PROTEIN ENGINEERING OF RHIZOPUS ORYZAE LIPASE FOR IMPROVED THERMOSTABILITY

by
OZGUR GUL

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

SABANCI UNIVERSITY
January 2010

APPROVED BY:

Assoc. Prof. Dr Uğur Sezerman .....
(Dissertation Supervisor)

Asst. Prof. Dr. Alpay Taralp .....

Assoc. Prof. Dr. Batu Erman .....

Prof. Dr. Yaşar Gürbüz .....

Prof. Dr. Dilek Kazan .....

DATE OF APPROVAL: .....

PROTEIN ENGINEERING OF RHIZOPUS ORYZAE LIPASE FOR IMPROVED THERMOSTABILITY

Ozgur Gul
Biological Sciences and Bioengineering, PhD Thesis, 2010
Thesis Advisor: Assoc. Prof. Dr Ugur Sezerman

Key words: Protein engineering, Molecular dynamics simulation, Thermostability, *Rhizopus oryzae* lipase, *Pichia pastoris* expression, Site directed mutagenesis, Microarray scale enzyme assays

ABSTRACT

Protein engineering provides tools for understanding the structure function relationship and deduces the rules governing this relationship. Enzymes catalyze several types of reactions and have wide range of applications ranging from the textile industry to the pharmaceutical industry. Today one of the most important topics in biotechnology industry focus on replacing chemicals with enzymes accordingly turning all the industrial processes into green industry. Especially lipases have great importance in such processes. Unfortunately currently available active lipases do not function well under the harsh industrial conditions such as high temperature and extreme pH. In this thesis I have focused on the thermostability of *Rhizopus oryzae* lipase and tried to deduce the rules that make it stable at high temperatures while preserving their activity. I have developed a protocol involving the computational analysis predicting beneficial point mutations, and analyzed the effects of predicted mutations *in silico*. I have cloned, expressed and purified the mutant proteins to verify the predictions. I showed that the P195G mutation in *Rhizopus oryzae* lipase sustained 80.5 percent of the activity at high temperatures in comparison to the native lipase.

RHIZOPUS ORYZAE LİPAZININ SICAKLIK DAYANIKLILIĞININ PROTEİN MÜHENDİSLİĞİ İLE GELİŞTİRİLMESİ

Özgür Gül
Biyoloji Bilimleri ve Biyomühendislik, Doktora Tezi, 2010
Tez Danışmanı: Doç. Dr. Uğur Sezerman

Key words: Protein mühedisliği, Molekuler dinamik simulasyonu, Sıcaklık dayanıklılığı, *Rhizopus oryzae* lipazı, *Pichia pastoris* expresyonu, Yönlendirilmiş nokta mutasyonu, Mikroarray ölceğinde enzim deneyi

ÖZET

Protein mühedisliği yapı-fonksiyon ilişkisini anlamak için kullanılan araçları temin eder ve bu ilişkiyi yöneten kuralları ortaya çıkarır. Enzimler muhtelif birçok reaksiyon katalizasyonunda rol alırlar ve tekstil endüstrisiden ilaç endüstrisine kadar geniş bir dizi uygulamada kullanılırlar. Günümüzde biyoteknoloji endüstrisinin en önemli konularından biri kimyasalların yerini enzimlerin almasını odaklamaktadır. Özellikle lipazların bu tür işlemlerde büyük önemleri vardır. Fakat mevcut aktif lipazlar endüstriyel uygulamaların gerektirdiği olağandışı koşullarda örnekle yüksek sıcaklık ve aşırı pH koşularında fonksiyon kaybı göstermektedirler. Bu nedenle bu tez çalışmasında *Rhizopus oryzae* lipazının yüksek sıcaklıkta kararlı hale getirilmesini ve bu sırada aktivitesini korumasını sağlayarak termokararlığı üstüne kurallar çkarılmasına yoğunlaşılmıştır. Öncelikle yararlı nokta mutasyonları tahmin eden hesaplamaya dayalı bir analiz içeren protokol geliştirilmiş ve tahmin edilen mutasyonların etkileri *in silico* analiz edilmiştir. Daha sonra bu tahminleri doğrulamak amacı ile mutant proteinler önce gen seviyesinde klonlanmış, ekpress edilmiş ve ardından saflaştırılmıştır. Sonucunda *Rhizopus oryzae* lipazindaki P195G mutasyonun proteinin tabi haline kıyasla yüksek sıcaklıkta aktivitede yüzde 80.5 daha fazla muhafaza sağladığı gözlemlenmiştir.

TABLE OF CONTENTS

ABBREVIATIONS

4MU: 4-Methylumbelliferyl

α-MF: Alpha-mating Factor

AOX : Alcohol OXidase

BLOSUM: BLOcks of Amino Acid SUbstitution Matrix

BMG: Buffered Minimal Glycerol

BMGlu: Buffered Minimal Glucose

BMM: Buffered Minimal Methanol

BSC: Barcelona Supercomputer Center

BTGL: Bacillus thermocatenulus Lipase

CALIPA: Candida antarctica Lipase A

CE (Alignment): Combinatorial Extension (Alignment)

DCM: Dichloromethane

DMF: Dimethylformamide

DNS: 1,3-dinitrosalicylic acid

DSF: Differential Scanning Fluorimetry

DTT: Dithiothreitol

EB: Elution Buffer

EDTA: ethylenediaminetetraacetic acid

ER: Endoplasmic Reticulum

FLD: Formaldehyde Dehydrogenase

FRODA: Framework Rigidity Optimized Dynamical Algorithm

GAP: Glyceraldehyde-3-phosphate Dehydrogenase

GPTS: 3-glycidoxypropyltrimethoxysilane

HT: High throughput

IPTG: isopropyl--D-thiogalactopyranoside

MD: Molecular Dynamics

MM: Minimal Methanol

MTP: Multi-titer Plate

OE-PCR: Overlap-extension PCR

PCR: Polymerase Chain Reaction

PIPES: piperazine-1,2-bis[2-ethanesulfonic acid]

PVDF: Polyvinylidene Fluoride

RE: Restriction Site

RFU: Relative Fluorescence Unit

RGYR: Radius of Gyration

RMSD: Root Mean Square Deviation

RNL: Rhizopus niveus Lipase

ROL: Rhizopus oryzae Lipase

rpm: revolution per minute

SDM: Site Directed Mutagenesis

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SMM: Small Molecule Microarray

TB: Transformation Buffer

TLC: Thin layer chromatography

TT: Transcription Termination

ULAKBIM: Ulusal Akademik Ağ ve Bilgi Merkezi (National Academic Network and Information Center)

YPD: Yeast extract – Peptone – Dextrose

LIST OF FIGURES

LIST OF TABLES

Chapter 1

1 INTRODUCTION

Many enzymes have potential uses in biotechnology, but their applicability is limited due to their lack of stability under extreme conditions; e.g. elevated temperatures and pH, exposure to organic solvents and various chemicals. Therefore, the proteins that are stable under such conditions attract attention as they are suitable for the industrial applications. Understanding the origins of instability may help in the design of thermostable and wide pH working biomolecules for technological applications such as in textile industry, pharmaceutical industry and detergency. Once the factors affecting the adaptation to the extreme conditions are diagnosed, one can perform the approach to mutate the enzyme to increase thermal stability and workable pH ranges.

Among all of the enzymes lipases are being used increasingly for a variety of industrial purposes and consequently lots of effort has been put into their cloning and expression as well as their study by the site-directed mutagenesis. Lipases (EC 3.1.1.3) catalyze reversibly the cleavage of ester bonds of triacylglycerol to yield free fatty acids, diacylglycerols, and monoacylglycerols. They have therefore been widely used in industrial applications, such as in chemical, food, pharmaceutical, and detergent industries. However, most of the reactions must be performed at high temperature due to efficiency and time concerns of industrial processes. Therefore in this thesis we aim to understand the rules governing the temperature stability for *Rhizopus oryzae* lipase (ROL) by the protein engineering protocol developed in this work so that we can modify the ROL lipase to operate at high temperatures while retaining its activity. We also have developed high throughput methodologies for cloning, expression, purification of lipases and for their activity assays.

The developed methodology and the results on ROL lipase all through the course of this work are explained in five chapters including this introductory chapter. The

literature survey and the background related to the topic of the chapter and the results are given separately in every chapter. Bibliography for each chapter is given at the end of each chapter.

The Chapter 2 is dedicated to the computational studies including the thermal stability related feature extraction, the flexibility analysis and the molecular dynamics simulations. After the introductory part, related background information are given. The structures, the software and related methodologies are explained in the Materials and Methods Sections. Our results on the computational studies concerning the *Rhizopus oryzae* lipase thermal stability are given in Results and Discussions section followed by the Conclusion.

The Chapter 3 is dedicated to the molecular biology studies covering cloning, mutagenesis, protein expression and purification, and enzyme assay studies depending on the results of Chapter 2. After the introductory part for Chapter 3, related background information is presented. Plasmids, strains, enzymes, chemicals and kits related to the molecular biology part of this thesis are given in the Materials Section followed by the Methods section containing the molecular biology related techniques and protocols. Our results on the experimental findings are presented in the Results and Discussion section followed by the conclusion related to Chapter 3.

The Chapter 4 is dedicated to the high throughput methodologies (linear cassette transformation, sorbitol co-feeding strategies in microtiter plate expression, and the silica gel based enzyme microarrays) developed in the time course of this thesis. After the introductory part for Chapter 4, related background information is presented. Methodologies, protocols and techniques employed in this thesis and related materials, enzymes and chemicals are presented in two consecutive sections, Materials and Methods. Our findings and optimization results are presented and discussed in the Results and Discussion section of this chapter followed by the Conclusion section.

The last chapter (Chapter 5) of this thesis is dedicated to the overall conclusion of our computational and molecular biology studies. The conclusion concerning the developed high throughput techniques is also included in this Chapter.

Chapter 2

2 RATIONAL DESIGN OF RHIZOPUS ORYZAE LIPASE

2.1 Introduction

The aim of this chapter is to analyze the thermal stability of the lipase from *Rhizopus oryzae* by introducing mutations at given positions, and analyze the impact of the proposed mutations on the protein stability via molecular dynamics (MD) studies. Temperature stabilizer motifs were determined by comparing the lipases from mesophilic organisms to proteins of the same family from thermophilic organisms. The rules learned from thermophilic proteins were a starting point for determining the motifs. Out of this motif library, an appropriate selection of mutations has been made for the lipase of interest, and 3D models have been generated via homology modeling. These 3D models were the input for the MD studies to determine the extent to which they stabilize/destabilize the system.

In this chapter, I have designed sets of experiment to analyze the thermostability properties of candidate mutations using Molecular Dynamics studies. First of all, I have determined the simulation temperature that can allow us to analyze mutations in reasonable time frame. After assigning the simulation temperature, I analyzed all of the candidate mutations using Molecular Dynamics. Although root mean squared deviation (RMSD) analysis can be a good measure for the thermal stability, I also analyzed the results in terms of the radius of gyration, and the catalytic residue distances.

Catalytic residue distance can be used as a measure for the activity during simulations. For ROL, the catalytic residues (145SER, 204ASP, and 257HIS) should be in a similar distance within the native structure. I used these distances in order to comment on the stability of the mutant proteins. Radius of gyration is a measure of how much the structure spreads out from its center. In another words it indicates the compactness of the structure. I used the radius of gyration values in order to evaluate the

mutant structures behavior during simulations at different temperatures by comparing the radius of gyration values of the moderately stable and the thermostable structures.

One can mutate protein residues to thermodynamically favor the folded state by either adjusting the S (entropy) or H (enthalpy) terms (a-1, and a-2). Another possibility is to mutate or modify the protein residues to kinetically trap the folded state and bypass any ground-state consequences of the modifications (b-1). These examples are explained in Figure 2.1. In this thesis our work focused on the adjustment of the ground state entropy term.

a-1: Adjusting the ground state entropy term (the activation energy of unfolding might also change, but it is not focused in this case). In the case of Proline to Glycine mutations, the gain of chain flexibility and freedom is more noticeable in the folded state than the unfolded state because the folded state has less freedom to begin comparing to the unfolded state. The absolute entropy of the protein will increase more in the folded state. Overall, the absolute energy term will be decreased little in the unfolded state and more in the folded state. This will cause an increase of the stability of the folded state; ΔG of the folding is more negative.

a-2: Adjusting the ground state enthalpy term (the activation energy of unfolding might also change, but it is not focused in this case). If one make an intelligent choice of mutation, intramolecular (protein-protein) and intermolecular (water-protein) non-covalent bonding is maximized in the folded state and any loss of freedom in the folded state is minimized. For instance, Threonine (which has one hydroxyl group in the side chain) residue buried in a hydrophobic pocket with no possibility of Hydrogen bonding can be mutated into Valine (which has methyl group instead of hydroxyl) residue. Under these circumstances, the absolute enthalpy of the unfolded state will rise a little because the interaction of water with the Valine sidegroup is less favorable than the Threonine sidegroup. The absolute enthalpy of the folded state will decrease a little because the interaction of the Valine sidechain will be more favorable (exothermic) with the hydrophobic pocket. The absolute energy term will be noted to increase a little in the unfolded state and to decrease in the folded state, corresponding to an increase of stability of the folded state; ΔG of the folding is more negative. If we focus on the entropic change related only to the protein component (protein flexing, rotations, etc.), we can assume to a first approximation that both proteins lose the same amount of freedom in proceeding from unfolded to folded state. If we focus on the water

component, more freedom is gained when the mutant protein folds. Overall, the mutant protein system actually loses less freedom in the course of folding than the wt protein system.

b-1: Mutation or modification of proteins to kinetically trap the folded state. For example, we can create a mutant with 20 new disulphide bridges (intramolecular) or we can mix a lysine-rich protein crystal with glutaraldehyde in order to obtain crosslinked molecules (intermolecular). Overall, the system does liberate heat to the environment when these changes are made, so the enthalpic term would favor the crosslinked forms if we compare the ground states before folding and after folding and crosslinking. However, we can also expect a significant loss of freedom when this crosslinking is made, which in theory should destabilize the ground state of the folded form. Which effect is dominant is neither clear in this case, nor is it important. The only important term to consider here is the lack of low-energy unfolding pathways! When we impose so many crosslinks to the protein, there is no pathway left with a reasonably low activation barrier to permit unfolding. So the protein is kinetically trapped in the folded state as a crosslinked entity.

In this thesis, I have focused on the flexibility increase by mutating the proline residues to the glycine residues around active site region. The adjustment of the ground state entropy term via rationally design mutations is explained in case a-1 above.

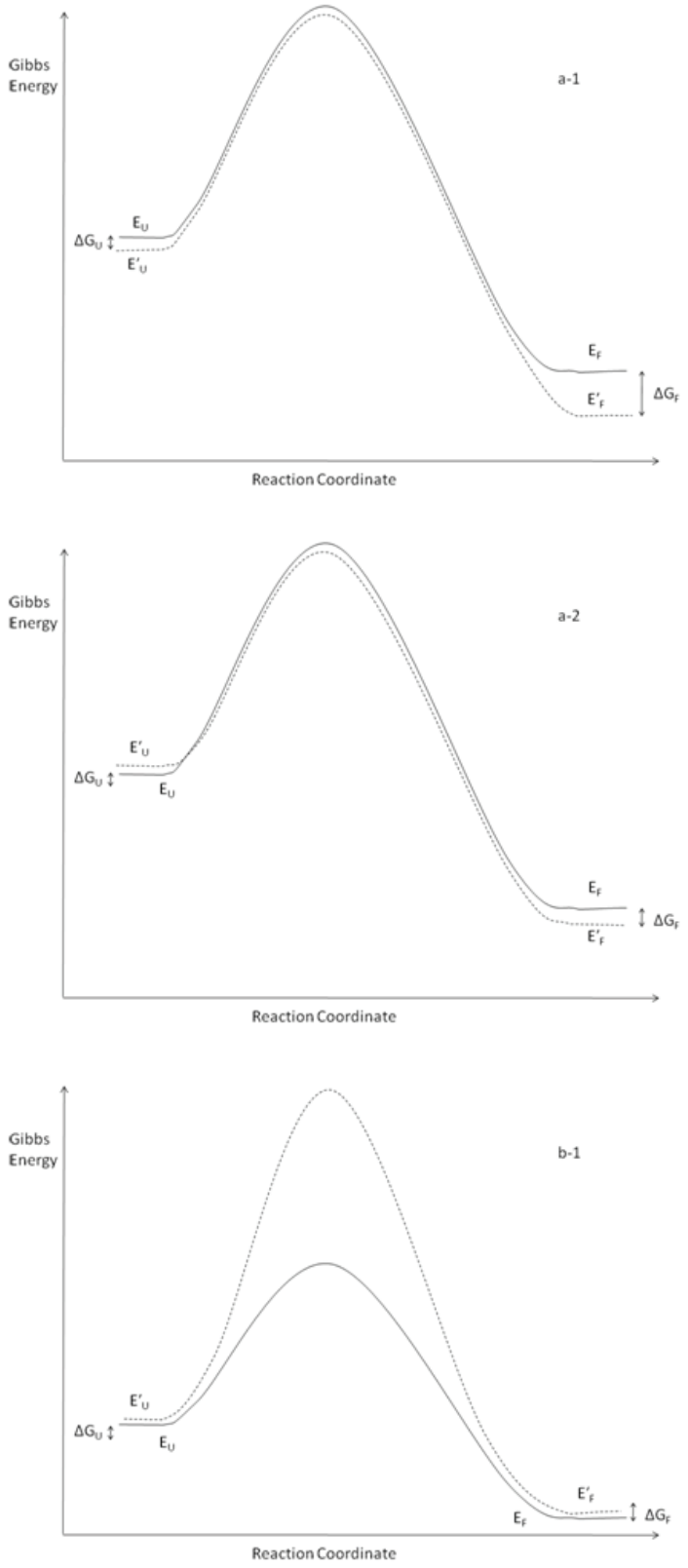


Figure 2.1 Energy diagrams of adjusting the ground state entropy term (a-1), adjusting the ground state enthalpy term (a-2), and mutation or modification of proteins to kinetically trap the folded state (b-1)

2.2 Background

2.2.1 Factors enhancing thermostability

Optimal activity at high temperatures and thermostability are inherent properties of thermophilic enzymes. Enzyme thermostability encompasses thermodynamic and kinetic stabilities. The free energy of stabilization (ΔG_{stab}) and melting temperature (*Tm,* the temperature at which 50 % of the protein is unfolded) determines the thermodynamic stability. Kinetic stability depends on the energy barrier to unfolding and regularly expressed as its half-life ($t_{1/2}$) at defined temperatures.

The free energy of stabilization (ΔG_{stab}, where $\Delta G_{stab} = \Delta H_{stab} - T\Delta S_{stab}$) of a protein is the difference between the free energies of the folded and the unfolded states of that protein where ΔH is the change in enthalpy and ΔS is the change in entropy between folded and unfolded states of the protein. The ΔG_{stab} of globular mesophilic proteins is typically between 5 and 15 kcal/mol at 25°C. ΔH_{stab} (the enthalpy of stabilization) and ΔS_{stab} (the entropy of stabilization) are large numbers changing almost linearly with temperature in the temperature range of the activities of most enzymes. The enthalpy and entropy changes between the folded and unfolded states of the protein are function of temperature

$$\frac{\partial \Delta H}{\partial T} = \Delta C_p$$

$$\frac{\partial \Delta S}{\partial T} = \Delta C_p / T$$

where ΔC_p is the is the change in the heat capacity of the protein between the folded and unfolded states. Usually, it is assumed to be constant in the temperature range relevant to protein stability studies.

In Figure 2.2, curve (c) represents the mesophilic protein and curves (a), (b) and (d) represent the thermophilic proteins. Protein (a) has the same temperature of maximal stability (T_s) as the mesophilic protein (c) and the ΔG vs. T curve of thermophilic protein is shifted upward. Protein (b) has the same T_s and ΔG of stabilization values as mesophilic protein (c) but protein (b) has clearly higher melting temperature (T_m). Thermophilic protein (d) and mesophilic protein (c) have different T_s values but have

the same ΔG of stabilization at their T_s. Protein (d) has clearly higher T_m with respect to mesophilic protein (c).

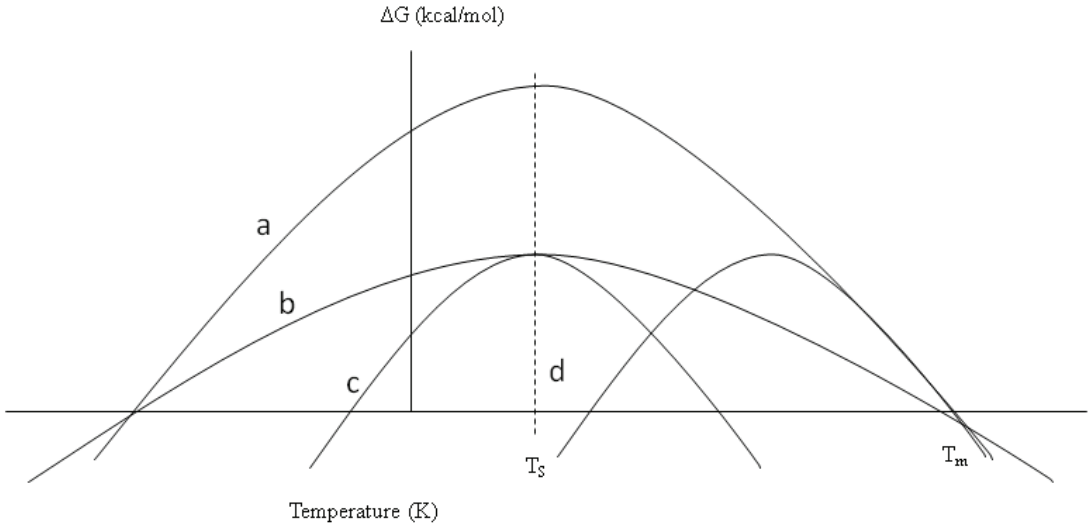


Figure 2.2 Comparison of theoretical ΔG values of mesophilic and thermophilic proteins. (c) is the ΔG vs. T curve for mesophilic protein and (a), (b) and (d) are the ΔG vs. T curve for thermophilic proteins.

Enough experimental evidence (e.g., sequence, mutagenesis, structure, and thermodynamics) has been collected over recent years about thermophilic proteins and conclusion could be that no single mechanism is responsible for the significant stability of these proteins. Increased thermostability must be found in a small number of highly specific mutations that often do not obey any obvious rules.

Amino Acid Composition

The comparison of residue contents in hyperthermophilic and mesophilic proteins based on the genome sequences of eight mesophilic and seven hyperthermophilic organisms shows only minor trends (Vieille and Zeikus, 2001). More charged residues are found in thermophilic proteins than in mesophilic proteins, mostly at the expense of uncharged polar residues. Thermophilic proteins also contain slightly more hydrophobic and aromatic residues than mesophilic proteins. As more experimental data accumulate, especially from complete genome sequences, it can be concluded that the thermophilic adaptation cannot be defined in terms of significant differences in the amino acid composition as Bohm et al. stated. (Bohm and Jaenicke, 1994).

Therefore, a bias in the amino acid composition of thermophilic protein could be evolutionarily relevant, rather than an indication for the adaptation to high temperatures. The distribution of the residues and their interactions in the protein are, most likely, more relevant to thermostability than amino acid composition.

Disulphide Bridges

Disulphide bonds can significantly stabilize the native structures of proteins. The effect is recognized to be due to a decrease in the configurational chain entropy of the unfolded protein (Matsumura et al., 1989). On the other hand, 100 °C was believed to be the upper limit for the stability of proteins containing disulphide bridges (Volkin and Klibanov, 1987), because of the sensitivity of cysteines and disulphide bridges to high temperatures. Nevertheless, not all disulphide bridges have equal susceptibility to heat denaturation since their position in protein backbone is also important.

Hydrophobic interactions

Hydrophobic interactions are a stabilization mechanism in thermophilic proteins. Hydrophobicity drives the protein to a collapsed structure from which the native structure is defined by the contribution of all types of forces, hydrogen bonds, ion pairs and Van der Waals interactions (Dill, 1990). Dill et al. reviewed their findings; (i) nonpolar solvents denature proteins; (ii) hydrophobic residues are typically sequestered into a core, where they largely avoid contact with water; (iii) residues and hydrophobicity in the protein core are more strongly conserved and related to structure than any other type of residue (replacements of core hydrophobic residues are generally more disruptive than other types of substitutions); and (iv) protein unfolding involves a large increase in heat capacity.

Given the vital role of the hydrophobic effect in protein folding, we can clearly assume that the hydrophobic effect is also the major force responsible for protein stability. Thermophilic and mesophilic proteins share some common properties; hydrophobic interactions and core residues are better conserved and many stabilizing substitutions are found in the solvent exposed surfaces. The high level of similarity found in the core of mesophilic and thermophilic proteins suggests that even mesophilic proteins are packed almost as efficiently as possible and that there is not much room left for stabilization inside the protein core. In conclusion, although hydrophobic interactions and protein core greatly influence or even determine the protein stability, small number of mutations are still key factors for the thermal stability of protein.

Hydrogen Bonds

RNase T1 contains 86 Hydrogen bonds with an average length of 2.95 Å and the effect of hydrogen bonds on RNase T1 stability has been carefully studied (Shirley et al., 1992). Hydrogen bonding contribution to RNase T1 stability was found to be comparable to the contribution of hydrophobic interactions. Since the identification of hydrogen bonds is highly dependent on the distance cutoff and because a number of thermophilic protein structures have not been refined to satisfactorily high resolutions, studying the role of hydrogen bonds in thermostability by structure analysis has not been completed yet (Vieille and Zeikus, 2001).

Ion Pairs

Ion pairs are not considered as a driving force in protein folding since they are not highly conserved and small numbers of ion pairs can be found in proteins (Dill, 1990). A single ion pair was calculated to be responsible for a 3 to 5-kcal/mol stabilization of T4 lysozyme (Anderson et al., 1990). The *P. furiosus* enzyme stability decreased at low pH values (where acidic residues are protonated and disrupt favorable ionic interactions) and at high salt concentrations (salts are known to destabilize protein ion pairs) suggesting that ion pairs are essential in maintaining this enzyme stability at high temperatures (Ogasahara et al., 1998).

Decreasing Entropy of Unfolding

The stability of protein can be increased by selected amino acid substitutions that decrease the configurational entropy of unfolding (Matthews et al., 1987). In the unfolded state, glycine has the highest conformational entropy since it does not have β-carbon. On the other hand, proline has the lowest conformational entropy and can adopt only a few configurations and restricts the configurations allowed for the preceding residue (Sriprapundh et al., 2000). Therefore, Gly → Xaa or Xaa → Pro mutations may decrease the entropy of unfolded state and stabilized the protein in a condition that given mutations do not introduce unfavorable strains in the protein structure.

Protein Engineering

The high conservation of the protein core, generally composed of α-helices and β-strands, in both thermophilic and mesophilic proteins suggest that the core is already optimized for stability. Therefore, mutations targeted to the protein core are often destabilizing. The most promising strategies for thermostabilization using site directed mutagenesis should be targeted to the surface of the protein of interest.

2.2.2 Enzyme Thermostability

Thermostable enzymes are gaining increased attention due to the fact that their inherent stability is well suited for harsh industrial processes. One simple but important advantage of conducting industrial processes at elevated temperatures is the reduction of the risk of contamination by common mesophilic organisms. Higher reaction rates due to the decreased viscosity and increased diffusion coefficient is another advantage at elevated temperatures. Besides these advantages, higher process yields can be achieved due to increased solubility of substrates and products and by a shift of thermodynamic equilibrium for endothermic reactions (Mozhaev, 1993). Thermostable enzymes are also valuable in terms of storage conditions and they also have increased stability against other denaturing conditions.

Amylolytic enzymes

Most industrial starch processes involve hydrolysis of starch into maltose, glucose and oligosaccharides. The starch polymer requires a combination of enzymes for complete hydrolysis such as α-amylases, glucoamylases or β-amylases (Guzman-Maldonado and Paredes-Lopez, 1995). The starch hydrolysis enzymes comprise 30 % of the world's enzyme consumption (van der Maarel et al., 2002). The enzymatic conversion of all starch polymers includes gelatinization which is achieved by heating starch with water and starch is water-soluble only at high temperatures depending on starch concentration and the source (Baks et al., 2007). For hydrolysis of starch to proceed immediately after gelatinization step, thus the enzyme has to be stable at high temperatures. Temperature stability of the enzymes used in the starch industry simply eliminates the cooling requirements among other operational difficulties.

Cellulosic enzymes

Cellulose is the most abundant and renewable nonfossil carbon source on Earth. The difference in the type of bond and the highly ordered crystalline form of the compound between starch and cellulose make cellulose more resistant to digest and hydrolyze. In order to attack the native crystalline cellulose, which is water insoluble and occurs as fibers of densely packed structures, however, thermostable cellulases active at high temperature and high pH are required. The enzymes required for the hydrolysis of cellulose include endoglucanases, exoglucanases and β-glucosidases (Matsui et al., 2000). Since the alkaline pretreatment of cellulose is performed at high

temperatures, thermostable cellulases should be the best candidate catalysts for cellulose degradation. Similarly, in the paper production process, pulping and bleaching steps are both performed at high temperatures and these processes require thermostable enzymes. Using thermostable enzymes in paper industry is not only feasible economically, but also reduces the waste generation providing more environmentally friendly industrial process.

Proteolytic enzymes

Proteases are generally classified into two categories, exopeptidases and endopeptidases. Exopeptidases cleave off the amino acids from the ends of the protein chain and endopeptidases cleave the peptide bonds within the protein. Proteases are very attractive in industrial processes, and constitute more than 65% of the world market (Rao et al., 1998). These enzymes are broadly used in the food, pharmaceutical, leather and textile industries (Mozersky et al., 2002). Availability of thermostable proteolytic enzymes is very important for detergent industry which requires elevated operational temperatures. On the other hand, there are attempts to use cold reactive enzymes in detergent industry to eliminate the need of high temperatures and to use the cold tap water in order to catalyze the reaction.

Lipolytic enzymes

Lipases of microbial origin are the most versatile enzymes and are known to bring about a range of bioconversion reactions such as hydrolysis, interesterification, esterification, alcoholysis, acidolysis and aminolysis (Jaeger et al., 1994). Lipases are used in the food industry in order to produce ester compounds such as flavor and aroma constitutes (Gandhi et al., 1995). Whereas long chain methyl and ethyl esters of carboxylic acid moieties provide important oleo chemicals that may function as fuel for diesel engines, esters of long chain carboxylic acid and alcohol moieties have applications as lubricants and additives in cosmetic formulations (Fjerbaek et al., 2009). Lipases from bacteria and fungi are the most commonly used enzymes for transesterification, and optimal parameters for the use of a specific lipase depend on the origin as well as the formulation of the lipase. In the paper industry, lipases involved in the process of removal of the pitch from pulp.

Lipases are generally used in the industrial processes at temperatures exceeding 45 °C. In order to be functional, the enzymes are required to exhibit an optimum temperature of around 50 °C. Most of the industrial processes in which lipases are

employed function at temperatures exceeding 45 °C. The enzymes, thus, are required to exhibit an optimum temperature of around 50 °C (Sharma et al., 2002). Some enzymatic processes for the physical refining of seed oils have some distinct requirements, including pH of 5.0 and optimal temperature of around 65 °C. The enzymatic reaction followed by the separation of the lysophosphatide from the oil at about 75 °C (Dahlke, 1998). These kinds of reactions, therefore, are enhanced through the utilization of thermostable lipases.

Among the desirable characteristics that commercially important lipases should exhibit, alkali tolerance and thermostability are the most important (Kulkarni and Gadre, 1999). Although, few lipases exist which are able to operate at 100 °C, their half-lives are reported to be short (Rathi et al., 2000).

