Rational and Evolutive Reengineering of Phosphatase: from Method Development to Understanding of Properties

by

Amol Vaijanathappa Shivange

A thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

in

Biochemical Engineering

Approved thesis committee

Prof. Dr. Ulrich Schwaneberg
Professor of Biochemical Engineering
RWTH Aachen University

Prof. Dr. Danilo Roccatano
Professor of Computational Biochemical Engineering
Jacobs University Bremen

Dr. Stefan Haefner
Senior Scientist
Fine Chemicals & Biocatalysis Research
BASF SE Ludwigshafen

Date of Defense: 16th December 2010

School of Engineering and Science

Dedicated to my beloved parents

Table of content

Acknowledgements……..……..……..……..……...1

Abstract……..……..……3

1. Phytase: Introduction and Overview 5

1.1 Introduction to phytase 5

1.2 Production of Phytases 5

1.3 Mechanism of Catalysis 5

1.3 Applications of Phytase 9

1.4 References 10

2. Advances in generating functional diversity for directed protein evolution 13

2.1 Introduction 13

2.2 Main challenges in managing functional diversity 14

2.3 Advancements in general theoretical concepts, computational programs and mutagenesis methods 15

2.3.1 Advancements in general theoretical concepts 15

2.3.2 Advancements in computational programs for restricting sequence space in focused mutant libraries 17

2.4 Advancements in smart (sequence-space restricting) methods for diversity generation 17

2.4.1 Mutagenesis methods for focused mutant library generation 17

2.4.2 Mutagenesis methods for random mutant library generation 20

2.4.3 Methods and programs, for identifying “key beneficial” mutations in recombination experiments with multiple parents 22

2.5 Conclusions 23

2.6 References 23

3. Insight into the rational protein design for enhancing thermostability 25

3.1 Recent Developments 25

3.2 Thermostabilization 26

3.3 Features Contributing to Thermostabilization 26

3.3.1 Hydrogen Bonding 27

3.3.2 Electrostatic Interactions 27

3.3.3 Hydrophobic Interactions 27

3.3.4 Disulphide bonds 27

3.3.5 Metal Binding 27

3.4 Conclusions 28

3.5 References 28

4. Directed evolution of a highly active *Yersinia mollaretii* phytase toward higher thermostability ... 31
4.1 Introduction ... 31
4.2 Material and Methods ... 33
4.2.1 Sequence similarity approach for sequence annotation ... 33
4.2.2 Sequence analysis ... 33
4.2.3 Cloning of *YmappA* into pALXtreme-5b ... 33
4.2.4 Construction of Ymphytase SeSaM library ... 34
4.2.5 High-throughput prescreening ... 34
4.2.6 Protein expression in 96-well plates ... 35
4.2.7 Screening for improved thermostability ... 36
4.2.8 Expression of Ymphytase in a shaking flask and purification ... 37
4.2.9 Characterization of wild-type Ymphytase and mutant ... 37
4.2.10 Molecular Modeling ... 39
4.3 Results ... 39
4.3.1 Sequence annotation and analysis ... 39
4.3.2 Quality of mutant library ... 41
4.3.3 High throughput prescreening of Ymphytase SeSaM library ... 41
4.3.4 Improved Ymphytase screening system ... 41
4.3.5 Characterization of wild-type and mutant Ymphytase ... 43
4.4 Discussion ... 45
4.5 Conclusions ... 48
4.6 References ... 48
5. Toward understanding an activity-thermostability relationship in *Yersinia mollaretii* phytase ... 57
5.1 Introduction ... 57
5.2 Material and Methods ... 58
5.2.1 Site-directed mutagenesis of *YmappA* ... 58
5.2.2 Construction of Ymphytase SeSaM library ... 59
5.2.3 Screening of Ymphytase mutants for thermostability (T_{50}) in test-tube culture ... 59
5.2.4 Homology modeling ... 60
5.2.5 MD simulations ... 60
5.2.6 Cluster Analysis ... 61
5.2.7 Principal component analysis ... 62
5.3 Results ... 63
5.3.1 Identifying important residues in Ymphytase ... 63
5.3.2 Screening a SeSaM mutant library ... 63
5.3.3 Combining “key beneficial” mutations in Ymphytase ... 63

5.3.4 Thermostability profiles 64
5.3.5 Determination of kinetic parameters 65
5.3.6 Homology modeling 66
5.3.7 MD simulations 66
5.3.8 Structural topology of Ymphytase 68
5.4 Discussion 68
5.5 Conclusions 72
5.6 References 72
6. Combinatorial assembly of site saturation in protein segment (ProCASTing): a novel method for protein engineering 81
6.1 Introduction 81
6.2 Material and Methods 82
6.2.1 Selection of protein segment 83
6.2.2 Primer Design 83
6.2.3 Step 1: Template generation 83
6.2.4 Step 2: Fragment amplification with degenerated primers 84
6.2.5 Step 3: Iodine cleavage, ligation and transformation 84
6.2.6 Colony PCR and DNA sequencing 85
6.2.7 Screening of ProCASTing mutant libraries 85
6.3 Results 85
6.3.1 General Concept of ProCASTing 85
6.3.2 ProCASTing library generation and sequencing 86
6.3.3 Screening of ProCASTing mutant library 87
6.4 Discussions 87
6.5 Conclusions 89
6.7 References 89
7. Multisite combinatorial assembly of site saturation (OmniCASTing): a novel method for directed evolution 93
7.1 Introduction 93
7.2 Material and Methods 94
7.2.1 Fragment amplification by PCR 95
7.2.2 Iodine treatment, ligation, and transformation 95
7.2.3 Colony PCR and DNA sequencing 96
7.2.4 Prescreening of OmniCASTing mutant libraries 97
7.2.5 Protein expression and screening in 96-well plates 97
7.2.6 Characterization of wild-type Ymphytase and mutant 97
7.3 Results 97
7.3.1 General Concept of OmniCASTing 97

7.3.2 Fragment amplification by PCR 98
7.3.3 Screening of OmniCASTing mutant library 99
7.3.4 Improving thermostability, activity and pH stability 99
7.4 Discussions 99
7.5 Conclusions 101
7.6 References 102
8. Protein consensus based surface engineering (ProCoS): A novel method for directed protein evolution 105
8.1 Introduction 105
8.2 Materials and Methods 106
8.2.1 Identification of conserved ancestral residues 106
8.2.2 Homology modeling 107
8.2.3 Selection and substitution of sites 107
8.2.4 Gene synthesis and subcloning 107
8.2.5 DNA recombination 107
8.2.6 Screening of ProCoS variants 108
8.2.8 Purification and characterization of wild-type Ymphytase and mutant 108
8.3 Results 108
8.3.1 General concept of ProCoS method 108
8.3.2 Ymphytase surface engineering by ProCoS 109
8.3.3 Characterization of Ymphytase ProCoS variant 110
8.4 Discussion 111
8.5 Conclusions 115
8.5 References 115
9. Conformational dynamics of active site loop in *Escherichia coli* phytase 121
9.1 Introduction 121
9.2 Materials and Methods 123
9.2.1 Atomic coordinates 123
9.2.2 MD simulations 123
9.2.3. Cluster analysis 124
9.2.4. Principal component analysis 124
9.2.5. DomainSelect analysis 124
9.3 Results and Discussion 125
9.3.1 General structural properties 125
9.3.2 Conformational changes of active site loop during simulation 125
9.3.3 Cluster analysis of *E. coli* phytase 126
9.3.4 Principal component analysis 127
9.3.5 DomainSelect analysis 128

9.3.6 Active site analysis of *E. coli* phytase 130
9.4 Conclusions 132
9.5 References 132
10. Water flowed heat inactivation aluminium block for screening of mutant libraries for improved thermostability 149
10.1 Introduction 149
10.2 Materials and Methods 149
10.3 Results and Discussion 150
10.4 Conclusions 151
10.5 References 151
11. Directed evolution platform and solid-phase screening system for *Klebsiella pneumoniae* phytase 153
11.1 Introduction 153
11.2 Materials and Methods 154
11.2.1 Cloning of *K. pneumoniae appA* 154
11.2.2 Expression of Kpphytase in 96-well microtiter plates 154
11.2.3 Solid phase screening system 155
11.2.4 Homology modeling 155
11.3 Results and Discussions 156
11.3.1 Sequence Analysis 156
11.3.2 Expression of Kpphytase in 96-well microtiter plates 158
11.3.3 Solid phase screening system 158
11.3.4 Structural topology of Kpphytase 159
11.4 Conclusions 159
11.5 References 160
Conclusions and future prospects 161
Curriculum vitae 165
Declaration 167

Acknowledgements

Sometimes words are hard to find. It surely is an overwhelming task if not a daunting one because no matter how perfect the lines may seem to me, it will never do justice in acknowledging all the individuals who have supported and guided me to reach this level. Although, their interest for my well-being can never be repaid by mere words, I acknowledge their solicited (and especially unsolicited) advice, efforts and guidance.

I consider myself immensely fortunate to be a student at Jacobs University Bremen (JUB), a pursuit which has been made infinitely gratifying with the expert and congenial guidance of my PhD advisor Prof. Dr. Ulrich Schwaneberg. His constant insistence on prioritizing and balancing of work has inculcated in me a profound awareness of time which has enabled me to meet and beat deadlines successfully. I thank him for his support, constant motivation and for the freedom of thought and freedom of generating scientific ideas that was vowed on me by him. He provided opportunities to get to know the scientific community in Germany and I am thankful for his enough trust on me. His training has been a wonderful learning experience for me and I shall always cherish my memories of working with him.

I express my sincere thanks to Prof. Dr. Danilo Roccatano for his proficient and amiable guidance to gain knowledge of the field of molecular dynamics simulations. It has been a great honor and pleasure to work with him. We have developed hypotheses (phytase loop hypothesis for example) based on computer simulations. Over the years, we have developed a good team to generate novel hypotheses. I am grateful for his constant and unlimited support in broadening my knowledge and understanding on computational simulations.

I express my respectful thanks to Dr. Stefan Haefner (BASF SE) for his fruitful discussion on phytase assay development and library screening. I also thank him for being a member of my PhD thesis committee. I would also like to thank Dr. Annegret Serwe (BASF SE) for constant support with the phytase screening system development. I am grateful to BASF SE for the financial support to accomplish this project.

I express my respectful thanks to Prof. Dr. Joachim Treusch, Director, JUB and Prof. Dr. Bernhard Kramer, Dean of School of Engineering and Science, JUB for providing the infrastructure to complete this work. I would also like to express my thanks to Prof. Dr. Ernst Schmachtenberg, Rector, RWTH Aachen University for providing infrastructure in Aachen.

I am thankful to Prof. Carsten Harms, Bremerhavener Institut of Biological Information Systems (BIBIS), Bremerhaven, Germany for providing genomic DNA of *Klebsiella pneumoniae.*

I convey my whole-hearted thanks to all the Schwaneberg group members who have made great contribution, making JUB and RWTH a fun place to work and for their cheerful company. I express my special thanks to Alex Schenk, Rajni, Radivoje, Milan and Jan for always being ready to go out of their way to help me, and that

too without having any expectations from me. I would also like to thank Hemanshu, Dominik, Güray, Arcan, Dragana, Alex Dennig, Tsvetan, Ronny, Leilei, Noor, Pravin, Ljubica, and Raluca for their friendly ambience generated during the tenure of my PhD. It takes a big heart and a very balanced mind to have this attitude and I deeply appreciate everything that they all have done for me to make my life in this lab a bit easier.

I am grateful to those bright students who worked with me: Alex Dennig, Anita Smarandache, Prashanta Shrestha, Sankar Sivakumar, and Kristin Rübsam.

I am auspicious enough to have lot of friends with whom I have shared the best moments of past three years. With great reverence, I express my debt of gratitude to Rajni, Sagar, Hemanshu, Krishna, Guray and Dominik for their valuable suggestions and help when ever sought for.

Lastly, I would like to apologize to all those who I have not mentioned by name but have asked them favors without inhibition whenever, whatsoever required. I am indebted to them for many ways they have helped me during the course of this work.

Finally and most importantly, I whole-souled thank to my parents who have forever been ‘the wind beneath my wings’. I am especially grateful to my father (Anna) for his UNCONDITIONAL love and support it is a matter of pride to be his Son. To put into words the contribution of my parents in my life, shall remain beyond my literary capacity…

Abstract

This thesis has been divided into five parts. Part-I (chapter 1 – 3) introduces phytase enzyme, advances in diversity generation and rational design approaches for thermostabilization. Phytases catalyze the release of inorganic phosphate by stepwise hydrolysis of phosphomonoester bonds in phytate (substrate). Phytases improves as a feed supplement the nutritional quality of phytate rich diets (e.g. cereal grains, legumes, and oilseeds) by increasing abdominal absorption of inorganic phosphates, minerals, and trace elements. Therefore, phytases have attracted increased interest from animal feed industry (Chapter 1). Directed protein evolution has over the last decades become a versatile and successful algorithm for tailoring enzymes and for advancing our understanding of structure-function relationships. For harvesting the power of directed evolution algorithms it is therefore important that generated mutant libraries are rich in diversity and enriched in active population. The recent advances in computational and experimental methods for high quality mutant library generation have been reviewed (Chapter 2). Various observations on thermostability improvement in enzymes have been discussed those can be used for rational design of phytase for improving thermostability (Chapter 3). Application of phytases faces considerable challenges of low activity and low thermostability, therefore, phytase from *Yersinia* sp. with high specific activity was selected for directed evolution.

Part-II (Chapter 4 – 5) of the thesis focuses on **directed evolution and structure function relationship** of phytase from *Yersinia mollaretii* (Ymphytase). The “key beneficial” mutations identified in the directed evolution have been iteratively combined. Ymphytase variant with 54% improvement in thermostability (58°C for 20 min) and 200 U/mg improved activity was achieved. MD simulations results showed decreased overall Ymphytase flexibility in thermostable variant with slight increase in active site loop flexibility. The decreased flexibility might be due to improved intra-protein interactions like hydrogen bonds (G187S, K289E) and salt-bridge interaction (T77K).

Three conceptually novel methods for protein engineering have been developed in part-III (Chapter 6 – 8). Combinatorial assembly of site saturation test in protein segment (**ProCASTing**), a sequence independent method was developed for parallel site saturation of more than one consecutive site (4 to 8) in any part of the protein. Multisite combinatorial assembly of site saturation test (**OmniCASTing**), practically simple method was developed for parallel site saturation of more than one site (five) regardless of positions in the gene. Using OmniCASTing, a variant with improved thermostability, pH stability and activity was obtained. Three properties improvement might be due to cooperative effects between the new combinations of mutations compare to parent combination. Protein consensus based surface engineering (**ProCoS**) method combining computational analysis and molecular

biology tools was developed. The utility of ProCoS method has been demonstrated by surface engineering of Ymphytase that yielded a variant with 34 amino acid substitution (20% negative polar amino acids) and 3.8 fold improvement in pH stability (pH 2.8).

Two hypotheses have been proposed in part-IV (chapter 8 – 9). The logic underlying ProCoS method was a hypothesis that the conserved residues belong to an ancestor and highly variable regions situated on the surface of protein can be targeted for surface engineering. Surface engineering of Ymphytase resulting in pH stability (affected by surface) improvement supports our first hypothesis. A long scale molecular dynamics studied *on E. coli* phytase fostered a hypothesis that the active site loop might be prerequisite for high activity and catalytic efficiency of phytases.

The last part-V (chapter 10 – 11) of the thesis highlights the miscellaneous work done. A water flowed aluminium block was developed for improving accuracy in the heat inactivation process in mutant library screening for improved thermostability. The aluminum block can be an inexpensive replacement for PCR cyclers (temperature deviation 0.6 °C) with low temperature deviation (0.1 °C). A directed evolution platform and solid phase screening system for *Klebsiella pneumoniae* phytase was developed. This enterobacteriaceae phytase with low catalytic activity was found to possess a α-helix in the active site replacing active site loop. The replacement by less flexible helical structure supports our second hypothesis on higher catalytic efficiency of enterobacteriaceae phytase might be due to identified highly flexible active site loop.

Phytase: Introduction and Overview

1.1 Introduction to phytase

Phosphatase or phytases (myo-inositol hexakis phosphohydrolase), catalyze release of phosphate by stepwise hydrolysis of phosphomonoester bonds in phytate (myo-inositol hexakisphosphate), the salt of phytic acid (Figure 1). Phytate is a major storage form of phosphate and inositol, predominantly occurring in cereal grains, legumes, and oilseeds [1]. Hydrolysis of phytic acid (phytate) to myco-inositol and phosphoric acid is considered as an important metabolic process in several bio-systems.

Phytic acid is considered to be anti-nutrient because of its polyanionic chelating property, that forms complexes with several divalent cations of major nutritional importance, e.g., Ca^{2+}, Mg^{2+}, Zn^{2+}, Cu^{2+}, Fe^{2+}, and Mn^{2+} [2]. Phytic acid can also form complexes with proteins and amino acids. Moreover, non-ruminants (monogastric animals) like human beings, dogs, birds, swine, and poultry have low or no phytase activity [3]. This leads to decrease in bioavailability of inorganic phosphate (Pi) to such animals and as a result undigested/unabsorbed phytate is excreted as large amounts of Pi into the environment [4]. This results in environmental pollution. However, for proper skeletal growth, these animals need Pi at suitable concentration. Thus phytases are considered of significant value in upgrading the nutritional quality of phytate rich feed and supplementation of phytase to the feed provides an alternative to tackle both these conditions effectively.

Society's awareness and increasingly demanding recent regulations world-over on controlling the agricultural pollution, particularly on phosphorus pollution and overcoming anti-nutritive property of phytate have intensified phytase research. Phytases are widespread in nature because they can be found in animals, plants, and microorganisms. Members of the histidine acid phosphatase (EC: 3.1.3.8) family have attracted interest from the animal feed industry because of their high specific activity and their ability to efficiently hydrolyze phytate [5].

1.2 Production of Phytases

Phytases are reported in variety of organism like calves [6], birds, reptiles, and fishes [7], as well as in plants like maize [8], rice [9], wheat [10], and soybean [11]. However, most of the investigation has been done on microbial phytases as listed in Table 1. The commercial phytase preparations authorized in the EU as feed additives are from *Aspergillus niger var. ficuum* (BASF), *Aspergillus awamori* (AB Enzymes) and *Peniophora lycii* (Novozymes) [12].

1.3 Mechanism of Catalysis

The histidine acid phosphatase family of enzymes has acidic pH optima and contains the conserved sequence motif “RHGXRXP” in the

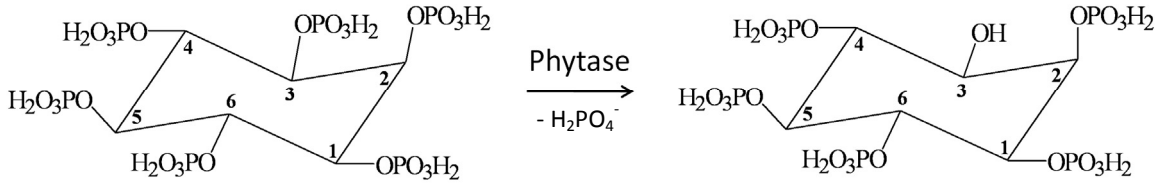


Figure 1: Reaction catalyzed by phytase enzyme.

active site. The crystal structures of several histidine acid phosphatases have been resolved, structures of human [13], rat [14], *Aspergillus niger* [15] and *Escherichia coli* [16] are found be nearly identical. These structures contain a conserved α/β-domain and a variable α-domain. The active site is located at the interface between the two domains. The mechanism of catalysis of acid phosphatase is summarized in Figure 2 and is supported by many studies [17, 18]. Reaction is accomplished by two consecutive phosphoryl transfer steps. In the first step; the enzyme reacts with the phosphate ester (phytate) to form the Michaelis complex. A nucleophilic attack by a group (Asp-OH) of the enzyme on the phosphate group leads to the formation of a covalent phosphoryl-enzyme and the alcohol is released.

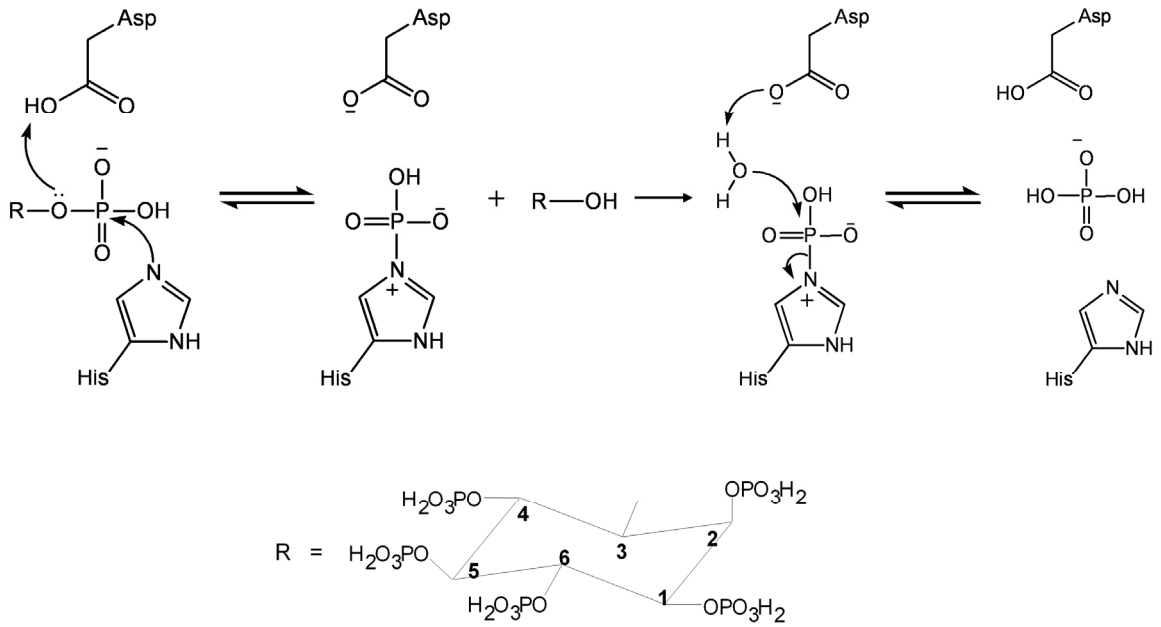


Figure 2: Reaction mechanism of phytase catalysis. Histidine (His) and Aspartic acid (Asp) are active site residues.

Table 1: Published phytase source and activity profile.

Organism	Phytase Source	Phytase Activity*	Reference
Bacteria	*Yersinia intermedia*	3960[a]	[19]
Bacteria	*Yersinia rohdei*	2456[a]	[20]
Bacteria	*Yersinia pestis*	2344[a]	[20]
Bacteria	*Citrobacter amalonaticus CGMCC 1696*	3548[a]	[21]
Bacteria	*Citrobacter braakii*	3,457[a]	[22]
Bacteria	Malaysian Waste–Water Bacterium	1105.8[a]	[23]
Bacteria	*Escherichia coli*	950	[24]
Bacteria	*Escherichia coli*	650	[25]
Bacteria	*Hafnia alvei*	356.7[a]	[26]
Bacteria	*Lactobacillus amylovorus*	146	[27]
Bacteria	*Escherichia coli*	117	[24]
Bacteria	*Escherichia coli*	114	[28]
Bacteria	*Escherichia coli*	105	[29]
Bacteria	*Enterobacter sp.*	70[a]	[30]
Bacteria	*Bacillus subtilis*	35	[31]
Bacteria	*Bacillus licheniformis*	28	[31]
Bacteria	*Mitsuokella jalaludinii*	13	[32]
Bacteria	*Bifidobacterium pseudocatenulatum ATCC 27919*	3.43[a]	[33]
Bacteria	*Bacillus laevolacticus*	2.957	[34]
Bacteria	*Pseudomonas syringae*	2.514[a]	[35]
Bacteria	*Bacillus amyloliquefaciens*	2	[36]
Bacteria	*Klebsiella sp.*	2	[37]
Bacteria	*Bifidobacterium infantis ATCC 15697*	1.011[a]	[38]
Bacteria	*Citrobacter braakii*	1	[39]
Bacteria	*Bacillus sp.*	<1	[40]
Bacteria	*Bacillus subtilis*	<1	[41]
Bacteria	*Klebsiella sp.*	<1	[42]
Bacteria	*Lactobacillus fructivorans*	<1	[43]
Bacteria	*Lactobacillus sanfranciscensis*	<1	[43]
Bacteria	*Megasphaera elsdenii*	<1	[44]
Bacteria	*Prevotella ruminicola*	<1	[44]
Bacteria	*Pseudomonas mendocina*	<1	[45]
Bacteria	*Pseudomonas putida*	<1	[45]
Bacteria	*Selenomonas ruminatum*	<1	[44]
Bacteria	*Weissela confusa*	<1	[43]
Bacteria	*Bifidobacterium catenulatum ATCC 27539*	0.1[a]	[33]
Bacteria	*Bifidobacterium breve ATCC 15700*	0.01[a]	[33]
Synthetic	Synthetic gene (*Peniophora lycii*)	10,540	[46]
Synthetic	Synthetic gene	17.6	[47]
Synthetic	Consensus	NA	[48]

Table 1: (continued)

Organism	Phytase Source	Phytase Activity	Reference
Fungi	*Aspergillus niger*	1008	[49]
Fungi	*Aspergillus awamori*	200	[50]
Fungi	*Neurospora crassa*	125[a]	[51]
Fungi	*Aspergillus niger*	108	[52]
Fungi	*Aspergillus niger*	64	[53]
Fungi	*Aspergillus fumigatus*	62	[50]
Fungi	*Aspergillus fumigatus*	55	[54]
Fungi	*Aspergillus fumigatus WY-2*	51[a]	[55]
Fungi	*Mucor hiemalis Wehmer*	46.7[a]	[56]
Fungi	*Aspergillus niger*	39	[57]
Fungi	*Mucor racemosus*	26	[58]
Fungi	*Aspergillus sp.*	17	[36]
Fungi	*Aspergillus ficuum*	15	[59]
Fungi	*Rhizopus oligosporus*	14	[60]
Fungi	*Mucor hiemalis*	12	[58]
Fungi	*Aspergillus niger*	8	[61]
Fungi	*Aspergillus niger*	7	[62]
Fungi	*Rhizopus oryzae*	6	[59]
Fungi	*Rhizopus oligosporus*	5	[59]
Fungi	*Aspergillus niger*	3	[53]
Fungi	*Rhizopus thailandensis*	3	[59]
Fungi	*Aspergillus oryzae*	2	[63]
Fungi	*Rhizopus microsporus*	1	[59]
Fungi	*Aspergillus oryzae*	<1	[64]
Fungi	*Aspergillus fumigatus*	NA	[48]
Fungi	*Aspergillus niger*	NA	[65]
Fungi	*Aspergillus terreus*	NA	[48]
Fungi	*Thermomyces lanuginosus TL-7*	NA	[66]
Fungi	*Cryptococcus laurentii ABO 510*	0.021[b]	[67]
Yeasts	*Kodamaea ohmeri BG3*	575.5	[68]
Yeasts	*Arxula adininivorans*	3	[69]
Yeasts	*Pichia anomala*	3	[70]
Yeasts	*Fellomyces fuzhouensis*	<1	[69]
Yeasts	*Pichia farinosa*	<1	[69]
Yeasts	*Rhodotorula gracilis*	<1	[71]
Yeasts	*Schwanniomyces occidentalis*	<1	[72]
Yeasts	*Schwanniomyces occidentalis*	<1	[69]
Yeasts	*Sporidiobolus johnsonii*	<1	[69]
Yeasts	*Sporobolimyces sp.*	<1	[69]
Yeasts	*Sterigmatosporus polymorphum*	<1	[69]

*Phytase activity: One unit is the amount of phytase required to liberate 1 μmol of inorganic phosphate per minute from phytate (Reported activity in U/ml)

[a]Reported activity in U/mg

[b]Km of the enzyme

In the next step, the phosphoryl-enzyme intermediate is hydrolyzed forming the non-covalent enzyme/phosphate complex which is then cleaved by nucleophilic enzyme group and water molecule in the active site.

Metal ions play an important role in the catalysis of some acid phosphatase. Based on electron denstity maps Lindqvist et al. showed that the metal species bound at the active site is monooxyanion, which interacts with Arg and His residues in the active site [73]. Moreover, selective modification of His residue in active site by iodoacetic acid and diethylpyrocarbonate was found to inactivate phytase [74]. In *E. coli* phytase, phytate binding induces conformational changes near the active site, confined to opening and closing of active site loop (R20-T25). Therefore, it is important to consider water molecules and metal ions in simulation studies.

1.3 Applications of Phytase

Besides animal nutrition, phytases have wide range of applications in human nutrition and chemical synthesis. Sandberg et al. showed that inositol hexaphosphate (IP6) as well as inositol pentaphosphate (IP5) have inhibitory effects on iron absorption [75] causing high prevalence of iron deficiency, e.g., in infants from developing countries, women of fertile age, or vegetarians. Haros et al. showed that phytase is an excellent bread-making improver [76] which reduced fermentation time. Lower phosphoric esters of *myo*-inositol were found to play vital role in calcium mobilization and transmembrane signaling processes in animal and plants. Hence, need for various inositol phosphate preparations were demanded. However, chemical synthesis was difficult [77] and use of phytase assisted preparation of lower phosphorylated myo-inositol [78]. *Myo*-inositol and *myo*-inositol phosphates were also found to be useful in treatment of atherosclerosis, inflammation and heart diseases by controlling hypercholesterolemia [79].

The supplementation of phytase in pigs and poultry diets has shown to be effective in improving utilization of phytate phosphate and zinc [80, 81]. However the expenses with the limited supply of the commercial phytase and the inactivation during feed pelleting (requires heating at high temperatures) preclude its practical use in animal industry [82]. Therefore, research in the field of improving phytase activity and thermostability has been augmented in recent years.

In the present study, phytases from *Yersinia* species was selected. As observable from Table 1, phytases from *Yersinia* species have been reported to have highest specific activity for phytate between ~2500 – 3960 U/mg compared to other reported phytases. *Y. intermedia* phytase had high activity from pH 2 to 6 and was resistant to pepsin and trypsin. Moreover, enzyme activity was >60% after treatment at 80°C for 10 min, and even 54% of the activity was retained at 80°C for 15 min [19]. Recently, phytase from *Y. kristeensenii* has been reported to be highly pH stability at pH 1.5-11.0 and thermostability [83]. Based on these desired characteristics found in the yersinia genus, we have selected a homolog *Y. mollaretii* phytase which has not yet been studied. Therefore, the selected phytases would be good candidate enzyme to study structure function

relationship responsible for higher activity and thesmostability of theses class of enzymes. Directed evolution and rational protein design approaches in protein engineering were applied to evolve the *Y. mollaretii* phytase. The evolved phytase can provide an efficient and simple system to economically produce phytases for the applications in food and feed industry and the hypotheses generated on structure function relationship might further broaden our understanding of catalytic efficiency of phytases and improve phytase activity and thermostability from fungi and yeast.

1.4 References

1. Cosgrove, D.J., *The chemistry and biochemistry of inositol polyphosphates.* Rev Pure Appl Chem, 1966. 16: p. 209-15.
2. Harland, B.F. and D. Oberleas, *Phytase in animal nutrition and waste management.* BASF Reference Manual, 1999: p. 237–40.
3. Bitar, K. and J.G. Reinhold, *Phytase and alkaline phosphatase activities in intestinal mucosae of rat, chicken, calf, and man.* Biochim Biophys Acta, 1972. 268(2): p. 442-52.
4. Wodzinski, R.J. and A.H. Ullah, *Phytase.* Adv Appl Microbiol, 1996. 42: p. 263-302.
5. Wyss, M., et al., *Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties.* Appl Environ Microbiol, 1999. 65(2): p. 367-73.
6. McCollum, E.V. and E.B. Hart, *On the occurrence of a phytin-splitting enzyme in animal tissues.* J. Biol. Chem., 1908. 4(6): p. 497-500.
7. Rapoport, S., E. Leva, and G.M. Guest, *Phytase in plasma and erythrocytes of various species of vertebrates.* J. Biol. Chem., 1941. 139(2): p. 621-32.
8. Hubel, F. and E. Beck, *Maize Root Phytase (Purification, Characterization, and Localization of Enzyme Activity and Its Putative Substrate).* Plant Physiol, 1996. 112(4): p. 1429-36.
9. Hayakawa, T., Y. Toma, and I. Igaue, *Purification and characterization of acid phosphatases with or without phytase activity from rice bran.* Agric Biol Chem, 1989. 53: p. 1475-83.
10. Nagai, Y. and S. Funahashi, *Phytase (myo-inositol hexaphosphate phosphohydrolase) from wheat bran.* Agric Biol Chem, 1962. 26: p. 794-803.
11. Hamada, J.S., *Isolation and identification of the multiple forms of soybean phytases.* J Am Oil Chem Soc, 1996. 73: p. 1143-51.
12. Haefner, S., et al., *Biotechnological production and applications of phytases.* Appl Microbiol Biotechnol, 2005. 68(5): p. 588-97.
13. Jakob, C.G., et al., *Crystal structure of human prostatic acid phosphatase.* Prostate, 2000. 42(3): p. 211-18.
14. Lindqvist, Y., G. Schneider, and P. Vihko, *Three-dimensional structure of rat acid phosphatase in complex with L(+)-tartrate.* J Biol Chem, 1993. 268(28): p. 20744-46.
15. Kostrewa, D., et al., *Crystal structure of Aspergillus niger pH 2.5 acid phosphatase at 2. 4 A resolution.* J Mol Biol, 1999. 288(5): p. 965-74.
16. Lim, D., et al., *Crystal structures of Escherichia coli phytase and its complex with phytate.* Nat Struct Biol, 2000. 7(2): p. 108-13.
17. Van Etten, R.L., *Human prostatic acid phosphatase: a histidine phosphatase.* Ann N Y Acad Sci, 1982. 390: p. 27-51.
18. Vincent, J.B., M.W. Crowder, and B.A. Averill, *Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions.* Trends Biochem Sci, 1992. 17(3): p. 105-10.
19. Huang, H., et al., *A novel phytase with preferable characteristics from Yersinia intermedia.* Biochem Biophys Res Commun, 2006. 350(4): p. 884-89.
20. Huang, H., et al., *A novel phytase from Yersinia rohdei with high phytate hydrolysis activity under low pH and strong pepsin conditions.* Appl Microbiol Biotechnol, 2008. 80(3): p. 417-26.
21. Luo, H., et al., *A novel phytase appA from Citrobacter amalonaticus CGMCC 1696: gene cloning and overexpression in Pichia pastoris.* Curr Microbiol, 2007. 55(3): p. 185-92.
22. Kim, Y.O.B., et al., *Phytase produced from citrobacter braakii*. 2006, Republic of National Fisheries Research and Develo, Gijang-gun (KR): US.
23. Greiner, R. and A.E. Farouk, *Purification and characterization of a bacterial phytase whose properties make it exceptionally useful as a feed supplement.* Protein J, 2007. 26(7): p. 467-74.
24. Stahl, C.H., D.B. Wilson, and X.G. Lei, *Comparison of extracellular Escherichia coli AppA phytases expressed in Streptomyces lividans and Pichia pastoris.* Biotechnol Lett, 2003. 25(10): p. 827-31.
25. Golovan, S., et al., *Characterization and overproduction of the Escherichia coli appA encoded bifunctional enzyme that exhibits both phytase and acid phosphatase activities.* Can J Microbiol, 2000. 46(1): p. 59-71.
26. Gu, W.N., et al., *Gene cloning, expression and characterization of a novel phytase from Hafnia alvei.* Sheng Wu Gong Cheng Xue Bao, 2007. 23(6): p. 1017-21.
27. Sreeramulu, G., D.S.N. Srinivasa, and R. Joseph, *Lactobacillus amylovorus as a phytase producer in submerged. culture.* Lett Appl Microbiol, 1996. 23: p. 385-88.
28. Rodriguez, E., Y. Han, and X.G. Lei, *Cloning, sequencing, and expression of an Escherichia coli acid phosphatase/phytase gene (appA2) isolated from pig colon.* Biochem Biophys Res Commun, 1999. 257(1): p. 117-23.
29. Miksch, G., et al., *Overexpression of the phytase from Escherichia coli and its extracellular production in bioreactors.* Appl Microbiol Biotechnol, 2002. 59(6): p. 685-94.
30. Kang, S.H., et al., *Cloning, sequencing and characterization of a novel phosphatase gene, phoI, from soil bacterium Enterobacter sp. 4.* Curr Microbiol, 2006. 52(4): p. 243-8.
31. Tye, A.J., et al., *Molecular cloning and the biochemical characterization of two novel phytases from B. subtilis 168 and B. licheniformis.* Appl Microbiol Biotechnol, 2002. 59: p. 190-97.
32. Lan, G.Q., et al., *Optimization of carbon and nitrogen sources for phytase production by Mitsuokella*

jalaludinii, a new rumen bacterial species. Lett Appl Microbiol, 2002. 35(2): p. 157-61.
33. Haros, M., M. Bielecka, and Y. Sanz, *Phytase activity as a novel metabolic feature in Bifidobacterium.* FEMS Microbiol Lett, 2005. 247(2): p. 231-39.
34. Gulati, H.K., B.S. Chadha, and H.S. Saini, *Production and characterization of thermostable alkaline phytase from Bacillus laevolacticus isolated from rhizosphere soil.* J Ind Microbiol Biotechnol, 2007. 34(1): p. 91-8.
35. Cho, J., et al., *Molecular cloning of a phytase gene (phy M) from Pseudomonas syringae MOK1.* Curr Microbiol, 2005. 51(1): p. 11-5.
36. Kim, Y.O., et al., *High-level Expression of a Recombinant Thermostable Phytase in Bacillus subtilis.* Biosci Biotechnol Biochem, 1999. 63: p. 2205-07.
37. Hwang, W.Z., *Screening of phytase-producing bacterial strains from soil and characterization of their phytase enzymes.* Guoli Zhongxing Daxue Nonglin Xuebao, 1999. 48: p. 15-25.
38. Haros, M., et al., *Myo-inositol hexakisphosphate degradation by Bifidobacterium infantis ATCC 15697.* Int J Food Microbiol, 2007. 117(1): p. 76-84.
39. Kim, H.W., et al., *Isolation and characterization of a phytase with improved properties from Citrobacter braakii.* Biotechnol Lett., 2003. 25: p. 1231-34.
40. Choi, Y.M., et al., *Isolation of a phytase-producing Bacillus sp. KHU-10 and its phytase production.* J Microbiol Biotechnol, 1999. 9: p. 223-26.
41. Powar, V.K. and V. Jagannathan, *Purification of phytase-specific phosphatase from Bacillus subtilis.* J Bacteriol, 1982. 151: p. 1102-08.
42. Shah, V. and L.J. Parekh, *Phytase from Klebsiella Sp. No. PG-2: purification and properties.* Indian J Biochem Biophys, 1990. 27: p. 98-102.
43. De Angelis, M., et al., *Phytase activity in sourdough lactic acid bacteria: purification and characterization of a phytase from Lactobacillus sanfranciscensis CB1.* Int J Food Microbiol., 2003. 87(3): p. 259-70.
44. Yanke, L.J., et al., *Phytase activity of anaerobic ruminal bacteria.* Microbiology, 1998. 144(6): p. 1565-73.
45. Richardson, A.E. and P.A. Hadobas, *Soil isolates of Pseudomonas spp. that utilize inositol phosphates.* Can J Microbiol, 1997. 43(6): p. 509-16.
46. Xiong, A.S., et al., *High level expression of a synthetic gene encoding Peniophora lycii phytase in methylotrophic yeast Pichia pastoris.* Appl Microbiol Biotechnol, 2006. 72(5): p. 1039-47.
47. Zou, L.K., et al., *Design and expression of a synthetic phyC gene encoding the neutral phytase in Pichia pastoris.* Acta Biochim Biophys Sin (Shanghai), 2006. 38(11): p. 803-11.
48. Mayer, A.F., et al., *An expression system matures: a highly efficient and cost-effective process for phytase production by recombinant strains of Hansenula polymorpha.* Biotechnol Bioeng, 1999. 63(3): p. 373-81.
49. Krishna, C. and S.E. Nokes, *Predicting vegetative inoculum performance to maximize phytase production in solid-state fermentation using response surface methodology.* J Ind Microbiol Biotechnol, 2001. 26(3): p. 161-70.
50. Martin, J.A., R.A. Murphy, and R.F. Power, *Cloning and expression of fungal phytases in genetically modified strains of Aspergillus awamori.* J Ind Microbiol Biotechnol, 2003. 30(9): p. 568-76.
51. Zhou, X.L., et al., *Biochemical properties of a thermostable phytase from Neurospora crassa.* FEMS Microbiol Lett, 2006. 258(1): p. 61-6.
52. Mandviwala, T.N. and J.M. Khire, *Production of high activity thermostable phytase from thermotolerant Aspergillus niger in solid state fermentation.* J. Ind. Microbiol. Biotechnol., 2000. 24(4): p. 237-43.
53. Han, Y., D.B. Wilson, and X.g. Lei, *Expression of an Aspergillus niger Phytase Gene (phyA) in Saccharomyces cerevisiae.* Appl. Environ. Microbiol., 1999. 65(5): p. 1915-18.
54. Rodriguez, E., E.J. Mullaney, and X.G. Lei, *Expression of the Aspergillus fumigatus phytase gene in Pichia pastoris and characterization of the recombinant enzyme.* Biochem Biophys Res Commun, 2000. 268(2): p. 373-78.
55. Wang, Y., et al., *Cloning, expression, and enzyme characterization of an acid heat-stable phytase from Aspergillus fumigatus WY-2.* Curr Microbiol, 2007. 55(1): p. 65-70.
56. Boyce, A. and G. Walsh, *Purification and characterisation of an acid phosphatase with phytase activity from Mucor hiemalis Wehmer.* J Biotechnol, 2007. 132(1): p. 82-87.
57. Xiong, A.S., et al., *Isolation, characterization, and molecular cloning of the cDNA encoding a novel phytase from Aspergillus niger 113 and high expression in Pichia pastoris.* J Biochem Mol Biol, 2004. 37(3): p. 282-91.
58. Bogar, B., et al., *Production of Phytase by Mucor racemosus in Solid-State Fermentation.* Biotechnol. Prog., 2003. 19(2): p. 312-19.
59. Bogar, B., et al., *Optimization of phytase production by solid substrate fermentation.* J Ind Microbiol Biotechnol., 2003. 30(3): p. 183-89.
60. Sabu, A., et al., *Solid-State Fermentation for Production of Phytase by Rhizopus oligosporus.* Appl Biochem Biotechnol., 2002. 103: p. 251-60.
61. van Hartingsveldt, W., et al., *Cloning, characterization and overexpression of the phytase-encoding gene (phyA) of Aspergillus niger.* Gene, 1993. 127(1): p. 87-94.
62. Hong, K., Y. Ma, and M. Li, *Solid-State Fermentation of Phytase from Cassava Dregs.* Appl. Biochem. Biotechnol., 2001. 93: p. 777-86.
63. Uchida, H., et al., *Expression of Aspergillus oryzae phytase gene in Aspergillus oryzae RIB40 niaD(-).* J Biosci Bioeng, 2006. 102(6): p. 564-7.
64. Shimizu, M., *Purification and Characterization of Phytase and Acid Phosphatase Produced by Aspergillus oryzae K1.* Bioscience, biotechnology, and biochemistry, 1993. 57(8): p. 1364-65.
65. Phillippy, B.Q. and E.J. Mullaney, *Expression of an Aspergillus niger Phytase (phyA) in Escherichia coli.* J. Agric. Food Chem., 1997. 45(8): p. 3337-42.
66. Gulati, H.K., B.S. Chadha, and H.S. Saini, *Production, purification and characterization of thermostable phytase from thermophilic fungus Thermomyces lanuginosus TL-7.* Acta Microbiol Immunol Hung, 2007. 54(2): p. 121-38.
67. van Staden, J., et al., *Phytase activity in Cryptococcus laurentii ABO 510.* FEMS Yeast Res, 2007. 7(3): p. 442-48.
68. Li, X.Y., Z.Q. Liu, and Z.M. Chi, *Production of phytase by a marine yeast Kodamaea ohmeri BG3 in an oats medium: Optimization by response surface methodology.* Bioresour Technol, 2008: p. (in press).
69. Sano, K., H. Fukuhara, and Y. Nakamura, *Phytase of the yeast Arxula adeninivorans.* Biotechnol Lett., 1999. 21: p. 33-38.
70. Vohra, A. and T. Satyanarayana, *A cost-effective cane molasses medium for enhanced cell-bound phytase production by Pichia anomala.* J. Appl. Microbiol., 2004. 97: p. 471-76.
71. Bindu, S., D. Somashekar, and R. Joseph, *A comparative study on permeabilization treatments for in situ determination of phytase of Rhodotorula gracilis.* Lett Appl Microbiol., 1998. 27: p. 336-40.

72. Lambrechts, C., et al., *Influence of culture conditions on the biosynthesis of Schwanniomyces castellii phytase.* Biotechnol Lett., 1993. 15(4): p. 399-404.
73. Lindqvist, Y., G. Schneider, and P. Vihko, *Crystal structures of rat acid phosphatase complexed with the transition-state analogs vanadate and molybdate. Implications for the reaction mechanism.* Eur J Biochem, 1994. 221(1): p. 139-42.
74. Wang, X.-Y., et al., *Kinetics of Inactivation of Phytase (phy A) During Modification of Histidine Residue by IAA and DEP.* Protein Pept Lett., 2006. 13: p. 565-70.
75. Sandberg, A.-S., et al., *Inositol phosphates with different numbers of phosphate groups influence iron absorption in humans.* Am J Clin Nutr, 1999. 70(2): p. 240-46.
76. Haros, M., C.M. Rosell, and C. Benedito, *Use of Fungal Phytase to Improve Breadmaking Performance of Whole Wheat Bread.* J. Agric. Food Chem., 2001. 49(11): p. 5450-54.
77. Potter, B.V.L., *Book Review: The Inositol Phosphates. Chemical Synthesis and Biological Significance. By D. C. Billington.* Angewandte Chemie Int Ed., 1993. 32(7): p. 1103-04.
78. Dvorakova, J., et al., *Formation of myo-inositol phosphates by Aspergillus niger 3-phytase.* Folia Microbiol (Praha), 2000. 45(2): p. 128-32.
79. Cavazza, C., *Carnitine and inositol phosphate-containing composition useful as dietary supplement or drug.* 2000, Sigma-tau Healthscience S.P.A. Cavazza, Claudio: WO.
80. Simons, P.C., et al., *Improvement of phosphorus availability by microbial phytase in broilers and pigs.* Br J Nutr, 1990. 64(2): p. 525-40.
81. Lei, X.G., et al., *Supplementing corn-soybean meal diets with microbial phytase linearly improves phytate phosphorus utilization by weanling pigs.* J. Anim Sci., 1993. 71(12): p. 3359-67.
82. Jongbloed, A.W. and P.A. Kemme, *Effect of pelleting mixed feeds on phytase activity and the apparent absorbability of phosphorus and calcium in pigs.* Anim. Feed Sci. Technol., 1990. 28(3-4): p. 233-42.
83. Fu, D., et al., *A highly pH-stable phytase from Yersinia kristeensenii: Cloning, expression, and characterization.* Enzyme Microb. Technol., 2008. 42(6): p. 499-505.
84. Sajidan, A., et al., *Molecular and physiological characterisation of a 3-phytase from soil bacterium Klebsiella sp. ASR1.* Appl. Microbiol. Biotechnol., 2004. 65: p. 110-18.

Advances in Generating Functional Diversity for Directed Protein Evolution

Amol V. Shivange[a,b], Jan Marienhagen[a], Hemanshu Mundhada[a,b], Alexander Schenk[b], and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Despite advances in screening technologies, only a very small fraction of theoretical protein sequence can be sampled in directed evolution experiments. At the current state of random mutagenesis technologies mutation frequencies have often been adjusted to values that cause a limited number of amino acid changes (often 1 – 4 amino acid changes per protein). For harvesting the power of directed evolution algorithms it is therefore important that generated mutant libraries are rich in diversity and enriched in active population. Insufficient knowledge about protein traits, mutational robustness of protein folds, and technological limitations in diversity generating methods are main challenges for managing the complexity of protein sequence space. This chapter covers computational and experimental advances for high quality mutant library generation that have been achieved in the last two years.

2.1 Introduction

The field of enzyme biotechnology is undergoing rapid change and diversification. Proteins, when properly configured, have a remarkable capacity for carrying out complex molecular processes with precision and efficiency such as high activity and stability under process conditions, desired substrate selectivity, high enantio-selectivity or enantio-specificity, pH range, etc [1]. However, wild-type enzymes often need to be optimized to fulfill these requirements. Two rather mutual tools can be used on a molecular level to create desired enzymes include directed evolution and rational protein design. The concept of directed evolution was introduced as early as 1967. It has been 40 years since Spiegelman and co-workers demonstrated how RNA molecules can be evolved in the test tube [2]. This result established evolution as a chemical process and paved the way for the many directed evolution experiments for successful tailoring of enzymes. In essence, directed evolution implements an iterative optimization process, whereby, the fittest variants are selected from a large pool of variants harboring random mutations. The great benefit of directed evolution techniques is that they require no prior structural knowledge of a protein, nor it is necessary to be able to predict what effect a given mutation will have. Certainly,

*Corresponding author

the results of directed evolution experiments are often surprising in that desired changes are often caused by mutations that no one would have expected.

A typical directed evolution campaign can be accomplished in five major steps [3] which comprise rational design, diversity generation, cloning of variants, screening for improved variants and isolating the gene encoding for the improved protein variant (Figure 1).

Directed protein evolution has over the last decades become a versatile and successful approach for tailoring protein properties to industrial demands and for advancing our understanding of structure-function relationships in biocatalysts. In iterative cycles of diversity generation and functional selection for improved variants, numerous success stories (e.g. enantioselectivity [4], enzymes for bioremediation [5], vaccines [6]) have been published.

Despite technology pushes in high-throughput screening technologies (in vitro compartmentalization [7], cell and phage display [8]) experimentalists are confronted with combinatorial complexity of protein sequence space. The theoretical protein sequence space of only a 10 amino acid peptide is 1024 x 10^{13} if every amino acid of this very short sequence can be changed to all others [9]. Let us assume that such a sequence is represented by the height of the Mount Everest. Screening one million variants would in relation represent an “impressive uphill climb” of 0.86 mm. Therefore, strategies and methods are required for understanding how many amino acid positions should be changed simultaneously for improving rapidly protein properties. The latter protein design principles will differ from protein fold to protein fold and from the targeted property. Key to discover those design principles is the generation of high quality mutant libraries which are enriched in the desired functional trait so that experimental screening demands can be matched and statistical relevant data can be generated. Figure 2 illustrates how progress in computational understanding of functional diversity generation, random mutagenesis methods, and recombination techniques deepens our understanding of high quality mutant library design. The progress in each area is presented and discussed in the context of the main challenges in functional diversity generation (Figure 2).

2.2 Main challenges in managing functional diversity

The main challenges for designing functionally enriched (high quality) mutant libraries are: I) Understanding of protein fold- and trait-dependent consequences of mutations on activity and stability of a protein. Hence, discovering principles for predicting optimal mutational loads in directed evolution experiments and developing theoretical concepts to predict mutational effects on protein folds are of great importance. II) In-depth knowledge of structure-function relationships are prerequisites for designing smart (sequence-space restricting) strategies for diversity generation in random mutagenesis and focused mutant libraries. III) Methods and programs for identifying “key beneficial” mutations [10] are

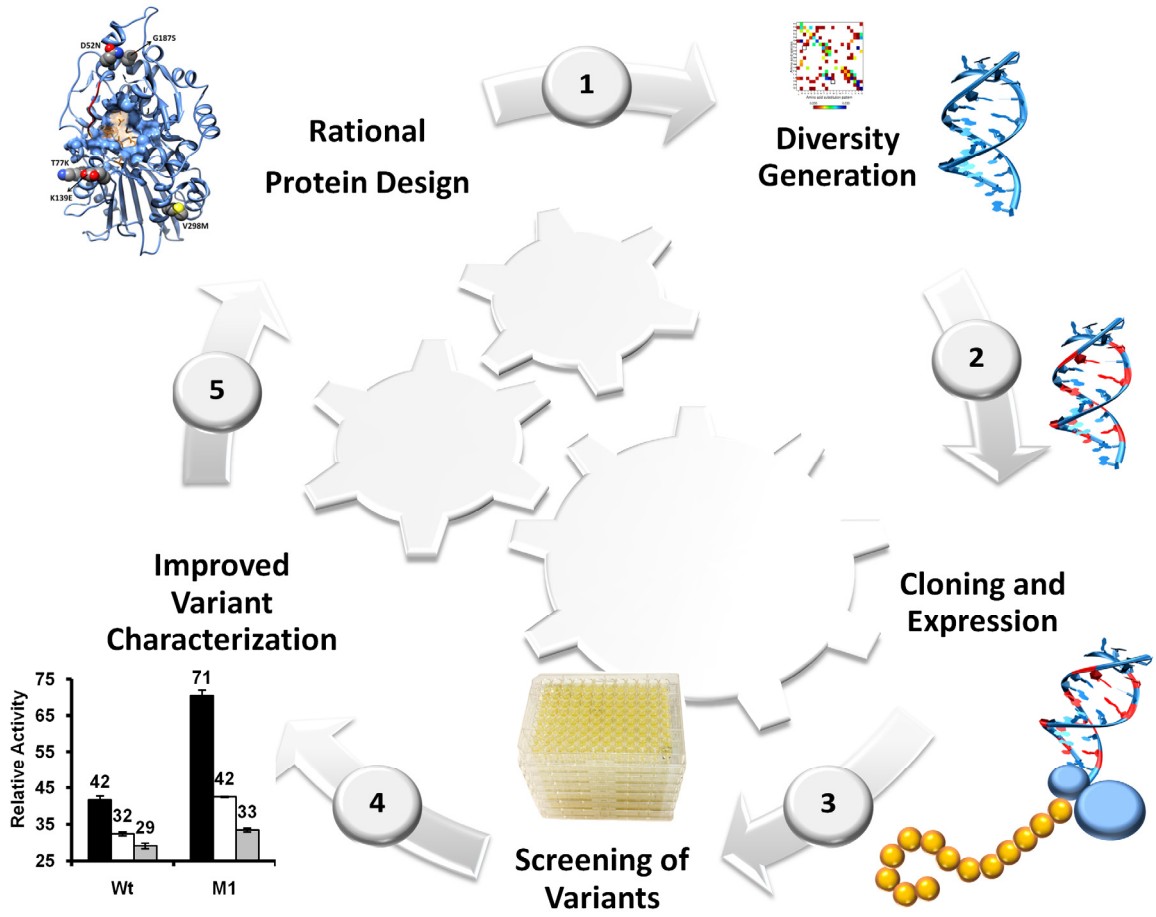


Figure 1: Schematic representation of typical directed evolution strategy.

important for providing the experimental data required to advance our theoretical concepts for reengineering proteins. Up to date many beneficial mutations are lost or not identified in recombination experiments with multiple parents due to technological and theoretical limitations.

2.3 Advancements in general theoretical concepts, computational programs and mutagenesis methods

2.3.1 Advancements in general theoretical concepts

Advances in general concepts to study and to understand mutational effects on protein properties comprise models for mutational thresholds (e.g. stability threshold model [11] and a bi-Gaussian model [12]). An important question for directed protein evolution is whether each mutation decreases the average fitness of a protein by the same amount or are several mutations more or less affective? A reduction higher than the decrease in average fitness is called negative epistasis, a lower one positive epistasis. For the design of high quality mutant libraries it is important to determine mutational thresholds for different protein folds and properties. Based on evolution of lactamases in presence of antibiotics [11] stability threshold

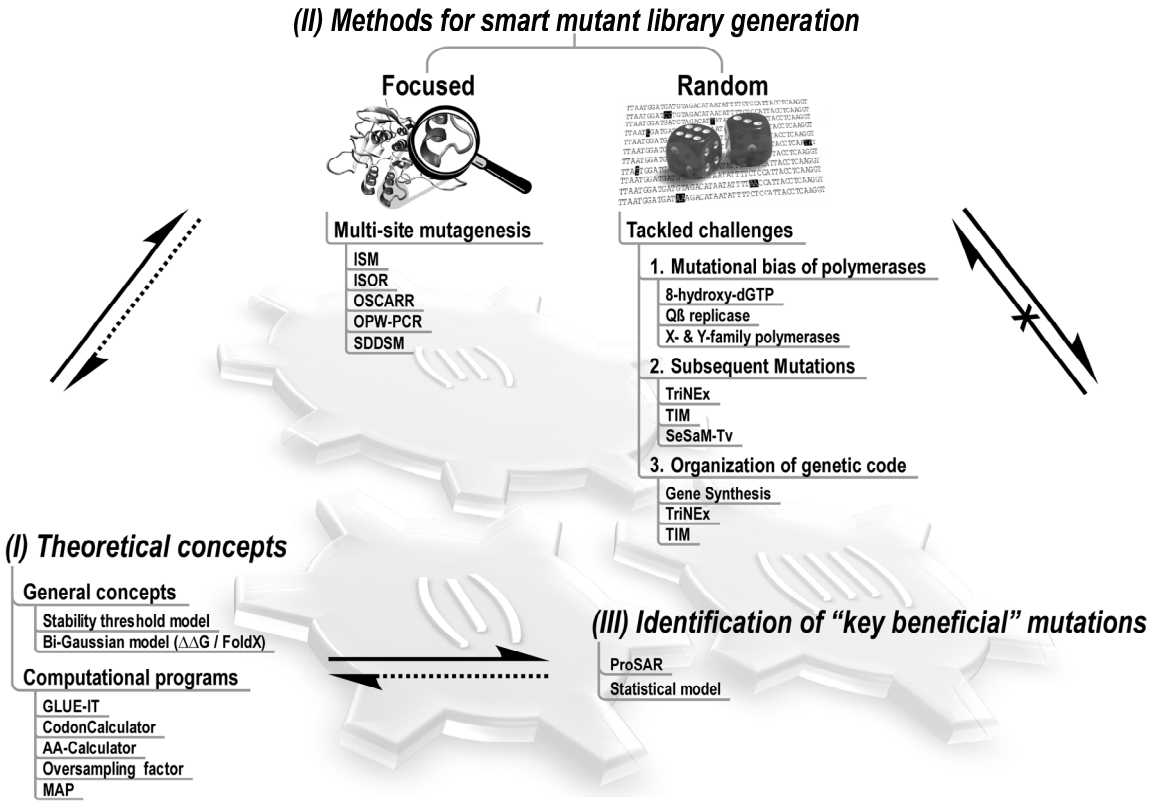


Figure 2: Progress in theoretical concepts (I) and methods for functional diversity generation (II-III). Arrows indicate the interplay between theoretical concepts, programs and methods managing and exploring protein sequence space (⟶ high impact; ⇢ low impact; ⟶✕ marginal impact).

models have been developed which predict that stable proteins are robust to the first few mutations before their fitness begins to decrease exponentially [11] or more than exponentially [13]. A main result of these analyses is that negative epistasis is observed for only the first few mutations, until the stability margin of the lactamase exhausted. Determining mutational thresholds for various protein folds and properties will provide first rules for adjusting mutation frequencies experimentally and analyzing the obtained amino acid substitutions patterns with the Mutagenesis Assistant Program MAP [14]. The latter might allow us to identify repeating patterns for understanding destabilization processes. The bi-Gaussian model predicts distributions of mutational stability effects based on thermodynamic stability (ΔΔG; [12]). Using the FoldX algorithm ΔΔG effects on various proteins were analyzed and similar distributions were found for various proteins which are unrelated in sequence and fold. Analysis of ΔΔG effects is useful to analyze and to predict the stability of muteins (protein variants differing from the respective wild-type by at least one amino acid substitution) using computational programs. It is further discussed that first improving stability of a protein might increase the success probabilities in finding improved muteins in small mutant libraries [13].

