ENGINEERING OF SYNTHETIC BIODEGRADATION PATHWAY TOWARDS ENVIRONMENTAL POLLUTANT 1,2,3-TRICHLOROPROPANE

by

LU YANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

Written under the direction of

Sagar D. Khare

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2018

ABSTRACT OF THE DISSERTATION

Engineering of Synthetic Biodegradation Pathway towards Environmental Pollutant 1,2,3-Trichloropropane

By LU YANG

Dissertation Director:

Sagar D. Khare

This dissertation describes the application of protein engineering and metabolic engineering approaches for rational redesign of metabolic pathway for biodegradation of important environmental pollutant 1,2,3-trichloropropane (TCP) utilizing both computational and experimental approaches.

A yeast surface display based assay was developed for the limiting enzyme in the pathway - haloalkane dehalogeanse DhaA, which enables high-throughput screening for designed mutants and various strategies were attempted enzyme tunnel engineering on haloalkane dehalogenase DhaA to obtain new starting point towards TCP biodegradation including slot tunnel redesign, *de novo* tunnel design and chimeric protein design with preserved tunnels and generated combinatorial design library subject to high-throuput screening, but unfortunately no design with both decent expression and enhanced activity was detected.

Apart from engineering, the TCP degradation pathway was optimized by metabolic engineering. A phosphorylation- and optically-responsive metabolon for the biodegradation of the environmental pollutant 1,2,3-trichloropropane (TCP) was constructed. The pathway efficiency improvement caused by co-localization in response to stimulus was demonstrated. The design method is modular and generalizable, and could enable spatio-temporal control over a wide variety of synthetic biotransformations.

To summarize, this dissertation presents an innovative concept for rational engineering of a synthetic biodegradation pathway for pollutant compound 1,2,3-trichloropropane and explores design approaches of protein engineering and metabolic engineering.

Acknowledgments

I want to thank my advisor, Prof. Sagar Khare, for his support, patience and encouragement throughout my graduate studies. His guidance, inspiration and passion have grown me into a better scientist.

I would like to thank rest of my thesis committee, Prof. Stephen Burley, Prof. Jean Baum and Prof. Vikas Nanda for their advices and comments guided me through my research towards this thesis. I would like to thank my collaborator Prof. Eduardo Sontag for his help.

I would like to thank current and former members in Khare Lab: Manasi Pethe, Nancy Hernandez, Kristin Blacklock, William Hansen, Elliott Dolan, Dr. Brahm Yachnin, Aliza Rubenstein, Dmitri Zorine, Dr. Srinivas Annavarapu, Sophia Tan, Jason Lin, and other current and past undergraduate researchers. The inspirational and intellectual contributions to my work are much appreciated.

Finally, I would like to thank my parents for providing me endless encouragement and love and my husband Peng for his help and support through my PhD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter1. Introduction

1.1 Metabolic pathway engineering: current strategies and technologies

1.1.1 Overview

Metabolism is a set of life-sustaining biochemical transformations within living cells catalyzed by enzymes which was divided into two categories: biodegradation and biosynthesis[1]. Enzymes catalyzing the cascade reactions do not work alone, instead, they will form enzyme metabolons catalyzing series of enzymatic reactions via metabolic pathways[2]. This field could provide possible solutions for many problems of mankind to overcome for valuable chemical production or for biodegradation of anthropogenic polluting compounds. Biocatalysts offer advantages over traditional catalyst thus like high specificity, high activity and mild operational conditions, which are very attractive for commercial applications[3]. As development of biotechnology, more information and knowledge were gained to understand the structure and function of basic parts of pathways[4][5] and lead to discovery and optimization of inefficient metabolic pathways[6][7].

A metabolic pathway could be synthesized and improved from three steps, the first step is to selection and design of pathway reactions required to convert starting material into desired product[8][9], the second step is to select appropriate natural

enzymes based on known information from database[10][11][12], and last, further pathway optimization and enhancement can be achieved by metabolic engineering and protein engineering both *in vivo* and *in vitro*[13][14](Figure 1.1). Numerous strategies could be used to carry out these three steps[8][15][16][17]. The increasing computational method and knowledge of biological structures allow the biologists and chemists to engineer metabolic pathways and even construct *de novo* biological parts facilitating the metabolic pathway[18][19]. This thesis will mainly focus on the pathway optimization part using metabolic engineering and protein engineering.

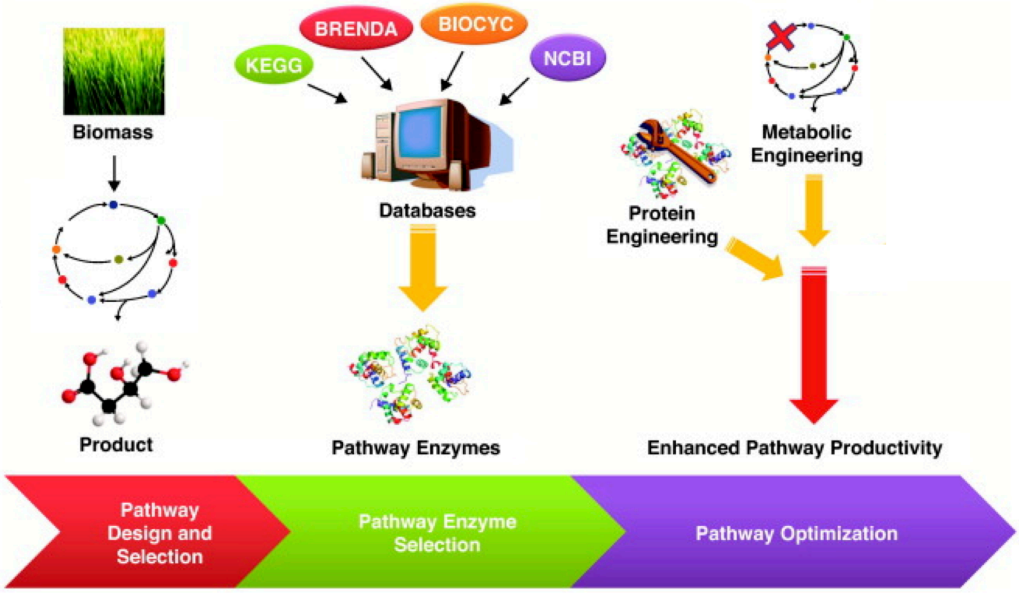

Figure 1.1 Workflow and tools for metabolic pathway engineering (Adopted and modified from reference [8])

1.1.2 Protein engineering

Enzyme is the building block of metabolic pathways, a highly efficient enzyme is able to increase whole pathway up to orders of magnitude. However, natural enzymes in nature can barely meet the requirements nowadays, especially when dealing with anthropogenic chemicals due to lack of time for natural evolution and selection[20]. This enzyme evolution process for "new" reactions could be accelerated in lab by protein engineering[21].

Driven by the need for pure chemicals and detachment from chemically burdened approaches for biosynthesis and biodegradation[22], lots of efforts have been devoted on protein engineering to meet the various demands[14][23], e.g. enhancing protein stability which is essential when applying enzyme in industry[24] improving or altering catalytic properties of enzymes[25] and understanding protein structure-function relationship[26][27]. As improvement of computation technology[28], protein engineering could also accomplish *de novo* enzyme design targeting on desired reactions[29][30]. Enzyme characteristics can be tuned through protein engineering with evolved kinetics and thermostable properties towards target reaction and the tuned enzyme could be further utilized for replacement of the slow biocatalyst in metabolic pathways.

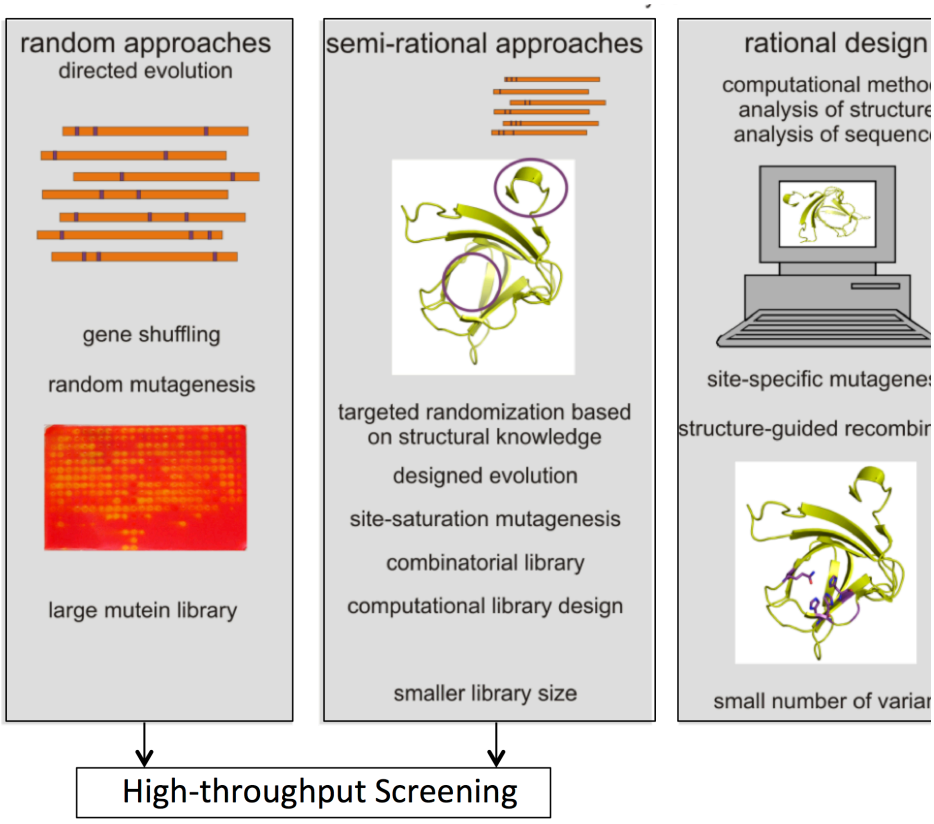

Figure 1.2 Overview of approaches for protein engineering by directed evolution, rational design and semi-rational design (adopted and modified from reference [22]).

Generally there are two distinct strategies for protein engineering: directed evolution and rational design. These two strategies could be combined to semi-rational design (or focused directed evolution) as the third strategy (Figure 1.2). These strategies were developed to allow identification of mutant enzymes obtaining desirable characteristics such as altering specificity, increasing activity, enhancing stability and so on[31][32].

Directed evolution has proven to be a powerful tool for enzyme engineering[33][34] especially when there is no available three-dimensional structure nor the catalytic

mechanism of the enzyme is known. There are two major approaches for directed evolution: introducing random changes in single protein sequence and randomly recombining a set of related sequences[35][36]. Directed evolution can introduce random mutations and result in unexpected protein variants, which could potentially improve the functions. However, these mutations will generate a huge number of mutations for screening, which is extremely time and labor consuming and a fast, cheap high-throughput assay is highly demanding and crucial for successful directed evolution. Commonly used high-throughput assays are usually based on selective growth assays, colorimetric assay or fluorescence activated cell sorting. In chapter 2, a high-throughput assay based on yeast surface display for haloalkane ligand binding to haloalkane dehalogenase DhaA will be discussed to enable designed library screening of haloalkane dehalogenase.

Rational design was established by studying protein structure-function relationship and amino acids interaction of enzymes at beginning of the design access. Instead of introducing random mutations, rational design usually generate proposed mutations towards desired function, which results a significant reduction of the library size. Only a small number of specific designs are required to be assayed, which is especially important when there is no high-throughput assay available.

As development of computational tools, availability of an increasing number of protein structures and reliable simulated models, computational rational design was rapidly developing[22]. Instead of using site-directed mutagenesis experimentally to make the mutants, mutants could be simulated and evaluated on a much larger scale using computational tools with powerful algorithms, e.g. Rosetta, HotSpot, Wizzard and FOLDX[37][38]. With help of computational tools, virtual designs could be screened *in silico* and top ranked designs could be subject to further evaluation experimentally. However, due to lack of thorough understanding of the structure-function relationship of protein, success rate of rational design might be low sometimes for certain enzyme systems.

Semi-rational design (or focused directed evolution) combines advantages of both directed evolution and rational design which results in smaller "smart" libraries that are easier to search through than directed evolution and more likely to yield positive results than rational design[39] by introducing random mutations or certain mutations in targeted regions.

Protein engineering has been shown success on engineering enzymes to enhance performance of the metabolic pathways[40][41]. All these success gives inspiration for improving efficiency of pathway using protein engineering especially when enzymes in the metabolic pathway have very limited activity towards the pathway reactions.

1.1.3 Metabolic engineering

Other than protein engineering, the target metabolic pathway could also be optimized by metabolic engineering[42]. This field was developing rapidly driven by the lack of energy reserves and toxic pollutant accumulation in the environment. Metabolic engineering can fulfill requirements of biopharmaceutical, biofuels, biomaterial production and biodegradation of anthropogenic pollutants. For years, both *in vivo* and *in vitro* metabolic engineering have been established and have been proven to be a powerful approach and are wildly used in academic research and industry manufacturing.

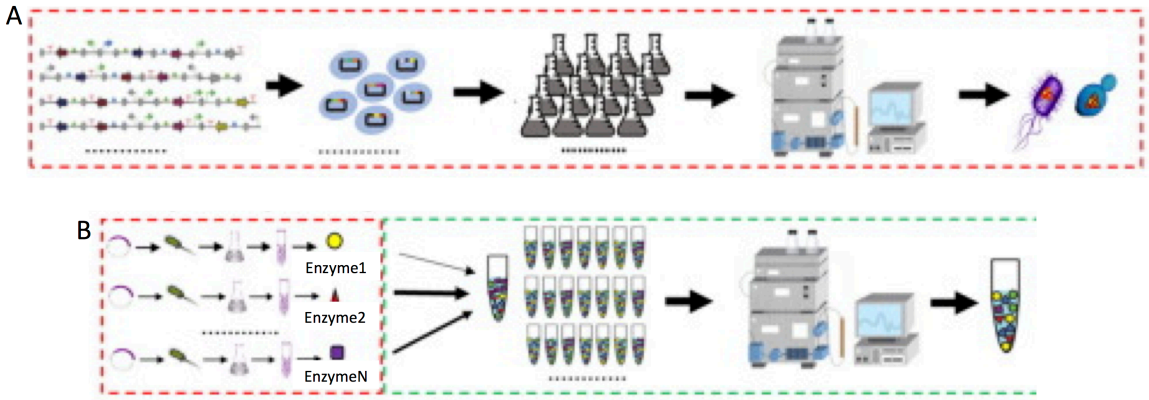

Figure 1.3 Schematic of A) *In vivo* metabolic engineering B) *In vitro* synthetic pathway engineering (adopted and modified from reference [48]).

***In vivo* metabolic engineering** Metabolic engineering could be accomplished *in vivo*[13][43][44] and can be obtained by replacing the inefficient original microorganism to a new host cell, controlling gene expression level by heterologous expression of recombinant enzyme and so on (Figure 1.3A). By optimizing genetic and regulatory

processes of cells, *in vivo* metabolic engineering can improve production yield of native products, increase the load of substrate, or assembly pathway enzymes originating from different organisms into same host and achieved a lot of success [45][46][47]. Despite the potential of utilizing microbial cell for biosynthesis and biodegradation pathway, *in vivo* metabolic engineering can be very complicated. The complexity made the field become highly interdisciplinary together with molecular biology, bioinformatics, microbiology, biochemistry, proteomics, genomics and transcriptomics. Due to the high complexity of *in vivo* metabolic engineering, optimization of rate-limiting metabolic pathways for yield improvement is often difficult to achieve[48].

The principle for ***in vitro*** **metabolic** engineering is using purified individual enzymes in the synthetic metabolic pathway to reconstitute the pathway and conduct its performance *in vitro* (Figure 1.3B). Many pathways have been successfully reconstituted *in vitro*. For example, the biodegradation pathway of 1,2,3-trichloropropane was reconstituted by purifying three enzymes *in vitro*[49] and will be discussed in detail later in this introduction chapter.

In vitro metabolic engineering allows fast pathway prototyping and becomes a powerful tool to study and reveal biochemical mechanisms and kinetics. *In vitro* metabolic engineering can be obtained by post-translation modification of

individual enzymes to improve stability and catalytic activity, e.g. immobilization and encapsulation[49], functional and spatial orchestration of enzyme cascades to achieve substrate channeling and enzyme co-localization effect[50]. Although cell-free synthetic enzyme engineering has been shown to be a powerful method for biochemistry research, tedious work of molecular biology and enzyme purification could be very time consuming and high cost of industry production could result in scale-up problems[48]. However, *in vitro* metabolic engineering system can serve as a model for spatial and functional arrangement of pathway enzymes by the convenience of both structure characterization and activity evaluation.

Although in recent years, metabolic engineering has been rapidly developing and successfully demonstrated the feasibility of carrying out desired metabolic reactions, many challenges still remain, e.g. the field of *in vitro* metabolic engineering still lacks temporal control for multi-enzyme cascade reactions. Tunable controls are essential characteristics for natural pathways, which allows the pathway to response to certain environmental changes as pH, light and temperature. General and optimal strategies establishment was still required to achieve maximum yield or functional control of desired pathway. Additionally, computational tools could also paly an important role to overcome constraints and help with pathway engineer process and is worth further development on metabolic pathway engineering.

1.2 Engineering of the synthetic metabolic pathway of 1,2,3-trichloropropane biodegradation

1.2.1 Overview of 1,2,3-trichloropropane biodegradation pathway

1,2,3-trichloropropane (TCP) is an anthropogenic compound known as pollutant to groundwater. TCP is produced in large amount and world widely used by chemical companies as a solvent, reagent for synthesis of other chemicals and is also formed as a byproduct during the synthesis of epichlorohydrin. Because of its massive production amount into ground water, removing this toxic and carcinogenic compound from the environment is needed. However, there is no existing organism isolated in nature that is capable to degrade TCP into non-toxic compound[51], probably due to TCP is an anthropogenic compound and just recently introduced to the environment[49]. Thus addressing the problem of no efficient way to degrade TCP is essential considering the toxicity of TCP to the environment.

Previous study showed that haloalkane dehalogenase is the enzyme to catalyze the first step which is the dehalogenation step in biodegradation pathways for halogenated pollutant, such as 1,2-dibromoethane and 1,2-dichloroethane[52]. The first haloalkane dehalogenase which is able to dehalogenate TCP (later assigned as DhaA) was isolated and characterized[53][54]. This dehalogenase could convert non-chiral TCP into equalmolar ratio of enamtiomers of 2,3-dichloropropan-1-ol

(DCP).

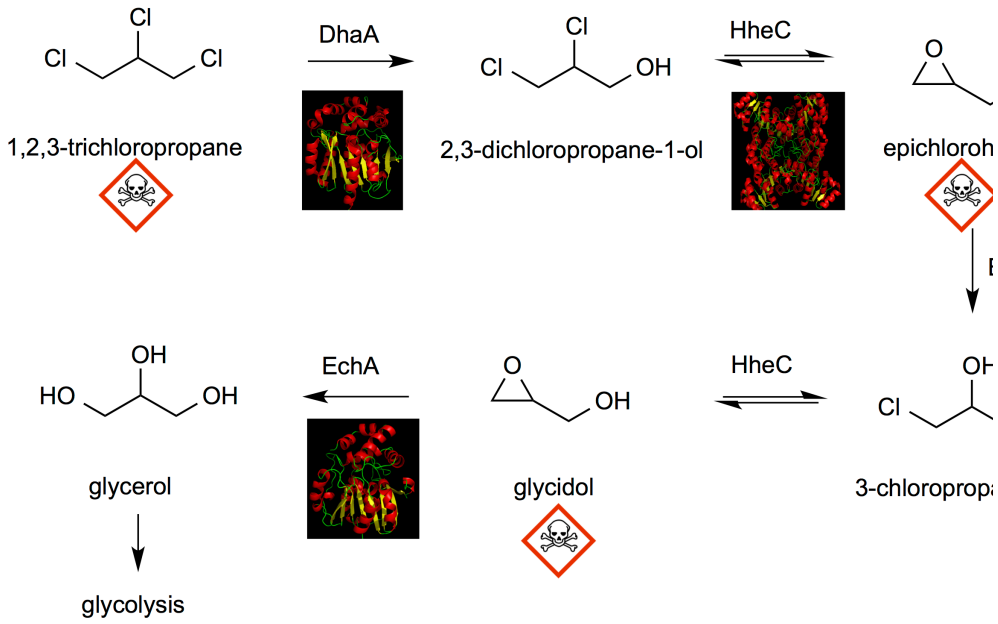

Figure 1.4 Synthetic pathway for the biodegradation of 1,2,3-trichloropropane

A bacterial strain known recently as Agrobacterium radiobacter AD1 was found to be able to convert dihalogenated alcohals, to ECH, then 3-chloropropane-1,2-diol (CPD), then glycidol (GDL) and finally result in non-toxic product glycerol (GDL). Haloalcohol dehalogenase (HheC) and epoxide hydrolase (EchA) were the enzymes catalyzing this series of reactions[55]. The first synthetic metabolic pathway capable of TCP degradation was constructed by heterologous expression wild type DhaA in A.radiobacter AD1 along with HheC and EchA[56]. However, the TCP conversion efficiency is very low due to the low activity of wild-type DhaA with the anthropogenic substrate TCP and the enantioselectivity of haloalcohol dehalogenase.

Rapid development of biotechnology results in more and more powerful approaches for metabolic pathway engineering in the following years, the successful experience and information allows better prediction and design. The development of high-throughput techniques like surface display followed by FACS, microfluidics allow screening for biocatalyst with desired activity and specificity. In this work, both protein engineering and metabolic engineering were attempted for the TCP degradation and will be discussed in the following chapters.

1.2.2 Protein engineering of haloalkane dehalogenase

Previous effects have been made on engineering haloalcohol dehalogenase HheC and epoxide hydrolase EchA[57][58][59] to improve activity, switch enantioselectivity, enhance stability or achieve broader substrate specificity using either rational design or directed evolution methods[60][61]. However, no success was obtained in term of TCP biodegradation. In contrast, engineering dehalogenation activity of haloalkane dehalogenase DhaA towards substrate TCP has been attempted and some achievements have been made. As DhaA has a low efficiency towards TCP dehalogenation, it is actually the rate-limiting step of the degradation pathway so it is essential to engineer haloalkane dehalogenase to achieve high efficiency and specificity towards TCP. This chapter will focus on the structure characterization of haloalkane dehalogeanse and previous work of protein engineering of haloalkane dehalogenase.

The structure and reaction mechanism of haloalkane dehalogeanse has been previously studied in detail. All these information also opens the door for structure-guided protein engineering of haloalkane dehalogenase to achieve desired activity and specificity. Haloalkane dehalogenases are composed by two domains: strictly conserved α/β-hydrolase bottom domain and the helical cap domain. The cap domain consists of several helices connected by loops and is inserted to the main domain by two flexible linkers (Figure 1.5).

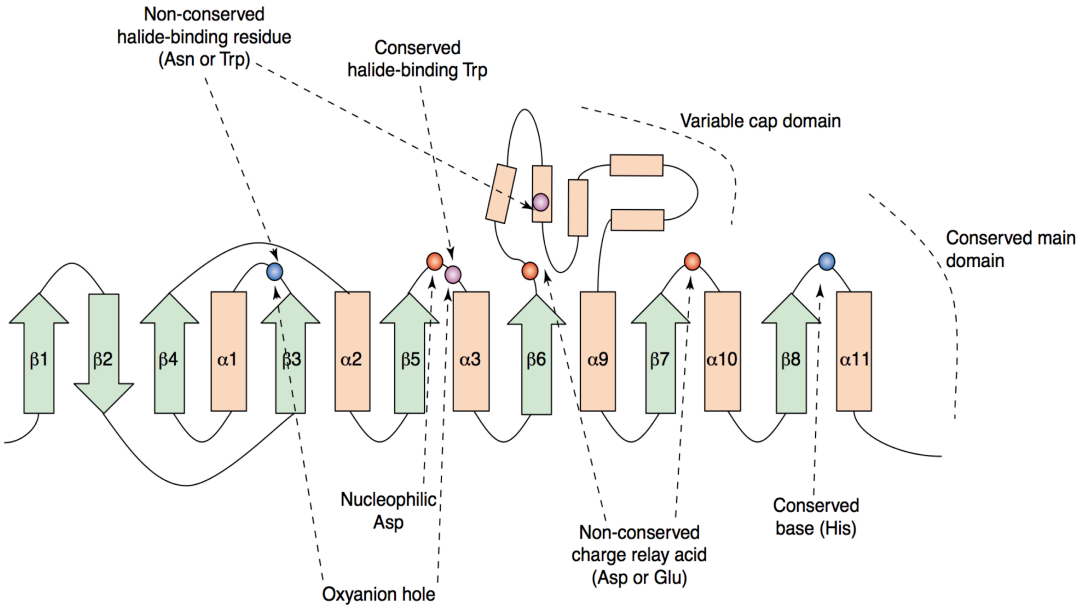

Figure 1.5 General topology of haloalkane dehalogeanse[62](Adopted from reference[62])

The active site of Haloalkane dehalogenase is located at the hydrophobic interface between the bottom domain and the cap domain. The active site is hidden from the

bulk solvent and stayed inside the hydrophobic pocket. A ligand-induced access tunnel controls the substrate entrance, product release and water molecule transport[63][64]. There are four steps in the catalytic mechanism: (1) substrate binding, (2) intermediate formation and halogen leaving, (3) hydrolysis of the alkyl-enzyme intermediate and (4) product release[65][66][67]. Generally,
the substrate binding and intermediate formation are fast and the rate-limiting step differs case by case.

Haloalkane dehalogeanse DhaA was previously engineered to improve conversion of 1,2,3-trichloropropane to 2,3-dichloropropane-1-ol (DCP), which is an essential step in the synthetic TCP biodegradation pathway. DhaA was first improved by error prone PCR, the DhaA variant (C176F) showed a 4-fold improvement in activity with TCP[68]. By using focused directed evolution of residues along the active-site access pathway, a mutant from the activity screening carried five mutations and showed ~30-fold improved catalytic efficiency towards TCP[69]. The importance of the active-site access tunnels for catalysis of haloalkane dehalogeanse was also highlighted in engineering LinB[63]where opening of the novel access tunnel resulted in a more proficient haloalkane dehalogenase.

