Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Ashley Bagwell Daugherty Date

Generating designer enzymes for therapeutic and industrial applications

By

Ashley Bagwell Daugherty
Doctor of Philosophy

Chemistry

_____

Dr. Stefan Lutz

Advisor

_____

Dr. David Lynn

Committee Member

_____

Dr. Dennis Liotta

Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D. Dean of the James T. Laney School of Graduate Studies

_____
Date

Generating designer enzymes for therapeutic and industrial applications

By

Ashley Bagwell Daugherty
B.S., University of South Carolina, 2008

Advisor: Stefan Lutz, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry

2013

Abstract

Generating designer enzymes for therapeutic and industrial applications

By Ashley Bagwell Daugherty

Enzymes are versatile catalysts that are important for their use in a variety of therapeutic and industrial processes. Often times enzyme's inherent properties are not suitable for a particular application and require engineering to improve substrate specificity, efficiency, stability, reactivity, enantioselectivity etc. In this dissertation semi-rational design strategies in combination with circular permuation were employed to alter the biocatalytic properties of three separate enzyme classes toward their respective applications. In each case, the engineering methods resulted in dramatically reduced mutant library size and variants were generated with enhanced designer properties.

Structural guided rational design and computational tools were used to reengineer the active site of the deoxyribonucleoside kinases dCK and *Dm*dNK to have specificity toward the 3'-modified nucleoside analogs ddT and kT. The top variants were characterized by kinetic analysis, which revealed a substrate specificity switch with a significant decline in activity for the natural substrate thymidine. Overall, this method proved successful for quickly tuning substrate specificity and identifying favorable substitutions for altering specificity towards nucleoside analogs using smaller focused libraries.

Secondly, Rosetta fixed backbone design was employed to redesign the subtilisin BPN' propeptide structure, as structural stabilization has been proposed to competitively inhibit protease self-cleavage. Computationally optimized propeptide sequences were

generated with increased structural integrity and enhanced thermostability. Molecular dynamics simulations verified that the synthetic propeptides retain structure in the absence of their cognate protease. Two designer propeptides were experimentally characterized via spectroscopic methods and protease inhibition assays. The data verified their structural stabilization and enhanced thermostability but the inhibition profiles suggest that propeptide stability alone is insufficient for effective inhibitor design. Overall, this approach offers a convenient highly tunable method for the design of future protease inhibitors.

Thirdly, a completely synthetic circular permutation library of Old Yellow Enzyme from *Saccharomyces pastorianus* (OYE1) was created using a whole gene synthesis method. Subsequently, a cell-free *in vitro* transcription/translation screen was employed to assess the impact of termini relocation throughout the protein sequence against three reference substrates: ketoisophorone, cinnamaldehyde and (*S*)-carvone. Library screening identified over 70 variants with enhanced catalytic activity and several showed over an order of magnitude improvement. This synthetic approach enables a robust screen to quickly identify lead circular permutants for a variety of industrial applications.

Interestingly, the location of the new termini in all of the improved library members were in the same four loop/lid regions near the active site. The data demonstrated that these structural regions are important for catalysis and that circular permutation presumably altered the conformational flexibility near these areas. Further structural investigation was completed on selected representatives from each of the four regions. Characterization revealed identical secondary structures, altered oliogmeric

states, and increased thermostability between the permutants studied. The crystal structure of a top variant, cpOYE303, was solved and verified a significantly more open and accessible active site with altered flexibility near the locations of new and old termini.

Generating designer enzymes for therapeutic and industrial applications

By

Ashley Bagwell Daugherty
B.S., University of South Carolina, 2008

Advisor: Stefan Lutz, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry

2013

Acknowledgments

I would first and foremost like to thank my Advisor, Dr. Stefan Lutz, for giving me the opportunity to work in his lab. He has been a supportive mentor, always eager to hear new science and constantly providing helpful advice throughout my graduate career. He has truly made my experiences at Emory unforgettable. I would also like to give a special thanks to my committee members, Dr. David Lynn and Dr. Dennis Liotta, for all of their helpful input with my research over the years. In addition, I would like to give a huge shout out to Ann Dasher for always providing an ear to listen and all the help she has given to me.

To all of the current and past Lutz lab members, I couldn't have asked for better colleagues and friends during this graduate school journey. Ya'll have really made coming to work everyday a joy! I thank Samantha Iamurri for being the perfect office mate on the last home stretch and Pravin Muthu who has provided me with continuous support and advice. Thanks you everyone for your patients and assistance, I have enjoyed every minute working with you guys.

I have made some of my best friends at Emory. I could have not made it through graduate school without Erin Schuler, Yoshie Narui, and Kevin Yehl. Each of you have been such a great support system and I am so lucky to have you all as friends! I will always cherish the memories that we have made.

To my parents and sisters: thanks you for providing continuous encouragement and support throughout everything. Last but not least, thank you so very much to my husband Michael and son Brayden. Michael has been my constant rock providing encouragement, support and love everyday. Thanks for all that you do!

List of frequently used abbreviations

Abbreviation	**Full name**
2Cda	2-chlorodeoxyadenosine
AZT	3'-azidothymidine
CD	Circular Dichroism
CP	Circular Permutation
dCK	Deoxycytidine Kinase
ddT	2',3'-dideoxythymidine
dGK	Deoxyguanosine Kinase
*Dm*dNK	*Drosophila melanogaster* Deoxyribonucleoside Kinase
dNKs	Deoxyribonucleoside Kinases
dNMPK	Deoxyribonucleoside Monophosphate Kinases
FMN	Flavin Mononucleotide
HB	Hydroxybenzaldehyde
HSV1-TK	*Herpes simplex* Virus Type1 Thymidine Kinase
IVTT	i*n vitro* Transcription/Translation
KIP	Ketoisophorone
kT	3'-Methylketo-thymidine
MD	Molecular Dynamics
NA	Nucleoside Analogs
NDPK	Nucleoside diphosphate Kinase

OYE	Old Yellow Enzyme
RMSD	Root Mean Square Deviation
RT	Reverse Transcriptase
TK1	Cytosolic Thymidine Kinase
TK2	Mitochondrial Thymidine Kinase

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

1.1 General Introduction

1.1.1 General applications of enzymes

Nature has evolved enzymes to catalyze a variety of chemical reactions, often performing with high substrate specificity as well as enantio and stereoselectivity, in living organisms under benign environmental conditions. Enzymes are the 'green' alternative to chemocatalyts. They perform under aqueous conditions, react with high product purity and selectivity, do not require organic solvents, and produce less waste.[1-3] These properties all make enzymes attractive candidates for biocatalytic applications in both therapeutic and industrial processes. The canonical classification of enzymes is by an Enzyme Commission Number (EC number), which is based off of the type of reaction they catalyzed.[4] This system consists of 6 primary groups: oxidoreductases (EC1), transferases (EC2), hydrolases (EC3), isomerases (EC4), lyases (EC5), and ligases (EC6);[4] as well as several subclasses that represent more specific mechanistic detail about the catalyzed reaction.[4] While a large number of native biocatalysts responsible for the production of high value chemical compounds reside in these classes, it is often desirable to have enzymes with designer properties.[5-7] In this thesis, I focus on three types of applications to alter enzyme function. Firstly, I investigated the engineering of highly specific nucleoside kinases (transferases) as phosphoryl transferases for nucleoside analog prodrugss for antiviral therapy. Secondly, I explored altering the tunablity of proteases (hydrolases) activity with direct application in consumer products. Finally, reductase variants (oxidoreductases) were generated with enhanced or altered catalytic properties toward the production of chiral synthons.

1.1.2 Protein engineering approaches

Often time's scientists want to change the properties of enzymes to make them more adaptable or ideal for a particular process or application of interest. For example, one might envision altering substrate specificity for a desired reaction or increasing the stability of the enzyme. Today, a variety of enzyme engineering strategies exist for tailoring the properties of the target protein. Several of the most common engineering methods are listed in Table 1.1 with the advantages and disadvantages highlighted. In the past, directed evolution was the most popular strategy, comprising a number of engineering techniques that follow the principals of Darwinian evolution.[7-9] Typically, these methods are an iterative process where a diverse library is initially created via random mutagenesis and *in vitro* recombination. Next, library members with improvements are identified by detecting a particular phenotype using selections and/or high throughput screening.[5,7,9-11] A major advantage of these techniques is that no structural information is required of the protein, as it is all based on mutations at the gene level.[6] Directed evolution strategies such as error-prone PCR, neutral drift libraries, as well as many recombination techniques, have been extremely successful in generating improved variants with enhanced catalytic activity, enantioselectivity, substrate specificity and more.[6,12] Although directed evolution has been very effective, there are also disadvantages, which prevent it from being a universal approach. For all directed evolution strategies, a selection or high throughput screening technique is necessary, which might not be feasible for all proteins.[6] Furthermore, error prone PCR suffers from low frequency of beneficial mutations. Greater than 30% of the mutations yields deleterious effects resulting in an inactive protein and only a small fraction of sequences

are sampled.[5,10,13] Neutral drift accumulates mutations under a selection pressure producing viable variants that maintain the original structure and function.[14] Such an approach can dramatically reduce the experimental library size. Gupta and coworkers use this technique to improve the activity of serum paraoxonase (PON1).[14] They fused the target protein to a green florescent protein probe and upon mutagenesis only the soluble proteins fluoresce and can be sorted from the unfolded proteins. This technique has demonstrated ~5000 hits per 10^6 cells.[14] Recombination techniques such as DNA shuffling have been successful for the introduction of novel function. These techniques rely on the combination of different gene fragments (typically with high homology) creating chimeras with altered performance.[13,15,16]

In general, scientists are gravitating towards engineering approaches that significantly reduce library size focusing on an enriched set of high probability targets. These targeted methods have demonstrated equal if not better performance compared to traditional directed evolution strategies.[17] The idea is to use the information available on protein structure, sequence and function in addition to predictive tools to target specific sites within the protein. This protein engineering approach offers quite a reduction in library size and eliminates the need for high throughput evaluation.[5,6,10,12] On the other hand, sequence and structural information is required for these methods. However, with this information and advances in whole gene synthesis, 'smarter' (smaller, more focused) libraries can be designed and synthesized that also represent a large number of sequences without the need to establish screening and selection protocols to process millions of library members. Overall, these semi-rational engineering approaches have the potential to achieve equal or greater improvements compared to several rounds of directed

evolution. In a recent example, Chen and coworkers employed a combination of multiple site-saturation mutagenesis and structure-based computational algorithms to improve the hydroxylation activity of P450 BM3 towards smaller alkanes for the production of the corresponding alcohol.[17] These products are high sought out chemical intermediates especially in the production of liquid fuels. Small libraries (≤ 1028 sequences) were compared to an earlier study using random mutagenesis for the same target. The most active variant isolated from the reduced libraries displayed a higher activity for the hydroxylation of propane compared to variants isolated after 10-12 rounds of directed evolution.[18]

There has been a great deal of focus on structure guided rational design and the use of computational tools to aid in both sequence analysis and amino acid/conformational sampling (Table 1.1). These semi-rational design strategies require sequence, structural and mechanistic knowledge, but all of this information allows specific sites to be targeted, which in turn translates into reduced screening efforts. In a recent study, Cerdobbel *et al.* increased the thermostability of sucrose phosphorylase, an important glycosylation catalyst that is used in higher temperature processes, through sequence and structural based mutagenesis.[19] They exploit two very common strategies that are used in the stabilization of proteins: B-factor analysis[20] and consensus design.[21] B-factor analysis indicates the flexibility of the atoms within the protein based on crystallographic data and consensus design uses conserved amino acids within a set of homologous proteins to predict beneficial mutations. Both of these strategies were used to identify combinations of mutations for two key positions resulting in an increased thermostability. Separately, crystal structural analysis was used to predict surface

residues to introduce salt bridges, also giving rise to a thermostable variant. When the beneficial mutations were combined from the two approaches, a variant was generated that had more than doubled half-life at 60 ºC compared to wild type.[19]

Several successes in protein engineering have also been due to the continuous development of new computational programs. These programs simulate enzyme properties and predict mutational hotspots that are important for a particular function. A few of the common computational methods that have been used in protein design efforts for improving enzyme activity, stability, ligand binding etc are listed in Table 1.1[17,22,23] RosettaDesign is one more popular methods that has been used to design novel enzymes, redesign active sites, increase binding affinity, and create novel protein folds.[23-25] This program is responsible for identifying the lowest energy scoring sequences for a particular target protein structure.[23-25] Recently, Gordon and coworkers identified an endopeptidase that is active under acidic conditions and subsequently used the Rosetta Software Suite to engineer the peptidase toward immunogenic elements that cause intolerance to gluten in celiac disease patients.[20] The group redesigned the active site of the peptidase to have specificity toward a PQ dipeptide (a specific motif that is protease resistant in celiac patients) while maintaining resistance to digestive proteases. The predicted mutations were inserted into the wild type peptidase and one set of mutations resulted in an activity increase of 116-fold for the desired peptide.[26] In summary, these few examples give a flavor of the protein engineering approaches that deal with a small number of variants but result in significantly improved designer enzymes.

The work in this dissertation focuses on two very different engineering strategies: semi-rational design approaches and circular permutation. A combination of site directed

mutagenesis based on structural information, Rosetta guided computational design, and the generation of a completely synthetic circular permuation library were all strategies used to tailor the properties of the biocatalysts in my work. Consequently, all engineering efforts resulted in improved biocatalysts.

Table 1.1. Common protein engineering methods

Method	Description	Advantages	Disadvantages	Ref
		Directed Evolution		
Error-prone PCR	- Introduction of random mutations into the parent gene sequence via PCR	- Easy - Cost efficient - No structural knowledge - No production necessary	- Limited amino acid substitution - Deleterious mutations - Polymerase bias - Need screen/selection - High throughput - Large libraries	5,11
Neutral drift libraries	- Accumulate mutations under a selection pressure	- Not completely random - Bias towards specific features - Maintain structure/function	- Can't design for unique function - Power lies on large sequence database	14
DNA shuffling	- Random recombination of digested DNA via primerless PCR	- Robust - Combinatorial	- Biased crossovers (unless synthetic) - Large amount of parent gene	15, 16
Saturation mutagenesis	- Random introduction of all	- Easy - Target specific	- Sequence information	6

	20 AA at a specific site	site - All 20 AA	required - Limited diversity	
Circular permutation	- Connecting termini via linker sequence and cleave peptide bonds throughout structure creating new termini	- Complete library - Library not too large	- Limited diversity - Linker design	67
Rational and Semi Rational Design				
Site-directed mutagenesis	- Introduction of specific mutations into a gene sequence via PCR	- Specific sequence target - No polymerase bias	- Diversity limited - Requires knowledge of protein structure and function	6, 10
CASTing (Combinatorial active-site saturation test)	- Saturation mutagenesis of residues near active site and using a hit from 1st round and applying process to a different site	- Decreases library size - Reduced screening	- Targets active site only - Combinatorial dependent	50
SCHEMA	- Structure guided recombination of protein fragments	- Doesn't disturb structure - Not limited to two parents	- Need 3D structure - Can't design for unique function	18
RosettaDesign	- Side chain replacement according to knowledge based distribution (can be modified to account for back bone flexibility)	- Enrich library - Filter poor sequences - Specifically design for features (structure, stability, specificity, etc) - Covers large sequence space	- Biased by score functions - Semi-empirical score function - Computer power - Need sequence/structure/function information	24
FoldX	- Crystal structures used to calculate free energy of protein	- Shown to correlate well with experimental stability	- Need 3D structure - Assumes absence of persistent	22

		- Web based	structure	
K* algorithm	- Uses molecular ensembles for protein-ligand redesign	- Flexible and dynamic ligand simulations - Structure based	- Need 3D structure - Biased by score functions	12

1.2 Deoxyribonucleoside kinases in nucleoside analog prodrug activation

1.2.1 Nucleoside analogs as anti viral and anti cancer therapies

There is a continuous need for the development of antiviral and cancer treatments. An expanding realm of therapeutics is nucleoside analog prodrugs. Ultimately, these drugs take advantage of cancer cells's low fidelity of nucleotide incorporation as the result of accelerated gene replication and the low fidelity of viral polymerases or reverse transcriptases. Structurally, the prodrugs are very similar to natural ribo and 2'-deoxyribonucleosides possessing modifications at the ribose and/or nucleobase moiety (Figure 1.1). Nucleoside analogs can be classified depending upon the location of the modification. For example, there are a number of analogs that have modifications on the nucleobase portion, substitutions on the ribose moiety at the 3'-position, as well as acyclic structures in place of the sugar. Analogs such as AZT and ddC (Figure 1.1) lack the 3' hydroxyl group, which is responsible for forming the phosphodiester bond during DNA elongation. Acyclic analogs such as GCV have a similar mode of action, as there is no 3' -end for nucleotide addition. On the other hand, NAs with modifications on the nucleobase portion such as 2-chloro-2'deoxyadenosine (2CdA) interfere with the enzymes involved in nucleotide metabolism.[27]

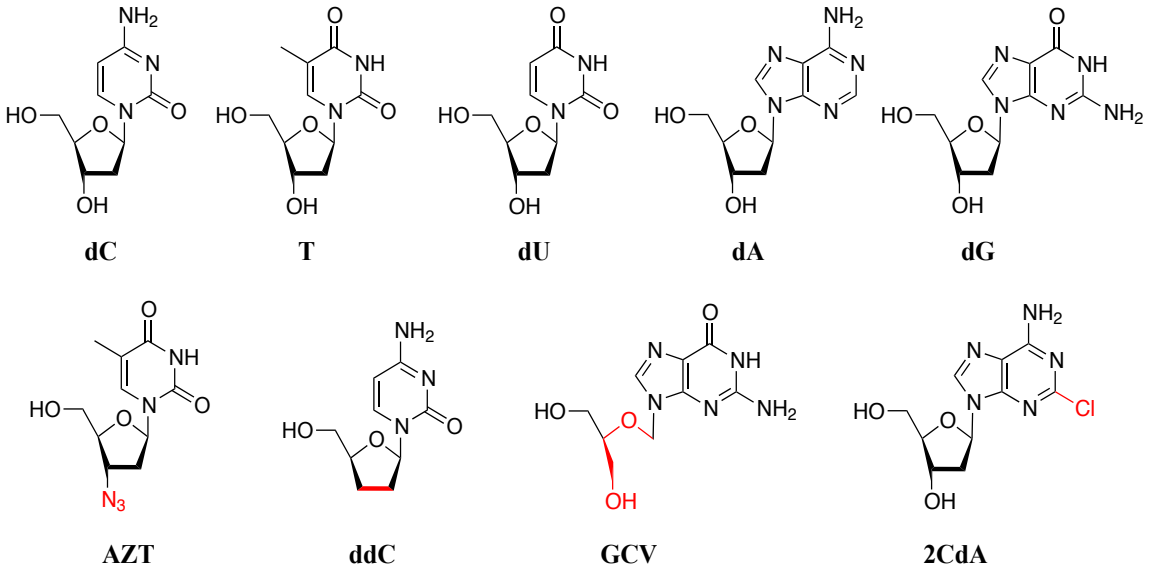


Figure 1.1. Nucleoside Analogs. Top row displays all of the natural deoxyribonucleosides and the bottom row shows representative nucleoside analogs possessing modifications belonging to the different classes.

In human cells, a number of nucleoside transporters aid in transporting the prodrugs into the cell where they are sequentially phosphorylated into their triphosphate anabolites by the cell's nucleoside salvage pathway.[28] The salvage pathway is responsible for phosphorylating deoxyribonucleosides that originate from apoptosis, phosphorylating intermediates produced from the *de novo* pathway and the phosphorylation of nucleoside analog prodrugs.[29] The salvage pathway is comprised of several kinases that are responsible for synthesis of deoxyribonucleotides from deoxyribonucleosides (Figure 1.2). In the first step, 2'-deoxyribonucleoside kinases (dNKs) catalyze the transfer of the first phosphate group onto the 5'-position of the ribose moiety.[30] For many nucleoside analogs this step is often the rate-limiting step.[31] The remaining two-phosphorylation

steps are catalyzed by deoxyribonucleoside monophosphate kinase (dNMPK) and nucleoside diphosphate kinase (NDPK), which are responsible for taking the nucleoside monophosphate to the triphosphate form. In addition, thymidine monophosphate that is produced by the cell's *de novo* pathway enter into the salvage pathway for the final phosphorylation steps (Figure 1.2).[31,32] Once the analogs are in their triphosphate state, they become substrates for reverse transcriptases and low fidelity polymerases found in viruses and cancer cells. Subsequently, their incorporation usually induces chain termination during replication, apoptosis, or induces the accumulation of mutations in viral progeny. [33] All of these mechanisms prevent further disease proliferation.

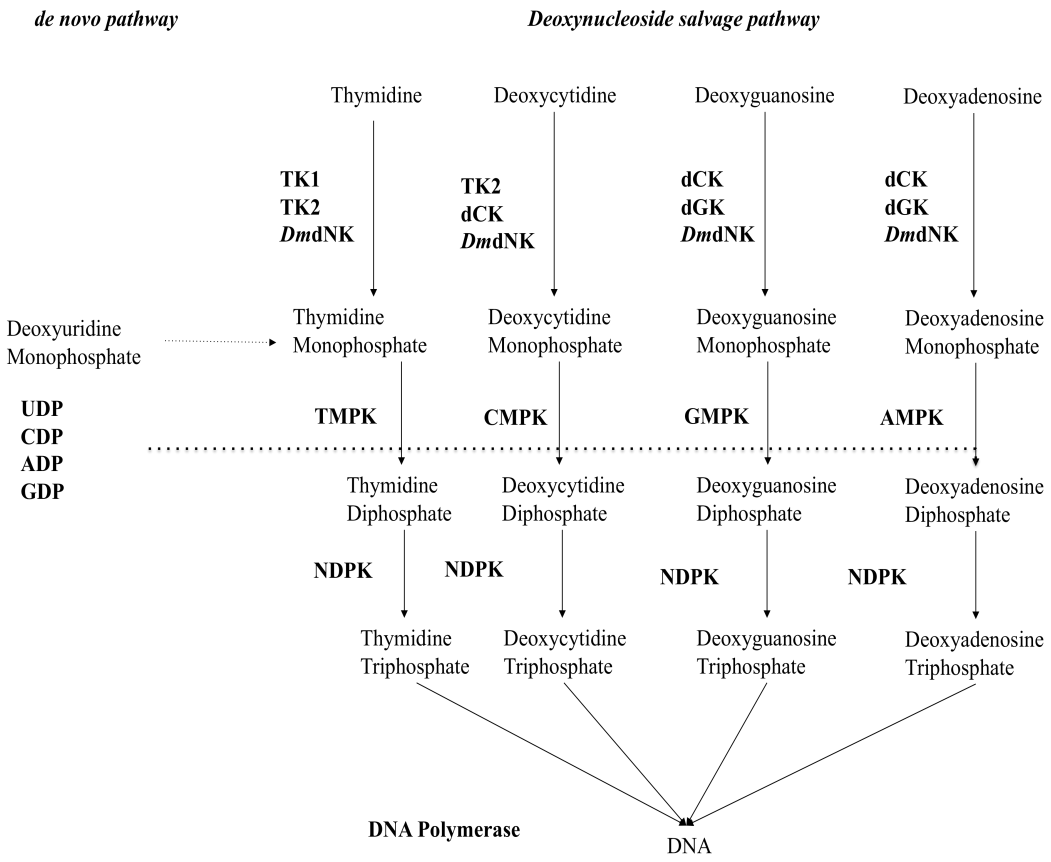


Figure 1.2. The Salvage Pathway in which dNTPs are synthesized and nucleoside analogs are activated. The de novo pathway feeds into the salvage pathway as indicated. thymidine kinase 1 (TK1), deoxycytidine kinase (dCK), *D. melanogaster* (*Dm*dNK), deoxyguanosine kinase (dGK), and thymidine kinase 2 (TK2), thymidine monophosphate kinase (TMPK), cytidine monophosphate kinase (CMPK), guanisine monophosphate kinase (GMPK), adenosine monophosphate kinase (AMPK), nucleoside diphosphate kinase (NDPK). Uridine diphosphate (UDP), cytidine diphosphate (CDP), Adenosine diphosphate (ADP), guanosine diphosphate (GDP).

1.2.2 dNK family

Deoxyribonucleside kinases have been a focal point of many gene therapy and biotech engineering efforts.[27,30,34,35] They are responsible for catalyzing the transfer of a γ-phosphoryl group from a donor molecule (commonly ATP) to the 5'-hydroxyl group of a 2'-deoxyribose forming the corresponding monophosphate anabolites. The transfer is accomplished via a general acid/base mechanism in which a carboxylate sidechain of a glutamate acts as a general base abstracting the proton from the 5'-OH and Mg^{2+} helps to position the substrate. There are also surrounding arginines in the active site to support the deprotonation by compensating for the build-up of negative charge.[36]

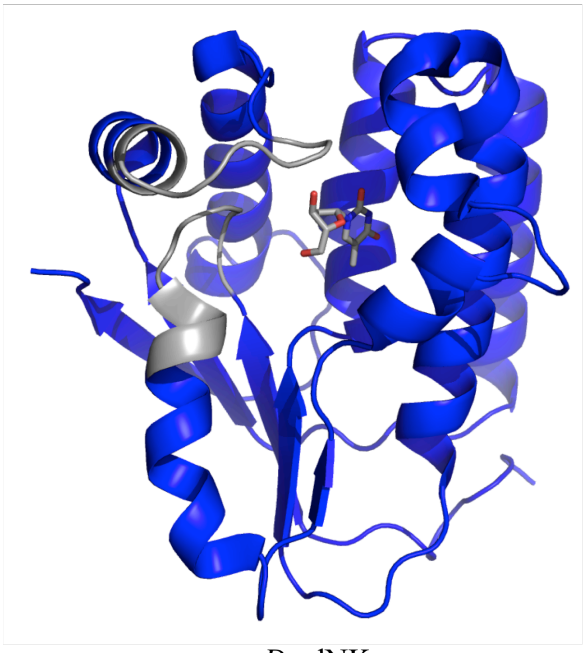
The number of dNKs, their substrate specificity, and cellular localization all vary depending upon the organism.[31] The best-characterized dNKs are from humans, a single deoxyribonucleoside kinase from *D. melanogaster* (*Dm*dNK), and thymidine kinase from *Herpes simplex* virus type1 (HSV1-TK). Humans contain four dNKs: thymidine kinase 1 (TK1) and deoxycytidine kinase (dCK) are located in the cytoplasm while deoxyguanosine kinase (dGK) and thymidine kinase 2 (TK2) reside in the mitochondria.[36] All four of the kinases have differing substrate specificities with some similarities between them (Figure 1.2). The fruit fly kinase, *Dm*dNK, is unique in that it is the only known kinase that is able to phosphorylate all four of the natural deoxyribonucleosides with relatively high turnover. HSV1-TK was the first kinase to be isolated and has been studied extensively for the activation of NAs.[37] In addition to its broad substrate specificity, this dNK can catalyze both the first and second phosphoryl transfers, functioning as a TK and TMPK, although the second reaction is less effeient.[37]

The deoxyribonuceoside kinases are classified into type-1 and type-2 kinases based on sequence and structural similarities. All of the human kinases (with the exception of TK1), *Dm*dNK and HSV1-TK reside in the type-1 family.[36-39] The focus of this dissertation will be only on dNKs that are members of the type-1 family specifically, *Dm*dNK and hdCK. Structurally, the type-1 dNKs assemble into homo dimers and each monomer is comprised of a β-sheet core structure flanked by several helices that compose an alpha/beta/alpha sandwhich fold (Figure 1.3). As far as sequence is concerned, there are two highly conserved motifs in the type-1 dNK family, the LID region and the P-loop. The LID region has a conserved sequence RXXXRXRXXE and is located between helices 7 and 8.[31] Upon substrate binding this flexible region undergoes a conformational change in order to position the Arg residues that are involved in the phosphoryl transfer reaction.[31,36] The P-loop motif is responsible for binding the γ-phosphate from the donor and has the consensus sequence GXXXXGKS/TT.[36]

1.2.4 Limitations with endogenous dNKs

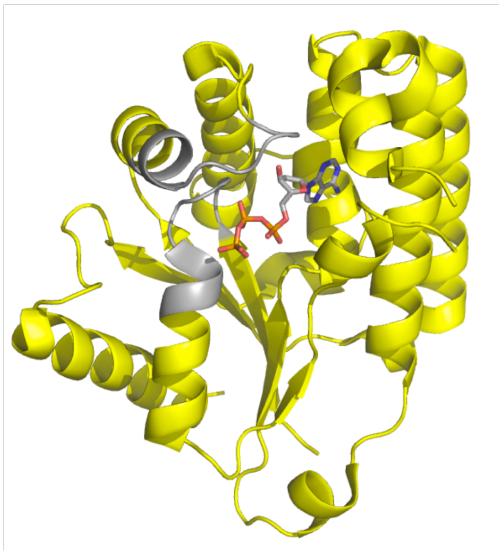
A major obstacle with nucleoside analog treatments is the presiding affinity of endogenous kinases presiding affinity for the natural substrate over the prodrug.[40] As a result, large doses of drug have to be administered introducing side effects for the patient due to the accumulation of cytotoxic intermediates.[34] An even bigger issue can be the inability of native dNKs to phosphorylate the prodrug *in vivo*. To circumvent these problems, efforts are being directed toward generating orthogonal NA kinases that are specific to the particular prodrug without interfering with the natural deoxyribonucleoside metabolism. This can be accomplished by co-administering an enzyme that can fulfill this

role *in vivo* with the drug through gene therapy. Gene therapy is a more targeted approach for treating tumor cells while sparing healthy cells. It allows for the delivery of a gene encoding for a specific enzyme responsible for activating the prodrug. Due to the gene only being expressed in the transected cells, the nucleoside analog is converted into the cytotoxic form in only those corresponding cells.[35]A hallmark example for this approach is the anticancer drug ganciclovir (Figure 1.1) in combination with HSV1-TK.[41-44] When HSV1-TK was introduced into lung cancer cell lines, it significantly increased the sensitivity of GCV showing a dependent toxicity of the prodrug toward this specific 'drug sensitivity' gene.[41-44] Recently, there has been extensive work towards creating or modifying the dNKs to be used as suicide genes via a variety of protein engineering tools.[45,46] My work investigates using structure-based enzyme design with the help of computational tools to generate deoxyribonuclesode analog kinases specifically for 3' modified nucleoside analogs. Key residues were identified that are important for tuning the substrate specificity of this class of prodrugs and can be helpful in the development of future gene partners.



*Dm*dNK

hdCK



hdGK

Figure 1.3. Type-1 dNKs. Structural comparison of Type 1 dNKs: DmdNK (PDB: 1OT338); hdCK (PDB: 2A3039); and hdGK (PDB: 2OCP29). The conserved P-loop and LID regions are highlighted in grey.

1.3 Subtillisin proteases and their propeptide component

In my second project, I focused on proteases that were historically among the first group of biocatalysts investigated for medicinal, industrial and synthetic applications.[47] These enzymes have a simple structural architecture, do not require expensive cofactors and are active under mild reaction conditions. Proteases are responsible for the hydrolysis of peptide bonds of protein and/or peptide substrates. Peptide bond hydrolysis is accomplished by a general acid/ base mechanism (Figure 1.4). Within the large realm of the protease family there are four main subclasses: serine, cysteine, aspartic, and metallo proteases which can be distinguished from one another based on the enzymes mechanistic details.[3] Proteases are also classified based on the peptide bond cleavage specificity with endopeptidases cleaving in the middle of a substrate and exopeptidases acting at the N (aminopeptidases) or C (carboxypeptidases) termini.[4] The majority of proteases fall into either the serine or cysteine class. Mechanistically, most serine and cysteine proteases utilize the classic catalytic triad (Ser/Cys-His-Asp) to facilitate the catalytic reaction. A common mechanism in which serine proteases proceed to catalyze the hydrolysis of a peptide bond is shown below (Figure 1.4) highlighting the enzyme active site containing a nucleophilic serine responsible for attack on the carbonyl of the substrate.

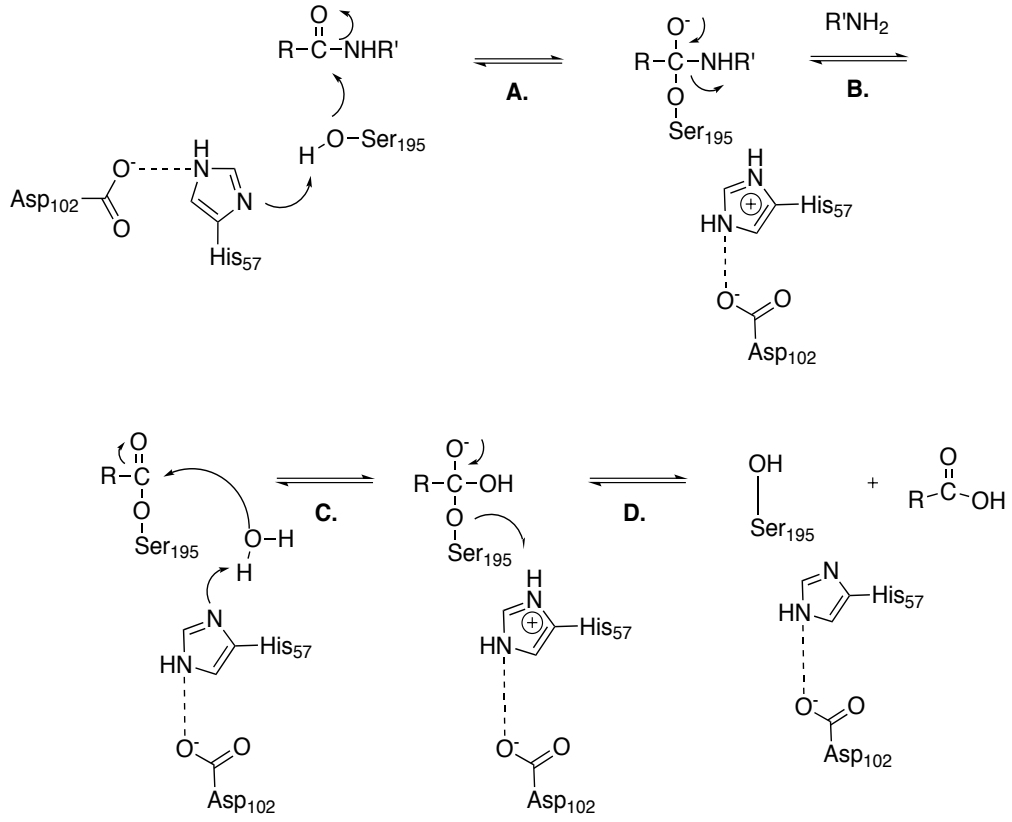


Figure 1.4. Mechanism of serine protease (numbering based on bovine α-chymotrypsin). The acylation part of the reaction occurs first with Ser195 attacking the carbonyl of the substrate, while His57 acts as a general base producing the tetrahedral intermediate. Asp102 stabilizes the protonated His57 and the oxyanion hole helps to stabilize the tetrahedral intermediate oxyanion. Next, the electrons move down, tetrahedral intermediate collapses, protonated His57 acts as a general acid, and the leaving group gets expelled yielding the acyl-enzyme intermediate. The process is repeated with water attacking the carbonyl again with the help of His57 and in the the final step both Ser195 and the carboxylic acid product is released.5

Subtilisin proteases are a class of serine endo-proteases isolated from *Bacillus* bacteria.[1] These enzymes share high sequence homology and are popular for their use in a variety of commercial applications such as laundry detergents, dishwashing soap,

cosmetics, leather processing and food processing industries.[2, 3] In fact, due to all of these uses, subtilisin comprises the largest commercial enzyme market world wide.[8,26]

Many of these proteases are unable to fold efficiently on their own and have evolved to produce an N-terminal propeptide segment that functions as a intramolcular chaperone in the maturation process of the protease.[4,5,6] The propeptide C-terminal region binds in the substrate-binding site of the protease and, upon correct folding, the protease either self-cleaves or existing active subtilisin removes the propeptide triggering unfolding and release of the pro segment. [1,7] The propeptide alone lacks secondary and tertiary structure but upon binding to its parent protease a defined tertiary structure is acquired.[4,7,6,8] In addition to functioning as a molecular chaperone, these propeptides can also function as a temporary inhibitor of the protease.[9,10] Over the years this system has been studied extensively, revealing several insights into the mechanism of the subtilisin activation pathway. The results from studies have suggested that protease function can be altered via 'pro-sequence engineering'. Mutations within the propeptide component can alter the folding, auto processing, and degradation of it's cognate protease (Figure 1.5).[48] As a result the pro-sequence provides an interesting starting point to engineer for the generation of novel propeptide variants with various beneficial functions to the protease such as stability.

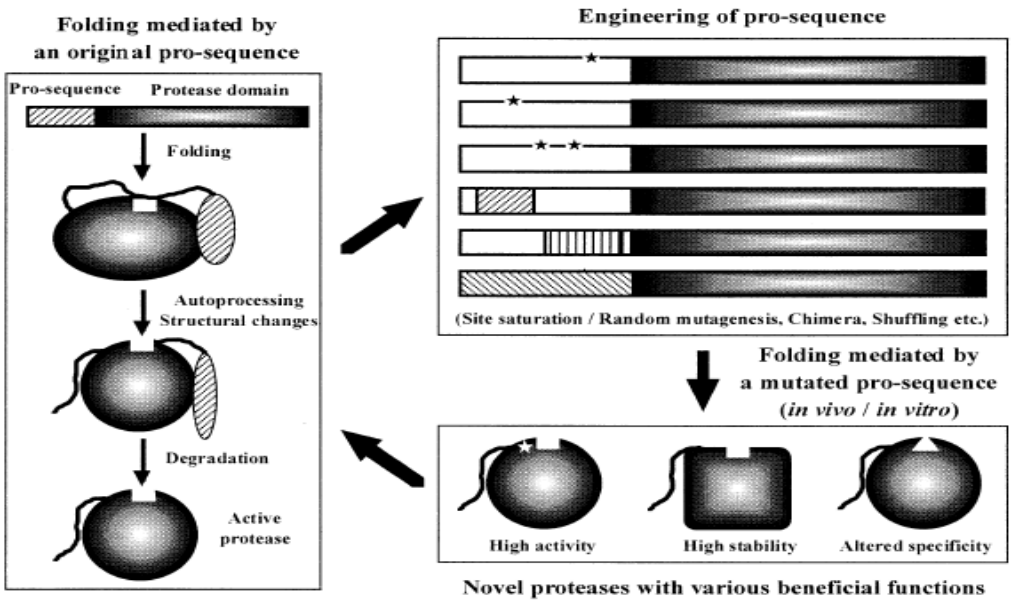


Figure 1.5. Concept of prosequence engineering.[48]

Previous work with this family of proteins has suggested that stabilization of the apo propeptide structure delays its release from the mature protease providing a potential mechanism for extending the protease lifetime.[1,11,12] Protease inhibitors have been found to exhibit high structural similarities with the folded propeptide, suggesting that the stabilization of a propeptide will facilitate its functional transition from molecular chaperon to protease inhibitor.[9] A particular protease and propeptide sequence that has been studied extensively in terms of their structure function relationship is subtilisin BPN' from *Bacillus amyloliquefaciens* (Figure 1.6). Prior work to increase the structural integrity of the BPN' pro sequence has been reported.[12,13,14,11] A notable example include the introduction of a disulfide bond in the propeptide sequence by Bryan and coworkers resulting in a dramatic stabilization of the peptide alone.[13] Separately, Kojima *et al* reported a triple propeptide BPN' mutant with an increase in stability determined by CD

spectroscopy and thermo-denaturation experiments.[11] These studies, along with many others, support the hypothesis that stabilization of the propeptide is one possible avenue for generating more resistant inhibitors.

From an industrial standpoint, the inhibitory function is highly desirable for laundry detergent applications, as the protease shouldn't be activated until use in order to extend the shelf life of the product. Currently, borate based inhibitors are included in formulations to prevent protease degradation but a greener alternative is in demand as potential environmental regulations will prohibit the application of borates in the foreseeable future. Protein engineering strategies can be used to further explore mutations in the propeptide region that will result in altered inhibitory properties. This could potentially be used as a mechanism for controlling the degradation of the mature protease and be a valuable option for generating propeptide based inhibitors. For such applications, it is desirable to have tunable binding affinity of the propeptide as too tight binding will prevent activation and low affinity inhibitors will compromise protease shelf-life. In this thesis, engineering efforts were focused on tuning the inhibitory properties by introducing propeptide structure. RosettaDesign was used to generate propeptide variants with increased structure and enhanced stability. The top scoring designs were synthesized and evaluated for structure, stability and its relationship to inhibitory function.

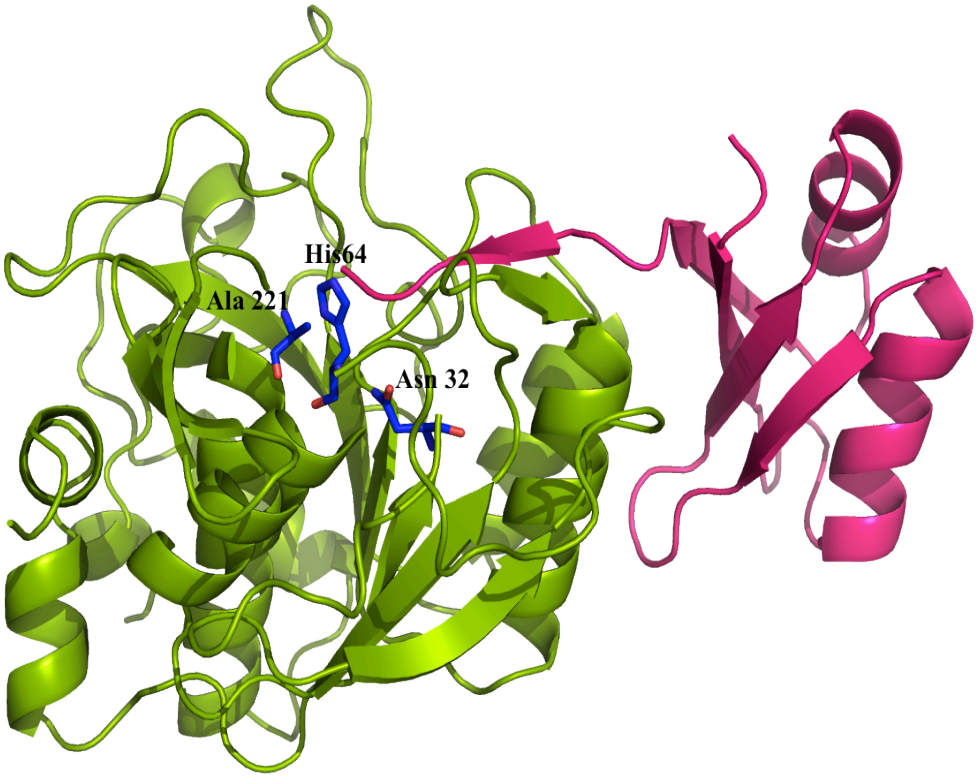


Figure 1.6. Crystal structure of Subtilisin BPN' with its propeptide bound (PDB: 1SPB). The catalytic triad (Asp32Asn, His64, Ser221Ala) is highlighted in blue and the C-terminus of the pro-sequence is bound in the product-binding site.

