

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:

Ying Yu

Date

**Enzyme Engineering by Circular Permutation:
Functional and Structural Investigation of Permutants**

By

Ying Yu
Doctor of Philosophy

Chemistry

Dr. Stefan Lutz
Advisor

Dr. David Lynn
Committee Member

Dr. Vincent Conticello
Committee Member

Accepted:

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

Date

**Enzyme Engineering by Circular Permutation:
Functional and Structural Investigation of Permutants**

By

Ying Yu

B.S., East China University of Science and Technology, 1999

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
2010

Abstract

Enzyme Engineering by Circular Permutation: Functional and Structural Investigation of Permutants

By Ying Yu

Enzyme engineering is an important tool to overcome the limitation of natural enzymes as biocatalysts. Although a number of significant achievements have been accomplished from evaluating large combinatorial libraries, more and more scientists and engineers have turned their attentions to small and high quality libraries. Circular permutation (CP) is a sequence rearrangement technique that has long been used for studying protein folding and sequence-function relationship. As an engineering tool, random CP does not require massive libraries due to its small theoretical size and thus provides opportunities for enzymes limited by the absence of high-throughput screening methods.

Previously, the engineering of *Candida antarctica* lipase B (CALB) by random CP yielded permutants with significant rate enhancements. In this dissertation, we extend the technique to three other industrially important enzymes including epoxide hydrolase (EchA), cutinase and xylanase (BcX). Screening of the BcX library identified 35 unique functional variants with new termini distributed across the whole sequence and subsequent characterization revealed permutants with up to 4-fold increased activity. In contrast, CP of EchA and cutinase only yielded permutants highly similar to wild-type enzymes. Our attempts to address this problem by optimizing the linker using combinatorial approaches were not successful.

To follow up our previous studies on permuted CALB, we evaluated the performance of our best permutant cp283 with a series of esters, as well as pure and complex triglycerides. In comparison with the wild-type enzyme, cp283 showed consistently higher catalytic activity (2.6- to 9-fold) for transesterification and interesterification when using 1-butanol and ethyl acetate as acyl acceptors, indicating the potential application of our cpCALB in biodiesel production. Furthermore, in order to better understand the consequences resulted from CP, we investigated the function and structure of the new termini of cp283 Δ 7, a variant of cp283 with truncated linker. Although the 17 residues at the C-terminus are not visible in the crystal structure, our results suggest that they are important for enzyme function and likely exist as a dynamic helical arrangement rather than being highly flexible and unstructured.

**Enzyme Engineering by Circular Permutation:
Functional and Structural Investigation of Permutants**

By

Ying Yu

B.S., East China University of Science and Technology, 1999

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
2010

Acknowledgments

This dissertation ends my long journey to obtain my doctoral degree. It is all the people who guided me, helped me and loved me made it possible.

First of all, my deepest gratitude is to my advisor, Dr. Stefan Lutz. I would not accomplish anything in this dissertation without his continuous support, constructive guidance, critical comments and enthusiastic inspiration. I am grateful to him for providing me opportunities to work on different projects, to become an independent thinker, to question thoughts and to express my ideas. His understanding and patience in the past five years helped me overcome many difficulties and finish this dissertation.

I would like to thank my committee members Dr. David Lynn and Dr. Dale Edmondson, as well as Dr. Vincent Conticello for their advices, criticisms and encouragements. Their wide knowledge and scientific way of thinking always encourage me to become better in research. I also appreciate their instrumental assistance with my experiments.

It was a great experience to collaborate with Dr. Steve Withers, Dr. Lawrence McIntosh and their lab members at University of British Columbia. Thanks for their hard and excellent work. I would like to also thank Dr. Tim Lian and his lab members for their help in our collaborations.

My time in the graduate school was made enjoyable in large part of due to the wonderful former and current Lutz lab members, as well as my friends. I appreciate their helps, fun discussions and warm friendships. Especially thanks to Kitty Liu, Dan Shi, Zhen Qian, Yichen Liu and Joey Licther for their caring, companion and entertainment.

I am so blessed to have a remarkably big family. Thanks to my entire extended family for building a loving environment since my childhood. My best and special thanks are to my in-laws (father, mother, brother, sister) for all of their support. Most importantly, my immediate family has been the love source and strength for me all these years, no matter where I am. Thanks my brother, for always being there for me. I can't express how grateful I am to my parents, Huogen Yu and Fanglan Qian, who lived a hard life to raise me and support my education. They always stand behind me and leave me free for all my decisions. Their unconditional love has helped to get me to where I am today. I appreciate every sacrifice they have made. To them I dedicate this dissertation.

Last, but not least, I wish to thank my husband and my daughter. Without my husband, Jianxiong Jiang, I would not go to graduate school and finish my Ph.D. Thanks him for all these years' companion, for caring me and for sharing the family responsibilities. My daughter, Aubree, came to me during my graduate career. She is the best project I ever received in my life and the project I will carry on throughout the rest of my life. I cherish all of the happiness she brings to me and I love her with all of my heart.

Table of Contents

1	Introduction Enzyme Engineering and Circular Permutation	1
1.1	Achievements of enzyme engineering	1
1.2	Enzyme engineering methods	4
1.3	Circular permutation (CP) of proteins	5
1.3.1	Circular permutation in nature	5
1.3.2	Laboratory circular permutation	7
1.3.3	Linker design in circular permutation	11
1.4	Functional and structural consequences of circular permutation	12
1.4.1	Influence of circular permutation on protein function	12
1.4.2	Influence of circular permutation on protein structure	13
1.5	Applying circular permutation in enzyme design	15
1.6	Circular permutation of CALB	17
2	Engineering Biocatalysts by Circular Permutation	20
2.1	Introduction	20
2.1.1	<i>Agrobacterium radiobacter</i> epoxide hydrolase (EchA)	21
2.1.2	<i>Fusarium solani</i> cutinase	22
2.1.3	<i>Bacillus circulans</i> xylanase (Bcx)	24
2.2	Materials and methods	25
2.2.1	Materials	25
2.2.2	Random circular permutation of <i>echA</i>	26
2.2.3	Preparation of CP vector for <i>echA</i>	29

2.2.4 Library Construction and selection	29
2.2.5 Construction of rational designed EchA permutants	30
2.2.6 Construction of EchA library comprising varying linkers	30
2.2.7 Screening of EchA library comprising varying linkers	32
2.2.8 Random circular permutation of cutinase	32
2.2.9 Screening of CP cutinase library	33
2.2.10 Construction and screening of random linker cutinase library	35
2.2.11 Circular permutation of cutinase with random linker.....	35
2.2.12 Random circular permutation of BcX	36
2.2.13 Screening of cpBcx library	37
2.3 Results and discussions	38
2.3.1 Random circular permutation of EchA	38
2.3.2 Characterization of rational designed EchA permutants	40
2.3.3 Optimization of EchA linker with combinatorial approaches	43
2.3.4 Random circular permutation of cutinase	46
2.3.5 Optimization of cutinase linker with combinatorial approaches	48
2.3.6 CP of cutinase with random linkers	49
2.3.7 Random circular permutation of Bcx	50
2.4 Conclusions	52

3 Improved Triglyceride Transesterification by Circular Permuted *Candida antarctica*

Lipase B	53
3.1 Introduction	53
3.2 Materials and methods	55
3.2.1 Materials	55
3.2.2 Kinetic analysis of lipases with chromogenic and fluorescence substrates	55

3.2.3 Kinetic analysis of transesterification reaction	56
3.2.4 Trans and interesterification of vegetable oil	56
3.3 Results and Discussions	57
3.3.1 Impact of acyl and alcohol moieties of substrates on the catalytic performance of cp283	57
3.3.2 Transesterification of triglyceride	60
3.3.3 Alcoholysis of complex triglyceride mixture and oils	62
3.3.4 Interesterification with ethyl acetate	64
3.3.5 Recycling of immobilized enzyme	65
3.4 Conclusions	65
4 Functional and Structural Investigation of the New Termini of Permuted CALB	67
4.1 Introduction	67
4.2 Materials and Methods	70
4.2.1 Materials	70
4.2.2 Construction of termini truncation and site-directed mutagenesis variants	72
4.2.3 Protein expression in <i>P. pastoris</i>	72
4.2.4 Protein purification by HIC and SE chromatography	72
4.2.5 Kinetic analysis of lipases activity with chromogenic and fluorescence substrates.	73
4.2.6 Circular dichroism analysis	73
4.2.7 Peptide synthesis and purification	74
4.2.8 Cloning and purification of sortase A	75
4.2.9 Peptide labeling and sortase-mediated protein ligation	76
4.2.10 Labeling of Cys mutants	76
4.3 Results and Discussions	77
4.3.1 Functional consequence of termini deletion	77

4.3.2 CD characterization of termini deletion variants.....	79
4.3.3 Role of the two leucine residues in the invisible C-terminus	80
4.3.4 Labeling of new C-terminus by protein-peptide ligation	82
4.3.5 Secondary structure of C-terminal peptide	85
4.3.6 Characterization and labeling of Cys mutants.....	87
4.4 Conclusions	89
5 Conclusions and Perspectives	90
5.1 Conclusions	90
5.2 Perspectives	92
References	93

List of Figures

1-1 Topology comparing of polyhydroxybutyrate depolymerase from <i>Penicillium funiculosum</i> with the canonical α/β hydrolase fold	6
1-2 Schematic diagram of random circular permutation	10
1-3 Ribbon diagrams of β B2-crystallin before and after circular permutation	15
1-4 Front and side view of wild-type CALB	18
2-1 Structure view of CALB, EchA, Cutinase and Bcx.....	23
2-2 Vector map of pSTBlue-1	28
2-3 Vector maps of pYY101 and pBAD-cp189-GFP	28
2-4 Vector maps (Invitrogen) of pPICZ α and pPIC9.....	34
2-5 Schematic diagram of circular permutation with random linkers	36
2-6 Sequencing analysis of selected members from native cpEchA library	39
2-7 Selection of functional epoxide hydrolase permutants using glycidyl phenyl ether	40
2-8 Rational designed EchA circular permutants.....	42
2-9 The use of GFP as a solubility reporter for screening of random linker EchA libraries	45
2-10 Screening of cp189-RLIT library using FACS	46
2-11 Random circular permutation of cutinase.....	47
2-12 Screening of cutinase libraries on tributyrin agar plate.....	49
2-13 Circular permutation of Bcx.....	51
2-14 Screening of Bcx libraries with Congo red plate assay.....	52
3-1 Time course of transesterification reactions for WT-CALB and cp283	63
3-2 Enzyme activity of immobilized WT-CALB and cp283 over multiple cycles of canola oil transesterification with ethyl acetate	66

4-1	Schematic overview of CALB engineering by CP and subsequent linker truncation	68
4-2	Comparison of the active-site accessibility in wild-type CALB and cp283 Δ 7.....	69
4-3	Far-UV CD spectra and melting temperatures of wild-type and termini deletion variants.....	79
4-4	Highlight of helix α 16 and α 17 in wild-type CALB structure	81
4-5	Gel filtration spectra of separation sortase-mediated ligation mixture	84
4-6	CD spectra of C-terminal peptide in different conditions	86
4-7	Five selected sites in wild-type CALB for fluorophore labeling	87

List of Tables

1-1 Examples of significantly improved enzyme properties by directed evolution	3
1-2 Examples of circularly permuted proteins generated in laboratory	8
1-3 X-ray crystallography studies on circular permuted proteins	14
3-1 Apparent kinetic constants for WT and cp283 with different ester substrates	59
3-2 Apparent kinetic constants for transesterification of triglycerides by WT and cp283	61
3-3 Composition of vegetable oils	61
4-1 Primers used to generate termini truncation and site-directed mutagenesis variants	71
4-2 Apparent kinetic constants with <i>p</i> -nitrophenyl butyrate and DiFMU octanoate as substrates for termini truncation variants	78
4-3 Apparent kinetic constants for Leu substitutions with <i>p</i> -nitrophenyl butyrate and DiFMU octanoate as substrates	82
4-4 Catalytic activity of Cys mutants with <i>p</i> -nitrophenyl butyrate as substrates	88

Chapter 1 Introduction

Enzyme Engineering and Circular Permutation

Enzymes are nature's catalyst, substantially accelerating the rates of a wide range of biochemical reactions. The use of enzymes as biocatalysts in industrial processes is ever increasing since it not only eliminates harsh conditions, but also offers advantages of exquisite selectivity, easy product recovery and reduced environmental impact. It has been estimated that the global market for industrial enzymes was around USD 2.9 billion in 2008, growing at an average of 3-8% annually (Freedonia Group Inc 2009). Major applications range from preparation of detergents, pulp and paper, animal feeds, food, production of biofuel to synthesis of fine chemicals and pharmaceuticals (Cherry and Fidantsef 2003). Despite the advantages over chemical catalysts, there are often discrepancies between the functions of naturally occurring enzymes and the specific requirements of practical applications. Hence, enzyme properties often need to be optimized for better performance and for reduced costs.

1.1 Achievements of enzyme engineering

For the past three decades, numerous enzymes from all types of families have benefited from laboratory evolution toward enhanced catalytic activity, stability or selectivity. Table 1-1 summarizes the most impressive engineering achievements reported in the literature. Among different engineering goals, obtaining an enzyme with increased activity stands for the largest category. Since the pioneering work by Pim Stemmer (Stemmer 1994), who improved β -lactamase activity by 32,000-fold after multiple rounds of DNA shuffling, the catalytic activities of a number of enzymes have been increased by 10-fold, 100-fold, or even more than a thousand-fold after one or more rounds of directed evolution. The engineered enzymes have diversified

structural folds and function, as reflected by some of the most outstanding improvements including halohydrin dehalogenase (Fox, Davis et al. 2007), glyphosate acetyltransferase (Castle, Siehl et al. 2004), α -amylase, (Wong, Batt et al. 2004) and phosphotriesterase (Hill, Li et al. 2003). Remarkable progress has also been accomplished for modification of enantioselectivity, substrate specificity, stability and solubility as discussed in several review papers (Bloom, Meyer et al. 2005; Johannes and Zhao 2006; Leisola and Turunen 2007; Turner 2009; Lutz 2010).

Recent optimization of substrate specificity and stability of a transaminase through structure-based rational design and directed evolution is one of the extraordinary examples how enzyme engineering is translated into practical advantages (Savile, Janey et al. 2010). This work aimed to replace the rhodium-based hydrogenation catalyst in the production of an antidiabetic drug, Sitagliptin, with an efficient biocatalyst for better selectivity and easier purification. Because the native transaminase does not accept bulky substrates like the precursors of sitagliptin, Savile *et al* first enlarged the binding site pocket by applying saturation mutagenesis on four residues that were identified by computation modeling. Subsequently, they carried out 10 rounds of directed evolution and screened with increasing stringency. The best variant as a result of this process, carrying a total of 27 mutations, achieved high productivity with excellent enantioselectivity and good efficiency in 50% organic solvent at 40°C. The mutations were found in the active site as well as at the enzyme's dimer interface.

Another significant advancement in enzyme engineering has been the de novo design of enzymes, catalyzing chemical reactions for which no natural biocatalyst exists (Damborsky and Brezovsky 2009). Rosetta, the leading algorithm in protein design, enables design of enzymes for multistep reactions and has led to several pioneering synthetic biocatalysts for the Kemp elimination (Rothlisberger, Khersonsky et al. 2008), retro-aldol reaction (Jiang, Althoff et al. 2008) and Diels-Alder reaction (Siegel, Zanghellini et al. 2010). In the latest example, Siegel and coworkers applied Rosetta to create a Diels-Alderase that stereoselectively catalyzes an intermolecular Diels-Alder cycloaddition, a reaction important in organic synthesis with no

Table 1-1. Examples of significantly improved enzyme properties by directed evolution

Property	Enzyme	Method	Result	Reference
Catalytic activity	Halohydrin dehalogenase	epPCR & gene shuffling (ProSAR)	4000-fold increase of productivity	(Fox, Davis et al. 2007)
	Glyphosate N-acetyltransferase	epPCR & DNA shuffling	10,000-fold increase	(Castle, Siehl et al. 2004)
	α -amylase	epPCR & DNA shuffling	1000-fold higher activity	(Wong, Batt et al. 2004)
	phosphotriesterase	cassette mutagenesis	Three orders of magnitude increase	(Hill, Li et al. 2003)
	β -lactamase	DNA shuffling	32,000-fold increase	(Stemmer 1994)
Thermostability	Lipase A	epPCR	T_m increased 15 °C	(Ahmad, Kamal et al. 2008)
	Pectate lyase	site saturation mutagenesis	T_m increased 16 °C	(Solbak, Richardson et al. 2005)
	Xylanase	site saturation mutagenesis	T_m increased 35 °C	(Palackal, Brennan et al. 2004)
Enantioselectivity	Lipase A	iterative saturation mutagenesis	E-value increased ~600 fold	(Reetz, Prasad et al. 2010)
	Cyclohexanone monooxygenase	epPCR	<i>ee</i> % for <i>R</i> -increased to 90% from 9%	(Reetz, Brunner et al. 2004)
	Epoxide hydrolase	epPCR & DNA shuffling	13-fold improvement	(van Loo, Spelberg et al. 2004)
Specificity	endopeptidase	epPCR	Three-million-fold selectivity	(Varadarajan, Gam et al. 2005)
Solubility	cytochrome P450 _{sca}	epPCR & DNA shuffling	30-fold increase	(Li, Guan et al. 2009)
Promiscuity	Carbonic anhydrase	mutagenesis & recombination	40-fold increase with 2-nahthyl acetate	(Gould and Tawfik 2005)

naturally biocatalyst. With the combination of computational modeling and experimental optimization, they identified enzymes that are able to catalyze intermolecular cycloaddition with high selectivity and activity comparable to the catalytic antibodies (Lutz 2010).

1.2 Enzyme engineering methods

Enzyme engineering is typically performed by either rational design or directed evolution. Rational design, which focuses on a small number of variants, relies on structural information and an in-depth mechanism understanding to identify specific residues that can be mutated in order to improve a desired property. In contrast, directed evolution follows a two-step iterative process of diversification and selection. A library of thousands or millions of mutants is created and functional variants are identified through high-throughput screening or selection (Dougherty and Arnold 2009). It is not necessary to have detailed knowledge of the relationship of sequence and structure to carry out directed evolution, but the successful implementation requires construction of sufficient sequence diversity and development of an efficient screening or selection assay.

Among all of the directed evolution techniques, random mutagenesis is the most straightforward and allows uncovering unexpected beneficial mutations. This approach randomly introduces mutations into the whole gene of an enzyme during error-prone PCR and often generates multiple mutations per gene. Using sequential rounds of random mutagenesis, followed by screening or selection, successes have been achieved for a wide range of engineering goals (Labrou 2010), including improved activity and enantioselectivity. However, it is unlikely for random mutagenesis to create multiple neighboring mutations simultaneously, which is generally required for tailoring substrate specificity or regioselectivity. Another type of directed evolution approach is homology-based recombination, which mimics nature's recombination events and introduces large sequence changes into proteins without the need for structural information. (Bloom, Meyer et al. 2005). DNA shuffling reassembles gene fragments from multiple parents

with high sequence homology, while ITCHY and SRATCH are independent of sequence identity. Using multiple rounds of DNA shuffling in combination with random mutagenesis has created variants with remarkable improvements in catalytic activity, stability, solubility and enantioselectivity (Leisola and Turunen 2007).

Although construction and evaluation of massive directed evolution libraries may provide a greater chance of finding variants with the desired improvements, it has been realized that even libraries with millions of members, created by directed evolution methods only represent a small fraction of the vast number of possible combinations in sequence space. Consequently, the emphasis of more recent methods has shifted towards the generation of smaller and higher quality libraries. Semi-rational design concentrates on a specific region in a protein and uses structural or biochemical data to choose several positions for simultaneous mutagenesis. It can dramatically reduce the library size and increase the ratio of positive variants, which allows tailoring enzymes' function for which no efficient screening assays are attainable. Impressive examples using semi-rational design have been achieved through combining the benefits of rational design and directed evolution (Chica, Doucet et al. 2005; Lutz 2010).

In this dissertation, we focused on circular permutation (CP), a conventional engineering method that has emerged as a useful tool to diversify enzyme sequence.

1.3 Circular permutation (CP) of proteins

1.3.1 Circular permutation in nature

A circularly permuted protein is a protein variant in which the native termini have been connected by a peptide sequence and new ends were introduced at another location in the sequence. Since the first observation of circular permutation in plant lectins in 1979, a large number of naturally occurring circular permuted proteins including saposins, aldolases, β -glucanases, DNA methyltransferases and depolymerase, have been reported in protein structural

and bioinformatics studies (Lindqvist and Schneider 1997; Jung and Lee 2001; Lo and Lyu 2008). Circular permuted proteins can emerge as a result of posttranslational modification, but a majority is believed to arise from gene duplication (Jeltsch 1999; Peisajovich, Rockah et al. 2006) or exon shuffling (Gilbert 1987) events during evolution.

An example of a circular permuted enzyme in nature is polyhydroxybutyrate depolymerase (PHD), a member of α/β hydrolase (Hisano, Kasuya et al. 2006). As shown in Figure 1-1, the topology of PHD (A) secondary structure resembles that of the canonical α/β hydrolase fold (B). However, its N- and C-termini are located within the α/β core fold which results in a discontinuity of the polypeptide chain in the middle of the central β -sheet

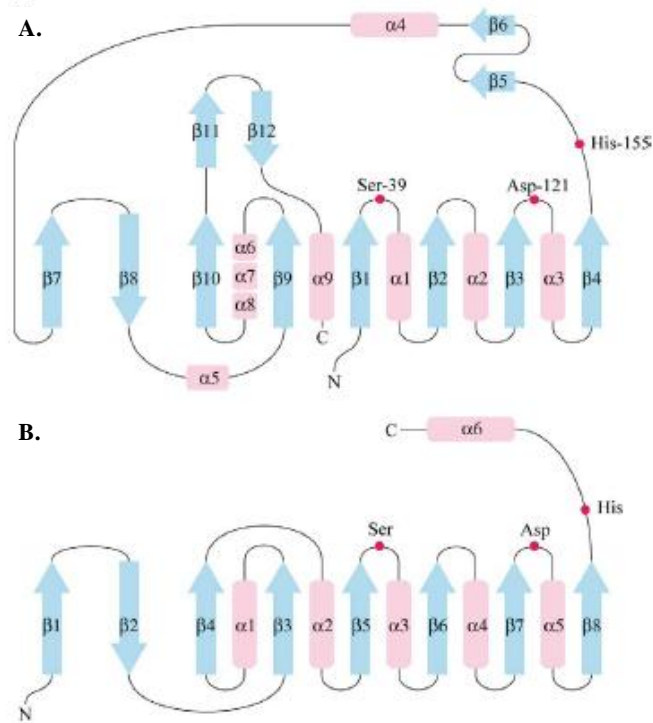


Figure 1-1. Topology comparing of (A) polyhydroxybutyrate depolymerase from *Penicillium funiculosum* with (B) the canonical α/β hydrolase fold (Hisano, Kasuya et al. 2006). Pink rectangles and blue arrows represent α -helices and β -sheets, respectively. Red dots indicate catalytic residues. (The reuse of the figure for this dissertation is permitted by Elsevier Limited. License number 2539710494920)

1.3.2 Laboratory circular permutation

In the laboratory, circular permutation involves connecting the native N- and C- termini either directly or with a short peptide and introducing new termini at an alternative position within the protein sequence. Unlike amino acid substitution, circular permutation alters the connectivity within the polypeptide chain without changing the amino acid composition. The first permuted protein was produced in 1983 by Goldenberg and co-workers (Goldenberg and Creighton 1983) who used direct chemical and enzymatic modification to create a circularly permuted bovine trypsin inhibitor. Since then, circular permutation has become a popular tool for evaluating the effects of primary sequence arrangement on the folding process, stability and function. A number of circularly permuted proteins with various structural folds have been constructed by rational design using genetic engineering approaches (summarized in Table 1-2). These artificially created permutants include monomeric single domain proteins such as xylanase, and multidomain proteins such as lipase B, T4 lysozyme and β -lactamase, as well as multimeric proteins such as aspartate transcarbamoylase and β -crystalline. The new termini were usually introduced in solvent exposed loops in order to retain the protein's structural and functional integrity. Indeed, most of the variants were able to retain proper structure although they often encountered reduced stability. The results from these studies demonstrate that the tendency to attain a stable conformation is not limited by the order of the secondary structural elements in the sequences.

In addition to site-directed circular permutation, random circular permutation was developed in 1996 (Graf and Schachman 1996) to explore the tolerance of termini relocation throughout the whole protein sequence. This technique involves the creation of a random circularly permuted gene library and the screening for functional variants. As illustrated in Figure 1-2, the linker sequence and a unique restriction site are introduced at the ends of the original gene by PCR. Then, the gene is circularized via intramolecular ligation and relinearized by limited DNaseI digestion. The resulting permuted DNA library is cloned into an expression vector for screening or selection.

Table 1-2. Examples of circularly permuted proteins generated in laboratory.

Protein	Estimated termini distance	Linker sequence	Reference
xylanase (<i>Bacillus circulans</i>)	3 Å ^b	- GG -	(Reitinger, Yu et al. 2010)
penicillin G acylase (<i>E. coli</i>)	5 Å ^a	random tetrapeptide	(Flores, Soberon et al. 2004)
cyanovirin-N (HIV)	7 Å	directly linked	(Barrientos, Louis et al. 2002)
bovine pancreatic trypsin inhibitor	8 Å	direct linked	(Goldenberg and Creighton 1983)
Hemoglobin α -globin (human)	8.5 Å	- G -	(Sanders, Lo et al. 2002)
spectrin (<i>Gallus gallus</i>)	10 Å ^a	- GTG -	(Viguera, Blanco et al. 1995)
ompX (<i>E. coli</i>)	9 Å	- GGSG -	(Rice, Schohn et al. 2006)
Streptavidin (<i>Streptomyces avidinii</i>)	10 Å	- GGGS -	(Chu, Freitag et al. 1998)
TEM- β -lactamase (<i>E. coli</i>)	10 Å ^a	Random linker	(Osuna, Perez-Blancas et al. 2002)
phosphoribosyl-anthranilate isomerase (<i>E. coli</i>)	11 Å	- YDPS -	(Luger, Hommel et al. 1989)
RNase T1 (<i>Aspergillus oryzae</i>)	11 Å	- GPG - Random linker	(Mullins, Wesseling et al. 1994; Kuo, Mullins et al. 1995)
T4 lysozyme (phage)	11 Å	- SGGGGS -	(Zhang, Bertelsen et al. 1993)
β -glucanase (<i>Bacillus macerans</i>)	11 Å	Direct linked	(Hahn, Piotukh et al. 1994; Heinemann, Ay et al. 1996)
Interleukin 4 toxin (Human)	11 Å	- GGNGG -	(Kreitman, Puri et al. 1994)
Glyceraldehyde dehydrogenase (<i>Bacillus stearothermophilus</i>)	11 Å	- GGGAG -	(Heinemann, Ay et al. 1996)
disulfide oxidoreductase (<i>E. coli</i>)	13-15 Å ^b	- GGGTG -	(Hennecke, Sebbel et al. 1999)

Protein	Estimated termini distance	Linker sequence	Reference
virulence factor Caf1 (<i>Yersinia pestis</i>)	14 Å	- TGSGNG -	(Chalton, Musson et al. 2006)
aspartate transcarbamoylase (<i>E. coli</i>)	14 Å ^b	- SGELM -	(Yang and Schachman 1993; Graf and Schachman 1996)
dihydrofolate reductase (<i>E. coli</i>)	15 Å	- GSG - Varied length	(Protasova, Kireeva et al. 1994; Iwakura and Nakamura 1998)
p300/CBP histone acetyltransferase	15 Å	- TGGGSGG -	(Karukurichi, Wang et al. 2010)
lipase B (<i>Candida antarctica</i>)	17 Å ^b	- GGTS GG -	(Qian and Lutz 2005)
avidin (<i>Gallus gallus</i>)	17 Å	- GGSGGS -	(Nordlund, Hytonen et al. 2005)
dihydrofolate reductase (<i>Mus musculus</i>)	19 Å	- GSRSVNL -	(Buchwalder, Szadkowski et al. 1992)
green fluorescence protein (<i>Aequorea victoria</i>)	19 Å ^b	- GGTGGS -	(Baird, Zacharias et al. 1999)
Myoglobin (sperm whale)	21 Å	- (GGGS) ₄ -	(Ribeiro and Ramos 2005)
alkaline phosphatase (<i>E. coli</i>)	27 Å	- (GGGS) ₃ -	(Kojima, Ayabe et al. 2005)
Barnase (<i>Bacillus amyloliquefaciens</i>)	28 Å	- (GGS) ₅ GTM - - C -	(Butler, Mitrea et al. 2009)
RNase A (<i>Bos taurus</i>)	30 Å	- GSGKPIE FLDLKAG -	(Plainkum, Fuchs et al. 2003)
5-aminolevulinate synthase (<i>Mus musculus</i>)	~50 Å ^b	Directly linked	(Cheltsov, Guida et al. 2003)

^a native termini were truncated to shorten distance for linker design

^b random circular permutation

Several enzymes have been subjected to random circular permutation. Such proteins include aspartate transcarbamoylase (Graf and Schachman 1996), GFP (Baird, Zacharias et al. 1999), disulfide oxidoreductase (Hennecke, Sebbel et al. 1999), 5-aminolevulinate synthase (Cheltsov, Barber et al. 2001), lipase B (Qian and Lutz 2005) and xylanase (Reitinger, Yu et al. 2010). These investigations found that the location for the new termini is not limited to flexible loop regions but that a remarkably large number of sites within regular secondary structures can produce functional variants. Random circular permutation is not only useful for understanding how primary sequence dictates protein folding and thermodynamics, but also for identifying valuable circularly permuted variants.