Despite all the improvements by directed evolution, the prediction of mutations on activity, specificity and stability of the enzymes remains to be challenging[70][71][72][73][74]. A well-developed general computational approach would

significantly reduce the experimental requirement and improve design efficiency towards degradation of toxic environmental pollutant 1,2,3-trichloropropane which is clearly limited by low catalytic properties of wild type DhaA or poor stabilities in organic solvent. To address this problem, attempts of computation-guided active-site access pathway design towards haloalkane dehalogenase DhaA will be discussed in Chapter 3 of this dissertation.

1.2.3Metabolic engineering of TCP degradation pathway

Despite enhancing individual enzyme activity by protein engineering, the TCP degradation pathway could also be improved by metabolic engineering. After the synthetic biodegradation pathway was initially assembled in A. radiobacter AD1, *in vivo* metabolic engineering towards this pathway has been attempted. Due to the enzymes (DhaA, HheC and EchA) in original TCP pathway do not need cofactors to catalyze the reactions in pathway, all of them have been shown enzymatic activities under *in vitro* conditions, thus engineering TCP degradation pathway *in vitro* is an alternative approach, which does not suffer from the limitations of engineered bacterium.

In vivo modular engineering of this biosynthetic multi-enzyme pathway has been carried out by combining duet vectors in single host cell[75]. The expression level of DhaA variants, wild type HheC and wild type EchA can be controlled by different promoter strength and copy number of different plasmids *in vivo*. Stoichiometry

optimization for multi-enzyme TCP pathway was also investigated *in vitro* using kinetic modeling to predict enzyme stoichiometry[76]. The study showed optimizing enzyme stoichiometry *in vitro* could improve the efficiency of the multi-enzyme TCP degradation pathway when using the same amount of enzyme.

In vitro assembly of the synthetic pathway for biodegradation of TCP has been reported by immobilizing enzymes (HheC, EchA and both DhaA and engineered DhaA31) into cross-linked enzyme aggregates and polyvinylalcohol LentiKats lenses[49]. Even though, this strategy is suitable for scale-up and compatible with high input amount of TCP for potential environmental applications, the entire metabolic pathway efficiency is not enhanced.

Although all these data and knowledge gained from previous study provided insights into TCP degradation pathway engineering, however, no efficiency enhancement will be achieved with given amount of enzyme without changing the stoichiometry. These previous studies demonstrated that the original TCP pathway offers advantage as a model system for metabolic engineering: first, all enzymes and metabolites are well characterized; second, no extra cofactors are needed to maintain the activity and last but not the least, the pathway yields well defined nontoxic product. These characteristics were proved to be necessary as a decent model for further strategy design and development on metabolic engineering, which could be generalized for other metabolic pathways and may lead to further

development of better biocatalysts. In Chapter 4, a modular design approach for TCP degradation pathway will be discussed. The designed strategy achieved efficiency improvement on TCP degradation flux and also successfully introduced stimulus-responsiveness to *in vitro* metabolic pathway engineering with temporal control.

1.3 Thesis summary

In this Thesis, both protein engineering and metabolic engineering strategies were applied on TCP biodegradation pathway as a model system. The thesis is composed of three areas of research. Chapter 2 presents the establishment of a high-throughput yeast surface display assay for Haloalkane dehalogeanse DhaA, which could facilitate library screening for enzyme engineering of DhaA. The enzyme engineering focus on active-site access pathway of DhaA is discussed in chapter 3 and is addressed in three different approaches: redesigning slot tunnel, de Novo tunnel design and tunnel transplantation into chimeric protein. Chapter 4 emphasizes on metabolic pathway engineering where we demonstrate that a multi-stimulus responsive multi-enzyme assembly was built using structure-guided and computationally designed modular protein-peptide interaction domains. This multi-enzyme assembly enhanced the whole pathway efficiency. In addition, no previous tunable control was achieved for *in vitro* multi-enzyme cascade

reactions[50]. These engineering strategies could be potentially generalized and applied to other metabolic pathways in the future.

References

[1] N. R. Pace, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 805–8.
[2] S. Schoffelen, J. C. M. van Hest, *Soft Matter* **2012**, *8*, 1736–1746.
[3] M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480–12496.
[4] J. D. WATSON, F. H. CRICK, *Cold Spring Harb. Symp. Quant. Biol.* **1953**, *18*, 123–31.
[5] L. PAULING, R. B. COREY, H. R. BRANSON, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205–11.
[6] M. Meselson, R. Yuan, *Nature* **1968**, *217*, 1110–4.
[7] W. Arber, S. Linn, *Annu. Rev. Biochem.* **1969**, *38*, 467–500.
[8] H. Dhamankar, K. L. Prather, *Curr. Opin. Struct. Biol.* **2011**, *21*, 488–494.
[9] C. H. Martin, D. R. Nielsen, K. V. Solomon, K. L. J. Prather, *Chem. Biol.* **2009**, *16*, 277–286.
[10] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498–2504.
[11] V. Hatzimanikatis, C. Li, J. A. Ionita, C. S. Henry, M. D. Jankowski, L. J. Broadbelt, *Bioinformatics* **2005**, *21*, 1603–1609.
[12] A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi, H. Kitano, *Proc. IEEE* **2008**, *96*, 1254–1265.
[13] J. Nielsen, M. Fussenegger, J. Keasling, S. Y. Lee, J. C. Liao, K. Prather, B. Palsson, *Nat. Chem. Biol.* **2014**, *10*, 319–322.
[14] C. M. Pirie, M. De Mey, K. L. J. Prather, P. K. Ajikumar, *ACS Chem. Biol.* **2013**, *8*, 662–672.
[15] W. B. Copeland, B. A. Bartley, D. Chandran, M. Galdzicki, K. H. Kim, S. C. Sleight, C. D. Maranas, H. M. Sauro, *Metab. Eng.* **2012**, *14*, 270–80.
[16] M. H. Medema, R. van Raaphorst, E. Takano, R. Breitling, *Nat. Rev. Microbiol.* **2012**, *10*, 191–202.
[17] C. N. S. Santos, G. Stephanopoulos, *Curr. Opin. Chem. Biol.* **2008**, *12*, 168–176.
[18] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, et al., *Science (80-.).* **2010**, *329*, 52–56.
[19] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, et al., *Science (80-.).* **2008**, *319*, 1387–1391.
[20] J. A. Gerlt, P. C. Babbitt, *Curr. Opin. Chem. Biol.* **2009**, *13*, 10–8.

[21] R. A. Chica, *Protein Sci.* **2015**, *24*, 431–3.
[22] K. Steiner, H. Schwab, *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209010.
[23] B. R. Glick, *Biotechnol. Adv.* **1995**, *13*, 247–61.
[24] V. G. H. Eijsink, A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. van den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105–120.
[25] J.-L. Jestin, P. A. Kaminski, *J. Biotechnol.* **2004**, *113*, 85–103.
[26] W. Schliebs, N. Thanki, R. Jaenicke, R. K. Wierenga, *Biochemistry* **1997**, *36*, 9655–9662.
[27] C. Nardella, A. Lahm, M. Pallaoro, M. Brunetti, A. Vannini, C. Steinkühler, *Biochemistry* **2004**, *43*, 1862–1873.
[28] J. Damborsky, J. Brezovsky, *Curr. Opin. Chem. Biol.* **2009**, *13*, 26–34.
[29] D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci.* **2001**, *98*, 14274–14279.
[30] J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11566–70.
[31] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.
[32] T. Davids, M. Schmidt, D. Böttcher, U. T. Bornscheuer, *Curr. Opin. Chem. Biol.* **2013**, *17*, 215–220.
[33] S. Brakmann, *Chembiochem* **2001**, *2*, 865–71.
[34] C. Neylon, *Nucleic Acids Res.* **2004**, *32*, 1448–59.
[35] R. C. Cadwell, G. F. Joyce, *PCR Methods Appl.* **1992**, *2*, 28–33.
[36] W. P. C. Stemmer, *Nature* **1994**, *370*, 389–391.
[37] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St.Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, et al., *Science (80-.).* **2010**, *329*, 309–313.
[38] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, et al., *Nature* **2008**, *453*, 190–195.
[39] H. Jochens, U. T. Bornscheuer, *ChemBioChem* **2010**, *11*, 1861–1866.
[40] K. Zhang, H. Li, K. M. Cho, J. C. Liao, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6234–9.
[41] E. Leonard, P. K. Ajikumar, K. Thayer, W.-H. Xiao, J. D. Mo, B. Tidor, G. Stephanopoulos, K. L. J. Prather, *Proc. Natl. Acad. Sci.* **2010**, *107*, 13654–13659.
[42] J. E. Bailey, *Science* **1991**, *252*, 1668–75.
[43] Y.-T. Yang, G. N. Bennett, K.-Y. San, *Electron. J. Biotechnol.* **1998**, *1*, 134–141.
[44] G. Stephanopoulos, *ACS Synth. Biol.* **2012**, *1*, 514–525.
[45] P. Nawabi, S. Bauer, N. Kyrpides, A. Lykidis, *Appl. Environ. Microbiol.* **2011**, *77*, 8052–8061.
[46] P. Xu, A. Vansiri, N. Bhan, M. A. G. Koffas, *ACS Synth. Biol.* **2012**, *1*, 256–266.
[47] J. Wu, G. Du, J. Zhou, J. Chen, *Metab. Eng.* **2013**, *16*, 48–55.
[48] W. Guo, J. Sheng, X. Feng, *Comput. Struct. Biotechnol. J.* **2017**, *15*, 161–167.

[49] P. Dvorak, S. Bidmanova, J. Damborsky, Z. Prokop, *Environ. Sci. Technol.* **2014**, *48*, 6859–6866.
[50] C. Schmidt-Dannert, F. Lopez-Gallego, *Microb. Biotechnol.* **2016**, *9*, 601–609.
[51] I. Conference, O. N. Chemistry, *Env. Sci Pollut Res* **2012**, *19*, 3067–3078.
[52] T. Koudelakova, E. Chovancova, J. Brezovsky, M. Monincova, A. Fortova, J. Jarkovsky, J. Damborsky, *Biochem. J.* **2011**, *435*.
[53] T. Yokota, T. Omori, T. Kodama, *J. Bacteriol.* **1987**, *169*, 4049–54.
[54] A. N. Kulakova, M. J. Larkin, L. A. Kulakov, *Microbiology* **1997**, *143*, 109–115.
[55] A. J. Van Den Wijngaard, D. B. Janssen, B. Witholt, *Microbiology* **1989**, *135*, 2199–2208.
[56] T. Bosma, E. Kruizinga, E. J. De Bruin, G. J. Poelarends, D. B. Janssen, *Appl. Environ. Microbiol.* **1999**, *65*, 4575–4581.
[57] R. Rink, M. Fennema, M. Smids, U. Dehmel, D. B. Janssen, *J. Biol. Chem.* **1997**, *272*, 14650–7.
[58] M. Nardini, I. S. Ridder, H. J. Rozeboom, K. H. Kalk, R. Rink, D. B. Janssen, B. W. Dijkstra, *J. Biol. Chem.* **1999**, *274*, 14579–86.
[59] J. E. van Hylckama Vlieg, L. Tang, J. H. Lutje Spelberg, T. Smilda, G. J. Poelarends, T. Bosma, A. E. van Merode, M. W. Fraaije, D. B. Janssen, *J. Bacteriol.* **2001**, *183*, 5058–66.
[60] L. Rui, L. Cao, W. Chen, K. F. Reardon, T. K. Wood, *J. Biol. Chem.* **2004**, *279*, 46810–7.
[61] L. Rui, L. Cao, W. Chen, K. F. Reardon, T. K. Wood, *Appl. Environ. Microbiol.* **2005**, *71*, 3995–4003.
[62] D. B. Janssen, *Curr. Opin. Chem. Biol.* **2004**, *8*, 150–159.
[63] J. Brezovsky, P. Babkova, O. Degtjarik, A. Fortova, A. Gora, I. Iermak, P. Rezacova, P. Dvorak, I. K. Smatanova, Z. Prokop, et al., *ACS Catal.* **2016**, *6*, 7597–7610.
[64] M. Otyepka, J. Damborský, *Protein Sci.* **2009**, *11*, 1206–1217.
[65] T. Bosma, M. G. Pikkemaat, J. Kingma, J. Dijk, D. B. Janssen, **n.d.**, DOI 10.1021/bi026907m.
[66] J. P. Schanstra, D. B. Janssen, *Biochemistry* **1996**, *35*, 5624–5632.
[67] Z. Prokop, M. Monincová, R. Chaloupková, M. Klvana, Y. Nagata, D. B. Janssen, J. Damborský, *J. Biol. Chem.* **2003**, *278*, 45094–100.
[68] K. A. Gray, T. H. Richardson, K. Kretz, J. M. Short, F. Bartnek, R. Knowles, L. Kan, P. E. Swanson, D. E. Robertson, *Adv. Synth. Catal.* **2001**, *343*, 607–617.
[69] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727–733.
[70] R. K. Kuipers, H.-J. Joosten, W. J. H. van Berkel, N. G. H. Leferink, E. Rooijen, E. Ittmann, F. van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, et al., *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, NA-NA.
[71] R. J. Floor, H. J. Wijma, D. I. Colpa, A. Ramos-Silva, P. A. Jekel, W. Szymański, B.

L. Feringa, S. J. Marrink, D. B. Janssen, *ChemBioChem* **2014**, *15*, 1660–1672.
[72] D. Bednar, K. Beerens, E. Sebestova, J. Bendl, S. Khare, R. Chaloupkova, Z. Prokop, J. Brezovsky, D. Baker, J. Damborsky, *PLOS Comput. Biol.* **2015**, *11*, e1004556.
[73] J. Bendl, J. Stourac, E. Sebestova, O. Vavra, M. Musil, J. Brezovsky, J. Damborsky, *Nucleic Acids Res.* **2016**, *44*, W479–W487.
[74] A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, et al., *Mol. Cell* **2016**, *63*, 337–346.
[75] P. Dvorak, L. Chrast, P. I. Nikel, R. Fedr, K. Soucek, M. Sedlackova, R. Chaloupkova, V. de Lorenzo, Z. Prokop, J. Damborsky, *Microb. Cell Fact.* **2015**, *14*, 201.
[76] P. Dvorak, N. P. Kurumbang, J. Bendl, J. Brezovsky, Z. Prokop, J. Damborsky, *ChemBioChem* **2014**, *15*, 1891–1895.

CHAPTER 2. High-throughput yeast surface display for haloalkane dehalogenase library screening.

2.1 Introduction

With help of computational protein engineering tools, the number of designs to be tested can be dramatically eliminated. However, given the limit of understanding of protein sequence-function relationships, it's still difficult to achieve desired improvements by testing only several mutants through rational design. A design library and directed evolution are still necessary for obtaining desired activity and specificity towards enzyme engineering. Thus, a high-throughput assay for haloalkane dehalogenase towards 1,2,3-trichloropropane degradation is desired to screen mutants from design library and directed evolution. Previous study has established colorimetric assays that can evaluate activity of haloalkane dehalogenase by measuring released halide ions using mercuric thiocyanate and ferric ion[1] or measuring decrease of pH of a weakly-buffered medium when protons and chloride ions are released by the enzyme[2]. Although the pH-indicated colorimetric assay could be performed in 96-well plate and screen a small size of library up to thousands[3], it's very labor and time consuming. Neither of them could fulfill the requirements to screen a larger size library up to 10^4 in an efficient way. Therefor, we developed a yeast surface display based high-throughput assay which enables large size mutant library screening for haloalkane

dehalogenase DhaA.

Yeast surface display is a particularly powerful platform for protein engineering and has been used to engineer a wide variety of proteins to improve binding affinity, stability and catalytic activity[4]. The yeast surface display library size can reach to 10^7 - 10^9 [5] and the display level of a protein of yeast indicates both thermal stability as well as soluble expression level[6] and the mutants can be easily screened by flow cytometry and isolated by fluorescence-activated cell sorting (FACS). In order to perform yeast display on haloalkane dehalogenase, the substrate need to stay bound to the active site of the enzyme instead of completely converting into the final product and getting released into the bulk.

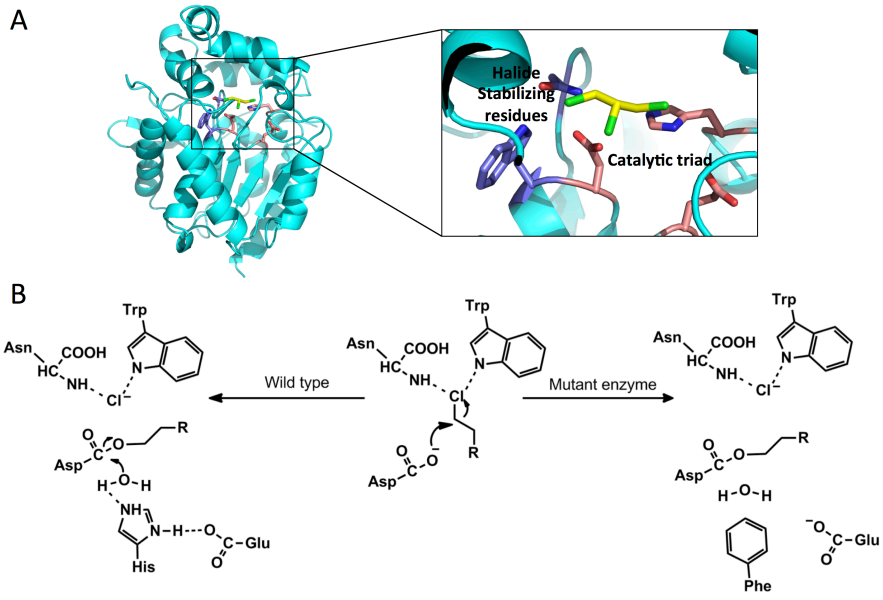

Figure 2.1 A) Structure of haloalkane dehalogenase DhaA (Protein Data Bank ID: 1BN6) and halide stabilizing residues and catalytic triad are highlighted. **B)** Dehalogenation mechanism of haloalkane dehalogenase DhaA for wild type and H272F mutation.

Aspartate carboxylate of the haloalkane dehalogeanse catalytic triad (Figure 2.1 A) attacks the halogen-substituted carbon atom, the leaving halogen were stabilized by stabilizing residues (Figure 2.1 A). The resulting covalent intermediate is an ester, which is hydrolyzed by a water molecule activated by His289[7]. By mutating catalytic histidine to phenylalanine, the water molecule will not be able
to be activated to hydrolase the ester bond between ligand and the enzyme and the ligand could covalently bind to haloalkane[8] (Figure 2.1B). With certain chemical modifications, e.g. biotinylation or attaching a fluorophore to the substrate molecule to enable staining of the substrate, we can detect the binding of the substrate to the enzyme on the substrate by yeast surface display. The surface display technique will allow us to perform screening on design libraries with large size using flow cytometry and conduct cell sorting utilizing FACS.

2.2 Experimental methods

Gene cloning and mutagenesis The wild type DhaA was a gift from Dr. Jiri Damborsky and was sub-cloned between NdeI/XhoI cloning sites of pETCON2 vector. The H272F mutation was introduced to DhaA_M construct using Kunkel mutagenesis and additional K175M/C176G/Y273L mutations were introduced using Kunkel mutagenesis to make DhaA_halo construct.

Yeast surface display Both DhaA_M and DhaA_halo were then chemically transformed into the EBY100 competent yeast strain and plated on selective complete (SC) media (20 g/L glucose) with a selective amino acid mix (–Trp, –Ura) at 30 °C. After 2 days of growth, a single colony was transferred to a 2 mL SC media culture tube supplemented with 2 μL of 1000 × antibiotics (carbenicillin, kanamycin) at 30 °C. The growth cultures were incubated for ~ 24 h (OD600 2.0–3.0) in a 30 °C shaking incubator. 1.5 × 10^7cells (OD600 ~ 0.5) were pelleted and resuspended in 2 mL induction media (20 g/L galactose, 2 g/L glucose) supplemented with 2 μL each of 1000 × antibiotics (carbenicillin, kanamycin). The induction cultures were grown overnight at 30 °C to an OD600 of 1–1.5. All spins in the protocol were done at 3000 rcf for 5 min. The induced cultures were pelleted and washed with 500 μL PBS followed by 500 μL PBS + 0.5% BSA. 5mM HaloTag PEG-Biotin Ligand (Promega, USA) was diluted in 1XPBS to 50 μM. 2 μL 50 μM HaloTag ligand was incubated on a rotater with 5 x 10^6 cells in 50 μL for two hours at room temperature. The samples were pelleted and resuspened with 500 μL PBS with 0.5% BSA to wash out free HaloTag ligand and the washing process was repeated one more time. For haloalkane inhibition assay, 5 x 10^6 cells in 50 μL were incubated with desired concentration of haloalkane for 2 hours before adding HaloTag ligand. No washing step was added between haloalkane and haloTag incubation. 1 μL of anti-cMyc-FITC (Genscript, USA), and 2 μL of Strptavidin-PE (Biolegend, USA) was incubated with 5 X 10^6 cells for 30 min at 4 °C. The samples were

resuspended by vortexing and incubated at room temperature for an additional 30 min. The cells were washed with 500 μL PBS with 0.5% BSA, pelleted and then resuspended in 500 μL 1 x PBS. Samples were diluted to achieve a final concentration of 10^6 cells/mL and then FITC (anti-cMyc) and PE (streptavidin) intensities were detected using a Flow Cytometer (Beckman Coulter Gallios). The signal gate was set based on unstained cells.

2.3 Experimental results

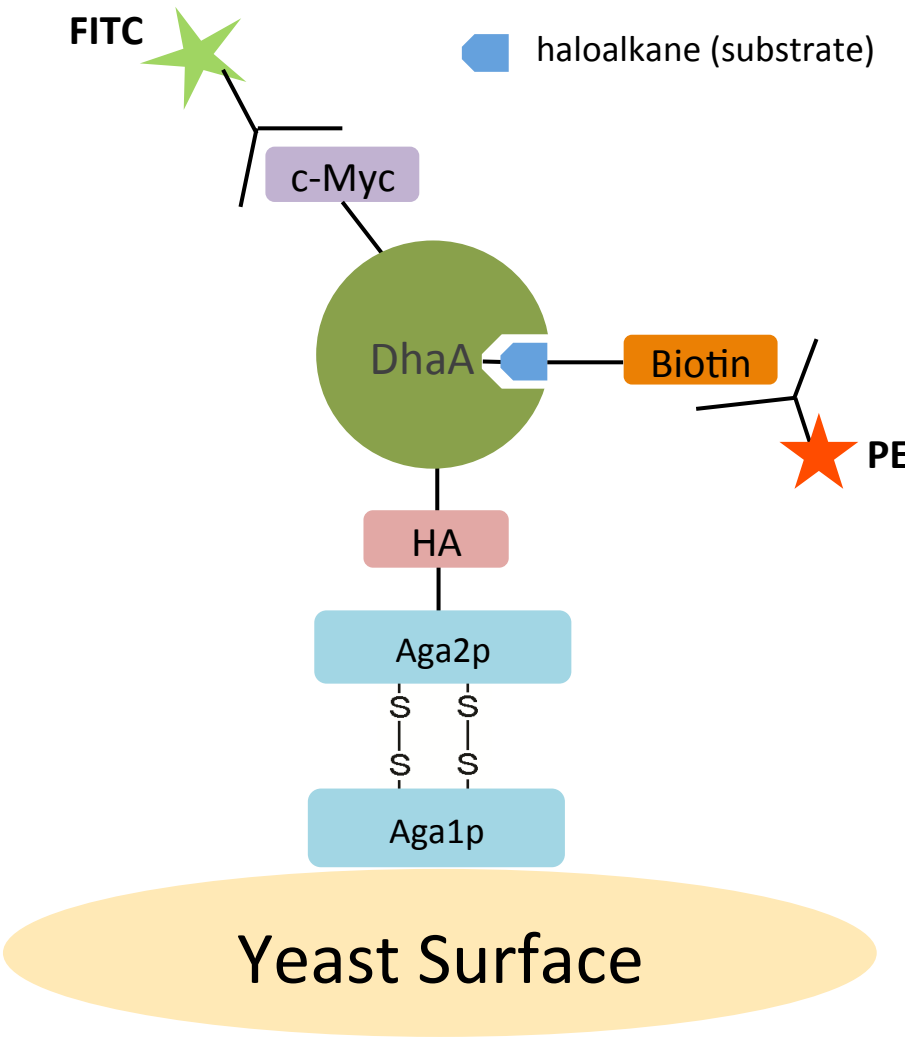

Figure 2.2 An illustration of yeast surface display of haloalkane dehalogenase. FITC intensity indicates protein expression level and PE intensity corresponds to biotinylated HaloTag ligand binding.

Binding of the commercial HaloTag ligand, which is a chloroalkane attached to a biotin via a PEG linker, to DhaA was probed by yeast surface display. The mutant DhaA_M with H272F mutation was subcloned to pETCON2 vector[9] and expressed on the surface of yeast. The anti-c-Myc-FITC signal intensity indicates the expression level of the protein on yeast surface and the streptavidin-PE signal intensity

indicates the binding level of the Halotag ligand to the enzyme (Figure 2.2).

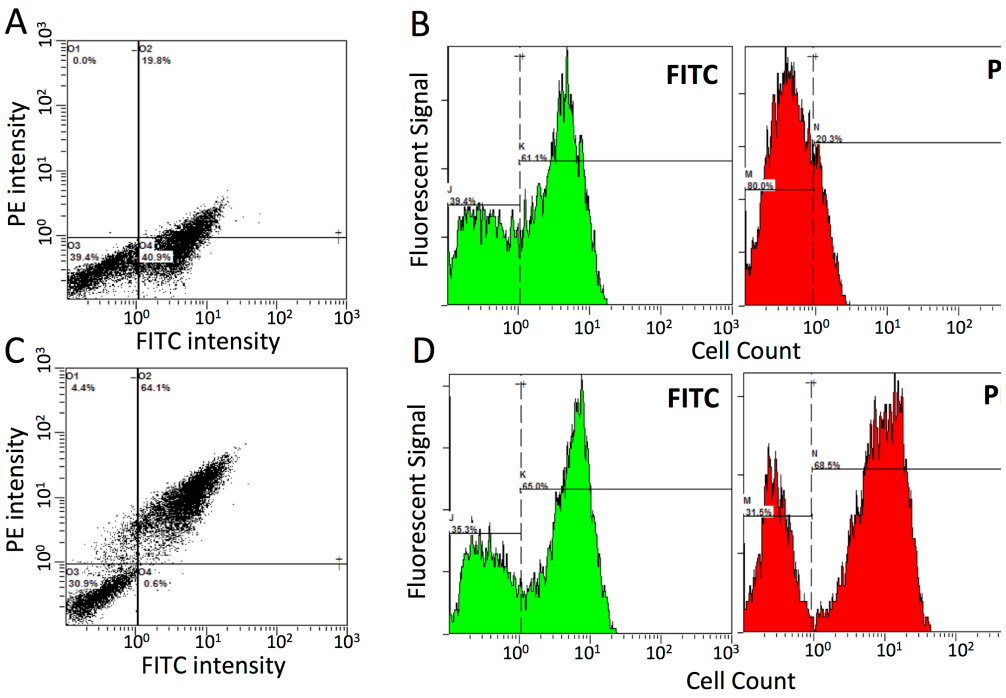

Figure 2.3 A) Flow cytometry analysis of HaloTag ligand binding to DhaA_M **B)** Flow cytometry histograms of Anti-c-Myc-FITC signal and streptavidin-PE signal of DhaA_M **C)** Flow cytometry analysis of HaloTag binding of DhaA_halo. **D)** Flow cytometry histograms of Anti-c-Myc-FITC signal and streptavidin-PE signal for DhaA_halo.