Lipases are of widespread occurrence throughout the microbial life. More abundantly they can be found in bacteria, fungi and yeasts (Wu et al., 1996). There is a continuous search for sources producing highly active lipolytic enzymes with stability towards pH, temperature, organic solvents and ionic strength.

Other enzymes

The polymerase chain reaction (PCR) is a process which gave huge advancement in genetic engineering due to its ability to amplify DNA. In the process, there are repeating cycles of denaturation (at a temperature of 95 °C), primer annealing (at a temperature around 55 °C) and elongation (at a temperature of 72 °C) (Mullis and Faloona, 1987). Development of this procedure has been facilitated by the availability of thermostable DNA polymerases, which catalyze the elongation step. In the earlier PCR procedures, DNA polymerases which were isolated from *E. coli* were utilized. These mesophilic enzymes, however, lost their enzymatic activities at elevated temperatures during denaturation step and, adding a new polymerase enzyme after each cycle before elongation steps was necessary. This was a time-consuming and costly way of DNA amplification. Taq polymerase from the bacterium *Thermus aquaticus* was the first thermostable DNA polymerase characterized (Chien et al., 1976) and repeated exposure to temperatures above 95 °C during cycles of PCR had little effect on the enzyme activity.

2.2.3 *Rhizopus oryzae* Lipase

Lipase enzymes catalyze not only the hydrolysis of ester bonds of triacylglycerols but also the synthesis of ester bonds and transesterification. Transesterification by lipases is particularly useful in industry, such as kinetic resolution of 1-phenylethanol (Schofer et al., 2001) or alpha-methylene beta-lactams through lipase catalyzed kinetic resolution (Adam et al., 2000).

The structures of lipases from various sources have been disclosed, such as *Rhizopus delemar* (Derewenda et al., 1994), *Candida antarctica* (Qian et al., 2009), or *Geotrichum candidum* (Schrag and Cygler, 1993). Although they have been isolated from different sources and have different amino acid sequences, most of them share common features. In general, lipases have a catalytic triad of three amino acids, Ser – His – Asp with an active Ser residue. Asp residue may be replaced by Glu in some cases. Also, lipases share a consensus motif of Gly – X – Ser – X – Gly, where X may be any amino acid and Ser is one of the amino acids in the catalytic core. Furthermore, lipases belong to the α/β hydrolyze family, which is a common 3-D fold in several other hydrolyzes (Ollis et al., 1992). The active site Ser residue is placed in a loop termed as the catalytic elbow. Moreover, lipase has a so-called flexible lid which shelters the catalytic center in solution in the resting form. Interfacial activation takes place in the presence of a substrate by the movement of a lid and exposure of the hydrophobic pocket and the active site structure (Brzozowski et al., 1991). The *Rhizopus oryzae* lipase (ROL) secreted naturally from the lipolytic fungus shows these common properties. The native ROL enzyme is a 392 amino acid protein, first 26 amino acid being signal a sequence, followed by a 97 amino acid pro-region and 269 amino acid mature protein (Beer et al., 1996). Although the 3-D structure of *Rhizopus oryzae* lipase has not been identified, *Rhizopus niveus* lipase (RNL) was crystallized by the hanging drop vapor diffusion at 2.5 A resolution, 1LGY (Kohno et al., 1996). Sequence alignment showed that these two lipases have 98.5 % sequence identity, only Ile254Leu and His134Asn are different. Isoleucine and Leucine residues have hydrophobic side chains and Histidine and Asparagine residues have polar side chains. Moreover, BLOSUM (BLOcks of Amino Acid SUbstitution Matrix) scores for these two pair of amino acids show that Ile-Leu and His-Asn substitutions are possible (Henikoff and Henikoff, 1992). We have decided to use 1LGY structure for computational studies

instead of performing homology modeling for ROL protein. The catalytic residues and the secondary structural features of 1LGY are presented in Figure 2.3.

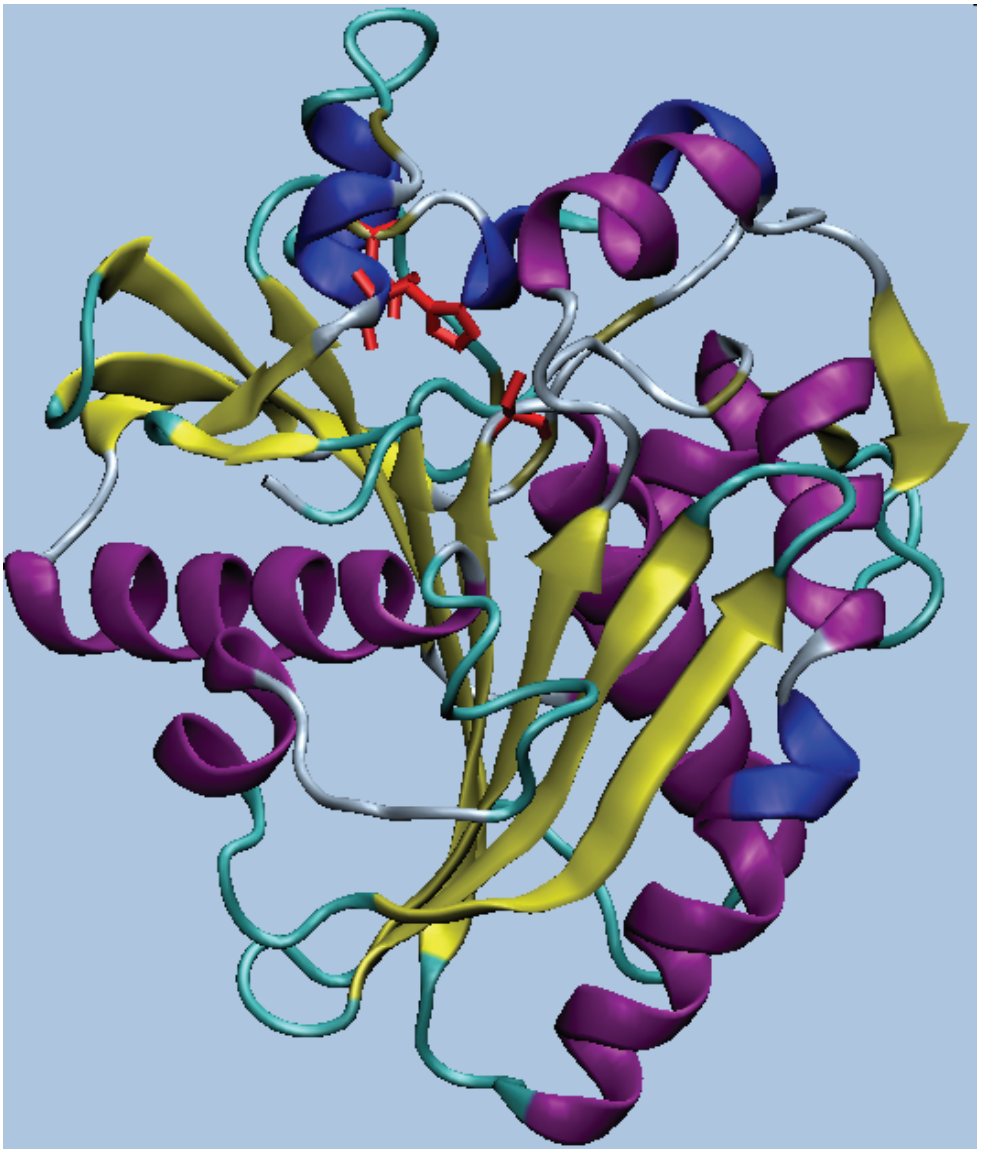


Figure 2.3 1LGY secondary structural features. Active site residues (SER 145, ASP 204, and HIS 257) are colored in red. Main secondary structures are colored as follows, Alpha helix: purple, 3_{10} Helix: blue, Extended sheet: yellow and coil: white.

SDS-PAGE analysis of the purified lipase showed a single protein band of 30 kDa similar to the size of the *E. coli* derived recombinant ROL (Minning et al., 1998). Although the ROL enzyme contains four common *N*-glycosylation sites of the sequence Asn-X-Ser/Thr, Minning et al., have showed no difference in the molecular weight of recombinant ROL before and after incubation with endo-β-N-acetylglycosamidase H,

suggesting that no glycosylation had occurred. Native enzyme has three disulphide bonds, between residues 29 and 268, 40 and 43, and the 235 and 244 (Figure 2.4).

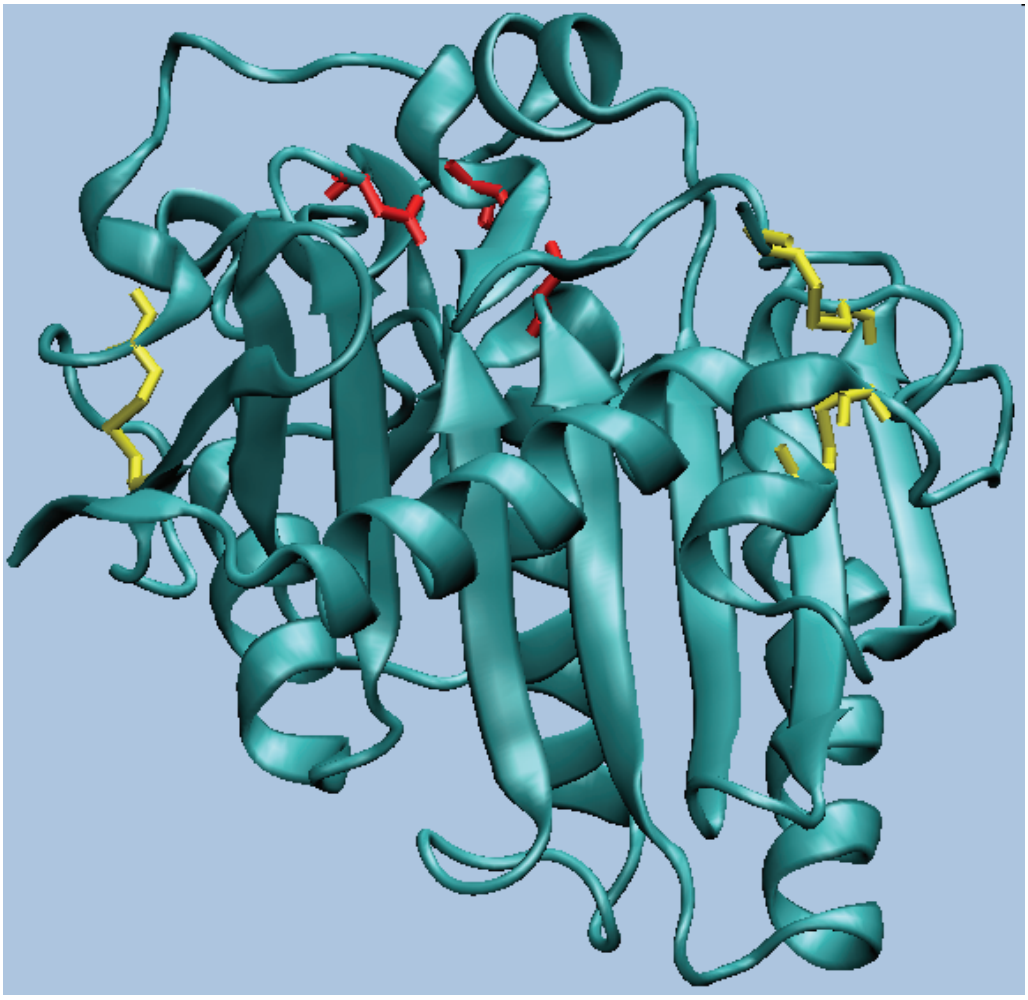


Figure 2.4 Structure of 1LGY. Active site residues (SER 145, ASP 204, and HIS 257) are colored in red. Disulphide bonds (29-268, 40-43, and 235-244) are colored in yellow.

2.2.4 Molecular Dynamics Simulations

With the selected subset of proteins, MD simulations can be performed to see the effect of the selected residues to overall protein stability. Several authors have previously utilized detailed MD studies to understand protein stability. Parthasarathy

and his coworkers have done a preliminary analysis suggesting that the crystallographically determined atomic displacement parameters might actually convey information regarding the thermal stability (Parthasarathy and Murthy, 2000). They have highlighted the variation of the crystallographic B-factors (Debye-Waller factors) from the core to the surface of proteins and found the amino acid compositions in high B-factor value regions of the polypeptide chains. Similarly, the mean square displacements of non-hydrogen atoms can be monitored by using Debye-Waller factors from the output of MD simulations (Baysal and Atilgan, 2005). In another study, the mean square displacements of each residue were investigated via MD to determine the thermal stability change in thioredoxins (Pedone et al., 1998). In that study, they have confirmed, via MD, the experimental observation that the replacement of residues reduces the backbone flexibility, thereby stabilizing the protein. Also by using MD, they were able to correlate the increase in protein stabilization caused by single amino acid replacements. Dynamic studies of different proteins performed by Dagget et al. using high temperature MD, in which they checked the mean square displacements of the residues, suggests that loop regions in proteins undergo largest deviations and that at higher temperatures these may be the regions of the protein that unfold first during thermal denaturation (Daggett and Levitt, 1993).

Molecular dynamics has provided many insights concerning the internal motions of biomolecules. Capacity of enabling investigation of biological motions that are often not accessible by experimental setups gives a power for this approach. Theory and experimental setups should be used in combination in order to understand structural features of proteins. Although, longer simulation times and larger systems requires more computational power, speed of computers and parallel computing algorithms are getting more advanced levels. There are many software packages available for MD simulations of biomolecules (Karplus and McCammon, 2002). We have used the NAMD software package for the molecular dynamics simulations of both native and mutant enzymes (Phillips et al., 2005).

2.3 Materials

2.3.1 Structures

1LGY and 1TIB (Derewenda et al., 1994; Kohno et al., 1996) were used throughout this study. Mutant structures for ROL enzyme were generated using 1LGY as a template using "Mutator Plugin, version 1.0" of VMD software.

2.3.2 Software

Flexweb: FRODA (Framework Rigidity Optimized Dynamical Algorithm) is a constrained geometric simulation method that models a protein as a number of rigid units which move relative to each other by an unrestricted motion about the dihedral bonds. A graph-theoretical algorithm called FIRST (Floppy Inclusion and Rigid Substructure Topography) is used to identify the rigid regions in a protein based on the non-covalent interactions present (http://flexweb.asu.edu/) (Jacobs et al., 2001; Rader et al., 2002).

NAMD software (version 1.6.1): "NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. Based on Charm++ parallel objects, NAMD scales to hundreds of processors on high-end parallel platforms and tens of processors on commodity clusters using gigabit ethernet. NAMD is distributed free of charge with source code"(http://www.ks.uiuc.edu/Research/namd/) (Phillips et al., 2005).

VMD software (version 8.6.): "VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics and built-in scripting. VMD supports computers running MacOS X, Unix, or Windows, is distributed free of charge, and includes source code" (http://www.ks.uiuc.edu/Research/vmd/) (Humphrey et al., 1996).

2.3.3 Parallel Computing Infrastructure

During the course of this thesis work, MD simulations were performed at the Barcelona Supercomputer Center (BSC), Sabanci University Servers, and ULAKBIM High Performance and Grid Computing Center.

2.4 Methods

2.4.1 *In silico* mutagenesis

In silico mutagenesis studies were carried out using Mutator plugin, v1.0 of VMD software using 1LGY as a template structure. Energy minimizations were performed on all of the created mutated structures using NAMD package. Briefly, plugin takes psf and pdb files generated using VMD psfgen Plugin, Version 1.4, as inputs and replace specific residue with desired amino acid. Since plugin does not support disulphide bonds, energy minimizations were performed for all of the mutated structures only after the DISU patch applied.

2.4.2 Preparation of structures for MD simulation

The MD runs have been performed for native and mutant proteins in water. Initial velocities were generated from Boltzmann distribution at the designated temperature. All the bonds of the protein and the water molecules were constrained by the RATTLE algorithm (Andersen, 1983). Prior to MD runs, the systems were equilibrated for 500ps with a timestep of 2 fs, while maintaining the temperature by velocity rescaling. The data collection stages were 10 ns length. At this stage the temperature control was established via Langevin Dynamics. Trajectory data were recorded at 2 ps intervals. The data sets were collected at 300 K, 350 K, 400 K, 450 K, 500 K, 550 K, and 600 K.

2.5 Results and Discussion

2.5.1 Flexibility Analysis

Although they have 30.8 % identity, Root Mean Square Deviation (RMSD) value for 1LGY and 1TIB structural alignment is 1.9 Å, and Z score is 7.2. To determine the factors those are important for stability, I have focused on the flexibility of the proteins. Based on the structural alignment I calculated the flexibility differences of each amino acid or domain between two structures using Flexweb software (Hemberg et al., 2006).

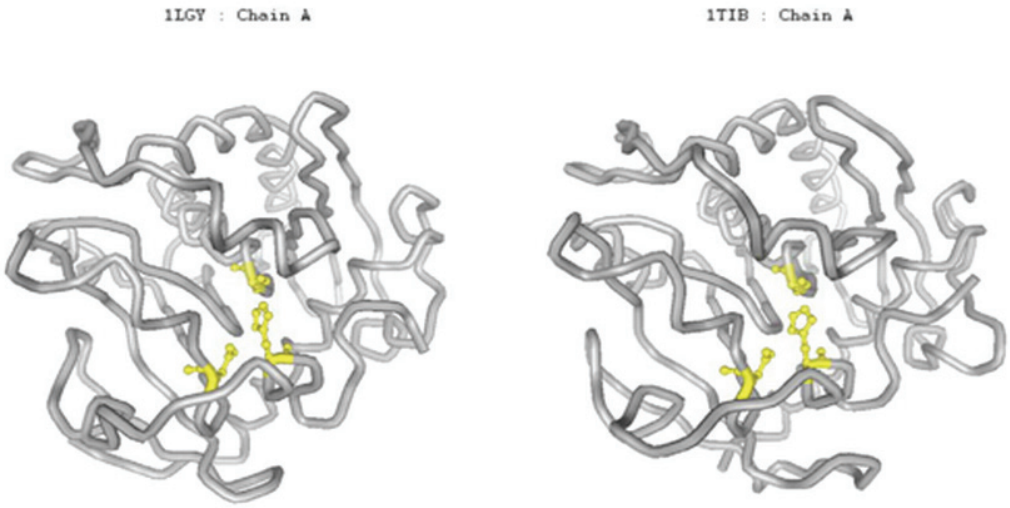


Figure 2.5 CE alignment of native structures of 1LGY and 1TIB. Active site residues (Serl45, Asp204, and His257 for 1LGY and Ser146, Asp201, and His258 for 1TIB) are shown in yellow.

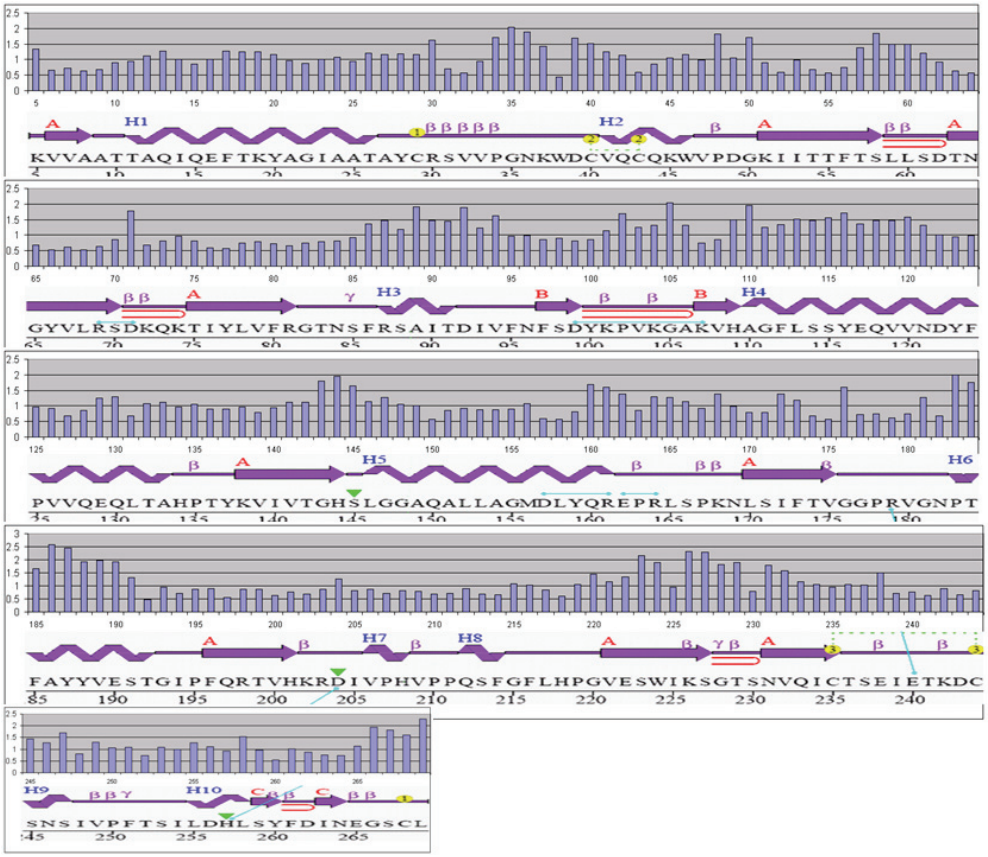


Figure 2.6 Flexibility analysis of 1LGY. Y-axis shows the RMSD value and X-axis shows amino acid sequence with secondary structural features. Green triangle shows the Catalytic residues (Serl45, Asp204, and His257) and yellow circles shows disulphide bond forming Cysteine residues.

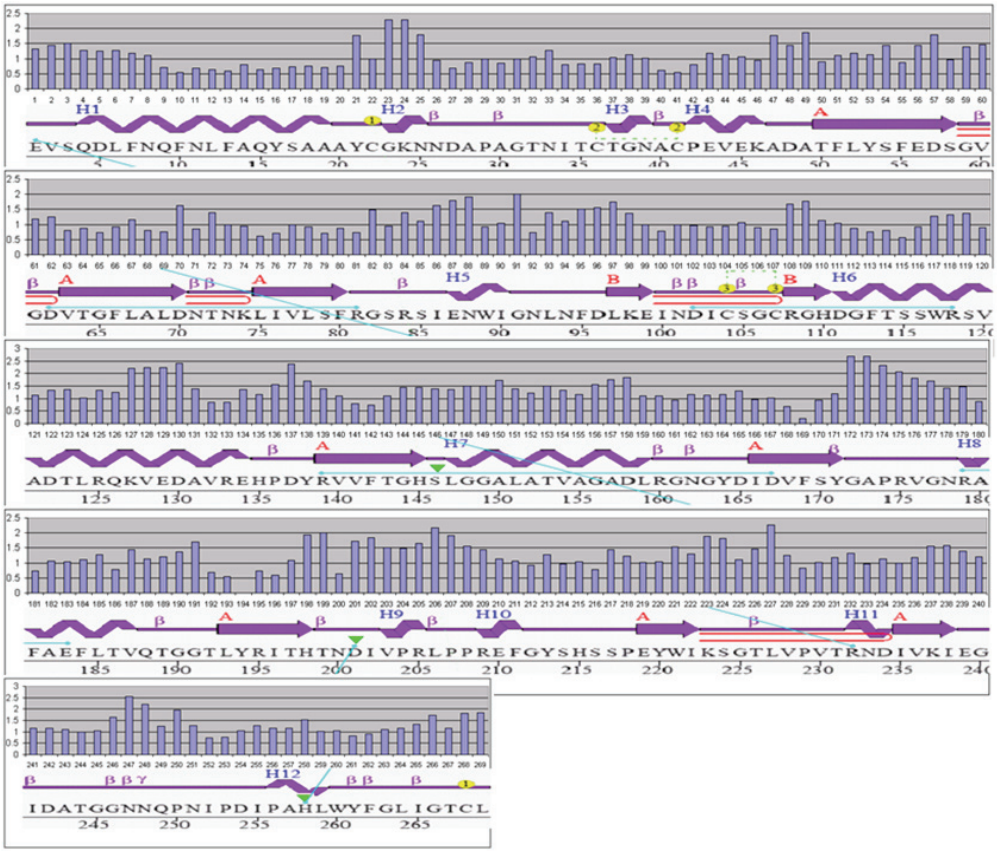


Figure 2.7 Flexibility analysis of 1TIB. Y-axis shows the RMSD value and X-axis shows amino acid sequence with secondary structural features. Green triangle shows the Catalytic residues (Ser146, Asp201, and His258) and yellow circles shows disulphide bond forming Cysteine residues.

Analysis of flexibility values for 1LGY and 1TIB showed that flexibility itself may play a role for the increased thermostability of 1TIB. Although overall flexibility of 1LGY and 1TIB is similar when I analyzed the all of the residues, there is a difference around the catalytic triad (Serl45, Asp204, and His257 for 1LGY and Ser146, Asp201, and His258 for 1TIB) both in 7 Å and 10 Å distances (Figure 2.8). The difference between the flexibility values of Ser145, Asp204 and His257 were high comparing to the overall flexibility values for 7 Å and 10 Å.

When I increase the radius, I observed that the flexibility values of the 14 Å distance became very similar to the overall flexibility. Therefore, I can conclude that the flexibility of residues around the catalytic sites influence the thermostability of the ROL (1LGY) enzyme.

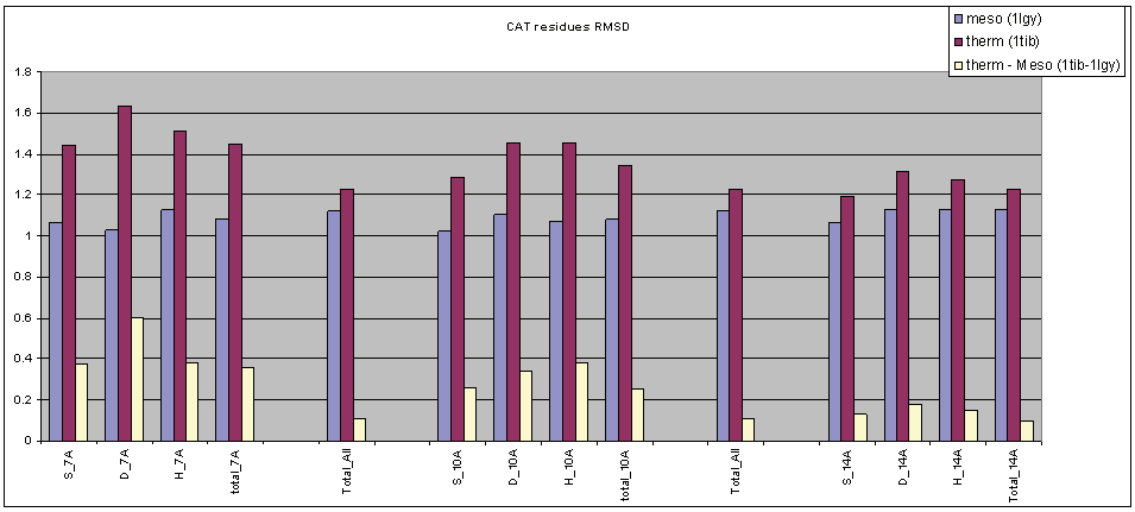


Figure 2.8 Comparison of RMSD values for 1LGY and 1TIB in 7 Å, 10 Å and 14 Å of catalytic residues (Ser145 (S), Asp204 (D) and His257 (H)) and overall RMSD for all of the residues. Red colored bar shows thermophilic enzyme (1tib), Blue colored bar shows mesophilic enzyme (1lgy) and yellow colored bar shows the difference.

2.5.2 Enthalpic and Entropic Contributions

Proposed mutations in this study can be explained using enthalpic and entropic contributions of each term, S (entropy) and H (enthalpy). Adjusting the S term of the ground state by mutating the residues to thermodynamically favorable state should stabilize the protein.

In the case of Proline to Glycine mutation at certain positions (P183G, P195G, P207G, and P219G) that increase in the flexibility of residues around active site, the gain of chain flexibility and freedom are more noticeable in the folded state than the unfolded state because the folded state has less freedom to begin. The absolute S term (entropy) of the ROL enzyme will increase more in the folded state. Overall, the absolute G term (Gibbs Energy) will be noted to decrease little in the unfolded state (E'_U in Figure 2.9) and more in the folded state (E'_F in Figure 2.9), corresponding to an increase of the stability of the folded state (ΔG of folding is more negative, $|E'_F - E'_U| > |E_F - E_U|$).

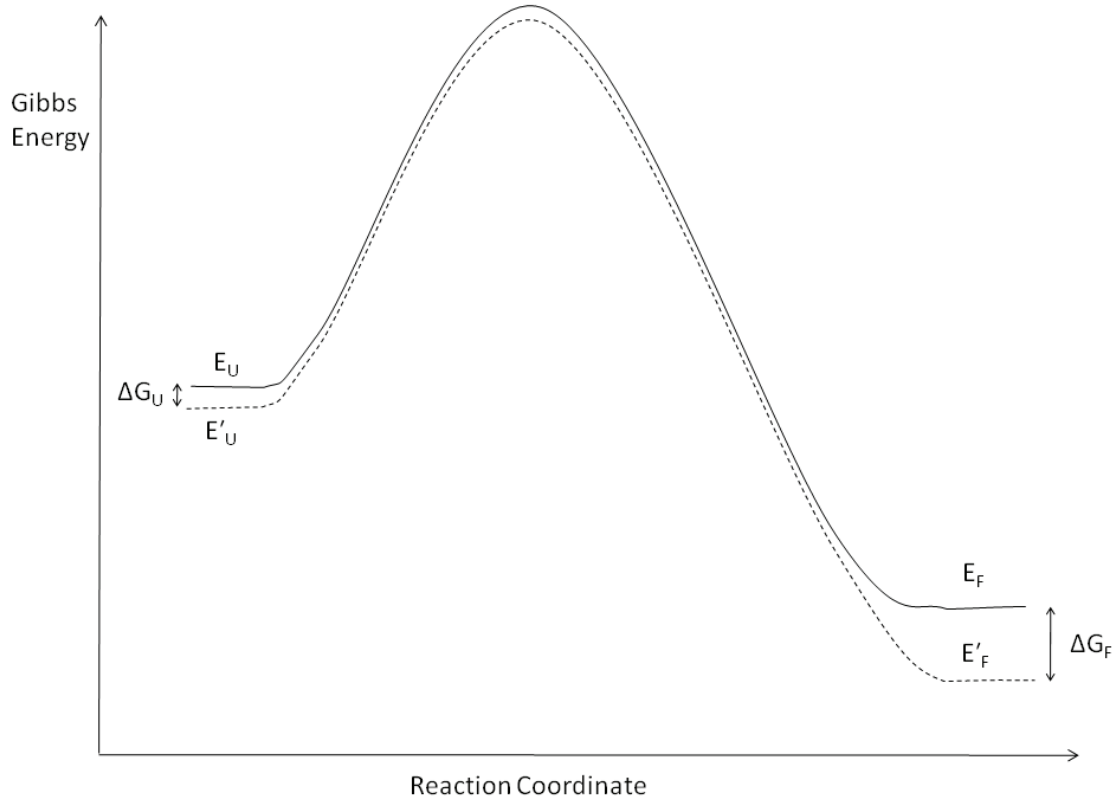


Figure 2.9 Effect of Flexibility: Gibbs Energy diagram of native and mutant ROL enzymes. The solid line represents the native structure and the dashed line represents the mutated structure. ΔG_U and ΔG_F are the Gibbs energy difference between mutated and native enzymes at unfolded and folded states respectively. Terms, E_U and E'_U, are the unfolded state Gibbs energies of the native and mutant enzymes respectively. Terms, E_F and E'_F, are the folded state Gibbs energies of the native and mutant enzymes respectively.