2.3.2 Advancements in computational programs for restricting sequence space in focused mutant libraries

Focused mutant libraries have been successfully used for improving localized protein properties [9] like enantioselectivity, activity and specificity if a reliable crystal structure was solved or a reliable model generated. Selection of one or multiple positions for saturation mutagenesis is often sufficient to alter enantioselectivity, activity or specificity to synthetic demands. Prerequisites for a successful saturation mutagenesis are the choice of degenerate codons (e. g. NNN/NNK (N: adenine/cytosine/guanine/thymine; K: guanine/thymine) and the number of variants that has to be screened for covering a significant fraction in the mutant library. For the latter purpose GLUE-IT was developed; GLUE-IT calculates an expected diversity and required library size to sample through a mutant library with a predetermined coverage [15]. By the same group the programs CodonCalculator and AA-Calculator were developed which assist experimentalists in choosing codons for randomized mutant library construction. CodonCalculator provides the number of each possible amino acid and codon variants upon input of a codon (standard codon of the genetic code), whereas AA-Calculator provides degenerated codon(s) upon amino acid input that are optimal for encoding all “desired” amino acids [15]. For achieving high quality mutant libraries it is important to restrict the generated protein sequence to a highly functional sequence space, eliminating stop codons and destabilizing amino acid substitutions such as glycine and proline in helical structures or tryptophans. Based on the GLUE-IT algorithm an ‘oversampling factor’ was introduced [16] which calculates a required number of transformants that has to be screened in site and multi-saturation mutagenesis experiments [16]. Covering 95% of the generated sequence space in three positions using conventional NNK randomization (32 codons; 20 amino acids; K: G, T) requires screening of 98163 variants compared to 5175 using a NDT-codon library (12 codons; D: A, G, T; balanced chemically diverse 12 amino acids, no stop codons) [13]. The quality of the NDT library, measured by the frequency of positive variants in a library of 5000 clones (NDT library 511 hits; NNK library – 38 hits), has proven to be 13-fold higher compared to the NNK library [16]. These findings are especially useful for experimentalists employing medium throughput screening systems in multiple site directed mutagenesis experiments.

2.4 Advancements in smart (sequence-space restricting) methods for diversity generation

2.4.1 Mutagenesis methods for focused mutant library generation

Combinatorial randomization of several targeted sites is an efficient way to use focused mutant libraries for improving enantioselectivity, activity, specificity and thermal resistance of proteins [9]. A recent conceptual advantage, the Iterative Saturation Mutagenesis (ISM), comprises iterative cycles of saturation mutagenesis at rationally chosen sites (four sites in [17]). The selected positions will individually be randomized and screened. The recombination strategy after initial saturation is

Table 1: Comparison of key performance parameters of seven random and five focused mutagenesis methods for generating high quality mutant libraries which are enriched in function. The main limitation in generating high quality mutant libraries are: A) Biased mutational spectra caused by DNA-polymerases (Transition bias; often chemically conservative substitutions // Transversion bias; chemically more diverse than transition bias [26]), B) Lack of subsequent mutations due to methodological limitations resulting in mutating often only one nucleotide in a codon, C) Organization of the genetic code which requires flexibility in bias generation for instance to avoid stop codons, and D) Uneven distribution of mutations over the gene of interest and tunable mutation frequency. Detailed analyses concerning main limitations in high quality mutant library generation are reported in [9] and [26].

Method	Uniform distribution of mutations	Unbiased mutational spectrum	Consensus mutations	Controllable mutation frequency	Codon based mutagenesis	Technically simple and robust	Limitations	Reference
Focused mutagenesis methods								
ISM	NR	NR	Yes	NR	Yes	Yes	• Selection of positions and iteratively used beneficial variants are based on researcher's intuition	[17]
OSCARR	NR	NR	Yes	Yes	Yes	Yes	• Primer quality dependent • Requires PCR optimization for each mutations	[19]
OPW-PCR	NR	NR	Yes	Yes	Yes	Yes	• Overlapping primers limit entire gene coverage	[20]
ISOR	NR	NR	Yes	Yes	Yes	No	• Primer concentrations need to be determined experimentally for adjusting mutation frequency	[18]
Rapid site directed domain scanning mutagenesis	Yes	Yes	Yes	Possible	Yes	No	• Primer quality dependent and relatively expensive • Restricted to domains only	[21]

Table 1: (continued)

Method	Uniform distribution of mutations	Unbiased mutational spectrum	Consensus mutations	Controllable mutation frequency	Codon based mutagenesis	Technically simple and robust	Limitations	Reference
Random mutagenesis methods								
epPCR followed by 8-hydroxy-dGTP	Yes	No	No	No	No	Yes	• Transitions favored	[28]
mRNA mutagenesis by Qβ replicase	Yes	Slightly biased	No	No	No	Yes	• Transitions favored • Transversions are underrepresented except GC→CG	[29]
epPCR by human X and Y family polymerases	Yes	No	No	No	No	Yes	• Transitions favored • Consecutive base deletions (frameshifts)	[30]
TriNEx	Yes	Yes	Yes	No	No but achievable	No	• Limited by preferential transposon integration sites • One mutation per gene	[32]
TIM	Yes	Yes	Yes	No	No but achievable	No		[33]
SeSaM-Tv	Yes	Yes and tunable	Yes	Yes	No	Yes	• <20% subsequent mutations • One mutation per gene	[34]
Complete saturation mutagenesis	Yes	Yes	Yes	Possible	Yes	No	• Primers for each mutation required • Laborious and relatively expensive	[31]

based on experimental results: the "best hit" is used as a template for saturation mutagenesis at the site from the second-best hit and continued in subsequent cycles. Alternative combinatorial methods for multi-site saturation mutagenesis that were recently published comprise: ISOR (Incorporating Synthetic Oligonucleotides via Gene Reassembly) [18] employs mutagenic primers to partially saturate multiple target positions in a protein sequence. OSCARR (One-pot Simple methodology for Cassette Randomization and Recombination) [19] is a technically simple one pot PCR for introducing multiple mutations is an OPW-PCR (overlap-primer-walk polymerase chain reaction; [20]) in which mutagenic primers are supplemented batch-wise to the same PCR (each 5 cycles) resulting in stepwise synthesis of the full length gene. OSCARR allows advantageously randomizing DNA fragments with high GC content which are barely mutated by conventional error-prone PCR due to polymerase biases. OPW-PCR is likely limited by the gene lengths and the quality of employed primers. Rapid site-directed domain scanning mutagenesis (SDDSM) is a QuikChange derived method to systematically generate a large number of in-frame insertion and deletion at chosen site throughout entire gene [21]. On the expense of high primer costs one can achieve compared to transposon derived methods a random insertion in the gene with adjustable linker sizes (e.g. protein purification tags).

2.4.2 Mutagenesis methods for random mutant library generation

Many structure function-relationships in proteins are not understood by computational means (pH-, organic solvent-, ionic liquid resistance, salt and detergent effects, substrate & product inhibitions) or are not traceable due to a lack of structural information. Directed evolution algorithms offer the possibility to improve biocatalysts for all properties that can be reflected in a medium throughput screening system and allow the discovery of novel structure-function relationships [22]. Current random mutagenesis methods are mainly PCR-based [23, 24] and suffer from three main challenges ([25]; Table 1): i. Mutational bias of employed polymerases, ii. Lack of subsequent mutations due to low mutation frequencies, and iii. Organization of the genetic code [26] in which amino acids are encoded by up to six codons whereas Trp and Met are only encoded by one codon. On an automated server MAP (Mutagenesis Assistant Program: http://map.jacobs-university.de/MAP.html) offers the possibility to analyze the consequences of these three limitations of 19 mutagenesis methods on protein substitution patterns [14]. As a result an average of less than 40% of all possible amino acid substitutions can be generated with standard epPCR conditions. Based on these MAP-analyses a benchmarking system has been developed which judges random mutagenesis methods not on DNA sequences but on the generated amino acid substitution patterns [14, 27].

Mutational bias of employed polymerases

Reported and reviewed methods such as Random Insertion and Deletion (RID) mutagenesis, Random Insertional-deletional Strand Exchange (RAISE) mutagenesis and Sequence Saturation Mutagenesis (SeSaM)

avoid mutagenic “hot spots” [9]. Recently reported methods that compensate or bypass mutational biases comprise: a) epPCR with unbalanced dNTP and Mn2+ concentrations followed by PCR amplification in presence of base analogue 8-hydroxy-dGTP [28], b) Mutagenesis of mRNA by error prone RNA dependent RNA polymerase ‘Qβ replicase’ [29] coupled with ribosome display technology to carry out mutagenesis and screening in one reaction, and c) PCR amplification employing error prone human X and Y family polymerases which show complementing biases on an expense of increased insertion and deletion frequency [30]. All these methods either do not introduce subsequent mutations or are limited in terms of modulating mutational spectra.

Subsequent mutations in a codon

A prerequisite for methods generating high quality mutant libraries is their ability to introduce subsequent mutations: By simultaneous exchanging two subsequent nucleotides of a codon 316 of 380 possible amino acid exchanges are accessible (all 20 amino acids replaced by the 19 other possible amino acids; 20 x 19 = 380) representing >83% of the natural diversity [26], whereas a single nucleotide mutation in a codon can only result in 150 amino acid substitutions (39.5% coverage).

As a proof of concept, comprehensive randomization of a target gene was achieved using a complete saturation mutagenesis protocol in which each position of a cutinase gene (214 amino acids) was randomized using the degenerate NNS codon (N: A, T, G or C and S: G or C). For improving activity and stability towards an anionic detergent 37,664 clones were screened and resulted in 24 sites with increased stability & activity; most of the mutated residues were caused by subsequent mutations biased towards transversions [31]. Recent advances in random mutagenesis methods generating subsequent mutations include the TriNEx (Trinucleotide exchange) method [32], the Transposon Integration mediated Mutagenesis (TIM; [33]) and the transversion enriched Sequence Saturation Mutagenesis-method (SeSaM-Tv; [34]). TriNEx [32] employs MuDel transposons to integrate trinucleotide (NNN) randomly with a frequency of one or two positions in the selected genes. As a proof of concept the effectiveness of TriNEx was demonstrated by a model protein TEM-1 ß-lactamase resulting in 64-fold increase in the activity towards ceftazidime which is much higher compare to a previously reported increase (~4 fold improvement; [35]). A conceptually related random mutagenesis method is the Transposon Integration mediated Mutagenesis (TIM) [33]. Transposon based methods show often sequence dependent integration preferences resulting in unequal distribution of mutations within the targeted sequence. SeSaM-Tv comprises as the SeSaM method four steps ([30]; Step 1: generating a pool of DNA fragments with random length; Step 2: adding universal/degenerate base(s) at the 3’-termini of the DNA fragment pool; Step 3: elongating the DNA fragment pool to full-length genes in a PCR and Step 4: replacing universal/degenerate bases by standard nucleotides). SeSaM-Tv has been improved compared to SeSaM by bypassing the requirement of ssDNA template and by employing Vent (exo-) and Deep Vent (exo-)

DNA polymerases which significantly increased PCR yields compared to previously used Taq DNA polymerase. The variation of DNA polymerases and buffers in Step 3 resulted in consecutive nucleotide exchanges by 105–106 fold compare to conventional epPCR [34]. The use of dNTPαS and three degenerate bases (P, K and I) in SeSaM-Tv allows to control and to adjust mutational biases in a flexible manner [34]. Very recently, DNA polymerases extending up to four consecutive mismatches at oligonucleotide 3´-ends were evolved by directed evolution and recombination employing an innovative compartmentalized self-replication screening system in emulsions [36]. These polymerases seem to be well-suited to solve the challenge of elongating subsequent mismatches in methods like SeSaM.

Organization of the genetic code

For protein engineers, the genetic code sets the upper limit of natural diversity that is accessible with an ideally designed random mutagenesis method. The number of codons encoding one amino acid varies however from six (Arg, Leu and Ser), four (Ala, Gly, Pro, Thr and Val), three (Ile), two (Asn, Asp, Cys, Gln, Glu, His, Lys, Phe and Tyr) to one (Met and Trp). The unequal distribution of codons and their biased occurrence in gene sequences limit significantly the diversity of generated protein sequence libraries [26]. The genetic code is interestingly optimized that transitions lead, in contrast to transversions, often to chemically similar or identical amino acids ([27]; fraction of preserved aa: 34.5% ideal transition bias; 14.2% ideal transversion bias). These consequences of the biased genetic code usage can be addressed by gene synthesis methods [37], transposon-based methods (TriNEx [32] and TIM [33]) and in site saturation mutagenesis experiments by partially degenerated codons (NDT or NWT; N: A, T, G or C /, D: G, A or T / W: A or T).

2.4.3 Methods and programs, for identifying "key beneficial" mutations in recombination experiments with multiple parents

A main challenge after various rounds of directed evolution is the identification of mutations contributing to the beneficial effect(s) in a pool of other mutations in the successful muteins. Various recombination strategies have been published a while ago to recombine and to identify such very important mutations (VIMs; e.g. in StEP-experiment with supplemented wild-type [38, 39]. Recently computational programs such as Protein Sequence Activity Relationship (ProSAR) have been developed to identify VIM-positions (Figure 2; [10, 40, 41]). ProSAR is based on QSAR (Quantitative Structure Activity Relationship) methods which are widely used in structure based drug design [40]. With ProSAR a 4,000 fold increase in volumetric productivity of halohydrin dehydrogenase was achieved [40, 42]. By another computation strategy based on sequence-activity-fitness statistical model and in the absence of any structural knowledge a >1500 fold improvement of NAD(P)H-dependant oxidoreductase (ChrR) activity was achieved [43]. These two computational programs are significant steps to identify key beneficial mutations minimizing experimental efforts.

Table 1 summarizes the advances in mutant library generation (random and focus) and compares main performance criteria of those

methods and programs that have been developed. We hope that this compilation inspires researchers to implement methods beyond epPCR and single site directed mutagenesis in their directed evolution experiments interconnecting theoretical concepts and sequence space exploration (Figure 2).

2.5 Conclusions

Significant progress has been achieved in addressing the main challenges for designing high quality mutant libraries in the last two years. First concepts are proposed to describe the impact of mutational loads on protein traits and computational tools for generating smart libraries showed remarkable success in restricting 'astronomically' vast sequence space to "screenable" sizes. Advancement in random and focused mutagenesis techniques like TriNEx, SeSaM-Tv and ISM are promising developments in generating functionally enriched diversity. Algorithms for systematic recombination of identified 'key beneficial' residues, offer an exciting promise to maximize trait improvements by reducing combinatorial complexity of sequence space to "screenable" sizes. Development of more accurate statistical models and algorithms quantifying and predicting even coupled mutational effects on protein functions would represent a significant step forward to an efficient reengineering of enzymes. The latter requires an intensified interplay between theoreticians and experiments as indicated in Figure 2.

Acknowledgement

The part of the chapter was "Reprinted from Current Opinion in Chemical Biology, 13, Shivange AV, Marienhagen J, Mundhada H, Schenk A, Schwaneberg U. Advances in generating functional diversity for directed protein evolution, 19-25, Copyright (2009), with permission from Elsevier".

2.6 References

1. Anwar, A. and M. Saleemuddin, *Alkaline proteases: A review.* Bioresour Technol., 1998. 64(3): p. 175-83.
2. Mills, D.R., R.L. Peterson, and S. Spiegelman, *An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule.* Proc. Nat. Acad. Sci. U.S.A., 1967. 58(1): p. 217-24.
3. Matsuura, T. and T. Yomo, *In vitro evolution of proteins.* J Biosci Bioeng., 2006. 101(6): p. 449-56.
4. Jaeger, K.E. and T. Eggert, *Enantioselective biocatalysis optimized by directed evolution.* Curr Opin Biotechnol, 2004. 15(4): p. 305-13.
5. Furukawa, K., *'Super bugs' for bioremediation.* Trends Biotechnol, 2003. 21(5): p. 187-90.
6. Locher, C.P., et al., *Development of novel vaccines using DNA shuffling and screening strategies.* Curr Opin Mol Ther, 2004. 6(1): p. 34-9.
7. Miller, O.J., et al., *Directed evolution by in vitro compartmentalization.* Nat Methods, 2006. 3(7): p. 561-70.
8. Aharoni, A., A.D. Griffiths, and D.S. Tawfik, *High-throughput screens and selections of enzyme-encoding genes.* Curr Opin Chem Biol, 2005. 9(2): p. 210-6.
9. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries.* Environ Microbiol, 2007. 9(11): p. 2645-59.
10. Fox, R.J. and G.W. Huisman, *Enzyme optimization: moving from blind evolution to statistical exploration of sequence-function space.* Trends Biotechnol, 2008. 26(3): p. 132-8.
11. Bloom, J.D., F.H. Arnold, and C.O. Wilke, *Breaking proteins with mutations: threads and thresholds in evolution.* Mol Syst Biol, 2007. 3: p. 76.
12. Tokuriki, N., et al., *The stability effects of protein mutations appear to be universally distributed.* J Mol Biol, 2007. 369(5): p. 1318-32.
13. Bershtein, S., et al., *Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein.* Nature, 2006. 444(7121): p. 929-32.
14. Wong, T.S., et al., *A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution.* J. Mol. Biol., 2006. 355(4): p. 858-71.
15. Firth, A.E. and W.M. Patrick, *GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries.* Nucleic Acids Res, 2008. 36(Web Server issue): p. W281-5.
16. Reetz, M.T., D. Kahakeaw, and R. Lohmer, *Addressing the numbers problem in directed evolution.* Chembiochem, 2008. 9(11): p. 1797-804.

17. Reetz, M.T. and J.D. Carballeira, *Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes.* Nat Protoc, 2007. 2(4): p. 891-903.
18. Herman, A. and D.S. Tawfik, *Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries.* Protein Eng Des Sel, 2007. 20(5): p. 219-26.
19. Hidalgo, A., et al., *A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution.* Protein Eng Des Sel, 2008. 21(9): p. 567-76.
20. Kumar, R. and K. Rajagopal, *Single-step overlap-primer-walk polymerase chain reaction for multiple mutagenesis without overlap extension.* Anal Biochem, 2008. 377(1): p. 105-7.
21. Deng, Q., W. Luo, and M.S. Donnenberg, *Rapid site-directed domain scanning mutagenesis of enteropathogenic Escherichia coli espD.* Biol Proced Online, 2007. 9: p. 18-26.
22. Kuper, J., et al., *Understanding a mechanism of organic cosolvent inactivation in heme monooxygenase P450 BM-3.* J Am Chem Soc, 2007. 129(18): p. 5786-7.
23. Sen, S., V. Venkata Dasu, and B. Mandal, *Developments in directed evolution for improving enzyme functions.* Appl Biochem Biotechnol, 2007. 143(3): p. 212-23.
24. Bustos-Jaimes, I. and C. Montiel, *Trends and Challenges in Directed Evolution.* Curr. Chem. Biol. , 2008. 2: p. 50-59.
25. Wong, T.S., D. Zhurina, and U. Schwaneberg, *The diversity challenge in directed protein evolution.* Comb Chem High Throughput Screen, 2006. 9(4): p. 271-88.
26. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Challenges of the genetic code for exploring sequence space in directed protein evolution.* Biocatal Biotransfor, 2007. 25(2): p. 229 - 241.
27. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain.* Biotechnol J, 2007. 2(1): p. 133-42.
28. Kamiya, H., M. Ito, and H. Harashima, *Induction of various mutations during PCRs with manganese and 8-hydroxy-dGTP.* Biol Pharm Bull, 2007. 30(4): p. 842-4.
29. Kopsidas, G., et al., *RNA mutagenesis yields highly diverse mRNA libraries for in vitro protein evolution.* BMC Biotechnol, 2007. 7: p. 18.
30. Emond, S., et al., *A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries.* Protein Eng Des Sel, 2008. 21(4): p. 267-74.
31. Brissos, V., et al., *Improving activity and stability of cutinase towards the anionic detergent AOT by complete saturation mutagenesis.* Protein Eng Des Sel, 2008. 21(6): p. 387-93.
32. Baldwin, A.J., et al., *Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx).* Nucleic Acids Res, 2008. 36(13): p. e77.
33. Hoeller, B.M., et al., *Random tag insertions by Transposon Integration mediated Mutagenesis (TIM).* J Microbiol Methods, 2008. 75(2): p. 251-7.
34. Wong, T.S., et al., *Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR.* Biotechnol J, 2008. 3(1): p. 74-82.
35. Fujii, R., M. Kitaoka, and K. Hayashi, *One-step random mutagenesis by error-prone rolling circle amplification.* Nucl. Acids Res., 2004. 32(19): p. e145.
36. d'Abbadie, M., et al., *Molecular breeding of polymerases for amplification of ancient DNA.* Nat Biotechnol, 2007. 25(8): p. 939-43.
37. Xiong, A.S., et al., *Chemical gene synthesis: strategies, softwares, error corrections, and applications.* FEMS Microbiol Rev, 2008. 32(3): p. 522-40.
38. Stemmer, W.P., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. 370(6488): p. 389-91.
39. Zhao, H., et al., *Molecular evolution by staggered extension process (StEP) in vitro recombination.* Nat Biotechnol, 1998. 16(3): p. 258-61.
40. Fox, R.J., et al., *Improving catalytic function by ProSAR-driven enzyme evolution.* Nat Biotechnol, 2007. 25(3): p. 338-44.
41. Evanko, D., *Efficient evolution.* Nat Meth, 2007. 4(4): p. 300-300.
42. Pelletier, J.N. and R. Lortie, *Sequence-activity relationships guide directed evolution.* Nat Biotechnol, 2007. 25(3): p. 297-8.
43. Barak, Y., et al., *Enzyme improvement in the absence of structural knowledge: a novel statistical approach.* ISME J, 2008. 2(2): p. 171-9.

Part of this chapter is

Published in: Current Opinion in Chemical Biology, Volume 13, Issue 1, February 2009, Pages 19-25.

Commented in: Kazlauskas R, Lutz S. Engineering enzymes by 'intelligent' design. Curr Opin Chem Biol, 2009. 13(1): p 1-2.

Insight into the rational protein design for enhancing thermostability

Amol V. Shivange[a,b], Danilo Roccatano[b] and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Rational design for tailoring enzymes is most dynamic and rapidly developing area of computational biology. Rational protein design is a knowledge based approach using 3D protein structural data and computer modeling for mutation design. Due to the natural complexity of protein structures and structure-function relationships, protein engineering greatly benefits from the current computational tools which can further be improved by development of novel methods. In this chapter, the recent developments in rational protein design and different approaches in protein thermostabilization have been briefly illustrated.

3.1 Recent Developments

Over the last years, various computational programs have been reported [1] for semi-rational design and rational design to understand the structure-function relationship and assist experimentalist in selection and design of directed evolution strategies. Mutagenesis assistant program (MAP) has been developed by Wong *at al*., which aid experimentalist for designing promising directed evolution strategies by comparing 19 mutagenesis methods [2] (http://map.iu-bremen.de/MAP.html) at amino acid level. Fischer et al. have developed a novel database specialized on protein super-families, integrating the information about protein sequence and structure with essential links to the protein function [3] (http://www.cyped.uni-stuttgart.de/). A new algorithm, 'CAVER' has also been developed for the identification of routes from buried active sites to the external solvent in static protein structures. This algorithm can be applied to static as well as dynamics protein structures as to the analysis of molecular dynamics (MD) simulation trajectories [4] (http://loschmidt.chemi.muni.cz/caver/). The HotSpot Wizard server (http://loschmidt.chemi.muni.cz/hotspotwizard) was developed to combine information from extensive sequence and structure database searches with functional data to create a mutability map for a target protein [5].

Quantitative Structural Activity Relationship (QSAR) based methods were also developed for protein design. One of them is COMparative BINding Energy (COMBINE) analysis which can predict binding enthalpy

*Corresponding author

and entropy for the protein ligand complex. General applicability of this method would be to estimate thermodynamic parameters of these complexes [6]. Recently, Braiuca et al. have developed Volsurf computational method based on QSAR [7]. QSAR model was built by calculating molecular descriptors of the protein surface and their role in the stabilization of the macromolecule has been analyzed. Regression model has shown a good predictivity and identified some structural requirements which correlated with protein thermostability. The advances and highlights in opportunities and limitation in rational and semi-rational design has been recently reviewed [8].

3.2 Thermostabilization

Protein thermostability is a parameter of fundamental importance for protein commercialization or in the selection of enzymes for biocatalytic processes. Therefore, massive increase of interest in protein stabilization has been observed in last decade. Understanding basic mechanisms of protein chemistry along with biological aspects of thermophilic organism would be one step towards getting insights into structure-function relationships of thermostable proteins. Enzyme thermostability encompasses thermodynamic and kinetic stabilities. Thermodynamic stability is defined by the enzyme's free energy of stabilization (ΔG_{stab}) and by its melting temperature (T_m, the temperature at which 50% of the protein is unfolded). ΔG_{stab} is the free energy change for the reaction folded↔unfolded under physiological conditions, which can be expressed as.

$$\Delta G_{stab} = \Delta H_{stab} - T\Delta S_{stab}$$

ΔG_{stab}, directly measures the thermodynamic stability of the folded protein. ΔH_{stab} (the stabilization enthalpy) and ΔS_{stab} (the stabilization entropy) are large numbers that vary almost linearly with temperature in the temperature range of the activities of most enzymes. Kinetic stability depends on the energy barrier to unfolding (i.e., the activation energy [E_a] of unfolding). An enzyme's kinetic stability is often expressed as its half-life ($t_{1/2}$) at defined temperatures [9].

In most cases the ΔG_{stab} of a thermostable protein (thermozyme) is 5–20 kcal/mol higher than that of mesophilic proteins at 25°C and ΔG_{stab} values as small as 3.0–6.5 kcal/mol can account for thermostability increases of up to 12°C [10]. Therefore, it is speculated that factors contributing to stability or increasing free energy of stabilization would be helpful to increase the thermostability of enzymes and to get an insight into structure function relationship.

3.3 Features Contributing to Thermostabilization

Till now many investigations revealed no structural motifs, particular amino acids, and covalent modifications which can explain the ability of some proteins to efficiently function at elevated temperature [11]. Thermostability was found to be a property acquired by a protein through a combination of many small structural modifications that are achieved with

the exchange of some amino acids for others and the modulation of some of the inter/intra molecular forces [12].

3.3.1 Hydrogen Bonding

Thermostability of proteins can be correlated with the number of intramolecular (buried within protein) and intermolecular hydrogen bonds (the fractional polar surface involved in hydrogen bonding with the surrounding water) [13]. The net energy gain for formation of a buried H-bond is found to be approximately 0.6 kcal/mol (http://www.cryst.bbk.ac.uk/PPS2/projects/day/TDayDiss/HBonds.html). Thermostability of proteinase [14], D-glyceraldehyde-3-phosphate dehydrogenase [15] and α-amylase [16] was found to be due to hydrogen bonding contribution.

3.3.2 Electrostatic Interactions

Enhancing electrostatic complementarity in the protein was correlated with thermostability. Interactions are mainly of three types: charge-charge, charge-dipole and dipole-dipole. Typically charge-charge interactions (salt bridges) are involved in thermal stabilization. Various studies [17] [18] [19] have shown that there is a clear preference for salt bridges in thermostabilization.

3.3.3 Hydrophobic Interactions

A direct correlation between the total hydrophobicity of a protein and its thermostability has been investigated. It was found that protein gains 1.3(+/- 0.5) kcal/mol in stability for each '-CH_2-' group being buried in the molecule during protein folding [20]. Enzymes like adenylate kinases [21], 3-isopropylmalate dehydrogenase [22] and ribonuclease HI [23] have been studied for contribution of hydrophobic interaction in thermostabilization.

3.3.4 Disulphide bonds

Disulfide bridges are believed to stabilize proteins mostly through an entropic effect, by lowering entropy of unfolding [24] which increases Gibbs free energy of the system (ΔG_{stab}) and ultimately thermostability. Site directed mutagenesis on serine proteases like substilisin E [25] and glucoamylase [26] have shown importance to disulphide bonds in protein stabilization.

3.3.5 Metal Binding

Metals have long been known to stabilize and activate enzymes. Presence of Ca^{++} showed high cooperativity of thermal unfolding and high intrinsic stability in cellobiohydrolase [27]. Addition of 0.2 M K^+ ions to wildtype enzyme resulted in a 10°C increase in thermal denaturation temperature [28]. Binding of calcium ions to *Bacillus amyloliquefaciens* phytase resulted in a dramatic increase in thermostability by ~30°C in melting temperature [29].

Other factors like α-helix stabilization, conformational strain reduction are also interesting properties to study for protein thermostabilization. Considering the complexity of a protein structure, it is improbable that any one universal stabilization mechanism is responsible for thermostability and relatively few structural features have been linked to thermostability Therefore, one

should consider all of the factors contributing to thermostability.

3.4 Conclusions

The molecular dynamics simulations were applied to protein unfolding from last decades. MD simulations can provide clues about how enzymes flexibility and which regions to target for stabilization. The comparison of MD simulations reproducing the thermal movements at room temperature and of MD simulations inducing unfolding allows the distinction between movements related to catalysis and movements related to unfolding. Thus, it can allow the identification of regions that could sustain stabilization without affecting activity.

3.5 References

1. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries.* Environ Microbiol, 2007. 9(11): p. 2645-59.
2. Wong, T.S., et al., *A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution.* J. Mol. Biol., 2006. 355(4): p. 858-71.
3. Fischer, M., et al., *The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family.* Bioinformatics, 2007. 23(15): p. 2015-17.
4. Petrek, M., et al., *CAVER: a new tool to explore routes from protein clefts, pockets and cavities.* BMC Bioinformatics, 2006. 7: p. 316.
5. Pavelka, A., E. Chovancova, and J. Damborsky, *HotSpot Wizard: a web server for identification of hot spots in protein engineering.* Nucleic Acids Res, 2009. 37(Web Server issue): p. W376-83.
6. Wang, T. and R.C. Wade, *Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics.* J Med Chem, 2002. 45(22): p. 4828-37.
7. Braiuca, P., et al., *Volsurf computational method applied to the prediction of stability of thermostable enzymes.* Biotechnol J, 2007. 2(2): p. 214-20.
8. Lutz, S., *Beyond directed evolution-semi-rational protein engineering and design.* Curr Opin Biotechnol, 2010. *Article in press.*
9. Vieille, C. and G.J. Zeikus, *Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability.* Microbiol Mol Biol Rev, 2001. 65(1): p. 1-43.
10. Nicholls, A., K.A. Sharp, and B. Honig, *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. 11(4): p. 281-96.
11. Fields, P.A., *Review: Protein function at thermal extremes: balancing stability and flexibility.* Comp. Biochem. Physiol. A: Mol. Int. Physiol., 2001. 129: p. 417-31.
12. Scandurra, R., et al., *Protein thermostability in extremophiles.* Biochimie, 1998. 80: p. 933-41.
13. Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs.* J. Mol. Biol., 1997. 269: p. 631-43.
14. Eijsink, G.H.V., et al., *Increasing the thermostability of a neutral protease by replacing positively charged amino acids in the N-terminal turn of {alpha}-helices.* Protein Eng., 1992. 5(2): p. 165-70.
15. Tanner, J.J., R.M. Hecht, and K.L. Krause, *Determinants of Enzyme Thermostability Observed in the Molecular Structure of Thermus aquaticus D-Glyceraldehyde-3-phosphate Dehydrogenase at 2.5 Å Resolution.* Biochemistry, 1996. 35(8): p. 2597-09.
16. Suvd, D., et al., *Crystal Structure of Bacillus stearothermophilus a-Amylase: Possible Factors Determining the Thermostability.* J Biochem, 2001. 129(3): p. 461-68.
17. Hakamada, Y., et al., *Identification of thermostabilizing residues in a Bacillus alkaline cellulase by construction of chimeras from mesophilic and thermostable enzymes and site-directed mutagenesis.* FEMS Microbiol Lett, 2001. 195(1): p. 67-72.
18. Karshikoff, A. and R. Ladenstein, *Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads.* Trends Biochem Sci, 2001. 26(9): p. 550-6.
19. Alsop, E., M. Silver, and D.R. Livesay, *Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis.* Protein Eng., 2003. 16(12): p. 871-74.
20. Pace, C.N., *Contribution of the hydrophobic effect to globular protein stability.* J Mol Biol, 1992. 226(1): p. 29-35.
21. Haney, P.J., M. Stees, and J. Konisky, *Analysis of Thermal Stabilizing Interactions in Mesophilic and Thermophilic Adenylate Kinases from the Genus Methanococcus.* J. Biol. Chem., 1999. 274(40): p. 28453-58.
22. Kirino, H., et al., *Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, Thermus thermophilus.* Eur J Biochem, 1994. 220(1): p. 275-81.
23. Ishikawa, K., et al., *Stabilization of Escherichia coli ribonuclease HI by cavity-filling mutations within a hydrophobic core.* Biochemistry, 1993. 32(24): p. 6171-8.
24. Matsumura, M., G. Signor, and B.W. Matthews, *Substantial increase of protein stability by multiple disulphide bonds.* Nature, 1989. 342(6247): p. 291-3.
25. Takagi, H., et al., *Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease.* J. Biol. Chem., 1990. 265(12): p. 6874-78.
26. Li, W.T., et al., *Thermodynamic Stability of Archaeal Histones.* Biochemistry, 1998. 37(30): p. 10563-72.
27. Kataeva, I.A., V.N. Uversky, and L.G. Ljungdahl, *Calcium and domain interactions contribute to the thermostability of domains of the multimodular cellobiohydrolase, CbhA, a subunit of the Clostridium*

thermocellum cellulosome. Biochem J, 2003. 372(Pt 1): p. 151-61.
28. Radfar, R., et al., *Cation Binding and Thermostability of FTHFS Monovalent Cation Binding Sites and Thermostability of N10-Formyltetrahydrofolate Synthetase from Moorella thermoacetica.* Biochemistry, 2000. 39(47): p. 14481-86.
29. Ha, N.C., et al., *Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states.* Nat Struct Biol, 2000. 7(2): p. 147-53.

Directed evolution of a highly active *Yersinia mollaretii* phytase toward higher thermostability

Amol V. Shivange[a,b], Annegret Serwe[c], Danilo Roccatano,[b] Stefan Haefner[c], and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany
[c]BASF SE, Fine Chemicals and Biocatalysis Research, GVF-A30, D-67056 Ludwigshafen, Germany

Phytases improves as a feed supplement the nutritional quality of phytate rich diets (e.g. cereal grains, legumes, and oilseeds) by increasing abdominal absorption of inorganic phosphates, minerals, and trace elements. Directed phytase evolution was reported with few phytases improving thermostability and activity by employing 96-well microtiter plate based absorption or fluorescence assays. In this chapter, a high throughput prescreening, based on filter paper assay and 384-well microtiter plate assay was developed to screen *Y. mollaretii* phytase mutant libraries (~8400 clones). A phytase from enterobacteriaceae family was selected for directed evolution due to its potential industrial applications. A diverse mutant library was generated using SeSaM method and cloned into an 'evolution plasmid' with high transformation efficiency. Subsequently, the screening resulted in a variant with ~20% improvement in thermostability (58° for 20 min) and 1.5 °C increase in melting temperature with slight loss of specific activity. Additionally, native polyacrylamide gel revealed that *Y. mollaretii* phytase has a tetrameric structure with positive cooperativity (Hill coefficient was 2.3) and a specific activity of 1073 U/mg.

4.1 Introduction

Phytases (myo-inositol hexakis phospho-hydrolase), catalyze the release of inorganic phosphate by stepwise hydrolysis of phosphomonoester bonds in phytate (myo-inositol-1,2,3,4,5,6 hexakisphosphate). Phytate is the major deposit of phosphate accounting for 50 to 80% of total phosphorous in cereal grains, legumes, and oilseeds [1]. However, phosphorous in this form is nearly unavailable to monogastric animals such as human beings, dogs, birds, swine, and poultry due to lower or negligible phytase activity [2]. Moreover, the anti-nutritive activity of phytate due to its polyanionic chelating property reduces the bioavailability of several divalent cations of major nutritional importance like Zn^{2+}, Cu^{2+}, Ca^{2+}, Mn^{2+} Fe^{2+}, and Mg^{2+} [3]. Phytate can

*Corresponding author

additionally be complexed to proteins and amino acids [4]. Consequently, the undigested or unabsorbed phytate is excreted as large amounts of inorganic phosphate (Pi) into the environment resulting into environmental pollution and hypereutrophication of streams, rivers, and lakes [1, 5], in addition to decreased bioavailability of Pi [6]. Supplemented phytase to food has been considered of significant value in upgrading nutritional quality and reducing pollution effectively.

Among several phytases widespread in nature, members of the histidine acid phosphatase (EC: 3.1.3.8) family have been attracted for the animal feed industry because of their ability to efficiently hydrolyze phytate with high specific activity [7]. Despite wide range of applications, widely used commercial phytases face the considerable challenge of low activity and thermostability. Adequate thermostability is requisite for feed application because of pelleting temperature. Recently, several phytases from enterobacteriaceae family [8] have been reported to possess higher specific activity (20 to 40 times more) compare to fungal phytases. Therefore, a member of enterobacteriaceae family, *Yersinia mollaretii* phytase (Ymphytase) was selected for directed evolution.

Protein engineering by directed evolution has over the past decades become a versatile tool to tailor the enzyme properties [9, 10]. Although, a theory on molecular principles governing thermostabilty improvement does not exist till date, the analyses of structural basis for protein stability have illustrated general factors contributing to protein thermostability. These include increase in ionic interactions (salt bridges), hydrogen bonds, disulphide bonds, hydrophobic interactions, reduction of conformational strain, and enhanced secondary structure propensity [11, 12]. Most of the success stories [13-15] of phytase thermostabilization involved improvement of the additional hydrogen bonding interactions due to diverse amino acid substitutions (eg. lysine to glutamate or methionine [13]). A statistical analysis illustrated that a transversion biased method can generate more diverse amino acid substitution patterns with a significantly different chemical composition than a transition bias method like error-prone PCR [16]. The success of any directed evolution is primarily depends on diversity generation enriched with functional trait. Therefore, a SeSaM method was used for generating functional diversity that incorporates transversions and subsequent mutations. Large pool of mutant library was generated using pALXtreme-5b plasmid (evolution plasmid), a modified pET-22b(+) plasmid that has high transformation efficiency due to its small size [17]. Another key in success of directed evolution is a reliable high throughput screening system. The high throughput prescreening system was developed based on filter paper assay and fluorescent substrate 4-MUP (4-methylumbelliferyl phosphate). The filter paper assay [18] was improved for high throughput activity screening of Ymphytase mutant library with no cross contamination possibility and negligible chances of false positive clones. The thermostability prescreening was accomplished by developing colony based screening in 384-well microtiter-plate format using 4-MUP. The final screening of variants was carried out in 96-

well microtiter plate format using both phytate and 4-MUP assay. In this study, we have optimized each step to accelerate evolution experiment by smart mutant library generation and robust prescreening method to improve thermostability of Ymphytase. To the best of our knowledge, this is a first report on directed evolution of highly active phytase from *Yersinia* species, biochemical characterization of *Y. mollaretii* phytase, use of SeSaM method for phytase evolution and use of combined screening system based on absorption and fluorescence assay for evolution.

4.2 Material and Methods

4.2.1 Sequence similarity approach for sequence annotation

Sequence annotation for *appA* gene in *Y. mollaretii* (ATCC 43969) unfinished genome (4.52773 Mbp) was performed using the genome sequence available in NCBI database under the accession no. AALD00000000 entitled 'Yersinia mollaretii ATCC 43969, whole genome shotgun sequencing project' (NCBI Genome Project ID: 16105). Total 146 contigs (AALD01000001-AALD01000146) were deposited in the NCBI database. DNA encoding phytase from *Yersinia mollaretii* was identified using sequence similarity approach. The amino acid sequence (GenBank Accession no. ABI95370) of a well characterized phytase [19] from homologous organism *Y. intermedia* was subjected to organism specific NCBI-BLASTp analysis [20] to find similar protein sequence in *Y. mollaretii*. Identified gene locus corresponding to *Y. mollaretii* hypothetical protein was then translated and subjected to pairwise sequence alignment using WATER program embedded in EMBOSS GUI (http://bips.u-strasbg.fr/EMBOSS/).

4.2.2 Sequence analysis

The annotated gene was subjected to domain prediction programs. The Conserved Domain Architecture Retrieval Tool (CDART) [21] implemented in NCBI and Pfam database search tools (http://pfam.sanger.ac.uk) were used to find the conserved domains in the protein sequence. Further, ExPASy-PROSITE (http://www.expasy.ch/prosite/) web server was used to find the functional domain in the amino acid sequence. The signal peptide was predicted using SignalP (http://www.cbs.dtu.dk/services/SignalP/) server that predicts presence and location of signal peptide cleavage sites in amino acid sequence [22]. Multiple sequence alignment and phylogenetic tree of protein sequences were constructed using the AlignX module (ClustalW based) of VectorNTI software using blosum62mt2 matrix with gap-opening penalty of 10 and gap-extension penalty of 0.05. Profile alignment was used using Ymphytase as a reference sequence. DiANNA web server [23] (http://clavius.bc.edu/~clotelab/DiANNA) was used to predict the disulphide bonds in the sequence.

4.2.3 Cloning of YmappA into pALXtreme-5b

The core region of *YmappA* gene was amplified from *Y. mollareti* (ATCC 43969) genomic DNA (60 ng) by PCR [94 °C for 3 min, one cycle; 94 °C for 15 sec, 58 °C for 30 sec (-0.5 °C/cycle), 68 °C for 90 sec, 20 cycles; 94 °C for 15 sec, 58 °C for 30 sec, 68 °C for 90 sec, 10 cycles; 68 °C for 180 sec] using forward primer

YmPNdeWtF (5'-CCG AAT TCA CAT ATG CGA TTA ACT GCA CTG GGC-3') or YmPNdeR2HF (5'-CCG AAT TCA CAT ATG CAC TTA ACT GCA CTG GGC-3') and reverse primer YmPNotR (5'-GAG TGC GGC CGC CTA TTA AAT ATG GCA TGC AGG TTC AAT C-3') containing *EcoRI* or *NdeI* and *NotI* (underlined) restriction enzyme sites respectively (0.4 mM each), Pfu DNA polymerase (0.025 U/µl) and dNTP mix (0.2 mM each). YmPNdeR2HF primer was used to incorporate R2H mutation in the wild-type *appA* gene. Both the wild-type and mutant (R2H) PCR products were digested with *NdeI* and *NotI* restriction enzymes (New England Biolabs GmbH, Frankfurt am Main, Germany). Digested PCR product was then cloned into the *E. coli* expression vector pALXtreme-5b (modified small pET-22b(+) vector; [17]) Sequencing of cloned genes (wild-type) revealed 100% identity with annotated DNA sequence.

4.2.4 Construction of Ymphytase SeSaM library

A SeSaM mutagenesis technique [24] was used to generate a random mutagenesis library for high mutational load. Briefly, the concentration of dATPαS was selected for forward-A (36%), reverse-A (37%), and dGTPαS for forward-G (27%) and reverse-G (28%) fragmentation based on the smearing pattern obtained (Supporting Information Figure S1). Four different fragmentation libraries (A_fwd_library, A_rev_library, G_fwd_library, and G_rev_library) were generated by incorporating dRTPαS (R=A or G) into *YmappA* gene by PCR. Following fragmentation, purification, and universal base addition, the full length gene synthesis was achieved using PCR by combining A_fwd_library with A_rev_library and G_fwd_library with G_rev_library. Finally universal bases were replaced by standard nucleotide. Equal amounts of both A and G libraries were combined and digested with *NdeI* and *NotI* restriction enzymes. Digested PCR product was cloned into pALXtreme-5b vector and transformed into *E. coli* BL21-Gold(DE3)laqI^{Q1} strain [17] for expression and screening.

4.2.5 High-throughput prescreening

Filter paper assay for activity prescreening

The filter paper assay [18] was modified to improve for high throughput activity screening of Ymphytase mutant library. *E. coli* colonies were transferred onto nitrocellulose membrane that lay on Luria-Bertani (LB) agar plate (22.5 cm × 22.5 cm) supplemented with 100 µg/ml ampicillin (Figure 2). Colony picking was performed using sterile toothpick and autoclavable agar plate lid containing 441 apertures (21 × 21, separated by 10 mm; designed and constructed in Werkstatt der Biologie, RWTH, Aachen) (Supporting Information Figure S2). Colonies were grown at 37 °C for 12 hours (colony diameter ~1 mm). Nitrocellulose membrane was withdrawn and the expression of Ymphytase was induced by spreading 1 ml of 40 mM isopropyl-β-D-thiogalactopyranoside (IPTG) onto the agar plate. Plates were incubated at 37 °C for 3 hours. After 3 hour, nitrocellulose membrane was washed by placing the membrane with colonies facing upward on Whatman filter paper presoaked in water for 10 min. The washing step was repeated twice and the NC membrane was similarly placed for 5 min on another Whatman

filter paper presoaked in the substrate solution (1 mM sodium phytate in 250 mM sodium acetate, pH 5.5, 1 mM $CaCl_2$, 0.01% Tween-20) for catalysis. The filter paper was developed by gently immerging in color developing solution (0.27% w/v ammonium molybdate and 1.08% w/v ascorbic acid in 0.32 M H_2SO_4), for 2 seconds and subsequently placed onto an absorbent paper to remove the excess of liquid until blue spots appeared.

Prescreening in 384-well plate for thermostability

Colonies grown on LB_{amp} plates were transferred, using manual schuett PickMaxX clone picking tool (Schuett-biotec GmbH, Göttingen Germany), into 384-well microtiter plates (black polystyrene plates, Greiner Bio-One GmbH, Frickenhausen, Germany) containing 250 mM sodium acetate buffer (50 µl; 1 mM $CaCl_2$, 0.01% Tween-20, pH 5.5). Each well was replicated using pin replicator (constructed in Werkstatt der Biologie, RWTH, Aachen) into a second series of 384-well microtiter plates containing ZYM-505 noninducing medium [25] (50 µl) supplemented with ampicillin (100 µg/ml). First series of 384-well microtiter plates were sealed with adhesive sealing film (MicroAmp Optical Adhesive Film, Applied Biosystems Deutschland GmbH, Darmstadt, Germany) and incubated into a modified metal block thermostat MBT plate heater (Entwicklungs- und Technologie Gesellschaft mbH, Ilmenau, Germany) at 60 °C for 20 min. Aluminium metal block in the plate heater was designed and constructed (Werkstatt der Biologie, RWTH, Aachen) for efficient heat transfer to each well of the plate (see Figure 2). Subsequently, plates were cooled down by keeping plates in 4 °C room on aluminium metal block. Since our primary screening was for the retention of structural integrity, the model substrate 4-methyl-umbelliferylphosphate (4-MUP) was used in the high-throughput screening. 4-MUP (50 µl; 5 mM) was then added to each well of 384-well plate using liquidator[96] (Steinbrenner Laborsysteme GmbH, Wiesenbach, Germany). The phytase activity was monitored for 15 minutes by the increase in the relative fluorescence (excitation wavelength, 360 nm; emission wavelength, 465 nm) using a Tecan Infinite M1000 microtiter plate reader (Tecan Group Ltd. Männedorf, Switzerland) due to the hydrolysis of 4-MUP to 4-methylumbelliferone. Three wild-type or empty vector controls were included in each plate. Relative activity for each clone (mutant) was defined by considering the average fluorescence (a time point in the linear range) or slope (increased fluorescence over time) of three different wt clones as 100% activity. The clones in the second set of plates were cultivated in a microtiter plate shaker (Multitron II, Infors GmbH, Einsbach, Germany; 37 °C, 900 rpm, 70% relative humidity) for ~24 hours and stored at -80 °C after addition of glycerol (50 µl; 50%).

4.2.6 Protein expression in 96-well plates

Selected clones from 384-well plates were transferred, using toothpick, into 96-well microtiter plates (V-bottomed, polystyrene plates; Corning GmbH, Kaiserslautern, Germany), containing ZYM-505 noninducing media (100 µl) supplemented with amplicillin (100 µg/ml). After 14 hours of cultivation in a microtiter plate shaker (Multitron II, Infors GmbH, Einsbach, Germany; 37 °C, 900 rpm,

70% relative humidity), each well was replicated into a second set of 96-well microtiter plates containing ZYM-5052 autoinduction medium [25] (150 µl) supplemented with ampicillin (100 µg/ml). The first set of 96-well plates were stored in -80 °C after addition of glycerol (100 µl; 50%) and second set of daughter plates were cultivated for 10 hours (Multitron II, Infors GmbH, Einsbach, Germany; 37 °C, 900 rpm, 70% relative humidity) for expression and subsequent screening.

4.2.7 Screening for improved thermostability

Sample Preparation

The harvested cell pellets were washed with ice-cold 10% glycerol to remove media (inorganic phosphate). Cell lysate was prepared for activity assay in 96-well plates by imbibing each cell pellet in 100 µL of Tris-HCl buffer (50 mM, 5 min, pH 7.4) and resuspended by shaking (1200 rpm for 5 min) using microplate shaker TiMix 2 (Edmund Bühler GmbH, Hechingen, Germany). Cell lysis was performed by adding 100 µl of lysozyme solution (2 mg/ml, 50 mM Tris, pH 7.4) followed by incubation (37°C, 1 h) and centrifugation (4°C, 3220 g, 20 min).

96-well plate-format phytase colorimetric screening assay

The diluted cell lysate (10 µl; 1:100 dilution in 250 mM acetate buffer, 0.01% Tween-20, pH 5.5) was transferred to 96-well plate for room temperature activity against natural substrate (phytate). The enzyme activity was initiated by adding phytate solution (1 mM; 70 µl) to each well and mixture was incubated (37 °C for 15 min). The reaction was stopped by addition of trichloroacetic acid (70 µl, 15%) and color development was performed by addition of color developing solution (70 µl, 0.54% w/v ammonium molybdate and 2.16% w/v ascorbic acid in 3.5% H_2SO_4) and incubation (50°C for 15 min) in a modified metal block thermostat MBT plate heater (Entwicklungs- und Technologie Gesellschaft mbH, Ilmenau, Germany). Absorbance was measured at 820 nm in a Tecan Infinite M1000 microtiter plate reader (Tecan Group Ltd., Männedorf, Switzerland).

96-well plate-format phytase fluorescent screening assay

The cell lysate (60 µl) was diluted in acetate buffer (140 µl; 250 mM sodium acetate, 1 mM $CaCl_2$, 0.01% Tween-20, pH 5.5) and diluted cell lysate (80 µl) was transferred to 96-well PCR plates (Thermowell 96-well polycarbonate PCR plates; Corning GmbH, Kaiserslautern, Germany) and PCR plates were sealed with sealing mat (Thermowell 96-well silicone rubber sealing mat, Corning GmbH, Kaiserslautern, Germany). PCR plates were incubated at 59 °C for 20 min in PCR machine (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) and cooled down to 4 °C for 20 min. Activity for room temperature sample (50 µl; diluted cell lysate) or heat inactivated samples (50 µl) was measured by adding 4-MUP (50 µl; 1 mM, 250 mM sodium acetate, 1 mM $CaCl_2$, pH 5.5, 0.01% Tween-20) solution. The phytase activity was determined (15 min) by the increase in the relative fluorescence (excitation wavelength, 360 nm; emission wavelength, 465 nm) using a Tecan Infinite M1000 microtiter plate reader (Tecan Group Ltd. Männedorf, Switzerland).

Standard deviation for both the assay was performed by transforming wild-type plasmid in

E. coli BL21-Gold(DE3)laqI^{Q1} strain and colonies from LB_{amp} agar plates were transferred to 96-well microtiter plate and screening was performed as described above. Apparent standard deviation was based on the absolute absorbance or relative fluorescence values obtained from the Ymphytase wild-type plate. The true standard deviation was calculated by subtracting the background absorbance value of the microtiter plate lacking Ymphytase from the apparent values.

4.2.8 Expression of Ymphytase in a shaking flask and purification

The purification was carried out at room temperature using ÄKTA system (ÄKTAprime plus, GE Healthcare Europe GmbH, Munich, Germany) and preparative LC column (YMC Europe GmbH, Dinslaken, Germany). Shaking flasks (2 L) containing autoinduction medium ZYM-5052 (400 mL) supplemented with ampicillin (100 µg/mL) were inoculated with a 1:100 dilution of pre-culture that was grown in ZYM-505 media for 8 hours at 37°C (*E. coli* BL21-Gold(DE3)LaqI^{Q1} harboring pALXtreme-5b-*YmappA*). After 16 h of expression, *E. coli* cells were harvested by centrifugation (4000 rmp, 15 min, 4 °C; Sorvall RC-6 Plus, Thermo Scientific) and resuspended in acetate buffer (20 mL, 25 mM sodium acetate, 0.1 mM $CaCl_2$, pH 5.5, 0.001% Tween-20). The resuspended culture was homogenized using probe sonicator (Bandelin Sonopuls HD 200, Bandelin electronic GmbH, Berlin, Germany) for 90 sec under ice cooling. The cell lysis of *E. coli* cells was subsequently carried out using a high-pressure homogenizer (~1300 bar, two cycles; EmulsiFlex C3, Avestin Europe GmbH, Mannheim, Germany). The disrupted cells were centrifuged (18350g, 45 min, 4 °C, Eppendorf 5810R), and the supernatant was further cleared by filtration through a low-protein-binding filter (0.45 µm; Minisart RC 25 disposable syringe filter; Sartorius, Hamburg, Germany). Wild-type Ymphytase and mutant M1 were subsequently purified by column chromatography. The filtered cell lysates were subjected to a Toyoperl SP-650C (Tosoh Bioscience GmbH, Stuttgart, Germany) cation-exchange column that was pre-equilibrated with acetate buffer (25 mM sodium acetate, pH 5.5, 0.1 mM $CaCl_2$, 0.001% Tween-20). Cell lysate (10 mL) was injected in the ÄKTA system and Ymphytases were eluted with NaCl gradient (0 to 60% for 100 ml, flowrate 3.5 ml/min) in acetate buffer (25 mM sodium acetate, pH 5.5, 0.1 mM $CaCl_2$, 0.001% Tween-20). The Ymphytase protein obtained by ion-exchange chromatography was desalted by dialysis using Spectra/Por dialysis membrane (Spectrum Lab Inc., CA, USA) against acetate buffer (2x, 1000 mL, 250 mM sodium acetate, pH 5.5, 1 mM $CaCl_2$, 0.01% Tween-20) and stored at 4°C. This was designated as the purified phytase preparation and used for further studies.

4.2.9 Characterization of wild-type Ymphytase and mutant

Thermostabilty assay

Purified phytases were diluted in acetate buffer (250 mM sodium acetate, pH 5.5, 1 mM $CaCl_2$, 0.01% Tween-20) to 40 ng protein per milliliter. The diluted enzymes were incubated at gradient temperature (47 °C to 67 °C) in PCR machine (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) for 20 min. The protein

samples were cooled down to 4 °C for 20 minutes. The residual phytase activity was measured using the phytase colorimetric assay at 37 °C and pH 4.5 as described in screening section.

Thermal shift assay

The melting temperature (Tm) of wild-type and mutant phytase was determined using real time PCR machine (iQ5 real time PCR detection system, Bio-Rad Laboratories, CA, USA). Both the phytase were diluted to 0.2 mg/ml concentration in acetate buffer (250 mM sodium acetate, pH 5.5, 1 mM $CaCl_2$, 0.01% Tween-20). Protein sample (20 µl) was mixed with 5 µl of 10x SYPRO Orange Protein Gel Stain solution (Sigma-Aldrich, St. Louis, Missouri) diluted in acetate buffer (250 mM sodium acetate, pH 5.5, 1 mM $CaCl_2$, 0.01% Tween-20). The proteins were scanned from 20 °C to 100 °C at a heating rate of 3 °C/min and background subtracted fluorescence was measured every 10 sec.

pH profile

The pH profile of wild-type and mutant phytase was determined at 37 °C using four different buffers: 0.25 M Glycine-HCL buffer for pH 2.0 to 3.2, 0.25 M sodium acetate buffer for pH 3.6 to 5.6, 0.25 M imidazole-HCL buffer for pH 6.0 to 7.0 and 0.25 M Tris-HCL for pH 7.4 to 9.0. Purified enzymes were diluted in specified buffers to 40 ng protein per milliliter and assay was carried out with 1mM phytate with corresponding pH.

Determination of kinetic parameters

The purified phytases were diluted to 0.2 µg/ml in acetate buffer (250 mM sodium acetate, pH 4.5, 1 mM $CaCl_2$, 0.01% Tween-20). The K_{half} and K_{cat} values were determined from initial-velocity data measured as a function of substrate (phytate) concentration. Enzyme reaction was carried out at 37°C in 96 well MTP (Flat bottom, polystyrene, Greiner Bio-One GmbH, Frickenhausen, Germany) by placing the plate on metal block thermostat MBT 250-4 heater (Entwicklungs- und Technologie Gesellschaft mbH, Ilmenau, Germany) with external temperature control using PT100 sensor. After 5 min of enzyme (50 µl) preincubation at 37 °C, the reaction was initiated by the addition of substrate solution (70 µl, 7.8 – 1000 µM of phytate, 250 mM sodium acetate, pH 4.5, 1 mM $CaCl_2$, 0.01% Tween-20) using liquidator[96] (Steinbrenner Laborsysteme GmbH, Wiesenbach, Germany). The enzymatic reaction was stopped by the addition of trichloroacetic acid (70 µl, 15%) to each row of MTP every 30 seconds. The color development was subsequently performed by addition of color developing solution (70 µl, 0.54% w/v ammonium molybdate and 2.16% w/v ascorbic acid in 3.5% H_2SO_4) and incubation (50°C for 15 min) in a modified metal block thermostat MBT 250-4 heater (Entwicklungs- und Technologie Gesellschaft mbH, Ilmenau, Germany). Absorbance was measured at 820 nm in a Tecan Infinite M1000 microtiter plate reader (Tecan Group Ltd. Männedorf, Switzerland). The initial velocity data obtained were fitted to equation

$$Y = V\max * X\wedge h / (K_{half}\wedge h + X\wedge h)$$

where Y is the initial activity, Vmax is the maximum velocity, X is the substrate concentration and K_{half} the concentration of substrate that will afford half maximal velocity

(see reference [26]) by using GraphPad Prism software (GraphPad software, San Diego, CA, USA). The K_{cat} was calculated from the ratio of Vmax and enzyme concentration. One enzyme unit (U) is defined as the amount of enzyme required to liberate 1 µmol of inorganic phosphate in 1 min.

4.2.10 Molecular Modeling

As a first step towards homology modeling, the template structure for modeling was identified from the Protein Data Bank (PDB) [27] using NCBI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). Structure with highest percentage identity and highest resolution was selected as a template structure for modeling (PDBID: 1DKM). Ymphytase structural model was constructed using automated protein modeling workspace in SWISS-MODEL server [28] (http://swissmodel.expasy.org). Model obtained from SWISS-MODEL server was inspected for accuracy and validity using 3D-profile programs (http://nihserver.mbi.ucla.edu/Verify_3D/) [29] that calculate a 3D to 1D compatibility score, and the graphical representation depicts the properly folded and misfolded regions in the protein structure by performing an Eisenberg analysis of the model. Stereochemistry of the model was further gauged using a Ramachandran plot [30] using PROCHECK server implemented in SWISS-MODEL server. The molecular graphics image in figure 5 was produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [31].

4.3 Results

4.3.1 Sequence annotation and analysis

The organism specific BLASTp of *Y. intermedia* phytase (Genbank Accession no. ABI95370) against *Y. mollaretii* protein database resulted in best hit sequence with 86% identity (hypothetical protein YmolA_01003233; Genbank Accession no. ZP_00824387). Consequently, to get the nucleotide sequence, genomic sequence was analyzed and the ORF for phytase hypothetical protein was found to be present on sequence contig NZ_AALD01000061 between 23849 – 25153 bp and the gene was encoded by the complementary strand. The obtained coding sequence was translated to amino acid sequence and subjected to local pairwise alignment (EMBOSS) that showed 86.2% identity and 92.9% similarity with *Y. intermedia* phytase and 100% identity with *Y. mollaretii* hypothetical protein. Sequence alignment of annotated phytase with phytases from fungi (*Aspergillus niger* and *A. fumigatus*), yeast (*Kodamaea ohmeri* BG3) and enterobacteriaceae family (*E. coli, Yesinia pestis, Y. rohdei, Y. intermedia* and *K. pneumoniae)* showed a presence of conserved active site “RHGXRXP” (Figure 1a) and “HD” sequence motif (Supporting Information, Figure S3).