Both DhaA_M and DhaA_halo were expressed on the surface of yeast and analyzed by flow cytometry to detect both expression and Halotag ligand binding activity. Both DhaA_M and DhaA_halo showed decent expression level on yeast surface, which agrees with the good expression level in *E. coli.* However, compared with DhaA_halo, DhaA_M has less expressed cells that passed the gate for PE intensity (Figure 2.3A, 2.3B), indicating a weaker binding affinity of DhaA_M. With additional K175M/C176G/Y273L mutations introduced to DhaA_M to make the main tunnel open, the binding affinity to HaloTag ligand can be dramatically improved[8] by the

construct DhaA_halo. With similar level of expression, DhaA_halo showed a much higher PE intensity than DhaA_M, almost all expressed cells passed gate (Figure 2.3C, 2.3D), which agrees with previous study[8], indicating that yeast surface display could used to detect binding affinity with haloalkanes.

Theoretically, all haloalkanes can be chemically modified to attach a biotin or a fluorophore, however, the customized modification sometimes can become tricky or expensive. Here we demonstrated a two-step HaloTag ligand inhibition assay which could detect the binding of haloalkane substrates without chemical modifications. Yeast cells with DhaA_halo expressed on surface were incubated with substrate of interest first, and then incubated with HaloTag ligand. If the active site is occupied by haloalkane substrate, it will covalently bind to haloalkne and will no longer be able to bind to HaloTag ligand. By probing inhibition of HaloTag ligand, the binding of haloalkane substrate could be detected on yeast surface.

In this case, 1,2,3-trichloropropane is the substrate of interest and 1-chlorobutane know having a higher binding affinity was chosen as a control. The flow cytometry results of the two-step inhibition assay showed that 1-chlorobutane completely inhibited HaloTag ligand binding (Figure 2.4A), while TCP could only partially inhibit HaloTag ligand binding (Figure 2.4B). The yeast surface display of HaloTag ligand inhibition result agrees with previous study that 1CB is a better substrate to DhaA and has higher binding affinity (lower Km) than TCP[10].

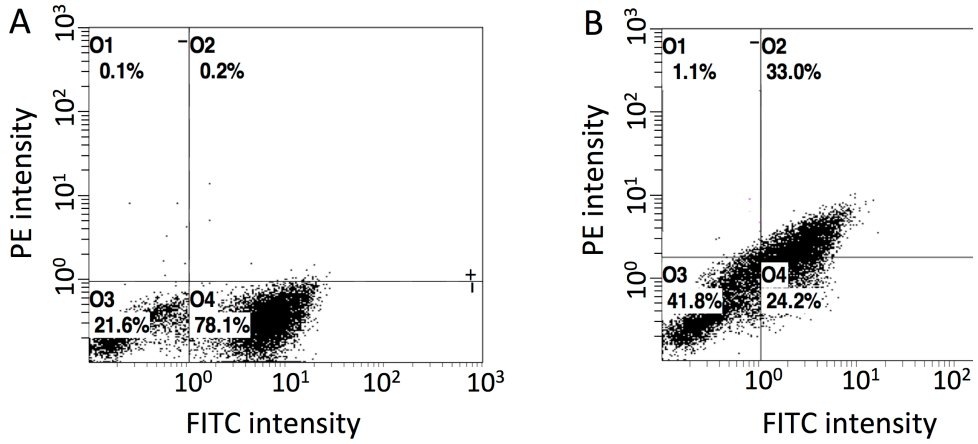

Figure 2.4 Flow Cytometry analysis of HaloTag ligand binding inhibition by **A)** 1-chlorobutane (1CB) and **B)** 1,2,3-tricloropropane (TCP) on DhaA_halo.

To further investigate the haloalkane inhibition kinetics, a range of apparent concentrations (20mM, 10mM, 4mM, 2mM, 1mM, 0.2mM) of 1-chlorobutane and TCP were used to incubate with yeast cells to inhibit HaloTag ligand binding. 1-chlorobutane started to show inhibition at a lower concentration compared with TCP (Figure 2.5). IC50 values for HaloTag ligand binding are 9.0mM for 1-chlorobutane and 21.4mM for TCP, which are proportional to the Km values for 1-chlorobutane (0.4mM) and TCP (2.2mM)[10], indicating this inhibition assay is able to evaluate haloalkane substrate based on their binding affinity to the enzyme. This inhibition assay enables high throughput screening for not only large size of DhaA mutants, but also on a variety of haloalkane substrates.

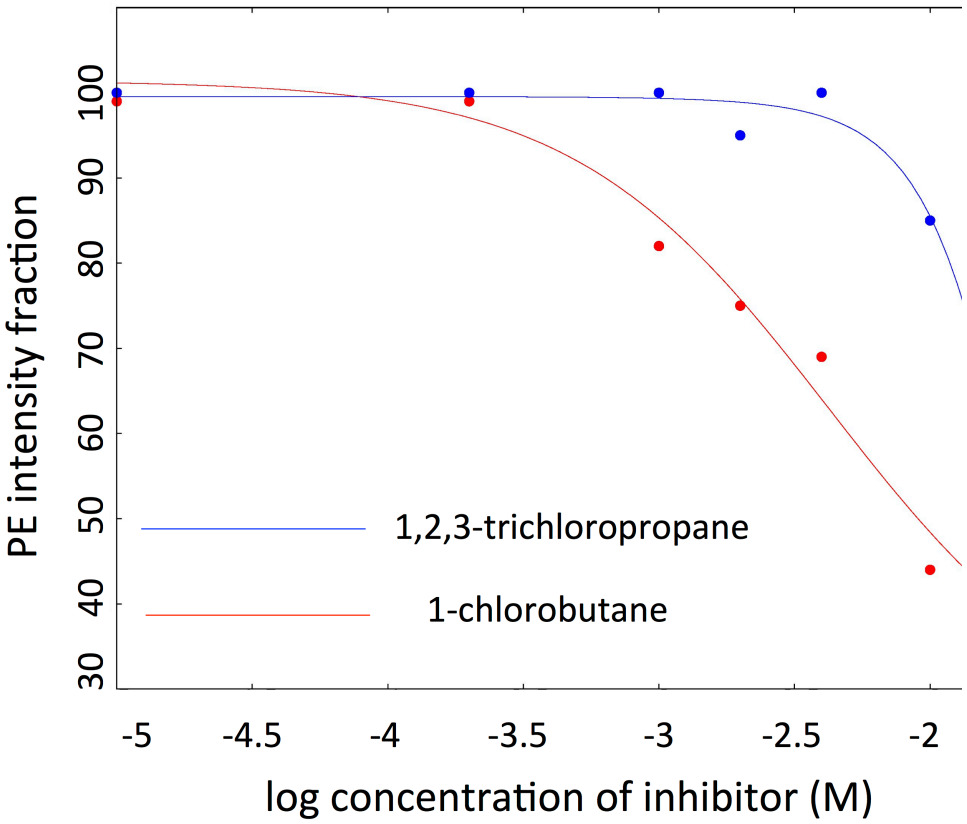

Figure 2.5. Inhibition profile of HaloTag ligand binding by 1-chlorobutane and 1,2,3-trichloropropane, respectively. IC50 was determined by Python Numpy.

2.4 Discussion

In this chapter, we established a yeast surface display assay for haloalkane dehalogeanse DhaA, which enables high-throughput screening for haloalkane dehalogenase engineering by flow cytometry. This high-throughput assay could successfully detect the mutant expression level and could also evaluate binding affinity level by using substrate attached with fluorophore. It also allows a two-step HaloTag ligand inhibition assay to detect binding of viable haloalkane substrates without chemical modifications. This yeast surface based high-throughput assay is rapid, quantitative and could fulfill the requirements to

screen a large size mutant library of DhaA up to $1x10^7$, which none of the colorimetric based assays could satisfy this requirement.

However, the yeast surface display could only evaluate binding affinity of substrate and could not provide information about product turnover, so it will not
be able to distinguish mutants with similar Km but different Kcat but this shortcoming could be overcome by measuring enzymatic kinetics of the mutants by colorimetric assay once the library size is dramatically eliminated after cell sorting. Second, to utilize the inhibition assay, mutations were introduced at main tunnel in DhaA_halo to accommodate the HaloTag ligand with high affinity for inhibition assay, which limits screening and engineering of the main tunnel region of DhaA and this could be solved by utilizing the fluorophore attached haloalkane substrates.

In summary, we developed a high-throughput yeast surface display which dramatically increased the screening efficiency for haloalkane dehalogeanse DhaA thus facilitating protein engineering of haloalkane dehalogenase towards biological pollutant degradation.

References

[1] I. Iwasaki, S. Utsumi, T. Ozawa, *Bull. Chem. Soc. Jpn.* **1952**, *25*, 226–226.
[2] P. Holloway, J. T. Trevors, H. Lee, **1998**, *32*, 31–36.
[3] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727–733.
[4] S. A. Gai, K. D. Wittrup, *Curr. Opin. Struct. Biol.* **2007**, *17*, 467–473.

[5] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, K. D. Wittrup, *Nat. Protoc.* **2006**, *1*, 755–768.
[6] E. V Shusta, M. C. Kieke, E. Parke, D. M. Kranz, K. D. Wittrup, *J. Mol. Biol.* **1999**, *292*, 949–956.
[7] J. Newman, T. S. Peat, R. Richard, L. Kan, P. E. Swanson, J. A. Affholter, I. H. Holmes, J. F. Schindler, C. J. Unkefer, T. C. Terwilliger, *Biochemistry* **1999**, *38*, 16105–14.
[8] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, et al., *ACS Chem. Biol.* **2008**, *3*, 373–382.
[9] K. Jacoby, M. Metzger, B. W. Shen, M. T. Certo, J. Jarjour, B. L. Stoddard, A. M. Scharenberg, *Nucleic Acids Res.* **2012**, *40*, 4954–64.
[10] T. Bosma, M. G. Pikkemaat, J. Kingma, J. Dijk, D. B. Janssen, **n.d.**, DOI 10.1021/bi026907m.

Chapter 3. Enzyme engineering of haloalkane dehalogenase DhaA: To create highly efficient and selective biocatalysts using computational enzyme active-site access pathway design

3.1 Overview of enzyme active site access design

Ligands transportation between bulk solvent and active site is essential for enzyme catalysis especially when active sites are buried. The transport tunnel allows substrate entry, product release and the physicochemical properties have significant impact on exchange rates of substrates or product between active sites and solvent environment [1][2] thus effecting activity and specificities. Tunnels could have also other sophisticated functions, for example, it could prevent enzyme damage from transition metal and toxic cellular intermediate poisoning[3][4], enables reactions that requires absence and water[5] and could also synchronize reactions involving multiple substrates, cofactors and intermediates[6][7]. Some enzyme tunnels could also show dynamic properties and shift between on and off status.

Previous study showed that enzymatic catalysis could be prompted by altering activity, specificity and stability via protein engineering [1][8][9][10][11][12][13], however, currently there are no existing general methods for rational enzyme engineering through protein tunnel design to overcome natural limitations. To address this problem, in this chapter, we tried three different strategies for enzyme tunnel engineering on DhaA haloalkane dehalogenase targeting on activity and specificity improvement towards dehalogenation of 1,2,3-trichloropropane: 1) Computer assisted tunnel redesign 2) Computational *de novo* tunnel design and 3) computational chimeric protein design with preserved enzyme active-site access

pathway.

3.2 Computer assisted tunnel redesign to modulate specificity

3.2.1 Introduction

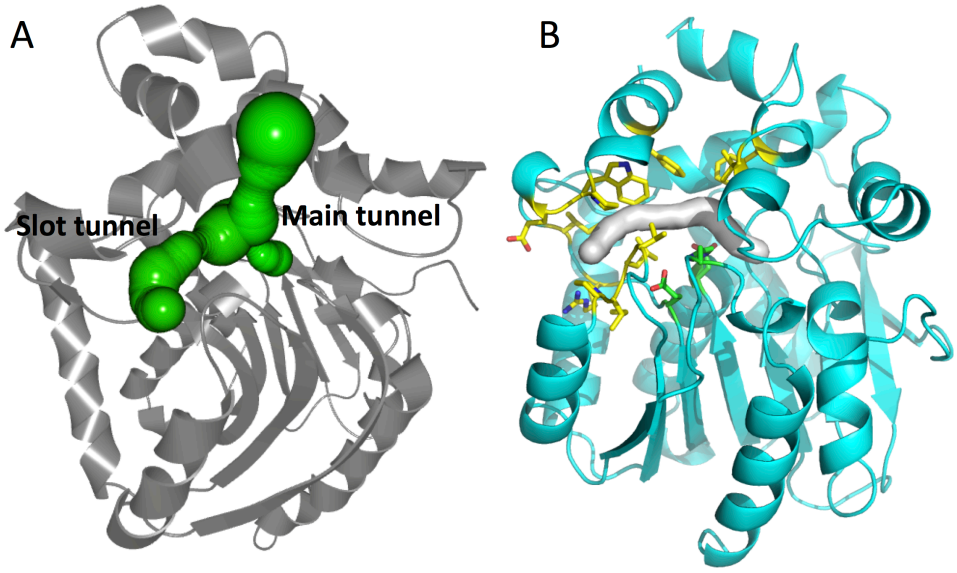

Figure 3.1 A) Active site access pathways observed in computer simulations with wild type DhaA utilizing Caver. Main tunnel and slot tunnel was highlighted in green. **B)** Slot tunnel (silver) surrounding residues were determined using Caver and were highlighted in yellow.

The active site of DhaA is a hydrophobic cavity located between cap domain and bottom domain[14]. There are two major tunnels leading to the active site: the main tunnel and the slot tunnel[15] (Figure 3.1A). In previous study, a double mutant (C176Y Y273F) was found to obtain 4 times more activity towards TCP by directed evolution[16]. Both mutations make the active site cavity less open to the bulk and block the main tunnel, further more, the efficiency could be improved by ~30 fold by narrowing the slot tunnel by saturated mutagenesis at focused region along with

the main tunnel double mutations[1]. It appears likely that the catalytic rate increase is caused by decreasing accessibility of active site for water molecules based the previous study. Based on this result, we hypothesized that by further engineering the slot tunnel to be more hydrophobic could improve the activity of the enzyme by the same mechanism.

3.2.2 Computational design

By calculating tunnels using CAVER[17], residues around the slot tunnel were located. *In silico* saturated mutagenesis was used to screen all tunnel neighbor positions based on total score, mutations that meet two requirement will be allowed for further design: 1) The mutation won't increase the total score compared with wild type enzyme thus unlikely to destabilize the structure, 2) The mutation should be bulky and hydrophobic. Designs with identified mutations were generated utilizing RosettaDesign[18]. Five thousands of generated decoys were scored based on active site interface energy and tunnel residue energy. Mutations from 10% top ranked designs were analyzed. Seven mutations were picked based on high percentage of appearance from the top ranked designs (Table 3.1) and will be further used to make combinatorial library and screened by yeast surface display assay described in chapter2 (Figure 3.2)

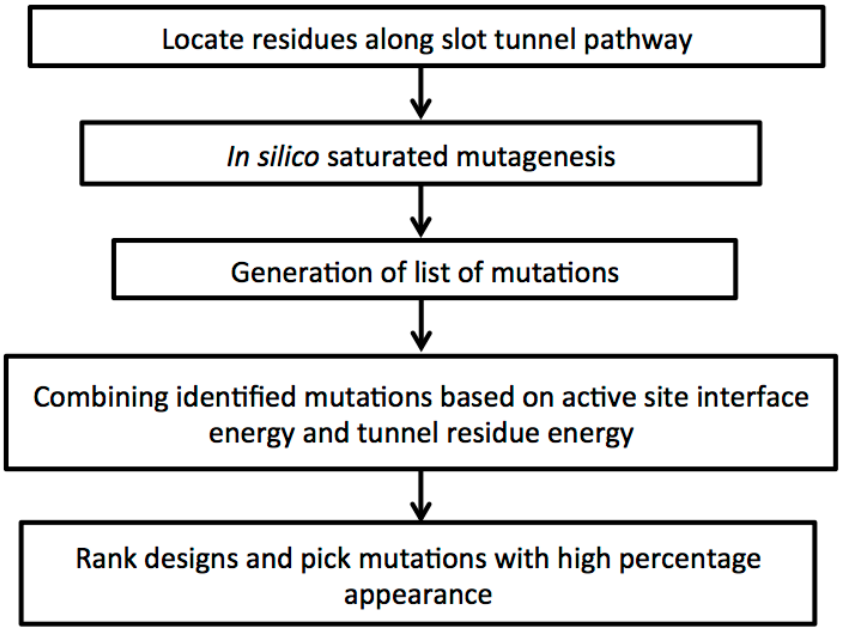

Figure 3.2 Flow chart of computational resign of DhaA slot tunnel

Table 3.1 Mutations with high percentage of appearance from the top ranked designs.

Mutation	Appearance%
I132M	46.43
R133W	100
I135L	52.9
E140Y	100
W141V	48.6
V245M	40.5
I247F	96.2

3.2.2 Experimental method

Library construction of gene inserts by Kunkel mutagenesis Beneficial mutations identified in the library were combined by Kunkel mutagenesis[19] using degenerate primers (Position 141, 140, 135, 133, 132: 5'-GGTCTCACGGGCGAATTC CGG**CMMWTM**GTCCCACGTCGG**GAK**AGG**CCRSAT**GAATTCCATACATGCAATACCTTTG ACCCGTTC-3'. Position 206, 208: 5'-CTCGCGGCTTCGGCCGGCGG**GAW**CAG**YAY**GCCG GGTGTGCCCCAGAAC-3'). At each mutagenized position, only the original amino acid and designed amino acids to those identified positions were allowed, resulting in a combinatorial library. DNA from each reaction was pooled, dialyzed in H_2O, and then transformed into electrocompetent *E. coli* strain BL21 (DE3) cells, yielding 2 x 10^7 transformants. Library plasmid DNA was isolated from expanded cultures using a miniprep kit. Gene insert was amplified from 30ng of library DNA by 25 cycles of PCR amplification using Q5 high-fidelity polymerase (NEB, Waltham, MA) to amplify genes between NdeI and XhoI along with overlap for pETCON2.

Library screening and sorting Yeast EBY100 cells were transformed with 4.0 μg of PCR-purified DNA insert and 1.0 μg of gel-purified pETCON2 vector digested with Nde1 and Xho1 using homologous recombination[20] , yielding 5 x 10^6 transformants. After transformation, cells were grown in 250 mL of selective complete media (20g/L glucose) with a selective amino acid mix (–Trp, –Ura) at 30 °C. After 2 days of growth, cells were collected and resuspended in 2mL induction media (20g/L

galactose, 2g/L glucose) to an OD600 of 0.5. The induction cultures were grown overnight at 30 °C to an OD600 of 1–1.5 and the yeast surface display screening by flow cytometry protocol is the same as previously described in chapter2.2.

Cells were subjected to two rounds cell sorting utilizing fluorescence activated cell sorting to isolate positive mutants with good expression and HaloTag ligand binding activity. For each round of sorting, cells were pre-incubated with Halotag Ligand and washed and then labeled with streptavidin-PE and anti-cMyc-FITC as described in Chapter 2.2. After two rounds of sorting, a portion of cells were plated and grown at 30°C. Plasmids from individual colonies were harvested using a Zymoprep Yeast Miniprep kit (Omega Bio-Tek Inc. Norcross, GA) and the 20 colonies were sequenced (Genscript Inc. Piscataway, NJ) from positive population.

Enzyme expression and purification Positive mutants were mapped on DhaA31 mutant in pET21b vector between NdeI and XhoI restriction sites by Kunkel mutagenesis and the sub-cloned plasmids were transformed to Competent cells of *E. coli* BL21 strain individually. Each transformed cell line was raised in 0.5L ZY Auto Induction media with 5ml preculture and grown at 37°C for 3 hours till cloudy, followed by dropping the temperature to 18°C and continued to incubate for 24 hours. Cells were collected and resuspended in 35 mL wash buffer (1XPBS, 30 mM imidazole, pH 7.4). 300 µL 0.1mM PMSF, Lysozyme and DNase(~2 mg/ml and ~0.2

mg/ml as final concentrations, respectively) were added to each resuspension. Resuspended cells were lysed via sonication on ice. The lysed cells were centrifuged at 20,000 rpm for 40 minutes. The his-tagged proteins were purified with Ni-NTA columns (Qiagen, Germany) and the protein was eluted with 10 times column volume of 1xPBS with 300mM imidazole. The elutions were dialyzed with 1XPBS (pH7.4) to remove imidazole. Purity of enzymes was analyzed by SDS-page gel and protein concentrations were measured by spectrophotometer (DS-11+, DeNovix)

pH indicated colorimetric assay validation for positive mutants A pH indicator dye-based colorimetric assay was adapted to measure DhaA and HheC activity quantitatively[21][22]. The buffer for the colorimetric dehalogenase assay consisted of 1 mM HEPES (pH 8.2), 20 mM sodium sulphate and 1 mM EDTA. Phenol red was added to a final concentration of 20 mg/ml. TCP was dissolved in the buffer to a concentration of 5mM. The assay was performed in a 96-well plate. Optical density was measured at 560nm using a Tecan Plate Reader Infinite 200. Each microplate well received 146 µL of the substrate-buffer solution and 4 µL of enzyme to a final enzyme concentration 1µM. The optical densities were determined quantitatively over a period of 30 minutes. A decrease in optical density was indicative of proton release. The progress of the reaction was also followed by visual observation as the buffer color changed from red through orange to yellow with decreasing pH.

3.2.3 Experimental results

Library screening and cell sorting The theoretical library size is 768 and the homologous recombination transformation efficiency is $5x10^6$ which is more than 600 times higher than the library size, indicating a good coverage of the design library. The induced yeast cells were subject to two rounds of cell sorting with a gate selecting mutants with both ~top 50% expression level and HaloTag ligand binding (Figure 3.3). After two rounds of sorting, two mutants were selected: R133W, W141V (M1) and E140Y (M2). These two mutants were picked up and further evaluated for TCP inhibition flow cytometry separately, and both showed HaloTag ligand binding (Figure 3.4A, 3.4D) and TCP inhibition (Figure 3.4B, 3.4E) with similar level as wild type.

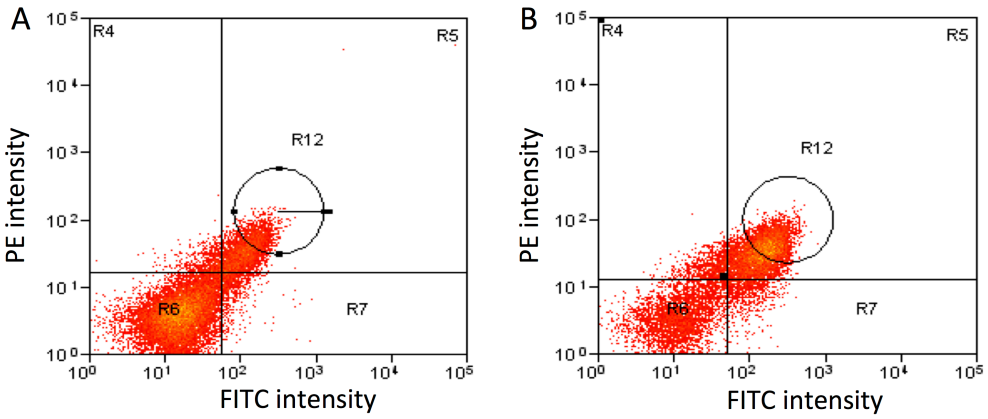

Figure 3.3 Selection gate of **A)** First round of cell sorting and **B)** second round of cell sorting. Cells in R12 gate with both good expression and Halotag ligand binding were sorted.

We then analyzed the slot tunnel in computational models of the two mutants by

CAVER, neither of the mutants blocked the slot tunnel: M2 had almost the same bottleneck radius as the wild type and M1 even made the bottleneck wider than the wild type. But both mutants made the slot tunnel mouth more hydrophobic than the wild type (Figure 3.4C, Figure 3.4F).