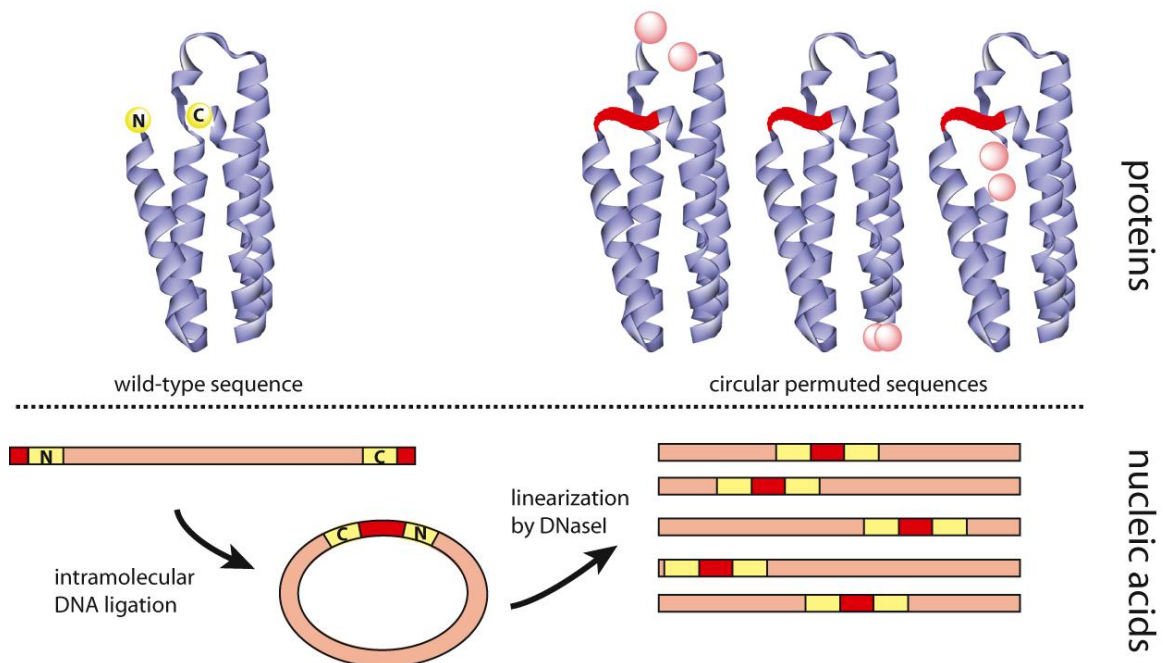


Figure 1-2. Schematic diagram of random circular permutation. The gene encoding for interested protein is self-ligated to create circular DNA which is then relinearized by limited DNaseI digestion to produce a library of permuted genes. The native termini are shown in yellow; the linker connects the native termini is shown in red; the new termini are shown in pink.

1.3.3 Linker design in circular permutation

An important consideration to obtain functional and stable circular permuted proteins is the linker peptide connecting the original N- and C- termini. The length and composition of the peptide linker can have a significant effect on protein activity and stability. As a general rule, the length is decided by ~ 3.5 Å per amino acid and glycines are often used to minimize the steric strain. Linking termini distance of less than 10 Å is usually a relative straightforward process. However, the analysis of native termini distances in all protein structures in the PDB reveals that less than 10% of protein termini are 10 Å apart while termini distances of 10-20 Å, 20-30 Å, and 30-40 Å make up the majority of cases, each occurring at a frequency of $\sim 20\%$ (Wang, Kaas et al. 2008). Linking these greater distances is more challenging and requires more careful design. One strategy is to truncate some of the residues at the termini to shorten the distance. When the study of proteins in the PDB expanded its termini proximity analysis to include entire terminal regions (10 residues on either end), the proteins with termini distances less than 20 Å became over 50% of the structures. Based on this data, the elimination of a few amino acid residues at the native protein termini could in many cases significantly shorten the required linker length and consequently simplify the design.

The other strategy that has been utilized by several groups is to optimize the length and composition of an existing linker by rational design or combinatorial approaches. Iwakura and co-workers constructed circular permuted dihydrofolate reductase (DHFR) with various poly-glycine linkers two to six residues in length. They concluded from biochemical and biophysical data that the use of five glycine residues is most favorable for joining the 15 Å distance between the native N- and C-terminal, while shorter linkers induce conformational strain on the permutants (Iwakura and Nakamura 1998). In a separate example, Akeman and co-workers generated a series of permuted enhanced green fluorescence protein (EGFP) to study how a given linker sequence affects the biophysical properties. The 24 Å termini distance was connected with linkers of one to 6 amino acids in length. The characterization showed that a 3-amino acid linker was sufficient to

form a functional chromophore, although the yield was lower compared to the permutants with longer linkers (Akemann, Raj et al. 2001).

In terms of the linker peptide composition, Garrett *et al* generated circularly permuted variants of ribonuclease T1 with a library of 3-amino acid peptides consisting of any combination of Pro, Asp, Asn, Ser, Thr, Tyr, Ala and His (Garrett, Mullins et al. 2003). Based on library screening and further characterization, they found that the amino acid content of the linker sequence has a significant influence on the thermodynamic stability of the permutants, favoring sequences with fewer glycine residues.

1.4 Functional and structural consequences of circular permutation

1.4.1 Influence of circular permutation on protein function

Artificially permuted proteins often retain their properly folded structure and biological functions. The activity of variants largely depends on the location of new termini. In some cases, the activities were comparable with those of their parental proteins. Examples include DHFR (Buchwalder, Szadkowski et al. 1992), DsbA (Hennecke, Sebbel et al. 1999), β -glucanase H (Hahn, Piotukh et al. 1994), TEM- β -lactamase (Osuna, Perez-Blancas et al. 2002), penicillin G acylase (Flores, Soberon et al. 2004), alkaline phosphatase (Kojima, Ayabe et al. 2005) and barnase (Butler, Mitrea et al. 2009).

Nevertheless, in a few cases, circular permutation was able to improve protein function. Kreitman and coworkers investigated if changing the location of the protein-protein junction could improve the function of fusion proteins. They initially generated a chimeric toxin in which interleukin 4 toxin (IL4) was fused to *Pseudomonas* exotoxin. However, the fusion protein bound to the IL4 receptor with only ~1% of the affinity of native IL4. Circular permutation of IL4 resulted in a fusion protein bound to the IL4 receptor with a 10-fold higher affinity than the initial

chimeric toxin (Kreitman, Puri et al. 1994). Another successful example comes from 5-aminolevulinatase synthase (ALAS), an enzyme that catalyzes the condensation of glycine and succinyl-CoA to form 5-aminolevulinic acid, CoA and carbon dioxide. Random circular permutation created a total of 9 unique active variants (Cheltsov, Barber et al. 2001). All of the variants showed either similar or better catalytic properties when compared to wild type enzyme. The best variant exhibited a 2 and 10-fold increase of catalytic efficiency to succinyl-CoA and glycine respectively. The most striking activity improvement was observed in the random circular permutation of *Candida antarctica* lipase B (CALB), which will be discussed later in detail.

1.4.2 Influence of circular permutation on protein structure

While circular permutation has a significant impact on protein function, its influence on structure is mainly limited to the old and new termini region as determined in spectroscopic and X-ray crystallographic studies (Table 1-3). Overall, permutants that retained any level of activity were shown to preserve the native-like folds after termini relocation.

An interesting phenomenon upon circular permutation is the change of quaternary structure. The wild-type rat eye lens β B2-crystallin, a two-domain protein, forms a dimer by domain swapping (Figure 1-3 A) (Wright, Basak et al. 1998). Circular permutation placed the new termini in the linker region of the two domains and converted the native intermolecular dimer into a two-domain monomer in which the two domains are now swapped back and pair with each other, resembling a homologue γ -crystallin (Figure 1-3 B). Similarly, circular permutation of HIV-inactivating protein cyanovirin-N converted the domain-swapped dimer protein to a monomer with significant reduced anti-HIV activity (Barrientos, Louis et al. 2002). Along the same line, *Yersinia pestis* vaccine antigen Caf1, a major protective antigen in the development of subunit vaccines against plague, was converted from the original polymer to a folded monomer by circular permutation (Chalton, Musson et al. 2006). *In vitro* and *in vivo* studies showed the

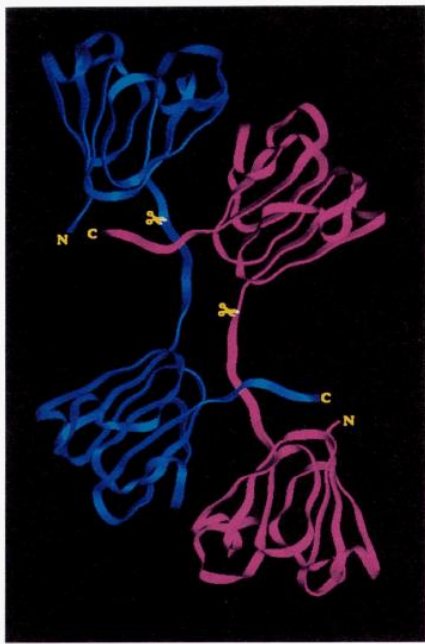
Table 1-3. X-ray crystallography studies on circular permuted proteins.

Protein	PDB	fold family	comments	Reference
xylanase (<i>Bacillus circulans</i>)	3LB9	β -jellyroll	CP variant highly similar to WT	(Reitinger, Yu et al. 2010)
lipase B (<i>Candida antarctica</i>)	3ICV 3ICW	α/β -hydrolase	WT monomer \rightarrow CP dimer (apo form & with bound suicide inhibitor)	(Qian, Horton et al. 2009)
lysozyme (T4 bacteriophage)	1P5C 2O4W		CP variant highly similar to WT	(Sagermann, Baase et al. 2004; Cellitti, Llinas et al. 2007)
β -glucanase (<i>Bacillus macerans</i> & <i>B. amyloliquefaciens</i>)	1CPM 1CPN	β -jellyroll	CP variant of <i>B.mac/B. amyl.chimera</i> & <i>B. macerans</i> . Both variants highly similar to WT	(Hahn, Piotukh et al. 1994)
β -lactamase (<i>Staphylococcus aureus</i>)	1ALQ	(β/α)8-barrel	CP variant highly similar to WT	(Pieper, Hayakawa et al. 1997)
β -crystallin (human)	1BD7	γ -crystallin like	WT dimer \rightarrow CP monomer	(Wright, Basak et al. 1998)
α -spectrin SH3 domain	1TUC 1TUD	SH3	two CP variants with WT-like structure	(Viguera, Serrano et al. 1996)
phosphoglycerate kinase (<i>Saccharomyces cerevisiae</i>)	1FW8		CP variant highly similar to WT	(Tougaard, Bizebard et al. 2002)
cyanovirin-N (<i>Nostoc ellipsosporum</i>)	1N02		WT dimer \rightarrow CP monomer	(Barrientos, Louis et al. 2002)
disulfide oxidoreductase (<i>E.coli</i>)	1UN2	thioredoxin	CP variant very close to WT	(Manjasetty, Hennecke et al. 2004)
streptavidin (<i>Streptomyces</i>)	1SWF		CP variant with complexed biotin; similar to WT	(Chu, Freitag et al. 1998)
avidin (<i>Gallus gallus</i>)	2AVI		CP variant with complexed biotin; highly similar to WT	(Maatta, Airenne et al. 2008)
ribosomal protein S6 (<i>Thermus thermophilus</i>)	2KJW	(β - α - β) ₂	CP variant very similar to WT	(Ohman, Oman et al. 2010)
RNase (Barnase) (<i>Bacillus amyloliquefaciens</i>)	1BRS		CP variant with bound inhibitor; similar to WT	(Butler, Mitrea et al. 2009)

monomeric antigen preserved the immunogenicity and protective efficacy, although antibody titers dropped 10-fold.

In the opposite direction, circular permutation of CALB created highly active monomeric permutants with compromised thermostability. Shortening the linker peptide forced the new N-terminal region of the permutant into an extended form that resulted in dimerization via domain swapping. The resulting cpCALB dimer retained the high activity and also regained some of the lost thermostability.

A.



B.



Figure 1-3. Ribbon diagrams of β B2-crystallin before and after circular permutation (Wright, Basak et al. 1998). **A:** Two subunits (blue and purple) of β B2-crystallin form a domain-swapped dimer. Native termini are shown in yellow. Permutation site is indicated by scissors.

B: Superposition of cp β B2 (pink) monomer with top half of a β B2 dimer in A. (blue). New termini are shown in pink. Connected native termini are shown in blue. (The reuse of the figure for this dissertation is permitted by Wiley & Son)

1.5 Applying circular permutation in enzyme design

Beyond the use of circular permutation as a tool for basic investigations of folding pathways, structure-function relationships and for tailoring enzyme activity or specificity, its integration for designing of biosensors and enzyme switches has produced several innovative solutions (Huang and Bystroff 2009; Carlson, Cotton et al. 2010). CP can be used to identify positions that are tolerant to insertion of other domains. The earliest example comes from GFP. Insertion of calmodulin into a circularly permuted GFP whose new termini are close to the chromophore resulted in a biosensor whose fluorescence can be enhanced several-fold upon the metal binding (Baird, Zacharias et al. 1999). Circularly permutation has also found great utility in the optimization of fluorescence resonance energy transfer (FRET)-based nanosensors for measurement of small molecule metabolites (Okada, Ota et al. 2009), in the improvement of bioluminescent probes (luciferases) for illumination of protein molecular dynamics (Kim, Sato et al. 2008) and in design of efficient ligand-binding switch (Ostermeier 2009). Additionally, Whitehead and coworkers found that CP reduces proteolysis and increases protein expression for recombinant thermosome (Group II chaperonin) and TEM-1 β -lactamase (Whitehead, Bergeron et al. 2009).

Another significant application of CP in enzyme design is the creation of zymogen by two different concepts. In the construction of a ribonuclease A zymogen, Raines and coworkers blocked the active site of the native enzyme by connecting the native termini with a peptide linker encoding a protease recognition sequence (Plainkum, Fuchs et al. 2003). The new termini were tested at 9 positions and the site that produced the most stable permutant was selected. The concept was tested with plasmepsin II, a highly specific protease of *Plasmodium falciparum*. In the presence of plasmepsin II, the RNase zymogen showed a 1000-fold increase in activity. Similar strategies were also used to develop sensors for HIV (Turcotte and Raines 2008) and hepatitis C virus NS3 proteases (Johnson, Lin et al. 2006). A slightly different mechanism for

creating artificial zymogenes by CP is alternate frame folding (AAF) (Stratton, Mitrea et al. 2008; Mitrea, Parsons et al. 2010). In AAF, either the N or C-terminal segment of the wild-type enzyme is duplicated and fused to the other end of the full-length sequence. The fusion effectively creates two interchangeable reading frames, the first representing the native enzyme and the second corresponding to a circular permutant. The protein can fold into either the active native form or the inactive permutant variant that carries a deleterious mutation. The protease cleavage site can be introduced into a surface loop that will become the new termini of the permutant upon cleavage of the duplicate fragment. In the presence of protease, cleavage induces the shift from wild-type to permutant and the switch is observed through the change in activity.

1.6 Circular permutation of CALB

Lipase B from *Candida antarctica* (CALB) is a versatile biocatalyst which can catalyze hydrolysis, esterification, transesterification and amidation reactions. Its high activity, stability and selectivity in both aqueous and organic solvents have made it a useful enzyme in the laboratory for asymmetric synthesis and in industry for a variety of biotechnological applications. Like many other lipases, CALB is a member of the α/β hydrolase fold family with a central, mostly parallel β -sheet, flanked on both sides by α -helices (Figure 1-4). The active site contains a Ser-His-Asp catalytic triad, an oxyanion hole, and the substrate binding site. Because of the important applications of CALB, studies have been carried out to modify its activity, stability, enantioselectivity and substrate specificity using rational design and directed evolution approaches (Lutz 2004).

In our group, we employed random circular permutation to engineer CALB and identified 63 unique functional variants from the library (Qian and Lutz 2005). The newly introduced termini were located not only in exposed loop regions, but also in secondary structure elements (Figure 1-4). A big portion of permutants were concentrated in the cap region, including the lid and the C-

terminal helix 16/17. Detailed characterization identified some variants with increased hydrolytic activity of 2 to 14-fold and 6 to 175-fold for *p*-nitrophenol butyrate and 6,8-difluoro-4-methylumbelliferyl octanoate, respectively (Qian, Fields et al. 2007). However, permutants suffered significant stability losses. The T_m value of the best permutant, cp283 dropped 14°C from 55°C for the wild type enzyme. To improve the protein stability and understand the structural changes contributed to the rate enhancements, incremental truncation was applied to the linker region of cp283. This secondary engineering recovered some of the lost thermostability while maintaining the high catalytic activity. Subsequently, the crystal structure provides further insight into the rate enhancement. Termini relocation to position 283 converted the narrow access tunnel to the catalytic site into a broad crevice for accelerated substrate entry and release.

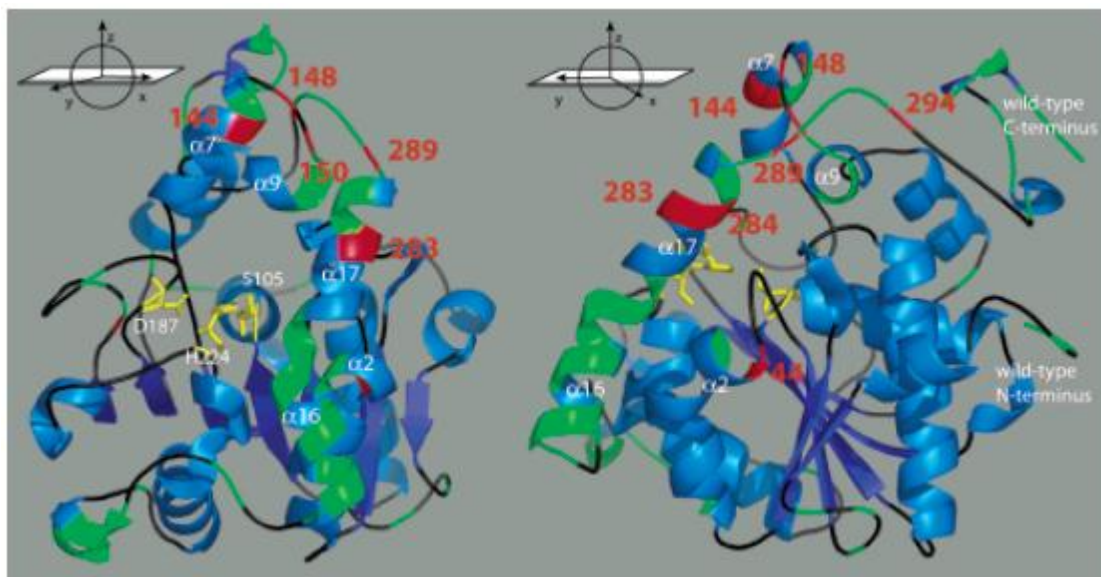


Figure 1-4. Front and side view of wild-type CALB (1TCA). Green indicates the permissible permutation sites; Red indicates further characterized candidates selected; Native termini and helices are labeled in white. Yellow indicates the catalytic residues. (The reuse of the Figure for this dissertation is permitted by American Chemical Society. License number 2540001449306)

In summary, protein engineering has made remarkable contributions to industry and disease treatment. It not only produces efficient biocatalyst, but also helps understand protein structure-function relationships in a basic level and provides insight into protein adaptive mechanisms (Arnold 2009). As a relatively new engineering tool, circular permutation produces small libraries (gene size equals theoretical library size) and therefore it is attainable for screening of the majority of diversification. Of course, there are limitations for this technique and it is not suitable for all the enzymes. Accumulation of more random CP examples would definitely be beneficial in understanding the scope of CP in protein engineering. Despite that, circular permutation is a very promising method of providing greater diversity and new features for future engineering studies.

Chapter 2 Engineering Biocatalysts by Circular Permutation

2.1 Introduction

As discussed in the previous chapter, circular permutation has been a useful tool to study polypeptide folding for more than three decades, but its capability and potential in protein engineering is not well-understood by the community. The success with *Candida antarctica* lipase B (CALB) encouraged us to explore random circular permutation toward other proteins in the same superfamily or proteins with other folds.

The α/β hydrolase fold family is one of the largest groups of structurally related enzymes. Members in this family include functionally versatile biocatalysts: lipase, esterase, peptidase, proteases, haloalkane dehalogenase, and epoxide hydrolase (Holmquist 2000). These enzymes contain a conserved eight-stranded α/β core fold that provides a stable scaffold for the catalytic triad, and a varied cap domain (also called lid or flap in some cases) that defines substrate specificity and regulates substrate accessibility. In some of the lipid hydrolases, the lids or flaps undergoes open/close conformational rearrangements during interfacial activation, a phenomenon that results in a sharp increase in enzyme activity when the substrate forms a lipid phase and thereby presenting to the enzyme an oil-water interface (Jaeger, Ransac et al. 1994; Holmquist 2000).

In this chapter, we first tested whether circular permutation can be used to improve the function of other α/β hydrolases. Two family members, *Agrobacterium radiobacter* epoxide hydrolase (EchA) and *Fusarium solani* cutinase were chosen for this study. While EchA has a defined cap domain (Figure 2-1B), cutinase (Figure 2-1C) only has a very small flap. In addition,

our studies of circular permutation were extended to xylanase from *Bacillus circulans*, a member of the glycoside hydrolase family with a β -jellyroll fold.

2.1.1 *Agrobacterium radiobacter* epoxide hydrolase (EchA)

Epoxides are chemically reactive and thus potentially cytotoxic compounds arising from xenobiotic metabolism. Epoxide hydrolases (EHs, EC 3.3.2.3) are a class of enzymes that hydrolyze epoxides to their corresponding vicinal diols by the addition of a water molecule. This function makes EHs important in cell detoxification (Fretland and Omiecinski 2000). For chemists, enantiopure epoxides and diols are valuable building blocks in the synthesis of enantiopure pharmaceuticals and versatile fine chemicals (Finney 1998; de Vries and Janssen 2003). Enantiopure epoxides can be prepared by asymmetric synthesis or kinetic resolution of a racemic epoxide, using a chemical catalyst or an EH (Steinreiber and Faber 2001; de Vries and Janssen 2003). So far, over 50 microbial EHs have exhibited moderate or high enantioselectivity toward approximately 70 epoxides, including monosubstituted epoxides, styrene oxide-type epoxides and disubstituted epoxides (Archelas and Furstoss 2001; Choi 2009)

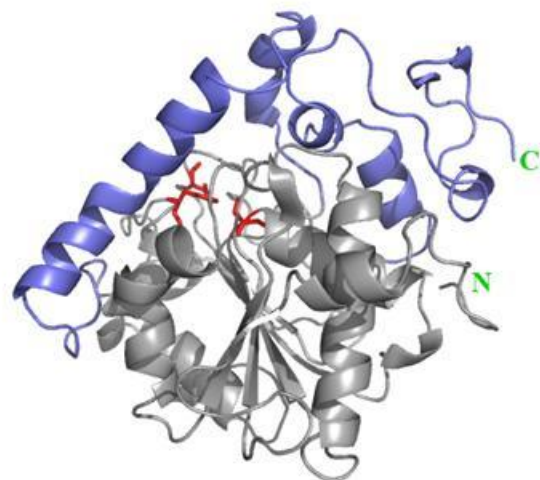
The EH from *Agrobacterium radiobacter* (EchA) is a 31 kDa monomeric protein. Similar to CALB, its active site contains an Asp-His-Asp catalytic triad and is located in a hydrophobic cavity between the α/β -fold core domain and the cap domain (Figure 2-1B). The reaction catalyzed by EchA follows a two-step mechanism involving an alkyl-enzyme intermediate, which is further hydrolyzed via the attack of a water molecule (Nardini, Ridder et al. 1999). Due to the promising applications of EHs in industry, it is desirable to expand their biocatalytic capabilities. Previously, some efforts have been conducted to tailor EchA's function (Lee and Shuler 2007). Site-directed mutagenesis resulted in a 5-fold increase in the enantiomeric ratio for (*R*)-*p*-nitrostyrene oxide (Rink, Lutje Spelberg et al. 1999). Using error-prone PCR and DNA shuffling followed by an agar plate screening, Loo and coworkers identified mutations around the active site that improve the enantioselectivity up to 13-fold toward *p*-nitrophenyl glycidyl ether (Loo,

Spelberg et al. 2004). Separately, the Wood group evolved EchA by directed evolution for chlorinated epoxyethanes and styrene oxide. In the former case, EchA was tuned to accept the unnatural substrate *cis*-1, 2-dichloroepoxyethane after accumulating beneficial mutations from three rounds of saturation mutagenesis at three selected active site residues (Rui, Cao et al. 2004). In the latter, DNA shuffling and saturation mutagenesis generated a variant with 2-fold enhanced activity and 5-fold enhanced enantioselectivity (Rui, Cao et al. 2005).

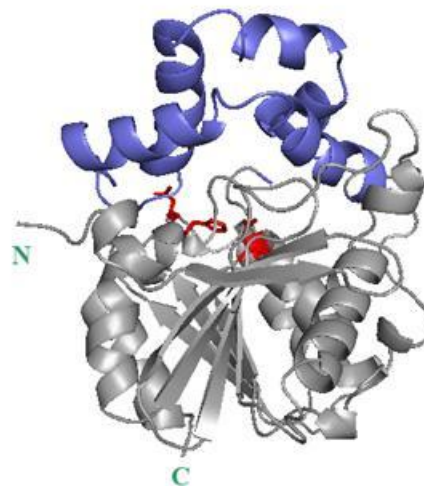
Despite all the progress in screening of new EHs or improving their function through engineering, challenges remain for the broader practical applications with respect to activity, cost and substrate scope. For example, few EHs accept trisubstituted epoxides or bulky internal epoxides such as *cis*-stilbene oxide and its derivatives as substrates (Zhao, Han et al. 2004). As part of our studies, we aimed to tailor EchA by random CP and explore its effect on activity and substrate specificity.

2.1.2 *Fusarium solani* cutinase

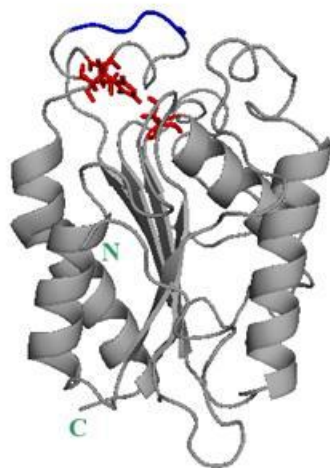
Cutinases are enzymes excreted by phytopathogenic fungi to hydrolyze ester bonds in the cutin polymer, an insoluble lipid polyester matrix covering the surfaces of plant leaves. They belong to the class of serine esterases and are also able to hydrolyze fatty acid esters and triglycerides as efficiently as lipases, but without displaying interfacial activation (Longhi and Cambillau 1999). Being the smallest member of the α/β hydrolase family, cutinases only have the α/β -fold domain, with the Ser-Asp-His catalytic triad on the surface exposed to the solvent. A 5-residue helical flap and a small binding loop partially block the entrance to the active site (Figure 2-1C) (Martinez, De Geus et al. 1992; Longhi, Czjzek et al. 1997). The easy accessibility of the active site feature allows cutinases to accept bulky polymers such as plastic for a substrate, making them alternative candidates for applications in green chemistry. Cutinase has also found applications in detergent, food and dairy industry, as well as chemical synthesis (Carvalho, Aires-Barros et al. 1999; Pio and Macedo 2009).