1.4 Enoate reductases in biocatalysis

In my third and final project, I explored enoate reducatases. These enzymes belong to the oxidoreductase enzyme family and are widely used in biocatalysis and pharmaceutical applications.[6,49-53] Members of the Old Yellow Enzyme family (OYE; E.C. 1.6.99.1) are the dominate enzymes used for trans-hydrogenation of activated alkene substrates due to their ability to provide up to two stereogenic centers in a single step.[6] Biocatalytic hydrogenation is advantageous over current synthetic methods as it offers metal free catalysis exhibiting high substrate specificity, as well as enantio and stereo-

selectivity.[53] OYEs are found in yeast, plants and bacteria and several have been characterized for their biocatalytic potential. In addition to OYEs, there have only been a few enoate reductases from the iron sulfur flavoproteins (E.C. 1.3.1.31) characterized for the Michael-type trans-hydrogenation.[6,54,55]

1.4.1 Old Yellow Enzyme family

Over 80 years ago, OYE1 was isolated from brewer's bottom yeast (*Saccharomyces carlsbergensis)* and was the first characterized flavoenzyme.[56] OYE is a gold standard in enzymology as several biochemical concepts, such as biological oxidations, have been demonstrated using this enzyme. The laboratory of Vince Massey has extensively characterized OYE, contributing detailed mechanistic, biochemical, and spectroscopic insight to the field.[56] They determined that OYE1 preferentially forms a homo-dimer, is reduced by NADPH and oxidized by molecular oxygen, and produced charge transfer bands upon binding phenolic compounds. The group also conducted a series of FMN cofactor analog studies.[56] The enzyme's physiological function is not known but it has taken precedence in the biocatalysis industry for the reduction of α,β-unsaturated alkene substrates. Preliminary sequence analysis suggests that OYE could have been evolved for sterol metabolism however this has not been verified.[52] Over the years there has been more than 20 OYE homolog's isolated from bacteria, plants and fungi and used for bioreductions.[57] Some of the most well studied enzymes include: OYE1 (brewers bottom yeast), YqjM[58] (*Bacillus subtillis),* pentaerythritol reductase[59] (PETNR; *Enterobacter cloacae* PB2), 12-oxophytodienoate reductase 1[60] (OPR1; tomato), morphinone reductase[61] (MR; *Pseudomonas putida* M10) and NAD(P)H-

dependent 2-cyclohexen-1-one reductase[62] (NCR). Members of the OYE family also show a fair amount of substrate promiscuity. These enzymes catalyze the reduction of the α,β- unsaturated aldehydes, ketones, nitroalkenes, carboxylic acids, imides, lactones and several other derivatives. Members have also been shown to reduce compounds found in explosive formulations such as nitroglycerin, nitroaromatics, nitrate esters, and cyclic triazines.[52]

1.4.2 Structural features and catalytic mechanism of OYE

The OYE family of enzymes can be classified into two separate classes based upon sequence and structural features from mesophilic and thermophilic organisms giving rise to the 'classical' OYEs and the 'thermostable-like' OYEs. Both classes share a characteristic $(\beta\alpha)_{8}$-[49] barrel (TIM barrel) core structure surrounded by several alpha helices and loop structures (Figure 1.7) with a non-covalently bound FMN cofactor. The two classes can be distinguished from one another by significant differences in their loop structures, amino acid composition and sequence.[63] Sequence comparison reveals only a 22 to 32 percent amino acid sequence identity between the two classes. Structurally, the largest difference is in the location and composition of their surface loops. This is shown by comparing the structures of OYE1 from *Saccharomyces pastorianus*, formally *Saccharomyces carlsbergensis* (classical), and YqjM from *Bacillus subtilis* (thermostable-like). The positioning of the loop regions in YqjM do not align well with OYE1 (Figure 1.7). The surface loop regions in the thermostable-like OYEs (Figure 1.7-orange) are shorter and the active sites are much more accessible compared to the classical subfamily (Figure 1.7-cyan). As a result, differences are observed in both

substrate specificity and reactivity between the OYE family members. In addition, changes in quaternary structure are observed with YqjM. Upon substrate binding domain swapping takes place repositioning the C-terminus into the active site.[49]

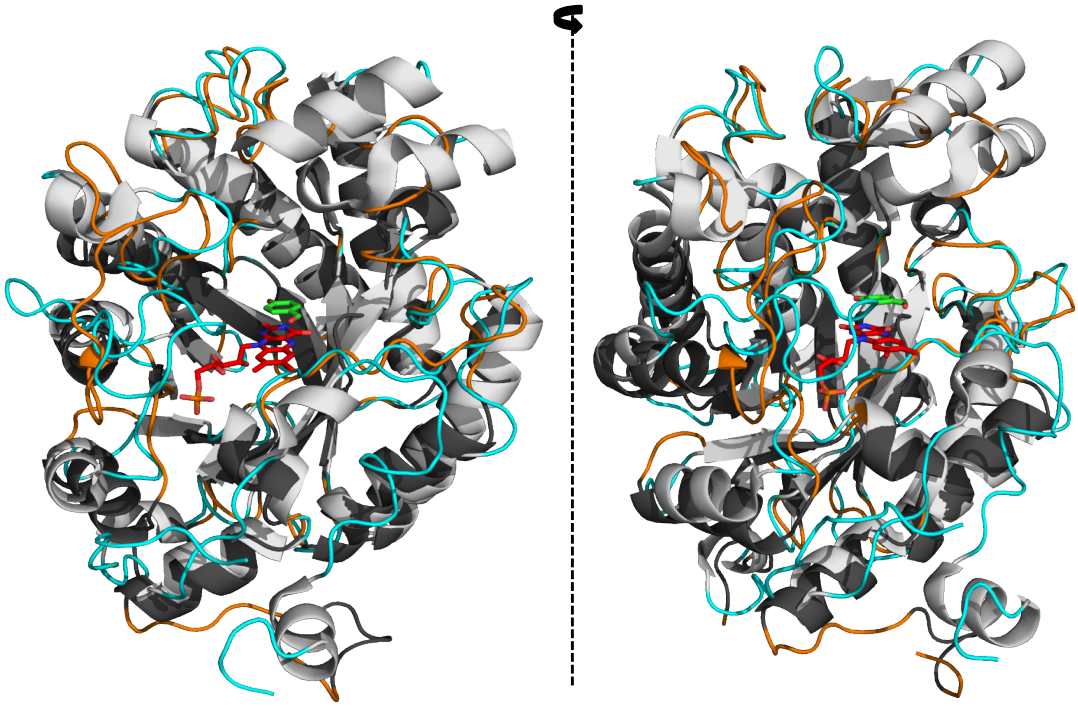


Figure 1.7. Structural comparison of OYE1 (light grey, 1OYB[64]) and YqjM (dark grey, 1Z42[58]) with two different views. The loops are highlighted (OYE1-cyan, YqjM-orange) for structural comparison of the two classes. The FMN (red) and pHB (green) is bound in the active site.

Mechanistically, OYEs catalyze the transhydrogenation of activated alkenes via an ordered bi-bi ping-pong mechanism: a reductive half and an oxidative half (Figure 1.8). In the reductive half FMN is reduced by NADPH followed by a hydride transfer from the reduced FMN to the substrate's β-carbon in the oxidative half. A tyrosine in the active site transfers the second proton to the activated alkene's α-carbon. In addition to

activated alkenes a number of compounds can serve as oxidants for OYEs including molecular oxygen, quinones, ferricyanide and methylene blue. A requirement for catalysis is that the substrates must possess an electron-withdrawing substituent. Substrates with a ketone moiety are activated by either two highly conserved histidines or an asparagine and one histidine in the active site. These two residues are also responsible for positioning the substrate in the correct orientation to the N5 atom of FMN.

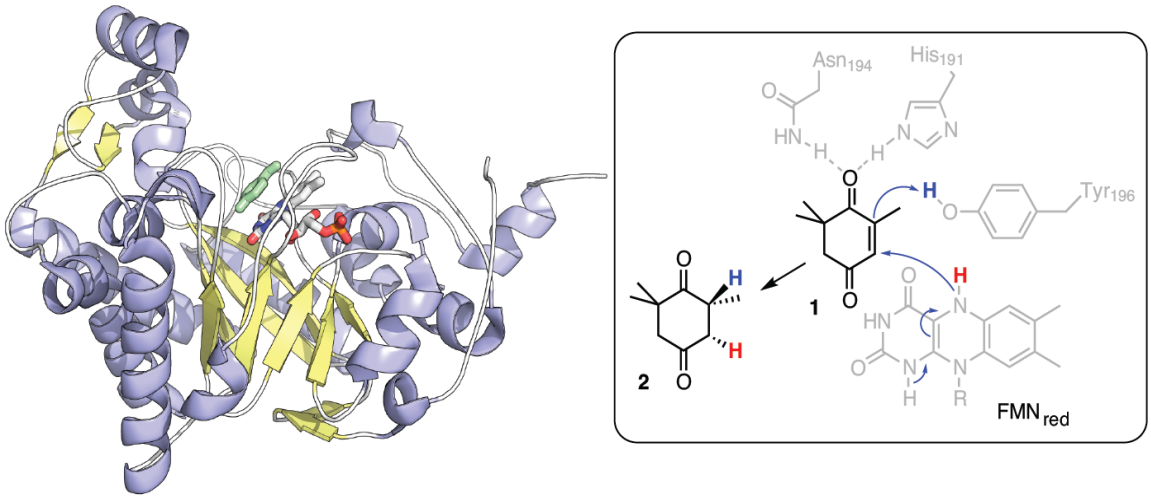


Figure 1.8. Reaction scheme of OYE1-catalyzed ketoisophorone (1) reduction to R-levodione (2). FMN is reduced by NADPH followed by a hydride transfer to the substrate's β-carbon (red) and the hydrogen transferred to the α-carbon (blue) comes from Tyr196.

1.4.3 OYE engineering

Efforts towards engineering OYE family members have been relatively limited mostly focusing on either site-saturation and/or site-directed mutagenesis to improve the biocatalytic functions toward a variety of industrially relevant substrates. Several OYE

homologues have been subjected to these approaches in the hopes of improving activity, as well as enantiopurity, for their respective substrates of interest. In the first example, Bougioukou and coworkers applied directed evolution via iterative CASTing to YqjM to improve enantioselctivity as well as broaden the substrate profile towards a variety of cyclohexenone and cyclopentenone derivatives.[50] From the crystal structure, over 20 sites were initially picked for site-saturation mutagenesis but there wasn't a screen in place to handle these numbers. A pooling strategy was applied enabling the creation of small mutant libraries, which reduced screening efforts. Variants were isolated from second-generation cycles that were improved for the reduction of 3-methylcyclohexenone with either the (*R)* or (*S)* selectivity.[50] Pompeu and coworkers performed site saturation mutagenesis on four active site residues of OYE from *Pichia stipitis* (OYE 2.6).[65] The four residues targeted were Thre35, Ile113, His188, His191, all sites that have previously shown importance for catalytic activity in the OYE family. The libraries were tested against three Baylis-Hillman adducts and the results identified His188/191 to be critical for function. None of the other mutants exhibited improved activity.[65] In a third example, Hulley *et al.* targeted key active site residues of pentraerythritol tetranitrate (PETN) reductase using site saturation mutagenesis.[51] The group developed a high throughput screening procedure involving automated colony picking, library culture generation, and robotic protein purification using Ni-magnetic beads to generate semi-pure protein for activity assays. Two active site mutants, T26S and W102, were identified that showed the most significant changes compared to wild type. Single mutant W102F displayed an increase in *ee* from 13 to 62 % towards methyl-*trans*-cinnamaldehyde and T26S

exhibited a switch in enantiopreference towards methyl-*trans*-cinnamaldehyde and (*E*)-2-aryl-1-nitropropene.[51]

In general, previous engineering efforts have avoided generating a large number of variants due to the lack of effective screening or selection methods. Before combinatorial approaches can be employed, major improvements are needed in the development of higher through put screens or selections as traditional methods using cell-based systems contain contaminating reductases that interfere with the chemistry of interest. In a few cases, mentioned above, screening strategies were developed but they still require some degree of purification to produce semi-pure protein before analysis. Ideally, a more effective library analysis strategy would circumvent this step. As part of this thesis, an *in vitro* trascription/translation protocol has been established that allows rapid screening of hundreds of variants without the need of a purification step. In addition, this screening platform enables the control of cofactor incorporation including the use of cofactor analogs in our screening efforts. This screening strategy was used for the analysis of our OYE1 circular permutation library.

The circular permutation engineering method was chosen for OYE1, over more traditional methods, as this technique has proven to be beneficial in proteins that contain multiple loop and α-helical features that aid in shaping the active site. This technique involves linking the original protein termini together and creating new N- and C- termini throughout the protein structure. In the past, this engineering approach was very successful in creating protein flexibility and active site accessibility translating into significantly improved catalytic activity with maintained enantioselectivity for desired substrates.[66,67] In the case of OYE1, we hypothesize that loop structures are key for

altering enzyme function and that relocation of termini in these regions will prove successful for generation of improved biocatalysts.

1.5 Aim and scope of the dissertation

This dissertation focuses on tailoring the biocatalytic properties of enzymes from three different families using protein-engineering strategies that dramatically reduce the experimental sample size. In each case, designer enzymes were generated that have great potential for applications in their respective fields.

In chapter 2, I report using structural guided rational design and computer-aided design to engineer dCK and *Dm*dNK to improve specificity toward 3'-modified nucleoside analogs over the natural substrate thymidine. Both catalytic and structural features were characterized providing a basis for the altered properties observed. These studies provide evidence for using structural based rational design strategies in tuning substrate specificity in addition to identifying residues that are critical for altering the specificity towards nucleoside analogs in this class.

In chapter 3, the Rosetta Molecular Modeling suite was used to redesign the propeptide of subtilisin BPN', generating synthetic peptides that exhibit increased and tunable structural stability. Molecular dynamics simulations provide supporting evidence that the artificial sequences retain structure in the absence of cognate protease, unlike the wild type propeptide. Experimental evaluation of two designer variants by spectroscopic methods display significantly enhanced thermostability. However their functional performance as protease inhibitors raises doubt that propeptide stability alone is sufficient for effective inhibitor design.

In chapter 4, circular permutation was employed to engineer OYE from *Saccharomyces pastorianus* (OYE1). The OYE1 circular permuted library was directly synthesized using a whole gene synthesis strategy and assayed against several α,β-unsaturated alkene substrates utilizing the PURExpress *in vitro* transcription/translation (IVT) technology, demonstrating a novel screening approach. Library evaluation using our IVT screening platform successfully identified ~30% of the OYE1 library having better activity than wild type. The results also clearly highlight three specific sectors in the OYE1 scaffold responsible for the most dramatic alterations in catalytic activity for all substrates evaluated. Representative permutants were selected for a more through kinetic characterization. In addition to the changes in substrate specificity as a result of permutation, chapter 4 also explores the substitution of FMN with cofactor analogs in CP variants as an orthogonal engineering strategy. I demonstrate alterations in catalytic activity, as well as the catalysis of the reverse reaction, as a result of changes in redox potential via tandem effects through both cofactor substitution and permutation. In chapter 5, I investigate the structural basis for the altered catalytic properties as a result of permutation using x-ray crystallography. In addition, selected variants are characterized for secondary structural changes, thermostablity, oligomeric state and changes in active site via mutagenesis.

1.6 References

1 Alcalde, M., Ferrer, M. & Plou, F. J. Environmental biocatalysis: From remediation with enzymes to novel green processes. *Biocatal. Biotransform.* **25**, 113-113, (2007).

2 Alcalde, M., Ferrer, M., Plou, F. J. & Ballesteros, A. Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends Biotechnol.* **24**, 281-287, (2006).

3 Woodley, J. M. New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends Biotechnol.* **26**, 321-327, (2008).

4 International Union of Biochemistry and Molecular Biology. Nomenclature Committee. & Webb, E. C. *Enzyme nomenclature 1992 : recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*. (Published for the International Union of Biochemistry and Molecular Biology by Academic Press, 1992).

5 Lutz, S. Beyond directed evolution-semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**, 734-743, (2010).

6 Toogood, H. S. & Scrutton, N. S. Enzyme engineering toolbox – a 'catalyst' for change. *Catal. Sci. Tech.*, (2013).

7 Wang, M., Si, T. & Zhao, H. M. Biocatalyst development by directed evolution. *Bioresour. Technol.* **115**, 117-125, (2012).

8 Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14**, 438-443, (2003).

9 Dalby, P. A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **21**, 473-480, (2011).

10 Goldsmith, M. & Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **22**, 406-412, (2012).

11 Brustad, E. M. & Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **15**, 201-210, (2011).

12 Damborsky, J. & Brezovsky, J. Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.* **13**, 26-34, (2009).

13 Rubin-Pitel, S. B., Cho, C. M. H., Chen, W. & Zhao, H. M. Directed Evolution Tools in Bioproduct and Bioprocess Development. *Bioprocessing for Value-Added Products from Renewable Resources: New Technologies and Applications*, 49-72, (2007).

14 Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939-942, (2008).

15 Stemmer, W. P. C. Rapid Evolution of a Protein in-Vitro by DNA Shuffling. *Nature* **370**, 389-391, (1994).

16 Stemmer, W. P. C. DNA Shuffling by Random Fragmentation and Reassembly - in-Vitro Recombination for Molecular Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10747-10751, (1994).

17 Chen, C. Y., Georgiev, I., Anderson, A. C. & Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3764-3769, (2009).

18 Chen, M. M. Y., Snow, C. D., Vizcarra, C. L., Mayo, S. L. & Arnold, F. H. Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng. Des. Sel.* **25**, 171-178, (2012).

19 Cerdobbel, A. *et al.* Increasing the thermostability of sucrose phosphorylase by a combination of sequence- and structure-based mutagenesis. *Protein Eng. Des. Sel.* **24**, 829-834, (2011).

20 Yuan, Z., Bailey, T. L. & Teasdale, R. D. Prediction of protein B-factor profiles. *Proteins: Struct. Funct. Bioinform.* **58**, 905-912, (2005).

21 Jäckel, C., Bloom, J. D., Kast, P., Arnold, F. H. & Hilvert, a. D. Consensus Protein Design Without Phylogenetic Bias. *J. Mol. Biol.* **399**, 541-546, (2010).

22 Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382-W388, (2005).

23 Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **6**, (2011).

24 Liu, Y. & Kuhlman, B. RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, W235-W238, (2006).

25 Kim, H. S., Quang, A. T. L. & Kim, Y. H. Development of thermostable lipase B from Candida antarctica (CalB) through in silico design employing B-factor and RosettaDesign. *Enzyme Microb. Technol.* **47**, 1-5, (2010).

26 Yuan, Z., Bailey, T. L. & Teasdale, R. D. Prediction of protein B-factor profiles. *Proteins-Structure Function and Bioinformatics* **58**, 905-912, (2005).

27 Gordon, S. R. *et al.* Computational Design of an alpha-Gliadin Peptidase. *J. Am. Chem. Soc.* **134**, 20513-20520, (2012).

28 Robak, T. New nucleoside analogs for patients with hematological malignancies. *Expert Opin. Invest. Drugs* **20**, 343-359, (2011).

29 Cihlar, T. & Ray, A. S. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.* **85**, 39-58, (2010).

30 Johansson, K. *et al.* Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. *Nat. Struct. Biol.* **8**, 616-620, (2001).

31 Jarchow-Choy, S. K., Sjuvarsson, E., Sintim, H. O., Eriksson, S. & Kool, E. T. Nonpolar Nucleoside Mimics as Active Substrates for Human Thymidine Kinases. *J. Am. Chem. Soc.* **131**, 5488-5494, (2009).

32 Arner, E. S. J. & Eriksson, S. Mammalian Deoxyribonucleoside Kinases. *Pharmacol. Ther.* **67**, 155-186, (1995).

33 Thelander, L. & Reichard, P. Reduction of Ribonucleotides. *Annu. Rev. Biochem.* **48**, 133-158, (1979).

34 Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discovery* **12**, 447-464, (2013).

35 Lutz, S., Liu, L. F. & Liu, Y. C. Engineering Kinases to Phosphorylate Nucleoside Analogs for Antiviral and Cancer Therapy. *Chimia* **63**, 737-744, (2009).

36 Ardiani, A. *et al.* Enzymes To Die For: Exploiting Nucleotide Metabolizing Enzymes for Cancer Gene Therapy. *Curr. Gene Ther.* **12**, 77-91, (2012).

37 Eriksson, S., Munch-Petersen, B., Johansson, K. & Eklund, H. Structure and function of cellular deoxyribonucleoside kinases. *Cell. Mol. Life Sci.* **59**, 1327-1346, (2002).

38 Wild, K., Bohner, T., Folkers, G. & Schulz, G. E. The structures of thymidine kinase from Herpes simplex virus type 1 in complex with substrates and a substrate analogue. *Protein Sci.* **6**, 2097-2106, (1997).

39 Mikkelsen, N. E. *et al.* Structural basis for feedback inhibition of the deoxyribonucleoside salvage pathway: Studies of the Drosophila deoxyribonucleoside kinase. *Biochemistry* **42**, 5706-5712, (2003).

40 Godsey, M. H., Ort, S., Sabini, E., Konrad, M. & Lavie, A. Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: Illuminating the role of main-chain reorganization. *Biochemistry* **45**, 452-461, (2006).

41 Knecht, W. *et al.* Drosophila deoxyribonucleoside kinase mutants with enhanced ability to phosphorylate purine analogs. *Gene Therapy* **14**, 1278-1286, (2007).

42 Culver, K. W. *et al.* In Vivo Gene Transfer with Retroviral Vector-Producer Cells for Treatment of Experimental Brain Tumors. *Science, New Series, Vol. 256, No. 5063 (Jun. 12,), pp.* **256**, 1550-1552, (1992).

43 Frobert, E. *et al.* Resistance of herpes simplex virus type 1 to acyclovir:Thymidine kinase gene mutagenesis study. *Antiviral Res.* **73** 147-150, (2007)

44 Knecht, W. *et al.* Drosophila melanogaster deoxyribonucleoside kinase activates gemcitabine. *Biochem. Biophys. Res. Commun.* **382**, 430-433, (2009).

45 Moolten, F. L. Tumor Chemosensitivity Conferred by Inserted Herpes Thymidine Kinase Genes:Paradigm for a Prospective Cancer Control Strategy. *Cancer Res.* **46**, 5276-5281, (1986).

46 Eriksson, S., Munch-Petersen, B., Johansson, K. & H. Eklundc. Structure and function of cellular deoxyribonucleoside kinases. *Cell. Mol. Life Sci.* **59**, 1327-1346, (2002).

47 He´brard, C., Dumontet, C. & Jordheim, L. Development of gene therapy in association with clinically used cytotoxic deoxynucleoside analogues. *Cancer Gene Ther.*, 1–10, (2009).

48 Bordusa, F. Proteases in organic synthesis. *Chem Rev* **102**, 4817-4868, (2002).

49 Takagi, H. & Takahashi, M. A new approach for alteration of protease functions: pro-sequence engineering. *Appl. Microbiol. Biotechnol.* **63**, 1-9, (2003).

50 Toogood, H. S., Gardiner, J. M. & Scrutton, N. S. Biocatalytic Reductions and Chemical Versatility of the Old Yellow Enzyme Family of Flavoprotein Oxidoreductases. *Chemcatchem* **2**, 892-914, (2010).

51 Bougioukou, D. J., Kille, S., Taglieber, A. & Reetz, M. T. Directed Evolution of an Enantioselective Enoate-Reductase: Testing the Utility of Iterative Saturation Mutagenesis. *Adv. Synth. Catal.* **351**, 3287-3305, (2009).

52 Hulley, M. E. *et al.* Focused Directed Evolution of Pentaerythritol Tetranitrate Reductase by Using Automated Anaerobic Kinetic Screening of Site-Saturated Libraries. *ChemBioChem* **11**, 2433-2447, (2010).

53 Williams, R. E. & Bruce, N. C. 'New uses for an Old Enzyme' - the Old Yellow Enzyme family of flavoenzymes. *Microbiol-SGM* **148**, 1607-1614, (2002).

54 Durchschein, K., Hall, M. & Faber, K. Unusual reactions mediated by FMN-dependent ene- and nitro-reductases. *Green Chem.* **15**, 1764-1772, (2013).

55 Buhler, M. & Simon, H. On the Kinetics and Mechanism of Enoate Reductase. *Hoppe-Seyler's Z. Physiol. Chem.* **363**, 609-625, (1982).

56 Giesel, H. & Simon, H. On the Occurrence of Enoate Reductase and 2-Oxo-Carboxylate Reductase in Clostridia and Some Observations on the Amino-Acid Fermentation by Peptostreptococcus-Anaerobius. *Arch. Microbiol.* **135**, 51-57, (1983).

57 Karplus, P. A., Fox, K. M. & Massey, V. Flavoprotein structure and mechanism .8. Structure-function relations for old yellow enzyme. *FASEB J.* **9**, 1518-1526, (1995).

58 Winkler, C. K., Tasnadi, G., Clay, D., Hall, M. & Faber, K. Asymmetric bioreduction of activated alkenes to industrially relevant optically active compounds. *J. Biotechnol.* **162**, 381-389, (2012).

59 Kitzing, K. *et al.* The 1.3 A crystal structure of the flavoprotein YqjM reveals a novel class of Old Yellow Enzymes. *J. Biol. Chem.* **280**, 27904-27913, (2005).

60 French, C. E., Nicklin, S. & Bruce, N. C. Sequence and properties of pentaerythritol tetranitrate reductase from Enterobacter cloacae PB2. *J. Bacteriol.* **178**, 6623-6627, (1996).

61 Biesgen, C. & Weiler, E. W. Structure and regulation of OPR1 and OPR2, two closely related genes encoding 12-oxophytodienoic acid-10,11-reductases from Arabidopsis thaliana. *Planta* **208**, 155-165, (1999).

62 French, C. E. & Bruce, N. C. Purification and Characterization of Morphinone Reductase from Pseudomonas-Putida M10. *Biochem. J.* **301**, 97-103, (1994).

63 Muller, A., Hauer, B. & Rosche, B. Asymmetric alkene reduction by yeast old yellow enzymes and by a novel Zymomonas mobilis reductase. *Biotechnol. Bioeng.* **98**, 22-29, (2007).

64 Reich, S., Kress, N., Nestl, B. M. & Hauer, B. Variations in the stability of NCR ene reductase by rational enzyme loop modulation. *J. Struct. Biol.*, (2013).

65 Fox, K. M. & Karplus, P. A. Old Yellow Enzyme at 2-Angstrom Resolution - Overall Structure, Ligand-Binding, and Comparison with Related Flavoproteins. *Structure* **2**, 1089-1105, (1994).

66 Pompeu, Y. A., Sullivan, B., Walton, A. Z. & Stewart, J. D. Structural and Catalytic Characterization of Pichia stipitis OYE 2.6, a Useful Biocatalyst for Asymmetric Alkene Reductions. *Adv. Synth. Catal.* **354**, 1949-1960, (2012).

67 Reitinger, S. *et al.* Circular Permutation of Bacillus circulans Xylanase: A Kinetic and Structural Study. *Biochemistry* **49**, 2464-2474, (2010).

68 Qian, Z. & Lutz, S. Improving the catalytic activity of Candida antarctica lipase B by circular permutation. *J. Am. Chem. Soc.* **127**, 13466-13467, (2005).

Chapter 2: Engineering Deoxyribonuceloside Kinases for 3'-Modified Nucleoside Analogs

2.1 Introduction

A number of nucleoside analogs (NA) have modifications at the 3' position of the sugar moiety and act by inhibiting DNA synthesis.[1-3] The analog is in its prodrug state until it is further activated to the triphosphate form where it can then exert its mode of action. A hallmark representative of this class is AZT, which is a thymidine analog containing a 3' azido group.[2] It was the first drug approved for HIV therapy and is a very effective RT inhibitor.[1] All though this analog proved to be very effective, activation of these prodrugs is often compromised by the deoxyribonucleoside kinase's presiding affinity for the natural substrate over the NA. One strategy to overcome this limitation is to design a deoxyribonucleoside kinase specifically for the phosphorylation of the analog.

The type-1 kinase *Dm*dNK is an ideal starting point for evolving orthogonal nucleoside analog kinases because it displays the highest activity and greatest promiscuity for both the natural 2'-deoxyribonucleosides and several NAs.[4,5] The phosphorylation rate of natural deoxyribonucleosides by *Dm*dNK has been reported to have a 4-20,000 fold greater turnover compared to the human kinases. In addition, compared to many of the human dNKs, this enzyme is easily obtainable through standard expression and purifications techniques and is relatively stable. All of these properties make *Dm*dNK a good template for rational and random mutagenesis experiments.[5]

Previous work in our laboratory led to the directed evolution of a nucleoside analog kinase for 2',3' –dideoxythymidine (ddT) (Figure 2.1), a representative NA that has proven to have compromised biological function due to lack of phosphorylation.[6] Directed evolution experiments rendered mutant R4.V3, containing four mutations T85M, E172V, Y179F and H193Y that displayed a switched substrate specificity for ddT

over Thy by 30-fold.[6] Reverse engineering was completed to assess the contribution of each of the four substations, suggesting that the two mutations responsible for the substrate specificity switch were E172V and Y179F.

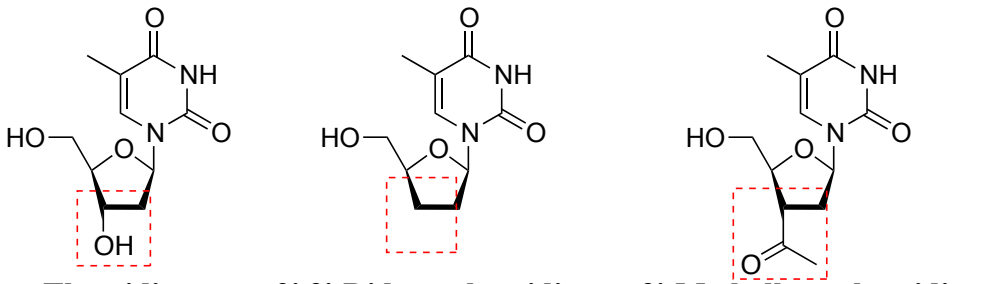


Figure 2.1. Deoxyribonucleosides described in this chapter.

In wild type *Dm*dNK both residues E172 and Y179 are in close proximity to the substrate ribose moiety and participate in hydrogen bonding with the 3' hydroxyl group. When the amino acids are mutated to residues with hydrophobic side chains, they are no longer able to form hydrogen bonds, thus gaining specificity for ddT. Although *Dm*dNK is the most robust deoxyribonucleoside kinase, there are limitations to using this enzyme in gene therapy approaches. The main concern is the risk of an undesirable immunogenic response in clinical applications.[3] For this reason, our lab was interested in exploring the transfer of these mutations to homologous positions in another highly similar type-1 dNK human family member, hdCK in order to see if the same specificity switch could be observed (Figure 2.2). This concept of transferring mutations between the two kinases is important for understanding if initial engineering can start out in a more compliant kinase and be transferred, at later stages, to a human enzyme for therapeutic purposes. This

strategy would be advantageous over directly evolving a human enzyme as *Dm*dNK does not suffer from low expression yields, has been shown to be relatively stable, and has demonstrated ability to tolerate multiple mutations.[4-7]

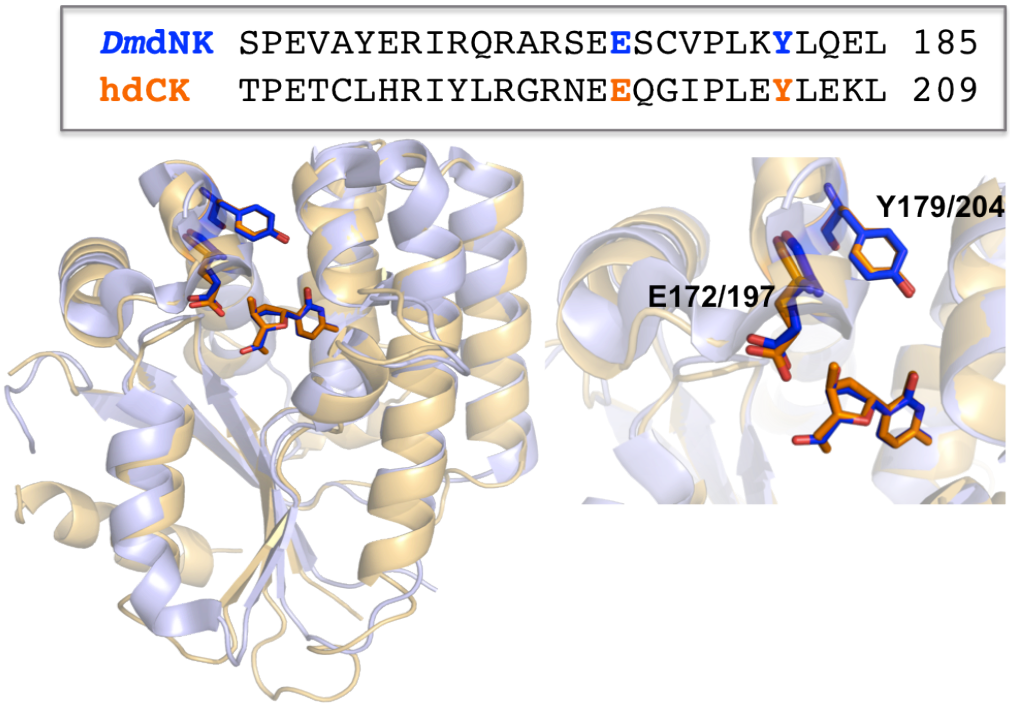


Figure 2.2. Sequence alignment of the highly conserved lid region of dCK (PDB code: 2A30) and DmdNK (PDB code: 1OT3). Residues responsible for ddT specificity switch highlighted in blue and homologous positions in dCK shown in orange.

In addition to ddT, I also focused on exploring 3'-methylketo-Thy (kT), a Thy analog that has a ketone group substituted at the 3'-hydroxyl group (Figure 2.2). The motivation for pursing this compound is that reverse transcriptase (RT) has a large 3' pocket suitable for analogs with a branched modification at the 3' position, such as kT, presumably allowing for a better fit in the polymerase active site.[8] This branched 3'

modification is expected to be advantageous over existing nucleosides because it is hypothesized to act more selectively with RT over other DNA polymerases lacking such a pocket. Liu *et al.* showed that kT-triphosphate was an effective HIV-RT inhibitor in an *in vitro* assay involving primer extension. However, *in vivo* tests of RT inhibition were unsuccessful, presumably due to lack of prodrug phosphorylation by nucleoside kinases. To create a kT-specific dNK, we chose computer-aided rational design to engineer *Dm*dNK. This strategy was chosen over more traditional protein engineering approaches, such as directed evolution, to allow for a more rapid search of sequence space. In addition, the activity of *Dm*dNK for kT is poor; therefore, computational tools are advantageous to accelerate the initial discovery process. Once modest enzyme activities are established, other strategies could be explored, hopefully requiring fewer evolutionary iterations.

In this chapter, I use a combination of semi-rational design strategies to engineer two dNKs to improve specificity toward 3'-modified nucleoside analogs ddT and kT. In the first part of the project, I transferred beneficial mutations between two deoxyribonucleoside kinases (*Dm*dNK, dCK) which resulted in a similar specificity switch, giving rise to a human ddT kinase. In the second project, we used the Rosetta software suite to investigate mutations that enabled the phosphorylation of kT by *Dm*dNK. Several iterations of Rosetta, in combination with rational design, resulted in functional variants with improved specificity.

2.2 Results and discussion

2.2.1 Transfer of beneficial mutations between DmdNK and hdCK

Comparison of the dNKs reveals that hdCK and *Dm*dNK share a sequence identity of 34.6 % and that their structures are superimposable with an RMSD of 1.45Å (PDB; *Dm*dNK: 1OT3,[9] hdCK: 2A30[10]). Based on the high similarities between the two kinases, we believe a similar specificity switch is obtainable upon the transfer of mutations to homologous E197 and Y204 positions in hdCK. Direct transfer of the mutations into wild type hdCK cannot be made because the enzyme naturally lacks Thy kinase activity. However, earlier studies by the Lavie *et al* and Lutz *et al* involved reengineering the active site of dCK to be an efficient Thy kinase.[11] Two top performing variants, dCK1 (A100V/R104M/D133S) and dCK2 (A100V/R104M/D133T), from these experiments were chosen as the templates for transfer experiments. Mutations E197V and Y204F were introduced into dCK1 and dCK2 both individually and as double mutants via site-directed mutagenesis. Initially, all genes were inserted into the pET-14b vector, but very little soluble protein was produced (< 5 mg/L), making subsequent characterization experiments difficult. Constructs were moved to the pMAL-c2X vector system, which provided a C-terminal maltose binding protein fusion that has been demonstrated as a solubility enhancer. This system significantly increased protein yields to (~15 mg/L) and did not require tag cleavage, as the tag did not interfere with subsequent characterization.

Kinetic experiments were completed with Thy and analog ddT (Table 2.1). Steady state kinetic parameters were determined using an established coupled enzyme assay that has been applied to numerous dNKs for characterization.[17] Kinetic parameters for Thy

were obtained for the E197V single mutant constructs, though no activity for ddT was detected, presumably due to increased enzyme instability. Mutant dCK1-E197V exhibited >1000-fold decline in Thy activity while > 2000-fold decline was observed for dCK2-E197V respectively. These decreases in catalytic performance could largely be attributed to increases in $K_{M.}$ For dCK1-E197V, k_{cat} values decreased ~ four-fold compared to dCK1 and dCK2-E197V decreased 10-fold compared to dCK2 with substrate Thy. The substitution of Y204F also resulted in an activity loss for Thy, but not as significantly as the previous substitution. Kinetic analysis could not be completed for the double mutants containing both E197V and Y204F as no activity was detected with Thy or ddT. Upon expression of the double mutant, soluble protein was produced but there was simply no activity observed. Consequently, we hypothesized that an amino acid with similar hydrophobic character would be more ideal for ddT specificity.

Table 2.1. Kinetic Properties of dCK variants with Thy and ddT[a]

	Thy			ddT			
Enzyme	**K_M (μM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($M^{-1}s^{-1}$) × 10^3**	**K_M (μM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($M^{-1}s^{-1}$) × 10^3**	**Specificity[b]**
dCK1	7.31 ± 1.02	0.36 ± 0.04	59	642.83 ± 42.65	0.092 ± 0.03	0.142	415
dCK2	5.62 ± 1.46	0.99 ± 0.1	177	687.86 ± 110.42	0.093 ± 0.09	0.135	1311
dCK1-E197V	1923.79 ± 807.31	0.095 ± 0.4	0.049	ND	ND	ND	--
dCK1-Y204F	86.41 ± 17.8	0.74 ± 0.07	8.6	354.31 ± 52.46	0.12 ± 0.01	0.34	25
dCK2-E197V	1251.87 ± 448.20	0.10 ± 0.2	0.082	ND	ND	ND	--
dCK2-Y204F	144.86 ± 16.82	0.25 ± 0.01	1.7	301.08 ± 82.27	0.11 ± 0.03	0.36	5

[a]Kinetic parameters were determined by a nonlinear fit to the Michaelis-Menten equation. Relative specificity changes compared to parent enzyme are shown in parentheses. ND=Non-detectable. [b]Ratio Thy activity to ddT activity.

2.2.2 Site saturation mutagenesis at position E197

Based on results from the kinetic characterization of the rational dCK mutants, a small library with various hydrophobic amino acids at position E197 in dCK1 was created. Using site-directed mutagenesis, hydrophobic substitutions were made to include amino acids I, L, A, F, M, S, and T. Separately, the variants were combined with the Y204F substitution to investigate further beneficial effects. A preliminary kinetic screen was conducted to determine the specific activity of the E197x mutants. Activity assays revealed four mutants with a preference for ddT over Thy (Table 2.2), with the top performer, dCK1-E197I, showing the largest specificity switch. The dCK1-E197I mutant was further characterized to obtain full kinetic parameters (Table 2.3). The kinetic data revealed that the E197I substitution in dCK1 results in a substrate specificity switch, eliminating Thy activity and maintaining a relatively high catalytic performance for ddT. In fact, the mutation lowered the K_M 10-fold for ddT suggesting a more favorable binding conformation for the preferred substrate.

In addition, earlier work demonstrated that the introduction of a Val at position 100 in the dCK1 mutant increased the catalytic performance for Thy by over 2-fold but decreased the thermal stability over 8 °C.[11] In a follow up experiment, the Val mutation at position 100 was reversed back to an Ala to see if this would help with the stability issues that were exhibited by the ssTK1A-E197x mutants. The reversal of V100 back to an Ala had a detrimental effect on the binding constant, lowering the overall catalytic performance by >7-fold. The second mutation that was transferred from *Dm*dNK, Y204F, was subsequently introduced into dCK1-E197I. However, negative effects on stability were observed that prevented further characterization. Presumably, the accumulation of

five active site amino acid substitutions in dCK1-E197I-Y204F destabilizes the enzyme, making it nonfunctional. We were also unable to assess the thermal denaturation of dCK1-E197I via CD measurements due to the fusion protein interference. Instead, an activity-based stability measurement was used to assess the impact of the E197I substitution on the dCK1 template. The enzymes were incubated at 4°C and 30°C for 7 min prior to measuring enzyme activity using the standard kinetic assay with ddT (V_{max} concentrations). The residual activity was determined from the V_{max} values calculated at the higher temperature compared to the lower. After incubation, dCK1-E197I maintained 80% activity compared to dCK1. These results suggest that even though a slight decrease is observed upon the introduction of E197I, the enzyme behaved significantly better than the E197V substitution. It is interesting to note that different hydrophobic amino acids at position 172 in *Dm*dNK had been explored in previous work, with the substitution of Ile also exhibiting the highest thermostability (~ 2-fold higher than E172V).[7] The rationalization behind this phenomenon is that the tighter packing of the Ile *sec*-butyl side chain with ddT compared to the isopropyl group of the Val side chain at E172/E179 in *Dm*dNK and dCK results in the increased stability (Figure 2.3).

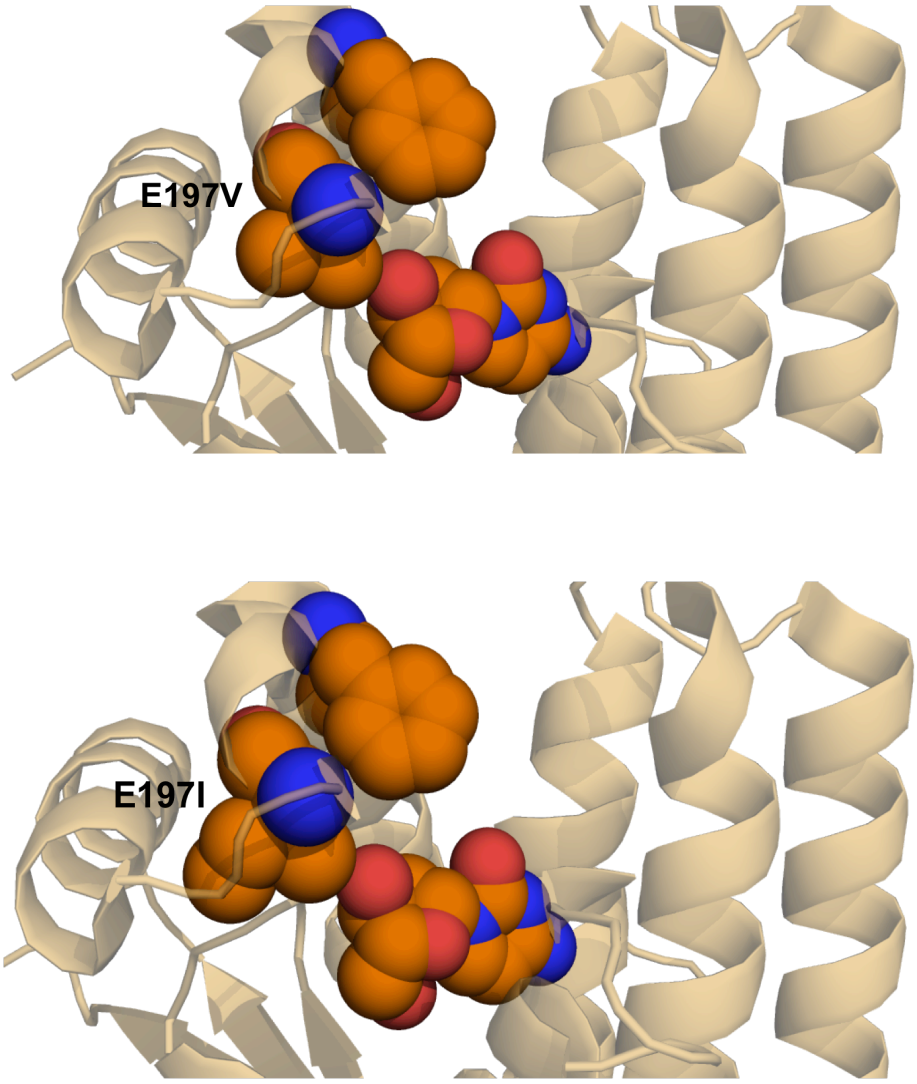


Figure 2.3. Structure of dCK (PDB code: 2A30) with ddT bound. The structural model suggests tighter packing with the branched extension of the Ile sec-butyl side chain a E197 compared to a isopropyl group of the Val with ddT .

Table 2.3. Limited site-saturation mutagenesis: specific activities at 500 μM Thy and ddT

Enzyme	Thy	ddT	Specificity[b]
WT-dCK	ND	6.5	ND
dCK1	550	77	0.1
dCK1-E197V	36	ND	ND
dCK1-E197I	11	66	6.0
dCK1-E197L	10	40	3.6
dCK1-E197A	31	34	3.1
dCK1-E197M	52	2	0.2
dCK1-E197F	ND	12	1.1
dCK1-E197S	24	42	3.8
dCK1-E197T	25	30	2.7

[a]Activities (μM/min per mg enzyme) were determined by spectrophotometric assay at a constant substrate concentration of 500 μM (standard error: ± 10%).[b]Ratio ddT activity to Thy activity.