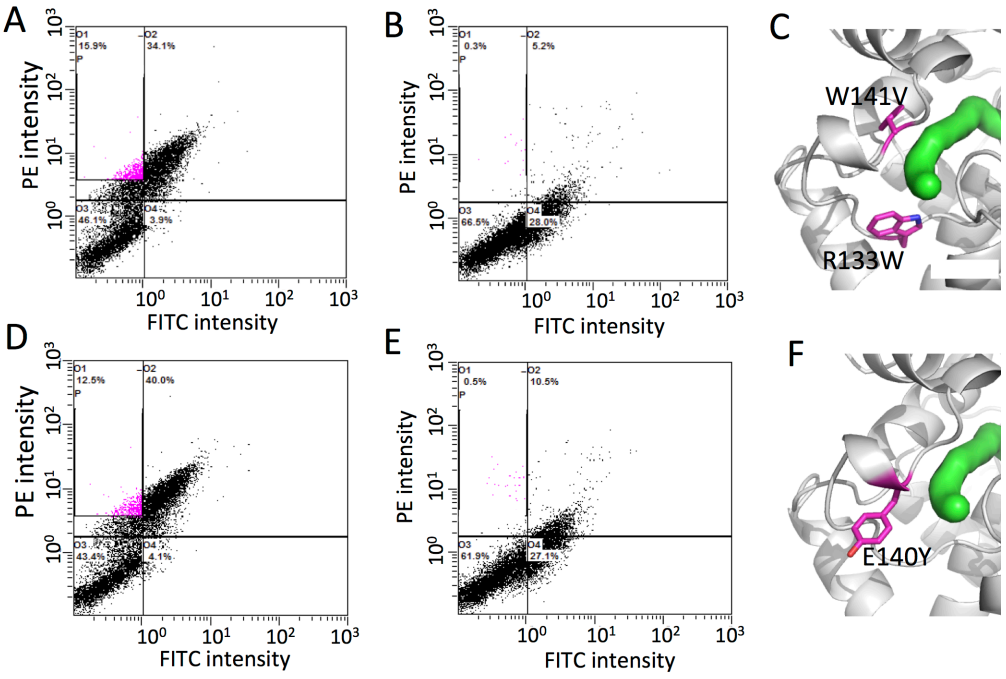

Figure 3.4 Flow cytometry analysis for HaloTag ligand binding of **A)** M1 (R133W, W141V) and **D)** M2 (E140Y); flow cytometry analysis for TCP inhibition of **B)** M1 (R133W, W141V) and **E)** M2 (E140Y); computational model of **C)** M1 (R133W, W141V) and **F)** M2 (E140Y). The slot tunnel was highlighted in green and mutations were highlighted in magenta.

Colorimetric assay on three mutants The R133W and E140Y mutation and combination of these two mutations were mapped on DhaA31, which is the most efficient DhaA mutant[1] in catalyzing TCP hydrolysis . W141V was not mapped to DhaA31 here because position 141 is a conserved mutation for DhaA31. Enzymatic activity towards to TCP dehalogenation of the three mapped mutants were

evaluated by pH indicated colorimetric assay, all three mutants showed similar level of absorbance decrease at 560nm (Figure 3.5), indicating no distinguishable improvement was detected for the three mutants compared with DhaA31.

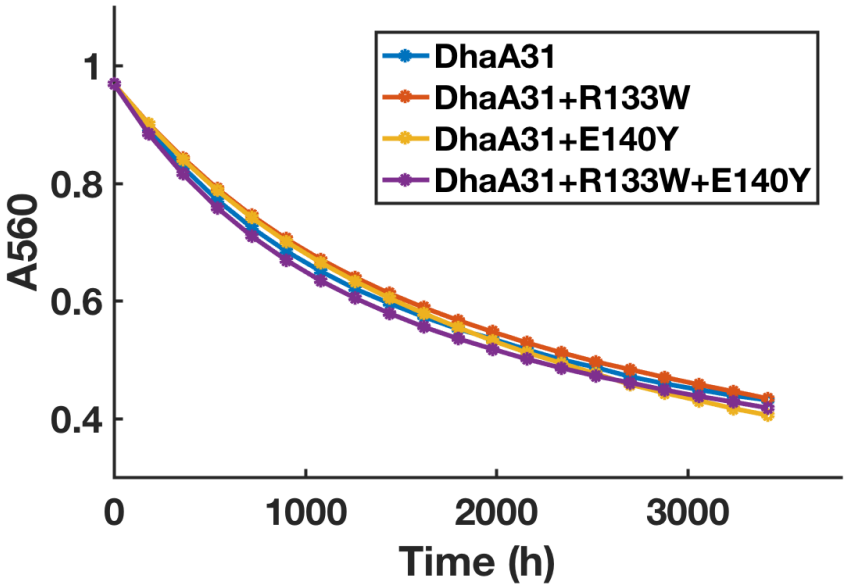

Figure 3.5 TCP conversion to DCP activity of DhaA31 by pH-indicated colorimetric assay.

3.2.4 Discussion

Computational slot tunnel engineering of haloalkane dehalogenase DhaA was executed to achieve TCP conversion activity enhancement based on the previous statement that decreasing accessibility of active site for water molecules caused increased catalytic rate[1]. A combinatorial mutant library with top ranked mutations were screened and sorted by yeast surface display and two mutations that actually made the tunnel more hydrophobic were isolated. However, no further enhancement towards TCP conversion is observed for these mutations. Two reasons may contribute to this result. First, for DhaA31, the physicochemical property of the

slot tunnel is no longer the limiting factor for enzymatic activity, thus no further enhancement will be achieved. Second, the conclusion addressed from previous study is only a hypothesis – decreasing accessibility of the slot tunnel may not be the actual reason for activity improvement for TCP.

For further study, saturated site-mutagenesis around slot tunnel could be employed to make larger combinatorial library, if any mutant with improvement are detected, the structure could be characterized to further explain the relationship between slot tunnel property and enzyme activity.

3.3 Computational De Novo Tunnel design for enzyme active-site access pathway

3.3.1 Introduction

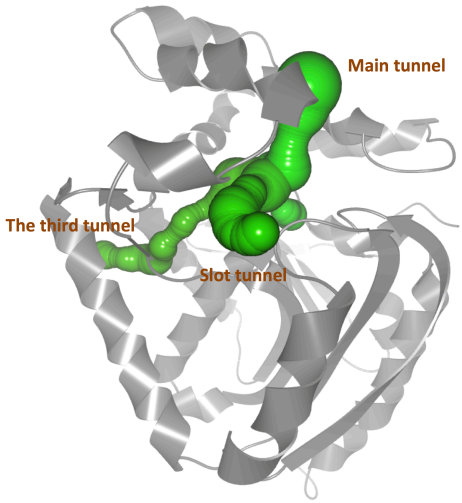

Figure 3.6 Tunnel observed by Caver calculation with haloalkane dehalogenase DhaA

As mentioned in the introduction of this chapter, ligand transportation tunnel are essential for catalysis, especially for enzymes with buried active sites[23]. Lack of optimized tunnel will be a huge barrier for enzymes. Except main tunnel and slot tunnel, which have been proven for substrate entrance and product release, DhaA haloalkane dehalogenase also has other paths which might be too narrow to be an functional access pathway (Figure 3.6). However, they could be engineered to

behave as a functional tunnel and the engineered tunnel might paly a significant effect on the substrate entry, product release and water molecule accessibility, thus affecting the activity and specificity of the enzyme. Despite the significance of active-site access pathway, there is no general computational approach to introduce a novel tunnel into existing enzymes. In the study, we address a computational strategy regarding *de novo* tunnel design for DhaA and the designs were screened by yeast surface display assay.

3.3.2 Computational design of *de novo* tunnel

Other than the two major tunnels leading to the active site on DhaA, there are other paths connecting the hydrophobic active site to the bulk, which could be further engineered to become a functional tunnel. Among all these tunnels, the third tunnel (Figure 3.6) is the best candidate cause it has the widest tunnel bottleneck and is located between the top domain and bottom domain as same as the major tunnels.

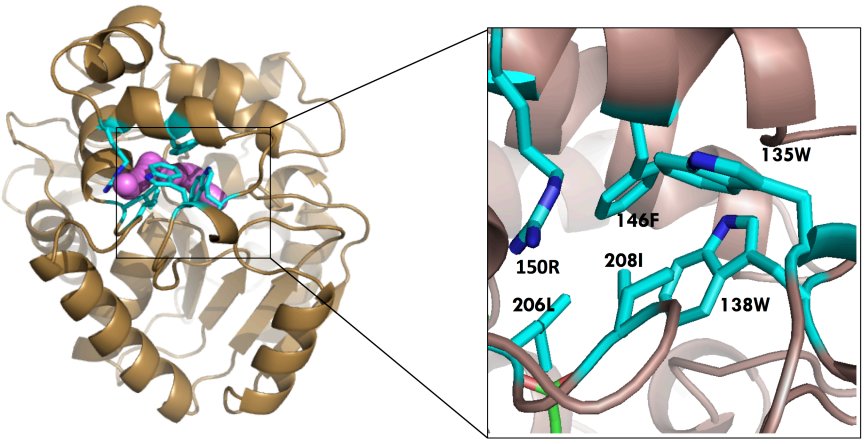

Figure 3.7 Structure of DhaA80 and the third tunnel (magenta). The tunnel neighbor residues of the third tunnel were defined by CAVER and highlighted in cyan.

For novel tunnel engineering, a negative construct with major tunnel closed is required. DhaA80 mutant (T148L, G171Q, A172V, C176F) is chosen as starting point because of two reasons. First, the major tunnels are closed on DhaA80. Second, DhaA80 is more thermostable compared with wild type DhaA, which makes it a suitable start point for design[12].

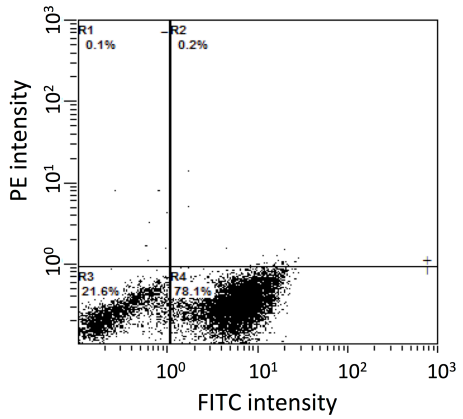

Figure 3.8 Flow cytometry analysis of DhaA80 to detect expression level and HaloTag ligand binding.

Flow cytometry results showed that DhaA80 expressed well on yeast surface, which is consistent with its good expression level in *E. coli.* No HaloTag ligand binding activity was detected for DhaA80, indicating that the major tunnel used for substrate binding was completely blocked. All these evidence confirmed that DhaA80 could serve as a negative starting point for *de novo* tunnel engineering of DhaA.

CAVER was employed to analyze the third tunnel on DhaA80 and located six residues along the tunnel as neighbor residues (135W, 138W, 146F, 150R, 206L, 208I) that contribute to the size and physicochemical properties of the tunnel (Figure 3.7). Evolutionary conservation of these six positions was estimated using ConSurf server[24]. All these positions are conserved to some extent but not extremely conserved, providing possibility for further engineering and evolution (Table 3.2).

Table 3.2 Consurf conservation form of bottleneck positions of DhaA

Position	Conserved Color	Amino Acid from ConSurf
135	4	F,S,W,Y,I,R,G,L
138	3	S,F,T,W,Y,V,M,I,R,L
146	5	S,A,F,T,W,Y,V,M,C,R,L
150	8	H,Q,N,K,R,L
206	5	A,F,M,I,Y,L,V
208	6	F,K,I,Y,L,V

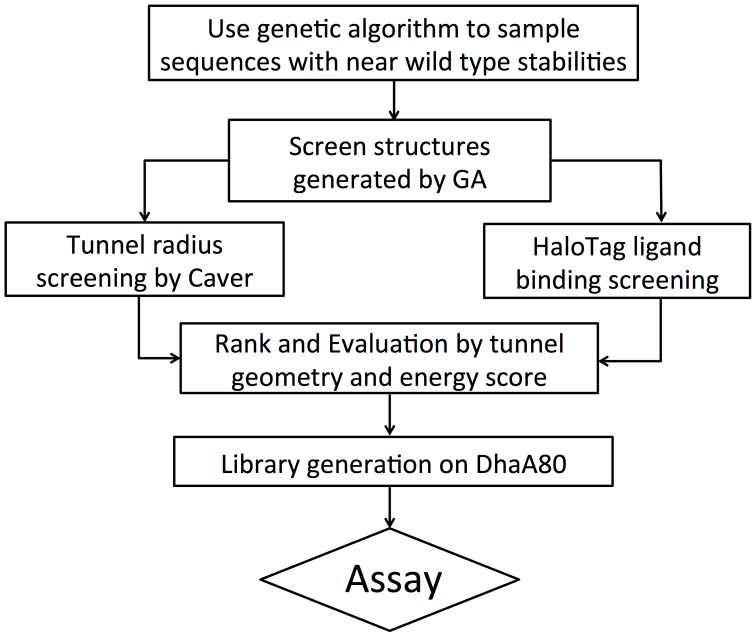

Figure 3.9 Flowchart of computational *de novo* tunnel design

Genetic algorithm was used to sample six tunnel neighbor positions to avoid being trapped in local optimal solution and to achieve an efficient sampling. Two sets of sequences were generated with biased distribution. The first set allows substitution using all amino acid while the second set only allows conserved amino acids listed on Consurf server (table 3.2). For both sets, smaller amino acid was set with a higher probability than bulky ones. 2500 different combinations of tunnel neighbor residues were generated for each set and subjected to screening by both Rosetta and CAVER. Mutation number distribution and total score energy distribution of the designs using conserved amino acids indicates a sufficient sampling using genetic algorithm for tunnel neighbor sequence generation (Figure 3.10).

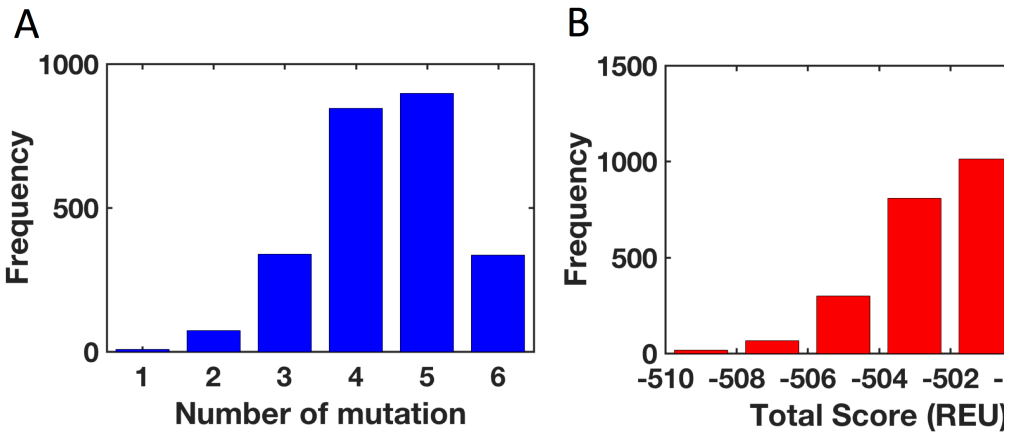

Figure 3.10 A) Mutation number distribution **B)** Total score distribution of sampled sequences generated by genetic algorithm using conserved amino acids.

Decoys generated by Genetic algorithm were screened by three different filters: First, the total score to maintain structure stability, which is calculated using Rosetta score function Talaris 2013. Second, size of tunnel bottleneck to keep the access pathway open (Figure 3.11B) and last, HaloTag ligand binding energy to favor ligand binding (Figure 3.11A). The essential idea for novel tunnel design here is to open a closed or narrow tunnel to accommodate the substrate or product transporting, so the bottleneck threshold of the newly opened tunnel measured by CAVER is set at 1.02 Å which is as the same as main tunnel in DhaA_Halo, a mutant known to bind HaloTag ligand with a high affinity. The binding energy was evaluated by comparing the energy of the protein bound with the HaloTag ligand and the apo-protein. All designs bound with and without HaloTag ligand constrained with catalytic geometry were subjected to relaxation via RosettaScripts[25] FastRelax[26]. The binding energy evaluation was validated by DhaA_Halo model

(PDB ID: 4KAA), giving a -2.9 binding energy in Rosetta Energy Unit. The threshold was then set to zero to filter out designs with positive binding energy.

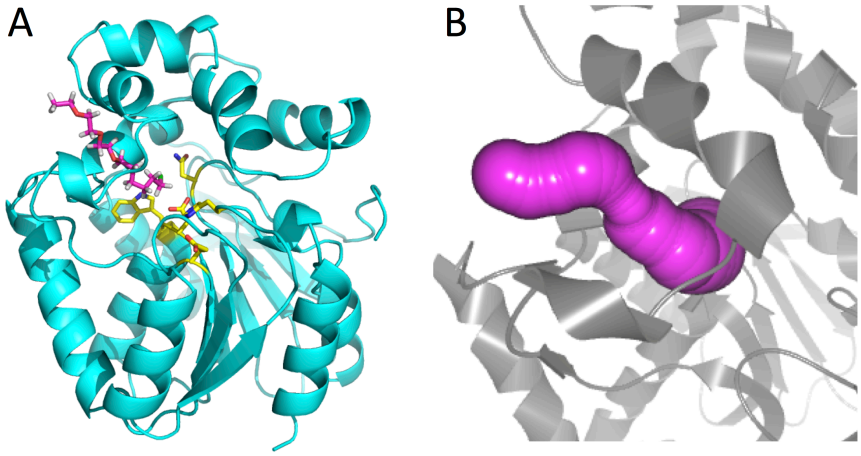

Figure 3.11 A) Structure of DhaA80 bound with HaloTag ligand in the third tunnel. B) Illustration of tunnel bottleneck of the new opened tunnel calculated by CAVER.

After filtering based on the three criteria, 6 sequences from the set using conserved amino acids (Figure 3.12 B) and 9 sequences from the set using all amino acids (Figure 3.12 A) passed all thresholds. Based on the logo plot, position 135 is the only position sharing no overlap for two design sets, while all the other five positions showed similar preference for both sets, position 146 even had exactly the same occurrence distribution (Figure 3.12C, D).

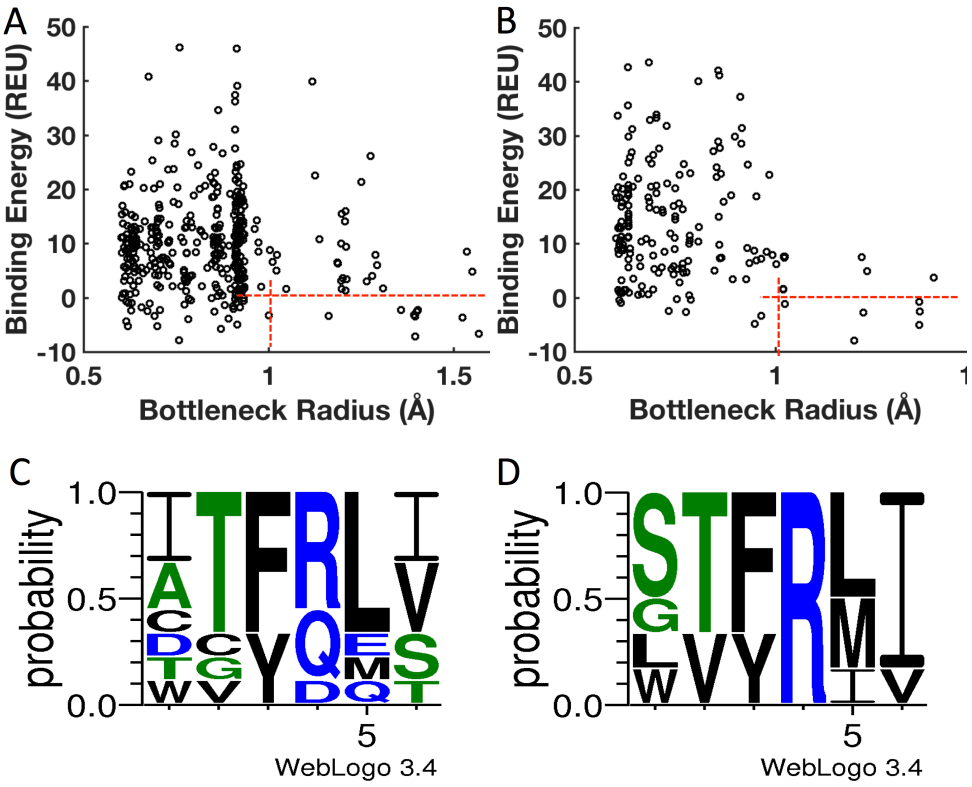

Figure 3.12 A) Plot of binding energy and tunnel bottleneck radius of sequences with lowest total score within 7.5 REU using all amino acids. **B)** Plot of binding energy and tunnel bottleneck radius of top 182 sequences with lowest total score within 5 REU utilizing only Consurf conserved amino acids. **C)** Logo plot of sequences with bottleneck radius > 1.02, Binding energy <0.**D)** Logo plot of sequences with bottleneck radius > 1.02, Binding energy <0. The size of the single letter amino acid code in each logo plot represents the occurrence of a particular amino acid at a particular position.

Instead of making only these 15 designs, a small library which cover wild type and all the designs as much as possible were built by Swiftlib[27], a program which can optimize a degenerate codon library to cover the desired set of amino acids at several positions within a limit size. This combinatorial library permits the examination of a wider range of mutation neighbors of the final designs. Yeast surface display discussed in Chapter2 will be employed for the designed library screening and sorting.

3.3.3 Experimental screening of designed library

Library construction of gene inserts by Kunkel mutagenesis Beneficial mutations identified in the library were combined by Kunkel mutagenesis using degenerate primers (Position 150, 146, 138, 135: 5'-GCGGT**KYS**GAAGGCCTG**GWA** CAGCTCACGGGCGAATTCCGG**TVY**TTCGTC**YVH**CGTCGGGATAGGCCGGATG-3'. Position 208, 206: 5'-GTTCGCGGGCTCACCGGC**TRY**GGG**YAK**CTCGTTGGGGAATCGC CACAGTGG-3'). At each mutagenized position, only the original amino acid and designed amino acids to those identified were allowed, resulting in a combinatorial library. DNA from each reaction was pooled, dialyzed in H2O, and then transformed into electrocompetent *E. coli* strain BL21 (DE3) cells, yielding 1.2 x 10^7 transformants. Library plasmid DNA was isolated from expanded cultures using a miniprep kit. Gene insert was amplified from 30ng of library DNA by 25 cycles of PCR using Q5 high-fidelity polymerase (NEB, Waltham, MA) to amplify genes between NdeI and XhoI along with overlap for pETCON2. Combinatorial library was generated using yeast homologous recombination using the same protocol as described as Chapter 3.2.2. The design library was subjected to screening and sorting following the same protocol described in Chapter 2.2 and Chapter 3.2.2.

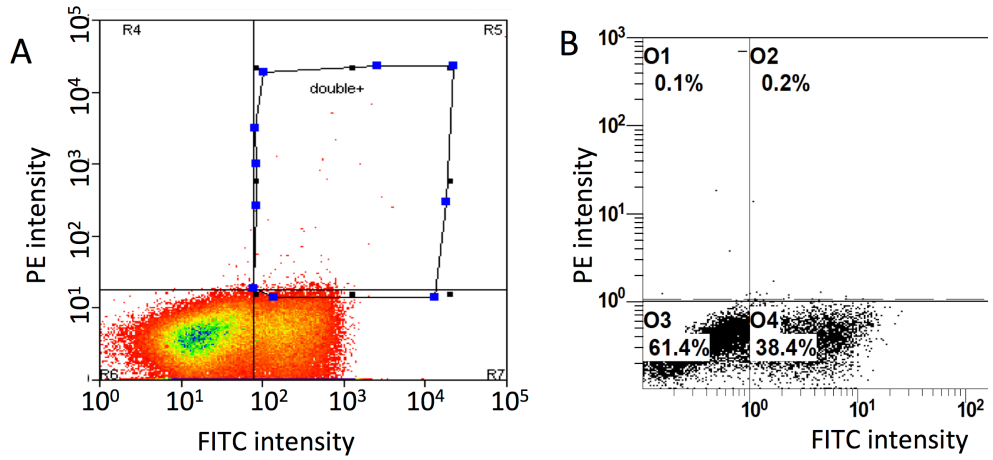

Figure 3.13 A) Fluorescence activated cell sorting of *de novo* design library. **B)** Flow cytometry analysis of mutants accumulated from sorting.

Results of library screening The theoretical library size is $1.5x10^4$ and the homologous recombination transformation efficiency is $5x10^6$, which is more than 30 times larger than the library size, indicating a good coverage of the designed library. 20 individual sequences were analyzed and the library variety was validated. The induced yeast cells were subject to fluorescent activated cell sorting. There is a large population of yeast cells with decent expression level and very small percentage of cells showing HaloTag ligand binding (Figure 3.13A). A double positive gate was set to select mutants with both good expression level and HaloTag ligand binding activity from cell sorting. However, the sorted cells can not be accumulated in double positive gate, indicating the positive PE signal from cell sorting could just be the noise (Figure 3.13B), which means no novel designed tunnel in this mutant library is able to accommodate HaloTag ligand.

3.3.4 Further library design by expanding designable shell

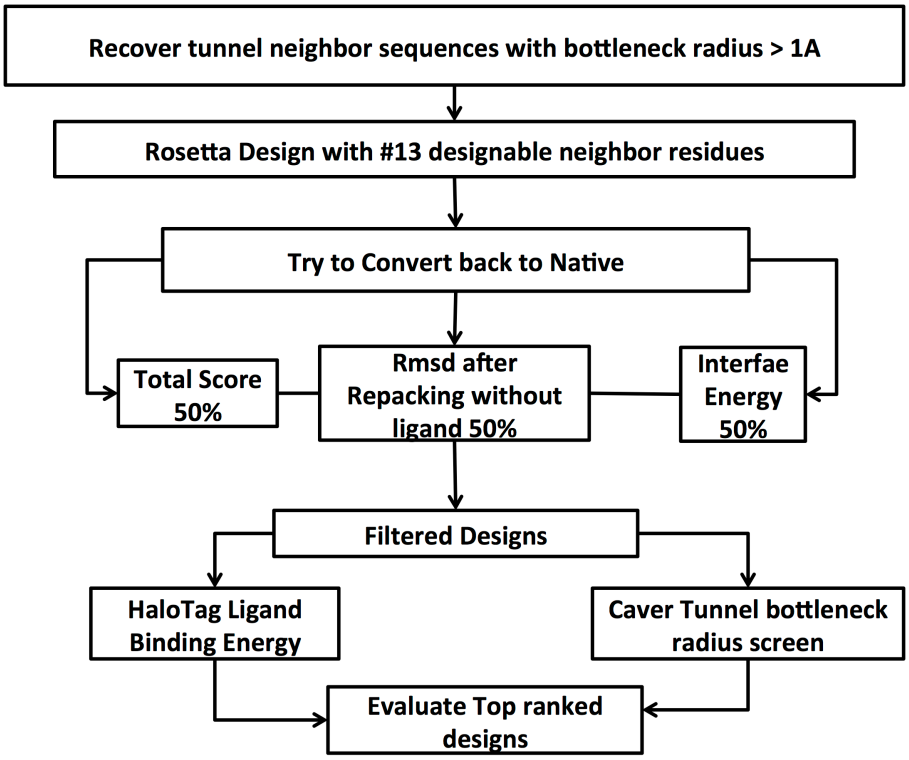

Figure 3.14 Flowchart of computational *de novo* tunnel design protocol with expanded designable shell

Computational Design Other than determining tunnel neighbor residues, surrounding residues of the tunnel neighbor residues may also need to be designed to find the optimal residue identities in order to stabilize the tunnel region and maintain the tunnel information. Designs from two genetic algorithm generated sets with >1 Å tunnel bottleneck of the third tunnel were recovered and subject to RosettaDesign[18], 13 positions within 6 Å of tunnel neighbor residues were allowed to be designed and were tried to converted back to native via GreedyOptMutation

mover via RosettaScript[25]. Designs were filtered by three criteria: Total score for stable structure, rmsd after repacking without ligand for preorganization and complementarity, and tunnel interface energy with HaloTag ligand. Only designs ranked in top 50% based on all three criteria were kept for further evaluation. 1500 designs passing the filters were screened by both tunnel bottleneck and HaloTag ligand binding energy, which is previously described in the first round of design. 85 designs with 39 specific sequences passed both thresholds (Figure 3.15). A $1X10^7$ library which cover wild type and all the 39 sequences at 11 positions were built by Swiftlib[27]. This combinatorial library permits the examination of a wider range of mutation neighbors of the final designs. Yeast surface display will be employed for the designed library screening and sorting.