A. CALB (PDB: 1TCA)
(Uppenberg, Hansen et al. 1994)



B. EchA (PDB: 1EHY)
(Nardini, Ridder et al. 1999)



C. Cutinase (PDB: 1AGY)
(Longhi, Czjzek et al. 1997)



D. Bcx (PDB: 1C5I)
(Joshi, Sidhu et al. 2000)

Figure 2-1. Structure view of CALB, EchA, Cutinase and Bcx. Cap domain or flap is indicated in blue; catalytic residue and native termini are indicated in red and green, respectively. For the cutinase, one residue at N-terminal and two residues at C-terminal are invisible.

Cutinase from *Fusarium solani*, a 197-amino acid protein with a molecular weight of 22 kD, is the mostly extensively-studied cutinase. Many sites including the substrate binding region and some charged residues were mutated to study this enzyme's electrostatic properties or structure and dynamics (Egmond and de Vlieg 2000). So far, engineering of cutinase has been only limited to site-directed mutagenesis. Mutations in the helical flap were found to shift the preference for the acyl chain of ester substrates, while introducing aromatic residues into the binding loop region led to higher activity with hydrophobic substrates (Egmond, de Vlieg et al. 1996). In another study, substituting an individual residue in the active site pocket to Ala achieved up to 5-fold enhanced activity toward larger polymers, such as polyethylene terephthalate and polyamide fibers (Araujo, Silva et al. 2007). Recently, cutinase was fused with a carbohydrate-binding module to increase enzyme concentration at the cellulose surface which then translated into increased hydrolysis of cellulose acetate (Matama, Araujo et al. 2010). Overall, further engineering of cutinase for improved activity and stability by directed evolution would greatly expand its potential for industrial and environmental applications.

2.1.3 *Bacillus circulans* xylanase (Bcx)

The endo- β -1,4 xylanase from *Bacillus circulans* is a member of the family 11 glycoside hydrolases. Xylanases catalyze the hydrolysis of the plant cell wall component xylan, the second most abundant polysaccharide in nature (Collins, Gerday et al. 2005). This ability renders xylanases a valuable tool in various biotechnology applications ranging from pulp and paper industry, animal feeds and food preparation and potentially biofuel production (Beg 2001). Consequently, xylanases have been subjected to numerous studies on the structural and functional level.

The Bcx is a 20 kDa protein that has served as a model for understanding the glycoside hydrolase mechanism. It has been extensively characterized kinetically, as well as structurally by X-ray crystallography and NMR spectroscopy (Wakarchuk, Campbell et al. 1994; Plesniak,

Wakarchuk et al. 1996). The β -jellyroll fold structure resembles a partially closed right-hand, with the active site sitting in the “palm” formed by β -pleated sheets (Figure 2-1D). Bcx degrades xylan by a retaining, double-displacement mechanism with the assistance of two conserved catalytic glutamic acids. Previously, studies of Bcx and its mutants performed by our collaborators has provided detailed insight into the mechanism of carbohydrate degradation (Lawson, Wakarchuk et al. 1996; McIntosh, Hand et al. 1996; Lawson, Wakarchuk et al. 1997; Joshi, Sidhu et al. 2000; Ludwiczek, Heller et al. 2007; Schubert, Poon et al. 2007). Nevertheless, further understanding of structural/function relationships could be obtained from circular permuted Bcx. We believe Bcx is a suitable candidate for CP since the β -jellyroll structure places its native termini in very close proximity.

2.2 Materials and methods

2.2.1 Materials

Genes: The gene of *Agrobacterium radiobacter* epoxide hydrolase (*echA*) and was a gift from Professor Dick Janssen at University of Groningen (Netherlands). The gene of *Fusarium solani* cutinase was a gift from DNA 2.0 (Menlo Park, CA). The gene of *Bacillus circulans* xylanase was provided by Professor Steve Wither at University of British Columbia (Canada). The gene of GFP was obtained from a vector pGFPuv (Clontech, Mountain View, CA).

Chemicals: Reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO) with the exception of arabinose, which was obtained from ACROS Organic (Morris Plains, NJ).

Plasmids: pSTBlue-1 and pET-27b were products from Novagen (Madison, WI). pBAD and pPIC9 were gifts from Professor Vincent Conticello’s lab and Professor Dale Edmondson’s lab (Emory University), respectively. pPICZ α was purchased from Invitrogen (Carlsbad, CA).

Strains: *E. coli* DH5 α was from laboratory stock. *E. coli* TOP10 was a gift from Professor Conticello’s lab. Chemical competent *E. coli* BL21 (λ DE3) was purchased from Stratagene (LA

Jolla, CA). *Pichia pastoris* strain GS115 (*his4*) was a gift from Professor Edmondson's lab. All of the electro-competent cells were prepared in the laboratory following the protocol in Molecular Cloning (Sambrook and Russell 2001).

Enzymes: Enzymes were purchased from New England Biolabs (Ipswich, MA) unless noted otherwise. DNaseI was acquired from Roche Biochemicals (Indianapolis, IN). Exonuclease III (ExoIII), T4 DNA ligase and T4 DNA polymerase were obtained from Promega (Madison, WI). *Pfu* DNA polymerase was purchased from Stratagene (La Jolla, CA). All of the reactions involving enzymes were performed in the manufacture buffers unless indicated otherwise.

Cloning: Oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA). *Pfu* DNA polymerase was used for all the cloning PCR. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit and DNA in agarose gel was recovered with QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). QIAquick PCR Purification Kit was used for purification of PCR product and enzyme treated reactions. All gene constructs were confirmed by DNA sequencing.

Media: Bacteria were grown under standard conditions in Luria-Bertani (LB) liquid media or on LB agar plates supplemented with appropriate antibiotics. Library harvest medium contains 8% tryptone, 5% yeast extract, 2.5% NaCl, 15% glycerol and 2% glucose, (pH 7.2). MMH- polycaprolactone plate (1.34% yeast nitrogen base, 0.004% biotin, 0.5% methanol and 0.05% polycaprolactone) and MM-tributyryn plate (0.34% yeast nitrogen base, 0.004% biotin, 0.5% methanol and 1% tributyrin) were used for cutinase library screenings. The media were homogenize with an IKA Ultra Turrax T18 (IKA works, Inc., Wilmington, NC) after autoclaving.

2.2.2 Random circular permutation of *echA*

Codons for a *SpeI* restriction site (underlined) and linker sequence (Italic) were introduced to the ends of wild-type *echA* by PCR with EHCP-For (5' - AATAACTAGTGGTGGTGGCGGC GCAATTCGACGTCCAGAAGAC -3') and EHCP-Rev (5' - CGCACTAGTACCGCCACCGCG GAACGCGGTTTTGATTCG -3'). The resulting PCR product was cloned into the *EcoRV* site of a high copy plasmid pSTBlue-1 (Figure 2-2) by blunt end ligation to produce pSTBlue-*echA*, which was prepared in a large amount as the starting material of random CP using Perfectprep Plasmid Maxi Kit (Eppendorf, Hamburg, Germany).

To perform random CP, over 30 µg of plasmid was digested with *SpeI* at 37°C overnight and the desired ~920 bp fragment was isolated via agarose gel electrophoresis. 5 µg of linker gene was then circularized by intramolecular ligation at the concentration of 2 µg of DNA/mL with 30 units/mL T4 DNA ligase at 16°C overnight. The ligation product was isolated by ethanol precipitation and treated with ExoIII (120 units/µg of DNA) at 37°C for 30 min to remove the remaining linear DNA. After heat inactivation of ExoIII at 72°C for 20 min, the circular DNA was recovered via QIAquick purification. Next, the circularized DNA was randomly cleaved by limited DNaseI (0.0005 units/µg of DNA) digestion in 50 mM Tris-HCl (pH 7.5) buffer containing 1 mM MnCl₂ at room temperature for 15 min. The reaction was quenched by adding 10 µl of 0.5 M EDTA, followed by QIAquick purification. Relinearized DNA was incubated in T4 ligase buffer with T4 DNA polymerase (1 unit/µg of DNA), T4 DNA ligase (2 units/µg of DNA) and 150 µM dNTP at room temperature for 1 hour to repair DNA nicks and facilitate blunt end ligation. The resulting permuted gene library was purified by agarose gel electrophoresis and cloned into pYY101-blunt.

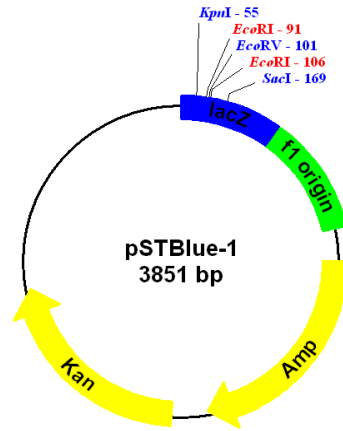


Figure 2-2. Vector map of pSTBlue-1 with used cloning site indicated.

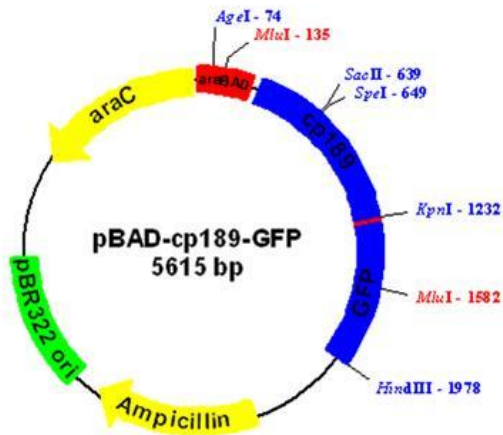
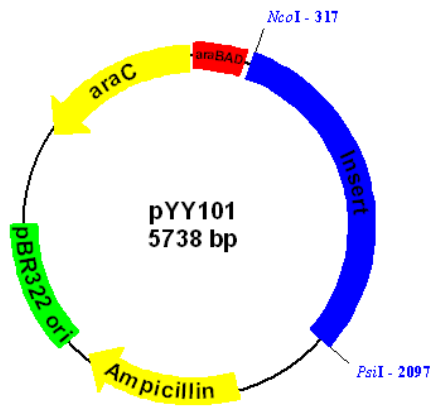


Figure 2-3. Vector maps of pYY101 (top) and pBAD-cp189-GFP (bottom) with cloning sites indicated.

2.2.3 Preparation of CP vector for *echA*

The CP vector for *echA*, pYY101 (Figure 2-3), was derived from pBAD-HisA and carries a *PsiI* restriction site introduced by PCR with primer: YY-pBAD-DmdNK-Rev (5'-CGCAAGCTTTTA TAATCAGGGCTGTTGGTTACTTGAGATG -3'). A ~1700 bp piece of DNA sequence cut from pET14-DmdNK (a construct available in the lab that carries *Drosophila melanogaster* kinase gene) was inserted between *NcoI* and *PsiI* to facilitate the separation of double digested vector from the single digested one, which could increase the background of subsequent blunt end ligation. To obtain blunt-ended vector and the start codon, the 5'-overhang from *NcoI* digestion was filled in by T4 DNA polymerase at 37°C for 5 min with the addition of 100 µM dNTP, followed by 10 min deactivation at 72°C and QIAquick purification. The resulting blunt-ended vector was dephosphorylated with alkaline phosphatase at 37°C for 30 min (named pYY101-blunt) and used for CP library cloning.

2.2.4 Library Construction and selection

The ligation mixture (inserts /vectors ≈ 3 to 1 ratio) was incubated at room temperature overnight and transformed into freshly prepared *E. coli* TOP10 cells by electroporation. Cells were spread on LB agar plate (ampicillin at 100 µg/mL) and incubated at 37°C for 10 h. The naïve library was harvested with library harvest medium (see section 2.2.1), resuspended and divided into aliquots for storage at -80°C. All the libraries discussed below were prepared by the same procedure.

For selection of functional EchA variants, colonies were grown on LB plate (100 µg/mL ampicillin and 0.02% arabinose) and then transferred to a selection plate containing 100 µg/mL carbenicillin, 0.02% arabinose and 3 mM glycidyl phenyl ether (GPE) with the help of a QPix2 colony picker (Genetix, San Jose, CA). Colonies that were able to grow after 2-4 days incubation were picked, regrown and analyzed by sequencing using primers pBAD-For (5'-ATGCCATAGCATT TTTATCC -3') and pBAD-Rev (5'-GATTTAATCTGTATCAGG -3').

2.2.5 Construction of rational designed EchA permutants

To confirm the results from our random CP and optimize the linker sequence, we created three permutants by rational design. Clones with new termini introduced at positions 167, 189 or 205 were generated via overlap PCR and placed into the pYY101-blunt vector (see section 2.2.3). Primers: cp167-for (5'- TCGATGCGCGAGGTGTGC -3'), cp167-rev (5'- GCCCACGACCTCAACGGC -3'), cp189-for (5'- TTGCTCACTGAGGAAGAACTTGAGG -3'), cp189-rev (5'- CTCATCCCGGTATGACCAGTG -3'), cp205-for (5'- CCTGATAACATTCAACGGAGG -3') and cp205-rev (5'- CTTCATACAGTTATCGACGTGAACC -3').

The two permutants composed of 10 and 11-amino acid long linkers were derived from cp205 by overlap PCR with primers: linker-for (5'- *GGCGGTACTAGTGGGGTGGCGGCGCAA* TTCGACGTCCAGAAGAC -3'), linker10-rev (5'- *ACCACCACTAGTACCGCCACCACCGCG* GAAAGCGGTCTTTATTCG -3') and linker11-rev (5'- *ACCACCACTAGTACCGCCACCACC* ACTGCGGAA AGCGGTCTTTATTCG -3'). The complementary sequences are indicated in italic. The fusion construct cp189-GFP (Figure 2-3) was created by two steps cloning. At first, cp189 was subcloned into pBAD using *MluI* and *KpnI* sites to remove the stop codon. The second step introduced the gene for GFP (taken from pGFPuv vector) at the C-terminus of cp189 gene using *KpnI* and *HindIII* sites.

2.2.6 Construction of EchA library comprising varying linkers

The composition of 9-amino acid linker was randomized with primers, samu-for (5'- AAGACCGCTTTCCGCGGTNNKNNKACTAGTNNKNNKNNKGGCGCAATTTCGACGTCC -3') and GFP-rev (5'- GTTAAGCTTTTATTTGTAGAGCTCATCCATC -3') using cp189-GFP as the template. The NNK codon (where N represents a 25% mix of the four nucleotides and K represents a 50% mix of thymine and guanine nucleotides) encodes all 20 amino acids but eliminate two of the three stop codons (Barbas 2001). The resulting PCR product was digested

2.2.7 Screening of EchA library comprising varying linkers

The 9-residue random linker library was screened by searching for glowing cells under UV light. Basically, the cells from the naïve library were spread on an LB agar plate (100 µg/mL ampicillin and 0.02% arabinose) and incubated at room temperature for 2 days or at 30°C for 30 h. The cells were inspected under the UV light and the glowing colonies were picked and regrown for sequencing analysis.

The linker library with various length and composition was subjected to FACS (fluorescence activated cell sorting) screening following a protocol described by Liu *et al.* (Liu, Li et al. 2009). Briefly, an aliquot of 20 µl cell culture from the naïve library was used to inoculate 10 mL LB media supplemented with ampicillin (50 µg/mL) and grown at 37°C for 2 h. Then 250 µl of this starter culture was used to inoculate 5 mL fresh LB medium and continued at 37°C until the OD (600 nm) reached ~0.5. Protein expression was induced with 0.02% arabinose and incubation temperature was lowered to 30°C. After 10 h, cells were centrifuged and the pellet was washed three times with PBS buffer (pH 7.5) before being resuspended in the PBS buffer to $\sim 1 \times 10^8$ cells/mL. Cell sorting was carried out on a FACSVantage flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). The event detection was set to forward and side scattering. Sorting was performed at < 2000 events/s with excitation by a UV laser (351–364 nm) and emission detection through a band pass filter (515 ± 20 nm). Cells were collected in tubes, plated on LB-agar plates and harvested using library harvest medium (see section 2.2.1).

2.2.8 Random circular permutation of cutinase

An 8-amino acid long linker was designed to bridge the 20 Å distance between the native termini. The linker was introduced to the ends of the wild-type cutinase gene with primers CUT-CPFor (5' - CATACTAGTGGTGCAGGATTAGGAAGAACAACAAGAGACGATC -3') and CUT-CPRev (5' - GTTACTAGTTCCAGCACCGGCTGAACCTCTAACAGCACGTAC -3'). The resulting PCR product was cloned into pSTBlue-1 by blunt end ligation and yielded construct

pSTBlue1-Cut, which was prepared in large quantities by QIAGEN Plasmid Midi Pre Kit (Qiagen, Valencia, CA). The latter steps adopted the same protocol as for CP of EchA (see section 2.2.3).

The CP vector (pYY201) was derived from pPICZ α A (Figure 2-4) and carried unique *SapI* and *Bsu36I* sites. The new cleavage sites (underlined) were introduced by PCR using pPICZ-*SapI*-For (5' - ATCTCTCGAGAAAAGACGAAAGAGCGGAGGCTGAAGCTGAATTC -3') and pPICZ-*Bsu36I*-Rev primer (5' -ATGAGTTTTTGTCTAGATTAGCTTACCTTAGGAAAGCTGGCGGCCGCGC -3'). To prepare the blunt-end vector for CP ligation, 5 μ g plasmid of pYY201 was sequentially digested with *SapI* and *Bsu36I* since the required digestion buffers for the two enzymes are not compatible. The DNA was recovered via agarose gel electrophoresis after every step of digestion. The double-digested vector was treated with T4 DNA polymerase at 37°C for 10 min in the presence of 100 μ M dNTP, followed by 10 min deactivation at 72°C and QIAquick purification. This fill-in step provides blunt ends, including the stop codon and the intact signal recognition site for α -factor signal peptide upon ligation. Finally, the vector was dephosphorylated with alkaline phosphatase at 37°C for 15 min (named pYY201-blunt) and used for CP library cloning. The permuted library was transformed into *E. coli* DH5 α by electroporation, yielding a library of $\sim 1.5 \times 10^5$ cfu.

2.2.9 Screening of CP cutinase library

Plasmids containing the CP cutinase library were extracted, linearized with *SacI*, transformed into electro-competent *P. pastoris* GS115 and plated on MM plate containing 0.05% polycaprolactone. Colonies expressing active hydrolases were identified by the formation of a halo surrounding the host cells. Positive colonies were picked and analyzed by DNA sequencing with pPIC9-for (5' - GACTGGTTCCAATTGACAAGC-3') and pPIC9-rev (5' - GCAAATGGCATTCTGACATCC-3'). Separately, the CP library was also subcloned into pPIC9 vector (Figure 2-4) using *XhoI* and *NotI* sites and screened on MM- tributyrin plates. Due to the

large number of positive colonies, colony PCR was carried out as the preliminary screen to DNA sequencing. The pre-screen confirmed that the majority of positive colonies were highly similar to the wild-type cutinase. Therefore, only those of colonies that were different from the wild type enzyme were submitted for sequencing analysis.

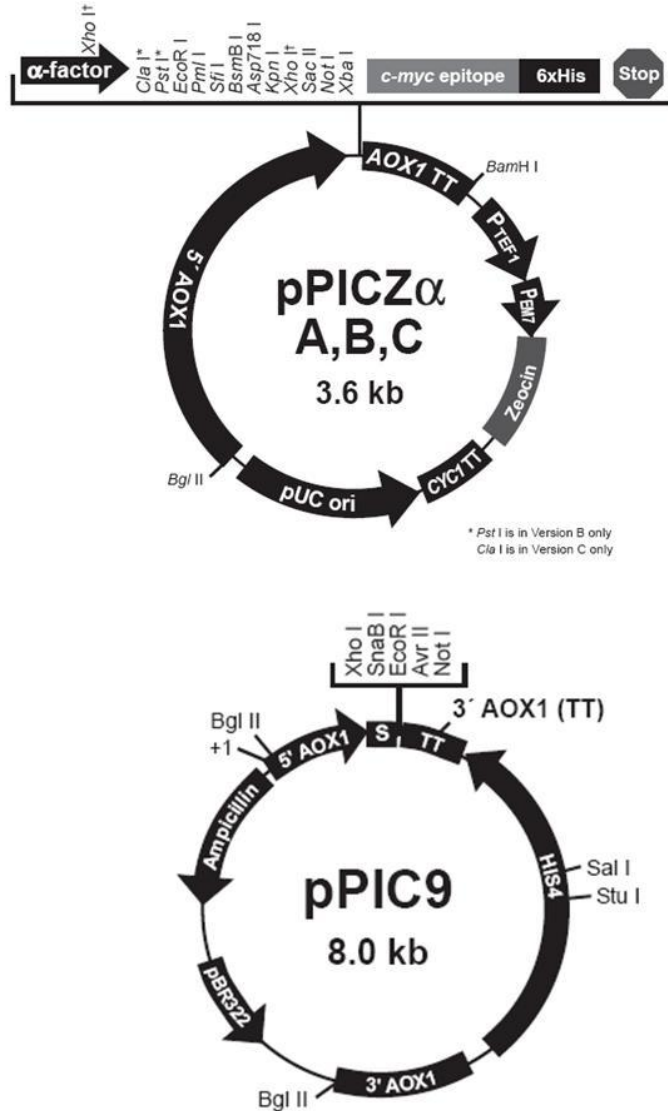


Figure 2-4. Vector maps (Invitrogen) of pPICZ α and pPIC9 used in cutinase engineering.

2.2.10 Construction and screening of random linker cutinase library

To optimize the linker peptide of cutinase, a library with various linker length and composition was prepared using a permutant (cpC31), which was identified in the previous CP library, as template. Library preparation closely followed the protocol described in section 2.2.6. Briefly, an 8-amino acid linker library (cp31-RLlib) was generated by overlap PCR with primers, CUT-RANLINK-For (5'- GCTGTTAGAGGTTCAGCCNNKNNKNNKACTAGT NNKNNKNNKTTAGGAAGAACAACAAGAG -3'), CUT-Rev (5'- GGCTGAACCTCTAAC AGCACG -3') and CUT-cp31-Rev (5'- GTAGCGGCCGCTTAGGAGGCAGAATTTCCATTT AT -3'). PCR product was digested with *XhoI/NotI* and ligated into pPIC9 predigested with the same enzymes. The ligation mixture was transformed into *E. coli* DH5 α by electroporation, producing approximately 1×10^6 colonies. The plasmid was isolated and subjected to incremental truncation for varying linker length as described in section 2.2.6. The final library (cp31-RL-ITlib) had 1×10^6 colony members.

2.2.11 Circular permutation of cutinase with random linker

As an alternative to finding suitable linkers for cutinase engineering, we performed circular permutation using random linkers instead of a defined linker. Library construction followed the scheme illustrated in Figure 2-5. Briefly, a permuted cutinase gene (cpcut_*EcoRI*) containing the previously used 8-residue linker was generated by rational design and cloned into pSTBlue1 vector. Next, six positions in the Gly-rich linker were randomized by overlap PCR with four primers: CUT-RANLINK-For, CUT-Rev, SP6 (5'- ATTTAGGTGACACTATAG -3') and T7-Rev (5'-GCTAGTTATTGCTCAGCG G 3'). The random linker genes were cloned into *KpnI* and *SacI* sites in pSTBlue (Figure 2-2), yielding a library of $\sim 5 \times 10^6$ cfu. In the third step, CP was carried out using the random linker library as starting materials through the same strategies as described in section 2.2.8. Finally, the permuted library was cloned into pPIC9 (library size: 1×10^6 cfu) and transformed into *P. pastoris* for MM-tributyryn plate screening.

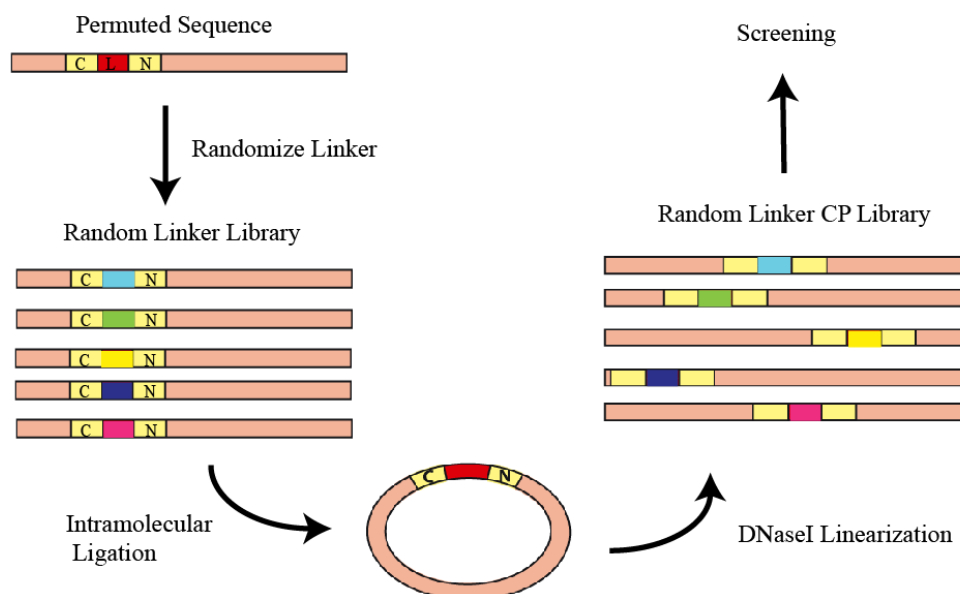


Figure 2-5. Schematic diagram of circular permutation with random linkers. A permuted gene containing a rational designed linker was firstly generated. The linker sequence was then randomized to create a random linker library, which was used as the starting material of following random circular permutation. Finally, the permuted gene library was cloned into an expression vector for screening.

2.2.12 Random circular permutation of BcX

Random circular permutation of BcX also utilized the same procedure. Briefly, a permuted gene (cpBcX_ *Nsi*I) was constructed as the starting material to avoid any wild type contamination using cpA123 as template, a permutant generated through overlap PCR by our collaborators in the study of suitable linkers. PCR with *Nsi*I-For (5' - TGAATGCATGGAAATCCCACGGTA TGAAC -3') and *Nsi*I-Rev (5' - TCCATGCATTCACGTGATTGGTGAACG -3') positioned the unique internal *Nsi*I restriction site (underlined) at both ends of cpBcX_ *Nsi*I gene which was then cloned into pSTBlue. Following plasmid amplification in *E. coli* DH5 α , the vector was digested with *Nsi*I, and the desired 560-bp DNA fragment was isolated via agarose gel electrophoresis.

The subsequent steps for the construction of the CP library were identical to the method described in section 2.2.2.

Separately, we prepared pET27-PP, a cloning vector derived from pET27 with unique *Pst*I and *Pac*I sites (underlined), as well as three stop codons (bold) in all possible reading frames. The new cleavage sites were introduced by overlap PCR with primer pET27b-PstI-For (5'-CTGCTCCTC GCTGCCAGCCGGCGATGGCCTGCAGATGGATATCGGAATTAATTCG -3') and pET27b-PacI-Rev (5'-GATCTCGAGTTAGTTAGTTAATTAAGCGGCCGCAAGCTTGTCGAC -3'). To prepare the blunt-end vector for the ligation, 5µg of pET27-PP was digested with *Pst*I and *Pac*I sites and recovered by agarose gel electrophoresis. Subsequently, the vector was treated with Klenow polymerase (1 unit/µg of DNA) in the presence of 100 µM dNTP for 15 min at 25°C to remove the 3'-overhangs, followed by QIAquick column purification.