Table 2.4. Kinetic properties of dCK1-E197I variants with Thy and ddT

Substrate	Enzyme	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$) × 10^3
Thy	dCK1-E197I	ND	ND	ND
	dCK1-A100-E197I	>1000	<0.01	---
ddT	dCK1-E197I	62 ± 10	0.06 ± 0.01	0.942
	dCK1-A100-E197I	502.7 ± 88.4	0.064 ± 0.004	0.127

[a]Kinetic parameters were determined by a nonlinear fit to the Michaelis-Menten equation. ND = Non-detectable.

2.2.3 Rosetta designer kinase for kT

Computational approaches present an alternative way to design kinase variants for experimental exploration.[12-15] *In silico* design allows for a faster and more comprehensive search of sequence space compared to other protein engineering methods.[12-15] The Rosetta software suite was used to redesign the active site of *Dm*dNK to better accommodate kT via favorable steric and hydrogen bonding interactions in order to allow for more efficient phosphorylation.[14,15] Using the crystal structure of *Dm*dNK with kT modeled in the active site, fixed backbone design (side chain replacement using crystal structure backbone coordinates) was applied allowing the computer to sample residues within 8 Å of the 3' position of the substrate. From the simulations, a set of five positions (L66H, M69R, Y70D, E172G, V175D) were identified around the substrate binding pocket and substitutions for these positions were chosen based the proposed interactions with the ketone moiety. The initial substitutions were rationalized via a predicted hydrogen-bonding network. The substitution of an Asp at position 70 creates a network of hydrogen bonding between side chains of residues L66H, Y70D, M69R, and Y179, enabling a tight fit around kT. A Gly at E172 introduces more side chain flexibility and His at L66 is in proximity to form hydrogen bonding with the carbonyl oxygen of the keto moiety. The corresponding RD-5 construct (L66H, M69R, Y70D, E172G, V175D) was created via site directed mutagenesis and cloned into the pET-14b. After purification, SDS-PAGE analysis revealed that all of the protein was eluted in the void volume of the gel filtration step, indicating aggregate formation. Closer examination of the active site revealed that M69R is in proximity to form a salt bridge with surrounding residues; therefore, it was eliminated from the design. To narrow down the remaining residues

responsible for aggregation, two constructs were tested that had the mutations in the same vicinity during initial creation of RD-5, DmdNK-172A-175D and DmdNK-66H-70D. After purification it was clear that aggregation was due to either 172A and/or 175D (Figure 2.4). Upon reversal of position 175 back to a Val, the triple mutant L66H-Y70D-E172A (Figure 2.5) was purified, eluting at the volume corresponding to a dimer according to the gel filtration standards (Figure 2.4). The destabilization is most likely a result of V175D, in which a highly conserved hydrophobic residue was mutated to a charged residue.

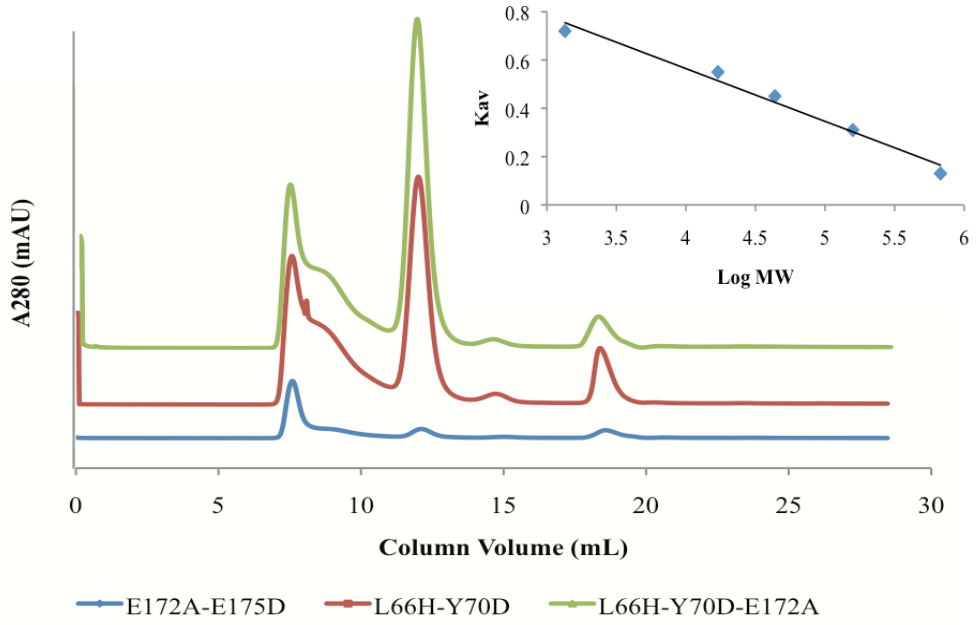


Figure 2.4. Gel filtration chromatograms of first generation designs DmdNK-172A-175D (blue) indicated almost complete aggregate eluting in the void volume at 7-8 mLs, DmdNK-66H-70D and the final triple mutant are in the correct dimer oliogomeric state eluted between 12-13 mLs (Elution Volume: ~12.1 mLs; MW ~150 kDa): The inset shows the standard molecular weight curve generated using Bio-Rad's gel filtration standards. The molecular weight of each peak was calculated using this curve.

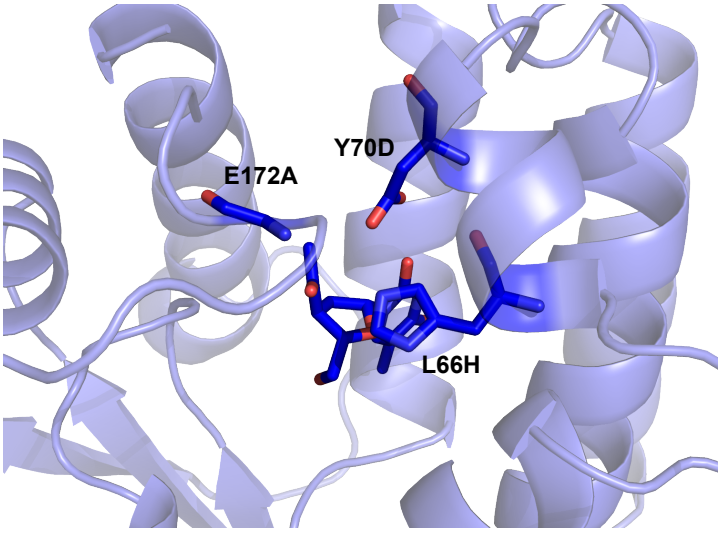


Figure 2.5. Rosetta model of the first generation design of DmdNK with kT bound after the residues responsible for aggregation were reversed back to their normal identity.

A second round of Rosetta simulations were conducted based upon the stability information obtained from the first set of constructs. Substitutions E172A and L66H were held constant and the other two residues were allowed to vary, but no changes were made to the conformation of the backbone. Four of the lowest scoring combinations were chosen for further study (Table 2.4).

Table 2.5. Summary of amino acid substitutions in second generation designs.

	Amino acid residue			
Variant	**66**	**70**	**172**	**175**
*Dmd*NK	L	Y	E	E
RD-1	H	M	A	
RD-2	H	M	A	A
RD-3	H	I	A	L
RD-4	H	V	A	L

Purification revealed that V175L mutation in RD-3 and RD-4 also resulted in aggregation, leaving RD, RD-1 and RD-2 for further characterization (Figure 2.6). Activity assays were completed using 3-methyl-keto-uridine (kU) as our model compound in place of kT. The subtle differences between the uracil and thymine nucleobases should not impact function, but allow for modifications at the nucleobase 5-position for generation of future analogs. The specific activity was measured for kU and Thy at 200 μM for the second-generation designs (Table 2.4). Among the substitutions, L66H and Y70D/M, lowered the activity for Thy by 43 to 83-fold. Small declines in kU activity are also observed, resulting in specific activities lower than desired. The thermal stability of the second-generation designs was measured via CD following the signal at 222 nm. We wanted to make sure that a dramatic decrease in stability was not observed as the first round of constructs resulted in aggregation problems. The melting temperatures were slightly lower for all variants with RD-2 displaying the lowest T_M of 48°C compared to *Dm*dNK at 58°C. Although progress was seen with the second round of designs, something was missing in the models to promote kT preference.

A. B.

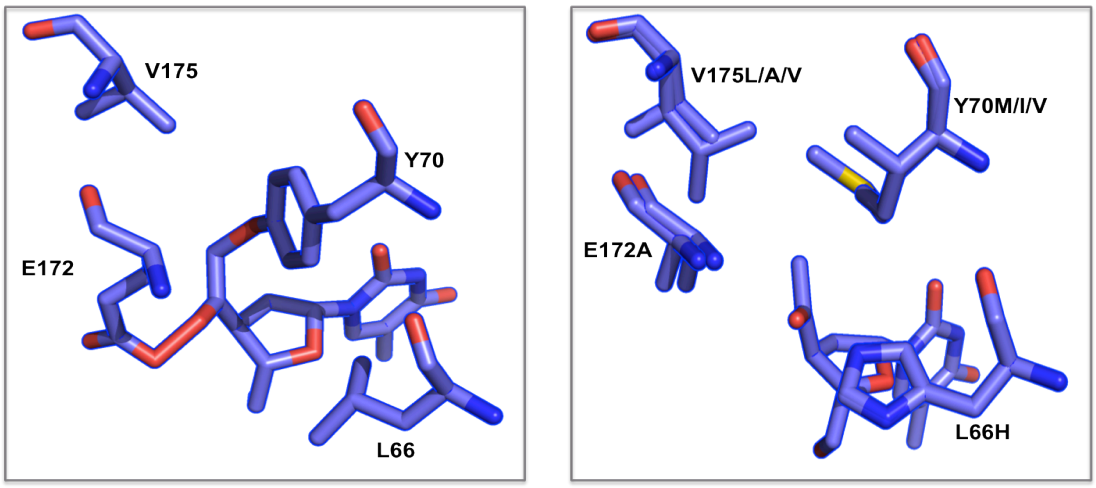


Figure 2.6. Second generation Rosetta models. A) Active site residues of wild type DmdNK (PDB: 1OT3) with kT modeled in the active site. The residues identified by Rosetta important for Thy to kT specificity switch are highlighted. B) Overlay of the suggested mutations resulting in favorable interactions and sterics with respect to kT.

Table 2.6. Second Generation Rosetta Designs: Substrate Specificities at 200 µM Thy and kU[a]

Variant	kU	Thy
DmdNK	275	15936
DmdNK-E172A	154	2052
DmdNK-66H-70D	<30	168
RD (L66H-Y70D-V175-E172A)	<30	387
RD-1 (L66H-Y70M-V175-E172A)	<30	197
RD-2 (L66H-Y70M-V175A-E172A)	<30	223

Activities (µM/min per mg enzyme) were determined by spectrophotometric assay at a constant substrate concentration of 200 µM (standard error: (10%).

2.2.4 Refinement of kT designer kinase

Several questions were raised before running additional simulations: 1) is ligand rotation playing a role, and if so, what is the preferred conformation of kT; 2) are other

residues being overlooked; and 3) is L66H really serving a H-bonding role to benefit kT binding? Ligand rotation was investigated as we hypothesized that the rotation of kT around the C3'-CO bond could substantially impact the predictions. Addressing this question could also account for the other concerns if an optimal binding orientation is found. For the third round of calculations, the substrate was docked in the active site at 90° rotations in each of the four orientations (Figure 2.7). The scores and sequences after 20 runs, in each orientation, were compiled and subsequently compared to construct a new model (Figure 2.8). The predictions favored the 180° orientation of the ketone moiety, as these were the lowest energy designs. The suggested substitution for E172 was a Gly and predictions for position 70 favored small hydrophobic residues. The E172G mutation creates more space for the branched 3'-group and suggestions for position 70 seemed to be largely compensatory for the bulky substitution predicted at position 175. Rosetta selected a new position (I18) that is in close proximity to the thymine group of the ketone. The predictions favored a Leu, creating additional space for this substrate orientation. Position 175 was also identified and the models consistently substituted a Trp. The Trp replacement permits tighter packing while the indole ring has the potential to hydrogen bond with the carbonyl of the ketone moiety, depending upon the rotation. The predictions eliminate L66H mutation as V175W takes on this role.

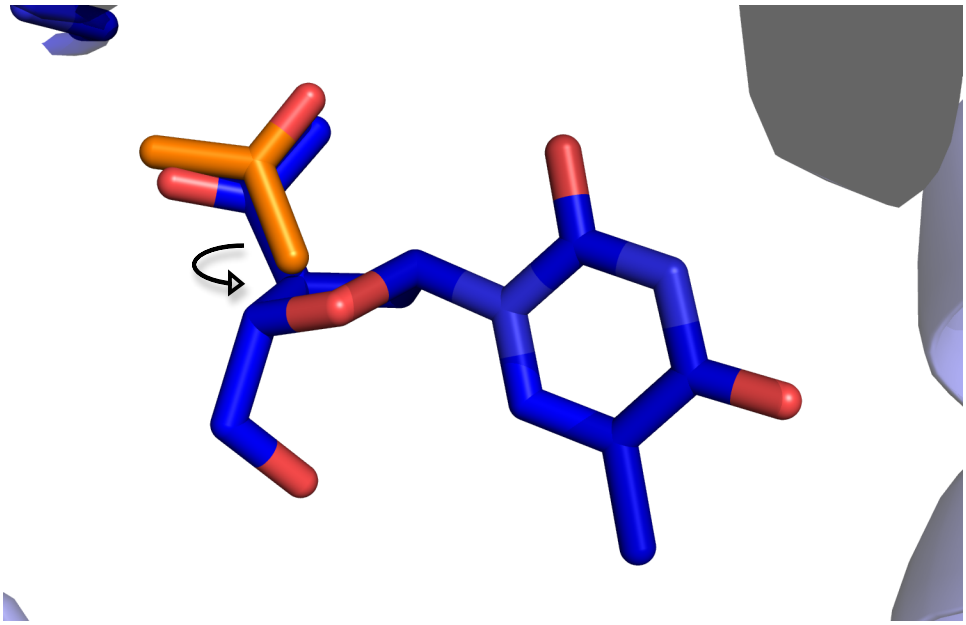


Figure 2.7. The lowest energy conformers (Blue 0° and orange 180°) for assessing the impact of kT rotation around the C3'-CO bond in the third round of computational simulations.

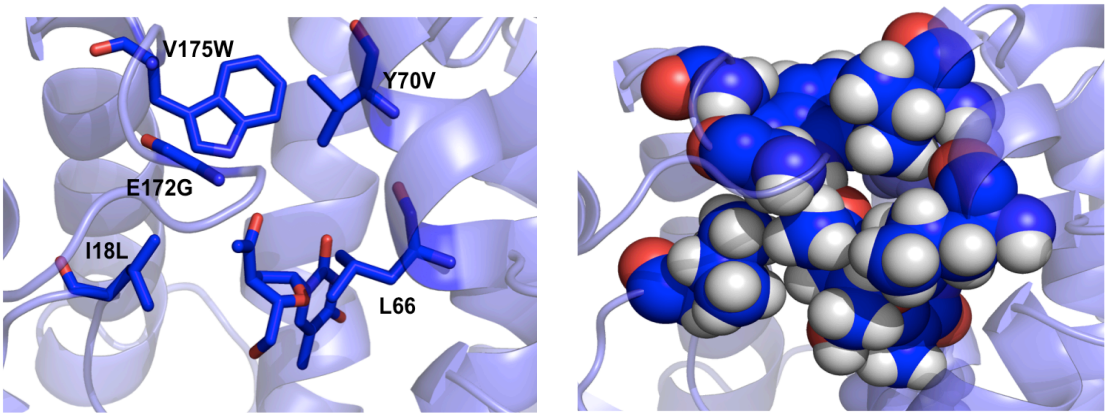


Figure 2.8. Third generation design. RD-4B, with kT bound in the most favorable orientation, 180°; Right- Space filling representation of RD-4B, representing the redesigned active site to selectively bind kT versus Thy.

2.2.5 In vitro characterization of final kT Mutants

The Rosetta guided design constructs were mutated in a stepwise fashion to evaluate the addition of the mutations in the following order: RD-E172G, RD-E172G-V175W, RD-E172G-V175W-Y70V, RD-E172G-V175W-Y70V-I18L (RD-4B). All four mutants were constructed, expressed and purified. The SDS-PAGE of the purification of the final construct RD-4B is shown in Figure 2.9. The enzyme was expressed and purified in high yields with no aggregate formation. The thermal denaturation measurements indicated that E172G had the largest impact on protein stability and each subsequent mutation reduces the T_M by 1-2 ° C (Table 2.5). To add another point of comparison, the third generation designs were tested with dU in addition to Thy. It is assumed that Thy and dU should display similar kinetic properties with the enzymes, as seen with *Dm*dNK. Due to the lack of kU availability, only specific activities were obtained at 200 μM (Table 2.5). The final design showed a 91-fold decline for Thy and >100-fold for dU. The final kU activity was decreased slightly from wild type but was ~ 8-fold higher compared to the second round of designed constructs. For dU, a substrate specificity switch was observed showing a 5-fold higher activity for kU. In contrast, Thy activity is was slightly higher, exhibiting a specificity constant (activity kU/activity Thy) of 0.7. Full kinetic parameters were determined for RD-4B with Thy and dU to determine the differences between binding constants and turnover rates between the natural substrates (Table 2.6). A 31-fold decrease in catalytic efficiency with dU compared to Thy was detected solely as a result of the 35-fold increase in the K_M value. The data suggests that there are more discrepancies than expected working with the uracil base in place of Thy. Upon redesigning the active site for kT, we successfully diminished both

Thy and dU activity, though the latter to a lesser degree. This could be explained by the differences in sugar conformation when the methyl group is present in Thy, and not in dU. Studies have been done to show that substitutions on the nucleobase influences the sugar pucker and certain conformations are more favorable to the kinase.[16] Nevertheless, our third generation designs were successful in predicting mutations that favor the keto group at the 3'position over the natural 3'hydroxyl group. In addition, obtaining the kT analog to test specific activity may result in higher activity as the active site was modeled with this substrate docked and discrepancies with the sugar conformation were unexpected. In summary, our study reveals a combination of positions responsible for altering substrate specificity in regard to the ribose moeity and provides a predictive framework for nucleoside analogs possessing modifications at the 3'-position. Applying other engineering approaches to the current kT designer kinase presents a promising option to evolve an enzyme with higher activity for kT.

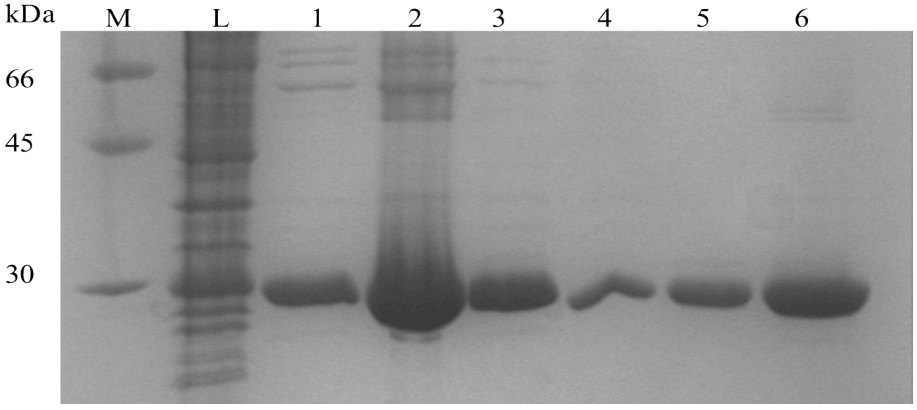


Figure 2.9. SDS-PAGE analysis of RD-4B purification. Lane M: molecular weight marker; lane L: cell lysate; lane 1-4: elution fractions affinity purification; lane 5 and 6: elution fractions gel filtration

Table 2.7. Final Rosetta Designs: Specific Activities at 200 μM Thy and kU and effects on Protein Stability[a]

Variant	kU	dU	Thy	T_M (°C)
DmdNK-E172G	144	1379	2266	55
DmdNK-E172G-V175W	95	108	258	54
DmdNK-E172G-V175W-Y70V	147	132	402	52
DmdNK-E172G-V175W-Y70V-I18L	129	26	175	51

[a]Activities (μM/min per mg enzyme) were determined by spectrophotometric assay at a constant substrate concentration of 200 μM (standard error: ± 10%).

Table 2.8. Kinetic Properties of *Dm*dNK and final Rosetta Design with Thy and 2'-dU[a]

Substrate	Enzyme	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$) × 10^3
Thy	*Dm*dNK	2.46± 0.61	8.2 ± 0.63	3341
	RD-4B	67.62 ± 11.2	0.047 ± 0.002	0.691
2'-dU	*Dm*dNK	9.39 ± 1.07	11.4 ± 0.45	1214
	RD-4B	2412 ± 711	0.052 ± 0.007	0.022

[a]Kinetic parameters were determined by a nonlinear fit to the Michaelis-Menten equation. ND=Non-detectable.

2.3 Conclusion remarks

The importance of this chapter lies in using semi-rational design to create orthogonal nucleoside analog kinases that have a presiding affinity for 3'-modified nucleoside analogs over natural substrates. The transfer of amino acid changes from *Dm*dNK into a human enzyme homolog proved successful. Although the substrate specificity switch in hdCK was not as effective as in *Dm*dNK, the same mutations were responsible. Consequently, transference was possible due to the high sequence and structural homology of the enzymes increased the likelihood of more readily indentifying nucleoside analog kinases for gene therapy applications. In the second study, the active

site of *Dm*dNK was redesigned using the Rosetta software suite to decrease affinity for the natural substrate Thy in favor of nucleoside analog, kT. Three rounds of Rosetta guided predictions revealed a set of mutations that proved to be synergistic for dramatically reducing Thy and maintaining kT activity. Ideally this would provide a starting point for applying additional engineering methods to enhance kT activity even further. Unfortunately, our experiments were limited by the availability of substrate and were not pursued further. The results from this study not only identified a set of powerful residues responsible for altering the activity of 3' modified nucleosides, but also support the success of this method for tuning substrate specificity.

2.4 Materials and methods

2.4.1 Materials

Enzymes used were purchased at New England Biolabs (Ipwich, MA). QIAquick purification kits (Qiagen, Valencia, CA) were used to purify all DNA samples according to their protocols. DNA extractions were completed using QIAprep miniprep purification kit (Qiagen, Valencia, CA). Primers used in the site-directed mutagenesis of dCK were purchased from Integrated DNA Technologies (Coralville, IA). DNA samples were sequenced by GENEWIZ Inc. (South Plainfield, NJ). Genes were subcoloned into the pET-14b (Novagen, Madison, WI) or pMAL-c2X (NEB, Beverly, MA). All plasmids were transformed into *Escherichia coli* strains BL21(DE3) and BL21(DE3)pLysS (Novagen, Madison, WI).

2.4.2 Site-directed mutagenesis dCK mutants

All mutations were introduced via primer overlap extension into templates dCK1 and dCK2. The primers used in the PCR reactions are listed in Table 2.7. Initial cloning was into the pET-14b vector and later all gene inserts were put into the pMAL-c2X for solubility enhancement. For pET-14b, PCR products were digested with *Nde*I and *Spe*I and ligated into pET-14b using T4 DNA ligase. For pMAL-c2X, resulting PCR products were subcloned into pMAL-c2X via *EcoR*I and *BamH*I. All ligation mixtures were transformed into *E. coli* strain DH5α and plated onto LB agar plates containing ampicillin (100 μg/mL) for selection. The plates were incubated overnight at 37 °C. Colonies from the plates were picked and cultured. Isolated plasmids were verified by DNA sequencing.

2.4.3 Site-Directed Mutagenesis DmdNK mutants

The amino acid substitutions were introduced using primer overlap extension into wild type *Dm*dNK. The primers used in the overlap PCR reactions are listed in Table 2.7. The PCR products and pET-14B vector were double digested at restriction sites *Nde*I and *Spe*I and were then ligated together. All ligation mixtures were transformed into *E. coli* strain DH5α and plated onto LB agar plates containing ampicillin (100 μg/mL) for selection. The plates were incubated overnight at 37 °C. Colonies from the plates were picked and cultured. Isolated plasmids were verified by DNA sequencing.

Table 2.9. Primers used to construct hdCK and *Dm*dNK mutants

Mutagenesis primers
E197I-for: GGA AGA AAT GAA ATT CAA GGC ATT CCT C
E197I-rev: GAG GAA TGC CTT GAA TTT CAT TTC TTC C
E197L-for: GGA AGA AAT GAA CTG CAA GGC ATT CCT C
E197L-rev: GAG GAA TGC CTT GCA GTT CAT TTC TTC C
E197A-for: GGA AGA AAT GAA GCG CAA GGC ATT CCT C
E197A-rev: GAG GAA TGC CTT GCG CTT CAT TTC TTC C
E197F-for: GGA AGA AAT GAA TTT CAA GGC ATT CCT C
E197F-rev: GAG GAA TGC CTT GAA ATT CAT TTC TTC C
E197M-for: GGA AGA AAT GAA ATG CAA GGC ATT CCT C
E197M-rev: GAG GAA TGC CTT GC ATTT CAT TTC TTC C
E197S-for: GGA AGA AAT GAA TCC CAA GGC ATT CCT C
E197S-rev: GAG GAA TGC CTT GG GATT CAT TTC TTC C
E197T-for: GGA AGA AAT GAA ACC CAA GGC ATT CCT C
E197T-rev: GAG GAA TGC CTT GG GTTT CAT TTC TTC C
L66H-M69R-for: GTAAATCTGCATGAGCTGCGTTACAAGGAT
L66H-M69R-rev: ATCCTTGTAACGCAGCTCATGCAGATTTAC
Y70D-for: GAGCTGATGGATAAGGATCCCAAG
Y70D-rev: CTTGGGATCCTTATCCATCAGCTC
E172A-V175D-for: CGTTCTGAGGCAAGCTGCGAT
E172A-V175D-rev: ATCGCAGCTTGCCTCAGAACG
Y70M-for: GAGCTGATGATGAAGGATCCCAAG
Y70M-rev: CTTGGGATCCTTCATCATCAGCTC
Y70I-for: GAGCTGATGATTAAGGATCCCAAG
Y70I-rev: CTTGGGATCCTTAATCATCAGCTC
Y70V-for: GAGCTGATGGTGAAGGATCCCAAG
I18L-for: GAGGGCAACCTGGGCAGCGGG
I18L-rev: CCCGCTGCCCAGGTTGCCCTC
Y70V-for: GAGCTGATGGTGAAGGATCCCAAG
Y70V-rev: CTTGGGATCCTTCACCATCAGCTC
E172G-V175W-for: GTTCTGAGGGCAGCTGCTGGCCGCTTA
E172G-V175W-rev: TAAGCGGCCAGCAGCTGCCCTCAGAAC
E172G-for: CGTTCTGAGGGCAGCTGCGTG
E172G-rev: CACGCAGCTGCCCTCAGAACG

2.4.4 Protein overexpression in *E. coli* BL21(DE3)pLysS

Proteins were overexpressed with a hexa-his tag on the N-terminus to allow for protein purification. Cell cultures were first grown at 37 ℃ in 50 mL of LB media containing antibiotics ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) until the

OD_{600} reached approximately 0.5. Protein expression was induced by adding 0.1 mM IPTG and expression took place at 28 ℃ for 4 hours. The cells were centrifuged for 20 minutes 4 ℃ at 4000g. The supernatant was disposed and the cell pellets were flash frozen in liquid nitrogen and stored in the - 80 ℃ until further purification.

2.4.5 His-tagged Protein Purification

Cells were thawed and resuspended in 5 mL of lysis buffer (10 mM imidazole, 0.3 M NaCl, 50 mM Tris-HCL (pH 8)). To the mixture, 7.5 µL lysozyme, 10 µL of protease inhibitor cocktail, and 5 µL of benzonase were added and the sample was stored on ice for ~15 minutes. Cells were then lysed using sonication (7 times for 10 seconds with a 20 pause). The suspension was centrifuged for 30 minutes at 4 ℃ and 10,000g. Separately, Ni-NTA agarose resin (Qiagen) was washed with lysis buffer and the clear supernatant from centrifugation was added. The resin was then incubated on the orbital shaker at 4 ℃ for 60-90 minutes. Next, the resin mix was loaded in a Prep-column (BioRad, Carlsbad, CA) and washed with 2 CV of lysis buffer. In order to remove unspecifically bound protein, the resin was further washed with 4 CV of wash buffer (20 mM imidazole, 0.3 M NaCl, 50 mM Tris-HCL pH 8) as well as 1 CV of wash buffer containing 3 mM ATP and 10 mM $MgCl_2$, and 4 CV wash buffer containing 50 mM imidazole. The protein was then eluted using elution buffer (250 mM imidazole, 0.3 M NaCl, 50 mM Tris-HCL (pH 8)). Samples were taken at each step during the purification and ran on a 12 % SDS-PAGE. The product fractions were combined and placed in an Amicon ultrafiltration column (Millipore, Bedford, MA) with protein storage buffer (2 mM DTT, 0.5 M NaCl, 5 mM $MgCl_2$, 50 mM Tris-HCl (pH 8)) and centrifuged at 4 ℃ and 4000g for 20-30 minutes.

The buffer exchange process was repeated 3 times. Samples were flash frozen in liquid nitrogen and stored at -80℃.

2.4.6 Protein overexpression in *E. coli* K12 TB1

Proteins were overexpressed as fusion proteins with maltose-binding protein at the N-terminus to assist in protein folding and purfication. The pMAL-c2X vector containing the gene of interest was transformed into the expression strain *E. coli* K12 TB1 and plated on LB agar plates containing ampicillin (100 μg/mL) for selection. The plates were incubated overnight at 37 ˚ C. Colonies from the plates were picked and cultured in 2 mL of LB media containing ampicillin (100 μg/mL). The overnight cell cultures were subcultured into 250 mL of glucose rich media containing ampicillin (100 μg/mL) and were grown at 37 ℃ until the OD_{600} reached approximately 0.5. Protein expression was induced using 0.3 mM IPTG at 28 ˚ C for 2-4 hours. The cells were centrifuged at 4 ˚ C for 20 minutes at 4,000g. The supernatant was disposed and the cell pellets were stored at - 20 ˚ C until further purification.

2.4.7 MBP-tagged protein purification

Cells were thawed and resuspended in 5 mL of column buffer (1 mM EDTA, 200 mM NaCl, 20 mM Tris-HCl (pH 7.5)) Per 1 gram cell pellet, 50 μL of protease inhibitor cocktail, 5 μL of benzonase and 1mM DTT were added to the pellet mixture and incubated on ice for ~15 minutes. Cells were lysed using sonication (8-10 times for 10 seconds with a 20 second pause in between) and the slurry was centrifuged at 4 ˚ C and 10,000g for 20-30 minutes. Separately, the amylose resin was washed with column

buffer. The clear supernatant was added to the resin and incubated on a orbital shaker at 4 ˚C for 60-90 minutes. The resin mix was loaded into the prep-column and washed with 4 CV of column buffer. In order to remove unspecifically protein, the resin was washed with 12 CV of column buffer. The fusion protein was then eluted using column buffer containing 10 mM maltose. Samples were taken at each step during the purification and ran on a 12 % SDS-PAGE to assess purity. The product fractions were combined and concentrated using an Amicon ultrafiltration column with protein exchange buffer (500 mM NaCl, 50 mM Tris-HCl (pH 7.5)) and centrifuged at 4 ˚C and 4000g for 20-30 minutes. Samples were flash frozen in liquid nitrogen and stored at -80˚C.

2.4.8 Kinetics

The activity of all kinase mutants was determined using a UV spectrophotometer coupled enzyme assay.[17] A 5X working stock solution was prepared with 5 mM DTT, 12.5 mM $MgCl_2$, 5 mM ATP, 250 mM Tris, 250 mM KCl, water, 1.05 mM PEP, and 0.9 mM NADH. The reaction solution was prepared with the 1X workstock solution, water, 2 units/mL lactate dehydrogenase and pyruvate kinase, and various substrate concentrations to reach a total volume of 1.6 mL. The experiments were performed in triplicate at 30℃ and 37℃ and a wavelength of 340 nm. Different amounts of enzyme were added to the reaction mixture after a baseline was recorded.

2.4.9 CD Spectroscopy

Proteins were buffer exchanged with 50 mM potassium phosphate buffer (pH 7.4), 0.5 M KF. The experiments were performed using a Jasco J-810 spectropolarimeter

and T_M measurements were be taken using the Peltier unit that is connected to the spectropolarimeter for temperature control. The path length of the cuvette used was 0.2 mm with enzyme concentration ranging from 2-4 μM. The CD scans were taken in the wavelength range of 190 nm-260 nm. The thermal denaturation of the proteins was measured from 4-85 °C at 222 nm at a temperature gradient of 1° C/min.

2.5 References

1 Lutz, S., Liu, L. F. & Liu, Y. C. Engineering Kinases to Phosphorylate Nucleoside Analogs for Antiviral and Cancer Therapy. *Chimia* **63**, 737-744, (2009).

2 Furman, P. A. *et al.* Phosphorylation of 3'-Azido-3'-Deoxythymidine and Selective Interaction of the 5'-Triphosphate with Human-Immunodeficiency-Virus Reverse-Transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 8333-8337, (1986).

3 Rossolillo, P., Winter, F., Simon-Loriere, E., Gallois-Montbrun, S. & Negroni, M. Retrovolution: HIV-driven evolution of cellular genes and improvement of anticancer drug activation. *PLoS Genet.* **8**, e1002904, (2012).

4 Munch-Petersen, B., Piskur, J. & Sondergaard, L. Four deoxynucleoside kinase activities from Drosophila melanogaster are contained within a single monomeric enzyme, a new multifunctional deoxynucleoside kinase. *J. Biol. Chem.* **273**, 3926-3931, (1998).

5 Munch-Petersen, B., Knecht, W., Lenz, C., Sondergaard, L. & Piskur, J. Functional expression of a multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster and its C-terminal deletion mutants. *J. Biol. Chem.* **275**, 6673-6679, (2000).

6 Liu, L. F., Li, Y. F., Liotta, D. & Lutz, S. Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. *Nucleic Acids Res.* **37**, 4472-4481, (2009).

7 Liu, L., Murphy, P., Baker, D. & Lutz, S. Computational design of orthogonal nucleoside kinases. *Chem Commun (Camb)* **46**, 8803-8805, (2010).

8 Liu, X. J., Xie, W. & Huang, R. H. Structure-based design, synthesis, and in vitro assay of novel nucleoside analog inhibitors against HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **15**, 3775-3777, (2005).

9 Mikkelsen, N. E. *et al.* Structural basis for feedback inhibition of the deoxyribonucleoside salvage pathway: Studies of the Drosophila deoxyribonucleoside kinase. *Biochemistry* **42**, 5706-5712, (2003).

10 Godsey, M. H., Ort, S., Sabini, E., Konrad, M. & Lavie, A. Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: Illuminating the role of main-chain reorganization. *Biochemistry* **45**, 452-461, (2006).

11 Iyidogan, P. & Lutz, S. Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity. *Biochemistry* **47**, 4711-4720, (2008).

12 Chen, C. Y., Georgiev, I., Anderson, A. C. & Donald, B. R. Computational structure-based redesign of enzyme activity (vol 106, pg 3764, 2009). *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7678-7678, (2009).

13 Damborsky, J. & Brezovsky, J. Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.* **13**, 26-34, (2009).

14 Leaver-Fay, A. *et al.* Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods in Enzymology, Vol 487: Computer Methods, Pt C*, 545-574, (2011).

15 Liu, Y. & Kuhlman, B. RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, W235-W238, (2006).

16 Marquez, V. E. *et al.* Experimental and structural evidence that herpes 1 kinase and cellular DNA polymerase(s) discriminate on the basis of sugar pucker. *J. Am. Chem. Soc.* **126**, 543-549, (2004).

17 Schelling, P., Folkers, G. & Scapozza, L. A spectrophotometric assay for quantitative determination of kcat of herpes simplex virus type 1 thymidine kinase substrates. *Anal. Biochem.* **295**, 82-87, (2001).

Chapter 3: Novel Protease Inhibitors via Computational Redesign of Subtilisin BPN' Propeptide

Adapted with permission from (Daugherty A.B., Muthu, P.J., Lutz S. Novel Protease Inhibitors via Computational Redesign of Subtilisin BPN' Propeptide. Biochemistry 51(41) 8247-8255; 2012.).

3.1 Introduction

Beyond the function of the propeptide as intramolecular chaperone in protease maturation, previous studies have indicated that the bound pro-sequence can also temporarily inhibit proteolytic activity.[1,2] A key component to sustaining inhibitory function is retention of the propeptide tertiary structure. It has been reasoned that these naturally effective but short-lived inhibitors can work for extended time periods if the structural integrity of the propeptide is preserved through enhanced protein stability, which reduces pro-sequence unfolding and susceptibility to proteolysis.[3-5] Such arguments are independently supported by structural analysis of natural protease inhibitors such as proteinase A inhibitor 1 (POIA1) from Pleurotus ostreatus (PDB accession codes: 1V5I & 1ITP) and yeast proteinase B inhibitor 2, which have been found to exhibit high structural similarities with the folded propeptide.[1,6-8]

The hypothesis has also been explored experimentally by a series of protein engineering studies involving subtilisin BPN' from Bacillus amyloliquefaciens and its pro-sequence. Rational protein design and directed evolution strategies led to the introduction of a disulfide bond in the propeptide sequence by Bryan and coworkers and translated into a dramatic stabilization of the propeptide.[9-12] Separately, Kojima et al reported a number of propeptide BPN' mutants with improved packing of its hydrophobic core that exhibited increased stability and inhibitory function.[4,13,14] However, the former design was not tested for inhibitory function while the data for the latter variants proved difficult to reproduce.

In recent years, advances in computational algorithms for in silico protein design have offered a promising new approach for improving protein stability.[15,16] In silico

methods have an advantage over directed evolution strategies in regards to speed and ability for more comprehensively sampling the vastness of sequence space. An extended study of nine globular proteins by Baker and coworkers demonstrated that their Rosetta design algorithm is an effective tool for identifying amino acid substitutions in proteins and peptides that will enhance their overall thermostability without significant structural pertubations.[17,18] In addition, the arbitrary energy scores assigned to individual designs could offer a convenient, quantitative tool for tuning protein stability. In the case of propeptide redesign into subtilisin inhibitors, such tunability of stability could ideally translate directly into tailoring inhibitory function.

In this chapter, the creation of artificial proteinaceous protease inhibitors in silico using the Rosetta design algorithm, as well as their structural evaluation and functional analysis against selected proteases is described. These latter studies assess changes in secondary structure content, thermostability, and inhibitory properties. Furthermore, the design features in the artificial propeptides are probed by site-directed mutagenesis and molecular dynamics simulations to further test and validate the design strategy. Through this study we are able to provide an alternative tool for the redesign of synthetic propeptides that have increased structural integrity and can be evaluated for their potency as protease inhibitors.

3.2 Results and Discussion

3.2.1 Computational propeptide redesign

The Rosetta Molecular Modeling program (v.3.1) was used for redesigning the propeptide domain of subtilisin BPN'.[19] An initial search of SCOP (Structural

Classification of Proteins; http://scop.mrc-lmb.cam.ac.uk/scop/) found four crystal structures for propeptides in the subtilase family; PDB accession codes: 1SPB (subtilisin BPN' from Bacillus amyloliquefaciens)[20], 1SCJ (subtilisin E from Bacillus subtilis)[3], 2Z30 (subtilisin from Thermococcus kodakarensis)[21] and 1T1E (pro-kumamolisin activation domain from Bacillus sp. MN-32)[22], as well as two structures of the native protease inhibitor POIA1 from Pleurotus ostreatus (PDB accession code: 1V5I, 1ITP)[7,8]. Based on the extensive amount of prior work with propeptide from subtilisin BPN,' which sets a benchmark for our own experiments, its crystallographic coordinates complexed with the natural binding partner subtilisin BPN' (PBD accession code: 1SPB) was selected to serve as template for three rounds of fixed backbone design (Fig. 3.1).

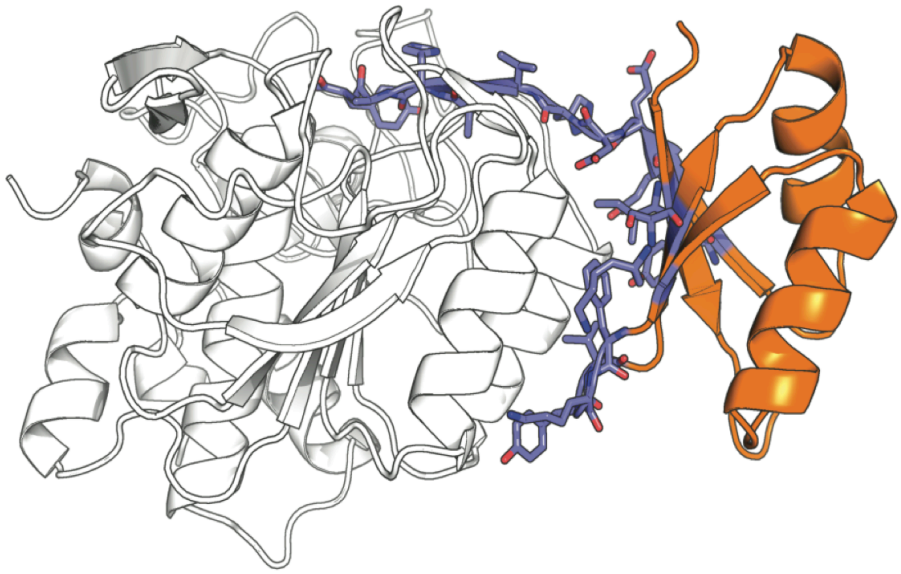


Figure 3.1. Protease propeptide redesign. Structure of wild type subtilisin BPN' with bound propeptide (PDB# 1SPB)9. The propeptide is highlighted in orange with residues at the propeptide-protease interface shown as purple sticks.

For the initial design, 54 non-interface residues of the 71-residue propeptide sequence were allowed to sample the side-chains of all natural amino acids (Table 3.1). To prevent perturbation of the propeptide-protease interface, changes in the remaining 17 residues that directly contact the protease core structure were limited to side-chain rotamers of the native amino acids. Over three rounds of Rosetta design, the propeptide sequence retained 40-50% sequence identity to the parental BPN' propeptide. Beyond the 17 originally conserved residues, an additional 12 to 18 positions showed a strong preference for the native amino acid while 36 to 42 positions experienced amino acid substitutions (Table 3.1).