2.5.3 Mutant structure analysis

After our findings about flexibility and thermostability relationship for 1LGY and 1TIB, I analyzed some mutant structures in terms of stability and tried to decide which mutations have positive impact on stability. In Figure 2.10, the structural positions of the proline residues are given. In Figure 2.11, RMSD values are plotted onto the protein sequence alignment of 1LGY and 1TIB. Protein alignment was generated using CE alignment algorithm (Shindyalov and Bourne, 1998), based on the structural alignment.

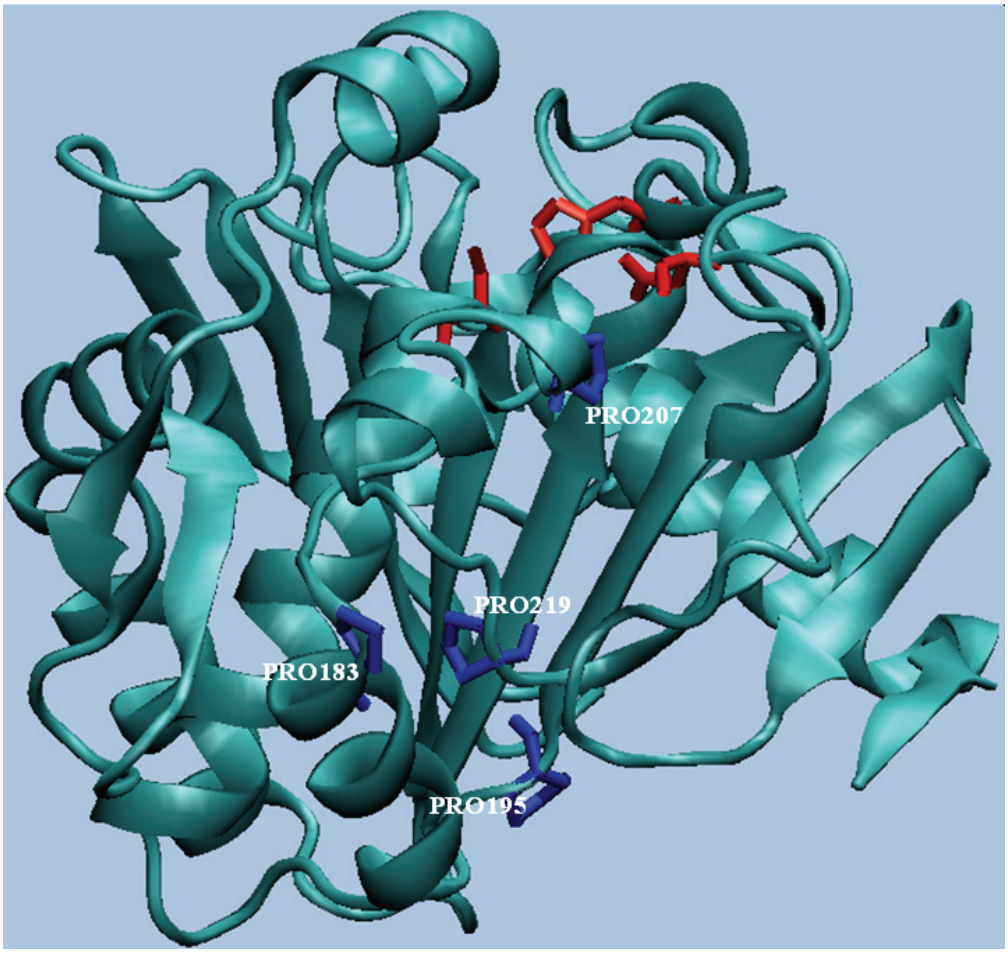


Figure 2.10 Structural positions of Proline residues mutated into Glycine residue. *In silico* mutagenesis methodology is explained in Section 2.4.1.

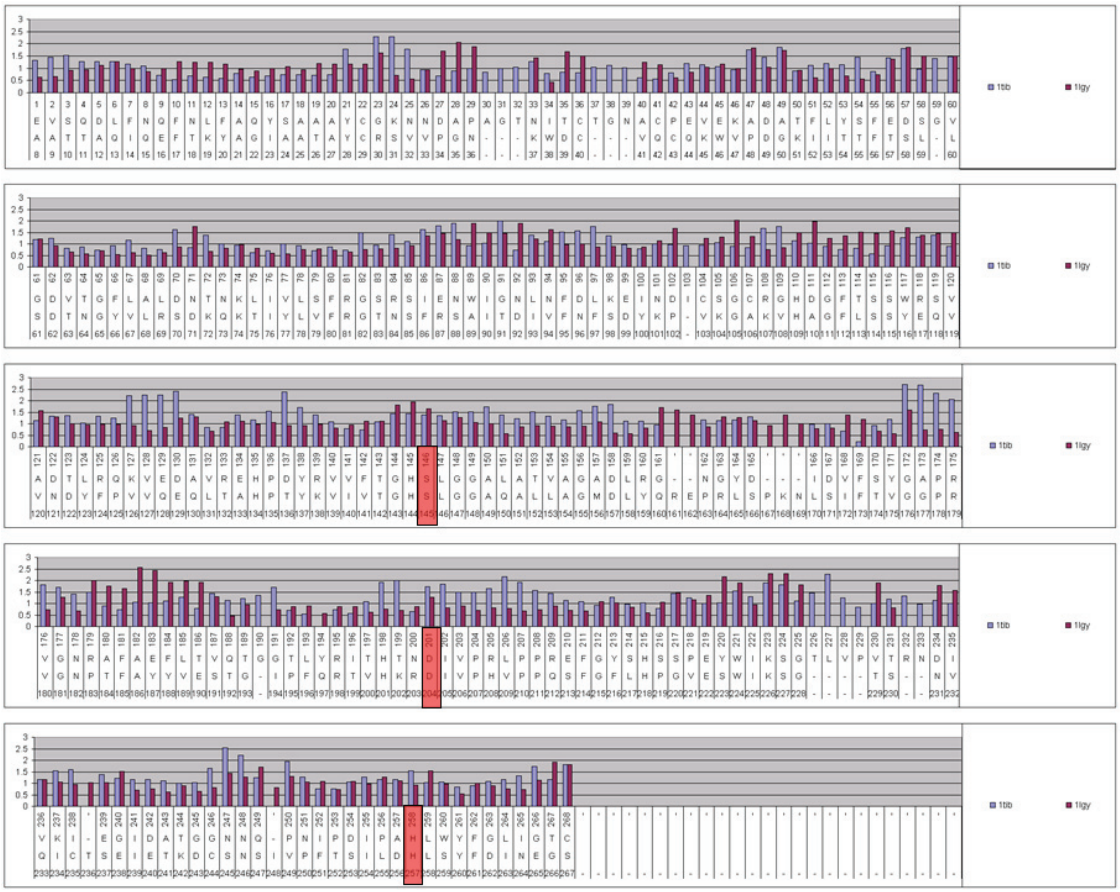


Figure 2.11 Flexibility differences of 1LGY and 1TIB on protein sequence. Catalytic residues are shown in red box. Y-axis shows the RMSD value and X-axis shows 1LGY and 1TIB sequence alignment. Blue colored bar shows thermophilic enzyme (1tib), Red colored bar shows mesophilic enzyme (1lgy).

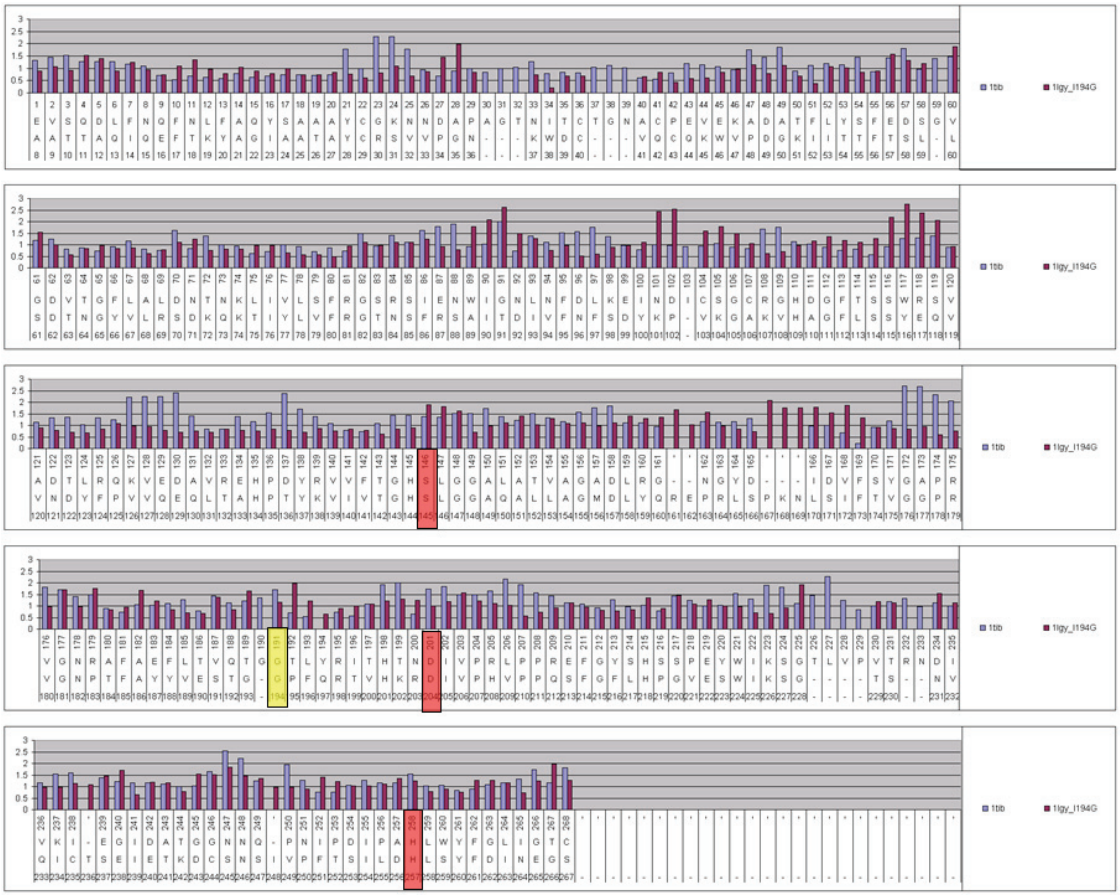


Figure 2.12 Flexibility analysis of P194G mutant structure. Mutation position is shown in yellow box; Catalytic residues are shown in red box. Y-axis shows the RMSD value and X-axis shows 1LGY and 1TIB sequence alignment. Blue colored bar shows thermophilic enzyme (1tib), Red colored bar shows mesophilic enzyme (1lgy).

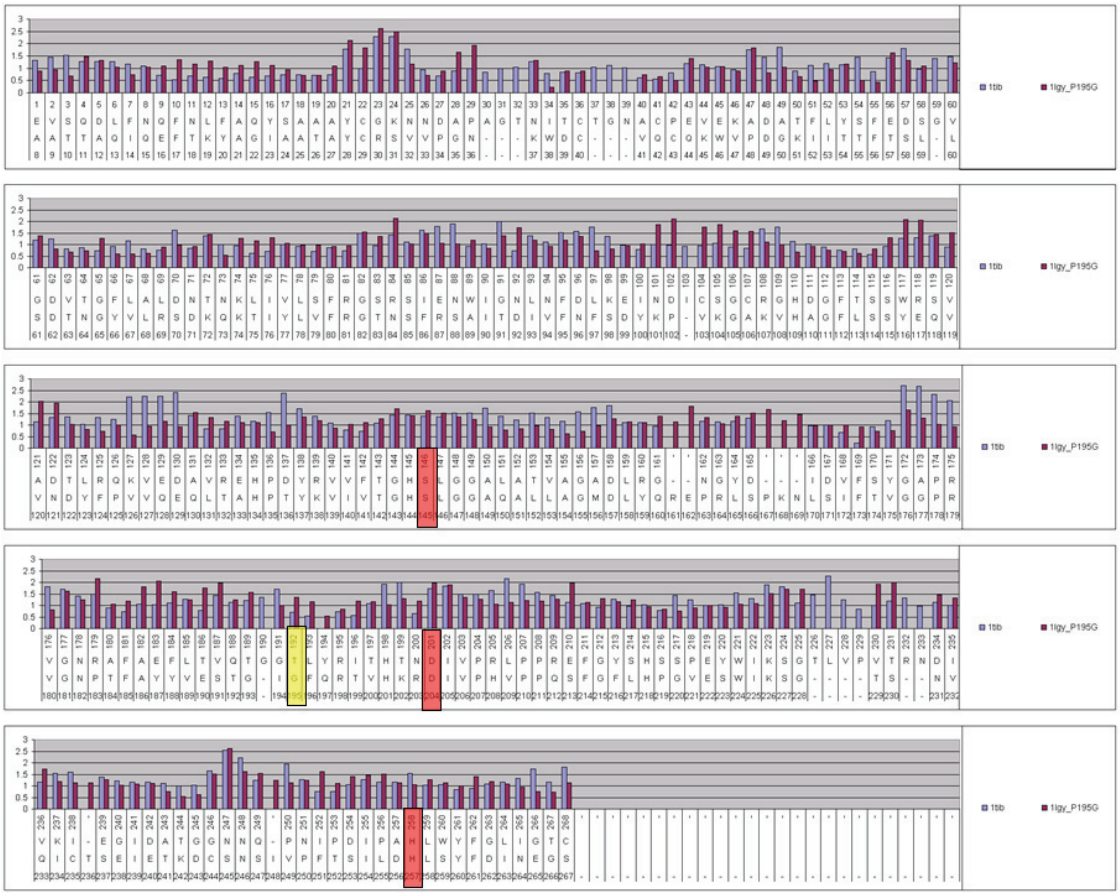


Figure 2.13 Flexibility analysis of P195G mutant structure. Mutation position is shown in yellow box; Catalytic residues are shown in red box. Y-axis shows the RMSD value and X-axis shows 1LGY and 1TIB sequence alignment. Blue colored bar shows thermophilic enzyme (1TIB), Red colored bar shows mesophilic enzyme (1LGY).

Although P194G mutation was beneficial in terms of flexibility around the active site residues (Figure 2.12), mutation caused a relatively higher flexibility difference in the region 170 to 190. On the other hand, P195G mutation increased the flexibility around the active site residues and did not change the other parts of the protein significantly (Figure 2.13). Mutations selected for further studies are shown on the PDBsum representation (Figure 2.14)

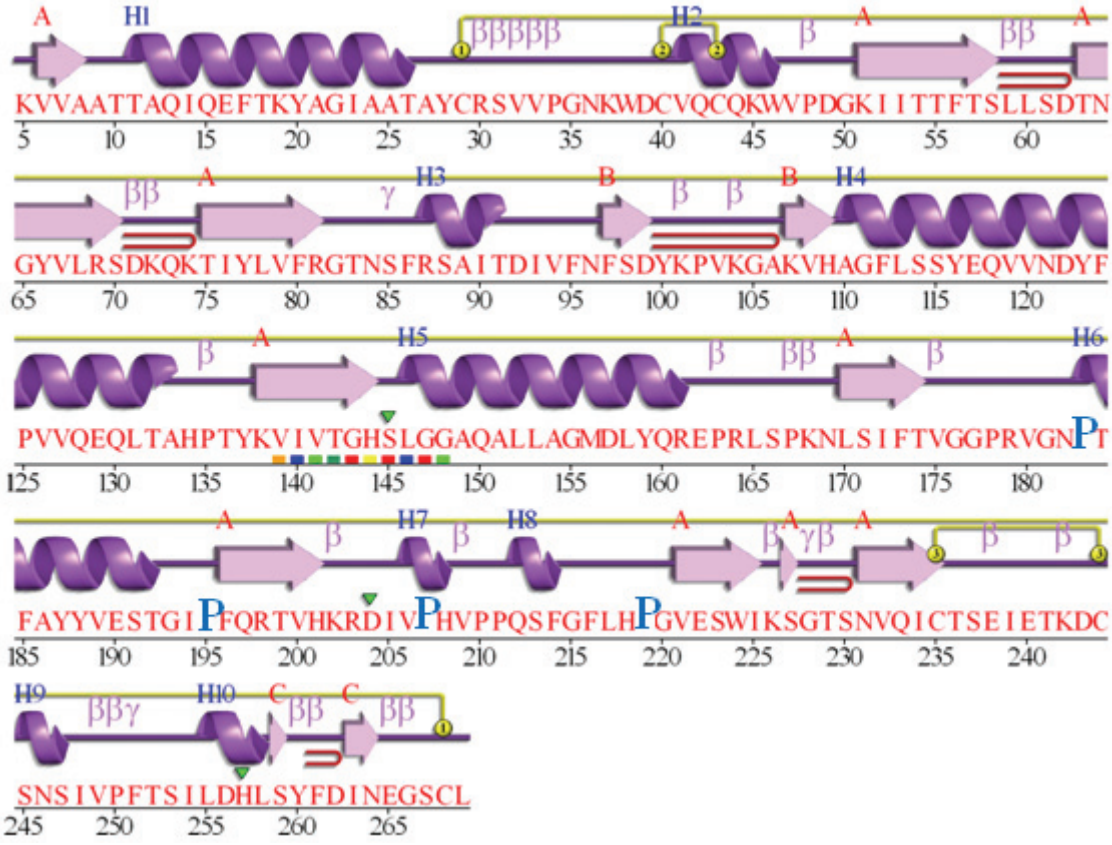


Figure 2.14 PDBsum presentation of 1LGY and mutations (183, 195, 207 and 219). Protein sequence is given in Red color and mutated residues are given in Blue color. Main secondary structures are colored as follows, helix: purple, Extended sheet: pink and coil: black line. Yellow lines and circles represent disulphide bonds and bond forming Cysteine residues.

2.5.4 MD Simulation Results

2.5.4.1 Simulation Temperature

We have performed a series of MD simulations at different temperatures (300K, 350K, 400K, 450K, 500K, 550K, and 600K) in order to find the suitable temperature for the mutation analysis. As seen in the contour plot (Figure 2.15), the root mean squared displacement (RMSD) values of 1LGY are plotted against the simulation time (x10 ps). In Figure 2.15, black regions represent high RMSD values and white regions represent low RMSD values. Briefly, RMSD values were used to determine the unfolding pattern of protein as calculated for each residue during simulation time. After analyzing the

outcomes of MD simulations at different temperatures, I have decided that 450K simulation temperature can be used to compare mutant structures, in terms of thermal stability. At 300 K, we did not observe any unfolding pattern and decided to employ this temperature as a reference simulation when searching the temperature for mutant comparison. At 550 K and 600 K, we have observed too much deviation and that could hinder the effect of mutations supposed to increase the thermal stability.

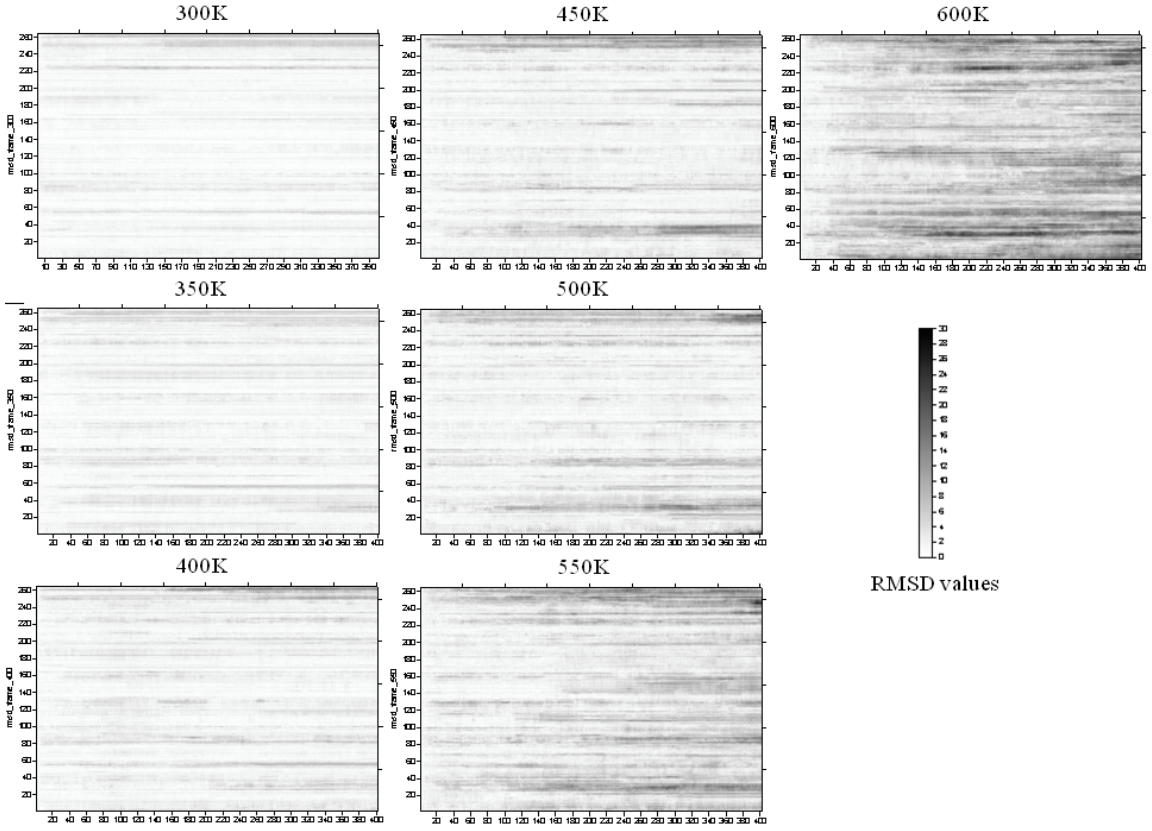


Figure 2.15 RMSD per residue over MD simulation for 1LGY. RMSD values for each residue were calculated taking the minimized structure as a reference. Values are plotted as contour graphs, showing residue number in Y-axis, and timeline (x10 ps) in X-axis. Color bar shows RMSD values in Angstrom, 30 A as Black in color and 0 A as White in color.

2.5.4.2 Active Site Stability

Active site residues for 1LGY (Serl45, Asp204, and His257) and 1TIB (Ser146, Asp201, and His258) were analyzed. Distance between C-alpha atoms of two residues was used as a measure of the thermostability during MD simulation. Ser – Asp residue distance did not change dramatically during simulations; nearly stable around 10 Angstrom (Figure 2.16). On the other hand, distance between Ser – His residues both in 1TIB (thermostable) and 1LGY (mesostable) have been increased during simulations. Although 300 K simulations for 1LGY showed some variance, simulation at 450 K has clearly indicated that distance getting larger during that temperature (Figure 2.17).

1LGY showed more than 20 Angstom change in distance, starting around 7 Angstrom. Similarly, Asp – His residue distance also lost its integrity (Figure 2.18).

Active site residue distance analysis showed that thermostability could be improved if some of the mutations stabilize the active site compactness. For this purpose, all of the mutations for the active site distances have been analyzed.

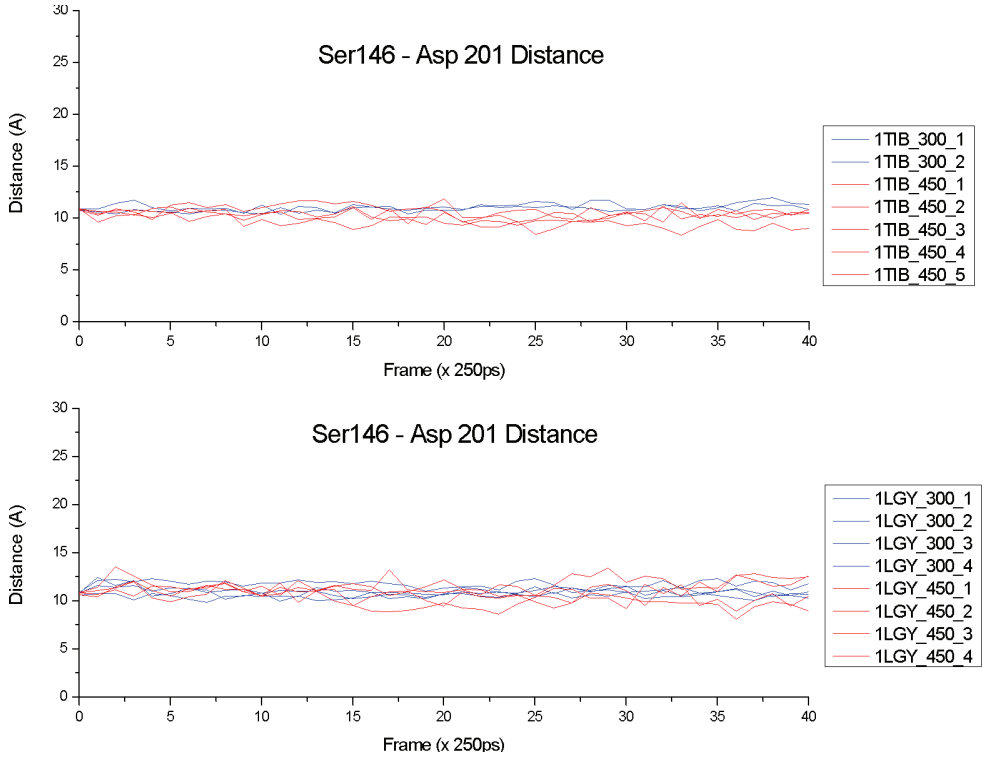


Figure 2.16 Distance between the catalytic Ser and Asp residues of 1LGY and 1TIB at 300 K and 450 K. Red lines represent distance values for 450 K and Blue lines represent distance values for 300 K. 1LGY_300_3 shows a distance between the active site Ser and Asp residues for 1LGY structure at 300 K and 3rd simulation, similarly, 1TIB_450_5 shows a distance between the active site Ser and Asp residues for 1TIB structure at 450 K and 5th simulation.

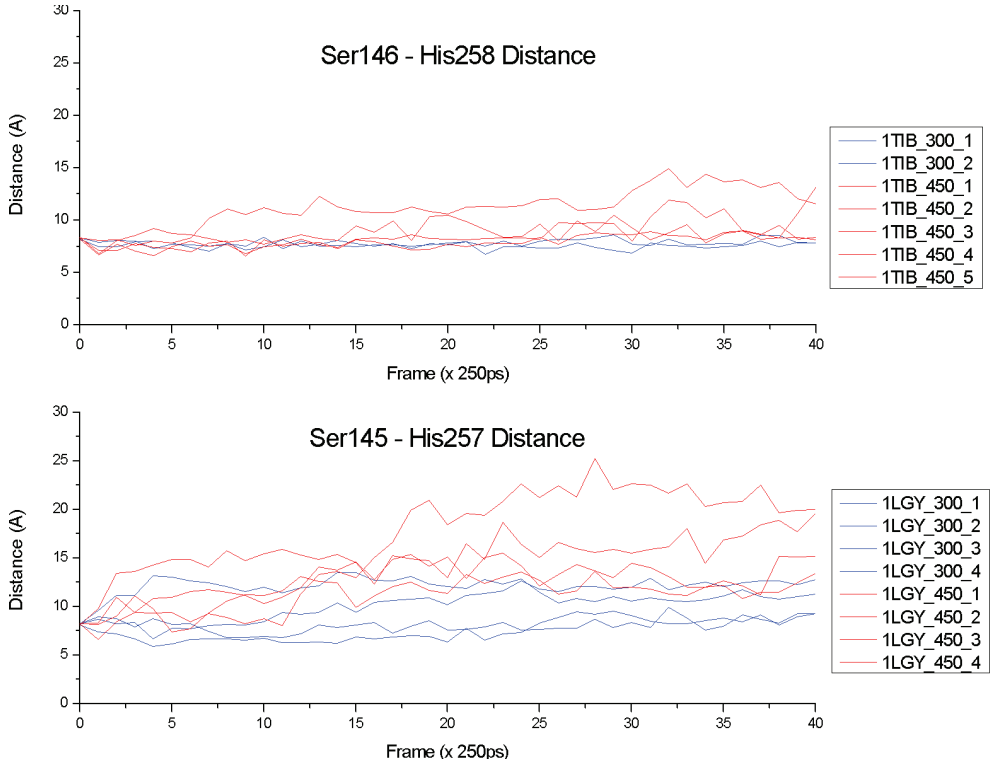


Figure 2.17 Distance between the catalytic Ser and His residues of 1LGY and 1TIB at 300 K and 450 K. Red lines represent distance values for 450 K and Blue lines represent distance values for 300 K. 1LGY_450_4 shows a distance between the active site Ser and His residues for 1LGY structure at 450 K and 4th simulation, similarly, 1TIB_300_3 shows a distance between the active site Ser and His residues for 1TIB structure at 300 K and 3rd simulation.

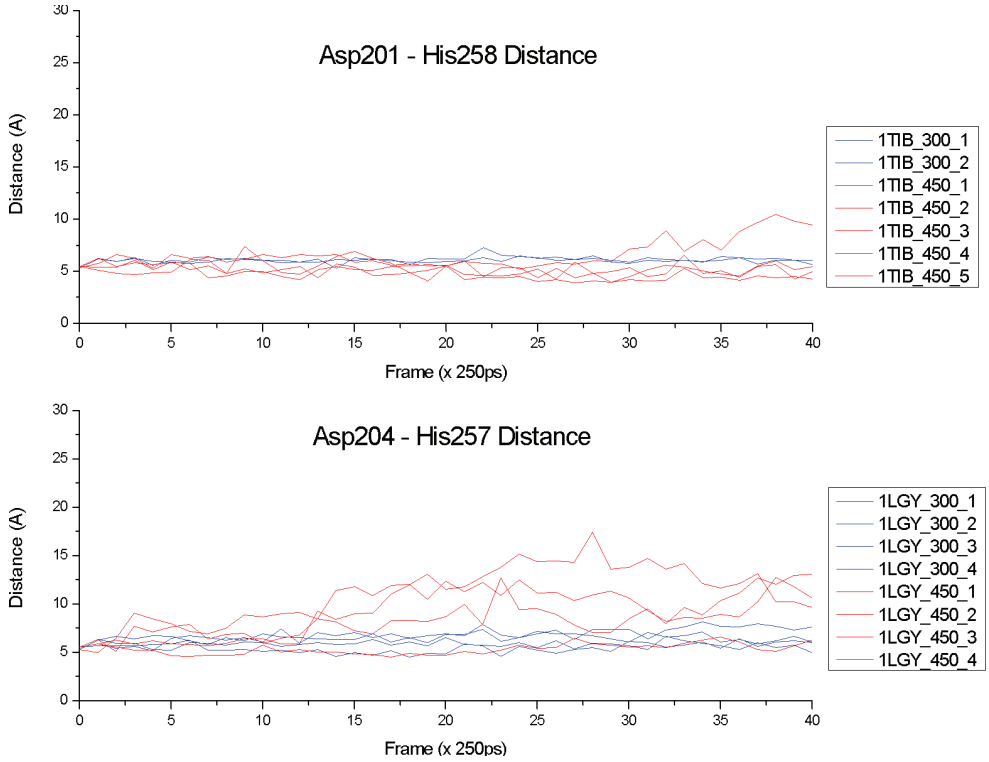


Figure 2.18 Distance between the catalytic Asp and His residues of 1LGY and 1TIB at 300 K and 450 K. Red lines represent distance values for 450 K and Blue lines represent distance values for 300 K. 1LGY_300_3 shows a distance between the active site Asp and His residues for 1LGY structure at 300 K and 3rd simulation, similarly, 1TIB_450_5 shows a distance between the active site Asp and His residues for 1TIB structure at 450 K and 5th simulation.