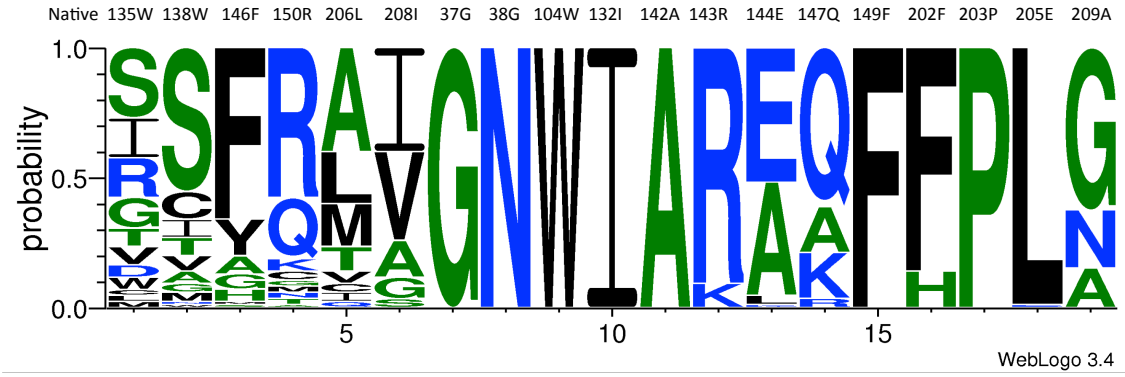

Figure 3. 15 Logo plot of top design sequences h expanded designable shell.

Library construction of gene inserts by polymerase cycling assembly To incorporate top ranked designs, polymerase cycling assembly[28] using degenerate primers and yeast homologous recombination were combined to construct the library. 27 DNA fragments were used to generate the DhaA mutants, resulting in a

combinatorial library. Size of assembled insert fragment was validated by agarose gel electrophoresis. Assembled insert was amplified from 30ng of library DNA by 25 cycles of PCR using Q5 high-fidelity polymerase (NEB, Waltham, MA) to amplify genes between NdeI and XhoI along with overlap for pETCON2. Methods for yeast homologous recombination and library screening and sorting are the same as described in Chapter 2 and Chapter 3.2. The transformation efficiency is 10^7.

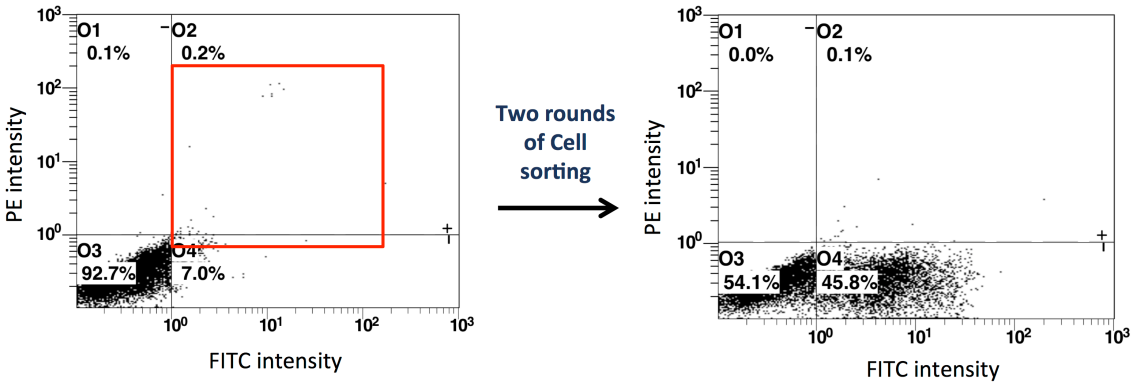

Figure 3. 16 Flow cytometry analysis of de novo design library with expanded designable shell before (left) and after (right) two rounds of cell sorting.

Most mutants in the library did not express on yeast surface based on flow cytometry analysis. The induced yeast cells were subject to fluorescent activated cell sorting anyway. Same double positive gate was set to select mutants with both good expression level and HaloTag ligand binding activity. However, only cells were good expression were accumulated, no mutants with HaloTag ligand binding activity were detected (Figure 3.16).

3.3.5 Discussion

This study demonstrated a strategy attempting for *de novo* tunnel design, unfortunately at this stage, we cannot accurately predict transport tunnel opening in haloalkane dehalogeanse DhaA. There are several possible reasons that could possibly cause the non-proficient design. First, there is an absence of information for evolutionary tunnel optimization in nature, which means there is clear relationship between enzyme function (activity and specificity) and tunnel properties, like tunnel position, shape, bottleneck radius and other physicochemical properties, causing difficult to evaluate the designs. Second, though transport tunnel is essential for access to the active site, it may not be the only contributor. Previous study showed dynamic motion of the cap domains of haloalkane dehalogenase could be relevant to the function by effecting flexibility of the cap domain and entrance tunnels[15]. For further research, more factors, like hydrophobicity level, should also be considered for design evaluation. Also instead of engineering on a static model, molecular dynamics simulations could be employed to confirm successful opening of the novel tunnel with defined structure and function. What's more, the sorted sequences isolated from this work with decent expression level and wider bottleneck can serve as new starting points for further characterization and engineering.

3.4 Computational chimeric protein design with preserved enzyme

active-site access pathway

3.3.1 Introduction

Driven largely by evolutionary pressures to increase the diversity of molecular functions[29][30], domain shuffling design stems from genomic duplication and recombination events that lead to the introduction of large jumps in protein sequence space, which in turn leaves the component domains susceptible to radical alterations in topology and function[30]. A particular topological characteristic that has become a prominent feature in a significant number of multi-domain proteins is the utilization of the interface between two or more domains to form binding or active sites, such as in the periplasmic binding proteins amino acid dehydrogenase , αβ hydrolase folds enzymes and serine proteases[31][32]. Thus Domain shuffling is an efficient protein engineering method to achieve protein functional evolution and is a particularly effective strategy to modify the function of multi-domain proteins.

Unlike most multi-domain proteins are formed by connecting domains through end-to-end linkages, αβ hydrolase superfamily featured a nested topology in which at least one domain is inserted into another[33] (Figure 1.5). Similar to the case of end-to-end multi-domain proteins, inserted domains provide tunable functionality to a protein complex by creating domain-domain interfaces capable of developing

novel activities and altered substrate specificities[34][35][36]. In previous study, chimeras were constructed from a pair of members of αβ hydrolase superfamily (a carboxylesterase and an acylpeptide hydrolase) by grafting their functional domains to the conserved main domain hydrolase and their activities on hydrolyze *p*-nitrophenyl esters with different acyl chain length were explored[37].

Previous results suggested that the substrate-binding domain is the dominant factor on substrate specificity indicating that cap domain shuffling could gain new activity and specificity by introducing new interface and tunnels to the enzyme. Members of αβ hydrolase superfamily catalyze diverse processes and have very different cap domains that could potentially provide improvements towards TCP dehalogenation through domain swapping.

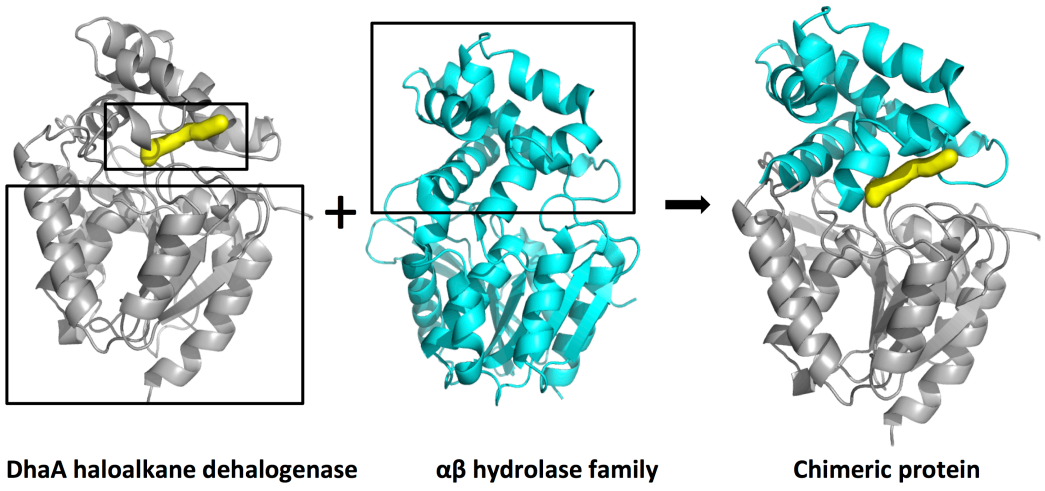

Figure 3.17. Schematic of Computational chimeric protein design with preserved enzyme active-site access pathway

In this chapter, we developed a computational approach to sample new activity and specificity based on novel domain fusions to obtain new starting points for haloalkane dehalogenase DhaA towards TCP degradation. We constructed chimeric protein models by grafting cap domains from other members of αβ hydrolase superfamily to DhaA bottom domain bound with haloalkane ligand and rebuild loops between two domains (Figure 3.17). This method can be used to introduce novel interface interactions and maintain tunnel geometry to accommodate target substrate, provide new starting point for further design and could also be generally applied to other multi-domain proteins, thus opening new avenues for enzyme design.

3.4.2 Computational design

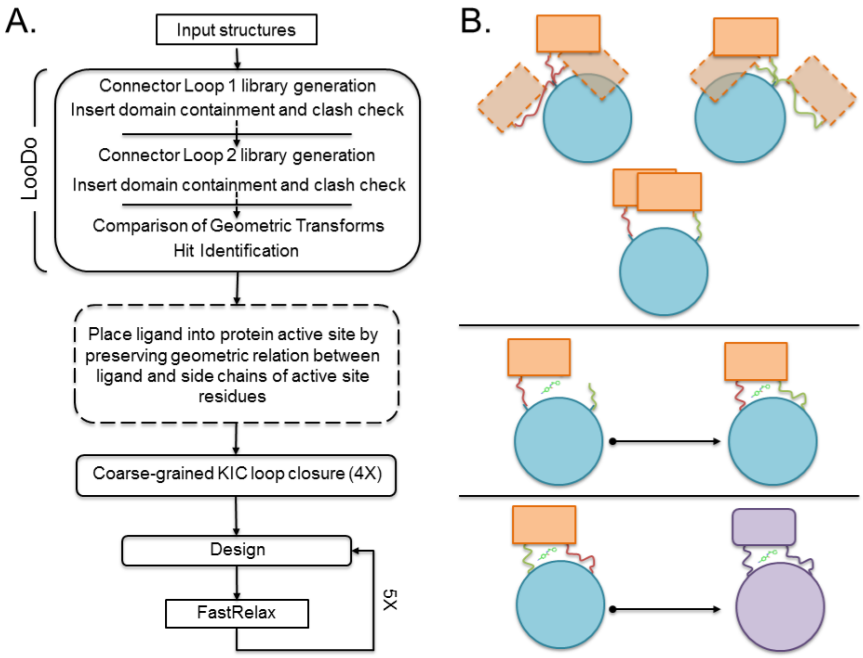

Figure 3.18 A. Flowchart of the design pipeline B. Cartoon diagram of the LooDo algorithm.

Table 3.3 List of cap domains from αβ hydrolase superfamily utilized for design

Cap PDB ID	Enzyme Name	Cap Domain
1A8S	Chloroperoxidase	126-202
1C4X	2-Hydroxyl-6-Oxo-6-Phenylhexa-2,4-Dienoic Acid Hydrolase	166-228
1D07	Haloalkane dehalogeanse LinB	138-207
1DWO	Hydroxynitrile Lyase	140-208
1IUN	2-Hydroxy-6-oxo-7-methylocta-2,4-dienoate hydrolase	135-190
1MT3	Prolyl aminopeptidase	133-218
1R3D	Pfam abhydrolase	120-186
1UK6	Meta-cleavage product hydrolase	134-189
1WOM	phosphatase	139-200
1WPR	Phosphatase	136-197
1X2B	Prolyl aminopeptidase	140-236
1XQV	tricorn protease	131-216
1Y37	Fluoroacetate Dehalogenase	113-180
1Y7H	Methylsalicylate esterase	139-216
1YAS	Hydroxynitrile lyase	113-177

Generation of input structure A set of 15 two-domain protein structures belonging to the αβ hydrolase superfamily was selected from the Protein Data Bank (Table 3.3) based on cap topology, which should be able to accommodate the haloalkane substrate. The insert cap domain region were determined by visual inspection via PyMOL, the parent domain (bottom domain) bound with the HaloTag ligand was chopped from DhaA_halo (PDB ID: 4KAA) also via visual inspection. Each parent domain and insert domain was then separately subjected to full-atom relaxation via the RosettaScripts[25] FastRelax[26] protocol in which side-chain

coordinate constraints were imposed on each the starting structures. A Placement Grid was generated from the native conformation of the 4KAA cap domain with respect to the parent domain, which consists of a set of voxels corresponding to the space occupied by the insert domain.

LooDo algorithm The cap domain was placed utilizing LooDo algorithm (Figure 3.18), which was developed by Kristin Blacklock in our lab (manuscript submitted to PROTEINS: structure, Function, and Bioinformatics). During the initial phase of the LooDo algorithm, libraries of 4,500 eight-residue linker fragments were culled from Rosetta's internal protein structure database (Vall) of idealized proteins[38] where the amino acid sequences and secondary structure sequences of the four terminal residues (two on either end of the fragment) were biased toward the amino acid and secondary structure identities of the two residues in the parent insertion site and insert attachment site.

Each fragment in the library generated for first insertion (Linker A) was superimposed by two N-terminal residues onto Parent Insertion Site A followed by superimposition of the insert domain by Insert Attachment Site A onto the two C-terminal residues of the library fragment, thus yielding a placement of the insert domain with respect to the parent domain completely directed by Linker A. This conformation was then evaluated for its containment within the Placement Grid

and its lack of Cα atom clashes with all heavy atoms in the parent domain. Insert A configurations passing both filters were saved for further analysis. Similar methods were used for the generation of insert domain placements dictated by library fragments generated for second insertion (Linker B), to yield a set of Insert B configurations. After both sets of insert domain placements were generated, every Insert B was compared to every Insert A and pairs that were deemed sufficiently close to one another in rotational and translational relation to the parent domain were identified as Insert Pairs. Each insert domain conformation in the Insert Pair was then assembled with both the parent domain segments and both Linkers such that the primary structure of the resulting decoy was complete. More decoys of cap placement were generated from the complete chimeric structures by subtle rotation and translation using spin mover written by Kristin Blacklock via RosettaScripts.

Generalized kinematic loop closure These assembled decoy protein conformations underwent iterative rounds of a generalized kinematic loop closure (KIC)-based algorithm[39][40] to idealize the peptide bonds at the four chain breakpoints. Closure was attempted first at the site at which the break was evaluated to be the largest, while rigid body movements of the insert domain with respect to the parent domain were allowed during minimization steps. The remaining chainbreak points were closed without rigid body movements of the

insert domain.

Sequence and structure optimization The assembled and closed structures were subjected to five cycles of sequence and structure optimization, which entailed amino acid identity and rotamer sampling at each of the protein-ligand interface region, the loop regions, and the loop-neighboring regions (only top domains were allowed to be designed) followed by structural refinement using Rosetta's FastRelax application within the RosettaScripts framework.

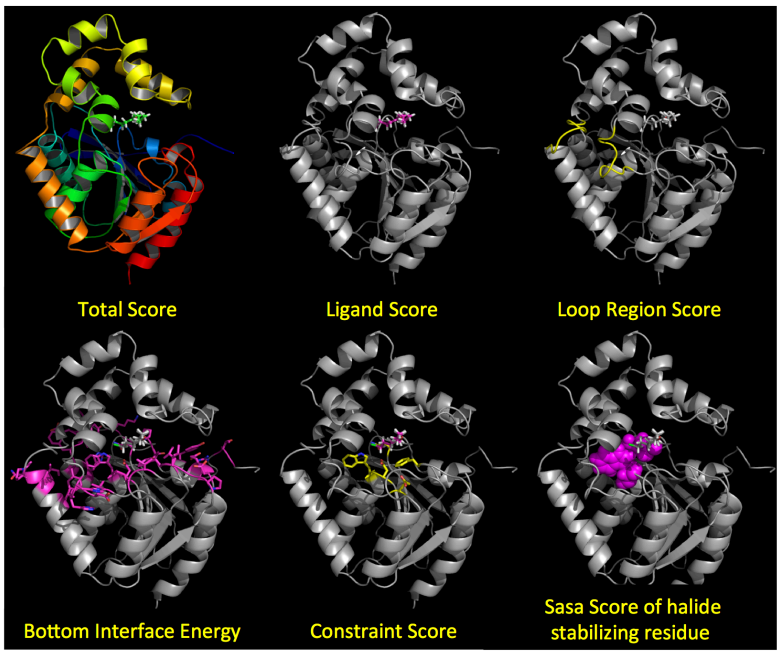

Figure 3.19. Score functions utilized for final design evaluation

Design evaluation Decoys were scored using the talaris2013 score function. All designs were subjected to several filtering criteria: total score, ligand score, loop

region score, bottom interface energy, contact number, constraint score and solvent accessible surface area score (Sasa score) separately to evaluate overall structure, designed region and catalytic geometry. The Sasa score is utilized to evaluate active site, which should be buried and hydrophobic as wild type DhaA. Designs with decent ranking in all scoring categories were selected for visual evaluation using Foldit and mutated residues were manually reverted back to their native scaffold identity if no stable or functional benefit could be obtained. Five top designs were selected as final designs for experimental test.

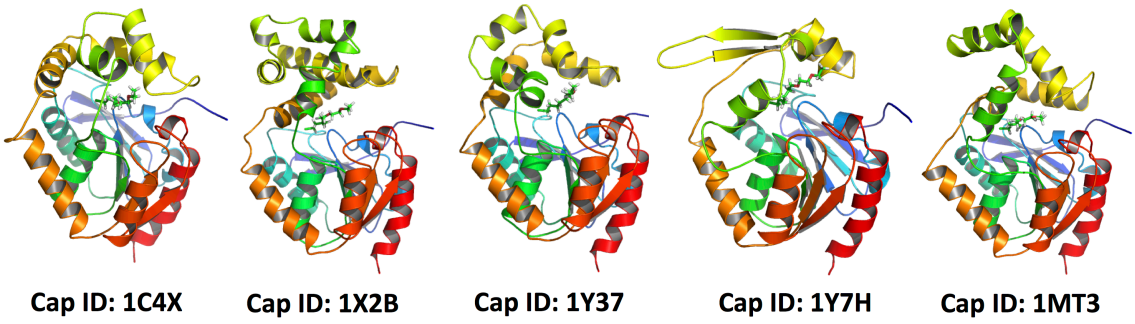

Figure 3.20. Modl structures of final designs in PyMOL

3.4.3 Experimental methods and results

Methods for yeast surface display Designed caps with overlap with DhaA_halo bottom domain were ordered from Integrated DNA technologies (Coralville, IA). DhaA_halo bottom in pETCON2 was linearized by PCR amplification. Designed cap and loops were inserted in DhaA_halo bottom domain utilizing Gibson assembly cloning protocol[41]. All designs were then chemically transformed into the EBY100 competent yeast strain as described in Chapter 2.2. The HaloTag ligand binding

assay by flow cytometry was conducted following same protocol described in Chapter 2.2 as well.

Results of Halotag ligand binding assay for final chimeric designs All five designs were subject to flow cytometry to detect expression level and HaloTag ligand binding affinity. Unfortunately, none of the five designs expressed on yeast surface, indicating no stable structures were folded (Figure 3.21).

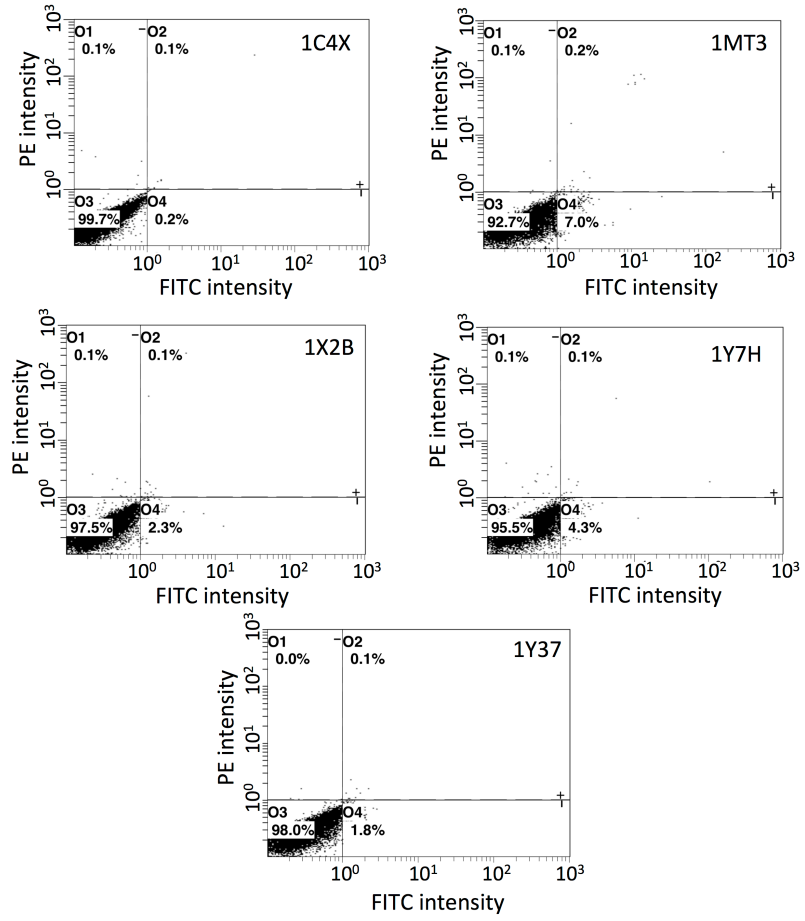

Figure 3.21. Flow cytometry analysis for five chimeric designs

Directed evolution by error-prone PCR An Error-Prone PCR based Mutant

Library was built to overcome the energy barrier. Each design was cloned using the forward primer (5′- GTA GCG GAG GCG GAG GGT CGG CTA GCC ATA TG -3′) and reverse primer (5′- GTC CTC TTC AGA AAT AAG CTT TTG TTC GGA TCC GCC CCC CTC GAG -3′). Error-prone PCR was performed using the design in pETCON2 vector as template and 0.25 nM primers each, 0.2 mM dATP and dGTP, 1 mM dCTP and dTTP, and Taq buffer containing 7 mM $MgCl_2$ and 0.2 mM $MnCl_2$. 1µl Taq DNA polymerase (NEB, Waltham, MA) was used. The temperature program used was 2 min at 95°C followed by 15 cycles of 20 s 95°C, 60 s 62°C, and 60 s 72°C and finished with 5 min at 72°C. The PCR product was purified by PCR purification kit (life technology). Yeast EBY100 cells were transformed with 4.0 µg of PCR-purified DNA insert and 1.0 µg of gel-purified pETCON2 vector digested with Nde1 and Xho1 using homologous recombination[20], yielding 1 x 10^7 transformants. The cells were screened for both expression and HaloTag ligand binding using flow cyotmetry and expressed population were subject to two rounds of cell sorting following the same protocol described in Chapter 2.2 and Chapter 3.2. Plasmids from individual colonies of each design were harvested using a Zymoprep Yeast Miniprep kit (Omega Bio-Tek Inc. Norcross, GA) and the 12 colonies were sequenced (Genscript Inc, Piscataway, NJ) for each design to check mutation rate of the directed evolution.

Results of directed evolution The average error rate is 4/900bp ranging between

2 to 9 based on sequencing for single colonies, indicating a sufficient amount of mutations were introduced. After two rounds of cell sorting, mutants with expression were isolated and amplified for four out of five starting designs, however, no HaloTag ligand binding activity was detected for all designs (Figure 3.22).

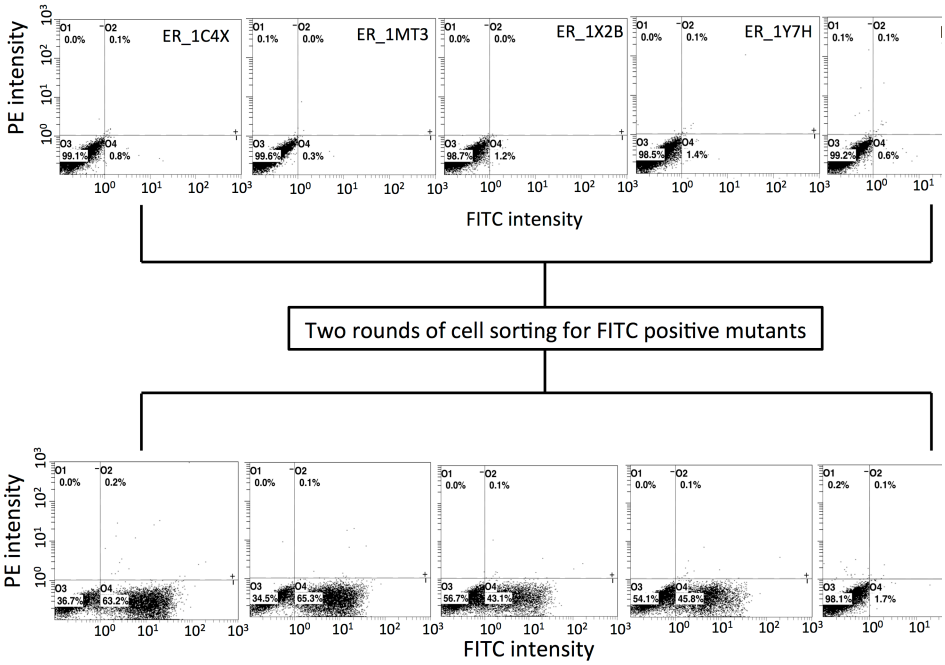

Figure 3.22. Flow cytometry analysis of library after directed evolution (Top) and amplified cell population after two rounds of cell sorting for mutants with positive

3.4.4 Discussion and further work

In this chapter we describe a computational method to enable the structure prediction and design of novel chimeric proteins towards TCP degradation, however the final designs showed no expression on yeast surface. Even after directed evolution, mutants recovered with decent expression showed no HaloTag ligand binding function were achieved.