Table 3.1. Summary of positional amino acid variation over three rounds of Rosetta design

AA Pos.	native AA	Round 1	Round 2	Round 3	ROS1	ROS3
1	E	ALLAA	T	T	T	T
2	K	ALLAA	L	L	L	L
3	K	ALLAA	FKL	FKL	L	F
4	Y	ALLAA	HTV	HT	H	H
5	I	NATRO	I	I	I	I
6	V	ALLAA	HISTV	LIVA	V	V
7	G	ALLAA	G	G	G	G
8	F	ALLAA	FY	FY	F	F
9	K	ALLAA	GR	KR	R	R
10	Q	ALLAA	IKLR	KRQN	K	K
11	T	ALLAA	NT	NT	N	N
12	M	ALLAA	GN	G	G	G
13	S	ALLAA	G	G	G	G
14	T	ALLAA	ART	AT	A	A
15	M	ALLAA	FRLTVY	FRLTVY	T	T
16	S	ALLAA	ADKLRS	ADKLRS	D	A
17	A	ALLAA	ES	AESQN	E	E
18	A	ALLAA	DENQ	ADENQ	N	E
19	K	ALLAA	EILQR	ENQR	E	E
20	K	ALLAA	WRK	W	W	W
21	K	ALLAA	EKR	EKR	E	E
22	D	ALLAA	DELNR	DELNR	N	R
23	V	ALLAA	ADEIT	ALVIT	A	A
24	I	ALLAA	ADEILN	ALVIN	L	L
25	S	ALLAA	KE	KE	K	K
26	E	ALLAA	KEQR	KEQR	Q	Q
27	K	ALLAA	HY	HY	H	H
28	G	ALLAA	G	G	G	G
29	G	ALLAA	G	G	G	G
30	K	ALLAA	D	DELNR	D	D
31	V	ALLAA	ADPRST	ALVIPST	P	P
32	Q	ALLAA	I	IQ	I	I
33	K	NATRO	K	K	K	K

AA Pos.	native AA	Round 1	Round 2	Round 3	ROS1	ROS3
34	Q	**ALLAA**	DHIR	DHRNQ	H	H
35	F	NATRO	F	F	F	F
36	K	NATRO	K	K	K	K
37	Y	NATRO	Y	Y	Y	Y
38	V	NATRO	V	V	V	V
39	D	**ALLAA**	DELP	DELIVAP	L	L
40	A	**ALLAA**	A	A	A	A
41	A	**ALLAA**	AST	A	A	A
42	S	NATRO	S	S	S	S
43	A	**ALLAA**	SA	A	A	A
44	T	**ALLAA**	IL	ILST	I	L
45	L	**ALLAA**	AELMQST	ALIVMST	L	L
46	N	**ALLAA**	S	STNQ	S	S
47	E	**ALLAA**	DQSW	DQSW	S	S
48	K	**ALLAA**	DEKLR	DEKLR	E	E
49	A	**ALLAA**	AST	A	A	A
50	V	**ALLAA**	HIKSVY	HILMSVY	V	V
51	K	**ALLAA**	DLQ	DENQ	Q	Q
52	E	**ALLAA**	EKLRW	EKRQ	K	K
53	L	**ALLAA**	AILST	AILVM	L	L
54	K	**ALLAA**	AM	AMKRQN	A	A
55	K	**ALLAA**	KL	KLRQN	L	L
56	D	**ALLAA**	DRS	D	D	D
57	P	**ALLAA**	ES	SP	S	S
58	S	**ALLAA**	DGHKNRT	DGHKRNT	R	R
59	V	**ALLAA**	HTV	VALI	V	V
60	A	**ALLAA**	K	K	K	K
61	Y	NATRO	Y	Y	Y	Y
62	V	**ALLAA**	HSTV	V	V	V
63	E	NATRO	E	E	E	E
64	E	NATRO	E	E	E	E
65	D	NATRO	D	D	D	D
66	H	NATRO	H	H	H	H
67	V	NATRO	V	V	V	V
68	A	NATRO	A	A	A	A
69	H	NATRO	H	H	H	H
70	A	NATRO	A	A	A	A
71	Y	NATRO	Y	Y	Y	Y

Note: [a] Throughout the table, naive amino acids at individual positions are colored in blue while amino acid substitutions are marked in orange. In round 1, the design algorithm was given only two choices: preservation of the WT residue by NATRO (natural rotamers) - allowing only rotamers of the native amino acid, or unlimited substitution by ALLAA (all amino acids) - allowing for rotamers of all 20 natural amino acids.

Those amino acid changes ranged from relatively conservative substitutions (i.e. K9R and I24L) to complete charge reversion (i.e. K19E, K21E and D22R), as well as functional shifts (i.e. K3F, M12G, S13G and K20W). Nevertheless, no dramatic changes in regard to overall favorable net surface charges, hydrogen bonding networks and hydrophobicity, nor correlation between Rosetta's score function and amino acid substitution patterns could be observed among third-round sequences. The exception is an intriguing design feature that emerged in the extended loop region between β1 and α1, as well as the lower portion of helix-1 (residues 9-20) near the N-terminus of the propeptide variants. This region shows elevated B-factors in the crystal structure of the WT propeptide, suggesting increased conformational flexibility, which could at least in part be responsible for the structural instability of the native domain. In all third-round variants, Rosetta replaced the -MSTxxxxKKK- region (residues 12-21) with -GGAxxxx(E/K)W(E/K)-. A review of the corresponding model structure indicates that the glycines provide the necessary conformational flexibility for a tight turn in the protein structure while the substitutions of T14A and K20W appear to translate into an expansion of the hydrophobic core of the propeptide (Fig. 3.2). The resulting structural integration of the extended loop region could be beneficial to overall protein stability. Moreover, the strategy to stabilize the propeptide through an expansion of the hydrophobic core is clearly distinct from the previously reported covalently-linked disulfide variant, offering in principle the opportunity for tunability of protein stability through modifications of non-covalent interactions in proteins including hydrophobic core and/or surface residues.

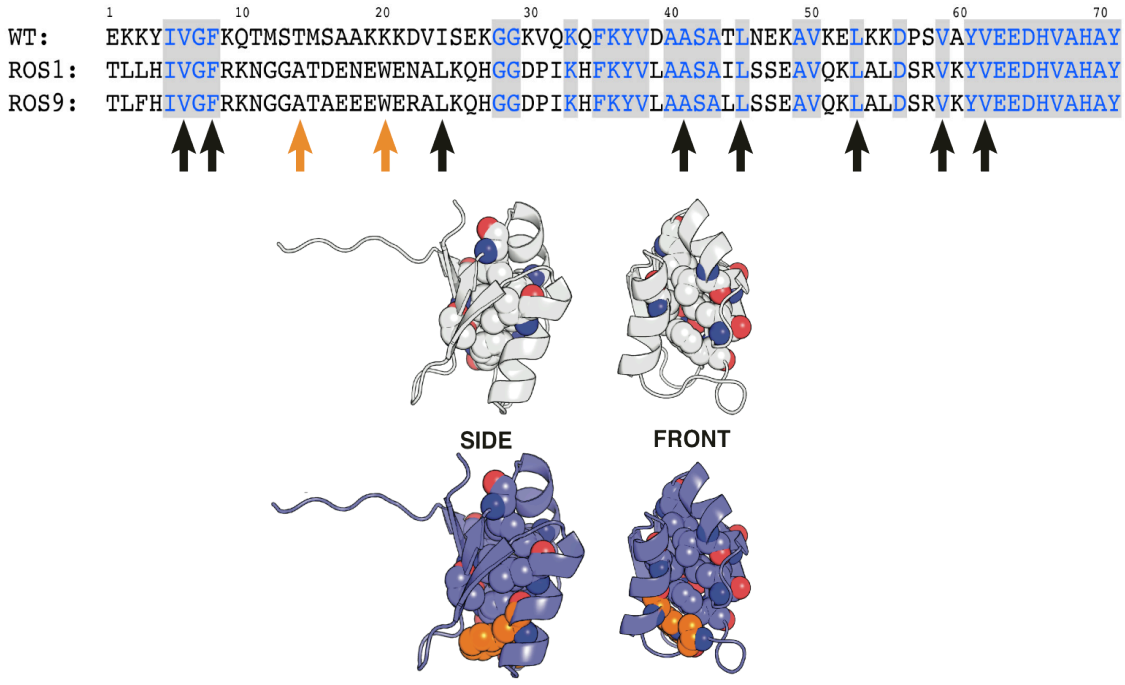


Figure 3.2. Propeptide redesign analysis. Top: Sequence alignment of native propeptide (WT) with Rosetta designs ROS1 and ROS3. Conserved residues are shown in shaded grey while residues of the hydrophobic core structure are marked with arrows. Positions 14 and 20 (orange arrows) represent the residues involved in the predicted core expansion. Bottom: Side and front view of the native propeptide with hydrophobic core residues shown in space-filling mode. Computational model of ROS1 with hydrophobic core expanded by T14A and K20W highlighted in orange.

Separately, we validated our Rosetta propeptide designs by assessing the structural integrity of propeptide variants via molecular dynamics (MD) simulation using a CHARMM22 force field. Rather than performing such simulations on all designs, representative sequences were selected based on a phylogenetic analysis of third round sequences using ClustalW2.[23] The analysis reveals a grouping of the propeptide sequences into eight clusters (Fig. 3.3). One representative (marked ROS 1-8) from each

clade was selected for MD simulation. The corresponding amino acid sequences for ROS1-8 are listed in Fig.3.3. The differences in sequence among the eight clusters are relatively small and limited to various combinations of amino acid changes in ten positions. In addition to wild type BPN' propeptide (WT) and the disulfide-linked variant by Bryan and coworkers10 (REF), a representative from each cluster was picked (ROS1-8, Fig. 3.4) and used for the MD simulations.

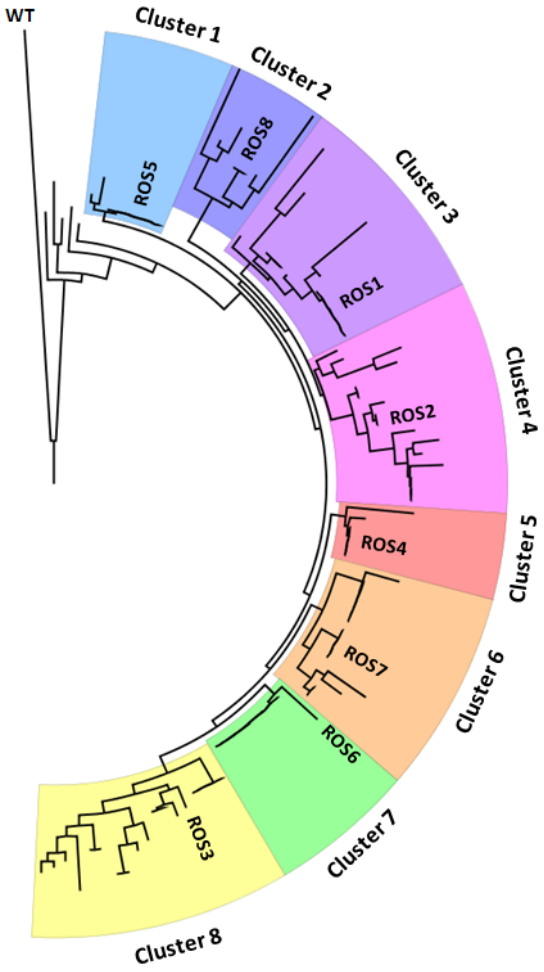


Figure 3.3. Phylogenetic analysis of third round Rosetta designs. (ClustalW2). The analysis grouped the designer propeptides into eight clusters (family 1-8).

Family 1 - ROS6 (-133.004)
TLLHIVGFRKNGGARREEEWEEALKEHGGDPIKHFKYVLAASAILSSEAVQKLALDSKVKYVEEDHVAHAY

Family 2 - ROS3 (-133.124)
TLFHIVGFRKNGGATAEEEWERALKQHGGDPIKHFKYVLAASALLSSEAVQKLALDSRVKYVEEDHVAHAY

Family 3 - ROS7 (-132.766)
TLLHIVGFRKNGGATAEQEWEEALKEHGGDPIKHFKYVLAASAILSSEAYQKLMLDSKVKYVEEDHVAHAY

Family 4 - ROS4 (-133.101)
TLLHIVGFRKNGGATAEQEWEEALKEHGGDPIKHFKYVLAASAILSSEAVQKLALDSKVKYVEEDHVAHAY

Family 5 - ROS2 (-133.322)
TLLHIVGFRKNGGATAEEEWERALKQHGGDPIKHFKYVLAASAILSSEAVQKLALDSRVKYVEEDHVAHAY

Family 6 - ROS1 (-133.335)
TLLHIVGFRKNGGATDENEWENALKQHGGDPIKHFKYVLAASAILSSEAVQKLALDSRVKYVEEDHVAHAY

Family 7 - ROS8 (-132.669)
TLLHIVGFRKNGGARREEEWEEALKEHGGDPIKHFKYVLAASAILSSEAYQKLMLDSKVKYVEEDHVAHAY

Family 8 - ROS5 (-133.042)
TLLHIVGFRKNGGATDENEWEEALKEHGGDPIKHFKYVLAASAILSSEAVQKLALDSKVKYVEEDHVAHAY

Figure 3.4. Amino acid sequences of representative family members (ROS1-8). The bracketed number is the arbitrary overall score assigned to each sequence by Rosetta.

The predicted difference in protein stability of the propeptide variants translates into distinct, time-dependent trajectories for their root mean square deviation (RMSD) after an arbitrary burn-in period of 200 picoseconds to reach the equilibrated structure (Fig. 3.5; Fig. 3.6). As shown in Fig. 3.5, the WT sequence does not appear to be stable, atrophying to the disordered structure. In contrast, the REF sequence seems very stable, as deviant conformers quickly recover to the equilibrated state. The artificial disulfide bond, covalently stitching critical secondary structures, may confer this increased stability. The MD data for the eight Rosetta designs fall in between these two extremes, showing mostly favorable, rebounding trajectories. These results are consistent with the anticipated improvement in structural stability of variants compared to WT propeptide, yet reflect greater conformational variance than REF (Fig. 3.5; Fig. 3.6). The only exception is ROS4 where simulation suggests a similar progression towards a disorderly state as seen for WT. Based on the computational predictions, we chose two Rosetta designs, the top-scoring ROS1 and the "second-tier candidate" ROS3, for in-depth experimental evaluation.

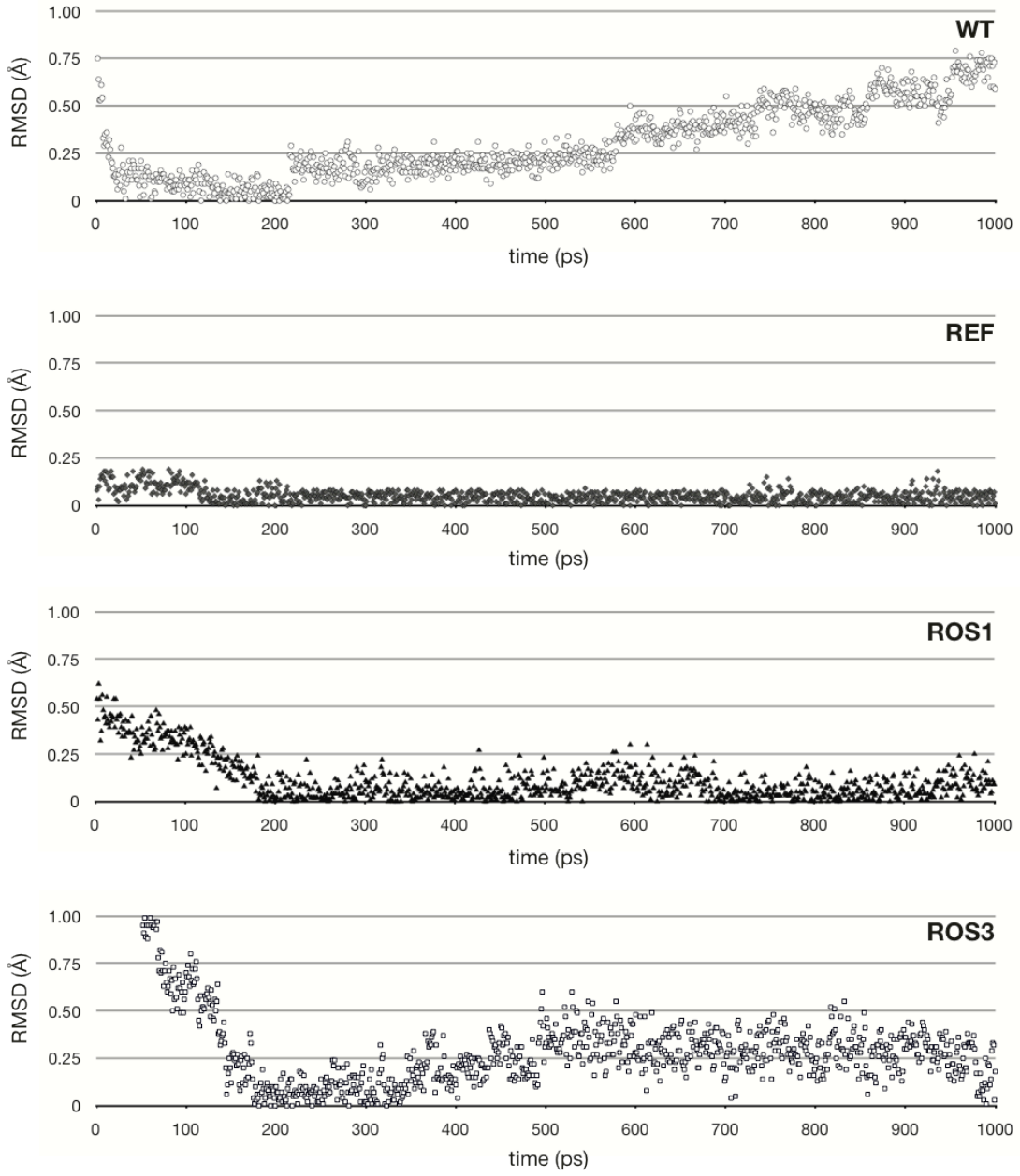


Figure 3.5. RMSD with respect to simulation time for 1 nanosecond MD simulation using wild type BPN' propeptide (WT) and the disulfide-linked variant (REF), as well as the top candidates ROS1 and ROS3.

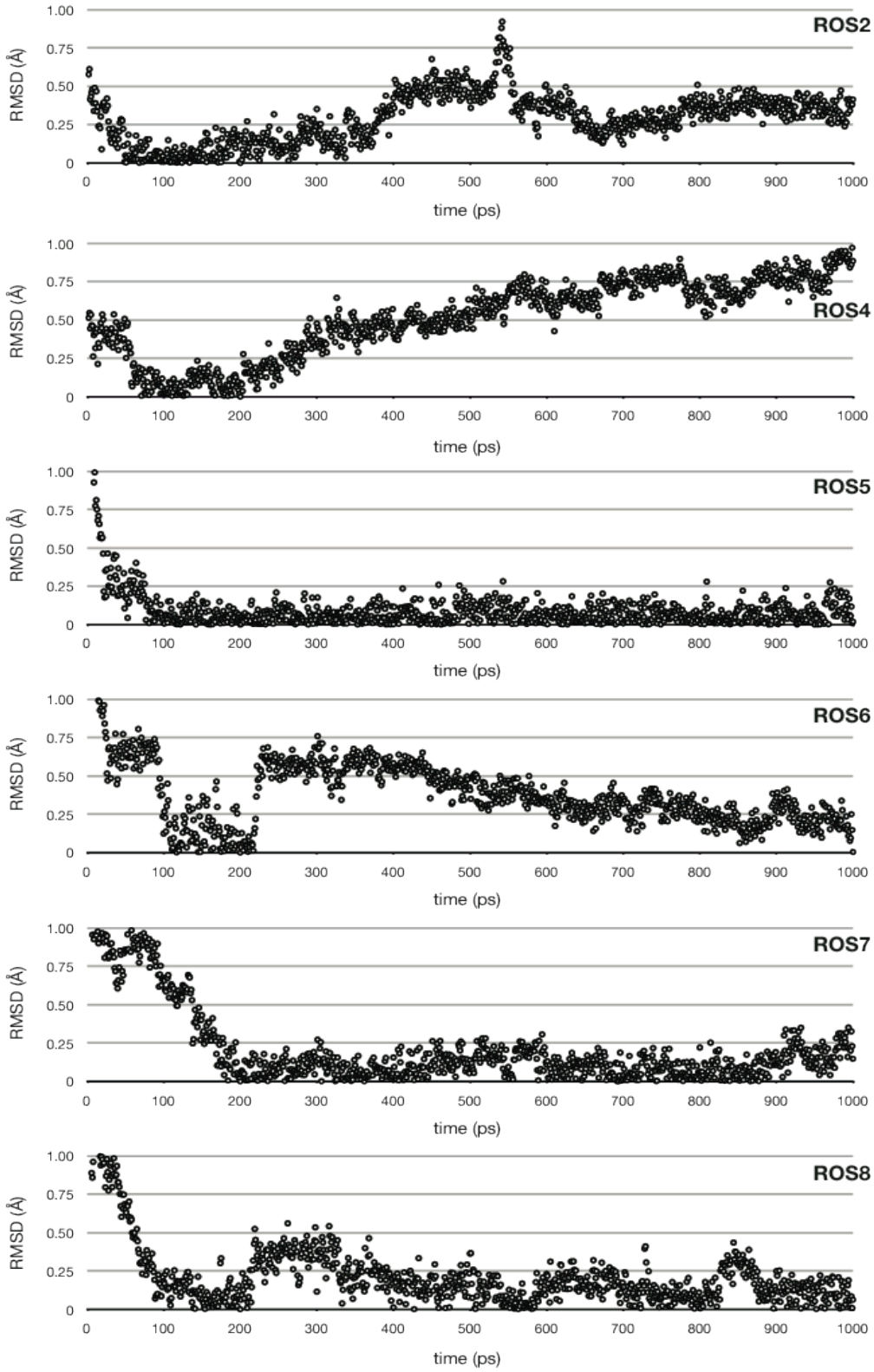


Figure 3.6. RMSD with respect to simulation time for 1 nanosecond MD simulations of representatives from family members in the other clusters.

3.2.2 Expression and purification of designer propeptides

The gene sequences corresponding to the predicted top-scoring propeptide designs ROS1 and ROS3, as well as WT propeptide, were prepared by whole-gene DNA synthesis. All nucleotide sequences were codon-optimized for protein expression in E. coli. Separately, we generated the disulfide-linked propeptide variant, which served as a positive control in our study. The four required mutations (A23C, K27E, V37L, and Q40C) were introduced in the WT sequence, yielding the reference propeptide (REF).

Initially, the constructs were inserted into pTXB1 vector system in order to take advantage of the IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag) technology. The system uses the self-cleavage activity of inteins to cleave the tag from the target protein upon addition of thiols in typically a single step without the use of proteases. A large amount of intein-propeptide soluble protein was detected after expression for all constructs. Purification and cleavage was completed as described below. SDS-PAGE analysis of the purification revealed double band character near the target peptide size of ~ 10 kDa suggesting multiple cleavage sites or truncation of the propeptides during the cleavage process (representative purification gel of REF shown in Fig 3.7). We also observed flow through of the fusion protein of ~38 kDa and the intein tag of ~ 28 kDa. In the absence of any obvious truncation sites the expression, purification, and cleavage conditions were varied to see if the double band could be eliminated. Various strategies to eliminate the double band included a.) changes in DTT concentration, b.) addition of protease inhibitors, c.) variations in buffer salt concentration and composition, d.) lower expression temperatures, and e.) usage of different reducing agents to initiate intein cleavage. Switching the reducing agent from

DTT to cysteine and 2-mercaptoethanol (bME) showed a change in the cleavage pattern. While bME failed to process the fusion protein all together, the addition of cysteine seemed to prevent the promiscuous processing, yielding only a single product (upper band of the doublet). Nevertheless, the cysteine-induced cleavage was less efficient and resulted in the covalent modification of the new C-terminus, an undesirable feature due to its possible interference upon protease binding. Further attempts to optimize the IMPACT system were hence abandoned in favor of a tag-less pET-based expression system.

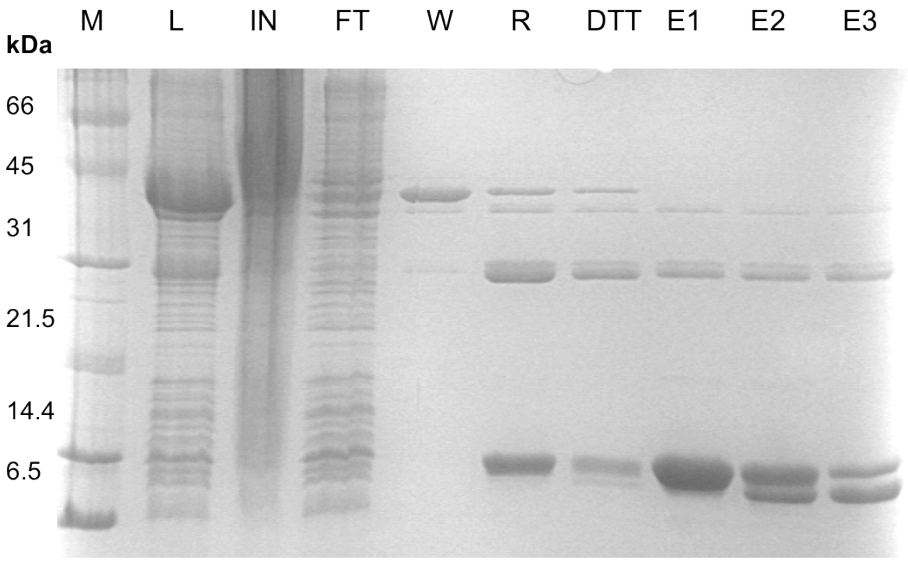


Figure 3.7. SDS-PAGE analysis of REF-Intein fusion. M, BioRad standard marker; L, cell lysate; IN, insoluble fraction; FT, flow-through, W, wash fraction; R, resin sample; DTT, DTT washing solution; E1-E3, elution fractions.

All constructs were inserted into the pET-14b vector as described in the methods section. Soluble expression of all four peptide sequences was readily accomplished in E. coli ArcticExpress (DE3) (for ROS1 and ROS3) or E. coli BL21(DE3) pLysS (for WT

and REF), yielding protein at ~5 mg/Lt of culture for ROS1 and ROS3, as well as 15-20 mg/Lt of culture for WT and REF, respectively. The difference in soluble protein expression levels for the Rosetta designs compared to the WT and REF sequences is significant and interesting in light of the notably improved thermostability of the former (see below). A possible explanation for this phenomenon could be Rosetta's focus on the final, folded propeptide structure without consideration of the actual folding pathway. Nevertheless, soluble protein could be purified by a combination of ion exchange and gel filtration chromatography, yielding the individual propeptides at >95% purity based on SDS-PAGE analysis (Fig. 3.8). The gel filtration experiments also confirmed previous reports of propeptide monomer/dimer equilibrium, indicating a 70:30 ratio in favor of monomeric ROS1, ROS3 and REF.

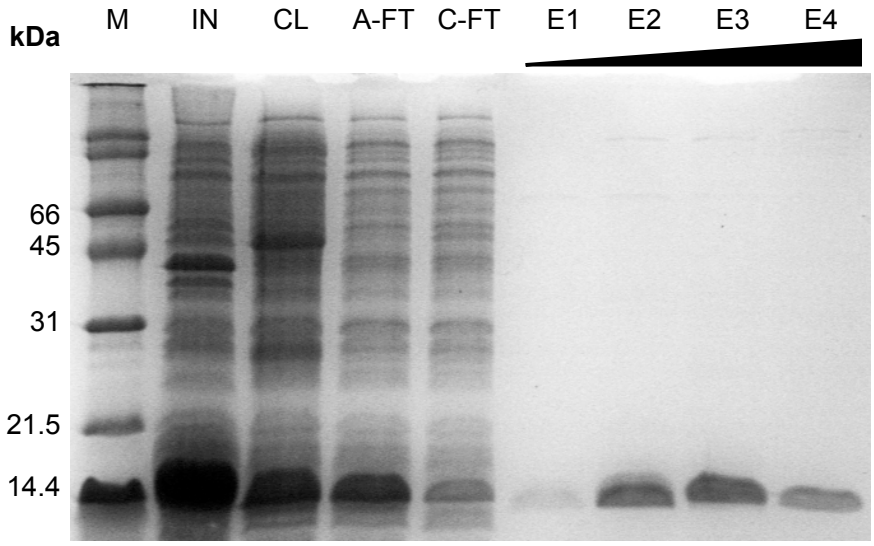


Figure 3.8. SDS-PAGE analysis of REF tag-less protein. M, BioRad standard marker; IN, insoluble fraction; CL, clarified lysate; A-FT, flow-through after anion exchange column; C-FT, flow-through after cation exchange column W, E1-E4, elution fractions from gel filtration column.

Intrigued by the idea of stabilizing the propeptide through the expansion of the hydrophobic core region, we were also curious whether extensive mutagenesis at over 50% of the variable positions, as suggested by Rosetta, was indeed necessary for enhanced protein stability. Alternatively, substantive improvements might be possible through fewer, more specific substitutions in the extended loop region. We thus decided to build a WT variant with only the two site-specific amino acid substitutions in positions 14 and 20 (T14A, K20W). In the absence of obvious steric problems to accommodate these changes, the two substitutions were introduced by site-directed mutagenesis of the WT gene sequence. Expression of the corresponding propeptide in *E. coli* ArcticExpress (DE3) was attempted at various temperatures and induction conditions, yet failed to produce any detectable amounts of soluble propeptide. Our results demonstrate that the simple double-mutant WT propeptide with the core expansion substitutions alone is insufficient to accomplish the anticipated propeptide stabilization and emphasizes the benefits of additional, compensatory substitutions necessary for expression of soluble protein.

3.3.3 Structural characterization of propeptides

The secondary structure content of the WT propeptide, as well as ROS1, ROS3, and REF was initially investigated by far UV circular dichroism (CD) spectroscopy (Fig. 3.9). Spectral scans were recorded from 190 to 260 nm at ambient temperature. In line with previous findings, the WT protein spectrum was consistent with random coil and is assumed to exist in a largely disordered state. In contrast, the spectra for ROS1, ROS3 and REF were consistent with mixed α-helix/β-sheet structures as predicted based on the

secondary structure elements observed in the BPN' propeptide crystal structure as part of the pro-protease structure (PDB accession code: 1SPB)[9].

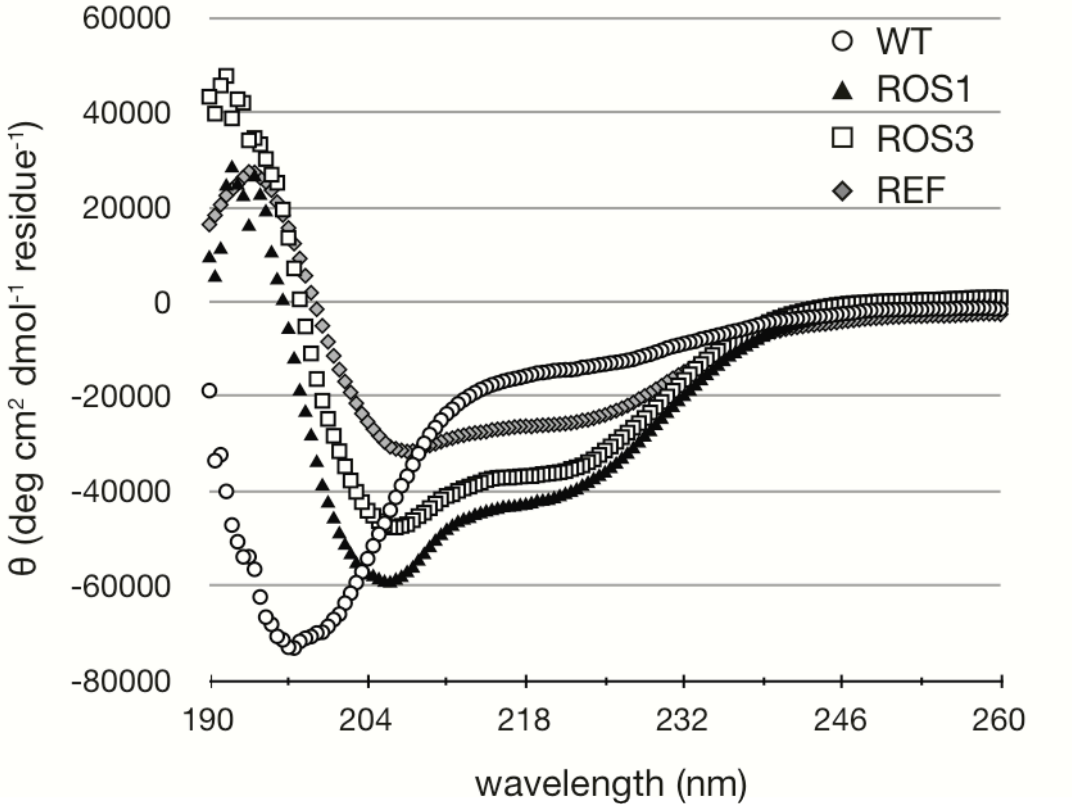


Figure 3.9. Far-UV circular dichroism spectra (190-260 nm) of wild type BPN' (WT), Rosetta design ROS1 and ROS3, and the disulfide-linked variant (REF). While the WT sample resembles a random coil, the spectra of ROS1, ROS3 and REF show clear signatures of mixed α-helices/β-sheet secondary structures. Spectra were recorded on protein samples (25-40 μM) in potassium phosphate buffer (pH 7-8) at ambient temperatures.

Next, the stability of all four propeptides was assessed by thermodenaturation (Fig. 3.10). Changes in secondary structure were monitored at 222 nm from 4˚C to 80˚C in the CD spectrophotometer. As predicted from the far UV CD spectrum, WT propeptide was unfolded at all temperatures with no significant melting transition while the remaining three protein variants showed temperature-induced denaturation profiles. The broad profile for REF indicates a complex unfolding pattern with a denaturation midpoint T_M of ~47˚C. In comparison, the unfolding transition for ROS1 and ROS3 was notably sharper and shifted to higher T_M-values of 55˚C and 53˚C, respectively. The experimentally determined trend in protein stability for the Rosetta designs is consistent with the computational predictions. In addition to the CD thermodenaturation experiments, we also studied tertiary structure changes in ROS1 and ROS3 at elevated temperatures via intrinsic tryptophan fluorescence (Fig. 3.10B). While WT and REF lack any tryptophan residue in their sequences, ROS1 and ROS3 carry a unique tryptophan residue in position 20. Based on our models, the residue is positioned at the edge of the propeptide hydrophobic core and should display an environment-dependent red-shift of its emission spectrum upon protein unfolding and increased solvent exposure. These predictions were confirmed experimentally for both propeptide variants (Fig. 3.10B). At ambient temperature, ROS1 and ROS3 had tryptophan emission maxima of 352 and 349 nm, respectively. Upon gradual increase in temperature, their emission maxima shifted exponentially towards higher wavelengths to reach ~362 nm at 70˚C. For comparison, the emission maximum of free tryptophan in the same buffer solution was near 360 nm at ambient temperature and showed only a small temperature-induced red-shift to 362 nm at 70˚C. If we assume that both ROS1 and ROS3 are largely unfolded at 70˚C, our

fluorescence data give estimated T_M values between 55˚C and 60˚C for the two proteins, consistent with the CD data for loss of secondary structure at elevated temperature. Overall, the CD and fluorescence data support the conclusion that the redesigned propeptide sequences ROS1 and ROS3 are stable and structurally well-defined proteins. c

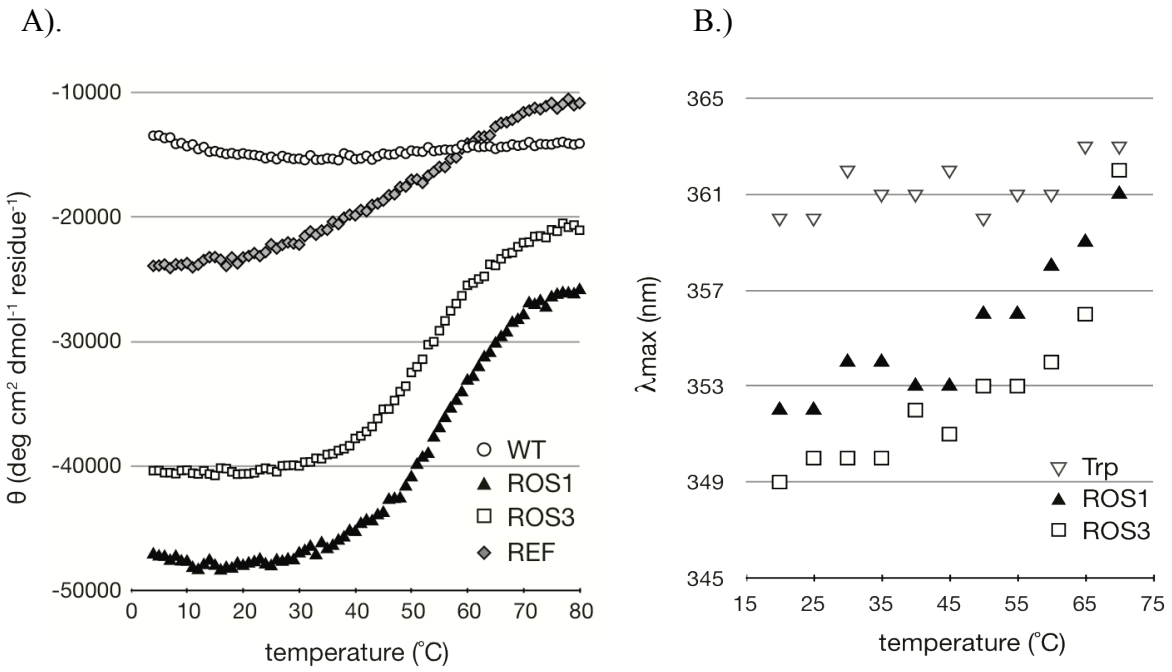


Figure 3.10. Irreversible thermal denaturation curves of wild type BPN' (WT), Rosetta design ROS1 and ROS3, and the disulfide-linked variant (REF). A.) Thermal unfolding detected at 222 nm by CD spectroscopy from 4˚C to 80˚C. While WT is unfolded at all temperatures, data for ROS1, ROS3 and REF were analyzed by curve fitting to a two-state model using the Boltzman equation in Origin7 (OriginLab, Northampton, MA). Spectra were recorded on protein samples (25-40 μM) in potassium phosphate buffer (pH 7-8). B.) Thermal unfolding detected by intrinsic tryptophan fluorescence via emission maximum shift of the unique W20 residue in ROS1 and ROS3. Spectra were recorded on protein samples (70-140 μM) in potassium phosphate buffer (pH 8) in 5˚C increments at the indicated temperatures. Free Trp (50 μM) in the same buffer served as control.

Lastly, preliminary one-dimensional ^{1}H-NMR spectroscopy experiments were conducted to further confirm the structural integrity of the redesigned propeptides (Fig 3.11). This technique has been used in the past to assess the rigidity of redesigned proteins by looking at differences in signal dispersion, line sharpness, and peaks indicative of β-sheet character.[18] Overall these properties can be used to characterize folded, conformationally well-defined proteins. In comparison with REF and ROS1, WT protein shows less signal dispersion, as highlighted in the displayed regions of the NMR spectrum. In addition, the small peaks in the 5 to 5.5 ppm range of REF are typically associated with residues in β-sheets, suggesting the presence of the predicted secondary structure elements. Signals in the same range were not detectable in ROS1 due to high noise levels, presumably due to low concentration of sample. Nevertheless, the observed peaks upfield of 0.6 ppm in REF and ROS1 are consistent with a defined tertiary structure, originating from methylene groups of amino acid side chains in unique chemical environments of proteins. Although these preliminary NMR results suggest a defined secondary structure for ROS1, higher resolution NMR studies need to be completed to make a more conclusive argument with this technique.

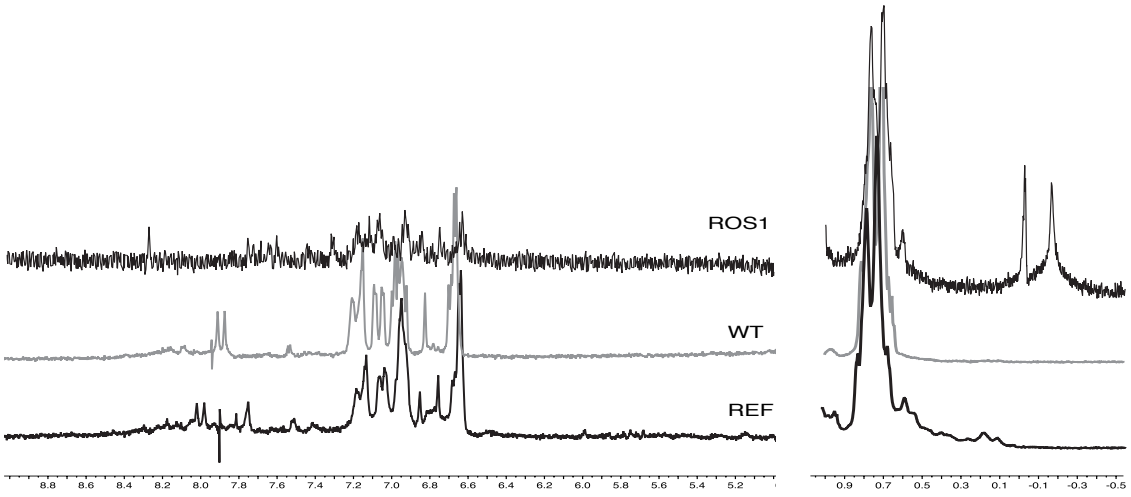


Figure 3.11. 1H-NMR spectroscopy of wild type BPN' (WT), Rosetta design ROS1 and the disulfide-linked variant (REF).

3.3.4 Functional characterization of propeptides

In parallel to the structural characterization of our designer propeptides, we evaluated the inhibitory activity of ROS1 and ROS3, as well as WT and REF against three serine proteases: subtilisin Carlsberg, Savinase and BLAP. Crystal structures among protease complexes have less than 0.4 Å interface rms and sequence similarity in this region maintain complementary patches of buried hydrophobic and charged interactions, suggesting similar driving forces among subtilisin family members. Subtilisin BPN' was not included in our study, even though it served as template for our redesign, as the protease is no longer commercially available. The performance of the propeptide variants was tested in a standard spectrophotometric AAPF activity assay, monitoring proteolytic activity of individual hydrolases in the presence of different molar ratios of propeptide inhibitor.[24]

As shown in Fig. 3.12, all three proteases were inhibited by WT, ROS1, ROS3 and REF albeit to various degrees. The one exception was WT propeptide, which did not affect Savinase activity. Against subtilisin Carlsberg, the WT propeptide proved an effective inhibitor with less than 15% proteolytic activity at 20:1 inhibitor:protease ratio. The result was only matched by REF at the highest ratio (50:1) tested in our assay. In contrast, ROS1 and ROS3 show only slight inhibitor activity (60-80% residual proteolytic activity at 50:1). For BLAP, the pattern of inhibitor performance changes significantly. The presence of REF lowers protease activity to <20% at a 10:1 ratio while ROS1 reduces AAPF hydrolysis to ~50% at 10:1 and ~40% at 50:1 inhibitor:protease. ROS3 and WT are less effective, only reaching 30-40% inhibition at a 50:1 ratio. In the case of Savinase, the two Rosetta designs and REF perform similarly, suppressing protease activity to ~70% at the 50:1 ratio. Despite their moderate inhibitory performance, the results for the three propeptide variants are relevant in light of no detectable inhibitory activity with WT.

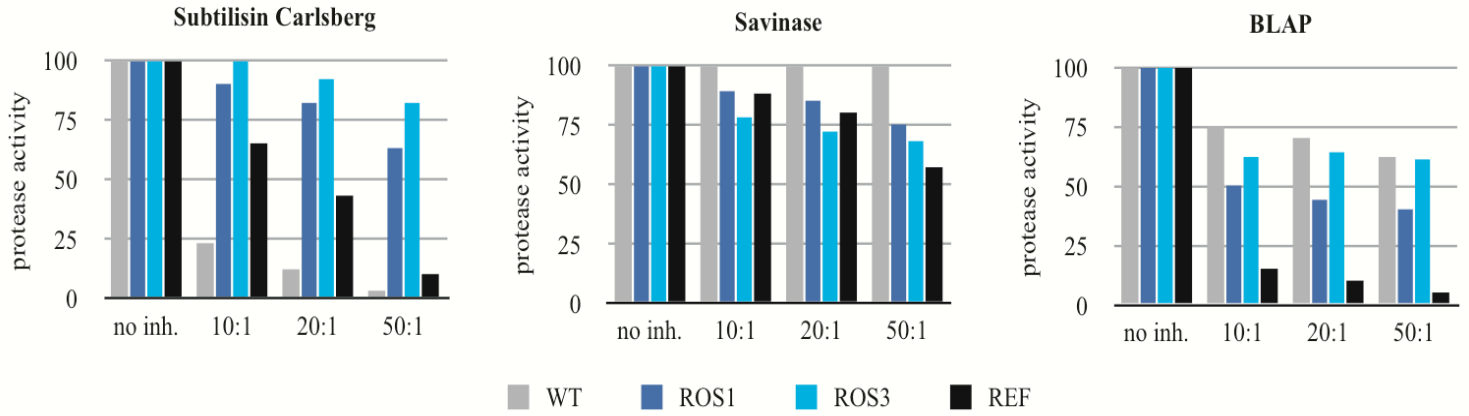


Figure 3.12. Comparison of inhibitory activity for ROS1 and ROS3 with WT and REF against three subtilisin proteases(Carlsberg, Savinase, and BLAP). Standard error is less that 20% for all inhibitory activities. Residual protease activity was determined by the following equation: Activity = (activity in presence of propeptide / activity in absence of propeptide) * 100%.