Conserved domains analyzed using NCBI-CDART revealed that the annotated protein has phosphoanhydride phosphorylase domain (NCBI Protein Cluster ID PRK10172), which is a

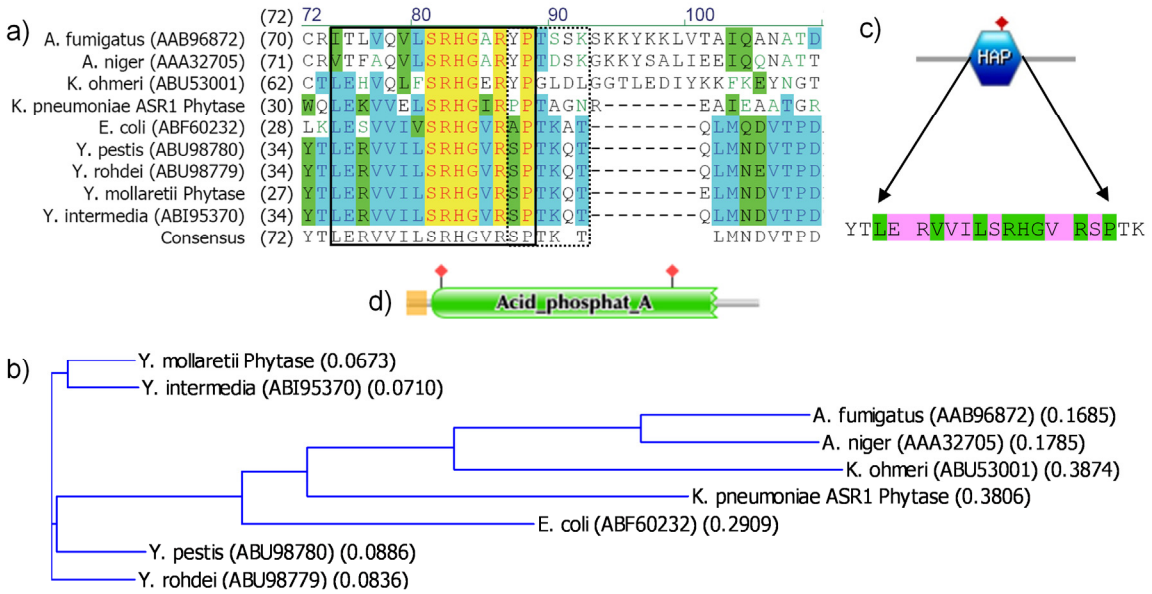


Figure 1: Sequence and conserved domain analysis of Ymphytase. Multiple sequence alignment (a) and phylogenetic tree (b) showing the evolutionary relatedness and level of homology between phytases from fungi (*A. niger* and *A. fumigatus*), yeast (*Kodamaea ohmeri* BG3) and enterobacteriaceae family (*E. coli, Yesinia pestis, Y. rohdei, Y. intermedia, Y. mollaretii* and *K. pneumoniae*). The histidine acid phosphatase (HAP) domain (c/d) is depicted in the box and active site loop residues are shown in dashed box (a). All sequences were retrieved from GenBank database. Alphanumeric number in bracket indicates accession number of the database. Identical, conserved and similar residues are colored in yellow, cyan and green color respectively.

member of the HP (histidine phosphatase) superfamily. Additionally, Pfam database search revealed a presence of Acid_phosphat_A domain (Figure 1c) and ExPASy PROSITE scan showed a presence of histidine acid phosphatase (HAP) domain (PROSITE Accession no. PS00616). The consensus pattern of the conserved domain in this family is

"[LIVM] - x(2) - [LIVMA] - x(2) - [LIVM] - x - R - H - [GN] - x - R - x - [PAS]"

(Square bracket: any one of the mentioned residues is conserved; x: any linker residue). The conserved domain in the *Y. mollaretii* annotated sequence was found to be between Leu 29 to Pro 43 represented by "LerVviLsRHGvRsP" sequence pattern, which showed the consensus residues in the domain are conserved in the annotated sequence (designated by capital letters) (Figure 1a and 1b). Despite of a very low sequence similarity (Figure 1d), fungal and bacterial phytases have a similar structural topology, the residues belonging to active site loop (Figure 1a dashed box; [32]) were highly conserved in bacterial phytases possessing high catalytic efficiency [32]. DiANNA analysis predicted the presence of 4 disulphide bonds in *Y. mollaretii* phytase (C97-C128, C153-C415, C198-C208 and C406-C432) as observed in *E. coli* phytase [33]. These results suggest that the annotated sequence is a *Y. mollaretii* phytase (Ymphytase) possesses

high structural similarity to other phytases from HAP family. Ymphytase has found to have a signal peptide and a cleavage site was found to be in between A17 – A18.

4.3.2 Quality of mutant library

The screening of ~6800 mutants from epPCR library (2 – 3 mutations/kb; transition biased method) did not succeed improving Ymphytase thermostability with 96-well microtiter plate absorbance assay (data not shown). epPCR method [Taq (+, G>>A=C=T)] was selected based on MAP analysis [34] that showed to incorporate more charged residues compare to other epPCR methods. A SeSaM protocol was used to introduce more mutations (5 – 6 mutations/kb). The quality of SeSaM mutant library was verified by sequencing 6 random clones (Supporting Information Table S1). Sequencing revealed 50% transitions (Ts), 41% transversions (Tv), 6.82% transeversion followed by transitions (TvTs subsequent mutations) and 2.27% deletion. The quality of SeSaM library was further validated by the distribution of mutations found in the six sequenced clones that ensured a uniform distribution of mutations throughout the *YmappA* gene (Supporting Information Figure S4).

4.3.3 High throughput prescreening of Ymphytase SeSaM library

Two different prescreening methods were employed for activity and thermostability screening of Ymphytase mutant library (Figure 2). The SeSaM library was prescreened for activity and thermostability using filter paper assay and 384-well colony based assay. A filter paper assay was modified to reduce cross contamination and chances of false positive clones that can be used for activity prescreening. In case of thermostability prescreening, a robust and simple colony based method was used by detecting 4-methylumbelliferone released after phytase catalysis. Approximately 1600 clones were screened for activity using filter paper assay and active clones (149) were transferred to 96-well microtiter-plates for further screening. About 800 clones were screened for improved thermostability using 384-well microtiter-plate based assay. Clones retaining at least 20% of the relative activity (wildtype average 100%) from 384-well plates (91 clones) were selected for further screening in 96-well microtiter plates.

4.3.4 Improved Ymphytase screening system

Although various assays for inorganic phosphate detection have been described [35], the narrow linear range of detection in microtiter plate format has been a limitation in high throughput screening of phytase mutant library. The key performance parameters in identifying improved variant through microtiter-plate-based screening system are linear detection range and true standard deviation of the assay. The ammonium molibdate assay was employed and optimized to broaden the linear detection range and reduce standard deviation. The microtiter-plate-based screening assay was modified for simplified handling by combining enzyme reaction and color development in the same assay plate and optimized concentrations of the ingredients. The linear detection range was broadened to 50 nmoles (optical density 3.0 at 820 nm) (Figure 3a) that was sufficient to identify the mutant with 2 – 3 fold improvement in activity. Color development stop solution and

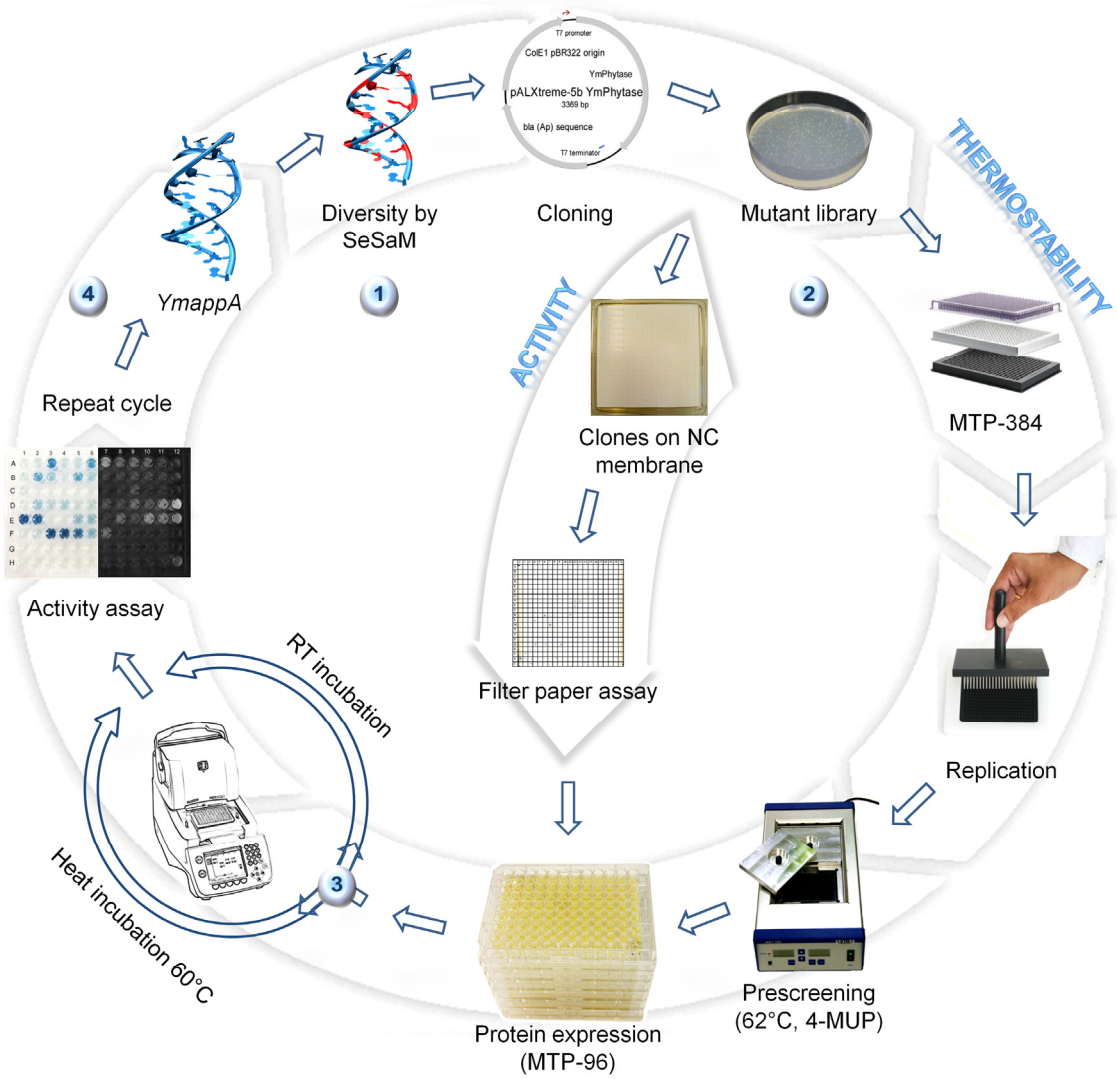


Figure 2: The directed evolution platform for phytase screening. Mutant library (colonies) generated from SeSaM (1) are prescreened (2) using either filter paper assay for activity or 384-well microtiter plate assay for thermostability. Hits from prescreening are expressed in 96-well microtiter plate and screened for activity and thermostability (3) using phytate and 4-MUP activity assay. The best mutant (4) obtained from first round of mutagenesis can be used for iterative cycles.

toxic arsenite [36] was omitted without affecting linear range at 820 nm. The standard deviation of the screening system was reduced by optimizing expression and screening conditions such as several expression medias (minimal and complex autoinduction medias; LB-IPTG medias, expression time and assay conditions such as color development in modified microtiter-plate heater, use of liquidator). The preferred combination resulted in a true standard deviation of 6.3%, 14.4% and 10.8% for cell growth; room temperature screening and heat inactivated screening respectively (Figure 3b). In case of 4-MUP assay, the standard deviations were reduced to 8.72% and 11.45% for the slope (production of fluorescent product

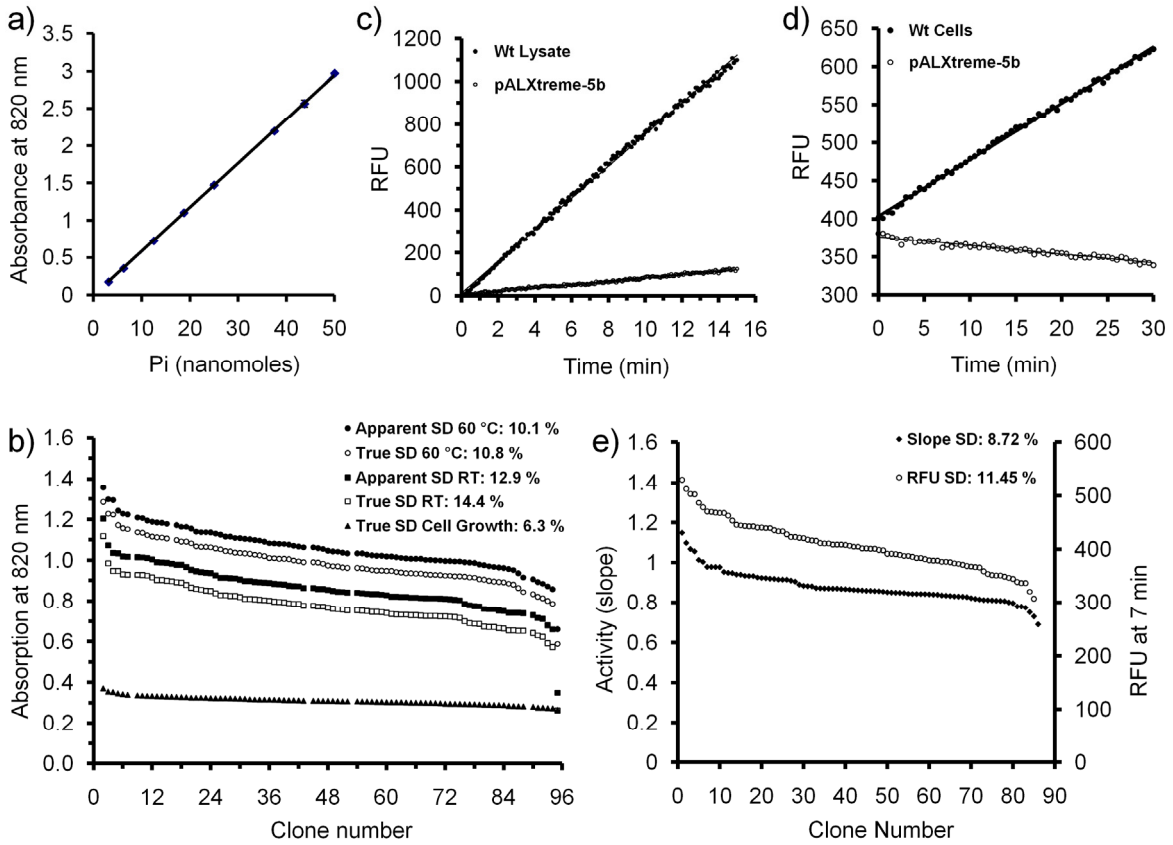


Figure 3: Calibration curve and linear detection range of 96-well microtiter plate format phytate (a) and 4-MUP (c) screening assay. Linear detection range of colony based 384-well microtiter plate prescreening assay (d). Activity values in descending order of the Ymphytase catalyzed conversion of phytic acid (b) and 4-MUP (e) in a 96-well microtiter plate using optimized screening system.

over time) and relative fluorescent unit (at single time point) respectively. The linear range for fluorescent product detection was achieved for 15 minutes with either colony or cell lysate screening (Figure 3c, 3d and 3e). Subsequently, thermostability of selected 240 clones was assayed in 96-well format screening system using both phytate and 4-MUP assay. Data was analyzed by considering the ratio of the activity at room temperature versus heat inactivated samples. Total 3 out of 16 identified variants with improved ratio were identical in both the assay. All 16 clones were then expressed in test tube and thermostability (T_{50}) was measured using PCR gradient. The best mutant (M1) was found to be a member of 3 clones that were identical in both the assays. The sequence analysis of M1 revealed presence of 7 mutations and 5 amino acid substitutions (Table 1).

4.3.5 Characterization of wild-type and mutant Ymphytase

Purification of Ymphytases

The heterologously expressed Ymphytases were purified to apparent homogeneity by a simple one step ion exchange chromatography.

Table 1: Mutational analysis (SeSaM Vs epPCR) of YmPhytase mutant M1

Substitution	Mutations	Ts/Tv	Substitution probability* with epPCR Taq (+, G>>A=C=T)
D52N	**G**AT --> **A**AT	Ts	Possible
T77K	A**C**A --> A**A**A	Tv	Not possible
K139E	**A**A**A** --> **G**A**G**	Ts	Possible
G187S	**G**GC --> **A**GC	Ts	Possible
V298M	**G**TG --> **A**TG	Ts	Not possible
R41R (Silent)	CG**T** --> CG**C**	Ts	Possible
R2H	C**GA** --> C**AC**	NA	NA

* Substitution probability was predicted using MAP server
Ts – transition, Tv – transversion, NA – not applicable

Table 2: Kinetic parameters and thermostbility of YmPhytase

Enzyme	Specific activity (U mg^{-1})	K_{half} (µM)	*K_{cat} (S^{-1})	Hills coefficient (h)	T_{50} (°C)	T_m (°C)
Wt	1073 (± 26)	292.75 (± 2.25)	846.74 (± 20.52)	2.3 (± 0.02)	56.5	63.0
M1	993.25 (± 3)	262.25 (± 0.65)	783.81 (± 2.33)	2.3 (± 0.002)	58.0	64.5

Phytase kinetics was determined in 0.25 M sodium acetate containing different sodium phytate (7.8 to 750 µM) at 37°C and pH 4.5 for both the phytases
Values in the bracket represent standard error.
*Catalytic center activity: enzyme concentration (moles) multiplied by the number of active centers (four).

SDS-PAGE of the purified enzymes showed a single band of ~47 kDa that is analogous to the predicted molecular weight based on amino acid sequence. However, a native PAGE showed a single band ~188 kDa for both the enzymes which reflects that the Ymphytase is a tetramer enzyme (Figure 4d).

Determination of kinetic parameters

The pH profile of the mutant M1 was remained unchanged compared to wild-type. Both the enzymes showed pH optima at pH 4.5. The phytase kinetics was determined at pH 4.5, using sodium phytate as a substrate that exhibited sigmoidal kinetics, suggesting that the product formation is driven by an allosteric interaction between the phytase monomers (Supporting Information Figure S5). The data was analyzed by the general allosteric model using the Hill equation. The Hill coefficient (h) value for wild-type and M1 enzyme was 2.3, which indicates positive cooperativity (Table 2). The K_{half} for wild-type and M1 were 292.75 µM and 262.25 µM respectively. The catalytic center activity (K_{cat}) was calculated by considering four active site centers per molecule of Ymphytase. The specific activity and K_{cat} of M1 was not significantly altered (7.4% less than wild-type) against phytate as a substrate, despite the use of non-natural substrate (selection pressure) in prescreening.

Thermostability of Ymphytase mutant

The mutant M1 showed improved heat stability compared to wild-type, as the incubation temperature was increased. The M1 displayed a residual activity with ~20% and ~11% improvement over that of wild-type enzyme after being heated at 58 °C and 55 °C for 20 min respectively. The thermostabilty determined by T_{50} was found to be improved by 1.5 °C compare to wild-type (PCR machine temperature homogeneity equals ≤ ±0.3 °C). The melting temperature (T_m) was determined by thermal shift assay that showed the mid-point of the thermal unfolding was increased by 1.5 °C (Figure 4).

4.4 Discussion

Despite a wide range of phytases from various organisms have been reported till date, a very few phytatses have been evolved using either phytate assay [13, 14] or 4-MUP assay [37]. In this study, we have reported a robust and high throughput screening system for directed evolution of one of the highly active phytase. A phytase from *Y. mollaretii* was annotated using sequence similarity approach to find out the coding sequence for *YmappA.* The amino acid sequence showed a presence of acid phosphatase domain that was further confirmed by activity assay. The mutation R2H was incorporated into the wild-type gene to improve *in-vivo* stability of phytase based on N-end rule [38] and used as a template for mutant library generation. We have previously proposed the conservation of the active site loop might be prerequisite for higher catalytic efficiency [32]. The loop was highly conserved in Ymphytase as seen in other highly active phytases from enterobacteriaceae family (Figure 1). Each step in the directed evolution experiment was optimized in a way to speed up the directed evolution of phytases. The main limiting factors in any directed evolution experiments are lack of diversity generation enriched with functional trait, difficulty in generating large pool of mutant library and lack of simple high throughput screening system. Screening of ~6800 mutants from epPCR library was not successful in improving thermostability of Ymphytase that might be due to transition biased method with low mutation frequency and use of only absorbance based screening assay. However, screening of only ~2350 mutant generated by SeSaM method improved the thermostability of Ymphytase. The mutational spectrum of SeSaM library showed a tranversion enriched mutant library that is responsible for 3 – 6 amino acid changes per clone. Although, there are no particular structural motifs, amino acid changes or covalent modification that can universally be employed to improve thermostability, most of the evolved enzymes have chemically different amino acid substitutions. The transition biased random mutagenesis method like epPCR with low mutational load mostly generate chemically similar amino acid changes [39]. The mutant libraries with high mutational load (epPCR; 15 to 30 mutations per gene) were reported to be enriched in improved sequences due to uniqueness and functional clones [40]. The functional diversity can be achieved with the mutagenesis method biased towards tranversions [10] like SeSaM. The MAP analysis [34] revealed two out of five substitutions found in M1 (T77K and V298M) were not achievable

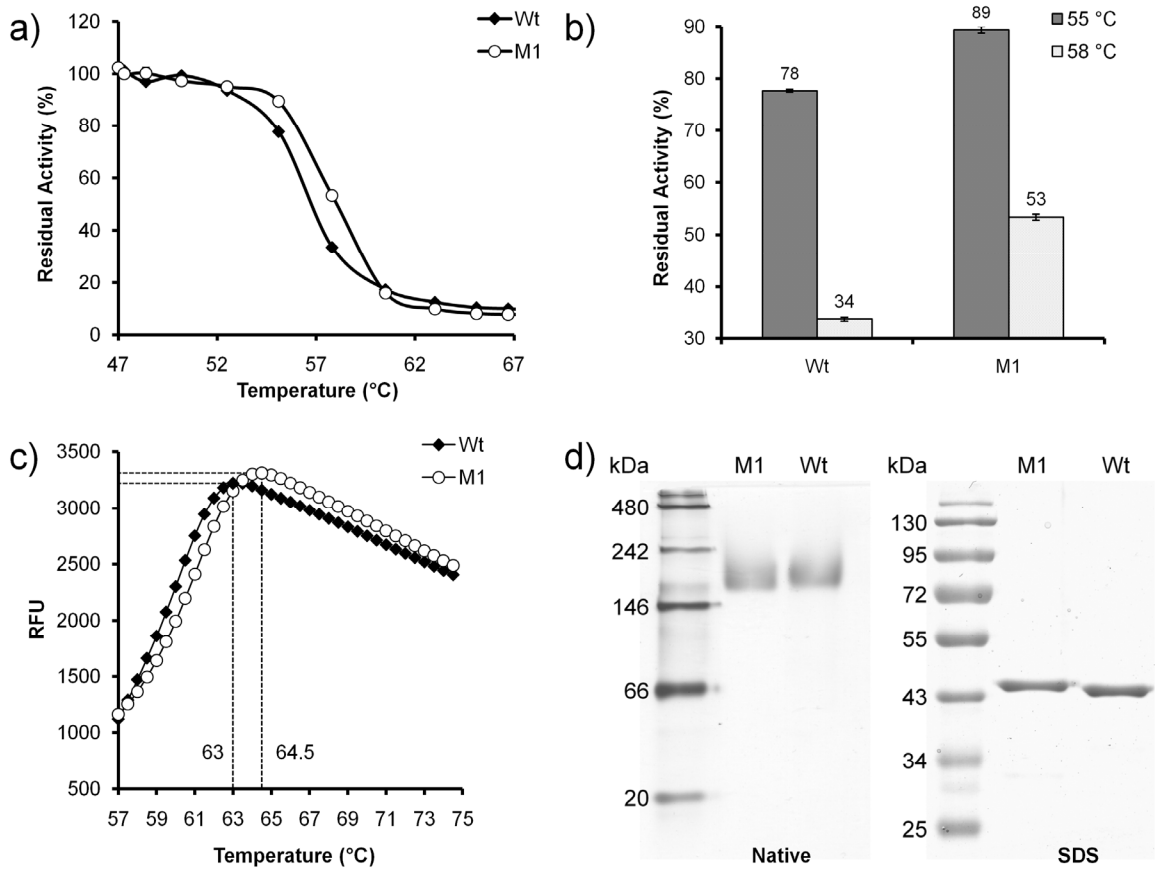


Figure 4: Residual enzyme activity of wild-type Ymphytase and M1 variants after being incubated at gradient temperatures (47 to 67 °C) for 20 min (a/b). Melting temperature (T_m) of wild-type (Wt) and M1 variant as determined using thermal shift assay. Purified wild-type (Wt) Ymphytase and M1 variant showed a tetrameric arrangement (native-PAGE) compare to denatured (SDS-PAGE) Ymphytase (d).

with epPCR conditions (Table 1; Supporting Information Figure S6). Additionally, a substitution T77K was a result of transversion mutation. The large pool of mutant library was generated using a modified small pET-22b(+) vector, pALXtreme-5b as a result of increased transformation efficiency [17]. The SeSaM method with increased mutational load was used to evolve Ymphytase combined with a simple high throughput prescreening system.

The screening system in 96-well format was optimized to broaden a linear range of phytate assay and reduce the true standard deviation. The epPCR library screened with only phytate assay with low mutational load, was not successful in improving thermostability of Ymphytase, and might be due to chemically similar substitutions. Therefore, a prescreening was employed in the directed evolution of Ymphytase using a non natural substrate (4-MUP). The selection pressure due to non-natural substrate used in prescreening might reduce the specific activity against natural substrate (phytate). Therefore, an activity

against phytate was also measured during 96-well format screening of mutant library to select the mutants that are active against phytate (natural substrate). Directed evolution protocol established for phytase has been validated by screening only 240 mutants in 96-well microtiter-plate format generated using SeSaM method. Improvement in thermostability of Ymphytase validates the simple and improved screening system combining both natural and non-natural substrates.

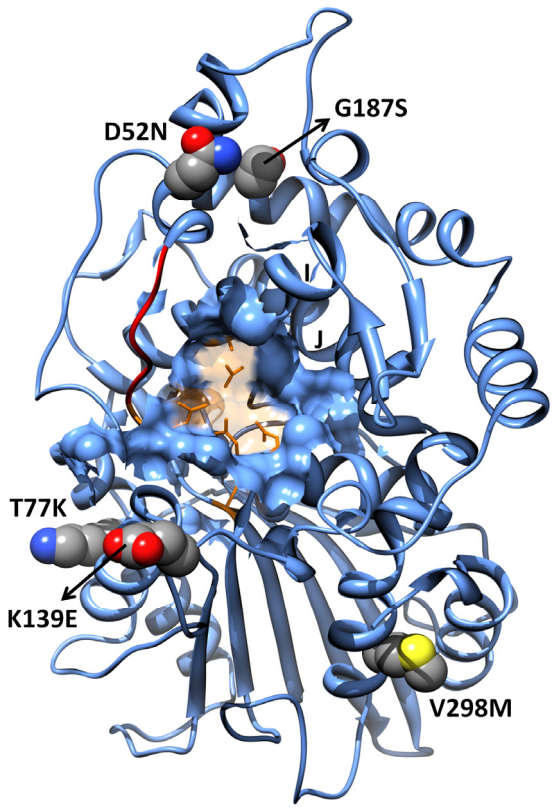


Figure 5: A structural model of Ymphytase. The amino acid substitutions in M1 variant are shown in space-filling model. Active site loop (Ser 42 – Thr 47) is highlighted by red color and catalytically important residues involved in substrate binding (Arg 37, Arg 41, Glu 241 and Asp 327) are shown in orange color in the active site that is depicted by molecular surface.

The T_{50} and T_m (melting temperature) values are often used to characterize thermostability, quantitatively [41, 42]. These values for M1 was increased by 1.5 °C compare to wild-type. Consistently, the residual activity was improved by ~20% after being heated at 58 °C for 20 min as a result of five substitutions (Figure 5). Both the wild-type and M1 were tetramer enzyme with a positive cooperativity between the monomers. A product inhibition was observed after 1 mM of substrate (phytate). The specific activity of partially purified enzyme solution for wild-type and M1 was 1073 ± 26 U/mg and 993.25 ± 3 U/mg, which is ~10 times higher than the *Aspergillus niger* phytase (100 U/mg) that is in widespread use [7].

The homology model for Ymphytase was constructed using *E. coli* phytase (PDB ID: 1DKM) as a template structure. The overall structural topology for Ymphytase was found to be similar to *E. coli* phytase [33] possessing conserved α/β-domain and a variable α-domain. The lower α/β-domain contains a centrally twisted seven stranded β-sheet of mixed topology with two α-helices on each side that is homologous to the α/β-domains of other phytases [33, 43, 44]. The α-domain of YmPhytase consists of two central α-helices (I, J: nomenclature as per *E. coli* phytase [33]) that partly define the substrate binding region with a number of helices and loops packed around them. The active site is located at the interface between these two domains composed by four residues (Arg 37, Arg 41, Glu 241 and Asp 327) involved in substrate binding and the active site loop (Ser 42 – Thr 47) based on *E. coli* phytase [11, 32]. All the five substitutions in M1 were distant from the active site except D52N. The substitution D52N was found to be close to the

active site loop of Ymphytase that might be the reason of small decrease in the activity of M1.

4.5 Conclusions

In summary, a rapid high throughput prescreening protocol and a directed evolution protocol has been developed for Ymphytase and validated by improving thermostability. We demonstrated the use of mutant library enriched with diverse substitution pattern obtained due to transversion biased mutagenesis method allow to improve phytase thermostability. A fast and simple prescreening method based on non-natural substrate (4-MUP) combined with natural substrate (phytate) screening can likely be employed to other phytases that would counteract the possibility of obtaining thermostable mutant with decreased activity towards natural substrate. Mutational analysis showed that the Ymphytase is reasonably robust towards mutations. The substitution T77K might be involved in ionic (salt bridge) interactions with D122 and G187S can make hydrogen bonding interactions with the adjacent loop. The reported improved screening platform for directed phytase evolution can likely be applied to other phytases for improving thermostability or activity.

Acknowledgements

We thank BASF SE for financial support and Dr. Alexander Schenk for providing pALXtreme-5b vector and *E. coli* BL21-Gold(DE3)laqI[Q1] expression strain.

4.6 References

1. Bohn, L., A.S. Meyer, and S.K. Rasmussen, *Phytate: impact on environment and human nutrition. A challenge for molecular breeding.* J Zhejiang Univ Sci B, 2008. 9(3): p. 165-91.
2. Bitar, K. and J.G. Reinhold, *Phytase and alkaline phosphatase activities in intestinal mucosae of rat, chicken, calf, and man.* Biochim Biophys Acta, 1972. 268(2): p. 442-52.
3. Harland, B.F. and D. Oberleas, *Phytase in animal nutrition and waste management.* BASF Reference Manual, 1999: p. 237–40.
4. Sebastian, S., S.P. Touchburn, and E.R. Chavez, *Implications of phytic acid and supplemental microbial phytase in poultry nutrition: a review.* World Poultry Sci J, 1998. 54(01): p. 27-47.
5. Vats, P., M.S. Bhattacharyya, and U.C. Banerjee, *Use of Phytases myo-Inositolhexakisphosphate Phosphohydrolases for Combatting Environmental Pollution: A Biological Approach.* Crit Rev Env Sci Tec, 2005. 35(5): p. 469 - 486.
6. Wodzinski, R.J. and A.H. Ullah, *Phytase.* Adv Appl Microbiol, 1996. 42: p. 263-302.
7. Wyss, M., et al., *Biophysical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): molecular size, glycosylation pattern, and engineering of proteolytic resistance.* Appl Environ Microbiol, 1999. 65(2): p. 359-366.
8. Huang, H., et al., *A novel phytase from Yersinia rohdei with high phytate hydrolysis activity under low pH and strong pepsin conditions.* Appl Microbiol Biotechnol, 2008. 80(3): p. 417-26.
9. Sen, S., V. Venkata Dasu, and B. Mandal, *Developments in directed evolution for improving enzyme functions.* Appl Biochem Biotechnol, 2007. 143(3): p. 212-23.
10. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Curr Opin Chem Biol, 2009. 13(1): p. 19-25.
11. Kumar, S., C.J. Tsai, and R. Nussinov, *Factors enhancing protein thermostability.* Protein Eng, 2000. 13(3): p. 179-91.
12. Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs.* J. Mol. Biol., 1997. 269: p. 631-43.
13. Kim, M.S. and X.G. Lei, *Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR.* Appl Microbiol Biotechnol, 2008. 79(1): p. 69-75.
14. Kim, M.S., J.D. Weaver, and X.G. Lei, *Assembly of mutations for improving thermostability of Escherichia coli AppA2 phytase.* Appl Microbiol Biotechnol, 2008. 79(5): p. 751-8.
15. Yun, C., H. Matsuda, and M. Kawamukai, *Directed evolution to enhance secretion efficiency and thermostability of chitosanase from Mitsuaria chitosanitabida 3001.* Biosci Biotechnol Biochem, 2006. 70(2): p. 559-63.
16. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain.* Biotechnol J, 2007. 2(1): p. 133-42.
17. Blanusa, M., et al., *Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method.* Anal Biochem, 2010. 406(2): p. 141-6.
18. Senn, A.M. and R.A. Wolosiuk, *A high-throughput screening for phosphatases using specific substrates.* Anal Biochem, 2005. 339(1): p. 150-6.
19. Huang, H., et al., *A novel phytase with preferable characteristics from Yersinia intermedia.* Biochem Biophys Res Commun, 2006. 350(4): p. 884-89.
20. Cummings, L., et al., *Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes.* FEMS Microbiol Lett, 2002. 216(2): p. 133-8.
21. Geer, L.Y., et al., *CDART: protein homology by domain architecture.* Genome Res, 2002. 12(10): p. 1619-23.

22. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* J Mol Biol, 2004. 340(4): p. 783-95.
23. Ferre, F. and P. Clote, *DiANNA: a web server for disulfide connectivity prediction.* Nucleic Acids Res, 2005. 33(Web Server issue): p. W230-2.
24. Wong, T.S., et al., *Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR.* Biotechnol J, 2008. 3(1): p. 74-82.
25. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. 41(1): p. 207-34.
26. Copeland, R.A., *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. 2nd ed. 2000: Wiley-VCH, New York.
27. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. 28(1): p. 235-42.
28. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. 22(2): p. 195-201.
29. Luthy, R., J.U. Bowie, and D. Eisenberg, *Assessment of protein models with three-dimensional profiles.* Nature, 1992. 356(6364): p. 83-5.
30. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* J Mol Biol, 1963. 7: p. 95-9.
31. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. 25(13): p. 1605-1612.
32. Shivange, A.V., U. Schwaneberg, and D. Roccatano, *Conformational dynamics of active site loop in Escherichia coli phytase.* Biopolymers, 2010. 93(11): p. 994-1002.
33. Lim, D., et al., *Crystal structures of Escherichia coli phytase and its complex with phytate.* Nat Struct Biol, 2000. 7(2): p. 108-13.
34. Wong, T.S., et al., *A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution.* J. Mol. Biol., 2006. 355(4): p. 858-71.
35. Gawronski, J.D. and D.R. Benson, *Microtiter assay for glutamine synthetase biosynthetic activity using inorganic phosphate detection.* Anal Biochem, 2004. 327(1): p. 114-8.
36. Gonzalez-Romo, P., S. Sanchez-Nieto, and M. Gavilanes-Ruiz, *A modified colorimetric method for the determination of orthophosphate in the presence of high ATP concentrations.* Anal Biochem, 1992. 200(2): p. 235-8.
37. Zhu, W., et al., *Modifying Thermostability of appA from Escherichia coli.* Curr Microbiol, 2010.
38. Varshavsky, A., *The N-end rule: Functions, mysteries, uses.* Proc. Nat. Acad. Sci. U.S.A., 1996. 93(22): p. 12142-49.
39. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Challenges of the genetic code for exploring sequence space in directed protein evolution.* Biocatal Biotransfor, 2007. 25(2): p. 229 - 241.
40. Drummond, D.A., et al., *Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins.* J Mol Biol, 2005. 350(4): p. 806-16.
41. Jaenicke, R. and G. Bohm, *The stability of proteins in extreme environments.* Curr Opin Struct Biol, 1998. 8(6): p. 738-48.
42. Zhao, H. and F.H. Arnold, *Directed evolution converts subtilisin E into a functional equivalent of thermitase.* Protein Eng, 1999. 12(1): p. 47-53.
43. Lindqvist, Y., G. Schneider, and P. Vihko, *Three-dimensional structure of rat acid phosphatase in complex with L(+)-tartrate.* J Biol Chem, 1993. 268(28): p. 20744-46.
44. Kostrewa, D., et al., *Crystal structure of Aspergillus niger pH 2.5 acid phosphatase at 2. 4 A resolution.* J Mol Biol, 1999. 288(5): p. 965-74.

Supporting Information: Chapter 4

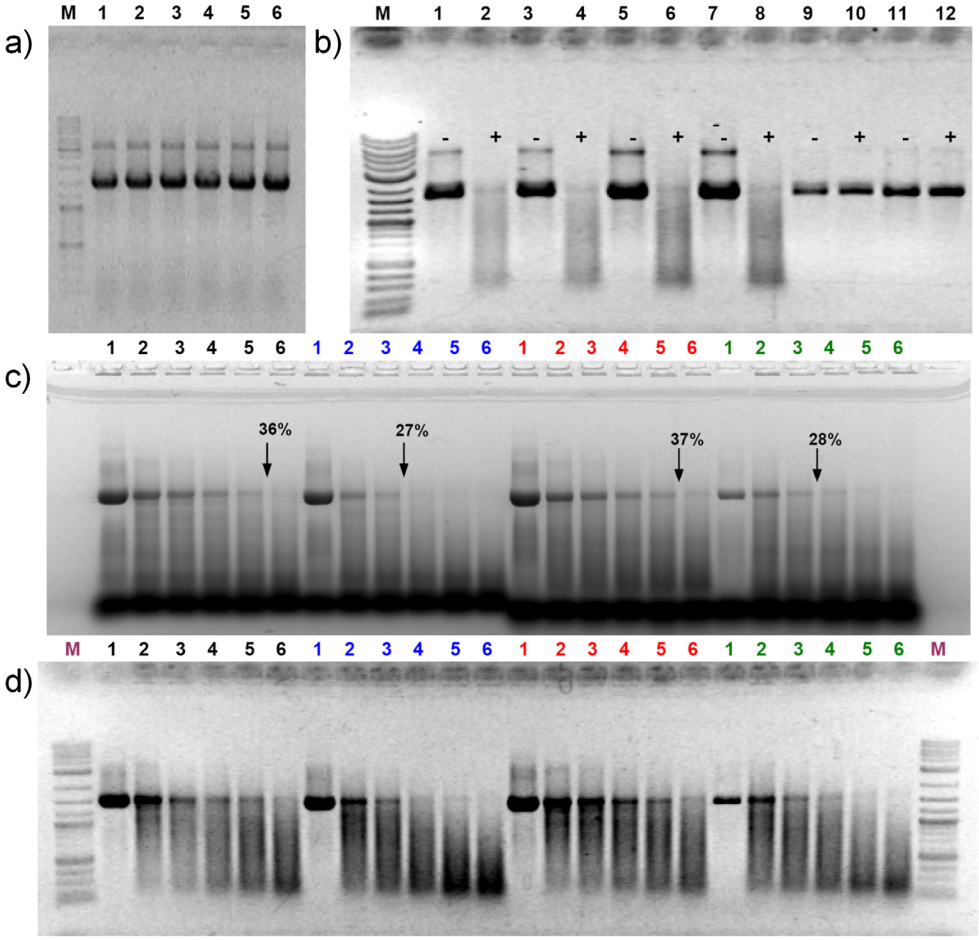


Figure S1: Electrophoretic analyses of steps in SeSaM technique used for *Y. mollaretii appA* library generation. (a) PCR amplification of SeSaM template; (b) fragmentation libraries, before (-) and after (+) iodine cleavage for A_fwd_library (lane – 1 & 2), G_fwd_library (lane – 3 & 4) A_rev_library (lane – 5 & 6), G_rev_library (lane – 7 & 8), forward positive controls (lane – 9 & 10) and reverse positive controls (lane – 11 & 12). [No αS was added to positive control]; (c/d) Determination of optimal phosphorothioate percentage using different concentrations of dATPαS or dGTPαS (lane1 = 0, lane2 = 20, lane3 = 25, lane4 = 30, lane5 = 35, lane6 = 40 percent) for A_fwd_library (black), G_fwd_library (blue) A_rev_library (red), G_rev_library (green) is determined using phosphorimager (c) and ethidium bromide staining (d).

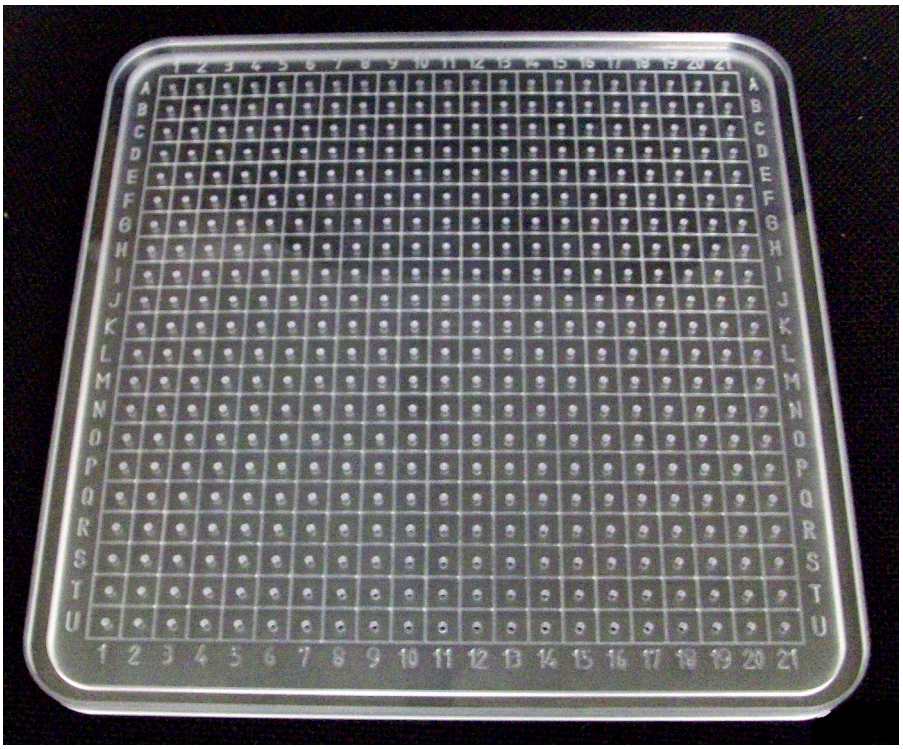


Figure S2: Autoclavable agar plate lid containing 441 apertures (21 × 21), separated by 10 mM.

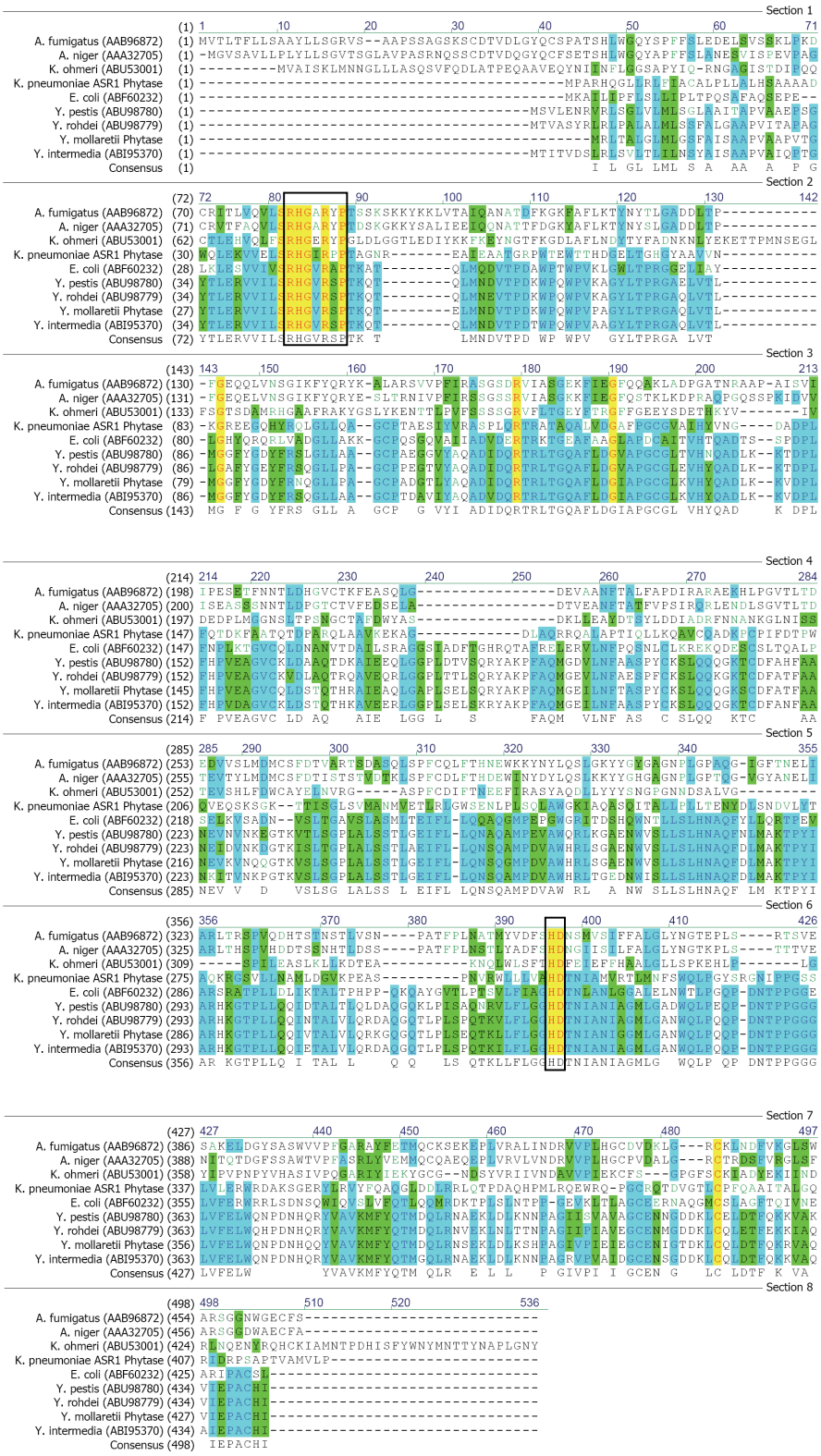


Figure S3: Multiple sequence alignment of fungal, yeast and bacterial phytases. Conserved sequence motifs “RHGXRXP” and “HD” are shown in black boxes.

Table S1: Mutational analysis of random clones from Ymphytase SeSaM mutant library

Mutant	P6-F1	P6-F2	P6-F3	P6-F4	P6-F5	P6-F8	All
Nucleotide changes	AC->GA G->T C->T T->C C->A	G->T T->C A->G G->C A->C C->T C->T	A->T G->C T->C C->T C->T T->A G->A A->T A->G C->T AA->G	T->C AT->CC C->G T->C T->C G->T G->A A->G A->G	A->T C->A A->T C->A T->C GT->AA	T->C A->T T->C G->C C->T C->A	
Amino acid changes	K139N, L201I, H433N	D56Y, A339V, T411I	K45STOP & frame shift…	V24A, D56A, T113S, L121P, V302M, Q304R	T44S, A211D, D277V, R287S, I333T, S393K	Q154H, T158I, T193N	
No. of nucleotide changes	5	7	11	9	6	6	
Single Ts	2	4	6	6	1	3	22
Single Tv	2	3	4	2	4	3	18
TvTs	1	0	0	1	1	0	3
Deletions	0	0	1	0	0	0	1

Ts – transitions, Tv – transversions, TvTs – transion followed by transversions

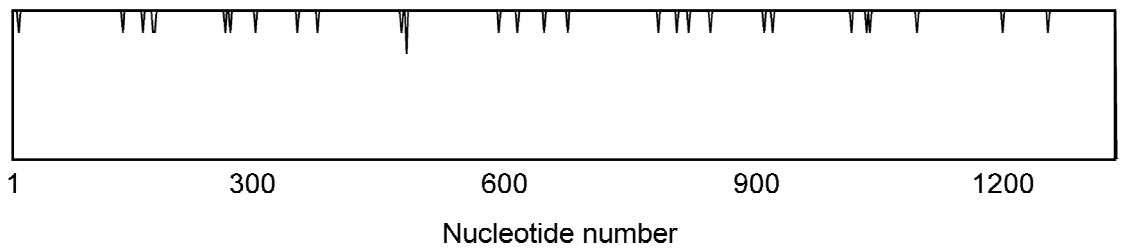


Figure S4: Distribution of mutations in six sequenced clones from SeSaM library. The down arrows indicate position of mutation in *YmappA*.

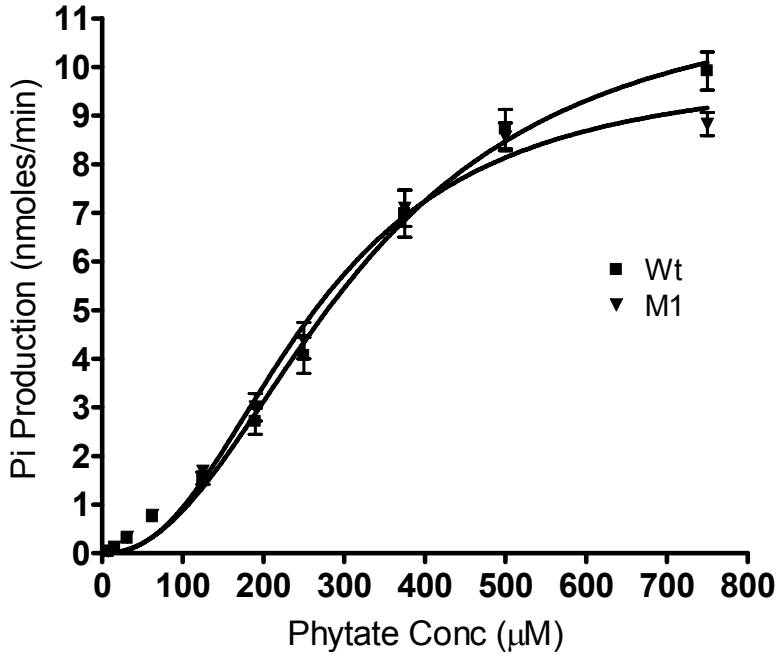


Figure S5: Enzyme velocity *vs* substrate plot of inorganic phosphate formation for wild-type (Wt) and mutant (M1) Ymphytase. A sigmoidal relationship between substrate concentration and product formation was observed, suggesting that Pi formation is driven by an allosteric interaction between monomers, showing positive cooperativity. The data was analyzed by Hill's equation.

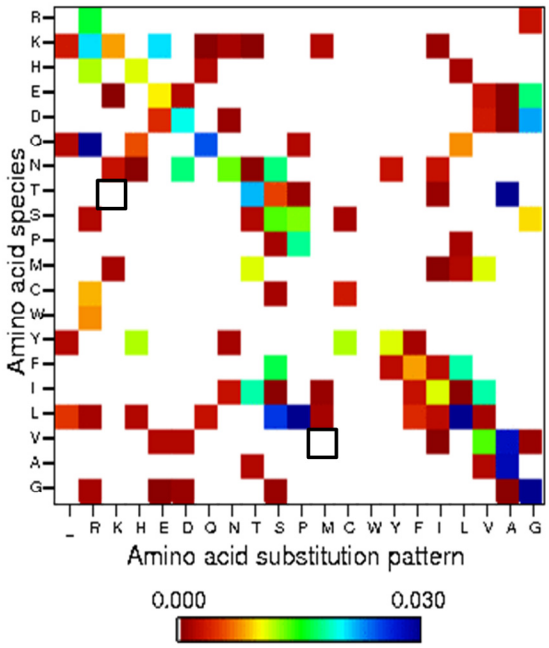


Figure S6: MAP analysis of amino acid substitution patterns of Ymphytase generated by an epPCR method [1] Taq (+, G>>A=C=T). The Y-axis shows the original amino acid species and the X-axis shows the substitution pattern. The color bar indicates the substitution pattern for 20 amino acid species from red (lowest probability) to blue (highest probability). Amino acid substitutions that do not occur are colored in white. Probability of T77K and V298M substitutions are highlighted in back box.

Reference

1. Wong, T.S., et al., *A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution.* J. Mol. Biol., 2006. 355(4): p. 858-71.

Toward understanding an activity-thermostability relationship in *Yersinia mollaretii* phytase

Amol V. Shivange[a,b], Danilo Roccatano[b] and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Bacterial phytase enzymes have attracted an increased interest from feed industry because of high catalytic efficiency. Prerequisites for any phytase for applications in industry are high activity and sufficient thermostability for feed pelleting. Therefore, it is important to study structure-function relationship of phytase influencing activity and thermostability. In chapter 5, we studied by molecular dynamics (MD) simulations, the effect of mutations on activity and thermostability of a bacterial phytase which are identified by directed evolution. The phytase variant (M2; T77K, G187S) was identified as a "most beneficial" combination from previously identified mutant (M1 in chapter 4) using site-directed mutagenesis. Sites found in second round of directed evolution using SeSaM were separately combined with M2 yielding thermostable variants with higher (M3; T77K, G187S, Q154H) and lower (M4; T77K, G187S, K289E) catalytic efficiency. Saturation mutagenesis of K289 resulted in a more thermostable variant (K289Q) but did not improve the activity. Finally, K289Q was combined with variant (M3) that revealed the reduced activity of M4 was compensated by a Q154H substitution. The best combination resulted in a variant M6 (T77K, G187S, K289Q, Q154H) with ~54% improvement in thermostability (58° for 20 min) and 3 °C increase in melting temperature without loss of specific activity and tetrameric structural arrangement. MD simulations results showed decreased overall phytase flexibility in thermostable variant with slight increase in active site loop flexibility. The decreased flexibility might be due to improved intra-protein interactions like hydrogen bonds (G187S, K289E) and salt-bridge interaction (T77K).

5.1 Introduction

Among several phytases widespread in nature, members of the histidine acid phosphatase (EC: 3.1.3.8) family have attracted increased interest from the animal feed industry because of their ability to efficiently hydrolyze phytate with high specific activity [1]. Despite wide range of applications, widely used commercial phytases like *Aspergillus niger* (BASF), *A. awamori* (AB Enzymes), face the

*Corresponding author

considerable challenge of low activity and thermostability. Adequate thermostability of phytase is requisite for feed application because diets for swine and poultry are usually pelleted at high temperature. Despite discovery of several phytases from different microorganism, only few have been evolved by directed evolution. However, only limited information is available concerning structural dynamics of this class of enzymes. Although, the universal principles governing thermostabilty improvement does not exist till date, the analyses of structural basis for protein stability have illustrated general factors responsible for protein stability. These include increase in ionic interactions (salt bridges), hydrogen bonds, disulphide bonds, hydrophobic interactions, reduction of conformational strain, and enhanced secondary structure propensity [2, 3]. Most of the success stories [4, 5] of phytase thermostabilization involved improvement of hydrogen bonding interactions predicted based on crystal structure conformation.

The previously indentified mutant SM2P3E4 (M1 in chapter 4) had five substitutions include D52N, T77K, K139E, G187S, and V298M). To identify the “key beneficial” residues responsible for improving thermostability, we have performed site directed mutagenesis on these positions in wild-type. The substitution T77K was found to be important for thermostability that was further improved in combination with G187S. Subsequently, sites identified by second round of directed evolution were combined to find the best combination of mutations improving thermostability. To gain insights into the possible molecular reasons for thermostability improvement in *Yersinia mollaretii* phytase (Ymphytase), MD simulations were performed on wild-type (20 ns) and two mutant enzymes (10 ns). A detailed comparison is made between all three system derived from MD simulation in water. Cluster analysis [6] was used to extract a conformation close to most abundant conformation in the trajectory. Principal component analysis [7] (PCA) was applied to separate the functionally important phytase flexibility from extraneous local fluctuations. To our knowledge, this is the first study of MD simulation of mutants on phytase class of enzymes for improved activity or thermostability. The Ymphytase variant M6 was obtained by combining “key beneficial” residues identified by directed evolution. MD simulations reveal an insight into intra-protein interactions and phytase flexibility that might be responsible for improved activity and thermostability of Ymphytase.

5.2 Material and Methods

5.2.1 Site-directed mutagenesis of YmappA

A modified method for site-directed mutagenesis (SDM) [8] was used to combine the key beneficial mutations. Two PCRs [98 °C for 30 sec, one cycle; 98 °C for 10 sec, 55 °C for 30 sec, 72 °C for 80 sec, 2 cycles] using either forward primer or reverse primer (0.4 mM each), Phusion Hot Start DNA polymerase (0.5 U in each; Finnzymes, Keilaranta, Finland) and dNTP mix (0.2 mM each) together with template (pALXtreme-5b harboring *YmappA*; 5 ng in each) were used. Both the PCRs were combined for final PCR [98 °C for 30 sec, one cycle; 98 °C for 10 sec, 55 °C for 30 sec, 72 °C for 80 sec, 14 cycles; 72 °C for 180 sec]

Table 1: Sequences of the primers used for site-directed mutagenesis and site saturation mutagenesis

Primer name	Sequence
Ymph_D52N_AS_F	CAGAGTTAATGAATAATGTCACACCGGATAAG
Ymph_T77K_AS_F	GGTGCGCAATTAGTGAAACTGATGGGCGGCTTC
Ymph_K139E_AS_F	CATTATCAGGCTGATTTGGAGAAGGTTGATCCACTG
Ymph_G187S_AS_F	CCATTTGCCCAGATGAGCGAGATTCTCAATTTC
Ymph_V298M_AS_F	GTTGTTGCAACAGATTATGACGGCGCTAGTGCTTC
YmG81E-F_AS	GTGACACTGATGGGCGAATTCTATGGTGATTATTTC
YmQ154H-F_AS	CCGGTGTGTGTCATCTAGATTCGACAC
YmL239V-F_AS	CACTCTCATCCACAGTGGGTGAAATCTTCTTG
YmK289E-F_AS	CTTATATCGCCCGTCATGAGGGAACTCCGTTGTTG
YmK289X-F_AS	CTTATATCGCCCGTCATNNKGGAACTCCGTTGTTG

YmPNdeR2HF primer was used to incorporate R2H mutation in the wild-type *appA* gene. Following the PCR, *DpnI* (20 U; New England Biolabs) was added, and the mixture was incubated overnight at 37 °C. The PCR products were transformed into *E. coli* BL21-Gold(DE3)LaqI^{Q1} for expression. Primers used for SDM are summarized in Table 1.

5.2.2 Construction of Ymphytase SeSaM library

A SeSaM mutagenesis technique [9] was used to generate a random mutagenesis libraries with high mutational load. Briefly, a SeSaM-Tv^{+} library for all four nucleotides in the coding strand of *YmappA* was generated, using pALXtreme-5b vector [10] specific primers SeSaM-fwd-F1 5' -CAC ACT ACC GCA CTC CGT CGC GAC TCA CTA TAG GGG AAT TGT GAG CGG A-3' and SeSaM-rev-R3 5'-GTG TGA TGG CGT GAG GCA GCC GGG CTT TGT TAG CAG CCG GAT CTC AG-3' for SeSaM template generation. Following *YmappA* fragmentation, purification, and universal base addition, the full length gene synthesis was achieved using PCR by combining A_fwd_library with A_rev_library and G_fwd_library with G_rev_library. The universal bases were replaced by standard nucleotide. Equal amounts of both A and G libraries were combined and digested with *NdeI* and *NotI* restriction enzymes. Digested PCR product was cloned into pALXtreme-5b vector and transformed into *E. coli* BL21-Gold(DE3)laqI^{Q1} strain [10] for expression and screening.

