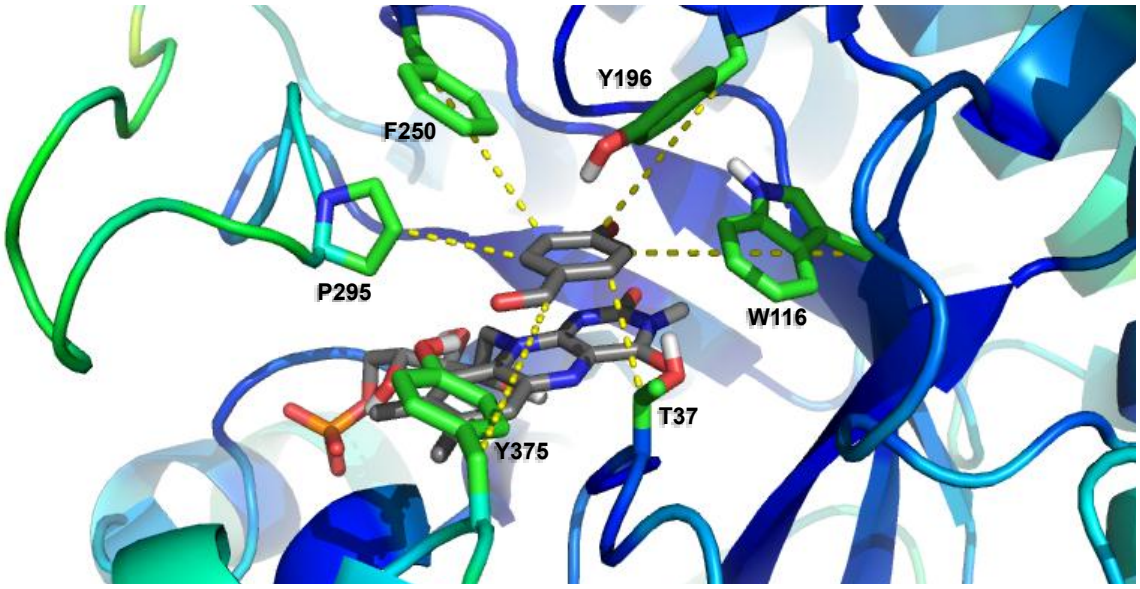


Figure 4-2. OYE 1 active site measurements.

This figure depicts the methodology used to conduct measurements listed in Table 4-1 using the crystal structure of OYE 1 (PDB: 1OYB).

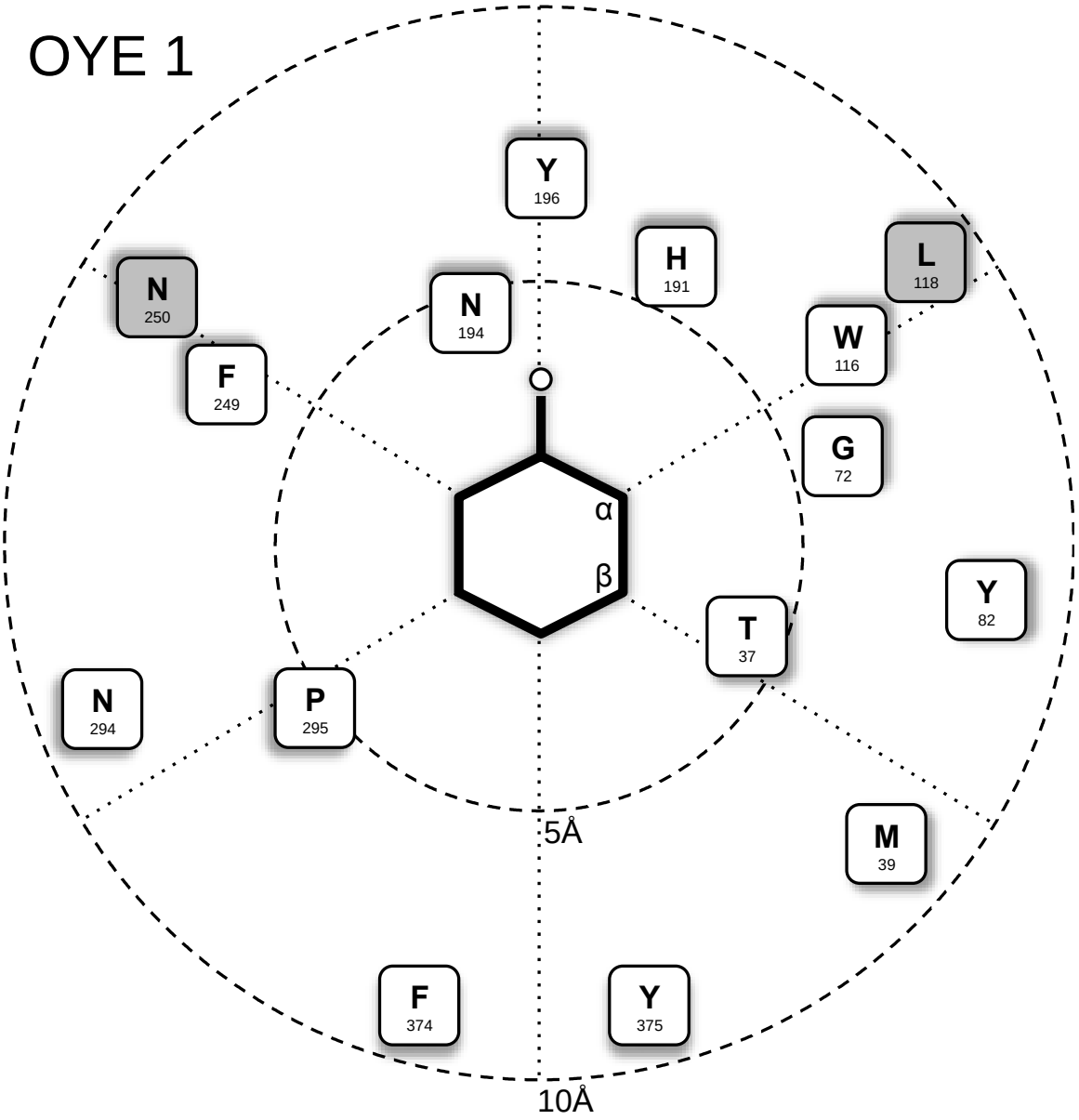


Figure 4-3. OYE active site schematic.

This figure depicts a 2-dimensional representation of the distribution of amino acid residues considered during this study. Despite their proximity to the modeled substrate, residues indicated in grey were not selected due to their obscuration by more closely positioned selections.

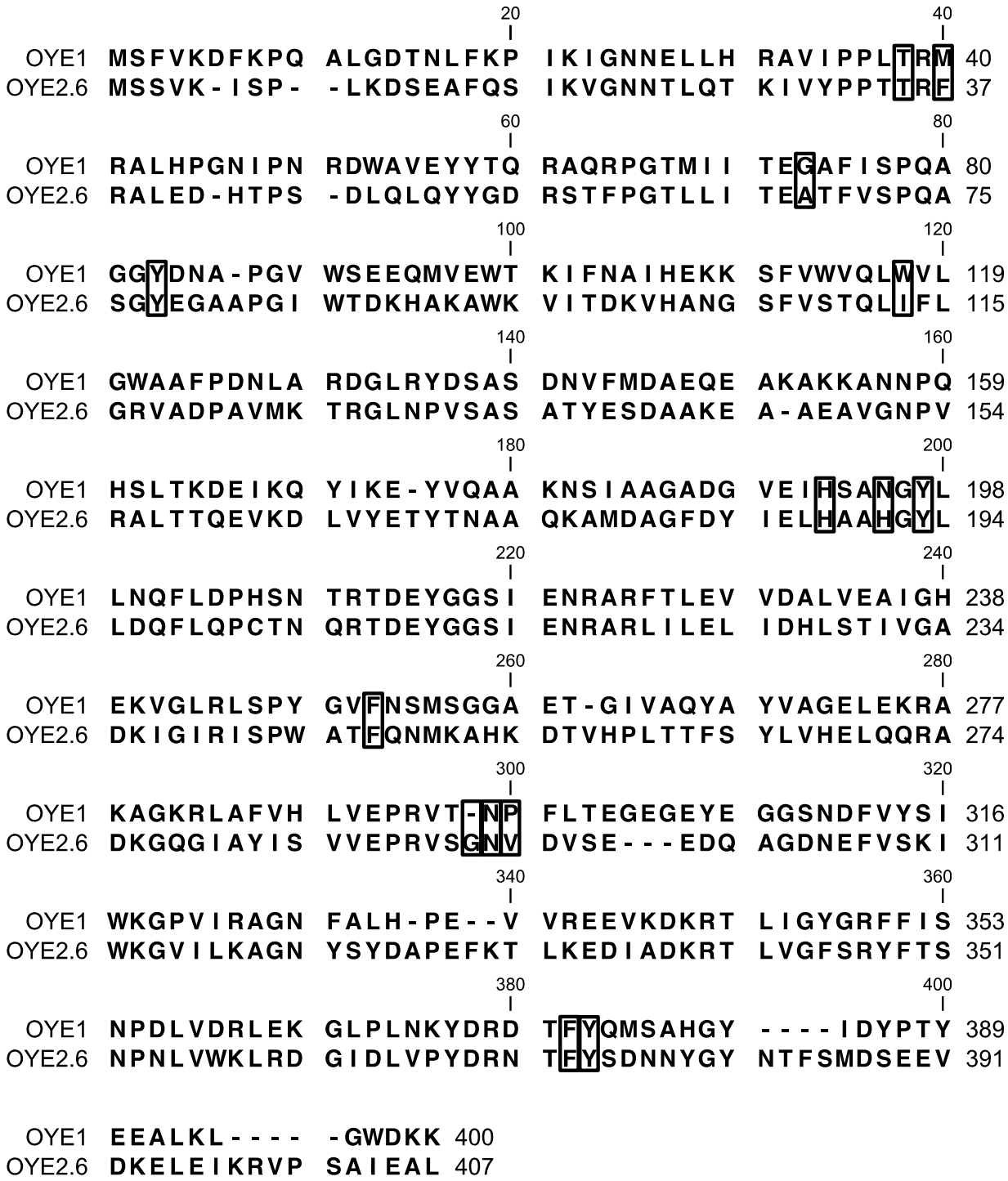


Figure 4-4. Sequence alignment of OYE 1 and OYE 2.6.

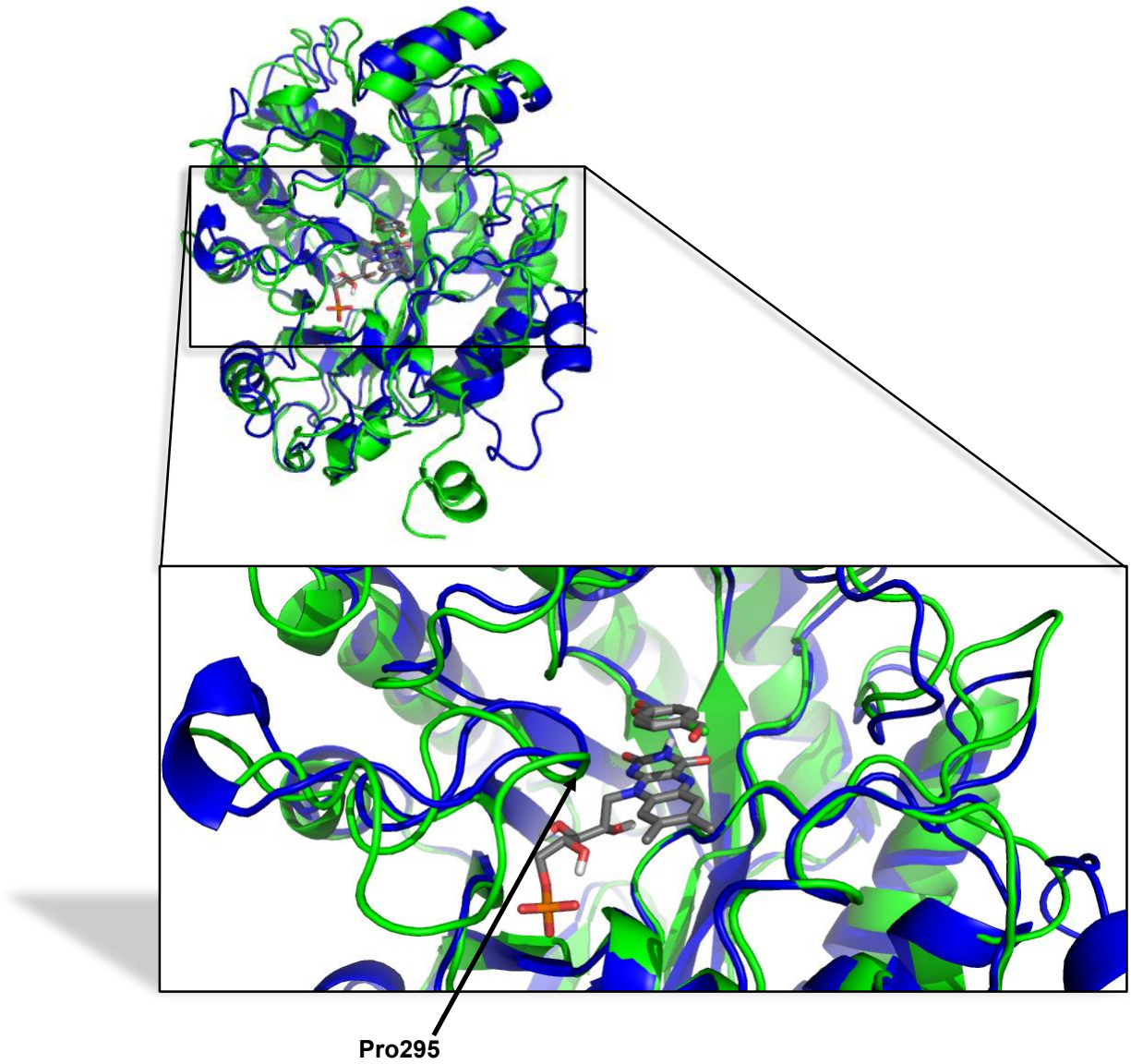


Figure 4-5. Overlay of OYE1 and OYE2.6 crystal structures

The ribbon diagram of OYE 1 (green, PDB: 1OYB) is shown aligned with that of OYE 2.6 (blue, PDB: 4DF2). Bound ligands, *p*-hydroxybenzaldehyde (OYE 1), *p*-chlorophenol (OYE 2.6), and FMN are depicted in stick form with carbons in grey. The position of proline 295 (OYE 1), which has no analogous residue in OYE 2.6, is indicated with a black arrow.

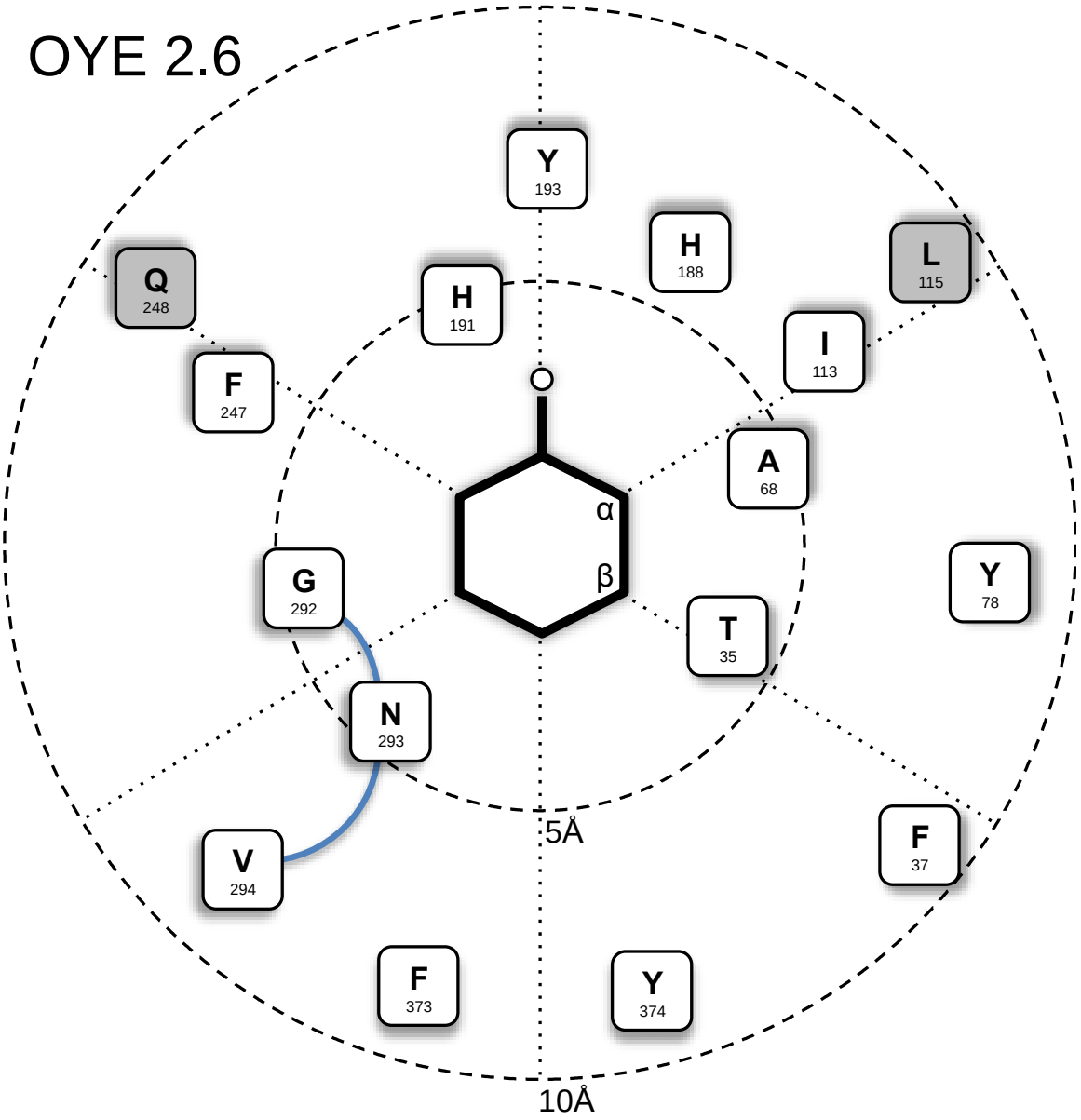


Figure 4-6. OYE 2.6 active site schematic

This figure depicts a 2-dimensional representation of the distribution of amino acid residues considered during this study. Despite their proximity to the modeled substrate, residues indicated in grey were not selected due to their obscuration by more closely positioned selections. A blue line connects the flexible loop segment (residues 292-294).