2.2.13 Screening of cpBcx library

The circularly permuted Bcx library was retransformed into the expression host *E. coli* BL21(λDE3) and spread out on LB agar plates containing 30 µg/mL kanamycin. After 12 h incubation at 37°C, colonies were replicated to fresh plates with 30 µg/mL kanamycin. While the master plates were stored at 4°C, the replica plates were grown for 3 h at 37°C prior to overlay with molten agar containing 0.4% agar, 0.4% birchwood xylan (Sigma, St. Louis, MO), 30 µg/mL kanamycin and 1 mM IPTG. Once the overlay agar solidified, plates were incubated at 37°C for another 8 h. To visualize xylanase activity, the agar plates were stained with 0.5% Congo Red solution for 15 min and washed with 1 M NaCl for 30 min (Teather and Wood 1982). Colonies expressing active glycoside hydrolases were identified by formation of clear zone surrounding the host cells. The corresponding positive colonies were picked from master plates and regrown for sequencing analysis with T7-forward primer (5'-TAATACGACTCACTATAGGG-3').

2.3 Results and discussions

2.3.1 Random circular permutation of EchA

Inspired by the success of CALB, we first chose EchA from the α/β hydrolase family as the new study target for random circular permutation. Starting from wild-type *echA*, a flexible 9-amino acid linker (- GGGTSGGGG -) was introduced to bridge the ~ 27 Å distance between the original termini, largely based on our experience with CALB and the general rules discussed in chapter 1. The permuted library genes were cloned into a modified vector pYY101-blunt, allowing for blunt-end ligation of the gene inserts next to a start codon and a stop codon. The DNA library was transformed into *E. coli* TOP10 cells, yielding approximately 5×10^5 colonies (theoretical library size ~ 900). Sequencing analysis of 54 randomly selected members in the naïve library confirmed the unbiased distribution of new termini over the entire protein sequence (Figure 2-6). It was noted that most of the sequences exhibited small amino acid extensions or deletions at their ends (± 10 amino acids), which is actually a phenomenon that has been observed in all the random circular permutation experiments. It is caused by multiple digestions of DNaseI and the fill-in treatment by T4 polymerase intended to repair the nicks and increase ligation efficiency.

Functional EchA variants were identified by colony selection on glycidyl phenyl ether (GPE) plates based on the cell toxicity of epoxides (Reetz and Wang 2006). As shown in Figure 2-7, cells containing wild-type *echA* had the ability to degrade GPE and grew on the selection medium after 2-4 days while the control cells containing empty vector did not grow at high GPE concentrations and grew a lot slower at low GPE concentrations. To reduce selection pressure and meanwhile minimize the background, a GPE concentration of 3 mM was chosen for the CP library selection. Unfortunately, screening of over 5000 colonies at room temperature or 30°C found only 5 functional variants with new termini in close proximity to the native termini.

Specifically, two of variants had new termini located in the linker region and the other three were at nucleotide positions 12, 18 and 837 of the wild-type sequence, respectively.

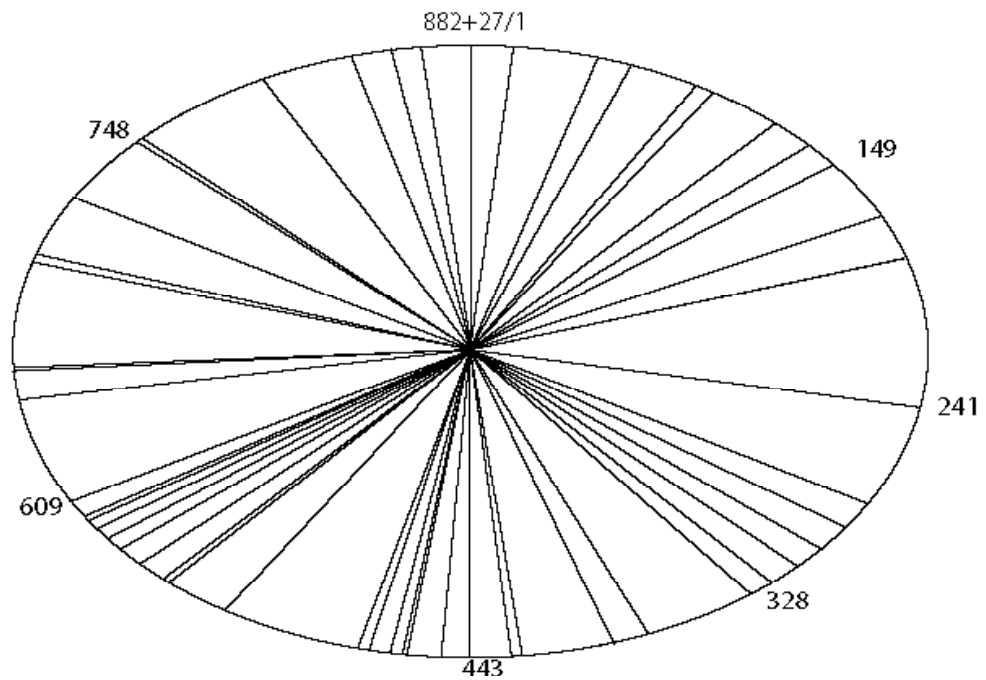
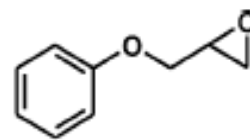
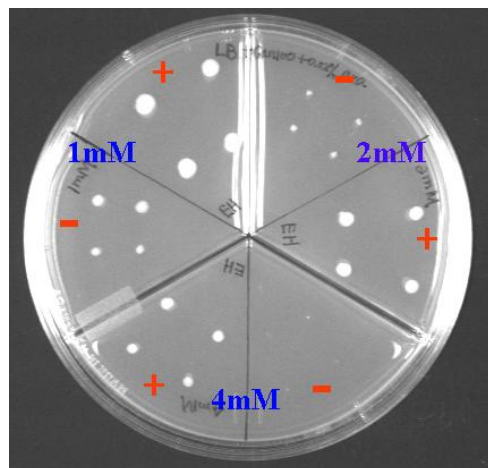


Figure 2-6. Sequencing analysis of selected members from native cpEchA library. The circle represents the circularized *echA* gene (1-882) with the linker (+27). Each line indicates one sequencing data.



Glycidyl phenyl ether

Figure 2-7. Selection of functional epoxide hydrolase permutants using glycidyl phenyl ether (GPE) as the substrate. Three different concentrations (1 mM, 2 mM and 4 mM) of substrate were added to the LB medium containing carbenicillin and arabinose. Cells with (+) or without (-) *echA* were spotted on the plate and incubated for three days at 30°C.

2.3.2 Characterization of rational designed EchA permutants

To understand why functional variants were limited to positions near the native termini, three circular permutants (cpS167, cpL189 and cpP205) were generated by rational design using the same linker used for the CP library. The selected sites are located within the exposed loops of the cap domain (Figure 2-8 A), regions that were shown to be most tolerated towards backbone breakage in the case of CALB. Consistent with the result of EchA library screening, the three variants were not active. Although they were expressed in good yields, all were found in the insoluble fraction upon induction at either 20°C or 37°C (Figure 2-8 B). A possible explanation was that placing the new termini in these positions could interfere with protein folding. Along the same line, the existing linker might not properly fit the termini gap and interferes with protein

folding and topology as reported in other permutation studies (Osuna, Perez-Blancas et al. 2002; Garrett, Mullins et al. 2003). In the case of CALB, the termini distance is around ~ 17 Å and both termini are in flexible loops. Truncation experiments showed that these termini could be even directly connected (Qian, Horton et al. 2009). In contrast, the crystal structure of EchA shows that the original termini are the extension of defined secondary structures and point away from each other, suggesting that any constraint of the termini could affect protein folding and stability. Supporting these observations, we found that the removal of the first 8 residues at the N-terminal of wild-type EchA led to a misfolded mutant.

As discussed in the section 1.3.3 of chapter 1, linker design is a critical step for circular permutation. The length and composition of the peptide linker can have a significant effect on protein activity and stability. Currently, there is not much information for designing linkers to bridge greater distances over 20 Å. Hence, we decided to tackle this problem using rational designed permutants to find suitable linkers for CP of EchA and to provide knowledge for guiding future linker design. In the case of permuted DHFR, it was observed that the strain induced by shorter linkers destabilized the conformation, leading to a flexible, less structured form (Iwakura and Nakamura 1998). In light of this, we tested whether longer linkers could help the solubility of the rational designed permutant. However, extending the initial 9-amino acid linker of cpP205 by one Gly or two Gly did not improve the protein solubility.

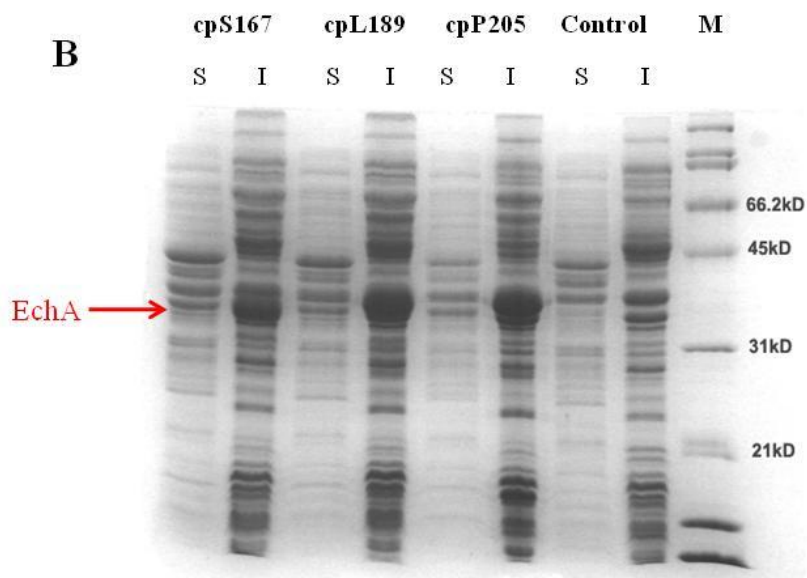
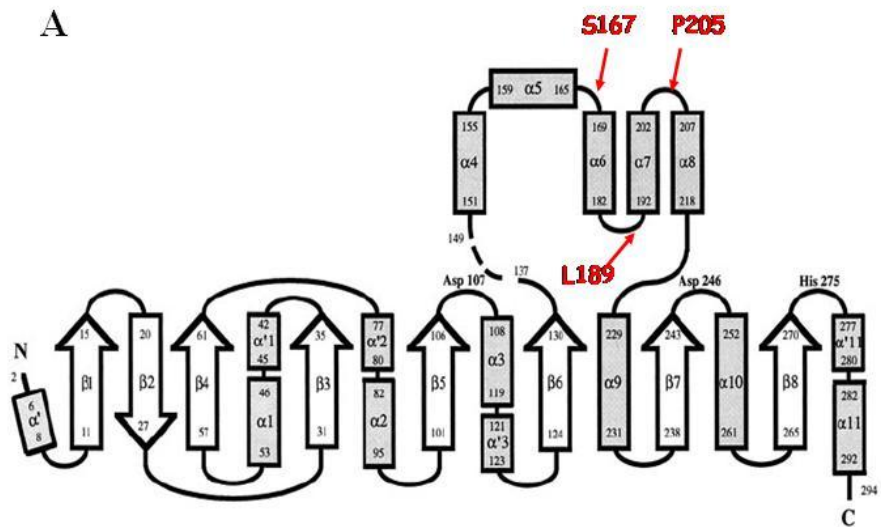


Figure 2-8. Rational designed EchA circular permutants. **(A)** Topology diagram of wild-type EchA. The rational chosen permutation sites (S167, L189 and P205) are indicated by the arrows and labels (red). **(B)** SDS-PAGE of expression of the three rational designed permutants (induced at 20°C for overnight). Red arrow points to the expected molecular weight for cpEchA.

2.3.3 Optimization of EchA linker with combinatorial approaches

Meanwhile, we also realized that although glycine linkers minimize potential strain from native termini topology, they are thermodynamically challenging and tend to burden the protein's overall stability. Accordingly, we randomized the linker sequence of cp189 with the hope of finding linker residues that were able to interact with other residues and therefore help stabilize the added sequence. Five of the positions (underlined) in the initial linker (- GGGTSGGG -) were randomized to any of the 20 possible amino acids. Thr and Ser were conserved as unique *SpeI* restriction site and Gly at each end was kept to facilitate some flexibility.

Considering the possibility that the soluble cp189 may not have the EH activity, green fluorescence protein (GFP) was fused to the C-termini of *echA* as a solubility reporter. As illustrated in Figure 2-9 (A and B), the signal of GFP can only be detected when cpEchA is soluble, and the fluorescence intensity would be proportional to the amount of soluble test protein (Waldo 2003; Cabantous, Pedelacq et al. 2005). To verify this method, the following three proteins: wild-type EchA, cp189, and cp205 were fused to GFP. As expected, cells carrying wild-type EchA showed fluorescence under UV light (FOTO/UV 15 Transilluminator, Fotodyne, Inc.), while cells carrying the two cpEchA variants remain dark (Figure 2-9 C). Using this method, 20 positive colonies were identified after screening of over 5000 colonies from the random linker library. However, some of them had a stop codon in the linker region and thus did not contain the full length *echA*. SDS-PAGE and western-blot analysis of the remaining positive hits revealed that the total expressed fusion proteins were low and a majority were found in the insoluble fraction. Therefore, we did not continue with these variants.

Expanding the search for a functional linker, we constructed a second library with variable linker length and amino acid compositions. The library was generated in two steps. The first step involved creating a library of a 30-amino acid linker in which 24 positions (indicated as X) were encoded by NNK (- XXXXXXXXXXXXXXXSGTSGSXXXXXXXXXXXX -). The second step utilized incremental truncation to gradually shorten the linker to different degrees leading to variations in

length. The final library was screened with a more powerful technique, FACS (fluorescence activated cell sorting). As seen in Figure 2-10, before sorting, the whole library population had lower fluorescence when compared with the negative control, cells carrying cp189-GFP plasmid. This could suggest most of the library members had incorrect reading frames or unfavourable linkers that led to even less fluorescence than the background of cp189-GFP produced. Nevertheless, after two rounds of enrichments, the fluorescence signal showed an obvious shift (purple curve). Around 300 colonies were recovered and the 15 colonies that showed green fluorescence under UV lamp were further characterized by DNA sequencing and protein expression. Unfortunately, 10 of them did not have the full length fusion protein, missing at least 10 amino acids which were caused by the incremental truncation process. SDS-PAGE and western-blot analysis of the expressed cells also did not observe significant soluble fusion protein.

In summary, we are still not convinced that EchA is intolerant to circular permutation within the interior of protein sequence. Although we were unable to find suitable linkers from the combinatorial libraries, we realized that it could be the consequence of the following three pitfalls. First of all, we relied on rational designed permutants for linker optimization, but the sites we chose were probably not good candidates if the connectivity of cap domain is important for cap domain folding. Secondly, while GFP fusion facilitates the screening, attaching another protein to the permutant definitely burdened the folding process. Additionally, there is also the problem with the library size. The average 10^6 library we obtained only represents a small portion of the theoretical random linker libraries.

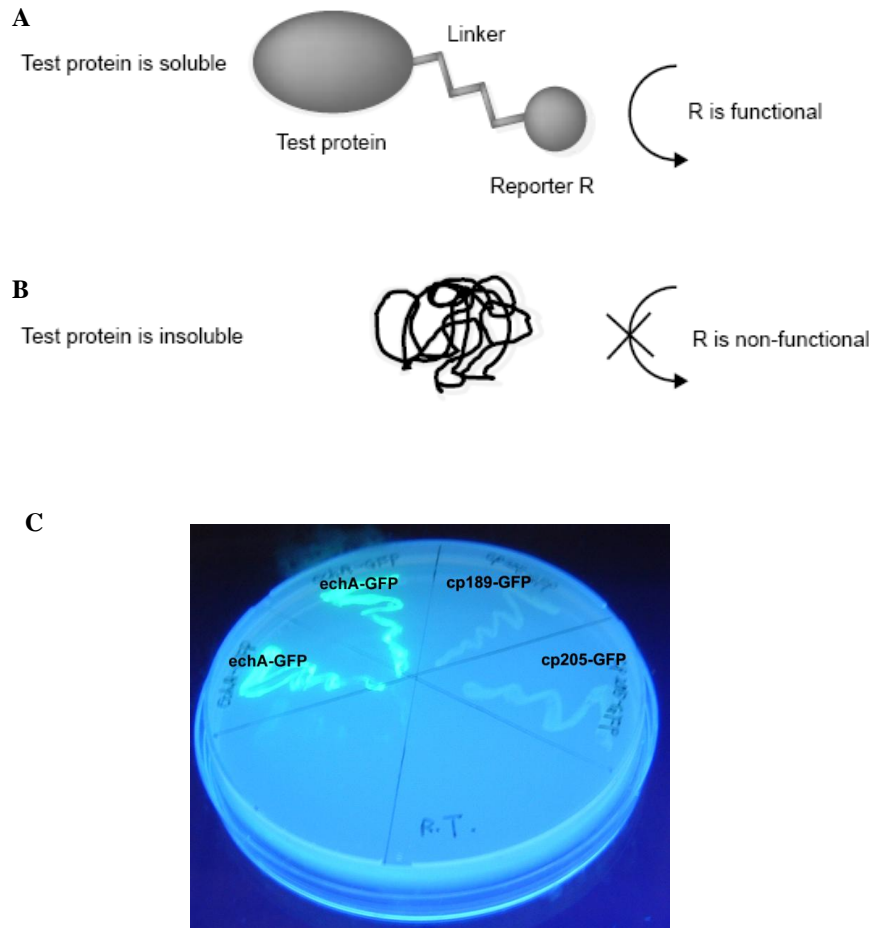


Figure 2-9. The use of GFP as a solubility reporter for screening of random linker EchA libraries. **(A)** When test protein (cp189 in our case) is soluble, the reporter is correctly folded and functional. **(B)** When test protein is insoluble, the reporter is misfolded as well and gives no signal. (Waldo 2003) **(C)** Under UV Transilluminator (312 nm), cells containing *echA*-GFP glow, while cells containing cp189-GFP and cp205-GFP remain dark.

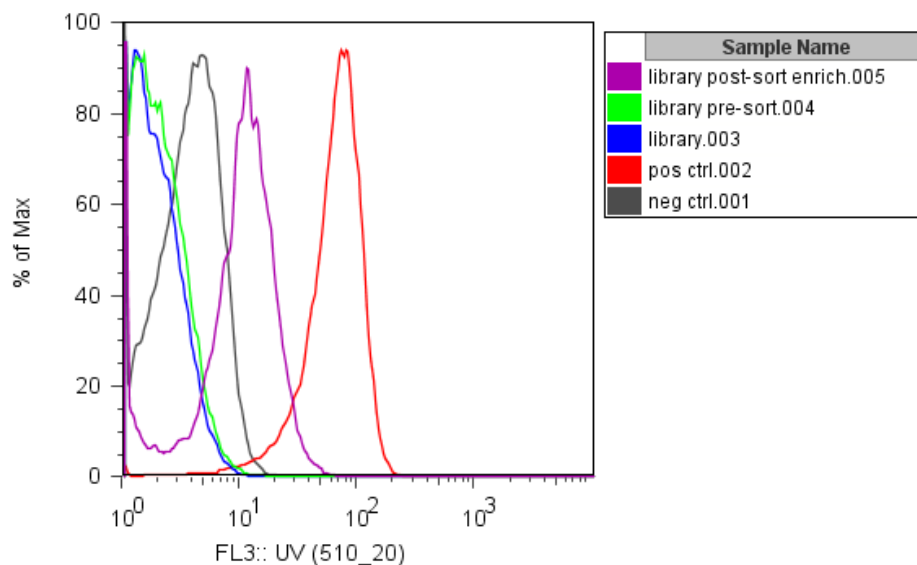


Figure 2-10. Screening of cp189-RLIT library using FACS. Cells were induced with 0.02% arabinose for overnight at 20°C and subjected to two rounds of sorting. Cells carrying *echA*-GFP plasmid (red curve) and carrying cp189-GFP plasmid (gray curve) were used as positive and negative controls. Initially, the linker library population (blue curve) had lower fluorescence than the negative control. After two rounds of enrichments, the fluorescence signal showed an obvious shift (purple curve).

2.3.4 Random circular permutation of cutinase

After the failure with epoxide hydrolase, we selected another α/β hydrolase, cutinase, for our circular permutation study. As opposed to CALB and EchA, cutinase is a small enzyme without a cap domain. The termini distance is around 20 Å termini, shorter than that of EchA. We designed an 8-amino acid linker - GAGTSGAG - to bridge this distance and generated the random CP library. Sequencing colonies from the naïve library indicated an unbiased distribution of new termini across the entire protein sequence (Figure 2-11, inner circle, red lines). The permuted proteins were expressed in *P. pastoris* under the control of the AOX1 promoter and secreted into the medium led by α -factor X signal sequence (Figure 2- 4), which enables screening on a larger

scale. The hydrolytic activity of cutinase variants toward polycaprolactone or tributyrin was visualized by the formation of a halo around colonies (Figure 2-12). Screening of 7000 colonies identified 22 unique new termini (Figure 2-11, middle circle, green lines). However, the new termini of these permutants were again skewed towards the native termini region.

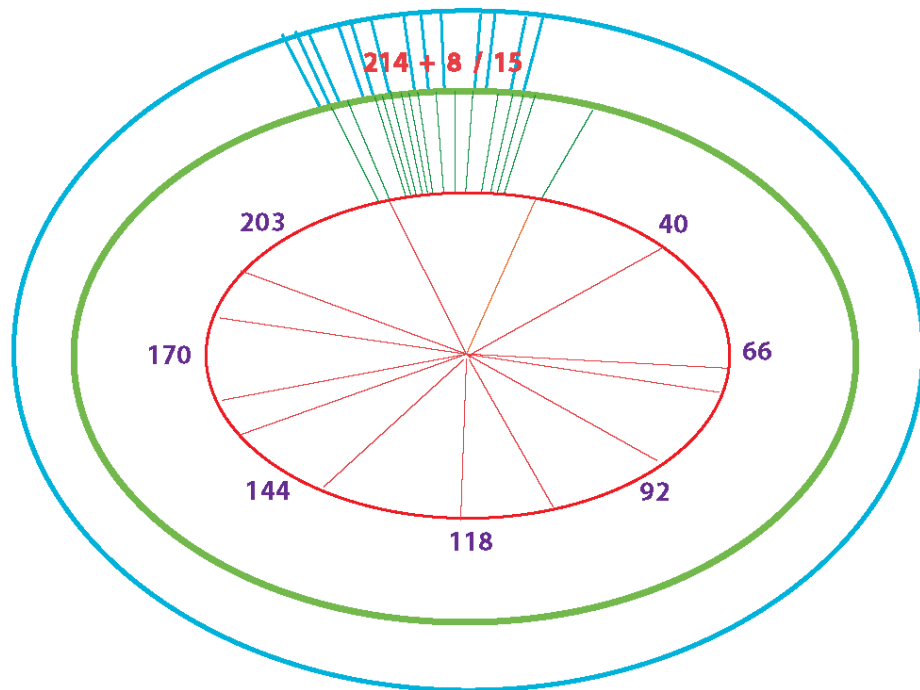


Figure 2-11. Random circular permutation of cutinase. Circles represent circularized amino acid sequence (15-214) and linker sequence (+8). Red lines represent the sequencing of randomly picked colonies from the native library. Green and blue lines indicate active variants identified from random CP library and linker library (CP-RL-IT lib), respectively.

2.3.5 Optimization of cutinase linker with combinatorial approaches

As in the case of EchA, we believe that the flexible linker was not favorable and should be optimized in order to obtain permutants with new termini located interior of the protein sequence. To avoid the similar risk of starting with an inactive permutant, an active permutant cpC31 from CP cutinase library, was chosen as the template for linker optimization based on two reasons. The new termini of cpC31 are closest to the interior of the protein among the 22 identified active permutants, and it only produces a halo at low screening temperatures. Our characterization also showed that this permutant has poor stability and solubility. We intended to improve this permutant through linker optimization and screening at elevated temperatures.

To do so, six positions (X) of the initial 8-amino acid linker (- XXXTSXXX-) were randomized and the length was gradually shortened by incremental truncation. We did not explore longer linker as in the case of EchA since we believe 8 amino acids is enough for the 20 Å distance and extending the length will significantly increase the theoretical library size. The random linker library was generated by adopting the same procedure for EchA and screened on tributyrin-plates at 22-30°C. Ten positive variants were identified from screening over 5000 colonies and sequencing showed that 8 of them still had 8-residue linkers, while 2 variants did not have a linker. However, we were unable to reveal any pattern for the amino acids that could benefit the linker design from this study.

2.3.6 CP of cutinase with random linkers

As an alternative to modify an existing linker, we decided to perform circular permutation using random linkers instead of a defined linker sequence. Following the procedures described in the method section (2.2.11), we obtained approximately $\sim 10^6$ cfu for the final CP library and $\sim 13,000$ colonies were screened with the tributyrin plate assay. Analysis of the active variants yielded 13 unique new termini that still focused around the original termini region (Figure 2-11, outer circle, blue lines).

Although it is possible we missed the suitable linkers from our libraries due to the number of colonies we screened were not comparable to the theoretical library sizes, maybe cutinase is just not tolerant to termini relocation to its interior region because of its minimal structure.

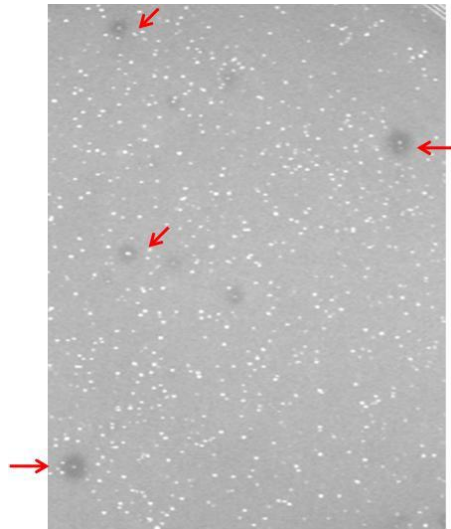


Figure 2-12. Screening of cutinase libraries on tributyrin agar plate. Transformants were plated and incubated at 30°C or room temperature for 3-5 days. Red arrows indicate some of the halos produced by active variants.

2.3.7 Random circular permutation of Bcx

To investigate whether circular permutation could extend beyond the α/β hydrolase family, we studied xylanase from *Bacillus circulans* (Bcx) in collaboration with the laboratories of Dr. Steve Withers and Dr. Lawrence McIntosh at University of British Columbia (Canada). Initially, three circular permutants with new termini located in exposed surface loops were generated by rational design. Following expression and purification of these proteins, steady-state kinetics revealed only minor differences in catalytic efficiency for hydrolysis of 2, 5-dinitrophenyl β -xylobioside (DNPX2) (0.3-1.3 fold) compared to wild-type enzyme. These results verified that Bcx could be circularly permuted (Reitinger, Yu et al. 2010). Subsequently, we created a combinatorial library of randomly circular permuted Bcx variants (2×10^5 colonies) using a double glycine linker to join the 5 Å distance between the original termini. DNA sequencing analysis of 32 randomly chosen members confirmed an even distribution of new termini across the entire protein sequence (Figure 2-13, red lines). The screening of glycoside hydrolase activity was facilitated by the *pelB* leader sequence in the vector which allowed protein secretion to the periplasmic space. Functional variants against xylan were identified using a Congo Red overlay assay (Figure 2-14). The examination of 3000 colonies yielded 59 active candidates, in which 35 had unique new termini in positions different from the wild type. Our data indicated that Bcx tolerates permutations in about 20% of positions across the whole length of the sequence. When mapped onto the wild-type Bcx structure, a large number of the new start sites occur in exposed loops. Surprisingly, the new termini of numerous permutants are also located within β -strands or even near the active site. The only region that active permutants were not found was in the single α -helix, which was reported to be important for protein stability. (Davoodi, Wakarchuk et al. 2007)

Additionally, our collaborators carried out enzymatic, structural, and dynamic characterization of a set of active permutants. Several exhibited comparable or improved activity over wild type Bcx. The best permutant, cpN35 showed a 4-fold rate enhancement for the

hydrolysis of DNPX2. NMR spectroscopy or X-ray crystallography on four permutants revealed that circular permutation had little impact on the overall structures. Changes were largely limited to increased local mobility near the glycine linker, as well as in the new termini region (Reitinger, Yu et al. 2010). This permuted library of Bcx did not only provide further mechanistic understanding, but it also provided excellent potential starting points for the specific incorporation of unnatural or labeled amino acids into the protein by semi-synthetic approaches.