Prescreening, screening, purification and characterization were performed as described in section 4.2.

5.2.3 Screening of Ymphytase mutants for thermostability (T_{50}) in test-tube culture

Culture tubes (10 mL) containing autoinduction medium [11] ZYM-5052 (2 mL) supplemented with ampicillin (100 µg/mL) were inoculated with a 1:100 dilution of pre-culture that was grown in ZYM-505 media for 8 hours at

37°C (*E. coli* BL21-Gold(DE3)LaqI[Q1] harboring pALXtreme-5b-*YmappA*). After ~12 h of expression, *E. coli* cells (1 ml culture) were harvested by centrifugation (16000 g, 1 min, Eppendorf 5424). Cell lysis was carried out by lysozyme digestion (1 ml, 1 mg/ml, 37 °C) and cell lysate was obtained by centrifugation (16000 g, 1 min, Eppendorf 5424). The cell lysate was diluted (1:5) and diluted cell lysates were incubated at gradually increased temperature (47 °C to 67 °C) in PCR cycler (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) for 20 min. The protein samples were cooled down to 4 °C for 20 minutes. The residual phytase activity was measured using the phytase fluorescent assay at 37 °C and pH 5.5 as described in screening section 4.2.7.

5.2.4 Homology modeling

Homology modeling was performed using MODELLER 9v3 software [12].

Template selection

The amino acid sequence of Ymphytase was subjected to NCBI-BLAST (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) to identify homologous structure for Ymphytase by searching the structural database of protein sequences in the Protein Data Bank [13] (PDB). A template, *E. coli* phytase (1DKM) was selected based on highest percentage identity and highest resolution of the crystal structure (Supporting Information Table S1).

Sequence alignment and model building

Sequence alignment of the query sequence with respective template was performed using MODELLER program which uses dynamic programming algorithm. This method provides an alignment slightly different from standard sequence-sequence alignment methods since it takes into account structural information from the template when constructing an alignment. This is done by using variable gap penalty function that tends to place gaps in solvent exposed and curved regions, outside secondary structure segments, and between two positions that are close in space (Supporting Information Figure S1). Subsequently, The alignmenst was used by the same program to model the Ymphytase prote using 1DKM as a template structure.

Validation of homology models

The accuracy and validity of the generated homology model was checked with both the PROCHECK server implemented in SWISS-MODEL server (http://swissmodel.expasy.org) and 3D-profile programs [14] (http://nihserver.mbi.ucla.edu/Verify_3D/). The latter calculates 3D to 1D compatibility score and performs Eisenberg analysis for depiction of the properly folded and misfolded regions in the protein structure.

Ymphytase was found to be a tetrameric structure in solution. Unfortunately, the *E. coli* phytase template structure used for the homology modeling is a monomeric state in both solution and in the crystal [15]. Therefore modeling tetrameric arrangement for Ymphytase was not feasible and all the molecular modeling studies were carried out on the monomer homology model of Ymphytase.

5.2.5 MD simulations

MD simulation was performed first on wild-type and a representative conformation of the

most abundant cluster of wild-type phytase was extracted (see section 5.2.6). Mutations in the representative Ymphytase conformation were incorporated using Swiss-Pdb Viewer software [16], rotamers with the highest fit probability were selected for substituted residues. The N-terminal loop (M1 – A17; not correctly folded) corresponds to a signal sequence which was deleted from all Ymphytase structures. MD Simulations were performed on *E. coli* phytase (PDB ID: 1DKL; reference simulation), Ymphytase wild-type, M5, and M6 using the GROMACS (version 3.3.1) software package [17, 18] and GROMOS96 (version 43a1) forcefield [19]. Phytases were protonated and solvated in dodecahedron boxes with water by stacking equilibrated boxes of the solvent to form a box large enough to contain phytase and at least 0.95 nm (nanometer) in 1DKL and 0.9 nm in Ymphytase of the solvent on all sides. Since the resulting protonation state of the molecule gave a total charge of -3 in Ymphytase wild-type and M6 systems and -4 in 1DKL and M5 systems, three or four sodium counterions were added in corresponding system to satisfy the condition of global electroneutrality in a simulation cell. Each simulation box of Ymphytases and 1DKL contained roughly 17,400 and 15,000 water molecules respectively, resulting in about 21,450 and 19,000 atoms in Ymphytases and 1DKL system respectively. Energy minimization of the entire system was carried out for 1,000 steps using the steepest descents minimization algorithm in order to remove any van der Waals clashes generated by solvent molecules.

The MD simulations were carried out at a constant temperature and pressure. The temperature was kept constant at 300 K using the weak coupling algorithm [20] with coupling constant of 0.1 ps while the pressure was kept constant at 1 bar by the weak coupling constant of 0.5 ps [20]. For water molecules, the SPC water model [21] was used. The SHAKE algorithm [22] was used to constrain all the bond lengths. For the water molecules, the SETTLE algorithm was used to constrain bond lengths and bond angles [23]. The non-bonded interactions (electrostatic and LJ) were calculated using the Particle Mesh Ewalds (PME) method [24]. For the calculation of the long-range interactions, a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation were used to compute the potential and forces in between grid points. A non-bonded pair-list cutoff of 0.9 nm was used and the pair-list was updated every 5 time steps. The potential energy of each system was first minimized using the steepest descent algorithm. Water molecules were allowed to adjust by performing 100 ps of molecular dynamics with position restraints on the protein. Initial velocities for each atom were generated with the Maxwell distribution at 300 K. Each system was equilibrated for 100 ps before trajectory data were produced for analysis. Twenty nanosecond production runs were performed on wild-type and 1DKL and 10 ns production runs were performed for M5 and M6.

5.2.6 Cluster Analysis

Cluster analysis was used to identify a representative phytase conformations corresponding to the most abundant clusters in

the trajectories. The RMSD-based GROMOS algorithm [6] was used for the analysis of the MD trajectories. For each structure in the trajectory, a least-square translational and rotational fit for wild-type and mutants were performed using backbone and protein atoms respectively. The clustering algorithm assigns all structures within these cutoffs as neighbors with the structure possessing the largest number of neighbors being assigned as the median of the cluster. This structure was extracted from the pool along with its neighbors and the process repeated until all the structures are assigned to clusters [6]. A cutoff of 0.15 nm was selected for wild-type and 0.175 nm was selected for both the mutants M5 and M6 phytase.

5.2.7 Principal component analysis

Principal component analysis or essential dynamics is a powerful tool to extract essential configurational subspace that comprises most of the positional fluctuations from the subspace of extraneous local fluctuations in a protein [7]. PCA simplifies a MD simulation dataset by separating this configurational subspace into essential degrees of freedom (representing principal components/eigenvectors) that are relevant to protein function from the irrelevant local fluctuations. The PCA of each Ymphytase positional fluctuations along the MD trajectories was carried out to detect correlated motions of large amplitude. This method allows the characterization of relevant collective motions (essential motions) occurring during the MD simulation and to distinguish them from thermal noise motions.

The analysis was performed by building a covariance matrix of positional fluctuations of the backbone atoms (Cα, N, C) obtained from MD simulations. Upon diagonalization of covariance matrix, a set of eigenvalues and eigenvectors were generated. The eigenvectors correspond to directions in a 3N-dimensional space (where N = 1251, are the number of atoms used for the analysis) along which collective fluctuations of atoms occur. Eigenvalues represent the total mean square fluctuation of the system along the corresponding eigenvector.

It is often found that first few principal components (eigenvectors) depict more than half of the total conformational fluctuations with further eigenvectors frequently representing smaller random fluctuations [7]. These functionally important motions in the phytase were calculated by projecting the trajectories onto corresponding five largest principal components. These conformations were separated by 1 ns resulting in 10 models. Structural superimposition of these models with corresponding median structure obtained by cluster analysis was performed using UCSF Chimera software [25]. All the sequences were first aligned pairwise using Needleman-Wunsch algorithm with blosum-62 matrix and then structures were superimposed by fitting the aligned residue pairs. The PCA was performed on last 10 ns MD simulation trajectories of Ymphytase wildtype and mutants.

All analyses were performed using the tools of GROMACS [26] package. The molecular graphics images were generated using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [25].

5.3 Results

5.3.1 Identifying important residues in Ymphytase

Ymphytase mutant, SM2P3E4 reported previously (M1 in chapter 4) showed an improved thermostability attributable upto six substitutions including R2H, D52N, T77K, K139E, G187S, and V298M (Figure 1). Therefore, SDM was employed to incorporate each mutation except R2H in wild-type. First round of SDM reveal that only one substitution (M1; T77K) was improving thermostability (T_{50}) by 1 °C and other substitutions did not contribute to improved Ymphytase thermostability. In the second round of SDM, T77K was independently combined with all other four sites (D52N, K139E, G187S, and V298M). Out of four combinations, only one combination (M2; T77K, G187S) showed improved thermostability (T_{50} = 1.5 °C) compare to M1. Ymphytase mutant M2 showed improved thermostability (equal to SM2P3E4) without losing activity.

5.3.2 Screening a SeSaM mutant library

SeSaM mutant library was generated using SM2P3E4 as a template with high mutational load. A previously reported (Figure 2 in chapter 4), robust 384-well microtiter plate based high-throughput prescreening method was used. Only ~1600 clones (four plates) were sampled for improved thermostability. In the prescreening, using 384-well plates, mutants showing a ~20% higher residual activity (62 °C for 20 min) than wild-type Ymphytase were selected. Subsequently, selected 72 clones were screened in a 96-well microtiter plate based screening assay. Considering activity and thermostability, 11 clones were classified into four categories (i) Improved thermostability and activity (SM3P1F3); (ii) Improved thermostability and reduced activity (SM3P3E4); (iii) Improved activity (SM3P1D5, SM3P1F10, SM3P1F1, SM3P1A5, SM3P1B2, SM3P1A6, SM3P1C1); (iv) Improved activity and thermostability (SM3P1A3 and SM3P1C12). Mutant SM3P3D20 was identified by screening the same library at 75 °C (20 min) using only 384-well microtiter plate based screening system. Four new sites (K289E, G81E, Q154H, and L239V) found in thermostable mutants (group i and ii) were selected to be combined with the Ymphytase M2 variant (Figure 1).

5.3.3 Combining "key beneficial" mutations in Ymphytase

Importance of each position was analyzed by incorporating each mutation in the wild-type Ymphytase. Important residues were further combined by SDM. Ymphytase variant, M2 was identified as a best combination in SM2P3E4 variant. Q154H was combined with M1 or M2 (named as M3) resulting in improvement of 1.5 °C or 2 °C thermostability (T_{50} 20 min) respectively. However, combining K289E with M1 or M2 (named as M4) showed 2 °C or 2.5 °C improved thermostability (T_{50} 20 min) but the activity of all the mutants possessing K289E substitution was lower than the wild-type. Mutant SM3P1B4 possessing K289E mutation, showed only 40 – 50% activity in cell lysate compared to wild-type.

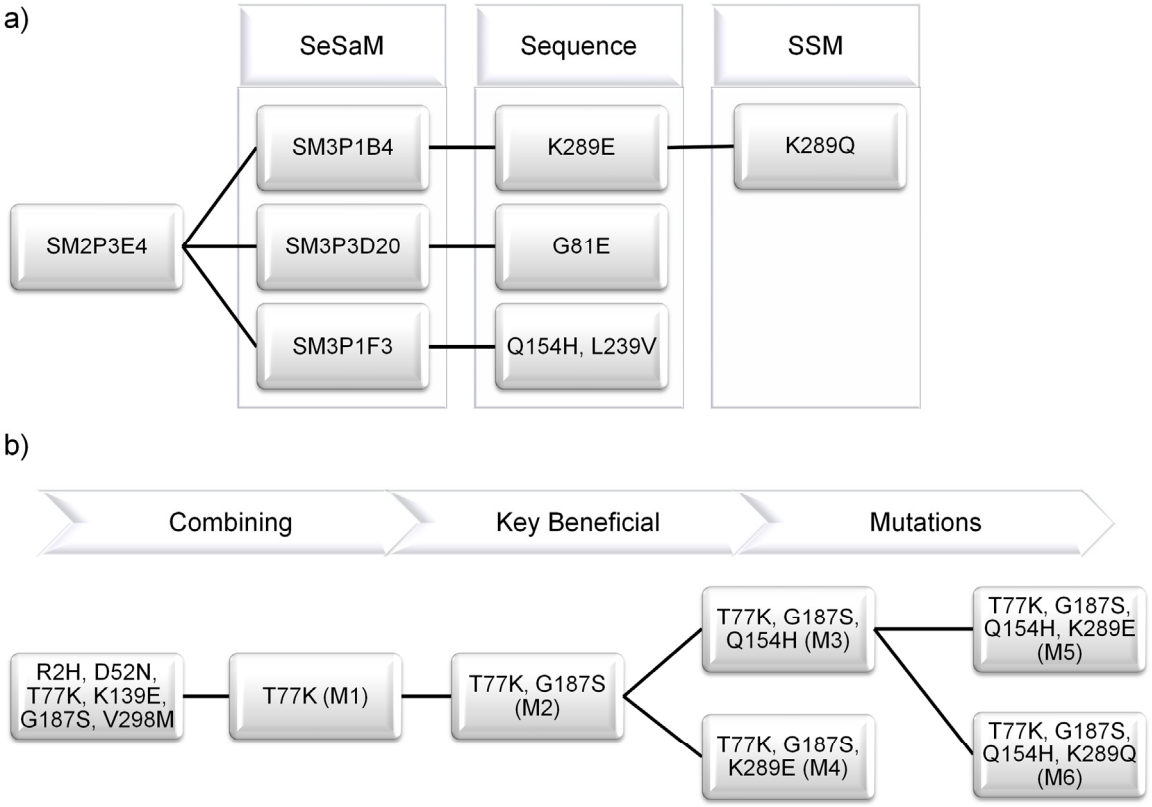


Figure 1: Schematic representation of mutations obtained in directed evolution (a); a flowchart of combining key beneficial mutations is depicted (b).

Site saturation mutagenesis

Position K289 was saturated using NNK codon degeneracy and mutant library (~200 clones) was screened in 96-well microtiter based screening system using 4-MUP as a substrate. Considering activity and thermostability 16 variants were selected for rescreening for T_{50} determination. All mutants except three clones (P1B4, P1A8, and P1C12) showed decreased activity compare to wild-type. Sequencing of these clones reveal that improved thermostability was achievable only with K289E or K289Q substitution whereas activity was only retained with wild-type amino acid K289K.

Ymphytase variant M3 (T77K, G187S, and Q154H) showed improved activity therefore M3 was combined with K289E (M5) and K289Q (M6). Ymphytase mutant M6 showed increased thermostability compare to M5 without a loss of activity when compared to wild-type Ymphytase.

5.3.4 Thermostability profiles

The most promising Ymphytase mutants M3 and M6 were purified and characterized. Both the mutants (M3 and M6) showed improved heat stability compared to wild-type, as the incubation temperature was increased. The M3 displayed a residual activity with ~42% and ~13% improvement over that of wild-type enzyme after

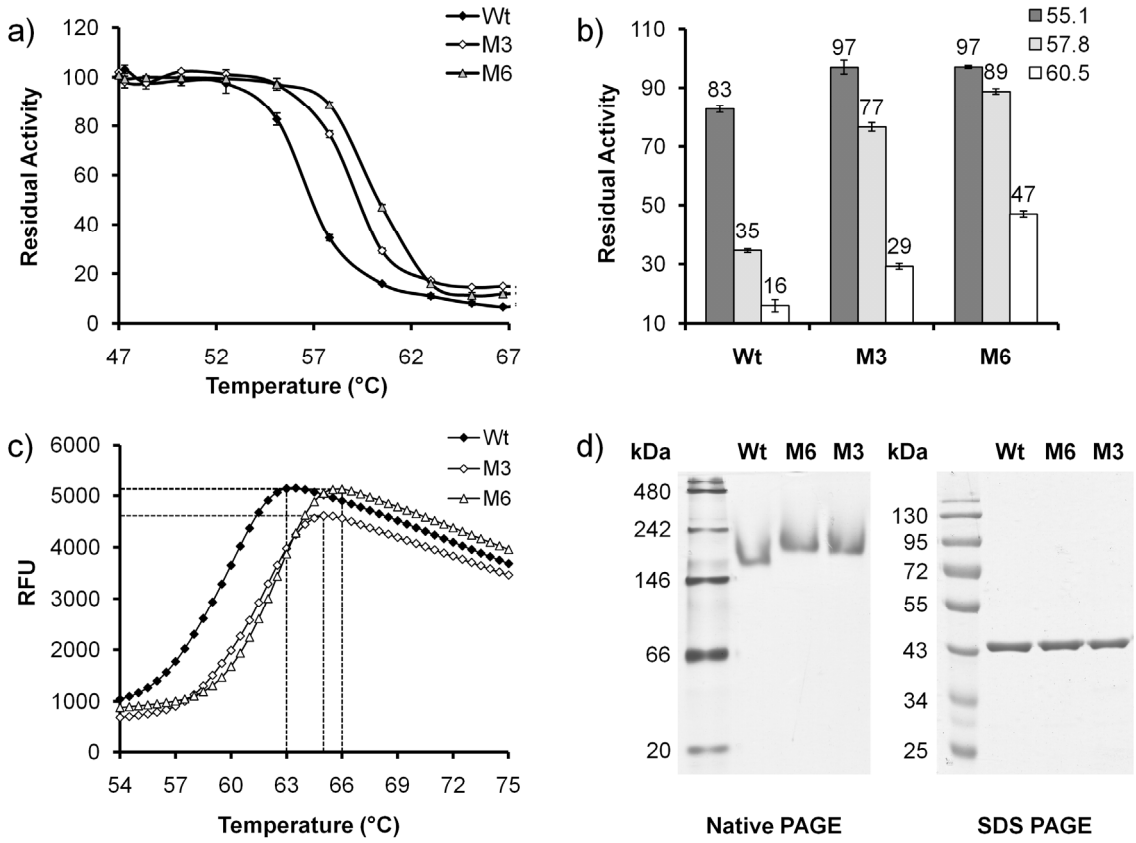


Figure 2: Residual enzyme activity of Ymphytase wild-type (Wt), M3 and M6 variants after being incubated at gradient temperatures (47 to 67 °C) for 20 min (a/b). Melting temperature (T_m) of wild-type (Wt), M3 and M6 variants as determined using thermal shift assay. Purified Ymphytase wild-type (Wt), M3 and M6 variants showed a tetrameric arrangement (native-PAGE) compare to denatured (SDS-PAGE) Ymphytase (d).

being heated at 58 °C and 60.5 °C for 20 min respectively. In case of M6, residual activity was improved by ~54% and ~31% compare to wild-type after being heated at 58 °C and 60.5 °C for 20 min respectively. The thermostability determined by T_{50} for M3 and M6 was found to be improved by 2.5 °C and 3.5 °C respectively compared to wild-type (Figure 2a and 2b; PCR cycler temperature homogeneity ≤ ±0.3 °C). The melting temperature (T_m), determined by thermal shift assay showed the mid-point of the thermal unfolding was increased by 2 °C and 3 °C for M3 and M6 respectively (Figure 2c).

5.3.5 Determination of kinetic parameters

The pH profile of the mutant M3 and M6 remained unchanged compared to wild-type. Both Ymphytase variants showed a pH optima at a pH 4.5. The phytase kinetic parameters were determined at pH 4.5, using sodium phytate as a substrate and generated data was analyzed by the general allosteric model using the Hill equation. The Hill coefficient (h) value for wild-type, M3 and M6 enzyme was 2.3, 2.05, and 2.65 indicating positive cooperativity (Table 2).

Table 2: Kinetic parameters and thermostbility of YmPhytase

Enzyme	Specific activity (U mg^{-1})	K_{half} (μM)	*K_{cat} (S^{-1})	Hills coefficient (h)	T_{50} (°C)	T_m (°C)
Wt	1073 (± 26)	292.75 (± 2.25)	846.74 (± 20.52)	2.3 (± 0.02)	56.5	63.0
M3	1274.5 (± 110)	353.65 (± 32.55)	1005.76 (± 87.2)	2.05 (± 0.12)	59.0	65.0
M6	1017.5 (± 0.5)	284.2 (± 7)	802.95 (± 0.4)	2.6485 (± 0.05)	60.0	66.0

Phytase kinetics was determined in 0.25 M sodium acetate containing different sodium phytate (7.8 to 750 μM) at 37°C and pH 4.5 for both the phytases
Values in the bracket represent standard error.
*Catalytic activity: enzyme concentration (moles) multiplied by the number of active centers (four).

The K_{half} for the wild-type, M3 and M6 were 292.8 μM, 353.7 μM and 284.2 μM respectively. The catalytic activity (K_{cat}) was calculated by considering four active site centers per molecule of Ymphytase. The specific activity of M3 was increased by ~200 U/mg compare to wild-type. In case of M6, the specific activity was not significantly altered (5% less than wild-type) against phytate.

5.3.6 Homology modeling

Quality of the constructed comparative protein model for Ymphytase was estimated using Ramachandran plot for YmPhytase (Supporting Information Figure S2) showed 98.9% residues in the allowed region, 0.5% in the generously allowed region, and only 0.5% in the disallowed region. The RMS deviation of the YmPhytase model with template *E. coli* phytase (1DKM), at the C-α region was calculated to be 0.23 Å and 0.28 Å in the backbone region. 3D compatibility assessed using 3D profile (Supporting Information Figure S3) revealed reliable folding in most of the regions of the Ymphytase model. This illustrates that the model was satisfactorily accurate for further structure-based analysis.

5.3.7 MD simulations

Stability and reliability of MD simulations on Ymphytase enzyme were accessed by analyzing root-mean-square deviations (RMSD) of backbone atoms from the starting structure coordinates of the *E. coli* phytase (Supporting Information Figure S4) and Ymphytase wild-type enzyme (Figure 3a). In case of *E. coli* phytase and Ymphytase wild-type, a stable plateau convergence occurred after ~10 ns simulation time.

Cluster analysis

Clustering of Ymphytase wild-type trajectory using RMSD-based cutoff of 0.15 nm on backbone atoms, resulted in 6 clusters. Two main clusters of Ymphytase conformations, covering 48% and 45% of the trajectory were obtained (Figure 3c). A representative median structure from first largest cluster was extracted from the trajectory. Mutations (M5 or M6) were incorporated into this force-field optimized Ymphytase structure for further MD simulations

of M5 and M6 variants. Ymphytase M5 and M6 showed an overall structural stability throughout simulation time (10 ns; Figure 3b). Cluster analysis on M5 and M6 using RMSD-based cutoff of 0.175 nm on all protein atoms, showed 10 and 18 clusters respectively. Approximately, 39% and 36% of the snapshots in the trajectory were represented by a largest cluster in M5 and M6 mutant respectively (Figure 3d and 3e).

Ymphytase flexibility analysis

Analysis of the root-mean-square fluctuations (RMSF) over last 10 ns simulation time reveal the two regions (L92 – T102, Q304 – Q317) in M5 and M6 to be less flexible compare to wild-type (Figure 4a). Overall flexibility of Ymphytase was decreased in order of M6, M5 and wild-type. PCA was performed in order to get more insight into flexibility of wild-type and mutant phytases. The 2D graph of the projection of Ymphytase wild-type, M5 and M6 trajectories on to corresponding first two eigenvectors obtained

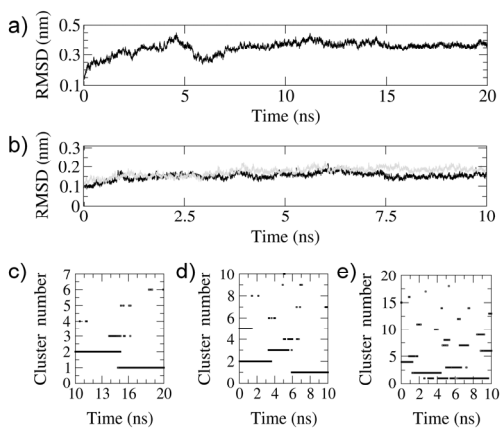


Figure 3: RMSD of backbone atoms from the starting structure over the length of MD simulation of Ymphytase wild-type (a), M5 (b; grey line) and M6 (b; black line) variants. Sampling of clusters over last 10 ns MD simulation trajectory of wild-type (c), M5 (d), and M6 (e) variant is shown.

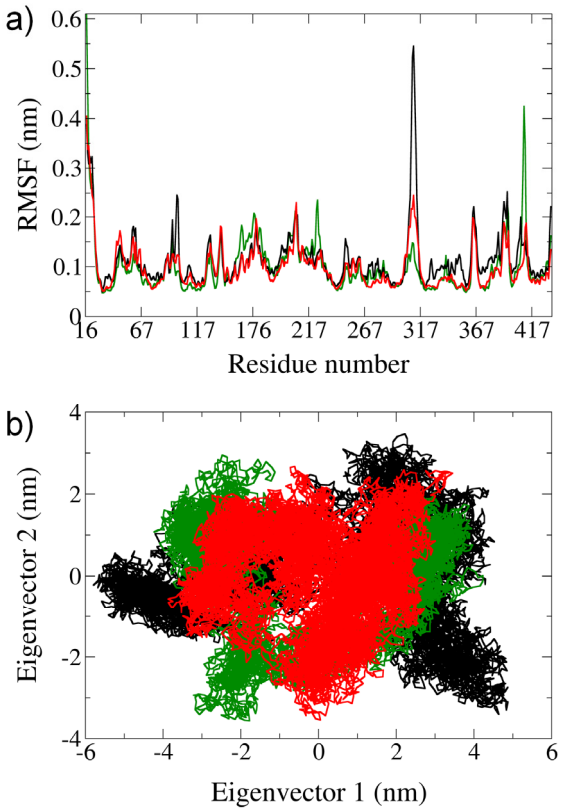


Figure 4: Root-mean-fluctuations over the last 10 ns simulation of wild-type (black), M5 (green), and M6 (red) is depicted (a); The 2D projection of wild-type (black), M5 (green), and M6 (red) trajectories onto the corresponding first two principle components obtained from all backbone atoms covariance matrix.

from backbone atoms is shown in Figure 4b. The phytase conformations in wild-type were found to be more scattered than M5 and the conformational sampling of Ymphytase M6 mutant was smallest compare to M5 and wild-type enzyme. First two eigenvectors covered only 48% and 25% of total fluctuations for wild-type and M6 respectively. Therefore, the trajectories were projected on to the corresponding first five eigenvectors (covering 60% and 42% of total fluctuation in wild-type M6 respectively) that showed a similar flexibility pattern. Flexibility in three regions (L92 - T102,

N247 – P252, and Q304 – Q317) has reduced compared to wild-type (Figure 5).

5.3.8 Structural topology of Ymphytase

Phytases are structurally conserved enzymes despite of low sequence similarity [27]. Structural topology of Ymphytase model (Figure 6) obtained by cluster analysis of MD simulation was found to be reasonably similar to phytases from enterobacteriaceae family [27, 28]. The active site of Ymphytase is located at the interface between conserved α/β-domain and a variable α-domain. The lower α/β-domain contains centrally twisted seven stranded β-sheet (1, 4, 10, 11, 12,13a, and 13b) of mixed topology with two α-helices (B, C and N, K) on each side that is homologous to the α/β-domains of other phytases [27, 29, 30]. The α-domain of YmPhytase consists of two central α-helices (I, J) that partly define the substrate binding pocket with a number of helices (A, D, E, F, G, L, and M), loops and a β-hairpin structure (6 and 7) packed around them.

5.4 Discussion

Site-directed mutagenesis was performed to identify beneficial residues, contributing thermostability of Ymphytase mutant (M1 in chapter 4 that is SM2P3E4) reveal T77K mutation (M1) to be essential. Mutation G187S did not show any improvement, however, G187S in combination with T77K (M2) increased thermostability that might be due to a cooperative effect between the two substitutions.

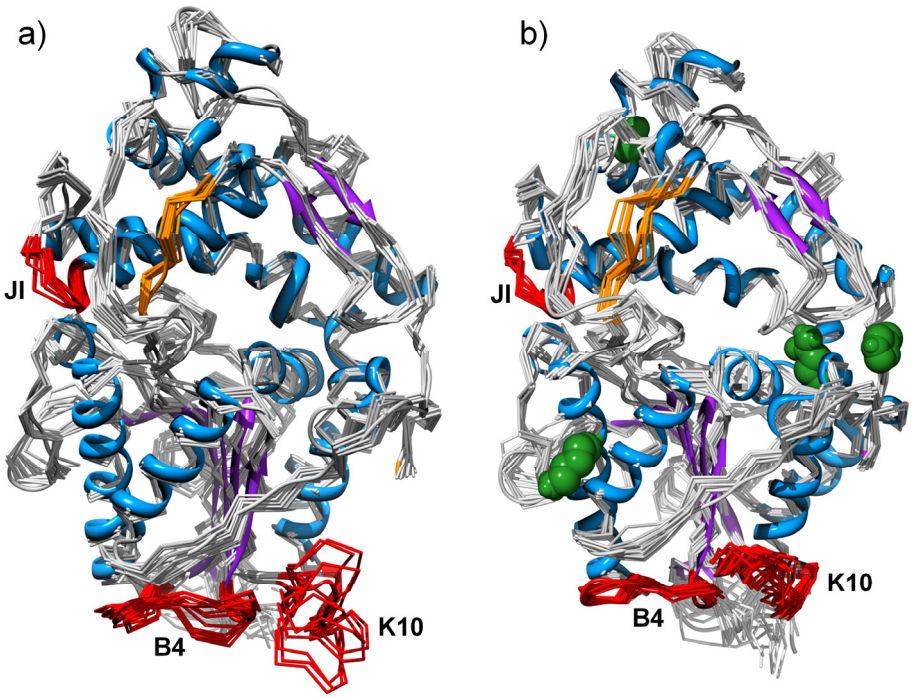


Figure 5: Structural superimposition of 10 configurations obtained by projecting the protein backbone motion onto the first five principle component of wild-type (a) or M6 variant (b), configurations are separated by 1 ns. Cartoon representation shows a most abundant conformation obtained by cluster analysis. Flexible loops in Ymphytase are shown in red color and active site loop is depicted by orange color.

The activity of M2 was not reduced compared to wild-type due to the absence of D52N mutation that was close to the active site loop (see section 4.4). Substitution R2H was not considered during recombination due to its contribution to *in-vivo* phytase stability improvement as a result of the N-end rule [31]. Subsequently, screening of another SeSaM library resulted in identification of new sites (K289E, G81E, Q154H, and L239V). Incorporation of individual sites in wild-type or M1 or M2 revealed K289E and Q154H to be the "key beneficial" substitutions for improving thermostability. Saturation mutagenesis of the K289 position did not improve the activity of phytase compare to K289E. Additionally, K289 showed an improved thermostability after substitution from lysine to glutamic acid with reduced activity. However, combining Q154H with M2, improved the activity of Ymphytase variant M3. All the key beneficial mutations were then combined together. Ymphytase variant M3 was combined with K289E (M5) and K289Q (M6) in a hope to compensate for reduced activity of Ymphytase variant M4, due to tha presence of K289E substitution. The thermostability for M6 was better than M5. As assumed, the reduced activity due to K289Q substitution was compensated after combining with M3, resulted in Ymphytase variant M6, possessing nearly same catalytic efficiency as wild-type.

To gain some structural insights for improved Ymphytase thermostability, we first constructed a reliable homology model. The validation of Ymphytase model indicated that the backbone dihedral angles (φ and ψ) were reasonably accurate with reasonable folding. Stability of the

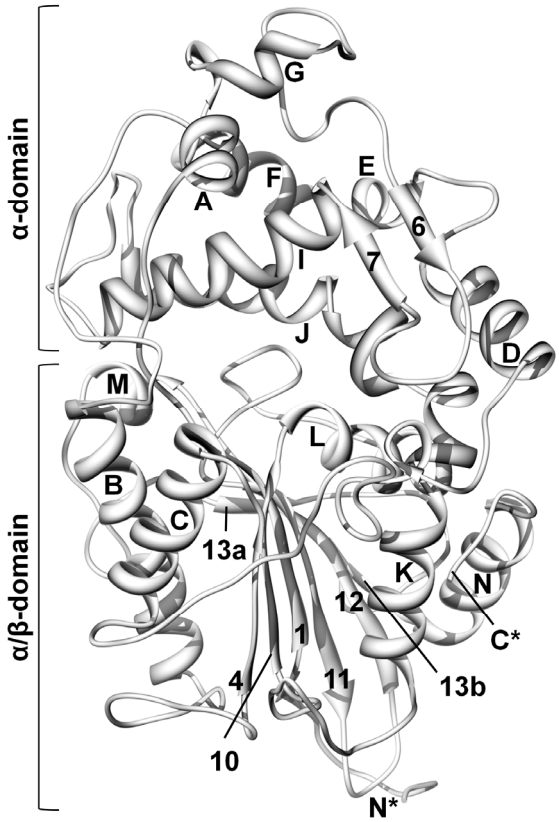


Figure 6: Structural topology of Ymphytase structural model. N-terminus (N*) and C-terminus (C*) are shown. Numbering was adopted from Lim *et. al.* [27].

trajectory assessed, using RMSD of backbone atoms as a function of simulation time, showed for all the four systems, an overall structural stability throughout the simulation time (last 10 ns). MD simulation on *E. coli* phytase was performed without crystal water molecules that served as a reference simulation for an employed forcefield and selected simulation parameters. The most abundant conformation of Ymphytase wild-type in the MD simulation trajectory was extracted using cluster analysis and mutations were incorporated to this median structure for further MD simulation of Ymphytase mutant M5 and M6.

Trajectory of the mutant M6 (most promising variant) was analyzed for intra-protein

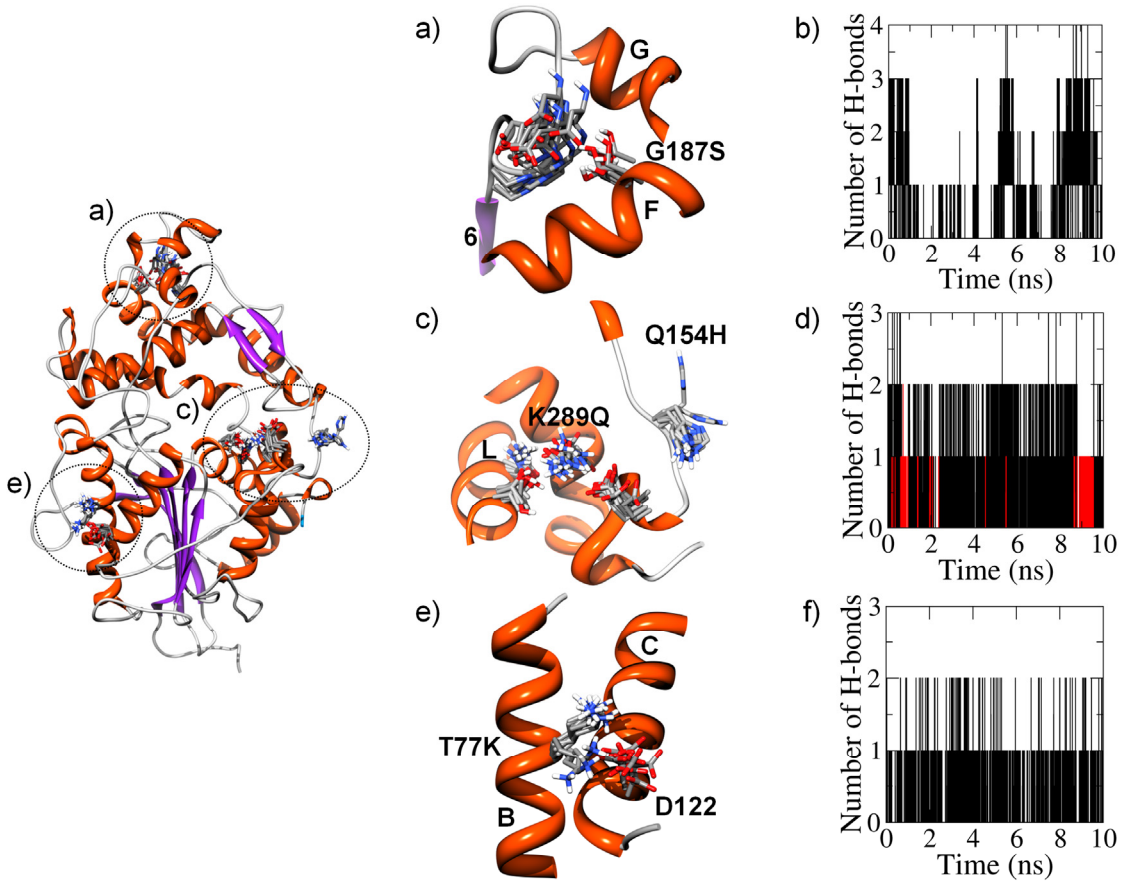


Figure 7: Hydrogen bonding network of Ymphytase variant M6 deciphered by MD simulation. Substitution G187S showed 4 hydrogen bonds with loop-G6 (a/b), K289Q showed 2 – 3 hydrogen bonds (black; d) with helix-L and 1 – 2 hydrogen bonds (red; d) with loop-CD (c/d), T77K was involved in salt-bridge interaction with D122 holding two helices (B and C) together (e/f). Side chain conformations shown are separated by 1 ns over the 10 ns MD simulation time.

interactions during the simulation time. Protein segments were considered for studying the interactions with the substitutions and compared with wild-type. Figure 7 shows a model of most abundant conformation from 10 ns simulation trajectory for mutant M6. Each mutation in M6 was analyzed for possible hydrogen bonding network in the phytase structure. Substitution G187S on helix-F was analyzed for possible hydrogen bonding interactions, that reveal formation of four hydrogen bonds with loop situated between helix-G and β-sheet-6 (Figure 7a and 7b). These interactions were not possible with glycine due to absence of hydroxyl side chain (Supporting Information Figure S5). In the case of K289Q substitution, a helix-L and loop (loop-CD) located between helix-C and helix-D were involved in hydrogen bonding network. Lysine 289 in wild-type trajectory showed 2 – 3 hydrogen bonding interactions with helix-L. However, total numbers of two hydrogen bonds for glutamine (K289Q) were more prominent compare to wild-type substitution. Additionally, hydrogen bonding interactions with loop-CD

were augmented when lysine residue was replaced with glutamine (Figure 7c and 7d) compare to wild-type (Supporting Information Figure S5). Substitution T77K showed a salt-bridge interaction with D122 throughout the simulation time (Figure 7e and 7f).

These improved intra-protein interactions might be responsible for improved thermostability. Overall flexibility as calculated by RMSF calculation was found to be reduced for both the Ymphytase mutants (M5 and M6). Further, the functional flexibility profile of the mutant and wild-type enzymes were studied by PCA. PCA is often used to extract the functionally important flexibility in the trajectory by eliminating irrelevant local fluctuations. Conformational sampling of the structures in the trajectory showed less scattered space for Ymphytase mutant M6 compare to wild-type enzyme. This reduced sampling space might be due to decreased Ymphytase flexibility in three main phytase segments (Figure 5; red colored). First region was a loop-B4 (L92 – T102) that was situated between helix-B and β-sheet-4. T77K mutation is located on helix-B and forms a salt-bridge interaction with D122 (on helix-C), holding the two helixes together. This reduced flexibility in loop-B4 following helix-B can be due to holding of two helices by salt-bridge interactions. The second region was the most flexible part of wild-type Ymphytase (Figure 4a) simulation. This loop-K10 (Q304 - Q317) between helix-K and β-sheet-10 showed reduced flexibility that might be due to K289Q mutation located on the helix-K. Additionally, flexibility in one more region (JI; N247 – P252), situated distant from mutations was decreased in M6 mutant compare to wild-type.

We have recently proposed a hypothesis on higher catalytic efficiency of enterobacteriaceae phytases which might be due to flexible active site loop [32]. The active site loop (S42 – T47) in Ymphytase mutant M6 was found to be more flexible than wild-type enzyme (Figure 5; orange colored). This increased flexibility in the active site loop might be a cooperative effect of mutations (T77K, G187S, and Q154H) that is improving specific activity of mutant M3 as determined experimentally (Table 2). The reduced activity of mutants containing K289Q/K289E substitution which is close to active site cavity (positive charge to negative or neutral charge) might be due to a different binding affinity of phytate (negatively charged substrate) compare to lysine substitution (positively charged) in wild-type Ymphytase.

An increase in fractional polar surface that results in increased number of hydrogen bonds with solvent has been proposed to improve the protein thermostability [3]. A side chain of glutamine residue (polar; neutral) at position 154 showed four hydrogen bonds with protein residues (Supporting Information Figure S6), making it unavailable to interact with solvent. Replacement glutamine by histidine (polar; positive) in Q154H, showed only two less frequent hydrogen bonds with Ymphytase during the simulation. This decreased intra-protein interactions of positively charged, surface residue histidine 154 presumably makes more hydrogen bonds to solvent (sodium acetate buffer; pH 5.5) than glutamine and increasing thermostability.

5.5 Conclusions

In summary, we have generated thermostable variants of Ymphytase enzyme. Variant with better catalytic efficiency (M3) and a variant compensating reduced catalytic efficiency (M6; activity equals to wild-type) were obtained by combining "key beneficial" mutations which were identified by directed evolution. During combination of mutations in Ymphytase only a small additive effect of thermostability improvement was observed. MD simulation studies illustrated that the increase in an intra-protein hydrogen bonding network is necessary for improving phytase thermostability. Solvent-exposed, charged amino acid substitution can improve the thermostability by decreasing intra-protein interaction in phytase. Increased active site loop flexibility, in support of proposed hypothesis [32] on phytase activity might be the reason for improved activity of Ymphytase mutant M3. The structure-function relationship revealed in this study can be a step forward to improve and understand thermostability and activity of other phytases.

5.6 References

1. Wyss, M., et al., *Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties.* Appl Environ Microbiol, 1999. 65(2): p. 367-73.
2. Kumar, S., C.J. Tsai, and R. Nussinov, *Factors enhancing protein thermostability.* Protein Eng, 2000. 13(3): p. 179-91.
3. Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs.* J. Mol. Biol., 1997. 269: p. 631-43.
4. Kim, M.S. and X.G. Lei, *Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR.* Appl Microbiol Biotechnol, 2008. 79(1): p. 69-75.
5. Kim, M.S., J.D. Weaver, and X.G. Lei, *Assembly of mutations for improving thermostability of Escherichia coli AppA2 phytase.* Appl Microbiol Biotechnol, 2008. 79(5): p. 751-8.
6. Daura, X., et al., *Peptide Folding: When Simulation Meets Experiment.* Angew Chem Int Edit, 1999. 38(1-2): p. 236-240.
7. Amadei, A., A.B.M. Linssen, and H.J.C. Berendsen, *Essential dynamics of proteins.* Proteins, 1993. 17(4): p. 412-425.
8. Hutchison, C.A., et al., *Mutagenesis at a specific position in a DNA sequence.* J Biol Chem, 1978. 253(18): p. 6551-60.
9. Wong, T.S., et al., *Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR.* Biotechnol J, 2008. 3(1): p. 74-82.
10. Blanusa, M., et al., *Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method.* Anal Biochem, 2010. 406(2): p. 141-6.
11. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. 41(1): p. 207-34.
12. Sali, A., *Comparative protein modeling by satisfaction of spatial restraints.* Immunotechnology, 1996. 2: p. 279-80.
13. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. 28(1): p. 235-42.
14. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure.* Science, 1991. 253(5016): p. 164-70.
15. Wyss, M., et al., *Biophysical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): molecular size, glycosylation pattern, and engineering of proteolytic resistance.* Appl Environ Microbiol, 1999. 65(2): p. 359-366.
16. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.* Electrophoresis, 1997. 18(15): p. 2714-23.
17. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation.* Comput Phys Commun, 1995. 91(1-3): p. 43-56.
18. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* J Mol Model, 2001. 7(8): p. 306-317.
19. Scott, W.R.P., et al., *The GROMOS Biomolecular Simulation Program Package.* J Phys Chem A, 1999. 103(19): p. 3596-3607.
20. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath.* J Chem Phys, 1984. 81(8): p. 3684-3690.
21. Berendsen, H.J.C., et al., *Interaction models for water in relation to protein hydration.* Intermolecular Forces, 1981: p. 331-342.
22. Ryckaert, J., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* J Comput Phys, 1977. 23(3): p. 327-341.
23. Miyamoto, S. and P. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models.* J Comput Chem, 1992. 13(8): p. 952-962.
24. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems.* J Chem Phys, 1993. 98(12): p. 10089-10092.
25. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. 25(13): p. 1605-1612.
26. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. 26(16): p. 1701-18.

27. Lim, D., et al., *Crystal structures of Escherichia coli phytase and its complex with phytate.* Nat Struct Biol, 2000. 7(2): p. 108-13.
28. Bohm, K., et al., *Crystal structure of Klebsiella sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme.* Febs J, 2010. 277(5): p. 1284-96.
29. Lindqvist, Y., G. Schneider, and P. Vihko, *Three-dimensional structure of rat acid phosphatase in complex with L(+)-tartrate.* J Biol Chem, 1993. 268(28): p. 20744-46.
30. Kostrewa, D., et al., *Crystal structure of Aspergillus niger pH 2.5 acid phosphatase at 2. 4 A resolution.* J Mol Biol, 1999. 288(5): p. 965-74.
31. Varshavsky, A., *The N-end rule: Functions, mysteries, uses.* Proc. Nat. Acad. Sci. U.S.A., 1996. 93(22): p. 12142-49.
32. Shivange, A.V., U. Schwaneberg, and D. Roccatano, *Conformational dynamics of active site loop in Escherichia coli phytase.* Biopolymers, 2010. 93(11): p. 994-1002.

Supporting Information: Chapter 5

Table S1: Available templates for homology modeling of YmPhytase (*selected template)

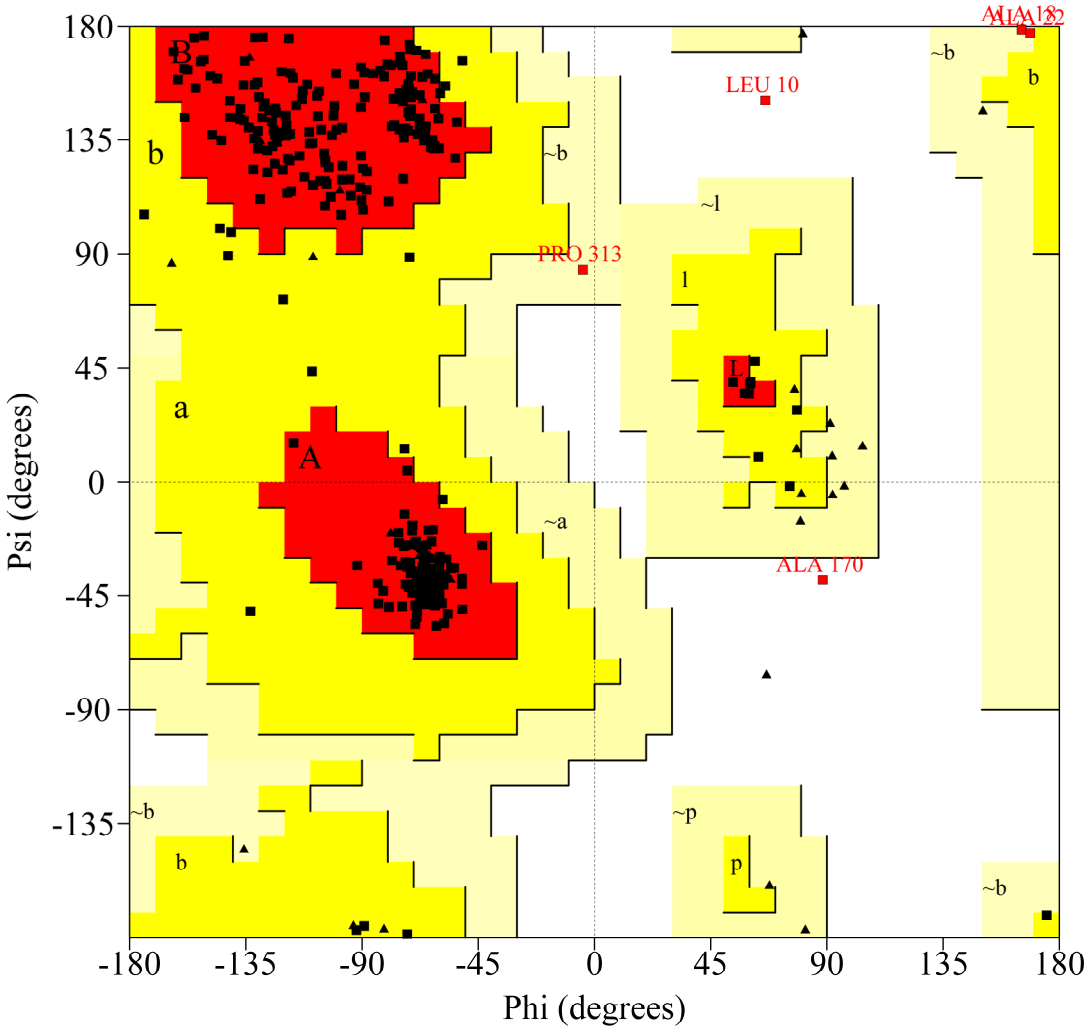
Template PDBID	% Identity	Resolution (Å)	Mutation (s)	Ligand
1DKL	45	2.30	No	No
1DKM*	45	2.25	A116T	No
1DKP	44	2.28	A116T, H17A	Phytate
1NT4	33	2.40	H18A	Glucose-1-Phosphate

```
 _aln.pos          10        20        30        40        50        60
1dkmA     ---------------------QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWPTWPVKLGWL
ymph      MRLTALGLMLGSFAISAAPVAAPVTGYTLERVVILSRHGVRSPTKQTELMNDVTPDKWPQWPVPAGYL
 _consrvd                              ** *** ****** *** * ** ***** ** ***  * *

 _aln.p   70        80        90       100       110       120       130
1dkmA     TPRGGELIAYLGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQ
ymph      TPRGAQLVTLMGGFYGDYFRNQGLLP-AGCPADGTLYAQADIDQRTRLTGQAFLDGIAPGCGLKVHYQ
 _consrvd ****  *    *          ***   ***  *     ** * *** ** **  * ** *   ** *

 _aln.pos  140       150       160       170       180       190       200
1dkmA     TDTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGHRQTAFRELERVLNFPQSNLCLKREK
ymph      ADLKKVDPLFHPVEAGVCQLDSTQTHRAIEAQLGAPLSELSQRYAKPFAQMGEILNFTASPYCKSLQQ
 _consrvd  *    **** *   ******      **    *             *      ***  *  *

 _aln.pos    210       220       230       240       250       260       270
1dkmA     QDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSL
ymph      QGKSCDFATFAANEVKVNQQGTKVSLSGPLALSSTLGEIFLLQNSQGMPDVAWHRLSGAENWVSLLSL
 _consrvd *  **         * **      *** *   * * * ******  ****   * *      *  ****

 _aln.pos      280       290       300       310       320       330       340
1dkmA     HNAQFYLLQRTPEVARSRATPLLDLIKTALTPHP-PQKQAYGVTLPTSVLFIAGHDTNLANLGGALEL
ymph      HNAQFDLMAKTPYIARHKGTPLLQQIVTALVLQRKGQGQTLPLSEQTKLLFLGGHDTNIANIGGMLGA
 _consrvd ***** *   **  **   ****  * ***      * *       *  **  ***** ** ** *

 _aln.pos        350       360       370       380       390       400
1dkmA     NWTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQQMRDKTPLSLN-TPPGEVKLTLAGCE
ymph      NWQLPQQPDNTPPGGGLVFELWQNPDNHQQYVAVKMFYQTMDQLRNSEKLDLKSHPAGIVPIEIEGCE
 _consrvd ** ** ********* **** *       *   *    **  * *    * *   * * *     ***

 _aln.p  410       420       430
1dkmA     ERNAQGMCSLAGFTQIVNEARIPACSL
ymph      NIGTDKLCQLDTFQKRVAQVIEPACHI
 _consrvd        * *  *   *     ***
```

Figure S1: Structure-sequence alignment of template and query sequence based on dynamic programming algorithm of MODELLER for YmPhytase.

PROCHECK

Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	339	93.1%
Residues in additional allowed regions [a,b,l,p]	21	5.8%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.5%
Residues in disallowed regions	2	0.5%
	---	------
Number of non-glycine and non-proline residues	364	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	39	
Number of proline residues	29	

Total number of residues	434	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure S2: PROCHECK-Ramchandran plot for the homology model of YmPhytase.

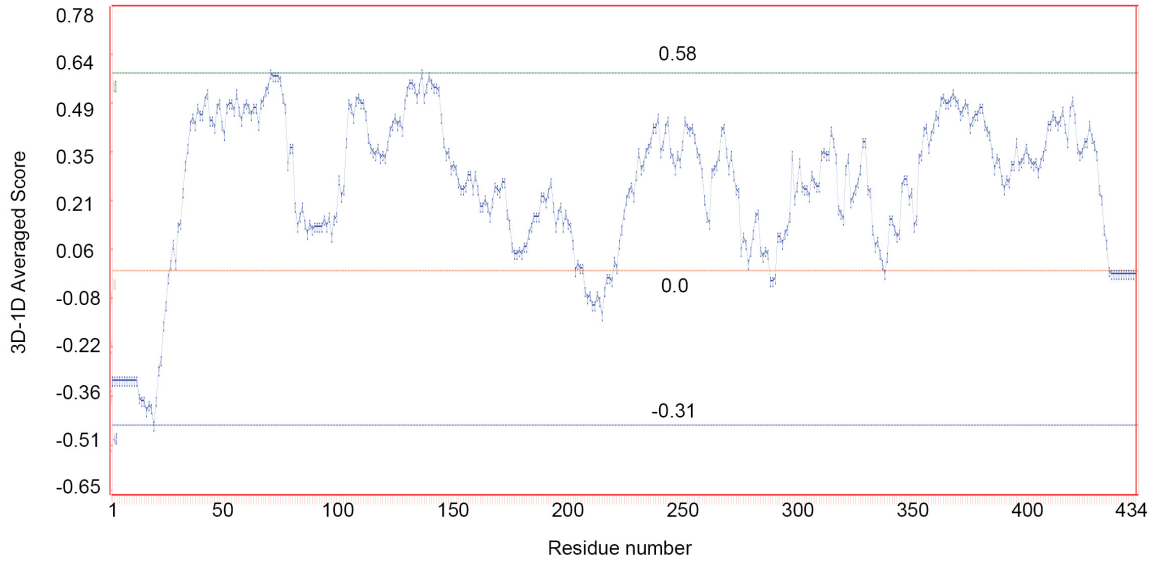


Figure S3: 3D-profile window plots for Ymphytase homology model. The vertical axis shows the average 3D-1D score for the residue and horizontal line indicate residue number in the model. The residues with average positive score in 3D-1D slide window are reasonably folded.

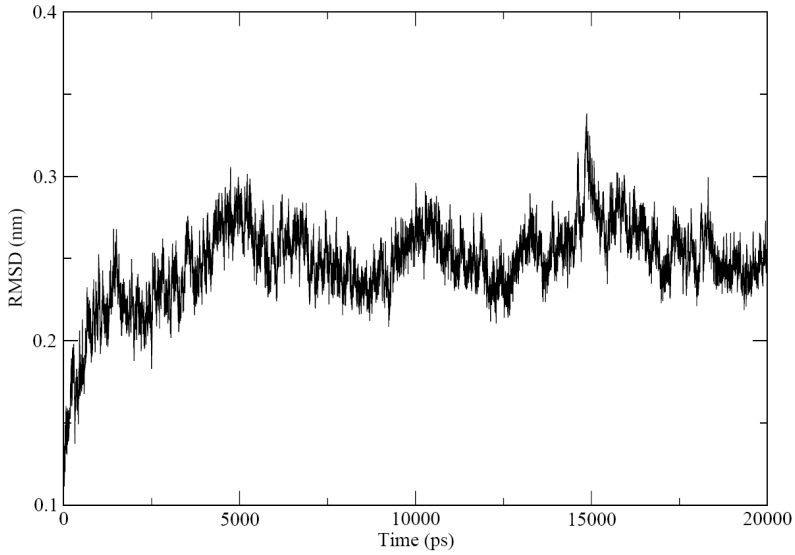


Figure S4: RMSD of backbone atoms from the starting structure over the length of MD simulation of *E. coli* phytase (1DKL) trajectory.

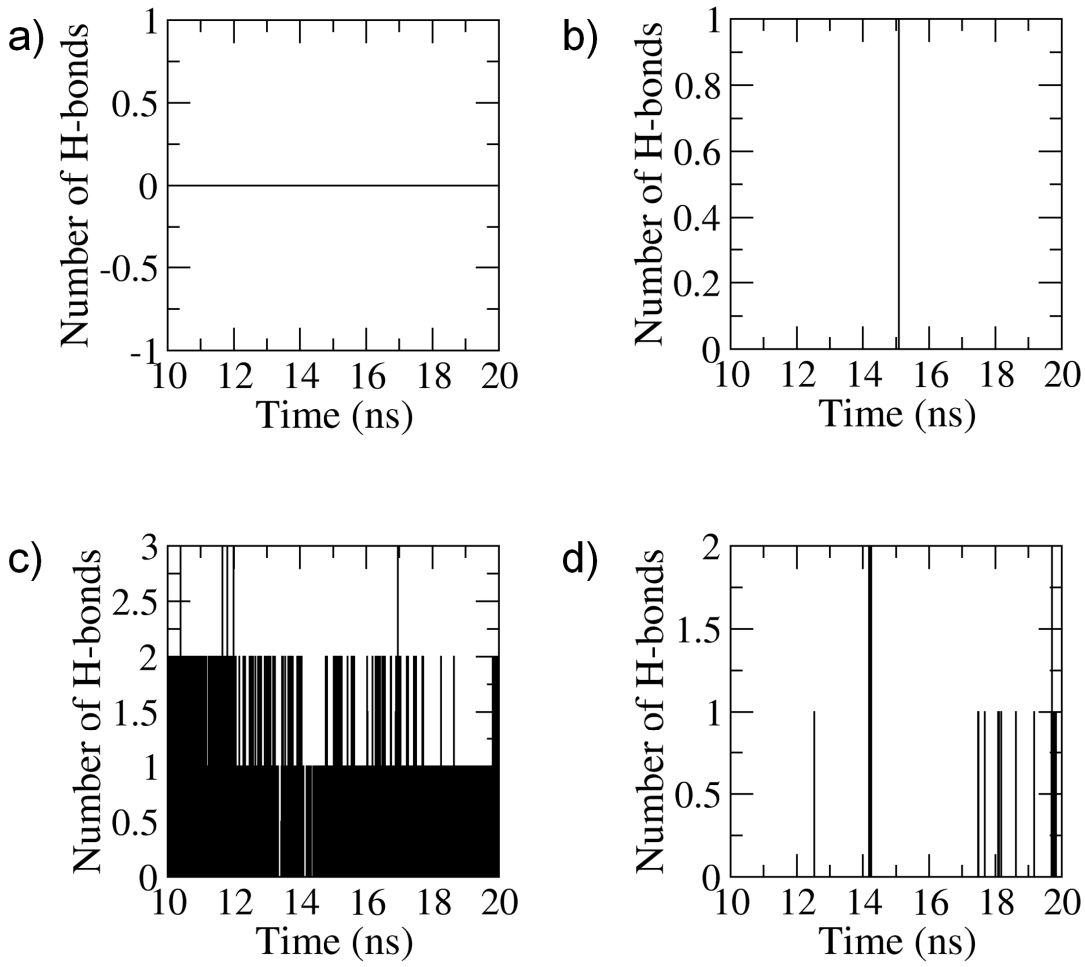


Figure S5: Hydrogen bonding network of Ymphytase wild-type deciphered by MD simulation. Amino acid residue G187 showed no hydrogen bonds with loop-G6 (a), T77 did not show any hydrogen bonds with D122 (b), K289 showed 2 – 3 hydrogen bonds with helix-L (c) but the frequency of second hydrogen bond was less compare to variant M6 and negligible hydrogen bonds with loop-CD (d).