The computational design approach could be further improved. The outstanding questions of how to choose the insertion site and catalytic residues remain to be investigated. Domain shuffling engineering could provide significant functional leap, while at the same time, it can greatly disrupt protein structure and stability. This strategy is difficult due to the inability to completely predict the three-dimensional protein structure of the chimeric design.

Despite the difficulty, further efforts could be made to give more insights towards this project. The cap library size used for shuffling is too small to provide sufficient topologies, an expanded library with 42 caps from αβ hydrolase were employed and more chimeric designed proteins were generated using the same strategy and could be further tested. The sequence identities of mutants sorted from directed evolution could further analyze, structure with promising sequence could be further characterized and serve as new stable starting point for function improvement. The ability to design new domain inserted proteins is expected to aid design efforts for generating novel protein functions at the resulting domain-domain interfaces.

References

[1] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727–733.

[2] L. Biedermannová, Z. Prokop, A. Gora, E. Chovancová, M. Kovács, J. Damborský, R. C. Wade, *J. Biol. Chem.* **2012**, *287*, 29062–29074.

[3] P.-P. Liebgott, F. Leroux, B. Burlat, S. Dementin, C. Baffert, T. Lautier, V. Fourmond, P. Ceccaldi, C. Cavazza, I. Meynial-Salles, et al., *Nat. Chem. Biol.* **2010**, *6*, 63–70.

[4] A. Volbeda, L. Martin, C. Cavazza, M. Matho, B. W. Faber, W. Roseboom, S. P. J. Albracht, E. Garcin, M. Rousset, J. C. Fontecilla-Camps, *JBIC J. Biol. Inorg. Chem.* **2005**, *10*, 239–249.

[5] R. E. Amaro, R. S. Myers, V. J. Davisson, Z. A. Luthey-Schulten, *Biophys. J.* **2005**, *89*, 475–487.

[6] Y. Fan, L. Lund, Q. Shao, Y.-Q. Gao, F. M. Raushel, *J. Am. Chem. Soc.* **2009**, *131*, 10211–10219.

[7] H.-X. Zhou, J. A. McCammon, *Trends Biochem. Sci.* **2010**, *35*, 179–185.

[8] † Xiangshi Tan, † Huay-Keng Loke, † and Shawn Fitch, †,‡ Paul A. Lindahl*, **2005**, DOI 10.1021/JA043701V.

[9] R. A. W. Frank, C. M. Titman, J. V. Pratap, B. F. Luisi, R. N. Perham, *Science (80-.).* **2004**, *306*, 872–876.

[10] R. Fedorov, R. Vasan, D. K. Ghosh, I. Schlichting, *Proc. Natl. Acad. Sci.* **2004**, *101*, 5892–5897.

[11] J. M. Jez, M. E. Bowman, J. P. Noel, *Proc. Natl. Acad. Sci.* **2002**, *99*, 5319–5324.

[12] T. Koudelakova, R. Chaloupkova, J. Brezovsky, Z. Prokop, E. Sebestova, M. Hesseler, M. Khabiri, M. Plevaka, D. Kulik, I. Kuta Smatanova, et al., *Angew. Chemie Int. Ed.* **2013**, *52*, 1959–1963.

[13] V. Liskova, D. Bednar, T. Prudnikova, P. Rezacova, T. Koudelakova, E. Sebestova, I. K. Smatanova, J. Brezovsky, R. Chaloupkova, J. Damborsky, *ChemCatChem* **2015**, *7*, 648–659.

[14] J. Newman, T. S. Peat, R. Richard, L. Kan, P. E. Swanson, J. A. Affholter, I. H. Holmes, J. F. Schindler, C. J. Unkefer, T. C. Terwilliger, *Biochemistry* **1999**, *38*, 16105–14.

[15] M. Otyepka, J. Damborský, *Protein Sci.* **2009**, *11*, 1206–1217.
[16] T. Bosma, J. Damborský, G. Stucki, D. B. Janssen, *Appl. Environ. Microbiol.* **2002**, *68*, 3582–7.
[17] M. Petřek, M. Otyepka, P. Banáš, P. Košinová, J. Koča, J. Damborský, *BMC Bioinformatics* **2006**, *7*, 316.
[18] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker, *PLoS One* **2011**, *6*, e19230.
[19] T. A. Kunkel, *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 488–92.
[20] L. Benatuil, J. M. Perez, J. Belk, C.-M. Hsieh, *Protein Eng. Des. Sel.* **2010**, *23*, 155–159.
[21] P. Holloway, J. T. Trevors, H. Lee*, *J. Microbiol. Methods* **1998**, *32*, 31–36.
[22] L. Tang, Y. Li, X. Wang, *J. Biotechnol.* **2010**, *147*, 164–168.
[23] J. Brezovsky, P. Babkova, O. Degtjarik, A. Fortova, A. Gora, I. Iermak, P. Rezacova, P. Dvorak, I. K. Smatanova, Z. Prokop, et al., *ACS Catal.* **2016**, *6*, 7597–7610.
[24] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **2010**, *38*, W529-33.
[25] S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E.-M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, et al., *PLoS One* **2011**, *6*, e20161.
[26] K. M. S. Misura, D. Baker, *Proteins Struct. Funct. Bioinforma.* **2005**, *59*, 15–29.
[27] T. M. Jacobs, H. Yumerefendi, B. Kuhlman, A. Leaver-Fay, *Nucleic Acids Res.* **2015**, *43*, e34–e34.
[28] W. P. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, H. L. Heyneker, *Gene* **1995**, *164*, 49–53.
[29] W. Gilbert, *Nature* **1978**, *271*, 501.
[30] M. Bashton, C. Chothia, *Structure* **2007**, *15*, 85–99.
[31] A. D. McLachlan, *J. Mol. Biol.* **1979**, *128*, 49–79.
[32] P. J. Baker, A. P. Turnbull, S. E. Sedelnikova, T. J. Stillman, D. W. Rice, *Structure* **1995**, *3*, 693–705.
[33] R. Aroul-Selvam, T. Hubbard, R. Sasidharan, *J. Mol. Biol.* **2004**, *338*, 633–41.
[34] A. M. Burroughs, K. N. Allen, D. Dunaway-Mariano, L. Aravind, *J. Mol. Biol.* **2006**, *361*, 1003–1034.
[35] L. Aravind, R. Mazumder, S. Vasudevan, E. V Koonin, *Curr. Opin. Struct. Biol.* **2002**, *12*, 392–9.
[36] M. L. Mascotti, M. Juri Ayub, N. Furnham, J. M. Thornton, R. A. Laskowski, *J. Mol. Biol.* **2016**, *428*, 3131–3146.
[37] X. Zhou, H. Wang, Y. Zhang, L. Gao, Y. Feng, *Acta Biochim. Biophys. Sin. (Shanghai).* **2012**, *44*, 965–973.
[38] D. Gront, D. W. Kulp, R. M. Vernon, C. E. M. Strauss, D. Baker, *PLoS One* **2011**, *6*, e23294.
[39] E. A. Coutsias, C. Seok, M. J. Wester, K. A. Dill, *Int J Quantum Chem Int. J.*

Quantum Chem. **2006**, *106*, 176–189.
[40] D. J. Mandell, E. A. Coutsias, T. Kortemme, *Nat. Methods* **2009**, *6*, 551–552.
[41] D. G. Gibson, in *Methods Enzymol.*, **2011**, pp. 349–361.

CHAPTER4. Computation-guided design of multi-stimulus responsive multi-enzyme supramolecular assembly towards biodegradation of 1,2,3-trichloropropane

4.1 Introduction

Many enzymes in cells form multi-enzyme complexes, bringing their active sites in close proximity to each other in a spatio-temporally regulated manner. Enzyme clustering can lead to higher metabolic flux as a result of substrate channeling, sequestration of toxic intermediates, and decreased competing side-reactions[1,2][3][4][5]. Based on this natural blueprint, the design of artificial multi-enzyme metabolons has been attempted both *in vitro* and *in vivo* using a host of co-localization strategies[6][7][8][9][10][11][12][13][14]. Although these strategies were successful in improving the efficiency of synthetic metabolic pathways to varying extents, no temporal control, characteristic of natural metabolic assemblies *e.g.,* the purinosome[15], on the formation of the multi-enzyme complexes has been designed. The ability to control synthetic metabolic pathways using external stimuli, such as phosphorylation, light, pH and temperature, would afford precise temporal control

over synthetic pathway flux and dynamics[2]. A modular design approach for introducing stimulus-responsiveness would allow a plug-and-play framework, where any chosen multi-enzyme pathway can be made responsive to multiple chosen stimuli. We, therefore, developed a scaffold-free, modular design approach for using phosphorylation and light to trigger the formation of a multi-enzyme assembly.

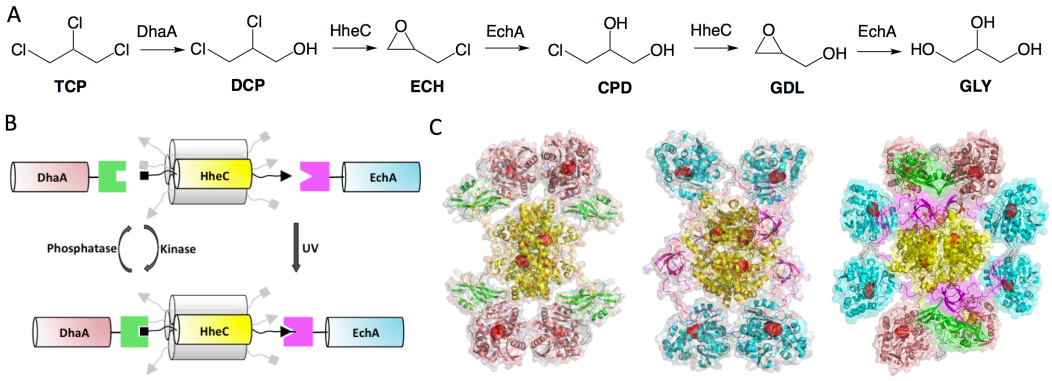

Figure 4.1. A) Synthetic pathway for the biodegradation of 1,2,3-trichloropropane. **B)** Schematic of the designed multi-enzyme assembly controlled by both phosphorylation and UV light. **C)** Illustration of designed multi-enzyme metabolons (active sites highlighted in red) Left: DhaA-HheC two-component assembly; middle: HheC-EchA two-component assembly; right: DhaA-HheC-EchA three-component assembly.

As a model system, we considered a synthetic metabolic pathway, composed of three enzymes, for the degradation of the toxic groundwater pollutant 1,2,3-trichloropropane (TCP) to the benign product glycerol (GLY) which has been previously assembled, both *in vivo* and *in vitro*[16–20][21]. Three pathway enzymes, haloalkane dehalogenase DhaA, haloalcohol dehalogenase HheC and epoxide

hydrolase EchA convert TCP to GLY in five steps (Figure 4.1A). We utilized a variant haloalkane dehaogenase DhaA31, which is ~30-fold more efficient than the wild type DhaA[22] in catalyzing TCP hydrolysis, and wild type EchA and HheC enzymes as our design starting points. Phosphorylation control on multi-enzyme assembly formation was achieved by fusing a previously engineered phosphotyrosine-binding Src homology 2 (SH2) domain[23] and its corresponding phosphorylatable binding peptide to DhaA31 and HheC, respectively. Optical control over enzyme localization was realized by developing a structure-guided computationally-designed Src Homology3 (SH3) domain which could covalently crosslink with methionine on its corresponding peptide under 350nm UV light (Figure 4.1B). We used computational simulations to construct atomic-resolution models of the designed synthetic metabolons to evaluate sites of fusion structures and to resolve the relative locations of active sites as well as geometric compatibility of the resulting inter-enzyme connections.

4.2 Phosphorylation responsive DhaA-HheC two-component assembly

4.2.1 Computational simulation of the DhaA-HheC two-component assembly

Five different SH2 domain-DhaA31 fusions (Figure 4.2) were modeled and the binding partner peptide GEPQEEI was fused at the C-terminal end of HheC via a Glycine-Serine (GS)-rich linker. Computational modeling, performed using Monte

Carlo conformational sampling simulations implemented with the Rosetta software, showed that among the five different fusions, the construct with the SH2 domain fused to the N-terminus of DhaA31 with a flexible linker gave the lowest energy distribution for this two-component assembly (Figure 4.2).

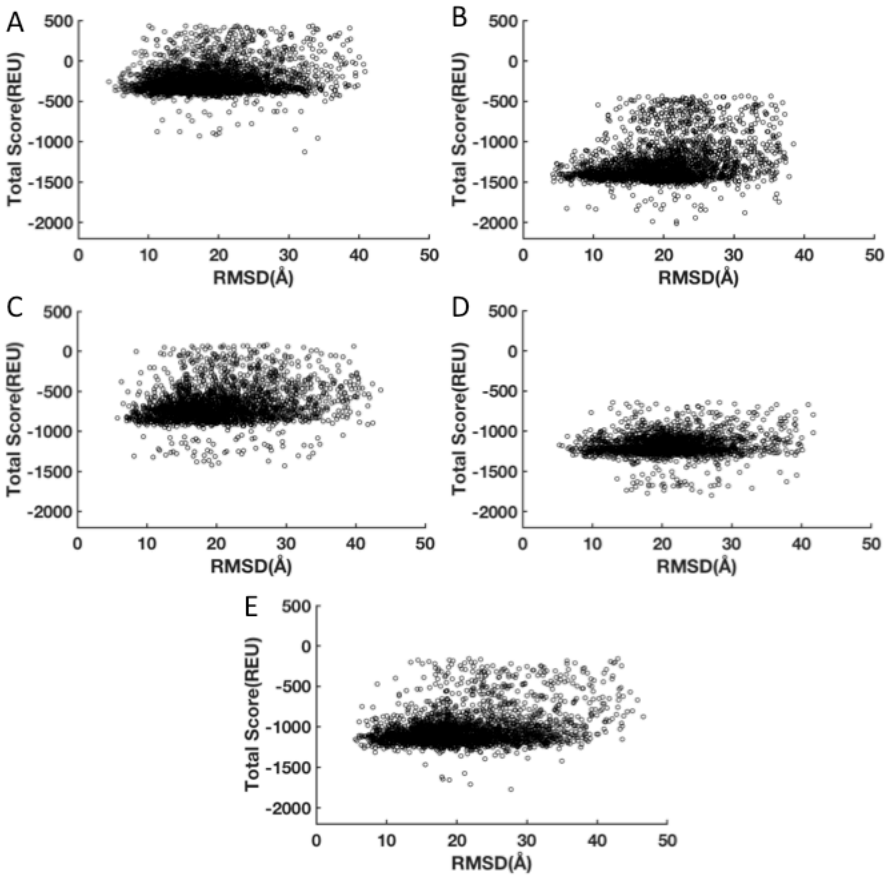

Figure 4.2. SH2-DhaA31-HheC-pYsh2tag two-component assembly energy distribution using HheC-sh2tag and A) DhaA31-SH2; B) SH2-DhaA31; C) DhaA31-FL-SH2; D) SH2-FL-DhaA31; E) DhaA31-HL-SH2.

This analysis also demonstrated that the addition of the SH2 domain at the N-terminus of DhaA31 and the binding peptide at the C-terminus of HheC allowed placement of the active sites within 5-10 nm of each other, and enabled the formation of a contiguous surface linking the two active sites in low-energy

conformations (Figure 4.3A, 4.3B).

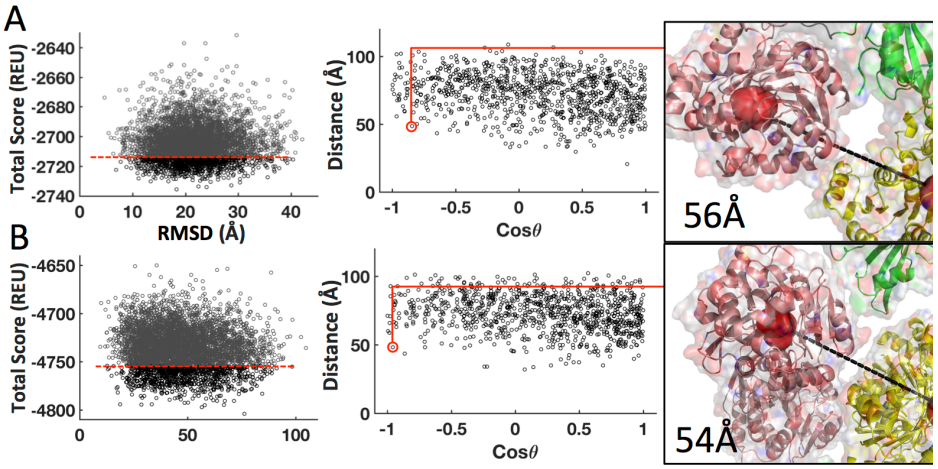

Figure 4.3. A) Asymmetric and **B)** Symmetric conformation simulation of the DhaA-HheC two-component assembly: total score distribution in Rosetta energy units *vs.* RMSD (left); conformations with 20% lowest energy were used to analyze active site distance and orientation between active site access tunnels (middle); Example structures are shown (right) (DhaA31: salmon, SH2 domain: green and HheC: yellow).

We hypothesized that if these low-energy conformations are indeed significantly populated, enzyme co-localization and substrate channeling between the two active sites, including by bounded diffusion on the contiguous surface, may occur[9][24][25][26][27][28]. Enhanced transfer of the intermediate, DCP, to the HheC active site would result in a lower apparent value for the Michaelis constant, K_m, for HheC, and may lead to an improvement in pathway flux. Therefore, we chose N-terminally fused SH2 domain (SH2-DhaA31, hereafter) for use in assemblies.

4.2.2 Experimental results and discussion

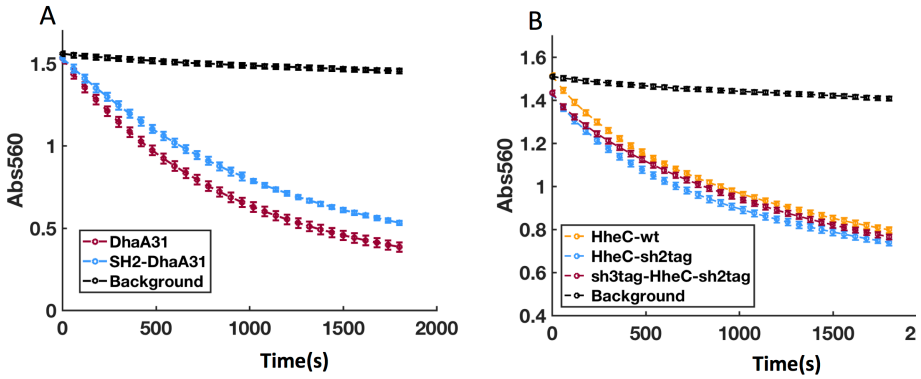

Figure 4.4. A) TCP conversion to DCP activity of SH2-DhaA and DhaA31 wild type B) DCP conversion to ECH activity of HheC, HheC-sh2tag and sh3tag-HheC-sh2tag

Both fusion enzymes, SH2-DhaA31 and HheC-sh2tag (HheC fused at its C-terminal end to the binding peptide for the SH2 domain), were expressed in *E. coli* and purified by standard Ni-NTA chromatography. At the concentration used for generating assemblies, SH2-DhaA31 was marginally worse than DhaA31 in catalyzing the conversion of TCP to DCP, while HheC-sh2tag was as active as wild type HheC (Figure 4.4), indicating that no significant negative effect was caused on the enzymatic activity of either enzyme by constructing fusions.

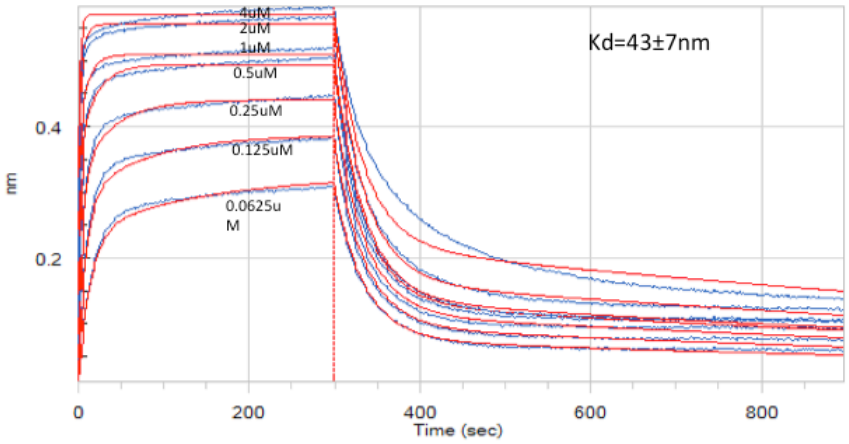

Figure 4.5. Biolayer interferometry (BLI) analysis of SH2-DhaA binding to HheC-pYsh2tag. Association and dissociation phases were 300s and 600s in length separately. HheC-pYsh2tag was biotinylated and bound to the streptavidin sensors as the loading phase. Analyte (SH2-DhaA) concentrations are noted for each trace. Reference-substracted raw data was fit to a global association and dissociation model. All data were measured and analyzed using Octet RED96 system (ForteBio, CA, USA).

The measured binding affinity between SH2-DhaA31 and HheC-pYsh2tag (dissociation constant, K_d = 43±7 nM) was moderately weaker (~10-fold) compared to the reported value for the free SH2 domain for the phosphopeptide[23](Figure 4.5), suggesting a small but measurable impact of enzyme fusion on affinity.

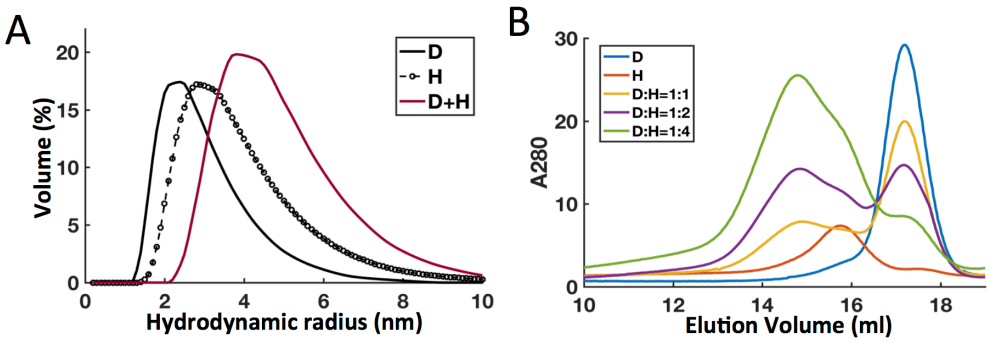

Figure 4.6. A) Volume distribution curves obtained from DLS of the SH2-DhaA31-HheC-pYsh2tag assembly formation. For for DLS measurements, the sample

concentration was 1μM for both SH2-DhaA31 and HhheC-pYsh2tag both as single components and in assembly. B) GFC elution profiles of single enzyme component 5μM SH2-DhaA31, 5μM HheC-pYsh2tag, and mixtures of the two at 1:1, 1:2, 1:4 ratio at fixed 5μM concentration of SH2-DhaA31. D=SH2-DhaA31, H=HheC-pYsh2tag.

Dynamic Light Scattering (DLS) and gel filtration chromatography (GFC) were used to characterize formation and phosphorylation-dependence of the two-enzyme assembly using Src kinase and Lambda phosphatase, and the observed increase (Figure 4.6A, 4.6B) and decrease (Figure 4.7) in the size of particles upon incubation with kinase and phosphatase enzymes, respectively, demonstrates that both assembly formation and disassembly can be controlled in a phosphorylation-dependent manner.

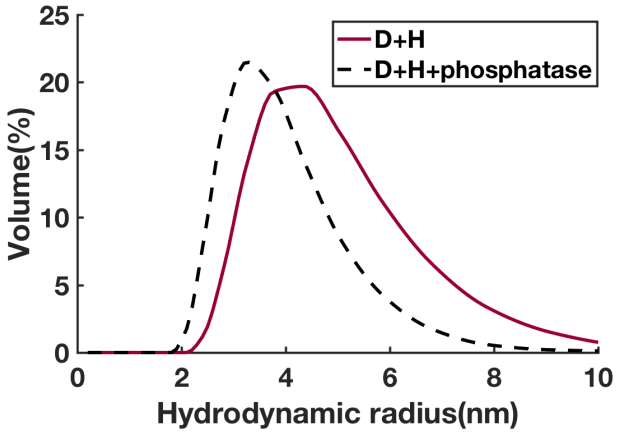

Figure 4.7. DLS characterization of the disassembly of SH2-DhaA31-HheC-pYsh2tag two-component assembly. The sample concentration is 1μM for both D, H in both single component and assembly for DLS measurement. D = SH2-DhaA31, H = HheC-pYsh2tag.

While DLS data showed emergence of particles with a higher hydrodynamic radius upon phosphorylation (Figure 4.6A), GFC traces showed incomplete assembly formation, possibly as a result of dilution on the column (Figure 4.6B). However, increasing molar ratios (monomer:monomer) of DhaA:HheC led to higher apparent

molecular weight species (Figure 4.6B), indicating that assembly formation equilibria are dynamic and optimizable.