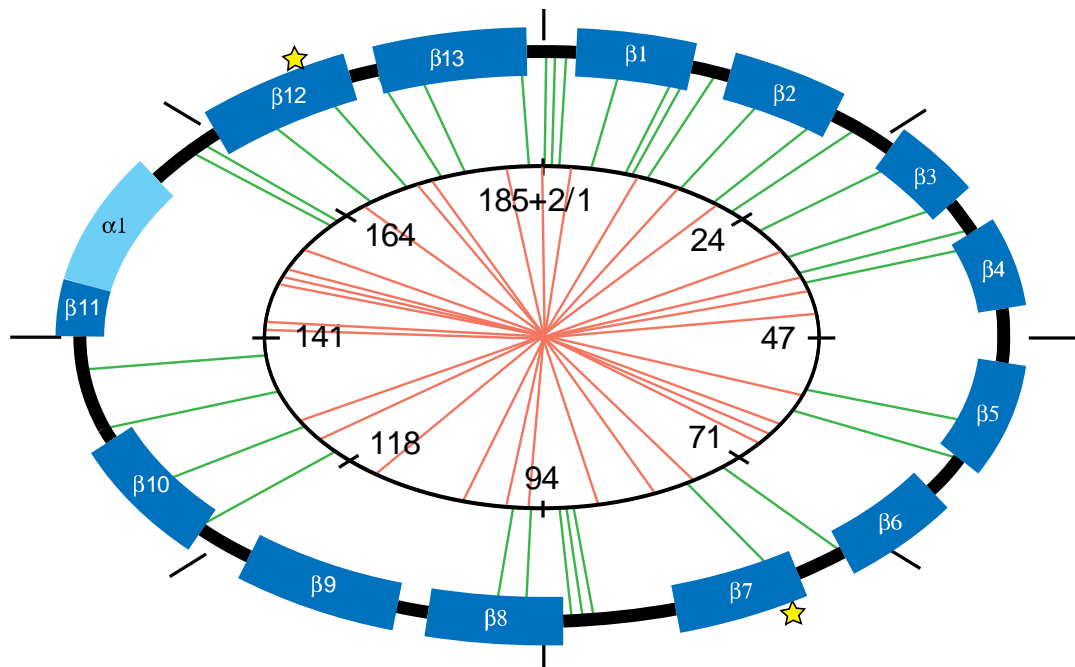


Figure 2-13. Circular permutation of Bcx. The unbiased library distributed was confirmed by the sequencing of 32 randomly selected library members (red lines of inner circle). New identify termini sites are indicated by green lines. Secondary structures are shown in shades of blue; the two catalytic residues are marked by yellow stars.

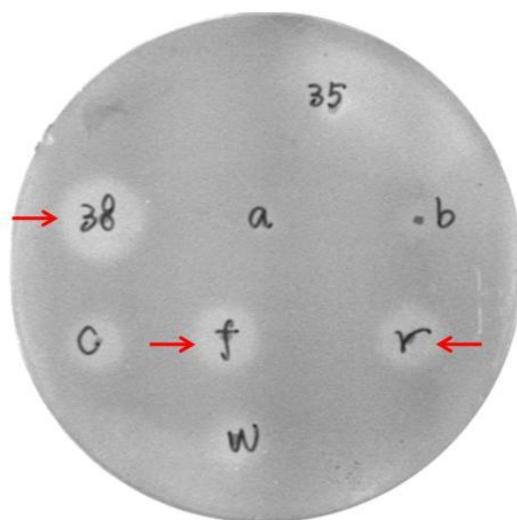


Figure 2-14. Screening of Bcx libraries with Congo red plate assay. Permuted Bcx enzymes were secreted into the medium containing xylan and functional permutants were identified by visualized halo (red arrows) after congo-red staining.

2.4 Conclusions

In conclusion, we applied circular permutation to three different enzymes that have important industrial applications. Our success with Bcx once again suggests that random circular permutation can reveal many unexpected permutation sites in different protein folds. Although we were not successful with EchA and cutinase, we learned the critic role of the linker and tested strategies that could be used to search for suitable linkers. In the future, in addition to experimental approaches, computational design would be very helpful for the linker design.

Chapter 3 Improved Triglyceride Transesterification by Circular Permuted *Candida antarctica* Lipase B

Published by Ying Yu, Stefan Lutz

(Biotechnology & Bioengineering, 2009, 105, 44-50)

This chapter is the reformatted published paper. The permission to use the full article for this dissertation is granted by the publisher John Wiley and Sons. License Number 2531960741755

3.1 Introduction

Enzymes are remarkable catalysts, often exhibiting superb selectivity and specificity for their natural substrates, as well as substantially accelerating chemical reactions under environmentally benign conditions. Lipases (EC 3.1.1.3) represent a particularly popular group of biocatalysts with applications in asymmetric synthesis but also industrial processes (Berglund 2001; Bornscheuer and Kazlauskas 2006; Salameh and Wiegel 2007). In the latter category, the production of biodiesel via transesterification of triglycerides with short-chain alcohols has been an active area of research. The conversion of locally generated waste products such as spent cooking oils and animal fat into biodiesel represents a renewable and eco-friendly alternative to petroleum-based diesel fuel. From an engineering perspective, the use of lipases for fuel production would be advantageous over the established process, chemical alcoholysis with NaOH or KOH, as the enzymatic route tolerates more diverse feedstock, minimizes side-reactions which complicate downstream processing, and conserves energy (Akoh 2007; Ranganathan 2008). A

number of enzymes have successfully been tested for biodiesel production including lipase B from *Candida antarctica* (CALB) (Jegannathan, Abang et al. 2008; Fjerbaek 2009). Nevertheless, high costs for the enzymes have made the process economically impractical. Higher productivity represents one approach for reducing costs and can be accomplished through the optimization of reaction conditions and the identification of better catalyst. The search for new enzymes in nature has been a promising strategy while engineering biocatalysts by rational design and directed evolution has produced enzymes with improved properties (Qian, Fields et al. 2007).

Recently, we employed a technique called circular permutation to explore the effects of termini relocation on catalytic performance of CALB (Qian and Lutz 2005). During the process of circular permutation, the natural N and C-termini of a protein are covalently linked by a short peptide and new termini are created elsewhere within the original sequence. As a result, the amino acid composition of the protein remains unchanged, yet the primary sequence is rearranged. Upon random circular permutation of CALB, we found that almost 20% of permuted enzymes retained ester hydrolase activity. More importantly, some permutants showed increases in hydrolytic activity of 11- and 175-fold for *p*-nitrophenol butyrate (*p*-NB) and 6, 8-difluoro-4-methylumbelliferyl octanoate (DiFMU octanoate), respectively (Qian, Fields et al. 2007). As part of our ongoing investigation of the effects of circular permutation on the structure and function of CALB, we were interested in exploring the performance of our engineered lipases in transesterification and biodiesel production.

In this study, we have focused on cp283, a circular permuted variant of CALB whose new N- and C-termini are located at positions 283 and 282 of the wild-type CALB sequence, respectively. In the native enzyme, these two amino acids are part of an extended alanine-rich helix in the enzyme's cap domain which forms part of the active site binding pocket. Although we do currently not fully understand the structural consequences of circular permutation, changes in local backbone flexibility and active site accessibility are believed to benefit the catalytic properties of the enzyme. Comparing the performance of cp283 with its parental CALB, our

transesterification experiments with individual triglycerides, as well as various vegetable oils in combination with different acyl acceptors show that the engineered catalyst consistently outperforms the native enzyme, making it a promising candidate for industrial applications including the production of biodiesel.

3.2 Materials and methods

3.2.1 Materials

Immobilized lipase B from *C. antarctica* (wild-type and circular permuted enzyme) was prepared as described (Qian, Fields et al. 2007). As enzyme loadings on resin vary among different enzymes and between batches, the amount of active lipase for each sample preparation was determined by active site titration (Qian, Fields et al. 2007). Paranitrophenolate esters and triglyceride (TAG) substrates (*p*-nitrophenyl butyrate (*p*-NB), *p*-nitrophenyl octanoate (*p*-NO), tributyrin, tripalmitin, and triolein), as well as analytical standards (butyl butyrate, butyl oleate) are commercially available from Sigma-Aldrich (St. Louis, MO) while 6,8-difluoro-4-methylumbelliferyl octanoate (DiFMU octanoate) and the reference standard 6,8-difluoro-7-hydroxy-4-methylcoumarin (DiFMU) were purchased from Invitrogen (Carlsbad, CA). Canola oil and corn oil were purchased at a local supermarket. The spent cooking oil was a gift from a local restaurant. All reagents and solvents were of analytical grade and were dried over molecular sieves (3Å) prior to use.

3.2.2 Kinetic analysis of lipases with chromogenic and fluorescence substrates

Lipase activity was determined by measuring the initial hydrolysis rate of *p*-NB, *p*-NO, and DiFMU octanoate at room temperature on a Synergy-HT microtiterplate reader (Bio-Tek Instruments, Winooski, VT). Stock solutions of *p*-NB (200 mM) and *p*-NO (200 mM), as well as DiFMU octanoate (5 mM) were prepared in DMSO. *p*-Nitrophenolate formation over a substrate range of 0–1.2 mM for *p*-NB and 0–120 mM for *p*-NO was measured in 50 mM K-phosphate

buffer (pH 7.5) at 400 nm (ϵ for p-nitrophenolate = $13,260\text{M}^{-1}\text{cm}^{-1}$). The rate of DiFMU octanoate hydrolysis was determined by measuring the DiFMU formation over a substrate range of 0–12 mM in 50 mM K-phosphate buffer (pH 7.0) at an excitation/emission wavelength of 360/460 nm. All experiments were repeated at least three times. Kinetic constants were calculated by fitting the initial rates to the Michaelis–Menten equation using Origin7 (OriginLab, Northampton, MA).

3.2.3 Kinetic analysis of transesterification reaction

The rates of triglyceride transesterification with 1-butanol were performed in 2-mL reaction volume containing enzyme resin (10–50 mg). Product formation was monitored by GC analysis. Specifically, reactions with tributyrin and triolein were performed in cyclohexane at room temperature while tripalmitin was tested at 50°C due to low substrate solubility. Substrate concentration was varied from 50 mM to 1 M. After 30 min of preincubation, the reaction was initiated by adding 1-butanol (final concentration = 1 M). At least five samples were taken over a period of 1–5 min; this limited overall substrate conversions to <5%. The samples were analyzed by gas chromatography (Agilent 6850 series instrument, connected to a HP-INNOWax capillary column (0.32mm \times 30 m, Agilent, Santa Clara, CA) and a flame ionized detector). For tributyrin, the oven temperature was kept at 50°C for 3 min and raised to 160°C at 218°C /min, held for 5 min. The injector and detector temperatures were set at 250 and 300°C. The retention time for butyl butyrate was 6.0 min. For tripalmitin and triolein, the oven temperature was kept at 210°C for 10 min and raised to 250°C at 15°C /min, held for 5 min. The injector and detector temperatures were set at 250 and 300°C. The retention times for butyl palmitate (16:0) and butyl oleate (18:1) were 4.4 and 7.4 min, respectively.

3.2.4 Trans and interesterification of vegetable oil

Trans and interesterification experiments of complex triglycerides (canola and corn oil, as well as spent cooking oil) with 1-butanol and ethyl acetate were carried at room temperature with

continuous stirring. Specifically, 5 g of oil (5.6 mmol; estimated MW: 885 g/mol) were mixed with enzyme resin (30 – 400 mg) and 1-butanol (1.6 mL, 17 mmol, threefold molar excess) or ethyl acetate (3.6 mL, 36 mmol, six fold molar excess). Samples were drawn from the reaction mixture over the indicated time period and analyzed by GC (system description see above). For the butyl esters, the same protocol as for tripalmitin and triolein was used. The retention times for butyl linoleate (18:2) and butyl linolenate (18:3) were 8.3 and 9.8 min, respectively. For ethyl esters, the oven temperature was kept at 120°C for 2 min and raised to 200°C at 30°C /min, held for 6 min, and finally raised to 250°C at 25°C /min and maintained for 6 min. The injector and detector temperatures were set at 250 and 300°C. The retention times for ethyl palmitate (16:0), ethyl oleate (18:1), ethyl linoleate (18:2), and ethyl linolenate (18:3) were 7.1, 9.3, 10.0, and 11.0 min, respectively.

3.3 Results and Discussions

3.3.1 Impact of acyl and alcohol moieties of substrates on the catalytic performance of cp283

The nature of the binding interactions between an ester's acyl and alcohol moieties and the active site binding pocket of a lipase critically affects enzyme performance. The perturbation of these interactions as a consequence of our enzyme engineering by circular permutation could potentially contribute towards the observed rate enhancements for cp283, yet is only poorly understood. To aid in our studies of various triglyceride substrates for cp283, we initially evaluated the influence of a substrate's acyl (butyrate vs. octanoate) and alcohol moieties (*p*-nitrophenolate vs. methylcoumarin) on catalytic performance by conducting additional steady-state kinetic experiments with *p*-nitrophenyl octanoate (*p*-NO) (Table 3-1). As a structural intermediate of the two previous substrates, *p*-NO shares the smaller alcohol moiety with *p*-NB and the larger acyl group of DiFMU octanoate. Overall, the kinetic data for *p*-NO showed a 30-

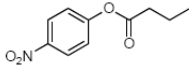
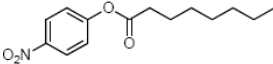
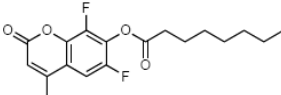
fold increase in the catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) of cp283 over the wild-type enzyme. Consistent with the results for the two other reference substrates, the increase is largely due to higher turnover rates of cp283 while the apparent binding constant remains unchanged.

When comparing the kinetic performances of CALB and cp283 with *p*-NB and *p*-NO, two substrates that distinguish themselves only by the size of their acyl portion (C4 to C8), we observed a drop in the apparent binding constants (K_{M}) by 6- and 15-fold for cp283 and WT-CALB, respectively. The lower K_{M} values are likely a reflection of the additional favorable binding interactions between the longer octanoate moiety and the enzymes. In contrast, the observed turnover rates (k_{cat}) showed opposite trends, declining twofold for the wild-type enzyme but increasing fourfold for cp283 which could be indicative of changes in the rate-determining steps of the two enzymes. Nevertheless, the precise nature of the slow steps in their respective catalytic cycles cannot be deduced from our measurements but would require presteady-state experiments. Separately, we investigated the impact of circular permutation on the alcohol binding site with *p*-NO and DiFMU octanoate, two substrates with identical acyl moieties but different alcoholic substituents. Here, the changes of the kinetic parameters for both enzymes were more consistent albeit varied in their magnitude. While the K_{M} values dropped another 10-fold, probably due to additional contacts with the substrate's alcohol moiety, the rates of hydrolysis decreased by 10- and 60-fold for cp283 and WT-CALB, respectively.

Concerning the original question regarding the influence of the substrate's acyl and alcohol moieties on the observed rate enhancements, we conclude from our kinetic data that both portions of the three esters tested in this study contribute similarly to the net activity gain in cp283. While the substitution of the acyl portion (*p*-NB vs. *p*-NO) resulted in an eightfold difference in rate change ($\Delta\Delta k_{\text{cat}} = [k_{\text{cat}}(p\text{-NO})/k_{\text{cat}}(p\text{-NB})\text{cp283}]/[k_{\text{cat}}(p\text{-NO})/k_{\text{cat}}(p\text{-NB})\text{WT}]$) in favor of the longer alkyl chains with cp283, the exchange of the alcohol group (*p*-NO vs. DiFMU octanoate) translated into a $\Delta\Delta k_{\text{cat}}$ of 6 in favor of the bulkier methylcoumarin with cp283. Furthermore, we noticed significantly less fluctuation in the k_{cat} values for the permuted enzyme (~10-fold)

across all three substrates compared to WT-CALB (~100-fold). The observation can be rationalized by an acceleration or shift in the rate-determining step for the engineered lipase relative to the parental enzyme. Independent evidence for such changes in the reaction cycle is provided by the observed structural differences in the active site binding pocket of the recently solved crystal structure of a CALB permutant (Qian, Horton et al. 2009). The termini relocation to position 282/283 creates a more accessible active site which could affect the rates of substrate binding and product release, as well as influence the protein's dynamics associated with conformational changes as part of the reaction cycle.

Table 3-1. Apparent kinetic constants for WT and cp283 with different ester substrates.

Substrate	Enzyme	K_M [μM]	k_{cat} [min^{-1}]	k_{cat}/K_M [$\text{min}^{-1}\mu\text{M}^{-1}$]	Relative catalytic efficiency
 <i>p</i> -nitrophenyl butyrate	WT	309±43	236±12	0.8	11
	cp283	130±17	1100±43	8.5	
 <i>p</i> -nitrophenyl octanoate	WT	20.4±4.2	128±10	6.3	30
	cp283	22.6±4.2	4200±267	186	
 DiFMU octanoate *	WT	2.6 ± 0.3	2 ± 0.1	0.8	175
	cp283	2.4 ± 0.4	340 ± 17	140	

WT = wild type CALB, cp283 = circular permutated CALB whose N terminus starts at amino acid 283 of the wild type sequence. Asterisk marks previously reported data (Qian 2007). Relative catalytic efficiency: $[k_{\text{cat}}/K_M (\text{cp283})] / [k_{\text{cat}}/K_M (\text{WT})]$.

3.3.2 Transesterification of triglyceride

Next, we expanded our study of cp283 to the transesterification reaction of three triglycerides (tributylin (C4), tripalmitin (C16:0), and triolein (18:1)) with 1-butanol. Butanol was selected over frequently used short-chain alcohols such as methanol and ethanol due to its higher solubility in organic solvents, facilitating a homogeneous reaction mixture for the determination of accurate kinetic parameters. It also reduces a common problem with enzyme denaturation at high concentrations of short-chain alcohols. To ensure that 1-butanol was not a limited factor in the kinetic analysis of the triglycerides, the alcohol concentration in the reaction mixture was kept at 1M, approximately 20-fold above its K_M value.

The kinetic data for the transesterification of triglycerides by wild-type CALB and cp283 are summarized in Table 2-2. For individual substrates, the circular permutant cp283 showed between 2.6- and 9-fold higher catalytic efficiency compared to wild-type enzyme. These gains in performance can be attributed entirely to faster turnover rates as the results showed little to no change in the K_M values with either enzyme. Furthermore, the comparison of k_{cat} values across the three substrates revealed a steady decline in activity with increasing length of the fatty acid chain. While the turnover rate for wild-type CALB dropped almost fivefold from C4 to C16:0 and another fourfold from C16:0 to C18:1, cp283's loss of activity is notably smaller, decreasing only approximately twofold per step. Overall, these findings are consistent with our results for the chromogenic and fluorogenic substrates discussed above.

Table 3-2. Apparent kinetic constants for transesterification of triglycerides by WT and cp283.

	Enzyme	K_M (TAG) [mM]	K_M (1-butanol) [mM]	k_{cat} [s ⁻¹]	k_{cat}/K_M (TAG) [s ⁻¹ M ⁻¹]	Relative catalytic efficiency
Tributyryn (C4:0)	WT	677±63	58±7	81±4	120	2.6
	cp283	659±47	42±5	206±7	313	
Tripalmitin (C16:0)	WT	526±51	nd	16.9±0.8	32	6
	cp283	552±88	nd	108±9	196	
Triolein (C18:1)	WT	633±98	nd	4.2±0.3	6.6	9
	cp283	685±79	nd	41.3±2.8	60.3	

nd = not determined, WT = wild type CALB, cp283 = circular permutated CALB whose N terminus starts at amino acid 283 of the wild type sequence. Asterisk marks previously reported data (Qian 2007). Relative catalytic efficiency: $[k_{cat}/K_M(\text{cp283})] / [k_{cat}/K_M(\text{WT})]$.

Table 3-3. Composition of vegetable oils.

Oil	Fatty acid composition (wt %)						Water content (wt %)
	16:0	18:0	18:1	18:2	18:3	Others	
Canola oil	4.7	2.1	62.2	21.1	9.7	--	0.5
Corn oil	11.4	1.7	28.6	57.3	0.9	--	0.3
Spent cooking oil	12.4	4.2	39.1	34.2	4.1	5.9	2.5

3.3.3 Alcoholysis of complex triglyceride mixture and oils

To further evaluate the performance of cp283 for biodiesel production, three vegetable oils (canola oil, corn oil, and spent cooking oil) were chosen for the transesterification with 1-butanol. The compositions of the different oils are summarized in Table 3-3. In contrast to the kinetic measurements with the triglycerides, these experiments focused solely on substrate conversion and were therefore conducted in solvent-free system. The latter conditions are preferred from a practical standpoint as they simplify product recovery.

Alcoholysis reactions with the three oil samples were run with relatively low enzyme loadings of 0.01% w/w (based on oil weight) at a molar ratio of 3:1 (1-butanol to oil). The course of the individual reactions was followed by GC analysis. Sample aliquots were taken over a 48-h time period as shown in Figure 3-1. Control experiments confirmed no detectable catalytic activity of the resin alone (Figure 3-1A). Consistent with previous reports for lipase-catalyzed transesterification, no significant differences in the overall reaction profile were noted despite of the distinct composition of the vegetable oils (Hernandez-Martin and Otero 2008). For canola oil, the transesterification with 1-butanol over 48 h reaches ~85% conversion with cp283 while the wild-type CALB shows approximately 20 – 25% conversion (Figure 3-1A). When comparing the initial rates for both enzymes, based on 10% and 20% feedstock conversion, the engineered lipase showed a six- to sevenfold higher performance over the native enzyme. Similar data were obtained for the conversion of corn oil with 1-butanol (Figure 3-1B). After 48 h, cp283 reached ~90% conversion, compared to 30% product formation with wild-type CALB. In terms of initial rates, cp283 performed approximately five- to six fold faster than CALB with this substrate. Beyond these pure vegetable oils, we tested the use of our engineered lipase with cheaper feedstock such as spent cooking oil (Figure 3-1D). Using a filtered oil sample from a local restaurant without further purification under the reaction conditions outlined above, the transesterification with 1-butanol and cp283 proceeded to about 70% conversion in 48 h while the native enzyme reached roughly 20% product formation. Based on the initial rates, the circular

permutated lipase exhibits approximately fivefold higher activity than the wild-type enzyme. We attribute the observed drop in total conversion from 90% for the pure vegetable oils to ~70% for the spent cooking oil to impurities which lower the overall content of triglycerides in the latter. In summary, our data indicate that the circular permutated lipase consistently outperformed the native CALB in converting three vegetable oils with different fatty acid composition to the corresponding butyl esters.

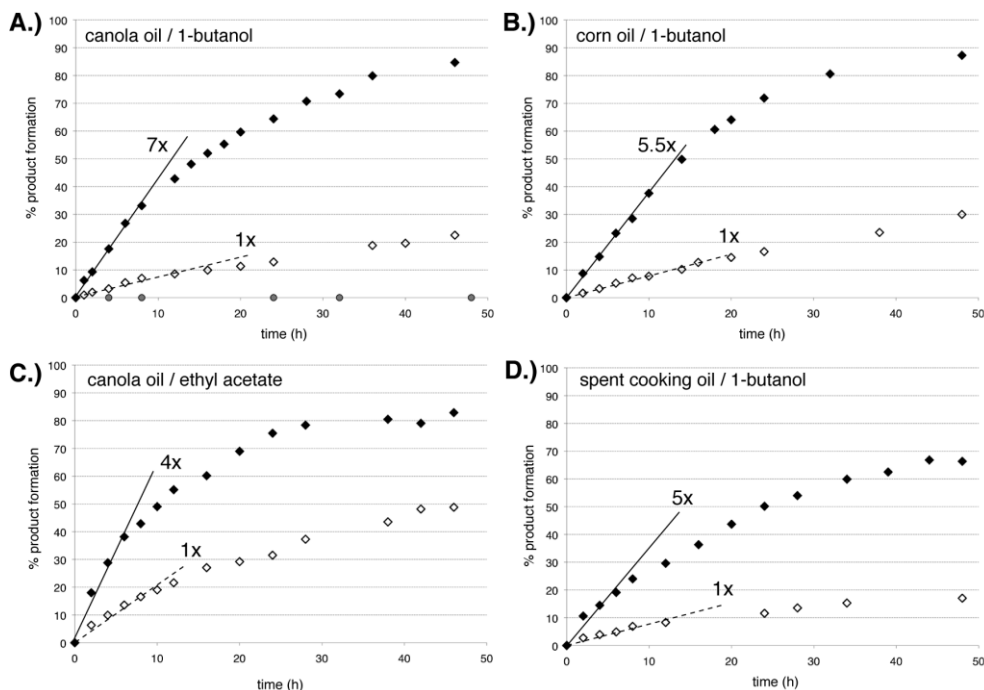


Figure 3-1. Time course of transesterification reactions for WT-CALB (◇) and cp283, the circular permutated CALB (◆) with different vegetable oils and acyl acceptors. **A:** Reactions of canola oil and 1-butanol (1:3 molar ratio; 0.01% w/w enzyme based on oil weight). A control experiment with resin-only (●) (10% w/w resin based on oil weight). **B:** Reaction of corn oil and 1-butanol (1:3 molar ratio; 0.01% w/w enzyme based on oil weight). **C:** Reaction of canola oil and ethyl acetate (1:6 molar ratio; 0.01% w/w enzyme based on oil weight). **D:** Reaction of spent cooking oil and 1-butanol (1:3 molar ratio; 0.01% w/w enzyme based on oil weight). The initial rate of conversion based on 10% and 20% product formation for cp283 relative to WT-CALB is indicated.

3.3.4 Interesterification with ethyl acetate

Short chain linear alcohols used as acyl acceptor in the transesterification have been found to be toxic to immobilized lipase. The activity loss is inversely proportional to the carbon chain length (Chen 2003) and can be controlled through stepwise addition of sub-stoichiometric amounts of alcohol. Alternatively, the use of methyl acetate or ethyl acetate as acyl acceptor for the interesterification of vegetable oils has been reported (Du 2004). This strategy not only eliminates the risk of enzyme deactivation by the free alcohol, but also prevents glycerol accumulation by derivatizing the side product to its corresponding acetate, thereby avoiding losses in enzyme activity due to glycerol absorb onto the resin surface (Xu 2005; Modi 2007).

(Note: We carried out the transesterification of canola oil with ethanol as acyl acceptor. The reaction was 4-fold slower for both wild-type and cp283 when compared with the reaction using 1-butanol, presumably due to the alcohol toxicity. This experiment was not included in the published paper.)

The performance of cp283 as a biocatalyst for interesterification was tested with canola oil and ethyl acetate. Reactions were run with enzyme loadings of 0.01% w/w (based on oil weight) at molar ratios of 6:1 (ethyl acetate to oil) and product formation was monitored by GC analysis of reaction samples over 48 h (Figure 3-1C). Compared to the corresponding transesterification reaction with 1-butanol, the initial conversion rate of canola oil with ethyl acetate by cp283 is approximately 1.5 times faster. Nevertheless, rates for wild-type CALB are also notably higher, lessening the gains of the engineered lipase over the native enzyme to about fourfold. After 30 h, cp283 has reached ~80% product formation while wild-type CALB shows about 40% ethyl ester formation. Under the present reaction conditions, interesterification levels off around 85% for the former while wild-type enzymes reaches ~50% product formation after 48 h. The results demonstrate that the observed improved catalytic performance of cp283 is not limited to transesterification reactions although some of the gains might be attributed to the greater molar excess of ethyl acetate compared to 1-butanol.

3.3.5 Recycling of immobilized enzyme

From an economic standpoint, a biocatalyst's high operational stability is important to maximize its reusability. As many engineered enzymes including cp283 exhibit reduced stability relative to their natural parents, the question of enzyme performance over extended time periods and multiple reaction cycles needed to be addressed. Using the same experimental conditions as for the interesterification reaction described above, we evaluated the conversion of canola oil and ethyl acetate by cp283 and wild-type CALB over seven 24-h reaction cycles or a total of 168 h (Figure 3-2). Based on our earlier results, cp283 reaches approximately 75% conversion after 24 h while wild-type CALB yields about 35% of the fatty acid ethyl ester product. At the end of each cycle, the immobilized biocatalyst was recovered by decanting of the spent reaction mixture, followed by addition of fresh substrates. The remaining lipase activity was determined in terms of ethyl ester formation for each cycle, relative to the starting activity.