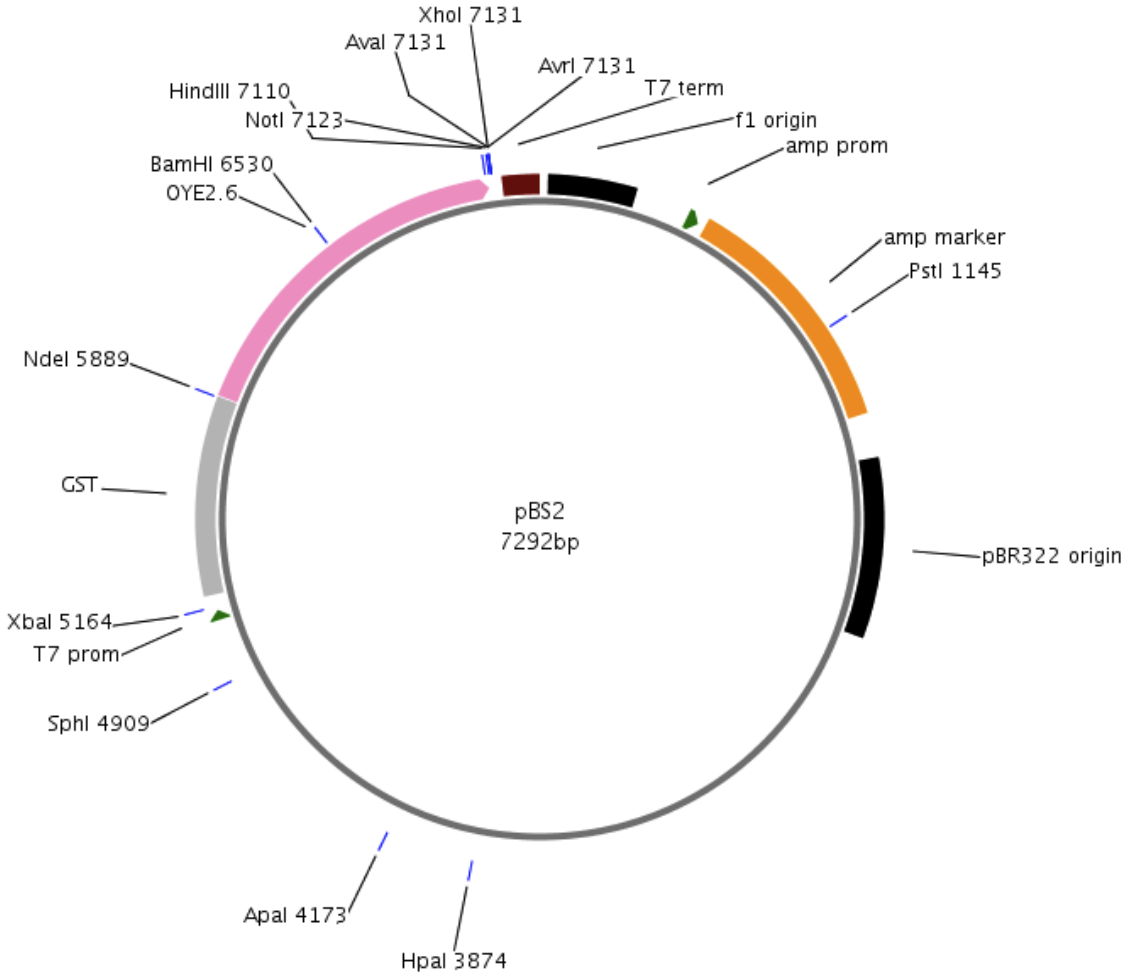
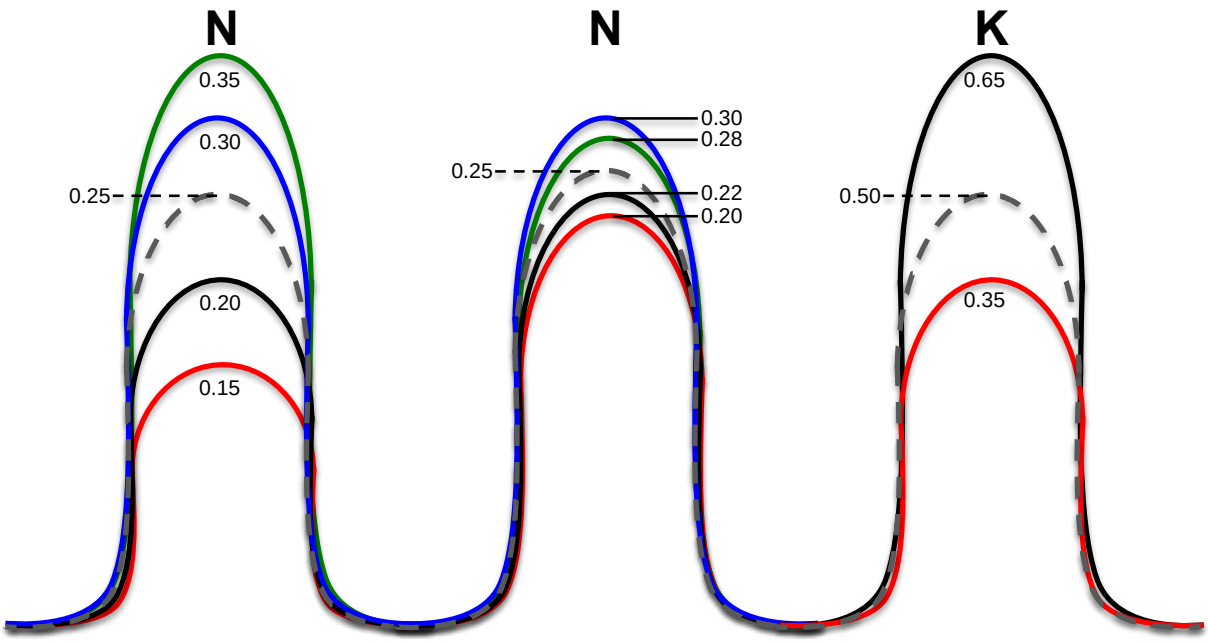


Figure 4-7. Plasmid map for pBS2 template.

Plasmid rendering done using PlasMapper v. 2.0 available at http://wishart.biology.ualberta.ca/PlasMapper/.



0.25 - |0.25 - 0.35| = 0.15
0.25 - |0.25 - 0.30| = 0.20
0.25 - |0.25 - 0.20| = 0.20
0.25 - |0.25 - 0.15| = 0.15

Q_N = 0.15 + 0.20
0.20 + 0.15 = **0.70**

0.25 - |0.25 - 0.30| = 0.20
0.25 - |0.25 - 0.28| = 0.22
0.25 - |0.25 - 0.22| = 0.22
0.25 - |0.25 - 0.20| = 0.20

Q_N = 0.20 + 0.22
0.22 + 0.20 = **0.84**

0.50 - |0.50 - 0.65| = 0.35
0.50 - |0.50 - 0.35| = 0.35

Q_K = 0.35 + 0.35 = **0.70**

$$Q_{pooled} = (0.4 \times 0.70) + (0.4 \times 0.84) + (0.2 \times 0.70) = \mathbf{0.76}$$

Figure 4-8. Sample calculation of Q_{pooled} from a degenerate sequencing chromat

NNK degeneracy is depicted for a sample codon. A (green), C (blue), G (black), T (red), and average (dashed) peaks are depicted where appropriate. Perfect degeneracy is defined as when all peaks overlap. Fractional peak values are indicated for each curve and the total value at any position sums to 1. Individual base peaks closer to the average yield a higher score for the base, the position, and theoretically the codon as a whole.

Table 4-2. Summary of mutagenesis development efforts

Reference (Method)	Primers overlap (5'-end)	Primers non-overlap (3'-end)	Q_{codon}	Notes
Bougioukou *et al.* (QuikChange®)	33	-	0.84	Primer concatemer inserts 15 amino acids
Edelheit *et al.* (SPRINP)	33	-	0.41	High parent template
Phusion® (SDM Kit)	-	27	0.42	High parent template "GGG" rich
Liu and Naismith	20	25	0.15	High parent template
Zheng *et al.*	25	15	0.69	17 amino acids

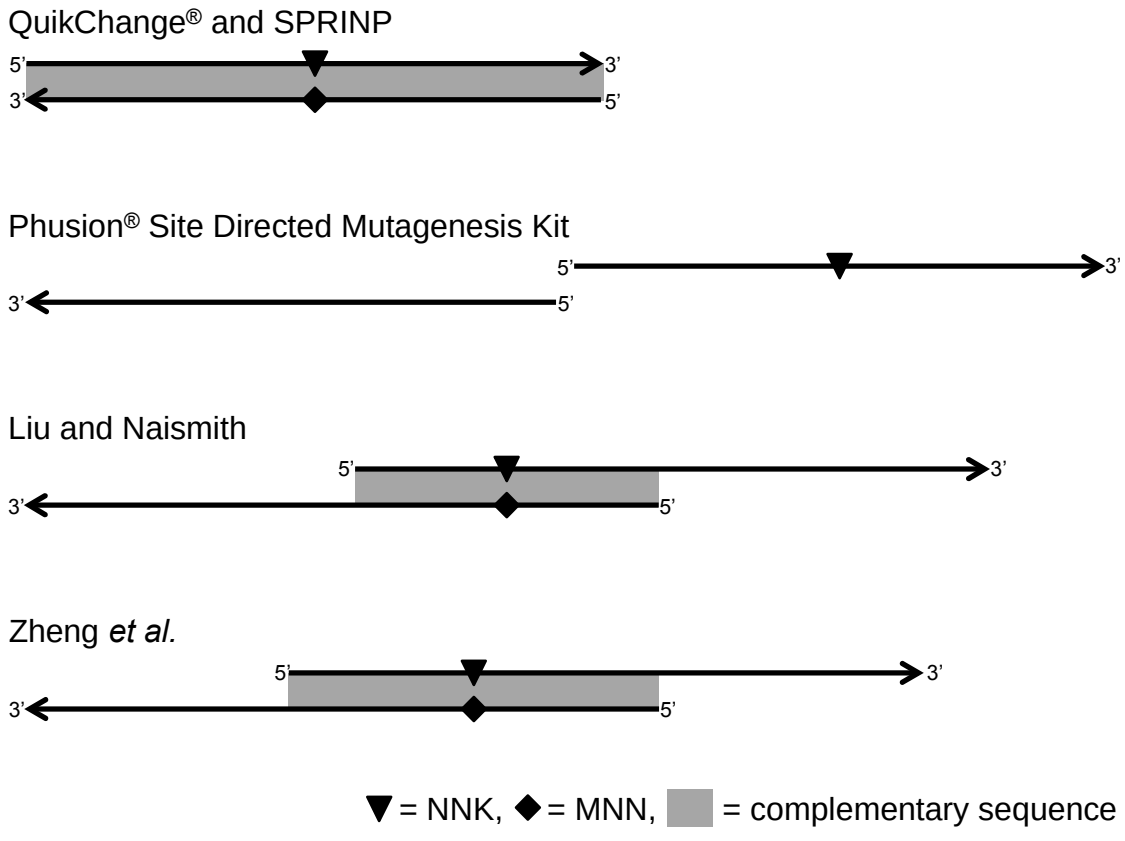


Figure 4-9. Schematic of primer designs used

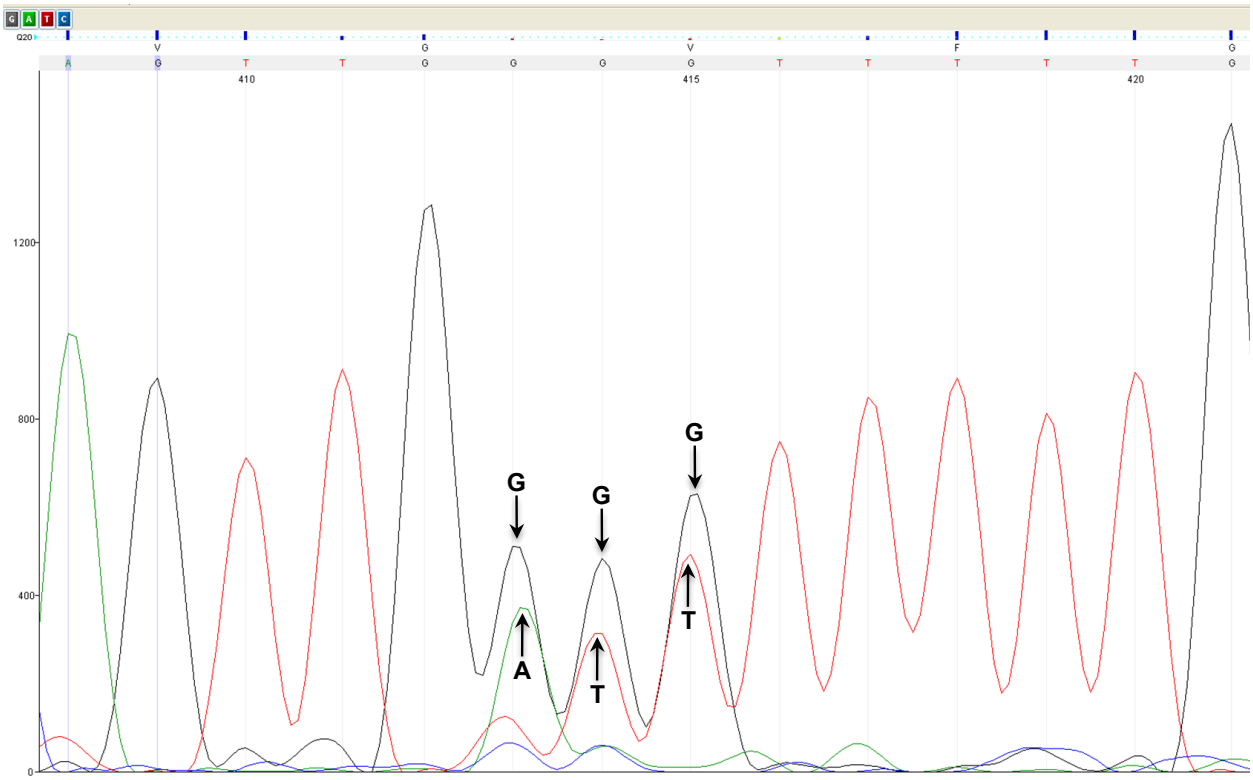


Figure 4-10. Pooled plasmid sequencing from Phusion® SDM Kit experiment

Screen-shot image of sequencing results from OYE 2.6 Ile113 NNK saturation mutagenesis using the Phusion® SDM Kit. G is the predominant species at each position of the codon while the wild type codon (ATT) shows below. While small peaks indicate a degree of degeneracy at the first two positions in the codon they are only slightly above background of the chromat. At our discretion this run was discarded after visual inspection and no Q value assessment was conducted.

Table 4-3. Amino acid bias.

Amino Acid	# codons	# observed	% observed
R	3	86	10.6%
L	3	76	9.4%
S	3	61	7.5%
G	2	116	14.3%
P	2	84	10.3%
A	2	62	7.6%
T	2	50	6.2%
V	2	48	5.9%
W	1	34	4.2%
Q	1	31	3.8%
E	1	22	2.7%
D	1	19	2.3%
H	1	19	2.3%
M	1	18	2.2%
Y	1	16	2.0%
K	1	14	1.7%
STOP	1	14	1.7%
N	1	13	1.6%
C	1	11	1.4%
I	1	10	1.2%
F	1	8	1.0%

The number of each amino acid observed is the sum of occurrences in all plate-sequenced libraries. Wild type codon appearance is not included in these totals. Amino acids encoded by NNK degeneracy with one, two, and three codons should theoretically provide 3%, 6% and 9% occurrence respectively.

Table 4-4. Assessment of first round library degeneracy.

Library	# amino acids	# codons	# sequenced	Q_{pooled}	Q_{codon}	ΔQ	% wt
T35X	18	30	79	0.66	0.71	0.05	20
F37X	20	30	74	0.80	0.83	0.03	3
I113X	17	26	84	0.69	0.76	0.07	17
H188X	19	31	92	0.72	0.79	0.07	15
H191X	16	28	91	0.84	0.78	-0.06	5
Y193X	17	24	87	0.63	0.70	0.07	14
F247X	19	30	93	0.76	0.80	0.04	13
N293X	18	27	92	0.76	0.72	-0.04	5
V294X	17	29	82	0.73	0.65	-0.08	22
F373X	16	24	90	0.64	0.60	-0.04	3
Y374X	15	22	87	0.62	0.44	-0.18	31

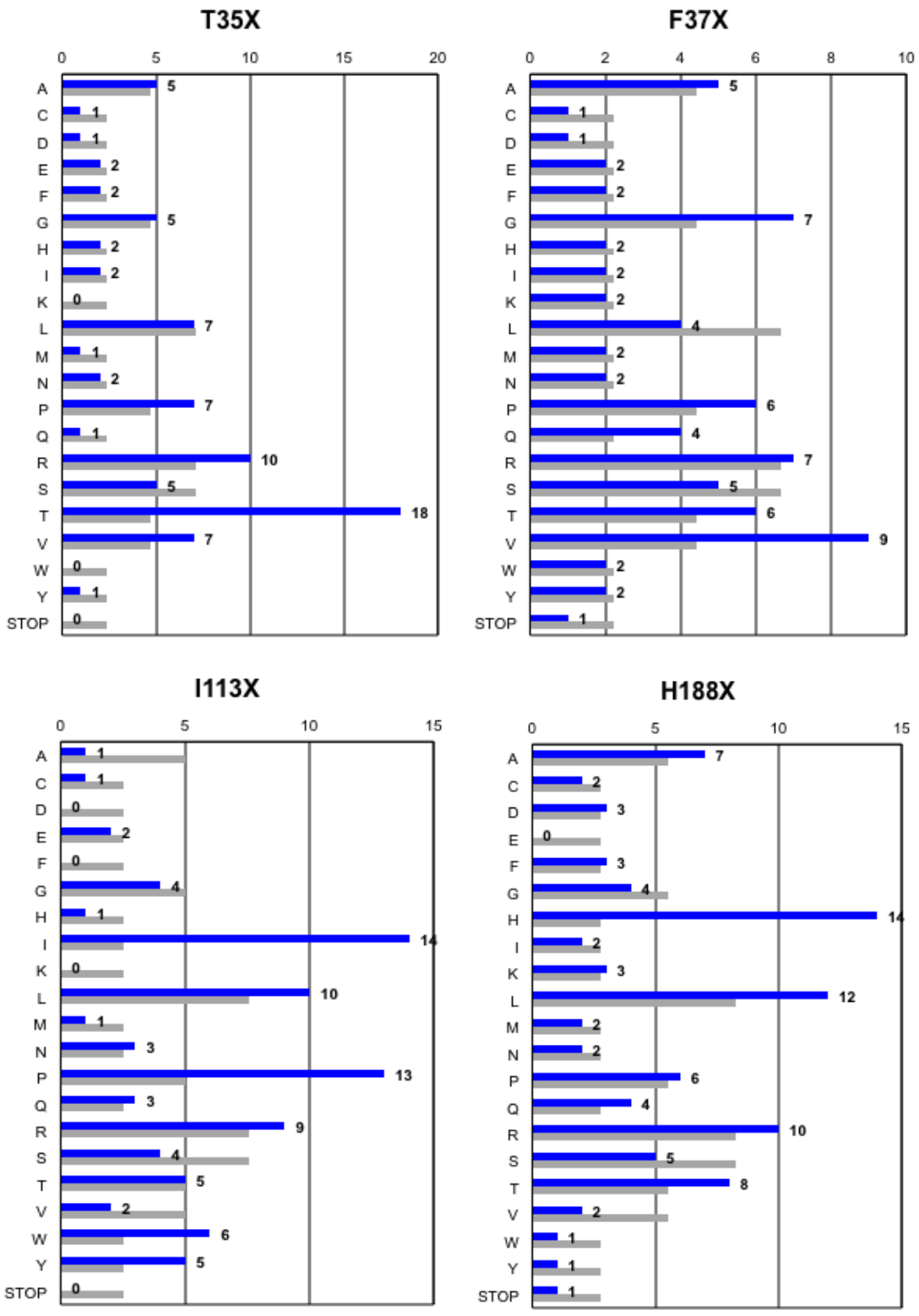


Figure 4-11. Amino acid distribution for T35, F37, I113 and H188 libraries

Blue bars and numbers indicate the observed count of each amino acid. Grey bars depict the theoretical distribution based on sample size and NNK degeneracy.

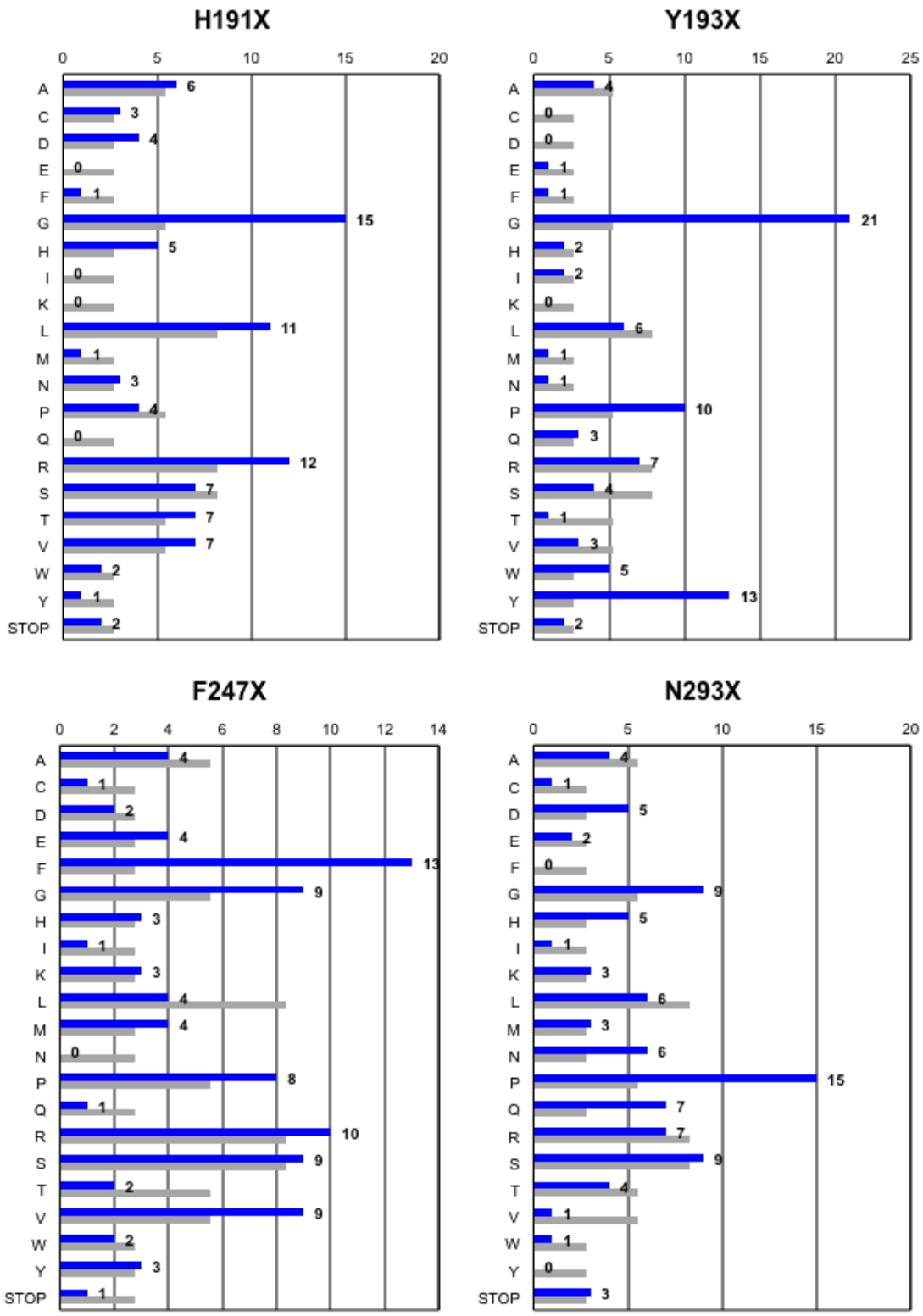


Figure 4-12. Amino acid distribution for H191, Y193, F247 and N293 libraries

Blue bars and numbers indicate the observed count of each amino acid. Grey bars depict the theoretical distribution based on sample size and NNK degeneracy.