Overall, the functional study demonstrates that all four propeptide variants can inhibit selected proteases. Nevertheless, there is no clear correlation between improved thermostability of the propeptides and their effectiveness as inhibitors in the present dataset. Beyond measuring the inhibitory function of the propeptide variants immediately after mixing with protease, we also explored the time-dependent change in inhibitor activity. Preincubation of individual propeptide variants with proteases for 20 minutes prior to substrate addition might reveal a stronger correlation between propeptide stability and inhibitor function as increased structural integrity of the propeptide can slow its degradation by the protease, in turn prolonging inhibitor activity. Subsequent AAPF hydrolysis measurements of the preincubated protein mixtures did detect a reduction in inhibitor activity among all tested propeptide, yet changes were generally within the margin of error (data not shown). Additional experiments exploring longer preincubation times for each protease:inhibitor pair might be necessary for detecting significant differences in inhibitor function in correlation to propeptide stability.

3.3 Conclusions

This study has demonstrated the successful application of computational protein design in combination with MD simulations for generating thermostable variants of the subtilisin BPN' propeptide domain. The experimental evaluation of the structural properties of ROS1 and ROS3 are consistent with relative rankings based on the arbitrary energy scores in RosettaDesign. Although the dataset will benefit from the experimental characterization of additional propeptide designs, our current results are consistent with the idea that in silico predictions correlate with propeptide stability. Despite substantial

improvements in the structural integrity as reflected by the increased thermostability, the protease inhibition studies nevertheless highlight the current discrepancies between structure and function. More specifically, the original hypothesis in the literature that stabilization of the propeptide will translate into more potent protease inhibitors is clearly not supported by our results. Even for WT and REF which share high sequence and structure identity, the inhibitory activity for different proteases varies significantly with no recognizable correlation to the elevated stability of REF. Comparison of the crystal structures for a disulfide-linked propeptide domain closely related to REF (PDB accession code: 3CNQ)[25] and the WT structure (PDB accession code: 1SPB)[20] shows an RMSD of 0.94Å over the entire length of the 71-residue propeptide. Yet despite these striking structural similarities, the inhibitory effects of these two propeptides on the proteases in our experiments are notably different. We attribute these uncertainties to discrepancies related to the origin of the propeptide model (subtilisin BPN') and the tested proteases, as well as from small differences in structure and dynamics not accurately captured in either the crystallographic information or the computational model. Improved future designs might be possible with the help of advanced Rosetta models considering flexible backbone and incorporation of interactions at the protease:propeptide interface. Separately, high-resolution NMR studies will be insightful for validating the accuracy of the Rosetta designs and investigating possible conformational differences between the native propeptide domain and the redesigned structures. Preliminary NMR experiments with the two designer propeptides do indicate defined secondary structure similar to REF. Finally, we postulate that the C-terminal portion of the propeptide, which was left unchanged for all variants in our study, might

play an important role in the context of the inhibitory function of these proteins. The region's compatibility with and affinity for the substrate recognition cleft flanking the active site of individual proteases, as well as conformational differences between these propeptides and native protease inhibitors such as POIA1, will be very helpful to consider in future studies.

3.4 Material and Methods

3.4.1 Materials

Whole-gene synthesis was performed by DNA2.0 (Menlo Park, CA) while individual oligodeoxynucleotides were ordered from Integrated DNA Technologies (Coralville, IA). Pfu Turbo DNA polymerase (Stratagene, La Jolla, CA) was used for DNA amplification. Restriction enzymes were purchased from New England Biolabs (Ipswich, MA) while proteases and reagents were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated.

3.4.2 Computational simulations

The Rosetta Molecular Modeling program (v.3.1) was used for redesign of the propeptide domain.[31] Modeling was based on the crystallographic coordinates of BPN' propeptide, complexed with its natural binding partner subtilisin BPN' (PBD accession code: 1SPB)[20]. We performed three rounds of fixed backbone design, idealizing amino acid sequences to the crystallized backbone torsions.[26] To account for backbone, rigidity, amino acid side-chain conformations were modeled using an expanded Dunbrack rotamer

library for increased flexibility during side-chain packing.[27] The first round consisted of 100 independent trajectories, followed by sequence profile analysis using ConSurf.[28] Amino acid substitutions at each position that were observed more than once (>1%) were chosen as sample set for the second round of optimization (Table 3.1). Eleven of the initially variable positions that showed very strong preference for either their native residue or one specific amino acid substitution were fixed in round 2. For the remaining 42 positions, two to seven amino acid changes were allowed. In addition, an expanded rotamer library was used to include chi-angles plus/minus one standard deviation away from the most commonly observed chi angles. As in the initial round, 100 independent trajectories were performed in the second round. Overall, the results from round 2 show few significant sequence changes compared to round 1, suggesting that the expanded conformational space for side chains and the more focused library offered no clear additional benefit. Consequently, the sample set of allowed amino acid substitutions was revised for the third and final round of design. The structural context of individual residues, as well as the predicted variance at each amino acid position, were evaluated and the sample set manually refined and expanded to include homologs of high-scoring residues and native amino acids (Table 3.1). Following 100 independent trajectories in round 3, sequences were analyzed based on the Rosetta energy score function.

3.4.3 Molecular Dynamics

The molecular dynamics simulations were performed by Pravin Muthu in the Lutz lab. The predicted atom coordinates from Rosetta fixed backbone designs were scored using the CHARMM22 all atom-force field.[29,30] Residues 66-71, which constitute the tail

region to the protease were omitted in our model. The structures were minimized using 1,000 steps of steepest decent and conjugant gradient. The system was heated from 50 K to 300 K over 30 ps, and then the molecular dynamics trajectory was recorded. RMSD calculations used an arbitrary reference at 200 ps to accounting for system equilibration. All calculations used the CHARMM22 force field, an implicit solvation model (SASA) and were performed using the charmm package. Simulations were run in replicate to assess reproducibility.

3.5.6 Gene synthesis

The genes encoding for the wild type BPN' propeptide sequence (WT), as well as the two selected designer sequences ROS1 and ROS3 were prepared by whole-gene synthesis (DNA2.0) and shipped in pJ201 DNA vector. The pJ201 vector containing the propeptide genes and the pTXB1 vector were double digested, separately, at restriction sites *Nde*I and *Spe*I and all three genes were ligated into the pTXB1 vector using T4 ligase. In order to switch to the pET-14b vector via same restriction sites were used. Primers had to be used to remove the end of the intein sequence and introduce a stop codon at the 3' end. Following amplification the inserts and vector were double digested and ligated into pET-14b with exclusion of the His-tag. Following transformation into E. coli DH5α, plasmid DNA from selected colonies was extracted and the sequences verified by DNA sequencing.

3.4.7 Site-directed mutagenesis

Site-specific amino acid substitutions in the propeptide sequences were introduced by primer overlap extension mutagenesis. The synthetic WT-BPN' gene served as template for introducing the four mutations (A23C, K27E, V37L, Q40C) in the disulfide-linked BPN' propeptide variant (REF)[10] and the hydrophobic core expansion variant (T14A, K20W). PCR products were cloned into pET-14b as described above and transformed into E. coli DH5α. Gene sequences were confirmed by DNA sequencing.

3.4.8 Protein expression and purification in the IMPACT system

Proteins were overexpressed as a fusion protein containing an intein tag on the C terminus to allow for protein purification. Cell cultures were first grown at 37 °C 250 mL of 2YT media containing antibiotic ampicillin (100 μg/mL)) until the OD_{600} reached approximately 0.8-1.0. Protein expression was induced by adding 0.3 mM IPTG at 20-37 °C for 4-8 hours depending on induction temperature. The cells were centrifuged for 20 minutes 4 °C at 4000g. The supernatant was disposed and the cell pellets were stored at -80 °C until purification. For purification, cell pellets were thawed and resuspended in column buffer (20 mM Tris-HCL 0.3 M NaCl (pH 8.5)) and volume was adjusted according to pellet weight. To the mixture 50 μL of protease inhibitor cocktail per gram of pellet (Sigma), and 5 μL of benzonase per gram of pellet (Novagen) were added and stored on ice for ~15 minutes. Cells were lyzed using sonication (8 times for 10 seconds with a 20 second pause in between). Cell mixture was centrifuged for 30 minutes 4℃ at 10000g. The Chitin column (Qiagen) was washed with column buffer and the supernatant from centrifugation was added to the chitin column at 0.5-1.0 mL/min. The column was

washed with at least 20 bed volumes of column buffer and then quickly washed with 3 column volumes of cleavage buffer (Column buffer containing either 50 mM DTT or Cysteine). The flow of the column was stopped for on-column cleavage to occur at 4 ℃ for 16-40 hours. The target peptides were eluted in the first four fractions by adding 1/3 column volume of column buffer. Elution fractions were combined and placed in an amicon ultrafiltration column to be concentrated before gel filtration. Size exclusion chromatography was performed to try and eliminate contaminates after the cleavage reaction using a Superdex 200, 10/300 GL column, equilibrated with 50 mM Tris-HCl pH 8, 300 mM NaCl buffer at a flow rate of 0.5 ml/min. Elution of propeptide was monitored by UV-detection at 280 nm and product fractions were combined and analyzed by SDS-PAGE. Propeptide samples were flash frozen in liquid nitrogen and stored at -80˚C.

3.4.9 Tag-less protein expression

The propeptide variants were expressed in E. coli BL21(DE3) pLysS or E. coli ArcticExpress (DE3) (Stratagene, La Jolla, CA), respectively. Briefly, individual colonies were cultured in 2 mL LB medium containing antibiotics gentamycin (20 μg/ml) and ampicillin (100 μg/ml) overnight at 37°C. An aliquot of overnight culture was used to inoculate 2YT medium without antibiotics and cultures were grown at 30°C for at least 3 h or until OD(600) reached 0.5 to 0.7. The cultures were cooled to 13°C and overexpression was induced by addition of IPTG to a final concentration of 1 mM. After overexpression for 24 h, cultures were centrifuged for 20 min at 4°C and 4000 g. The supernatant was disposed and cell pellets were stored at - 80°C.

3.4.10 Tag-less protein purification

The purification of the propeptides varies due to the difference in their isoelectric points (pI_{WT} 9.85; pI_{ROS1} 6.62; pI_{ROS3} 7.9; pI_{REF} 9.13). In a typical purification, the cell pellet from 500 mL culture was resuspended in 12 mL of buffer A (WT: 20 mM Tris-HCl (pH 8.5), 50 mM NaCl; ROS1, ROS3 and REF: 50 mM sodium phosphate (pH 6), 50 mM NaCl). To the mixture, 100 µL of protease inhibitor cocktail (Sigma) and 10 µL of benzonase (Novagen) were added and stored on ice for 15 min. Cells were lyzed using sonication (8x with 10 sec pulses and 20 sec pauses). After centrifugation for 30 min at 4 ℃ and 10,000 g, the clear lysate was further purified via two-step ion-exchange chromatography. Initially, lysate was filtered through a HiTrap DEAE FF column (GE Healthcare). Next, the flow-through was collected and loaded onto a HiTrap SP FF column, preequilibrated with buffer B (WT: 50 mM HEPES (pH 8), 50 mM NaCl; ROS1, ROS3 and REF: 50 mM sodium phosphate (pH 6), 50 mM NaCl). The column was washed with 7 column volumes (CV) of the corresponding buffer B. A linear gradient from 0 to 50% buffer C (buffer B with 1 M NaCl) over 25 CV was applied. Product fractions were combined and concentrated to ~1 mL using a Millipore filter unit (MWCO: 3 kDa). In a final polishing step, propeptides were purified by size exclusion chromatography (Superdex 200, 10/300 GL column, equilibrated with buffer D (50 mM Tris-HCl pH 8, 300 mM NaCl); flow rate: 0.5 ml/min). Elution of propeptide was monitored by UV-detection at 280 nm and product fractions were combined and analyzed by SDS-PAGE. Propeptide samples were flash frozen in liquid nitrogen and stored at -80˚C.

3.4.11 Circular dichroism spectroscopy

The experiments were performed using a Jasco J-810 spectropolarimeter equipped with a Peltier unit for temperature control. The path length of the cuvette used was 1 mm with protein concentration ranging from 25 - 40 μM as determined by the absorbance at 280 nm. All samples were prepared in 50 mM potassium phosphate buffer (pH 7-8 depending on the pI). Spectra were collected in the far UV wavelength range at 25 °C using a scan rate of 20 nm/min, response time of 2 s, and bandwidth of 2 nm. The spectrum reported represents an average of three accumulation scans. For thermal denaturation experiments, the change in ellipticity at 222 nm was monitored from 4 to 80 ˚C with a temperature gradient of 1.0 ˚C/min.

3.4.12 Intrinsic tryptophan fluorescence

All measurements were conducted on a FluoroMax-3 spectrophotometer (Horbia Jobin Yvon, Edison, NJ), equipped with a thermostatted cell holder and a NESLAB RTE7 water bath (Thermo Electron Cooperation, Waltham, MA). Purified protein samples of ROS1 and ROS3 (70-140 μM) were prepared in 50 mM potassium phosphate buffer (pH 8). As a control, tryptophan (50 μM) was used in the same buffer. Samples were excited at 295 ± 0.5 nm and their emission spectra were measured with 1-nm bandwidth resolution at wavelengths between 300 and 400 nm. Sample scans were taken in 5˚C increments from 20˚C to 75˚C with 3-minute sample equilibration time. All spectral data were analyzed in Origin7 to determine the maximum peak intensity and emission wavelength.

3.4.13 Nuclear magnetic resonance spectroscopy

Standard 1H-NMR spectra were recorded on a Varian Inova600 using water presaturation. Protein samples (0.8 - 1 mM) in 50 mM sodium phosphate buffer (pH 7.0) were lyophilized twice and resuspended in D2O prior to spectra acquisition.

3.4.14 Protease inhibition

Protease inhibition was monitored via hydrolysis of AAFP (N-succinyl-L-Ala-L-Ala-L-Pro-L-Phe p-nitroanilide) by selected proteases in the presence of various amounts of propeptide.[37] Assays were conducted with protease concentrations remaining consistent for each data set (Savinase 1.8 nM, BLAP 0.38 nM, Carlsberg 0.56 nM) and 1 to 50-fold molar ratios of propeptide at 25˚C. Briefly, the propeptide was mixed with protease and preincubated for 0 or 20 min, followed by addition to 0.111 mM AAFP in reaction buffer (0.1 M Tris-HCl (pH 8.6), 0.1% Brij35) and a total of 1 mL reaction volume. Upon addition of enzyme mix, substrate conversion was monitored spectrophotometrically via change in absorbance at 410 nm over time. Substrate hydrolysis for each propeptide-protease ratio was measured in triplicates.

3.5 References

1 Kojima, S., Iwahara, A. & Yanai, H. Inhibitor-assisted refolding of protease: A protease inhibitor as an intramolecular chaperone. *FEBS Lett.* **579**, 4430-4436, (2005).

2 Ohta, Y. *et al.* Pro-Peptide as an Intermolecular Chaperone - Renaturation of Denatured Subtilisin-E with a Synthetic Pro-Peptide. *Mol. Microbiol.* **5**, 1507-1510, (1991).

3 Jain, S. C., Shinde, U., Li, Y. Y., Inouye, M. & Berman, H. M. The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 angstrom resolution. *J. Mol. Biol.* **284**, 137-144, (1998).

4 Kojima, S., Minagawa, T. & Miura, K. The propeptide of subtilisin BPN' as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity. *FEBS Lett.* **411**, 128-132, (1997).

5 Kojima, S., Yanai, H. & Miura, K. Accelerated refolding of subtilisin BPN ' by tertiary-structure-forming mutants of its propeptide. *Journal of Biochemistry* **130**, 471-474, (2001).

6 Kojima, S., Deguchi, M. & Miura, K. Involvement of the C-terminal region of yeast proteinase B inhibitor 2 in its inhibitory action. *J. Mol. Biol.* **286**, 775-785, (1999).

7 Lee, W. C., Kikkawa, M., Kojima, S., Miura, K. & Tanokura, M. Crystal structure of serine protease inhibitor POIA1 in complex with subtilisin BPN'. *Data deposited in PDB database only.*, (2005).

8 Sasakawa, H., Yoshinaga, S., Kojima, S. & Tamura, A. Structure of POIA1, a homologous protein to the propeptide of subtilisin: implication for protein foldability and the function as an intramolecular chaperone. *J. Mol. Biol.* **317**, 159-167, (2002).

9 Ruan, B., Hoskins, J. & Bryan, P. N. Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant. *Biochemistry* **38**, 8562-8571, (1999).

10 Ruan, B., Hoskins, J., Wang, L. & Bryan, P. N. Stabilizing the subtilisin BPN ' pro-domain by phage display selection: How restrictive is the amino acid code for maximum protein stability? *Protein Sci.* **7**, 2345-2353, (1998).

11 Ruvinov, S. *et al.* Engineering the independent folding of the subtilisin BPN' prodomain: analysis of two-state folding versus protein stability. *Biochemistry* **36**, 10414-10421, (1997).

12 Wang, L., Ruan, B., Ruvinov, S. & Bryan, P. N. Engineering the independent folding of the subtilisin BPN ' pro-domain: Correlation of pro-domain stability with the rate of subtilisin folding. *Biochemistry* **37**, 3165-3171, (1998).

13 Kojima, S., Hisano, Y. & Miura, K. Alteration of inhibitory properties of Pleurotus ostreatus proteinase A inhibitor 1 by mutation of its C-terminal region. *Biochem. Biophys. Res. Commun.* **281**, 1271-1276, (2001).

14 Kojima, S., Minagawa, T. & Miura, K. Tertiary structure formation in the propeptide of subtilisin BPN ' by successive amino acid replacements and its close relation to function. *J. Mol. Biol.* **277**, 1007-1013, (1998).

15 Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. Computational thermostabilization of an enzyme. *Science* **308**, 857-860, (2005).

16 Shah, P. S. *et al.* Full-sequence computational design and solution structure of a thermostable protein variant. *J. Mol. Biol.* **372**, 1-6, (2007).

17 Dantas, G. *et al.* High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J. Mol. Biol.* **366**, 1209-1221, (2007).

18 Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332**, 449-460, (2003).

19 Leaver-Fay, A. *et al.* Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods in Enzymology, Vol 487: Computer Methods, Pt C*, 545-574, (2011).

20 Gallagher, T., Gilliland, G., Wang, L. & Bryan, P. The Prosegment-Subtilisin Bpn Complex - Crystal-Structure of a Specific Foldase. *Structure* **3**, 907-914, (1995).

21 Tanaka, S., Matsumura, H., Koga, Y., Takano, K. & Kanaya, S. Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: insight into structural changes during maturation. *J. Mol. Biol.* **372**, 1055-1069, (2007).

22 Comellas-Bigler, M. *et al.* 1.2 A crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase. *Structure* **12**, 1313-1323, (2004).

23 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948, (2007).

24 DelMar, E. G., Largman, C., Brodrick, J. W. & Geokas, M. C. A sensitive new substrate for chymotrypsin. *Anal. Biochem.* **99**, 316-320, (1979).

25 Ruan, B., London, V., Fisher, K. E., Gallagher, D. T. & Bryan, P. N. Engineering substrate preference in subtilisin: structural and kinetic analysis of a specificity mutant. *Biochemistry* **47**, 6628-6636, (2008).

26 Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368, (2003).

27 Dunbrack, R. L., Jr. & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661-1681, (1997).

28 Berezin, C. *et al.* ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* **20**, 1322-1324, (2004).

29 Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545-1614, (2009).

30 Brooks, B. R. *et al.* Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **4**, 187-217, (1983).

Chapter 4: Circular Permutation of Old Yellow Enzyme:

Characterization of a Complete Synthetic Library

4.1 Introduction

The reduction of alkenes is a widely used and effective strategy for the asymmetric synthesis of chiral building blocks. While traditional catalysts for these reactions include chiral transition metal complexes [1-3] and organocatalysts[4], enzymes offer an attractive highly selective, evolvable and sustainable alternative.[5-7] More specifically, ene-reductases of the Old Yellow Enzyme (OYE) family [EC 1.3.1.31] are known to catalyze the highly stereoselective trans-hydrogenation of α,β-unsaturated aldehydes, ketones, carboxylates, nitriles and nitroalkenes.[8-14] The redox chemistry of OYEs is facilitated by a non-covalently bound flavin mononucleotide (FMN) cofactor. Following reduction of the alkene substrate, FMN is regenerated via hydride transfer from NADPH (Figure 1.1).

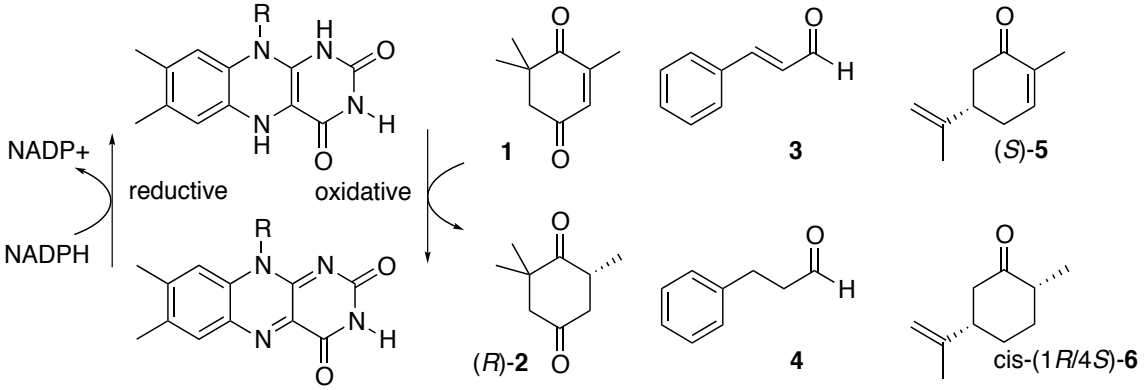


Figure 4.1. Catalytic reaction of OYE with substrates investigated in this chapter.

The popularity of OYEs in biotechnological and pharmaceutical applications has resulted in isolation and characterization of family members from a variety of organisms, providing a rich source of oxidoreductases for the conversion of alkenes.[8,9,13,15-19] At the

same time, these biocatalysts represents attractive targets for protein engineers to further improve and customize their functional performances. A number of structure-guided site-directed and site-saturation mutagenesis studies, as well as directed evolution experiments, have yielded OYE variants with improved catalytic properties including enhanced turnover rates, reversed enantioselectivity and increased stability.[20-25] Unfortunately, reported gains in catalytic activity were generally modest (two to four-fold). However, potential benefits of amino acid substitution in OYE family members are highlighted by the complete reversal of enantioselectivity as a result of a single amino acid change.[23] A review of past OYE engineering studies also identified a major experimental challenge. Following standard heterologous expression of OYEs in *E. coli*, the functional characterization of the exogenous oxidoreductase is complicated by the host's endogenous reductase activity. Therefore, additional purification steps are required prior to functional evaluation of OYE libraries.[22] Consequently, library analysis has typically been limited to either small numbers of variants or to substrates with low background from host reductases.[20,23] Future efforts involving larger, more comprehensive protein libraries for tailoring OYE family members would clearly benefit from a more effective library preparation and analysis protocol.

Our OYE engineering efforts were motivated by a number of biochemical studies and crystallographic structure analyses that suggested substantial conformational changes as part of the enzyme's catalytic cycle. Spectroscopic evidence for such conformational changes in kinetic measurements and structural differences in crystals of apo-form and substrate-bound OYEs, as well as more recent data from an engineering study of *Zymomonas mobilis* NCR enoate reductase all support the hypothesis that structural

rearrangements of loops and domains play an important and possibly rate-limiting role in the catalytic function of these enzymes.[24,26-31] We therefore decided to test the functional role of various flexible regions in OYE1 from *Saccharomyces pastorianus* (formerly *S. carlsbergensis*), the most extensively studied member of the OYE family, by a protein engineering strategy called circular permutation (CP). During CP, a protein's original amino and carboxyl-termini are covalently linked by a peptide linker and new termini are introduced elsewhere in the protein structure through the breakage of a peptide bond.[32] While termini relocation leaves the amino acid composition of the protein unchanged, the sequence reorganization has been shown to affect a protein's local conformational flexibility. For new termini positioned near an enzyme active site, the altered protein dynamics can significantly impact catalytic performance by translating into greater active site accessibility and modifying rate-determining structural changes.[33-35]

In addition, Old Yellow Enzyme provides a suitable scaffold for exploring FMN analogs and their impact on the reactivity.[36] Substitutions at positions 7 and 8 of the flavin are responsible for modulating the redox potential of the cofactor (Figure 4.2).[37] Previous work has demonstrated that upon altering the redox potential of the cofactor, properties such as oxygen sensitivity as well as the catalysis of new chemistries can be established.[38-42]

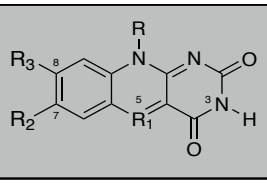
 R = D-ribityl	R_1 =	R_2 =	R_3 =	redox potential ($E_{1/2}$)
7,8-di(trifluoromethyl) riboflavin	N	$-CF_3$	$-CF_3$	+20 mV
8-cyano riboflavin	N	$-CH_3$	-CN	-50 mV
7-chloro riboflavin	N	-Cl	$-CH_3$	-127 mV
8-chloro riboflavin	N	$-CH_3$	-Cl	-152 mV
riboflavin	**N**	**$-CH_3$**	**$-CH_3$**	**-207 mV**
5-deaza-5-carboriboflavin	C	$-CH_3$	$-CH_3$	-311 mV
8-amino riboflavin	N	$-CH_3$	$-NH_2$	-330 mV

Figure 4.2. Examples of riboflavin analogs with altered redox potentials.

The cofactors with more negative redox potentials such as 5-deaza and 8-amino riboflavin are of interest to address oxygen sensitivity of OYEs.[36,43] In aerobic conditions undesirable side reactions take place with oxygen and FMN producing hydrogen peroxide. Furthermore, superoxide radicals can go on to react with the enzyme and substrates producing undesirable side products. To circumvent this issue experiments are typically performed under anaerobic conditions.[22] The requirement for anaerobic conditions can be very undesirable from a biocatalyst standpoint as it limits its use to special environmental conditions and scale. These two analogs are unique in that they are both two-electron mediators disallowing the one-electron reduced semiquinone state that has been associated with formation of reactive oxygen species.[36] On the other end of the spectrum, more positive redox cofactors such as 8-CN-riboflavin has been shown to catalyze the desaturase reaction of several substrates (reverse of reduction).[42] The ability of OYE to catalyze new reactions would broaden it's utility for the production of chiral synthons.

In this chapter, we report on the identification and characterization of catalytically improved OYE1 variants via CP. To address the aforementioned practical limitations for working with combinatorial libraries of OYE, we have opted for a whole-gene synthesis approach that gives access to an idealized gene library with maximal (theoretical) diversity at minimal size. Subsequent synthesis of the corresponding cpOYE variants is accomplished via *in vitro* transcription/translation (IVTT) using the chemically defined PURE system which almost completely eliminates background reductase activity.[44,45] This *ex vivo* protein engineering strategy has proved highly effective for the parallel synthesis of hundreds of OYE library members whose catalytic activity can be assessed without further need for purification. The evaluation of our new protocol for OYE1 library screening on three reference substrates is described. The approach has helped to identify a number of candidates with significantly (>10-fold) enhanced catalytic performance. Evidence in support of the critical role of loop and domains near the active site are found in the systematic analysis of protein variants with multiple substrates and detailed characterization of selected cpOYE variants. The complete OYE1 library in combination with the PURE screening system also enables supplementation of each library member with the cofactor analog of interest in a high through put manner. This combination affords a platform for two-dimensional engineering, simultaneously altering protein scaffold and redox cofactor.

4.2 Results and Discussion

4.2.1 OYE1 library generation

The structurally and functionally well-characterized OYE1 from *S. pastorianus* was chosen to test our hypothesis that CP of an OYE family member could yield variants with improved catalytic performance. Initially, we analyzed representative structures of OYE1 (PDB access code: 1OYA[27] and 1K03[26]) to determine a suitable linker sequence for connecting the native amino and carboxyl termini. Trimming the C-terminus by three residues shortened the termini distance from 13Å to 6Å without impacting the catalytic performance of the wild type enzyme. The remaining (shorter) gap was bridged by a flexible, hydrophilic three-amino acid residue linker (-Gly-Thr-Ser-).

Taking advantage of whole-gene synthesis, we abandoned the previously employed random circular permutation protocol [35,46] to instead prepare a fully synthetic DNA library of circular permuted OYE1 (cpOYE) variants. A tandem-OYE1 gene sequence was prepared by chemical DNA synthesis that served as template for PCR amplification of individual cpOYE variants (Figure 4.3). This strategy dramatically reduces the size of the circular permutation library by eliminating out-of-frame and inversely cloned representatives in the sequence pool. Furthermore, parallel cloning of individual genes into an appropriate DNA vector for protein expression, followed by verification of the correct DNA sequence for each library member creates a chemically defined collection of enzyme variants. Such up-front efforts during library generation are advantageous as it dramatically simplifies library analysis by eliminating the need for oversampling.

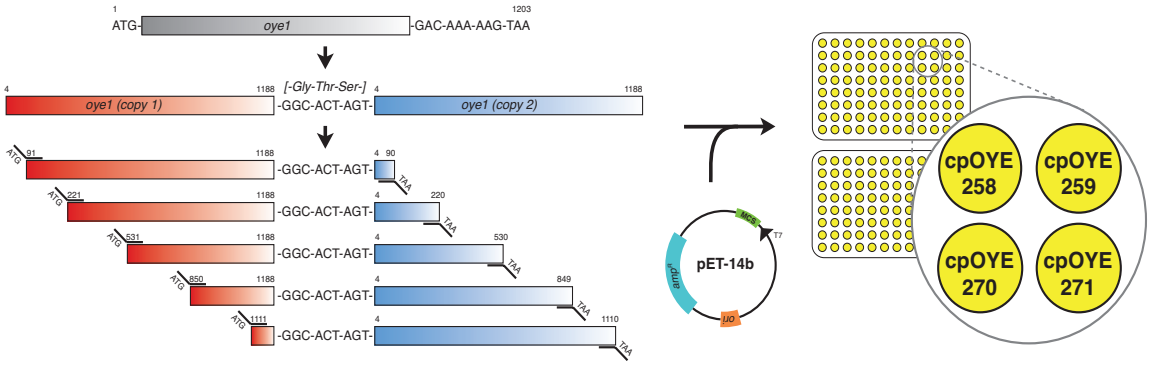


Figure 4.3. Assembly of cpOYE library via whole-gene synthesis. Two copies of truncated oye1 (marked in red and blue) were prepared that lack the three native C-terminal amino acids and the Start/Stop codons. The two genes were assembled in tandem and connected by a 9-nucleotide sequence encoding the -Gly-Thr-Ser- linker. PCR amplification with site-specific primers and the tandem repeat as template yielded individual members of the CP library and reintroduced Start/Stop codons, as well as flanking restriction sites for the subsequent cloning into pET-14b. Correct gene constructs (cpOYE#) were confirmed by DNA sequencing, sorted and stored as purified plasmids in microtiterplates.

For practical reasons, we created a focused cpOYE library by synthesizing only every other possible variant, starting with even-numbered positions at the N-terminus of OYE1. In addition, variants within ± 6 amino acids to the native termini were left out as they are predicted to at best show wild type-like activity. Following the initial functional evaluation of this primary library, odd-numbered variants in regions of interest could quickly be prepared by the same PCR-based method. Over the course of this entire study, we synthesized and screened a total of 228 cpOYE variants, identifying roughly 70 members (~30%) in our circular permutation library with equal or better than wild type

activity for reduction of ketoisophorone (**1**), a widely used reference substrate for the functional evaluation of OYE family members (see below).

4.2.2 Development of a high throughput IVTT screen

Building on the idea of a synthetic protein engineering approach, we employed the PURE system for functional analysis of the cpOYE library. PURE distinguishes itself from other IVTT systems in that all of its components are isolated and purified individually, followed by reconstitution of a chemically fully defined functional protein synthesis machinery (Figure 4.4).[44,45]

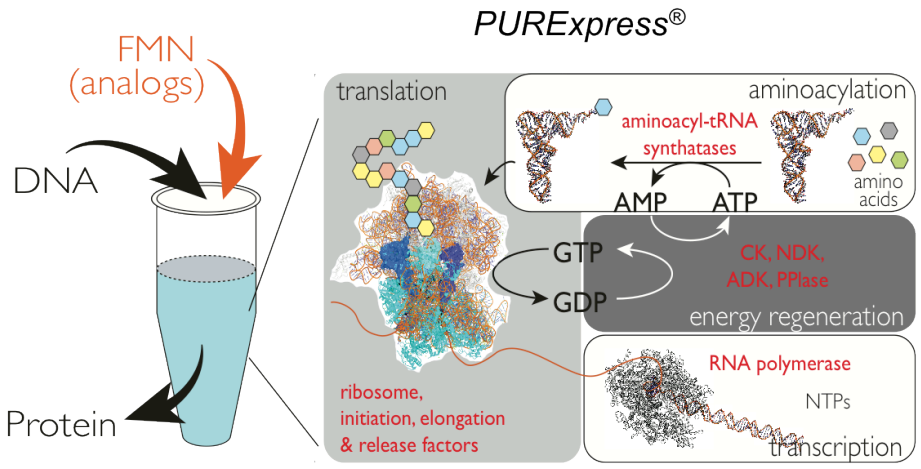


Figure 4.4. Schematic of the PURE system. Upon supplementation of the PURE reagents with FMN (or FMN analog) and DNA, protein is produced within 2 hours. (Illustration adapted from genefrontier.com)

As such, PURE dramatically lowers the background signal caused by contaminating endogenous reductases in cell-based *in vitro* and *in vivo* expression

systems and allows for direct assaying of enzyme activity. Initial control experiments were completed upon establishing the IVTT screen. Two IVTT kits were tested for comparison, the PURExpress *In vitro* Synthesis kit from New England Biolabs and a second cell-based *In vitro* kit from Promega. As expected, protein synthesis by the PURE system resulted in dramatically lower background signal due to the absence of the contaminating reductases unlike in the cell-based system (Figure 4.5). The signal-to-noise ratio was improved for OYE expression enabling a platform for the direct functional analysis of native and engineered OYEs.

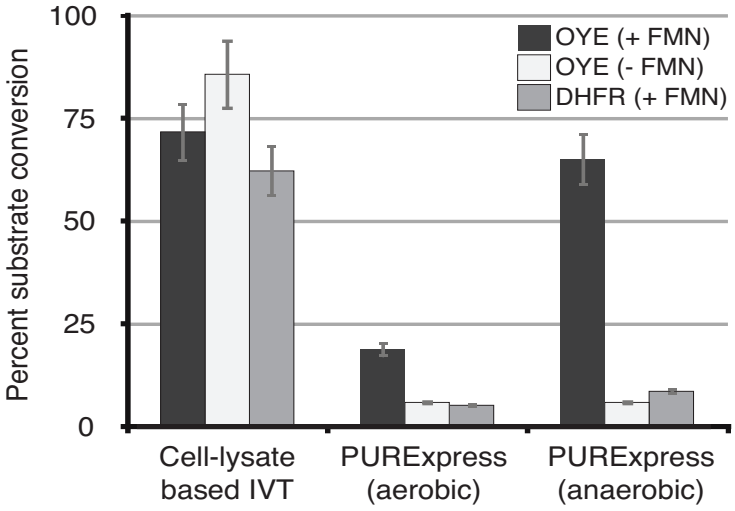


Figure 4.5. Validation of in vitro transcription/translation systems with wild type OYE1. Percent conversion of ketoisophorone (1) to R-levodione (2) by OYE in the presence or absence of FMN and DHFR (- control) synthesized by a cell-lysate based IVTT and PURExpress.

Furthermore, PURE is highly scalable and allows for parallel synthesis of library members in 96-well microtiterplate format. Expression level variability is inherent in a

library and can have a major influence on subsequent activity assay. To test this, OYE and five randomly selected members of the cpOYE library were chosen to evaluate protein yield via SDS-PAGE analysis. Analysis of the gel image based on densitometry indicated that protein yields vary by ~ 2-fold with the exception of cp120 which has slightly lower yield presumably due to stability or folding issues (Figure 4.6).

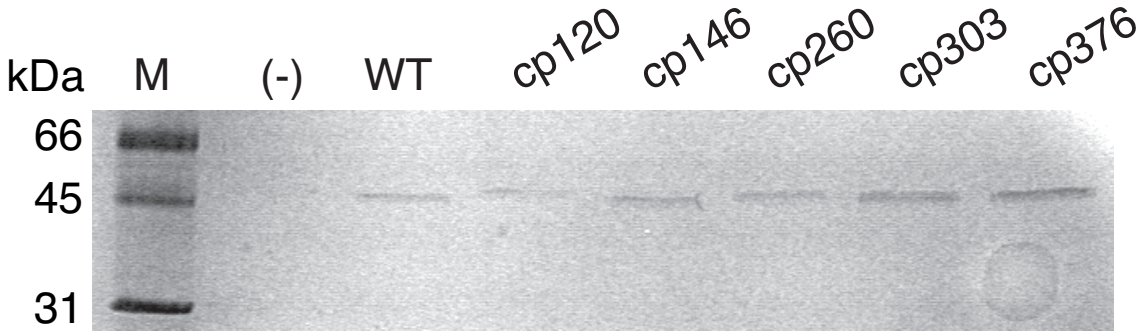


Figure 4.6. Expression level analysis PURE. SDS-PAGE analysis of the protein from a 25 uL PURE reaction of WT (OYE1) and five cpOYE library members. M: molecular weight maker, (-) negative control with DHFR containing plasmid.

In subsequent activity assays, a 10-μL PURE reaction produced enough enzyme for detecting reduction of **1** by OYE1 variants with as little as 10% of wild type activity. Activity assays performed under anaerobic conditions maximize the signal-to-noise ratio, presumably eliminating side reactions and enzyme inactivation by reactive oxygen species generated in the presence of reduced FMN.[22] Finally, expression of catalytically competent OYE required supplementation of the IVTT reaction mixture with FMN. Maximum enzyme activity was reached at [FMN] >1 μM which is approximately 100-fold above the reported dissociation constant for FMN in OYE1 (Figure 4.7).[47] We

further raised the [FMN] to 100 μM to ensure saturation of enzyme with cofactor, even for engineered OYEs with potentially lower cofactor binding affinity.

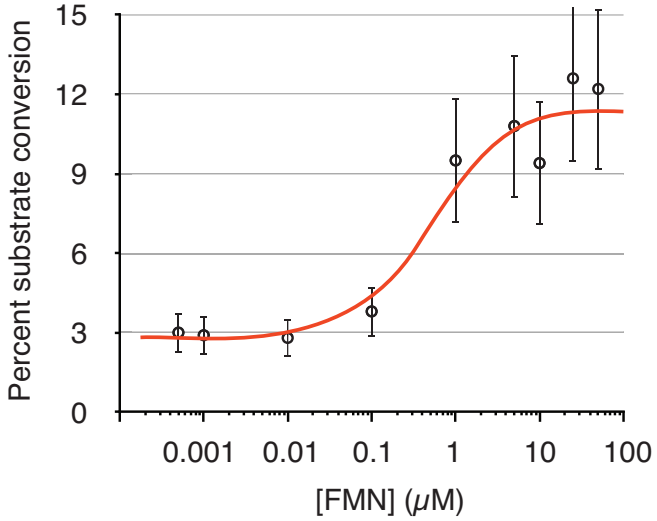


Figure 4.7. Percent conversion of ketoisophorone by wild type OYE1 at [FMN] ranging from 0.5 nM to 100 μM.

4.2.3 Library Screening-KIP

Our initial functional evaluation of the cpOYE library focused on the stereoselective reduction of **1** to *R*-**2**, one of the most prominent industrial application for ene-reductases and a widely used standard for the functional evaluation of OYE family members.[15,19,48-51] Preliminary screening with PURE quickly identified ~70 cpOYE variants with catalytic activity equal or better than wild type enzyme. The location of the new termini in these active variants fell into four distinct sectors within the primary sequence (Figure 4.8).

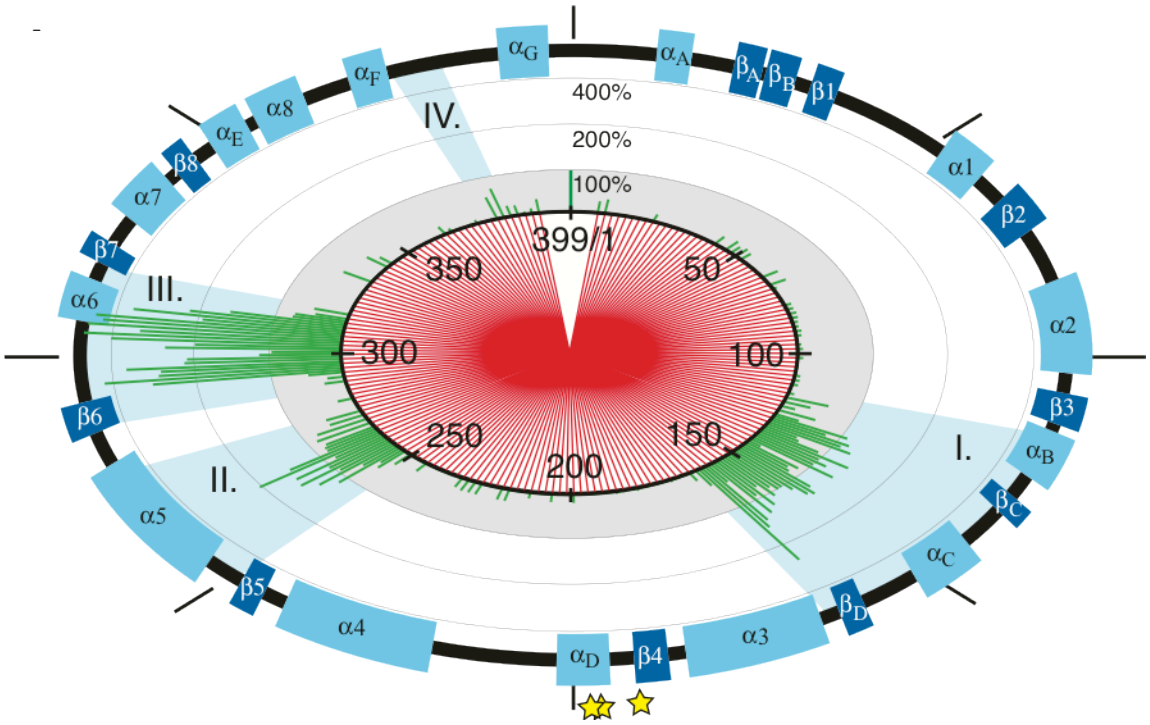


Figure 4.8. Primary screening data of cpOYE library for ene-reductase activity on ketoisophorone (1) to levodione (R-2). Whole-gene synthesis created a native cpOYE library with perfect distribution (red lines in inner circle). Catalytic activity of library members for each substrate was measured by semi-quantitative assay and is reflected in the length of the green lines. Wild-type activity is indicated by the grey-shaded area. The outer-most circle marks the secondary structure elements (shades of blue) and active site residues (yellow stars) of OYE1 (399 amino acid residues). For all three substrates, four sectors (I - IV) in the protein sequence with activity equal or better than wild-type were identified.

Sector I covers most of the exterior helical subdomain (red: amino acid residues 125-160; numbering based on OYE1) while sectors II and III include loop/helix regions 5 (orange: residues 250-265) and 6 (yellow: residues 290-310), respectively (Figure 4.9).

Sector IV represents a short loop (green: residues 375-380) near the native C-terminus. The activity assay suggested that permutants with new termini in sector III were particularly beneficial for effective reduction of **1**, showing >400% of wild type activity for multiple cpOYE variants. Analysis of the reaction mixtures by chiral GC confirmed unchanged *R*-enantioselectivity for all active cpOYE variants with native-like ee_R-values of >98%.