As summarized in Figure 3-2, the activity of both lipases steadily declines by approximately 10% per 24-h cycle. After seven cycles, about 40% of the original activity remains. More importantly, the loss of activity for both immobilized enzymes is within margins of error of one another, suggesting no adverse effects due to the lower thermostability of cp283. Similar recycling experiments with wild type CALB by others (Hernandez-Martin and Otero 2008) suggest that the optimization of the reaction conditions could probably result in further improvements of the long-term performance of the biocatalysts.

3.4 Conclusions

Our study of transesterification reactions catalyzed by cp283 has not only addressed basic questions concerning changes in enzyme–substrate interactions but directly translated into a better understanding of the biocatalytic process for production of biodiesel. The initial kinetic experiments with wild-type CALB and cp283 point toward an alleviation or shift in the rate-

determining step for the engineered enzyme, translating into uniformly higher turnover rates. In the context of transesterification of plant oil and animal fat waste products to biodiesel fuel, the calculated four- to sevenfold improvement in activity for the engineered biocatalyst can be rationalized by the same phenomenon; an acceleration of fatty acid ester release.

The observed enhancement in performance is an important step towards raising the competitiveness of the enzymatic route for biodiesel production which has been limited by the high costs of the biocatalyst. A recent review estimates the costs for lipase at 0.14 US\$ per kg of ester, approximately 23-fold higher than the costs for NaOH, the catalyst for the chemical route, which runs at 0.006 US\$ per kg of ester product (Fjerbaek 2009). The reported increase in productivity with cp283 marks a significant step towards reducing the price gap.

Finally, our present study was aimed towards establishing a benchmark for cp283 by comparing its performance on a broad range of potential substrates with wild-type CALB. Future experiments, exploiting process engineering to find optimal reaction conditions, will likely be able to further raise productivity, potentially making lipases a true competitor for alternative fuel production.

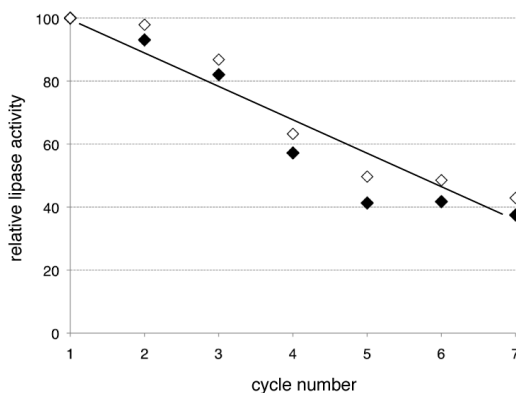


Figure 3-2. Enzyme activity of immobilized WT-CALB (\diamond) and cp283 (\blacklozenge) over multiple cycles of canola oil transesterification with ethyl acetate (1:6 molar ratio; 0.01% w/w enzyme based on oil weight). Each cycle represents a 24-h period after which the enzyme is recovered and added directly to a new batch of substrate.

Chapter 4 Functional and Structural Investigation of the New Termini of Permuted CALB

4.1 Introduction

In chapter 3, we observed that our circular permuted CALB cp283 consistently outperformed the wild-type enzyme. The rate enhancements toward various substrates can be explained by the crystal structure of cp283 Δ 7. The original narrow access tunnel to the active site is converted into a broad crevice, thus suggesting faster substrate entry and product release. cp283 Δ 7 is a variant identified from the cp283 linker truncation library (Figure 4-1) (Qian, Horton et al. 2009), in which five residues of the original hexapeptide linker together with the two residues in the native C-termini were truncated. This modification had no impact on catalytic activity, but resulted in an increase in the temperature of unfolding from 40°C to 45°C. The thermostability gain was attributed to the unexpected quaternary-structural change from a monomer, as seen in wild-type CALB and cp283, to a new homodimer.

To understand the nature of dimer formation, the crystal structure of cp283 Δ 7 was solved in the apo-state (PDB code: 3ICV) and the inhibitor-bound state (PDB code: 3ICW). When superimposed with wild-type CALB (Figure 4-2A), the core folds of the two proteins overlay very well. The most significant change is the swap of the two subunits' N-terminal segment. As the result of linker truncation, the extended N-terminus of one subunit points away from the rest of protein and orients itself onto the other subunit in a way that its N-terminus precisely connects with the other subunit (Figure 4-2B). Also noticeable, in the apo protein, electron density is not observed for the first five residues at the new N-terminus, as well as the 17 amino acids that make up the new C-terminus, suggesting these regions are relatively mobile. Co-crystallization of

cp283 Δ 7 with a phosphonate inhibitor yielded good electron density for the missing N-terminus, which rotates $\sim 45^\circ$ away from the active site (Figure 4-2 B). However, the 17 residues at the C-terminus remained invisible. In the wild-type enzyme, these residues belong to α -helix 16 (amino acids P268-L277) and a portion of α -helix 17 (amino acids L278-A282). In cp283 Δ 7, their conformation could range from a completely unstructured tether to a defined helical structure. In this chapter, the function and structure of the invisible C-terminal segment was investigated by biochemical and biophysical methods. We carried out termini deletion and site-directed mutagenesis to understand the relationship of these invisible residues to the catalytic activity. To gain insights into the structure of the missing C-terminus, we attempted to conduct NMR studies, but our protein ligation experiments failed to incorporate the isotopically-labeled peptide fragment onto the protein. We then conducted the peptide CD measurements and initiated the time-resolved fluorescence anisotropy decay study to understand the structural dynamics of the C-terminal segment.

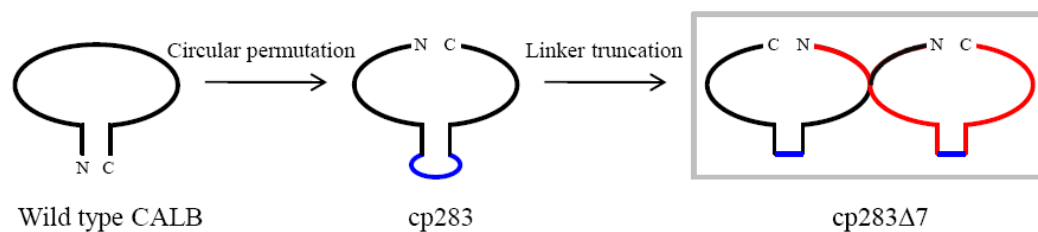


Figure 4-1. Schematic overview of CALB engineering by CP and subsequent linker truncation. CP created cp283 with native termini connected by a 6-amino acid linker (blue). Incremental truncation of the linker region changed the quaternary structure, shifting it from a strictly monomer to a domain-swapped dimer (Adapted from (Qian, Horton et al. 2009)).

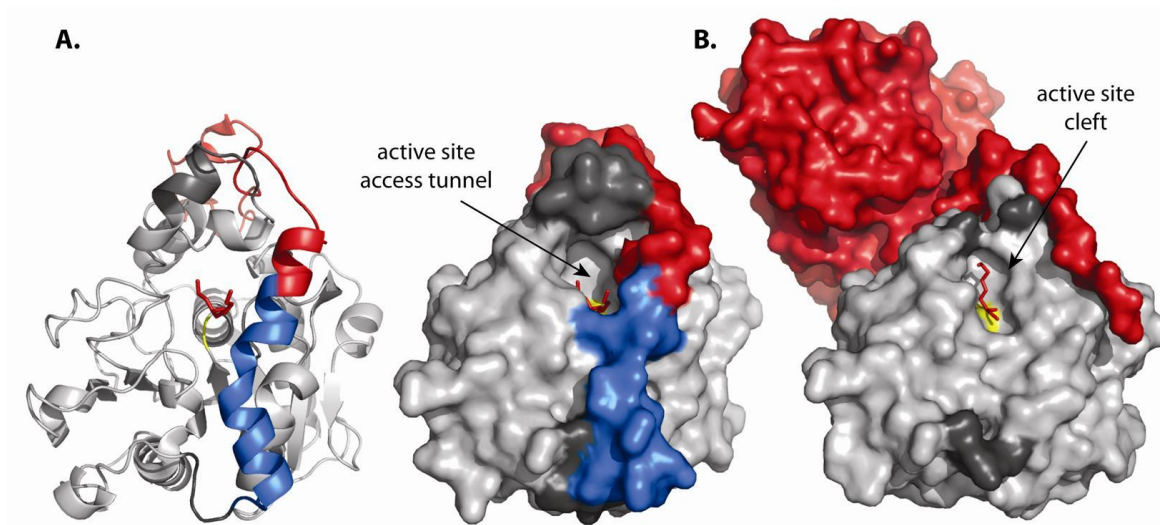


Figure 4-2. Comparison of the active-site accessibility in wild-type CALB (PDB code: 1LBS) (A) and cp283 Δ 7 (PDB code: 3ICW) (B). The structure of WT-CALB is shown in both ribbon and surface representations. In all structures, the active site is marked in yellow and the bound phosphonate inhibitor is shown in red sticks. The dark-gray area above the active site marks the enzyme's lid region, which could not be assigned in cp283 Δ 7 due to a lack of electron density. Cleavage of wild-type at Ala283 (position at red–blue crossover) results in widening of the active-site pocket. The new N-terminus (red) of cp283 Δ 7 rotates 45° away from the active site. 17 amino acids (blue in WT) at the new C-terminus were not present due to a lack of electron density in cp283 Δ 7.

(The reuse of the figure for this dissertation is permitted by Elsevier Limited. License number 2538540311165)

4.2 Materials and Methods

4.2.1 Materials

Gene: The gene of *Staphylococcus aureus* sortase A was cloned from *Staphylococcus aureus* genomic DNA, ATCC BAA-1556 (ATCC, Manassas, VA).

Chemicals: Reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless indicated otherwise. DiFMU octanoate (6, 8-difluoro-4-methylumbelliferyl octanoate) was purchased from Invitrogen (Carlsbad, CA). Fluorescein-5-maleimide (FM) and TCEP (tris (2-carboxyethyl) phosphine) are commercially available from ThermoFisher Scientific (Rockford, IL). Enzymes were purchased from New England Biolabs (Ipswich, MA).

Strains and media: *E. coli* DH5 α (Invitrogen, Carlsbad, CA) was used for cloning. Bacteria were grown under standard conditions in Luria-Bertani (LB) liquid media or on LB agar plates supplemented with 100 μ g/mL ampicillin. Lipase variants were transformed into *Pichia pastoris* strain GS115 (*his4*) cultured in YPD medium (1% yeast extract, 2% peptone and 2% glucose) and plated on MD plate (1.34 % yeast nitrogen base, 0.004% biotin, 2% glucose and 1.5% agar). Protein overexpression was performed in BMGY medium (1% yeast extract, 2% peptone, 1.34% yeast nitrogen base, 0.004% biotin, 1% glycerol and 50 mM potassium phosphate buffer, pH 6.0) and BMMY medium (1% yeast extract, 2% peptone, 1.34% yeast nitrogen base, 0.004% biotin, 0.5% methanol and 50 mM potassium phosphate buffer, pH 6.0).

Cloning: Oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA). *Pfu* DNA polymerase (Stratagene, LA Jolla, CA) was used for the cloning PCR. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit and PCR products were purified with QIAquick PCR Purification Kit (Qiagen, Valencia, CA).

Table 4-1. Primers used to generate termini truncation and site-directed mutagenesis variants

Variants	Primer Name	Sequence
NΔ5	cp288-for	5'- CGCCTCGAGAAAAGAGAGGCTGAAGCTGGTCCAAAGCAGAACT GCGAG – 3'
NΔ5CΔ3	A279-rev	5'- GTAGCGGCCGCTTACGCCAGGAGCGCAGCCGCGGGCG -3'
NΔ5CΔ4	L278-rev	5'- GTAGCGGCCGCTTACAGGAGCGCAGCCGCGGGCGACC -3'
NΔ5CΔ5	L277-rev	5'- GTAGCGGCCGCTTAGAGCGCAGCCGCGGGCGACCTTTTG -3'
NΔ5CΔ7	A275-rev	5'- GTAGCGGCCGCTTAAGCCGCGGGCGACCTTTTGCTCG -3'
NΔ5CΔ10	V272-rev	5'- GTAGCGGCCGCTTAGACCTTTTGCTCGGGAGTCAGATC -3'
NΔ5CΔ14	P268-rev	5'- GTAGCGGCCGCTTAGGGAGTCAGATCATTGGCGGGAAAG -3'
NΔ5CΔ17	D265-rev	5'-CGCGCGGCCGCTTAATCATTGGCGGGAAAGAGGGTTGC-3'
L278A	L278A-rev	5'-GTAGCGGCCGCTTATGCGAGCGCAGCCGCGGGCGACCTT-3'
L278S	L278S-rev	5'-GTAGCGGCCGCTTATGAGAGCGCAGCCGCGGGCGACC TT-3'
	cp283-for	5'-CGCCTCGAGAAAAGAGAGGCTGAAGCTGCAGCCATCGTGCGGG GTCCA-3'
WT- V272C	for	5'- CTGACTCCCCGAGCAAAAGTGCGCCGCGGGCTGCGCTCCTG -3'
	rev	5'- CTTTTGCTCGGGAGTCAGATC -3'
WT- A275C	for	5'- GAGCAAAAGGTCGCCGCGTGTGCGTCCTGGCGCCGGCG -3'
	rev	5'- CGCGGGCGACCTTTTGCTCGGG -3'
cp283- V272C	rev	5'- GTAGCGGCCGCTTAAGCCGCGGGCGCCAGGAGCGCAGCCGCGG CGCACTTTTGCTCGGGAGTCAG -3'
cp283- A275C	rev	5'- GTAGCGGCCGCTTAAGCCGCGGGCGCCAGGAGCGCACACGCGG CGACCTTTTG -3'
cp283- A279C	rev	5'- GTAGCGGCCGCTTAAGCCGCGGGCACAGGAGCGCAGCCGCG G-3'
cp283- V286C	for	5'- CCCCTCGAGAAAAGAGAGGCTGAAGCTGCAGCCATCTGCGCGG -3'
Sequencing	pPIC9-for	5'- GACTGGTTCCAATTGACAAGC -3'
	pPIC9-rev	5'-GCAAATGGCATTCTGACATCC-3'

4.2.2 Construction of termini truncation and site-directed mutagenesis variants

The termini truncation and L278 mutation variants were generated by one-step PCR using cp283 Δ 7 as the template and primers summarized in Table 4-1. Following purification with QIAquick column, the PCR products were digested with *Xho*I and *Not*I and ligated into pPIC9 digested with the same enzymes. The ligation products were transformed into *E. coli* DH5 α cells. After confirming their DNA sequences, the correct constructs were linearized by *Sac*I digestion and electroporated into *P. pastoris* strain GS115.

Five of the Cys mutants for time-resolved fluorescence anisotropy decay experiments, WT-A29C, WT-V286C, WT-G288C, cp283V286C and cp283G288C were generated by Lingfeng Liu. The rest of Cys mutants were created by either one-step or overlap PCR using WT-CALB and cp283 as the templates and primers in Table 4-1. Mutated genes were cloned into pPIC9 as above.

4.2.3 Protein expression in *P. pastoris*

The overexpression and purification of proteins followed the protocols described in our published paper (Qian and Lutz 2005). Briefly, a starter culture containing 25 mL BMGY was inoculated with a single colony from MD plate and was shaken at 250 rpm at 30°C until OD₆₀₀ reached 2-6. The cells were collected and resuspended in BMMY medium to an OD₆₀₀ of 1. Protein expression was induced by addition of methanol to a final concentration of 0.5% (v/v) every 24 h. CP variants were induced for 2 days at 20°C, while wild-type variants were induced for 2-3 days at 30°C. Lipase activity was tested by adding 2 μ l of *p*-nitrophenol butyrate (*p*-NB) to 200 μ l culture supernatant. Finally, the culture containing lipases was harvested by centrifugation at 1500 \times g for 10 min at 4°C.

4.2.4 Protein purification by HIC and SE chromatography

All proteins were purified to homogeneity using a combination of hydrophobic interaction (HIC) and size exclusion (SE) chromatography. The clear culture supernatant was mixed with 2 M (NH₄)₂SO₄ and 1 M potassium phosphate (KPi) buffer (pH 7.0) to a final concentration of 1 M

(NH₄)₂SO₄ and 50 mM KPi, respectively. For volumes > 150 mL, the culture supernatant was concentrated by ultrafiltration to < 30 mL (UF system with stirred cell, Millipore, Billerica, MA), prior to loading on a butyl-sepharose 4 resin column (Amersham Biosciences, Piscataway, NJ). The column was pre-equilibrated with starting buffer (1 M (NH₄)₂SO₄, 50 mM KPi, pH 7.0). After all the supernatant entered into the resin, the column was rinsed with 3 column volumes (CV) of starting buffer, followed by a stepwise reduction of the (NH₄)₂SO₄ concentration in the KPi buffer (0.2 M increments, 3 CV per step). Lipase typically eluted at (NH₄)₂SO₄ concentration of 0.2 M and the elution process was monitored with *p*-NB hydrolysis which gave visible yellow product. Fractions containing the desired activity were combined and concentrated by ultrafiltration (Amicon Ultra-15 centrifugal filter unit; Millipore, Bedford, MA). The protein obtained from this purification step is usually of 85% purity as determined by SDS-PAGE.

To obtain higher purity, the concentrated protein was loaded on a Superdex-200 10/300 GL column (Amersham Biosciences, Piscataway, NJ) attached to an ÄKTA purifier system (GE Healthcare, Piscataway, NJ) (150 mM NaCl and 50 mM KPi, pH 7.0). The eluent was monitored at 280 nm absorbance. Fractions with hydrolytic activity for *p*-NB and correct retention time (~15.3 min for wild-type and cp283 variants, ~13.5 min for cp283Δ7 variants) were combined and concentrated by Amicon Ultra-15 centrifugal filter unit. The final products showed > 95% purity as determined by SDS-PAGE analysis. Protein concentrations were calculated via the absorbance at 280 nm using $\epsilon = 33,000 \text{ M}^{-1}\text{cm}^{-1}$ (Rotticci, Norin et al. 2000).

4.2.5 Kinetic analysis of lipases activity with chromogenic and fluorescence substrates

Michaelis-Menten steady-state parameters for the hydrolysis of *p*-nitrophenyl butyrate and DiFMU octanoate were determined as described in section 3.2.2.

4.2.6 Circular dichroism analysis

Circular dichroism (CD) spectra were measured on a JASCO J-810 spectropolarimeter (JASCO, Easton, MD). The purified proteins were adjusted to 1 mg/mL in 50 mM KPi buffer (pH

7.0) and loaded to a 0.1 mm path length cell. Far-UV CD spectra (190 - 260 nm, 0.5 nm increments) were recorded at 10°C at a scanning speed of 20 nm/min with a response time of 4 sec and bandwidth of 2 nm. Three spectra were acquired and averaged for each sample. Data were corrected for buffer absorbance and converted to mean residue ellipticity.

Thermal denaturation curves were recorded by monitoring the signal at 222 nm as a function of temperature, increasing at a rate of 1°C/min from 10 to 80°C. The midpoint unfolding temperature (T_M), was calculated by nonlinear least-squares data fitting to a standard equation describing a two-state conformational equilibrium using Origin7 (OriginLab, Northampton, MA).

4.2.7 Peptide synthesis and purification

Peptide sequences were synthesized on a Liberty CEM Microwave Automated Peptide Synthesizer (CEM Corporation, Matthews, NC) using a 30 mL reaction vessel at a scale of 0.1 mmol. The Fmoc Wang resin (AnaSpec, Fremont, CA) was initially swollen using dimethylformamide (DMF) /dichloromethane (DCM) (50:50 v/v) for 15 minutes. Fmoc deprotection was achieved by addition of 20% piperidine, 0.1 M N-hydroxybenzotriazole (HOBt) in DMF for 180 sec with microwave power set to 45-55°C, followed by 3× flushing with DMF. Each coupling step was activated with 0.1 M 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), and 0.2 M N,N-diisopropylethylamine (DIEA) in DMF. Coupling temperatures were maintained for 300 sec between 75-82°C by optimizing microwave power. After coupling, the resin was rinsed with three aliquots of DMF.

Peptides were cleaved from resin in 10 mL trifluoroacetic acid (TFA) / thioanisole /1,2-ethanedithiol/ anisole (90:5:3:2, v/v/v/v) for 4 hours at room temperature. The solution was filtered and precipitated by dropwise addition of cold (-20°C) diethyl ether. Precipitated product was collected by centrifugation for 10 min at 4000 rpm at 4°C and washed with cold diethyl ether for three times. The crude peptides were dried overnight under vacuum.

For peptide purification, the crude peptide was dissolved in the minimal volume of 40 % acetonitrile + 0.1% TFA and purified by HPLC using a C18-reverse phase column with detection at 222 nm. A gradient of acetonitrile from 10% to 60% in water was run for 50 min at flow rate 10 mL/min. The desired peptide eluted at ~30.7 min and the molecule weights (Na⁺ peak: 1716 g/mol for wild-type peptide, 1749 g/mol for cys containing peptide) were confirmed by MALDI-TOF using a 2,5-dihydroxybenzoic acid matrix. Purified peptide was concentrated on the rotavap, followed by lyophilization.

4.2.8 Cloning and purification of sortase A

Primer SrtAΔN-For (5'-GAACATATGCAAGCTAAACCTCAAATTCCG-3') and SrtAΔN-REV (5'-GTGCTCGAGTTTGACTTCTGTAGCTACAAA-3') were used to amplify the sortase AΔN sequence corresponding to amino acids 60-204 from genomic *Staphylococcus aureus*. The PCR fragment was digested with *NdeI*/*XhoI* and cloned into pET21b (Novagen, Madison, WI) to express the N-terminal truncated sortase A. After DNA sequencing confirmation, the plasmid was transformed into *E. coli* BL21(DE3). The protein expression was induced with 0.2 mM isopropyl β-D-thiogalactoside (IPTG) and continued for 4 h at 30°C.

Sortase purification was facilitated by the 6×His tag attached to the C-terminus of the gene. Cell paste was suspended in lysis buffer (150 mM NaCl, 10 mM imidazole and 50 mM Tris-HCl, pH7.5) and lysed by sonication. Benzonase was added to remove the nucleic acids and reduce the viscosity. After centrifugation, the clear lysate was applied to Ni-NTA agarose resin (Qiagen, Valencia, CA) and washed with 10 resin volumes of wash buffer (150 mM NaCl, 20 mM imidazole and 50 mM Tris-HCl, pH7.5). Finally, the protein was eluted with 2 resin volumes of elute buffer (150 mM NaCl, 250 mM imidazole and 50 mM Tris-HCl, pH7.5) and exchanged into the storage buffer (150 mM NaCl and 50 mM Tris-HCl, pH7.5) by Amicon Ultra-15 centrifugal filter unit.

4.2.9 Peptide labeling and sortase-mediated protein ligation

In order to detect the protein-peptide ligation product, we synthesized a peptide with A280C mutation and labeled it with IAEDANS (5-({2-[(iodoacetyl) amino] ethyl}amino) naphthalene-1-sulfonic acid). Briefly, the peptide (300 μM) was treated with a 10-fold molar excess of TCEP agarose beads for 30 min at room temperature to reduce the introduced Cys, prior to adding a 10-fold molar excess of IAEDANS dye dissolved in DMF. Labeling was left to proceed at room temperature overnight and the labeled peptide was purified by HPLC with conditions described in section 4.2.7. Peak showing absorbance at 222 nm (peptide) and at 330 nm (fluorophore) was collected and concentrated on the rotavap, followed by lyophilization.

Sortase-mediated ligation followed the protocols in the literature (Mao, Hart et al. 2004). 50 μM CALB and 1 mM peptide was mixed with 50 μM sortase in the sortase buffer (50 mM MOPS (pH7.5), 10 mM CaCl_2 and 150 mM NaCl) and incubated for 20 h at 30°C. For the ligation reaction with IAEDANS-labeled peptide, the ligation mixture was applied to Superdex-200 10/300 GL column and separated as described in section 4.2.4.

4.2.10 Labeling of Cys mutants

To label the Cys mutants with fluorescein-5-maleimide (FM), the protein was diluted to 25 μM (20 mM KPi, 100 mM NaCl, pH 7.0) and treated with DTT (final concentration = 0.25 mM) for 30 min at room temperature. After removal of the remaining DTT by gel filtration over a PD-25 column, a 10-fold molar excess of FM, freshly dissolved in DMF, was added to the protein solution. The reaction was left to proceed for 2 h at room temperature and applied over a PD-25 column to remove unreacted FM. The labeling efficiency was calculated via UV absorption using $\epsilon_{(\text{FM})} = 83,000 \text{ M}^{-1}\text{cm}^{-1}$. The fluorescence intensity of labeled proteins was measured on a FluoroMax-3 Spex spectrofluorometer (HORIBA Jobin Yvon Inc., Edison, NJ) at room temperature. Excitation and emission slits were set to 5 nm. The excitation wavelength was set at 491 nm.

4.3 Results and Discussions

4.3.1 Functional consequence of termini deletion

To determine whether the invisible residues at the new termini of cp283 Δ 7 apo form are relevant to the enzyme function, we initially removed all of the invisible amino acids and generated variants N Δ 5 (five N-terminal amino acids deleted) and N Δ 5C Δ 17 (five N-terminal and 17 C-terminal amino acids deleted). The kinetic data with *p*-NB and DiFMU octanoate showed that N Δ 5 had no noticeable effect on the enzyme activity; however, removing the 17 residues at the C-terminus (N Δ 5C Δ 17) caused a significant drop in catalytic efficiency (k_{cat}/K_M), 5-fold for *p*-NB and 12-fold for DiFMU octanoate, respectively. This data suggests that the C-terminus plays an important role in enzyme function.

To more specifically pinpoint which residues in the invisible C-terminus are responsible for the activity drop, we gradually truncated amino acids from the carboxyl end and characterized the variants with regard to catalytic activity and overall thermodynamic stability. The kinetic data summarized in Table 4-2 revealed an interesting pattern for both *p*-NB and DiFMU octanoate. Variants with deletions of the first three (N Δ 5C Δ 3) and four (N Δ 5C Δ 4) residues maintained catalytic performance similar to that of cp283 Δ 7. However, truncation of one more residue L278 (N Δ 5C Δ 5) resulted in a sharp decline in the catalytic efficiency, 2.5-fold for *p*-NB and 5-fold for DiFMU octanoate, respectively. The removal of two or five additional amino acids (N Δ 5C Δ 7 and N Δ 5C Δ 10) did not cause any more significant decline in activity. When comparing the kinetic parameters for *p*-NB, slightly higher K_M and lower k_{cat} values contribute to the net activity loss. For DiFMU octanoate, the K_M values of all the mutants are within the 2-5 μ M range, while the k_{cat} values show greater changes and are mainly responsible for the activity drop. To complete this study, deletion of four more residues (N Δ 5C Δ 14) resulted in a further increased the K_M value and decreased k_{cat} value for *p*-NB, which translates into another 2-fold drop in catalytic efficiency. Similarly, the k_{cat}/K_M value with DiFMU octanoate decreased as well.

Our results strongly indicated that remaining some of the residues in the invisible C-terminal segment is necessary for full catalytic activity, suggesting that this segment may not exist in an unstructured, highly flexible conformation. Based on the above data, we speculate that these functionally relevant residues fold into an ordered helical secondary structure. Our subsequent studies aim to investigate this hypothesis.

Table 4-2. Apparent kinetic constants with *p*-nitrophenyl butyrate and DiFMU octanoate as substrates for terminally truncated variants

Enzyme	Sequence (N/C)	<i>p</i> -nitrophenyl butyrate (<i>p</i> -NB)			DiFMU octanoate		
		K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\text{min}^{-1}\mu\text{M}^{-1}$)	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\text{min}^{-1}\mu\text{M}^{-1}$)
WT	L1/P317	309±43	236±12	0.8	4.9±0.9	5.7±0.3	1.2
cp283Δ7	A283/A282	204±31	1318±50	6.5	4.7±0.5	820±34	174
NΔ5	G288/A282	202±20	1409±26	7.0	3.7±1.0	594±49	160
NΔ5CA3	G288/A279	214±14	1279±96	6.0	4.8±0.5	865±26	180
NΔ5CA4	G288/L278	232±32	1444±64	6.2	2.9±0.2	484±11	167
NΔ5CA5	G288/L277	371±29	910±24	2.5	2.3±0.3	71±3	31
NΔ5CA7	G288/A275	332±44	708±36	2.1	4.9±0.8	176±9	36
NΔ5CA10	G288/V272	380±53	937±46	2.5	2.5±0.3	64±2	26
NΔ5CA14	G288/P268	608±49	617±14	1.0	4.6±0.9	76±5	17
NΔ5CA17	G288/D265	553±62	704±64	1.3	3.9±0.9	57±4	15

4.3.2 CD characterization of terminal deletion variants

In parallel with the kinetic analysis, we also performed circular dichroism (CD) scans and thermal denaturation experiments with the deletion variants to explore the impact of the invisible terminal segments on protein secondary structure and stability. CD spectroscopy in the far-UV region was consistent for all truncation variants maintaining similar α -helix/ β -sheet spectral signatures with little intensity variations at 209 nm and 222 nm (Figure 4-3). However, the effect of terminal deletions is more significant in the temperature of unfolding. Although removing the five residues in the N-terminus had no impact on enzyme activity, it caused a $\sim 2^\circ\text{C}$ decrease in the T_M value. Similarly, the melting temperature gradually drops with the deletion of C-terminal residues, from 44°C for cp283 Δ 7 to 39°C for N Δ 5C Δ 17 (Figure 4-3). The loss in thermostability could result from the exposure of hydrophobic regions to the solvent after removal of these surface residues.