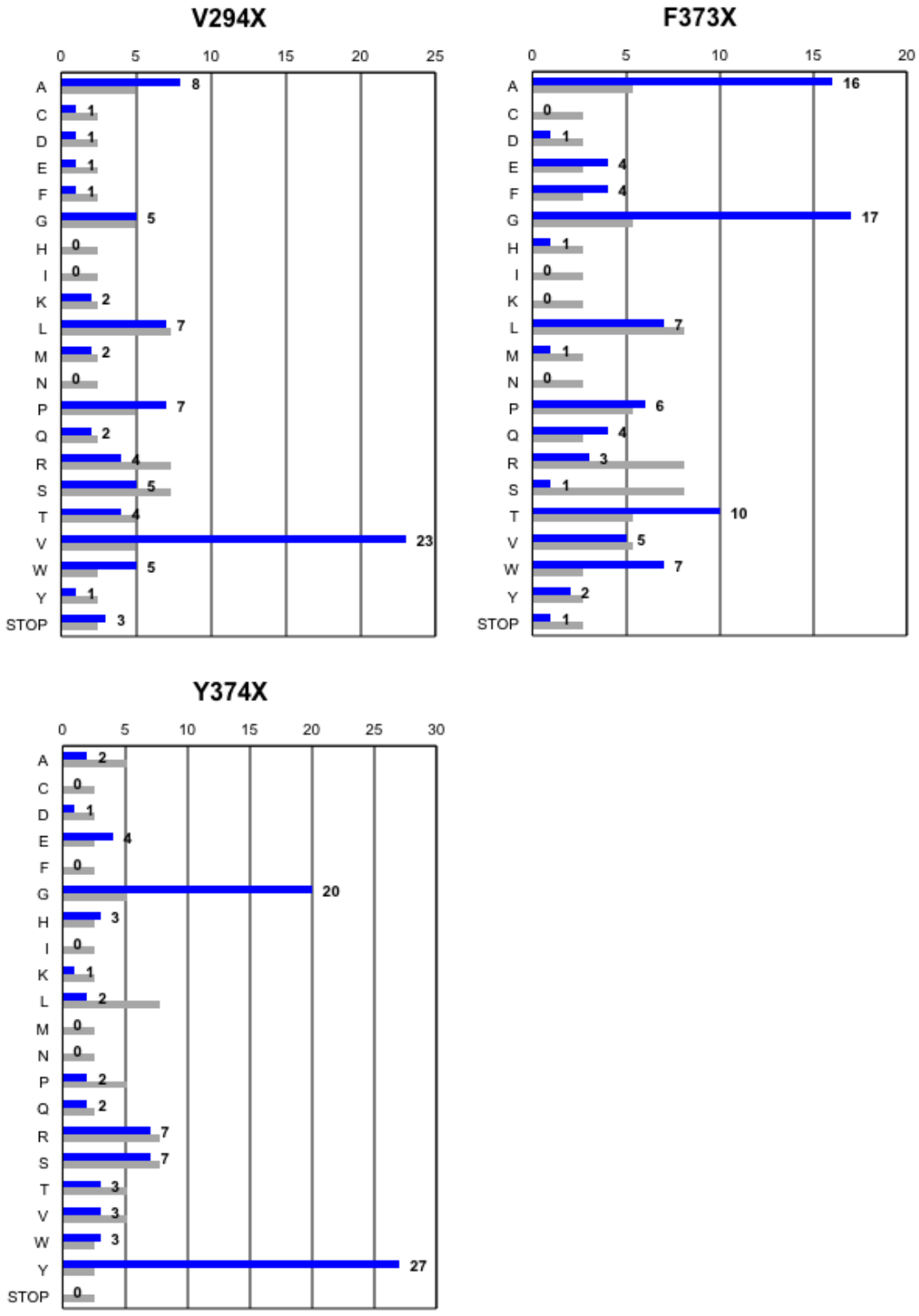


Figure 4-13. Amino acid distribution for V294, F373 and Y374 libraries

Blue bars and numbers indicate the observed count of each amino acid. Grey bars depict the theoretical distribution based on sample size and NNK degeneracy.

CHAPTER 5
ITERATIVE SATURATION MUTAGENESIS: PUTTING IT ALL TOGETHER

Introduction

In order to complete the evolutionary cycle, the libraries created by saturation mutagenesis needed to be screened and assessed. The best mutants identified in first-generation libraries become template DNA for a second round of saturation mutagenesis, hence the term Iterative Saturation Mutagenesis. This chapter discusses the screening, assessment and selection phases of the laboratory evolutionary cycle as summarized in Figure 5-1.

Experimental Strategy

Rather than increase the degree of oversampling required to achieve saturation at a given library position, we decided to accept the risk associated with keeping our screening libraries at 95 members. In short, we assumed that only 17 of 20 amino acids were likely present in any given saturated library. For libraries that display sufficiently interesting screening results, we proposed to sequence the 95-member library to identify missing members. Amino acids not represented would then be individually constructed using site-directed mutagenesis.

In order to screen mutant libraries, we decided to use a whole cell assay technique for the benefits of simplicity and time. Expression strains containing unique mutant plasmids can be cultured in a 96-well plate format and induced. Rather than conducting cell lysis and providing exogenous co-factor and the necessary recycling system, as done previously[124], this approach allows us to keep the cells intact and use their intracellular NADPH pool and physiologic recycling mechanism. Previous experience[130,131] led us to conclude that redox reactions involving esters and cyclic

ketones are possible in the whole cell format. Some approaches to whole cell screening include the addition of small amounts of cofactor, regeneration enzymes and additives to enhance the permeability of cell membranes.[126,132] Our methodology is more straightforward and requires only the addition of glucose to the reaction buffer.

Induction of mutant proteins is also further simplified using auto-induction media as described by Studier.[133] Auto-induction media is inoculated and allowed to grow to saturation overnight. Care must be taken to provide adequate aeration in order to maximize expression in our host strain. Following induction, the cell material is isolated and mixed with reaction buffer containing substrate and glucose and the reaction is allowed to proceed for six hours. Conveniently, the whole cycle takes approximately 24 hours from inoculation to the start of our analysis phase by GC-FID.

GC analysis is conducted as described in Chapter 1, Experimental Procedures. Total time for a single 96-well plate GC analysis is approximately 32 hours for substrates **1** and **2**, and 38 hours for substrate **3**. Induction and analysis cycles were staggered, where possible, to maximize throughput in our lab.

Once first round screening was complete for all libraries, the best results were noted and used as “anchors” for second round library creation. Specifically, plasmid DNA from desirable mutants was isolated and subject to randomization at a site other than its original mutation that also gave an interesting or desirable result. The resulting second round library was subject to screening as previously described. This process was to be repeated until our desired outcome was achieved.

Given that OYE 2.6 reduces all three of our Baylis-Hillman adducts efficiently and with a strong preference for the (*S*)-enantiomer, our primary consideration for a positive

result was any mutant that could shift the reduction product toward racemic or even preferentially to the (*R*)-enantiomer. At the same time total conversion had to be taken into consideration. Results giving less than 10% conversion, regardless of apparent ee%, were not considered for follow-up as second round "anchor" candidates.

Experimental Procedures

Degenerate libraries were constructed and arrayed as described in Chapter 4 into a 96-well format. These libraries were screened as constructed with the exception of the Library Master Plate (LMP).

Library Master Plates

First round libraries that were sequenced (Table 5-1) to determine the level of degeneracy actually present were arrayed into a consolidated format. These LMPs were constructed in accordance with the template depicted in Figure 5-2. The idea is that each codon (if available) is present only once. This eliminates duplicate screening reactions and makes it possible for a single micro-titer plate to contain three different saturated libraries. The resulting plates were then replicated and stored as previously described in Chapter 4, Experimental Procedures.

Auto Induction

Aliquots (20 µL) from each well of freshly prepared library plates or stabs from library plates stored at -80°C were then used to inoculate a new plate containing 600 µL ZYP-5052 auto-inducing media (10 g/L tryptone, 5 g/L yeast extract, 1 mM $MgSO_4$, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4, 5 g/L glycerol, 0.5 g/L anhydrous glucose, 2 g/L α-lactose monohydrate) supplemented with 200 mg/L ampicillin.

Plates were then mounted in a locally fabricated growth apparatus (Figure 5-3) designed to facilitate maximal oxygen transfer[134] and were shaken at 300 rpm and 37°C

for 16-18 hours. To further ensure maximum oxygen transfer rate (OTR), cells were grown under these conditions in square well plates.[135] Following auto-induction, the cell pellets were harvested by centrifugation at 3000 rpm for 30 minutes and the media was removed by aspiration.

Whole Cell Assays

Cell pellets were resuspended in 300 µL of 50 mM phosphate buffer, pH 7.0, augmented with 100 mM glucose and 10 mM substrate **1**, **2**, or **3**. The plates were shaken at 250 rpm and room temperature for 6 hours. 500 µL of ethyl acetate was then added to each well and the plates were shaken as above for 30 minutes followed by a short centrifugation to facilitate separation of the aqueous and organic phases. The organic layer was then analyzed by chiral-phase GC as described in Chapter 1.

Site-directed Mutagenesis

The I113D, I113F, I113K, Y193C, Y193D, Y193K, Y78W/I113C, Y78W/I113W, and Y78W/I113Y mutants were prepared by a modification of the method previously described in Chapter 4, Experimental Procedures, based on their absence from sequenced plate libraries. Each PCR reaction (total volume 100 µL) containing 5 × Phusion® HF Buffer (20 µL), pBS2 (20 ng), forward and reverse primers (0.5 µM each), dNTPs (200 µM each), and Phusion® Hot Start II High-Fidelity DNA Polymerase (1 U) was subjected to an initial denaturation step of 98°C (30 s) followed by 25 cycles of 98°C (10 s) and 72°C (4 min) followed by a final incubation at 72°C (7 min). Amplicons were purified by DNA spin columns, digested with *Dpn*I at 37°C (10 U for 4 h followed by an additional 10 U for 4 h) and then purified by an additional DNA spin column. Aliquots (5 µL) were used to transform *E. coli* JM109 (75 µL) by electroporation. SOC medium was added (600 µL), then samples were incubated for 1 h at 37°C prior to

selection on LB medium supplemented with ampicillin. Plasmid DNA was purified (spin columns) from randomly chosen colonies and was subsequently analyzed by DNA sequencing to identify the desired OYE 2.6 variants. Once the desired mutations were identified the plasmid DNA was further transformed into *E. coli* BL21 Gold (DE3) for the purpose of expression and screening.

Results and Discussion

Table 5-1 lists all libraries screened in this study. Each was screened against all three Baylis-Hillman adducts discussed in Chapter 1. Eleven completely sequenced first round libraries were screened in LMP formats while the others were screened "blind". In the latter case, only clones yielding interesting results were sequenced. As previously discussed, only the F37X library was complete, containing all 20 amino acid residues after degenerate primer PCR. We decided to use manual methods to make the missing members of the I113X, Y193X, and Y78W/I113X libraries based on the results of screening (*vide infra*).

First Round Screening.

Figures 5-4 thru 5-7 graphically depict first round screening results from LMP libraries. Libraries A68X, Y78X, and G292X were screened "blind" and results are reported by exception only. In very general terms, the total conversion for substrate **1** > substrate **2** > substrate **3**. Based on these data, we identified several mutations that influenced the stereoselectivity of OYE 2.6. These results are summarized in Table 5-2. The use of LMPs also provided us with useful information about the active site of OYE 2.6 by comparing those results to mutagenesis results on analogous positions in the active site of OYE 1.

T35X

OYE 2.6 Thr35 is analogous by sequence alignment to Thr37 in OYE 1. In the case of OYE 2.6, only the T37S mutation retained activity similar to that of the wild type enzyme. This is logical given the structural similarity of serine and threonine residues, both with a structurally similar hydroxyl group. Partial activity was also present when this position contained Cys, Met, and Ala. It is likely that Thr35 in OYE 2.6 also possibly helps to stabilize charge on the reduced flavin as does the analogous Thr37 in OYE 1 described previously.[32]

Given the relatively poor results observed, we did not use this position in further rounds of mutagenesis. In the alkene reductase, YqjM, Cys26 aligns with the analogous active site position. Bougioukou *et al*. found that this enzyme tolerated several residues at this position (tryptophan, alanine, asparagine, aspartate, glycine, and valine) and these allowed for slight improvements in enzymatic activity and stereoselectivity when screened against the model substrate used in this study.[124] When the glycine and aspartate mutations were combined with additional changes elsewhere in the protein, pronounced effects in stereoselectivity were observed.

H188X and H191X

Positions His188 and His191 are analogous by alignment to His191 and Asn194 in OYE 1. The side-chains of these residues likely serve as hydrogen bond donors to the electronegative carbonyl oxygen that polarize the alkene and stabilize the enol(ate) intermediates.[31] Interestingly, there appears to be no suitable substitute for His188 that retains enzyme activity. Partial conversion is recovered when His191 is substituted with asparagine. This is logical given that OYE 1 has an asparagine at the same position;

however, it is clear that OYE 2.6 has a distinct preference for histidine at that same position.

A detailed analysis of the role of these residues in the OYE homolog pentaerythritol tetranitrate (PETN) reductase from *Enterobacter cloacae* PB2 has recently been conducted. Initial results indicated that, for a wide range of α,β-unsaturated carbonyl compounds, NADPH-mediated alkene reduction was largely eliminated in nearly all members of site-saturation mutant libraries at the analogous positions (His181 and His184).[136] These results are consistent with our observations for OYE 2.6. A more focused study by the same group on the H181N and H184N mutants revealed that the enantioselectivity for the specific compound 2-phenyl-1-nitropropene (and its *p*-chloro derivative) could be enhanced from ~50% ee (wild type enzyme) to ~90% ee (either H181N or H184N).[137] Unfortunately, a similar improvement in enantioselectivity was not observed in our case with OYE 2.6. However, the suitability of asparagine, at least at position 191 (OYE 2.6 mutant) and position 194 (OYE 1 wild type) is consistent with our results. Further investigation of these positions in the *E.cloacae* enzyme revealed that multiple substitutions at positions 181 (A, C, and N) and 184 (A, F, I, N, K, and R) were capable of similar enantioselectivity shifts with the model nitro compound.[138] On the other hand, these results did not extend to (*R*)-carvone, further supporting our observations and lending some hope that a broader substrate screen with our mutants may reveal altered enantiopreferences.

Y193X

OYE 2.6 Tyr193 is analogous to Tyr196 in OYE 1. According to Kohli and Massey, the role of Tyr at this position is to provide a proton to the stabilized enolate intermediate.[13] Without an appropriately positioned proton donor, catalysis should not

occur. In our investigation this would appear to be largely true with the exceptions of Y193C and Y193T. While it seems plausible that cysteine could provide an acidic proton at that position, it is difficult to imagine that threonine is sufficiently acidic to carry out the same task. Given that these mutations provide near-racemic products it seems plausible that the anionic intermediate undergoes protonation by an alternate mechanism.

Using the purified mutants, Y193C and Y193T, we conducted isolated enzyme screening of these mutants against a broader list of potential substrates (Figure 5-8). In general terms, these results show dramatically reduced activity for these two mutants as compared to the wild type enzyme. When screened against 2-methyl substituted cyclic enones the stereoselectivity is still diminished but shows a slight preference for the (*R*)-enantiomer and thus a "normal" binding mode. When screened against (*R*)- and (*S*)-carvone the enantiomeric excess of the products were reduced by 10-20% from that of the wild type enzyme. These results indicate that the formation of racemic reduction products is not an absolute function of the mutant enzyme. At this point it is not clear if the formation of racemic reduction products is due to altered substrate binding or an alternate protonation mechanism, either protein or solvent in origin.

I113X

We expected that variations at Ile113 of OYE 2.6 would yield similar results as those observed for the analogous mutations at Trp116 of OYE 1 (results discussed in detail in Chapter 1). This was essentially true for **1** and **2**; however, **3** did not display a wide range of enantioselectivity as observed with OYE 1. I113W provided a shift to 11% ee for (*R*)-**5** and I113D to 13% (S)-**4** qualifying those positions as an "anchors" for our second round investigations.

Summary of first round results

Table 5-2 summarizes the best results of our first round screening efforts. Y78W was the only mutation identified as useful in our "blind" screening experiments. None of the mutants identified in this round of screening exhibited any propensity to alter the enantioselectivity for **3**. We chose Y78W, I113D, I113W, and V294P as anchors for second round library construction. We deliberately excluded the Phe247 mutants listed in the table because of their relatively low impact on enantioselectivity. We also chose to exclude the Tyr193 mutants given our estimation that the source of poor enantioselectivity is most likely an alternate protonation mechanism as opposed to an alternate binding mode that we believe is the key to stereoselectivity.

Second Round Screening

Second round libraries were constructed with the best mutant "anchors" identified in the first round. Those mutations and their pooled degeneracy assessments are listed in Table 5-1. Table 5-3 summarizes our second round screening efforts. In several cases, the resulting libraries displayed a complete loss of activity. Others retained the stereoselectivity of the original parent and in some cases we saw a significant degradation in conversion. These libraries are indicated in Table 5-3 as "no change" or "low conversion" respectively. In two libraries we saw results we classified as varied. For the library anchored with I113W and randomized at Tyr78, varied results in enantioselectivity were observed for substrate **2** only. However, for all members that displayed activity, the results lay between racemic and high in (*S*)-character. Given its lack of improvement in (*R*)-enantioselectivity, this library was not sequenced.

Y78W

The second round library for Y78W randomized at Ile113 yields significant and varied results for substrates **2** and **3**. Figure 5-9 summarizes the data for this screening. Interestingly, there is no general change in enantioselectivity when this library was assayed against substrate **1**, although a wide variance in data is observed for **2** and **3**. These results are depicted graphically in Figures 5-10 and 5-11. These results are similar to the results observed for the OYE 1 W116X library against substrate **3**. A range of both conversion and enantioselectivity over the full spectrum from (*R*) to (*S*) was observed for both enzymes. However, when OYE 2.6 is altered with a Y78W mutation and position 113 is varied this same effect is seen for both **2** and **3**. The bottom line is that changing Tyr78 to Trp makes OYE 2.6 behave more like OYE 1.

We used the recently solved X-ray crystal structure of OYE 2.6 to understand the results of mutagenesis. Overlaying the structure of OYE 1 W116I with substrate **2** bound in the active site in both productive and non-productive binding modes (PDB 3RND) with the structure of wild-type OYE 2.6 (PDB 4DF2) makes the role of Tyr78 evident. Figure 5-12 depicts the relative position of both tyrosine residues in the active site with respect to the bound ligand. The distance from the hydroxymethyl oxygen to the tyrosyl oxygen in OYE 2.6 is 1.8 Å. The distance between the same ligand and the analogous residue Tyr82 in OYE 1 W116I is 5.4 Å. This three-fold increase in distance is a function backbone positioning within the active site. It is a reasonable hypothesis that when Tyr78 is mutated to a tryptophan, the lack of the potential hydrogen bond with the hydroxymethyl moiety of the Baylis-Hillman substrate weakens the affinity for the "normal" (*S*)-binding mode. As a result, changes in the nature of the residue at position

113 have larger effects and are able to now provoke substrate binding in the alternate mode.

Y78W, F247X

At the same time, the apparent proximity of the hydroxyl moiety of Tyr78 to the modeled substrate (Figure 5-12) offers an explanation as to OYE 2.6 wild type enzyme's proficiency in yielding high enantiomeric excess of (*S*)-hydroxymethyl products. Our results suggested that mutation of Tyr78 to Trp is a *prerequisite* to achieving altered binding of hydroxymethyl substrates. It is then a concern that other positions in the active site that may have influence on substrate binding, analogous to that observed by variations at position 113, may have been overlooked during first round screening.

To test this hypothesis, an additional library consisting of OYE 2.6 Y78W further randomized at position 247 was prepared. Phe247 is located in the active site in close proximity to where the substrate hydroxyl would be positioned in a pro-(*R*) binding mode. In theory, if the (*S*)-stabilizing influence of Tyr78 were removed by mutation to Trp then it may be able to stabilize the opposite binding mode with an appropriate side chain substitution. Initial screening of this library against **1** and **3** revealed dramatically reduced conversion among most variants and minimal shifts in enantiomeric excess where observed. However, when screened against **2**, ten variants with greater than 90% conversion and enantiomeric excess approaching 90% were observed. Sequencing revealed that these mutants were F247H (89% (*R*)), F247A (87% (*R*)), and F247W (85% (*R*)). These mutants have yet to be purified for further characterization.

These observations support our hypothesis that stabilizing the alternate binding mode is possible when the primary mode is disfavored. The observation that three

variants lead to increased enantiopreference mimics the range of possible solutions observed when variations are performed at position 113. The varied side chain properties of these three variants (His, Ala, and Trp) seem to imply that the stabilizing effect is more general than specific in nature. The single mutation (Y78W) parent enzyme in this screening only displays altered stereoselectivity for **2**. Since this mutant offers no change in enantioselectivity for **1**, and the enzyme is inactive with **3,** it is not surprising that further mutations at position 247 have little effect on these substrates. It follows that a more divergent approach to evolving enantiopreference may be required to further improve results for substrates **1** and **3**.

Purified enzyme screening results

OYE 2.6 variants displaying the greatest degree of (*R*)-stereoselectivity identified during the systematic investivgation of iterative mutagenesis were purified by GST affinity column and screened for activity as in Chapter 1. Screening results are summarized in Table 5-4. Under the same conditions, wild-type OYE 2.6 offers high (*S*)-selectivity and full conversion for 10 mM **1** and **2** in less than one hour and within 3 hours for **3**.