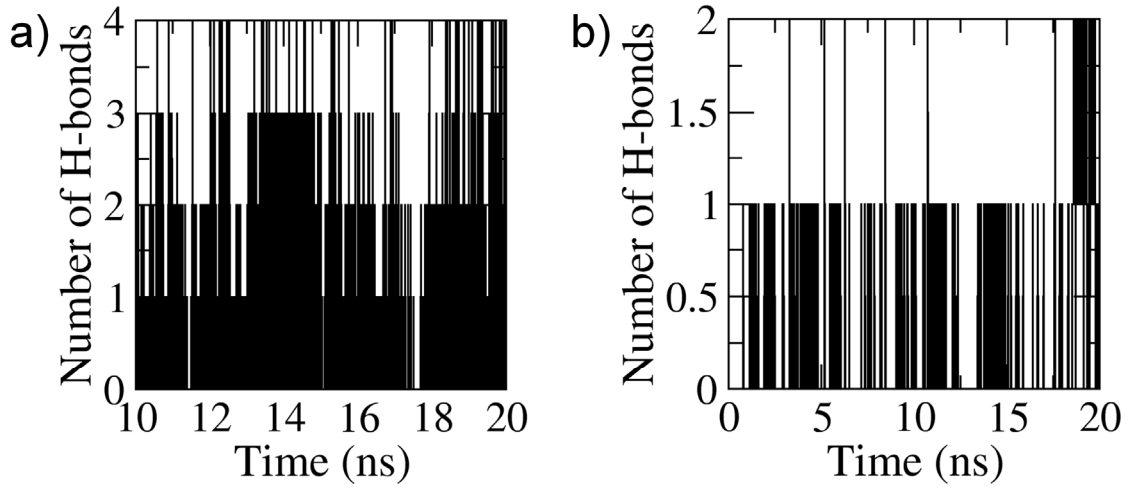


Figure S6: Hydrogen bonding network of Ymphytase residue Q154 in wild-type (a) and substitution Q154H in variant M6 over the time of MD simulation. Hydrogen bonds were calculated between residue 154 and the protein molecule.

Combinatorial assembly of site saturation in protein segment (ProCASTing): a novel method for protein engineering

Amol V. Shivange[a,b] Rajni Verma[b], Danilo Roccatano[b], and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Combinatorial assembly of site saturation test in protein segment (ProCASTing) is a novel method for parallel site saturation of more than one consecutive sites in any part of the protein. ProCASTing can be accomplished within one day and comprises three steps: template generation, amplification of DNA fragments using phosphothioated primers, and chemical cleavage of phosphothioate bonds in an iodine/ethanol solution, ligation, and transformation. ProCASTing is absolutely sequence independent method. The amino acid residues in any segment of protein structure can easily be saturated to replace with other amino acids. Additionally, ProCASTing ligation can be accomplished within 5 min making it robust and reliable. Using *Yersinia mollaretii* phytase (Ymphytase) enzyme as a model system, we proved by saturating a segment in Ymphytase. A protein segment (8 amino acids; Gly 307 – Ser 315) in Ymphytase was saturated using NDT (N: A, T, G or C; D: A, T or G) codon degeneracy for mutating eight subsequent amino acids or reducing the protein segment to four consecutive substitutions. Sequencing random clones from both the libraries showed a saturation mutagenesis of all eight or four amino acids in the selected region of Ymphytase. This method can be used in combination with high throughput screening to harness the sequence space efficiently.

6.1 Introduction

Enzymes with proper configurations possess outstanding capacity for carrying out complex molecular reactions making them highly versatile and proficient catalyst. Enzymes optimized by evolution over millions of years, can perform reactions with extreme precision and efficiency such as activity, stability, enantio-selectivity or enantio-specificity, substrate selectivity, pH range [1-3]. Protein engineering by directed evolution has over the last decades become successful tool for tailoring the catalytic, biophysical and molecular recognition properties of target proteins and for exploring structure-function relationship. It consists of four major

*Corresponding author

steps that comprise random gene mutagenesis (diversity generation), cloning of variants, screening (or selecting) for improved variants (best hit) for desired functional property and isolation of gene for next iterative cycle of directed evolution. The process is repeated until the desired level of improvement has been achieved [4]. The approach of directed evolution can be challenging as even mutant libraries with millions of variants still sample only a tiny fraction of astronomically vast sequence space possible for average protein size. Advances in smart library design for restricting the sequence space have been already discussed [5, 6]. However, biases due to experimental methods and organization of genetic code further restrict the library design [7, 8]. Therefore, many researchers are moving beyond traditional directed evolution, towards developing new strategies for smaller and high quality library design.

Protein engineering by semi-rational, rational, focused or knowledge-based library design is an attractive immersing filed for tailoring enzymes for desired needs. These approaches are based mainly on computational algorithms that utilize the protein sequence-structure and function information. Until now, most of these rational design approaches focused on single amino acid substitution and therefore limited by mutating specific protein segment due to experimental methodological limitations. Site-directed mutagenesis has been enormously used to study structure-function relationship by mutating the residues identified by directed evolution or by rational design [9-11]. In addition, a method of site-saturation mutagenesis has proven to be a promising tool to explore the position in the enzyme to understand the importance of the specific site [12] and identify the best fit residue. Several success stories have been reported using these methods to identify beneficial, neutral and deleterious mutation influencing the enzyme property [13] and improving enzyme property by iteratively combining mutations [14]. From practical perspective, focusing the directed evolution in part of the protein would significantly increase the efficiency of biocatalyst tailoring by moving the field from discovery-based towards hypothesis-driven protein engineering.

In this study, we have developed ProCASTing, a novel method is developed for parallel mutagenesis of subsequent amino acid in a protein segment. Moreover, ProCASTing can be applied to saturate amino acids with desired degenerate codons to incorporate specific type of amino acids in a protein segment (eg. VRV for charged amino acid substitutions; V: G, C or A, R: G or A). To our knowledge, this is the first sequence independent method for parallel saturation mutagenesis of subsequent amino acid residues.

6.2 Material and Methods

All chemicals used were of analytical reagent grade or higher quality and purchased from Sigma-Aldrich Chemie GmbH (Munich, Germany), Applichem GmbH (Darmstadt, Germany) and Carl Roth GmbH (Karlsruhe, Germany). A thermal cycler (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) and thin-wall PCR tubes (Mμlti-Ultra tubes; 0.2 ml; Carl Roth, Germany) were used in all PCRs. The

PCR volume was always 50 µl and the amount of DNA was quantified using NanoDrop photometer (PEQLAB Biotechnologie GmbH, Erlangen, Germany).

6.2.1 Selection of protein segment

The temperature factor reported in atomic resolution structures (PDB files) or structures obtained from MD simulation was used to measure local residue flexibility. The temperature factor reflects the degree of dispersal of atomic electron density around the equilibrium position due to thermal motion and positional disorder. The perl based script was developed (Supporting Information Figure S1) to identify the protein segments with higher flexibility in Ymphytase based on temperature factor. Ymphytase structure obtained from MD simulation (see section 5.2.5) containing temperature factor was used as an input for the script. The GROMACS software package [15] was used to extract the temperature factor in a pdb file during simulation time (10ns) and a protein segment (G307 – S315) was selected.

6.2.2 Primer Design

Two universal primers (F1PTO-F and F1PTO-R) were designed that can be used to most of the pET vector series. The complementary phosphothioated region of these primers was 12 nucleotides at 5' end. Primer NDT-F was designed in a way that it contain phosphothioated nucleotides (12, complementary to phosphothioated part of VecPTO1 primer) followed by degenerate nucleotides (NDT; N: A, T, G, or C, D: G, A, or T; 12 or 24) and finally standard nucleotides (annealing region with template, 12). All the phosphothioated primers possess phosphothioate bonds in 12 nucleotides at 5´ end. All primerss were purchased from Eurofins MWG Operon (Ebersberg, Germany) in salt-free form. Oligonucleotides were diluted in Milli-Q water to a final concentration of 100 µM. All oligonucleotides and primers used in this study are summarized in Table 1.

Two mutant libraries were generated for saturating four consecutive amino acids (4-NDT) and eight consecutive residues (8-NDT) using 4-NDT-F and 8-NDT-F primer respectively.

6.2.3 Step 1: Template generation

Template DNA fragments for step 2 were generated using PCR. The *pro*tein *seg*ment (ProSeg in Figure 1) targeted for saturation mutagenesis was deleted in step 1 by using VecA and VecB primers. Template-A was amplified from pALXtreme-5b-*YmappA* construct (50 ng) by PCR [98 °C for 30 sec, one cycle; 98 °C for 10 sec, 55 °C for 15 sec, 72 °C for 15 sec, 30 cycles; 72 °C for 180 sec] using forward F1PTO-F primer and reverse VecA primer (0.4 µM each), Phusion Hot Start DNA polymerase

Table 1: Sequences of the primers used for the ProCASTing library generation

Name	Length (bp)	Tm (°C)	Sequence (5´→ 3´)
F1PTO-F	27	60.2	gactcactatagGGGAATTGTGAGCGG
F1PTO-R	35	58.4	ctatagtgagtcGTATTAATTTCGAACATGTGAGC
VecA	25	63.5	CTTACGCTGAAGCACTAGCGCCGTC
VecB	25	58.1	GAGCAGACCAAACTCCTTTTCCTTG
VecPTO1	25	63.5	cttacgctgaagCACTAGCGCCGTC
4NDT-F	37	63	cttcagcgtaagNDTNDTNDTNDTGAGCAGACCAAAC 3'
8NDT-F	49	64.8	cttcagcgtaagNDTNDTNDTNDTNDTNDTNDTNDTGAGCAGACCAAAC

*T_m for degenerate primers have been reported as mean value; capital letter – normal nucleotides, small letter – phosphothioated nucleotides

(0.02 U; Finnzymes, Keilaranta, Finland) and dNTP mix (0.2 mM each). Template-B was also amplified from pALXtreme-5b-*YmappA* construct (50 ng) by PCR [98 °C for 30 sec, one cycle; 98 °C for 10 sec, 55 °C for 15 sec, 72 °C for 35 sec, 30 cycles; 72 °C for 180 sec] using forward VecB primer and reverse F1PTO-R primer (0.4 µM each), Phusion Hot Start DNA polymerase (0.02 U; Finnzymes, Keilaranta, Finland) and dNTP mix (0.2 mM each). Both the PCR products were column purified (NucleoSpin Extract II kit Macherey-nagel, Düren, Germany) and template plasmid DNA was digested overnight with FastDigest *DpnI* (1 FDU) enzyme at 37 °C.

6.2.4 Step 2: Fragment amplification with degenerated primers

The deleted protein segment in first step is replaced by degenerated codons in fragment-B using touchdown PCR. Fragment-A was amplified from template-A (50 ng) using PCR [98 °C for 30 sec, one cycle; 98 °C for 10 sec, 55 °C for 15 sec, 72 °C for 15 sec, 30 cycles; 72 °C for 180 sec] using forward F1PTO-F primer and reverse VecPTO1 primer (0.8 µM each), Phusion Hot Start DNA polymerase (0.02 U; Finnzymes, Keilaranta, Finland) and dNTP mix (0.2 mM each). The fragment-B was generated by touchdown PCR [98 °C for 120 sec, one cycle; 98 °C for 10 sec, 63 °C for 15 sec (-0.5 °C/cycle), 72 °C for 35 sec, 30 cycles; 72 °C for 180 sec] using forward 4-NDT-F or 8-NDT-F primer and reverse F1PTO-R primer (0.4 µM each), template-B (10 ng), Phusion Hot Start DNA polymerase (0.02 U; Finnzymes, Keilaranta, Finland), and dNTP mix (0.2 mM each). All the fragments were purified by gel extraction (NucleoSpin Extract II kit Macherey-nagel, Düren, Germany) and subsequently used for step 3.

6.2.5 Step 3: Iodine cleavage, ligation and transformation

DNA fragments were diluted with Milli-Q water to 0.01 pmol/µl (fragment-B) and 0.03 pmol/µl (fragment-A). Cleavage reaction was performed in thin walled PCR tubes (Mµlti-Ultra tubes; 0.2 ml; Carl Roth, Germany) containing 4 µl of DNA sample (fragments) and 1 µl of iodine digestion mix. The samples were incubated (70 °C for 5 min) in a PCR machine (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) and subsequently kept on ice until further use. The

iodine digestion mix was prepared by adding 3 µl of 100 mM iodine/ethanol solution in 5 µl of digestion buffer (0.5 M Tris-HCl, pH 9.0) and 2 µl of Milli-Q water.

Enzyme free ligation was achieved by mixing digested fragment-A (5 µl) with digested fragment-B (5 µl) and incubating the reaction mixture at room temperature for 5 min. The reaction mix was transformed directly into chemically competent *E. coli* BL21-Gold(DE3)laqI^{Q1} strain [16].

6.2.6 Colony PCR and DNA sequencing

Colony PCR of mutant clones was performed with slight modification in standard PCRs. The master mix of PCR containing F1 primer (5'-CGA CTC ACT ATA GGG GAA TTG TGA GCG GA-3') and R3 primer (5'-CGG GCT TTG TTA GCA GCC GGA TCT CAG-3') (0.4 µM each), Phusion Hot Start DNA polymerase (0.02 U; Finnzymes, Keilaranta, Finland), and dNTP mix (0.2 mM each) was prepared and dispensed (20 µl) in thin-walled PCR tubes. Colonies were directly transferred from LB-agar plate to PCR tubes and PCR [98 °C for 3 min, one cycle; 98 °C for 10 sec, 58 °C for 15 sec, 72 °C for 25 sec, 25 cycles; 72 °C for 180 sec] was performed and PCR products (5 µl) were subsequently analyzed on agarose-TAE gel. Few random clones from both 4-NDT and 8-NDT library were sequenced at GATC Biotech (Konstanz, Germany).

6.2.7 Screening of ProCASTing mutant libraries

The mutant libraries for 4-NDT and 8-NDT clones were prescreened using 384-well microtiter plate based screening assay for improved thermostability as described in section 4.2.5.

6.3 Results

6.3.1 General Concept of ProCASTing

ProCASTing method is based on phosphothioate nucleotide chemistry. ProCASTing is a sequence independent method that consist three steps (Figure 1). In the first step (Figure 1), PCRs are performed to amplify linear DNA templates (A and B) using primers either containing phosphothioate bonds or standard nucleotides. The protein segment targeted for saturation mutagenesis is deleted in this step from the gene-vector construct.

The second step includes an incorporation of degenerate codons (in protein segment) into the 5'-end of the DNA template-B. Fragment-B is synthesized using forward primer containing 12 phosphothioated nucleotides followed by degenerate codons and standard nucleotides and a reverse phosphothioated primer. Fragment-A is amplified using two set of phosphothioated primers. The phosphothioated primers are designed to replace phosphodiester bonds at 5´end with the phosphothioate bonds. Sites for saturation mutagenesis were incorporated between standard nucleotides and phosphothioated nucleotides in NDT-F primer. The forward primer (NDT-F) was designed in a way that the phosphothioated nucleotide region is complimentary to reverse primer of other DNA fragment (eg. 12 nucleotides at 5´ end of primer NDT-F are complimentary to 12 nucleotides at 5´ end of VecPTO1 primer in Figure 1). Similarly, 5'-end of primer set (F1PTO-F and

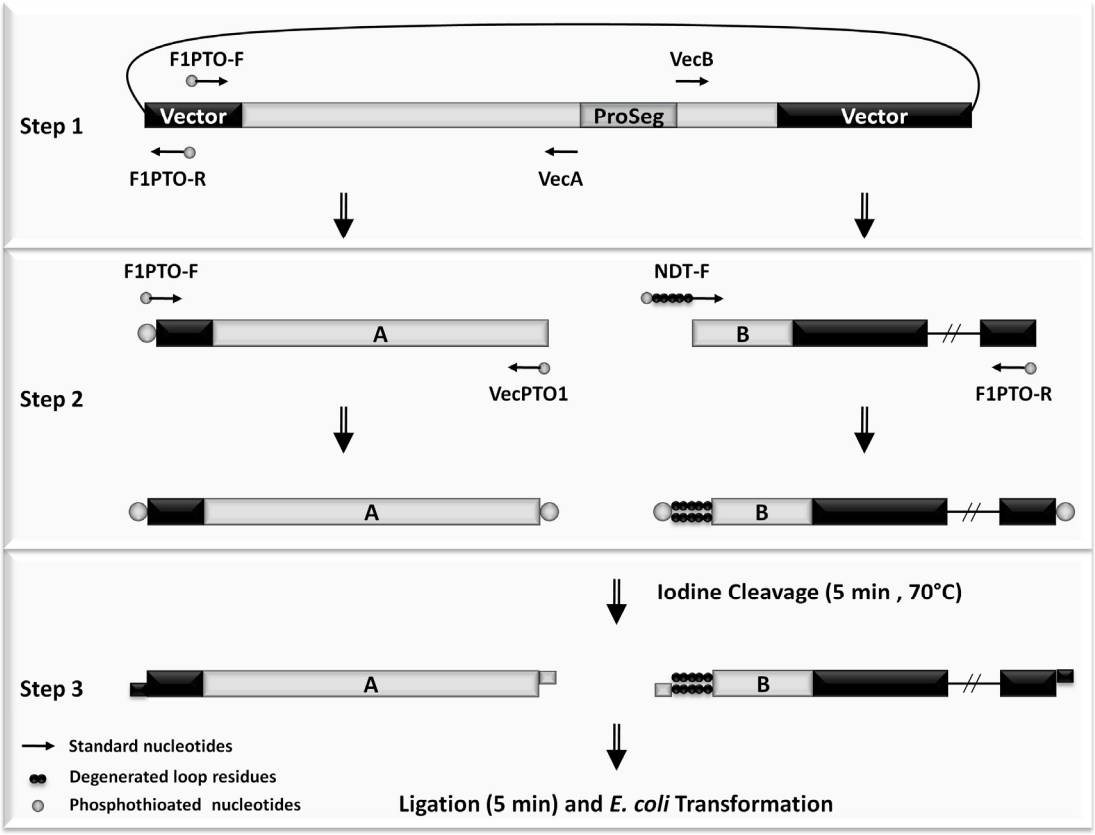


Figure 1: Schematic representation of the ProCASTing method that comprises three steps. Step 1: Template generation, Step 2: Generation of DNA fragments with phosphotioated primers, Step 3: Iodine cleavage, ligation and transformation.

F1PTO-R) used for fragment-A amplification is complimentary.

PCR amplification is performed with standard nucleotide so that the phosphorotioated nucleotides are incorporated in single strand of double helix DNA. Phophothioate bond cleavage is then performed in step 3 by iodine/ethanol solution at high temperature resulting in sticky end overhangs of 12 nucleotides that are complimentary to sticky ends in other DNA fragment. Subsequently, the enzyme-free ligation is performed by mixing both the DNA fragments at room temperature (5 min) and standard *E. coli* transformation is performed.

6.3.2 ProCASTing library generation and sequencing

In the first step, a PCR was performed to linearize the circular plasmid construct and the protein segment targeted for saturation was deleted. In the second step, PCR was performed to amplify the fragment-A and fragment-B (Figure 2). Fragment-A amplified using F1PTO-F and VecPTO1 primers showed a single band (1 kb) on agarose gel. Two types of fragment-B were synthesized using either 4-NDT-F or 8-NDT-F primer and F1PTO-R primer (2.3 kb). The 4-NDT-F and 8-NDT-F primers possess four and eight degenerate codons (NDT) respectively replacing the targeted protein

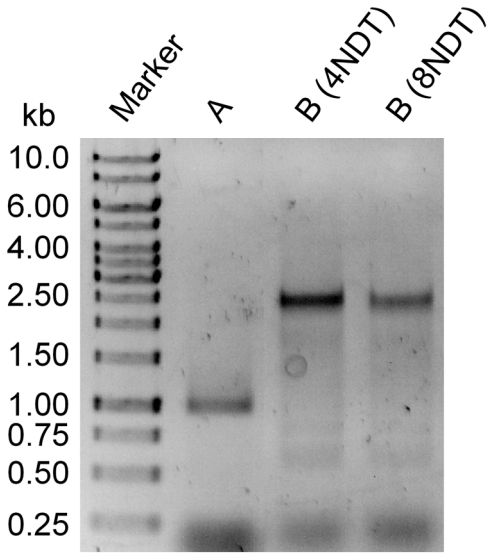


Figure 2: Agarose gel: PCR amplification of fragment-A (A) and fragment-B with 4NDT or 8NDT degeneracy.

segment. Same PCR conditions were used for amplification of 4-NDT and 8-NDT fragment-B. In the third step, fragment-A and fragment-B (4NDT or 8NDT) were digested, ligated and transformed.

Validation of ProCASTing method was performed using colony PCR and sequencing of few random clones from the mutant library. Colony PCR using vector specific primer (F1 and R3) showed presence of insert in the cloned plasmid (Supporting Information Figure S2). Four clones out of five in 4-NDT and five out of five clones in 8-NDT mutant library showed presence of insert that reveal 90 – 100% ligation efficiency. Further, four and five clones from 4-NDT and 8-NDT library respectively were sequenced. Sequencing of random clones validated the current protocol of ProCASTing, the target protein segment was successfully reduced in size and four subsequent amino acids were saturated. In addition, amino acid residues in the complete protein segment (8 amino acids) were replaced with other amino acids (Figure 3). All the five sequenced clones from 8-NDT library showed no nucleotide deletion in the loop region, whereas, in case of 4-NDT library one clone showed a single nucleotide deletion in third amino acid position. No significant bias of amino acid substitution pattern was observed in the sequenced clones from both the libraries.

6.3.3 Screening of ProCASTing mutant library

In the first round of mutant library screening, approximately, 1125 ProCASTing mutant clones from each 4-NDT and 8-NDT library were prescreened using a high throughput screening method based on 384-well microtiter plate. In the prescreening for thermostability, using 384-well plates, mutants showed high rate of enzyme inactivation. In case of 8-NDT ProCASTing library, only 25% of the mutants showed activity after heat incubation. However, 4-NDT ProCASTing library showed 35% of active clones. In both cases, the activity of Ymphytase was drastically affected, most of the variant showed activity between 0 – 15% relative activity compare to wild-type Ymphytase. Further, about 1900 mutant clones from 4-NDT ProCASTing library were further screened for improved thermostability. Unfortunately, after purification of three hits from the screening did not show any improvement in thermostability compare to wild-type Ymphytase.

6.4 Discussions

In order to validate the ProCASTing method is working we have generated a diversity in the protein segment. The protein segment targeted

a)

Q R K H R S G V L N D E Q T
CAG CGT AAG CAT CGT AGT GGT GTT CTT AAT GAT GAG CAG ACC

Q R K Y H F H R S G G E Q T
CAG CGT AAG TAT CAT TTT CAT CGT AGT GGT GGT GAG CAG ACC

Q R K F N D G D L L C E Q T
CAG CGT AAG TTT AAT GAT GGT GAT CTT CTT TGT GAG CAG ACC

Q R K V V H N F L I N E Q T
CAG CGT AAG GTT GTT CAT AAT TTT CTT ATT AAT GAG CAG ACC

Q R K F D L D S N I C E Q T
CAG CGT AAG TTT GAT CTT GAT AGT AAT ATT TGT GAG CAG ACC

Q R K G F G V N S G G E Q T
CAG CGT AAG GGT TTT GGT GTT AAT AGT GGT GGT GAG CAG ACC

b)

Q R K R N Y G E Q T
CAG CGT AAG CGT AAT TAT GGT GAG CAG ACC

Q R K R H X G E Q T
CAG CGT AAG CGT CAT –TT GGT GAG CAG ACC

Q R K N S H N E Q T
CAG CGT AAG AAT AGT CAT AAT GAG CAG ACC

Q R K I I L C E Q T
CAG CGT AAG ATT ATT CTT TGT GAG CAG ACC

Figure 3: Multiple sequence alignment of sequenced clones from 8-NDT (a) and 4-NDT (b) ProCASTing mutant library. Translated amino acid for each codon is shown on top of each codon.

for mutagenesis was successfully replaced with other amino acid residues. Deletion of this highly flexible protein segment identified by Perl script was improving thermostability with drastically affecting activity of the Ymphytase (Supporting Information Figure S3). The targeted segment of 8 residues was easily reduced to 4 amino acids with substitutions at each position and saturated to posses 8 subsequent amino acid substitutions, reveled by sequencing (Figure 3). Codon degeneracy of NDT (N: A, T, G or C; D: A, T or G) was used to reduce the screening efforts [17]. NDT degenerate codon codes for twelve amino acids covering all chemically different amino acids (no stop codons) compare to NNK degeneracy that codes for all twenty amino acids including a stop codon. Statistical analysis using GLUE-IT program [18] showed screening of 10,000 clones was able to cover 0.38 expected amino acid completeness with four NDT codons (completeness = 1 is all possible amino acid substitutions in given sites), whereas, with NNK degeneracy, only 0.05 expected amino acid completeness was predicted.

First round of screening revealed that the activity ratio in 4-NDT ProCASTing library was higher than 8-NDT library after heat treatment. Therefore, further mutant clones from 4-NDT ProCASTing library were screened. The activity of most of the variants were drastically affected compare to wild-type activity and screening of only few thousand clones did not improve thermostability that might be due to negligible

expected amino acid completeness of the screened library (4-NDT, 1000 clones, 0.05 completeness).

Perl based script was developed to identify highly flexible protein segment during molecular dynamics simulations or protein crystallization (Supporting Information Figure S1). A program B-FITTER [14] was developed to identify the flexible regions in protein crystal structure. However, B-FITTER program cannot be applied to structures obtained by molecular dynamics simulations. The flexibility in the protein structure is a key feature of protein stability. Protein stability was intensively studied and reduced flexibility is found to be one of the key features in improving stability(Chapter 5 and [19]), making the protein more rigid so that unfolding in extreme environment is less favored.

The ProCASTing method is technically simple and sequence independent that can be used to insert, delete or saturate any segment of a protein molecule. The standard site directed mutagenesis protocol can be used to saturate the few subsequent amino acids that is limited by a binding bias of the primers. The primers with wild-type nucleotides will bind preferentially compare to degenerate nucleotides. ProCASTing method was developed to overcome the limitation of binding bias making it sequence independent method. Insertion or deletion of amino acid stretch can easily be achieved with ProCASTing due to its robustness and simplicity.

The ProCASTing libraries with long stretch of protein segment will be suitable for ultrahigh throughput screening systems based on fluorescent activated cell sorting [20] or microfluidics [21] to cover a significant fraction of sequence space. ProCASTing method can be used to selectively saturate few important positions in the same protein segment that will reduce the screening efforts.

6.5 Conclusions

In summary, we have developed ProCASTing, a novel method for directed evolution that can be used for insertion, deletion or saturation mutagenesis of any segment of an enzyme. The Perl based script was developed to assist experimentalist for accurate selection of flexible regions in the protein structure extracted from MD simulation or crystal structure. ProCASTing can be used to generate large mutant libraries with low empty vector background for high throughput screening due to its robustness, reliability and simplicity. As a proof of principle, we achieved saturation mutagenesis of four or eight subsequent site in Ymphytase enzyme. The ProCASTing method can easily be applied to active site modification by parallel site saturation mutagenesis of more than one amino acid to broaden our understanding of structure-function relationship on enzyme reactions and promiscuity.

6.7 References

1. Reetz, M.T., et al., *Directed evolution of an enantioselective epoxide hydrolase: uncovering the source of enantioselectivity at each evolutionary stage.* J Am Chem Soc, 2009. 131(21): p. 7334-43.
2. Zhu, L., et al., *Directed evolution of an antitumor drug (arginine deiminase PpADI) for increased activity at physiological pH.* Chembiochem, 2010. 11(5): p. 691-7.
3. Kim, M.S. and X.G. Lei, *Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR.* Appl Microbiol Biotechnol, 2008. 79(1): p. 69-75.
4. Matsuura, T. and T. Yomo, *In vitro evolution of proteins.* J Biosci Bioeng., 2006. 101(6): p. 449-56.

5. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Curr Opin Chem Biol, 2009. 13(1): p. 19-25.
6. Bershtein, S. and D.S. Tawfik, *Advances in laboratory evolution of enzymes.* Curr Opin Chem Biol, 2008. 12(2): p. 151-8.
7. Wong, T.S., D. Zhurina, and U. Schwaneberg, *The diversity challenge in directed protein evolution.* Comb Chem High Throughput Screen, 2006. 9(4): p. 271-88.
8. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Challenges of the genetic code for exploring sequence space in directed protein evolution.* Biocatal Biotransfor, 2007. 25(2): p. 229 - 41.
9. Wagner, C.R. and S.J. Benkovic, *Site directed mutagenesis: a tool for enzyme mechanism dissection.* Trends Biotechnol, 1990. 8(9): p. 263-70.
10. Alam, M., D.E. Vance, and R. Lehner, *Structure-function analysis of human triacylglycerol hydrolase by site-directed mutagenesis: identification of the catalytic triad and a glycosylation site.* Biochemistry, 2002. 41(21): p. 6679-87.
11. Yuen, C.M. and D.R. Liu, *Dissecting protein structure and function using directed evolution.* Nat Methods, 2007. 4(12): p. 995-7.
12. Padhi, S.K., D.J. Bougioukou, and J.D. Stewart, *Site-saturation mutagenesis of tryptophan 116 of Saccharomyces pastorianus old yellow enzyme uncovers stereocomplementary variants.* J Am Chem Soc, 2009. 131(9): p. 3271-80.
13. Fox, R.J., et al., *Improving catalytic function by ProSAR-driven enzyme evolution.* Nat Biotechnol, 2007. 25(3): p. 338-44.
14. Reetz, M.T. and J.D. Carballeira, *Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes.* Nat Protoc, 2007. 2(4): p. 891-903.
15. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. 26(16): p. 1701-18.
16. Blanusa, M., et al., *Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method.* Anal Biochem, 2010. 406(2): p. 141-6.
17. Reetz, M.T., D. Kahakeaw, and R. Lohmer, *Addressing the numbers problem in directed evolution.* Chembiochem, 2008. 9(11): p. 1797-804.
18. Firth, A.E. and W.M. Patrick, *GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries.* Nucleic Acids Res, 2008. 36(Web Server issue): p. W281-5.
19. Vihinen, M., *Relationship of protein flexibility to thermostability.* Protein Eng, 1987. 1(6): p. 477-80.
20. Yang, G. and S.G. Withers, *Ultrahigh-throughput FACS-based screening for directed enzyme evolution.* Chembiochem, 2009. 10(17): p. 2704-15.
21. Agresti, J.J., et al., *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution.* Proc Natl Acad Sci U S A, 2010. 107(9): p. 4004-9.

Supporting Informaton: Chapter 6

```
#!/usr/bin/perl

#print "Enter the path of pdb file:";
print "Enter the path of PDB file:\n";
#contain the path of output file
$in=<STDIN>;
chomp $in;
open IN,"< $in" or die "error";
#print "Enter the path of output file:";
print "Enter the path of output file:\n";
#contain the path of output file
$out=<STDIN>;
chomp $out;

open OUT,">$out" or die "error";

#open IN,"test.pdb" or die "error";
@ar=<IN>;
#To read the file line by line
my $c;
my @c;
my $parse;
foreach $ar(@ar)
{
#extract coordinates
if($ar=~m/^ATOM/g)
{
    $bfact=substr($ar,60,6);
#$bfact=~s/\s+//g;
#@aa=split(/\s/,$ar);

#print "$bfact\n";
    $atomname = substr($ar,12,3);
    $resname = substr($ar,17,3);
    $resnum = substr($ar,22,4);
    #$chain = substr($ar,21,1);
    $parse="$atomname\t$resname\t$resnum\t$bfact";
    $parse=~s/^\s+//g;
push (@beta,$parse);
#sprint $parse;
}
}
my $ar1;
my $ar2;
my $total;
my $count;
my $T;
#print
for($i=0;$i<scalar @beta;$i++)
{
@ar1=split("\t",$beta[$i]);
$count=0;
if($ar1[0]=~/\bN\b/)
{
for($j=$i;$j<scalar @beta;$j++)
{
#print "$beta[$i]####$beta[$j]\n";
@ar2=split("\t",$beta[$j]);
#print "$ar1[3]\t$ar2[3]\n";
if ($ar1[2] == $ar2[2])
{
$total+=$ar2[3];
$count++;
#$tc = "$ar1[1]\t$ar1[2]\t$ar1[3]\t$total $count";
}
next if($ar1[2] != $ar2[2]);
#print "$tc\n";
}
$tc = "$ar1[1]\t$ar1[2]";
#$final="$tc##\n";
$avg=$total/$count;
$avg1=sprintf("%.2f",$avg);
if (length $avg1 <= 4)
{
$avg2= " $avg1";
$c="$tc\t$avg2\n";
push(@c,$c);
}
else
{
$c="$tc\t$avg1\n";
push(@c,$c);
}
$total=0;
#print "$ar1[$i]\n";
}
}
#decending order
@sort_ar = reverse sort {substr($a,-7) <=> substr($b,-7)} @c;
foreach $sort(@sort_ar)
{
print OUT "$sort";
}
```

Figure S1: A perl script used to identify the felxible part of the Ymphytase structure based on temperature factor obtained from MD simulation.

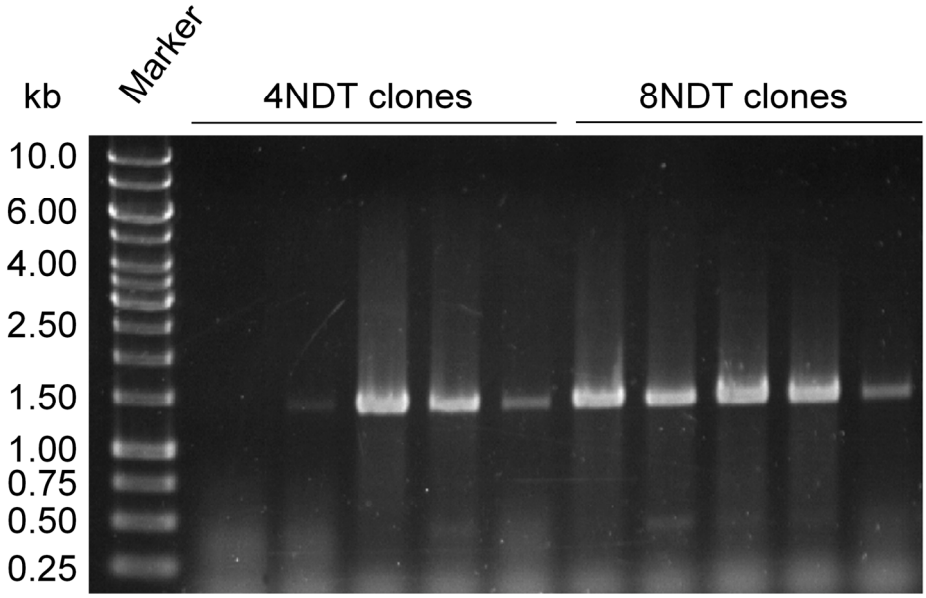


Figure S2: Agarose gel: Colony PCR showing the amplification of insert in randomly picked clones from 4-NDT and 8-NDT ProCASTing mutant library.

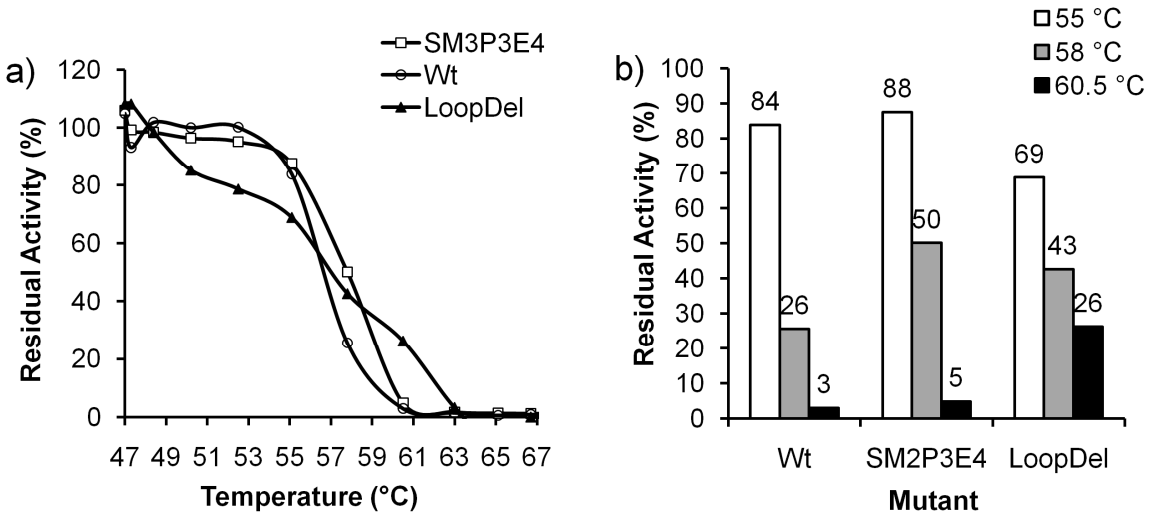


Figure S3: Residual enzyme activity of Ymphytase wild-type (Wt), SM2P3E4 (See Chapter 4) and LoopDel (Δ G307 – S315) variants after being incubated at gradient temperatures (47 to 67 °C) for 10 min (a/b).

Multisite combinatorial assembly of site saturation (OmniCASTing): a novel method for directed evolution

Amol V. Shivange[a,b], Alexander Dennig[a], and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Multisite combinatorial assembly of site saturation test (OmniCASTing) is a conceptually novel method for parallel site saturation of more than one site. Despite advances in few methods for multiple site directed mutagenesis, most of the methods suffer from tedious and long procedures. OmniCASTing can be accomplished within one day and comprises three simple steps: generating DNA fragments using phosphothioated primers, chemical cleavage of phosphothioate bonds in an iodine/ethanol solution, stepwise ligation, and transformation. OmniCASTing is absolutely sequence independent method and any combination of selected sites for saturation is possible. Additionally, OmniCASTing protocol does not require DNA purification step and ligation can be accomplished within 30 min (5 min/fragment pair) making it robust and reliable. Using *Yersinia mollaretii* phytase enzyme as a model system, we proved by sequencing 48 genes the concept of the OmniCASTing method and achieved 32% improvement in thermostability (58° for 20 min) and 2 °C increase in melting temperature compare to wild-type without loss of specific activity. Interestingly, the pH stability was improved by 3 fold at pH 2.8 compare to parent mutant combination. *Y. mollaretii* phytase showed a positive cooperativity (Hill coefficient was 2.4) and a specific activity of 1034 U/mg that is ~10 times higher than the commercial phytase like *Aspergillus niger* phytase.

7.1 Introduction

The field of white biotechnology is enduring rapid diversification. Properly configured enzymes have an outstanding capacity for carrying out complex molecular reactions with extreme precision and efficiency such as activity, stability, enantio-selectivity or enantio-specificity, substrate selectivity, pH range [1-3]. Directed evolution has over the last decades become successful tool for tailoring proteins and exploring structure-function relationship. It consists of four major steps that comprise random gene mutagenesis (diversity generation), cloning of variants, screening (or selecting) for improved variants (best hit) for

*Corresponding author

desired functional property and isolation of gene for next iterative cycle of directed evolution. The process is repeated until the desired level of improvement has been achieved [4, 5]. It is an *in-vitro* Darwinian evolution that does not require prior thorough knowledge of protein structure, making it robust in contrast to other approaches of protein engineering such as rational protein design and *de novo* design. Most of the directed evolution experiment results in few amino acid changes in the gene that leads to improvement. Site-directed mutagenesis has been enormously used to study structure-function relationship by mutating the residues identified by directed evolution or by rational design [6-8].

In addition, a method of site-saturation mutagenesis has proven to be a promising tool to explore the position in the enzyme to understand the importance of the specific site [9] and identify the best fit residue. Several success stories have been reported using these methods to identify beneficial, neutral and deleterious mutation influencing the enzyme property [10] and improving enzyme property by iteratively combining mutations [11]. However, all these methods suffer from a limitation of exploring the effect of one mutation on other mutations, a cooperative effect between these mutations. These cooperative effects can be responsible for a synergetic effect arising from a specific combination of mutations. In last decade, few methods for multiple site-directed mutagenesis were developed. Most of the methods for multiple site-specific mutations suffer from tedious intermediate agarose gel or PAGE purification procedures [12, 13], several primers annealing at once making it sequence dependant [14] and modification of overlap extension PCR to perform several PCRs of amplified DNA fragment that needs PCR optimization for fusion of each DNA fragment [15]. In this study, we have developed OmniCASTing, a novel and technically simple method of saturation mutagenesis of multiple positions that overcomes a limitation of exploring the cooperative effects of mutations. OmniCASTing can be applied to saturate amino acids with desired degenerate codons to incorporate specific type of amino acids (eg. VRV for charged amino acid substitutions; V: G, C or A, R: G or A). The concept of OmniCASTing has been applied to an already identified thermostable *Yersinia mollaretii* phytase (Ymphytase) enzyme variant with five substitutions. Subsequently, screening of OmniCASTing library of Ymphytase resulted in improvement of three enzyme properties that are thermostability, specific activity and pH stability with new combinations of mutations in Ymphytase variant.

7.2 Material and Methods

All chemicals used were of analytical reagent grade or higher quality and purchased from Sigma-Aldrich Chemie GmbH (Munich, Germany), Applichem GmbH (Darmstadt, Germany) and Carl Roth GmbH (Karlsruhe, Germany). A thermal cycler (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) and thin-wall PCR tubes (Mµlti-Ultra tubes; 0.2 ml; Carl Roth, Germany) were used in all PCRs. The amount of DNA was quantified using NanoDrop photometer (PEQLAB Biotechnologie GmbH, Erlangen, Germany). All oligonucleotides were

purchased from Eurofins MWG Operon (Ebersberg, Germany) in salt-free form. Oligonucleotides were diluted in Milli-Q water to a final concentration of 100 µM. All oligonucleotides and primers used in this study are summarized in Table 1.

7.2.1 Fragment amplification by PCR

Plasmid isolation was accomplished by using QIAGEN QIAprep Spin Miniprep Kit (Hilden, Germany). A PCR set up contained 1x *Pfu*S buffer, dNTP (0.2 mM each), 0.4 µM fragment specific forward and reverse primer, 5 U *PfuS* DNA polymerase and 20 ng of DNA template in a final volume of 50 µl. All PCR amplifications were carried out in Eppendorf Mastercycler® proS PCR cycler (Hamburg, Germany) and in thin-walled PCR tubes (Sarstedt, Nuembrecht, Germany). The extension reaction for all fragments was initiated by pre-heating sample mixtures to 94°C for 3 min followed by 25 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 30 s (fragments A, B, C, and D) or 1.5 min (vector fragment). For final elongation one cycle was performed at 72 °C for 1 min (fragment A, B, C, and D) or 3 min (vector fragment). PCR products were separated on 2% (w/v) Tris-acetate-EDTA (TAE) agarose gel to confirm the size of the respective PCR products. Specific degradation of methylated template DNA was performed by adding 1 µl (20 U) of *Dpn*I to each PCR sample. After incubation at 37°C for 4 h, PCR products were purified using the Macherey-Nagel NucleoSpin Extract-II DNA purification kit (Dueren, Germany) and eluted in 50 µl Milli-Q water. Purified PCR products were quantified using a Nanodrop spectrophotometer.

7.2.2 Iodine treatment, ligation, and transformation

All steps were performed on ice unless stated otherwise. Purified vector fragment and fragments A, B, C, and D were diluted to 0.01 pmol/µl or 0.11 pmol/µl respectively using Milli-Q water. The PCR-products were cleaved by iodine at alkaline conditions as described [16] with few modifications. Briefly, 4 µl of each PCR fragment was combined with 2 µl iodine cleavage buffer [1 µl of 0.5 M Tris-HCl (pH 9.0); 0.6 µl of 100 mM iodine in 99% ethanol; 0.2 µl Milli-Q water] in PCR tubes followed by incubation at 70°C for 5 min in a thermocycler. Cleaved DNA samples were kept on ice for additional 5 min. Annealing of cleaved fragments was performed gradually by first adding 6 µl fragment A cleavage mix to 6 µl of vector fragment cleavage mix. Hybridization of both fragments was performed at room temperature (RT) for 5 min. Subsequently, cleavage mixtures of fragments B, C, and D were gradually added in the same way until entire plasmid DNA was combined. The mixture was kept on ice for further 5 min. A vector fragment self-hybridization control that would also result in ampicillin-resistance, consisted of 6 µl iodine cleaved vector fragment, gradually combined with 4 times of 4 µl iodine cleavage buffer and 2 µl Milli-Q water. Finally, 4 µl of both hybridization set ups were transformed with no further purification step into 100 µl chemical competent *E. coli* BL21-Gold (DE3) lacI^{Q1} cells via heat shock following a standard protocol [17]. Transformation mixtures were spread on agar plates with Luria Bertani (LB) medium supplemented with 100 mg/ml ampicillin and incubated over night.

Table 1: Sequences of the primers used for the OmniCASTing libraries using NNK degeneracy

Oligoname	Sequence	% GC	T_m (°C)*
F1PTO_AS_F	gactcactatagGGGAATTGTGAGCGG	51.9	60.2
F1PTO_AS_R	ctatagtgagtcGTATTAATTTCGAACATGTGAGC	37.1	58.4
YmD52_PTO_F	ccggataagtggCCTCAATGGCCGGTAC	60.7	65.8 59.8
YmD52_PTO_R	ccacttatccggTGTGACMNNATTCATTAACTCTG	44.3	61.7 63.9
YmT77_PTO_F	ggcggcttctatgGTGATTATTTCCG	50.0	59.7 64.3
YmT77_PTO_R	catagaagccgccCATCAGMNNCACTAATTGCGC	54.4	66.7 68.9
YmK139_PTO_F	gatccactgtttcACCCCGTCGAAG	56.0	61.3 60.4
YmK139_PTO_R	gaaacagtggatcAACCTTMNNCAAATCAGCCTG	45.6	62.7 64.9
YmG187_PTO_F	ctcaatttcactgCTTCCCCCTATTGC	48.1	59.8 60.4
YmG187_PTO_R	cagtgaaattgagAATCTCMNNCATCTGGGCAAATG	43.1	62.5 64.4
YmV298_PTO_F	ctagtgcttcagCGTAAGGGGCAAG	56.0	61.4 63.3
YmV298_PTO_R	ctgaagcactagCGCCGTMNNAATCTGTTGCAAC	51.5	65.3 67.6

*T_m for degenerate primers have been reported as minimum, mean and maximum value; capital letter – normal nucleotides, small letter – phosphothioated nucleotides

7.2.3 Colony PCR and DNA sequencing

Colony PCRs were performed to verify the correct reassembling of mutant plasmids. Therefore, 20 colonies were randomly picked into liquid LB media containing 100 µg/ml ampicillin. Cells were grown overnight at 37°C and 250 rpm. From liquid cultures 10 µl cells were lysed in 50 µl Milli-Q water at 95°C for 15 min. After centrifugation at 16,000 rpm for 2 min, 5 µl of the resulting supernatant served as template DNA for colony PCR-amplifications. Phosphatase gene amplification was performed in a final volume of 35 µl. The PCR set up contained 1x *Taq* DNA polymerase buffer, 0.2 mM dNTP mix, 0.4 µM forward (F1: 5´-CGA CTC ACT ATA GGG GAA TTG TGA GCG GA-3´) and reverse (R3: 5´-CGG GCT TTG TTA GCA GCC GGA TCT CAG-3´) primers, 5 U *Taq* DNA polymerase and 5 µl of centrifuged cell lysate. Colony PCR was carried out using the same cycling program as for vector fragment amplification. Only the elongation time in each cycle was increased to 1 min and in final elongation step to 3 min. PCR samples were separated on 1% (w/v) TAE agarose gels to verify the size of each PCR product. Sequencing was performed at GATC Biotech (Konstanz, Germany) and sequence analysis was done by

using Clone Manager 9 Professional Edition software (Scientific & Educational Software, Cary, NC, USA).

7.2.4 Prescreening of OmniCASTing mutant libraries

A high throughput prescreening was performed as described in section 4.2.5 using 384-well microtiter plates. Relative activity for each mutant was defined by considering the average fluorescence (a time point in the linear range) or slope (increased fluorescence over time) of three different wt clones as 100% activity. Clones exceeding more than 20% residual activity were selected for screening.

7.2.5 Protein expression and screening in 96-well plates

Screening of OmniCASTing mutant library was performed as described in section 4.2.7 using fluorescent based assay.

7.2.6 Characterization of wild-type Ymphytase and mutant

Ymphytase wild-type and mutant were purified as described in section 4.2.9 and characterized for thermostability and pH stability.

pH profile and stability

The pH profile and stability of wild-type and mutant phytase was determined at 37 °C using four different buffers: 0.25 M Glycine-HCL buffer for pH 2.0 to 3.2, 0.25 M sodium acetate buffer for pH 3.6 to 5.6, 0.25 M imidazole-HCL buffer for pH 6.0 to 7.0 and 0.25 M Tris-HCL for pH 7.4 to 9.0. Purified enzymes were diluted in specified buffers to 40 ng protein per milliliter and assay was carried out with 1mM phytate with corresponding pH, for pH stability phytases were incubated at 37 °C for 3 hours and residual activity was measured using phytase absorbance assay. Non pH treated enzyme activity was considered as 100% activity.

Determination of kinetic parameters

The kinetic parameters for wild-type and Omni1 variant were determined as described in section 4.2.9. The initial velocity data obtained were fitted to equation

$$Y = V\max * X\wedge h / (K_{half}\wedge h + X\wedge h)$$

where Y is the initial activity, Vmax is the maximum velocity, X is the substrate concentration and K_{half} the concentration of substrate that will afford half maximal velocity (see reference [18]) by using GraphPad Prism software (GraphPad software, San Diego, CA, USA). The K_{cat} was calculated from the ratio of Vmax and enzyme concentration. One enzyme unit (U) is defined as the amount of enzyme required to liberate 1 µmol of inorganic phosphate in 1 min.

7.3 Results

7.3.1 General Concept of OmniCASTing

OmniCASTing method is based on phosphothioate nucleotide chemistry. In the first step (Figure 1), PCRs are performed to amplify each DNA fragments (A, B, C, D and vector) using two primers containing 12 – 13 phosphothioate bonds. These primers are designed to replace phosphodiester bonds at 5´end with the phosphothioate bonds. Sites for saturation (MNN that is NNK in direct strand) were incorporated between normal nucleotides and phosphothioated nucleotides in a reverse primer. The forward primer was designed in a way that the phosphothioated nucleotide region

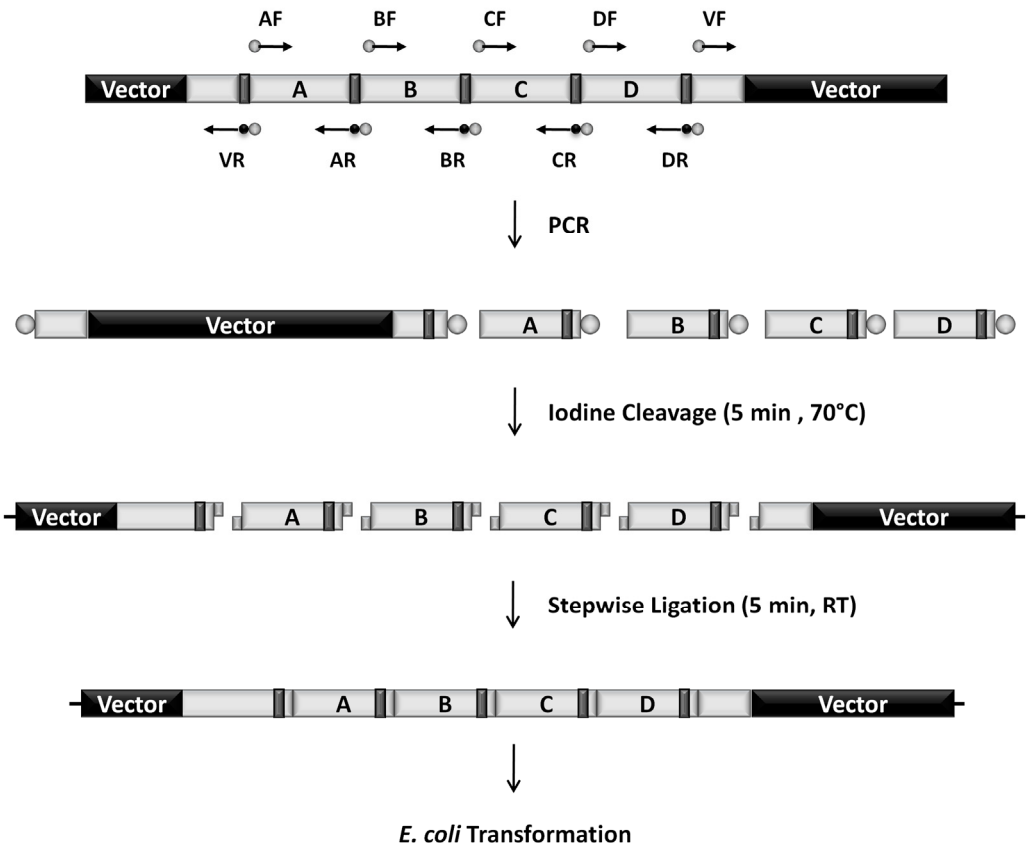


Figure 1: Schematic representation of the OmniCASTing method that comprises three steps. Step 1: Generation of DNA fragments with phosphotioated primers. Step 2: Iodine cleavage. Step 3: Stepwise ligation.

is complimentary to reverse primer of adjacent fragment (eg. 5´ end of primer AR is complimentary to 5´ end of primer BF in Figure 1). PCR amplification was performed with standard nucleotide so that the phosphorotioated nucleotides are incorporated in single strand of double helix DNA. Phophothioate bond cleavage is then performed by iodine/ethanol solution at high themperature resulting in sticky end overhangs of 12 to 13 nucleotides that are complimentary to sticky ends in other DNA fragment. The next step is a stepwise DNA hybridization (ligation) of fragments at room temperature (5 min/fragment pair) and standard *E. coli* transformation.

7.3.2 Fragment amplification by PCR

The fragments amplification of PCR was analyzed on the agarose gel that showed a correct size single band for each fragment of desired size. Further sequencing of 48 clones revealed that all the sites were saturated and except site Asp 52, all sites possess chemically different amino acids.

7.3.3 Screening of OmniCASTing mutant library

Approximately, 1100 OmniCASTing mutant clones were prescreened using a high throughput screening method based on 384-well microtiter plate. In the prescreening for thermostability using 384-well plates, mutants showing more than 80% residual activity (62 °C for 20 min) than Ymphytase wild-type were selected. Subsequently, selected 227 clones were screened in 96-well microtiter plate based screening system. Considering activity and thermostability, 19 clones were rescreened in test tube and mutant P17D8 (Omni1) was selected for further purification and characterization.

7.3.4 Improving thermostability, activity and pH stability

Several mutant with increased or descreased thermostability were identified however, the most promising Ymphytase mutant Omni1 showed improved heat stability compared to wild-type and mutant M1 (Chapter 4). Ymphytase mutant M1 had five substitutions include D52N, T77K, K139E, G187S, and V298M. OmniCASTing of these five sites resulted in Ymphytase mutant Omni1 harboring four substitutions D52E, K139T, G187S, and V298F. The Omni1 mutant displayed a residual activity with 32% improvement over that of wild-type enzyme after being heated at 58 °C for 20 min (Figure 2) and the melting temperature determination showed the mid-point of the thermal unfolding was increased by 2 °C. A new combination of mutation in Omni1 mutant also showed improved activity (~40 U/mg) compare to M1. Hill coefficient (h) value for wild-type, and Omni1 mutant enzyme was 2.24 and 2.44 that indicate positive cooperativity (Table 2). The K_{half} for wild-type and Omni1 mutant were 268 µM and 251 µM respectively. A threefold improvement in pH stability of Ymphytase Omni1 was observed at pH 2.8 compare to M1 (Figure 2).

7.4 Discussions

In order to validate the OmniCASTing method is working we have sequenced 48 clones that show parallel site saturation of all five selected sites. Sequencing only active clones (48) reveal that the site Asp 52 was not saturated completely. This might be due to its location close to active site loop (Figure 3) that is proposed to play an important role in activity of phytases [19].

Table 2: Kinetic parameters and thermostbility of YmPhytase

Enzyme	Specific activity (U mg^{-1})	K_{half} (µM)	*K_{cat} (S^{-1})	Hills coefficient (h)	T_{50} (°C) improvement	T_m (°C) improvement	Ref
Wt	1043.5 (± 26.5)	268.1 (± 18.2)	823.47 (± 20.91)	2.24 (± 0.12)	-	-	This study
M1	993.25 (± 3)	262.25 (± 0.65)	783.81 (± 2.33)	2.3 (± 0.002)	1.5	1.5	Chapter 4
M2	1034.25 (± 39.75)	250.75 (± 9.65)	816.17 (± 31.37)	2.44 (± 0.09)	2	2	This study

Phytase kinetics was determined in 0.25 M sodium acetate containing different sodium phytate (7.8 to 750 µM) at 37°C and pH 4.5 for both the phytases. Values in the bracket represent standard error.
*Catalytic center activity: enzyme concentration (moles) multiplied by the number of active centers (four).

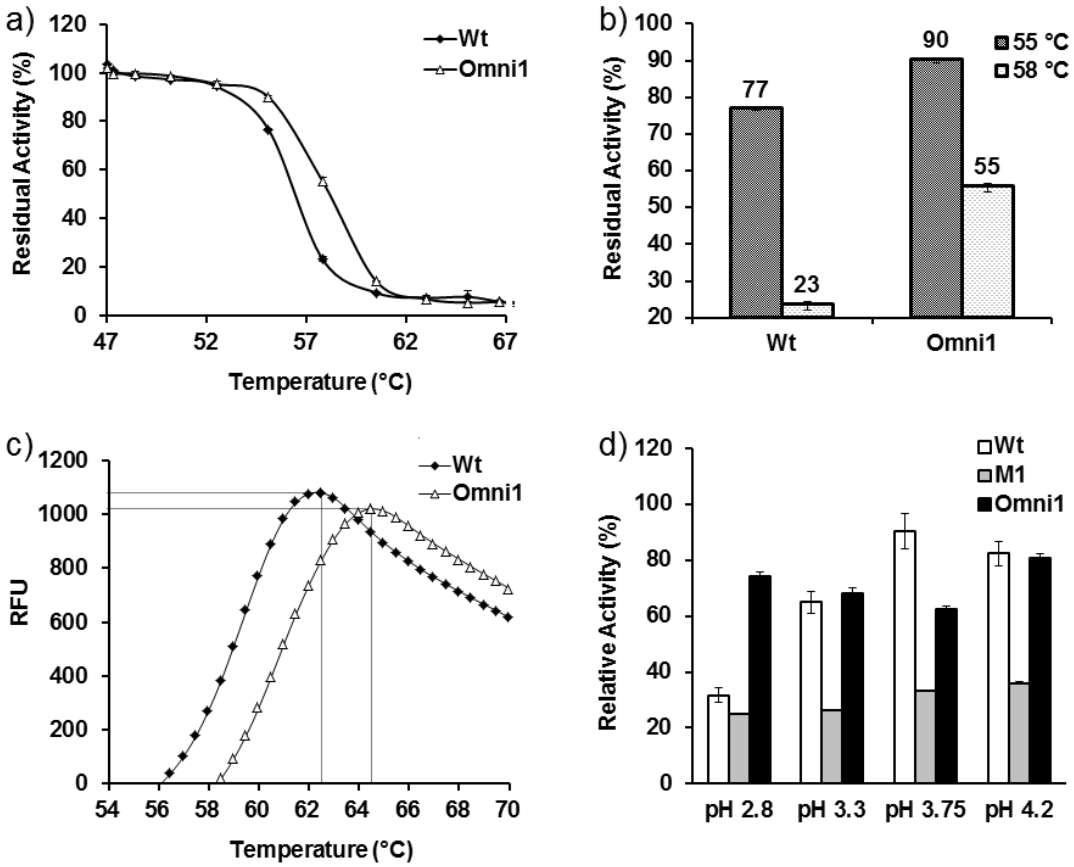


Figure 2: Residual enzyme activity of wild-type Ymphytase and Omni1 variants after being heated at gradient temperatures (47 to 67 °C) for 20 min (a/b). Melting temperature (T_m) of wild-type (Wt) and Omni1 variant as determined using thermal shift assay (c). Phytase pH stability was determined after incubating enzymes at different pH for 3 hour at 37°C (d).