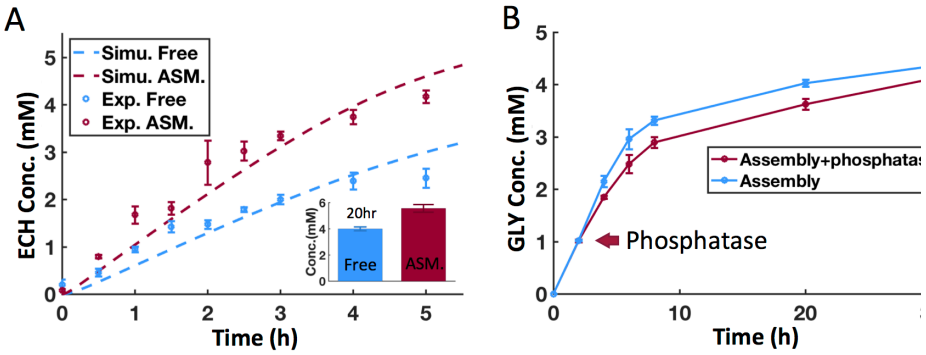

Figure 4.8 A) Simulated and experimentally measured time Course of product ECH converted from 10mM TCP substrate. Bar graph showed TCP to ECH conversion after 20 hours. **B)** TCP conversion activity by SH2-DhaA-HheC-pYsh2tag two-component assembly (1μM:2μM) and free EchA-SH3(1μM) with and without phosphatase(0.1μM) added after 2 hours as determined by production of GLY.

One-pot, two-enzyme conversion of TCP into epichlorohydrin (ECH) was investigated by measuring ECH using gas chromatography (GC), and the assembly showed a faster ECH production rate than a mixture of unassembled enzymes (Figure 2E). Numerical simulations based on Michaelis-Menten kinetics[29] were performed to analyze this two-enzyme in-series reaction system. We found that HheC is inhibited by its product and the ECH inhibition constant was, therefore, included in the simulation (see experimental method chapter). Simulated kinetics agree well with the experimental data obtained for both free enzymes and the assembly, and show that the assembled enzyme system leads to an approximately 3-fold lower apparent K_m for HheC compared to the free enzyme system (Figure 4.8A). While structure-based conformational landscape mapping (Figure 4.3A, 4.3B)

indicates that this decrease in apparent K_m may arise from adoption of channeling-competent geometries in the lowest energy conformers, proximity-based enhancement[30] cannot be entirely ruled out without high-resolution structural characterization. To investigate the stimulus responsiveness of the designed assembly, we performed an experiment

in which the SH2-DhaA31—HheC-sh2tag complex was assembled using phosphorylation, and substrate TCP and free EchA were added at t = 0 hours. At t = 2 hours, excess phosphatase or an equal volume of buffer was added to samples and controls, respectively. The observed decrease in the rate of GLY production after adding phosphatase compared to the controls shows that this two-enzyme system can be dynamically controlled in response to chosen external stimulus (Figure 4.8B).

4.3 UV light responsive two-component assembly

Having established a phosphorylation-mediated assembly between DhaA and HheC enzymes using modular SH2-peptide interactions, we next introduced UV-light-based control of multi-enzyme assembly formation using similarly modular peptide-binding domain interactions. We utilized *p*-benzoyl-L-phenylalanine (pBpa), an unnatural photo-reactive amino acid which has been shown to participate in proximity-mediated covalent bioconjugation with proteins and peptides triggered by 350nm UV light [31–33].

4.3.1 Computational design of photo-reactive SH3 domain

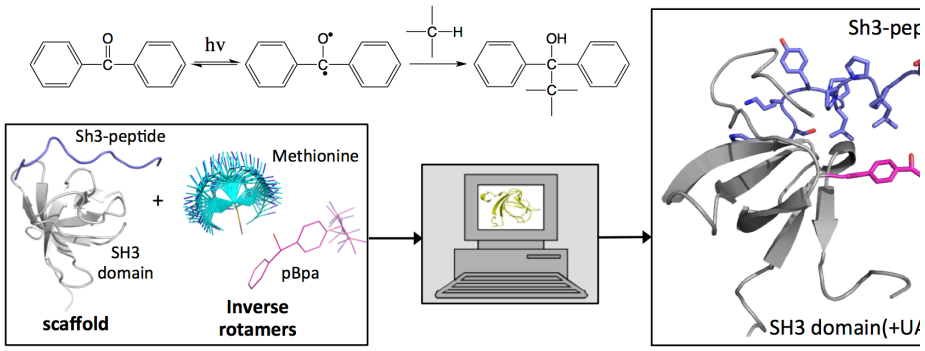

Figure 4.9. A) Schematic of the computational design approach. Design model highlighting the unnatural amino acid (UAA), *p*-benzoyl-L-phenylalanine, in the SH3 domain and methionine in the binding peptide.

Based on the amino acid preferences and the mechanism of crosslinking of pBpa with C-H bonds[32], the Rosetta Enzyme Design approach[34] was utilized to obtain a redesigned SH3 domain (Protein Databank (PDB) code:2OI3) featuring pBpa-methionine crosslink.

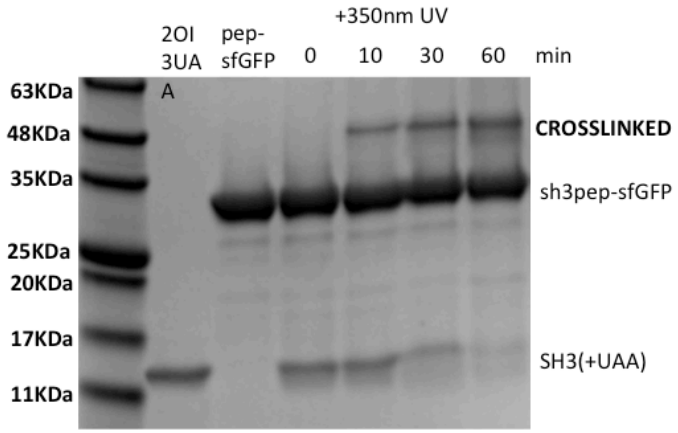

Figure 4.10. SDS-PAGE gel depicting the cross-linking between SH3 (+UAA) and sh3tag-sfGFP under 350nm UV for the indicated time intervals (0-120min).

Protein expression was performed using the pEvol[33] system and characterization of the binding properties of the isolated redesigned SH3 domain showed that it was able to covalently photo-crosslink with its binding peptide upon exposure to 350 nm UV light (Figure 4.10).

4.3.2 Experimental results and discussion

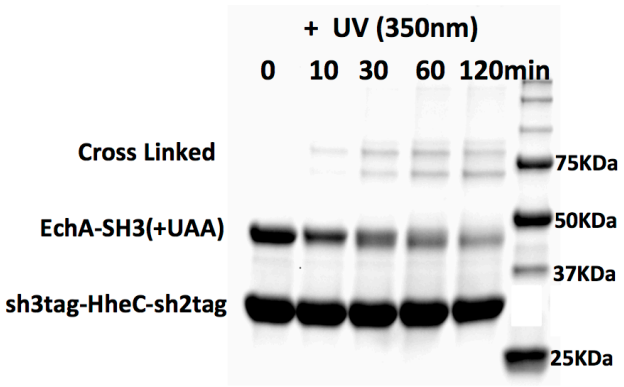

Figure 4.11. SDS-PAGE gel depicting the cross-linking between EchA-SH3 with UAA and sh3tag-HheC-sh2tag under 350nm UV light irradiation for the indicated time intervals (0-120min).

The computationally-designed pBpa-containing SH3 domain was fused to the C-term of EchA *via* a GS-rich linker, and the corresponding designed binding peptide (HSKYPLPPLPSM) was fused to the N-terminus of HheC-sh2tag with another GS-rich linker. Proteins were expressed and purified as described above, and the formation of a photocrosslinked product between EchA-SH3 (44.8KDa) and sh3tag-HheC-sh2tag (31.5KDa) was observed by SDS-PAGE within two hours (Figure 4.11). The crosslinking efficiency was estimated to be ~20%, which is not unlike previous reports where incomplete crosslinking was observed[31].

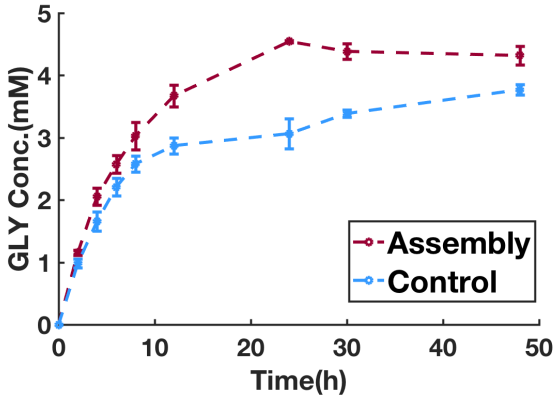

Figure 4.12. DCP conversion activity using preassembled sh3tag-HheC-sh2tag-EchA-SH3 assembly (2μM:1μM) and free enzymes (2μM HheC wild type and 1μM EchA-SH3) was determined by measuring produced GLY.

As the photocrosslinking reaction is proximity-mediated[32], increase in crosslinking efficiency may be achieved by designing a higher affinity variant of the SH3-peptide interface. Nonetheless, these results indicate that our designed SH3 domain-peptide tag can be used to covalently crosslink cargo enzymes fused to them. Despite the modest observed crosslinking efficiency, the yield of GLY (in four steps) from DCP after 12 hours using the crosslinked enzymes was 28% higher compared to a control non-crosslinked enzyme mixture (Figure 4.12).

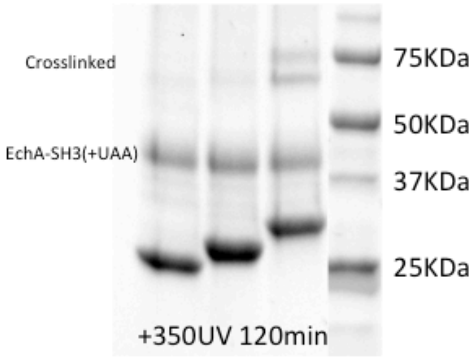

Figure 4.13. SDS gel depicting the cross-linking between 10μM EchA-SH3 and 20μM HheC wild type (lane 1) ; 10μM EchA-SH3 (+UAA) and 20μM HheC-sh2tag; 10μM EchA-SH3 and 20μM sh3tag-HheC-sh2tag (lane 3).

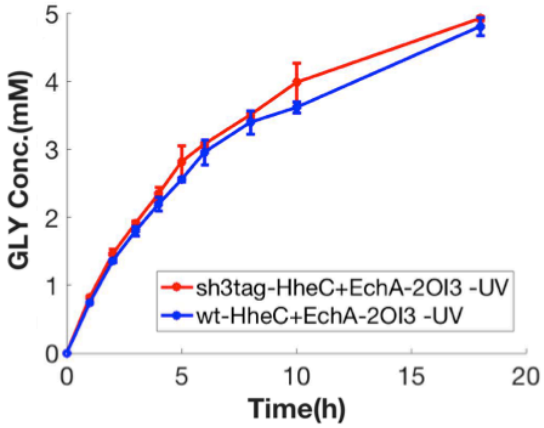

Figure 4.14. DCP conversion activity was determined by measuring GLY concentration on sh3tag-heC-sh2tag+EchA-SH3 (red) and sh3tag-HheC-sh2tag was replaced by HheC wild type as a control (blue). HheC: EchA = 1:2μM.

Photo-induced crosslinking was shown to be strictly dependent on the designed interaction between the SH3-domain and its binding tag fused to the HheC protein: no observable crosslinking between EchA-SH3 and wild type HheC or HheC-sh2tag variants (Figure 4.13) occurred. Further, without crosslinking, no activity improvement was detectable (Figure 4.14).

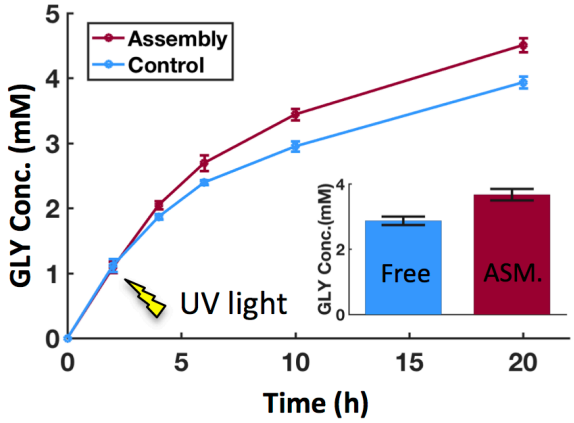

Figure 4.15. DCP conversion activity was determined by measuring GLY concentration produced by assembly (2μM sh3tag-HheC-sh2tag and 1μM EchA-SH3) and free enzymes (2μM HheC wild type and 1μM EchA-SH3). A 350nm UV light stimulus was given at 2h and

lasted for two hours. Bar graph (inset) shows GLY yield after 12 hours with preassembled sh3tag-HheC-sh2tag-EchA-SH3 two-enzyme assembly compared to free enzymes.

Stimulus-responsiveness of the sh3tag-HheC-sh2tag—EchA-SH3 assembly was probed by shining, for 2 hours, 350-nm light after 2 hours of addition of the substrate DCP to component enzymes, and, gratifyingly, a 14% increase in GLY production yield was obtained after 20 hours, compared to the control (EchA-SH3 and wild type HheC) in which no crosslinking was possible (Figure 4.15).

4.3.3 Computational simulation of the HheC-EchA two-component assembly

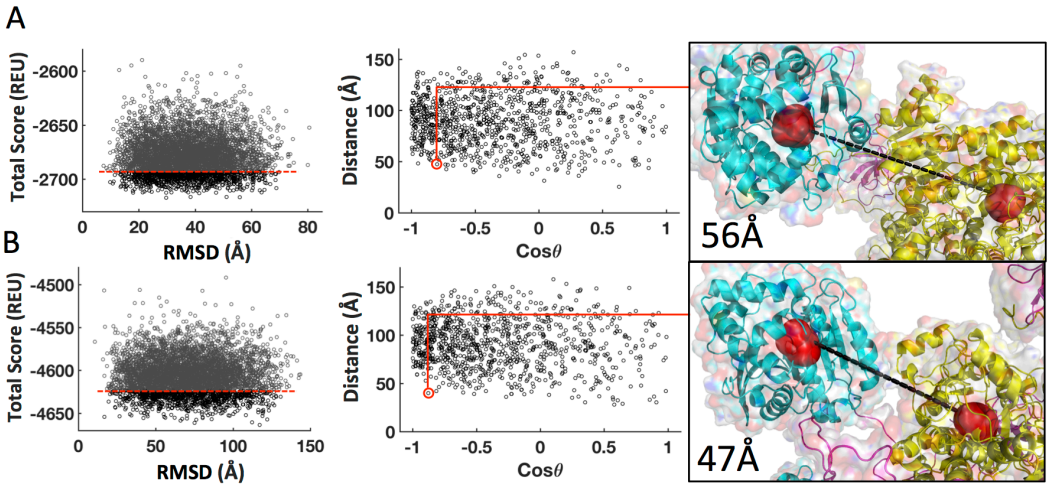

Figure 4.16. A) Asymmetric and **B)** Symmetric conformational simulation of the sh3tagHheCsh2tag-EchA-SH3 two-component assemblies.

To investigate the source of the observed increase in pathway efficacy, computational modeling (conformational sampling) was performed using the same strategy as developed for the DhaA-HheC assembly. In the resulting landscape, we

found that a majority of the distances between the enzyme active site tunnel mouths are within 15 nm and the inter-active site tunnel vector angle (see methods), a measure of the direction of the two active sites with respect to one another, showed a higher density towards 180°, indicating that in the low-energy structures, the active sites are close to each other and are oriented favorably for metabolite channeling (Figure 3D, 3E).

4.4 Multi-stimulus responsive three-component assembly

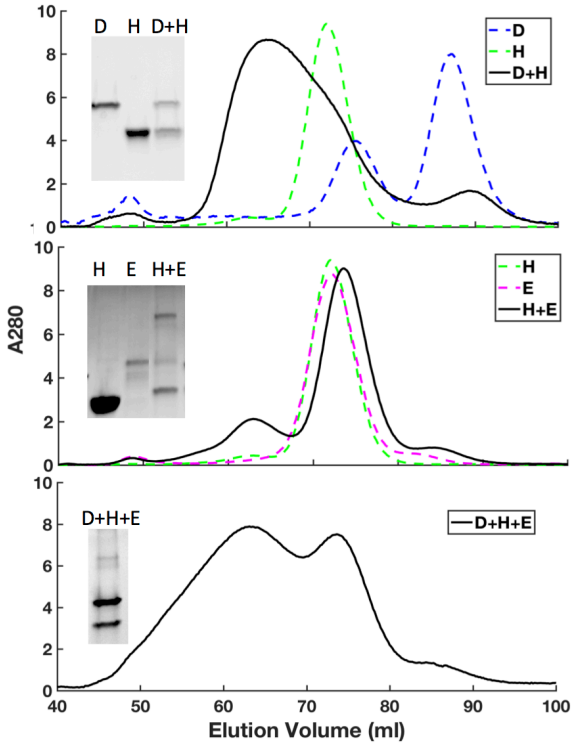

Figure 4.17. Gel filtration elution profiles of SH2-DhaA31, HheC-pYsh2tag and their mixture (top); sh3tag-HheC-sh2tag-EchA-SH3 and their mixture after UV light exposure (middle); and the three-component assembly (bottom). Fractions from the main elution peaks were validated by SDS-PAGE (insets).

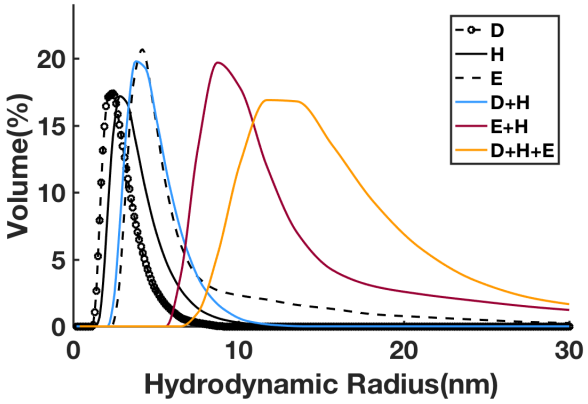

Figure 4.18. Volume distribution curves obtained from DLS of free single enzyme, two-enzyme assembly and three-enzyme assembly formation (1μM, 2μM and 1μM for D, H and E). D=SH2-DhaA, H=HheC-pYsh2tag, E=EchA-SH3.

To combine the phosphorylation and UV-light mediated co-localization, we mixed SH2-DhaA31, EchA-SH3 and sh3tag-HheC-sh2tag together, thus both SH2-DhaA31 and EchA-SH3 were expected to colocalize with the HheC tetramer upon phosphorylation and irradiation with UV light. A mixture of the three enzymes subjected to the two stimuli showed a larger hydrodynamic radius compared to either two-enzyme assembly based on DLS and GFC (Figure 4.17, Figure 4.18).

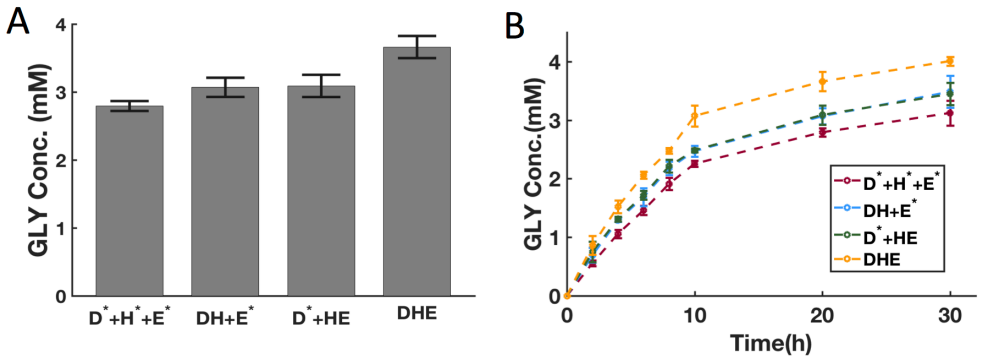

Figure 4.19. A) TCP conversion activity determined by measuring GLY yield after 20 hours. **B)** TCP conversion time course was determined by measuring GLY concentration.

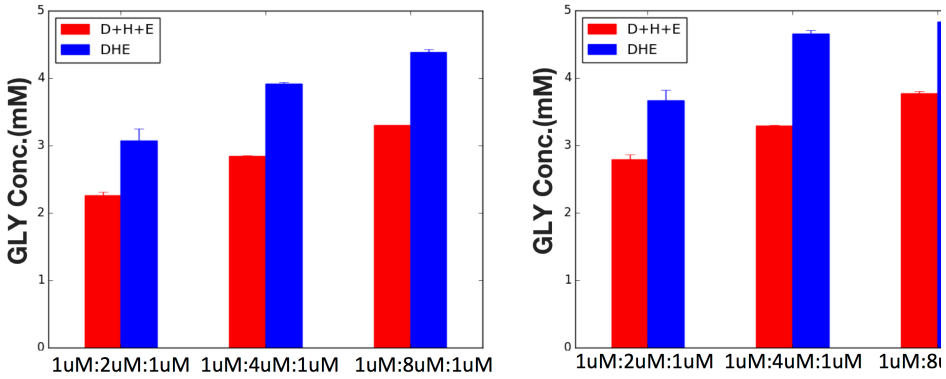

Figure 4.20. Activity of three component system at different ratios of DhaA:HheC:EchA (1μM:2μM:1μM; 1μM:4μM:1μM; 1μM:8μM:1μM) was investigated. GLY was measured after A) 10 hours and B) 20 hours reaction. After 10 hours reaction, the TCP to GLY conversion showed 36%, 37%, 33% improvement for 1:2:1, 1:4:1, 1:8:1 separately.

A comparison between TCP conversion activity catalyzed by (1) free enzymes, (2) SH2-DhaA31-HheCpYsh2tag assembly + free EchA-SH3, (3) free SH2-DhaA31 + sh3tag-HheC-EchA-SH3 assembly, and (4) the three-component assembly, showed that the three-component assembly had the highest conversion efficiency while the free, unassembled enzymes gave the lowest conversion efficiency (Figure 4.19A, Figure 4.19B). The increase observed due to co-localization was insensitive to changes in enzyme stoichiometry used for assembly, indicating that assembly formation was robust (Figure 4.20).

To evaluate if co-localization of enzymes affords other functional benefits such as enhanced thermotolerance, an important parameter for the utility of synthetic

metabolic pathways[6][24], we evaluated the ability of assembled enzymes and unassembled controls to tolerate a heat shock. GLY production yield at room temperature was measured after 30 minutes of exposure to a range of temperatures, and we found that the three-component assembly maintains activity to a greater extent compared to unassembled enzymes (Figure 4.21).

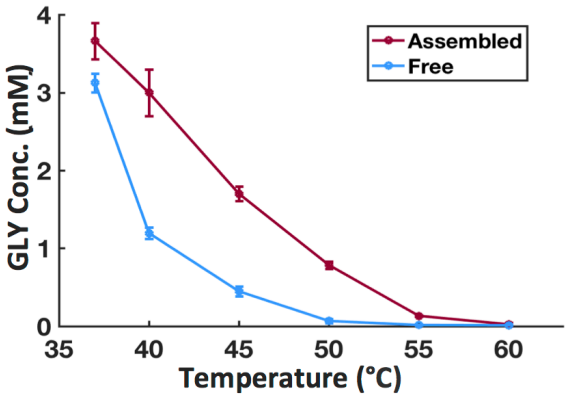

Figure 4.21. Thermotolerance profiles of the three-component assembly and free enzyme mixture determined by residual activity (GLY yields in 20 hours) after 30 minutes of exposure to indicated temperatures.

Thus, apart from improvements in the pathway flux, the designed enzyme assembly also appears to impart greater robustness to harsher environments. Further engineering and optimization of assemblies for improving pathway activity at higher temperatures may benefit the utilization of these enzymes in wastewater treatment[16], and improvements in thermostability may be possible by the stabilization of the individual enzymes of the dehalogenation pathway[35].

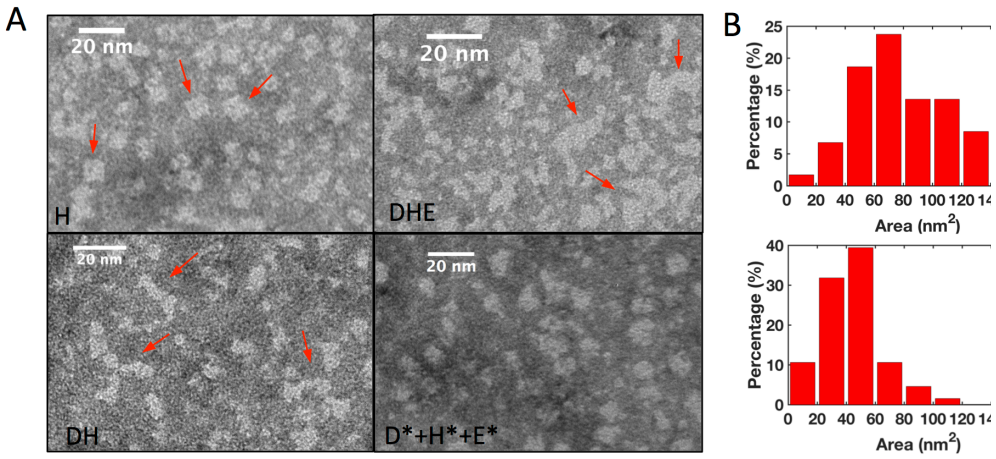

Figure 4.22. A) TEM micrographs of single sh3tag-HheC-sh2tag (top left); SH2-DhaA31-HheCpYsh2tag assembly (bottom left), three-component assembly (top right) and a mixture of free enzymes. B) ImageJ software was used to perform size distribution analysis on the micrographs for three-component assembly (top) and free enzyme mixture (bottom). In all figures D=SH2-DhaA, H=HheC-pYsh2tag, E=EchA-SH3.

Electron microscopy of assembled and unassembled enzymes was used to investigate the diversity of structures and structural changes upon assembly formation. Square shapes of size 7-8nm, presumably corresponding to the HheC tetramer, were the predominant species in the micrographs of single component sh3tag-HheC-sh2tag (Figure 4.22). The fraction of these square shapes decreased dramatically in micrographs of the two-component assembly and several larger and elongated shapes were observed, indicating attachment of SH2-DhaA31 to HheC-sh2tag. The three-component assembly showed a highly diverse shape distribution, including structures that are significantly larger compared to those observed for the two-component assembly, as well as the shapes observed in a mixture of the three free enzymes (Figure 4.22), indicating that the designed interactions, and not non-specific aggregation, are responsible for the detected

shapes. While the micrographs show that assemblies are formed as intended, the large size, polydispersity (partly due to low crosslinking efficiency), and conformational diversity of the assemblies precludes characterization of the observed structures using higher-resolution techniques. Further computationally-guided stabilization of the newly generated interfaces and replacement of GS-rich linkers with structured elements may aid in obtaining more uniform shapes and sizes, suitable for characterization with high-resolution structure-determination techniques.