None of the second-generation mutants yielded altered stereo selectivity for **1**. OYE 2.6 Y78W and Y78W/I113M provide for full reduction of **2** to **5** with an enantiomeric excess of 75% (*R*) within 6 or 8 hours respectively. In contrast, Y78W/I113C, Y78W/I113F, Y78W/I113L, and Y78W/I113V yielded 90% (*R*)-**6** to a limit of approximately 50% conversion within 24 hours. This is similar to the results obtained when screening OYE 1 W116V against **3**, implying that the mutation of Tyr78 is an absolute requirement to make OYE 2.6 more like OYE 1 in active site topography.

Conclusions

Foremost, we have both developed and demonstrated an approach to saturation mutagenesis that significantly reduces the library size and subsequent screening effort. These tools for directed evolution are therefore now more accessible to non-specialist high-throughput laboratories like our own.

Significant progress was made toward our goal of achieving enantiocomplementarity for the OYE 2.6 enzyme. For substrates **2** and **3**, this result is essentially complete; however, additional work is needed to make these catalysts industrially viable. For reasons unknown, substrate **1** did not yield progress toward achieving the (*R*)-enantiomer. In this case, it is plausible that the cyclohexenone substrate, when allowed to bind in the "flipped" binding mode, proceeds through a dismutation mechanism resulting in enzyme inactivation with a bound phenol(ate) product. This would be especially pronounced if the mutant being evaluated had a reduced affinity for the NADPH substrate.

We have demonstrated that an iterative approach to saturation mutagenesis focused on the active site binding pocket residues in an alkene reductase is capable of altering the enantioselectivity of the parent enzyme (Figure 5-13). In our case, this result was dependent on the identities of three critical active site residues, Ile113, Tyr78 and Phe247. Previous experience with OYE 1 implicated position 113 as a focal point for altering the enantioselectivity of 2-substituted cyclic enones. This work led to identifying the role of Tyr78 in governing the enantioselectivity of the enzyme toward 2-hydroxymethyl substituted cyclic enones. These results were achieved by systematic cross combination of our best results from first round screening. Conjecture led to the further investigation of Phe247. Further combinations of variants at positions 113 and

247 have a high probability of success in further increasing (*R*)-stereoselectivity for hydroxymethyl-substituted substrates. Fundamentally, we have demonstrated the ability to direct the binding of functionalized enones within an alkene reductase enzyme.

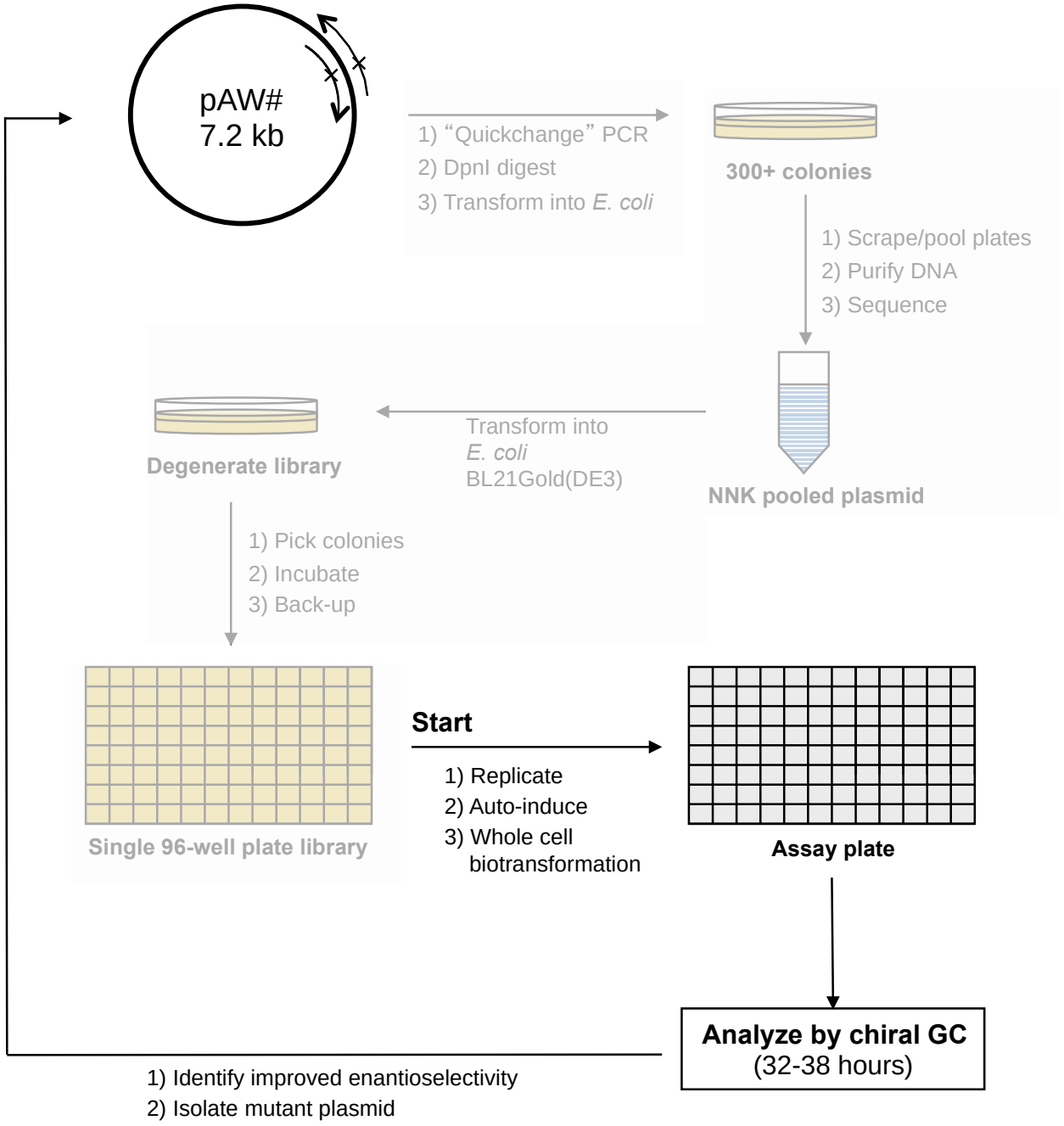


Figure 5-1. Summary of Chapter 5 experimental objectives and efforts

	Position #1								Position #2								Position #3							
	1		2		3		4		5		6		7		8		9		10		11		12	
A	A	GCG	H	CAT	P	CCG	S	TCT	A	GCG	H	CAT	P	CCG	S	TCT	A	GCG	H	CAT	P	CCG	S	TCT
B	A	GCT	I	ATT	P	CCT	T	ACG	A	GCT	I	ATT	P	CCT	T	ACG	A	GCT	I	ATT	P	CCT	T	ACG
C	C	TGT	K	AAG	Q	CAG	T	ACT	C	TGT	K	AAG	Q	CAG	T	ACT	C	TGT	K	AAG	Q	CAG	T	ACT
D	D	GAT	L	CTG	R	AGG	V	GTG	D	GAT	L	CTG	R	AGG	V	GTG	D	GAT	L	CTG	R	AGG	V	GTG
E	E	GAG	L	CTT	R	CGG	V	GTT	E	GAG	L	CTT	R	CGG	V	GTT	E	GAG	L	CTT	R	CGG	V	GTT
F	F	TTT	L	TTG	R	CGT	W	TGG	F	TTT	L	TTG	R	CGT	W	TGG	F	TTT	L	TTG	R	CGT	W	TGG
G	G	GGG	M	ATG	S	AGT	Y	TAT	G	GGG	M	ATG	S	AGT	Y	TAT	G	GGG	M	ATG	S	AGT	Y	TAT
H	G	GGT	N	AAT	S	TCG	Stop	TAG	G	GGT	N	AAT	S	TCG	Stop	TAG	G	GGT	N	AAT	S	TCG	Stop	TAG

Figure 5-2. Template for construction of Library Master Plates

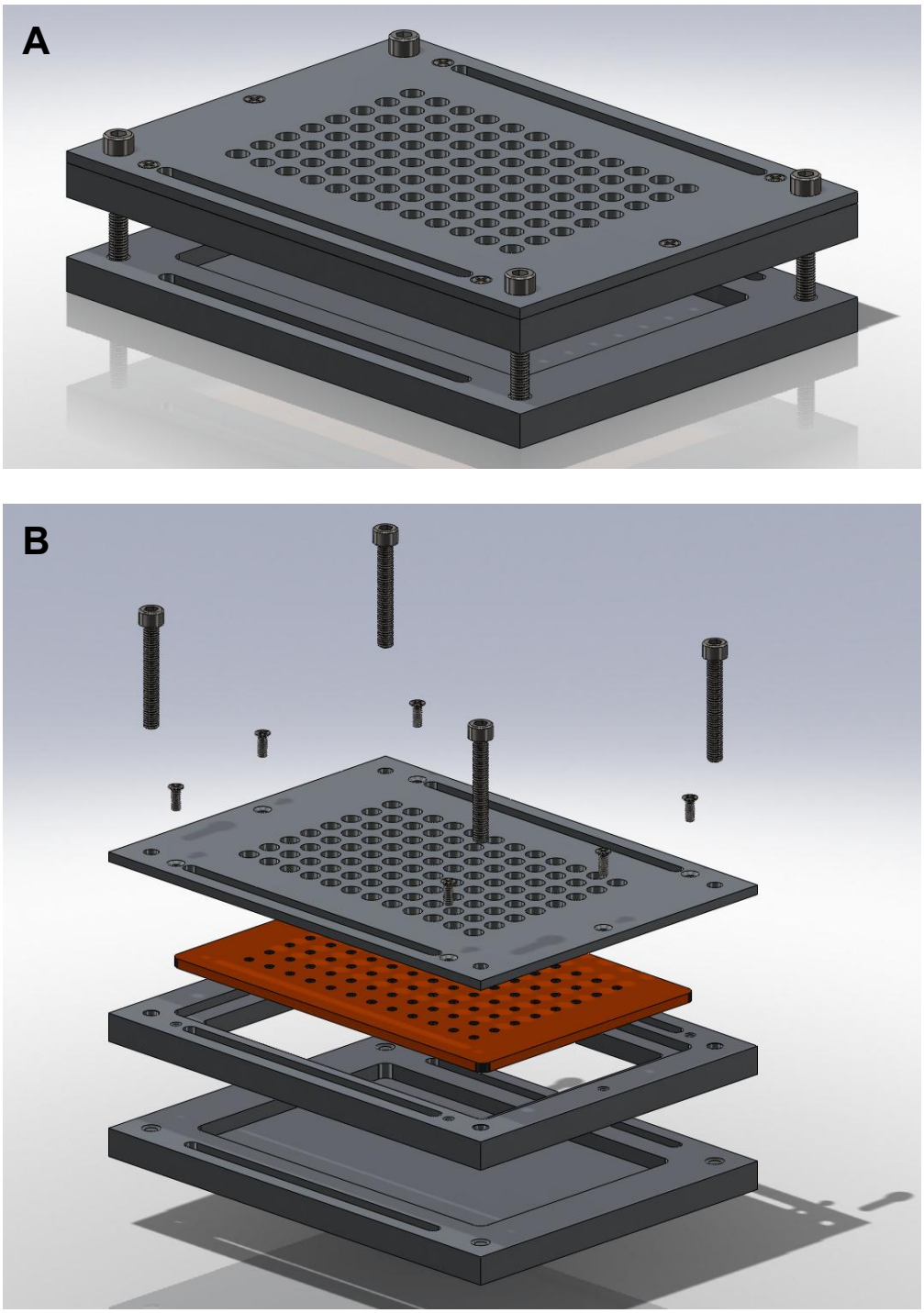


Figure 5-3. Growth apparatus for auto-induction of *E. coli* cells

Layers are machined from 12.4 mm × 16.7 mm aluminum sheets and are anodized black to finish. In the assembled view (A) a 96-well microtiter plate (not shown) is mounted between the top assembly and the base plate by long screws (shown). This assembly can then be mounted to a shaker platform by screws (not shown). In the exploded view (B) a layer of glass wool is inserted between the top plate and the red silicone rubber layer prior to assembly and autoclave sterilization. Drawings courtesy of Todd Prox, Department of Chemistry Machine Shop, University of Florida.

Table 5-1. Libraries screened in this study.

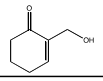
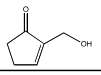
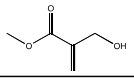
Library	Q_{pooled}	Sequenced	Complete
T35X	0.66	Yes	No
F37X	0.80	Yes	Yes*
A68X	0.62	No	Unknown
Y78X	0.75	No	Unknown
I113X	0.69	Yes	Yes
H188X	0.72	Yes	No
H191X	0.84	Yes	No
Y193X	0.63	Yes	Yes
F247X	0.76	Yes	No
G292X	0.72	No	Unknown
N293X	0.76	Yes	No
V294X	0.73	Yes	No
F373X	0.64	Yes	No
Y374X	0.62	Yes	No
Y78W, I113X	0.69	Yes	Yes
Y78W, F247X	0.63	No	Unknown
Y78W, V294X	0.77	No	Unknown
I113D, Y78X	0.87	No	Unknown
I113D, V294X	0.72	No	Unknown
I113W, Y78X	0.72	No	Unknown
I113W, V294X	0.76	No	Unknown
V294P, Y78X	0.90	No	Unknown
V294P, I113X	0.78	No	Unknown

*F37X was the only library to provide all 20 amino acids by NNK saturation mutagenesis alone. All other libraries were completed by site directed mutagenesis where indicated.

Table 5-2. Selected first round screening results.

OYE 2.6	Substrate 1		Substrate 2	
mutant	% conv.	% ee	% conv.	% ee
Y78W			99	63 (*R*)
I113W			20	11 (*R*)
Y193C	49	10 (*R*)	26	2 (*R*)
Y193T	15	12 (*R*)		
I113D	20	13 (*S*)		
V294P	11	61 (*S*)		
I113C	99	81 (*S*)	99	83 (*S*)
F247Y			99	84 (*S*)
F247W			99	86 (*S*)

Substrate **3** did not display altered enantioselectivity in any of the first round library screenings.

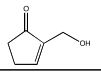
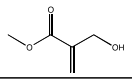
T35X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K						
-	D	-	-	-	-	-	-
	E	-	-	<5%	n.d	-	-
polar	C	14%	>95%	28%	>90%	-	-
	M	13%	>95%	12%	>90%	-	-
	H	-	-	-	-	-	-
	N	-	-	-	-	-	-
	Q	10%	>95%	18%	>90%	-	-
	S	100%	>95%	100%	>90%	13%	>95%
	T	100%	>95%	100%	>90%	61%	>95%
non-polar	P	-	-	-	-	-	-
	G	-	-	-	-	-	-
	A	45%	>95%	90%	>90%	-	-
	V	5%	>95%	<5%	n.d	-	-
	L	12%	>95%	12%	>90%	-	-
	I	-	-	<5%	n.d	-	-
aromatic	F	-	-	-	-	-	-
	Y	-	-	-	-	-	-
	W						

F37X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	9%	n.d.	12%	>95%	-	-
	K	57%	>95%	31%	>95%	-	-
-	D	-	-	-	-	-	-
	E	6%	n.d.	29%	>95%	-	-
polar	C	100%	>95%	100%	>95%	8%	>95%
	M	100%	>95%	100%	>95%	15%	>95%
	H	100%	>95%	58%	>95%	-	-
	N	29%	>95%	66%	>95%	-	-
	Q	26%	>95%	45%	>95%	-	-
	S	19%	>95%	28%	>95%	-	-
	T	53%	>95%	61%	>95%	-	-
non-polar	P	-	-	-	-	-	-
	G	43%	>95%	76%	>95%	-	-
	A	32%	>95%	30%	>95%	-	-
	V	57%	>95%	59%	>95%	-	-
	L	100%	>95%	100%	>95%	25%	>95%
	I	41%	>95%	59%	>95%	-	-
aromatic	F	100%	>95%	100%	>95%	70%	>95%
	Y	100%	>95%	100%	>95%	6%	>95%
	W	100%	>95%	90%	>95%	-	-

I113X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K	-	-	-	-	-	-
-	D	20%	**13% (*S*)**	<5%	n.d.	-	-
	E	10%	n.d.	9%	n.d.	-	-
polar	C	100%	**81% (*S*)**	100%	**83% (*S*)**	28%	>95%
	M	100%	>95%	100%	>90%	63%	>95%
	H	100%	>95%	100%	>90%	24%	>95%
	N	100%	>95%	100%	>90%	100%	>95%
	Q	100%	>95%	100%	>90%	92%	>95%
	S	97%	>95%	100%	>90%	23%	>95%
	T	100%	>95%	100%	>90%	38%	>95%
non-polar	P	-	-	-	-	-	-
	G	100%	>95%	100%	>90%	11%	>95%
	A	100%	>95%	100%	>90%	69%	>95%
	V	97%	>95%	100%	>90%	61%	>95%
	L	99%	>95%	100%	>90%	36%	>95%
	I	100%	>95%	100%	>90%	75%	>95%
aromatic	F	92%	>95%	64%	>90%	35%	>95%
	Y	20%	>95%	97%	>90%	-	-
	W	3%	>95%	20%	**11% (*R*)**	-	-

Figure 5-4. First round screening results for T35X, F37X and I113X

Shaded blocks indicate residues missing from the library. Bold text indicates deviation from wild type observations. N.d. = not determined. All e.e.% are (*S*) unless otherwise indicated.

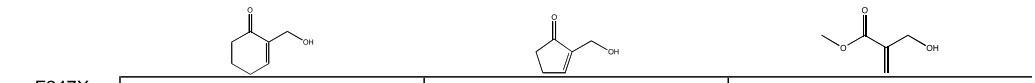
H188X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K	-	-	-	-	-	-
-	D	<5%	n.d	-	-	-	-
	E						
polar	C	-	-	-	-	-	-
	M	<5%	n.d	-	-	-	-
	H	100%	>95%	100%	>90%	83%	>95%
	N	13%	<75%	6%	n.d.	-	-
	Q	10%	<75%	-	-	-	-
	S	17%	<75%	9%	n.d.	-	-
	T	<5%	n.d	-	-	-	-
non-polar	P	-	-	-	-	-	-
	G	<5%	n.d	-	-	-	-
	A	<5%	n.d	-	-	-	-
	V	-	-	-	-	-	-
	L	-	-	-	-	-	-
	I	-	-	-	-	-	-
aromatic	F	-	-	-	-	-	-
	Y	-	-	-	-	-	-
	W	-	-	-	-	-	-

H191X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K						
-	D	-	-	-	-	-	-
	E						
polar	C	<5%	n.d	-	-	-	-
	M	-	-	-	-	-	-
	H	100%	>95%	100%	>90%	83%	>95%
	N	73%	>95%	75%	>90%	<5%	n.d.
	Q						
	S	9%	n.d.	6%	n.d.	-	-
	T	-	-	-	-	-	-
non-polar	P	-	-	-	-	-	-
	G	10%	n.d.	8%	n.d.	-	-
	A	<5%	n.d	-		-	-
	V	-	-	-	-	-	-
	L	-	-	-	-	-	-
	I						
aromatic	F	-	-	-	-	-	-
	Y	-	-	-	-	-	-
	W	-	-	-	-	-	-

Y193X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K	-	-	-	-	-	-
-	D	-	-	-	-	-	-
	E	-	-	-	-	-	-
polar	C	49%	**10% (*R*)**	26%	**2% (*R*)**	-	-
	M	-	-	-	-	-	-
	H	-	-	-	-	-	-
	N	-	-	-	-	-	-
	Q	-	-	-	-	-	-
	S	-	-	-	-	-	-
	T	15%	**12% (*R*)**	-	-	-	-
non-polar	P	-	-	-	-	-	-
	G	-	-	-	-	-	-
	A	-	-	-	-	-	-
	V	-	-	-	-	-	-
	L	-	-	-	-	-	-
	I	-	-	-	-	-	-
aromatic	F	-	-	-	-	-	-
	Y	100%	>95%	100%	>90%	100%	>95%
	W	-	-	-	-	-	-

Figure 5-5. First round screening results for H188X, H191X and Y193X

Shaded blocks indicate residues missing from the library. Bold text indicates deviation from wild type observations. N.d. = not determined. All e.e.% are (*S*) unless otherwise indicated.