Most of the substitutions were chemically similar substitutions that reflect that the position Asp 52 in Ymphytase might be important for activity. Additionally, the identified mutant Omni1 showed to posses D52E substitution. The OmniCASTing mutant library was enriched in thermostable mutant compare to wild-type mutant library. Screening of limited amount of clones (~1100) resulted in higher percentage of mutant withstanding elevated temperature. Best combination in ~1100 Ymphytase mutants was Omni1 mutant that had four substitutions (Figure 3).

The thermostability of Ymphytase Omni1 mutant showed to improve by 2°C whereas M1 showed 1.5 °C (Chapter 4). A new combination of mutations identified by OmniCASTing showed no drastic improvement. However, improvement in three properties (activity, thermostability and pH stability) was achieved compare to parent combination of mutations discovered by directed evolution. D52 was changed to chemically similar amino acid (D52E). Improved in Ymphytase thermostability might be due to improved intra-protein interactions. G187S

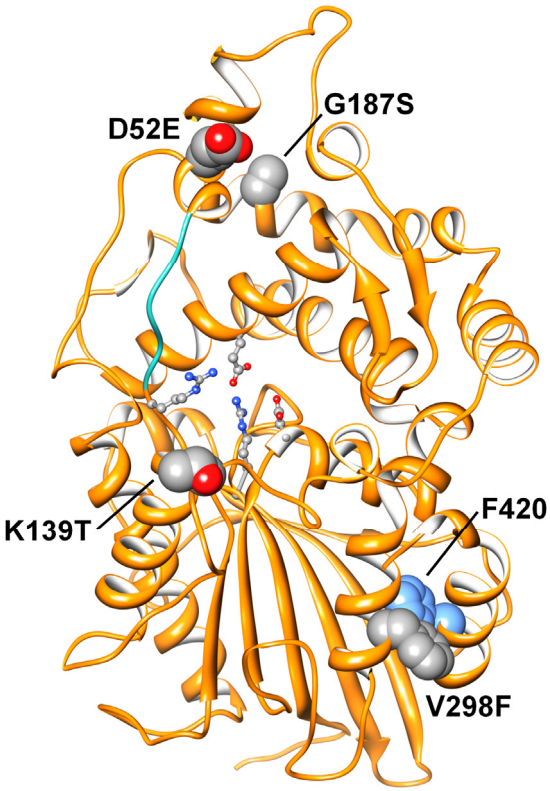


Figure 3: The amino acid substitutions in Ymphytase Omni1 variant are shown in space-filling model colored by element; F420 is shown in blue color. Active site loop (Ser 42 – Thr 47) is highlighted by cyan color and catalytically important residues involved in substrate binding (Arg 37, Arg 41, Glu 241 and Asp 327) are shown in ball and stick model in the active site. Image produced using UCSF Chimera software [20].

substitution showed hydrogen bonding interactions with adjacent loop as reveal by molecular dynamics simulations on homology model (Chapter 5). Further improved intra-protein interactions in Ymphytase homology model were studied based on the distances between the side chain atoms using protein interaction calculator [21]. K139T substitution (best fit rotamer) in Ymphytase can form side chain – side chain hydrogen bonds with D137 that was not observed in wild-type. In case of V298F substitution, aromatic-aromatic interaction was observed between V298F and F420 (Supporting Information Figure S1). Aromatic-aromatic interaction between 4.5 – 7 Å was proposed to be a mechanism of protein stabilization [22]. This aromatic-aromatic interaction between two phenylalanines might be the reason for improved Ymphytase stability. Site directed mutagenesis on position T77 to incorporate Lysine in wild-type was shown to improve thermostability (Chapter 5). However, the obtained Ymphytase Omni1 mutant showed wild-type substitution at this position. Combining T77K with the identified Omni1 mutant would reveal importance of T77K in combination all other substitution. As a proof of principle we have shown the importance of co-operative effect of mutations in directed evolution experiment that can be applied to other class of enzymes to study structure-function relationship and improve several desired biocatalyst property.

7.5 Conclusions

In summary, we have reported OmniCASTing, a novel method for multiple site saturation that can be true alternative to already reported method for multiple site directed mutagenesis due to its robustness, reliability and simplicity. This method can be used to generate large mutant libraries with low empty vector background for high throughput screening. A cooperative effect of distant mutations can easily be studied with this method that is not feasible with current computational advances. As a proof of principle, we have improved three properties of already identified mutant from Ymphytase.

This method can easily be applied to active site modification to broaden our understanding for enzyme reactions and promiscuity.

7.6 References

1. Reetz, M.T., et al., *Directed evolution of an enantioselective epoxide hydrolase: uncovering the source of enantioselectivity at each evolutionary stage.* J Am Chem Soc, 2009. 131(21): p. 7334-43.
2. Zhu, L., et al., *Directed evolution of an antitumor drug (arginine deiminase PpADI) for increased activity at physiological pH.* Chembiochem, 2010. 11(5): p. 691-7.
3. Kim, M.S. and X.G. Lei, *Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR.* Appl Microbiol Biotechnol, 2008. 79(1): p. 69-75.
4. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Curr Opin Chem Biol, 2009. 13(1): p. 19-25.
5. Bershtein, S. and D.S. Tawfik, *Advances in laboratory evolution of enzymes.* Curr Opin Chem Biol, 2008. 12(2): p. 151-8.
6. Wagner, C.R. and S.J. Benkovic, *Site directed mutagenesis: a tool for enzyme mechanism dissection.* Trends Biotechnol, 1990. 8(9): p. 263-70.
7. Alam, M., D.E. Vance, and R. Lehner, *Structure-function analysis of human triacylglycerol hydrolase by site-directed mutagenesis: identification of the catalytic triad and a glycosylation site.* Biochemistry, 2002. 41(21): p. 6679-87.
8. Yuen, C.M. and D.R. Liu, *Dissecting protein structure and function using directed evolution.* Nat Methods, 2007. 4(12): p. 995-7.
9. Padhi, S.K., D.J. Bougioukou, and J.D. Stewart, *Site-saturation mutagenesis of tryptophan 116 of Saccharomyces pastorianus old yellow enzyme uncovers stereocomplementary variants.* J Am Chem Soc, 2009. 131(9): p. 3271-80.
10. Fox, R.J., et al., *Improving catalytic function by ProSAR-driven enzyme evolution.* Nat Biotechnol, 2007. 25(3): p. 338-44.
11. Reetz, M.T. and J.D. Carballeira, *Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes.* Nat Protoc, 2007. 2(4): p. 891-903.
12. Meetei, A.R. and M.R. Rao, *Generation of multiple site-specific mutations in a single polymerase chain reaction product.* Anal Biochem, 1998. 264(2): p. 288-91.
13. Peng, R.H., A.S. Xiong, and Q.H. Yao, *A direct and efficient PAGE-mediated overlap extension PCR method for gene multiple-site mutagenesis.* Appl Microbiol Biotechnol, 2006. 73(1): p. 234-40.
14. Seyfang, A. and J.H. Jin, *Multiple site-directed mutagenesis of more than 10 sites simultaneously and in a single round.* Anal Biochem, 2004. 324(2): p. 285-91.
15. An, Y., et al., *A rapid and efficient method for multiple-site mutagenesis with a modified overlap extension PCR.* Appl Microbiol Biotechnol, 2005. 68(6): p. 774-8.
16. Blanusa, M., et al., *Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method.* Anal Biochem, 2010. 406(2): p. 141-6.
17. Hanahan, D., J. Jessee, and F.R. Bloom, *Plasmid transformation of Escherichia coli and other bacteria.* Methods Enzymol, 1991. 204: p. 63-113.
18. Copeland, R.A., *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. 2nd ed. 2000: Wiley-VCH, New York.
19. Shivange, A.V., U. Schwaneberg, and D. Roccatano, *Conformational dynamics of active site loop in Escherichia coli phytase.* Biopolymers, 2010. 93(11): p. 994-1002.
20. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. 25(13): p. 1605-1612.
21. Tina, K.G., R. Bhadra, and N. Srinivasan, *PIC: Protein Interactions Calculator.* Nucleic Acids Res, 2007. 35(Web Server issue): p. W473-6.
22. Burley, S.K. and G.A. Petsko, *Aromatic-aromatic interaction: a mechanism of protein structure stabilization.* Science, 1985. 229(4708): p. 23-8.

Supporting Information: Chapter 7

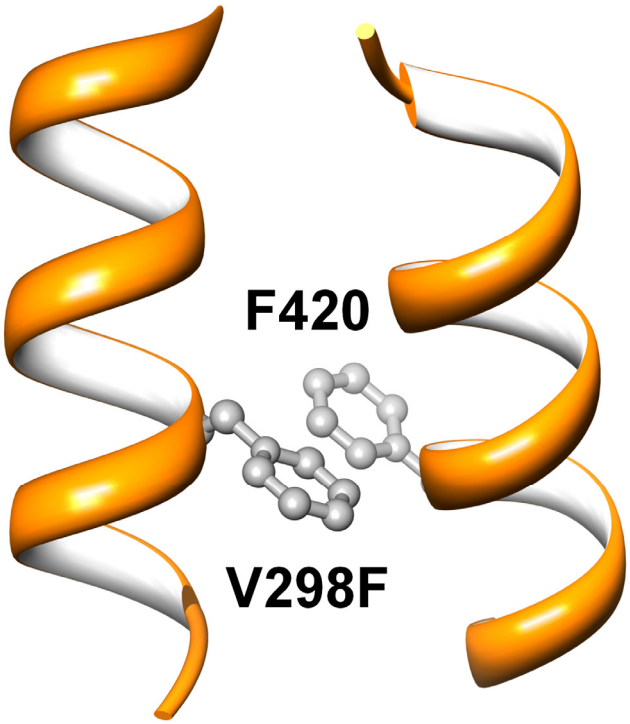


Figure S1: Aromatic-aromatic interactions between V298F and F420 residue in Ymphytase Omni1 variant.

Protein consensus based surface engineering (ProCoS): A novel method for directed protein evolution

Amol V. Shivange[a,b], Wolfgang Hoeffken[c], Stefan Haefner[c], and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany
[c]BASF SE, Fine Chemicals and Biocatalysis Research, GVF-A30, D-67056 Ludwigshafen, Germany

Protein consensus based surface engineering (ProCoS) is a novel method for directed protein evolution combining computational analysis with molecular biology tools. ProCoS is based on a hypothesis that the conserved residues belong to an ancestor and highly variable regions situated on the surface of protein can be targeted for surface engineering. ProCoS comprise mainly of four steps: identification of conserved ancestral and highly variable regions, gene synthesis of substituted surface residues in highly variable region, StEP recombination of synthetic genes, and screening for active variant. ProCoS is technically simple method for surface mutagenesis of a protein of interest without requirement of high computation power. Multiple sequence alignment was used for selection of surface residues using homology model. To demonstrate the utility of ProCoS method for directed evolution, the method was used to engineer the surface of phytase enzyme from *Yersinia mollaretii*. Screening of ProCoS mutant library yielded an enzyme with 34 amino acid exchanges whose pH stability at pH 2.8 was increased by 3.8 fold compare to wild-type phytase that might be due to significantly increased (20%) negatively charged amino acid substitutions.

8.1 Introduction

Directed protein evolution is a well established, versatile and successful algorithm for tailoring protein properties to industrial demands and for advancing our understanding of structure-function relationships in biocatalysts. Directed evolution entails accumulation of beneficial mutations in iterative cycles of mutagenesis and screening or selection for improved enzyme variants. These accumulations of mutations, a mutational path mostly lead a downhill on the fitness landscape plot of fitness versus sequence. Downhill mutational path eventually lead to unfolded or inactive variant of an enzyme. It is becoming clear that there exists two pathways (i) a widely known pathway of accumulating single amino acid change at each cycle of mutagenesis and

*Corresponding author

screening to select a improved enzyme [1] and (ii) a pathway where a cooperative effects in a combination of mutations leading to synergistic or additive improvement affecting the desired enzyme property due to a high number of mutations. These mutations can act in cooperation with each other. Little has been reported on the latter pathway for considering it as a major player in directed enzyme evolution. This study is evidence that the second pathway also plays a major role in directed evolution. Change in the protein surface property has become an interesting tool due to its wide potential applications in delivery [2], immobilization [3], solubilization, and stabilization of proteins. Modification of enzyme surface charge has been shown to improve thermostability [4]. Different approaches have been developed using chemical modifications [3] and using polymers [2] to change the surface property of the protein.

In this study, based on the second pathway of directed evolution, we have developed ProCoS, a method that can be used to incorporate more than 30 amino acid substitutions in an enzyme by surface modifications. A computational analysis based on conservation of amino acid was used to decipher the functionally important region and highly variable region in a model protein, phytase from *Yersinia mollaretii* (Ymphytase). A combination of computational and molecular biology tools were used in surface engineering of Ymphytase. Our main aim was to see if large number of mutations can be incorporated in the protein by maintaining its functionality. Therefore, a ProCoS method was developed to engineer the surface property of Ymphytase. The Ymphytase variant retaining (~40%) catalytic phytase activity was identified with 34 amino acid exchanges situated on the surface of Ymphytase. Interestingly, the pH stability of the Ymphytase was improved by 3.8 fold (pH 2.8) compare to wild-type that might be due to significant increase (20%) in negative surface substitutions in the identified Ymphytase variant.

8.2 Materials and Methods

8.2.1 Identification of conserved ancestral residues

As a first step towards identification of functionally important residues, available phytase amino acid sequences from enterobacteriaceae family of bacteria were retrieved from ExPASy proteomics server (www.expasy.org). A total 25 sequences of phytase enzymes were obtained from enterobacteriaceae family of bacteria from different genus include *Escherichia* (6), *Yersinia* (7), *Klebsiella* (5), *Pectobacterium* (1), *Shigella* (1), *Obesumbacterium* (2), and *Citrobacter* (3). Sequence alignment of all 25 sequences was performed using VectorNTI suite 10 software (Invitrogen, Darmstadt, Germany) using blosum62mt2 matrix with gap-opening penalty of 10 and gap-extension penalty of 0.05. A phylogenetic tree was obtained from the sequence alignment of 25 enterobacteriaceae phytases. Residues belonging to conserved, similar or identical sets of amino acids were considered in calculating consensus sequence that includes all conserved and ancestral residues in the enterobacteriaceae family.

8.2.2 Homology modeling

Homology model of Ymphytase wild-type enzyme was constructed as reported in section 5.2.4.

8.2.3 Selection and substitution of sites

Amino acid sites in Ymphytase were selected based on conservation of each amino acid in a family of enterobacteriaceae bacteria. Protein sequence was divided into functionally important region that include all conserved ancestral residues and variable region. Each residue in the variable region of sequence alignment was checked for frequency of occurrence in all the species. Positions of all these residues were then analyzed in homology model of Ymphytase using VMD software [5]. Residues belonging to loop and surface regions were selected. Selected residues were substituted with other amino acid based on three criteria (i) frequently occurring residues in the sequence alignment were identified and selected for substitution, (ii) Chemically similar amino acids were preferred, and (iii) Sterically favorable residue was incorporated.

8.2.4 Gene synthesis and subcloning

Three synthetic genes were designed based on mentioned protocol and codon optimization was perfomed using GeneDesign server for *E. coli* expression [6]. *YmappA* variant genes with restriction sites (*NdeI* at 5' end and *NotI* at 3' end) were synthesized commercially at GENEART (Regensburg, Germany). All synthetic constructs were digested with *NdeI* and *NotI* restriction enzymes. Digested synthetic genes were cloned into pET-22b(+) vector and transformed into *E. coli* BL21-Gold(DE3) strain for expression.

8.2.5 DNA recombination

A StEP recombination protocol [7] was optimized for recombination of synthetic genes of *YmappA* variants. Template for StEP recombination was generated from each pET-22b(+)-*YmappA* variant construct (1 ng/µl) by PCR (50 µl; 98 °C for 3 min, one cycle; 98 °C for 10 sec, 58 °C for 15 sec, 72 °C for 25 sec, 25 cycles; 72 °C for 180 sec) using pET-22b(+) vector specific F1 primer (5'- CGA CTC ACT ATA GGG GAA TTG TGA GCG GA-3') and R3 primer (5'-CGG GCT TTG TTA GCA GCC GGA TCT CAG-3') (0.4 µM each), Pfu DNA polymerase (0.025 U/µl), and dNTP mix (0.2 mM each) in thin-walled PCR tubes. All PCR products were column purified (NucleoSpin Extract II kit Macherey-nagel, Düren, Germany).

Vector specific primers F1 and R3 were used for StEP recombination. Each generated template (5 ng) was mixed together and three PCRs were performed using different annealing/extension time at 55 °C. StEP PCR was performed using linear amplified template (20 ng), 0.15 µM of each primer, 1x *Taq* buffer, dNTP (0.2 mM each), and 2.5 U Taq polymerase. PCR program: 94 °C for 30 sec (denaturation), and 55 °C for 1 sec, 5 sec or 10 sec (annealing/extension). StEP was perfomed using Mastercycler gradient (Eppendorf AG, Hamburg, Germany). PCR products (~1.4 kb) were gel extracted and purified (NucleoSpin Extract II kit Macherey-nagel, Düren, Germany). PCR products obtained from 5 or 10 sec annealing/extension time were

mixed together. Both the purified PCR products (1 sec or 5-10 sec) were digested with *NdeI* and *NotI* restriction enzymes (New England Biolabs GmbH, Frankfurt am Main, Germany). Digested PCR products were cloned into the *E. coli* expression vector pET-22b(+) and transformed into *E. coli* BL21-Gold(DE3) strain for expression. From now onwards we will refer mutant library obtained 1 sec annealing/extension as ProCoS mutant library-A and library obtained from 5 or 10 sec annealing/extension as ProCoS mutant library-B.

8.2.6 Screening of ProCoS variants

ProCoS mutant libraries were expressed in 96-well microtiter plates as described in 4.2.6. A two 96-well microtiter plate based screening assay was used. Briefly, 10 µl of cell lysate was incubated with 140 µl of substrate solution (0.6% phytic acid in 250 mM acetate buffer, 0.01% Tween-20, pH 5.5) for 1 hour at 37 °C. The reaction was stopped by the addition of 150µl of 15% trichloroacetic acid. Inorganic phosphate released was quantified by addition of 20 µl of stopped reaction mixture to 280 µl color mix solution (0.27% w/v ammonium molybdate and 1.08% w/v ascorbic acid in 0.32 M H_2SO_4), and the absorption was measured at 820 nM.

8.2.8 Purification and characterization of wild-type Ymphytase and mutant

Purification and kinetic parameters determination for Ymphytase wild-type and ProCoS-2 variant were performed as described in section 4.2.8 and 4.2.9 respectively.

pH stability

The pH stability of wild-type and mutant phytase was determined at 37 °C using four different buffers: 0.25 M Glycine-HCL buffer for pH 2.0 to 3.2, 0.25 M sodium acetate buffer for pH 3.6 to 5.6, 0.25 M imidazole-HCL buffer for pH 6.0 to 7.0 and 0.25 M Tris-HCL for pH 7.4 to 9.0. Purified enzymes were diluted in specified buffers to 40 ng protein per milliliter and incubated at 37 °C for three hours. Activity assay was carried out with 1 mM phytate using phytase at pH 4.5 and at 37 °C using 96-well plate-format phytase colorimetric screening assay as described in section 4.2.7. Non pH-treated enzyme activity was considered as 100% activity and residual (relative) activity was calculated.

8.3 Results

8.3.1 General concept of ProCoS method

Functional proteins (enzymes) have some common features (sequence motifs) that reflect their functionality and these functional sequence motifs are conserved throughout different species. A correct configuration (folding) of a protein gives an ability of being a catalyst. Amino acid residues present in the interior part, buried inside the protein are important for correct folding. The underlying logic for the ProCoS engineering is a hypothesis that the highly conserved residues in the protein sequence belongs to ancestor in a family of species and the highly variable regions in the protein situated on the protein surface can be targeted for surface modification. These protein surface residues that are not conserved during the evolution of a specific class of protein

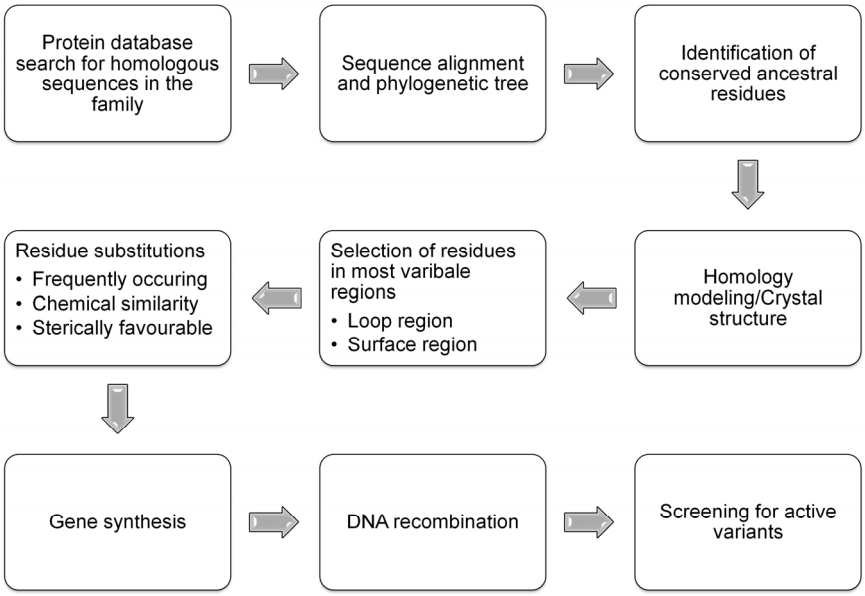


Figure 1: Schematic representation of the steps in ProCoS method

throughout the species can be targeted to alter the surface properties of a biomolecule.

In ProCoS method (Figure 1), first a target protein sequence is divided into functional and variable regions based on sequence alignment of the enzymes from homologous species. The sequences for homologousproteins from a single family of species are retrieved from protein database. A phylogenetic tree is then constructed based on sequence alignment to identify the most variable residues. The positions of these residues are then analyzed on the protein structure to find out the location. Residues situated on surface of the protein and in the loop region are selected. These residues are then substituted to the another residues based on three criteria (i) Frequently occurring residues in the sequence alignment are preferred for substitution (ii) Chemical similarity: Charged amino acids are introduced either by swapping two charged amino acids that might involve in salt bridge interactions or by analyzing the surrounding area to avoid accumulation of positive or negative charge (iii) Sterically favorable: size of the amino acid was considered for substitution to avoid steric clashes after substitution. Three synthetic genes are designed based on above criteria and synthesized commercially. All synthetic gene constructs were recombined using StEP recombination and mutant library is screened for active variants.

8.3.2 Ymphytase surface engineering by ProCoS

Three synthetic genes with 31 (SyntheticGene-1), 34 (SyntheticGene-2), and 40 (SyntheticGene-3) substitutions were synthesized. As expected, due to high mutational load on a protein, all three synthetic Ymphytase variant were inactive after protein expression at 37 °C. A little phytase activity was

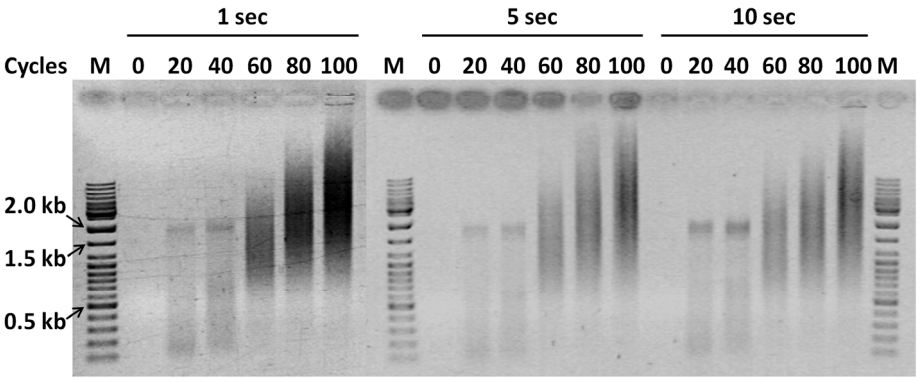


Figure 2: Electrophoresis gel showing a single DNA band observed during StEP recombination of synthetic genes with different annealing/extension time between 20 – 40 cycles.

detected in an order of SyntheticGene-2 > SyntheticGene-3 > SyntheticGene-1, after expression at 30 °C. StEP recombination of all three synthetic genes without wild-type Ymphytase was performed. Templates using vector specific primers were amplified to incorporate a universal primer binding site (F1 and R3) in all the synthetic genes to be recombined. Three different annealing/extension time were used for StEP PCR resulted in a discrete electrophoresis band of the correct size (~1.4 kb) after 20 and 40 cycles in all three cases. Higher number of cycles resulted in a smear that was increasing with PCR cycles (Figure 2).

The cloned ProCoS mutant libraries of Ymphytase variants were screened for activity. Approximately, 1050 clones were screened using 96-well microtiter plate based assay. A total 552 and 460 mutant clones were screened from ProCoS mutant library-A and ProCoS mutant library-B respectively. At initial screening, mutants showing at least 20% of relative activity of wild-type Ymphytase were selected. ProCoS mutant library-A was found to have more active clones (26%) than ProCoS mutant library-B (20%). Six best ProCoS variants were retransformed and activity was assayed using test-tube expression (Supporting Information Figure S2). Sequences of Ymphytase variant, ProCoS-3 and ProCoS-5 were identical. A ProCoS-2 variant with highest activity among six clones was selected for further purification and characterization.

8.3.3 Characterization of Ymphytase ProCoS variant

The pH profile of the ProCoS-2 variant showed broad pH optimum from pH 3 to pH 4.5 compared to wild-type (pH optima at pH 4.5). The phytase kinetics parameters were determined at pH 4.5, using sodium phytate as a substrate and data generated was analyzed by the general allosteric model using the Hill equation. The Hill coefficient (h) value for wild-type, ProCoS-2 enzyme was 2.2 and 2.9 indicating positive cooperativity (Table 1). The K_{half} for the wild-type and ProCoS-2 variant were 268 μM and 261 μM respectively. The catalytic activity (K_{cat}) was calculated by considering four active site centers per molecule of Ymphytase.

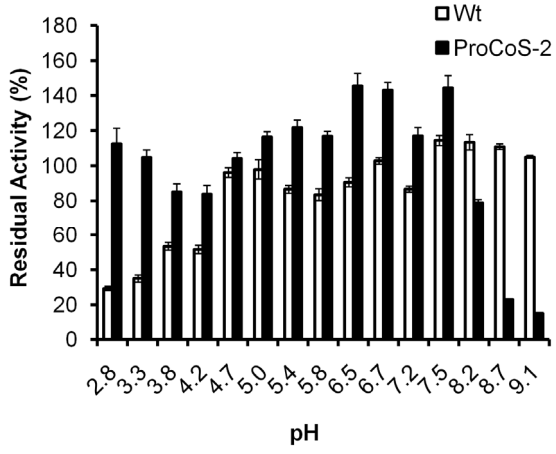


Figure 3: Residual enzyme activity of Ymphytase wild-type (Wt), and ProCoS-2 variant after being incubated at pH 2.8 (37 °C for 3 hours).

The ProCoS-2 variant was found to posses ~40% of wild-type activity after being having 34 amino acid substitutions. Ymphytase ProCoS-2 variant showed overall improved pH stability between pH 2.8 to 7.5 than wild-type, a 3.8 fold and 3 fold pH stability (37 °C/3 hours) was improved at pH 2.8 and pH 3.3 respectively (Figure 3).

8.4 Discussion

The validation of the ProCoS method was done by engineering the surface of a phytase from *Y. mollaretii*. Functionally important conserved residues belonging most likely to ancestor were identified by phylogenetic analysis based on multiple sequence alignment (Supporting Information Figure S1). Amino acid sequences were selected for alignment based on the evolutionary relatedness. The relatedness of the seven groups of species from enterobacteriaceae family of bacteria is shown by phylogenetic tree in figure 4. Sequences from these species were 25% to 85% identical with Ymphytase. Most of the sequences were ~50% identical to Ymphytase except *Yersinia* (80% – 86%) and *Klebsiella* (25-30%) species. *Yersinia* phytases have been shown to possess high catalytic efficiency (1000 – 3900 U/mg) [8] whereas *Klebsiella* phytases possess lower activity [9]. Therefore, 7 and 5 sequences out of 25 for *Yersinia* and *Klebsiella* respectively were used in consensus sequence calculation. Despite of very low sequence identity (25% - 85%) between the used sequences, significant regions in the phytase sequence were highly conserved in enterobacteriaceae phytases. Based on consensus sequence calculation, the highly variable region in Ymphytase sequence was identified. The highly variable regions situated on the surface of the Ymphytase were targeted for mutagenesis.

Synthetic genes possessing 30 – 40 amino acid substitutions were expressed in *E. coli* that revealed the tiny activity after expression at lower temperature (25 – 30 °C), whereas at 37 °C no detectable activity was observed that might be due to lower protein stability or inclusion bodies. Therefore, StEP recombination was employed to shuffle the synthetic genes. StEP recombination has been successfully used to improve thermostability of subtilisin E enzyme [10]. Sequence information within the parental sequences is exchanged due to template switching events in StEP recombination. The unique sequence information (mutations) in a parent sequence can pass to other parent sequence and mutations are distributed in the chimeras (ProCoS mutant library) obtained from StEP experiment.

The selection pressure of activity was employed to screen ProCoS mutant library,

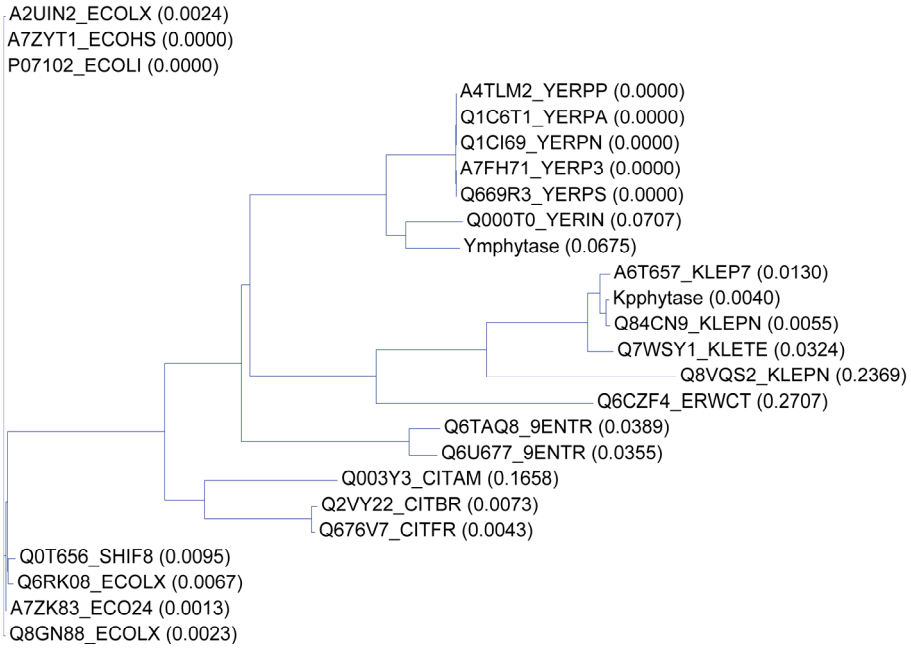


Figure 4: Phylogenetic tree showing the evolutionary relatedness and level of homology between phytases used for synthetic gene design by sequence alignment

Table 1: Kinetic parameters of YmPhytase wild-type and ProCoS-2 variant.

Enzyme	Specific activity (U mg^{-1})	K_{half} (μM)	*K_{cat} (S^{-1})	Hills coefficient (h)
Wild-type	1043.5 (± 26.5)	268.1 (± 18.2)	823.47 (± 20.91)	2.24 (± 0.12)
ProCoS-2	401.08 (± 10.92)	260.55 (± 9.55)	316.51 (± 8.62)	2.93 (± 0.2155)

Phytase kinetics was determined in 0.25 M sodium acetate containing different sodium phytate (7.8 to 750 μM) at 37°C and pH 4.5 for both the phytases
Values in the bracket represent standard error.
*Catalytic center activity: enzyme concentration (moles) multiplied by the number of active centers (four).

screening revealed higher number of active clones in ProCoS mutant library-A compare to library-B. High ratio of active clones might be due to high recombination efficiency at 1 sec extension/aanealing in StEP PCR compare to 5 or 10 sec extension/annealing conditions. Sampling of only ~1000 clones resulted in six variants with phytase activity. Phylogenetic tree (Figure 5) using multiple sequence alignment of six ProCoS variants, synthetic genes and wild-type Ymphytase reveal the evolutionary pathway of the ProCoS variants. Sequences were grouped according to activity. ProCoS-2 variant was grouped close to SyntheticGene-2 which is the most active synthetic gene among all other synthetic genes when expressed at low temperature. The ProCoS-2 variant was recombined to possess 6 amino acids substitutions at the C-terminal of Ymphytase compare to SyntheticGene-2 of Ymphytase

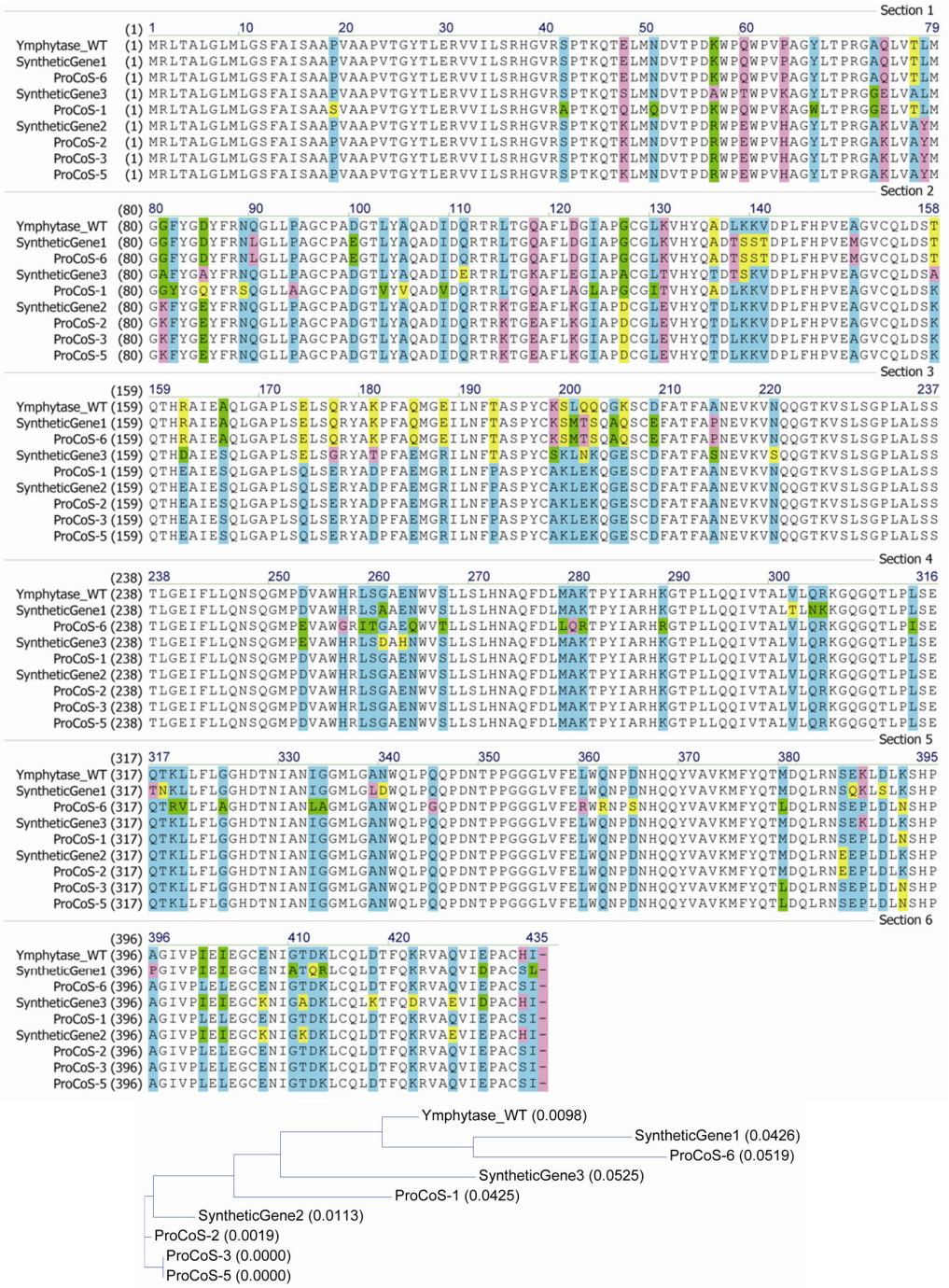


Figure 5: Multiple sequence alignment (above) and phylogenetic tree (below) showing the relatedness and level of homology between the obtained ProCoS variants of Ymphytase and designed synthetic genes.

(I401L, I403L, K407E, K411T, E426Q, and H433S). However, active clones obtained from ProCoS mutant library screening were distributed in a phylogenetic tree that shows the designed sequences to be quite close to the active clones obtained. In surface engineering of

Table 2: Analysis of the amino acid composition on 34 substitutions obtained in ProCoS-2 variant.

Amino acid	Wild-type		ProCoS-2		Chemical Property
	Total numbers	Substitutions (%)	Total numbers	Substitutions (%)	
Glu	3	9	10	29	Negative, Polar
Lys	6	18	8	24	Positive, Polar
Ala	2	6	2	6	Neutral, Non-polar
Arg	1	3	2	6	Positive, Polar
Asp	2	6	2	6	Negative, Polar
Leu	2	6	2	6	Neutral, Non-polar
Pro	1	3	2	6	Neutral, Non-polar
Ser	2	6	2	6	Neutral, Polar
Gln	7	21	1	3	Neutral, Polar
His	1	3	1	3	Positive, Polar
Thr	3	9	1	3	Neutral, Polar
Tyr	0	0	1	3	Neutral, Polar
Gly	2	6	0	0	Neutral, Non-polar
Ile	2	6	0	0	Neutral Non-polar

Ymphytase using ProCoS, most of the substitutions that are chemically non-similar were situated at N-termini (Figure 5; highlighted by magenta color) despite the active site of Ymphytase is situated close to N-termini. Total 34 amino acid positions were substituted in ProCoS-2 variant compare to wild-type. In case of wild-type, a total 21% of amino acids were represented by Glutamin (Neutral polar amino acid) that was reduced to 3% in ProCoS-2 variant (Table 2). Numbers of lysines, a positively charged amino acid were increased from 18 to 24 in ProCoS-2 and numbers of negatively charged amino acids, Glutamine were increased from 9 to 29 in ProCoS-2 Ymphytase variant. This reflects increase in negative polar surface on Ymphytase enzyme.

In order to get the fraction of surface changed due to all substitutions, a thorough analysis on type of substitution was performed (Table 3). In case of wild-type, out of 34

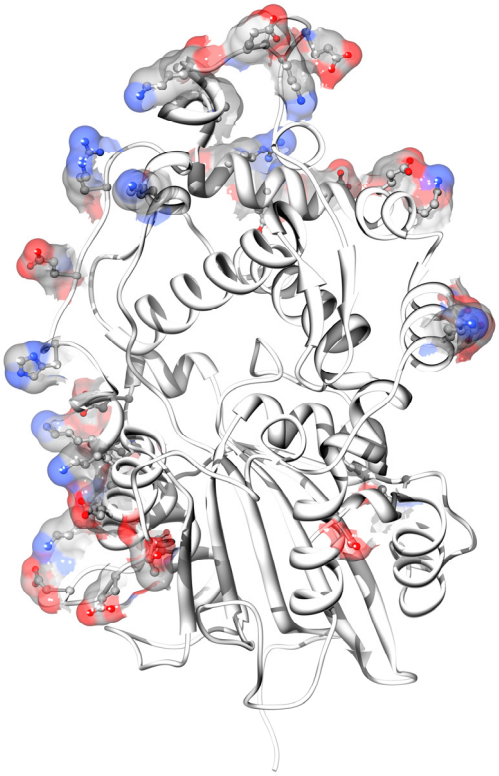


Figure 6: Amino acid substitutions (34) in Ymphytase ProCoS-2 variant are shown in the homology model of Ymphytase

Table 3: Fraction of surface residues changed in Ymphytase ProCoS-2 variant illustrated by chemical property of substitutions.

Amino acids	Fraction of substituted residues (%)	
	Wild-type	ProCoS-2
Negative, Polar	15	35
Positive, Polar	24	32
Neutral, Polar	35	15
Neutral, Non-polar	26	18
Polar	74	82
Non-polar	26	18

residues substituted, negative polar surface was represented by 15% that was increased by 20% in ProCoS-2 variant. Surface contributed by either positive polar or polar residues was increased by 8%. Both the neutral non-polar or non-polar residues were decreased by 8% and Neutral polar surface was decreased by 20% of substituted residues. Overall, the polar surface was increased by 8% and non-polar surface of Ymphytase wild-type was decreased by 8%. The residues substituted in ProCoS-2 variant are shown in figure 6. The increase in pH stability of ProCoS-2 variant might be due to increased fraction of polar surface, a significant increase in negative surface (20%) was achieved by decreasing neutral surface (20%) of the Ymphytase.

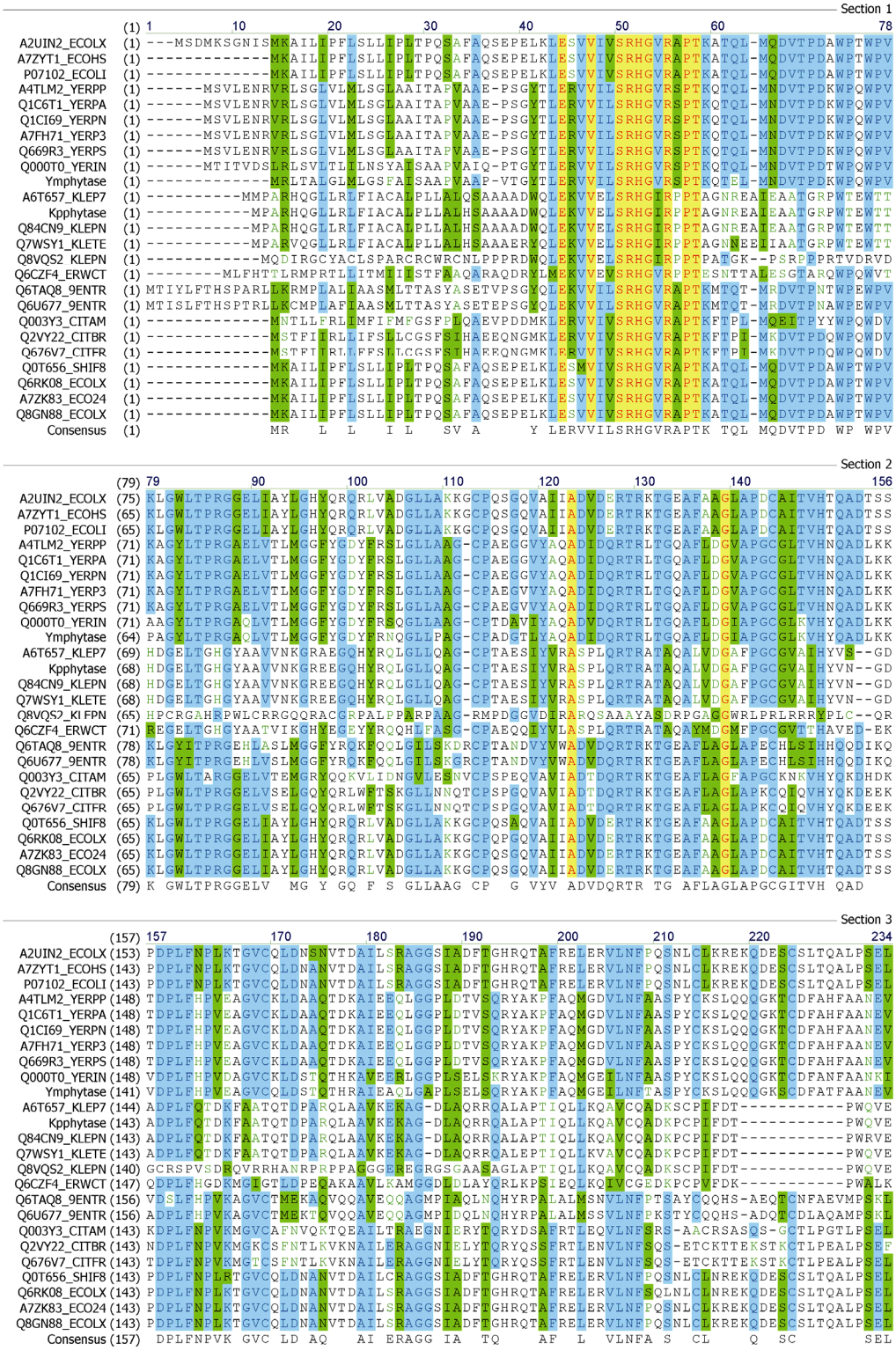
8.5 Conclusions

ProCoS, a novel method combining computational analysis and molecular biology techniques has been developed for protein surface engineering. ProCoS method was based on the hypothesis that the highly conserved residues in the protein belong to ancestor in a family of species and the highly variable regions in the protein situated on the protein surface can be targeted for protein surface modification. These residues were successfully substituted (34 sites) with other amino acids in Ymphytase using ProCoS method. Active variant was identified with 3.8 fold improved pH stability. Improved pH stability might be due to increased negatively charged amino acid substitutions on Ymphytase surface. ProCoS method can be applied to other biocatalyst for improving properties that are influenced by surface properties of the protein like solubility, thermostability, organic solvent stability.

8.5 References

1. Tracewell, C.A. and F.H. Arnold, *Directed enzyme evolution: climbing fitness peaks one amino acid at a time.* Curr Opin Chem Biol, 2009. 13(1): p. 3-9.
2. Veronese, F.M., *Enzyme surface modification by polymers for improved delivery.* J Controlled Release, 1994. 29(1-2): p. 171-176.
3. Montes, T., et al., *Chemical modification of protein surfaces to improve their reversible enzyme immobilization on ionic exchangers.* Biomacromolecules, 2006. 7(11): p. 3052-8.
4. Gribenko, A.V., et al., *Rational stabilization of enzymes by computational redesign of surface charge-charge interactions.* Proc Natl Acad Sci U S A, 2009. 106(8): p. 2601-6.
5. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. 14(1): p. 33-8, 27-8.
6. Richardson, S.M., et al., *GeneDesign: rapid, automated design of multikilobase synthetic genes.* Genome Res, 2006. 16(4): p. 550-6.
7. Aguinaldo, A.M. and F.H. Arnold, *Staggered extension process (StEP) in vitro recombination.* Methods Mol Biol, 2003. 231: p. 105-10.
8. Huang, H., et al., *A novel phytase from Yersinia rohdei with high phytate hydrolysis activity under low pH and strong pepsin conditions.* Appl Microbiol Biotechnol, 2008. 80(3): p. 417-26.
9. Sajidan, A., et al., *Molecular and physiological characterisation of a 3-phytase from soil bacterium Klebsiella sp. ASR1.* Appl Microbiol Biotechnol, 2004. 65(1): p. 110-8.
10. Zhao, H., et al., *Molecular evolution by staggered extension process (StEP) in vitro recombination.* Nat Biotechnol, 1998. 16(3): p. 258-61.

Supporting Information: Chapter 8



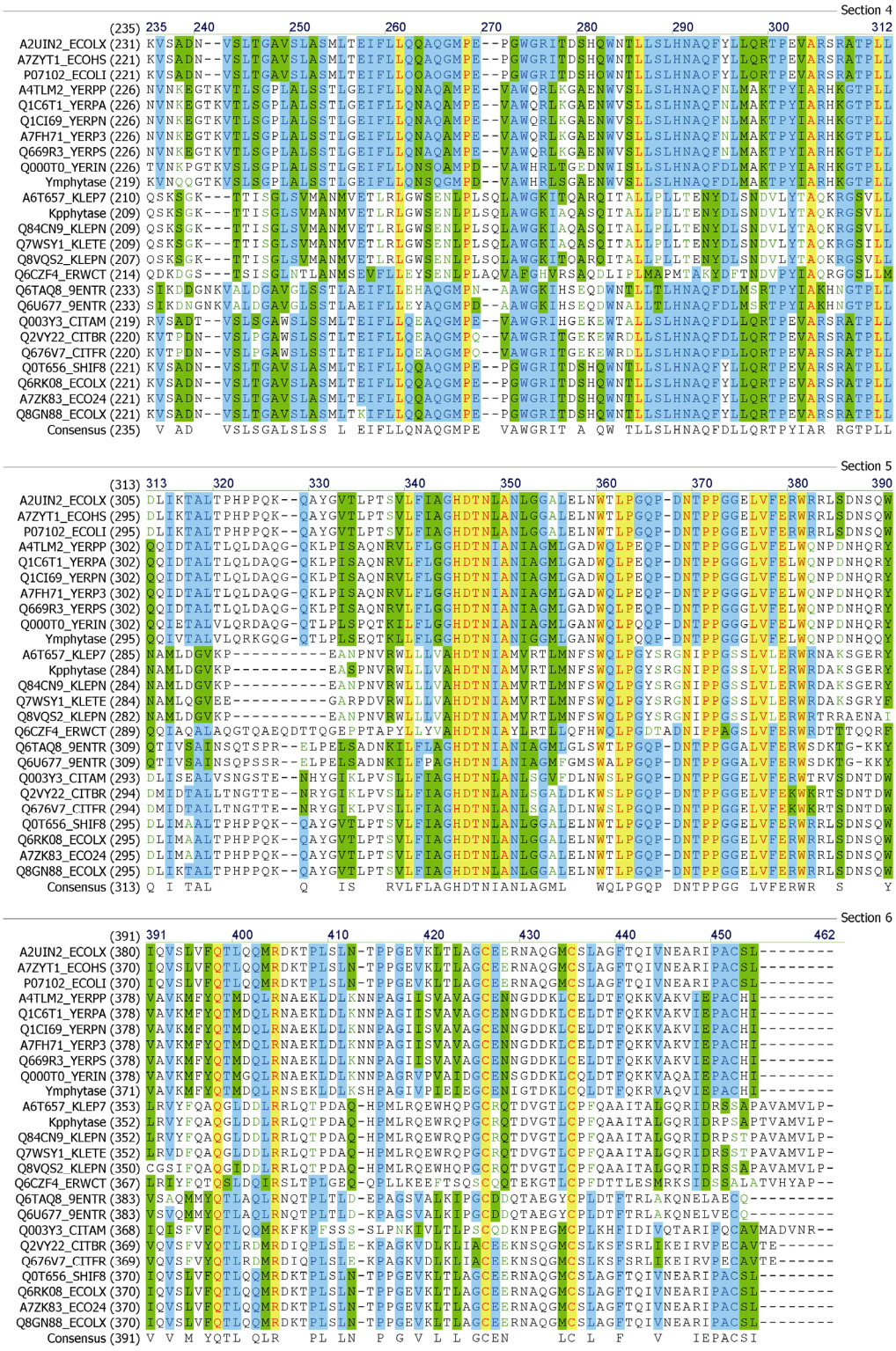
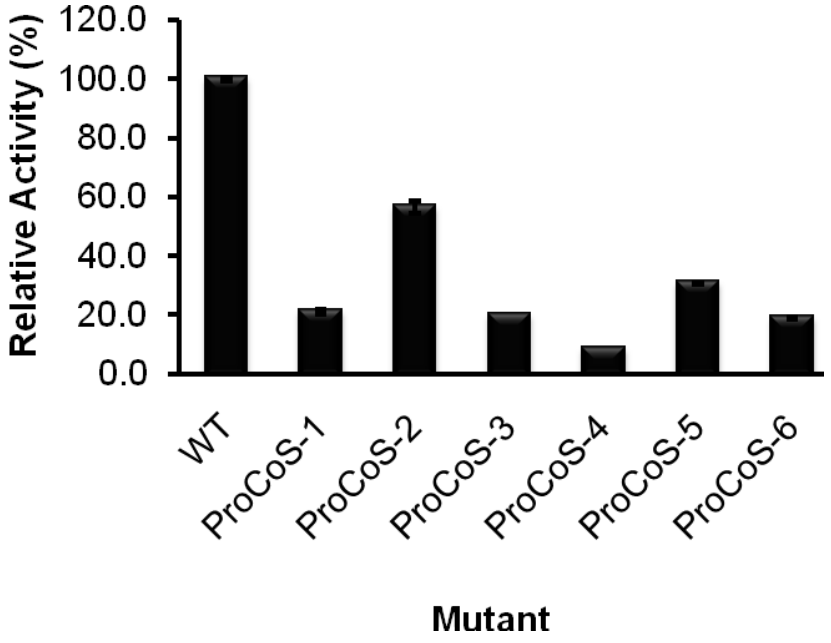


Figure S1: Multiple sequence alignment of phytases from enterobacteriaceae family of bacteria.



Variant	Relative Activity
WT	100
ProCoS-1	21
ProCoS-2	57
ProCoS-3	20
ProCoS-4	8
ProCoS-5	31
ProCoS-6	19

Figure S2: Relative activity of selected six cloned from ProCoS mutant library. Activity assay was perfomed for phytate as a substrate in cell lysate.

Conformational dynamics of active site loop in *Escherichia coli* phytase

Amol V. Shivange[a,b], Ulrich Schwaneberg[a,b] and Danilo Roccatano[a]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Phytases catalyze the release of phosphate by stepwise hydrolysis of phytate, a major source of phosphate in cereal grains, legumes and oilseeds. Phytase improves as a feed supplement the nutritional quality of phytate rich diets and eventually reduce environmental pollution. Recently, phytases from enterobacteriaceae family have attracted industrial interest due to their high specific activity (2500 – 4000 U/mg). However, only limited information is available concerning structural dynamics of this class of enzymes. In this study, 50 nanosecond molecular dynamics simulation was performed on two *E. coli* phytase structures (closed and open active site loop) to investigate conformational dynamics of the active site loop. Cluster analysis and principal component analysis (PCA) reveal significant difference in the conformational dynamics of active site compared to reported crystal structure. Molecular dynamic studies indicated that the movement in the active site of *E. coli* phytase is mainly confined by the active site loop resulted in wider opening of the loop in absence of phytate. The molecular dynamics studies highlight the possible role of loop residues as prerequisite for highly active phytases.

9.1 Introduction

Phytases (myo-inositol hexakis phosphohydrolase), catalyze the release of phosphate by stepwise hydrolysis of phosphomonoester bonds in phytate (myo-inositol-1,2,3,4,5,6-hexakisphosphate). Phytate is the major storage form of phosphate and inositol, predominantly occurring in cereal grains, legumes and oilseeds [1]. However, the phosphorous in this form is essentially unavailable to monogastric animals such as human beings, dogs, birds, swine and poultry due to lower or negligible phytase activity [2]. Furthermore, phytate is considered to be anti-nutrient because of its polyanionic chelating property that forms complexes with several divalent cations of major nutritional importance like Zn^{2+}, Cu^{2+}, Ca^{2+}, Mn^{2+} Fe^{2+}, and Mg^{2+}. Phytate can also form complexes with proteins and amino acids [3]. Undigested phytate leads to decrease in bioavailability of inorganic phosphate (Pi) and nutritionally important elements to the animals and as a result undigested/unabsorbed phytate is excreted as large amounts of Pi into the environment resulting in environmental pollution [4]. For

*Corresponding author

proper skeletal growth, these animals need Pi at suitable concentration, thus phytases are considered of a significant value in upgrading the nutritional quality of phytate rich feed, supplementation of phytase to the feed additive provided an alternative to tackle both these conditions effectively.

In addition to its major application in the animal nutrition, phytases are also used in human nutrition, inhibitory effect of inositol hexaphosphate (phytate) as well as pentaphosphate on iron absorption has been reported [5], causing the high prevalence of iron deficiency in infants from developing countries, women of fertile age, or vegetarians. The lower phosphoric esters of myo-inositols might be useful in controlling atherosclerosis, inflammation and heart diseases [6]. Some lower myo-inositols were successfully synthesized using *E. coli* phytase [7] that would overcome the limitation of difficult chemical synthesis of phosphorylated myo-inositols [8].

Phytases are widespread in nature, found in animals, plants, and microorganisms. Members of the histidine acid phosphatase (EC: 3.1.3.8) family have attracted increased interest from the animal feed industry because of their high specific activity and their ability to efficiently hydrolyze phytate [9]. Fungal phytases from *Aspergillus niger* (var. *ficuum*) (BASF) and *A. awamori* (AB Enzymes) are widely used as a commercial feed additive [10]. Despite wide range of applications, phytases face the considerable challenge of the low activity. Recently, several phytases from enterobacteriaceae family [11] have been reported to possess higher specific activity (2500 – 4000 U/mg) compare to fungal phytases [10] towards phytate including the *E. coli* phytase. However, inadequate information is available on structural dynamics of this class of enzymes. At present, only 30 crystal structures of phytases are present in the Protein Data Bank (PDB) [12]: 6 from *E. coli,* 11 from *Selenomonas ruminantium*, 5 from *A. fumigatus*, 1 from *A. ficuum*, 5 from *Bacillus amyloliquefaciens*, 1 from *Mitsuokella multacida* and 1 from *Debaryomyces castellii*.

E. coli phytase crystal structures have been solved in the presence or absence of phytate [13]. Phytate binding has been demonstrated to induce the conformational changes near the active site of phytases resulting in two different conformation of an active site loop (Arg 20 – Ala 25). Open loop conformation was observed in the absence of phytate which closes upon binding of phytate [13]. Several highly conserved residues in the active site, comprise Glu 219, Asp 304 and residues Arg 16, Arg 20 (part of a conserved sequence motif “RHGXRXP” of histidine acid phosphatase) are important for substrate binding and catalysis [13]. In order to investigate the conformational dynamics of the phytase active site, open loop conformation (PDB ID: 1DKL) and closed loop conformation (PDB ID: 1DKP) structures were selected. A detailed comparison is made between collective loop fluctuations in *E. coli* phytase derived from an extensive molecular dynamics (MD) simulations in water. Two simulations, each of 50 nanoseconds (ns) were performed for both the open and closed loop conformation X-ray structures. Principal component analysis [14] (PCA) was applied to separate the functionally

important positional fluctuations from extraneous local fluctuations. Cluster analysis [15] and DomainSelect analysis [16] was used to investigate the direction and extent of active site loop opening in *E. coli* phytase. To our knowledge, this is the first study of MD simulation on the enzyme class of phytase and the first use of DomainSelect to investigate loop dynamics. Results reveal a significant dynamic flexibility in phytase active site which can primarily be attributed to an active site loop that leads to a wider opening of loop in the absence of phytate. Rapid cycling of different active site conformations has been reported to be essential for higher enzymatic activity [17] and therefore the high flexibility in the active site confined by the active site loop in *E. coli* phytase might be prerequisite for higher catalytic efficiency of phytases.

9.2 Materials and Methods

9.2.1 Atomic coordinates

The starting structures (PDB ID: 1DKP and 1DKL) were retrieved from Protein Data Bank [12]. The ligand (phytate) and Hg^{++} ions were removed from the closed loop conformation structure (1DKP). A116T and H17A mutations were incorporated in the open loop conformation structure (1DKL) so that the sequences of open and closed loop structures were identical. Mutations were incorporated using PyMOL software, [18] rotamers with the highest fit probability were selected for substituted residues. *E. coli* phytase is a monomer protein [19] therefore chain B from 1DKL was deleted. The missing loop between Arg181 and Ser187 in 1DKL was modeled by performing a manual modeling using the corresponding loop coordinate obtained from another unliganded *E. coli* phytase structure (PDB ID: 1DKO) after superimposing with 1DKL (see Figure S1 in the Supporting Information). Both the structures were energy optimized by energy minimizations and subsequently used for two MD simulations. From now on we will refer to them as $1DKL_{open}$ (open loop) and as $1DKP_{closed}$ (closed loop).