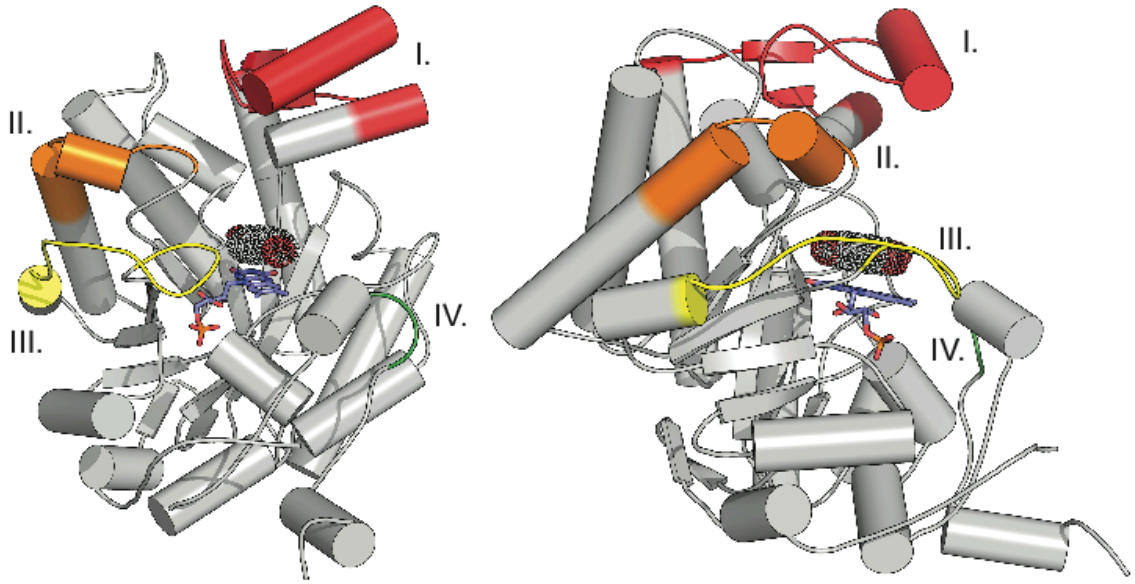


Figure 4.9. Schematic of OYE1 structure (based on PDB access code 1K03 26). The FMN cofactor is shown in blue sticks while the substrate-binding site is occupied by p-hydroxybenzaldehyde (dotted cloud). The location of permutation sites which resulted in catalytic improvements are highlighted in red (sectors I), orange (sectors II), yellow (sectors III) and green (sectors IV), respectively.

4.2.4 Detailed characterization of cpOYE variants

From a practical perspective, we were interested in whether the observed activity differences in PURE were reliable predictors of catalytic activity measured with purified enzyme. To assess the predictability of activity in PURE, we selected nine cpOYE variants (cpOYE146, 154, 160, 257, 260, 291, 292, 303, 306 and 378; the number indicates the location of a variant's new N-terminal based on the numbering of residues in wild type enzyme) which showed 1 to 4-fold activity gains over OYE1 in the PURE reactions with **1**. Enzyme samples of these nine CP variants and OYE1 were obtained through (traditional) heterologous expression and purification, followed by kinetic measurements of individual conversion rates for **1** (Table 4.1). Determination of the Michaelis-Menten parameters was not possible as enzyme saturation could not be reached due to limited solubility of **1**. The kinetic data showed rate increases of up to 19-fold (for cpOYE303) over wild type enzyme. The data also indicate that PURE systematically underestimates the rates compared to purified enzyme. Nevertheless, the observed changes in the IVTT system were overall proportional - the top performers in the PURE system also showed the highest catalytic gains with purified enzyme. These results suggest that the IVTT results are semi-quantitative and can be used as reliable predictors of enzyme activity.

Table 4.1. Comparison of biocatalytic rates for reduction of ketoisophorone (1), cinnamaldehyde (3), and S-carvone (5).

Variant	IVT(**1**)[a]	EA(**1**)[b] (min^{-1})	EA(**3**)[c] (min^{-1})	EA(**5**)[c] (min^{-1})
OYE1	1x	6.5 (1x)	53 (1x)	1.9 (1x)
cpOYE146	1.3x	19 (3x)	60 (1.1x)	1.4 (0.8x)
cpOYE154	0.8x	12.2 (2x)	83 (1.5x)	6 (3x)
cpOYE160	1.6x	10.1 (1.6x)	85 (1.6)	9.5 (5x)
cpOYE257	1x	13 (2x)	44 (0.8x)	5.7 (3x)
cpOYE260	2x	25.4 (4x)	113 (2x)	7.7 (4x)
cpOYE291	3.2x	36 (6x)	21 (0.4x)	3.4 (1.8x)
cpOYE303	5.3x	122 (19x)	24 (0.4x)	22.5 (12x)
cpOYE306	4.2x	47.2 (7x)	42 (0.8x)	23.6 (13x)
cpOYE378	0.7x	nd	57 (1x)	5.9 (3x)

[a] Fold activity change in based on GC analysis. Conversion rate (fold-change) of regular enzyme activity (EA) measurements via [b] GC or [c] UV analysis. Standard error for IVT data ±100%; EA data ±10%. nd = not determined.

4.2.5 Exploring other substrates: cinnamaldehyde and S-carvone

The experimental data for reduction of **1** led to three questions concerning the structure-function relationship in these engineered OYE variants: a.) will the same four sectors be identified upon screening with other substrates, b.) will sector III always be the preferred site for termini relocation, and c.) will native *R*-enantioselectivity be preserved for other substrates? We explored the first question by screening the cpOYE library for improvements in ene-reductase activity with cinnamaldehyde (**3**) and *S*-carvone (**5**) as substrates (Figure 4.10). Both cinnamaldehyde and carvone are important intermediates in the fragrance and flavor industry. Library screening with these two substrates identified functional variants in the same four sectors of the protein sequence. For conversion of **3**, the catalytic improvements upon CP appeared moderate, an observation that was confirmed in activity assays with purified enzyme, which detected up to 2-fold rate increases in catalytic efficiency for cpOYE260 (Table 4.1). We hypothesize that the already high catalytic efficiency of OYE1 for reduction of **3** makes it more likely for additional functional gains to be countered by undesirable structural perturbations upon CP. In contrast, the screening of our cpOYE library with **5** showed significant catalytic improvements for variants in all four sectors. These functional gains were independently verified in subsequent activity assays with purified enzyme, showing 3 to 13-fold increases in the rate of conversion (Table 4.1). As for **1**, the limited substrate solubility of **5** prevents the determination of detailed kinetic parameters and no changes in the native enzyme's *R*-selectivity for the conversion of *S*-**5** to *cis*-1*R*/4*S*-**6** were detected. Summarizing the screening results for **1**, **3** and **5**, all active candidates among the OYE1 variants had their new protein termini located in the same four sectors.

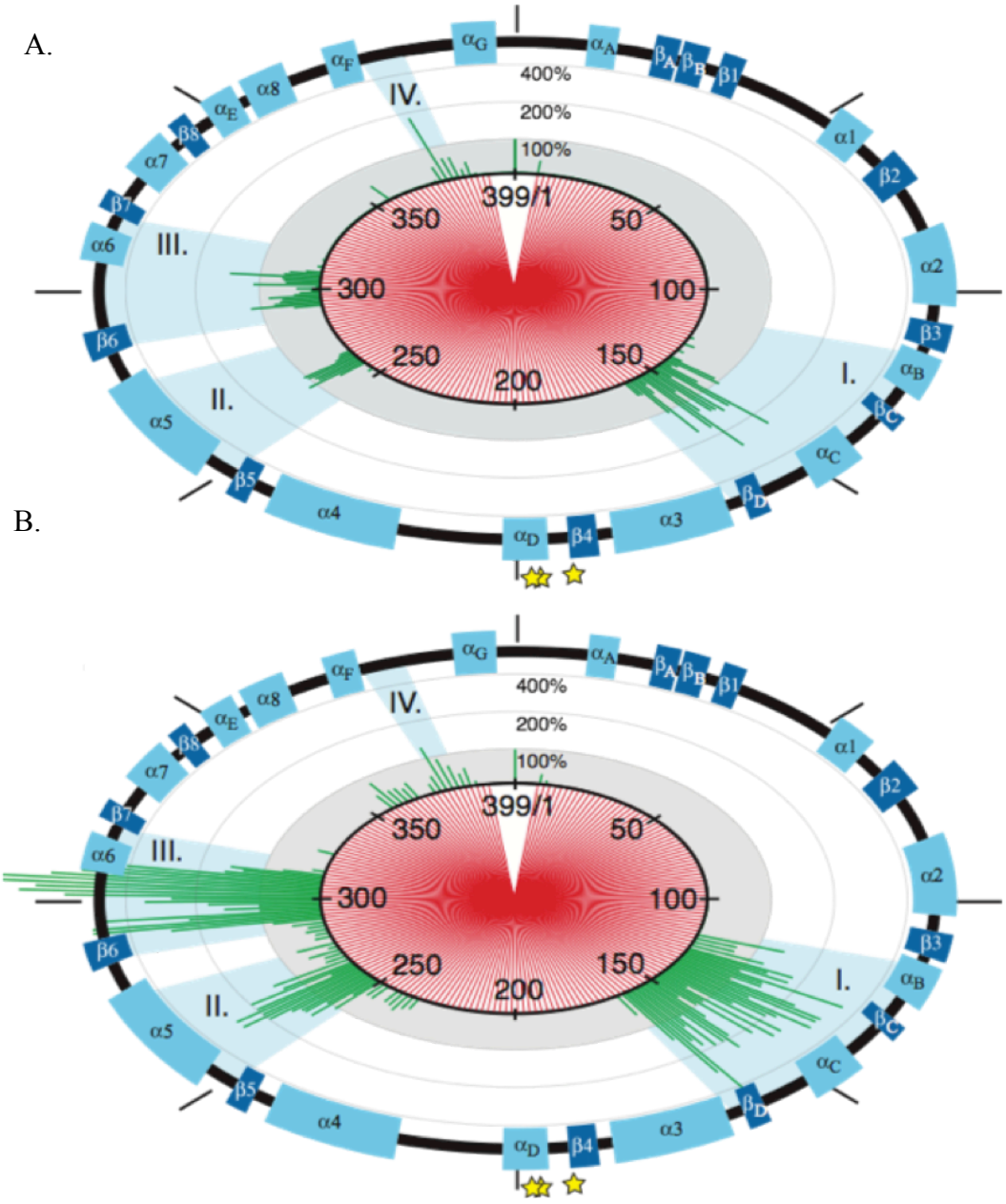


Figure 4.10. Primary screening data of cpOYE library for ene-reductase activity on A.) cinnamaldehyde (3) to dihydrocinnamaldehyde (4), and B.) carvone (S-5) to dihydrocarvone (cis-(1R,4S)-6).

The exclusivity of functional variants in these sectors might at least in part associated with protein foldability upon termini relocation. Nevertheless, their proximity to the active site and in some cases significant improvement in catalytic activity over the native enzyme suggests additional benefits arising from changes in conformational flexibility and active site accessibility. The same data also showed that relative catalytic performance of cpOYE variants differs with individual substrates. Although variants in sector III are clearly the best biocatalysts for 1, sectors I and IV offer slight activity gains for 3 (Table 4.2) while variants from all four sectors show substantive functional benefits for reduction of 5.

The observed preference of the three substrates for protein termini in certain sectors and those sectors' location relative to the substrate-binding site could provide clues to the underlying structure-function relationship. Future in-depth structural studies on selected cpOYE variants via x-ray crystallography will help to address these questions in more detail. In summary, the relocation of protein termini in OYE1 introduced structural and conformational changes in distinct portions of the active site, in turn affecting substrate binding affinity, orientation and catalytic turnover.

Table 4.2. Complete Michealis-Menten kinetics for reduction of cinnamaldehyde (**3**) by wild type OYE1 and selected cpOYE variants.

Variant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M x10^3 (M^{-1} s^{-1})
OYE1	0.97 ± 0.04	18 ± 3	53 (1x)
cpOYE146	0.94 ± 0.04	16 ± 3	60 (1.1x)
cpOYE154	1.50 ± 0.05	30 ± 4	50 (1x)
cpOYE257	1.00 ± 0.04	70 ± 13	14 (0.3x)
cpOYE260	2.20 ± 0.09	44 ± 7	50 (1x)
cpOYE306	0.98 ± 0.04	70 ± 11	14 (0.3x)
cpOYE378	1.43 ± 0.04	14 ± 2	105 (2x)

4.2.6 Rapid reaction kinetics

To further investigate the nature of the catalytic rate enhancements, we conducted a series of rapid reaction kinetic experiments with wild type OYE1 and cpOYE303, the variant with the most significant functional improvements (Figure 4.11). The enzyme's catalytic cycle can be split into two steps; a reductive half reaction representing the NADPH-driven reduction of the oxidized flavoprotein and an oxidative half reaction involving reoxidation of the reduced flavin cofactor upon substrate conversion (Figure 4.1). Based on studies by Massey and coworkers, the reductive half reaction can be further divided into three sub-steps: i) initial binding of NADPH to enzyme (E)•FMN_{ox} followed by ii) conformational changes and repositioning of the nicotinamide cofactor from E•FMN_{ox}•NADPH to E*•FMN_{ox}•NADPH and finally iii) hydride transfer.[29] For wild type OYE1, formation of E•FMN_{ox}•NADPH is observed in stopped-flow experiments as the adduct generates a short-lived charge-transfer complex with a

characteristic red-shifted FMN absorption peak near 460 nm. Subsequent reorientation of the flavin (E*•FMN_{ox}•NADPH) and reduction of the flavin cofactor results in the disappearance of the 460-nm band. In agreement with Massey's data, we detected the intermediate in our reaction catalyzed by wild type OYE1, yet the global fit analysis of spectral data for the engineered variant cpOYE303 did not indicate accumulation of an observable charge-transfer complex prior to reduction (Figure 4.11B). At the same time, the overall rates of the reductive half reaction for OYE1 (k_{red} = 5.1 ± 0.2 s^{-1}) and cpOYE303 (k_{red} = 5.4 ± 0.2 s^{-1}) remained largely unchanged (Figure 4.11C). Given these observations, we rationalized the absence of an intermediate in cpOYE303 with an increased dissociation rate for NADPH from E•FMN_{ox}•NADPH. Such an explanation is also consistent with previous reports from structural studies that emphasized the importance of conformational changes in loop region 5 (amino acid positions 290-310) as part of the reductive half reaction.[27] The location of the new protein termini in cpOYE303 in this region is likely responsible for the lower NADPH binding affinity, reflecting the increased local conformational flexibility upon cleavage of the polypeptide sequence which has been observed in numerous structure studies with circularly permutated proteins.[34,52-55] In summary, protein sequence reorganization by CP does impact the NADPH binding step of the reductive half reaction but at least in the case of cpOYE303 leaves the overall rate of the reductive half reaction unchanged.

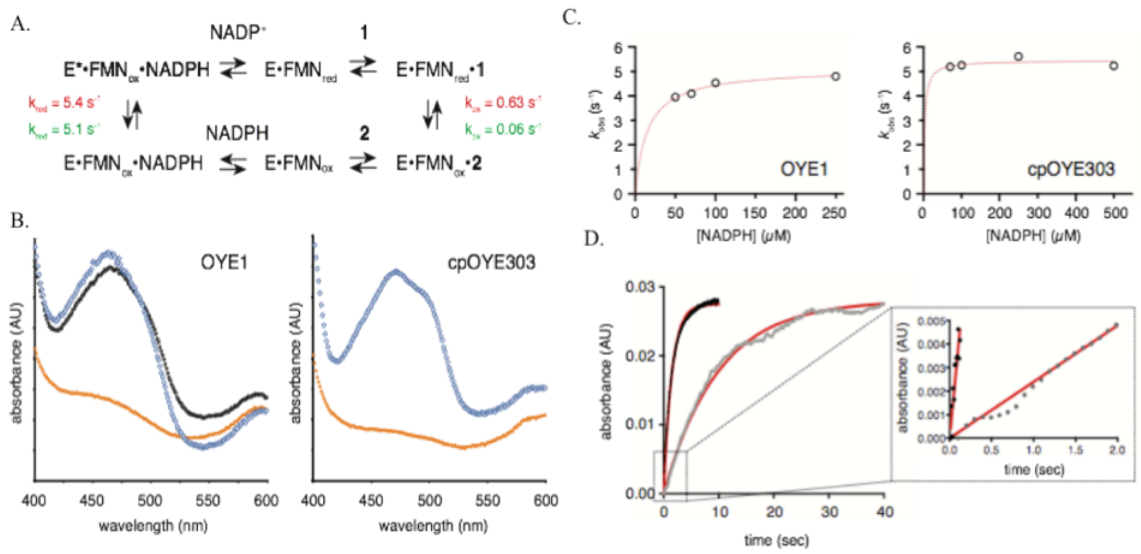


Figure 4.11. Rapid reaction kinetics for wild type OYE1 and cpOYE303 A.) Complete reaction cycle for OYE with rate constants for the reductive and oxidative half reactions of OYE1 (green) and cpOYE303 (red). B.) Changes in flavin spectra for wild type and cp variant during reductive half reaction. The characteristic absorption band of FMNox near 460 nm (blue diamonds) disappears upon conversion to FMNred (orange circles). For OYE1, global fit analysis indicates formation of a short-lived intermediate with a slightly red-shifted maximum near 460 nm (black squares) corresponding to the charge-transfer complex in E•FMNox•NADPH. For cpOYE303, no such intermediate was detectable. C.) Observed rate for the reductive half reaction as a function of [NADPH] based on the disappearance of the FMNox absorption band. Given experimental restraints to [NADPH] >50 μM, data analysis was limited to determination of maximum rates (kred). D.) Flavin reoxidation (kox) upon rapid mixing of E•FMNred with 1 (0.5 mM) under anaerobic conditions followed single-exponential kinetics. Experimental data are shown for the overall and initial (insert) oxidative half reaction catalyzed by wild type OYE1 (grey) and cpOYE303 (black). Curve fits are displayed as red lines.

In contrast, the observed rate for the oxidative half reaction increases significantly upon CP and largely accounts for the gains in catalytic activity measured by steady-state

kinetics. Following pre-incubation of the two OYEs with NADPH to fully reduce the enzyme-bound FMN, the rates of cofactor reoxidation (k_{ox}) were measured after mixing with **1** (Figure 4.11D). For both enzymes, the data could be fitted to a single-exponential function, yielding initial rates of conversion for cpOYE303 and OYE1 of 0.63 ± 0.09 s^{-1} and 0.06 ± 0.01 s^{-1}, respectively. The data indicate that FMN reoxidation with **1** is significantly slower than cofactor reduction in both OYEs and likely represents the rate-limiting step in the reaction cycle. More importantly, we observe an eleven-fold rate increase for cpOYE303 over OYE1 which accounts for most of the gains in catalytic activity observed for the OYE variant in our steady-state kinetic measurements (Table 1). Although these initial studies do not reveal detailed information regarding the effects of CP on the individual steps along the reaction coordinate, the results clearly show that rate enhancement in the OYE-catalyzed reduction of **1** by cpOYE303 is due to changes in the oxidative half reaction. However, it seems unlikely that the observed functional effect is general. Previous studies have shown that the nature of the rate-limiting step in OYE can change as a function of substrate.[29,30] While we believe that the catalytic gains measured for the reduction of *S*-(**5**) originate from similar beneficial effects on the oxidative half reaction, the lack of change for **3** could be explained by differences in the rate-determining step.

4.2.7 *p*-Hydroxybenzaldehyde Binding Studies

In the absence of detailed crystallographic information, we investigated perturbations of the active site environment as a result of CP via changes in the spectral properties of the flavin cofactor. Upon binding of *p*-hydroxybenzaldehyde to OYE1,

Massey and coworkers observed distinct long wavelength charge-transfer bands ranging from 560 to 590 nm with λ_{max} values depending on hydrogen-bonding interactions of the phenolate and the FMN redox potential.[47,56] The spectral properties of cpOYE variants can therefore serve as sensitive probes for the active site environment. Our analysis of representative variants from all four sectors with *p*-hydroxybenzaldehyde did indicate subtle but significant active site changes (Table 4.3 & Figure 4.12). While small, two-fold differences in dissociation constants for the phenolate are consistent with overall integrity of the active site binding pocket, the shifts in λ_{max} values of the charge-transfer bands from 564 nm to 582 nm (relative to 578 nm for wild type OYE1) reflect changes in the FMN redox potential of ~20 mV which suggest small conformational and environmental perturbations near the cofactor. Nevertheless, the change in redox potentials for the cpOYE variants did not show a clear correlation with catalytic activity for our three substrates.

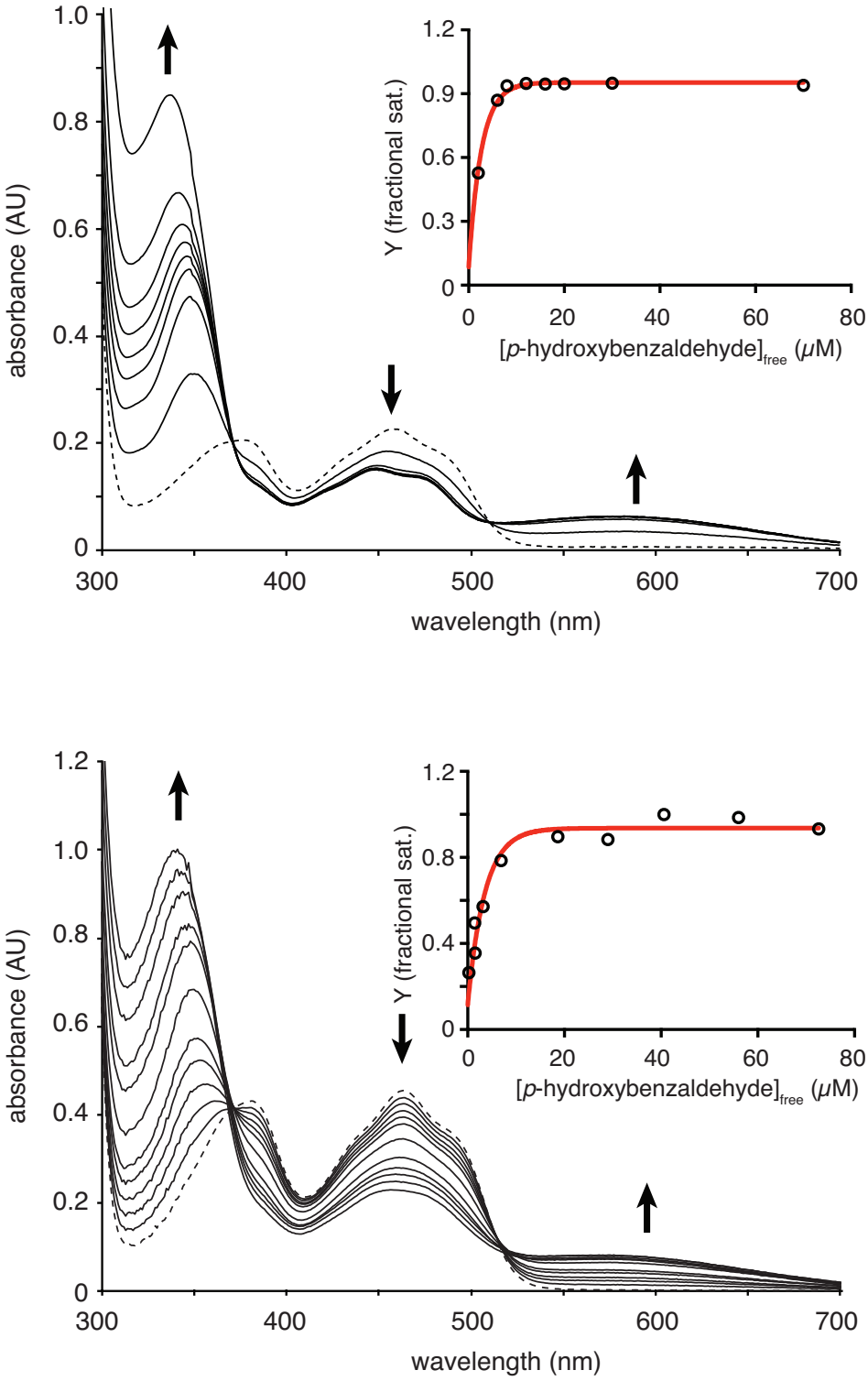


Figure 4.12. Changes in absorbance spectra upon titration of p-hydroxybenzaldehyde to OYE1 and cpOYE303. Dashed line represents spectrum in the absence of phenolic inhibitor.

Table 4.3. Spectral property changes upon inhibitor binding

Variant	$K_D(\mu M)$[a]	λ_{max}(CT; nm) ($\varepsilon = M^{-1}cm^{-1}$)	λ_{max}(FMN; nm) ($\varepsilon = M^{-1}cm^{-1}$)	$E^{\circ\prime}$ (mV)[b]
OYE1	1.6 ± 0.4	578 (4335)	460 (10600)	-207
cpOYE154	1.0 ± 0.3	582 (3753)	457 (12784)	-196
cpOYE160	0.6 ± 0.2	572 (2971)	462 (11913)	-210
cpOYE257	1.9 ± 0.7	566 (2485)	463 (11341)	-214
cpOYE260	1.1 ± 0.4	580 (2973)	460 (11979)	-200
cpOYE291	4.0 ± 1.0	564 (2956)	464 (12260)	-216
cpOYE303	1.9 ± 0.2	569 (2685)	463 (13147)	-212
cpOYE306	2.9 ± 0.9	565 (2756)	458 (10774)	-215
cpOYE378	2.3 ± 0.9	578 (3702)	458 (12224)	-207

[a] Dissociation constant for *p*-hydroxybenzaldehyde. [b] E°'-values were estimated based on correlation of λ_{max}(CT) band with FMN E°'-values.[47] Standard error: ±5 mV

4.2.8 Secondary Engineering Cofactor Analogs

Initial exploratory experiments with cofactor analogs involved looking at the two extremes of the redox spectrum compared to FMN (-207 mV): the more positive 7,8-dichloro FMN (-120 mV) and the more negative 8-amino FMN (-330 mV) and 5-deaza FMN (-311 mV). All three cofactor analogs were synthesized by Dr. Edmondson to the riboflavin form. Briefly, 7,8-dichloro riboflavin and 8-amino riboflavin were synthesized

as reported by Kasai *et al* and Lamboy *et al*.[57,58] The synthesis of 5-deaza riboflavin was performed as reported by Yoneda and coworkers except triphosgene was substituted with the Vilsmeier reagent.[59,60] The riboflavin analogs were converted to the FAD form using purified FAD synthetase from *Corynebacterium ammoniagenes*, followed by the hydrolysis of the adenyl moiety with phosphodiesterase to arrive at FMN.[61] Both reactions were monitored by TLC, UV and fluorescence spectroscopy to assure full conversion. Final product was confirmed by mass spectroscopy. After synthesis, apo OYE and permutants were prepared for subsequent FMN analog incorporation. The purified enzyme was placed in a dialysis bag with stripping buffer for 2-3 days to remove native FMN.[47] The enzyme was reconstituted with the appropriate analog and successful binding was indicated by a red shift in the UV spectrum.[62]

The first set of cofactor analog experiments was conducted with 5-deaza and 8-amino FMN to probe changes in oxygen sensitivity with the altered redox potentials.[63] We hypothesized that by substituting FMN with cofactors that have lower oxygen sensitivity it would simplify the reaction procedure and potential side reactions would be avoided. The catalytic performance was investigated by assaying enzyme activity in both anaerobic and aerobic conditions with substrate **1.** Several reactions were performed comparing OYE wild type, 8-amino-OYE, and 5-deaza-OYE in both the presence and absence of oxygen. The 5-deaza experiments were quickly dismissed as the analog was very photo reactive producing inconstant results from photo reduction taking place. In comparison, 8-amino-OYE conversions were very reproducible and showed that, in general, the reductase activity for **1** is lower upon substituting OYE with 8-amino FMN. However, catalytic function of 8-amino-OYE is still compromised when the reaction is

subjected to aerobic conditions (Table 4.4). The data suggests that other reactions compete with the reduction of **1** in the presence of oxygen with a much lower redox potential. Further studies will need to be conducted to measure the specific side reactions that take place with wild type OYE and compare these rates with the 8-amino-OYE to see if they are slower or not detected. For example, OYE has NADPH oxidase activity and in the absence of alkene substrates oxygen can reoxidize reduced FMN.[64] This is one reason reactions are performed anaerobically, as their competing activity is eliminated.[64] Experiments could explore the differences in oxidase activity between wild type OYE and 8-amino-OYE.

Table 4.4. Conversion of KIP for 8-amino-OYE compared to Wt-OYE at various enzyme concentrations in aerobic and anaerobic conditions.

Reaction	Aerobic		Anaerobic	
	Conversion %	ee %	Conversion %	ee %
wt-OYE 1 μM	55.1	90.7	61.4	89.7
wt -OYE 3 μM	84.9	88.8	90.6	89.9
wt -OYE 6 μM	97.1	89.3	98.8	89.4
8-amino-OYE 1 μM	7.4	100.0	12.0	100.0
8-amino-OYE 3 μM	16.3	100.0	26.0	100.0
8-amino-OYE 6 μM	27.1	100.0	46.8	91.7

In addition to using unnatural FMN analogs to alter oxygen reactivity, changing redox potential may also enable the enzyme to catalyze the reverse reaction. Massey and coworkers were successful in converting OYE from a reductase to an oxygen dependent desaturase by replacing native FMN (-207 mV) with the 8-CN-FMN (-50 mV).[42] In turn, changing the redox potential relative to the substrate reversed the flow of electrons,

allowing the reverse reaction to occur. Mechanistically, the oxidized form of 8-CN-OYE reacts with the substrate yielding reduced cofactor, which is reoxidized by molecular oxygen (Figure 4.13).[42] The reconstituted enzyme displayed efficient desaturase activity toward a number of carbonyl compounds. The highest activity was seen with **4** (dihydrocinnamaldeyde) to yield **3** (cinnamaldehyde. The k_{cat} of the reduction of **3** by native OYE is 125 min^{-1} and upon substitution with 8-CN FMN the k_{cat} was reduced to 3.5 min^{-1} and the established desaturase activity was 11.1 min^{-1}.[42]

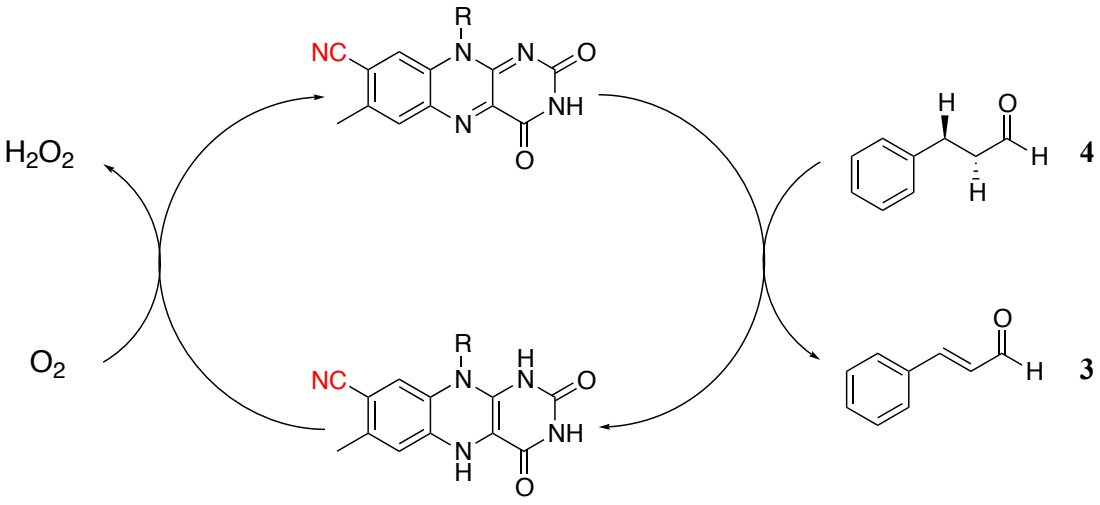


Figure 4.13. The oxygen dependent oxidation of carbonyl compounds to their corresponding α,β-unsaturated alkene by 8-CN-OYE.

A cofactor with a slightly more negative redox potential, 7,8-dichloro FMN (-120 mV), was used in our studies. A large stock of previously synthesized 7,8-dichloro FMN was available unlike the 8-CN-FMN which would have to be synthesized. The 7,8-dichloro analog has never been tested with OYE for desaturase chemistry and it wasn't clear at what point on the redox scale at which desaturase activity would be introduced. After reconstitution of wt-OYE with 7,8-dichloro FMN, binding was apparent by monitoring a red shift in the UV spectrum (Figure 4.14).[62] It appears that binding affinity of the analog to apoOYE is very similar to FMN. Activity assays with **4** showed moderate desaturase activity for 7,8-dichloro-OYE and no activity was detected for native OYE. Full Michaelis-Menten kinetics were determined in order to establish the catalytic performance compared to wild type enzyme (Table 4.5). The 7,8-dichloro-OYE displayed a 30-fold decline in the reduction of **3** and moderate desaturase activity was introduced with a k_{cat} of 1.2 min^{-1}. Next, we wanted to see if desaturase activity varied between the permutants and wildtype OYE with **4**. Two permutants from different structural regions, cpOYE154 and cpOYE303 were chosen to test the impact of both permutation coupled with the substitution of cofactor analog. The purified permutants reconstituted with 7,8-dichloro FMN were able to catalyze the reverse desaturase reaction for **4** with cpOYE154 performing 2-fold better than cpOYE303 and about the same as wt-OYE. This initial experiment showed that possible synergist effects were resulting from the variants (Figure 4.15 & Table 4.6).

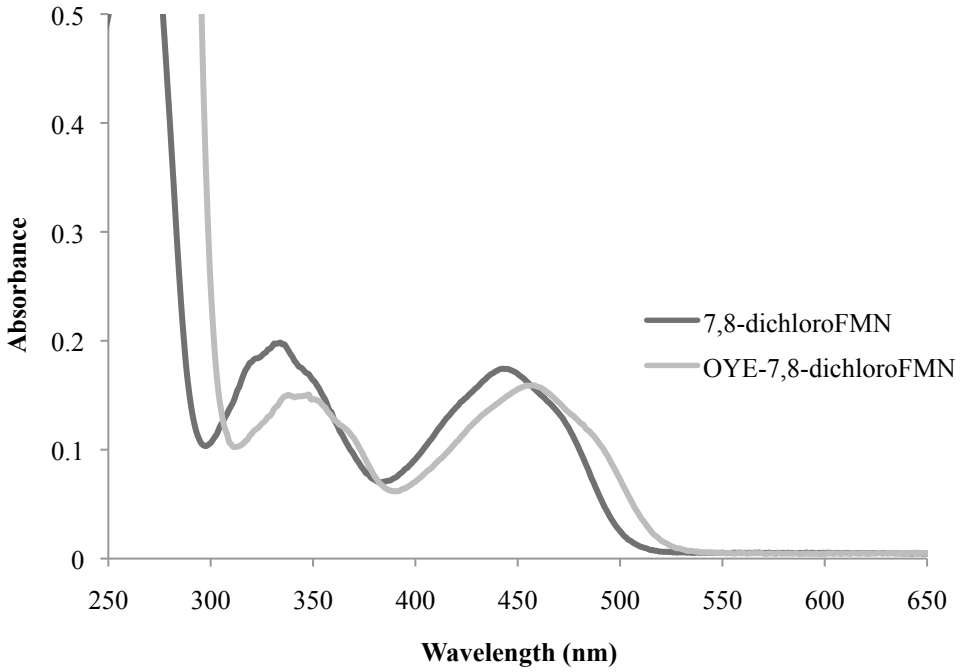


Figure 4.14. UV spectrum of 7,8-dichloro FMN in solution and bound to native OYE

Table 4.5. Michaelis-Menten parameters for wt-OYE and 7,8-dichloro-OYE with 3 (substrate for reduction) and 4 (substrate for desaturase).

Cofactor	**Cinnamaldehyde Reduction**		**Dihydrocinnamaldehyde Desaturase**	
	k_{cat} (min^{-1})	K_M (μM)	k_{cat} (min^{-1})	K_M (μM)
wt-OYE	58.2 ± 0.04	18.2 ± 3.0	ND	ND
7,8-dichloro-OYE	1.9 ± 0.03	12.1 ± 2.5	1.2 ± 0.09	23 ± 6.3

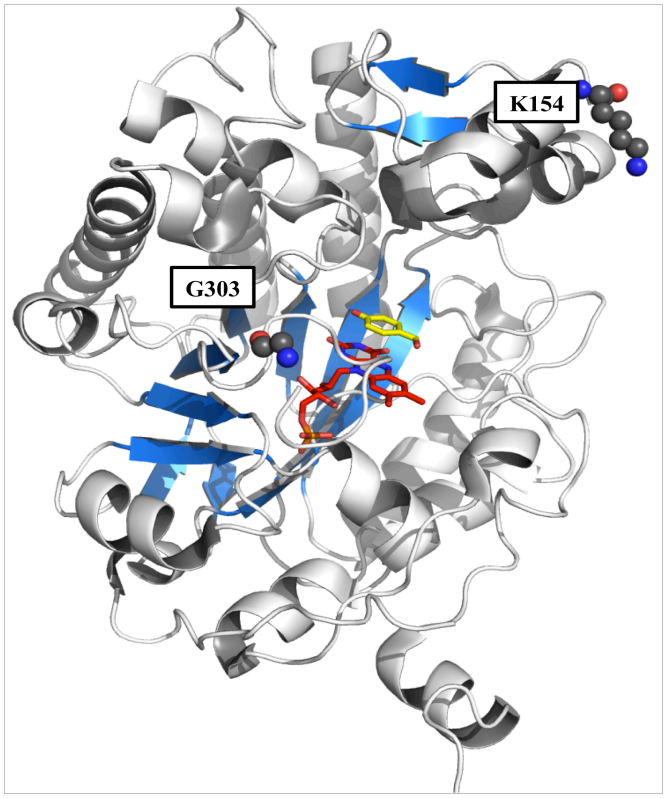


Figure 4.15. OYE permutants selected for 7,8-dichloro FMN incorporation. The blue highlights the core structure and the locations where the new termini reside are labeled.

Table 4.6. Turnover of selected permutants with dihydrocinnamaldehyde

Enzyme	**Turnover number @ 200 µM O_2 saturated buffer (min^{-1})**
OYE-FMN	0
7,8-dichloro-OYE	0.96
7,8-dichloro-cpOYE154	1.1
7,8-dichloro-cpOYE303	0.40

In order to test the effects of both permutation and cofactor replacement on a larger scale the PURE screen was employed for cofactor replacement evaluation in a high throughput format. The PURE system simplifies the evaluation of cofactor analogs as it eliminates the time consuming and labor-intensive dialysis step of purified enzymes. The stripping procedure of purified enzyme is straightforward but it is very time consuming especially when looking at multiple proteins. Initial attempts to express cpOYEs with unnatural FMNs using the commercially available PURE reaction were unsuccessful as the UV traces revealed modification of the flavin. After further investigation it became clear that the reducing agents dithiothreitol (DTT) and beta-mercaptoethanol (bME) in the formulation caused a substitution of a thiol for the chlorine group at the 8-position of the flavin.[65] Optimization revealed that tris(2-chloroethyl) phosphate (TCEP) could be used in place of DTT and the bME could be reduced by half to 3.5 mM providing the same fidelity of the original PURE system without flavin modification. New England Biolabs generated the modified formulation, which was successful for the production of OYE with 7,8-dichloroFMN incorporation. A limited set of permutants (3-5 variants) was selected from each of the four regions to screen for desaturase activity. The production of enzyme in the PURE system with the 7,8-dichloro FMN analog was substantially lower than with the native FMN cofactor. Desaturase activity was still detectable at these enzyme levels and in several cases the permutants exhibited higher activity compared to 7,8-dichloro-OYE (Figure 4.16). Looking at the results its seems that the same trend in activity improvements between the sectors carried over with the reverse reaction and synergistic effects emerge from the library evaluation. Regions IV and I displayed the highest improvement in desaturase activity compared to 7,8-dichloro-OYE. Interesting

when these regions are compared to the previously estimated redox potentials from the phenol binding studies (Table 4.3), the regions that displayed the more positive shifts in redox potential also showed the highest activity for **4**. Future studies need to be conducted to further explore this correlation.

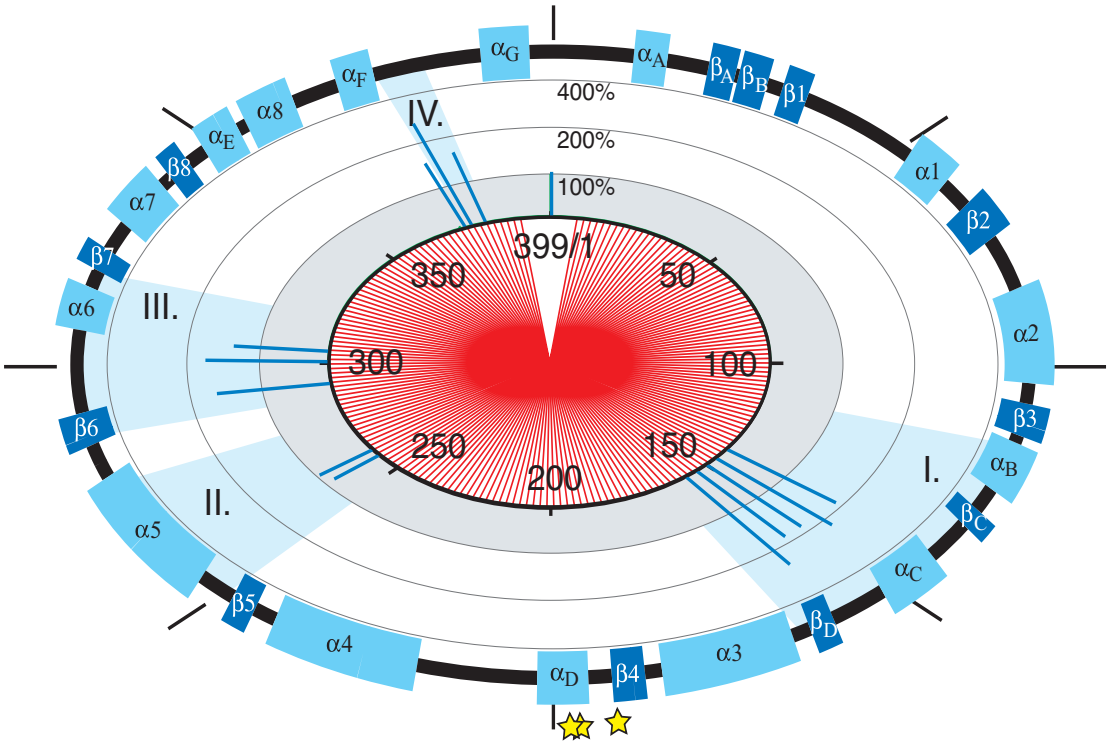


Figure 4.16. Primary screening data of selected cpOYE library for desaturase activity on dihydrocinnamaldehyde (4) to cinnamaldehyde (3).

4.3 Conclusions

In conclusion, we have identified a series of circularly permuted OYE variants with enhanced ene-reductase activity for two representative enones and an enal. For all

three substrates, the new protein termini of catalytically active and improved enzymes fall in the same four structural regions, including three loops and a helical subdomain near the active site. The functional benefits of new termini in these regions varied with substrate, presumably reflecting subtle differences in protein structure and interaction with the substrates. The native enzyme's high enantioselectivity was preserved in all cpOYE variants and catalytic activity for **1** and *S*-**5** was improved by >10-fold. The rapid reaction kinetics analysis of the oxidative and reductive half reactions in wild type OYE1 and cpOYE303 suggests that the functional gains can almost exclusively be attributed to improvements of the rate-limiting oxidative half reaction with **1**. Further in-depth studies of the individual steps in the reaction cycle will be required to explore whether the observed rate enhancements are more specifically due to changes in binding of **1**, release of (*R*)-**2**, or the catalytic step. Initial evidence in support of beneficial changes to substrate/product affinity upon CP was identified by the absence of the characteristic charge-transfer complex (E•FMN_{ox}•NADPH) in the reductive half reaction of cpOYE303. The accumulation of the complex in wild type OYE1 has been associated with a required conformational change involving loop 5; the same loop where the new protein termini in cpOYE303 are located. The increased conformational flexibility of a protein's termini region is well established in the literature and could explain the improved catalytic activity in our engineered OYE variants. Crystallographic and in-depth biophysical studies explored in the next chapter help to further rationalize the observed differences in catalytic performance. Beyond the exploration of fundamental aspects of enzyme catalysis, our results demonstrate that CP offers an effective strategy for improving the catalytic performance of OYE1. Although a direct comparison with

other mutagenesis-based engineering approaches is complicated by differences in reaction conditions and substrates, the rate enhancements for our cpOYE variants are of similar magnitude or better than for previously reported OYE variants.[20,66] For future experiments, the structural and functional similarities among enzymes of the OYE family suggests that CP could result in similar functional gains in other family members. Furthermore, selected cpOYE variants do not necessarily mark the end point of a protein engineering project but can serve as a novel templates for subsequent mutagenesis experiments, allowing for even greater diversity in tailoring these biocatalysts to specific conditions and substrates. On the technical side, the use of a chemically defined cpOYE gene library in combination with the PURE system dramatically accelerated the process of identifying catalytically improved cpOYE variants for any substrate of choice. As technological advances continue to raise efficacy and lower costs, we envision such synthetic protein engineering approaches to become a popular new tool for synthetic biologists. For engineering OYE and cofactor-dependent enzymes in general, the use of IVTT systems also offers unprecedented opportunities for combinatorial approaches to exploring the functional impact of cofactor substitutions. Exploration of selected cpOYE variants with 7,8-dichloro FMN identified permutants that were better at performing the reverse of reducatase chemistry, desaturase, compared to native OYE. The results suggest that synergistic effects occur as a result of both permutation to the OYE scaffold and replacement of native FMN with a more positive redox potential cofactor.