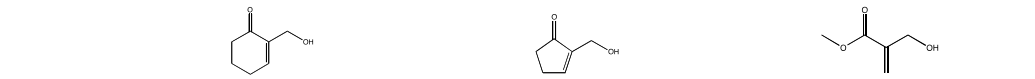
F247X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	12%	>95%	12%	>90%	-	-
	K	29%	>95%	13%	>90%	-	-
-	D	16%	>95%	-	-	-	-
	E	23%	>95%	20%	>90%	-	-
polar	C	90%	>95%	77%	>90%	<5%	n.d.
	M	98%	>95%	96%	>90%	8%	n.d.
	H	100%	>95%	100%	>90%	25%	>95%
	N						
	Q	59%	>95%	32%	>90%	-	-
	S	69%	>95%	42%	>90%	<5%	n.d.
	T	70%	>95%	23%	>90%	-	-
non-polar	P	-	-	-	-	-	-
	G	38%	>95%	32%	>90%	-	-
	A	98%	>95%	77%	>90%	<5%	n.d.
	V	63%	>95%	33%	>90%	-	-
	L	59%	>95%	41%	>90%	-	-
	I	64%	>95%	33%	>90%	-	-
aromatic	F	100%	>95%	100%	>90%	70%	>95%
	Y	100%	>95%	100%	**84% (*S*)**	100%	>95%
	W	100%	>95%	100%	**86% (*S*)**	55%	>95%

N293X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	100%	>95%	22%	>90%	<5%	n.d.
	K	100%	>95%	28%	>90%	23%	>95%
-	D	100%	>95%	100%	>90%	100%	>95%
	E	100%	>95%	100%	>90%	42%	>95%
polar	C	100%	>95%	78%	>90%	62%	>95%
	M	100%	>95%	37%	>90%	69%	>95%
	H	100%	>95%	100%	>90%	51%	>95%
	N	100%	>95%	100%	>90%	81%	>95%
	Q	100%	>95%	69%	>90%	39%	>95%
	S	100%	>95%	95%	>90%	54%	>95%
	T	100%	>95%	96%	>90%	100%	>95%
non-polar	P	100%	>95%	82%	>90%	26%	>95%
	G	100%	>95%	85%	>90%	16%	>95%
	A	100%	>95%	61%	>90%	22%	>95%
	V	100%	>95%	81%	>90%	85%	>95%
	L	100%	>95%	65%	>90%	100%	>95%
	I	100%	>95%	94%	>90%	76%	>95%
aromatic	F						
	Y						
	W	100%	>95%	63%	>90%	100%	>95%

V294X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	100%	>95%	100%	>90%	47%	>95%
	K	100%	>95%	100%	>90%	52%	>95%
-	D	100%	>95%	83%	>90%	48%	>95%
	E	100%	>95%	96%	>90%	88%	>95%
polar	C	100%	>95%	100%	>90%	48%	>95%
	M	98%	>95%	100%	>90%	68%	>95%
	H						
	N						
	Q	100%	>95%	83%	>90%	47%	>95%
	S	100%	>95%	75%	>90%	27%	>95%
	T	100%	>95%	78%	>90%	47%	>95%
non-polar	P	11%	**61% (*S*)**	-	-	-	-
	G	100%	>95%	78%	>90%	5%	n.d.
	A	100%	>95%	79%	>90%	46%	>95%
	V	100%	>95%	100%	>90%	85%	>95%
	L	100%	>95%	88%	>90%	69%	>95%
	I						
aromatic	F	100%	>95%	93%	>90%	29%	>95%
	Y	100%	>95%	100%	>90%	34%	>95%
	W	-	-	-	-	-	-

Figure 5-6. First round screening results for F247X, N293X and V294X

Shaded blocks indicate residues missing from the library. Bold text indicates deviation from wild type observations. N.d. = not determined. All e.e.% are (*S*) unless otherwise indicated.



F373X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K						
-	D	7%	n.d.	-	-	-	-
	E	-	-	-	-	-	-
polar	C						
	M	100%	>95%	31%	>90%	-	-
	H	11%	>95%	-	-	-	-
	N						
	Q	13%	>95%	-	-	-	-
	S	14%	>95%	-	-	-	-
	T	5%	n.d.	-	-	-	-
non-polar	P	-	-	-	-	-	-
	G	-	-	-	-	-	-
	A	10%	n.d.	-	-	-	-
	V	10%	n.d.	-	-	-	-
	L	73%	>95%	31%	>90%	<5%	n.d.
	I						
aromatic	F	100%	>95%	100%	>90%	100%	>95%
	Y	100%	>95%	19%	>90%	-	-
	W	64%	>95%	25%	>90%	34%	>95%

Y374X		conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	44%	>95%	8%	n.d.	-	-
	K	75%	>95%	9%	n.d.	-	-
-	D	25%	>95%	10%	n.d.	-	-
	E	59%	>95%	14%	>90%	-	-
polar	C						
	M						
	H	33%	>95%	9%	>90%	-	-
	N						
	Q	83%	>95%	28%	>90%	-	-
	S	100%	>95%	16%	>90%	<5%	n.d.
	T	100%	>95%	21%	>90%	<5%	n.d.
non-polar	P	7%	n.d.	-	-	-	-
	G	51%	>95%	11%	>90%	-	-
	A	100%	>95%	14%	>90%	-	-
	V	54%	>95%	17%	>90%	-	-
	L	53%	>95%	20%	>90%	7%	n.d.
	I						
aromatic	F						
	Y	100%	>95%	100%	>90%	100%	>95%
	W	49%	>95%	13%	>90%	10%	n.d.

Figure 5-7. First round screening results for F373X and Y374X

Shaded blocks indicate residues missing from the library. Bold text indicates deviation from wild type observations. N.d. = not determined. All e.e.% are (*S*) unless otherwise indicated.

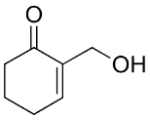
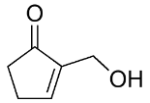
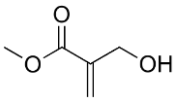
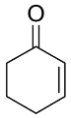
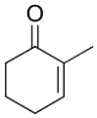
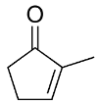
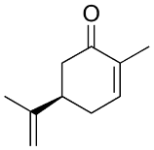
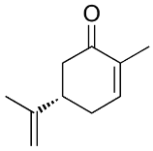
	wild type		Y193C		Y193T	
	% conv.	% ee (de)	% conv.	% ee (de)	% conv.	% ee (de)
	99	>95 (*S*)	50	3 (S)	44	5 (S)
	99	92 (*S*)	15	2 (R)	6	n.d.
	99	>95 (*S*)	-	-	-	-
	99	n/a	14	n/a	43	n/a
	99	>95 (*R*)	21	17 (R)	13	50 (R)
	99	97 (*R*)	3	n.d.	0	-
	99	93 (1*R*,4*R*)	15	80 (1*R*,4*R*)	21	78 (1*R*,4*R*)
	99	92 (1*R*,4*S*)	12	70 (1*R*,4*S*)	20	68 (1*R*,4*S*)

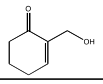
Figure 5-8. Selected substrate screening results for OYE 2.6 Tyr193 mutants

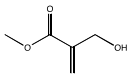
Screening reactions were conducted under conditions described in Chapter 1. Substrate loading is 10 mM in all trials.

Table 5-3. Summary second round screening results.

Anchor Mutation	% ee*	Substrate*	2nd round randomized	Screened substrate 1	2	3
Y78W	63 (*R*)	2	I113X	n.c.	varied	varied
			F247X	n.c.	varied	low conv.
			V294X	n.c.	n.c.	-
I113W	11 (*R*)	2	Y78X	n.c.	varied	-
			V294X	-	low conv.	-
I113D	13 (*S*)	1	Y78X	-	-	-
			V294X	-	-	-
V294P	61 (*S*)	1	Y78X	-	low conv.	-
			I113X	low conv.	-	-

% ee listed is from first round screening for substrates indicated. Varied results from the Y78W anchor are depicted in Figures 5-9, 5-10, and 5-11. Varied results from the I113W anchor result in no improvements in % ee and were not sequenced. Dashes (-) indicate complete loss of activity. n.c. = no change observed in enantioselectivity across the screened library. low conv. = less than 5% conversion in any screened well, insufficient to determine % ee. *Results from first round screening.





	Y78W-I113X	conv.	e.e.	conv.	e.e.	conv.	e.e.
+	R	-	-	-	-	-	-
	K	-	-	-	-	-	-
-	D	-	-	-	-	-	-
	E	-	-	-	-	-	-
polar	C	42%	**79% (*S*)**	25%	**33% (*R*)**	21%	**>90% (*R*)**
	M	47%	>95%	99%	**73% (*R*)**	1%	n.d.
	H	79%	>95%	99%	**52% (*S*)**	56%	**44% (*S*)**
	N	66%	>95%	99%	**59% (*S*)**	32%	**23% (*R*)**
	Q	68%	>95%	66%	**40% (*S*)**	11%	>95%
	S	29%	>95%	34%	**47% (*S*)**	2%	n.d.
	T	18%	>95%	41%	**9% (*S*)**	6%	n.d.
non-polar	P	-	-	-	-	-	-
	G	15%	>95%	16%	>90% (*S*)	-	-
	A	29%	>95%	33%	**42% (*S*)**	6%	n.d.
	V	37%	>95%	79%	**46% (*R*)**	45%	**>90% (*R*)**
	L	28%	>95%	78%	**34% (*R*)**	21%	**>90% (*R*)**
	I	40%	>95%	99%	**73% (*R*)**	-	-
aromatic	F	15%	>95%	60%	**21% (*R*)**	14%	**>90% (*R*)**
	Y	-	-	-	-	-	-
	W	-	-	-	-	-	-

Figure 5-9. Second round screening results for Y78W-I113X

Shaded blocks indicate residues missing from the library. Bold text indicates deviation from wild type observations. N.d. = not determined. All e.e.% are (*S*) unless otherwise indicated.

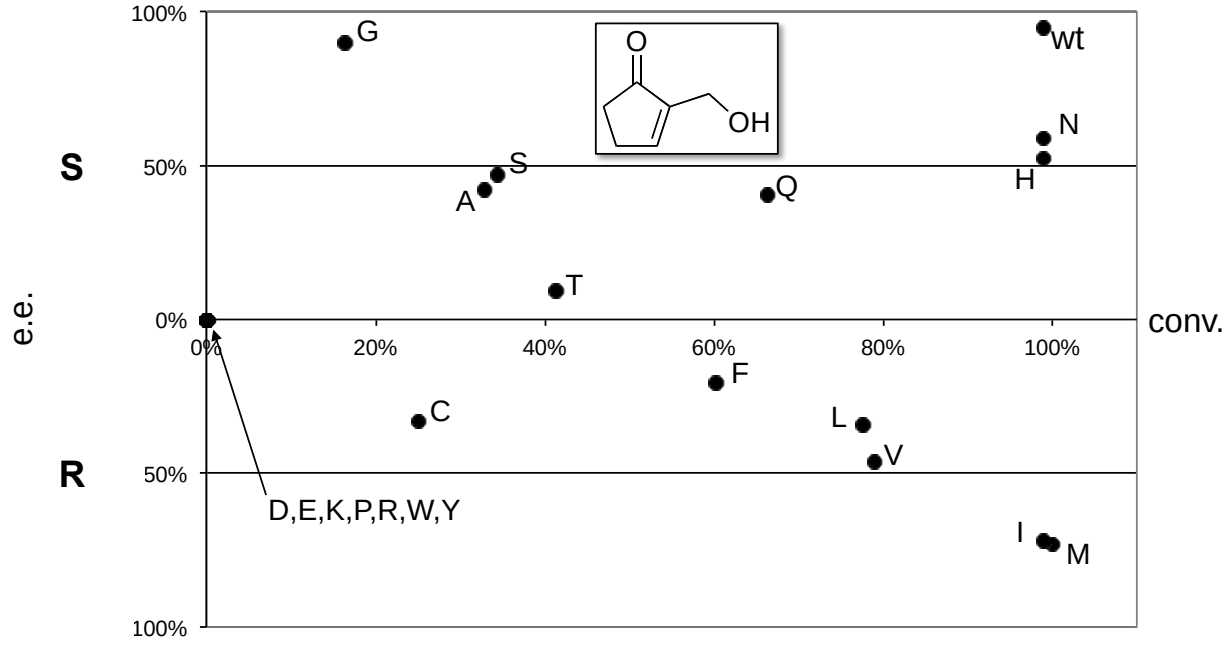


Figure 5-10. Results of Y78W-I113X screening with 2-(hydroxymethyl)cyclopentenone

Select data from Figure 5-9 with e.e.% on the vertical axis and total conversion on the horizontal axis.

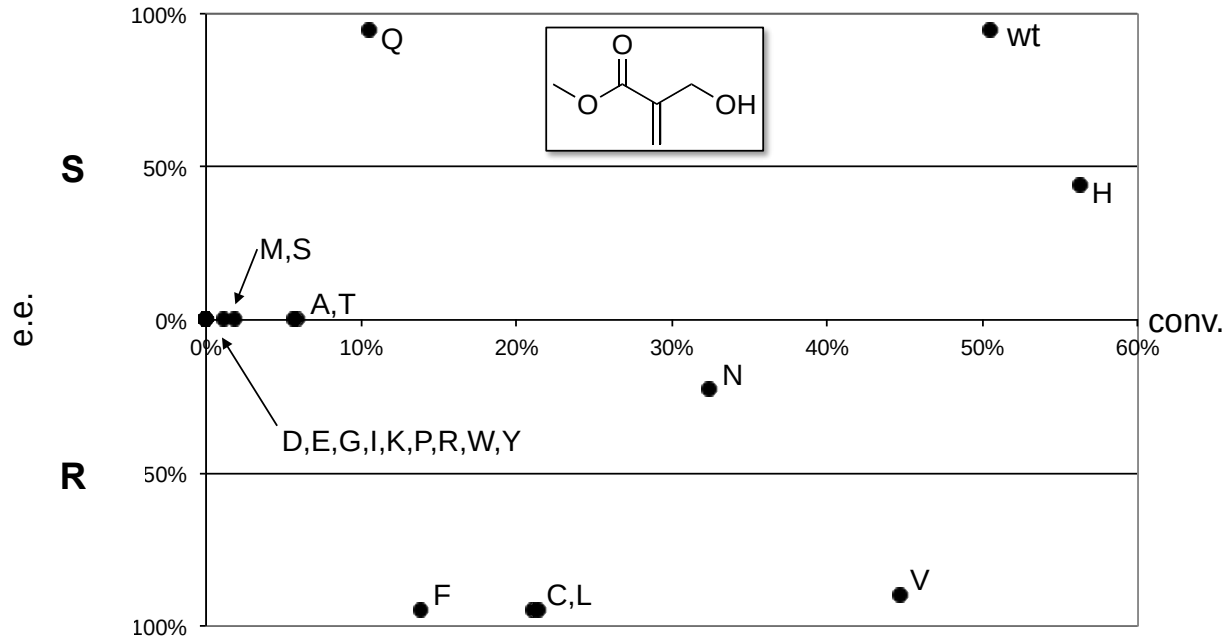


Figure 5-11. Results of Y78W-I113X screening with methyl 2-(hydroxymethyl)acrylate

Select data from Figure 5-9 with e.e.% on the vertical axis and total conversion on the horizontal axis.

Table 5-4. Summary purified enzyme screening results.

Enzyme	Mutation(s)	Substrate	Time (h)	% conv	% ee
OYE 2.6	none	1	<1	99	99 (*S*)
OYE 2.6	none	2	<1	99	95 (*S*)
OYE 2.6	none	3	3	99	99 (*S*)
OYE 2.6	Y78W	2	6	99	75 (*R*)
OYE 2.6	Y78W, I113M	2	8	99	75 (*R*)
OYE 2.6	Y78W, I113C	3	24	57	90 (*R*)
OYE 2.6	Y78W, I113F	3	24	48	90 (*R*)
OYE 2.6	Y78W, I113L	3	24	64	85 (*R*)
OYE 2.6	Y78W, I113V	3	24	47	87 (*R*)
OYE 1	W116V	3	24	52	86 (*R*)

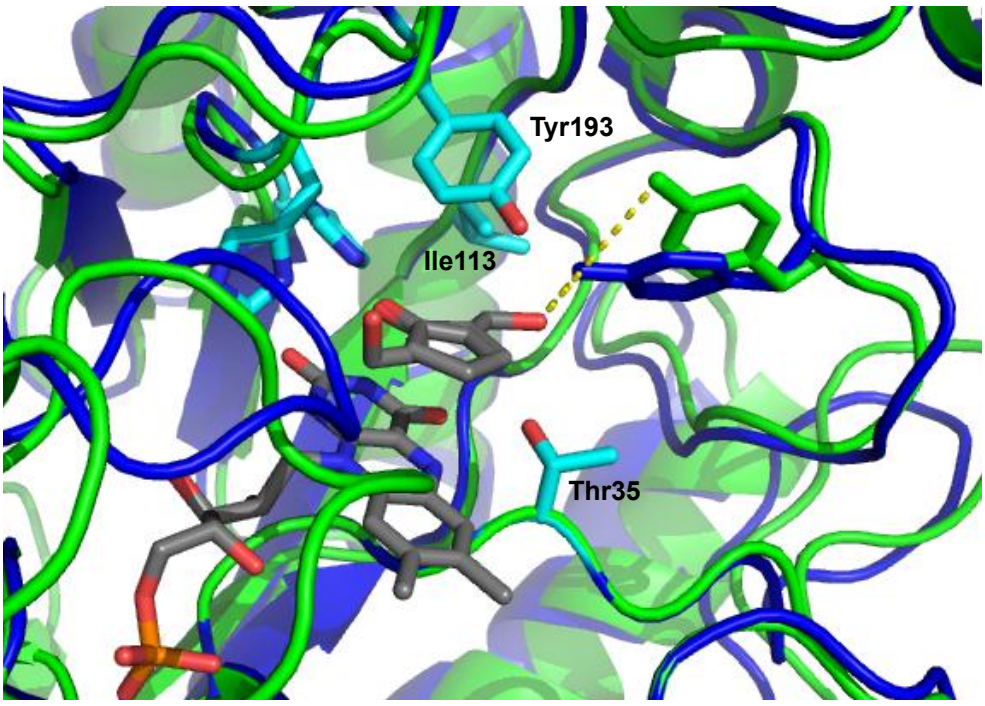


Figure 5-12. The role of Tyr78/Tyr82 in the active site

OYE 2.6 (4DF2) is shown in blue and OYE 1 W116I (3RND) is shown in green. Selected residue side chain carbons are shown in cyan. Ligand carbons are shown in grey. Distance between hydroxymethyl and tyrosyl oxygens is 1.8 Å for OYE 2.6 and 5.4 Å for OYE1 W116I.