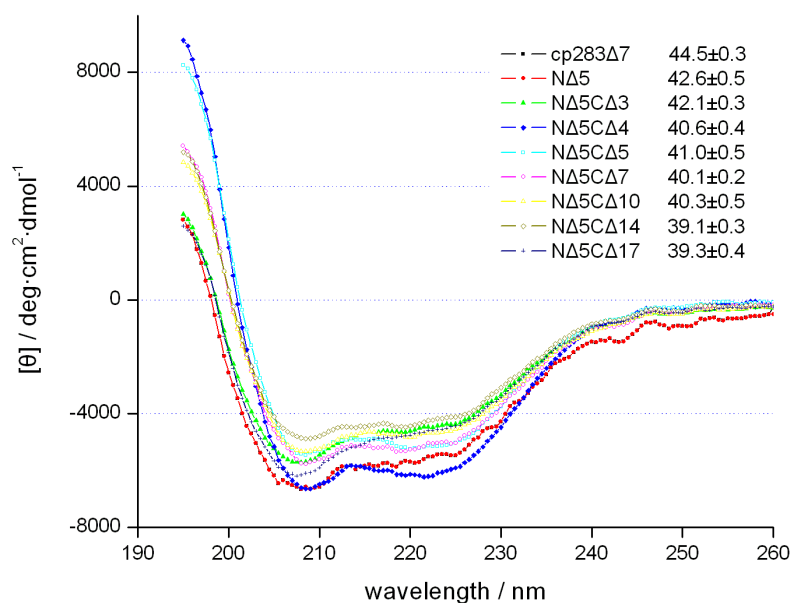


Figure 4-3. Far-UV CD spectra and melting temperatures of wild-type and terminal deletion variants.

4.3.3 Role of the two Leu residues in the invisible C-terminus

Since the N-terminal residues are visible in the inhibitor-bound crystal structure (PDB code: 13CV) and have no effect on enzymatic activity, our studies focused on the invisible C-terminus and more specifically on the detailed function of L278, due to its critical role in the earlier truncation experiments. In the wild type CALB crystal structure, L278 is located at the “kink” between α -helix 16 and α -helix 17 (Figure 4-4) and points toward the active site pocket. Based on these observations, we proposed two possible roles for L278: a) that its side chain interacts with the substrate during catalysis or b) that it acts as a capping residue for helix 16. A simple experiment to test these two ideas was to replace L278 with either Ala (N Δ 5C Δ 4-L278A) or Ser (N Δ 5C Δ 4-L278S) in N Δ 5C Δ 4. If L278 functions as a capping residue, substitution of the side chain should not interfere with its role in stabilizing helix 16 through the NH group in the peptide backbone (Aurora and Rose 1998; Rose 2006). The kinetic data for *p*-NB and DiFMU octanoate showed slightly improved catalytic efficiency for both variants compared to the L278 deletion mutant (N Δ 5C Δ 5), but it is still 2-fold (*p*-NB) and 5-fold (DiFMU octanoate) lower than cp283 Δ 7 (Table 4-3). The data suggests that the side chain of L278 is critical to the enzymatic function and that the residue is therefore not exclusively a helix capping residue. Instead, L278 is more likely involved in substrate binding.

In a separate mutational experiment of amino acids in position 264-266, substitutions in position 264 and 265 were well tolerated, showing little effect on catalysis, while replacement of L266 with Gly decreased activity by approximately 3-fold and 7-fold for *p*-NB and DiFMU octanoate, respectively. In fact, the activity of the 264-266 triple mutant closely matched that of the N Δ 5C Δ 14 deletional mutant. Since the side chain of L266 also points toward the protein core fold, but not directly to the active site pocket, we speculated that together with L278, L266 helps

position the C-terminus of cp283 Δ 7 during catalysis. This also suggests the invisible C-terminal segment is not completely free. Instead, it makes contact with the rest of protein.

The results from the truncation and mutation studies suggest that the C-terminal region, despite its lack of electron density, is structured and contributes to the overall enzyme performance. After termini relocation in cp283 Δ 7, part of the original α -helix17 (amino acids A283-A287) becomes the new N-terminus and folds into a helical arrangement as seen in the inhibitor-bound structure. The other half of the helix 17 (amino acids A279-A282) turns into the new C-terminus and its removal seems to have no effect on enzyme activity. However, the original α -helix 16 (amino acids P268-L278) might retain its integrity and directly contributes to enzyme catalysis (Figure 4-4).

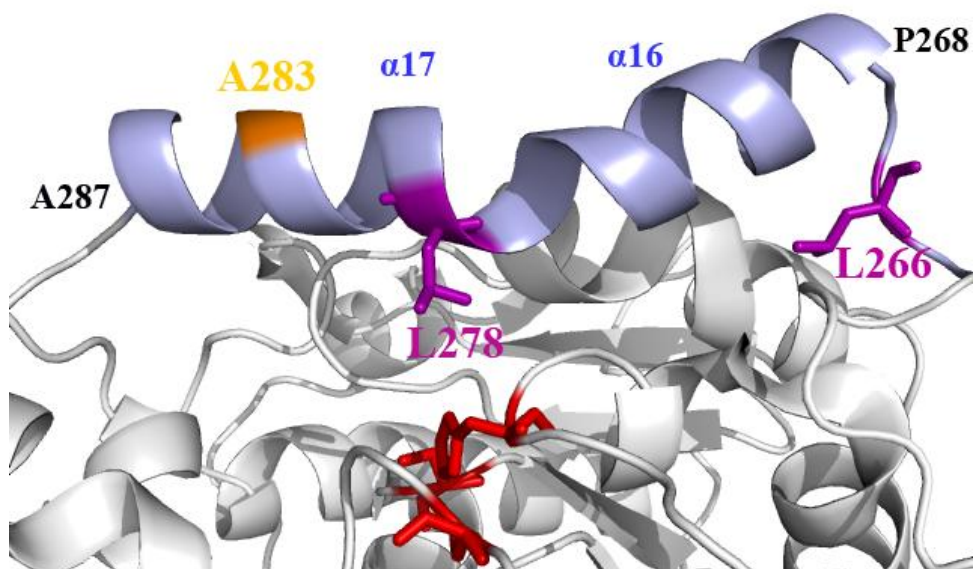


Figure 4-4. Highlight of α - helix 16 (P268-L277) and α - helix 17 (L278-A287) in wild-type CALB structure (1TCA). The corresponding invisible terminal residues in cp283 Δ 7 are colored in light blue. Catalytic residues are indicated in red. A283 and the two leucine residues are indicated in orange and purple, respectively.

Table 4-3. Apparent kinetic constants for Leu substitutions with *p*-nitrophenyl butyrate and DiFMU octanoate as substrates

Enzyme	Sequence (N/C)	<i>p</i> -nitrophenyl butyrate (<i>p</i> -NB)			DiFMU octanoate		
		K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\text{min}^{-1}\mu\text{M}^{-1}$)	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\text{min}^{-1}\mu\text{M}^{-1}$)
cp283 Δ 7	A283/A282	204 \pm 31	1318 \pm 50	6.5	4.7 \pm 0.5	820 \pm 34	174
N Δ 5C Δ 4-L278A	G288/A278	345 \pm 61	1016 \pm 50	2.9	2.4 \pm 0.6	126 \pm 12	53
N Δ 5C Δ 4-L278S	G288/S278	364 \pm 56	1258 \pm 70	3.4	2.5 \pm 0.6	143 \pm 11	57
cp283 Δ 7-N264TD265G	A283/A282	236 \pm 32	1364 \pm 58	5.8	3.6 \pm 0.9	507 \pm 47	141
cp283 Δ 7-N264TD265GL266G	A283/A282	448 \pm 71	872 \pm 51	1.9	3.0 \pm 0.7	53 \pm 4	18
N Δ 5C Δ 14	G288/P268	608 \pm 49	617 \pm 14	1.0	4.6 \pm 0.9	76 \pm 5	17

4.3.4 Labeling of new C-terminus by protein-peptide ligation

To further test the conformation of the C-terminus of cp283 Δ 7, we sought to investigate the structure of this invisible segment using NMR in conjugation with segmental isotope labeling. Conceptually, we planned to synthesize the relevant sequence on a peptide synthesizer using isotopically labeled amino acids and then to connect the fragment to the target protein through peptide ligation. The most common and successful protein ligation technique is intein-mediated protein ligation (IPL) and hence was first tested (Skrisovska, Schubert et al. 2010). IPL is also

called expressed protein ligation and allows ligation of a synthetic peptide with an N-terminal Cys residue to the C-terminus of a protein resulting in the formation of a peptide bond (Muir, Sondhi et al. 1998; Severinov and Muir 1998). This technique requires fusion of an engineered intein next to a Cys at the C-terminus of the target protein. The modified synthetic peptide is ligated to the protein upon induced intein self-cleavage. IPL has been applied to many proteins expressed in bacteria (Xu, Ayers et al. 1999; Shekhtman 2005; Berrade and Camarero 2009) but rarely to proteins expressed in yeast. We initially fused the bacterial intein to the C-terminus of cp283 Δ 7-C Δ 17, the construct missing all of the invisible C-terminal residues. However, no clear expression of the fusion protein was observed. Considering the codon usage bias in the yeast host strain, we replaced the bacterial intein with the VMA intein from *Saccharomyces cerevisiae* (Chong and Xu 1997). Still, we were unable to express the fusion protein.

As an alternative, we explored a relatively new technique called sortase mediated protein ligation (SML), which does not require a fusion protein (Tsukiji and Nagamune 2009). Sortases are transpeptidases found in most gram-positive bacteria. They are responsible for covalently anchoring a variety of surface proteins to the cell wall. The *Staphylococcus aureus* sortase A (SrtA) recognizes and cleaves a short C-terminal motif (LPXTG) on the target protein followed by the formation of a native amide bond. Due to its simplicity and specificity, SrtA has become a useful alternative to native chemical ligation (Kobashigawa, Kumeta et al. 2009). It has been applied to a variety of protein engineering and bioconjugation projects, including incorporation of new functionalities into peptides or proteins, creation of conjugates, protein circularization and labeling of cell surface proteins (Proft 2010). To perform SML, we cloned sortase A from *Staphylococcus aureus* as described in section 4.2.8 (Ton-That, Liu et al. 1999; Ilangovan, Ton-That et al. 2001). Separately, we introduced the LPXTG recognition motif into cp283 Δ 7 by mutating two residues directly before the invisible C-terminal segment and confirmed that these mutations had no obvious impact to enzyme activity.

Unfortunately, our ligation reactions were unable to yield any product after multiple trials. Initially, determination of the ligation product by SDS-PAGE was problematic since the attached peptide is only 1.6 kD, which makes it impossible to differentiate ligated from unligated proteins. Therefore, we introduced a Cys into the peptide and labeled it with the fluorophore IAEDANS. If the peptide is attached to the protein, we should see fluorescently labeled protein after the removal of free dye. However, no labeled protein was observed. As shown in the gel filtration spectrum (Figure 4-5), the peak corresponding to cp283Δ7 had neither dye absorbance (red curve) nor fluorescence. The failure could arise from any one of the steps in SML, recognition of the cleavage site, cleavage of the old sequence and ligation of the new peptide to the protein (Mao, Hart et al. 2004; Kobashigawa, Kumeta et al. 2009).

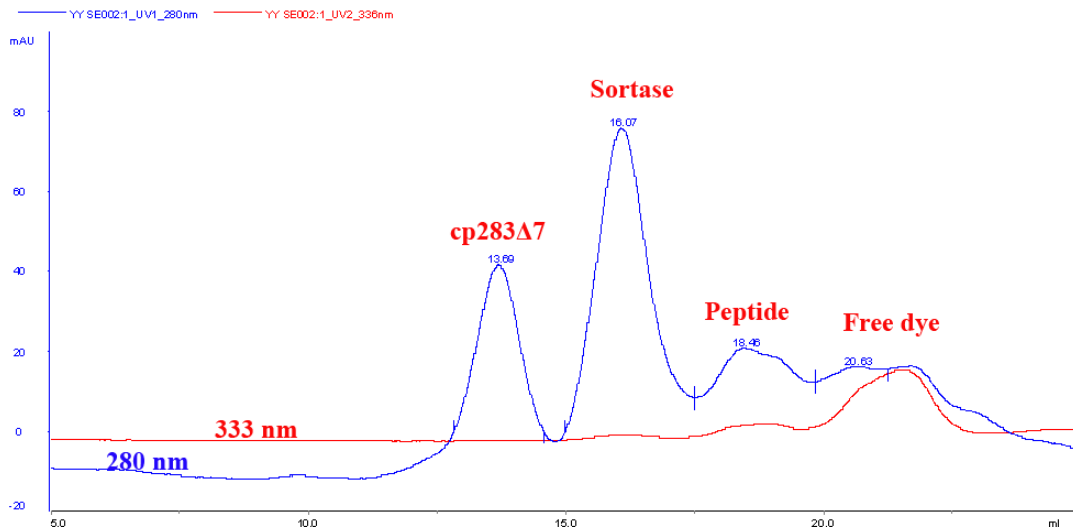


Figure 4-5. Gel filtration spectra of the sortase-mediated ligation mixture. Absorbance at 280 nm for protein (blue curve) and at 333 nm for dye (red curve) was monitored during the separation.

4.3.5 Secondary structure of C-terminal peptide

For the protein ligation experiments, we chemically synthesized an 18-amino acid peptide (GLTPEQKVAAAALLAPAA), whose sequence is almost identical to the invisible C-terminal segment except for the Gly which replaced Asp for peptide ligation purposes. We were curious if this peptide could provide any structural information about the C-terminus, and thus, carried out CD studies. In pure phosphate buffer, the spectrum showed a mostly random coil CD signature with a weak α -helical signal between 220-230 nm (Figure 4-6 A). To test if the signal is truly produced by helical structure, we then added 50% TFE, a well-known promoter of helix formation. As predicted, we observed a change in the spectral properties towards helical characteristics.

Next, the reversibility of helix folding was tested by heating the peptide to 70°C and then cooling it back to room temperature. As shown in Figure 4-6 B, the helix signal at 220-230 nm disappeared at 70°C, indicative of unfolding. Once the peptide was cooled back to room temperature, the helical signal reappeared, suggesting that this helix folding is reversible. This peptide study supports our hypothesis that the invisible C-terminal segment of cp283 Δ 7 is partially structured. We further predicted that the helical structure of this peptide is very dynamic, as the majority of the signal is still random coil. It is probably also true in the case of the permuted proteins, in which these sequences are dynamic and undergo folding and unfolding processes.

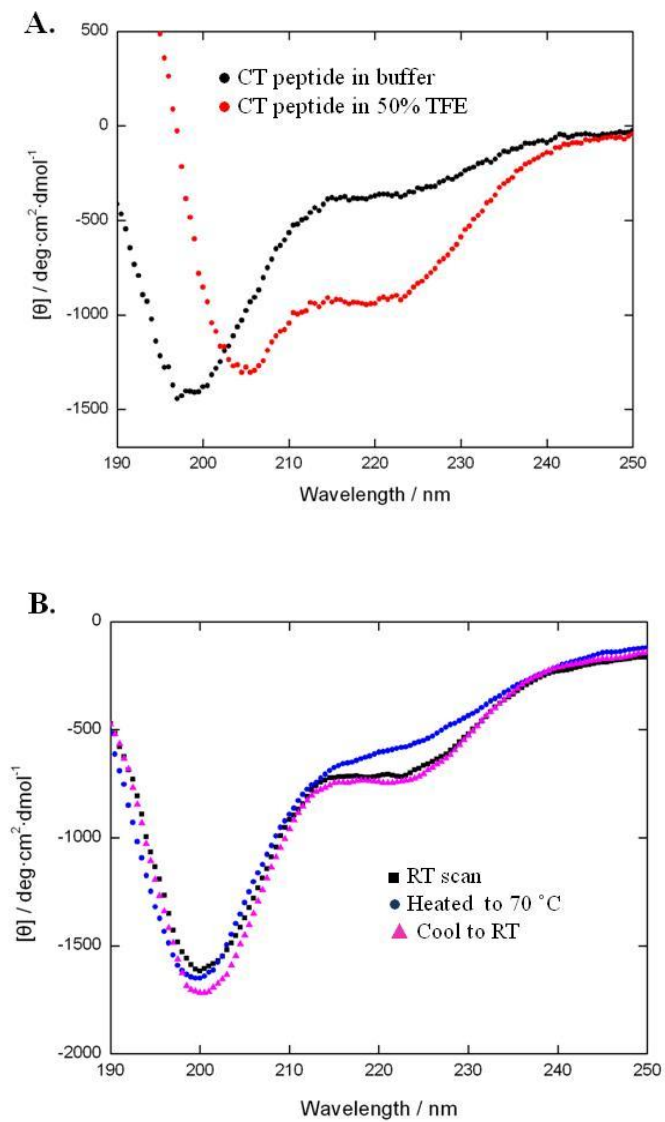


Figure 4-6. CD spectra of the C-terminal peptide in different conditions. **(A)** TFE (trifluoroethanol) induced an α -helical signal. **(B)** The helical signal disappeared when heated to 70°C, but reappeared once it was cooled down.

4.3.6 Characterization and labeling of Cys mutants

Since it was not possible for us to obtain the structure of the invisible C-terminal segment, we decided to study the dynamics of this segment using site-directed labeling in combination with time-resolved fluorescence anisotropy decay (TRFA), a biophysical approach to investigate protein dynamics (Bucci and Steiner 1988; Lakowicz 2006). We used the monomeric cp283 because we were concerned that the dimeric cp283 Δ 7 may complicate the experiment. Five positions on cp283 (V272, A275, A279, V286 and G288, Figure 4-7) and the five corresponding positions on wild-type CALB were selected as labeling sites. All 10 Cys-substituted mutants were generated, expressed and purified to homogeneity. Kinetic assays with *p*-nitrophenol butyrate confirmed that the Cys replacement had no obvious impact on enzyme activity, as shown in Table 4-4.

We chose fluoresceine-5-maleimide (FM), a fluorophore often used in TRFA studies for our experiments. The data measurements are still ongoing.

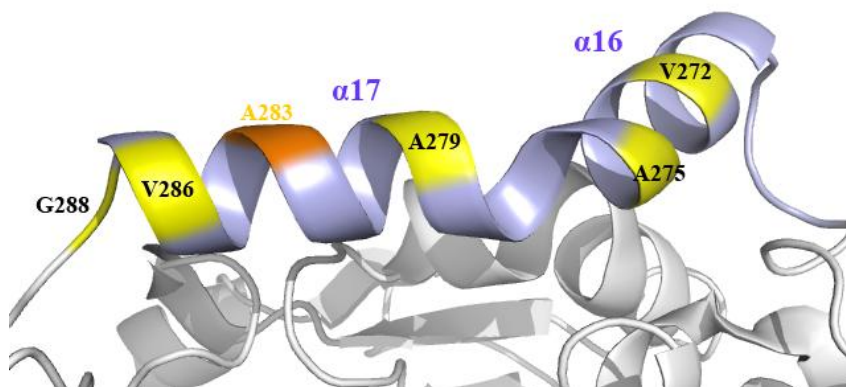


Figure 4-7. Five selected sites in wild-type CALB (PDB: 1TCA) for fluorophore labeling. G288, V286, A279, A275 and V272 are indicated in yellow. Invisible segment and A283 is indicated in light blue and orange, respectively.

Table 4-4. Catalytic activity of Cys mutants with *p*-nitrophenyl butyrate as the substrate

Enzyme	K_M (μM)	k_{cat} (min^{-1})	$\frac{k_{\text{cat}}}{K_M}$ ($\mu\text{M}^{-1}\text{min}^{-1}$)
WT	309 \pm 43	232 \pm 12	0.75
WT-V272C	301 \pm 31	228 \pm 9	0.76
WT-A275C	487 \pm 77	409 \pm 28	0.84
WT-A279C	328 \pm 53	248 \pm 15	0.76
WT-V286C	317 \pm 39	344 \pm 16	1.1
WT-G288C	409 \pm 45	467 \pm 29	1.1
cp283	130 \pm 17	1078 \pm 33	8.3
cp283-V272C	124 \pm 13	1038 \pm 24	8.4
cp283-A275C	117 \pm 14	892 \pm 25	7.6
cp283-A279C	160 \pm 16	1390 \pm 44	8.7
cp283-V286C	100 \pm 14	809 \pm 25	8.1
cp283-G288C	167 \pm 22	1425 \pm 52	8.5

4.4 Conclusions

From these deletion, mutagenesis and peptide studies, we have obtained evidence that supports our hypothesis that the invisible C-terminus in cp283 Δ 7 is structured. It is clear that this segment plays a role in maintaining enzyme activity and stability. We observed from the peptide experiments that the segment itself had some helical character and believe that the same is true when in the content of the protein. The invisible C-terminal segment is composed of 5 amino acids from the original α -helix 17 (amino acids A279-A282) and the entire α -helix 16 (amino acids P268-L278). Since our data showed that the remaining α -helix 17 is not important to enzyme performance, it is proposed to be unstructured. In contrast, α -helix 16 retains its helical structure and integrity, a characteristic that is critical for full enzyme activity. This conclusion agrees with previous results from the characterization of different circularly permuted CALBs (Qian, Fields et al. 2007). Variants cp284 and cp268, in which α -helix 16 remains intact, showed catalytic efficiencies similar to cp283. However, variants cp277 and cp278 have activities similar to the deletion variant N Δ 5C Δ 5 and the mutated L278A/S variants, providing additional evidence for our hypothesis. Consistent with the flexibility indicated in the crystal structure, we propose that the entire segment is flexible but not unstructured. The helical structure is likely dynamic, moving along with the whole segment, constantly unfolding and refolding. The ongoing time-resolved fluorescence anisotropy decay experiments shall provide more insight into the dynamical nature of the unsolved, but critical residues.

Chapter 5 Conclusions and Perspectives

5.1 Conclusions

Since the first use of site-directed mutagenesis to modify the active site of an enzyme three decades ago, protein engineering has made a huge impact on industry, medicine and scientific research. Traditional engineering methods usually rely on amino acids substitutions to tailor protein properties, yet have overlooked the potential benefits of rearranging the polypeptide chain of a protein. Random circular permutation, first developed in 1996 (Graf and Schachman 1996), has not been a well-established and recognized protein engineering tool, possibly due to its technical complexity and lack of successful examples.

Our previous studies with *Candida antarctica* lipase B (CALB) demonstrated that random CP can uncover unexpected permutation sites which translate into significantly improved enzymatic activity. With the purpose of expanding the use of random CP and investigating if this technique could be a universal engineering tool, we applied random CP to three industrially relevant enzymes including epoxide hydrolase (EchA), cutinase and xylanase (Bcx). While functional Bcx variants have their new termini broadly distributed across the whole protein, the new termini of functional EchA and cutinase variants were all near the amino and carboxyl-termini of their wild-type enzymes. Concluding from our studies, random CP is not a universal engineering tool, just like any other protein engineering method. The success of obtaining improved variants largely depends on the protein itself and is not limited by the type of structural fold. As in our cases, EchA and cutinase share a common α/β fold with CALB, yet the outcomes are completely different from CALB. In contrast, the β -jellyroll fold in Bcx is tolerant to termini relocation in many sites.

Although we still do not fully understand what criteria are required for successful engineering of proteins using random CP, it is true that proteins like Bcx with termini proximity have a good chance to tolerate CP within the interior of the fold, while proteins with termini far away from each other encounter more challenges. However, this rule did not exclude the use of CP in proteins with long termini distance. Random CP of 5-aminolevulinate synthase, whose termini are 50 Å away from each other, yielded variants with up to 10-fold increased activity toward glycine (Cheltsov, Barber et al. 2001) as discussed in section 1.4.1. In this dissertation, we explored possible experimental strategies for designing suitable linkers. By constructing combinatorial libraries, the linker length and composition can be optimized to connect greater distance without interfering with protein folding and stability.

To continue our studies with CALB and investigate the practical applications of permuted CALB beyond the laboratory substrates, we assessed the performance of cp283 for converting vegetable oils into fatty acid esters that make up biodiesel. We were excited to see that in comparison with wild-type CALB, cp283 showed consistently higher catalytic activity in transesterification and interesterification with pure and complex triglycerides. Lipases with a faster rate could speed up the reaction and increase the productivity, and thus narrow the gap between the cost for enzymes and alkaline catalysts that are currently used in biodiesel production. Next, to better understand the function and structure of the new termini created in circular permutation, we focused on the C-terminal segment of cp283 Δ 7 that is missing in the crystal structure due to the lack of electron density. Using truncation and site-directed mutagenesis, we investigated the role of the invisible C-terminus and identified the residues critical to enzyme activity. Furthermore, we gained insight into the dynamic structure of this segment and speculated that part of the residues folds into a helical structure.

5.2 Perspectives

From the exploration of four enzymes, we have demonstrated that random circular permutation could be a useful tool for protein engineering. Nevertheless, we also realized that linker design is a critical component for this technique. Although our efforts using combinatorial approaches to identify suitable linkers were not successful, this aim could still be tackled with the assistance of computational algorithms. In fact, I recommended that the future linker design, especially for connecting long termini distances, should be guided or evaluated by computational methods if possible. Meanwhile, I expect to see additional use of random CP for tailoring protein functions. The accumulation of more examples could provide insight into the range of random CP in protein engineering.

The goal of enzyme engineering is always to identify candidates that could perform better in industrial or therapeutic processes. This dissertation presented one possible application of our engineered CALB, for the future, it would be interesting to test the efficiency of the permuted CALB in other types of reactions, such as polymerization and amidation, since wild-type CALB is known for broad substrate specificity.

Another aspect for CALB study could be additional engineering to further improve the permutants. As a matter of fact, majority of the substantial accomplishments in enzyme engineering were the results of multiple rounds of directed evolution, as described in chapter 1. In the case of CALB, incremental truncation of the linker region was able to help the thermostability. It is likely that the permutants could benefit from other engineering efforts as well. For example, it was found that swapping the lid of wild-type CALB with the lid regions from its homologues led to increased enantioselectivity (Skjot, De Maria et al. 2009). This idea could also be applied to permuted CALB and tested to see if the similar effect could be observed in the permutant.

References

- Ahmad, S., M. Z. Kamal, et al. (2008). "Thermostable *Bacillus subtilis* lipases: in vitro evolution and structural insight." J Mol Biol **381**(2): 324-340.
- Akemann, W., C. D. Raj, et al. (2001). "Functional characterization of permuted enhanced green fluorescent proteins comprising varying linker peptides." Photochem Photobiol **74**(2): 356-363.
- Akoh, C., Chang SW, Lee GC, Shaw JF (2007). "Enzymatic approach to biodiesel production." J Agric Food Chem. **55**: 8995-9005.
- Araujo, R., C. Silva, et al. (2007). "Tailoring cutinase activity towards polyethylene terephthalate and polyamide 6,6 fibers." J Biotechnol **128**(4): 849-857.
- Archelas, A. and R. Furstoss (2001). "Synthetic applications of epoxide hydrolases." Curr Opin Chem Biol **5**(2): 112-119.
- Arnold, F. H. (2009). "How proteins adapt: lessons from directed evolution." Cold Spring Harb Symp Quant Biol **74**: 41-46.
- Aurora, R. and G. D. Rose (1998). "Helix capping." Protein Sci **7**(1): 21-38.
- Baird, G. S., D. A. Zacharias, et al. (1999). "Circular permutation and receptor insertion within green fluorescent proteins." Proc Natl Acad Sci U S A **96**(20): 11241-11246.
- Barbas, C., III, Burton DR, Scott JK, Silverman GJ (2001). Appendix 1, in Phage display: a laboratory manual. . Phage display: a laboratory manual. . Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press: A1.1–A1.2.
- Barrientos, L. G., J. M. Louis, et al. (2002). "Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N." Proteins-Structure Function and Genetics **46**(2): 153-160.