4.4 Materials and Methods

4.4.1 Materials

Whole-gene synthesis was performed by DNA2.0 (Menlo Park, CA). PURExpress kits were purchased from New England Biolabs (Ipswich, MA). All reagents, substrates, and reference materials were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated.

4.4.2 cpOYE1 library synthesis

The gene encoding OYE1 from *S. pastorianus* (formerly *S. carlsbergensis*; NCBI access number: X53597.1) served as starting sequence for our experiments. The CP library was prepared by PCR amplification of gene sequences encoding the individual variants using an *oye1* tandem repeat as template (Figure 4.3). The tandem repeat was created by whole-gene synthesis, linking two copies of *oye1* head-to-tail via an oligonucleotide sequence encoding a three amino acid linker (-Gly-Ser-Thr-). The Start and Stop codons of the 399-amino acid enzyme were eliminated in the tandem repeat. Subsequently, gene sequences corresponding to individual circular permuted enzymes were obtained by PCR amplification using the tandem repeat as template in combination with sequence-specific oligonucleotide primers. In the 5'-regions of the forward primers, a ATG start codon was included while reverse primers carried a TAA termination codon. The PCR products were subcloned into pET-14b (Novagen) via flanking type-IIs restriction endonuclease cleavage sites (*Bsa*I) to enable ligation of PCR products into the vector's *Nco*I/*Xho*I sites independent of variations in the *oye1* library sequences. Following transformation of individual ligation reactions into *E. coli*, plasmid DNA from selected colonies was

extracted and gene sequences were verified by DNA sequencing. Aliquots of confirmed plasmids carrying specific library members were stored separately in 96-well microtiter plates.

4.4.3 Primary library screening

All members of the cpOYE library were initially evaluated for catalytic function by *in vitro* transcription/translation (IVTT) followed by an in situ ene-reductase activity assay. Overall, the IVTT reactions were assembled using the PURExpress *in vitro* protein synthesis kit (PURE) following manufactures protocol with a few adjustments (see below) optimized for our application. Reactions were assembled in a 96-well microtiter plate on a 10-μL scale containing PURExpress solution A, PURExpress solution B, 100 μM FMN, 10 units RNase Inhibitor (NEB), 100 ng DNA template, and nuclease-free H_2O. All reactions from the cpOYE library screening were ran simultaneously with wild type OYE1 (positive control) and dihydrofolate reductase (DHFR; negative control; provided with the PURExpress kit) under the same experimental conditions. The reactions were incubated at 37˚C for 2.5 h to allow for adequate protein synthesis prior to the activity assay. The reactions were cooled to 4˚C to stop the reaction. Subsequently, ene-reductase activity was assessed under anaerobic conditions (Coy Laboratory Products, Grass Lake, MI) at ambient temperature utilizing glucose dehydrogenase (GDH) from *Thermoplasma acidophilum* for NADPH regeneration. The IVTT reaction was mixed with substrate (200 μM ketoisophorone (**1**), 200 μM cinnamaldehyde (**3**), or 1 mM *S*-carvone (**5**)), 200 μM $NADP^+$, 2 U GDH, and 100 mM glucose. The 30-μL assay ran for 2.5 h (due to high turnover, reaction time for **3** was reduced to 30 min) and was

quenched by mixing with an equal amount of ethyl acetate containing 1 mM cyclohexanone as internal standard. A sample of the organic phase was injected on an Agilent technologies 6850 GC machine equipped with a chiral CycloSil-B column (Agilent, Santa Clara, CA) and an FID detector. The percent conversions and enantiomeric excess was calculated from substrate and product integration areas and were quantified using standard curves generated using known amounts of the substrate and product.

4.4.4 Protein expression and purification

Selected members of the cpOYE library were chosen for a more detailed *in vitro* characterization in order to confirm and further characterize interesting variants from the primary library screening analysis. Individual plasmid DNA encoding the corresponding library members was initially transformed in *E. coli* BL21(DE3)pLysS for expression. Colonies were cultured in 2 mL LB medium containing chloramphenicol (34 µg/ml) and ampicillin (100 µg/ml) overnight at 37˚C. The overnight culture was used to inoculate 2YT medium containing the same antibiotics and cultures were grown at 37˚C until the OD(600) reached 0.5 to 0.7. Overexpression was induced by addition of IPTG to a final concentration of 0.4 mM and cultures were overexpressed for 18 h at 20˚C. Subsequently, cultures were centrifuged for 20 min at 4˚C and 4000*g* and cell pellets were stored at -20˚C until purification. The purification of wild type OYE1 and variants followed the same purification procedure. Cell pellets from a 250 mL culture was resuspended in 6 mL of buffer A (40 mM Tris-HCl (pH 8.0), 20 mM NaCl). To the mixture, 75 µL of protease inhibitor cocktail (Sigma), 7.5 µL of benzonase (Novagen), and 10 µM PMSF were

added and stored on ice for 30 min. Cells were lyzed using sonication (8x with 10 sec pulses and 20 sec pauses). After centrifugation for 30 min at 4˚C and 10,000*g*, the clear lysate was further purified via anion-exchange chromatography (HiTrap Q HP-5mL column pre-equilibrated with buffer A). The column was washed with 2 column volumes (CV) of buffer A, followed by a linear gradient to 100% buffer B (40 mM Tris-HCl (pH 8.0), 1 M NaCl) over 10 CV. Product fractions were combined and concentrated to ~1 mL using a Millipore filter unit (MWCO: 10 kDa). In a final polishing step, proteins were purified by size exclusion chromatography (Superdex 200, 10/300 GL column equilibrated with buffer C (40 mM Tris-HCl (pH 8.0), 300 mM NaCl); flow rate: 0.5 mL/min). Elution of protein was monitored by UV-detection at 280 nm and 460 nm and product fractions were combined. SDS-PAGE analysis of the final product showed >95% purity.

4.4.5 Spectral properties of OYE1 variants

All spectral analysis of OYE1 and permutants was completed using a Varian Cary 100 spectrophotometer. The extinction coefficient for each variant was determined by recording the absorbance spectrum of the protein-bound FMN solution in 10 mM Tris-HCl (pH 7.5) buffer. A 25 µL aliquot of 10% SDS was added to denature the protein and spectra were recorded until changes were no longer observed. The extinction coefficient for each enzyme-bound FMN was calculated using the following equation: $\varepsilon = \varepsilon_{free(FMN)}$ (12,500 $M^{-1}cm^{-1}$) x (absorbance (at λ_{max}) of FMN_{bound} (prior to SDS treatment) / absorbance (at λ_{max}) of FMN_{bound} (after SDS)).[67]

Phenolate inhibitor binding assays were completed at room temperature in 100 mM potassium phosphate buffer pH 7 using a cuvette with a 1-cm path length (Figure 4.12) The initial enzyme concentration (10-30 μM) in each experiment was calculated by the corresponding FMN extinction coefficient. Aliquots of *p*-hydroxybenzaldehyde were added to the protein solution and spectra were recorded after 3 min of equilibration until apparent saturation was reached. The long wavelength charge transfer maximum values for each data set were determined by first derivative analysis and extinction coefficients were calculated based on the absorbance values. Separately, dissociation constants were calculated by plotting $[ligand]_{free}$ versus Y (fractional saturation of OYE) and analysis by non-linear curve fit using GraphPad Prism6 (GraphPad, La Jolla, CA) (Fig.S1 inserts).

4.4.6 Activity assays

The enzyme concentration for each cpOYE variant was determined based on the extinction coefficient of protein-bound FMN. Purified enzyme was assayed in an anaerobic chamber (Coy LabProducts) by incubating 0.1-1 μM enzyme with 1 mM **1**, 200 μM $NADP^+$, 5 U GDH, 100 mM glucose in 500-μL total reaction volume. The reaction was monitored over 45 min (timed to achieve less than 30% product formation) by taking aliquots and mixing with an equal amount of ethyl acetate and analyzed via GC (see above). The product was quantified using a standard curve generated using known amounts of the product (*R*)-levodione (**2**) and 1 mM cyclohexanone as internal standard. Alternatively, turnover was monitored spectrophotometrically at ambient temperatures with the Bioteck EPOCH plate reader housed in the anaerobic chamber. Purified enzyme was assayed following the decrease in absorbance of NADPH at 340 nm (molar

extinction coefficient 6,220 $M^{-1}cm^{-1}$). The 300-µL reaction contained 100 mM sodium phosphate buffer (pH 7.0), 1 mM of the appropriate substrate (from 20 mM stock in ethanol), 100 µM NADPH and 0.1-1 µM enzyme. For **3**, where substrate saturation was not a problem, kinetic constants were determined by fitting initial rates to the Michaelis-Menten equation through a non-linear curve fit using Origin7 (OriginLab, Northampton, MA) (Table 4.2). The stereoselectivity displayed by the permutants for reducing (*S*)-**5** was determined by incubating 25 µg enzyme with 5 mM substrate, 200 µM $NADP^+$, 5 U GDH, and 100 mM glucose in 500-µL total reaction volume. The reaction was carried out for 24 h at room temperature. After quenching, the reaction was extracted with an equal amount of ethyl acetate and the organic layer analyzed by chiral GC.

4.4.7 Stopped flow experiments

Measurements were made on an OLIS RSM-1000 (Olis Inc., Bogart, GA) stopped flow spectrophotometer under anaerobic conditions in 50 mM Tris-HCl (pH 7.5) at 25 °C. For the reductive half reaction, OYE1 or cpOYE303 (10 µM) were mixed in the spectrophotometer with NADPH (final concentration: 50 - 500 µM). The oxidative half reaction was studied with **1** as substrate. Enzyme (13 µM) was preincubated with 20 µM NADPH to completely reduce FMN, followed by mixing with **1** (0.25 mM and 0.5 mM for OYE1; 0.5 mM and 1 mM for cpOYE303). Spectra were recorded with delays of 10 sec and 100 sec after mixing. All data analysis was performed in GraphPad Prism 6 (GraphPad, La Jolla, CA).

4.4.8 Cofactor analog incorporation

Native OYE and variants were stripped of all FMN via three rounds of dialysis (dialysis membrane tubing MWCO: 10 kDa) using FMN stripping buffer (100 mM potassium phosphate (pH 5), 2 M KBr, 6.7 mM EDTA, and 10 μM PMSF) over a period of 24 h. After removal of FMN, the apo protein was dialyzed 100 mM Tris-HCl, 20 mM NaCl (pH 7.6) until all of the KBr was removed (typically 3-4 rounds of dialysis). For reconstitution, binding reactions were set up containing a 5:1 ratio of analog to protein. The reaction was incubated at 4 °C for 2 h. Following incubation; ultra filtration (Millipore filter unit MWCO: 10 kDa) was used to wash out any unbound cofactor. The binding and concentration of analog was monitored via UV. For reconstitution using the PURE system the cofactor analog was supplemented during IVTT reaction assembly at a final concentration of 200 μM.

4.4.9 Desaturase activity assays

For reconstituted purified enzyme, activity was measured under aerobic conditions at ambient temperature using the UV spectrophotometer and following the absorbance maximum for **3** (292 nm). Reactions cuvettes contained substrate **4** in oxygen saturated 0.02 M pyrophosphate buffer, pH 8.5 at a final volume of 500 μL. The final concentration of enzyme per reaction was 1-2 μM (calculated according to 7,8-dichloro absormance at 460 nm). For the IVTT reactions, activity was assessed under aerobic conditions at ambient temperature. The IVTT reaction was mixed with substrate 200 μM **4** in oxygen saturated 0.02 M pyrophosphate buffer, pH 8.5. The 30-μL assay ran for 6 h and was quenched by mixing with an equal amount of ethyl acetate containing 1 mM

cyclohexanone as internal standard. A sample of the organic phase was injected on an Agilent technologies 6850 GC machine and conversions were determined from a set of know standards.

4.5 References

1 Lu, S. M. & Bolm, C. Highly chemo- and enantioselective hydrogenation of linear alpha,beta-unsaturated ketones. *Chemistry* **14**, 7513-7516, (2008).

2 Martin, N. J. A. & List, B. Highly enantioselective transfer hydrogenation of alpha,beta-unsaturated ketones. *J. Am. Chem. Soc.* **128**, 13368-13369, (2006).

3 Taylor, C. J. S. N. & Jaekel, C. Enantioselective Hydrogenation of Enones with a Hydroformylation Catalyst. *Adv. Synth. Catal.* **350**, 2708-2714, (2008).

4 Tuttle, J. B., Ouellet, S. G. & MacMillan, D. W. C. Organocatalytic transfer hydrogenation of cyclic enones. *J. Am. Chem. Soc.* **128**, 12662-12663, (2006).

5 Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185-194, (2012).

6 Nestl, B. M., Nebel, B. A. & Hauer, B. Recent progress in industrial biocatalysis. *Curr. Opin. Chem. Biol.*, (2010).

7 Wohlgemuth, R. Biocatalysis--key to sustainable industrial chemistry. *Curr. Opin. Biotechnol.* **21**, 713-724, (2010).

8 Adalbjornsson, B. V. *et al.* Biocatalysis with Thermostable Enzymes: Structure and Properties of a Thermophilic 'ene'-Reductase related to Old Yellow Enzyme. *ChemBioChem* **11**, 197-207, (2010).

9 Hall, M. *et al.* Asymmetric bioreduction of activated C = C bonds using Zymomonas mobilis NCR enoate reductase and old yellow enzymes OYE 1-3 from yeasts. *Eur. J. Org. Chem.*, 1511-1516, (2008).

10 Kawai, Y., Inaba, Y. & Tokitoh, N. Asymmetric reduction of nitroalkenes with baker's yeast. *Tetrahedron-Asymmetry* **12**, 309-318, (2001).

11 Kosjek, B., Fleitz, F. J., Dormer, P. G., Kuethe, J. T. & Devine, P. N. Asymmetric bioreduction of alpha,beta-unsaturated nitriles and ketones. *Tetrahedron-Asymmetry* **19**, 1403-1406, (2008).

12 Swiderska, M. A. & Stewart, J. D. Stereoselective enone reductions by Saccharomyces carlsbergensis old yellow enzyme. *J. Mol. Catal. B: Enzym.* **42**, 52-54, (2006).

13 Toogood, H. S. *et al.* Structure-Based Insight into the Asymmetric Bioreduction of the C=C Double Bond of alpha,beta-Unsaturated Nitroalkenes by Pentaerythritol Tetranitrate Reductase. *Adv. Synth. Catal.* **350**, 2789-2803, (2008).

14 Toogood, H. S., Gardiner, J. M. & Scrutton, N. S. Biocatalytic Reductions and Chemical Versatility of the Old Yellow Enzyme Family of Flavoprotein Oxidoreductases. *Chemcatchem* **2**, 892-914, (2010).

15 Hall, M. *et al.* Asymmetric bioreduction of C=C bonds using enoate reductases OPR1, OPR3 and YqjM: Enzyme-based stereocontrol. *Adv. Synth. Catal.* **350**, 411-418, (2008).

16 Pompeu, Y. A., Sullivan, B., Walton, A. Z. & Stewart, J. D. Structural and Catalytic Characterization of Pichia stipitis OYE 2.6, a Useful Biocatalyst for Asymmetric Alkene Reductions. *Adv. Synth. Catal.* **354**, 1949-1960, (2012).

17 Schittmayer, M. *et al.* Old Yellow Enzyme-Catalyzed Dehydrogenation of Saturated Ketones. *Adv. Synth. Catal.* **353**, 268-274, (2011).

18 Stueckler, C. *et al.* Bioreduction of alpha-methylcinnamaldehyde derivatives: chemo-enzymatic asymmetric synthesis of Lilial (TM) and Helional (TM). *Dalton Trans.* **39**, 8472-8476, (2010).

19 Winkler, C. K., Tasnadi, G., Clay, D., Hall, M. & Faber, K. Asymmetric bioreduction of activated alkenes to industrially relevant optically active compounds. *J. Biotechnol.* **162**, 381-389, (2012).

20 Bougioukou, D. J., Kille, S., Taglieber, A. & Reetz, M. T. Directed Evolution of an Enantioselective Enoate-Reductase: Testing the Utility of Iterative Saturation Mutagenesis. *Adv. Synth. Catal.* **351**, 3287-3305, (2009).

21 Hall, M. & Bommarius, A. S. Enantioenriched compounds via enzyme-catalyzed redox reactions. *Chem. Rev. (Washington, DC, U. S.)* **111**, 4088-4110, (2011).

22 Hulley, M. E. *et al.* Focused Directed Evolution of Pentaerythritol Tetranitrate Reductase by Using Automated Anaerobic Kinetic Screening of Site-Saturated Libraries. *ChemBioChem* **11**, 2433-2447, (2010).

23 Padhi, S. K., Bougioukou, D. J. & Stewart, J. D. Site-saturation mutagenesis of tryptophan 116 of Saccharomyces pastorianus old yellow enzyme uncovers stereocomplementary variants. *J. Am. Chem. Soc.* **131**, 3271-3280, (2009).

24 Reich, S., Hoeffken, H. W., Rosche, B., Nestl, B. M. & Hauer, B. Crystal Structure Determination and Mutagenesis Analysis of the Ene Reductase NCR. *ChemBioChem* **13**, 2400-2407, (2012).

25 Toogood, H. S. *et al.* A site-saturated mutagenesis study of pentaerythritol tetranitrate reductase reveals that residues 181 and 184 influence ligand binding, stereochemistry and reactivity. *ChemBioChem* **12**, 738-749, (2011).

26 Brown, B. J., Hyun, J. W., Duvvuri, S., Karplus, P. A. & Massey, V. The role of glutamine 114 in old yellow enzyme. *J. Biol. Chem.* **277**, 2138-2145, (2002).

27 Fox, K. M. & Karplus, P. A. Old yellow enzyme at 2 A resolution: overall structure, ligand binding, and comparison with related flavoproteins. *Structure* **2**, 1089-1105, (1994).

28 Karplus, P. A., Fox, K. M. & Massey, V. Flavoprotein structure and mechanism. 8. Structure-function relations for old yellow enzyme. *FASEB J.* **9**, 1518-1526, (1995).

29 Massey, V. & Schopfer, L. M. Reactivity of Old Yellow Enzyme with Alpha-Nadph and Other Pyridine-Nucleotide Derivatives. *J. Biol. Chem.* **261**, 1215-1222, (1986).

30 Niino, Y. S., Chakraborty, S., Brown, B. J. & Massey, V. A New Old Yellow Enzyme of Saccharomyces-Cerevisiae. *J. Biol. Chem.* **270**, 1983-1991, (1995).

31 Vaz, A. D. N., Chakraborty, S. & Massey, V. Old Yellow Enzyme - Aromatization of Cyclic Enones and the Mechanism of a Novel Dismutation Reaction. *Biochemistry* **34**, 4246-4256, (1995).

32 Yu, Y. & Lutz, S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.* **29**, 18-25, (2011).

33 Qian, Z., Fields, C. J. & Lutz, S. Investigating the structural and functional consequences of circular permutation on lipase B from Candida antarctica. *ChemBioChem* **8**, 1989-1996, (2007).

34 Qian, Z., Horton, J. R., Cheng, X. & Lutz, S. Structural redesign of lipase B from Candida antarctica by circular permutation and incremental truncation. *J. Mol. Biol.* **393**, 191-201, (2009).

35 Qian, Z. & Lutz, S. Improving the catalytic activity of Candida antarctica lipase B by circular permutation. *J. Am. Chem. Soc.* **127**, 13466-13467, (2005).

36 Walsh, C. *et al.* Chemical and Enzymatic Properties of Riboflavin Analogs. *Biochemistry* **17**, 1942-1951, (1978).

37 Ghisla, S. & Edmondson, D. E. in *Flavins and Flavoproteins.* eds S. Ghisla, P. Kroneck, P. Macheroux, & H. Sund) 71-76 (Agency for Scientific Publications).

38 Ghisla, S. & Massey, V. New Flavins for Old - Artificial Flavins as Active-Site Probes of Flavoproteins. *Biochem. J.* **239**, 1-12, (1986).

39 Ghisla, S., Massey, V. & Yagi, K. Preparation and Some Properties of 6-Substituted Flavins as Active-Site Probes for Flavin Enzymes. *Biochemistry* **25**, 3282-3289, (1986).

40 Massey, V., Ghisla, S. & Yagi, K. 6-Azidoflavins and 6-Aminoflavins as Active-Site Probes of Flavin Enzymes. *Biochemistry* **25**, 8095-8102, (1986).

41 Massey, V., Ghisla, S. & Yagi, K. 6-Thiocyanatoflavins and 6-Mercaptoflavins as Active-Site Probes of Flavoproteins. *Biochemistry* **25**, 8103-8112, (1986).

42 Murthy, Y. V. S. N., Meah, Y. & Massey, V. Conversion of a flavoprotein reductase to a desaturase by manipulation of the flavin redox potential. *J. Am. Chem. Soc.* **121**, 5344-5345, (1999).

43 Koder, R. L., Haynes, C. A., Rodgers, M. E., Rodgers, D. W. & Miller, A. F. Flavin thermodynamics explain the oxygen insensitivity of enteric nitroreductases. *Biochemistry* **41**, 14197-14205, (2002).

44 Ueda, T. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751-755, (2001).

45 Ying, B. W. & Ueda, T. in *Protein Engineering Handbook* Vol. 2 (eds S. Lutz & U. T. Bornscheuer) 515-535 (Wiley-VCH, Weinheim, 2009).

46 Reitinger, S. *et al.* Circular permutation of Bacillus circulans xylanase: a kinetic and structural study. *Biochemistry* **49**, 2464-2474, (2010).

47 Abramovitz, A. S. & Massey, V. Interaction of Phenols with Old Yellow Enzyme - Physical Evidence for Charge-Transfer Complexes. *J. Biol. Chem.* **251**, 5327-5336, (1976).

48 Fryszkowska, A. *et al.* Asymmetric Reduction of Activated Alkenes by Pentaerythritol Tetranitrate Reductase: Specificity and Control of Stereochemical Outcome by Reaction Optimisation. *Adv. Synth. Catal.* **351**, 2976-2990, (2009).

49 Kataoka, M. *et al.* Old Yellow Enzyme from Candida macedoniensis catalyzes the stereospecific reduction of the C=C bond of ketoisophorone. *Biosci. Biotechnol. Biochem.* **66**, 2651-2657, (2002).

50 Stuermer, R., Hauer, B., Hall, M. & Faber, K. Asymmetric bioreduction of activated C=C bonds using enoate reductases from the old yellow enzyme family. *Curr. Opin. Chem. Biol.* **11**, 203-213, (2007).

51 Wada, M. *et al.* Production of a doubly chiral compound, (4R,6R)-4-hydroxy-2,2,6-trimethylcyclohexanone, by two-step enzymatic asymmetric reduction. *Appl. Environ. Microbiol.* **69**, 933-937, (2003).

52 Ay, J. *et al.* Crystal structures and properties of de novo circularly permuted 1,3-1,4-beta-glucanases. *Proteins* **30**, 155-167, (1998).

53 Chu, V., Freitag, S., Le Trong, I., Stenkamp, R. E. & Stayton, P. S. Thermodynamic and structural consequences of flexible loop deletion by circular permutation in the streptavidin-biotin system. *Protein Sci.* **7**, 848-859, (1998).

54 Manjasetty, B. A., Hennecke, J., Glockshuber, R. & Heinemann, U. Structure of circularly permuted DsbA(Q100T99): preserved global fold and local structural adjustments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 304-309, (2004).

55 Pieper, U., Hayakawa, K., Li, Z. & Herzberg, O. Circularly permuted beta-lactamase from Staphylococcus aureus PC1. *Biochemistry* **36**, 8767-8774, (1997).

56 Matthews, R. G., Massey, V. & Sweeley, C. C. Identification of p-hydroxybenzaldehyde as the ligand in the green form of old yellow enzyme. *J. Biol. Chem.* **250**, 9294-9298, (1975).

57 Kasai, S., Fritz, B. J. & Matsui, K. An Improved Synthesis of 8-Amino-8-Demethylriboflavin. *Bull. Chem. Soc. Jpn.* **60**, 3041-3042, (1987).

58 LAMBOOY, J. P., SCALA, R. A. & HALEY, E. E. The Biological Activities of 6-Methyl-7-chloro-9-(l'-D-ribityl) isoalloxazine and Dichlororiboflavin. *J. Nutr.* **74**, 466-472, (1961).

59 Murthy, Y. V. S. N. & Massey, V. Syntheses and applications of flavin analogs as active site probes for flavoproteins. *Vitamins and Coenzymes, Pt J* **280**, 436-460, (1997).

60 YONEDA, F. Syntheses of 5-Deazaflavins. *Methods Enzymol.* **66**, 267-277, (1980).

61 Murthy, Y. V. S. N. & Massey, V. Synthesis and properties of SCN-flavin nucleotide analogs and studies with flavoproteins. *J. Biol. Chem.* **273**, 8975-8982, (1998).

62 Murthy, Y. V. S. N. & massey, V. Synthesis and Applications of Flavin Analogs as Active Site Proves for Flavoproteins. *Methods Enzymol.* **280**, 436-460, (1997).

63 Massey, V. Activation of Molecular-Oxygen by Flavins and Flavoproteins. *J. Biol. Chem.* **269**, 22459-22462, (1994).

64 Fryszkowska, A. *et al.* A surprising observation that oxygen can affect the product enantiopurity of an enzyme-catalysed reaction. *FEBS J.* **279**, 4160-4171, (2012).

65 Schopfer, L. M., Massey, V. & Claiborne, A. Active-Site Probes of Flavoproteins - Determination of the Solvent Accessibility of the Flavin Position 8 for a Series of Flavoproteins. *J. Biol. Chem.* **256**, 7329-7337, (1981).

66 van den Heuvel, R. H., van den Berg, W. A., Rovida, S. & van Berkel, W. J. Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *J. Biol. Chem.* **279**, 33492-33500, (2004).

67 Brown, B. J., Deng, Z., Karplus, P. A. & Massey, V. On the active site of Old Yellow Enzyme. Role of histidine 191 and asparagine 194. *J. Biol. Chem.* **273**, 32753-32762, (1998).

Chapter 5: Structural Exploration of Representative OYE Permutants

5.1 Introduction

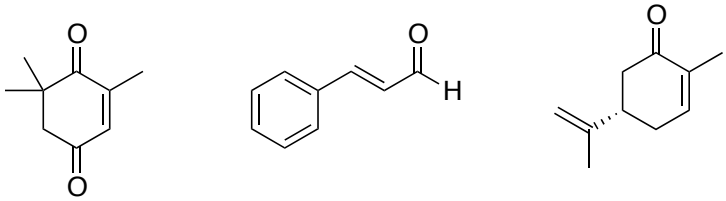
Circular permutation (CP) has been used to study protein folding and to a lesser extent as a protein engineering method to manipulate protein structure and function over the years.[1] Many lessons have been learned from the exploration of circular permutated variants, providing a relatively good understanding of the general impact of termini relocation on structure and function.[1-6] For example, general rules of linker design have been established and aid in more successful design of future CP targets.[7-10] Crystallographic studies have shed light on the consequences of CP, both tertiary and quaternary structure of proteins.[3,4,6,7,11-17] Overall, it seems that CP variants generally resemble the wild type structure with changes that are limited to locations near the old and new termini.[3,9,18] In a few cases, more substantial alterations in structure were observed resulting in significantly altered protein stability and catalytic properties. For ribosomal protein S6 from *Thermus thermophilus*, permutation in an 18- residue loop region did not affect the overall 3-D –structure, but the overall stability of the protein was increased presumably due to reduced conformation dynamics in the permutant.[16] On the other hand, lipase B from *C. antarctica* (CALB) showed changes to tertiary and quaternary structure. For the CALB variant, the termini were relocated in the center of an α-helical region that was part of the substrate-binding pocket.[9] As revealed in the crystal structure, the relocation had a large impact on active site accessibility and was presumably responsible for the large catalytic rate enhancements observed over wild type enzyme.[9] In addition, changes at the quaternary level were observed. While the wild type enzyme is monomeric, the termini relocation resulted in dimeric cpCALB variants.[9] Further investigations concluded that both the composition and length of the linker

sequence in cpCALB were the driving forces for altering the equilibrium between monomer and domain-swapped dimers.[9] These previous results have built a repertoire of potential impacts of CP that can help with the structural and functional characterization of the OYE1 permutants created in the chapter 4.

In chapter 4, we implemented a highly efficient approach for cell-free cpOYE library preparation by combining whole-gene synthesis with *in vitro* transcription/translation. The versatility of such an *ex vivo* system was further demonstrated by the rapid and reliable functional evaluation of library members under variable environmental conditions with three reference substrates: ketoisophorone, cinnamaldehyde and (*S*)-carvone (Figure 5.1). Library analysis identified over 70 functional OYE1 variants, with several biocatalysts exhibiting over an order of magnitude improved catalytic activity. Although catalytic gains of individual cpOYE library members vary by substrate, locations of new protein termini in functional variants for all tested substrates fell within the same four distinct loop/lid regions near the active site (Figure 5.2).

In this chapter, I have focused on investigating the impact of termini relocation in these four specific sectors with respect to structure, stability, and enantioselectivity. In order to understand the observed gains in catalytic activity and changes in substrate preferences between the different regional variants, a more thorough understanding of the structural impacts of OYE1 permutation is necessary. Initial CD spectroscopy and gel filtration studies reveal identical secondary structures but altered oligomeric states and enhanced thermostability. Subsequent x-ray crystallography of the top-performing variant, cpOYE303, indicates similar tertiary structure except at the critical loop region

near the new termini. Comparison with wild type OYE1 shows a considerably more accessible and enlarged active site cavity. Overall, the combined findings from chapter 4 and 5 support our hypothesis of conformational flexibility as a limiting factor for catalysis in wild type OYE.



ketoisophorone cinnamaldehdye (*S*)-carvone

Figure 5.1. Substrates characterized in chapter 4.

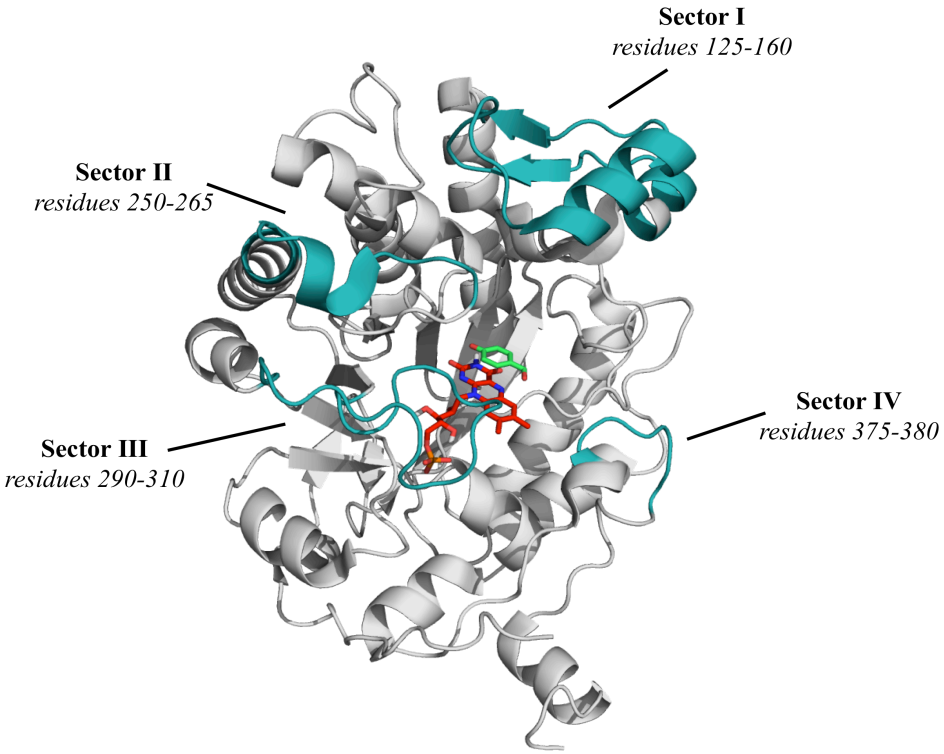


Figure 5.2. Regions that resulted in improved catalytic activity. OYE (PDB: 1OYB19) with the four sectors highlighted in teal which resulted in catalytic improvements.

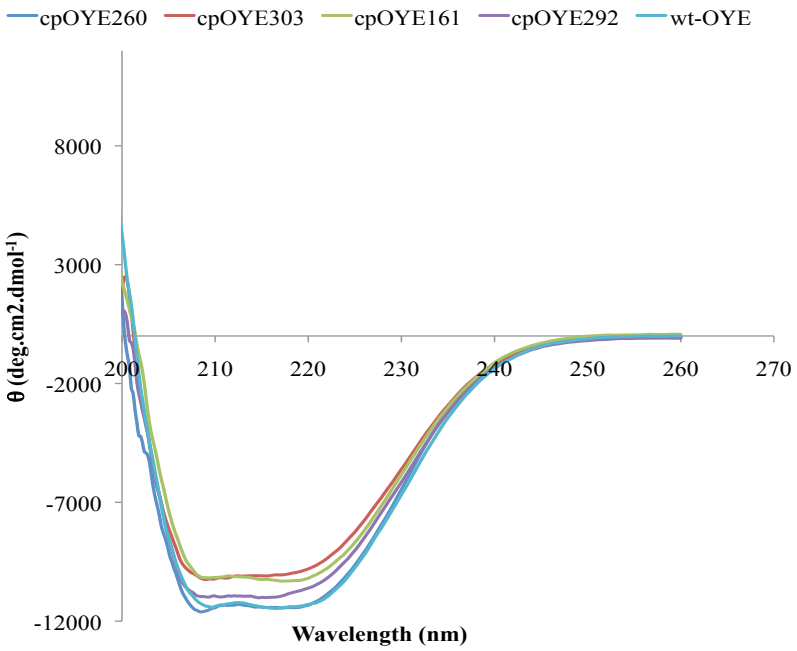
5.2 Results and discussion

5.2.1 Investigation of secondary structure and thermal denaturation of permutants

Circular dichroism (CD) spectroscopy was used to assess the secondary structural features of representative permutants from each of the four sectors and wild type OYE. We hypothesized that the relocation of the termini in these regions would not have a significant impact on secondary structure as most changes are localized in the region of old and new termini. The far-UV CD spectra of ten selected permutants were obtained and compared to wild type OYE (Figure 5.3A). All spectra have a similar overall shape characteristic of mixed α/β structures. These results suggest that the permutation has little effect on the overall structure of the protein. This is in agreement with the observations from structural studies on other circular permuted proteins in the literature.[11-13,17,20]

Thermostability is a very important property of biocatalysts as many industrial processes involve reaction conditions including elevated temperatures.[21] Higher temperatures can improve solubility, reduce product inhibition, reduce competing reactions as well as lower unwanted thermodynamic product yields.[21] The stability of the OYE1 variants was explored by following the change in α- helical content at 222 nm from 4° to 80°C (Figure 5.3B). The thermodenaturation measurements indicate that permutants in sector III maintain a similar T_m compared to wild type OYE (~ 49°C) and in some cases the permutation caused up to an 8 °C increase in protein stability. Permutant cpOYE260 from sector II displayed a < 3°C decline in stability while sector I variants showed the most dramatic increase in T_m by 16°C.

A.



B.

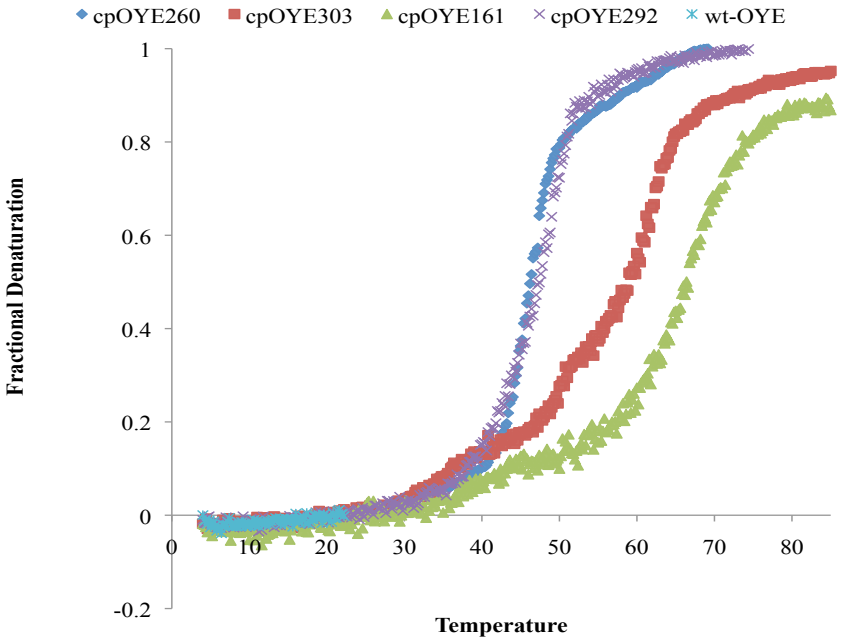


Figure 5.3. Circular dichroism analysis. A.) Far-UV circular dichroism spectra of representative permutants. Spectra were recorded on protein samples (10-15 μM) in potassium phosphate buffer (pH 8). B.) Thermal denaturation curves of permutants from 4° to 80°C.

Based on CD data no obvious correlation can be made between the sectors in which the termini are relocated and the observed changes observed in the thermodenaturation experiments. It is clear that in this particular case the activity enhancements observed are not correlated with the difference is stability. Presumably, the increases in thermostability could be attributed to the differences in entropic contributions from the translocation of new N- and C- termini. The observations with OYE1 are unlike many other studies where circular permutation compromised the thermostability compared to wild type.[6,8,9,11,13-15,17] With OYE, the location of the new termini did not destabilize the enzyme whereas in the case of CALB it was demonstrated that the destabilization was from the linker length connecting the old termini.[17] Experiments on CALB were conducted to reduce linker length and in turn it translated into an increase in stability of the variants. This discrepancy can be explained by favorable entropy of loop closure, from connecting the old OYE termini, and/or the creation of the new termini in these sectors.

It is interesting to note that one of the major differences between the thermostable-like and classical OYE classes is their loop size, sequence and orientation (Detailed in Chapter 1).[21] The thermostable-like class has more open accessible active sites as a result of shorter loops.[21] Loop region number 3 is the main loop that differs between the two classes and is also the region encompassing our most active permutants in OYE1 (cpOYE library region III-residues 288-309). In this region, thermostability increases are observed by up to 8 °C for cpOYE303. This overall stability increase suggests that breakage of the peptide bond alters the surface loop and in turn increases stability. It could be that the altered loop creates a more thermostable-like OYE

confirmation. Recently, Reich and coworkers targeted this same loop region via rationally designing two variants of NCR to have shorter loops in region 3 and 4.[22] The shorter loop 3 variant showed both a slight increase in activity-based thermostability as well as increased tolerance to organic solvents.[22] They rationalized that the shorter loop increased the rigidity towards a more thermostable-like OYE loop.[22] The structural investigation of cpOYE303 suggests that it is not necessarily increased rigidity but altered conformational dynamics contributing to the observed enhancements in thermostability.

5.2.2 Oligomeric state as a result of permutation

Sequence reorganization from circular permutation has been shown to cause significant changes in the quaternary structure of some proteins.[23,24] The oligomeric state within the OYE family varies and different members exist as a monomer (pentaerythritol reductase- PETNR), dimer (OYE1), or even a tetramer (YqjM).[21] In order to test if relocation of termini for OYE changed the quaternary structure the oligomeric state of selected permutants was determined using size exclusion chromatography (Table 5.1). The chromatographs were compared to a set of know protein standards in order to assign the oligomeric state. The results indicated that most of the permutants exist in mixed populations, forming either monomeric or dimeric structures. One noteworthy exception is cpOYE260, in which only the monomeric species could be observed. Furthermore, variants in sector I formed not only monomers and dimmers but also assembled into higher order structures (trimer or larger). The oligomeric state and thermodenaturation measurements were compared to see if there was a correlation between these properties.

Oligomerization has previously shown to be a mechanism of thermostabilization.[25,26] The dramatic increases in the stability of permutants from region I could be a result of the higher order oligomeric state as this is where the trimer population was observed. Further studies are needed to explore the stability of isolated individual oligomeric forms, as well as test for possible correlation with activity. Such experiments would reveal if the trimeric forms were associated with increases in thermostability.

Table 5.1. Summary of selected permantant's Tm and Oliogomeric state

Permutant	**T_m (°C)**	**Oligomeric State**
wt-OYE	49.4	dimer
cpOYE307	55.3	monomer/dimer
cpOYE306	50.5	monomer/dimer
cpOYE305	56.6	monomer/dimer
cpOYE303	57.1	monomer/dimer
cpOYE292	46.8	dimer
cpOYE291	43.2	dimer
cpOYE260	46.5	monomer
cpOYE161	65.9	monomer/dimer/trimer
cpOYE160	66	monomer/dimer/trimer
cpOYE146	64	monomer/dimer/trimer

5.2.3 Crystallographic analysis of cpOYE variants

X-ray crystallography was employed to study the structural changes of our top OYE permutant cpOYE303. Initial crystallization trials were set up to mimic the same conditions under which wild type OYE1 protein was crystallized.[19] Attempts to obtain crystals under the original conditions (0.1 M HEPES pH 8.3, 35% PEG400, 0.2 M $MgCl_2$) were unsuccessful. Screening was completed around these conditions by altering buffer, PEG400 composition, as well as protein concentration, but did not result in any improvements. Moving forward, five crystallization screens were set up to find the optimal conditions for cpOYE303. All screens were assembled using three different protein concentrations (10, 20, and 40 mg/mL) giving a total of 1440 conditions. Crystals appeared from ten conditions after 2-3 days at 16°C (Table 5.2). The conditions produced two different types of crystals: a more needle-like and a chuncky crystal (Figure 5.2). It wasn't clear what caused the two different morphologies but a commonality between most of the conditions was precipitant PEG 3350 (20-25%) and 0.2 M $MgCl_2$. To obtain single crystals with better diffraction several optimization screens centered around these conditions were performed.

Table 5.2. Crystallization screening conditions yield crystals for cpOYE303

Kit (Hampton)	**Buffer (pH)**	**Salt**	**Precipitant**
B-Main	-	0.2 M magnesium acetate	20% PEG 3350
	-	0.2 M magnesium formate	20% PEG 3350
	-	0.2 M KCl	20% PEG 3350
	0.1 M Bis-Tris (6.5)	0.2 M $MgCl_2$	25% PEG 3350
	0.1 M Hepes (7.5)	0.2 M $MgCl_2$	25% PEG 3350
	0.1 M Tris-HCl (8.5)	0.2 M $MgCl_2$	25% PEG 3350
	0.1 M Tris-HCl (8.5)	0.2 M lithium sulfate	30% PEG 4000
	0.1 M MES (6.5)	0.2 M ammonium sulfate	30% PEG 5kMME
Crystal Screen	0.1 M Tris-HCl (8.5)	0.2 M lithium sulfate	30% PEG 4000
Index	0.1 M Tris-HCl (8.5)	0.2 M ammonium acetate	45% MPD

A. B. C.

Figure 5.4. Representative crystals from cpOYE303 crystallization screensA) 0.1M Hepes (pH 7.5), 0.2 M MgCl2, 25% PEG 3350 B) 0.2M Magnesium Formate, PEG 3350 (20%) C) 0.1M Hepes (pH 7.5), 0.2 M MgCl2, 18% PEG 3350, 0.25% glucoside.

The first round of screens all used the small sitting drop crystallization method. In the subsequent trials both larger sitting drop and hanging drop techniques were tested for comparison. These two vapor diffusion methods differ in the vertical orientation of the

protein drop within the overall system and can result in different crystallization behaviors depending on the protein.[27] Other parameters that were varied throughout this process were: pH, buffering agent, protein concentration, PEG 3350 %, and $MgCl_2$ concentration. A small range of conditions was determined for the 'best' and most reproducible crystals (Table 5.3). Although conditions for optimal crystal growth were not found, parameters were identified that generated crystals that could be broken into single pieces (Figure 5.4).