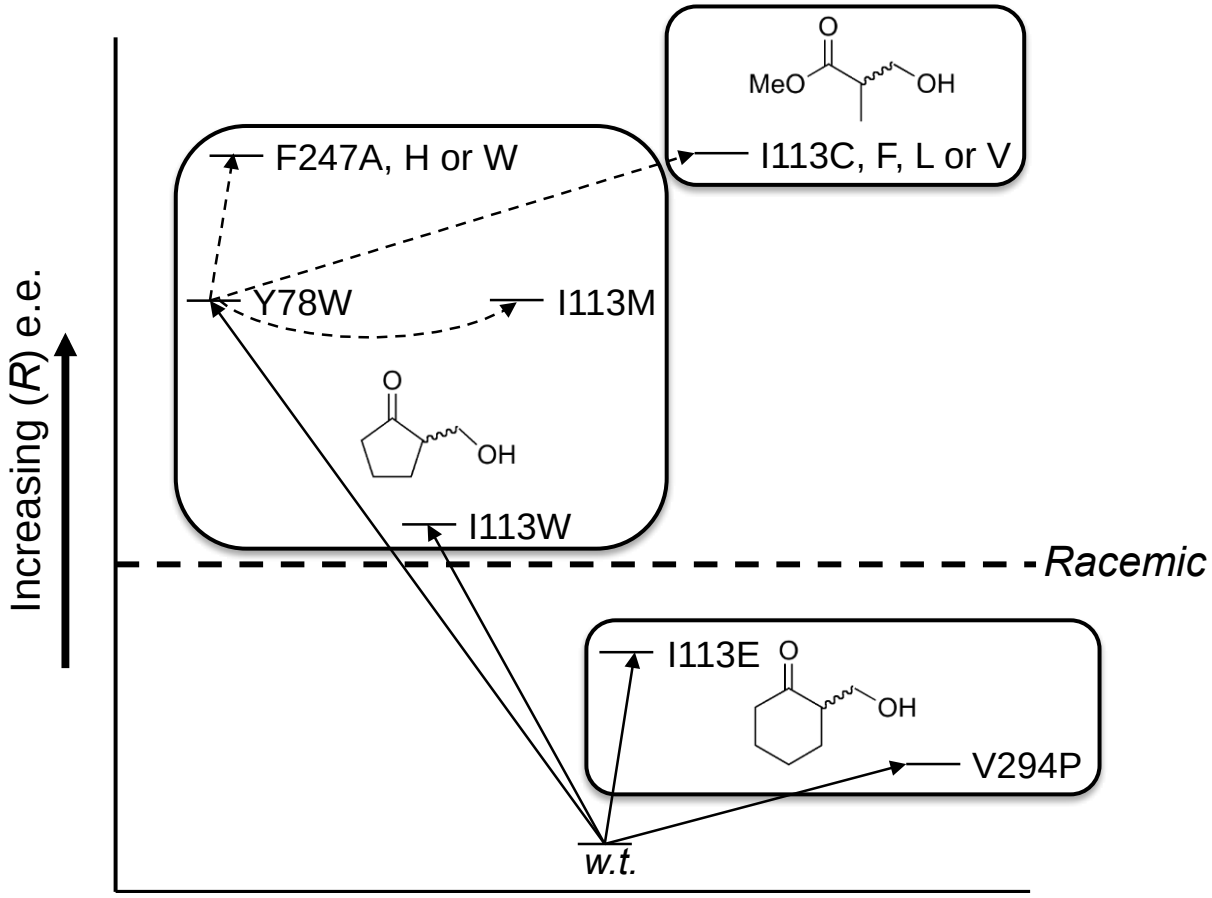


Figure 5-13. Final ISM pathways to observed best results

APPENDIX A
DEGENERATE LIBRARY SEQUENCING DATA

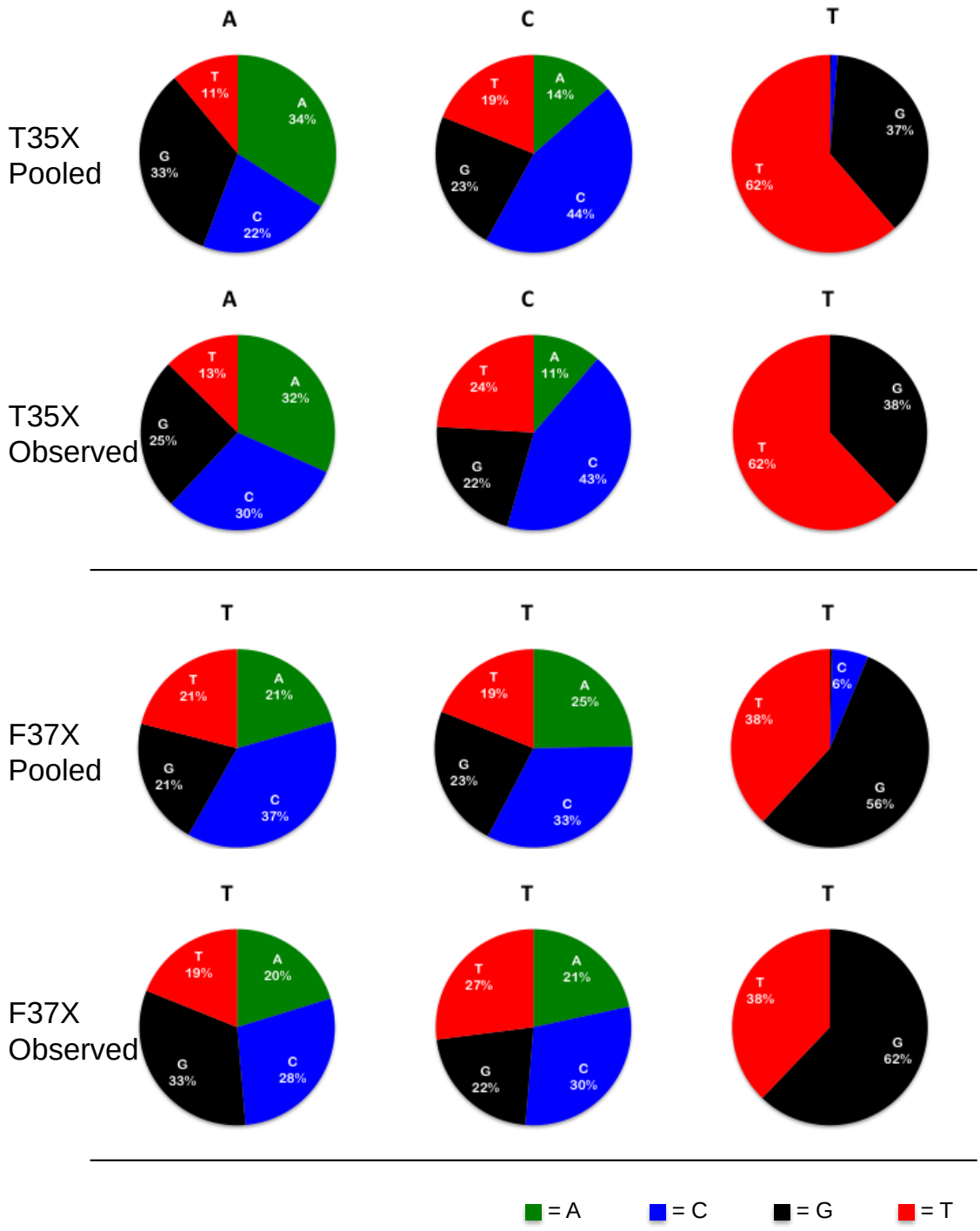


Figure A-1. Pooled and Observed Sequencing for T35X and F37X

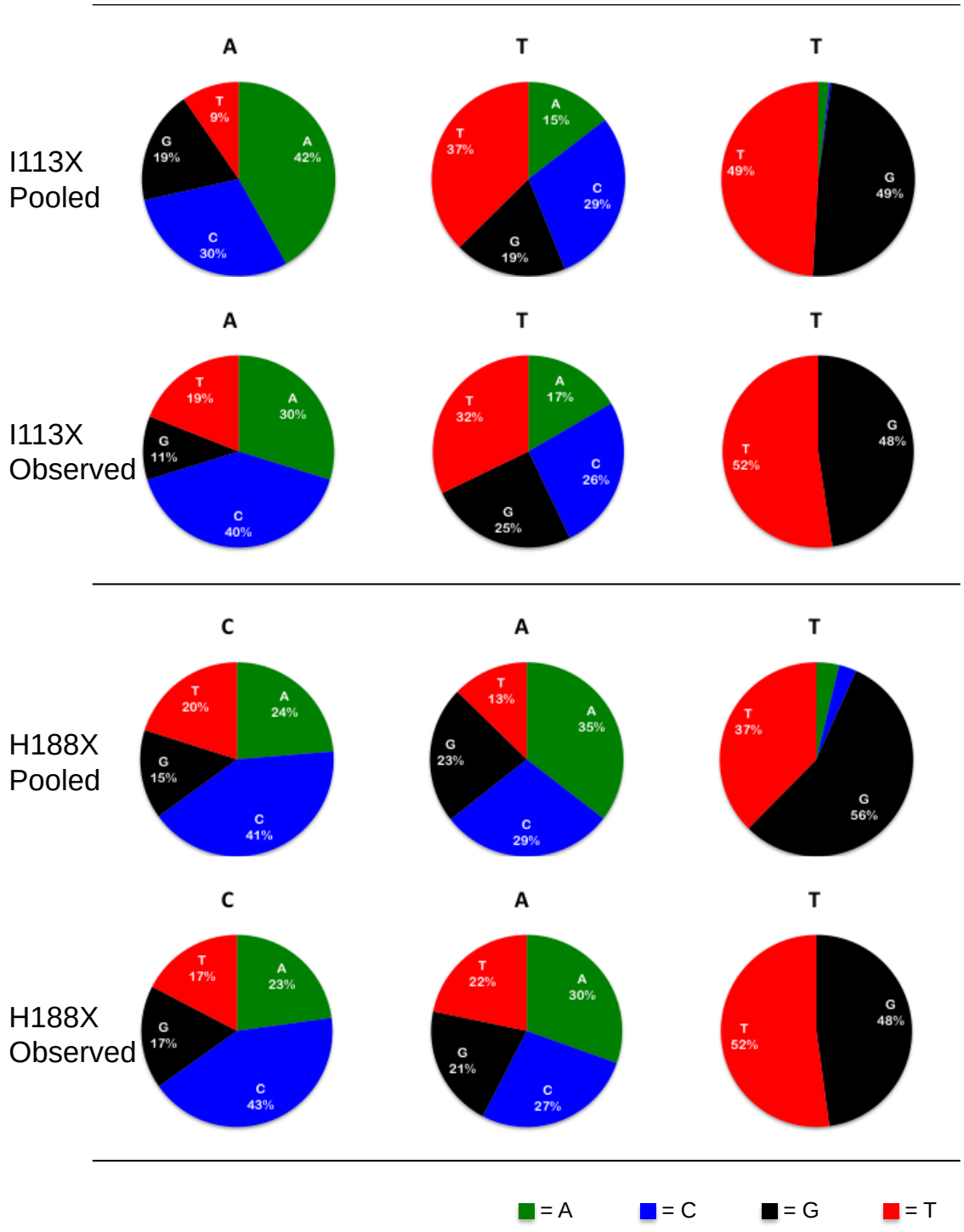


Figure A-2. Pooled and Observed Sequencing for I113X and H188X

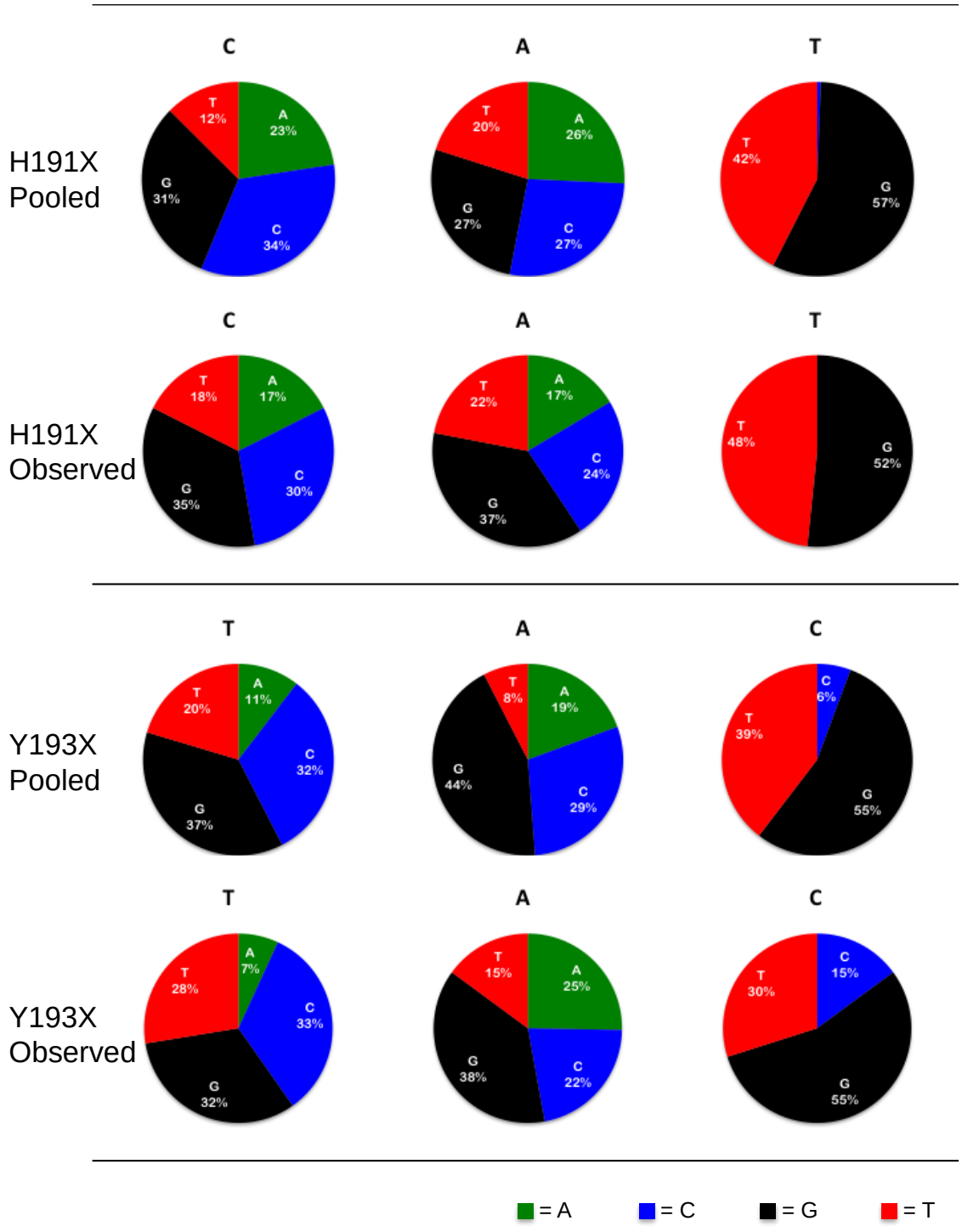


Figure A-3. Pooled and Observed Sequencing for H191X and Y193X

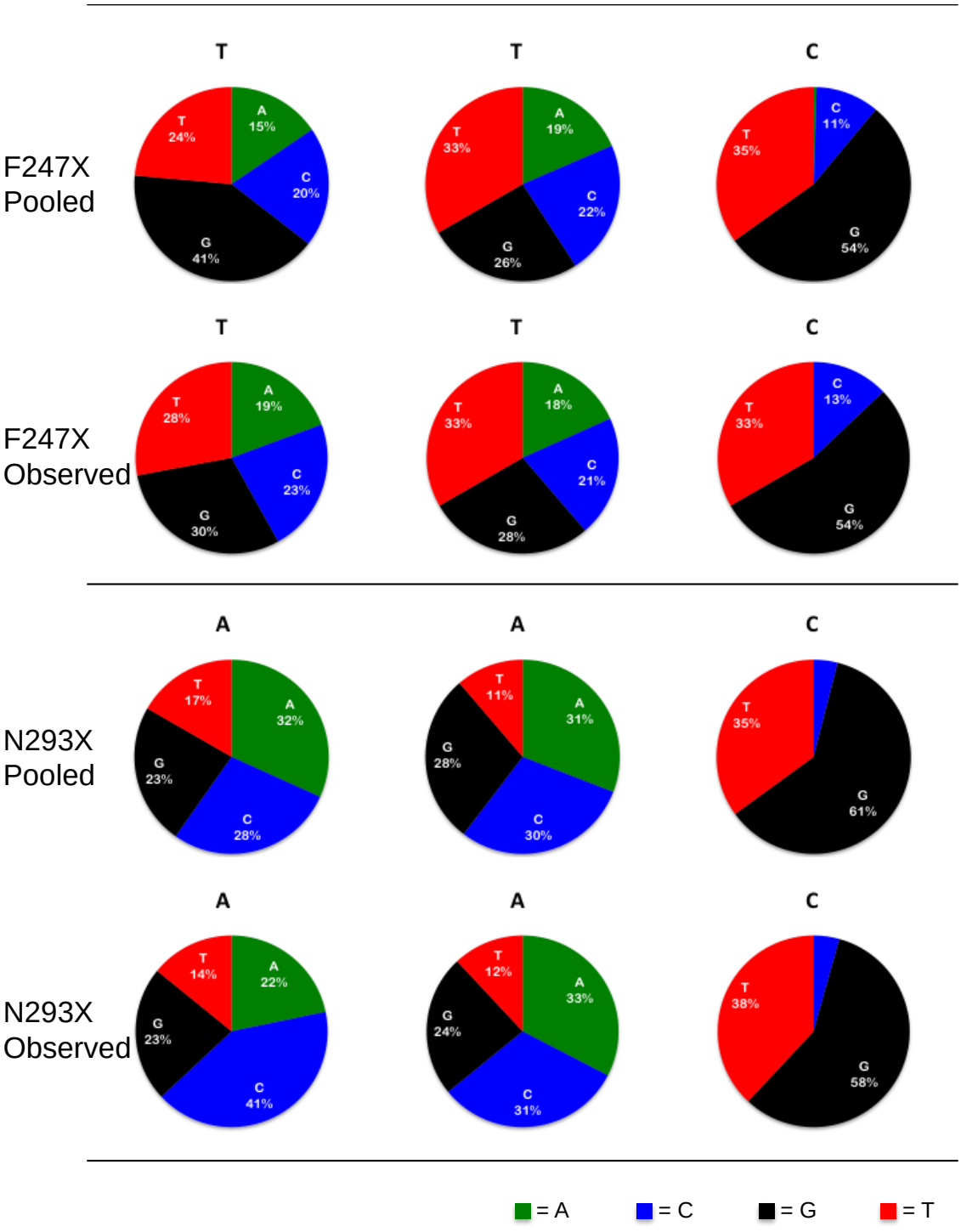


Figure A-4. Pooled and Observed Sequencing for F247X and N293X

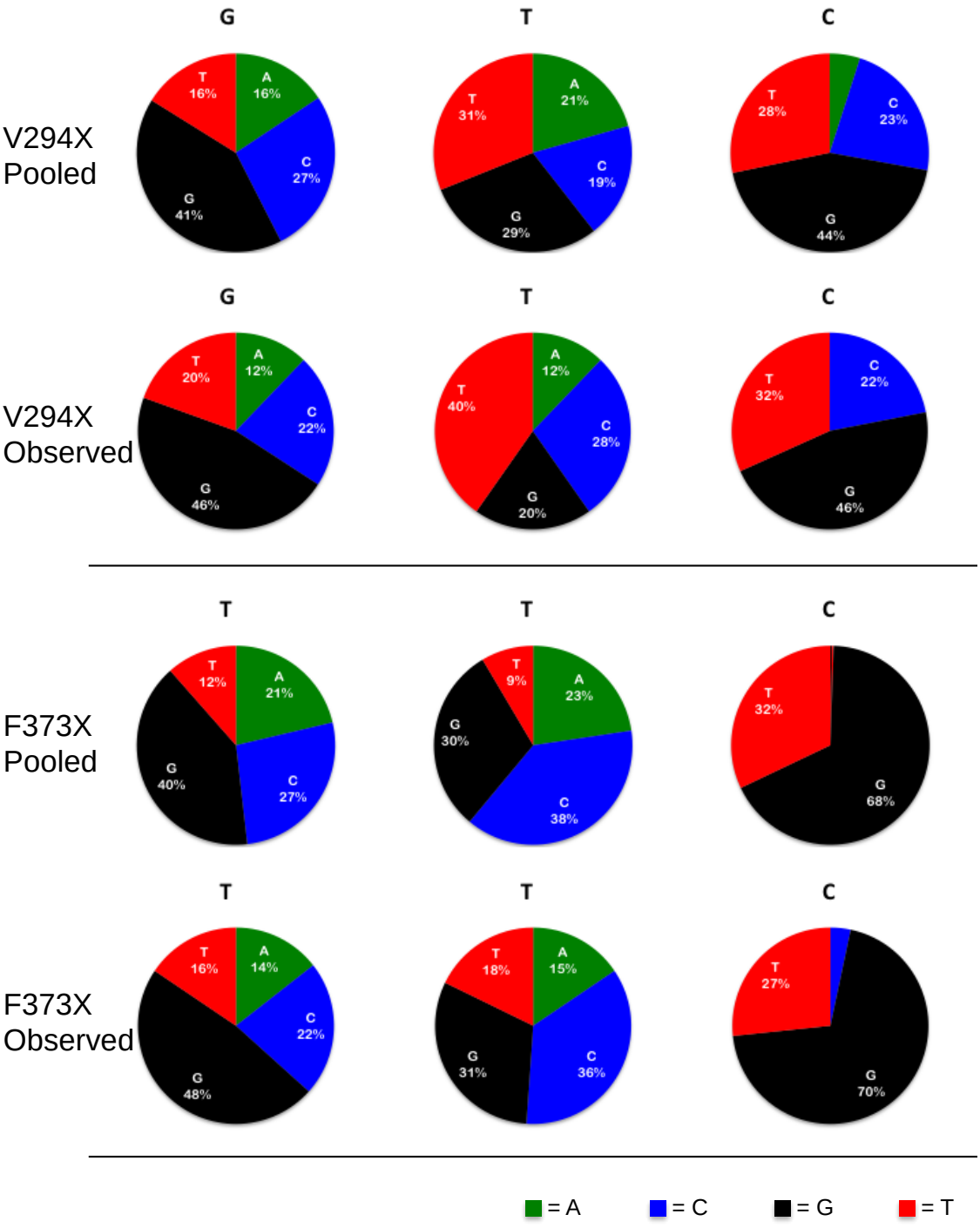


Figure A-5. Pooled and Observed Sequencing for V294X and F373X

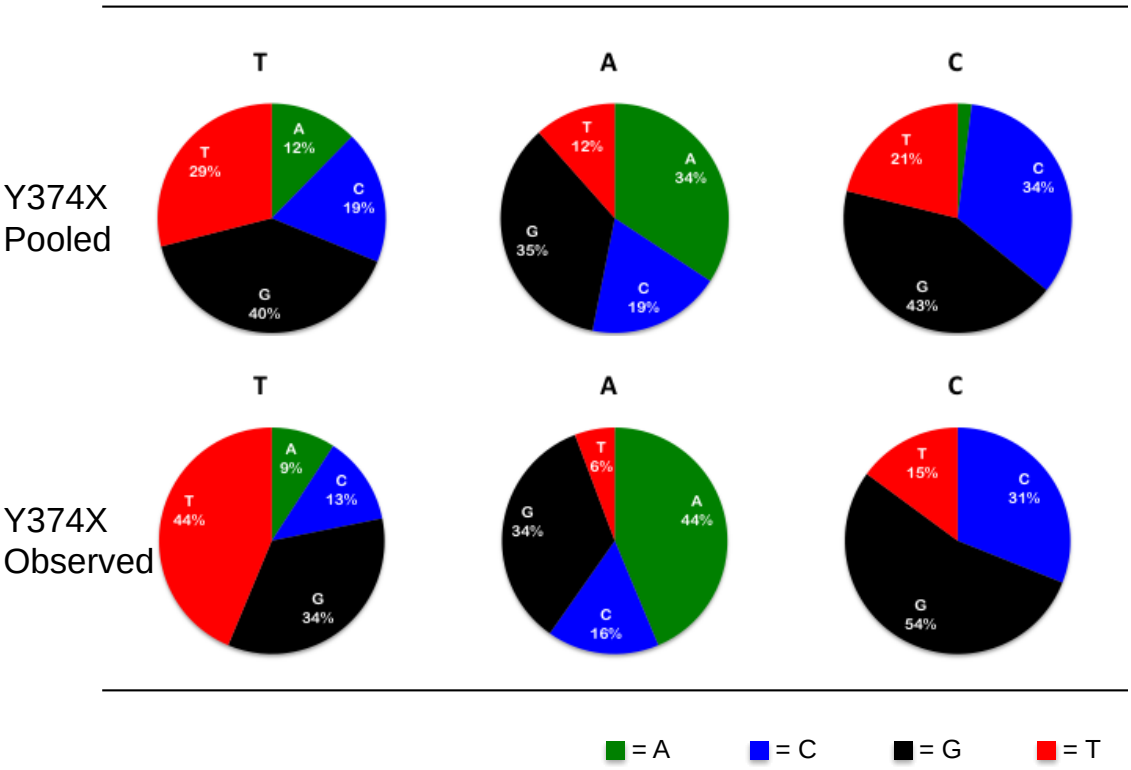


Figure A-6. Pooled and Observed Sequencing for Y374X

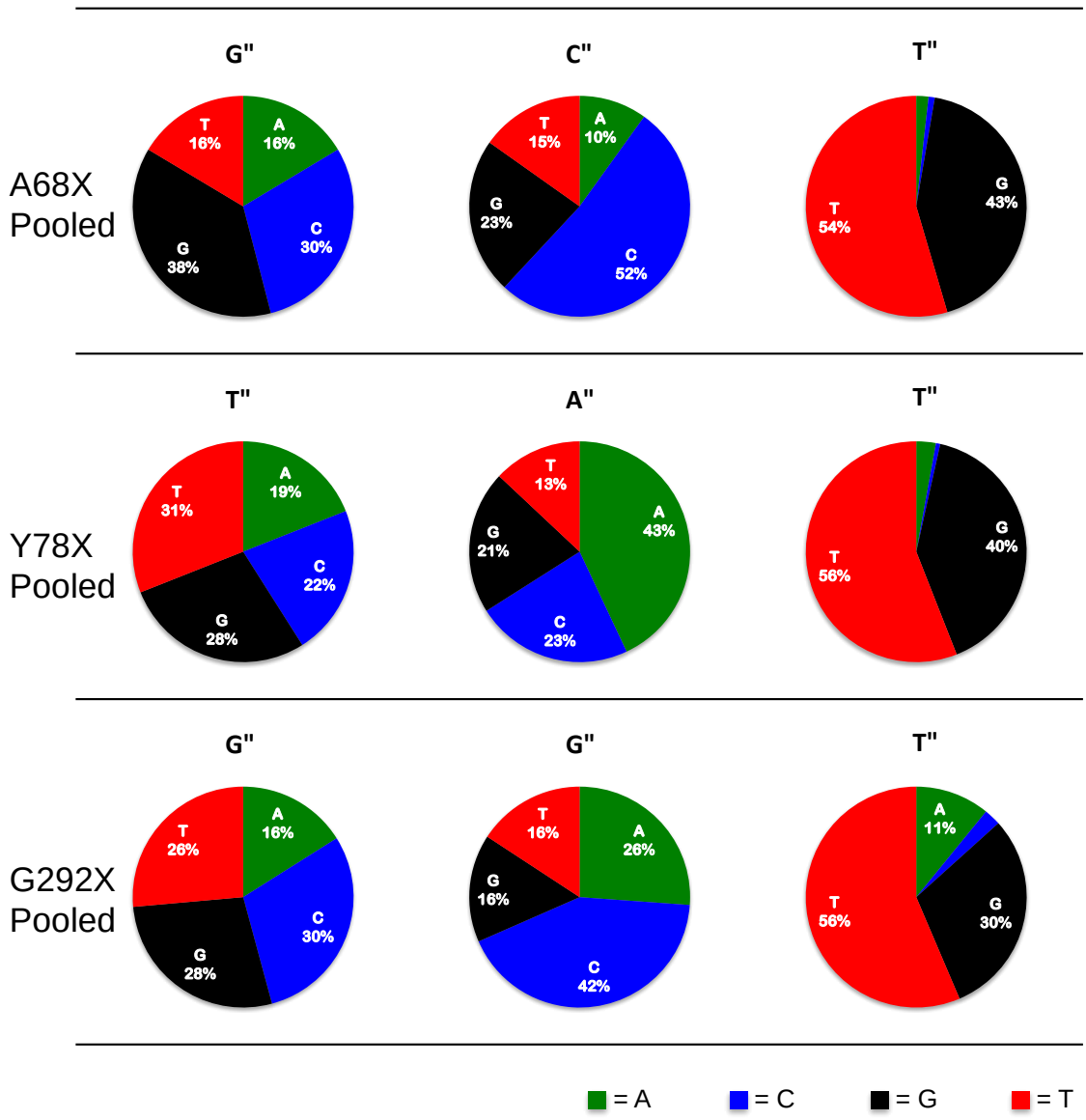


Figure A-7. Pooled Sequencing for A68X, Y78X, and G292X

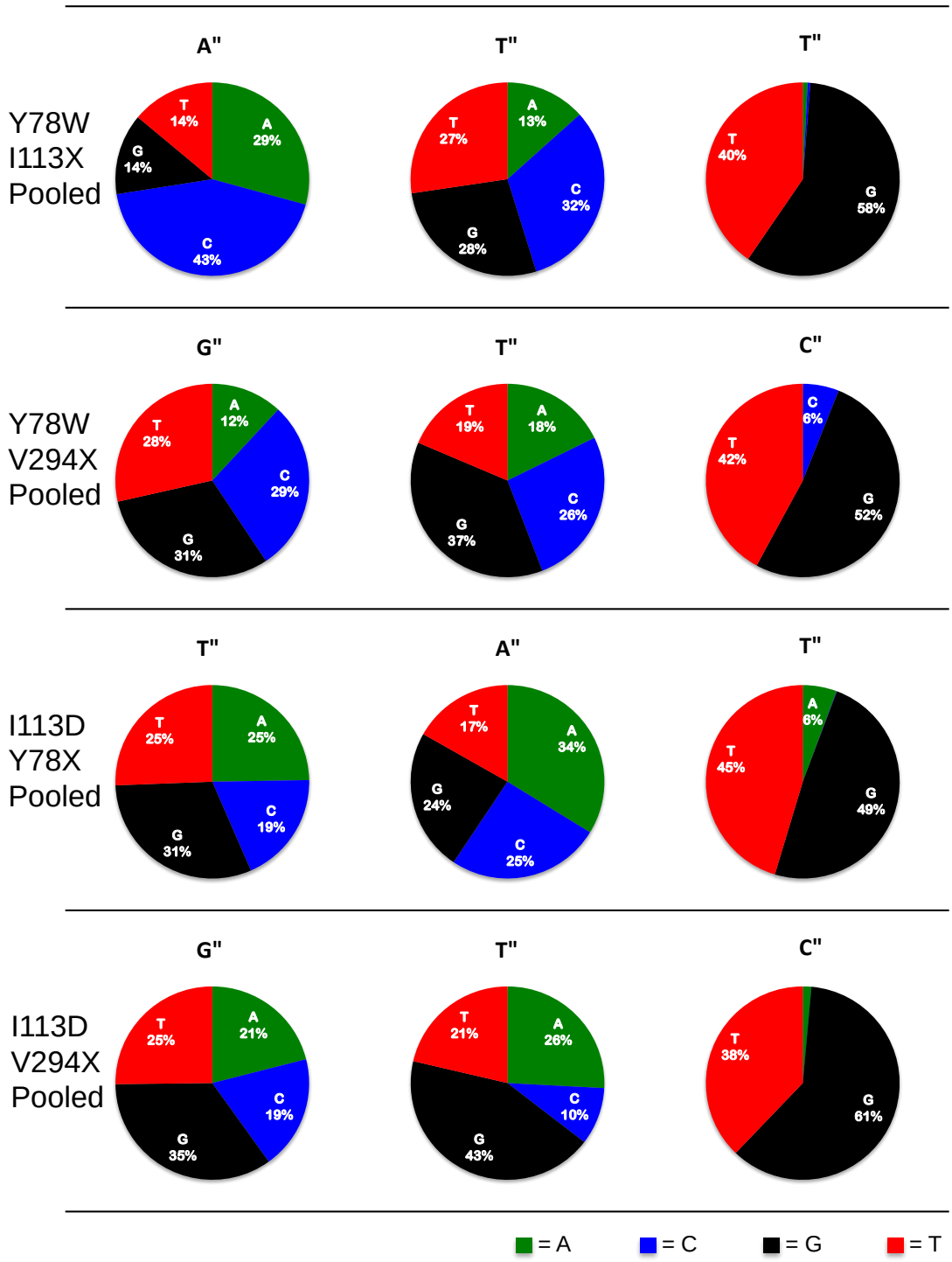


Figure A-8. Pooled Sequencing for Y78W and I113D Second Round Libraries

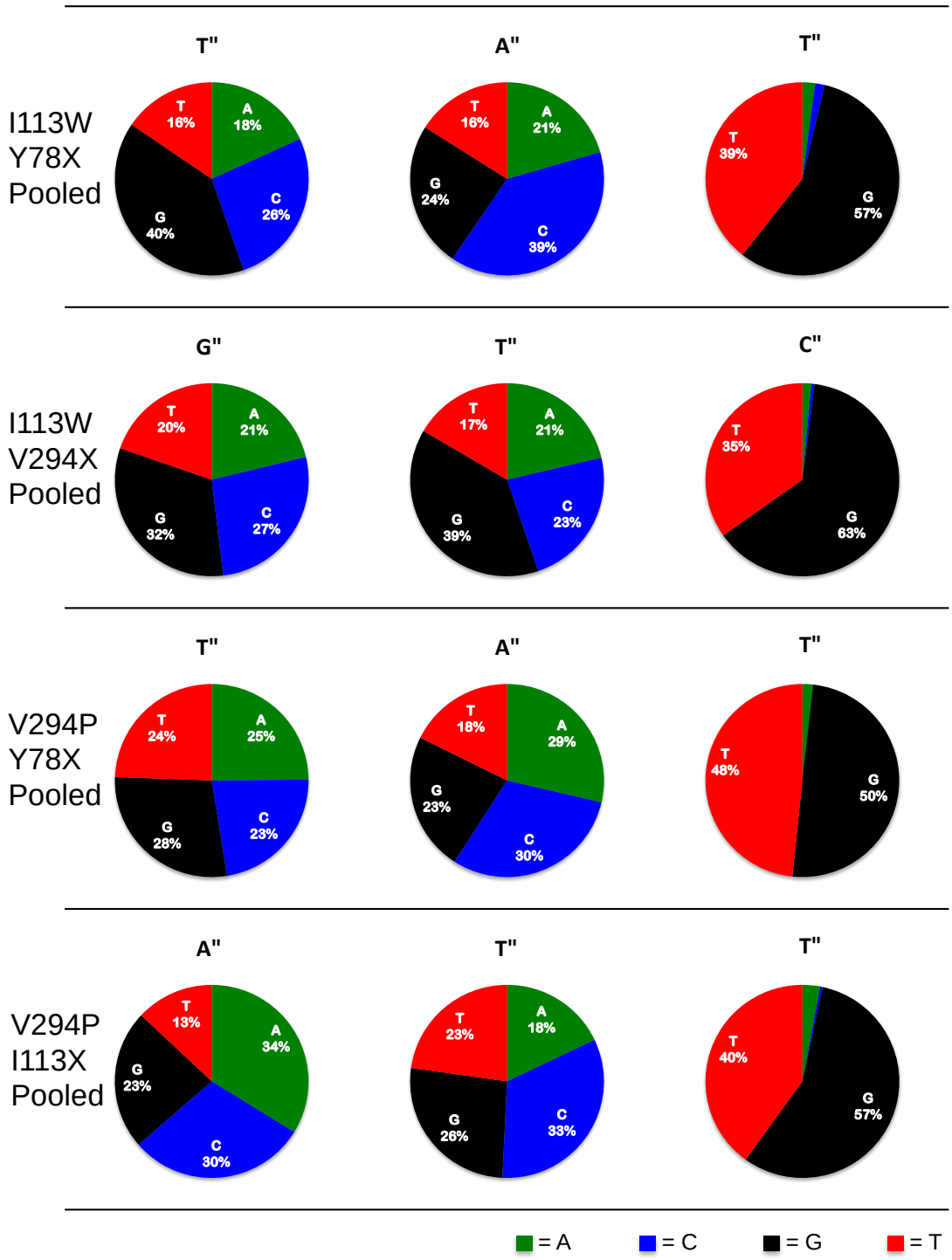


Figure A-9. Pooled Sequencing for I113W and V294P Second Round Libraries

APPENDIX B
SEQUENCE DATA

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGC
GTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCG
CCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAG
TGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCG
CCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCC
GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAA
ATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTT
TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCA
ATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAG
ATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAG
TTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACT
TGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG
CAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGA
CCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG
AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGC
AACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATA
GACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGT
TTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCC
AGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAA
CGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAA
GATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCA
GACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT
TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCT
TTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCG
TAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGT
TACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTT
ACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGA
ACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAG
GGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGT
CGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC
TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAG
CGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATC
TGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAG
TTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCC
AACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTG

Figure B-1. Sequence of pBS2 template plasmid.

(continued)

ACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGC
TGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTC
CAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGG
GCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAA
TGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTT
ACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACT
CAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATC
CTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGA
AACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCG
CTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTA
GCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAAT
GGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGC
AAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGC
CGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCAT
AAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCT
CACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAA
CAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGC
CCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGG
TATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCG
CATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTC
AGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCG
GCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGA
ACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCC
AGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAA
GAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGG
ATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCT
TCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATT
TAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAG
CAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATC
GCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAA
CGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCAC
CACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCG
ATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGT
AGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACA
GTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTG
GCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCG
CCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAAT
ACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAAC
TTTAAGAAGGAGATATACATAATGACCAAGTTACCTATACTAGGTTATTGGAAAATTAAGGGCC
TTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGA
GCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTT
CCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTG

Figure B-1. Sequence of pBS2 template plasmid (continued).

(continued)

ACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGC
GGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAA
GTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAA
CATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGT
TTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATT
GAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGG
GCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCATCTGGTTCCGCGTCA
TATGCCCATGTCTTCAGTCAAAATTTCTCCATTGAAGGATTCTGAAGCATTCCAGTCTATCAAA
GTTGGTAACAACACTCTTCAAACCAAGATTGTCTATCCACCAACTACTAGATTTAGAGCTTTAG
AAGACCACACTCCTTCTGATTTGCAATTGCAGTACTATGGCGACAGATCCACTTTCCCAGGTAC
TTTGCTTATCACTGAAGCTACTTTTGTCTCTCCTCAAGCCTCTGGTTATGAAGGTGCTGCTCCA
GGTATTTGGACTGACAAGCACGCTAAAGCATGGAAGGTTATTACTGATAAAGTTCATGCCAACG
GTTCTTTCGTTTCAACCCAGTTGATTTTTTTGGGAAGGGTTGCAGATCCAGCTGTTATGAAGAC
CCGTGGGTTGAATCCAGTTTCTGCCTCTGCTACTTATGAAAGTGATGCCGCTAAAGAAGCTGCC
GAAGCAGTTGGTAACCCTGTTAGAGCTTTGACTACCCAAGAAGTCAAGGATCTTGTTTACGAGG
CTTACACCAACGCTGCTCAGAAGGCCATGGATGCTGGTTTCGACTATATTGAACTCCATGCTGC
TCATGGCTACCTTTTAGATCAATTTTTGCAACCATGCACCAATCAAAGAACTGATGAATACGGT
GGATCCATTGAGAACAGAGCCAGGTTAATTCTTGAGTTGATTGACCATTTGTCTACCATTGTCG
GTGCTGACAAGATTGGTATCAGAATCTCTCCATGGGCTACTTTCCAAAACATGAAGGCTCACAA
GGACACTGTTCACCCATTGACTACTTTCTCTTACTTGGTCCACGAATTGCAACAGAGAGCTGAC
AAGGGTCAAGGTATTGCCTACATTTCTGTCGTTGAGCCTCGTGTAAGTGGTAACGTCGACGTCT
CTGAAGAAGACCAAGCTGGTGACAACGAATTTGTCTCCAAGATCTGGAAGGGTGTTATCTTGAA
GGCAGGTAACTACTCCTACGATGCTCCAGAGTTCAAGACATTGAAGGAAGATATCGCTGACAAG
CGTACATTAGTTGGCTTCTCCAGATACTTCACCTCGAATCCTAACTTGGTTTGGAAATTGCGTG
ATGGAATTGACTTGGTGCCATACGACAGAAACACGTTCTACAGTGACAATAACTATGGTTACAA
TACCTTTTCTATGGATTCCGAAGAGGTTGATAAAGAATTAGAAATCAAGAGAGTTCCTTCGGCC
ATTGAAGCTTTGTGATGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCT
AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCC
TTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

Figure B-1. Sequence of pBS2 template plasmid (continued).

APPENDIX C
MUTAGENIC PRIMERS

Table C-1. List of mutagenic primers.

Mutation	Sequence
T35X Fwd	5'-ATCCACCAACT*NNK*AGATTTAGAGCTTTAGAAGACCACAC-3'
T35X Rev	5'-GCTCTAAATCT*MNN*AGTTGGTGGATAGACAATCTTGGTTT-3'
F37X Fwd	5'-CAACTACTAGA*NNK*AGAGCTTTAGAAGACCACACTCCTTC-3'
F37X Rev	5'-TCTAAAGCTCT*MNN*TCTAGTAGTTGGTGGATAGACAATCT-3'
A68X Fwd	5'-TTATCACTGAA*NNK*ACTTTTGTCTCTCCTCAAGCCTCTGG-3'
A68X Rev	5'-GAGACAAAAGT*MNN*TTCAGTGATAAGCAAAGTACCTGGGA-3'
Y78X Fwd	5'-AAGCCTCTGGT*NNK*GAAGGTGCTGCTCCAGGTATTTGGAC-3'
Y78X Rev	5'-GCAGCACCTTC*MNN*ACCAGAGGCTTGAGGAGAGACAAAAG-3'
I113X Fwd	5'-CAACCCAGTTG*NNK*TTTTTGGGAAGGGTTGCAGATCCAGC-3'
I113X Rev	5'-CTTCCCAAAAA*MNN*CAACTGGGTTGAAACGAAAGAACCG-3'
I113N Fwd	5'-CAACCCAGTTG*AAT*TTTTTGGGAAGGGTTGCAGATCCAGC-3'
I113N Rev	5'-CTTCCCAAAAA*ATT*CAACTGGGTTGAAACGAAAGAACCG-3'
H188X Fwd	5'-ATATTGAACTC*NNK*GCTGCTCATGGCTACCTTTTAGATCA-3'
H188X Rev	5'-CCATGAGCAGC*MNN*GAGTTCAATATAGTCGAAACCAGCAT-3'
H191X Fwd	5'-TCCATGCTGCT*NNK*GGCTACCTTTTAGATCAATTTTTGCA-3'
H191X Rev	5'-AAAAGGTAGCC*MNN*AGCAGCATGGAGTTCAATATAGTCGA-3'
Y193X Fwd	5'-CTGCTCATGGC*NNK*CTTTTAGATCAATTTTTGCAACCATG-3'
Y193X Rev	5'-TGATCTAAAAG*MNN*GCCATGAGCAGCATGGAGTTCAATAT-3'
Y193C Fwd	5'-CTGCTCATGGC*TGT*CTTTTAGATCAATTTTTGCAACCATG-3'
Y193C Rev	5'-TGATCTAAAAG*ACA*GCCATGAGCAGCATGGAGTTCAATAT-3'
Y193D Fwd	5'-CTGCTCATGGC*GAT*CTTTTAGATCAATTTTTGCAACCATG-3'
Y193D Rev	5'-TGATCTAAAAG*ATC*GCCATGAGCAGCATGGAGTTCAATAT-3'
Y193K Fwd	5'-CTGCTCATGGC*AAG*CTTTTAGATCAATTTTTGCAACCATG-3'
Y193K Rev	5'-TGATCTAAAAG*CTT*GCCATGAGCAGCATGGAGTTCAATAT-3'
F247X Fwd	5'-CATGGGCTACT*NNK*CAAAACATGAAGGCTCACAAGGACAC-3'
F247X Rev	5'-TTCATGTTTTG*MNN*AGTAGCCCATGGAGAGATTCTGATAC-3'
G292X Fwd	5'-CTCGTGTAAGT*NNK*AACGTCGACGTCTCTGAAGAAGACCA-3'
G292X Rev	5'-ACGTCGACGTT*MNN*ACTTACACGAGGCTCAACGACAGAAA-3'
N293X Fwd	5'-GTGTAAGTGGT*NNK*GTCGACGTCTCTGAAGAAGACCAAGC-3'
N293X Rev	5'-GAGACGTCGAC*MNN*ACCACTTACACGAGGCTCAACGACAG-3'
V294X Fwd	5'-TAAGTGGTAAC*NNK*GACGTCTCTGAAGAAGACCAAGCTGG-3'
V294X Rev	5'-TCAGAGACGTC*MNN*GTTACCACTTACACGAGGCTCAACGA-3'
F373X Fwd	5'-ACAGAAACACG*NNK*TACAGTGACAATAACTATGGTTACAA-3'
F373X Rev	5'-TTGTCACTGTA*MNN*CGTGTTTCTGTCGTATGGCACCAAGT-3'
Y374X Fwd	5'-GAAACACGTTC*NNK*AGTGACAATAACTATGGTTACAATAC-3'
Y374X Rev	5'-TTATTGTCACT*MNN*GAACGTGTTTCTGTCGTATGGCACCA-3'

APPENDIX D
MUTAGENIC PLASMIDS

Table D-1. List of plasmids used in this study.

Plasmid	Parent	Mutation	Description
pDJB32			OYE2.6-GST fusion protein
pBS2	pDJB32	Y368 silent	deletion of NdeI restriction site
pBS9	pBS2	I113D	single mutation
pBS10	pBS2	I113F	single mutation
pBS11	pBS2	I113K	single mutation
BTS unnamed	pAW3	I113C	double mutation
BTS unnamed	pAW3	I113W	double mutation
BTS unnamed	pAW3	I113Y	double mutation
pAW1	pBS2	I113W	isolated from 1st round screening
pAW2	pAW7	I113N	double mutation
pAW3	pBS2	Y78W	isolated from 1st round screening
pAW4	pBS2	I113E	isolated from 1st round screening
pAW5	pBS2	V294P	isolated from 1st round screening
pAW6	pBS2	I113N	isolated from 1st round screening
pAW7	pBS2	F247Y	isolated from 1st round screening
pAW8	pAW3	I113F	isolated from 2nd round screening
pAW9	pAW3	I113L	isolated from 2nd round screening
pAW10	pAW3	I113M	isolated from 2nd round screening
pAW11	pAW3	I113V	isolated from 2nd round screening
pAW12	pBS2	Y193C	single mutation
pAW13	pBS2	Y193D	single mutation
pAW14	pBS2	Y193K	single mutation
pAW15	pBS2	Y193T	isolated from 1st round screening

LIST OF REFERENCES

(1) Williams, R.; Bruce, N. *Microbiol.* **2002**, *148*, 1607–1614.