2.5.4.3 Active Site Analysis for mutant structures

The effect of mutations on the catalytic residue distances during simulations at different temperatures was analyzed. As, Ser – Asp distance was not affected during simulation at 300 K and 450 K, the same results for mutations P183G, P195G, P207G and P219G was observed (Figure 2.19). As seen in Figure 2.20, some mutations had positive impact on stability by decreasing the distance of catalytic residues. Detailed analysis of the effects of mutations on Ser – His distance is given in Figure 2.21. Nearly all of the runs for P195G and most of the P207G runs were below the 1LGY runs at 450 K. This finding can be interpreted as P195G and P207G mutations have increased the thermostability. Although I have one run for P219G, I can also comment that P219G mutation has positive impact on stability. On the other hand, P183G mutation had

similar trend with 1LGY, so that it may or may not increase the stability. Conclusion from active site residue distance for P183G was inconclusive. Similar results and conclusions were also valid for Asp – His distance (Figure 2.22).

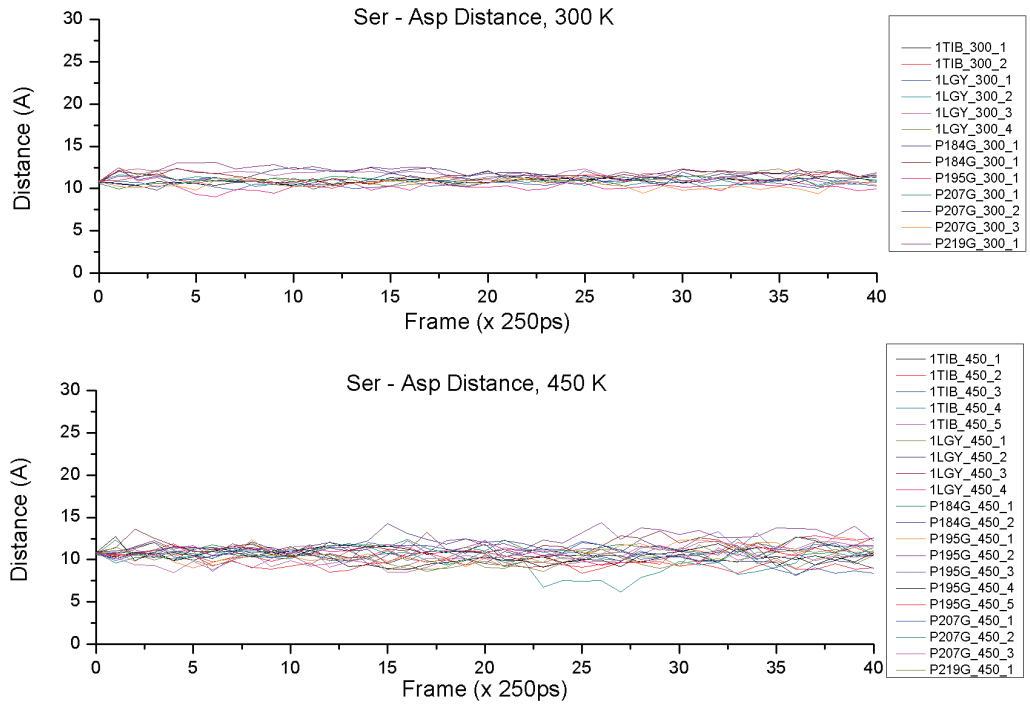


Figure 2.19 Ser – Asp distance over trajectory. All the mutations have been analyzed. P195G_300_1 shows a distance between the active site Ser and Asp residues for P195G mutant structure at 300 K and 1st simulation; similarly, P207G_450_3 shows a distance between the active site Ser and Asp residues for P207G mutant structure at 450 K and 3rd simulation.

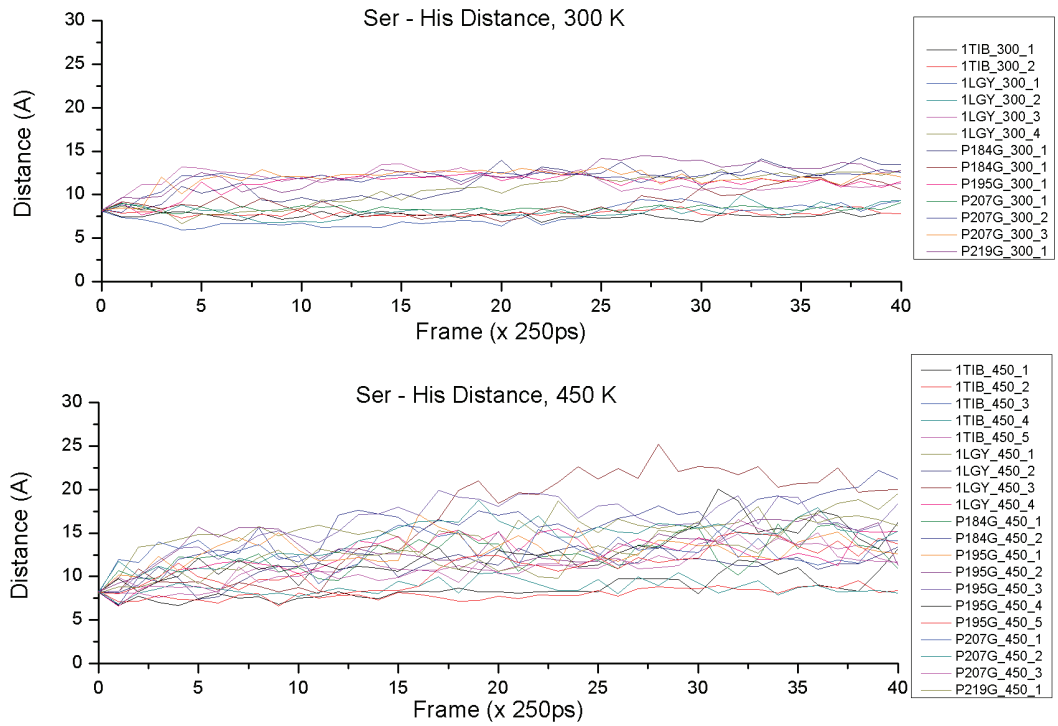


Figure 2.20 Ser – His distance over trajectory. All the mutations have been analyzed. P195G_300_1 shows a distance between the active site Ser and His residues for P195G mutant structure at 300 K and 1st simulation; similarly, P207G_450_3 shows a distance between the active site Ser and His residues for P207G mutant structure at 450 K and 3rd simulation.

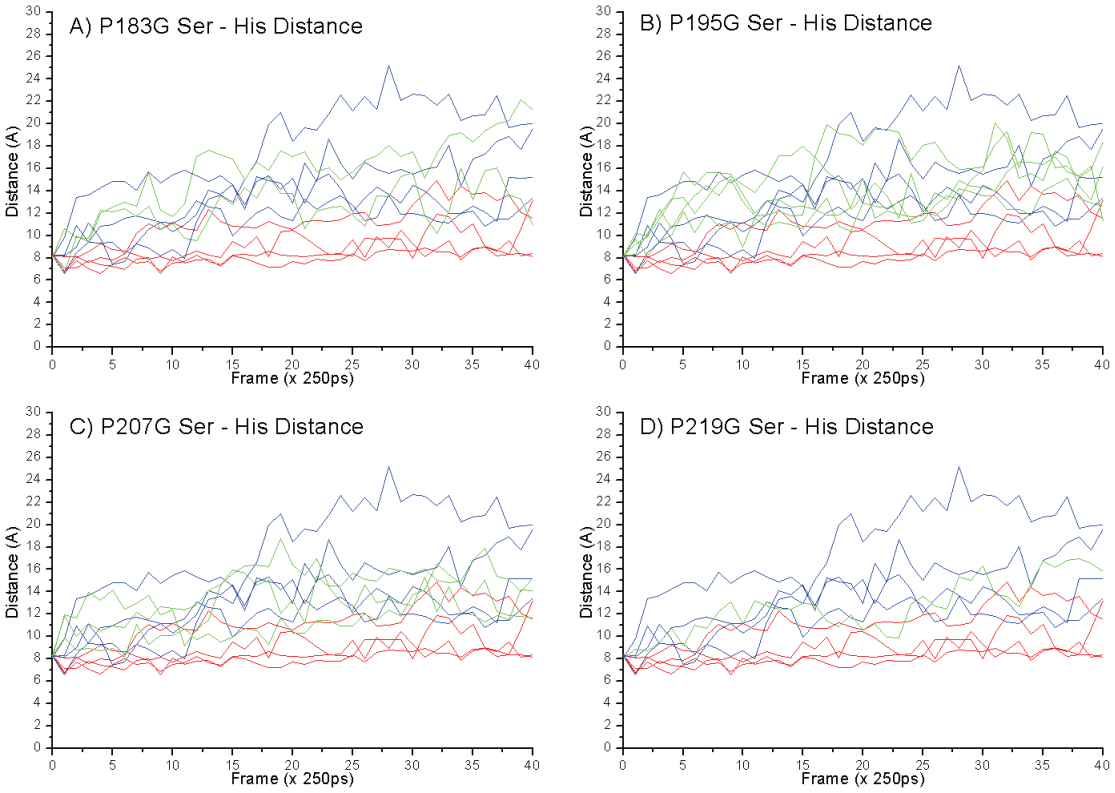


Figure 2.21 Ser – His distance for the mutations. Blue line represents 1LGY at 450 K, Red line represents 1TIB at 450 K and Green lines represents mutations A) P183G, B) P195G, C) P207G, and D) P219G at 450 K. Each line for the same color represents different simulations at given temperature.

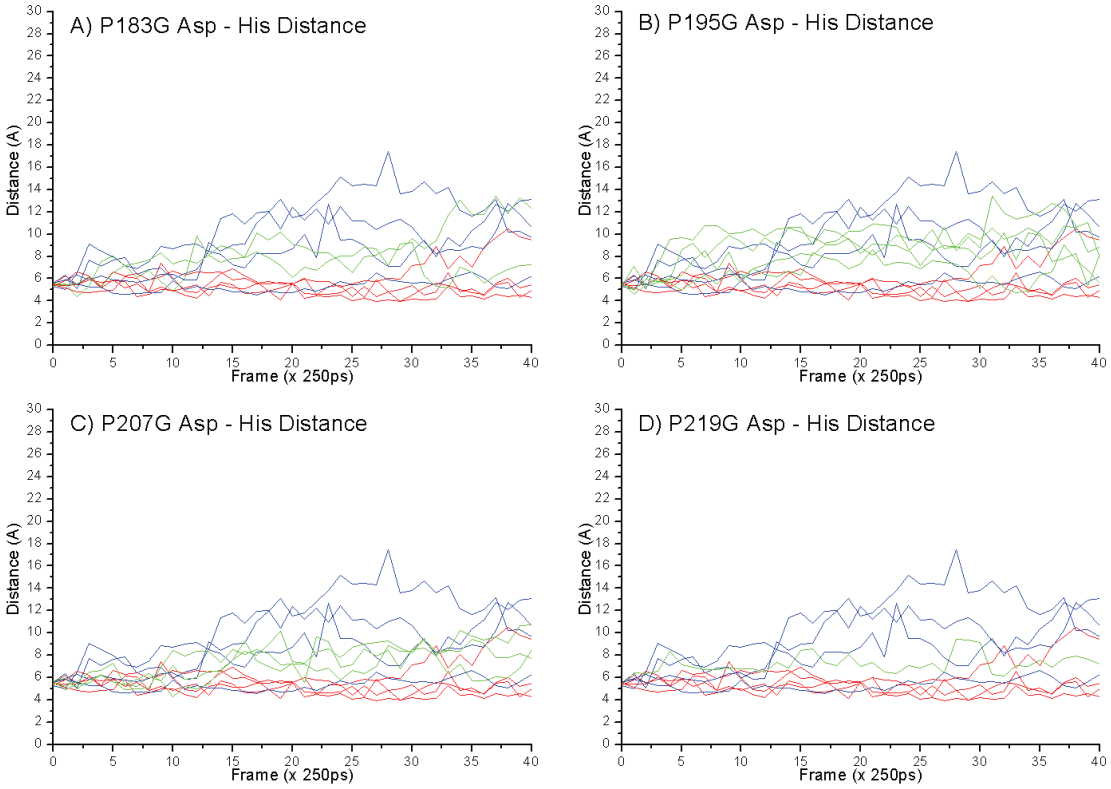


Figure 2.22 Asp – His distance for the mutations. Blue line represents 1LGY at 450 K, Red line represents 1TIB at 450 K and Green lines represents mutations A) P183G, B) P195G, C) P207G, and D) P219G at 450 K. Each line for the same color represents different simulations at given temperature.

2.5.4.4 Radius of gyration

The radius of gyration (RGYR) was calculated using the measure function in the TCL/TK command window within VMD. The formula used to calculate the radius of gyration is given by Equation (2.1):

$$rgyr = \left(\sum_i m(i)(r(i) - \bar{r})^2 \right) \Big/ \left(\sum_i m(i) \right) \tag{2.1}$$

where the sums run over the selected group of atoms, $m(i)$ is the mass of atom i, $r(i)$ is the position of atoms i and $\bar{r}$ is the geometric center of the selected group of atoms.

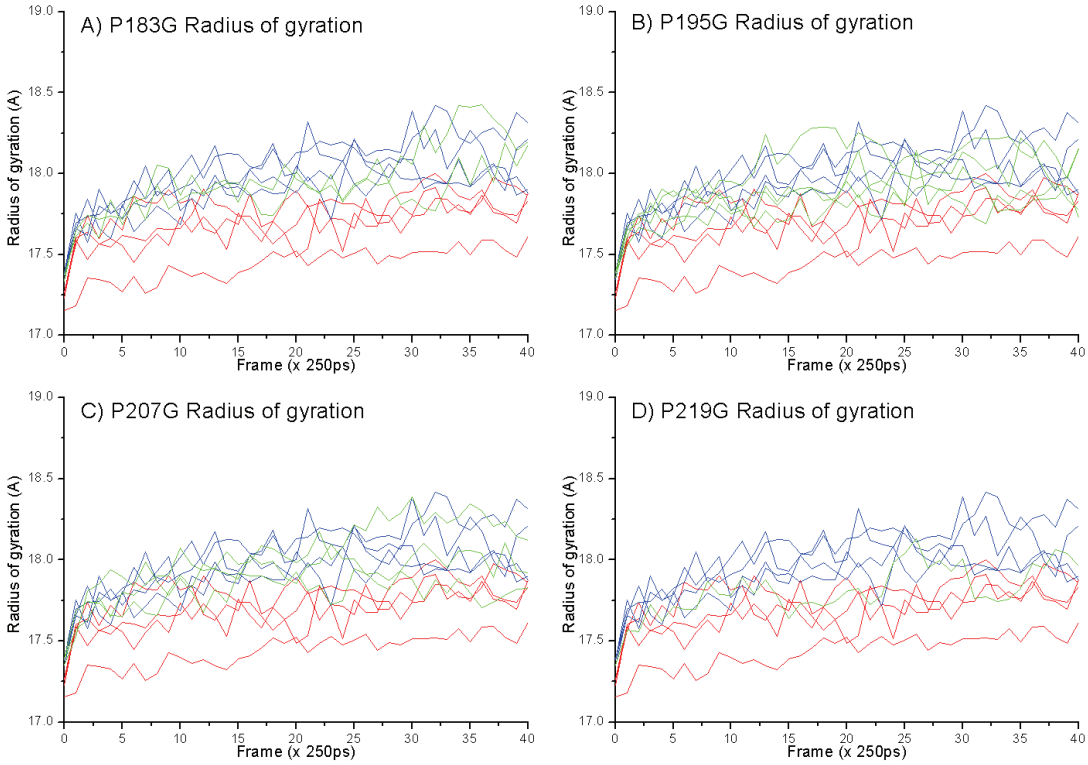


Figure 2.23 Radius of gyration analysis of the mutant structures. Blue line represents 1LGY at 450 K, Red line represents 1TIB at 450 K and Green lines represents mutations A) P183G, B) P195G, C) P207G, and D) P219G at 450 K. Each line for the same color represents different simulations at given temperature.

2.5.4.5 Root Mean Square Deviation

The root mean square deviation (RMSD) is a widely used measure of distance between two structures. In molecular dynamics simulations, RMSD is the measure of root mean square deviation of the positions of the backbone atoms over the course of the dynamics simulation, compared to the initial (equilibrated) structure. RMSD calculation gives an idea about the conformational stability of the protein during simulation.

The formula used to calculate RMSD is given by Equation (2.2):

$$RMSD(N;x,y)=\left[\frac{\sum_{i=1}^{N} w_i \left\|x_i - y_i\right\|^2}{N\sum_{i=1}^{N} w_i}\right]^{\frac{1}{2}} \qquad (2.2)$$

where, N is the atom positions from structure x and the corresponding structure y , $w(i)$ is the weighting vector.

In case of molecular dynamics simulations, each given structure at defined simulation time should be transformed in such a way that RMSD values between given structure and reference structure should be minimized by moving one structure onto another.

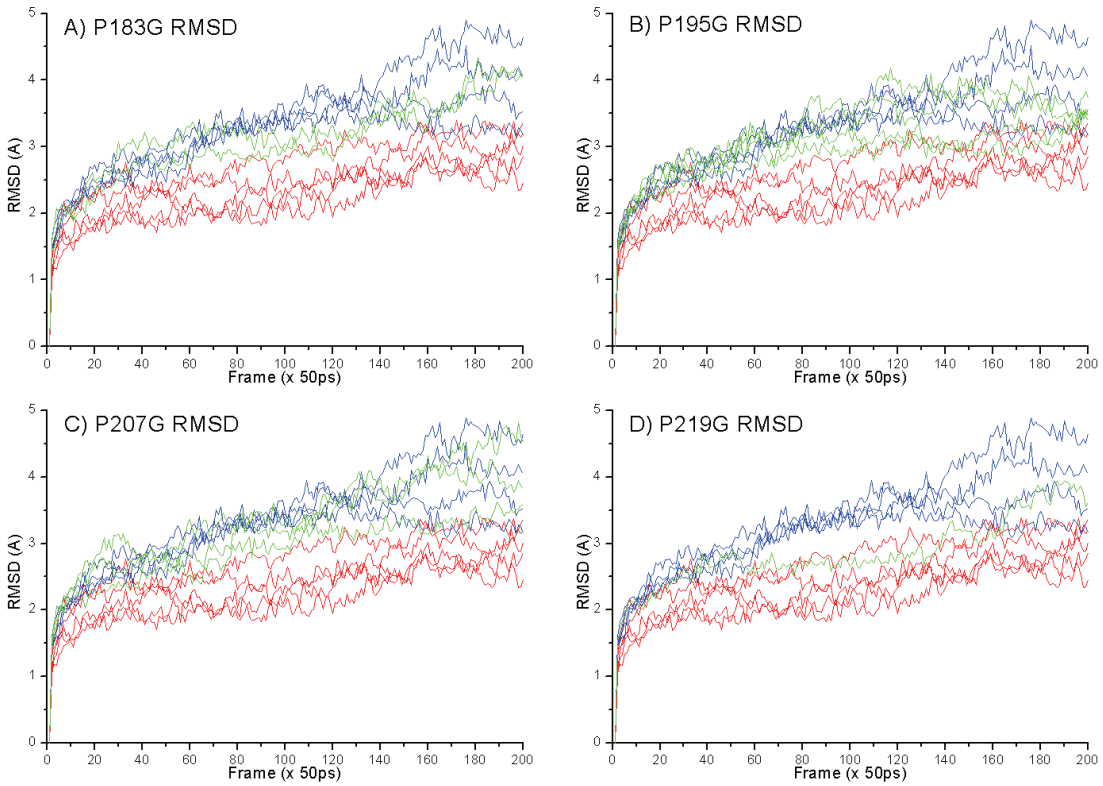


Figure 2.24 RMSD calculation over trajectory. Blue line represents 1LGY at 450 K, Red line represents 1TIB at 450 K and Green lines represents mutations A) P183G, B) P195G, C) P207G, and D) P219G at 450 K.

For P183G mutation, RMSD measure was similar to the native structure (1LGY) and showed an increase after 7.5 ns as native structure did (Figure 2.24 – A). This suggests that changing Proline to Glycine at position 183 is not an improvement in terms of thermostability of the protein. Similar observations obtained for P219G

mutation. RMSD values for the mutant P219G increased after 7.5 ns, showing conformational instability (Figure 2.24 – D). P207G mutation always followed the same pattern of conformational stability as native structure and showed an increase after 8 ns suggesting a conformational instability (Figure 2.24 – C). In conclusion, P183G, P207G and P219G mutant structures did not show any stability improving properties when I analyzed the RMSD values.

On the other hand, RMSD analyses showed P195G mutation minimized the RMSD especially at the end of the runs (Figure 2.24 – B). This finding clearly suggests that P195G mutation has beneficial in terms of lipase thermostability. RMSD values can be interpreted as the measure of the thermostability in our case.

2.6 Conclusion

Prediction of the mutations beneficial for enzyme improvements for protein engineering using existing sequence and structural based knowledge about proteins can be valuable tool for developing more industrially suitable enzymes.

Different point mutations were analyzed in terms of their impact on flexibility of the catalytic residues. Mutations that increase the flexibility around the catalytic residues and do not radically decrease the flexibility of any other domain in the overall structure were selected as a candidate sites for MD simulations. These mutations were P183G, P195G, P207G, and P219G. We have analyzed the active site residue distances and found that 1LGY which is moderately stable and 1TIB which is more stable behaved differently under simulation conditions. At high temperature (450 K), catalytic triad residues have kept the distance for stable protein (1TIB) while homologous residues of moderately stable protein (1LGY) have lost their distances. We have observed similar stability behavior for P195G, P207G and P219G mutants at 450 K. On the other hand for P183G mutant showed a trend similar to the 1LGY, native moderately stable structure.

The radius of gyration measurement was another parameter towards finding beneficial mutations in terms of thermostability. The results obtained from the radius of gyration analysis showed a trend similar to the active site distance measurements. Thermostable protein showed more compactness during simulation compared to the less stable native protein at 450 K simulation temperature. All of the mutated proteins

behaved similarly according to the radius of gyration measure. Although we could not comment on the stability effects of the mutations clearly, none of them showed a negative impact on thermostability. It is difficult to translate simulation values and/or behaviors to the melting temperature values since the molecular dynamics simulations were carried out at elevated temperatures in nanosecond scale. On the other hand, comparing the active site residue distances, root mean squared deviations, and radius of gyration values, we can predict the effect of mutations on protein thermal stability.

2.7 References

Adam, W., P. Groer, H. U. Humpf and C. R. Saha-Moller. "Synthesis of optically active alpha-methylene beta-lactams through lipase-catalyzed kinetic resolution."(2000) J Org Chem **65**(16): 4919-4922.

Andersen, H. C. "Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations."(1983) Journal of Computational Physics **52**(1): 24-34.

Anderson, D. E., W. J. Becktel and F. W. Dahlquist. "pH-Induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme."(1990) Biochemistry **29**(9): 2403-2408.

Baks, T., M. E. Bruins, A. E. M. Janssen and R. M. Boom. "Effect of Pressure and Temperature on the Gelatinization of Starch at Various Starch Concentrations."(2007) Biomacromolecules **9**(1): 296-304.

Baysal, C. and A. R. Atilgan. "Relaxation Kinetics and the Glassiness of Native Proteins: Coupling of Timescales."(2005) **88**(3): 1570-1576.

Beer, H. D., G. Wohlfahrt, R. D. Schmid and J. E. McCarthy. "The folding and activity of the extracellular lipase of Rhizopus oryzae are modulated by a prosequence."(1996) Biochem J **319 (Pt 2)**: 351-359.

Bohm, G. and R. Jaenicke. "Relevance of sequence statistics for the properties of extremophilic proteins."(1994) Int J Pept Protein Res **43**(1): 97-106.

Brzozowski, A. M., U. Derewenda, Z. S. Derewenda, G. G. Dodson, D. M. Lawson, J. P. Turkenburg, F. Bjorkling, B. Huge-Jensen, S. A. Patkar and L. Thim. "A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex."(1991) Nature **351**(6326): 491-494.

Chien, A., D. B. Edgar and J. M. Trela. "Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus."(1976) J. Bacteriol. **127**(3): 1550-1557.

Daggett, V. and M. Levitt. "Protein unfolding pathways explored through molecular dynamics simulations."(1993) J Mol Biol **232**(2): 600-619.

Dahlke, K. "An Enzymatic Process for the Physical Refining of Seed Oils."(1998) Chemical Engineering & Technology **21**(3): 278-281.

Derewenda, U., L. Swenson, Y. Wei, R. Green, P. M. Kobos, R. Joerger, M. J. Haas and Z. S. Derewenda. "Conformational lability of lipases observed in the absence of an oil-water interface: crystallographic studies of enzymes from the fungi Humicola lanuginosa and Rhizopus delemar."(1994) J Lipid Res **35**(3): 524-534.

Dill, K. A. "Dominant forces in protein folding."(1990) Biochemistry **29**(31): 7133-7155.

Fjerbaek, L., K. V. Christensen and B. Norddahl. "A review of the current state of biodiesel production using enzymatic transesterification."(2009) Biotechnology and Bioengineering **102**(5): 1298-1315.

Gandhi, N. N., S. B. Sawant and J. B. Joshi. "Studies on the lipozyme-catalyzed synthesis of butyl laurate."(1995) Biotechnology and Bioengineering **46**(1): 1-12.

Guzman-Maldonado, H. and O. Paredes-Lopez. "Amylolytic enzymes and products derived from starch: a review."(1995) Crit Rev Food Sci Nutr **35**(5): 373-403.

Hemberg, M., S. N. Yaliraki and M. Barahona. "Stochastic Kinetics of Viral Capsid Assembly Based on Detailed Protein Structures."(2006) **90**(9): 3029-3042.

Henikoff, S. and J. G. Henikoff. "Amino acid substitution matrices from protein blocks."(1992) Proc Natl Acad Sci U S A **89**(22): 10915-10919.

Humphrey, W., A. Dalke and K. Schulten. "VMD: visual molecular dynamics."(1996) J Mol Graph **14**(1): 33-38, 27-38.

Jacobs, D. J., A. J. Rader, L. A. Kuhn and M. F. Thorpe. "Protein flexibility predictions using graph theory."(2001) Proteins **44**(2): 150-165.

Jaeger, K.-E., S. Ransac, B. W. Dijkstra, C. Colson, M. Heuvel and O. Misset. "Bacterial lipases."(1994) FEMS Microbiology Reviews **15**(1): 29-63.

Karplus, M. and J. A. McCammon. "Molecular dynamics simulations of biomolecules."(2002) Nat Struct Mol Biol **9**(9): 646-652.

Kohno, M., J. Funatsu, B. Mikami, W. Kugimiya, T. Matsuo and Y. Morita. "The crystal structure of lipase II from Rhizopus niveus at 2.2 A resolution."(1996) J Biochem **120**(3): 505-510.

Kulkarni, N. and R. V. Gadre. "A novel alkaline, thermostable, protease-free lipase from Pseudomonas sp."(1999) Biotechnology Letters **21**: 897-899.

Matsui, I., Y. Sakai, E. Matsui, H. Kikuchi, Y. Kawarabayasi and K. Honda. "Novel substrate specificity of a membrane-bound [beta]-glycosidase from the hyperthermophilic archaeon Pyrococcus horikoshii."(2000) FEBS Letters **467**(2-3): 195-200.

Matsumura, M., G. Signor and B. W. Matthews. "Substantial increase of protein stability by multiple disulphide bonds."(1989) Nature **342**(6247): 291-293.

Matthews, B. W., H. Nicholson and W. J. Becktel. "Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding."(1987) Proceedings of the National Academy of Sciences of the United States of America **84**(19): 6663-6667.

Minning, S., C. Schmidt-Dannert and R. D. Schmid. "Functional expression of Rhizopus oryzae lipase in Pichia pastoris: high-level production and some properties."(1998) Journal of Biotechnology **66**(2-3): 147-156.

Mozersky, S. M., O. Dale Allen and W. N. Marmer. "Vigorous proteolysis: Reliming in the presence of an alkaline protease and bating (Post-Liming) with an extremophile protease."(2002) Journal of the American Leather Chemists Association **97**(4): 150-155.

Mozhaev, V. V. "Mechanism-based strategies for protein thermostabilization."(1993) Trends Biotechnol **11**(3): 88-95.

Mullis, K. B. and F. A. Faloona (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology. W. Ray, Academic Press. **Volume 155:** 335-350.

Ogasahara, K., E. A. Lapshina, M. Sakai, Y. Izu, S. Tsunasawa, I. Kato and K. Yutani. "Electrostatic Stabilization in Methionine Aminopeptidase from Hyperthermophile Pyrococcus furiosus†."(1998) Biochemistry **37**(17): 5939-5946.

Ollis, D. L., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag and et al. "The alpha/beta hydrolase fold."(1992) Protein Eng **5**(3): 197-211.

Parthasarathy, S. and M. R. N. Murthy. "Protein thermal stability: insights from atomic displacement parameters (B values)."(2000) Protein Eng. **13**(1): 9-13.

Pedone, E. M., S. Bartolucci, M. Rossi and M. Saviano. "Computational analysis of the thermal stability in thioredoxins: a molecular dynamics approach."(1998) J Biomol Struct Dyn **16**(2): 437-446.

Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten. "Scalable molecular dynamics with NAMD."(2005) Journal of Computational Chemistry **26**(16): 1781-1802.

Qian, Z., J. R. Horton, X. Cheng and S. Lutz. "Structural redesign of lipase B from Candida antarctica by circular permutation and incremental truncation."(2009) J Mol Biol **393**(1): 191-201.

Rader, A. J., B. M. Hespenheide, L. A. Kuhn and M. F. Thorpe. "Protein unfolding: rigidity lost."(2002) Proc Natl Acad Sci U S A **99**(6): 3540-3545.

Rao, M. B., A. M. Tanksale, M. S. Ghatge and V. V. Deshpande. "Molecular and biotechnological aspects of microbial proteases."(1998) Microbiol Mol Biol Rev **62**(3): 597-635.

Rathi, P., S. Bradoo, R. K. Saxena and R. Gupta. "A hyper-thermostable, alkaline lipase from Pseudomonas sp. with the property of thermal activation."(2000) Biotechnology Letters **22**(6): 495-498.

Schofer, S. H., N. Kaftzik, U. Kragl and P. Wasserscheid. "Enzyme catalysis in ionic liquids: lipase catalysed kinetic resolution of 1-phenylethanol with improved enantioselectivity."(2001) Chemical Communications(5): 425-426.

Schrag, J. D. and M. Cygler. "1.8 A refined structure of the lipase from Geotrichum candidum."(1993) J Mol Biol **230**(2): 575-591.

Sharma, R., S. K. Soni, R. M. Vohra, L. K. Gupta and J. K. Gupta. "Purification and characterisation of a thermostable alkaline lipase from a new thermophilic Bacillus sp. RSJ-1."(2002) Process Biochemistry **37**(10): 1075-1084.

Shindyalov, I. N. and P. E. Bourne. "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path."(1998) Protein Eng **11**(9): 739-747.

Shirley, B. A., P. Stanssens, U. Hahn and C. N. Pace. "Contribution of hydrogen bonding to the conformational stability of ribonuclease T1."(1992) Biochemistry **31**(3): 725-732.

Sriprapundh, D., C. Vieille and J. G. Zeikus. "Molecular determinants of xylose isomerase thermal stability and activity: analysis of thermozymes by site-directed mutagenesis."(2000) Protein Eng. **13**(4): 259-265.

van der Maarel, M. J. E. C., B. van der Veen, J. C. M. Uitdehaag, H. Leemhuis and L. Dijkhuizen. "Properties and applications of starch-converting enzymes of the [alpha]-amylase family."(2002) Journal of Biotechnology **94**(2): 137-155.

Vieille, C. and G. J. Zeikus. "Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability."(2001) Microbiol. Mol. Biol. Rev. **65**(1): 1-43.