9.2.2 MD simulations

Simulations were performed using the GROMACS (version 3.3.1) software package [20, 21] and GROMOS96 (version 43a1) forcefield [22]. Both $1DKL_{open}$ and $1DKP_{closed}$ proteins were protonated and solvated in dodecahedron boxes with water by stacking equilibrated boxes of the solvent to form a box large enough to contain phytase and at least 0.9 nm (nanometer) of the solvent on all sides. Since the resulting protonation state of the molecule gave a total charge of –4 in $1DKP_{closed}$ and $1DKL_{open}$ system, four sodium counterions were added to satisfy the condition of global electroneutrality in a simulation cell. Each simulation box contained roughly 15,000 water molecules, resulting in about 19,000 atoms. Energy minimization of the entire system was carried out for 1000 steps using the steepest descents minimization algorithm in order to remove any van der Waals clashes generated by solvent molecules.

The MD simulations were carried out at a constant temperature and pressure. The temperature was kept constant at 300 K using the weak coupling algorithm [23] with coupling constant of 0.1 ps while the pressure was kept constant at 1 bar by the weak coupling constant

of 0.5 ps [23]. For water molecules, the SPC water model [24] was used. The SHAKE algorithm [25] was used to constrain all the bond lengths. For the water molecules, the SETTLE algorithm was used to constrain bond lengths and bond angles [26]. A dielectric permittivity, ε_r=1, and a time step of 2 fs were used. The non-bonded interactions (electrostatic and LJ) were calculated using the Particle Mesh Ewalds (PME) method [27]. For the calculation of the long-range interactions, a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation were used to compute the potential and forces in between grid points. A non-bonded pair-list cutoff of 0.9 nm was used and the pair-list was updated every 5 time steps. The potential energy of each system was first minimized using the steepest descent algorithm. Water molecules were allowed to adjust by performing 100 ps of molecular dynamics with position restraints on the protein. Initial velocities for each atom were generated with the Maxwell distribution at 300 K. Each system was equilibrated for 100 ps before trajectory data were produced for analysis. Fifty nanoseconds production runs were performed on the open loop ($1DKL_{open}$) and closed loop ($1DKP_{closed}$) systems.

9.2.3. Cluster analysis

Clustering was used to study different loop conformations sampled in trajectories. The RMSD-based GROMOS algorithm [15] was used for the analysis of the MD trajectories. For each structure in the trajectory, a least-square translational and rotational fit was performed using backbone atoms, and the RMSD for the backbone atoms of either the loop residues or entire protein residues were calculated. A cutoff of 0.15 nm was selected for both $1DKL_{open}$ and $1DKP_{closed}$ proteins. Clustering of active site loop conformations was also performed only on the active loop residues by combining both $1DKL_{open}$ and $1DKP_{closed}$ trajectories and using a cutoff of 0.08 nm.

9.2.4. Principal component analysis

Principal component analysis or essential dynamics was used to investigate the collective motions of the protein during the simulation. The analysis was performed on the backbone atoms (Cα, N, C) as described in Ref. 14. The number of atoms where N = 1230 in $1DKP_{closed}$ and N = 1227 in $1DKL_{open}$ are the number of atoms used for the analysis. Each principal component describes a collective motion, which can be represented by two structures, one at the maximum value, and the other at the minimum value of the collective variable when the trajectory is projected onto it. These maximum and minimum structures (extreme structures) were extracted by projecting either $1DKP_{closed}$ or $1DKL_{open}$ trajectory on corresponding first principal component and superimposed using UCSF Chimera software [28]. The analysis was performed on 50 ns MD simulation trajectories of $1DKP_{closed}$ or $1DKL_{open}$.

9.2.5. DomainSelect analysis

DomainSelect program [16] was used to study conformational changes occurring in the active site loop during simulation. DynDom and DomainSelect programs have been successfully used for analyzing user specified domain fluctuations [29-33]. Here, we used DomainSelect to analyze loop opening by taking

into consideration all protein residues except loop residues as domain α (fixed domain; Ser1 – Arg19, Leu27 – Leu409) and loop residues as domain β (moving domain; Ala20 – Gln26). DomainSelect superimposes two conformations of domain α (all protein residues except loop residues) by least-squares fit and then calculates the angle of rotation required to superimpose the two conformations of the active site loop (domain β). The angle of rotation calculated is the angle between active site loop in reference conformation and in another conformation. A representative median structure of first three clusters of active site loop from each $1DKP_{closed}$ and $1DKL_{open}$ simulation were extracted from the trajectory and used as an input for comparison with X-ray structures. Two extreme structures calculated using principal component analysis, were also used as an input for DomainSelect analysis.

All analyses were performed using the tools of GROMACS [34] package. The molecular graphics images in Figure 2 was generated using VMD program [35] and in Figure 4 and 5 were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [28].

9.3 Results and Discussion

9.3.1 General structural properties

The stability and reliability of the simulations were accessed by a series of standard analyses for both the trajectories. Root-mean-square deviations (RMSD) of backbone atoms from the starting structure coordinates during the MD simulation (Figure 1), radius of gyration (Rg) (Supporting Information, Figure S2) and secondary structure (calculated using the Kabsch-Sander method [36], Supporting Information, Figure S3 and S4,) as a function of time, showed for both the phytase structures, an overall structural stability throughout the simulations time. The secondary structural elements were found to be almost constant throughout the simulation time and radius of gyration showed a formation of a more compact structure with respect to the starting one. A slightly higher average RMSD value using $1DKP_{closed}$ as a reference structure was observed, which illustrates a rough estimation that the conformations in $1DKP_{closed}$ trajectories were more similar to $1DKL_{open}$ crystal structure. The detail analysis is reported in supporting information. The convergence to a stable plateau for both RMSD and Rg curves occurred after approximately 20 ns simulation time, however, the open structure maintains a slightly more extended structure than the closed one (see Figure 2c).

9.3.2 Conformational changes of active site loop during simulation

The analysis of the active site loop residues (Arg 16 – Thr 26) was performed by concatenating the trajectory of $1DKP_{closed}$ and $1DKL_{open}$. RMSD for only active site loop was calculated with respect to either open loop (grey line) or closed loop (black line) reference structure for $1DKP_{closed}$ (Figure 1b) and $1DKL_{open}$ (Figure 1c) trajectory. RMSDs calculated with respect to corresponding reference structures from closed and open phytase starting structures reached similar values. The latter result indicates that at the end of both

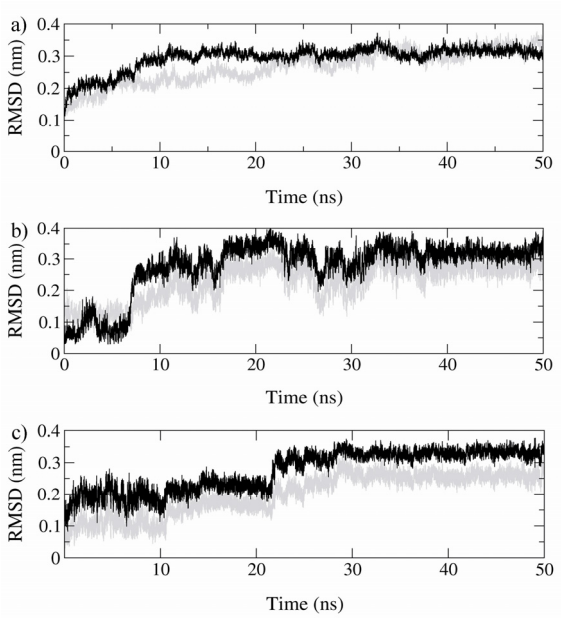


Figure 1: RMSD of backbone atoms from the starting structure over the length of MD simulation of $1DKP_{closed}$ (black line) and for $1DKL_{open}$ (grey line) (a); Overlap of RMSD calculation of active site loop residues in the trajectory of $1DKP_{closed}$ (b) and $1DKL_{open}$ (c) using either open loop conformation (grey line) or closed loop conformation (black line) structure as a reference structure.

simulations, the equilibrium conformations converge to a similar conformation. To gain further insights into conformational changes for both open and closed loop structures, cluster analysis of trajectories and DomainSelect analysis of the median structures of the most abundant clusters were performed.

9.3.3 Cluster analysis of E. coli phytase

Cluster analysis on only loop residues and on all protein residues was carried out.

Cluster analysis of loop conformations

The trajectories of $1DKL_{open}$ and $1DKP_{closed}$ were concatenated for better comparison of clusters. Clustering of concatenated trajectory (only loop residues) using RMSD-based cutoff of 0.08 nm resulted in a total of 75 clusters. First 50 ns in concatenated trajectory correspond to the clustering of $1DKP_{closed}$ trajectory, whereas last 50 ns represent clusters of $1DKL_{open}$ trajectory (Figure 2a). Three main clusters for both $1DKL_{open}$ and $1DKP_{closed}$ covering most of the trajectory were selected. Approximately, 33 % of the snapshots from concatenated trajectory were represented in cluster 1, 4 and 7 (renamed to 1, 2 and 3) correspond to $1DKP_{closed}$ trajectory and 36 % were represented in cluster 2, 3 and 5 (renamed to 1´, 2´ and 3´) correspond to $1DKL_{open}$ trajectory. A representative structure of each cluster from $1DKP_{closed}$ (Figure 2b) or $1DKL_{open}$ (Figure 2c) were superimposed with starting conformation of either $1DKL_{open}$ or $1DKP_{closed}$. Transition from the closed loop conformation of starting structure or cluster 3 ($1DKP_{closed}$) to the open loop conformations observed in cluster 2 and 1 is shown in Figure 2b. The most closed loop structures were observed in cluster 3 ($1DKP_{closed}$) which represents only ~5.4 % of concatenated trajectory. Whereas, ~28 % of concatenated trajectory corresponds to cluster 1 and 2 ($1DKP_{closed}$), was found to be open conformation. In case of $1DKL_{open}$, more opening of open loop conformation was observed (cluster 1´, 2´ and 3´).

Cluster analysis of protein structures (all protein residues) has been discussed in supporting information. Clustering of all protein residues also showed a similar trend of active site loop opening. Intermolecular crystal contacts bury a significant portion of solvent accessible area and these contacts were found to influence the backbone conformation of the protein [37]. Crystal contact analysis on all the

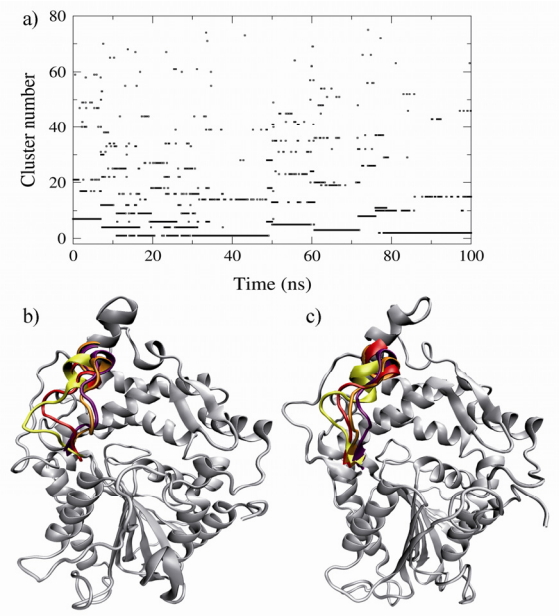


Figure 2: Sampling of active site loop clusters over 50 ns of concatenated trajectory of $1DKP_{closed}$ (0 to 50 ns) and $1DKL_{open}$ (50 to 100 ns) MD simulation (a); Representative conformations of the active site loop from the clustering of loop residues generated from 50 ns of $1DKP_{closed}$ (b) and $1DKL_{open}$ (c) MD simulation. Loop conformations of $1DKP_{closed}$ in cluster 1 (yellow), cluster 2 (red), cluster 3 (orange) and starting structure (purple) are depicted. The loop conformations of $1DKL_{open}$ in cluster 1´ (yellow), cluster 2´ (red), cluster 3´ (orange) and starting structure (purple) are shown.

available *E. coli* phytase crystal structure using CryCo web server (http://ligin.weizmann.ac.il/cryco) and PyMOL software [18] showed involvement of active site loop residues in crystal contact (Supporting Information, Table S1 and Figure S5). Therefore, the tendency of the loop to have more open conformation in the trajectory can be the result of a lack of crystal contacts with neighboring protein replicas. The missing exploration of the open structure conformations from the $1DKP_{closed}$ trajectory can be the result of an inadequate sampling or a trapping of the loop in local conformational minima with higher barrier.

9.3.4 Principal component analysis

Functionally relevant correlated motions, significant to loop opening were identified using PCA of $1DKP_{closed}$ and $1DKL_{open}$ trajectories. For both trajectories, PCA was carried out for protein backbone atoms that showed at least 80 % of the overall positional fluctuation was confined by the first 27 and 22 eigenvectors out of ~3690 eigenvectors for the $1DKP_{closed}$ and $1DKL_{open}$ structure respectively. First eigenvectors contributed 34 % and 37 % and the second for 13 % and 10 % of the total functional fluctuations of $1DKP_{closed}$ and $1DKL_{open}$ trajectory respectively, thus showing the first eigenvector to dominate the total motions in both phytase structures (Supporting Information, Figure S6).

The 2D graph of the projection of both $1DKP_{closed}$ (green) and $1DKL_{open}$ (black) trajectory onto the first two corresponding eigenvectors obtained from all backbone atoms (Figure 3a) or active site loop backbone atoms (Figure 3b) is reported. For 2D projection of the trajectories, in both the cases, the covariance matrices were generated using $1DKL_{open}$ as a reference structure and analyzed. In the same graph, the positions for the conformation of X-ray open, X-ray closed and conformation of the median structure from each clusters obtained from cluster analysis of active site loop residues are reported. In both $1DKP_{closed}$ and $1DKL_{open}$ protein, 2D projection of the trajectory has found to have three main clusters of conformation sampling (Figure 3b). These clusters were originating from one closed loop cluster observed at the beginning of simulations,

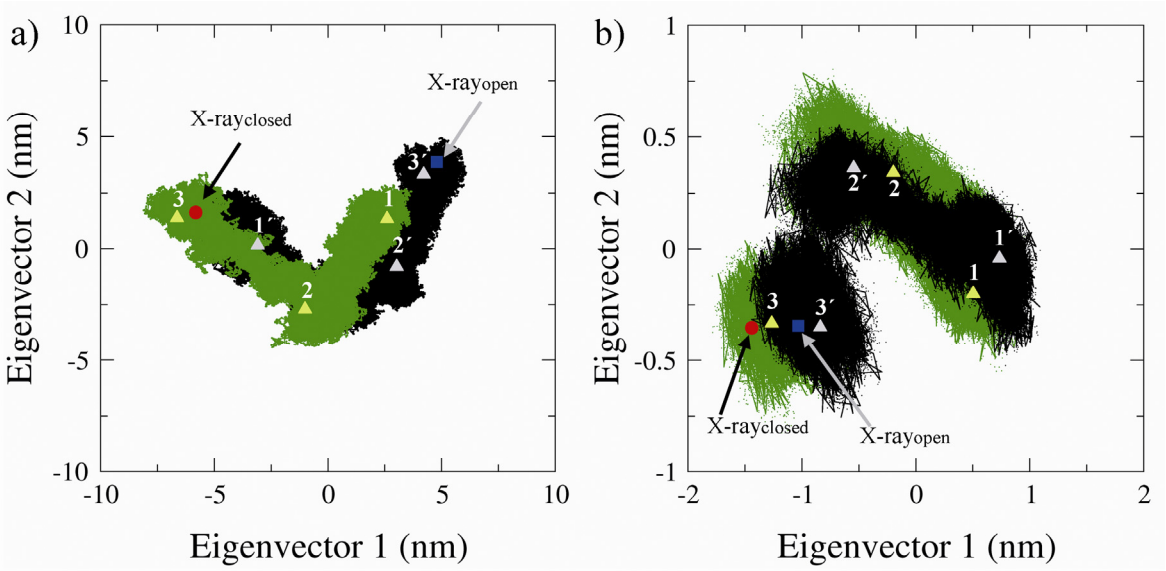


Figure 3: The 2D projection of $1DKP_{closed}$ (green) or $1DKL_{open}$ (black) trajectory onto the corresponding first two principle components obtained from all backbone atoms covariance matrix (a) or active site loop backbone atoms covariance matrix (b). The X-ray $1DKP_{closed}$ (red circle) and $1DKL_{open}$ (blue square) loop conformation structures are highlighted. Median structure from each clusters of $1DKP_{closed}$ (1, 2 and 3) and $1DKL_{open}$ (1´, 2´ and 3´) analyzed using the active site loop cluster analysis are shown.

that was possessing cluster 3 and 3´ of the $1DKP_{closed}$ and $1DKL_{open}$ trajectory respectively. The clusters 1 ($1DKP_{closed}$) and 1´ ($1DKL_{open}$) were located in close proximity to each other, which indicates that at the end of both the simulations, the protein conformations converge to a similar more open loop conformation. Finally, both the trajectories, projected on corresponding first eigenvector evidenced that the largest motion was occured in the active site loop area of both $1DKP_{closed}$ and $1DKL_{open}$ systems (Supporting Information, Figure S7).

9.3.5 DomainSelect analysis

DomainSelect analysis was performed to analyze the motion of active site loop during the MD simulations. The effective hinge axis was calculated between each of the representative structure from first three largest clusters of active site loop residues and corresponding starting conformation of $1DKP_{closed}$ or $1DKL_{open}$ structure. DomainSelect program assumes that the displacement in moving active site loop is a rigid body movement, therefore the average internal RMSD between loop residues of first three clusters from both the trajectory was calculated. The average RMSD for the loop residues from both the trajectories was 0.059 nm, covering ~70 % of the concatenated trajectory (Supporting Information, Table S2). The structures from each cluster showed lower RMSD in corresponding cluster conformation therefore the movement in active site loop is a hinge bending motion. The effective hinge axis between median structure of first largest cluster 1 or 1´ and starting conformation of $1DKP_{closed}$ or $1DKL_{open}$ respectively, is shown in Figure 4 and the directional motion of loop opening is shown

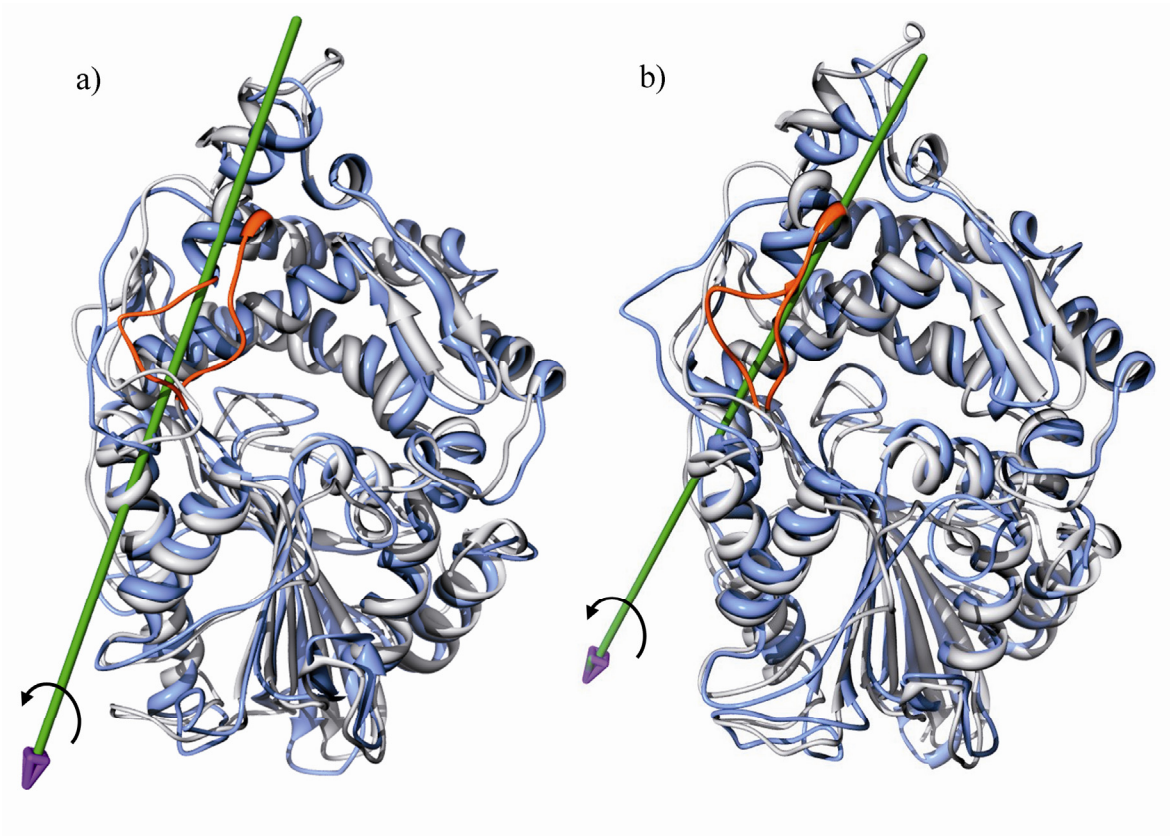


Figure 4: DomainSelect analyses of difference between closed loop conformation in starting experimental X-ray structure (grey) and open loop conformation in representative structure (blue) of most abundant cluster 1 from $1DKP_{closed}$ trajectory (a). Conformational changes in open loop conformation of starting X-ray structure (grey) and loop conformation in representative structure (blue) of most abundant cluster 1´from $1DKL_{open}$ trajectory (b) are depicted using DomainSelect analysis. Active site loop residues (moving part; Ala20 – Gln26) are shown in red color. Green arrow depicting effective hinge axis of loop transition to open conformation in MD simulation. Black curved arrow specifies the opening direction of active site loop. (**COVER PAGE**)

in black curved arrow. Table 1 summarizes the results of DomainSelect analysis on all active site loop conformations compared with starting conformation. The $1DKP_{closed}$ simulation shows a 19-degree loop opening in first ~7 ns that reached to 127-degree of hinge bending angle in median structure of most abundant cluster 1 of the trajectory. In case of $1DKL_{open}$, open loop was found to be more opened from 23-degree opening in first ~6 ns, which reached 47-degree angle of rotation when compared to open loop conformation as a reference structure. The median structure from cluster 1´ (abundant cluster of $1DKL_{open}$) showed 117-degree angle of rotation when compared to closed loop X-ray crystal structure ($1DKP_{closed}$) as a reference structure, which is almost similar opening in cluster 1 corresponds to $1DKP_{closed}$ trajectory. Furthermore, angle of rotation between two extreme structures obtained from PCA of $1DKP_{closed}$ trajectory showed 124-degree opening. This suggests that the conformational dynamics of the active site loop leads to a “more

Table 1: Relative active site loop motion along most abundant active site loop conformations from first three clusters of 1dkp or 1dkl trajectories (fixed domain α: Ser1 – Arg19, Leu27 – Leu409; moving domain β (loop): Ala20 – Gln26)

Enzyme model	1dkp§ (Close Structure)				1dkl§ (Open Structure)		
	Cluster 3	Cluster 2	Cluster 1	Cluster 2´	Cluster 3´	Cluster 2´	Cluster 1´
Angle of rotation (°)	18.7	106.5	126.8	116.9	23.4	65.8	47.4
Translation along axis (Å)	1.189	2.353	4.56	3.755	0.605	1.893	4.204
Distance* (Å)	0.137	1.980	2.349	2.819	1.972	2.273	3.781
Percentage closure motion	28.144	49.66	48.515	62.62	39.068	40.136	64.78

*Distance between hinge axis and line connecting centers of mass of domain α and domain β

§Reference conformation used for comparison

open" conformation compared to X-ray crystal structure (see above).

9.3.6 Active site analysis of E. coli phytase

In the absence of substrate (phytate), MD simulation showed that the *E. coli* phytase has increased flexibility in active site in both closed and open loop conformation enzyme trajectory. To investigate the active site cavity during the simulation, distances between the catalytically important residues were measured. Two extreme structures calculated by projecting $1DKP_{closed}$ trajectory onto first principal component were aligned with X-ray crystal structure (Figure 5). Distances between the Cα atoms of residues Asp 304, Glu 219, Arg 20 and Arg 16 are shown in Table 2. Residue Arg 16 and Arg 20 were found to be involved in coordination with scissile phosphate of phytate, whereas Asp 304 and Glu 219 are involved in catalysis [13]. Distances measured between the backbone atoms (Cα) of these potentially important catalytic residues were found to be not deviating much from the reported experimental X-ray crystal structure. This suggests that the active site cavity of *E. coli* phytase in both the trajectory was not altered during simulation and only flexible part was composed of active site loop.

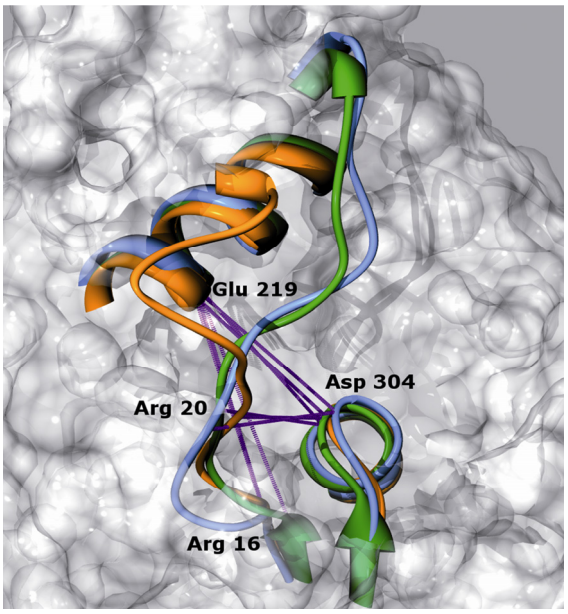


Figure 5: Structural alignment of $1DKP_{closed}$ X-ray crystal structure (green) with two extreme structures (blue/orange) obtained by projecting $1DKP_{closed}$ MD simulation trajectory onto the first principle component. Important active site residues involved in substrate binding are shown, purple colored dotted line indicate distance between Cα atoms of active site residues. Active site cavity is behind the loop (green/blue) that makes loop acting as a door to binding pocket.

Table 2: Distances measured between Cα atom of the active site residues of X-ray structure and two extreme structures obtained from PCA of 1DKP simulation.

Enzyme Model	Distance between Asp 304 and Glu 219 (Å)	Distance between Arg 16 and Glu 219 (Å)	Distance between Asp 304 and Arg 20 (Å)
1DKP (X-ray structure)	11.21	13.64	12.60
Extreme Structure 1 (close)	11.29	14.42	11.69
Extreme Structure 2 (open)	11.69	12.94	11.13

PCA and clustering of the simulation reveal the loop to open, causing loop to move back away from the active site, exposing the active site to the bulk solvent. Furthermore, loop residues Thr 23 and Thr 24 are frequently found far from the binding pocket, these residues were reported to flip inside upon substrate binding. Our results in support of proposed mechanism of action of phytase catalysis [13] suggest that the only flexible region in active site comprised of active site loop that might enhance the catalysis upon binding of phytate by adopting several favorable conformations that would help in binding with desired substrate conformation into the active site cavity.

Despite a very low sequence identity, phytases from fungi, bacteria and mammals have nearly identical structural topology with two domains, a conserved α/β domain and α domain with small variations near the active site [13]. To investigate conservative patterns of active site loops, we have performed multiple sequence alignment of *E. coli* phytase with other enterobacteriaceae, fungal (commercially used) and yeast phytases, only Pro 22 and Thr 23 were found to be highly conserved and other amino acids were variable (Supporting Information, Figure S8). Pro 22 belongs to the conserved sequence motif “RHGXRXP” of the histidine acid phosphatase family. The active site loop is conserved to a great extent in highly active phytases from the enterobacteriaceae family (*E. coli* and *Yersinia* species) in contrast to phytases showing lower catalytic efficiencies (phytases from fungi, yeast and *Klebsiella*). During the time of manuscript revision, the low active phytases from enterobacteriaceae family *(Klebsiella* phytase) has been crystallized [38]. Interestingly, the *Klebsiella* phytase (PDBID: 2WNH) bear an similar overall structural topology to *E. coli* phytase, however, the active site loop was replaced by an α-helix in the active site (Supporting Information, Figure S9). Fungal phytases are about 30-fold less active than enterobacteriaceae phytases that possesses active site loop (Supporting Information, Table S3). Higher activity of phytases from enterobacteriaceae family might be due to the identified flexible loop. A deeper understanding of dynamics during the phytase catalysis could be possible by further MD simulations of equilibrated open structures obtained from these trajectories in presence of substrate and different loop residue substitutions. Additional saturation mutagenesis experiments of this loop residues might further broaden our understanding of catalytic efficiency of phytases

and improve phytase activity from fungi, yeast or *Klebsiella* species.

9.4 Conclusions

The large-scale MD simulations were performed on two different conformations of *E. coli* phytase in absence of phytate and different analysis techniques were used to investigate functionally important movement in *E. coli* phytase. Cluster analysis and PCA of trajectories reveal that the most abundant conformation during MD simulation was opening of active site loop, which moves away from the active site in absence of phytate. The movement in the active site is mainly confined by the active site loop. Our results in support of proposed mechanism of action of phytase catalysis [13] suggest that the flexible active site loop might enhance catalysis by adopting favorable conformations that would help in binding with desired substrate conformation into the active site cavity. The high flexibility might be prerequisite for higher catalytic efficiency of phytases. Our active site loop dynamics studies illustrate a step towards understanding higher catalytic efficiency in different phytases.

Acknowledgement

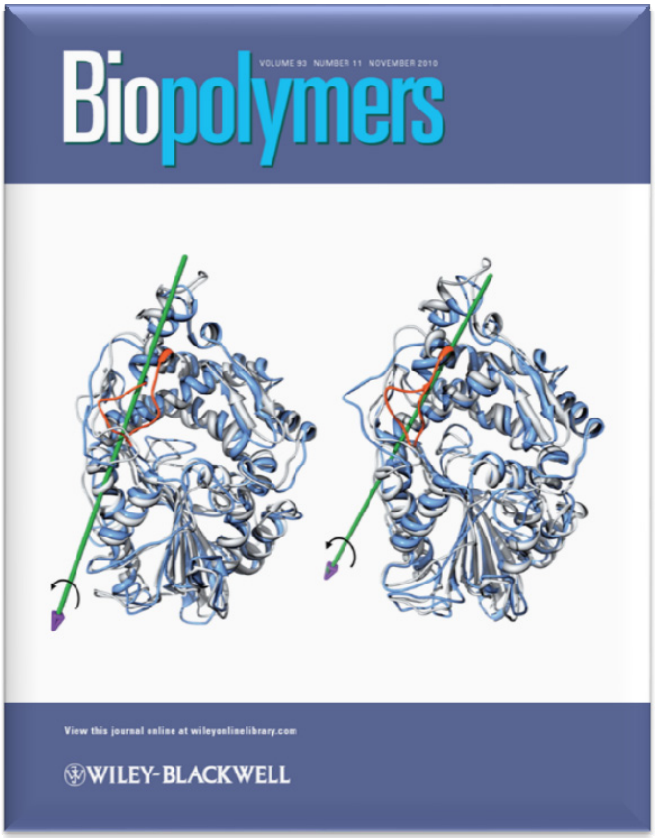
The part of the chapter was “Reprinted from Biopolymers, 93, Shivange AV, Schwaneberg U, Roccatano D., Conformational dynamics of active site loop in *Escherichia coli* phytase., 994-1002, Copyright (2010), with permission from John Wiley and Sons”.

9.5 References

1. Bohn, L., A.S. Meyer, and S.K. Rasmussen, *Phytate: impact on environment and human nutrition. A challenge for molecular breeding.* J Zhejiang Univ Sci B, 2008. 9(3): p. 165-91.
2. Bitar, K. and J.G. Reinhold, *Phytase and alkaline phosphatase activities in intestinal mucosae of rat, chicken, calf, and man.* Biochim Biophys Acta, 1972. 268(2): p. 442-52.
3. Sebastian, S., S.P. Touchburn, and E.R. Chavez, *Implications of phytic acid and supplemental microbial phytase in poultry nutrition: a review.* World Poultry Sci J, 1998. 54(01): p. 27-47.
4. Wodzinski, R.J. and A.H. Ullah, *Phytase.* Adv Appl Microbiol, 1996. 42: p. 263-302.
5. Sandberg, A., et al., *Inositol phosphates with different numbers of phosphate groups influence iron absorption in humans.* Am J Clin Nutr, 1999. 70(2): p. 240-246.
6. Cavazza, C., *Carnitine and inositol phosphate-containing composition useful as dietary supplement or drug.* 2000, Sigma-tau Healthscience S.P.A. Cavazza, Claudio: WO.
7. Greiner, R. and U. Konietzny, *Construction of a bioreactor to produce special breakdown products of phytate.* J Biotechnol, 1996. 48(1-2): p. 153-9.
8. Billington, D.C., *The Inositol phosphates: chemical synthesis and biological significance*, ed. D.C. Billington. 1993, Weinheim: VCH Verlagsgesellschaft.
9. Wyss, M., et al., *Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties.* Appl Environ Microbiol, 1999. 65(2): p. 367-73.
10. Haefner, S., et al., *Biotechnological production and applications of phytases.* Appl Microbiol Biotechnol, 2005. 68(5): p. 588-97.
11. Huang, H., et al., *A novel phytase from Yersinia rohdei with high phytate hydrolysis activity under low pH and strong pepsin conditions.* Appl Microbiol Biotechnol, 2008. 80(3): p. 417-26.
12. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. 28(1): p. 235-42.
13. Lim, D., et al., *Crystal structures of Escherichia coli phytase and its complex with phytate.* Nat Struct Biol, 2000. 7(2): p. 108-13.
14. Amadei, A., A.B.M. Linssen, and H.J.C. Berendsen, *Essential dynamics of proteins.* Proteins, 1993. 17(4): p. 412-425.
15. Daura, X., et al., *Peptide Folding: When Simulation Meets Experiment.* Angew Chem Int Edit, 1999. 38(1-2): p. 236-240.
16. Hayward, S. and H.J.C. Berendsen, *Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme.* Proteins, 1998. 30(2): p. 144-154.
17. Tsou, C.L., *The role of active site flexibility in enzyme catalysis.* Biochemistry (Mosc), 1998. 63(3): p. 253-8.
18. DeLano, W.L., *The PyMOL Molecular Graphics System.* 2002: Palo Alto, CA, USA.
19. Wyss, M., et al., *Biophysical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): molecular size, glycosylation pattern, and engineering of proteolytic resistance.* Appl Environ Microbiol, 1999. 65(2): p. 359-366.
20. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation.* Comput Phys Commun, 1995. 91(1-3): p. 43-56.

21. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* J Mol Model, 2001. 7(8): p. 306-317.
22. Scott, W.R.P., et al., *The GROMOS Biomolecular Simulation Program Package.* J Phys Chem A, 1999. 103(19): p. 3596-3607.
23. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath.* J Chem Phys, 1984. 81(8): p. 3684-3690.
24. Berendsen, H.J.C., et al., *Interaction models for water in relation to protein hydration.* Intermolecular Forces, 1981: p. 331-342.
25. Ryckaert, J., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* J Comput Phys, 1977. 23(3): p. 327-341.
26. Miyamoto, S. and P. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models.* J Comput Chem, 1992. 13(8): p. 952-962.
27. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems.* J Chem Phys, 1993. 98(12): p. 10089-10092.
28. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. 25(13): p. 1605-1612.
29. Roccatano, D., A.E. Mark, and S. Hayward, *Investigation of the mechanism of domain closure in citrate synthase by molecular dynamics simulation.* J Mol Biol, 2001. 310(5): p. 1039-1053.
30. Hayward, S., *Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements.* J Mol Biol, 2004. 339(4): p. 1001-21.
31. Hayward, S. and A. Kitao, *Molecular dynamics simulations of NAD+-induced domain closure in horse liver alcohol dehydrogenase.* Biophys J, 2006. 91(5): p. 1823-31.
32. Roccatano, D., et al., *Dynamical Aspects of TEM-1 β-Lactamase Probed by Molecular Dynamics.* J Comput-Aided Mol Des, 2005. 19(5): p. 329-340.
33. Coiro, V.M., et al., *Molecular dynamics simulation of the interaction between the complex iron-sulfur flavoprotein glutamate synthase and its substrates.* Protein Sci, 2004. 13(11): p. 2979-91.
34. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. 26(16): p. 1701-18.
35. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. 14(1): p. 33-8, 27-8.
36. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. 22(12): p. 2577-637.
37. Eyal, E., et al., *The limit of accuracy of protein modeling: influence of crystal packing on protein structure.* J Mol Biol, 2005. 351(2): p. 431-42.
38. Bohm, K., et al., *Crystal structure of Klebsiella sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme.* Febs J, 2010. 277(5): p. 1284-96.

Published in Biopolymers, 2010, Volume 93, Issue 11, pages 994–1002 (COVER PAGE)



Supporting Information: Chapter 9

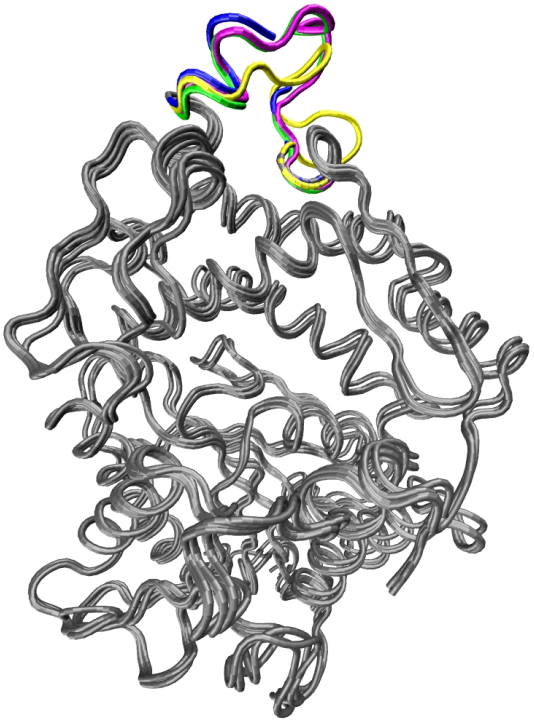
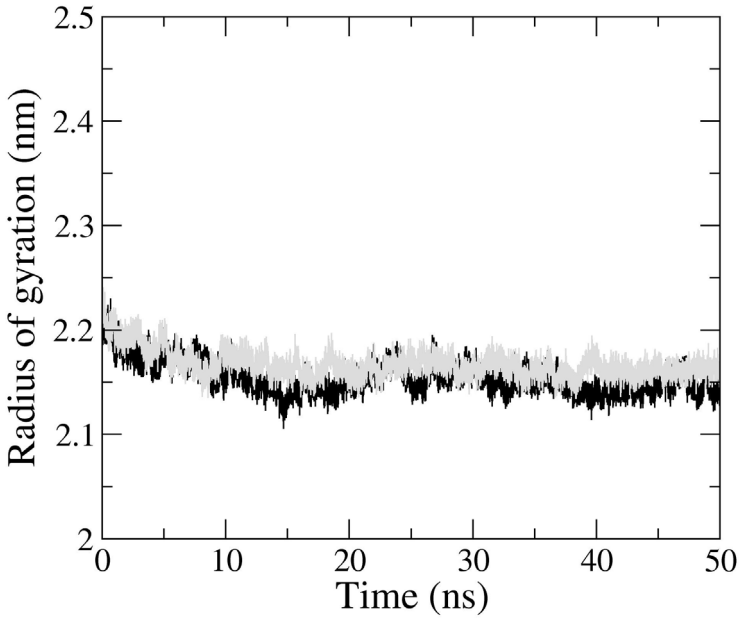


Figure S1: Superimposition of *E. coli* phytase crystal structures. A missing loop (Arg181 - Ser187) in 1DKL crystal structure (blue) was modelled using ModLoop server (yellow) showed undesirable conformation compare to loop modelled by providing dummy atoms (magenta) from another unliganded phytase conformation (green, PDB ID: 1DKO) The figure was generated using the molecular graphics program VMD [1].



Time (ns)	Average Rg Values (nm)		Variation (%)	
	1DKP	**1DKL**	**1DKP**	**1DKL**
StartConf	2.19	2.20	0.00	0.00
0 to 20	2.16	2.17	-1.20	-1.17
20 to 50	2.15	2.16	-1.61	-1.64

Figure S2: Radius of gyration of the backbone atoms (Cα, N, C) of *E. coli* phytase protein throughout the simulation of closed active site loop structure 1DKP (black) and open active site loop structure 1DKL (grey). Average Rg values are shown for both simulations in the table and compared with the Rg value of corresponding starting experimental crystal structure conformation (StartConf). Variation (%) shows the percent decrease in Rg value during simulation with respect to corresponding starting structure.

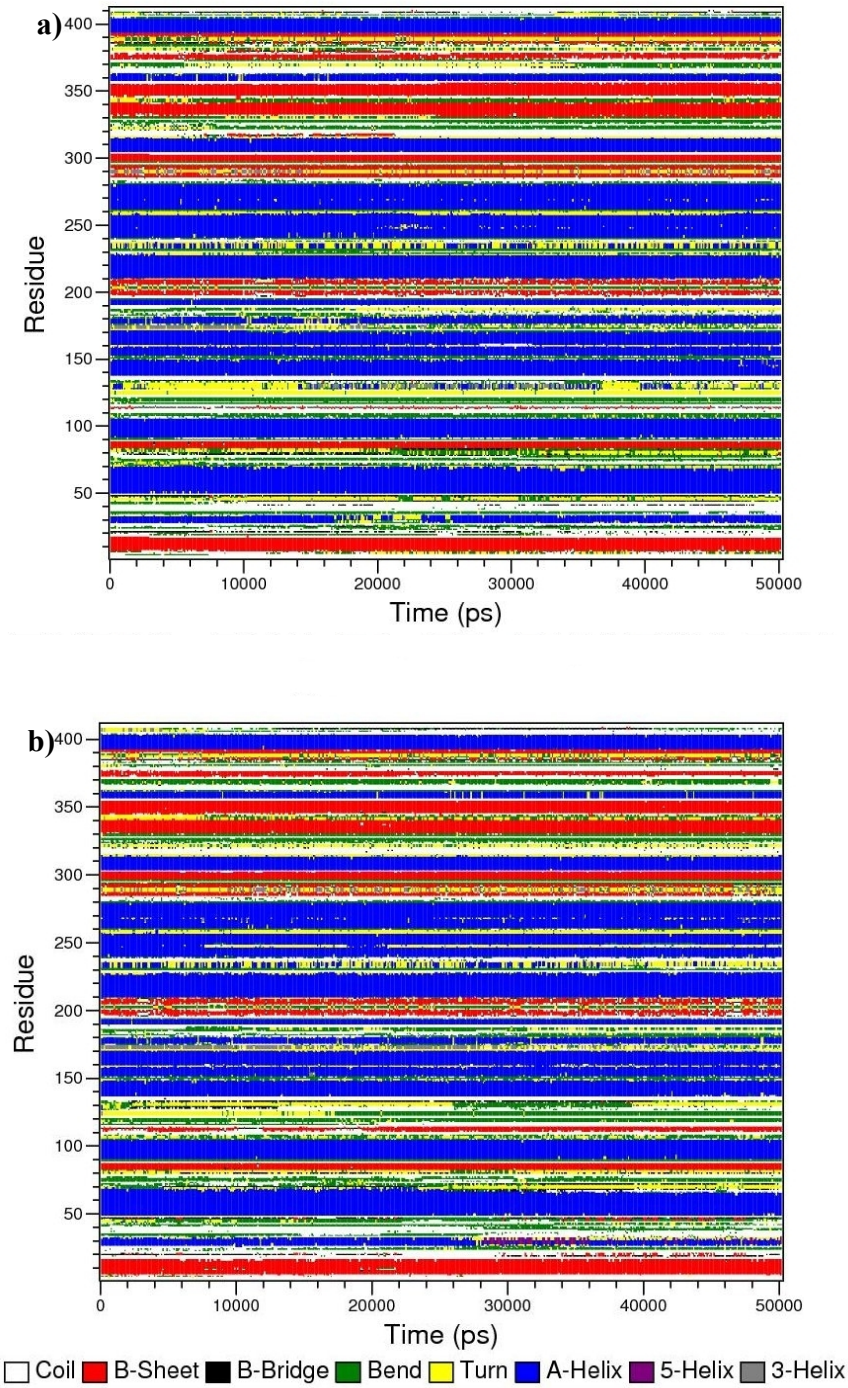
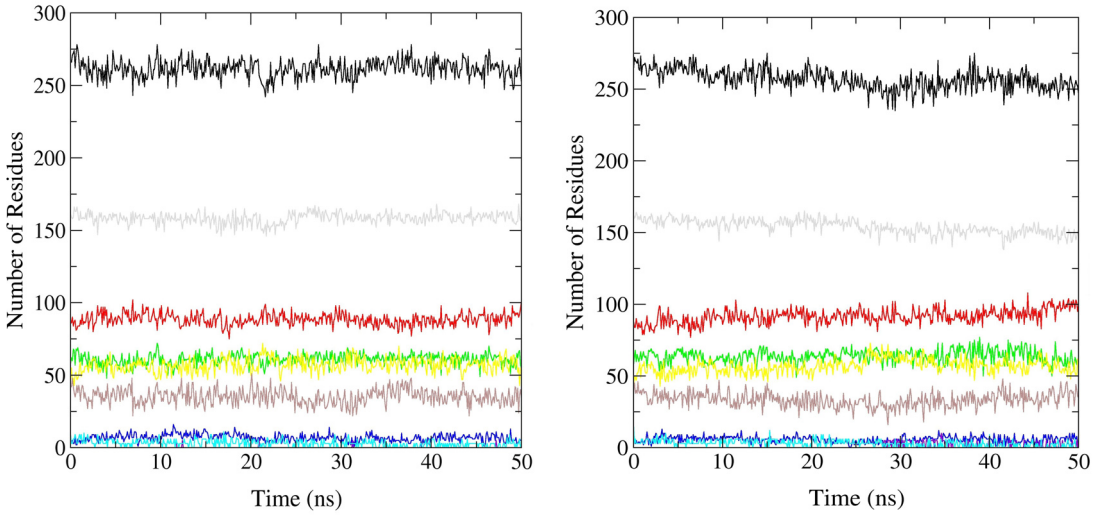


Figure S3: Secondary structures along the MD simulation of 1DKP (a) and 1DKL (b) protein. The different secondary structure elements are depicted using different colors. Secondary structures were computed along the trajectory using DSSP program [2].



Seconday Structural Element	Average Residues		Average Residues (%)		Standard Deviation (%)	
	1DKP	1DKL	1DKP	1DKL	1DKP	1DKL
All (Structure)	261.42	256.83	63.76	62.80	± 2.41	± 2.83
B-Sheet	61.03	63.82	14.88	15.60	± 6.40	± 7.09
A-Helix	158.30	153.99	38.61	37.65	± 2.36	± 3.18

Figure S4: Total number of residues belonging to each specified secondary structures type along the MD simulation of 1DKP (a) and 1DKL (b) protein. The different types of secondary structural elements are shown using different colors [structure (■), coil (■), B-sheet (■), B-bridge (■), bend (■), turn (■), A-helix (■), 5-helix (■) and 3-helix (■)]. Average number of residues involved in secondary structure was shown in the table with % standard deviation during the simulation. Secondary structures were computed along the trajectory using DSSP program [2].

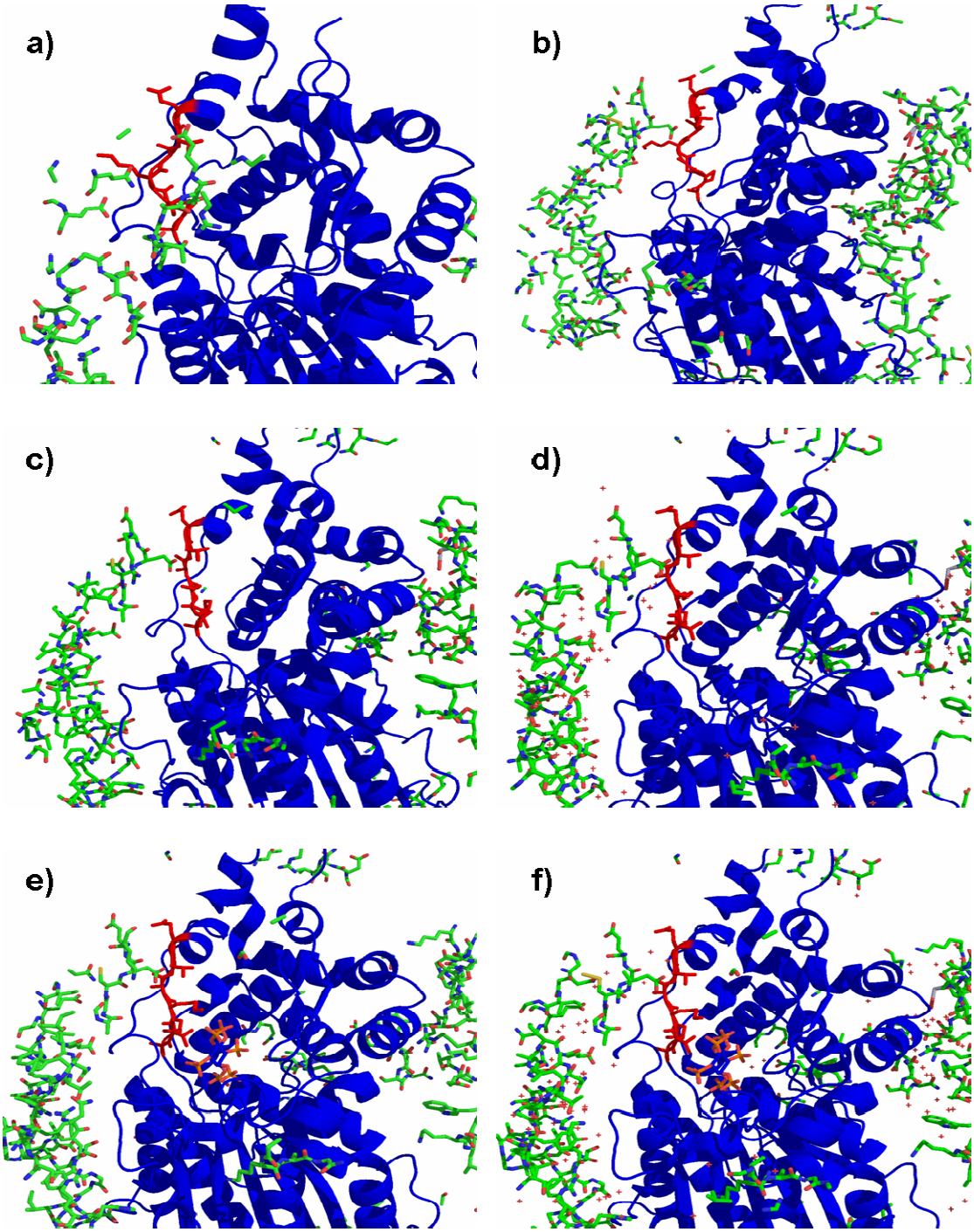


Figure S5: Crystal contact analysis for *E. coli* phytase crystal structure 1DKL (a), 1DKM (b), 1DKN (c), 1DKO (d), 1DKP (e) and 1DKQ (f). Crystal symmetry was generated using PyMOL software [3] and residues in symmetry molecules (green) within the area of 1 nm (cutoff) around the phytase (blue) is shown. The active site loop residues are displayed in red color.

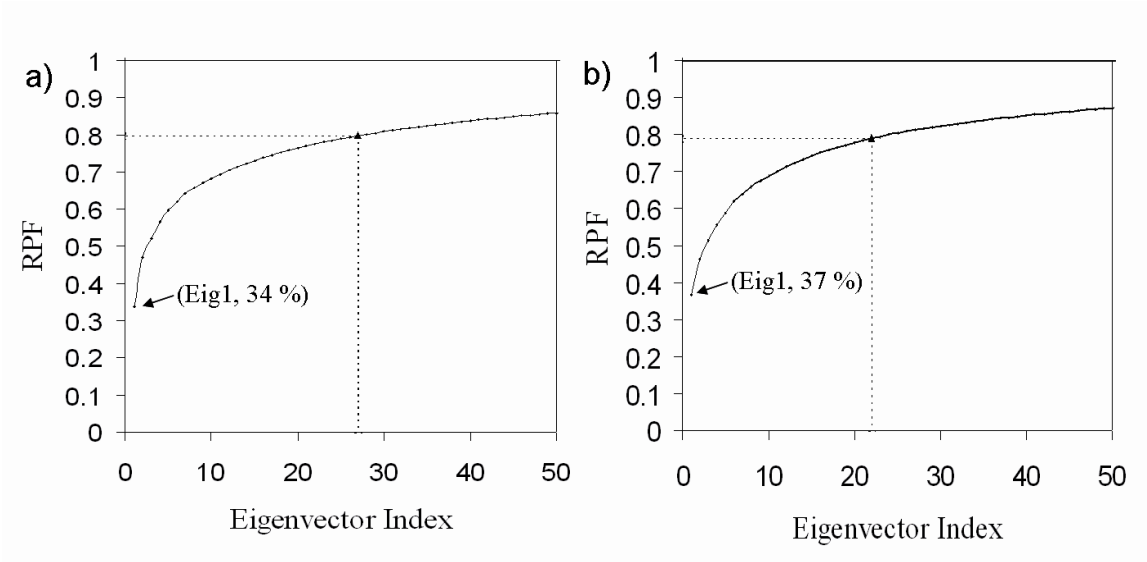


Figure S6: Relative positional fluctuation (RPF) of the motion along the eigenvectors obtained from backbone coordinates covariance matrix of $1DKP_{closed}$ (a) and $1DKL_{open}$ (b). Dotted line indicates eigenvectors needed to cover 80% of the overall positional fluctuation, positional fluctuation contribution of first eigenvector is shown in the bracket.

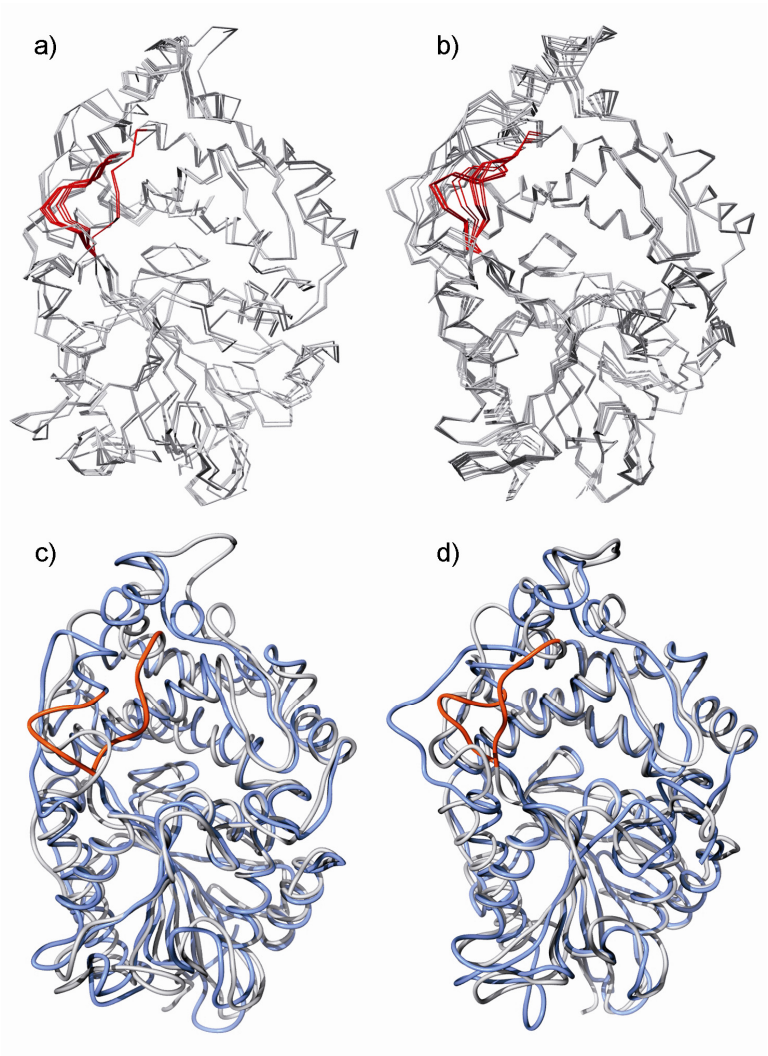


Figure S7: Superimposition of 11 configurations obtained by projecting the protein backbone motion onto the first principle component of 1DKPclosed (a) or 1DKLopen (b), configurations are separated by 5ns; Structural alignment of two extreme structures acquired by projecting 1DKPclosed (c) or 1DKLopen (d) MD simulation trajectory on the corresponding first principle component. Residues of active site loop (Ala20 – Gln26) are highlighted in red color.

The largest directional motion was observed in the active site loop of both $1DKP_{closed}$ and $1DKL_{open}$ trajectories. In the case of $1DKP_{closed}$, the active site loop was first opened due to absence of substrate (phytate) that reached to similar conformation as open loop crystal structure ($1DKL_{open}$) and high positional fluctuations were observed after opening, which are similar to positional fluctuations of the active site loop in $1DKL_{open}$ trajectory.

In both the cases one extreme conformation of active site loop was similar to starting conformation of either $1DKP_{closed}$ or $1DKL_{open}$ structure, and other conformation was more open active site loop.

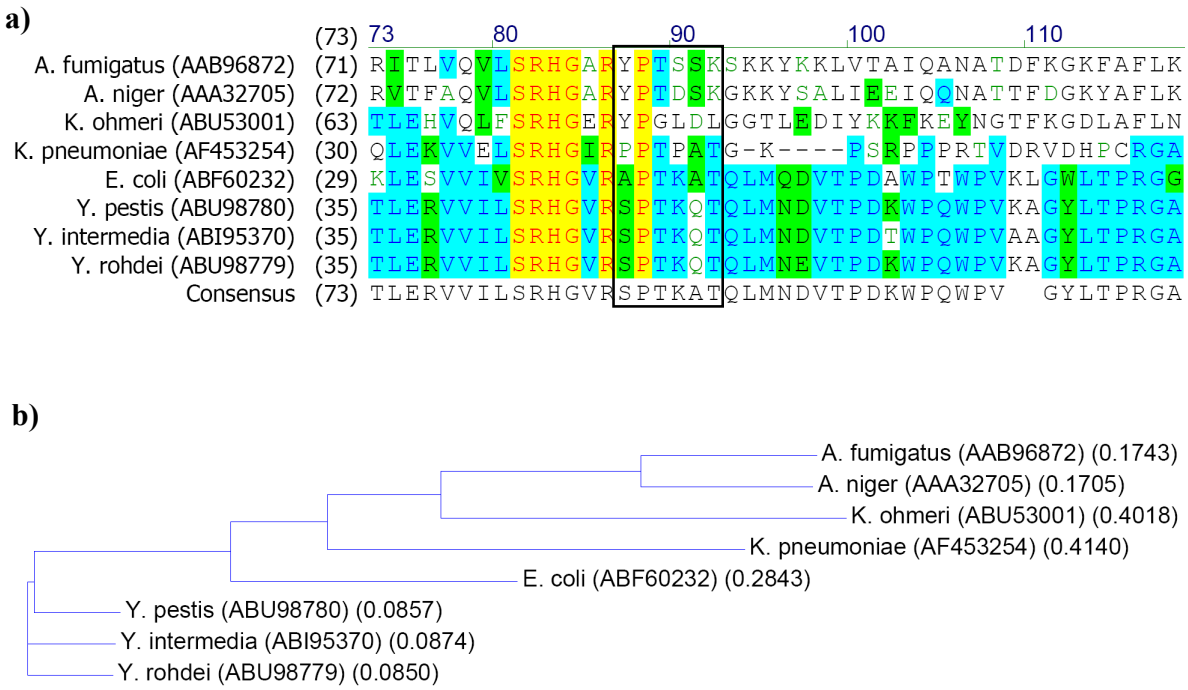


Figure S8: Multiple sequence alignment (a) and phylogenetic tree (b) showing the evolutionary relatedness and level of homology between phytases from fungi (*Aspergillus niger* and *A. fumigatus*), yeast (*Kodamaea ohmeri* BG3) and enterobacteriaceae family (*E. coli, Yesinia pestis, Y. rohdei, Y. intermedia* and *K. pneumoniae)*. Active site loop residus are depicted in the box. All sequences were retrived from GenBank database. Alphanumeric number in bracket indicates accession number of the database. Identical, conserved and similar residues are colored in yellow, cyan and green color respectively.