(2) Toogood, H. S.; Gardiner, J. M.; Scrutton, N. S. *ChemCatChem* **2010**, *2*, 892–914.

(3) Hollmann, F.; Arends, I. W. C. E.; Holtmann, D. *Green Chem.* **2011**, *13*, 2285–2314.

(4) Warburg, O.; Christian, W. *Naturwissenschaften* **1932**, *20*, 688–688.

(5) Theorell, H.; Akeson, A. *Arch. Biochem. Biophys.* **1956**, *65*, 439–448.

(6) Matthews, R.; Massey, V. *J. Biol. Chem.* **1969**, *244*, 1779–&.

(7) Abramovitz, A.; Massey, V. *J. Biol. Chem.* **1976**, *251*, 5327–5336.

(8) Nishina, Y.; Kitagawa, T.; Shiga, K.; Watari, H.; Yamano, T. *J. Biochem.* (9) Abramovitz, A.; Massey, V. *J. Biol. Chem.* **1976**, *251*, 5321–5326.

(10) Saito, K.; Thiele, D.; Davio, M.; Lockridge, O.; Massey, V. *J. Biol. Chem.* **1991**, *266*, 20720–20724.

(11) Stott, K.; Saito, K.; Thiele, D.; Massey, V. *J. Biol. Chem.* **1993**, *268*, 6097–6106.

(12) Vaz, A.; Chakraborty, S.; Massey, V. *Biochem.* **1995**, *34*, 4246–4256.

(13) Kohli, R.; Massey, V. *J. Biol. Chem.* **1998**, *273*, 32763–32770.

(14) Swiderska, M. A.; Stewart, J. D. *J. Mol. Catal. B-Enzym.* **2006**, *42*, 52–54.

(15) Mueller, A.; Hauer, B.; Rosche, B. *Biotechnol. Bioeng.* **2007**, *98*, 22–29.

(16) Müller, A.; Stürmer, R.; Hauer, B.; Rosche, B. *Angew. Chem. Int. Edit.* **2007**, *46*, 3316–3318.

(17) Bougioukou, D. J.; Stewart, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7655–7658.

(18) Padhi, S. K.; Bougioukou, D. J.; Stewart, J. D. *J. Am. Chem. Soc.* **2009**, *131*, 3271–3280.

(19) Hall, M.; Stueckler, C.; Hauer, B.; Stuermer, R.; Friedrich, T.; Breuer, M.; Kroutil, W.; Faber, K. *Eur. J. Org. Chem.* **2008**, 1511–1516.

(20) Stueckler, C.; Mueller, N. J.; Winkler, C. K.; Glueck, S. M.; Gruber, K.; Steinkellner, G.; Faber, K. *Dalton T.* **2010**, *39*, 8472–8476.

(21) Winkler, C. K.; Stueckler, C.; Mueller, N. J.; Pressnitz, D.; Faber, K. *Eur. J. Org. Chem.* **2010**, 6354–6358.

(22) Stueckler, C.; Winkler, C. K.; Bonnekessel, M.; Faber, K. *Adv. Synth. Catal.* **2010**, *352*, 2663–2666.

(23) Brenna, E.; Gatti, F. G.; Manfredi, A.; Monti, D.; Parmeggiani, F. *Eur. J. Org. Chem.* **2011**, 4015–4022.

(24) Korpak, M.; Pietruszk, J. *Adv. Synth. Catal.* **2011**, *353*, 1420–1424.

(25) Stueckler, C.; Winkler, C. K.; Hall, M.; Hauer, B.; Bonnekessel, M.; Zangger, K.; Faber, K. *Adv. Synth. Catal.* **2011**, *353*, 1169–1173.

(26) Meah, Y.; Massey, V. *P. Natl. Acad. Sci.* **2000**, *97*, 10733–10738.

(27) Meah, Y.; Brown, B.; Chakraborty, S.; Massey, V. *P. Natl. Acad. Sci.* **2001**, *98*, 8560–8565.

(28) Williams, R. E.; Rathbone, D. A.; Scrutton, N. S.; Bruce, N. C. *Appl. Environ. Microbiol.* **2004**, *70*, 3566–3574.

(29) Swiderska, M. A.; Stewart, J. D. *Org. Lett.* **2006**, *8*, 6131–6133.

(30) Fox, K.; Karplus, P. *Structure* **1994**, *2*, 1089–1105.

(31) Brown, B.; Deng, Z.; Karplus, P. *J. Biol. Chem.* **1998**.

(32) Xu, D.; Kohli, R. M.; Massey, V. *P. Natl. Acad. Sci.* **1999**, *96*, 3556.

(33) Jeffries, T.; Grigoriev, I.; Grimwood, J.; Laplaza, J.; Aerts, A.; Salamov, A.; Schmutz, J.; Lindquist, E.; Dehal, P.; Shapiro, H. *Nature Biotechnol.* **2007**, *25*, 319–326.

(34) Bougioukou, D. Ph.D. Thesis, University of Florida, Gainesville, FL, 2006.

(35) Walton, A. Z.; Conerly, W. C.; Pompeu, Y.; Sullivan, B.; Stewart, J. D. *ACS Catal.* **2011**, *1*, 989–993.

(36) Casas, J.; Sunden, H.; Cordova, A. *Tetrahedron Lett.* **2004**, *45*, 6117–6119.

(37) Chen, A.; Xu, J.; Chiang, W.; Chai, C. L. L. *Tetrahedron* **2010**, *66*, 1489–1495.

(38) Di Bugno, C.; Colombani, S.; Dapporto, P.; Garzelli, G.; Giorgi, R.; Paoli, P.; Subissi, A.; Turbanti, L. *Chirality* **1997**, *9*, 713–721.

(39) Crimmins, M. T.; Jacobs, D. L. *Org. Lett.* **2009**, *11*, 2695–2698.

(40) Mohapatra, D. K.; Bhattasali, D.; Gurjar, M. K.; Khan, M. I.; Shashidhara, K. S. *Eur. J. Org. Chem.* **2008**, *2008*, 6213–6224.

(41) Xing, Y.; O'doherty, G. A. *Org. Lett.* **2009**, *11*, 1107–1110.

(42) Guo, H.; O'doherty, G. *Org. Lett.* **2005**, *7*, 3921–3924.

(43) Mlynarski, J.; Ruiz-Caro, J.; Fürstner, A. *Chem.-Eur. J.* **2004**, *10*, 2214–2222.

(44) Rezgui, F.; Gaied, El, M. *Tetrahedron Lett.* **1998**, *39*, 5965–5966.

(45) Mase, N.; Inoue, A.; Nishio, M.; Takabe, K. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3955–3958.

(46) Coutrot, P.; Grison, C.; Bomont, C. *Tetrahedron Lett.* **1994**, *35*, 8381–8384.

(47) Coutrot, P.; Grison, C.; Bômont, C. *J. Organometal. Chem.* **1999**, *586*, 208–217.

(48) Corey, E.; Pyne, S.; Su, W. *Tetrahedron Lett.* **1983**, *24*, 4883–4886.

(49) Kar, A.; Argade, N. P. *Synthesis* **2005**, 1234–1236.

(50) Drewes, S.; Loizou, G.; Roos, G. *Synthetic Commun.* **1987**, *17*, 291–298.

(51) Lees, W. J.; Whitesides, G. M. *J. Org. Chem.* **1993**, *58*, 1887–1894.

(52) Oertling, H.; Reckziegel, A.; Surburg, H.; Bertram, H.-J. *Chem. Rev.* **2007**, *107*, 2136–2164.

(53) Nie, Y.; Niah, W.; Jaenicke, S.; Chuah, G.-K. *J. Catal.* **2007**, *248*, 1–10.

(54) Etzold, B.; Jess, A. *Chem. Eng. Technol.* **2008**, *31*, 1282–1289.

(55) Trasarti, A. F.; Marchi, A. J.; Apesteguia, C. R. *J. Catal.* **2007**, *247*, 155–165.

(56) Wolken, W.; Have, Ten, R.; Van Der Werf, M. *J. Agr. Food Chem.* **2000**, *48*, 5401–5405.

(57) Gooding, O. W.; Voladri, R.; Bautista, A.; Hopkins, T.; Huisman, G.; Jenne, S.; Ma, S.; Mundorff, E. C.; Savile, M. M. *Org. Process Res. Dev.* **2010**, *14*, 119–126.

(58) Mwangi, M. T.; Runge, M. B.; Hoak, K. M.; Schulz, M. D.; Bowden, N. B. *Chem.-Eur. J.* **2008**, *14*, 6780–6788.

(59) Sheldon, R. A. *Biochem. Soc. T.* **2007**, *35*, 1583–1587.

(60) Bougioukou, D. J.; Walton, A. Z.; Stewart, J. D. *Chem. Commun.* **2010**, *46*, 8558–8560.

(61) Yamamoto, T.; Ujihara, H.; Watanabe, S.; Harada, M.; Matsuda, H.; Hagiwara, T. *Tetrahedron* **2003**, *59*, 517–524.

(62) Bergmann, J.; Lofstedt, C.; Ivanov, V.; Francke, W. *Eur. J. Org. Chem.* **2001**, 3175–3179.

(63) Riordan, J.; Vallee, B. *Meth. Enzymol.* **1972**, *25*, 494–499.

(64) Riebel, B.; Gibbs, P.; Wellborn, W.; Bommarius, A. *Adv. Synth. Catal.* **2002**, *344*, 1156–1168.

(65) Bratbak, G.; Dundas, I. *Appl. Environ. Microbiol.* **1984**, *48*, 755.

(66) Arnold, F. *Acc. Chem. Res.* **1998**, *31*, 125–131.

(67) Shao, Z.; Arnold, F. *Curr. Opin. Struc. Biol.* **1996**, *6*, 513–518.

(68) Neylon, C. *Nucleic Acids Res.* **2004**, *32*, 1448–1459.

(69) Stemmer, W. *P. Natl. Acad. Sci.* **1994**, *91*, 10747–10751.

(70) Stemmer, W. *Nature* **1994**, *370*, 389–391.

(71) Zhao, H.; Arnold, F. H. *Nucleic Acids Res.* **1997**, *25*, 1307.

(72) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. *Nature Biotechnol.* **1998**, *16*, 258–261.

(73) Coco, W. M.; Levinson, W. E.; Crist, M. J.; Hektor, H. J.; Darzins, A.; Pienkos, P. T.; Squires, C. H.; Monticello, D. J. *Nature Biotechnol.* **2001**, *19*, 354–359.

(74) Moore, G.; Maranas, C.; Lutz, S.; Benkovic, S. *P. Natl. Acad. Sci.* **2001**, *98*, 3226–3231.

(75) Zha, D.; Eipper, A.; Reetz, M. T. *ChemBioChem* **2003**, *4*, 34–39.

(76) Ostermeier, M. *Trends Biotechnol.* **2003**, *21*, 244–247.

(77) Ostermeier, M.; Nixon, A.; Benkovic, S. *Bioorg. Med. Chem.* **1999**, *7*, 2139–2144.

(78) Sieber, V.; Martinez, C.; Arnold, F. *Nature Biotechnol.* **2001**, *19*, 456–460.

(79) Cadwell, R. C.; Joyce, G. F. *Genome Res.* **1992**, *2*, 28–33.

(80) Cirino, P.; Mayer, K.; Umeno, D. *Meth. Mol. Biol.* **2003**, *231*, 3–10.

(81) Wong, T.; Zhurina, D.; Schwaneberg, U. *Comb. Chem. High T. Scr.* **2006**, *9*, 271–288.

(82) Wong, T. S.; Roccatano, D.; Zacharias, M.; Schwaneberg, U. *J. Mol. Biol.* **2006**, *355*, 858–871.

(83) Murakami, H.; Hohsaka, T.; Sisido, M. *Nature Biotechnol.* **2002**, *20*, 76–81.

(84) Reidhaarolson, J.; Sauer, R. *Science* **1988**, *241*, 53–57.

(85) Hughes, M. D.; Nagel, D. A.; Santos, A. F.; Sutherland, A. J.; Hine, A. V. *J. Mol. Biol.* **2003**, *331*, 973–979.

(86) Georgescu, R.; Bandara, G.; Sun, L. *Meth. Mol. Biol.* **2003**, *231*, 75–84.

(87) Ho, S.; Hunt, H.; Horton, R.; Pullen, J.; Pease, L. *Gene* **1989**, *77*, 51–59.

(88) Weiner, M. P.; Costa, G. L.; Schoettlin, W.; Cline, J.; Mathur, E.; Bauer, J. C. *Gene* **1994**, *151*, 119–123.

(89) Kretz, K. A.; Richardson, T. H.; Gray, K. A.; Robertson, D. E.; Tan, X.; Short, J. M. *Meth. Enzymol.* **2004**, *388*, 3–11.

(90) Hogrefe, H.; Cline, J.; Youngblood, G.; Allen, R. *Biotechniques* **2002**, *33*, 1158–.

(91) Kirsch, R. D.; Joly, E. *Nucleic Acids Res.* **1998**, *26*, 1848–1850.

(92) Tseng, W.-C.; Lin, J.-W.; Wei, T.-Y.; Fang, T.-Y. *Anal. Biochem.* **2008**, *375*, 376–378.

(93) Sanchis, J.; Fernandez, L.; Carballeira, J. D.; Drone, J.; Gumulya, Y.; Hoebenreich, H.; Kahakeaw, D.; Kille, S.; Lohmer, R.; Peyralans, J. J. P.; Podtetenieff, J.; Prasad, S.; Soni, P.; Taglieber, A.; Wu, S.; Zilly, F. E.; Reetz, M. T. *Appl. Microbiol. Biotechnol.* **2008**, *81*, 387–397.

(94) Reetz, M.; Carballeira, J. *Nature Protocols* **2007**.

(95) Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A. *Angew. Chem. Int. Edit.* **2005**, *44*, 4192–4196.

(96) Reetz, M. T.; Kahakeaw, D.; Lohmer, R. *ChemBioChem* **2008**, *9*, 1797–1804.

(97) Reetz, M. *Angew. Chem. Int. Edit.* **2002**, *41*, 1335–1338.

(98) Arnold, F. H. *Nature* **2001**, *409*, 253–257.

(99) Bornscheuer, U.; Pohl, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 137–143.

(100) Powell, K. A.; Ramer, S. W.; Del Cardayré, S. B.; Stemmer, W. P. C.; Tobin, M. B.; Longchamp, P. F.; Huisman, G. W. *Angew. Chem. Int. Edit.* **2001**, *40*, 3948–3959.

(101) Zhao, H.; Chockalingam, K.; Chen, Z. *Curr. Opin. Biotech.* **2002**, *13*, 104–110.

(102) Turner, N. *Trends Biotechnol.* **2003**, *21*, 474–478.

(103) Cherry, J.; Fidantsef, A. *Curr. Opin. Biotech.* **2003**, *14*, 438–443.

(104) Jaeger, K.; Eggert, T. *Curr. Opin. Biotech.* **2004**, *15*, 305–313.

(105) Robertson, D.; Steer, B. *Curr. Opin. Chem. Biol.* **2004**, *8*, 141–149.

(106) Lutz, S.; Patrick, W. *Curr. Opin. Biotech.* **2004**, *15*, 291–297.

(107) Bommarius, A. S.; Blum, J. K.; Abrahamson, M. J. *Curr. Opin. Chem. Biol.* **2011**, *15*, 194–200.

(108) Reetz, M. T. *Angew. Chem. Int. Edit.* **2011**, *50*, 138–174.

(109) Paramesvaran, J.; Hibbert, E. G.; Russell, A. J.; Dalby, P. A. *Protein Eng. Des. Sel.* **2009**, *22*, 401–411.

(110) Zhang, J. H.; Dawes, G.; Stemmer, W. P. C. *P. Natl. Acad. Sci.* **1997**, *94*, 4504.

(111) Parikh, M. R.; Matsumura, I. *J. Mol. Biol.* **2005**, *352*, 621–628.

(112) Reetz, M. T.; Torre, C.; Eipper, A.; Lohmer, R.; Hermes, M.; Brunner, B.; Maichele, A.; Bocola, M.; Arand, M.; Cronin, A.; Genzel, Y.; Archelas, A.; Furstoss, R. *Org. Lett.* **2004**, *6*, 177–180.

(113) Chen, C.; Fujimoto, Y.; Girdaukas, G.; Sih, C. *J. Am. Chem. Soc.* **1982**, *104*, 7294–7299.

(114) Zocher, F.; Enzelberger, M.; Bornscheuer, U.; Hauer, B.; Schmid, R. *Anal. Chim. Acta.* **1999**, *391*, 345–351.

(115) Reetz, M.; Becker, M.; Klein, H.; Stockigt, D. *Angew. Chem. Int. Edit.* **1999**, *38*, 1758–1761.

(116) Reetz, M. T.; Wang, L.-W.; Bocola, M. *Angew. Chem. Int. Edit.* **2006**, *45*, 1236–1241.

(117) Reetz, M. T.; Zheng, H. *ChemBioChem* **2011**, *12*, 1529–1535.

(118) Prasad, S.; Bocola, M.; Reetz, M. T. *ChemPhysChem* **2011**, *12*, 1550–1557.

(119) Reetz, M.; Zonta, A.; Schimossek, K.; Liebeton, K.; Jaeger, K. *Angew. Chem. Int. Edit.* **1997**, *36*, 2830–2832.

(120) Liebeton, K.; Zonta, A.; Schimossek, K.; Nardini, M.; Lang, D.; Dijkstra, B.; Reetz, M.; Jaeger, K. *Chem. Biol.* **2000**, *7*, 709–718.

(121) Reetz, M.; Wilensek, S.; Zha, D.; Jaeger, K. *Angew. Chem. Int. Edit.* **2001**, *40*, 3589–.

(122) Zha, D.; Wilensek, S.; Hermes, M.; Jaeger, K.-E.; Reetz, M. T. *Chem. Commun.* **2001**, 2664–2665.

(123) Reetz, M. T.; Prasad, S.; Carballeira, J. D.; Gumulya, Y.; Bocola, M. *J. Am. Chem. Soc.* **2010**, *132*, 9144–9152.

(124) Bougioukou, D. J.; Kille, S.; Taglieber, A.; Reetz, M. T. *Adv. Synth. Catal.* **2009**, *351*, 3287–3305.

(125) Pompeu, Y.; Stewart, J. D. *Unpublished Results*.

(126) Kille, S. Ph.D Dissertation, Max Planck Institut fur Kohlenforschung, Mulheim an der Ruhr, 2010.

(127) Zheng, L.; Baumann, U.; Reymond, J. *Nucleic Acids Res.* **2004**, *32*, E115.

(128) Edelheit, O.; Hanukoglu, A.; Hanukoglu, I. *BMC Biotechnol.* **2009**, *9*, 61.

(129) Liu, H.; Naismith, J. *BMC Biotechnol.* **2008**, *8*, 91.

(130) Walton, A. Z.; Stewart, J. D. *Biotechnol. Progr.* **2002**, *18*, 262–268.

(131) Walton, A. Z.; Stewart, J. D. *Biotechnol. Progr.* **2004**, *20*, 403–411.

(132) Schwaneberg, U.; Otey, C.; Cirino, P. C.; Farinas, E.; Arnold, F. H. *J. Biomol. Screen.* **2001**, *6*, 111–117.

(133) Studier, F. *Protein. Expres. Purif.* **2005**, *41*, 207–234.

(134) Duetz, W.; Ruedi, L.; Hermann, R.; O'connor, K.; Buchs, J.; Witholt, B. *Appl. Environ. Microbiol.* **2000**, *66*, 2641–2646.

(135) Duetz, W.; Witholt, B. *Biochem. Eng. J.* **2004**, *17*, 181–185.

(136) Hulley, M. E.; Toogood, H. S.; Fryszkowska, A.; Mansell, D.; Stephens, G. M.; Gardiner, J. M.; Scrutton, N. S. *ChemBioChem* **2010**, *11*, 2433–2447.

(137) Fryszkowska, A.; Toogood, H.; Sakuma, M.; Stephens, G. M.; Gardiner, J. M.; Scrutton, N. S. *Catal. Sci. Technol.* **2011**, *1*, 948–957.

(138) Toogood, H. S.; Fryszkowska, A.; Hulley, M.; Sakuma, M.; Mansell, D.; Stephens, G. M.; Gardiner, J. M.; Scrutton, N. S. *ChemBioChem* **2011**, *12*, 738–749.

BIOGRAPHICAL SKETCH

Lieutenant Colonel Adam Z. Walton graduated from the University of California, Davis in 1994 with a Bachelors of Science in Chemistry. Upon graduation he was awarded a commission as a 2nd Lieutenant in the US Army, branched Chemical Corps, via his completion of the Reserve Officers Training Course at U.C. Davis. In 2001 he was selected to participate in the U.S. Army Advanced Civilian Schooling (ACS) program where he studied under Dr. Jon Stewart at the University of Florida and earned a Masters of Science in Chemistry. After leaving Florida he served as an Instructor and Assistant Professor in the Department of Chemistry and Life Science at the United States Military Academy, West Point, New York. During his military career he has served in various positions in Louisiana, Alabama, Kentucky, Kansas and New York. Most recently he deployed to Afghanistan in support of Operation Enduring Freedom with the historic 101st Airborne Division. His military decorations include the Bronze Star Medal, the Meritorious Service Medal, and the Joint Meritorious Unit Award. In 2009 Lieutenant Colonel Walton was once again selected to participate in the U.S. Army ACS program to pursue a PhD in Chemistry at the University of Florida. Upon completion of these studies he will return to West Point and the Department of Chemistry and Life Science where he will continue to coach, teach, and mentor our Nation's future leaders.