Volkin, D. B. and A. M. Klibanov. "Thermal destruction processes in proteins involving cystine residues."(1987) Journal of Biological Chemistry **262**(7): 2945-2950.

Wu, X. Y., S. Jääskeläinen and Y.-Y. Linko. "An investigation of crude lipases for hydrolysis, esterification, and transesterification."(1996) Enzyme and Microbial Technology **19**(3): 226-231.

Chapter 3

3 CLONING, EXPRESSION AND CHARACTERIZATION OF MUTANT ENZYMES

3.1 Introduction

In this chapter, the results of the expression and the analysis of the mutant enzymes resulting from the rational design studies explained in Chapter 2 are presented. Molecular modeling studies showed that P195G, P207G, and P219G had slightly increased thermal properties. On the other hand the P183G mutation showed similar responses compared to native structure.

P. pastoris was used as an expression system for mutant enzymes. Secretion of expressed proteins performed by the help of the α-leader signal sequence, and functional lipase expressions from the KM71H *P. pastoris* cells for all the mutations predicted for this thesis work have been achieved.

I analyzed the temperature stability of these mutant proteins (Native ROL, and mutated proteins P183G, P195G, P207G and P219G) by performing the residual activity assays and the results of these experiments are presented in this chapter. Although my ultimate aim was to purify all proteins near homogeneity and investigate the structural features that were related to mutations predicted in Chapter 2, I could not achieve this goal. I have tried to purify expressed and secreted ROL mutants in functional form using anion exchange chromatography. Unfortunately, I have not been able to purify these proteins to a desired homogeneity level. Then I have decided to add 6XHis tag to the C-terminal of the proteins since it has been reported that N-terminal tags somehow negatively affects the thermostability of ROL protein (Di Lorenzo et al., 2005). In that case we should have observed any his-tag on the expressed proteins after western blot analysis or nickel column purification yet we could not. I have also used *E. coli* expression system in order to functionally express ROL proteins. Although

previous research on ROL expression in *E. coli* host has showed that it can only be expressed as an inclusion body form (Joerger and Haas, 1993), recently Di Lorenzo et. al reported the functional expression of ROL enzyme in the Origami strain of *E. coli* that has mutations in both the thioredoxin reductase (trxB) and glutathione reductase (gor) genes, which greatly enhances disulphide bond formation in the cytoplasm (Di Lorenzo et al., 2005). Origami strain provides the necessary conditions for disulphide bond formation which might be the most important factor for folding of ROL as the structure of the enzyme holds three different disulphide bonds (Figure 2.4). Nevertheless the previous findings on soluble ROL expression in *E. coli* my expression studies using Origami 2 and Rosetta-gami 2 *E. coli* strains did not yield in soluble ROL production. I have analyzed a range of different temperatures and induction conditions for soluble ROL expression.

3.2 Background

3.2.1 *Pichia pastoris* Expression system

Pichia pastoris expression system, with the commercially available vectors, host strains and well established transformation, selection and expression methodologies, is very useful in the expression of proteins especially those require for post translational modifications such as glycosylation, disulphide bond formation and proteolytic processes (Daly and Hearn, 2005). Furthermore, since *P. pastoris* is a haploid organism, any genetic modification is revealed in the phenotype of the next generations (Cregg et al., 2000). Another advantage of expressing proteins in *P. pastoris* is that heterologously expressed proteins can be secreted into the medium through the use of the secretion signal sequences. The α-mating factor pre-pro leader sequence (α-leader) of *Saccharomyces cerevisiae* can be used to produce proteins with secretion signals in frame to the gene of interest (Cereghino and Cregg, 2000). Secretion of the expressed protein eliminates the requirement of cell disruption and isolation of the expressed protein from cell extracts.

3.2.1.1 Strains

Based on their ability to grow on methanol, *P. pastoris* strains can be divided in three main catagories, Mut^S, Mut^+, Mut^-. The Mut^S strain (*m*ethanol *ut*ilization *s*low) has disrupted AOX1 gene, consequently its capacity to grow on methanol is affected, with a diminished growth rate on methanol and by relying on AOX2 gene. Alternative gene, AOX2, has the same specific activity as AOX1 but its expression depends on weaker promoter (Cereghino and Cregg, 2000; Daly and Hearn, 2005). Some widely used *P. pastoris* expression strains were listed in Table 3–1.

Table 3–1 *P. pastoris* expression host strains (Cregg et al., 2000; Cregg, 2007b)

Strain Name	Genotype	Phenotype	Reference
GS115	*his4*	Mut^+ His^-	(Cregg et al., 1985)
KM71	*aox1Δ::SARG4 his4 arg4*	Mut^s His^-	(Tschopp et al., 1987)
MC100-3	*aox1Δ::SARG4 aox2Δ::Phis4 his4 arg4*	Mut^- His^-	(Cregg et al., 1989)
SMD1168	*pep4Δ his4*	Mut^+ His^-, protease deficient	(Gleeson et al., 1998)
SMD1165	*prb1 his4*	Mut^+ His^-, protease deficient	(Higgins and Cregg, 1998)
SMD1163	*pep4 prb1 his4*	Mut^+ His^-, protease deficient	(Higgins and Cregg, 1998)
X-33			Invitrogen
Y-11430		Wild type	Northern Regional Research Laboratories Peoria, IL

3.2.1.2 Promoters

There are several promoters for protein expression in *P. pastoris*, but the most cited promoter is the AOX (alcohol oxidase) promoter. AOX promoter is the strongest promoter for heterologous protein expression and typically results in more than 30 % of the soluble protein in *P. pastoris* grown with methanol as a carbon source (Cregg et al., 1993). The AOX promoter is tightly repressed when growing with glycerol or glucose, allowing a robust control of the induction of heterologous protein expression. This regulation is similar to GAL1 promoter of S. cerevisiae in which cells are in repressed state when grown in the presence of glucose and induced in the absence of glucose with galactose (Mumberg et al., 1994).

Use of methanol induction protocols may not be suitable for some applications, i.e. food industry because of health and fire hazards of methanol. Alternative promoters for *P. pastoris* system have been developed. One is a strong constitutive promoter, in the presence of glucose, derived from the *P. pastoris* glyceraldehyde-3-phosphate dehydrogenase (GAP) gene. In addition methanol does not require for induction, GAP promoter system does not require any shifting of carbon source (Waterham et al., 1997). However, this promoter can cause a problem during expression of foreign gene products which are toxic to *P. pastoris*.

The *FLD1* gene encodes a glutathione-dependent formaldehyde dehydrogenase, a key enzyme required for the methanol and methylated amine metabolism (Shen et al., 1998). *FLD1*is required for growth of *P. pastoris* on either methanol or certain methylated amines, i.e., methylamine, as a sole nitrogen source. Heterologous expression under control of FLD1 promoter is repressed in glucose containing media but can be induced in media containing methanol or methylamine. The *FLD1* gene promoter yields expression levels similar to the *AOX1* promoter in some cases (Sunga and Cregg, 2004; Cos et al., 2005).

There are also several weakly or moderately expressing promoters available for *P. pastoris.* PEX8 (PER3) (Liu et al., 1995), YPT1 (Sears et al., 1998), TEF1 (Ahn et al., 2007), PGK1 (de Almeida et al., 2005), ICL1 (Menendez et al., 2003).

3.2.1.3 Genomic Integration

Although there are some newly developed episomal vectors (Lee et al., 2005), expression vectors are usually integrated into *P. pastoris* genome to obtain stable expression strains. Transformants are stable in the absence of selective pressure and generally host multi-copy of integrated vectors.

Linearized vectors can generate stable transformants via homologous recombination between vector and host genome shared sequences, this event is similar to the well established S. cerevisiae integration (Vedvick, 1991; Romanos, 1995).

Gene insertion events at the *AOX1* (GS115) or *aox1::ARG4* (KM71) loci arise from a single crossover event between the loci and any of the three *AOX1* regions on the vector: the *AOX1* promoter, the *AOX1* transcription termination region (TT), or the sequences even further downstream of *AOX1* (3′ *AOX1*). This results in the insertion of one or more copies of the vector upstream or downstream of the *AOX1* or the *aox1::ARG4* genes. The phenotype of such a transformant is His^+ Mut^+ in GS115 or His^+ Mut^S in KM71. Linearized recombinant vectors at a restriction site in the 5′*AOX1* regions, can generate Mut^+ or Mut^S recombinants, depending on the host strain used.

The Figure 3.1-A shows the result of an insertion of the plasmid 3′ to the intact AOX1 locus (Mut^+) and the gain of AOX1 promoter, gene of interest, and *HIS4* (expression cassette). This event could also happen at the 5′ AOX1 regions of the plasmid and genome with the resulting insertion positioned 5′ to an intact AOX1 locus. This also occurs with intact vectors, even though at a lower frequency.

In either GS115 (Mut^+) or KM71 (Mut^S), gene insertion events at the *his4* locus arise from a single crossover event between the *his4* locus in the chromosome and the *HIS4* gene on the vector. This results in the insertion of one or more copies of the vector at the *his4* locus. Since the genomic *AOX1* or *aox1::ARG4* loci are not involved in this recombination event, the phenotype of such a His^+ transformant has the same Mut phenotype as the parent strain. By linearizing the recombinant vector at a restriction enzyme site located in *HIS4* gene, Mut^+ or Mut^S recombinants can be conveniently generated depending on the host strain used. In the Figure 3.1-B, insertion of the plasmid between duplicated copies of the *HIS4*/*his4* genes creats two his genes, one still mutant, and the other wild type.

Multiple gene insertion events at a single locus in a cell do occur spontaneously with a low, but detectable frequency--between 1 and 10% of all selected His+ transformants (Clare et al., 1991). Multi-copy events can occur as gene insertions either at the *AOX1*, *aox1::ARG4*, or *his4* loci (Figure 3.1-C). This results in a Mut^{+} phenotype in GS115 and a Mut^{S} phenotype in KM71. Quantitative dot blot analysis, Southern blot analysis, and differential hybridization can detect multiple gene insertion events (Cregg et al., 1985).

In a *his4* strain such as GS115, a gene replacement event arises from a double crossover event between the *AOX1* promoter and 3′ *AOX1* regions of the vector and genome (Figure 3.1-D). This results in the complete removal of the *AOX1* coding region (i.e. gene replacement). The resulting phenotype is His+ MutS. His+ transformants can be readily and easily screened for their Mut phenotype, with Mut^{S} serving as a phenotypic indicator of integration via gene replacement at the AOX1 locus. The net result of this type of gene replacement is a loss of the AOX1 locus (Mut^{S}) and the gain of an expression cassette containing AOX1 promoter, gene of interest, and *HIS4*.

The site and the type (insertion or replacement) of the integration events can be confirmed by southern blot analysis hybridized with a probe generated from the AOX1 promoter region (Paifer et al., 1994) or even simplified PCR based methodologies (Burdychova et al., 2002). Selection of the high producers is important especially when the expression level of heterologous protein is low.

There are generally two ways of generating multicopy transformants. One method involves the construction of a vector with several head-to-tail copies of an expression cassette (Brierley, 1998). Although the method is straightforward, integration of large vectors into the genome of *P. pastoris* is difficult. On the other hand, screening of large number of colonies for the multi-copy integration using DNA hybridization techniques could be helpful to find the high producer cells (Romanos et al., 1991). Similarly, screening for high protein expression using SDS-PAGE, immunoblotting, or colony immunoblotting can be used (Wung and Gascoigne, 1996). Recently, real-time PCR method has been adapted for the identification of copy number of expression cassettes in *P. pastoris* (Mack et al., 2009; Zhu et al., 2009a; Zhu et al., 2009b).

Similar selection strategies for the multi-copy transformants have also been developed using antibiotic resistance gene on the expression cassettes. Currently, the Zeocin and Blasticidin resistance genes are the only dominant selectable markers that

can be used for primary selection of transformants. Multi-copy transformants can be selected on solid medium containing increasing amount of the antibiotics. Antibiotic resistance gene dosage (i.e. G418) is proportional to the expression cassettes number in the genome (Lin-Cereghino et al., 2008). Another recently developed technology uses the ADE2 gene encodes phosphoribosylaminoimidazole carboxylase, which catalyzes the sixth step in the *de novo* biosynthesis of the purine nucleotides (Som et al., 2005). In *Saccharomyces cerevisiae, Pichia pastoris* and other yeast strains, mutations in *ADE2* lead to the accumulation of purine precursors in the vacuole, which causes the colony to be red in color. The pigmentation phenotype can be used as for selection and screening, like in Invitrogen's pPINK system. The color of the colonies indirectly indicates the relative expression levels of protein of interest as the color of the colony depends on the copy number of the plasmid. The pink colonies express very little *ADE2* gene product, while the white colonies express higher amounts of the *ADE2* gene product, suggesting that those colonies have more copies of the integrated construct (Lin Cereghino et al., 2001).

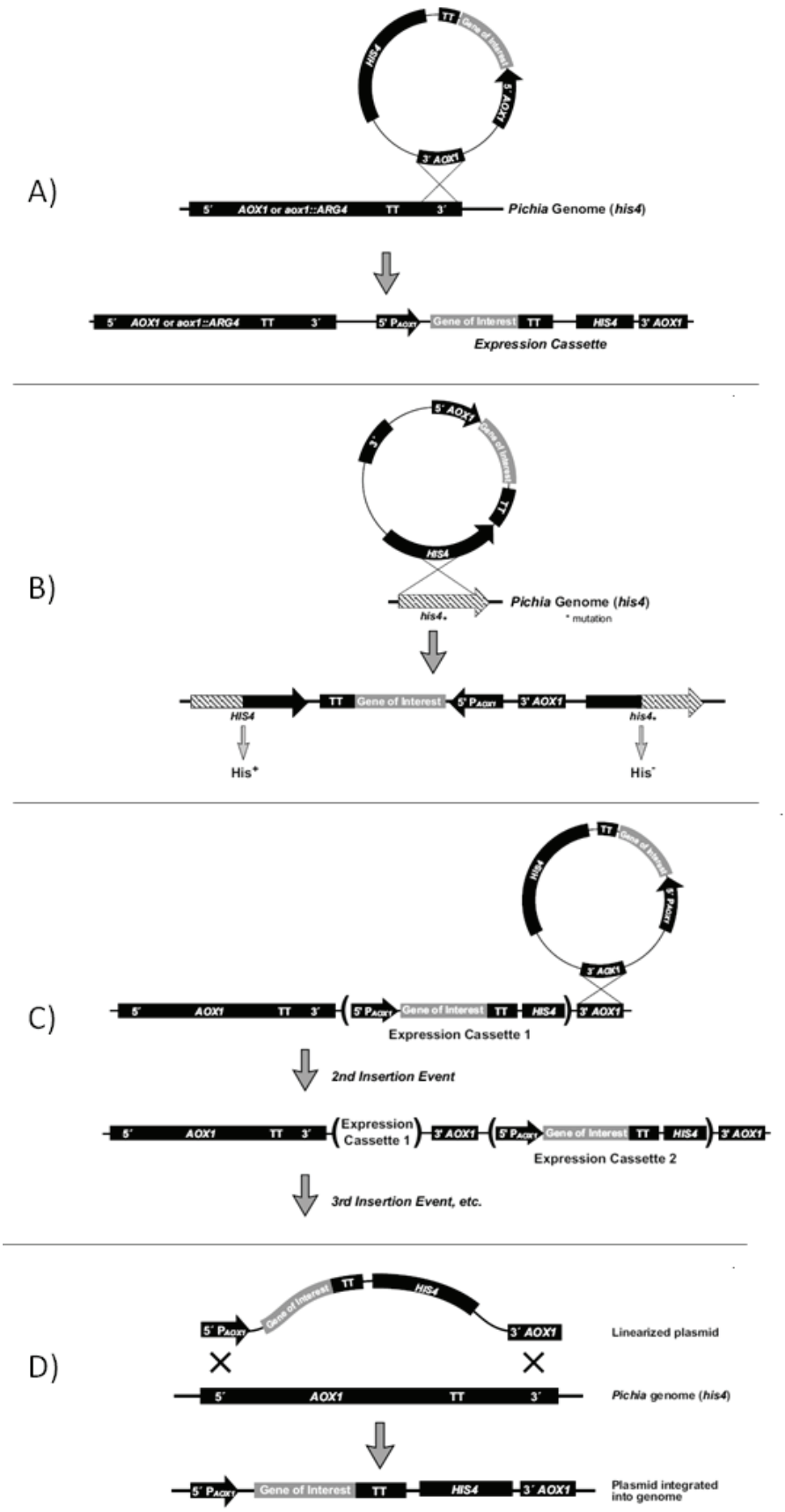


Figure 3.1 Integration of vectors into the *P. pastoris* genome (Adapted from Pichia manual (Invitrogen))

3.2.1.4 Transformation

Molecular basis of the genetic manipulation of any organism lies in the successful introduction of the gene into the host followed with the durable maintenance of the gene of interest. Similar to well-known microbial host systems renowned for their ease of adaptation to genetic manipulations the DNA-shuttling methods for the host *P. pastoris* is fitting to those described for other hosts, mostly *Saccharomyces cerevisiae*. Several studies have been carried out regarding the genetic manipulation of the species *P. pastoris*. The methods developed so far differ in the transforming DNA type, linear or circular. Moreover these methods have been issued for their efficiency and the cost-effectiveness aspects. Thus, in brief, relevant literature has existed for the implementation of the best transformation method for the *P. pastoris* species (Cregg, 2007a). Among the several described methods for the pastoris species, the transformation could be provided with an autonomously replicating DNA element or integration into host genome. Exploiting either one of these two DNA introduction principles 4 different transformation methods have been improved. Namely these methods are spheroplast generation, electroporation, PEG and LiCl mediated transformations. These methods vary in convenience, efficiency and generation of gene copies. The spheroplast method also referred as the spheroplast generation-polyethylene glycol-$CaCl_2$. This method apart from being well-characterized is also known to yield a high transformation frequency (a transformation efficiency of ~$10^5/\mu g$) compared to the rest of the techniques. Yet the mentioned method is said to be labor intensive due to the necessity of the transformed colonies retrieval from agar embeds (Cregg et al., 1985). This method utilizes the spheroplasts generated from the intact cells followed with chemical mediated DNA transformation. The remaining three transformation methods on the other hand do not require any spheroplasting of the cells and thus possess comparably convenience. Nevertheless bypassing the time-consuming spheroplast formation the chemical transformation methods have shown to have very low efficiencies in the previous reports (Invitrogen. 1999. Multi-copy Pichia Expression Kit, Version E. Carlsbad, CA., (Thompson et al., 1998)). Among these three methods harboring the whole-cells processing the electroporation technique yields comparably high transformation frequencies and thus it is the most widely used one (Cregg et al., 1985). The electroporation technique is well characterized in this host system (Wu and

Letchworth, 2004). The method was improved by the pretreatment of cells with Li-Acetate and Dithiothreitol (DTT) to a comparable efficiency of transformation up to 636 ± 208.0 × 10^4. The aforesaid method for transformation is therefore the most preferred one for *P. pastoris* transformation where the required equipments are provided.

3.2.1.5 Protein Secretion

The *S. cerevisiae* alpha-mating factor pre-pro signal sequence (α-MF) and the Glu-Ala repeats are the most widely used in order to secrete heterologously expressed proteins in *P. pastoris* (Daly and Hearn, 2005). The α-MF sequence consists of a pre-sequence signal peptide of 19 amino acids and a 60 hydrophilic amino acid pro-peptide (Brake et al., 1984). The signal sequence is responsible for translocation of the pro-protein into the ER and is subsequently cleaved by signal peptidase during the translocation process. The pro-protein is transported to the Golgi where the pro-region is cleaved by kex2 protease to release the mature protein. Kex2 protease recognizes basic pairs of amino acids such as Lys – Arg or Arg – Arg. After generation of mature protein, further trimming of the N-terminal Glu – Ala repeats occurs in Golgi by Ste13 dipeptidyl aminopeptidase (Anna-Arriola and Herskowitz, 1994).

3.2.2 Purification of ROL

ROL enzyme was heterologously expressed first time in *E. coli* (Beer et al., 1996). Although the enzyme was successfully expressed, the protein was found to be as inclusion body form. Di Lorenzo et al., have achieved functional expression of ROL in *E. coli* Origami strain (Di Lorenzo et al., 2005). On the other hand ROL enzyme was expressed in *P. pastoris* under the control of the AOX1 promoter. Enzyme was secreted into the medium using alpha-mating factor pre-pro signal sequence of *S. cerevisiae* (Minning et al., 1998). Same group has achieved 2.5 fold increase in productivity by introducing a transition phase that involved the simultaneous feeding of glycerol and methanol followed by a single methanol feed (Minning et al., 2001). They have

achieved homogeneity level purification by a simple two-step purification method, containing 10 kDa membrane filtration and cation exchange chromatography.

3.2.3 Enzyme Assays for Activity

Most lipases are water soluble enzymes that hydrolyze ester bonds of water insoluble substrates. A number of protocols have been developed to assay these enzymes. The majority of triacylglycerol lipases act on primary ester bonds such as those in the *sn*-1 and *sn*-3 positions of triacylglycerol (Gilham and Lehner, 2005). A variety of synthetic esters are developed for lipase activity assays. Fluorescent detection method is one of the most popular methods. Reaction products became fluorescent upon hydrolysis. Widely used fluorescent reporter is methylumbelliferyl (methylcoumarin) molecule (Jacks and Kircher, 1967). Different kinds of 4-methylumbelliferyl esters are commercially available (Beisson et al., 2000). This compound becomes highly fluorescent (Ex: 365, Em: 450) after hydrolysis of ester linkage. There are some possible drawbacks of this kind of substrates: (i) the substrates more closely resemble monoacylglycerol rather than triacylglycerol and (ii) it has been reported that these substrates can spontaneously hydrolyze with or without albumin at pH 8.8 and higher (Nyfeler et al., 2003). On the other hand sensitivity and high throughput compatibility of these substrates are important factors for screening studies comparing to the titrimetric methods (Brockman, 1981).

3.3 Materials

Detailed protocols followed during the course of this thesis for cloning, mutagenesis, expression, purification, and enzyme assays, are given in this section.

3.3.1 Instruments and Software

Instruments and software used in this chapter were listed in Appendix A with their catalog and model numbers.

3.3.2 Enzymes

All enzymes used in this chapter are listed in Appendix C with the catalog numbers. Some enzymes were included in kit content; these kits were listed in Appendix D.

3.3.3 Chemicals

Chemicals used in this chapter were listed in Appendix B with their catalog numbers. All chemicals and enzymes were used without any further purification.

3.3.4 Kits

All kits used in this study are listed in Appendix D with the catalog numbers. Manufacturer's protocols were followed using these kits unless otherwise noted in Methods sections.

3.3.5 Vectors and Microorganisms

Microorganisms and vectors used in this thesis are listed in Table 3–2 and Table 3–3. Vector maps are illustrated in Appendix E.

Table 3–2 Microorganisms

Organism	Relevant Genotype / Phenotype	Company
XL-1 Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)]	Stratagene
Top 10	F– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG	Invitrogen
Rosetta-gami 2(DE3)pLysS	Δ(ara–leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F'[lac+ lacI q pro] gor522::Tn10 trxB pLysSRARE2 (CamR, StrR, TetR)	Novagen
Origami 2(DE3)pLysS	Δ(ara–leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F'[lac+ lacI q pro] (DE3) gor522::Tn10 trxB pLysS (CamR, StrR, TetR)	
KM71H	arg4, aox1::ARG4 / Muts, Arg+	Invitrogen

Table 3–3 Vectors

Vector	Relevant Marker	Induction	Company
pPICZalphaA	Zeocin	Methanol	Invitrogen
pET22b(+)	Ampicillin	IPTG	Novagen

3.3.6 Oligonucleotides

All oligonucleotides were designed using VectorNTI software and were purchased from Microsynth (CH) and used without further purification. Oligonucleotides were resuspended in molecular biology grade water at 100 µM concentration and stored at -20 °C. Working solutions were prepared at 10 µM and aliquots were stored at -20 °C.

Table 3–4 Oligonucleotides for DNA amplification, SDM, colony PCR and sequencing

Oligo Name	Usage	Sequence (5' to 3')
F_AOX	PCR	GACTGGTTCCAATTGACAAGC
R_AOX	PCR	GCAAATGGCATTCTGACATCC
F_alpha	PCR	TACTATTGCCAGCAT TGCTGC
F_ROL_EcoRI	Cloning	GCT GAAGCTGAATTCTCTGATGG
R_ROL_XbaI	Cloning	CCGTCTAGAACCAAACAGCTTCCTTCGTTGATAT
R_ROL_Stop	Cloning	TAGAAAGCTGGCGGCCGCTTATT
F_P183G	SDM	GGTCCTCGTGTTGGTAACGGCACCTTTGCTTACTATGT
R_P183G	SDM	ACATAGTAAGCAAAGGTGCCGTTACCAACACGAGGACC
F_P195G	SDM	TATGTTGAATCTACCGGTATTGGTTTCCAACGTACCGTTC AC
R_P195G	SDM	GTGAACGGTACGTTGGAAACCAATACCGGTAGATTCAA CATA
F_P207G	SDM	CCGTTCACAAGAGAGATATCGTTGGTCACGTTCCTCCT
R_P207G	SDM	AGGAGGAACGTGACCAACGATATCTCTCTTGTGAACGG
F_219G	SDM	TCAATCCTTCGGATTCCTTCATGGCGGTGTTGAATCTTGG
R_219G	SDM	CCAAGATTCAACACCGCCATGAAGGAATCCGAAGGATT GA
RT_AOX_F	qPCR	ACATCCACAGGTCCATTC
RT_AOX_F	qPCR	GGTGTTAGTAGCCTAATAGAAG
RT_ARG_R	qPCR	GTCTATTAAGTCCATTCCGTCAAC
RT_ARG_R	qPCR	TAGCCAGCATATCCATAGTTAGAG
F_ROLpET22	In-fusion	TAAGAAGGAGATATACATATGTCTGATGGTGGTAAGG
R_ROLpET22	In-fusion	GTGGTGGTGGTGGTGCTCGAGCAAAC AGCTTCCTTCGTTGAT

3.3.7 PCR conditions

PCR conditions used throughout this thesis are given in Table 3–5, Perkin-Elmer 9700 model was used for all PCR amplifications.

Table 3–5 PCR conditions for DNA amplification

***P. pastoris* Colony PCR**				
Steps				
1	95 °C, 5 min	-	-	1
2	95 °C, 30 sec	53 °C, 30 sec	72 °C, 90 sec	32
3	-	-	72 °C, 7 min	1
***E. coli* Colony PCR**				
Steps				
1	95 °C, 10 min	-	-	1
2	95 °C, 30 sec	53 °C, 30 sec	72 °C, 90 sec	32
3	-	-	72 °C, 7 min	1
ROL PCR for Cloning				
Steps				
1	95 °C, 5 min	-	-	1
2	95 °C, 30 sec	55 °C, 30 sec	72 °C, 60 sec	34
3	-	-	72 °C, 7 min	1
Mutagenesis PCR				
Steps				
1	95 °C, 10 min	-	-	1
2	95 °C, 30 sec	55 °C, 30 sec	72 °C, 90 sec	32
3	-	-	72 °C, 7 min	1

3.3.8 Lipase Substrates

Table 3–6 Substrates for Lipase Activity Assays

Substrate		
4MU-Butyrate	Sigma	4 Carbon
4MU-Caprylate	Sigma	6 Carbon
4MU-Capric	Sigma	
4MU-Oleate	Sigma	

3.4 Methods

3.4.1 Plasmid Isolation

6 ml of the overnight culture of bacteria in Terrific Broth carrying the desired plasmid was centrifuged (13200 rpm, 10 min) and the pellet treated according to the protocol in the QIA Prep Miniprep kit (QIAGEN, Hilden, Germany). The purified plasmid DNA was eluted in 52 µl of elution buffer as indicated by the manufacturer's instructions. For applications such as RE digestions for analytical purposes we have used alkaline lysis miniprep method. Briefly, the cells were collected from 5 ml of overnight LB culture by centrifugation for 1 min at max rpm. The cell pellet was resuspended in 200 µl of solution 1 (100 mM Tris-HCl, pH 7.5, 10 mM EDTA, 400 µg/ml RNase I); the cells were then lysed by addition of 200 µl solution 2 (1M NaOH, 5 % SDS) followed by mild mixing. After adding 200 µl of solution 3 (Potassium Acetate, 5 M) and inverting the tube several times, the proteins and the genomic DNA were precipitated and collected by centrifugation (5 min, max rpm, RT). The plasmid DNA in the supernatant, was precipitated with isopropanol, and then resuspended in 20 µl EB (Qiagen).

3.4.2 Electrophoresis of DNA

The agarose gel electrophoresis was carried out according to the protocol in Sambrook et al. (Sambrook and Russell, 2001). DNA fragments were separated using 1 % agarose gels. The buffer TBE was used as electrophoresis buffer, and 6X DNA loaing dye was used (glycerine 30% (m/v), bromophenol blue (BPB) 0,2% (m/v), EDTA 25 mM, pH 7.5). The voltage applied was 120 V for 1% gels. The DNA was visualized by ethidium bromide staining, and the bands on the gels were photographed under an UV-lamp.

3.4.3 Gel Extraction

The extraction of the DNA from the agarose gel was carried out according to the instructions of the manufacturer (QiaQuick Gel Extraction Kit, Qiagen). The method is based on the specific binding of the DNA on a silica gel matrix. Briefly, part of gel containing desired fragment was sliced out as small as possible, and dissolved in QG buffer at 55 °C, and directly applied onto binding column, after washing step, DNA was eluted from column using EB buffer.

3.4.4 Ethanol Precipitation of DNA

1 vol. of DNA solution was mixed with 1/10 vol. of 3M Sodium Acetate, pH 5.2, and 2.5 vol. of absolute Ethanol and kept at -20 °C for 20 min for efficient DNA precipitation. After centrifugation at max speed for 15 min, supernatant was discarded and pellet was washed with 70 % Ethanol and centrifuged at max speed for 5 min. Pellet was dried in 60 °C oven for 5 min, and dissolved in EB buffer (Qiagen).

3.4.5 Cloning

Restriction Enzyme Based Cloning

The DNA digestions with restriction enzymes were performed under the reaction conditions specific for every enzyme (Section 3.3.2), as suggested by the

manufacturer's protocol. To avoid self-ligation of vectors cut only with an enzyme, the 5'-phosphate groups were eliminated by treatment with alkaline phosphatase. After purification of the digested DNA fragment and the vector backbone, using either Qiaquick Gel Extraction kit or Qiaquick PCR purification kit (Section 3.3.4), DNAs were analyzed semi quantitatively using Gel documentation system (Section 3.3.1), and the amount of fragment and backbone for ligation reaction were determined. The fragments and the vector backbones were generally ethanol precipitated and dissolved in minimum volume (~ 6 ul) to increase the DNA concentrations.

Ligation reactions were carried out in 10 ul volume by following the Fermentas Quick Ligation kit Protocol except that the ligation reactions were incubated for 2 hours at room temperature.

"In-fusion" Based Cloning

Mutated ROL genes were cloned into pET22b(+) vector using "In-fusion Advantage PCR cloning Kit" following the manufacturer's suggested protocol. Primers for in-fusion reaction were designed using kit manufacturer's guidelines and provided web based primer design tool (http://bioinfo.clontech.com/infusion/). Briefly, ROL mutants (Section 3.4.8 and Section 3.5.2) were amplified using F_ROL_pET22 and R_ROL_pET22 primers using "in-fusion PCR" PCR protocol (Table 3–5). pET22b(+) vector was digested with *NdeI* and *XhoI* restriction enzymes and gel extracted using protocol given in Section 3.4.3. 10 µl reaction mixtures were prepared according to "spin-column purified PCR inserts protocol". Reaction mixtures were incubated for 15 min at 37 °C, followed by 15 min at 50 °C using GeneAmp PCR system 9700. After the completion of reaction, the mixtures were 10 fold diluted using 1X TE buffer and 5 µl of reaction was transformed into Top10 *E. coli* competent cells using protocol given in Section 3.4.6.