4.5 Conclusion

In conclusion, we demonstrate that a multi-stimulus responsive multi-enzyme assembly was built using structure-guided and computationally designed modular protein-peptide interaction domains. Reversible and irreversible co-localizaton can be achieved *via* phosphorylation and UV photocrosslinking, respectively. Both phosphorylation and photocrosslinking-driven two-enzyme systems showed improved reaction efficiency compared to non-crosslinked enzyme mixture controls, and the three-component assembly showed a further enhancement of product yields compared to the two-enzyme assemblies. The magnitudes of increase in pathway efficacy were smaller compared to previous spatial localization based studies, which typically involve a reversible step so that co-localization has a large thermodynamic and kinetic boost[36]. However, in contrast to co-localization with

only spatial arrangement, the multi-stimulus responsive enzyme assembly approach provides additional temporal control and may be generally applicable to other pathway systems, including those with a reversible step. *In vivo*, temporal control over synthetic metabolic pathways should allow regulating the levels of often toxic pathway intermediates[11,16] – thereby, making the process more efficient. Overall, we show that our approach for controllably constructing a multi-enzyme complex provides a viable methodology for artificial and dynamic metabolon design using computational structure-based design approaches**.**

4.6 Methods

Computational protocol for conformational sampling of designed assemblies

Five different SH2/DhaA31 fusion were constructed: SH2 domain directly fused to the N-terminus and C-terminus separately; SH2 domain fused to the N-terminus and C-terminus separately with a flexible 6-residue linker and the SH2 domain fused to C-terminus with a helical linker. Its' respective binding peptide GEPQEEI was fused at the C-terminal end of HheC via a GC linker.Computational modeling for the best protein fusion construct was performed with the RosettaScripts application in the Rosetta software suite. Initial SH2-DhaA31-HheC-pYsh2tag assembly constructs of the fusion proteins were modeled by fusing the protein models (protein, linker, and fused domain) together with the Generalized Kinematic Closure application, followed by four repeats of the FastRelax protocol [REF] for minimization. The

energy landscape of each fusion protein was determined by sampling torsion angles of the linker regions between each protein domain with RosettaScripts [REF]. Sampling was accomplished through repeated torsion angle sampling of the linker regions in full atom mode, followed by minimization of all residues within 11 Å of any developed protein interfaces in centroid mode to find energetically favorable contacts in new protein-protein interfaces. This protocol was performed symmetrically on each of the SH2-DhaA fusion proteins bound to the HheC tetramer to ensure no clashing between the fusion proteins. After many repeats of this protocol, the lowest energy structure generated was then run through a symmetric FastRelax protocol [REF] with constraints to minimize the energy further without drastically changing the final conformations.

An altered protocol from above was used to perform a more thorough sampling of the SH2-FL-DhaA31 fusion when bound to the HheC tetramer. Utilizing RosettaScripts and the previously developed constructs of the fusion proteins, eight additional repeats of the FastRelax protocol with relaxed constraints were performed on the fusion protein models to reach an even lower energy minimum. The energy landscape of the HheC-bound fusion protein was performed in a symmetric and an asymmetric manner (four or one SH2-DhaA fusion proteins bound to HheC, respectively). The sampling was accomplished through repeated torsion angle sampling of the linker regions in full atom mode followed by full atom

rotamer sampling and backbone minimization with constraints by residues within 8 Å of any developed protein interfaces. The torsion angle sampling was performed a number two times the amount of linker residues sampled for each decoy and was wrapped in a Generic Monte Carlo function that would perform each successive sampling, accepting the lowest energy sampled structure with a temperature of 10, ultimately recovering the lowest sampled energy structure. The final decoy was then run through a FastRelax protocol with constraints, generating a final output conformation.

Using the generated set of symmetric and asymmetric models of the SH2-DhaA31-HheC-pYsh2tag complexes were generated, the relative distances of the active sites and the orientation of their respective tunnels was determined. For each model, four tunnel vectors were determined for the HheC tetramer, from the geometric center of the carbon alphas of the three residues in the active site of the monomer of the HheC tetramer, to the geometric center of the carbon alphas of three residues that make up the opening to the tunnel to the active site. For the DhaA enzyme a single vector was determined from the geometric center of the carbon alpha atoms of the three residues that form the active site of the DhaA enzyme, to the geometric center of the carbon alpha atoms of the three residues that make up the entrance to the tunnel to the active site. The distance between the enzymatic tunnels is determined with a simple distance formula between geometric

centers of the tunnel entrance of DhaA and HheC, with the angle between the vector sets determined by the cross product formula. The same method is performed for both symmetric and asymmetric models. The input files were attached at the end of the supplementary information file (For SH2-DhaA31-HheC-pYsh2tag assemblies).

Gene synthesis and cloning

The DhaA wild type in pET21b (between NdeI and XolI restriction site), HheC wild type in pET21b (between NdeI and BamHI restriction sites) and EchA wild type in pET28b(between NcoI and HindIII restriction sites) genes were kindly provided by Dr. Jiri Damborsky from Masaryk University, Czech Republic. SH2 and SH3 genes were commercially synthesized (Genscript, USA). DhaA31 was obtained by performing site-directed mutagenesis on DhaA wild type. HheC wild type was subcloned into pET29b between NdeI and XolI to insert a hexa-His tag for NTA-Ni purification. DhaA-SH2 fusions and EchA-Sh3 fusion were constructed using the Gibson assembly cloning method. HheC-sh2tag and sh3tag-HheC-sh2tag were made using RF cloning using HheC wild type as the template.

Enzyme expression and purification

Competent cells of *E. coli* BL21 strain were transformed individually with vectors pET21b::DhaA31, pET29::DhaA31SH2_fusions, pET29b::HheC, pET29b::HheC-sh2tag, pET29b::sh3tag-HheC-sh2tag, and pET28b::EchA. Each

transformed cell line was raised in 0.5L ZY Auto Induction media with 5ml preculture and grown at 37°C for 3 hours till cloudy, followed by dropping the temperature to 18 degree and continued to incubate for 24 hours. Cells were collected and resuspended in 35 mL wash buffer (1XPBS, 30 mM imidazole, pH 7.4). 300 μL 0.1mM PMSF, Lysozyme and DNase(~2 mg/ml and ~0.2 mg/ml as final concentrations, respectively) were added to each resuspension. Resuspended cells were lysed via sonication on ice. The lysed cells were centrifuged at 20,000 rpm for 40 minutes. The his-tagged proteins were purified with Ni-NTA columns (Qiagen, Germany) and the protein was eluted with 10 times column volume of 1xPBS with 300mM imidazole. The elutions were dialyzed with 1XPBS (pH7.4) to remove imidazole. Purity of enzymes was analyzed by SDS-page gel and protein concentrations were measured by spectrophotometer (DS-11+, DeNovix)

Competent cells of *E. coli* BL21 strain were dually transformed with EchA-SH3::pET29b and pEvol-pBpF, which was a gift from Peter Schultz (addgene plasmid #31190). Resulting colonies were used to inoculate 5 mL starting cultures in Terrific Broth (Fisher Scientific) supplemented with the appropriate antibiotic (100 μg/mL chloramphenicol for pEvol, 50 μg/mL kanamycin for pET29b). 500 mL of TB was prewarmed with 1 mM pBpa (Ark Pham, Inc), which was prepared by dissolving 135 mg of pBpa in 550 μL 1M NaOH, and adding it to the prewarmed medium[37]. All 5 mL of the saturated starter culture was then used to inoculate a 0.5

L culture with the same concentration of antibiotics, and the culture was grown until cell density reached an O.D.600 of 0.7-0.8. Then, L-arabinose was added to a final concentration of 0.02% to induce pEvol. After an additional 30 mins of shaking, the UAA-containing protein was induced with 1mM IPTG for 4 hrs at 37°C. Cells were harvested by centrifugation at 6,000 rcf for 15 minutes. The cell pellet was resuspended in 35 mL of wash buffer (1XPBS, 30mM Imidazole, pH7.4) with 300 μL. 0.1mM PMSF, Lysozyme and DNase (~2 mg/ml and ~0.2 mg/ml respectively) Then, samples were sonicated on ice. Lysate was centrifuged at 20,000 rpm for 40 minutes. The his-tagged proteins were purified with Ni-NTA column (Qiagen, Germany) and the protein was eluted with 10-column volume of elution buffer (1XPBS, 300 mM imidazole, pH 7.4). The samples then were dialyzed in 1XPBS (pH7.4) to get rid of imidazole. The purity of enzymes was analyzed on SDS-page gel and the protein concentrations were measured on Spectrophotometer (DS-11+, DeNovix).

Phosphorylation and dephosphorylaiton

HheC-sh2tag and sh3tag-HheC-sh2tag were phosphorylated in a buffer of 5 mM MOPS, 2.5 mM glycerol 2-phosphate, 5 mM MgCl2, 1 mM EGTA, and 0.4 mM EDTA and 2mM ATP. Just prior to use, DTT was added to a final concentration of 0.05 mM. Protein was diluted in phosphorylation buffer to a final concentration at 40μM. 2.5 μL LCK kinase (0.1μg/μL, GST-tagged LCK PRECISIO Kinase from Sigma-Aldrich)

was added. Total reaction volume is 250 μL and reaction was at 4°C for 24 hours. The phosphorylated proteins were then buffer exchanged into 1XPBS (pH7.4) using PD-10 desalting columns.

Phosphorylated protein was dephosphorylated using the protein tyrosine phosphatase (Lambda Protein Phosphatase, New England Biolabs, Inc)

in a buffer of 50 mM HEPES, 100 mM NaCl, 2 mM DTT, 0.01% Brij 35 and 1mM $MnCl_2$, pH 7.5. 30μM phosphorylated protein in 400 μL buffer was dephosphorylated by 5 μL phosphatase at 30°C for 4 hours. The dephosphorylated proteins were then buffer exchanged in 1XPBS (pH 7.4) using PD-10 desalting columns.

Assays for preassembled multi-enzyme activity

All the multi-enzyme conversion was assayed in 5mL of a 50mM tris-SO4 buffer (pH 8.5) in glass vials incubated in 37°C shakers (150 rpm). Enzymes were added to initiate the reaction[16].

TCP to ECH conversion was measured using Gas Chromatography. 500 μL samples was withdrawn from the reaction mixture into a 2 ml glass vial at different time points and mixed with 500μL ethyl acetate for extraction. Anhydrous Na2SO4 was added to dry the extracted samples. 2 μL of the extracted sample in ethyl acetate was analyzed by gas chromatography for detection using external standard calibration. A Gas Chromatograph Clarus 680 with flame ionization detector

(PerkinElmer, USA) and a capillary column DB-VRX, 20m x 0.18mm x 1.0μM (Agilent Technologies, USA) was used to detect ECH yield. Samples (2 μL) were injected into the GC with an inlet temperature of 250°C and split ratio of 20:1. The operating conditions for the column were: helium carrier gas with an initial flow of 1mL/min for 1 min, followed by a flow gradient from 1 to 1.5 mL/min (ramp 0.1 mL/min), temperature program was set to give an initial column temperature of 50°C for 1 min, followed by a temperature gradient from 50 to 150°C (15°C/min ramp), followed by another temperature gradient from 150 to 225°C (25°C/min ramp) and hold for 1min. The temperature of the FID detector was 250°C.

TCP to GLY conversion and DCP to GLY conversion was measured using a Glycerol assay Kit (Sigma-Aldrich, USA). 100 μL samples were taken from 5mL reaction mixtures and heated at 99°C for 10 minutes to destroy enzyme activity, centrifuged at 14000rpm for 1min, diluted in assay buffer accordingly and finally analyzed following the manufacturer's protocol. Concentrations of GLY were calculated from absorbance at 570 nm.

Assays for multi-enzyme stimulus-responsiveness

1μM:2μM preassembled SH2-DhaA31-HheC-pYsh2tag two-enzyme assembly and 1μM EchA-SH3 were added to 5mM TCP to start the reaction. At 2 hours time point, excess (0.1μM) phosphatase in 1x reaction buffer was added to the reaction, while

for the control, same amount of buffer was added to the reaction. Samples were withdrawn and prepared as described before and concentrations of GLY were calculated from absorbance at 570nm using Glycerol Assay Kit.

2μM:1μM EchA-SH3 + sh2tag-HheC-sh3tag assembly and control (EchA-SH3 + HheC-wt) were added to 5μM DCP to start the reaction. At 2 hours time point, the reaction was exposed inside a photoreactor equipped with 350nm UV lamp for two hours. Samples were withdrawn and prepared as described before and concentrations of GLY were calculated from absorbance at 570nm using Glycerol Assay Kit.

Dynamic Light Scattering measurements

Dynamic light scattering (DLS) measurements were performed using a Zetasizer Nano ZS (Malvern Instruments, U.K.). Data was collected with a 3 mw He−Ne laser at a 633 nm wavelength. This unit collects back scattered light at an angle of 173° and contains a built-in Peltier element temperature controller.

Transmission Electron Microscopy

Single enzyme component was filtered by 0.22μM filter before mixing to form assembly. Protein concentration was 1μM for sh3-HheC-sh2; 1 SH2-DhaA31+ 2μM sh3-HheC-pYsh2 for two-component assembly; 1μM Sh2-DhaA31 + 2μM

sh3-HheC-pYsh2 + 1μM EchA-SH3 (+UAA). Carbon-coated FCF400-Cu grids (Electron Microscopy Sciences, Hatfield, PA) were glow-discharged for two hours under UV light with shining side facing up to render the grids hydrophilic and adsorptive. A drop of sample (~5 μL) was dropped on a piece of wax film, grid was put onto the sample droplet for absorption for two minutes. Excess sample solution was removed with a filter paper from the side. A drop (~5 μL) of 1% uranyl acetate was dropped on the wax paper, the grid is put onto the staining solution droplet for two minutes to stain. Excess staining solution was removed by filter paper from the side, and the grids were allowed to air dry for two minutes. Images were collected on JEOL 1200EX electron microscope with AMT-XR41 digital camera. Size distribution analysis of the objects on the TEM image was done by ImageJ.

Computational protocol for design of photoactive SH3

RosettaMatch was used to find the optimal position on the peptide to place the methionine and the optimal position on the SH3 domain to place the pBpa so that the geometric constraints between the pBpa-Met interaction is preserved. The solutions generated by RosettaMatch were then subjected to conformation and sequence optimization through Rosetta repacking and design, respectively. The designs were then evaluated and ranked based on the energies of the domain-ligand interface, deviation from ideal geometries of the pBpa-Met interaction, and the

individual catalytic residues (i.e. methionine and pBpa).

Activity measurement for single components

A pH indicator dye-based colorimetric assay was adapted to measure DhaA and HheC activity quantitatively[38][39]. The buffer for the colorimetric dehalogenase assay consisted of 1 mM HEPES (pH 8.2), 20 mM sodium sulphate and 1 mM EDTA. Phenol red was added to a final concentration of 20 mg / ml. For DhaA, TCP was dissolved in the buffer to a concentration of 5mM, For HheC, DCP was dissolved in the buffer to a concentration of 5mM. The assay was performed in a 96-well plate. Optical density was measured at 560nm using a Tecan Plate Reader Infinite 200. Each microplate well received 146 μL of the substrate-buffer solution and 4 μL of enzyme to a final enzyme concentration 1μM. The optical densities were determined quantitatively over a period of 30 minutes. A decrease in optical density was indicative of proton release. The progress of the reaction was also followed by visual observation as the buffer color changed from red through orange to yellow with decreasing pH.

Michaelis-Menten Kinetic based simulation for SH2-DhaA31-HheC-pYsh2tag two-enzyme system

We built an ordinary differential equation (ODE) model using ode15s, a built-in MATLAB program to simulate ODE's. Simulations provided ECH product time

course profiles over time, for various given concentrations and ratios of enzymes. The parameters set for DhaA and HheC are based on Michaelis-Menten kinetics from a previous study[29].

We also discovered that the product Epichlorohydrin (ECH) exhibited product inhibition to the enzyme HheC. The Ki was experimentally measured by a pH-indicated colorimetric assay for HheC. Initial velocity was measured under different substrate concentration and inhibitor concentration using 1μM HheC. Plots of 1/v against the concentration of inhibitor at each concentration of substrate (Dixon Plot) converge above the x axis, indicating the product inhibition is competitive inhibition and the value of [I] where the plots intersect is –Ki. The measured Ki is around 0.7mM.

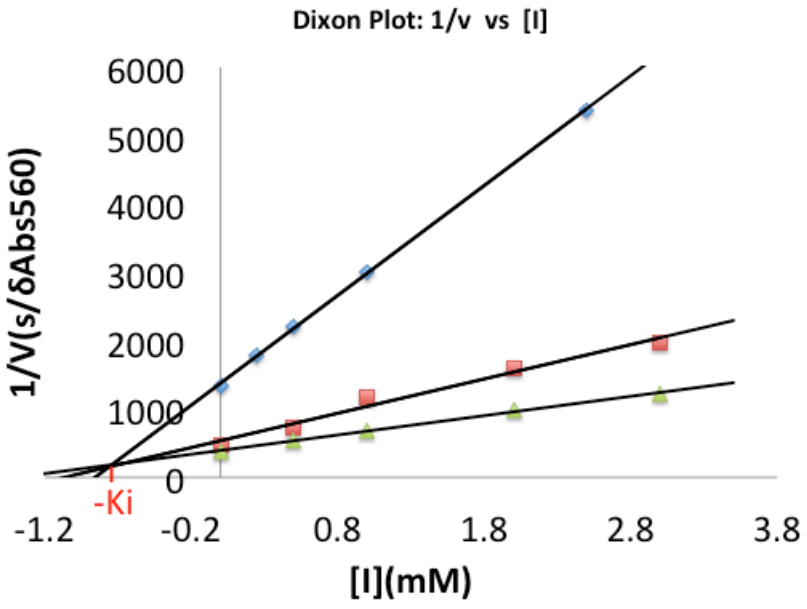

Figure 4.23. Dixon plot of the inhibition of HheC by product ECH in the presence of different substrate concentration (blue square: 1mM DCP, red square: 2mM DCP, green triangle: 5mM DCP) and inhibitor concentrations (0, 0.5, 1, 2,3 mM).

Below are the equations used for the simulation to fit the experimental data for free enzymes. To fit the time course profile of ECH production time course by free enzymes, the Ki was set to 0.2mM which has the same magnitude as the experimentally measured Ki.

```
k1 = 0.58 (S-1);        %k1 = kcat, TCP, (R)-DCP
k2 = 1.79 (mM);         %k2 = km, TCP
k3 = 0.47 (S-1);        %k3 = kcat, TCP, (S)-DCP
k4 = 1.81 (S-1);        %k4 = kcat, (R)-DCP
k5 = 2.49 (mM);         %k5 = km, (R)-DCP
k6 = 0.08 (S-1);        %k6 = kcat, (S)-DCP
k7 = 3.33 (mM);         %k7 = km, (S)-DCP
ki=0.2 (mM)
dCtcpdt = - (k1*Cdhaa*Ctcp)/(Ctcp + k2) - (k3*Cdhaa*Ctcp)/(Ctcp + k2);
dCrdcpdt = (k1*Cdhaa*Ctcp)/(Ctcp + k2) - (k4*Chhec*Crdcp)/(Crdcp + k5*(1+(Cech/ki)));
dCsdcpdt = (k3*Cdhaa*Ctcp)/(Ctcp + k2) - (k6*Chhec*Csdcp)/(Csdcp + k7*(1+(Cech/ki)));
dCechdt = (k4*Chhec*Crdcp)/(Crdcp + k5*(1+(Cech/ki)))+ (k6*Chhec*Csdcp)/(Csdcp + k7*(1+(Cech/ki)))
```

To fit the simulated ECH time course to experimental data catalyzed by two-enzyme assembly, while keeping all other parameter values fixed, the apparent Km for both

(R)-DCP and (S)-DCP needs to be set at 1/3 of the original value, which are 0.6225mM and 0.8325mM separately.

References

[1] I. Wheeldon, S. D. Minteer, S. Banta, S. C. Barton, P. Atanassov, M. Sigman, *Nat. Chem.* **2016**, *8*, 299–309.

[2] C. Schmidt-Dannert, F. Lopez-Gallego, *Microb. Biotechnol.* **2016**, *9*, 601–609.

[3] R. J. Conrado, J. D. Varner, M. P. Delisa, W. Bentley, M. Betenbaugh, *Curr. Opin. Biotechnol.* **2008**, *19*, 492–499.

[4] R. Chen, Q. Chen, H. Kim, K.-H. Siu, Q. Sun, S.-L. Tsai, W. Chen, *Curr. Opin. Biotechnol.* **2014**, *28*, 59–68.

[5] H. Lee, W. C. DeLoache, J. E. Dueber, *Metab. Eng.* **2012**, *14*, 242–251.

[6] X. Gao, S. Yang, C. Zhao, Y. Ren, D. Wei, *Angew. Chemie Int. Ed.* **2014**, *53*, 14027–14030.

[7] O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman, I. Willner, *Nat. Nanotechnol.* **2009**, *4*, 249–254.

[8] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V Ullal, K. L. J. Prather, J. D. Keasling, *Nat. Biotechnol.* **2009**, *27*, 5–8.

[9] J. Fu, M. Liu, Y. Liu, N. W. Woodbury, H. Yan, *J. Am. Chem. Soc.* **2012**, *134*, 5516–5519.

[10] C. You, Y.-H. P. Zhang, *ACS Synth. Biol.* **2013**, *2*, 102–110.

[11] J. V. Price, L. Chen, W. B. Whitaker, E. Papoutsakis, W. Chen, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12691–12696.

[12] F. Liu, S. Banta, W. Chen, *Chem. Commun.* **2013**, *49*, 3766.

[13] Y. Liu, D. P. Hickey, J.-Y. Guo, E. Earl, S. Abdellaoui, R. D. Milton, M. S. Sigman, S. D. Minteer, S. C. Barton, **n.d.**, DOI 10.1021/acscatal.6b03440.

[14] M. Park, Q. Sun, F. Liu, M. P. Delisa, W. Chen, D. Flavia Hozbor, *PLoS One* **2014**, *9*, DOI 10.1371/journal.pone.0097103.

[15] C. Y. Chan, H. Zhao, R. J. Pugh, A. M. Pedley, J. French, S. A. Jones, X. Zhuang, H. Jinnah, T. J. Huang, S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 1368–73.

[16] P. Dvorak, S. Bidmanova, J. Damborsky, Z. Prokop, *Environ. Sci. Technol.* **2014**, *48*, 6859–6866.

[17] T. Bosma, E. Kruizinga, E. J. De Bruin, G. J. Poelarends, D. B. Janssen, *Appl. Environ. Microbiol.* **1999**, *65*, 4575–4581.

[18] N. P. Kurumbang, P. Dvorak, J. Bendl, J. Brezovsky, Z. Prokop, J. Damborsky, *ACS Synth. Biol.* **2014**, *3*, 172–181.

[19] T. Bosma, J. Damborský, G. Stucki, D. B. Janssen, *Appl. Environ. Microbiol.* **2002**, *68*, 3582–7.
[20] M. I. Arif, G. Samin, J. G. E. van Leeuwen, J. Oppentocht, D. B. Janssen, *Appl. Environ. Microbiol.* **2012**, *78*, 6128–36.
[21] P. Dvorak, L. Chrast, P. I. Nikel, R. Fedr, K. Soucek, M. Sedlackova, R. Chaloupkova, V. de Lorenzo, Z. Prokop, J. Damborsky, *Microb. Cell Fact.* **2015**, *14*, 201.
[22] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R. C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, *Nat. Chem. Biol.* **2009**, *5*, 727–733.
[23] T. Kaneko, H. Huang, X. Cao, X. Li, C. Li, C. Voss, S. S. Sidhu, S. S. C. Li, *Sci. Signal.* **2012**, *5*, ra68.
[24] Y.-H. P. Zhang, *Biotechnol. Adv.* **2011**, *29*, 715–725.
[25] M. Pröschel, R. Detsch, A. R. Boccaccini, U. Sonnewald, *Front. Bioeng. Biotechnol.* **2015**, *3*, 168.
[26] F. Lopez-Gallego, C. Schmidt-Dannert, *Curr. Opin. Chem. Biol.* **2010**, *14*, 174–183.
[27] F. M. Raushel, J. B. Thoden, H. M. Holden, *Acc. Chem. Res.* **2003**, *36*, 539–548.
[28] H. O. Spivey, J. Ovádi, *Methods* **1999**, *19*, 306–321.
[29] P. Dvorak, N. P. Kurumbang, J. Bendl, J. Brezovsky, Z. Prokop, J. Damborsky, *ChemBioChem* **2014**, *15*, 1891–1895.
[30] M. Castellana, M. Z. Wilson, Y. Xu, P. Joshi, I. M. Cristea, J. D. Rabinowitz, Z. Gitai, N. S. Wingreen, *Nat. Biotechnol.* **2014**, *32*, 1011–1018.
[31] J. C. Kauer, S. Erickson-Viitanen, H. R. Wolfe, W. F. DeGrado, *J. Biol. Chem.* **1986**, *261*, 10695–10700.
[32] A. Wittelsberger, B. E. Thomas, D. F. Mierke, M. Rosenblatt, *FEBS Lett.* **2006**, *580*, 1872–1876.
[33] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020–4.
[34] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker, *PLoS One* **2011**, *6*, e19230.
[35] D. Bednar, K. Beerens, E. Sebestova, J. Bendl, S. Khare, R. Chaloupkova, Z. Prokop, J. Brezovsky, D. Baker, J. Damborsky, *PLOS Comput. Biol.* **2015**, *11*, e1004556.
[36] K.-H. Siu, R. P. Chen, Q. Sun, L. Chen, S.-L. Tsai, W. Chen, *Curr. Opin. Biotechnol.* **2015**, *36*, 98–106.
[37] C. C. Liu, L. Qi, C. Yanofsky, A. P. Arkin, *Nat Biotech* **2011**, *29*, 164–168.
[38] P. Holloway, J. T. Trevors, H. Lee, **1998**, *32*, 31–36.
[39] W. Xiong, *J. Biotechnol.* **2014**, *147*, 164–168.