In addition to crystallization trials with cpOYE303 in the presence of only FMN, we also attempted the co-crystallization with *p*-hydroxybenzaldehyde (pHBA), a tight binding competitor inhibitor of OYE1.[28] The same conditions were used to generate a large number of crystals for soaking experiments, as co-crystallizaiton trials did not result in crystal growth. The crystals were soaked in a final solution of 11 mM pHBA for 10 min prior to freezing. There was a slight color change from yellow to greenish-brown upon uptake of the inhibitor due to the formation of the charge transfer band.[28] All crystals were soaked in mother liquor containing ethylene glycol (20%) as cryo protectant and frozen in liquid nitrogen. A high quality x-ray data set was collected at the synchrotron (Argonne National Laboratory) for cpOYE303 with and without inhibitor.

Table 5.3. Optimized crystallization conditions via the hanging drop method for cpOYE303. Protein concentration of 40 mg/mL produced the best crystals. The conditions in which the final dataset was collected from are shown in bold.

Conditions	PEG 3350 %	$MgCl_2$	Glucoside	HEPES pH 7.5
A	16	0.2 M	--	0.1 M
B	16	0.2 M	0.25 %	0.1 M
C	18	0.2 M	--	0.1 M
D	**18**	**0.2 M**	**0.25 %**	**0.1 M**

5.2.4 cpOYE303-full length crystal structure

The crystals for cpOYE303 with and without the pHBA inhibitor had very low symmetry and belonged to space group P1 (Table 5.4). The structures were solved by molecular replacement using OYE1 (PDB: 1OYA) as the search model. As predicted, the structure correctly contains the new termini in loop region 4 residing near the active site cavity. When the inhibitor is not present in the active site, a sulfate or phosphate group occupies the space over the flavin. This is consistent with density observed in other OYE structures.[19,21,29] Electron density was missing for several residues near the new termini. The missing density included the first 3 amino acids at the N-terminus and 10 amino acids at the C-terminus and is due to the high flexibility in this region. Given their conformational flexibility, we postulated that truncation of these residues would effect on the stability and catalytic activity of our CP variant.

Table 5.4. X-ray crystallographic parameters for cpOYE303 structures

Parameters	**wt-OYE (PDB:1OYA[19])**	**cpOYE303 no I**	**cpOYE303 + pHB**	**cpOYE303-truncated no I**
Space Group	P 43 21 2	P1	P1	P 1 21 1
Resolution (Å)	2.00	2.68	2.47	1.55
Unit Cell a,b,c (Å)	142.88, 142.88, 43.01	45.66, 87.64, 113.57	45.59, 88.08, 105.62	64.25, 73.00, 84.74
α,β,γ (°)	90.00, 90.00, 90.00	69.32, 82.56, 90.02	89.45, 81.43, 89.89	90.00, 101.27, 90.00

5.2.5 Creation of truncation mutant and activity analysis

A truncation mutant was constructed by deleting all of the invisible residues at the new N and C- termini, respectively. I was interested in comparing the catalytic activity of the truncation mutant with the full-length protein in order to see if these unresolved regions play a large role in catalysis. If the impact is minimal, then the conclusions made from the original crystal structures should be valid. The truncation of the protein termini in cpOYE303 did not affect protein expression or purification, yielding product of similar stability although at slightly lower yields.

Table 5.5. Activity of cpOYE303 and cpOYE303-truncated variants with KIP

Variant	**KIP Conversion %**	**ee %**
cpOYE303	16.7 (30 min)	>99
cpOYE303-truncated	18.1 (30 min)	>99
cpOYE303	34.7 (1 h)	>99
cpOYE303-truncated	35.9 (1h)	>99

The rate of conversion and enantioselectivity for the reduction of ketoisophorone were the same as for the parental protein, supporting the hypothesis that this region is not important for catalysis (Table 5.5). The turnover numbers detected by the NADPH spectrophotometric assay were also very similar with cpOYE303 turnover number at 71.1 min^{-1} and cpOYE303-truncated displaying a slightly lower turnover of 60.3 min^{-1}. These results suggest that the largely flexible termini do not significantly contribute to catalysis.

5.2.6 Crystallization of truncated cpOYE303

The previous crystallization experiments of full-length cpOYE303 did not produce single crystals and the final data set collected only diffracted to 2.6 Å. By eliminating the flexible regions in the truncation mutant, it was hypothesized that the protein may be able to crystallize better. This could translate into higher quality crystals displaying better resolution. Initially, the same crystallization conditions were tried as in Table 5.3 but with the truncated variant. These conditions did not produce crystals. Previously successful crystallization screens B-Main and index (Hampton Research) were screened using three different protein concentrations (10, 20, and 40 mg/mL) giving a total of 576 conditions. Crystals appeared from ten conditions after 2-3 days at 16 °C (Table 5.6). The conditions produced crystals with different morphology compared to the full-length protein under its respective conditions. The single crystals appeared to be of two types possessing different space groups (Figure 5.3). The conditions shown in Table 5.6 (bold) produced the highest quality crystals belonging to space group P 1 21 1 with a final resolution of 1.55 Å. Again, all crystals were soaked in mother liquor containing

ethylene glycol (20%) as cryo protectant and frozen in liquid nitrogen before data collection.

Table 5.6. Crystallization screening conditions yield crystals for cpOYE303-truncated. Protein concentration of 40 mg/mL produced the best crystals. The conditions in which the final dataset was collected from are shown in bold.

Kit	Buffer (pH)	Salt	Precipitant
Index	0.1 M sodium acetate trihydrate (4.5)	-	2.0 M ammonium sulfate
	0.1 M BIS-Tris (5.5)	-	2.0 M ammonium sulfate
	0.1 M BIS-Tris (6.5)	-	2.0 M ammonium sulfate
	0.1 M HEPES (7.5)	-	2.0 M ammonium sulfate
	0.1 M Tris-HCl (8.5)	-	2.0 M ammonium sulfate
B-Main	**0.1 M HEPES (7.5)**	**-**	**20% PEG 10000**

A. B.

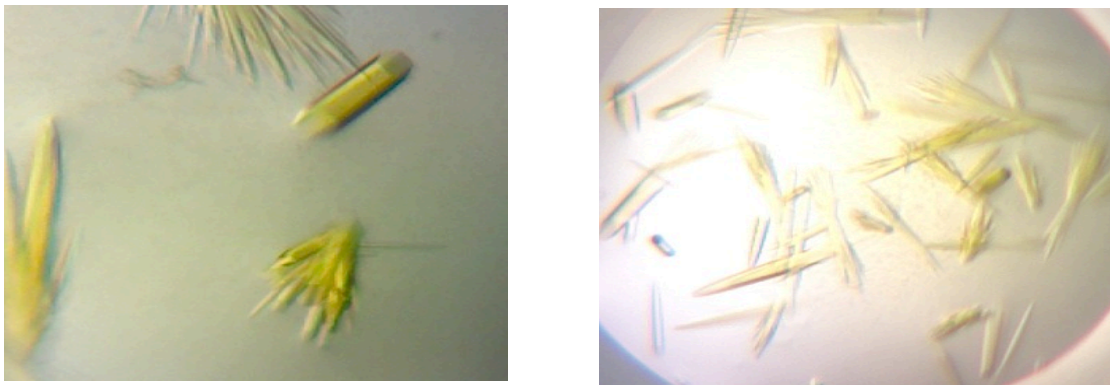


Figure 5.5. Representative OYE crystals from cpOYE303-truncated crystallization screens.

5.2.7 cpOYE303-truncated crystal structure

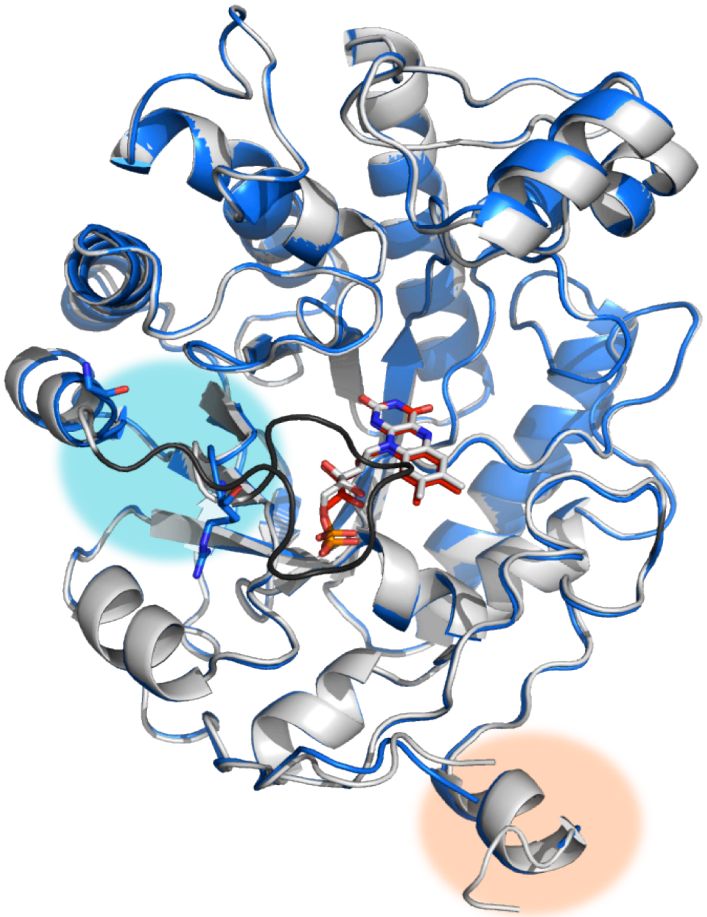
The crystals for the truncated variant diffracted to very high resolution and belonged to space group P1 21 1. The space group was different from both the full-length cpOYE303 protein and wt-OYE[19] (Table 5.6). In addition, the truncated variant was crystallized without the pHB inhibitor since we already had one cpOYE303 full-length structure containing pHB. A HEPES molecule from the crystallization buffer was found above the FMN in the substrate-binding site. Eliminating the residues that were not visible in the full-length protein had no effect on the overall structure. The density for all of the residues at the new truncated N and C- termini was visible. Due to the much higher resolution of this structure and the fact that there are no observed changes between its parental protein, it was used in the subsequent structural analysis.

5.2.8 Comparison of cpOYE303 crystal structures with wt-OYE

Structural analysis of all structures was completed to conclude the affects of CP on OYE. The structural alignment of wt-OYE and cpOYE303 shows that the permutant has an identical core structure as the wild type enzyme. The protein maintains the eight-stranded $(\beta\alpha)_8$ barrel structure with a non-covalently bound FMN cofactor residing on top of the β barrel (Figure 5.6A).[19] The main differences are seen at the old termini, which are connected by a three-residue linker and loop region 3 where the peptide bond was broken (between residue G302 and G303) (Figure 5.6A) Otherwise the two structures are superimposable. These localized differences are consistent with the observations from other circular permuted structures in the literature.[1,3,9,11,16] Circular permutation significantly increased the flexibility near the new termini, which is

responsible for the missing density in this loop region that forms part of the substrate binding site (Figure 5.6B).

A.



B.

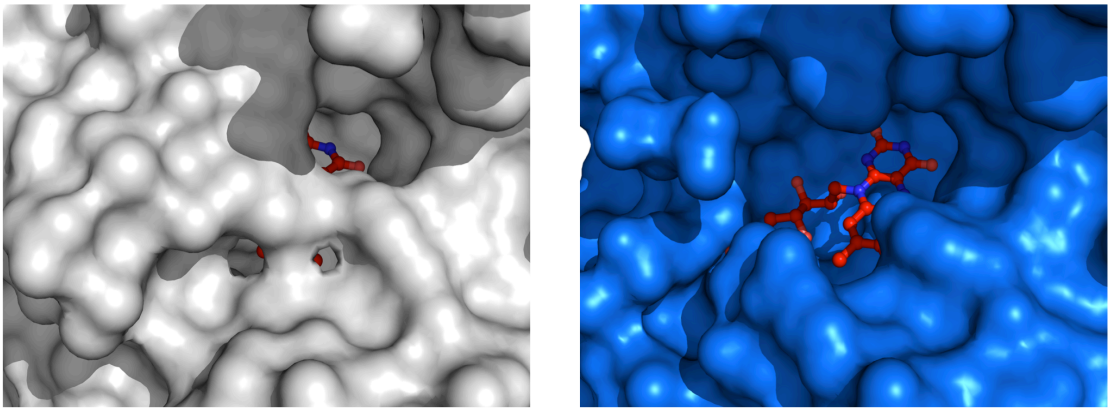


Figure 5.6. Comparison of OYE and cpOYE303 structures. A.) Structural overlay of OYE1 (grey) with FMN bound (1OYA) and cpOYE303 (blue) with FMN bound. The old termini are highlighted in the orange region and the new termini with the blue region. B) Zoom in of active the site cavity for OYE1 and cpOYE303.

The black colored loop region (residues 293-305) in the OYE1 structure (Figure 5.6A) is not present in cpOYE303 and allows for greater accessibility of the substrate to the active sites (Figure 5.6B). Our working hypothesis, is that these features might contribute to the >10-fold rate increases observed for KIP and *S*-carvone in chapter 4. Furthermore, the truncation of these missing residues revealed no change in the catalytic activity with **1**.

B-factor analysis was used to look at regions of altered flexibility within the cpOYE303 structure (Figure 5.7). B-factor analysis is often used as a semi-quantitative measure of the mobility of individual atoms. It can provide useful information on the structural flexibility and dynamics of a protein.[30] However, the analysis can only be used to compare B-factor differences within the same structure and not between other structures. The surface display of the active site cavity is much more open in cpOY303 compared to wild type and the FMN cofactor is clearly more exposed. This again is from the increased local chain flexibility from the termini relocation so close to the active site pocket. The flexibility in the area of the new termini is represented through higher B-factors in this region, which is also observed near the old termini. Furthermore, cpOYE303 displays lower temperatures in the exterior subdomain (residues 125-160, based on OYE1) as well as a nearby α-helical region (residues 273-282). The flexibility in these areas seems to have also been impacted by circular permutation.

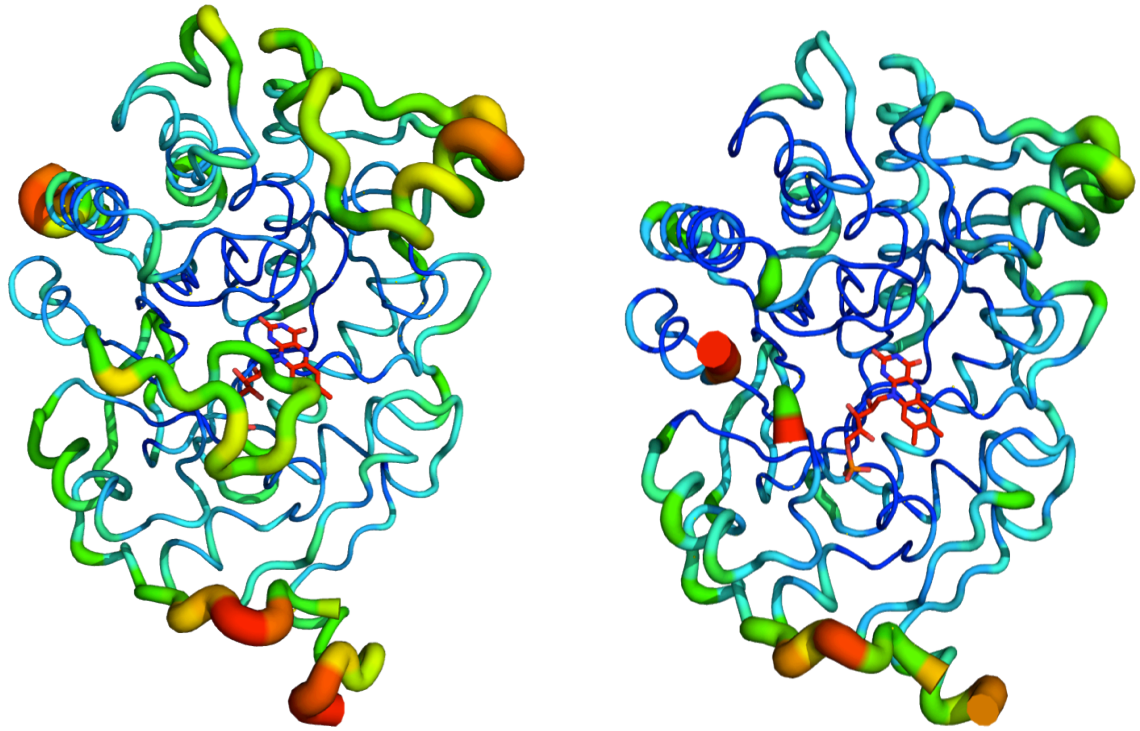


Figure 5.7. B-factor comparison. Surface representation of left- wt-OYE (1OYA19) and right- cpOYE303 (unpublished results) colored according to B-factors. Regions are colored using a blue, green, yellow, orange, red spectrum with blue and red regions representing the areas of lowest and highest temperature factors respectively.

A closer look investigation of the active site of cpOYE revealed an almost identical alignment with the residues from the original OYE1 structure with bound pHBA[19] (Figure 5.8). The active site was also compared to OYE1 without inhibitor, but the ligand bound structure provided a closer match. This is due to the cpOYE303 structure containing a HEPES molecule bound in the substrate binding site; therefore, the key residues that typically undergo repositioning when the inhibitor is bound also adjusts for HEPES binding. The three catalytic residues His191, Asn194 and Tyr196 show no

obvious differences between the active sites. On the other hand, the three residues that undergo conformational adjustments upon pHBA binding in OYE1, Try375, Phe296, and Phe250, move in a similar manner for HEPES binding in cpOYE303. Residues Try375 and Phe250 are slightly more tilted upon HEPES binding and Phe296 does not exist in this structure. The residues that exhibit the most change form the substrate-binding pocket above the *si*-face of the flavin in OYE1.[19] Interestingly, the loop that is missing in all of the cpOYE303 structures contains Phe296 and this region showed the most significant changes in the crystal structures of OYEs. This analysis supports our original hypothesis that conformational flexibility is key for the enhanced catalysis exhibited by cpOYE303.[19]

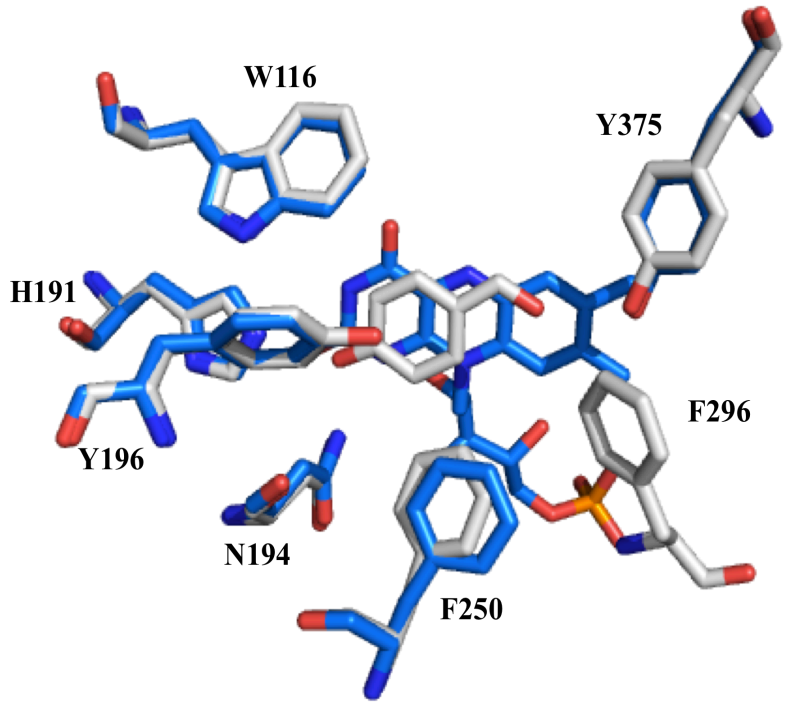


Figure 5.8. Superpositon of the active site residues of OYE1 (grey) with pHB bound (1OYB19) and cpOYE303 (blue).

5.3 Conclusions

In summary, the structural characterization of cpOYE variants revealed identical secondary structure but slightly altered tertiary and quaternary structure as a result of termini relocation. Surprisingly, permutation did not lower the stability of the proteins and in most cases increased the T_m by up to 16 °C. Crystallization of the top performer cpOYE303 clearly provides explanations for the significantly altered catalytic properties reported in chapter 4. The structure has a much larger active site cavity with greater accessibility compared to wild type OYE. The structural characterization promotes a better understanding of how permutation in a highly flexible loop region near the entrance of the active site cavity can significantly influence the reduction of α,β-unsaturated alkenes. In addition, the crystal structure provides a valuable data set for guiding future OYE engineering studies. Overall, the CP engineering approach was very successful in generating improved ene-reductase biocatalysts. It introduced beneficial variability in critical regions that play an important role in catalysis and maintained the critical interactions needed for protein stability.

5.4 Materials and Methods

5.4.1 Circular Dichroism Spectroscopy

All CD measurements were completed on representative permutants using a Jasco J-810 spectropolarimeter equipped with a Peltier unit for temperature control. All samples were prepared in 50 mM potassium phosphate buffer (pH 7.5) with protein concentrations ranging from 10 to 25 μM as determined by the absorbance at 280 nm. The spectra were collected in the far-UV range using a cuvette with a 1 mm path length

at 25 °C with a scan rate of 20 nm/min, a response time of 2 s, a bandwidth of 2 nm and an average of three accumulation scans. For thermal denaturation experiments, the change in ellipticity at 222 nm was monitored from 4 to 85 °C with a temperature gradient of 1.0 °C/min and data pitch of 0.2 nm.

5.4.2 Determination of Oligomeric State

Oligomeric state for representative permutants was determined using the Superdex 200, 10/300 GL column as well as the Bio SEC-3 HPLC (Agilent, Santa Clara, CA) column for confirmation. The protein samples were prepared at 50 μM in 40 mM Tris-HCl (pH 8.0), 300 mM NaCl buffer and all traces were compared to protein standards of known molecular weight.

5.4.3 Protein Expression and Purification for Crystallization

The same procedures as described in the Materials and Methods in chapter 4, sections 4. were performed. The purity of all protein used for crystallization was ≥ 95% pure which was established via SDS-PAGE analysis.

5.4.4 cpOYE303 Crystallization

After purification protein samples were exchanged into the protein crystallization buffer (10 mM Tris (pH8.0), 10 mM NaCl, 10 uM PMSF) and concentrated to ~ 40 mg/mL by ultrafiltration. Initial crystallization trials were completed following the conditions reported for wild type OYE. After several unsuccessful attempts,

crystallization screens were conducted first and conditions were optimized from the hits. For crystallizing both cpOYE303 wild type and truncated, the following crystallization kits were used: B-main, Index, Peg-ion, Crystal Screen B, and Synergy. All trials were set up in 96-well plates per the sitting drop method at 16°C. In the optimization trials the crystallization method, buffers, salt, precipitant and protein concentration were varied (see Results and Discussion for details). In most cases crystals appeared within 48-72 hours from assembling. Prior to data collection all crystals were soaked in mother liquor containing 20% ethylene glycol and then frozen in liquid nitrogen. For cpOYE303 full length with inhibitor, crystals were soaked in 11 mM parahydroxybenzaldehyde for 10 min and then cryoprotected as described.

5.4.5 Crystal data collection

All crystal data sets were collected by Dr. John Horton from Dr. Xiaodong Cheng's lab. The crystals for full length cpOYE303 and truncated belong to a different space group compared to each other and those of wild type OYE reported. The final structures were solved using molecular replacement based on the original crystal structure of OYE (PDB: 1OYA/B). All structures will be deposited in the Protein Data Bank once completed.

5.4.6 Construction of cpOYEG303 truncation

The full-length cpOYE303 gene sequence was used to create the truncated sequence. Primers (Tables 5.7) were used to remove the codons encoding for 4 amino acids at the 5' end and 10 amino acids at the 3'end. Following amplification, the insert

and pET-14b vector were double digested using restriction sites *NcoI* and *XhoI*. The gene was ligated into the pET-14b vector using T4 ligase with exclusion of the His-tag. Following transformation into *E. coli* DH5α, plasmid DNA from the selected colonies was extracted and sequences were confirmed.

5.4.7 Construction of cpOYE-W116I mutants

The W116I amino acid substitutions were introduced into OYE, cpOYE154, cpOYE260, cpOYE303 templates by primer overlap extension mutagenesis. The primers used for the reactions are listed in Table 5.7. Standard T7 promoter and terminator primers were used as the flanking primers for OYE and cpOYE303. Gene specific primers were used for cpOYE154 and cpOYE260 due to a sequence conflict with the restriction enzyme. PCR products were cloned into pET-14b as described above and transformed *E. coli* DH5α. Gene sequences were confirmed by DNA sequencing.

Table 5.7. Primers used for mutagenesis in Chapter 5

Primers:
303-truncation-for: CGCCCATGGGAGGTAGCAACGATT 303-truncation-rev: CGCCTCGAGTTAACGAGGTTCAACCAA Forward Primer cp154: GCGCCCATGGCTAAGGCCAACAACCCACAAC Reverse Primer cp154: GCGCTCGAGTTACTTGGCCTTAGCTTC Forward Primer cp260: GCGCCCATGGCTACCGGCATTGTTGCCC Reverse Primer cp260: GCGCTCGAGTTACTCGGCACCACCAGA W116I for: GGGTTCAGTTAATAGTTTTGGGTTG W116I rev: CCCAAGTCAATTATCAAAACCCAAC

5.4.8 Kinetic analysis of variants

The enzyme concentration for each cpOYE variant was determined based on the extinction coefficient of protein-bound FMN. Conversion for cpOYE303 full length and truncated enzymes were determined by incubating 0.1 μM enzyme with 200 μM KIP, 200 μM $NADP^+$, 5 U GDH, 100 mM glucose in 200-μL total reaction volume in an anaerobic chamber. The reaction was monitored over 1 hour by taking aliquots, mixing them with an equal amount of ethyl acetate, and analyzing via chiral GC. The product was quantified using a standard curve generated using known amounts of the product (*R*)-levodione and 1 mM cyclohexanone as internal standard. Alternatively, turnover was

monitored spectrophotometrically at ambient temperatures with a Bioteck EPOCH plate reader housed in the anaerobic chamber. Purified enzyme was assayed following the decrease in absorbance of NADPH at 340 nm (molar extinction coefficient 6,220 $M^{-1}cm^{-1}$). The 300-µL reaction contained 100 mM sodium phosphate buffer (pH 7.0), 1 mM of KIP (from 20 mM stock in ethanol), 100 µM NADPH and 1 µg enzyme. The conversion and stereoselectivity displayed by the cpOYEW116I permutants for reducing (*S*)-carvone was determined by incubating 25 µg enzyme with 5 mM substrate, 200 µM $NADP^+$, 5 U GDH, and 100 mM glucose in 500-µL total reaction volume under anaerobic conditions. The reaction was carried out for 24 h at room temperature. After quenching, the reaction was extracted with an equal amount of ethyl acetate and the organic layer was analyzed by chiral GC.

5.5 References

1 Yu, Y. & Lutz, S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.* **29**, 18-25, (2011).

2 Qian, Z. & Lutz, S. Improving the catalytic activity of Candida antarctica lipase B by circular permutation. *J. Am. Chem. Soc.* **127**, 13466-13467, (2005).

3 Reitinger, S. *et al.* Circular Permutation of Bacillus circulans Xylanase: A Kinetic and Structural Study. *Biochemistry* **49**, 2464-2474, (2010).

4 Tougard, P., Bizebard, T., Ritco-Vonsovici, M., Minard, P. & Desmadril, M. Structure of a circularly permuted phosphoglycerate kinase. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 2018-2023, (2002).

5 Yu, Y. & Lutz, S. Improved triglyceride transesterification by circular permuted Candida antarctica lipase B. *Biotechnol. Bioeng.* **105**, 44-50, (2010).

6 Zhang, T., Bertelsen, E., Benvegnu, D. & Alber, T. Circular permutation of T4 lysozyme. *Biochemistry* **32**, 12311-12318, (1993).

7 George, R. A. & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.* **15**, 871-879, (2002).

8 Nagi, A. D. & Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des.* **2**, 67-75, (1997).

9 Qian, Z., Horton, J. R., Cheng, X. D. & Lutz, S. Structural Redesign of Lipase B from Candida antarctica by Circular Permutation and Incremental Truncation. *J. Mol. Biol.* **393**, 191-201, (2009).

10 Wriggers, W., Chakravarty, S. & Jennings, P. A. Control of protein functional dynamics by peptide linkers. *Biopolymers* **80**, 736-746, (2005).

11 Fishburn, A. L., Keeffe, J. R., Lissounov, A. V., Peyton, D. H. & Anthony-Cahill, S. J. A circularly permuted myoglobin possesses a folded structure and ligand binding similar to those of the wild-type protein but with a reduced thermodynamic stability. *Biochemistry* **41**, 13318-13327, (2002).

12 Hahn, M., Piotukh, K., Borriss, R. & Heinemann, U. Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10417-10421, (1994).

13 Manjasetty, B. A., Hennecke, J., Glockshuber, R. & Heinemann, U. Structure of circularly permuted DsbA(Q100T99): preserved global fold and local structural adjustments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 304-309, (2004).

14 Mullins, L. S., Wesseling, K., Kuo, J. M., Garrett, J. B. & Raushel, F. M. Transposition of Protein Sequences - Circular Permutation of Ribonuclease-T1. *J. Am. Chem. Soc.* **116**, 5529-5533, (1994).

15 Ni, X. & Schachman, H. K. In vivo assembly of aspartate transcarbamoylase from fragmented and circularly permuted catalytic polypeptide chains. *Protein Sci.* **10**, 519-527, (2001).

16 Ohman, A., Oman, T. & Oliveberg, M. Solution structures and backbone dynamics of the ribosomal protein S6 and its permutant P54-55. *Protein Sci.* **19**, 183-189, (2010).

17 Qian, Z., Fields, C. J. & Lutz, S. Investigating the structural and functional consequences of circular permutation on lipase B from Candida antarctica. *ChemBioChem* **8**, 1989-1996, (2007).

18 Iwakura, M. & Nakamura, T. Effects of the length of a glycine linker connecting the N-and C-termini of a circularly permuted dihydrofolate reductase. *Protein Eng.* **11**, 707-713, (1998).

19 Fox, K. M. & Karplus, P. A. Old Yellow Enzyme at 2-Angstrom Resolution - Overall Structure, Ligand-Binding, and Comparison with Related Flavoproteins. *Structure* **2**, 1089-1105, (1994).

20 Guntas, G., Kanwar, M. & Ostermeier, M. Circular Permutation in the Omega-Loop of TEM-1 beta-Lactamase Results in Improved Activity and Altered Substrate Specificity. *PLoS One* **7**, (2012).

21 Adalbjornsson, B. V. *et al.* Biocatalysis with Thermostable Enzymes: Structure and Properties of a Thermophilic 'ene'-Reductase related to Old Yellow Enzyme. *ChemBioChem* **11**, 197-207, (2010).

22 Reich, S., Kress, N., Nestl, B. M. & Hauer, B. Variations in the stability of NCR ene reductase by rational enzyme loop modulation. *J. Struct. Biol.*, (2013).

23 Barrientos, L. G., Louis, J. M., Ratner, D. M., Seeberger, P. H. & Gronenborn, A. M. Solution structure of a circular-permuted variant of the potent HIV-inactivating protein cyanovirin-N: structural basis for protein stability and oligosaccharide interaction. *J. Mol. Biol.* **325**, 211-223, (2003).

24 Wright, G., Basak, A. K., Wieligmann, K., Mayr, E. M. & Slingsby, C. Circular permutation of betaB2-crystallin changes the hierarchy of domain assembly. *Protein Sci.* **7**, 1280-1285, (1998).

25 Eijsink, V. G. H., Gaseidnes, S., Borchert, T. V. & van den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **22**, 21-30, (2005).

26 Tanaka, Y. *et al.* How oligomerization contributes to the thermostability of an Archaeon protein - Protein L-isoaspartyl-O-methyltransferase from Sulfolobus tokodaii. *J. Biol. Chem.* **279**, 32957-32967, (2004).

27 McRee, D. E. *Practical Protein Crystallography*. 1-23 (Academic Press, 1993).

28 Abramovitz, A. S. & Massey, V. Interaction of phenols with old yellow enzyme. Physical evidence for charge-transfer complexes. *J. Biol. Chem.* **251**, 5327-5336, (1976).

29 Opperman, D. J. *et al.* Crystal structure of a thermostable old yellow enzyme from Thermus scotoductus SA-01. *Biochem. Biophys. Res. Commun.* **393**, 426-431, (2010).

30 Yuan, Z., Bailey, T. L. & Teasdale, R. D. Prediction of protein B-factor profiles. *Proteins* **58**, 905-912, (2005).

Chapter 6: Conclusions and Perspectives

6.1 Summary

Enzymes are extremely versatile catalysts that are commonplace in both therapeutic and biotech processes.[1-5] From activation of therapeutic molecules to production of fine chemicals, enzymes can greatly accelerate chemical reactions with high specificity as well as enantio and stereoselectivity. Natural enzymes are often not capable of a particular reaction or their properties are not well suited for a process. However, advancements in protein engineering methods over the years have it made it possible to create designer enzymes with tailored catalytic and biophysical properties for desired applications. In this dissertation, I addressed three separate applications of enzymes using semi-rational design approaches and circular permutation to alter their properties. The underlying theme throughout my work was using targeted engineering approaches to alter enzymatic performance and further investigate the impact of the change at a molecular level by conducting a variety of biophysical and structural characterizations.

6.2 A starting point for evolving future 3' modified nucleoside analog kinases

In chapter 2, homologous mutations from a previously evolved *Dm*dNK kinase for ddT were transferred into a hdCK variant resulting in a similar specificity switch.[6] These results provide further evidence for their highly similar active site environments and could potentially be a starting point for future efforts to create human deoxyribonucleoside analog kinases. Separately, the *Dm*dNK active site was redesigned with the help of the Rosetta Software Suite to more efficiently incorporate 3'-Methylketo-thymidine (kT), a potential RT inhibitor.[7] The final designer enzyme

contained mutations that significantly decreased the steric constraints created by the branched ketone group and also introduced new hydrogen bonding interactions with the carbonyl oxygen. All of these changes resulted in decreased activity for the natural substrate Thy while maintaining moderate activity for kT. Both of these studies involved generating NA analog kinases using minimally invasive screening with improved catalytic activity for their respective 3' modified nucleoside analog.

6.3 Future investigation of subtilisin BPN' propeptide

Prior work in the field has demonstrated that the inhibitory function of the propeptide domain of subtilisin is directly related to the stability and retention of the propeptide secondary and tertiary structure.[8-15] Bryan *et al* and Kojima *et al* have experimentally explored engineering the prosequence of subilisin BPN' for increased stability and have further conferred an increase in the inhibitory function in the latter.[9-12,16-18] In Chapter 3, I focused on generating novel protease inhibitors in silico using Rosetta software suite. The only stipulations given to the computer were to leave the 17 interface residues untouched as these presumably only affect direct interactions with the protease. Analysis of the primary amino acid sequence revealed that computational simulations generated peptide sequences possessing a striking 50 % sequence identity change compared to the parent propeptide scaffold. The top two variants, ROS1 and ROS3, were picked for experimental evaluation and both demonstrated a strong emergence of secondary structure and increased thermostability. Furthermore, MD simulations of the selected sequences provided additional support that the predicted propeptides retained structure over time. Surprisingly, the inhibitory performance of the designer propeptides only

showed modest improvements in inhibition for some proteases and none in others. These results suggest that the stabilization of the propeptide alone is not directly proportional to the potency of the inhibitor.

Future studies could explore the discrepancies between the structure and function of the propeptide domains. For instance, the original models were based off of the subtilisin BPN' propeptide domain and the designs were tested for inhibitory function against non-parent proteases. The difference in origins between the two could explain the poor function and increased inhibition might become apparent if the native propeptides were used in future designs. Although my experiments excluded making changes at interfacial residues, these contacts could also be altered by other substitutions introduced throughout the sequence. It would be interesting to include these residues in future propeptide designs to capture interactions at this interface. Overall this work provides a powerful tool for designing synthetic propeptide domains with increased structure and stability. The designer peptides display altered inhibition profiles for a variety of proteases offering a highly tunable approach for the design of potential protease inhibitors.

6.4 Continuing to teach Old Yellow 'new tricks'

Novel engineering strategies for creating altered biocatalysts is in constant demand.[4,19-21] A highly impactful discovery of this dissertation is the creation of a completely synthetic circular permutation library and the development of a new screening strategy for OYE1 which led to the identification of >70 improved biocatalysts. There have only been a few examples in which CP was used to enhance the catalytic

performance of an enzyme and even fewer have actually been successful.[22-25] The flavin depend oxidoreductase, OYE1, is considered to be a great candidate for circular permutation because the core structure is surrounded by a number of α-helices and loops that aid in forming the active site. We therefore hypothesized that relocations of the protein's N and C- termini in these regions would not affect protein folding but would increase the flexibility and active site accessibility, in turn increasing the catalytic activity for industrially unsaturated α,β-alkene substrates. Based on this hypothesis, a completely synthetic CP library was synthesized unlike the traditional random CP approach. This synthetic approach is becoming possible for more scientists to explore as the price of whole gene synthesis is decreasing. The OYE library provides a powerful demonstration for synthetic CP libraries yielding biocatalysts with improved catalytic function.

A subsequent *in vitro* screening platform was developed using the PURE system to allow for high throughput screening of oxidoreductase activity that did not require some form of purification and enabled the simple incorporation of FMN and its analogs, respectively. The entire library was tested against three reference substrates, identifying several variants that displayed over one order of magnitude improved catalytic activity. Interestingly, location of the new termini in all of the 'hits' fell within the same four α-helical and loop regions near the active site for all substrates. From library screening and further activity analysis we learned that all variants maintained *R*-enantioselectivtiy and the activity benefits between the regions varied with substrate.

Further mechanistic investigations with one of the top permutants, cpOYE303, revealed that catalytic improvements for ketoisophorone can be attributed to rate improvements for the rate-limiting oxidative half reaction. All of the finding from chapter

4 support our hypothesis that conformational flexibility in these regions of permutations is critical for catalysis. The structural similarities between the OYE family suggests that the CP engineering approach would be a fruitful endeavor for other members potentially resulting in gains in catalytic activity. Besides exploring other OYE homolog's it would also be interesting to test more structurally demanding substrates with the cpOYE library. There are several industrially relevant substrates that were reported turnover by native OYE at poor conversion rates and it would be interesting to see if permutation enables higher performance.

In addition, chapter 4 touches on exploration of FMN analogs in our library system. The concept of incorporating the analogs in the *in vitro* transcription/translation system is a novel idea and opens endless opportunities for future work. Through replacement of FMN with 7,8-dichloro FMN in selected library members, I was able to show preliminary results for beneficial rate enhancements for the reverse desaturase reactions. The enhancements are presumably due to synergistic effects from both permutation and altered redox potential. These experiments are just the start for continuing to engineer OYE1, as well as other cofactor dependent enzymes in a combinatorial protein engineering approach. In summary this work highlights: 1) a method for creating a complete and synthetic CP library, 2) an IVTT screening system that is high throughput, free of contaminating background, and enables cofactor supplementation. 3) initiates the concept of 2-D engineering showing simultaneous changes from both permutation and cofactor analog replacement.

In Chapter 5, I report on the consequences of CP on structure and stability by characterizing several cpOYE representatives from each of the four previously identified

regions. Circular dichroism measurements showed identical secondary structures for all tested permutants but size exclusion chromatography suggested varying quarternary structures. Thermostability assessments revealed that the variants either maintain or have an increased thermo-denaturation temperature indicating no negative effects from CP. Crystallographic studies on top performer cpOYE303 verifies a significantly more open and accessible active site with altered flexibility near the locations of new and old termini. The crystal structures solved in this chapter not only help rationalize the observed differences in catalytic performance in chapter 4 but also provide a solid structural framework for subsequent mutagenesis experiments.

6.5 References

1 Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discovery* **12**, 447-464, (2013).

2 Damborsky, J. & Brezovsky, J. Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.* **13**, 26-34, (2009).

3 Goldsmith, M. & Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **22**, 406-412, (2012).

4 Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14**, 438-443, (2003).

5 Toogood, H. S. & Scrutton, N. S. Enzyme engineering toolbox – a 'catalyst' for change. *Catal. Sci. Tech.*, (2013).

6 Liu, L. F., Li, Y. F., Liotta, D. & Lutz, S. Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. *Nucleic Acids Res.* **37**, 4472-4481, (2009).

7 Liu, X. J., Xie, W. & Huang, R. H. Structure-based design, synthesis, and in vitro assay of novel nucleoside analog inhibitors against HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **15**, 3775-3777, (2005).

8 Jain, S. C., Shinde, U., Li, Y. Y., Inouye, M. & Berman, H. M. The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 angstrom resolution. *J. Mol. Biol.* **284**, 137-144, (1998).

9 Kojima, S., Deguchi, M. & Miura, K. Involvement of the C-terminal region of yeast proteinase B inhibitor 2 in its inhibitory action. *J. Mol. Biol.* **286**, 775-785, (1999).

10 Kojima, S., Hisano, Y. & Miura, K. Alteration of inhibitory properties of Pleurotus ostreatus proteinase A inhibitor 1 by mutation of its C-terminal region. *Biochem. Bioph. Res. Co.* **281**, 1271-1276, (2001).

11 Kojima, S., Minagawa, T. & Miura, K. The propeptide of subtilisin BPN' as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity. *Febs. Lett.* **411**, 128-132, (1997).

12 Kojima, S., Minagawa, T. & Miura, K. Tertiary structure formation in the propeptide of subtilisin BPN ' by successive amino acid replacements and its close relation to function. *J. Mol. Biol.* **277**, 1007-1013, (1998).

13 Ruan, B., Hoskins, J., Wang, L. & Bryan, P. N. Stabilizing the subtilisin BPN ' pro-domain by phage display selection: How restrictive is the amino acid code for maximum protein stability? *Protein Sci.* **7**, 2345-2353, (1998).

14 Shaw, A. & Bott, R. Engineering enzymes for stability. *Curr. Opin. Struc. Biol.* **6**, 546-550, (1996).

15 Shinde, U., Li, Y. Y., Chatterjee, S. & Inouye, M. Folding Pathway Mediated by an Intramolecular Chaperone. *P. Natl. Acad. Sci. U. S. A* **90**, 6924-6928, (1993).

16 Bryan, P. *et al.* Catalyis of a Protein-Folding Reaction - Mechanistic Implications of the 2.0 Angstrom Structure of the Subtilisin-Prodomain Complex. *Biochemistry* **34**, 10310-10318, (1995).

17 Kojima, S., Iwahara, A. & Yanai, H. Inhibitor-assisted refolding of protease: A protease inhibitor as an intramolecular chaperone. *Febs. Lett.* **579**, 4430-4436, (2005).

18 Kojima, S., Yanai, H. & Miura, K. Accelerated refolding of subtilisin BPN ' by tertiary-structure-forming mutants of its propeptide. *J. Biochem-Tokyo* **130**, 471-474, (2001).

19 Wang, M., Si, T. & Zhao, H. M. Biocatalyst development by directed evolution. *Bioresour. Technol.* **115**, 117-125, (2012).

20 Toogood, H. S., Gardiner, J. M. & Scrutton, N. S. Biocatalytic Reductions and Chemical Versatility of the Old Yellow Enzyme Family of Flavoprotein Oxidoreductases. *Chemcatchem* **2**, 892-914, (2010).

21 Woodley, J. M. New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends Biotechnol.* **26**, 321-327, (2008).

22 Qian, Z., Horton, J. R., Cheng, X. & Lutz, S. Structural redesign of lipase B from Candida antarctica by circular permutation and incremental truncation. *J. Mol. Biol.* **393**, 191-201, (2009).

23 Qian, Z. & Lutz, S. Improving the catalytic activity of Candida antarctica lipase B by circular permutation. *J. Am. Chem. Soc.* **127**, 13466-13467, (2005).

24 Reitinger, S. *et al.* Circular permutation of Bacillus circulans xylanase: a kinetic and structural study. *Biochemistry* **49**, 2464-2474, (2010).

25 Yu, Y. & Lutz, S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.* **29**, 18-25, (2011).