3.4.6 *E. coli* Transformation

XL-1 Blue and Top10 *E. coli* competent cells were prepared according to literature (Nakata et al., 1997; Sambrook and Russell, 2006) with slight modifications. Cells were grown to OD600 0.6 in LB medium, and collected using centrifuge at 11000rpm at 4C for 10 min. Cells were resuspended in transformation buffer (TB; 55

mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl, 10 mM PIPES) and stored on ice for 30 min. Cells were collected by centrifugation and resuspended again in TB-DMSO (7 % DMSO in TB). 200 microliter aliquots were prepared and immediately put in liquid nitrogen and stored at -80 °C until use.

Generally, all 10ul of ligation reaction was mixed with 200 ul of competent cells and 42 °C heat shock was applied after 30min of incubation period on ice. After one hour shaking at 200 rpm, 37 °C, cells were spreaded onto LB plate containing appropriate antibiotics using 3mm glass beads (Worthington et al., 2001).

3.4.7 Glycerol Stocks

Overnight cultures of *E. coli* and *P. pastoris* cells were mixed with 30 % glycerol under sterile conditions in screw cap tubes. After brief vortexing, tubes were stored at -80 °C until use.

3.4.8 Site Directed Mutagenesis

QuickChange Site Directed Mutagenesis kit (Stratagene) was used to generate mutant constructs. Briefly, using appropriate primer pairs (listed in Table 3–4, usage labeled as "SDM"), PCR reaction was performed using 10ng of pPICZalphaA-ROL plasmid as the template. Detailed PCR conditions were given in Section 3.3.7, "mutagenesis PCR" .After the PCR reaction, the products were treated with DpnI enzyme to remove parental plasmids. DpnI treated samples were extracted using Ethanol precipitation (Section 3.4.4), and transformed (Section 3.4.6) into XL-1 Blue chemical competent cells. 3 – 5 colonies from each transformation were sequenced (Section 3.4.9).

3.4.9 Sequencing

Constructs were sequenced by MCLab (Molecular Cloning Laboratories, US) using F_AOX and R_AOX primers. Chromatograms were analyzed using FinchTV software and aligned using VectorNTI program (Section 3.3.1). Qia Prep miniprep kit

was used for all plasmid isolations and the quality of minipreps was analyzed using Nanodrop spectrophotometer by OD measuremet.

3.4.10 *Pichia pastoris* Transformation

Electroporation of Lithium Acetate and DTT pretreated Pichia cells protocol (Wu and Letchworth, 2004) was followed with slight modifications.

Pichia pastoris KM71H cells were grown in YPD broth to OD600 1.6 and pelleted by centrifugation at 2500 x g for 10 min at 4 °C and resuspended in 0.8 times the initial culture volume equivalent competent buffer (100mM LiAc, 10mM DTT, 0.6M Sorbitol, 10mM Tris-HCl pH 7.5). After 30 min incubation at room temperature, cells were pelleted again as before and resuspended in 1.5 times of the initial culture volume of 1M ice-cold sorbitol. Cells were washed three times in 1M ice-cold sorbitol and finally resuspended in 1M ice-cold sorbitol at a final concentration of 10^{10} cells / mL. Competent cells were treated with liquid nitrogen and stored at -80 °C until use. We did not observe any significant loss in competency for 12 months.

Thawed cells were mixed 3 – 100 ng of DNA (dissolved in EB buffer) and incubated on ice for 10 min after transferring to electroporation cuvette. The electroporating pulse was applied at 1.5kV, 25uF, 175ohm using ElectroCell Manipulator BTX 630 during 4 – 6 ns and cells were immediately resuspended in 1 mL of 1M ice-cold sorbitol and incubated at 30 °C while shaking at 250 rpm. After 1 hour incubation, cells were pelleted and resuspended in the possible smallest volume and spreaded onto YPD-Zeo plates using spreading beads (Worthington et al., 2001) and plates were incubated at 30 °C.

3.4.11 *Pichia pastoris* Colony PCR

Colony PCR's for yeast colonies were performed using Taq polymerase. The primers used were F_alpha and R_AOX (Table 3–4).

P. pastoris colonies were picked using 10ul tip, and placed at the bottom of the PCR tubes. After 3 min microwaving at highest settings, PCR master mix (containing dNTP's, Taq, primers etc) was delivered into tubes and "Colony PCR" program

(Section 3.3.7) was run. Whole PCR reaction was analyzed by analytical Agarose gel electrophoresis.

3.4.12 Selection of transformants

One of the selections was carried out on the rhodamine plates (Kouker and Jaeger, 1987). The transformed *Pichia pastoris* culture was plated out onto rhodamine containing plates. The functional expression was subsequently screened by the observation of the appearance of the pale pink fluorescent around the colony. The medium was consisted of 15% agar, 0.1 M potassium phosphate buffer; ph 6.0, 1.34% yeast nitrogen base with ammonium sulfate, 0.00004% (w/v) biotin, 0.003% (w/v) rhodamine, 1.2% (v/v) Kırlangıç Extra Virgin Oil, 0.5% (v/v) methanol. In the course of the preparation of the medium firstly and solely agar was autoclaved. After the sterile 15% agar solution cooled to 50-60°C the rest of the ingredients were added in the order written above. The homogenous distribution of the rhodamine and the oil into the medium was observed. Vigorous shaking and mixing at this step is required to prevent oil from forming drops in the solution. Once uniformity of the solution is ensured the immediate pouring is carried out.

The plates were kept at 4°C for no more than 2 weeks. Nevertheless methanol at a final concentration of 0.5% is added into the medium before incubation 300 µl of methanol were dropped onto each plate's lid. Pink fluorescent color development was observed during experiments, and images were taken using Bio-Rad DNA gel documentation system.

3.4.13 Real-Time PCR for copy number detection

A single colony was inoculated into 2.5 ml YPD. The culture was left at 30 °C with continuous shaking for overnight. The isolation of the genomic DNA from *P. pastoris* is performed following the procedure described in Current Protocols in Molecular Biology (Hoffman, 1997). 1ml of the overnight culture of about 30 OD 600 was harvested at 13,200 rpm for 5 min. In order to remove contaminating medium, the pellet was resuspended in 500 µl ddH_2O and pelleted again. The supernatant was removed without disturbing the pellet. The cells were resuspended in 200 µl breaking

buffer (2 % Triton X-100 (v/v), 1 % SDS (v/v), 100 mM NaCl, 10 mM Tris–Cl, pH 8.0, 1 mM EDTA, pH 8.0). 200 µl of glass beads and 200 µl phenol:chloroform:isoamyl alcohol (25:24:1) were added onto suspension. The cell suspension was subjected to vortexing until the solution becomes transparent. Usually the time was around 2-3 minutes. Then 200 µl TE buffer (10 mM Tris.Cl pH 8.0 and 1 mM EDTA pH 8.0) was added followed with a brief mixing. The solution was centrifuged at 13,200 rpm for 5 minutes. The upper aqueous phase (400 µl) was transferred to a new tube where ethanol precipitation of the genomic content was carried out. The contaminating RNA was removed by the addition of RNase A (1 mg/ml). The solution was incubated at 37 °C for 5 min. Another round of ethanol precipitation was done. The pellet was air dried and dissolved in sterile ddH_2O. The quantification was done spectrophotometrically. The quality of the genomic DNA was also visualized in agarose gel.

3.4.14 Pichia Expression

Starter culture was inoculated into 5 ml YPD from single colony. The second day the 5 ml culture was transferred to 50 ml YPD. The culture was centrifuged at 3500 rpm for 5 min. The supernatant was discarded. The cells were then washed with sterile ddH_2O and pelleted again. The expression was initiated in 50 ml BM medium by the final concentration of 1 % methanol and 1 % sorbitol. The starting OD_{600} of the culture was adjusted to 30. 1 ml of the starting culture was sampled and entitled as t_0. Every 24 hours 1 ml sampling continued for 4 days. 1 ml medium was taken and centrifuged at highest speed for 5 min. The supernatant was then transferred to another tube and kept -20°C. Methanol addition without changing its final concentration was also repeated during each sampling due to the evaporation of the available methanol in the medium.

3.4.15 *E. coli* Expression

The *E. coli* expression strains used were Rosettagami2 and Origami2 (Strain backgrounds). The transformants strains holding the recombinant pET22b(+) plasmids were grown in 5 ml Luria Bertani medium provided with the required antibiotics for selection of the expression strain and the transformants at 37 °C for continuous shaking.

The overnight culture was used to inoculate the starting culture. The starting OD600 was set to 0.02 and the induction was carried out using isopropyl--D-thiogalactopyranoside (IPTG) at final concentration of 0.1 mM when the cell density reaches the OD600 interval of 0.5-0.6. The incubation conditions were altered to check to recombinant expression. The temperature was gradually decreased to 30°C, 25°C, 20°C, 23°C, 15°C and 4°C. The IPTG concentration was decreased to 0.05 mM at 30°C. The induction OD600 was increased to 0.9-1.0 at 25°C. 1 ml of sample was collected and centrifuged for 3 min at 13200 rpm for different time intervals starting from the time of induction. An optical density measurement at 600 nm for each sampling time was carried out. The pellets were kept at -20°C.

3.4.16 *E. coli* Disruption and Fractioning the Soluble Fraction

The pellets were resuspended in Y-PER solution purchased from PIERCE. The required volume for each sample was calculated proportionally to the corresponding OD600 measurement data. The solutions were then incubated in 37°C with shaking at 200 rpm. Another cell disruption was also carried out using sonication. The pellets were dissolved in 1X PBS and exposed to sonication (10 sec., 40% for 5 times). The samples were kept on ice during process.

After the disruption was ensured, the samples were centrifuged for 5 minutes at 13,200 rpm to remove soluble fraction. The supernatants and pellets were separately kept at -20°C for further analyses. The insoluble fraction was re-solubilized using 1X Laemnli Sample Buffer. The soluble fraction and insoluble fraction were subjected to molecular size separation in 12% polyacrylamide gels which were visualized by PageBlue stain (Fermentas). The soluble fraction of the 25°C expression samples harvested from Rosettagami2 strain was separated in 12% polyacrylamide gel. Semi-dry transfer method was used for blotting of proteins to PVDF membrane at 220 milli-amperes for 1 hour. Overnight blocking in 5% Blotto (Rockland) was followed with 1 hour of α-histidine (Roche) incubation at room temperature. Subsequent signal generation was achieved using Lumi-Light^{Plus} Western blotting substrate (Roche) and blot was developed on X-ray film (Roche).

3.4.17 Enzyme Assays

The activity of lipase was measured in a Corning brand 96-well black microtiter plates (total volume of 200 µL) using 4-methylumbelliferyl-caprylate substrate. The stock solutions of the substrate were 50 µM, dissolved in methanol and aliquots were kept at -20°C. The assay solution consisted of 250 µM substrate and 100 mM Tris-Cl pH: 7.25. The quantification of the fluorescent compound 4-MU using an excitation wavelength of 355 nm and an emission wavelength of 460 nm was performed in Gemini XS Spectrofluorimeter in kinetic analysis mode. The measurement performed every minute for 1 hour at room temperature. The reaction rate values of the listed lipases were calculated accordingly. One unit of lipase activity was defined as the amount of enzyme releasing RFU / Sec under the assay conditions. Activity values were corrected for auto hydrolysis which was determined in the absence of enzyme. All measurements were performed in duplicates.

3.5 Results and Discussion

3.5.1 Cloning

The gene fragment for the mature ROL enzyme were successfully cloned between EcoRI and NotI sites with two stop codons and designated as pPICZaROL_stop. Similarly, PCR amplified fragment containing mature ROL was cloned between EcoRI and SalI restriction sites of pPICZαA vector and construct were labeled as pPICZaROL_w/his.

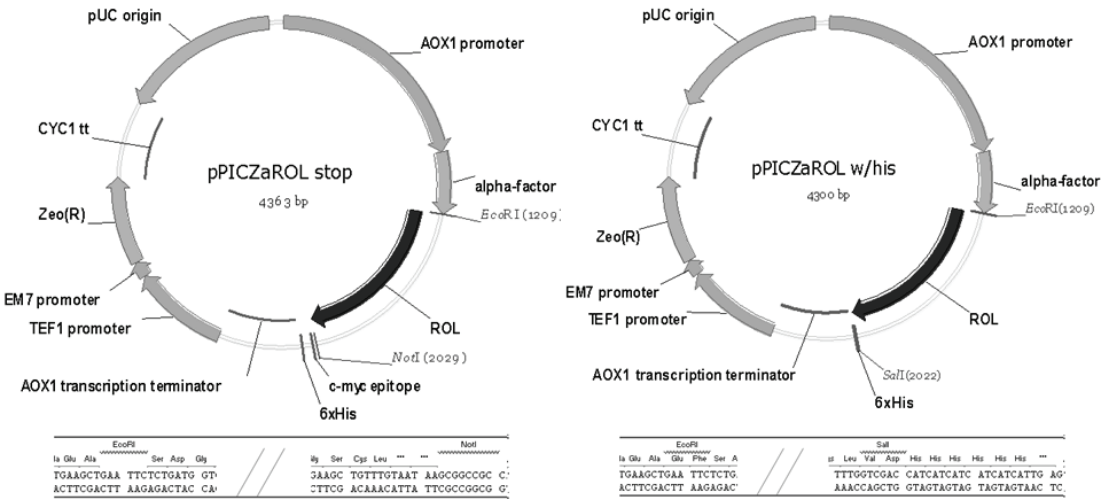


Figure 3.2 Cloning of ROL into pPICZalphaA with His-tag and without His-tag

3.5.2 Site Directed Mutagenesis

We have successfully mutated the ROL gene using site directed mutagenesis kit. We have used pPICZalphaA-ROL as a template for mutagenesis studies. pPICZalphaA vector is about 3000 bp and suitable for SDM technique. Although we have used DNA polymerase with high proof reading capacity, we cannot guarantee that there are no mutations in the other parts of vectors, as we have had the sequences only the region of construct that contain gene of interest. After sequencing of the constructs, we have re-cloned mutant genes into empty pPICZalphaA vector. Primers used to generate mutant enzymes are listed in Table 3–4.

3.5.3 *Pichia pastoris* Transformation and Selection

Colony PCR method was used to analyze the colonies after transformation and positive colonies were streaked at least 4 times using YPD-Zeo plates. As KM71H cells do not have any homologous sequence of α-leader sequence (derived from *S. cerevisiae*), we have used F_alpha (Table 3–4) as a forward primer and R_AOX (Table 3–4) as a reverse primer in colony PCR. As seen in Figure 3.3, I found at least one hit for each mutant and these colonies were selected for further studies.

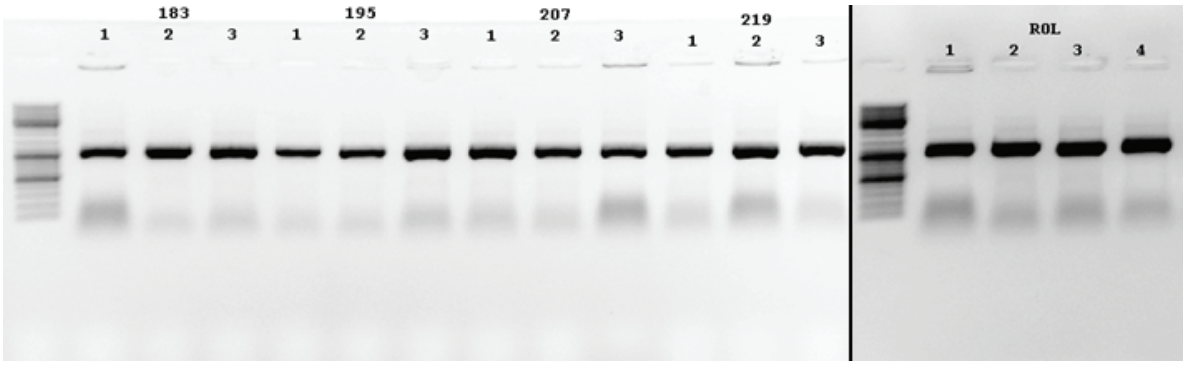


Figure 3.3 Identification of the positive transformants using Colony-PCR

3.5.4 Real-time PCR analysis of clones

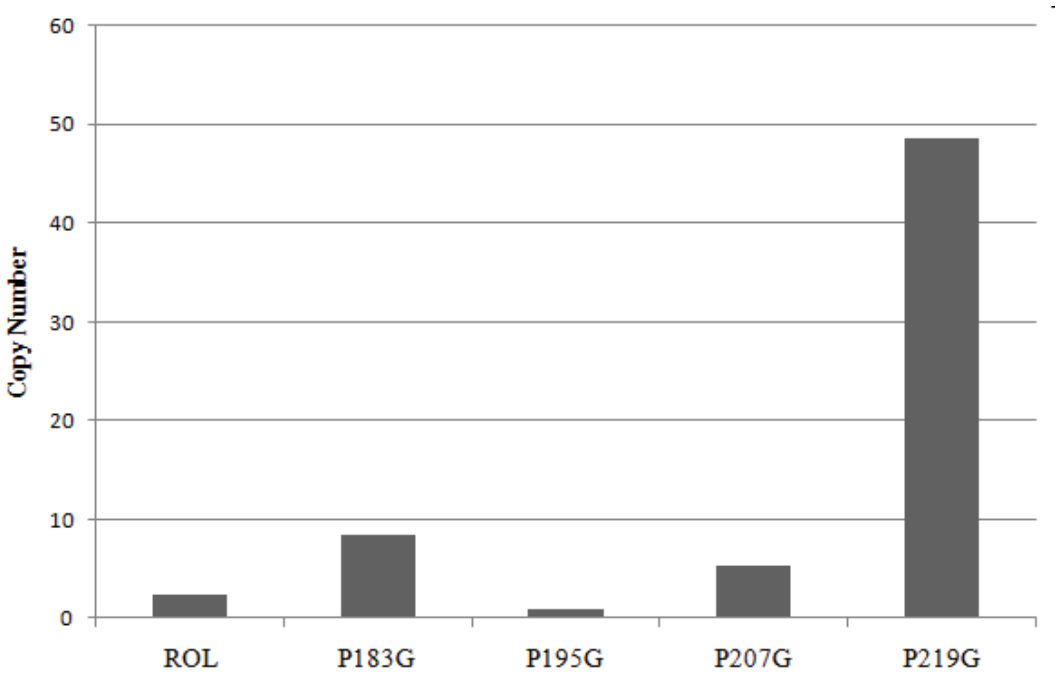


Figure 3.4 Copy number analysis of the genome integrated vectors for native and mutant enzymes.

Copy numbers of expression cassettes integrated into genome for each selected transformant containing one of the mutant lipases were analyzed by real-time PCR and results are presented in Figure 3.4.

3.5.5 Plate Assay

For all of the mutant enzymes, we have selected high enzyme producers using plate assay. Representative rhodamine-MM plates for the transformants selected for the shake flask expression studies can be seen in Figure 3.5. Native KM71H *P. pastoris* cells were also streaked onto rhodamine-MM plates as a negative control. All plates were supplemented with Methanol by putting 200 µl onto the lid of petri dish daily.

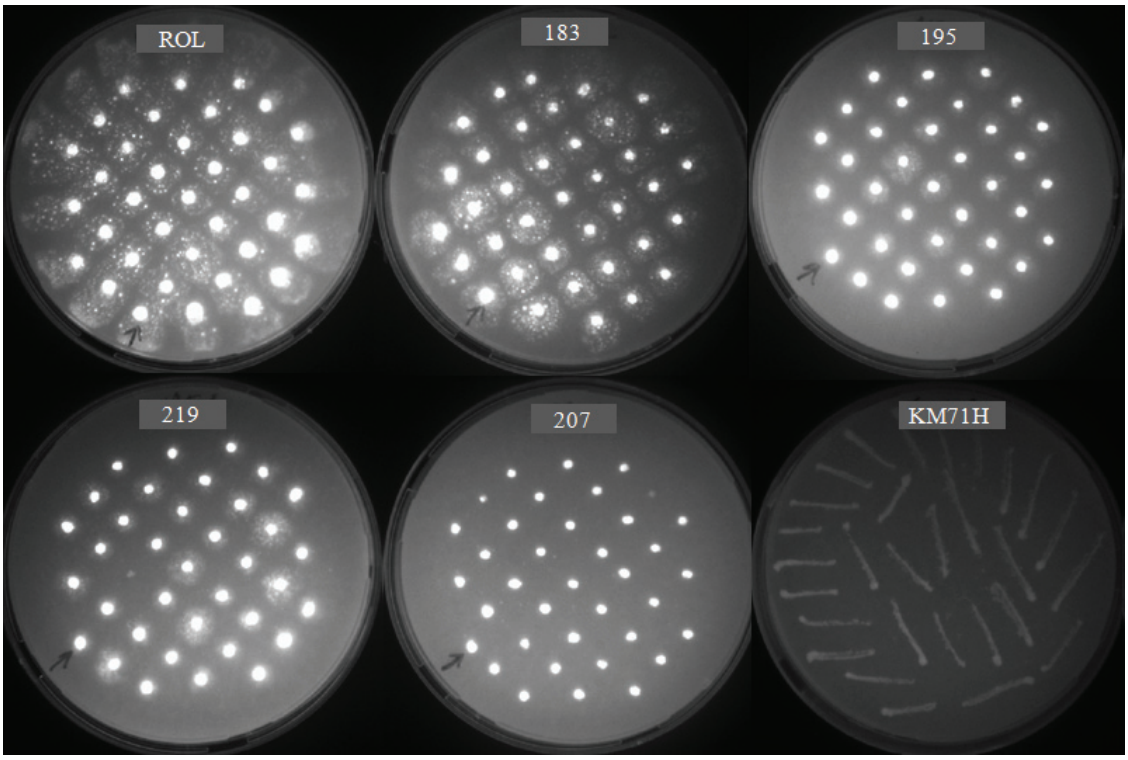


Figure 3.5 Rhodamine-MM plate assay for *P. pastoris* transformants showing fluorescent halo. KM71H cells were used as a negative control.

3.5.6 Expression and Purification

P. pastoris Expression

We have established an expression protocol for with and without his-tagged ROL enzyme using *P. pastoris* system. We have used pPICZαA plasmid for all cloning works. It has a signal sequence (alpha leader) which directs the proteins to the extracellular matrix. We have functionally expressed ROL mutants in KM71H Mut^{s}

cells using baffled flasks with shaking at 30 °C. SDS-PAGE and activity analysis of the results showed that we have successfully expressed all mutant proteins and achieved the secretion into the medium. After this finding we have tried to optimize the purification procedure for ROL enzyme. Unfortunately, we have not achieved desirable homogeneity and the remaining protein amount was not satisfactory for further structural analysis after ion exchange and size exclusion chromatography.

We have also established an expression procedure for C-terminal 6X-his tagged ROL enzymes in order to purify proteins using metal affinity chromatography (Figure 3.6). Surprisingly, metal affinity chromatography analysis showed that his-tagged ROL enzyme does not bind to the affinity column. After trying different binding and elution parameters for his-tagged ROL enzyme, we could not achieved good purification yields.

Since his-tagged ROL expression gave relatively pure enzymes compared to without tag clones (Figure 3.8), I decided to use his-tagged enzymes for further studies. In order to investigate thermal stability of these two different clones (his-tagged and without tag) I performed the activity assays (Figure 3.7). My results showed that, although the stability of his-tagged enzyme is somehow lower than the native enzyme, it can be used for comparison studies in order to investigate point mutation effects on stability of the ROL enzyme (Section 3.5.7).

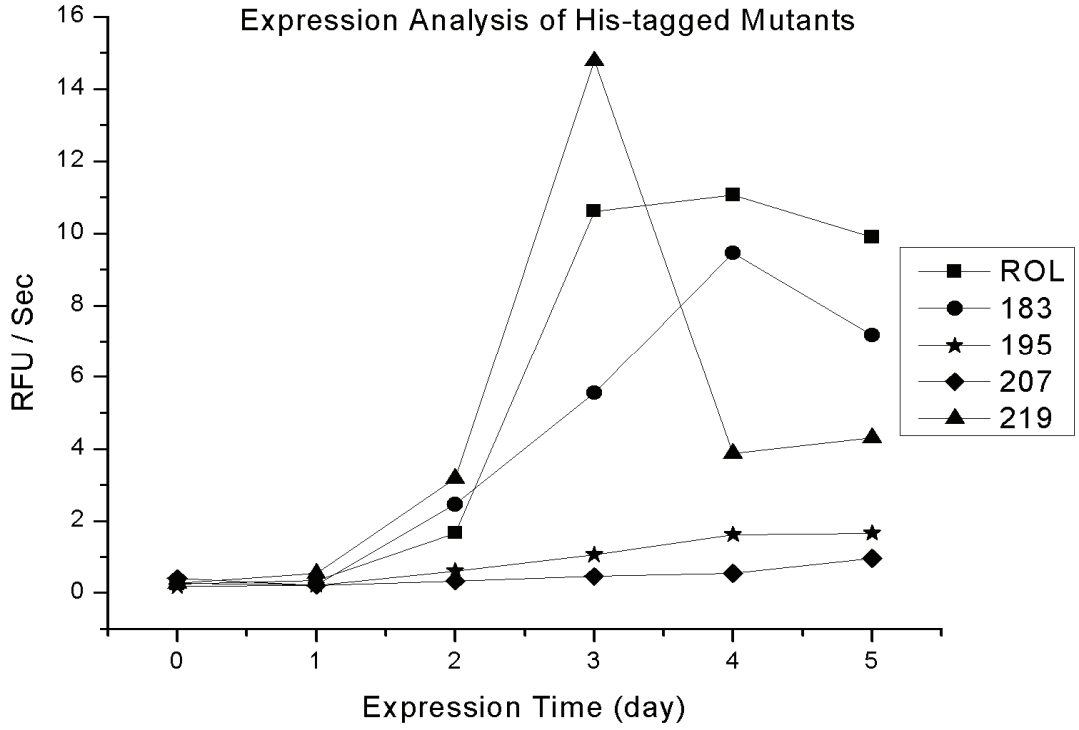


Figure 3.6 Activity assay of the expressed lipases during induction period. X-axis shows initial velocity of reaction in (RFU/sec) and Y-axis shows the time of the expression taking induction time as zero.

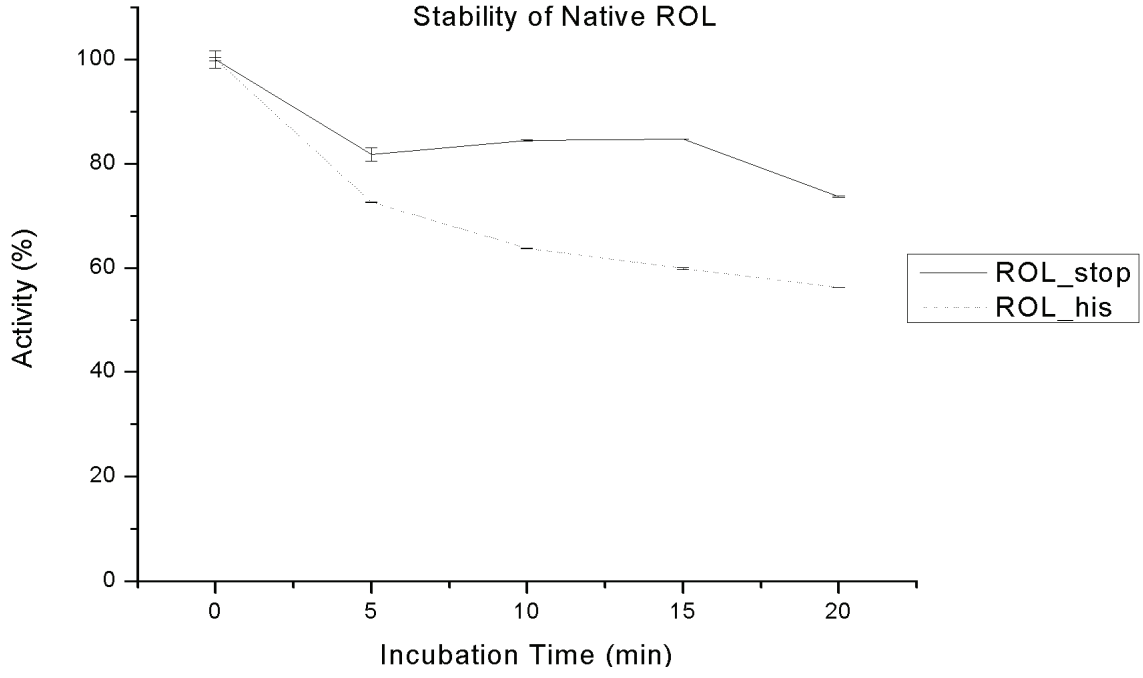


Figure 3.7 Stability analysis of His tagged and without his tagged native ROL enzymes. X-axis shows remaining activity after incubation at 50 °C for indicated period of time.

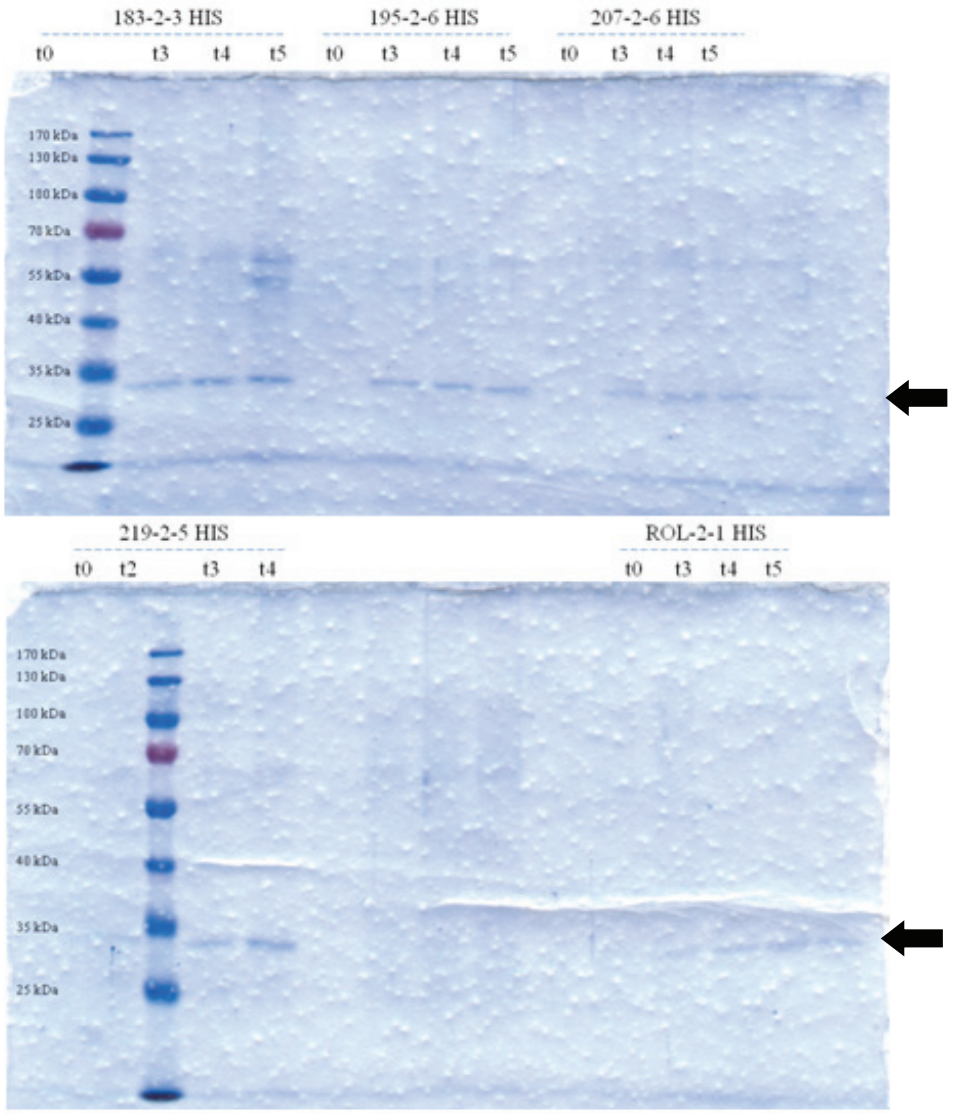


Figure 3.8 SDS-PAGE analysis of mutant lipase expressing *P.pastoris* transformants. Arrows shows secreted native and mutated enzymes around 30 kDa molecular weight.