- Beg, Q. K., Kapoor, M., Mahajan, L., Hoondal, G.S. (2001). "Microbial xylanases and their industrial applications: a review." Appl Microbiol Biotechnol **56**: 326-338.
- Berglund, P. (2001). "Controlling lipase enantioselectivity for organic synthesis." Biomol Eng **18**(1): 13-22.
- Berrade, L. and J. A. Camarero (2009). "Expressed protein ligation: a resourceful tool to study protein structure and function." Cell Mol Life Sci **66**(24): 3909-3922.
- Bloom, J. D., M. M. Meyer, et al. (2005). "Evolving strategies for enzyme engineering." Curr Opin Struct Biol **15**(4): 447-452.
- Bornscheuer, U. and R. Kazlauskas (2006). Hydrolases in organic synthesis: Regio- and stereoselective biotransformations, Weinheim: Wiley-VCH.
- Bucci, E. and R. F. Steiner (1988). "Anisotropy decay of fluorescence as an experimental approach to protein dynamics." Biophys Chem **30**(3): 199-224.
- Buchwalder, A., H. Szadkowski, et al. (1992). "A fully active variant of dihydrofolate reductase with a circularly permuted sequence." Biochemistry **31**(6): 1621-1630.
- Butler, J. S., D. M. Mitrea, et al. (2009). "Structural and thermodynamic analysis of a conformationally strained circular permutant of barnase." Biochemistry **48**(15): 3497-3507.
- Cabantous, S., J. D. Pedelacq, et al. (2005). "Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*." J Struct Funct Genomics **6**(2-3): 113-119.
- Carlson, H. J., D. W. Cotton, et al. (2010). "Circularly permuted monomeric red fluorescent proteins with new termini in the beta-sheet." Protein Sci.
- Carvalho, C. M., M. R. Aires-Barros, et al. (1999). "Cutinase: from molecular level to bioprocess development." Biotechnol Bioeng **66**(1): 17-34.

- Castle, L. A., D. L. Siehl, et al. (2004). "Discovery and directed evolution of a glyphosate tolerance gene." Science **304**(5674): 1151-1154.
- Cellitti, J., M. Llinas, et al. (2007). "Exploring subdomain cooperativity in T4 lysozyme I: structural and energetic studies of a circular permutant and protein fragment." Protein Sci **16**(5): 842-851.
- Chalton, D. A., J. A. Musson, et al. (2006). "Immunogenicity of a Yersinia pestis vaccine antigen monomerized by circular permutation." Infect Immun **74**(12): 6624-6631.
- Cheltsov, A. V., M. J. Barber, et al. (2001). "Circular permutation of 5-aminolevulinate synthase. Mapping the polypeptide chain to its function." J Biol Chem **276**(22): 19141-19149.
- Cheltsov, A. V., W. C. Guida, et al. (2003). "Circular permutation of 5-aminolevulinate synthase: effect on folding, conformational stability, and structure." J Biol Chem **278**(30): 27945-27955.
- Chen, J., Wu W (2003). "Regeneration of immobilized Candida antarctica lipase for transesterification." J Biosci Bioeng. **95**: 466-469.
- Cherry, J. R. and A. L. Fidantsef (2003). "Directed evolution of industrial enzymes: an update." Curr Opin Biotechnol **14**(4): 438-443.
- Chica, R. A., N. Doucet, et al. (2005). "Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design." Curr Opin Biotechnol **16**(4): 378-384.
- Choi, W. J. (2009). "Biotechnological production of enantiopure epoxides by enzymatic kinetic resolution." Appl Microbiol Biotechnol **84**(2): 239-247.
- Chong, S. and M. Q. Xu (1997). "Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs." J Biol Chem **272**(25): 15587-15590.
- Chu, V., S. Freitag, et al. (1998). "Thermodynamic and structural consequences of flexible loop deletion by circular permutation in the streptavidin-biotin system." Protein Sci **7**(4): 848-859.

- Collins, T., C. Gerday, et al. (2005). "Xylanases, xylanase families and extremophilic xylanases." FEMS Microbiol Rev **29**(1): 3-23.
- Damborsky, J. and J. Brezovsky (2009). "Computational tools for designing and engineering biocatalysts." Curr Opin Chem Biol **13**(1): 26-34.
- Davoodi, J., W. W. Wakarchuk, et al. (2007). "Mechanism of stabilization of *Bacillus circulans* xylanase upon the introduction of disulfide bonds." Biophys Chem **125**(2-3): 453-461.
- de Vries, E. J. and D. B. Janssen (2003). "Biocatalytic conversion of epoxides." Curr Opin Biotechnol **14**(4): 414-420.
- Dougherty, M. J. and F. H. Arnold (2009). "Directed evolution: new parts and optimized function." Curr Opin Biotechnol **20**(4): 486-491.
- Du, W., Xu Y, Liu D, Zeng J. (2004). "Comparative study on lipase-catalyzed transformation of soybean oil for biodiesel production with different acyl acceptors." J.Mol.Catal. B-Enzyme **30**: 125-129.
- Egmond, M. R. and J. de Vlieg (2000). "Fusarium solani pisi cutinase." Biochimie **82**(11): 1015-1021.
- Egmond, M. R., J. de Vlieg, et al. (1996). Strategies and design of mutations in lipases. Dordrecht, Kluwer Academic Publishers.
- Finney, N. S. (1998). "Enantioselective epoxide hydrolysis: catalysis involving microbes, mammals and metals." Chem Biol **5**(4): R73-79.
- Fjerbaek, L., Christensen KV, Norddahl B (2009). "A review of the current state of biodiesel production using enzymatic transesterification." Biotechnol Bioeng. **102**: 1298-1315.
- Flores, G., X. Soberon, et al. (2004). "Production of a fully functional, permuted single-chain penicillin G acylase." Protein Science **13**(6): 1677-1683.
- Fox, R. J., S. C. Davis, et al. (2007). "Improving catalytic function by ProSAR-driven enzyme evolution." Nat Biotechnol **25**(3): 338-344.
- Freedonia Group Inc (2009). "Enzymes Market Research Reports."

- Fretland, A. J. and C. J. Omiecinski (2000). "Epoxide hydrolases: biochemistry and molecular biology." Chem Biol Interact **129**(1-2): 41-59.
- Garrett, J. B., L. S. Mullins, et al. (2003). "Effect of linker sequence on the stability of circularly perinuted variants of ribonuclease T1." Bioorganic Chemistry **31**(5): 412-424.
- Gilbert, W. (1987). "The exon theory of genes." Cold Spring Harb Symp Quant Biol **52**: 901-905.
- Goldenberg, D. P. and T. E. Creighton (1983). "Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor." J Mol Biol **165**(2): 407-413.
- Gould, S. M. and D. S. Tawfik (2005). "Directed evolution of the promiscuous esterase activity of carbonic anhydrase II." Biochemistry **44**(14): 5444-5452.
- Graf, R. and H. K. Schachman (1996). "Random circular permutation of genes and expressed polypeptide chains: application of the method to the catalytic chains of aspartate transcarbamoylase." Proc Natl Acad Sci U S A **93**(21): 11591-11596.
- Hahn, M., K. Piotukh, et al. (1994). "Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis." Proc Natl Acad Sci U S A **91**(22): 10417-10421.
- Heinemann, U., J. Ay, et al. (1996). "Enzymology and folding of natural and engineered bacterial beta-glucanases studied by X-ray crystallography." Biological Chemistry **377**(7-8): 447-454.
- Hennecke, J., P. Sebbel, et al. (1999). "Random circular permutation of DsbA reveals segments that are essential for protein folding and stability." J Mol Biol **286**(4): 1197-1215.
- Hernandez-Martin, E. and C. Otero (2008). "Different enzyme requirements for the synthesis of biodiesel: Novozym 435 and Lipozyme TL IM." Bioresour Technol **99**(2): 277-286.
- Hill, C. M., W. S. Li, et al. (2003). "Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site." J Am Chem Soc **125**(30): 8990-8991.

- Hisano, T., K. Kasuya, et al. (2006). "The crystal structure of polyhydroxybutyrate depolymerase from *Penicillium funiculosum* provides insights into the recognition and degradation of biopolyesters." J Mol Biol **356**(4): 993-1004.
- Holmquist, M. (2000). "Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms." Curr Protein Pept Sci **1**(2): 209-235.
- Huang, Y. M. and C. Bystroff (2009). "Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein." Biochemistry **48**(5): 929-940.
- Ilangovan, U., H. Ton-That, et al. (2001). "Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*." Proc Natl Acad Sci U S A **98**(11): 6056-6061.
- Iwakura, M. and T. Nakamura (1998). "Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase." Protein Engineering **11**(8): 707-713.
- Iwakura, M. and T. Nakamura (1998). "Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase." Protein Eng **11**(8): 707-713.
- Jaeger, K. E., S. Ransac, et al. (1994). "Bacterial lipases." FEMS Microbiol Rev **15**(1): 29-63.
- Jegannathan, K., S. Abang, et al. (2008). "Production of biodiesel using immobilized lipase--a critical review." Crit Rev Biotechnol. **28**(4): 253-264.
- Jeltsch, A. (1999). "Circular permutations in the molecular evolution of DNA methyltransferases." J Mol Evol **49**(1): 161-164.
- Jiang, L., E. A. Althoff, et al. (2008). "De novo computational design of retro-aldol enzymes." Science **319**(5868): 1387-1391.
- Johannes, T. W. and H. Zhao (2006). "Directed evolution of enzymes and biosynthetic pathways." Curr Opin Microbiol **9**(3): 261-267.

- Johnson, R. J., S. R. Lin, et al. (2006). "A ribonuclease zymogen activated by the NS3 protease of the hepatitis C virus." FEBS J **273**(23): 5457-5465.
- Joshi, M. D., G. Sidhu, et al. (2000). "Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase." J Mol Biol **299**(1): 255-279.
- Jung, J. and B. Lee (2001). "Circularly permuted proteins in the protein structure database." Protein Sci **10**(9): 1881-1886.
- Karukurichi, K. R., L. Wang, et al. (2010). "Analysis of p300/CBP histone acetyltransferase regulation using circular permutation and semisynthesis." J Am Chem Soc **132**(4): 1222-1223.
- Kim, S. B., M. Sato, et al. (2008). "Circularly permuted bioluminescent probes for illuminating ligand-activated protein dynamics." Bioconjug Chem **19**(12): 2480-2486.
- Kobashigawa, Y., H. Kumeta, et al. (2009). "Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method." J Biomol NMR **43**(3): 145-150.
- Kojima, M., K. Ayabe, et al. (2005). "Importance of terminal residues on circularly permuted Escherichia coli alkaline phosphatase with high specific activity." J Biosci Bioeng **100**(2): 197-202.
- Kreitman, R. J., R. K. Puri, et al. (1994). "A circularly permuted recombinant interleukin 4 toxin with increased activity." Proc Natl Acad Sci U S A **91**(15): 6889-6893.
- Kuo, J. M., L. S. Mullins, et al. (1995). "Circular permutation of RNase T1 through PCR based site-directed mutagenesis." Techniques in Protein Chemistry Vi **6**: 333-340
- 585.
- Labrou, N. E. (2010). "Random mutagenesis methods for in vitro directed enzyme evolution." Curr Protein Pept Sci **11**(1): 91-100.
- Lakowicz, J. R. (2006). Principles of Fluorescence Spectroscopy, Springer.

- Lawson, S. L., W. W. Wakarchuk, et al. (1996). "Effects of both shortening and lengthening the active site nucleophile of *Bacillus circulans* xylanase on catalytic activity." Biochemistry **35**(31): 10110-10118.
- Lawson, S. L., W. W. Wakarchuk, et al. (1997). "Positioning the acid/base catalyst in a glycosidase: studies with *Bacillus circulans* xylanase." Biochemistry **36**(8): 2257-2265.
- Lee, E. Y. and M. L. Shuler (2007). "Molecular engineering of epoxide hydrolase and its application to asymmetric and enantioconvergent hydrolysis." Biotechnol Bioeng **98**(2): 318-327.
- Leisola, M. and O. Turunen (2007). "Protein engineering: opportunities and challenges." Appl Microbiol Biotechnol **75**(6): 1225-1232.
- Li, P., H. Guan, et al. (2009). "Heterologous expression, purification, and characterization of cytochrome P450sca-2 and mutants with improved solubility in *Escherichia coli*." Protein Expr Purif **65**(2): 196-203.
- Lindqvist, Y. and G. Schneider (1997). "Circular permutations of natural protein sequences: structural evidence." Curr Opin Struct Biol **7**(3): 422-427.
- Liu, L., Y. Li, et al. (2009). "Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting." Nucleic Acids Res **37**(13): 4472-4481.
- Lo, W. C. and P. C. Lyu (2008). "CPSARST: an efficient circular permutation search tool applied to the detection of novel protein structural relationships." Genome Biol **9**(1): R11.
- Longhi, S. and C. Cambillau (1999). "Structure-activity of cutinase, a small lipolytic enzyme." Biochim Biophys Acta **1441**(2-3): 185-196.
- Longhi, S., M. Czjzek, et al. (1997). "Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis." J Mol Biol **268**(4): 779-799.
- Loo, V., B., J. H. Spelberg, et al. (2004). "Directed evolution of epoxide hydrolase from *A. radiobacter* toward higher enantioselectivity by error-prone PCR and DNA shuffling." Chem Biol **11**(7): 981-990.

- Ludwiczek, M. L., M. Heller, et al. (2007). "A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase." J Mol Biol **373**(2): 337-354.
- Luger, K., U. Hommel, et al. (1989). "Correct folding of circularly permuted variants of a beta alpha barrel enzyme in vivo." Science **243**(4888): 206-210.
- Lutz, S. (2004). "Engineering lipase B from *Candida antarctica*." Tetrahedron: Asymmetry **15**: 2743-2748.
- Lutz, S. (2010). "Beyond directed evolution-semi-rational protein engineering and design." Curr Opin Biotechnol.
- Lutz, S. (2010). "Biochemistry. Reengineering enzymes." Science **329**(5989): 285-287.
- Maatta, J. A. E., T. T. Airene, et al. (2008). "Rational modification of ligand-binding preference of avidin by circular permutation and mutagenesis." Chembiochem **9**(7): 1124-1135.
- Manjasetty, B. A., J. Hennecke, et al. (2004). "Structure of circularly permuted DsbA(Q100T99): preserved global fold and local structural adjustments." Acta Crystallogr D Biol Crystallogr **60**(Pt 2): 304-309.
- Mao, H., S. A. Hart, et al. (2004). "Sortase-mediated protein ligation: a new method for protein engineering." J Am Chem Soc **126**(9): 2670-2671.
- Martinez, C., P. De Geus, et al. (1992). "Fusarium solani cutinase is a lipolytic enzyme with a catalytic serine accessible to solvent." Nature **356**(6370): 615-618.
- Matama, T., R. Araujo, et al. (2010). "Functionalization of cellulose acetate fibers with engineered cutinases." Biotechnol Prog **26**(3): 636-643.
- McIntosh, L. P., G. Hand, et al. (1996). "The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a ¹³C-NMR study of bacillus circulans xylanase." Biochemistry **35**(31): 9958-9966.
- Mitreá, D. M., L. S. Parsons, et al. (2010). "Engineering an artificial zymogen by alternate frame protein folding." Proc Natl Acad Sci U S A **107**(7): 2824-2829.

- Modi, M., Reddy JR, Rao BV, Prasad RB (2007). "Lipase-mediated conversion of vegetable oils into biodiesel using ethyl acetate as acyl acceptor." Bioresour Technol. **98**: 1260-1264.
- Muir, T. W., D. Sondhi, et al. (1998). "Expressed protein ligation: a general method for protein engineering." Proc Natl Acad Sci U S A **95**(12): 6705-6710.
- Mullins, L. S., K. Wesseling, et al. (1994). "Transposition of Protein Sequences - Circular Permutation of Ribonuclease-T1." Journal of the American Chemical Society **116**(13): 5529-5533.
- Nardini, M., I. S. Ridder, et al. (1999). "The x-ray structure of epoxide hydrolase from *Agrobacterium radiobacter* AD1. An enzyme to detoxify harmful epoxides." J Biol Chem **274**(21): 14579-14586.
- Nordlund, H. R., V. P. Hytonen, et al. (2005). "Tetravalent single-chain avidin: from subunits to protein domains via circularly permuted avidins." Biochemical Journal **392**: 485-491.
- Ohman, A., T. Oman, et al. (2010). "Solution structures and backbone dynamics of the ribosomal protein S6 and its permutant P(54-55)." Protein Sci **19**(1): 183-189.
- Okada, S., K. Ota, et al. (2009). "Circular permutation of ligand-binding module improves dynamic range of genetically encoded FRET-based nanosensor." Protein Sci **18**(12): 2518-2527.
- Ostermeier, M. (2009). "Designing switchable enzymes." Curr Opin Struct Biol **19**(4): 442-448.
- Ostermeier, M., A. E. Nixon, et al. (1999). "Combinatorial protein engineering by incremental truncation." Proc Natl Acad Sci U S A **96**(7): 3562-3567.
- Ostermeier, M., J. H. Shim, et al. (1999). "A combinatorial approach to hybrid enzymes independent of DNA homology." Nat Biotechnol **17**(12): 1205-1209.
- Osuna, J., A. Perez-Blancas, et al. (2002). "Improving a circularly permuted TEM-1 beta-lactamase by directed evolution." Protein Eng **15**(6): 463-470.
- Palackal, N., Y. Brennan, et al. (2004). "An evolutionary route to xylanase process fitness." Protein Sci **13**(2): 494-503.

- Peisajovich, S. G., L. Rockah, et al. (2006). "Evolution of new protein topologies through multistep gene rearrangements." Nat Genet **38**(2): 168-174.
- Pieper, U., K. Hayakawa, et al. (1997). "Circularly permuted beta-lactamase from *Staphylococcus aureus* PC1." Biochemistry **36**(29): 8767-8774.
- Pio, T. F. and G. A. Macedo (2009). "Cutinases: properties and industrial applications." Adv Appl Microbiol **66**: 77-95.
- Plainkum, P., S. M. Fuchs, et al. (2003). "Creation of a zymogen." Nat Struct Biol **10**(2): 115-119.
- Plesniak, L. A., W. W. Wakarchuk, et al. (1996). "Secondary structure and NMR assignments of *Bacillus circulans* xylanase." Protein Sci **5**(6): 1118-1135.
- Proft, T. (2010). "Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation." Biotechnol Lett **32**(1): 1-10.
- Protasova, N. Y., M. L. Kireeva, et al. (1994). "Circularly Permuted Dihydrofolate-Reductase of *Escherichia-Coli* Has Functional-Activity and a Destabilized Tertiary Structure." Protein Engineering **7**(11): 1373-1377.
- Qian, Z., Field C, Lutz S (2007). "Investigating the structural and functional consequences of circular permutation on lipase B from *Candida antarctica*." Chembiochem **8**: 1989-1996.
- Qian, Z., C. J. Fields, et al. (2007). "Investigating the structural and functional consequences of circular permutation on lipase B from *Candida antarctica*." Chembiochem **8**(16): 1989-1996.
- Qian, Z., C. J. Fields, et al. (2007). "Recent progress in engineering alpha/beta hydrolase-fold family members." Biotechnol J **2**(2): 192-200.
- Qian, Z., J. R. Horton, et al. (2009). "Structural redesign of lipase B from *Candida antarctica* by circular permutation and incremental truncation." J Mol Biol **393**(1): 191-201.
- Qian, Z. and S. Lutz (2005). "Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation." J Am Chem Soc **127**(39): 13466-13467.

- Ranganathan, S., Narasimhan SL, Muthukumar K (2008). "An overview of enzymatic production of biodiesel." Bioresour Technol. **99**: 3975-3981.
- Reetz, M. T., B. Brunner, et al. (2004). "Directed evolution as a method to create enantioselective cyclohexanone monooxygenases for catalysis in Baeyer-Villiger reactions." Angew Chem Int Ed Engl **43**(31): 4075-4078.
- Reetz, M. T., S. Prasad, et al. (2010). "Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods." J Am Chem Soc **132**(26): 9144-9152.
- Reetz, M. T. and L. W. Wang (2006). "High-throughput selection system for assessing the activity of epoxide hydrolases." Comb Chem High Throughput Screen **9**(4): 295-299.
- Reitinger, S., Y. Yu, et al. (2010). "Circular permutation of Bacillus circulans xylanase: a kinetic and structural study." Biochemistry **49**(11): 2464-2474.
- Ribeiro, E. A., Jr. and C. H. Ramos (2005). "Circular permutation and deletion studies of myoglobin indicate that the correct position of its N-terminus is required for native stability and solubility but not for native-like heme binding and folding." Biochemistry **44**(12): 4699-4709.
- Rice, J. J., A. Schohn, et al. (2006). "Bacterial display using circularly permuted outer membrane protein OmpX yields high affinity peptide ligands." Protein Sci **15**(4): 825-836.
- Rink, R., J. H. Lutje Spelberg, et al. (1999). "Mutation of Tyrosine Residues Involved in the Alkylation Half Reaction of Epoxide Hydrolase from Agrobacterium radiobacter AD1 Results in Improved Enantioselectivity." Journal of the American Chemical Society **121**(32): 7417-7418.
- Rose, G. D. (2006). "Lifting the lid on helix-capping." Nat Chem Biol **2**(3): 123-124.
- Rothlisberger, D., O. Khersonsky, et al. (2008). "Kemp elimination catalysts by computational enzyme design." Nature **453**(7192): 190-195.

- Rotticci, D., T. Norin, et al. (2000). "An active-site titration method for lipases." Biochim Biophys Acta **1483**(1): 132-140.
- Rui, L., L. Cao, et al. (2004). "Active site engineering of the epoxide hydrolase from *Agrobacterium radiobacter* AD1 to enhance aerobic mineralization of cis-1,2-dichloroethylene in cells expressing an evolved toluene ortho-monooxygenase." J Biol Chem **279**(45): 46810-46817.
- Rui, L., L. Cao, et al. (2005). "Protein engineering of epoxide hydrolase from *Agrobacterium radiobacter* AD1 for enhanced activity and enantioselective production of (R)-1-phenylethane-1,2-diol." Appl Environ Microbiol **71**(7): 3995-4003.
- Sagermann, M., W. A. Baase, et al. (2004). "Relocation or duplication of the helix A sequence of T4 lysozyme causes only modest changes in structure but can increase or decrease the rate of folding." Biochemistry **43**(5): 1296-1301.
- Salameh, M. and J. Wiegel (2007). "Lipases from extremophiles and potential for industrial applications." Adv Appl Microbiol **61**: 253-283.
- Sambrook, J. and D. Russell (2001). Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press.
- Sanders, K. E., J. Lo, et al. (2002). "Intersubunit circular permutation of human hemoglobin." Blood **100**(1): 299-305.
- Savile, C. K., J. M. Janey, et al. (2010). "Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture." Science **329**(5989): 305-309.
- Schubert, M., D. K. Poon, et al. (2007). "Probing electrostatic interactions along the reaction pathway of a glycoside hydrolase: histidine characterization by NMR spectroscopy." Biochemistry **46**(25): 7383-7395.
- Severinov, K. and T. W. Muir (1998). "Expressed protein ligation, a novel method for studying protein-protein interactions in transcription." J Biol Chem **273**(26): 16205-16209.

- Shekhtman, A. (2005). "Protein chemical ligation as an invaluable tool for structural NMR." Protein Pept Lett **12**(8): 765-768.
- Siegel, J. B., A. Zanghellini, et al. (2010). "Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction." Science **329**(5989): 309-313.
- Skjot, M., L. De Maria, et al. (2009). "Understanding the plasticity of the alpha/beta hydrolase fold: lid swapping on the *Candida antarctica* lipase B results in chimeras with interesting biocatalytic properties." Chembiochem **10**(3): 520-527.
- Skrisovska, L., M. Schubert, et al. (2010). "Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins." J Biomol NMR **46**(1): 51-65.
- Solbak, A. I., T. H. Richardson, et al. (2005). "Discovery of pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton fabric." J Biol Chem **280**(10): 9431-9438.
- Steinreiber, A. and K. Faber (2001). "Microbial epoxide hydrolases for preparative biotransformations." Curr Opin Biotechnol **12**(6): 552-558.
- Stemmer, W. P. (1994). "Rapid evolution of a protein in vitro by DNA shuffling." Nature **370**(6488): 389-391.
- Stratton, M. M., D. M. Mitrea, et al. (2008). "A Ca²⁺-sensing molecular switch based on alternate frame protein folding." ACS Chem Biol **3**(11): 723-732.
- Teather, R. M. and P. J. Wood (1982). "Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen." Appl Environ Microbiol **43**(4): 777-780.
- Ton-That, H., G. Liu, et al. (1999). "Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif." Proc Natl Acad Sci U S A **96**(22): 12424-12429.

- Tougaard, P., T. Bizebard, et al. (2002). "Structure of a circularly permuted phosphoglycerate kinase." Acta Crystallogr D Biol Crystallogr **58**(Pt 12): 2018-2023.
- Tsukiji, S. and T. Nagamune (2009). "Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering." Chembiochem **10**(5): 787-798.
- Turcotte, R. F. and R. T. Raines (2008). "Design and characterization of an HIV-specific ribonuclease zymogen." AIDS Res Hum Retroviruses **24**(11): 1357-1363.
- Turner, N. J. (2009). "Directed evolution drives the next generation of biocatalysts." Nat Chem Biol **5**(8): 567-573.
- van Loo, B., J. H. Spelberg, et al. (2004). "Directed evolution of epoxide hydrolase from *A. radiobacter* toward higher enantioselectivity by error-prone PCR and DNA shuffling." Chem Biol **11**(7): 981-990.
- Varadarajan, N., J. Gam, et al. (2005). "Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity." Proc Natl Acad Sci U S A **102**(19): 6855-6860.
- Viguera, A. R., F. J. Blanco, et al. (1995). "The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics." J Mol Biol **247**(4): 670-681.
- Viguera, A. R., L. Serrano, et al. (1996). "Different folding transition states may result in the same native structure." Nat Struct Biol **3**(10): 874-880.
- Wakarchuk, W. W., R. L. Campbell, et al. (1994). "Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase." Protein Sci **3**(3): 467-475.
- Waldo, G. S. (2003). "Genetic screens and directed evolution for protein solubility." Curr Opin Chem Biol **7**(1): 33-38.
- Wang, C. K., Q. Kaas, et al. (2008). "CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering." Nucleic Acids Res **36**(Database issue): D206-210.

- Whitehead, T. A., L. M. Bergeron, et al. (2009). "Tying up the loose ends: circular permutation decreases the proteolytic susceptibility of recombinant proteins." Protein Eng Des Sel **22**(10): 607-613.
- Wong, D. W., S. B. Batt, et al. (2004). "High-activity barley alpha-amylase by directed evolution." Protein J **23**(7): 453-460.
- Wright, G., A. K. Basak, et al. (1998). "Circular permutation of betaB2-crystallin changes the hierarchy of domain assembly." Protein Sci **7**(6): 1280-1285.
- Xu, R., B. Ayers, et al. (1999). "Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies." Proc Natl Acad Sci U S A **96**(2): 388-393.
- Xu, Y., Du W, Liu D (2005). "Study on the kinetics of enzymatic interesterification of triglycerides for biodiesel production with methyl acetate as the acyl acceptor." J Mol Catal B-Enzyme **32**: 241-245.
- Yang, Y. R. and H. K. Schachman (1993). "Aspartate transcarbamoylase containing circularly permuted catalytic polypeptide chains." Proc Natl Acad Sci U S A **90**(24): 11980-11984.
- Zhang, T., E. Bertelsen, et al. (1993). "Circular permutation of T4 lysozyme." Biochemistry **32**(46): 12311-12318.
- Zhao, L., B. Han, et al. (2004). "Epoxide hydrolase-catalyzed enantioselective synthesis of chiral 1,2-diols via desymmetrization of meso-epoxides." J Am Chem Soc **126**(36): 11156-11157.