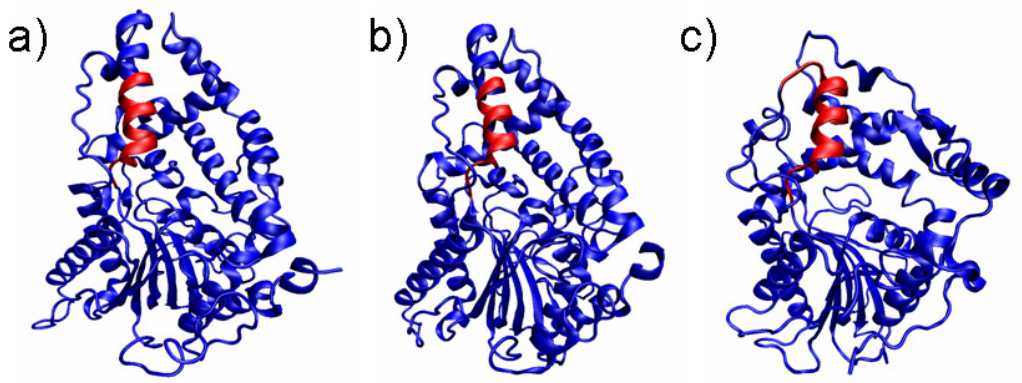


Figure S9: Active site loop region in fungal phytases [a (PDBID: 1SK8); b (PDBID: 1IHP)] and *Klebsiella* phytase [c (PDBID 2WNH)]

Note: Crystal structure of *K. pneumoniae* phytase (PDBID: 2WNH) was released. The active site loop was found to be replaced by an α-helix, this in support of our hypothesis that the active site loop is necessary for higher catalytic efficiency.

Table S1: Active site loop residues participating in crystal contact and area involved in crystal contact as calculated using CryCo program [4]

Note: 1DKL has two monomers in PDB database, CryCo analysis on a single monomer could not identify any residues involved in crystal contact

Residue Number	Crystal contact area (Å^2)					
	1DKL	**1DKM**	**1DKN**	**1DKO**	**1DKP**	**1DKQ**
Thr 23	-	13.6	-	0.5	0.5	-
Lys 24	-	70.0	18.7	31.5	7.8	1.3
Ala 25	-	4.1	16.2	22.6	23.1	9.4
Thr 26	-	31.0	24.3	22.5	24.4	25.1
Gln 27	-	6.0	2.9	12.5	9.9	15.8

Table S2: RMSD values for the active site loop residues in structures of each cluster

Phytase trajectory	Cluster number	Total no. of structures in cluster	Percent trajectory coverage (%)	RMSD (nm)
1DKP	1	2017	20.2	0.057
	2	766	7.7	0.065
	3	543	5.4	0.059
1DKL	1´	1842	18.4	0. 058
	2´	969	9.7	0.057
	3´	748	7.5	0.059
Average active site loop RMSD				0.059

Table S3: Reported specific activity of phytases from fungi*, yeast$ and bacteria§ for phytate as a substrate

Source of Phytase	Specific Activity (U mg^{-1})	Reference
*A. niger**	99±4.2	[5]
*A. fumigatus**	43	[6]
K. ohmeri$	114.6	[7]
E. coli§	3191±186	[5]
Y. pestis§	2344±163	[5]
Y. intermedia§	3960±348	[5]
Y. rohdei§	2456±197	[5]
K. pneumoniae§	169	[8]

General structural properties

Total numbers of residues belonging to different types of secondary structural elements were calculated as a function of MD simulation time. The percentage of each secondary structural element throughout the simulation was found to be almost constant as compared to starting conformation in both $1DKP_{closed}$ and $1DKL_{open}$ trajectories. The secondary structures calculated using DSSP program [2] showed an average of 63.8% (± 2.4%) and 62.8% (± 2.8%) of the residues contributed to all secondary structures (α-Helix, β-Sheet, β-Bridge and Turn,) in $1DKP_{closed}$ and $1DKL_{open}$ simulation respectively. Total residues forming β-sheets were 14.9% (± 6.4 %) and 15.6% (± 7.1%) and α-helices were 38.6% (± 2.4%) and 37.65% (± 3.2%) for $1DKP_{closed}$ and $1DKL_{open}$ trajectory respectively (Figure S4). Average backbone RMSD values of 0.29 nm and 0.27 nm was found for the $1DKP_{closed}$ and $1DKL_{open}$ respectively. The average RMSD reached, after first 20 ns in each simulation, a value of 0.31 nm and 0.30 nm in $1DKP_{closed}$ and $1DKL_{open}$ trajectory respectively. The comparison with respect to the same starting open structure ($1DKL_{open}$) showed an average RMSD value of 0.29 nm and 0.30 nm for last 30 ns of $1DKP_{closed}$ and $1DKL_{open}$ simulation respectively. Slightly higher average RMSD value using $1DKP_{closed}$ as a reference structure illustrates a rough estimation that the conformations in $1DKP_{closed}$ trajectories are more similar to open ($1DKL_{open}$) crystal structure. The average Rg value was decreased by 1.61% and 1.64% after 20 ns (from 2.19 nm to 2.15 nm in $1DKP_{closed}$ and from 2.20 nm to 2.16 nm in $1DKL_{open}$) with respect to corresponding starting experimental structure for $1DKP_{closed}$ and $1DKL_{open}$ respectively (Figure S2). Similar trend of radius of gyration in both the cases showed formation of a more compact structure with respect to the starting one.

Cluster Analysis of E. coli phytase

Cluster analysis of protein structures

Clustering of each trajectory resulted in a total of 25 clusters for the $1DKL_{open}$ and 21 clusters for the $1DKP_{closed}$ simulation. Figure S10 shows sampling and transition between clusters of $1DKP_{closed}$ (Figure S10a) and $1DKL_{open}$ (Figure S10d) as a function of simulation time.

In the case of the $1DKP_{closed}$ simulation, it is clearly shown in the Figure S10a that the first three clusters represent ~85% of the generated conformations. Furthermore, the cumulative curve shows also that the number of clusters decreased after 20 ns, total 17 out of 21 clusters corresponds to first 20 ns and only 4 clusters were obtained in last 30 ns of the simulation (Figure S10b). The median structures representing each of the first three clusters were extracted and superimposed. Transition from closed loop conformation (cluster 3) to open loop conformation (cluster 2 or 1) is shown in Figure S10c. Loop conformation in cluster 3 (orange) was similar to loop conformation in crystal structure (purple) whereas in case of cluster 2 and 1, (red and yellow respectively) open loop conformation was observed.

In case of $1DKL_{open}$, the cluster analysis shows a more distributed conformation variety. In fact, the first three clusters covered only 64% of the MD conformations and 7 clusters were needed to reach the 85% (see cumulative cluster curve in Figure S10e). Therefore, for this trajectory four major clusters were selected (~72% coverage of the trajectory) and median representative structures were compared with the starting structure. The superimposition of these structures is reported in Figure S10f. The conformational changes represented by these structures indicate a tendency of the loop to adopt "more open" conformations than the crystallographic one. For further investigation of conformational changes in both $1DKP_{closed}$ and $1DKL_{open}$, clustering of only loop region was carried out for both $1DKP_{closed}$ and $1DKL_{open}$ trajectories.

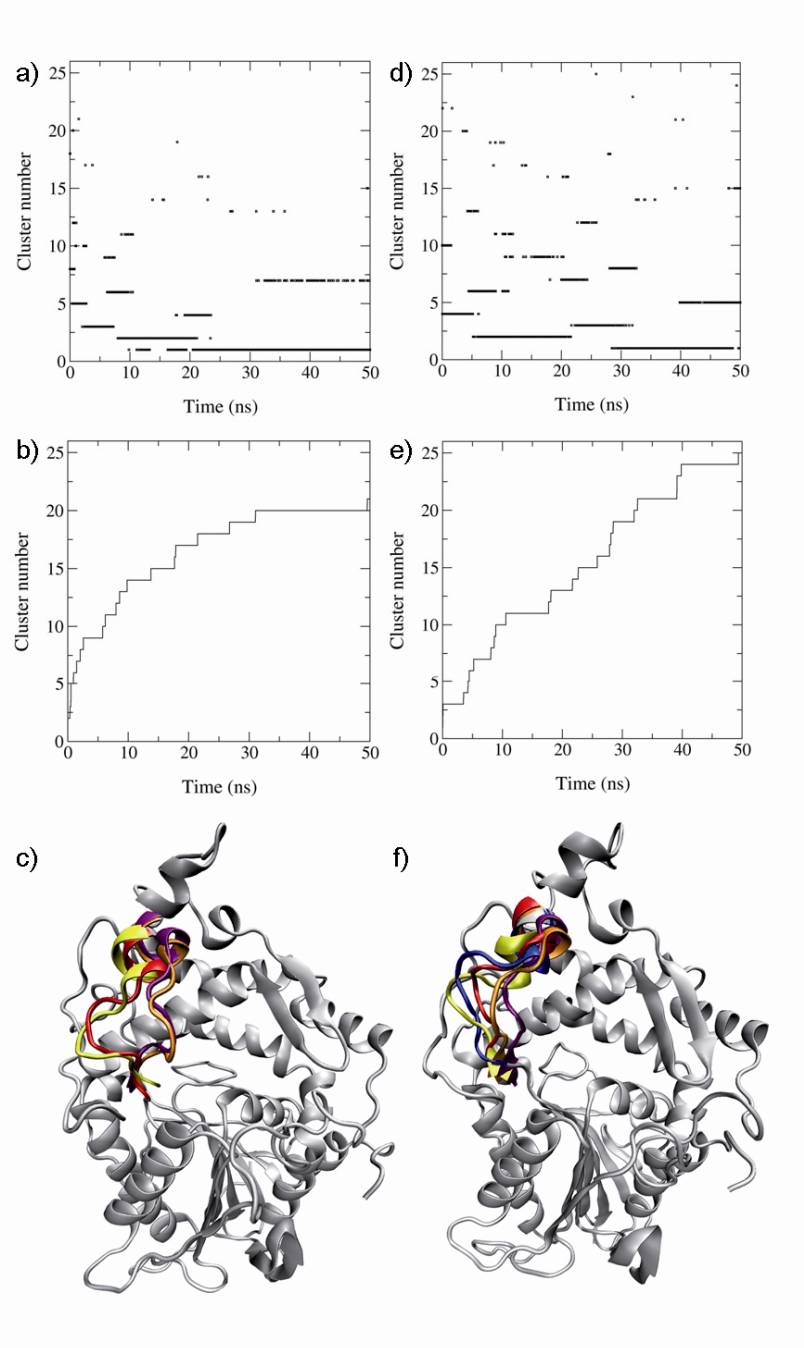


Figure S10: Sampling of clusters over 50 ns MD simulation of $1DKP_{closed}$ (a) and $1DKL_{open}$ (d); Cumulative sum of clusters for $1DKP_{closed}$ (b) and $1DKL_{open}$ (e); The representative conformations of the active site loop from the clustering of protein structures generated from 50 ns of $1DKP_{closed}$ (c) or $1DKL_{open}$ (f) MD simulation. Loop conformations of $1DKP_{closed}$ in cluster 1 (yellow), cluster 2 (red), cluster 3 (orange) and starting structure (purple) are depicted. The loop conformations of $1DKL_{open}$ in cluster 1 (yellow), cluster 2 (red), cluster 3 (blue), cluster 4 (orange) and starting structure (purple) are shown after superimposition [For clarity only one model outside active site loop is shown].

References

1. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. 14(1): p. 33-8, 27-8.
2. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. 22(12): p. 2577-637.
3. DeLano, W.L., *The PyMOL Molecular Graphics System.* 2002: Palo Alto, CA, USA.
4. Eyal, E., et al., *The limit of accuracy of protein modeling: influence of crystal packing on protein structure.* J Mol Biol, 2005. 351(2): p. 431-42.
5. Huang, H., et al., *A novel phytase from Yersinia rohdei with high phytate hydrolysis activity under low pH and strong pepsin conditions.* Appl Microbiol Biotechnol, 2008. 80(3): p. 417-26.
6. Rodriguez, E., E.J. Mullaney, and X.G. Lei, *Expression of the Aspergillus fumigatus phytase gene in Pichia pastoris and characterization of the recombinant enzyme.* Biochem Biophys Res Commun, 2000. 268(2): p. 373-8.
7. Li, X., et al., *Purification and characterization of extracellular phytase from a marine yeast Kodamaea ohmeri BG3.* Mar Biotechnol (NY), 2008. 10(2): p. 190-7.
8. Sajidan, A., et al., *Molecular and physiological characterisation of a 3-phytase from soil bacterium Klebsiella sp. ASR1.* Appl Microbiol Biotechnol, 2004. 65(1): p. 110-8.

Water flowed heat inactivation aluminium block for screening of mutant libraries for improved thermostability

Amol V. Shivange[a,b], Alexander Dennig[b], and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

10.1 Introduction

Directed protein evolution has become a versatile tool for improving enzyme properties. Several enzymes have been evolved for improved thermostability using directed evolution [1-3]. Improving thermostability of a thermostable enzyme is challenging task and most of the time such enzymes evolve slowly. In each round of directed evolution a small improvement in thermostability (T_{50} = 0.5 to 1 °C) is usually obtained. High throughput screening platforms based on 96-well microtiter plate have been reported. One of the key to success in improving thermostability is a highly homogenous heat inactivation method throughout 96 wells of microtiter plate. Most of the success stories have employed a PCR cycler for heat inactivation of mutant proteins to screen for higher thermostability.

In this study we have designed a 96-well format water flowed aluminium block for heat inactivation of mutant library. The designed model of aluminium block is an inexpensive alternative to the PCR cycler with high homogeneity in temperature throughout the 96-well compare to PCR cycler. During the screening of mutant libraries, we have observed that the heat transfer to the corner wells of PCR cycler is less compare to wells centered in the 96 well PCR plate due to significant standard deviation (0.5 °C) in temperature homogeneity. Therefore, a water flowed aluminium block was designed for heat inactivation that showed high homogeneity (standard deviation = 0.1 °C) for heat transfer throughout all the 96 wells.

10.2 Materials and Methods

The temperature deviation of randomly chosen wells containing 100 µl of buffer () were measured using Digital TFA Dostmann LT-101 thermometer (Wertheim, Germany). The thermometer was equally arranged into the well of PCR cycler (Mastercycler pro S, Eppendorf AG, Hamburg, Germany) by being fixed by a stand. Values are noted when displayed temperature is stable for 1 min.

Phytase enzyme activity as described in section 4.2.7 using 4-methylumbelliferyl phosphate as a substrate. Diluted cell lysates were directly dispensed into pre-heated

*Corresponding author

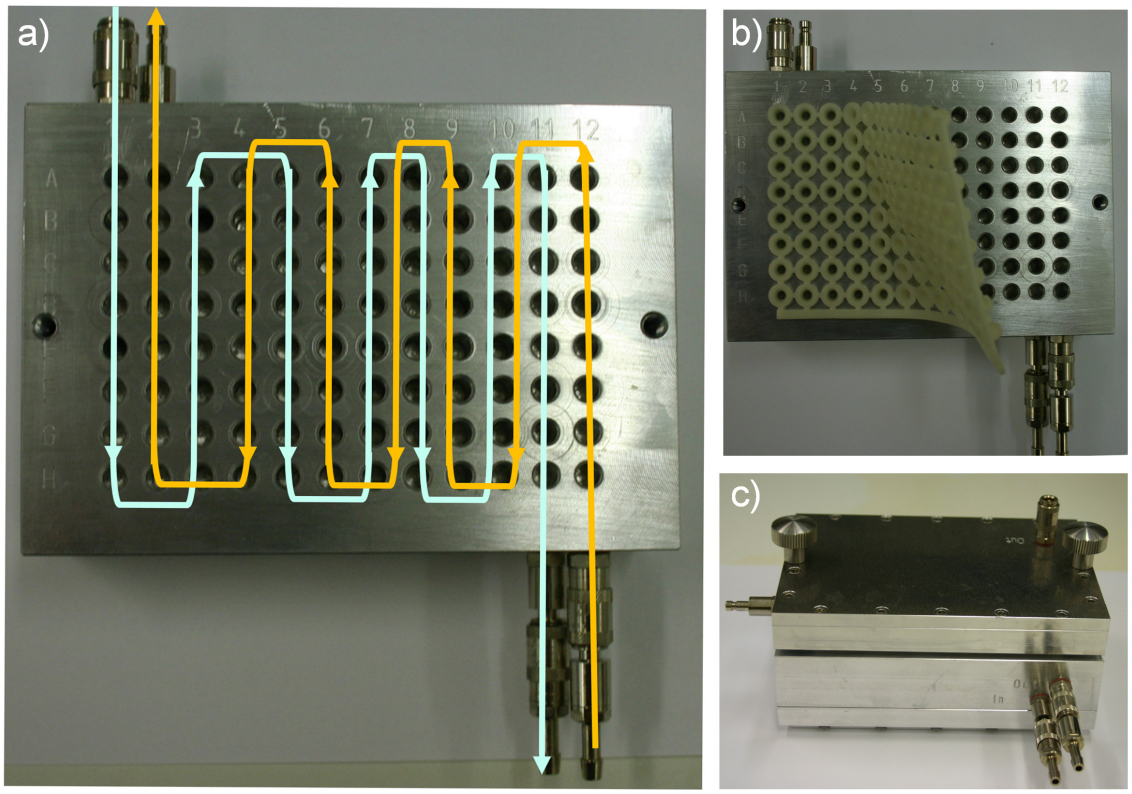


Figure 1: A water flowed aluminium block for heat inactivation. The top view and design of cross flow is depicted (a). Top view with sealing mat (b) and side view (c) is shown.

aluminium block (water flow rate equals 300 ml/min) using 96-well Liquidator (Steinbrenner Laborsysteme GmbH, Wiesenbach, Germany). To prevent evaporation, the wells were closed with a standard PCR plate sealing mat and tightly screwed with a upper aluminium block. The temperature of water bath was kept 5 °C more than the desired temperature in the aluminium block. After 20 minutes, the cold water (4 °C) was flowed into the block for refolding of the phytase. The phytase activity was monitored for 15 minutes by the increase in the relative fluorescence (excitation wavelength, 360 nm; emission wavelength, 465 nm) using a Tecan Infinite M1000 microtiter plate reader (Tecan Group Ltd. Männedorf, Switzerland) due to the hydrolysis of 4-MUP to 4-methylumbelliferone.

10.3 Results and Discussion

96-well format water flowed aluminium block was designed based on a cross flow water circulation system for better heat transfer (Figure 1). Two inlets and two outlets were integrated to the side of aluminium block. One inlet is paired with outlet for cross flow of the water in the block. The wells are designed to be tightly sealed with the sealing mat and hold at least 100 µl liquid protein sample. An upper

Table 1: Temperature deviation of the PCR cycler and water flowed aluminium block.

Instrument	Temperature of water-bath [°C]	Average liquid temperature in wells [°C]	Standard deviation [°C]	Standard deviation [%]
PCR cycler	59	57.31	0.57	1.01
Aluminium block	65	60.9	0.09	0.15

aluminium block has one sided water flow with screws to tighten both upper and lower block together. Upper block is integrated to minimize evaporation of the sample present in the well.

PCR cycler showed high deviation in heat transfer than the designed aluminium block (Table 1). Approximately, 0.6 °C deviation in the temperature homogeneity was observed in PCR cycler compare to aluminium block that was less than 0.1 °C deviation. This indicates that the PCR cyclers can only be used for screening of mutant libraries where improvement is expected to be higher per round of directed evolution for improving thermostability. Additionally, a maximum difference between two wells (A1 and D6) of PCR cycler was determined to be 1.4 °C.

In contrast, the water flowed aluminium block showed more homogeneous temperature distribution over the 96-well of plate when preheated. A very low temperature deviation was measured covering the whole block. Further, no significant differences between specific wells were detected.

10.4 Conclusions

The water flowed aluminium block for heat inactivation of the protein sample is developed for high accuracy in the heat transfer. This block can be an inexpensive replacement for PCR cyclers that are commonly used in the screening of mutant libraries for improving thermostability. A high temperature screening can easily be performed with the aluminium block using oil instead of water for heating.

10.5 References

1. Kim, M.S., J.D. Weaver, and X.G. Lei, *Assembly of mutations for improving thermostability of Escherichia coli AppA2 phytase.* Appl Microbiol Biotechnol, 2008. 79(5): p. 751-8.
2. Zhao, H. and F.H. Arnold, *Directed evolution converts subtilisin E into a functional equivalent of thermitase.* Protein Eng, 1999. 12(1): p. 47-53.
3. Eijsink, G.H.V., et al., *Increasing the thermostability of a neutral protease by replacing positively charged amino acids in the N-terminal turn of {alpha}-helices.* Protein Eng., 1992. 5(2): p. 165-70.

Directed evolution platform and solid-phase screening system for *Klebsiella pneumoniae* phytase

Amol V. Shivange[a,b] and Ulrich Schwaneberg[a,b]*

[a]Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[b]School of Engineering and Science, Jacobs University Bremen gGmbH, Campus Ring 1, 28759 Bremen, Germany

Phytases catalyze the release of phosphate by stepwise hydrolysis of phytate, a major source of phosphate in cereal grains, legumes and oilseeds. Phytase have been used to improve as a feed supplement the nutritional quality of phytate rich diets and eventually reduce environmental pollution. Recently, phytases from enterobacteriaceae family have attracted industrial interest due to their high specific activity (2500 – 4000 U/mg) towards phytate. However, a phytase from *Klebsiella pneumoniae* showed broad substrate specificity with lower catalytic efficiency (169 U/mg). To establish a screening system for improving desired property and studying structure function relationship, a phytase from *Klebsiella pneumoniae* cloned, expressed and a high throughput solid-phase screening system was developed. The reported technically simple agar plate prescreening in combination with the developed 96-well microtiter plate screening assay can be used to screen large amount of mutant libraries.

11.1 Introduction

Phytases (myo-inositol hexakis phosphohydrolase) are a class of phosphatases that catalyze the release of phosphate by stepwise hydrolysis of phosphomonoester bonds in phytate (myo-inositol-1,2,3,4,5,6-hexakisphosphate). Phytate is the major storage form of phosphate and inositol, predominantly occurring in cereal grains, legumes and oilseeds and accounts for 60-90 % of total phosphorus content in mature seeds [1, 2]. Phytases are widespread in nature, found in animals, plants, and microorganisms. Members of the histidine acid phosphatase (EC: 3.1.3.8) family have attracted increased interest from the animal feed industry because of their high specific activity and their ability to efficiently hydrolyze phytate [3]. Fungal phytases from *Aspergillus niger* (var. *ficuum*) (BASF) and *A. awamori* (AB Enzymes) are widely used as a commercial feed additive [4]. Despite wide range of applications, phytases face the considerable challenge of the low activity. Recently, several phytases from enterobacteriaceae family [5] have been reported to possess higher specific activity (2500 – 4000 U/mg) compare to fungal phytases [4] towards phytate. A phytase from

*Corresponding author

Klebsiella pneumoniae (Kpphytase) has been reported to posses broad substrate specificity nearly 13 different phosphatase substrates were degraded by Kpphytase.

In this study, we have cloned, expressed and established a high throughput prescreening system based on agar plate. Large amounts of mutant colonies can be screened for activity. Additionally, 96-well microtiter plate based assay has been developed.

11.2 Materials and Methods

11.2.1 Cloning of K. pneumoniae appA

The *appA* gene from *K. pneumoniae (Klebsiella sp. ASR1)* has been cloned and expressed in *E. coli* [6]. *K. pneumoniae appA* (*phyK*) gene was cloned from *K. pneumoniae* genomic DNA (DSMZ no. DSM15998), which was obtained from Prof. Carsten Harms (Bremerhavener Institut of Biological Information Systems (BIBIS), Bremerhaven, Germany). Primers were designed using reported *KpappA* gene sequence (GenBank accession no. AY091638). The core region of *appA* was amplified by PCR using forward primer KpPh_NdeF (3'- CC*G AAT TC*A CAT ATG CCT GCA AGA CAT CAG GGG CTG-5') and reverse primer KpPh_NotR (5'-GAG TGC GGC CGC CTA CTA CGG CAG GAC CAT GGC TAC-3') containing *Nde*I or *EcoR*I and *Not*I restriction enzyme sites respectively. *EcoR*I site has been added to optionally choose the *pelB* signal sequence (periplasmic localization) from pET-22b(+) expression vector (Novagen).

The purified PCR product was digested with *Nde*I and *Not*I restriction enzymes (New England Biolabs) and cloned into two different *E. coli* expression vector pET-22b(+) and pALXtreme-5b (modified small pET-22b(+) vector) [7]. Confirmation of correct construct of *aapA* gene from *K. pneumoniae* was done by commercial sequencing.

11.2.2 Expression of Kpphytase in 96-well microtiter plates

The pET-22b(+)-*KpappA* and pALXtreme-5b-*KpappA* constructs were transformed into *E. coli* BL21-Gold(DE3) and *E. coli* BL21-Gold(DE3)lacI^{Q} or *E. coli* BL21-Gold(DE3)lacI^{Q1} strain respectively. Expression of *K. pneumoniae* phytase has been checked in all three, TYM-5052, MD-5052 and LS-5052 autoinduction medias and a higher expression was observed in TYM-5052, then in MD5052 and lower in LS5052. The LS5052 autoinduction media was selected and checked for homogeneity of the expression. Homogeneity of the expression was checked with total three different expression systems. The construct pALXtreme-5b-*KpappA* was transformed into *E. coli* BL21-Gold(DE3)laqI^{Q} and *E. coli* BL21-Gold(DE3)laqI^{Q1}. Both the transformants and *E. coli* Bl21-Gold(DE3) pET-22b(+)-*KpappA* were then grown on TYM-505 non-inducing media for 8-10 hour at 37°C (900 rpm, 70 % relative humidity) and saturated culture was inoculated to 150µl LS-5052 autoinduction media and grown for 8 hour at 37°C (900 rpm, 70 % relative humidity). Culture was centrifuged and cell pellet was resuspended in 200 µl of lysozyme solution (1mg/ml). Assay was performed with 10µl of cell lysate and standard deviations were calculated.

Table 1: Available templates for homology modeling of KpPhytase (*selected template)

Template PDBID	% Identity	Resolution (Å)	Mutation (s)	Ligand
1NT4*	28	2.40	H18A	Glucose-1-Phosphate
1DKM	25	2.25	A116T	No
1DKN	25	2.40	A116T	No
1DKO	25	2.38	A116T	No
1DKL	25	2.30	No	No
1DKP	24	2.28	A116T, H17A	Phytate

11.2.3 Solid phase screening system

Colonies were grown on LB-agar plate containing 100 µg/ml ampicillin. Overlay mixture (4ml, 0.6% Low melting temperature agarose; 2.5 ml 0.1% 1-Naphthyl phosphate; preheated at 40 °C) was added over the surface of the plate to completely cover the cells. After the agar solidified in 5 to 10 minutes, fresh Garnet GBC solution (0.1%) was added to the plates and kept for 15 min for color development. Ability of the assay to distinguish between active and inactive clones was determined by plating cells harboring pET-22b(+)-KpappA construct and empty pET-22b(+) vector.

11.2.4 Homology modeling

The molecular modeling study of KpPhytase was carried out on a workstation with the openSUSE linux operating system. Homology modeling was performed using automated comparative protein modelling server [8] (SwissModel: http://swissmodel.expasy.org/) and MODELLER 9v3 software [9].

Template Selection

As a first step towards 3D structure prediction using homology modeling, the amino acid sequences of KpPhytase (GenBank accession no. AAM23271) was obtained from the NCBI protein database. NCBI-BLAST (http://www.ncbi.nlm.nih.gov/ blast/Blast.cgi) was used to identify homologous structure for Kpphytase by searching the structural database of protein sequences in the Protein Data Bank [10] (PDB). A template, *E. coli* glucose-1-phosphatase (1NT4) for Kpphytase was selected based on higest percentage identity and highest resolution of the crystal structure (Table 1).

Sequence Alignment and Model Building

Sequence alignment of the query sequence with respective template was performed using MODELLER program that uses dynamic programming algorithm (sequence-structure alignment). Alignment obtained in PAP format is shown in Figure 1 for KpPhytase modeling. Subsequently, kpphytase protein was modeled using 1NT4 as a template.

Validation of Homology Models

The stereo chemistry of the homology model was gauged using Ramachandran plot within PROCHECK server of Swiss-Model server (http://swissmodel.expasy.org/). The final model was inspected for accuracy and validity using

```
 _aln.pos         10        20        30        40        50        60
1NT4A     Q-----------------------TVPEGYQLQQVLMMSRANLRAPLANNGSVLEQSTPNKWPEWDVP
kpph      MPARHQGLLRLFIACALPLLALHSAAAADWQLEKVVELSRHGIRPPTAGNREAIEAATGRPWTEWTTH
 _consrvd                                ** *   **   * * * *    *  *   * **

 _aln.p   70        80        90       100       110       120       130
1NT4A     GGQLTTKGGVLEVYMGHYMREWLAEQGMVKSGECPPPYTVYAYANSLQRTVATAQFFITGAFPGCDIP
kpph      DGELTGHGYAAVVNKGREEGQHYRQLGLLQA-GCPTAESIYVRASPLQRTRATAQALVDGAFPGCGVA
 _consrvd  * **  *    *  *          *      **     *  *  **** ****    ******

 _aln.pos  140       150       160       170       180       190       200
1NT4A     VHHQE---KMGTMDPTFNPVITDDSAAFSEQAVAAMEKELSKLQLTDSYQLLEKIVNYKDSPACKEKQ
kpph      IHYVNGDADPLFQTDKFAATQTDPARQLAAVKEKAGDLAQRRQALAPTIQLLKQAVCQADKPCPIFDT
 _consrvd  *              *    **           *         *    ***   *   * *

 _aln.pos    210       220       230       240       250       260       270
1NT4A     QCSLVDGKNTFSAKYQQEPGVSGPLKVGNSLVDAFTLQYYEGFPMDQVAWGEIKSDQQWKVLSKLKNG
kpph      PWQVEQSKSGKT--------TISGLSVMANMVETLRLGWSENLPLSQLAWGKIAQASQITALLPLLTE
 _consrvd        *               * *    *    *   *  *  * *** *    *   *  *

 _aln.pos      280       290       300       310       320       330       340
1NT4A     YQDSLFTSPEVARNVAKPLVSYIDKALVTDRTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERT
kpph      NYDLSNDVLYTAQKRGSVLLNAMLDGVKPEASPNVRWLLLVAHDTNIAMVRTLMNFSWQLPGYSRGNI
 _consrvd   *        *      *                    ** ** ***   *   *

 _aln.pos        350       360       370       380       390       400
1NT4A     PIGGKIVFQRWHDSKANRDLMKIEYVYQSAEQLRNADALTLQAPAQRVTLELSGCPIDA-DGFCPMDK
kpph      PPGSSLVLERWRDAKSGERYLRVYFQAQGLDDLRRLQTPDAQHPMLRQEWRQPGCRQTDVGTLCPFQA
 _consrvd * *   *  ** * *            *    **       * *  *      **        **

 _aln.p  410       420       430
1NT4A     FDSVLNEAVK------------
kpph      AITALGQRIDRPSAPTVAMVLP
 _consrvd     *
```

Figure 1: Structure-sequence alignment of template and query sequence based on dynamic programming algorithm of MODELLER KpPhytase.

3D-profile programs [11] (http://nihserver.mbi.ucla.edu/Verify_3D/) which calculates 3D to 1D compatibility score and performs Eisenberg analysis for depiction of the properly folded and misfolded regions in the protein structure.

11.3 Results and Discussions

11.3.1 Sequence Analysis

As observable from Figure 2, *appA* sequence has found to have some dissimilarity in 6 different positions (codons: highlighted by green line). To confirm that the dissimilarities are due to cloning mutations or inaccuracy in reported sequence, we have selected phylogenetically close other *Klebsiella* species (Figure 2) and checked for the corresponding nucleotide. It was observed that the reported sequence may have some point mutations. As most of the dissimilar nucleotides (5 out of 6) in sequenced *K. pneumoniae appA* gene compare to reported sequence (GenBank accession no. AY091638) are conserved in other *Klebsiella* species.

First dissimilarity was at position 603 base i.e. TTT (Phe)→TTC (Phe), which is silent mutation and also the nucleotide 'C' was found to be conserved. Second dissimilarity was at 617, CGG (Arg)→CAG (Gln), where 'A' was conserved; Third dissimilarity was at 884

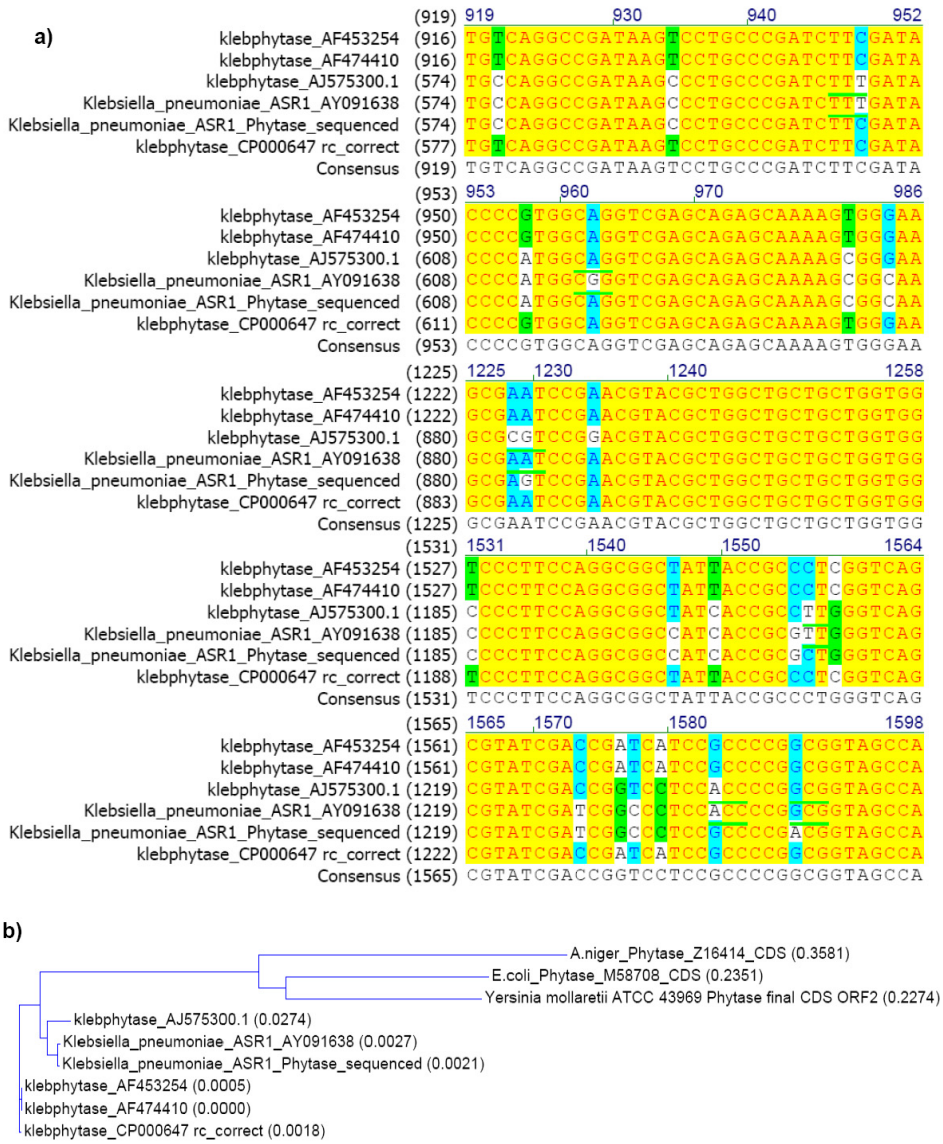


Figure 2: Multiple sequence alignment of the dissimilar regions in *appA* gene sequence from phylogenetically similar *Klebsiella* species (a). Phylogenetic tree based on nucleotide multiple sequence alignment of *Klebsiella sp*. and other known phytases (b). Alphsnumeric numbers shown in title are the NCBI Genbank accession number.

position where replacement was AAT (Asn)→AGT (Ser), 'G' was not conserved but probabilities are there, as 'G' was present in nearest homolog from *Klebsiella* (AJ575300.1) (Figure 2). Also this residue was found to be far away from the active site observed in homology model. Nucleotide at position 1237, TTG (Leu)→CTG (Leu) where mutation was silent and also 'C' was conserved. Nucleotide 'G' at 1237 was also conserved and the corresponding residue 'Ala' was also found to be conserved, AAC (Thr)→ GCC (Ala). Last dissimilarity may be due to the sequencing error as the nucleotide position 1243 was at 3' of the sequence. Based

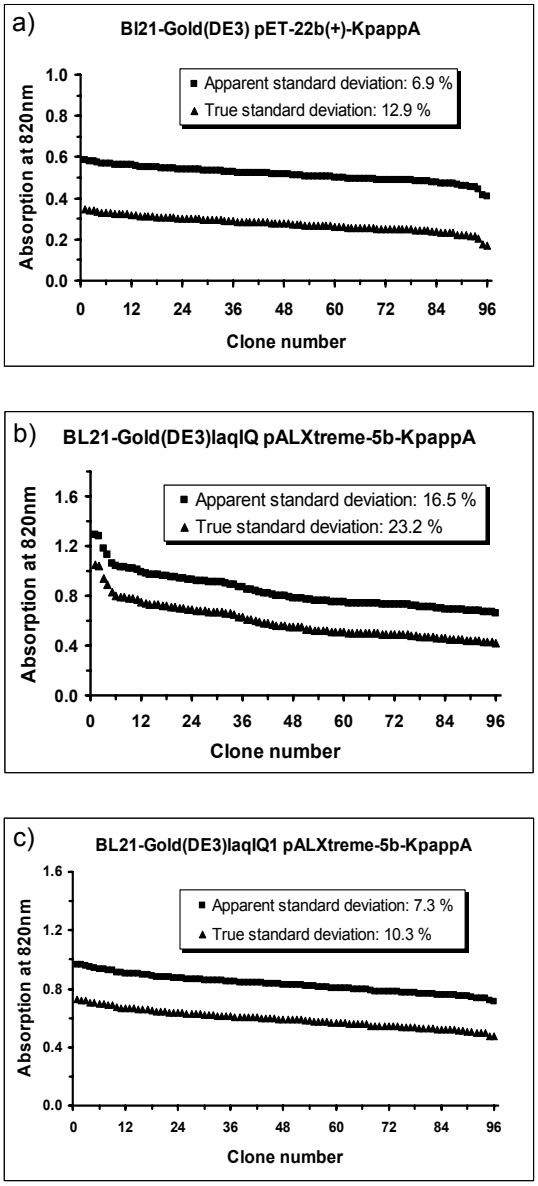


Figure 3: Activity values in descending order of the phytase catalyzed conversion of phytic acid in a 96-well plate for Bl21-Gold(DE3) pET-22b(+)-KpappA (a), BL21-Gold(DE3)laqI^{Q} pALXtreme-5b-KpappA (b) and BL21-Gold(DE3)laqI^{Q1} pALXtreme-5b-KpappA (c). The apparent standard deviation is calculated without subtracting the background and the true standard deviation after background subtraction.

on these observation and the enzyme was found to be active, it was concluded that the cloned gene is wildtype *appA* gene from *Klebsiella* ASR1 strain.

11.3.2 Expression of Kpphytase in 96-well microtiter plates

The preferred combination of 150µL and LS-5052 media in 96 well MTP plates at 900 rpm in a microplate shaker resulted in a true standard deviation or coefficient of variance (*CV* = [standard deviation/mean] × 100%) of 12.9% for Bl21-Gold(DE3) pET-22b(+)-KpappA, 23.2% for BL21-Gold(DE3)laqI^{Q} pALXtreme-5b-KpappA and 10.3% for BL21-Gold(DE3)laqI^{Q1} pALXtreme-5b-KpappA after subtracting the background (Figure 3). The background was calculated for *E. coli* BL21-Gold(DE3) pET22b(+). Therefore, BL21-Gold(DE3)laqI^{Q1} pALXtreme-5b-KpappA can be used for mutant library screening.

11.3.3 Solid phase screening system

To assist in screening mutant library a prescreening agar plate assay method has been developed. It has been reported that Fast Garnet GBC sulfate can used to screen for phosphatase active clones on agar plate provided a synthetic substrate 1-Naphthyl phosphate, which forms brick red insoluble product inside the cell [12]. Assay has been modified to overcome the cross contamination between the colonies on agar plate. It was observed that cells possessing empty pET22b(+) vector (Figure 4a) showed yellow colored colonies after color development and colonies having pET-22b(+)-KpappA showed red color (Figure 4b). A clear differentiation between phytase producing colonies versus nonproducing colonies can be seen in Figure 4c, plating has been done using mixture of cells possessing both the empty vector pET-22b(+) and pET-22b(+)-KpappA.

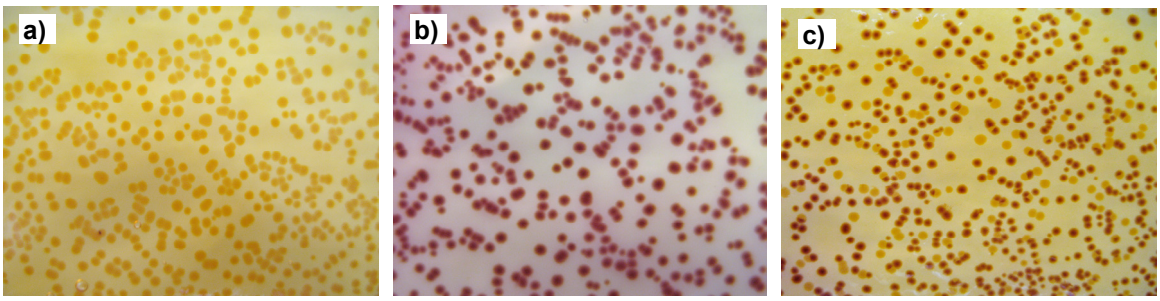


Figure 4: Prescreening agar plate assay for *K. pneumoniae* phytase. (a) *E. coli* BL21-Gold(DE3) colonies harboring empty vector [pET22b(+)], (b) *E. coli* BL21-Gold(DE3) colonies harboring pET-22b(+)-KpappA construct, red color observed due to formation of azo-coupled product by active phytase protein, (c) *E. coli* BL21-Gold(DE3) colonies having pET22b(+) [yellow] or pET-22b(+)-KpappA construct [red].

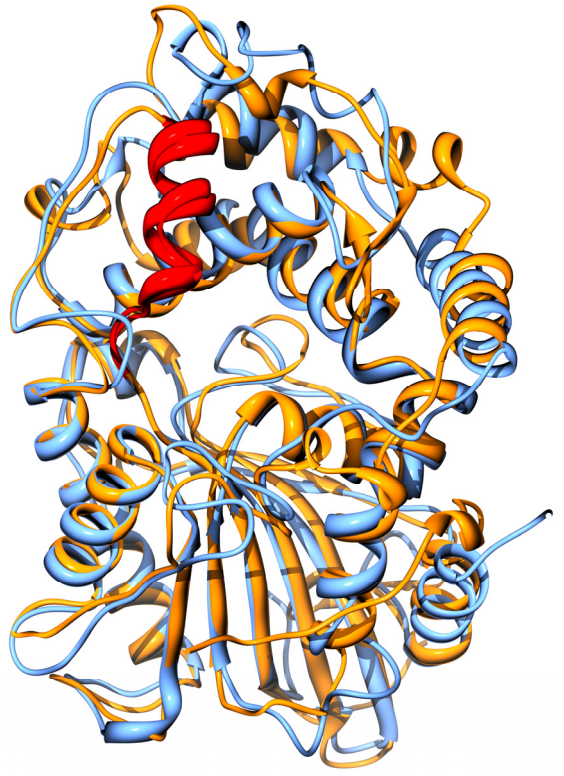


Figure 5: Structural superimposition of *K. pneumoniae.* phytase homology model (blue) and crystal structure (2WNH; orange). Active site loop replaced by helical structure is shown in red color.

11.3.4 Structural topology of Kpphytase

Kpphytase model developed by SwissModel was found be less accurate based on validation compare to the models built by MODELLER software. The overall structure of the Kpphytase is divided into two domains (α-domain and α/β-domain) by a central cavity. In case of KpPhytase this β-hairpin was found to be absent in homology model. Recently, a crystal structure of Kpphytase was solved [13]. To validate our homology, we have performed a structure alignment with Kpphytase crystal structure (PDB ID: 2WNH) using Swiss-pdb viewer software. A calculated RMSD between homology model and crystal structure was 1.36 Å at C-α region and 1.39 Å at backbone and all atom region of the Kpphytase. Interestingly, the overall topology of the homology model was quite similar to other phytases and crystal structure (Figure 5). The active site loop important for higher catalytic efficiency [14] has been found to be replaced by an helical structure in both homology model and crystal structure. This helical structure in the active site might be a reason for lower catalytic efficiency (discussed in section 9.3.6).

11.4 Conclusions

In summary, we have established a directed evolution platform for *K. pneumoniae* phytase. A rapid high throughput prescreening protocol has

been developed for screening mutant libraries. A developed screening protocol can be used for mutant library screening for improving activity of Kpphytase.

11.5 References

1. Bohn, L., A.S. Meyer, and S.K. Rasmussen, *Phytate: impact on environment and human nutrition. A challenge for molecular breeding.* J Zhejiang Univ Sci B, 2008. 9(3): p. 165-91.
2. Lott, J.N.A., et al., *Phytic acid and phosphorus in crop seeds and fruits: a global estimate.* Seed Sci Res, 2000. 10(01): p. 11-33.
3. Wyss, M., et al., *Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties.* Appl Environ Microbiol, 1999. 65(2): p. 367-73.
4. Haefner, S., et al., *Biotechnological production and applications of phytases.* Appl Microbiol Biotechnol, 2005. 68(5): p. 588-97.
5. Huang, H., et al., *A novel phytase from Yersinia rohdei with high phytate hydrolysis activity under low pH and strong pepsin conditions.* Appl Microbiol Biotechnol, 2008. 80(3): p. 417-26.
6. Sajidan, A., et al., *Molecular and physiological characterisation of a 3-phytase from soil bacterium Klebsiella sp. ASR1.* Appl. Microbiol. Biotechnol., 2004. 65: p. 110-18.
7. Blanusa, M., et al., *Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method.* Anal Biochem, 2010. 406(2): p. 141-6.
8. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. 22(2): p. 195-201.
9. Sali, A., *Comparative protein modeling by satisfaction of spatial restraints.* Immunotechnology, 1996. 2: p. 279-80.
10. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. 28(1): p. 235-42.
11. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure.* Science, 1991. 253(5016): p. 164-70.
12. Paunio, I.K., *Fractionation of Enzymes from Human Dental Plaque Hydrolyzing agr-Naphthyl Phosphate.* J Dent Res, 1969. 48(6): p. 1243-47.
13. Bohm, K., et al., *Crystal structure of Klebsiella sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme.* Febs J, 2010. 277(5): p. 1284-96.
14. Shivange, A.V., U. Schwaneberg, and D. Roccatano, *Conformational dynamics of active site loop in Escherichia coli phytase.* Biopolymers, 2010. 93(11): p. 994-1002.

Conclusions and future prospects

Directed protein evolution has become a well established, versatile and successful algorithm for tailoring protein properties to industrial demands and for advancing our understanding of structure-function relationships in biocatalysts. Despite advances in 'smart' enzyme evolution, a tiny fraction of theoretical sequence space can be sampled using the current technology. Therefore, recent advances in smart diversity generation have been reviewed (Chapter 2). Concepts of mutational load on protein traits, computational and experimental advances in generating smart libraries are discussed. Development of more accurate statistical models and algorithms quantifying and predicting even coupled mutational effects on protein functions would represent a significant step forward to an efficient reengineering of enzymes. Strategies and methods should be developed for understanding how many amino acid positions should be changed simultaneously for improving rapidly protein properties. The latter requires rules for mutational thresholds for different protein folds and properties.

The second part of the thesis describes structure activity-thermostability relationship of phytase enzyme (chapter 4 – 5). Directed evolution of phytase has been achieved for improved thermostability and catalytic activity using novel 384-well micortiter plate prescreening combined with improved screening assay system. A molecular dynamic simulation studies on wild-type and variant revealed that the increase in intra-protein hydrogen bonding network is necessary for improving phytase thermostability resulting in overall decreased phytase flexibility. The increased active site loop flexibility, in support of proposed hypothesis on phytase activity (chapter 9) might be the reason for improved activity of Ymphytase variant. A structure-function relationship revealed in this study can be a step forward to improve and understand thermostability and activity of other phytases.

Third part of the thesis focused on development of three conceptually novel methods for protein engineering. First a combinatorial assembly of site saturation test in protein segment (**ProCASTing**) was developed for parallel site saturation mutagenesis of any part of the protein (chapter 6). A successful site saturation of 8 or 4 subsequent amino acids in phytase was achieved using ProCASTing. Perl based script was developed to assist experimentalist for accurate selection of flexible regions in a protein during MD simulation or from the crystal structure. The ProCASTing method can easily be applied to active site modification by parallel site saturation mutagenesis of more than one amino acid to broaden our understanding of structure-function relationship on enzyme reactions and promiscuity. Second a multisite combinatorial assembly of site saturation test (**OmniCASTing**), practically simple method was developed for parallel site saturation of more than one site regardless of positions in the gene (chapter 7). This method can be used to

generate large mutant libraries with low empty vector background for high throughput screening. A cooperative effect of distant mutations can easily be studied with OmniCASTing method that is not feasible with current computational advances. As a proof of principle, we have improved three properties of already identified mutant from Ymphytase using OmniCASTing. Protein consensus based surface engineering (**ProCoS**) method combining computational analysis and molecular biology tools was developed (chapter 8). The active Ymphytase surface engineered variant with 34 amino acid substitution was identified with 3.8 fold improved pH stability (pH 2.8). Improved pH stability might be due to increased negatively charged amino acid substitutions on Ymphytase surface. ProCoS method can be applied to other biocatalyst for improving properties that are influenced by surface properties of the protein like solubility, thermostability, organic solvent stability.

Two hypotheses have been proposed in forth part of the thesis (chapter 8 – 9). A hypothesis that the conserved residues belong to an ancestor and highly variable regions situated on the surface of protein can be targeted for surface engineering was proposed. Based on this hypothesis, we have developed a method entitled ProCoS, improving phytase pH stability due to surface modifications, supports a proposed hypothesis on protein surface engineering. Second hypothesis was based on long scale molecular dynamic simulation studies *on E. coli* phytase (chapter 9). This study fostered a hypothesis that the active site loop might be prerequisite for high activity and catalytic efficiency of phytases. MD simulation studies on Ymphytase variant showed higher flexibility in the active site loop (chapter 5). Improved activity in this variant was compensated by a K289Q substitution, however variant lacking this substitution was improved in catalytic activity of phytase (200 U/mg). The high flexibility of the active site loop in Ymphytase variant during MD simulation supports our hypothesis on importance of this loop in higher catalytic efficiency. Additionally, a helical structure replacing active site loop in the crystal structure of phytase (low catalytic efficiency) from enterobacteriaceae species (*Klebsiella pneumoniae*) supports our second hypothesis on phytases.

The last part of the thesis summarizes development of water flowed aluminium block (chapter 10) and directed evolution screening assays for *K. pneumoniae* phytase (chapter 11). The water flowed aluminium block can be an inexpensive replacement for PCR cyclers that are commonly used in the screening of mutant libraries for improving thermostability. A high temperature screening can easily be performed with the aluminium block using oil instead of water for heating. A robust and high throughput solid phage screening assay reported for *K. pneumoniae* phytase can be used for improving phytase properties or for studying mutational threshold on a protein fold (α/β-hydrolase) due to technical simplicity and reproducibility of the assay.

In summary, to the best of our knowledge for the first time we have reported

1. Directed evolution of highly active phytase from *Yersinia* species, biochemical characterization of *Y. mollaretii* phytase, use of SeSaM method for phytase evolution and

use of combined screening system based on absorption and fluorescence assay for evolution (chapter 4).

2. MD simulation studies on mutants of phytase class of enzymes for improved activity or thermostability (chapter 5).
3. Two sequence independent methods for parallel saturation mutagenesis of subsequent amino acid residues (chapter 6) and for distant sites in a protein (chapter 7).
4. A novel method for protein surface engineering (chapter 8).
5. MD simulation studies on the enzyme class of phytase and the first use of DomainSelect to investigate loop dynamics (chapter 9).
6. Water-flowed heat inactivation device imbedded in aluminium block for screening of mutant library to improve thermostability (chapter 10)

PERSONAL DETAILS

Name:	Amol Vaijanathappa Shivange
Contact:	Jacobs University Bremen, College Ring 1, 28759 Bremen – Germany Telephone: +49 421 200 3634
Email:	a.shivange@biotec.rwth-aachen.de
Webpage:	http://amolshivange.googlepages.com
Gender:	Male
Date of Birth:	20.01.1982
Marital Status:	Single
Nationality:	Indian

ACADEMIC PROFILE

Since 11/07	PhD Fellow of Biochemical Engineering Jacobs University Bremen, Bremen, Germany Expected Date of Graduation Dec 2010
07/07 – 10/07	Scientist, Computer Aided Drug Design Connexios Life Sciences Pvt. Ltd., Bangalore, India
06/05 –06/07	M. S. (Pharm.) Biotechnology, (CGPA 9.62/10) National Institute of Pharmaceutical Education and Research, Punjab, India
09/03 – 05/05	B. Pharm. (74.92%; first class with distinction) University of Pune, Maharashtra, India
05/00 – 06/02	D. Pharm. (77.20%; first class with distinction) Channabasweshwar Pharmacy College, Latur, Maharashtra, India

PUBLICATIONS

Total Impact Factor: 17.978
Average citation per year: 6
Average citation per article: 4.8
Sum of times cited: 24 + 2

1. **Shivange AV**, Schwaneberg U, Roccatano D. Conformational dynamics of active site loop in *Escherichia coli* phytase. Biopolymers. 2010 Nov;93(11):994-1002. [**COVER PAGE**]
2. Kaur S, **Shivange AV**, Roy N. Structural analysis of trypanosomal sirtuin: an insight for selective drug design. Mol Divers. 2010 Feb;14(1):169-78.
3. **Shivange AV**, Marienhagen J, Mundhada H, Schenk A, Schwaneberg U. Advances in generating functional diversity for directed protein evolution. Curr Opin Chem Biol. 2009 Feb;13(1):19-25. [Review]
4. Prusty D, Mehra P, Srivastava S, **Shivange AV**, Gupta A, Roy N, Dhar SK. Nicotinamide inhibits *Plasmodium falciparum* Sir2 activity in vitro and parasite growth. FEMS Microbiol Lett. 2008 May;282(2):266-72.
5. Kadam RU, **Shivange AV**, Roy N. *Escherichia coli* versus *Pseudomonas aeruginosa* deacetylase LpxC inhibitors selectivity: surface and cavity-depth-based analysis. J Chem Inf Model. 2007 May-Jun;47(3):1215-24.

GENBANK SUBMISSION

1. **Shivange AV**, Agrawal V, Roy N. (2007) *Entamoeba histolytica* HM-1:IMSS: Synthetic construct Sir2 family transcriptional regulator, genomic sequence. Direct submission GenBank Accession no. EF202834.

CONFERENCE SPEAKER

1. Delivered talk on "Phosphatase engineering by directed evolution and insights into activity-thermostability relationship deciphered by molecular dynamics simulations" at Indo-German Partnering Workshop on Bio-economy, 23-24 Nov, 2010, New Delhi, India.
2. Delivered talk on "Germany: Scientific and Social Acquaintance" at Indo-German Partnering Workshop on Bio-economy, 23-24 Nov, 2010, New Delhi, India.
3. Delivered a lecture on "Diversity generation: Advances in managing the combinatorial complexity of protein sequence space" at Dechema Summer School 2010, 6 – 10 Sep, 2010, Greifswald, Germany.*
4. Delivered talk on "Steering directed protein evolution by managing the combinatorial complexity of mutant libraries" at ProkaGENOMICS 2009, 4-7 Oct, 2009, Göttingen, Germany.*
5. Delivered talk on "Steering Directed Protein Evolution: Strategies to Benchmark and to Manage Combinatorial Complexity of Mutant Libraries " at Enzyme Engineering XX, 20-24 Sep, 2009, Groningen, The Netherlands.*
6. Delivered talk on "Computational Protein Modeling and Sequence Analysis" and trained the participant for sequence analysis and protein modeling at Workshop on Pharmacoinformatics: Tools for Drug Target Identification Mar 2007, NIPER, Mohali, Punjab, India.

*On behalf of Prof. Ulrich Schwaneberg

CONFERENCES/ABSTRACTS

1. Active participant in BIOTECHNICA 2008, exhibition on biotechnology. Hannover, Germany. Oct 2008.
2. Active participant in Workshop on Directed Protein Evolution. Jacobs University Bremen, Bremen, Germany. Aug 2008.
3. Shivange, A. V.; Agrawal, V. and Roy, N. Structure Based Discovery of Potential Inhibitors for Sirtuin: A Novel Target of Entamoeba histolytica; Proceedings of the 4th Annual USHUPO Conference, Bethesda, Maryland, USA, 2008.
4. Active member of organizing committee in Workshop on Pharmacoinformatics: Tools for Drug Target Identification. NIPER, Mohali, Punjab. Mar 2007.
5. Shivange, A. V.; Agrawal, V. and Roy, N. A Comparative Study of Yeast Sirtuin and its Entamoeba Histolytica Homologue; Proceedings of the International Conference on the Biology of Yeast and Filamentous Fungi, Pune, India, 2007; P54, 101.
6. Kadam, R. U.; Shivange, A. V. and Roy, N. E. coli vs P. aeruginosa Deacetylase LpxC inhibitors Selectivity: Surface and Cavity depth based analysis; Proceedings of the 4th Joint Sheffield Conference on Chemoinformatics, University of Sheffield, Sheffield, UK, 2007.
7. Active participant in NIPER-NANO 2006, conference on nanotechnology in advanced drug delivery. NIPER, Mohali, Punjab. Feb 2006.

AWARDS & ACHIEVEMENTS

1. Qualified Graduate Aptitude Test in Engineering (GATE-2005) conducted by the Indian Institute of Technology (IIT) Mumbai, secured 98.74 percentile.
2. First prize in scientific project exhibition at S. G. R. S. college of pharmacy, Saswad entitled "Small Scale Rotary Tablet Machine" (2004-05).
3. First prize in scientific project exhibition at S. G. R. S. college of pharmacy, Saswad entitled "Small Scale Liquid Manufacturing Unit" (2003-04).
4. Sir Ratan Tata Trust's merit scholarship was awarded in year 2004.
5. Secured first rank in order of merit during 3rd and final years of Bachelor of Pharmacy.
6. U. A. Mashruwala prize was awarded in S. G. R. S College of pharmacy, Saswad. Pune.
7. Secured second rank in order of merit during both years of Diploma in Pharmacy.
8. Registered Pharmacist of state pharmacy council, Maharashtra. India (Registration no. 63978; PPP No. N/002629).

REFERENCES

Prof. Dr. Ulrich Schwaneberg
Head of the Institute
Lehrstuhl für Biotechnologie
RWTH Aachen University
Worringer Weg 1
52056 Aachen, Germany
Tel.: +49-241-80-24176
Fax: +49-241-80-22387
E-Mail: u.schwaneberg@biotec.rwth-aachen.de
Web: http://www.biotec.rwth-aachen.de

Prof. Dr. Danilo Roccatano
Assistant Professor
School of Engineering and Science
Jacobs University Bremen
Campus Ring 1
28759 Bremen, Germany
Tel: +49-421 200-3144
Fax: +49-421-200-3249
Email: d.roccatano@jacobs-university.de
Web: http://www.jacobs-university.de/schools/ses/droccatano

Declaration

I hereby declare that this thesis was written solely by me. The scientific work embodied in this thesis was my own research findings or results from my collaborative work with other scientists. Permission to use any material described in this thesis should be requested to me or corresponding authors. This work has not been submitted in part or in full to any other university or institute for award of any degree or diploma.

Date: 15th November 2010

Place: Bremen, Germany.

Amol Vaijanathappa Shivange

School of Engineering & Science

Jacobs University Bremen

Germany.