E. coli Expression

Functional expression of ROL in *E. coli* has not been achieved using other expression strains except Origami strain from Novagen (Di Lorenzo et al., 2005). Unfortunately despite the presence of the previous report on the ROL expression in *E. coli* the expression was not in the soluble fraction. The *E. coli* expression using the modified strains Rosetta-gami 2 and Origami 2 was attempted.

The expression at 37°C (Figure 3.9) was initially performed and for both of the strains the protein was found as an inclusion body form. Therefore the temperature of the incubation was decreased gradually to 30°C (Figure 3.10), 25°C (Figure 3.10) ,

20°C, 23°C, 15°C and 4°C (Figure 3.11) with continuous shaking. Nevertheless a wider temperature than the first report on ROL expression in *E. coli* was monitored, and no expression in the soluble fraction was observed at any of these incubation temperatures. The insoluble fraction was concurrently analyzed. There was no expression of the ROL at lower temperatures than 30°C for the insoluble fraction (Figure 3.11). In other words aside from a gain in the soluble expression the overall expression was lost at lower temperatures than 30°C. The results comprise an obvious conflict with the previous *E. coli* expression of ROL. The strain differences between both expression works might be leading this discrepancy. Di Lorenzo et.al., have used Rosetta and Origami strains since the modified versions of those strains have been transformed in this expression. The expression strains Rosetta-gami 2 and Origami 2 are unique to have the ability to form disulphide bridges in recombinant proteins. Though being modified versions these strain differences might have failed to reproduce the previous findings.

Apart from temperature optimizations during induction phase other critical factors such as IPTG concentration and initial cell density were altered to monitor the expression levels. The IPTG concentration was decreased to 0.5 in order to lessen the induction rate and thus increase the solubility of the expression at 30°C. Moreover the cell population during induction was increased to 1 OD600 at 25°C in order to initiate an overall expression. Yet both of the alterations did not advance the expression at all. (Data not shown)

The same culture was induced and induction performed at a temperature gradient of 30°C, 25°C, 20°C, 23°C, 15°C and 4°C for Rosetta-gami 2 strain using the ROL lipase and its mutant form annotated as 183. The expression was again monitored at the 30 and 37°C only in the insoluble form. The expression products were visualized using PageBlue staining following instruction's manual.

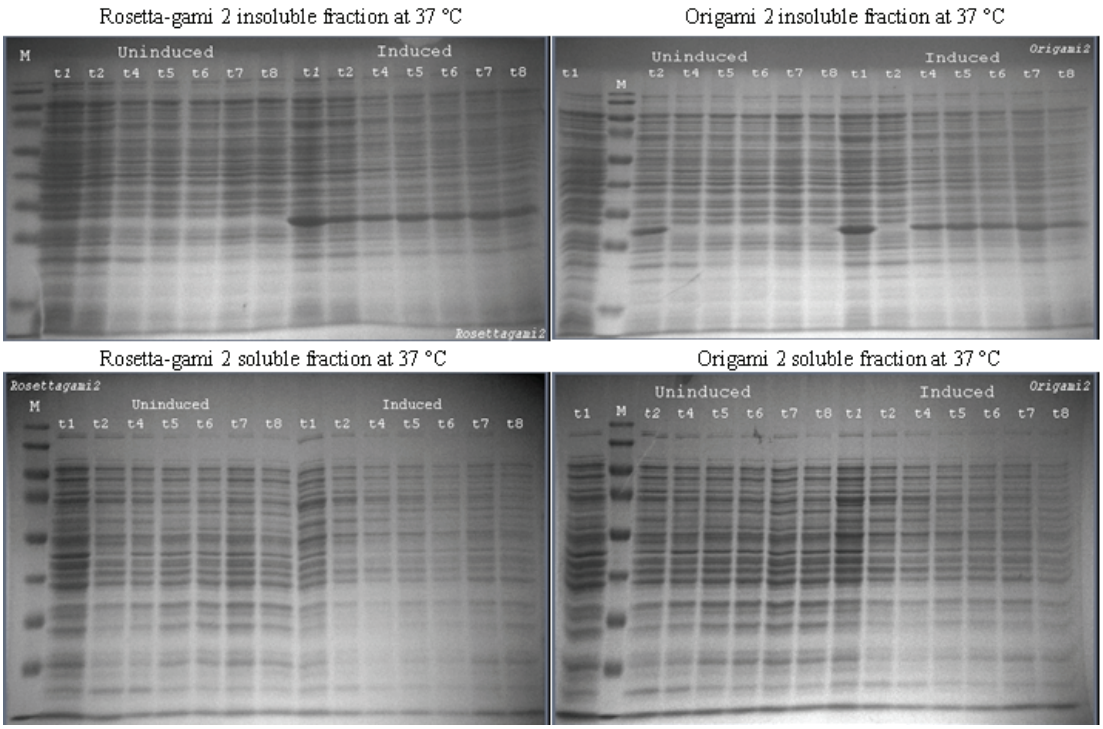


Figure 3.9 *E. coli* expression of ROL in Rosetta-gami 2 and Origami 2 strains at 37 °C. Soluble and insoluble fractions were loaded for both induced and uninduced cells. *Origami 2 insolube uninduced t2 lane and induced t2 samples were misloaded. t2 means 2 hour sample after induction point.

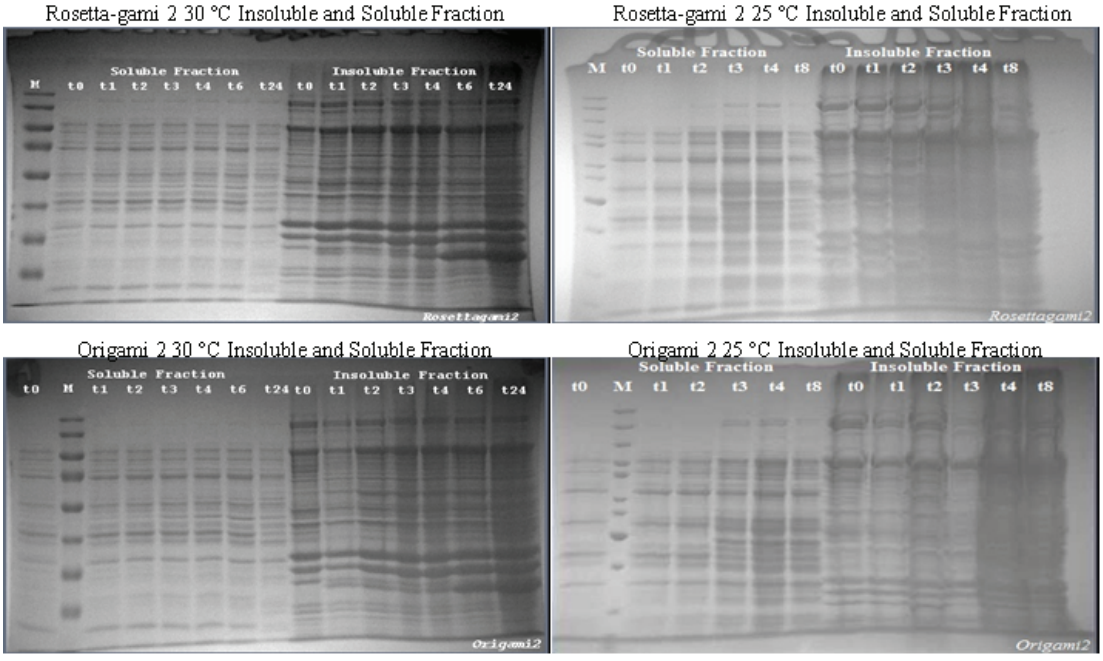


Figure 3.10 *E. coli* expression of ROL in Rosetta-gami 2 and Origami 2 strains at 30 °C and 30 °C. Soluble and insoluble fractions were loaded for both induced and uninduced cells. t2 means 2 hour sample after induction point.

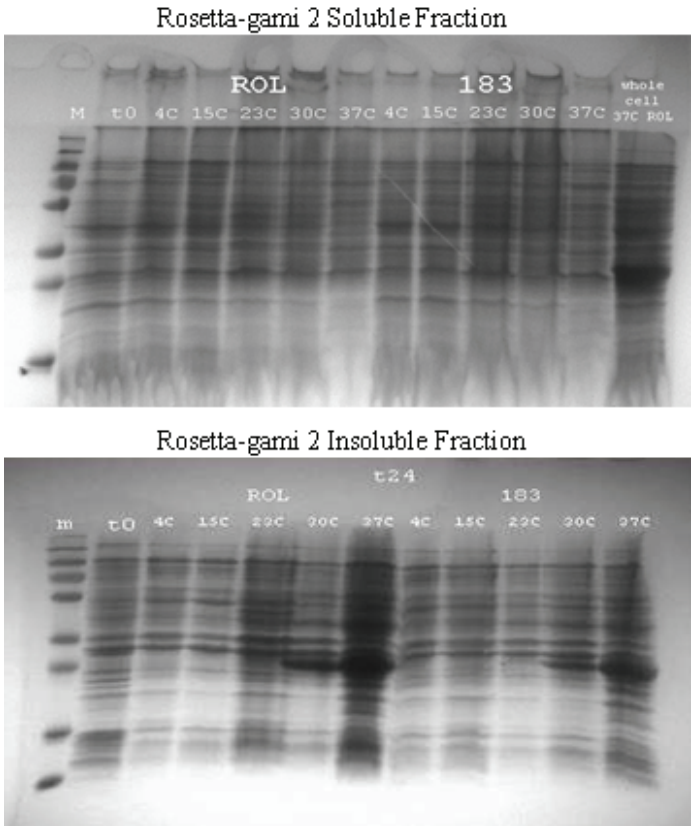


Figure 3.11 Soluble and Insoluble fractions of Rosetta-gami 2 containing ROL gene were loaded after 24 hour induction at different temperatures.

3.5.7 Enzyme Thermal Stability

Enzyme stability assessments were done by incubating enzyme solutions at various temperatures for predefined time intervals. All enzymes were placed onto ice after incubation period, and immediately assayed against 4MU-Caprylate.

As seen in Figure 3.12, P195G mutation has thermostability enhancing effect on ROL enzyme. Although P207G and P219G mutations have relatively positive impact on the stability of ROL enzyme, increasing in the thermostability of ROL remains small comparing to the P195G. On the other hand, P183G mutation has decreased the stability of ROL enzyme and activity was nearly 25 % after incubation at 50 °C for 20 min only.

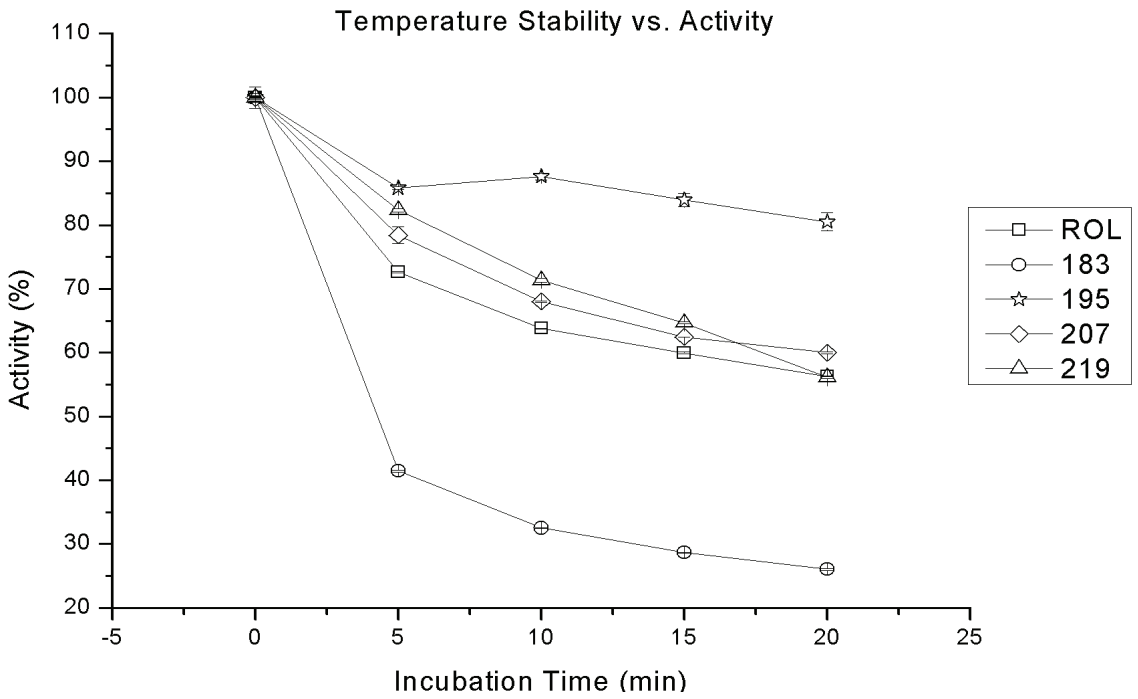


Figure 3.12 Residual lipase activity after incubation at 50 °C. X-axis shows remaining activity after incubation at 50 °C for indicated time period. Activities are given in percentage form taking the initial activity as 100 %.

3.6 Conclusion

I have successfully expressed the native and mutant ROL enzymes in *P. pastoris* expression system with or without 6XHis tag. All of the mutations predicted in Chapter 2 were created using site directed mutagenesis and constructs containing mutant enzymes were transformed into *Pichia pastoris* KM71H Mut^{s} cells for secreted expression. I have used electroporation technique complemented with rhodamine-MM plate screening assay to select enzyme producing transformants. Unfortunately, we could not isolate enzymes to homogeneity level purity using cation exchange chromatography. Although we have re-cloned all of the native and mutant genes into an expression vector containing C-terminal 6XHIS tag, after metal affinity chromatography we could not obtain high purity proteins. Expressed proteins using *E. coli* strains, Origami 2 and Rosetta-Gami 2, have been found as insoluble form.

On the other hand, I performed stability assays using 6XHis tagged native and mutant ROL enzymes using highest possible purity expression. I have found that, P195G mutation has increased the stability of ROL enzyme at 50 °C. P195G mutant enzyme has kept its activity around 80 % after incubation at 50 °C for 20 min, while native enzyme had only 56 % of its activity. Among the predicted mutations, P207G and P219G have been showed similar stability as native ROL enzyme.

I observed that P183G mutation had negative impact on the thermostability of ROL enzyme; residual activity of P183 after 20 min was around 26 % which was nearly 30 % less than native ROL. My findings after molecular biology studies showed similar thermostability patterns as I concluded after molecular dynamics simulations.

3.7 References

Ahn, J., J. Hong, H. Lee, M. Park, E. Lee, C. Kim, E. Choi and J. Jung. "Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter."(2007) Appl Microbiol Biotechnol **74**(3): 601-608.

Anna-Arriola, S. S. and I. Herskowitz. "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase."(1994) Yeast **10**(6): 801-810.

Beer, H. D., G. Wohlfahrt, R. D. Schmid and J. E. McCarthy. "The folding and activity of the extracellular lipase of Rhizopus oryzae are modulated by a prosequence."(1996) Biochem J **319 (Pt 2)**: 351-359.

Beisson, F., A. Tiss, C. Rivière and R. Verger. "Methods for lipase detection and assay: a critical review."(2000) European Journal of Lipid Science and Technology **102**(2): 133-153.

Brake, A. J., J. P. Merryweather, D. G. Coit, U. A. Heberlein, F. R. Masiarz, G. T. Mullenbach, M. S. Urdea, P. Valenzuela and P. J. Barr. "Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae."(1984) Proceedings of the National Academy of Sciences of the United States of America **81**(15): 4642-4646.

Brierley, R. A. "Secretion of recombinant human insulin-like growth factor I (IGF-I)."(1998) Methods Mol Biol **103**: 149-177.

Brockman, H. L. "Triglyceride lipase from porcine pancreas."(1981) Methods Enzymol **71 Pt C**: 619-627.

Burdychova, R., V. Ruzicka and M. Bartos. "PCR-based method for identification of integration events in the Pichia pastoris genome."(2002) Biotechniques **33**(6): 1214-1216, 1218.

Cereghino, J. L. and J. M. Cregg. "Heterologous protein expression in the methylotrophic yeast Pichia pastoris."(2000) FEMS Microbiol Rev **24**(1): 45-66.

Clare, J. J., F. B. Rayment, S. P. Ballantine, K. Sreekrishna and M. A. Romanos. "High-Level Expression of Tetanus Toxin Fragment C in Pichia Pastoris Strains Containing Multiple Tandem Integrations of the Gene."(1991) Nat Biotech **9**(5): 455-460.

Cos, O., D. Resina, P. Ferrer, J. L. Montesinos and F. Valero. "Heterologous production of Rhizopus oryzae lipase in Pichia pastoris using the alcohol oxidase and formaldehyde dehydrogenase promoters in batch and fed-batch cultures."(2005) Biochemical Engineering Journal **26**(2-3): 86-94.

Cregg, J., J. Cereghino, J. Shi and D. Higgins. "Recombinant protein expression in Pichia pastoris."(2000) Molecular Biotechnology **16**(1): 23-52.

Cregg, J. M. (2007a). DNA-Mediated Transformation. **389:** 27-42.

Cregg, J. M. (2007b). Pichia protocols. Totowa, N.J., Humana Press.

Cregg, J. M., K. J. Barringer, A. Y. Hessler and K. R. Madden. "Pichia pastoris as a host system for transformations."(1985) Mol Cell Biol **5**(12): 3376-3385.

Cregg, J. M., K. R. Madden, K. J. Barringer, G. P. Thill and C. A. Stillman. "Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris."(1989) Mol Cell Biol **9**(3): 1316-1323.

Cregg, J. M., T. S. Vedvick and W. C. Raschke. "Recent advances in the expression of foreign genes in Pichia pastoris."(1993) Biotechnology (N Y) **11**(8): 905-910.

Daly, R. and M. T. Hearn. "Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production."(2005) J Mol Recognit **18**(2): 119-138.

de Almeida, J. R., L. M. de Moraes and F. A. Torres. "Molecular characterization of the 3-phosphoglycerate kinase gene (PGK1) from the methylotrophic yeast Pichia pastoris."(2005) Yeast **22**(9): 725-737.

Di Lorenzo, M., A. Hidalgo, M. Haas and U. T. Bornscheuer. "Heterologous Production of Functional Forms of Rhizopus oryzae Lipase in Escherichia coli."(2005) Appl. Environ. Microbiol. **71**(12): 8974-8977.

Gilham, D. and R. Lehner. "Techniques to measure lipase and esterase activity in vitro."(2005) Methods **36**(2): 139-147.

Gleeson, M. A., C. E. White, D. P. Meininger and E. A. Komives. "Generation of protease-deficient strains and their use in heterologous protein expression."(1998) Methods Mol Biol **103**: 81-94.

Higgins, D. R. and J. M. Cregg (1998). Pichia protocols. Totowa, N.J., Humana Press.

Hoffman, C. S. "Preparation of Yeast DNA."(1997) Current Protocols in Molecular Biology **39**: 13.11.11 - 13.11.14.

Jacks, T. J. and H. W. Kircher. "Fluorometric assay for the hydrolytic activity of lipase using fatty acyl esters of 4-methylumbelliferone."(1967) Analytical Biochemistry **21**(2): 279-285.

Joerger, R. and M. Haas. "Overexpression of a Rhizopus delemar lipase gene in Escherichia coli."(1993) Lipids **28**(2): 81-88.

Kouker, G. and K. E. Jaeger. "Specific and sensitive plate assay for bacterial lipases."(1987) Appl Environ Microbiol **53**(1): 211-213.

Lee, C. C., T. G. Williams, D. W. S. Wong and G. H. Robertson. "An episomal expression vector for screening mutant gene libraries in Pichia pastoris."(2005) Plasmid **54**(1): 80-85.

Lin-Cereghino, J., M. D. Hashimoto, A. Moy, J. Castelo, C. C. Orazem, P. Kuo, S. Xiong, V. Gandhi, C. T. Hatae, A. Chan and G. P. Lin-Cereghino. "Direct selection of Pichia pastoris expression strains using new G418 resistance vectors."(2008) Yeast **25**(4): 293-299.

Lin Cereghino, G. P., J. Lin Cereghino, A. J. Sunga, M. A. Johnson, M. Lim, M. A. Gleeson and J. M. Cregg. "New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of Pichia pastoris."(2001) Gene **263**(1-2): 159-169.

Liu, H., X. Tan, K. A. Russell, M. Veenhuis and J. M. Cregg. "PER3, a Gene Required for Peroxisome Biogenesis in Pichia pastoris, Encodes a Peroxisomal Membrane Protein Involved in Protein Import."(1995) Journal of Biological Chemistry **270**(18): 10940-10951.

Mack, M., M. Wannemacher, B. Hobl, P. Pietschmann and B. Hock. "Comparison of two expression platforms in respect to protein yield and quality: Pichia pastoris versus Pichia angusta."(2009) Protein Expression and Purification **66**(2): 165-171.

Menendez, J., I. Valdes and N. Cabrera. "The ICL1 gene of Pichia pastoris, transcriptional regulation and use of its promoter."(2003) Yeast **20**(13): 1097-1108.

Minning, S., C. Schmidt-Dannert and R. D. Schmid. "Functional expression of Rhizopus oryzae lipase in Pichia pastoris: high-level production and some properties."(1998) Journal of Biotechnology **66**(2-3): 147-156.

Minning, S., A. Serrano, P. Ferrer, C. Sola, R. D. Schmid and F. Valero. "Optimization of the high-level production of Rhizopus oryzae lipase in Pichia pastoris."(2001) Journal of Biotechnology **86**: 59-70.

Mumberg, D., R. Muller and M. Funk. "Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression."(1994) Nucl. Acids Res. **22**(25): 5767-5768.

Nakata, Y., X. Tang and K. K. Yokoyama (1997). Preparation of Competent Cells for High-Efficiency Plasmid Transformation of Escherichia coli**:** 129-137.

Nyfeler, E., J. Grognux, D. Wahler and J.-L. Reymond. "A Sensitive and Selective High-Throughput Screening Fluorescence Assay for Lipases and Esterases."(2003) Helvetica Chimica Acta **86**(8): 2919-2927.

Paifer, E., E. Margolles, J. Cremata, R. Montesino, L. Herrera and J. M. Delgado. "Efficient expression and secretion of recombinant alpha amylase in <I>Pichia pastoris</I> using two different signal sequences."(1994) Yeast **10**(11): 1415-1419.

Romanos, M. "Advances in the use of Pichia pastoris for high-level gene expression."(1995) Current Opinion in Biotechnology **6**(5): 527-533.

Romanos, M. A., J. J. Clare, K. M. Beesley, F. B. Rayment, S. P. Ballantine, A. J. Makoff, G. Dougan, N. F. Fairweather and I. G. Charles. "Recombinant Bordetella pertussis pertactin (P69) from the yeast Pichia pastoris: high-level production and immunological properties."(1991) Vaccine **9**(12): 901-906.

Sambrook, J. and D. W. Russell (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Sambrook, J. and D. W. Russell. "The Inoue Method for Preparation and Transformation of Competent E. Coli: "Ultra-Competent" Cells."(2006) Cold Spring Harbor Protocols **2006**(2): pdb.prot3944-.

Sears, I. B., J. O'Connor, O. W. Rossanese and B. S. Glick. "A versatile set of vectors for constitutive and regulated gene expression in <I>Pichia pastoris</I>."(1998) Yeast **14**(8): 783-790.

Shen, S., G. Sulter, T. W. Jeffries and J. M. Cregg. "A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris."(1998) Gene **216**(1): 93-102.

Som, I., R. N. Mitsch, J. L. Urbanowski and R. J. Rolfes. "DNA-bound Bas1 recruits Pho2 to activate ADE genes in Saccharomyces cerevisiae."(2005) Eukaryot Cell **4**(10): 1725-1735.

Sunga, A. J. and J. M. Cregg. "The Pichia pastoris formaldehyde dehydrogenase gene (FLD1) as a marker for selection of multicopy expression strains of P. pastoris."(2004) Gene **330**: 39-47.

Thompson, J. R., E. Register, J. Curotto, M. Kurtz and R. Kelly. "An improved protocol for the preparation of yeast cells for transformation by electroporation."(1998) Yeast **14**(6): 565-571.

Tschopp, J. F., P. F. Brust, J. M. Cregg, C. A. Stillman and T. R. Gingeras. "Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris."(1987) Nucleic Acids Res **15**(9): 3859-3876.

Vedvick, T. S. "Gene expression in yeast: Pichia pastoris."(1991) Current Opinion in Biotechnology **2**(5): 742-745.

Waterham, H. R., M. E. Digan, P. J. Koutz, S. V. Lair and J. M. Cregg. "Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter."(1997) Gene **186**(1): 37-44.

Worthington, M. T., R. Q. Luo and J. Pelo. "Copacabana method for spreading E. coli and yeast colonies."(2001) Biotechniques **30**(4): 738-740, 742.

Wu, S. and G. J. Letchworth. "High efficiency transformation by electroporation of Pichia pastoris pretreated with lithium acetate and dithiothreitol."(2004) Biotechniques **36**(1): 152-154.

Wung, J. L. and N. R. Gascoigne. "Antibody screening for secreted proteins expressed in Pichia pastoris."(1996) Biotechniques **21**(5): 808, 810, 812.

Zhu, T., M. Guo, C. Sun, J. Qian, Y. Zhuang, J. Chu and S. Zhang. "A systematical investigation on the genetic stability of multi-copy Pichia pastoris strains."(2009a) Biotechnol Lett **31**(5): 679-684.

Zhu, T., M. Guo, Z. Tang, M. Zhang, Y. Zhuang, J. Chu and S. Zhang. "Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast Pichia pastoris."(2009b) J Appl Microbiol **107**(3): 954-963.

Chapter 4

4 HIGH THROUGHPUT CLONING, EXPRESSION AND PURIFICATION IN PICHIA PASTORIS FOR MICROARRAY SCALE ENZYME ASSAYS

4.1 Introduction

Generation of large mutant gene libraries could be performed through various techniques (Arnold and Georgiou, 2003), such as chemical mutagenesis, error-prone mutagenesis, site saturation mutagenesis, shuffling, use of degenerate oligos or mutator strains (Brakmann and Lindemann, 2005). As a result of these developments, mutant libraries could be generated using almost routine methods. On the other hand, outcome of created libraries, through rational or random approaches, importantly depends on the expression and screening studies (Olsen et al., 2000). Although many important enzymes used in industrial processes have been successfully improved in *E. coli* (Farinas et al., 2001), functional expression in this host has its own bottlenecks. *P. pastoris* expression system with its post-translational modification abilities and secretion properties for heterologously expressed proteins could be beneficial for large scale expression and screening investigations.

We have started with a cloning strategy which can be used for cloning and transformation of mutant libraries efficiently into *P. pastoris*. Then we have focused on the expression strategy in a high throughput manner that should be reproducible and easy to handle. Finally, we have developed the microarray based lipase assays that can be used to analyze the expressed proteins and be compatible with the automated systems for the high throughput applications.

Preferably, the high throughput methodologies should be complementary to each other. In other words, one high throughput system should be followed with another one. In this thesis, we have studied the possibility of combining the high throughput

strategies covering cloning, expression, purification and enzymatic analysis (Figure 4.1).

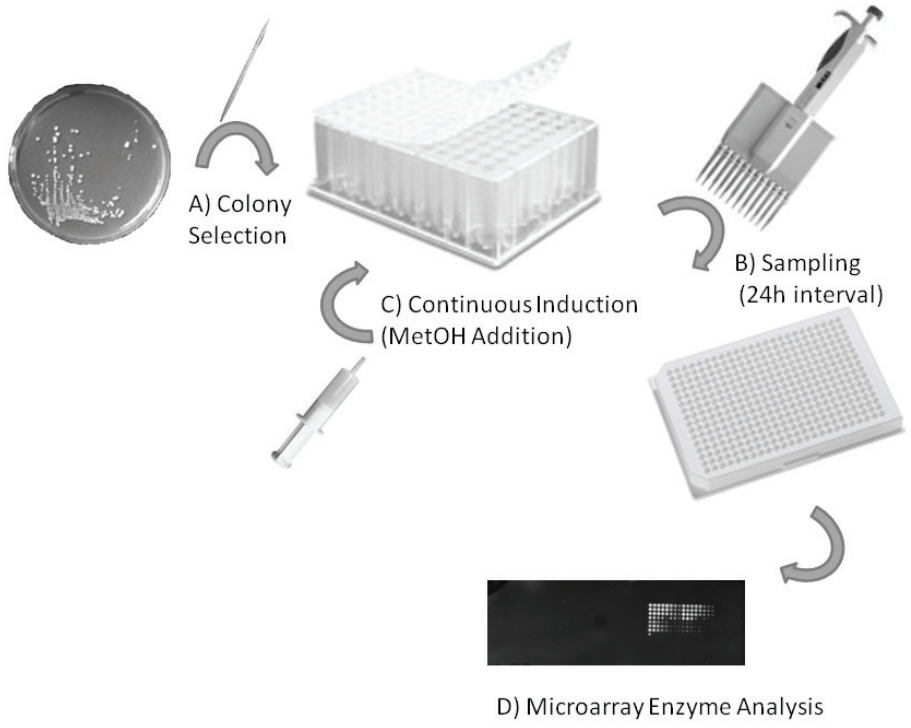


Figure 4.1 Proposed high throughput methodology for heterologous expression of proteins in *P. pastoris*. A) Colony selection for target protein expression, B) Sampling at predefined time intervals for expression analysis, C) Addition of Methanol in order to continue the expression, D) Microarray scale enzymatic analysis of samples.

Linear cassette transformation

Traditionally, the expression of a recombinant protein requires a preliminary cloning step of the target sequences into an expression vector before introducing into the expression hosts. Homology-based approaches in cloning techniques offer several advantages in high throughput procedures by eliminating the use of the restriction endonucleases and the ligases with many purification and isolation steps. Ligation independent cloning methods (Aslanidis and de Jong, 1990; Haun et al., 1992) and the use of recombination enzymes (Hartley et al., 2000) are the most common methods for structural genomic centers (Gräslund et al., 2008). Expression plasmids have to be constructed in *E.coli* before *P. pastoris* transformation. Generally, linearized vectors digested in AOX promoter region are used for genomic integration in Pichia system.

PCR based methods for generating expression cassettes can eliminate ligation, *E. coli* transformation and plasmid isolation steps. Overlap-extension PCR (OE-PCR) methodology can be applied to the *P. pastoris* expression system. Although, OE-PCR methods are generally used to create mutant or fusion proteins (Ho et al., 1989; Horton et al., 1989; Heckman and Pease, 2007; Simionatto et al., 2009), it can be used to generate linear expression cassettes that have the promoter, the gene of interest, and the antibiotic resistance gene (Figure 4.2). Exactly the same linear vector can be created using OE-PCR as well as minimized linear expression cassette which do not carry bacterial origin of replication.

We have first shown that overlap-extension (OE-PCR) PCR method can be applicable for *P. pastoris* transformation. We have designed several primers for PCR and amplified fragments for OE-PCR to test the efficiency of different parts of AOX promoter.

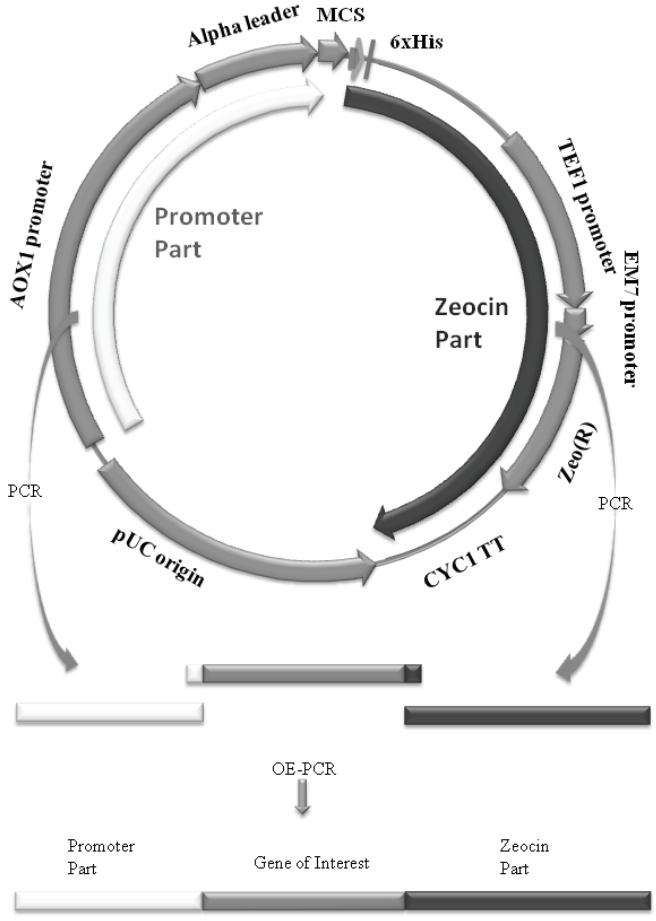


Figure 4.2 OE-PCR primers designed to generate linear expression cassettes.

High throughput expression for P. pastoris

It is very important to develop new technologies for the protein expression especially in high throughput manner. Although, there are some protocols about high the throughput protein expression in *P. pastoris* (Boettner et al., 2002; Bottner and Lang, 2004; Weis et al., 2004), we could not find any data about the non-repressing carbon source implementation in high throughput expression. Inan et al. reported that either one of alanine, sorbitol, mannitol or trehalose can be used as a sole carbon and energy source for *P. pastoris*, although the doubling time on trehalose was very long (Inan and Meagher, 2001). Xie et al. achieved highest angiostatin production of 191 mg/l using lactic acid. Lactic acid accumulation did not interfere with angiostatin production, indicating that lactic acid to be a non-repressive carbon source (Xie et al., 2005). Ramon et al. have used sorbitol/methanol mixed feeding strategy for *R. oryzae* lipase expression and this strategy resulted in a 2.2-fold increase in the heterologous protein level, compared with the methanol-only feeding strategy. In addition to the significant increase in the overall bioprocess volumetric productivity, co-feeding avoided the drastic decrease of the specific production rate observed after the start of the induction phase when methanol was used as sole carbon source. (Ramón et al., 2007). As a result of these findings, the screening of *P. pastoris* Mut^{s} strain for different non-repressing carbon sources in a high throughput approach may improve the system in terms of followings;

- Higher levels of heterologous protein production can be achieved so that number of the analytical methods investigated after expression could be increased.

- Deviations resulting from high methanol demand could be minimized so that more reliable screening of transformants could be achieved.

- Using vectors that allow secretion of heterologously expressed proteins could be used directly in some applications such as micro scale enzyme assays (Babiak and Reymond, 2005), functional binding assays (Zeder-Lutz et al., 2006), or even thermostability analysis using differential scanning calorimetry (Torres et al., 2004) or differential scanning fluorimetry (Niesen et al., 2007).

- Secreted proteins can be easily purified using existing protocols in high throughput manner without any cell disruption or buffer exchange interventions.

- Easily handled and robust induction system can be developed without growing cells to high densities before induction phase which is a necessary condition to get

reasonable expression levels when using Mut^{s} transformants as *P. pastoris* cells continue to grow feeding on non-repressing carbon sources.

We have used; Glucose as the repressing, Sorbitol, Mannitol- sugar alcohols, Trehalose-disaccharide and Alanine as the non-repressing carbon sources in our study. KM71H Mut^{s} cells containing pPICZα-ROL, pPICZα-BTGL and pPICZα-CALIPA vectors were used in the expression studies.

Microarray scale enzyme assays

After the first report of the microarray analysis of quantitative gene expression (Schena et al., 1995), more than seven thousand articles were published containing "microarray" in their title only. Usage of the high throughput and multiplex methods are increasing with developing technologies that, mostly results from other disciplines such as chemistry, electronics, photonics, bioinformatics etc. Even this technology was used for the gene synthesis and the quality assessment (Tian et al., 2004). Zhu et al. reported first protein microarray spotting 5800 overexpressed and purified S. cerevisiae proteins onto slides and they have identified new calmodulin and phosholipid interacting proteins (Zhu et al., 2001).

Tissue microarrays of tumor specimens has been developed, and *in situ* detection of DNA, RNA and protein targets were achieved (Kononen et al., 1998). Many immuno-based microarrays were developed using monoclonal antibodies to detect biologically important molecules from different samples. Cardiovascular biomarkers in human serum (Gul et al., 2007), cytokines from patient's serum (Huang et al., 2001), potential lung cancer biomarkers (Han et al., 2009), serum proteins associated with pancreatic cancer (Orchekowski et al., 2005).

Small molecules have been successfully used in the microarray applications. Diversity oriented synthesis was used to create structurally complex and diverse small molecules, and these molecules were printed onto glass slides, namely small molecule microarray (SMM) (MacBeath et al., 1999; Kuruvilla et al., 2002). Besides that kind of binding and inhibition screening studies, Zhu et al, achieved an enzymatic profiling on glass surface (Zhu et al., 2003). Recently, Babiak et al, showed that enzyme assays can be carried out on silica gel plates for lipase substrate profiling (Babiak and Reymond, 2005).

Screening of enzymes for improved properties is very important for both rational design and directed evolution approaches. Although rationally designed mutations,

insertions or deletions generally require a few numbers of mutants, it is still beneficial to use high throughput and/or multiplex enzyme assays.

We have developed substrate microarray strategies for enzyme assays. Our approach is coating TLC plates (microscope slides) with fluorogenic lipase substrates and incubating coated slides with spotted enzymes using microarray spotter. Less than 1 nl of enzyme solution was spotted onto silica gel slide coated with different lipase substrates (Figure 4.3). Although we have only demonstrated chain length selectivity and metal inhibition studies in this thesis, same silica gel surface can be investigated for other enzymes that have fluorescent substrates.

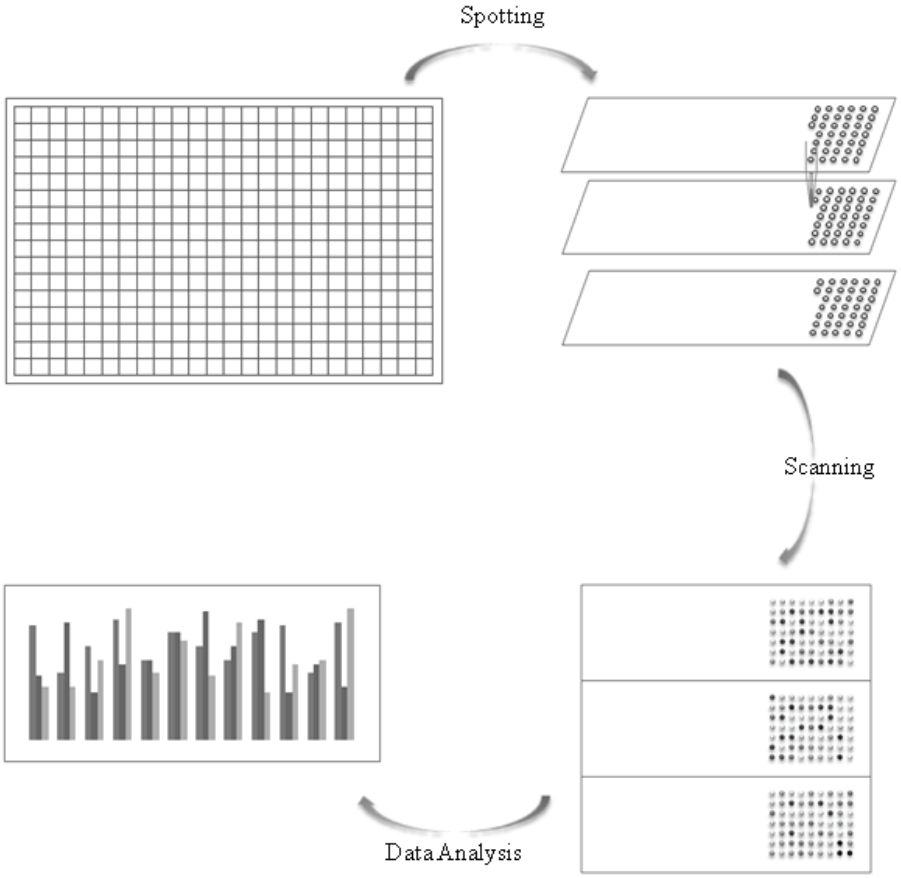


Figure 4.3 Microarray scale enzyme profiling

4.2 Background

4.2.1 *Pichia pastoris* Expression System

P. pastoris can easily be grown to high cell densities using defined minimal media (Higgins and Cregg, 1998). Commonly, tightly controlled and highly inducible promoter, alcohol oxidase 1 (AOX 1), controls the expression (Cereghino and Cregg, 2000). On the other hand, strains with a Mut^s phenotype have a non-functional AOX1 but a functional AOX2 enzyme. AOX2 enzyme has a lower expression level; accordingly methanol is utilized slowly giving rise to a "methanol utilization slow" phenotype (Cereghino and Cregg, 2000). Proteins can be expressed as intracellular or secreted products. Low levels of the native proteins secreted by the host cell and no added proteins in culture medium, secretion may be considered as a rough purification step. Secretion of heterologously expressed proteins is limited and generally, the proteins that are normally secreted by their native hosts are targeted to the secretory pathway. Although several different secretion signal sequences have been used successfully, the secretion signal sequence from the *S. cerevisiae* α-factor prepro peptide has been used with the most success (Cregg et al., 2000).

Several high throughput expression approaches have been reported using *P. pastoris*. Although Lang *et al.* has established a method for expressing proteins in microtiter plates (Boettner et al., 2002; Bottner and Lang, 2004), the deviations in expression levels have not been analyzed. Weis et al. wrote that "*but its value for comparative high-throughput screening for protein engineering has to be doubted due to the lack of data addressing standard deviations of expression levels across the whole plate*." about the previous research (Weis et al., 2004).

4.2.2 Overlap Extension PCR

Gene splicing by overlap extension is a technique that was developed to generate fusion proteins (Horton et al., 1989). Same group also introduced mutations into the gene using similar technique with mutagenic primers and overlap extension procedure,

namely, the site directed mutagenesis by overlap extension (Ho et al., 1989). After developing this technique, it has been used in diverse applications, such as exon fusion (Jeya et al., 2009), site directed mutagenesis (Simionatto et al., 2009). Also modified or improved versions of this technique has been applied, i.e., GC-rich overlap sequences improved the overlap hybridization (Cha-aim et al., 2009)

In separate reactions, the fragments are amplified with one flanking primer (a or d) and one internal primer (b or c). Overlapping fragments are generated (AB or CD) after the first PCR reaction, since the primers b or c overlap at their 5' end. When the fragments AB and CD are denatured and re-annealed, the complementary regions (added by primers b or c) can hybridize. In the second round PCR reaction, each overlapping strands acts as a primer on the other fragment and extension of overlap by the DNA polymerase results in a full length fusion product, AD (Figure 4.4).

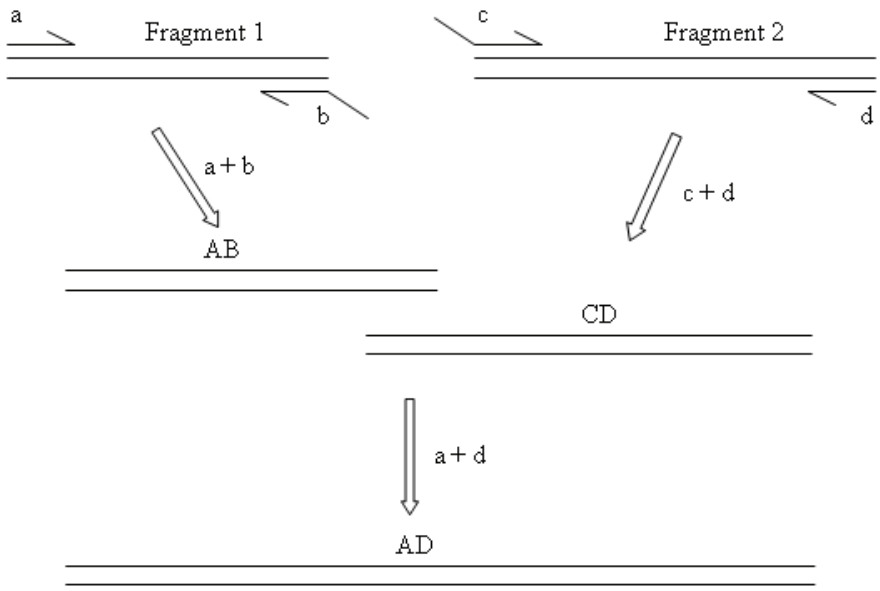


Figure 4.4 Gene fusion using Overlap Extension PCR. Fragment 1 and Fragment 2 were amplified separately using the primer pairs a / b and c /d. Products AB and CD were used as a template and the primers for each other in the second round PCR reaction. Primers (a / d) were also added to the mixture.

4.2.3 HT Expression Systems

Bacterial expression is the most widely used expression system for the production of the recombinant proteins (Baneyx, 1999). Recently, researchers from 14 different laboratories published a review paper describing consensus 'what to try first' strategy based on the analysis of the expression and purification of over 10,000 different proteins focused on structural genomics (Gräslund et al., 2008). pET vectors (Merck-EMD) are the most commonly used vector systems which T7 RNA polymerase promoter (Studier et al., 1990) and lac operator (Dubendorf and Studier, 1991) control the expression of recombinant proteins. Copy of T7 RNA polymerase gene was integrated into genomic DNA of λDE3 lysogen strains of *E. coli*. Under repressive conditions, this genome integrated gene which is under the control of lac repressor, does not produce T7 RNA polymerase and the transcription of target gene in vector is negligible (Dubendorf and Studier, 1991).

4.2.4 High throughput assays

Enzyme assays in the high throughput formats are essential tools for enzyme engineering and enzyme discovery. Recently, low volume enzyme assay format on silica gel plates for esterase and lipases were developed (Babiak and Reymond, 2005). Thin layer chromatography (TLC) plates were used as an assay platform. Small amount of enzyme solutions (~1 ul) were deposited onto TLC plates impregnated with fluorogenic substrates (Goddard and Reymond, 2004). Another high throughput enzyme assay was reported using indirect tagging strategy for product detection (Grognux and Reymond, 2006). Lipase substrate microarray was generated by printing aliphatic monoesters on bovine serum albumin coated glass slide and 1,2-diol products after enzymatic cleavage were converted to aldehyde using sodium periodate. On-chip generated aldehydes were marked with a fluorescent tag, rhodamine B sulfohydrazide.

High throughput detection methods are not limited to the activity assays, proteins which do not show any enzymatic activity can also be detected using antibodies. Zeder-Lutz et al. were applied dot-blot immunodetection method for G-protein coupled receptors (GPCR) expressed in *P. pastoris* (Zeder-Lutz et al., 2006). They were

detected Flag-tagged GPCRs from various samples, such as crude *P. pastoris* extracts to solubilized receptors.

High throughput technologies do not target only the functional properties only, structural properties of proteins or enzymes can also be analyzed. Recently developed technique, differential scanning fluorimetry (DSF), measures the fluorescence of a dye in which the intensity of emitted light changes upon interaction with its environment (Pantoliano et al., 2001). Lo et al. showed that DSF technique can be performed using available real-time PCR machines with appropriate filter sets (Lo et al., 2004), and detailed protocols for the screening of ligands that bind and stabilize the purified proteins (Niesen et al., 2007).

4.3 Materials

4.3.1 Instruments and Software

Instruments and software used in this chapter were listed in Appendix A along with their catalog numbers.

4.3.2 Chemicals

Chemicals used in this chapter were listed in Appendix B with their catalog numbers. All chemicals and enzymes were used without any further purification unless otherwise stated.

4.3.3 Enzymes

All enzymes were listed in Appendix C with their catalog numbers. Some enzymes were included in kit content; these kits were listed in Appendix D.

4.3.4 Kits

All kits were used in this study were listed in Appendix D with catalog numbers. Manufacturer's protocols were followed using these kits unless otherwise noted in Methods sections.

4.3.5 Oligonucleotides

Oligonucleotides used for oligo-mediated substrate immobilization studies were given Table 4–1. All oligonucleotides were purchased from Microsynth.. Oligonucleotides for cloning of *Bacillus thermocatenulus* Lipase (BTGL) and *Candida antarctica* Lipase A (CALIPA) enzymes are also given in Table 4–1. Oligonucleotide names and sequences for overlap extension PCR reactions to generate linear cassettes are listed in Table 4–1.

All oligonucleotides were designed using VectorNTI software (Invitrogen) and were purchased from Microsynth (CH) and used without any further purification. Oligonucleotides were resuspended in molecular biology grade water at 100 µM final concentrations and stored at -20 °C. Working solutions were prepared at 10 µM and aliquots were stored at -20 °C.

Table 4–1 Oligonucleotides for cloning and Overlap Extension PCR

Oligo Name	Usage	Sequence (5' to 3')
F_BTGL	Cloning	AGCTGAATTCGCGGCATCCCCACGCGCCA
R_BTGL	Cloning	TGTTCTAGAACAGGCCGCAAACTCGC
F_CALIPA	Cloning	GCTGAATTCGCGGCGCTGCCCAACCCCT
R_CALIPA	Cloning	TGTTCTAGAGGTGGTGTGATGGGGCCAAAG
F_partA	OE-PCR	CTAACATCCAAAGACGAAAGGT
F_partA_1	OE-PCR	CTCGCTCATTCCAATTCCTT
F_partA_2	OE-PCR	GAACCTAATATGACAAAAGCGT
R_partA	OE-PCR	GAATTCAGCTTCAGCCTCTCT
F_partC	OE-PCR	TCTAGAACAAAAACTCATCTCAGAAG
R_partC	OE-PCR	AGCTTGCAAATTAAAGCCTT
F_partD	OE-PCR	GAAGGCTTTAATTTGCAAGCTAGATCTAACATC CAAAGACGA
R_partD_1	OE-PCR	CTCCAATCAAGCCCAATAACTG
R_partD_2	OE-PCR	CAAGACAGCGTTTAAACTGTCA
F_ROL_OEPCR	OE-PCR	AAAAGAGAGGCTGAAGCTGAATTC
R_ROL_OEPCR	OE-PCR	CTTCTGAGATGAGTTTTTGTTCTAGA

4.3.6 Lipase Enzymes

Enzymes used for substrate and enzyme arrays studies are listed in Table 4–2. All enzymes were used directly without any further purification unless otherwise stated. Enzymes were dissolved in distilled water and impurities were removed by centrifugation. Enzyme concentrations were measured by the Bradford assay.

Table 4–2 Lipase enzymes used for microarray studies

	Lipases	Source	Catalog No.
AL	Lipase from *Aspergillus*	Sigma	84205
CAL	Lipase from *Candida antarctica*	Sigma	65986
CCL	Lipase from *Candida cylindracea*	Sigma	62316
MML	Lipase from *Mucor miehei*	Sigma	62298
PCL	Lipase from *Pseudomonas cepacia*	Sigma	62309
PFL	Lipase from *Pseudomonas fluorescens*	Sigma	95608
RAL	Lipase from *Rhizopus arrhizus*	Sigma	62305
RNL	Lipase from *Rhizopus niveus*	Sigma	62310
HPL	Lipase from hog pancreas	Sigma	62300
ROL	*Rhizopus oryzae* Lipase	*P. pastoris*	NA
BTGL	*Bacillus thermocatenulus* Lipase	*P. pastoris*	NA
CALIPA	*Candida antarctica* Lipase A	*P. pastoris*	NA
CALA	*Candida antarctica* Lipase A	ChiralVision B.V.	CV-CALAY
CALB	*Candida antarctica* Lipase B	ChiralVision B.V.	CV-CALBY
CRL	*Candida rugosa* Lipase	ChiralVision B.V.	CV-CRL1

4.4 Methods

The detailed methodology for the linear cassette transformation, the high throughput enzyme expression using the non-repressing carbon sources and the production protocols, assay settings and storage conditions for enzyme arrays are given in this section.

4.4.1 Vector construction

pPICZαA-ROL was a gift from Francesco Valero (UAB, Spain) and Rolf D. Schmid (Universität Stuttgart, Germany) (Minning et al., 1998) and all DNA manipulations were explained in the Chapter 3.

1214 bp fragment from genomic DNA of *Bacillus thermocatenulatus* (DSM 730) mature lipase (without 28-amino acid signal sequence) was amplified using the oligonucleotides, F_BTGL and R_BTGL, with EcoRI and XbaI restriction enzyme (RE) sites, respectively. After RE digestion, the fragment was cloned into pPICZαA backbone, cut by EcoRI and XbaI and dephosphorylated with alkaline phosphatase in order to prevent self ligation.

For *Candida antarctica* lipase gene, the same methodology were followed, but EcoRI and XhoI restriction enzymes were used. Genomic DNA was extracted from *Candida antarctica* (DSM 70725) and directly used as a template in the PCR reaction. Although there is no DNA sequence for CALA enzyme in the NCBI genomic database, we have modified the primers according to the published article (Pfeffer et al., 2007).

All constructs were sequenced using F_alpha and R_AOX primers (Table 3–4) and compared to the relevant database. Accession numbers for BTGL are GenBank: CAA64621.1 and PDB code: 1TCA. For CALIPA, we have compared the cloned sequence with recently released PDB structure (PDB code: 2VEO, Jan 25, 2008) and have found 100 % similarity.

We have found only one mutation in cloned BTGL gene, which was C355Y according to CAA64621 sequence (Jan 24, 1996). Recently, the X-ray diffraction data of the BTGL was released (PDB code: 2W22, Jun 18, 2009) and it has W (tryptophan) at residue 355. As the BTGL enzyme retains its lipase activity and we did not observe any irregularities in its expression in both *P. pastoris* and *E. coli*, we have used this enzyme.

4.4.2 Overlap Extension PCR

Three different parts of the linear cassette, i.e. the promoter part, the gene of interest part and the resistance gene (zeocin) part, were amplified separately in different

PCR reactions. After purification of the PCR products using the Gel Extraction Kit (Qiagen), the overlap extension PCR amplifications were performed. After the first 15 cycles, primers were added to the PCR reaction with the polymerase, the buffer, and the dNTP mixture as the final concentration of 1X.

4.4.3 *P. pastoris* transformation

The protocols in Section 3.4.10 were followed without any changes for transformation of pPICZαA vectors harboring BTGL, ROL and CALIPA.

For the linear-cassette transformation, various amounts of the PCR amplified cassettes were transformed into the *P. pastoris* KM71H strain using the same protocol.

4.4.4 Rhodamine Plate Assay

After transformation, YPD-Zeo plates were incubated at 30 °C and randomly selected colonies were transferred onto Rhodamine-MM plates. Rhodamine plates were prepared according to the given protocol (Section 3.4.12). Fluorescent halo formations were followed by UV lamp.

4.4.5 48-well growth and expression

In order to ensure the preliminary requirements of the high throughput cultivation and the expression *P. pastoris* KM71H holding BTGL clone was used. A 48-deep well storage (Abgene, US) was used. 2 ml of medium was inoculated into each well in order to provide air space for the microbial growth. For the determination of the effect of the glass beads, 5 mm in diameter, on shaking cultivation was performed with 3 glass beads in each well. The incubation conditions for the plate were adjusted to 260 rpm at 30°C. The plate was covered with a breathable (AeraSeal, RPI corp.) adhesive seal in order not to limit oxygen amount needed for the growth. The regarding strain was grown in YPD in shake flask and required amount of the cells were harvested and inoculated into the 48-deep well plate as resuspended in BMM. The starting optical density was 0.25 at 600 nm. 200 µl of sample was taken from each well and each well was fed with increasing concentration of sorbitol and 1% of methanol unless otherwise stated. 2

different sampling intervals; 12 hours and 24 hours were tried. In order to validate the optimized parameters for the high throughput cultivation and the expression we have also applied this system to the remaining two clones ROL, CALIPA. Sampling was allowed for 36 hours and 48 hours for 12-hour-interval and 24-hour-interval-sampling respectively. The samples were centrifuged at 4000 rpm for 5 minutes. The supernatants were kept at -20°C for further enzymatic assays. Apart from the 48-deep well plate single trial was also performed to verify the correct relationship between the plate well number and the volume of the medium. 96-well microplate of the same brand was used for 1 ml of the clone BTGL inoculums. The sampling volume and fed volume were also adjusted accordingly.

4.4.6 ELISA reader OD measurements

The OD measurements using ELISA reader depends on several parameters. The volume of the solution and the shape of the wells directly effects the measurement, therefore, we have only used same brand of 96-well plates for OD 590 measurements. In order to normalize OD measurements obtained using ELISA reader, we have compared the OD values of serial diluted *P. pastoris* cells using 96-well plate and standard 1 ml plastic cuvette with 10 mm path length. All OD measurements were normalized according to the linear fit slope of the comparison data (Figure 4.5)

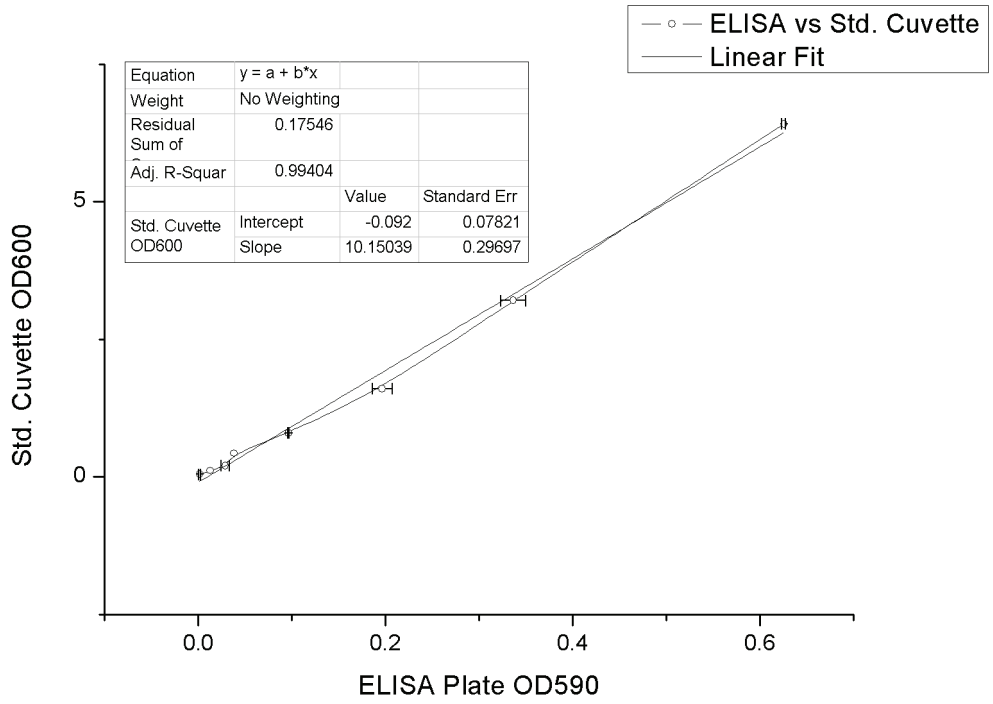


Figure 4.5 Normalization of the OD values using 96-well plate

4.4.7 DNS Glucose Assay

The glucose assay was performed using a published protocol (Xiao et al., 2005), with slight modifications. Briefly, 100 µl of cell free medium of centrifuged samples were mixed with 100 µl of DNS assay reagent (1% (w/v) 1,3-dinitrosalicylic acid, 0·05% (w/v) sodium sulfite, 1% (w/v) sodium hydroxide, 10% (w/v) sodium potassium tartrate, 0.2 % phenol) and incubated at 95 °C for 10 min using thermal cycler in 96-well plates. 100 µl of heated samples were transferred to round bottom 96-well plates and OD measurements were taken at 540 nm. For standard curve generation, serial dilutions of 2 % glucose containing BMGlu were prepared and assayed.

4.4.8 Protein purification using His-plate

After the expression of recombinant BTGL enzyme with Histidine tag, cells were centrifuged and the supernatant containing recombinant proteins were collected. Briefly, 48-well plates were centrifuged at 4500 rpm with swing-out rotor at 4 °C for 15 min, and the supernatants were transferred into another 48-well plate. 8X washing buffer and 1M imidazole were added to each well to final concentrations of 1X washing buffer and 40 mM imidazole.

1X washing buffer and 40 mM imidazole solution were used to equilibrate the 96-well his-plate, and the protein solutions from previous step were applied onto his-plate. After incubation at room temperature for 30 min, plates were centrifuged and flow-through was discarded. After three consecutive washes using 1X washing buffer with 40 mM imidazole, captured proteins were eluted in elution buffer (1X washing buffer with 500 mM imidazole). Purified proteins were subjected to the fluorescent lipase assay, zymogram and SDS-PAGE analysis.

4.4.9 Enzyme Microarrays

In this section, detailed protocols about preparation of enzyme arrays were given.

4.4.9.1 TLC microscope slide coating

The substrates were dissolved in Methanol or DMF at 50 mM concentration as a stock solution and kept at -20 °C until use. 1 mM concentrations of each substrate were prepared in dry DCM. The slides were incubated in glass petri dishes with 1 mM solutions for 10 min. at room temperature with agitation or until uniform color change has been developed by visual inspection. Substrate adsorbed slides were air-dried for 10 min. and kept in dark under argon gas until use.

4.4.9.2 Enzyme Spotting and Assays

Enzyme stock solutions were prepared by dissolving powders in distilled water at 1 mg/ml concentration and aliquots were stored at -20 °C until use. Two-fold serial dilutions of enzymes were prepared in different reaction buffers. Actual reaction buffers and enzyme concentrations were discussed in Section 4.5.

Two-fold diluted enzymes were dispensed into 384-well plate (Telechem Int.) and briefly centrifuged before loading into Omnigrid Accent Array Spotter. 30 % humidity level was kept stable during spotting process. 4 ms of spotting time (touch) was used for all microarray scale enzyme assays.

4.4.9.3 Slide Scanning and Data Analysis

Slides were scanned directly after the printing procedure. Fuji LAS-4000 Imaging System was used for scanning silica gel slides. Ethidium bromide filter was selected and different exposure times were optimized for each slide. Images were saved as TIFF format at 16 bit as a one color. Slides were analyzed using ArrayWorx scanner's software after converting image files into necessary input file format for ArrayWorx software.

4.4.9.4 Substrate Design and Synthesis

Besides using commercially available 4-methylumbelliferyl esters listed in Table 3–6, we have also used custom synthesized substrates, Figure 4.6. Substrates were synthesized by Advanced Chemtech/Thuramed (KY, USA).

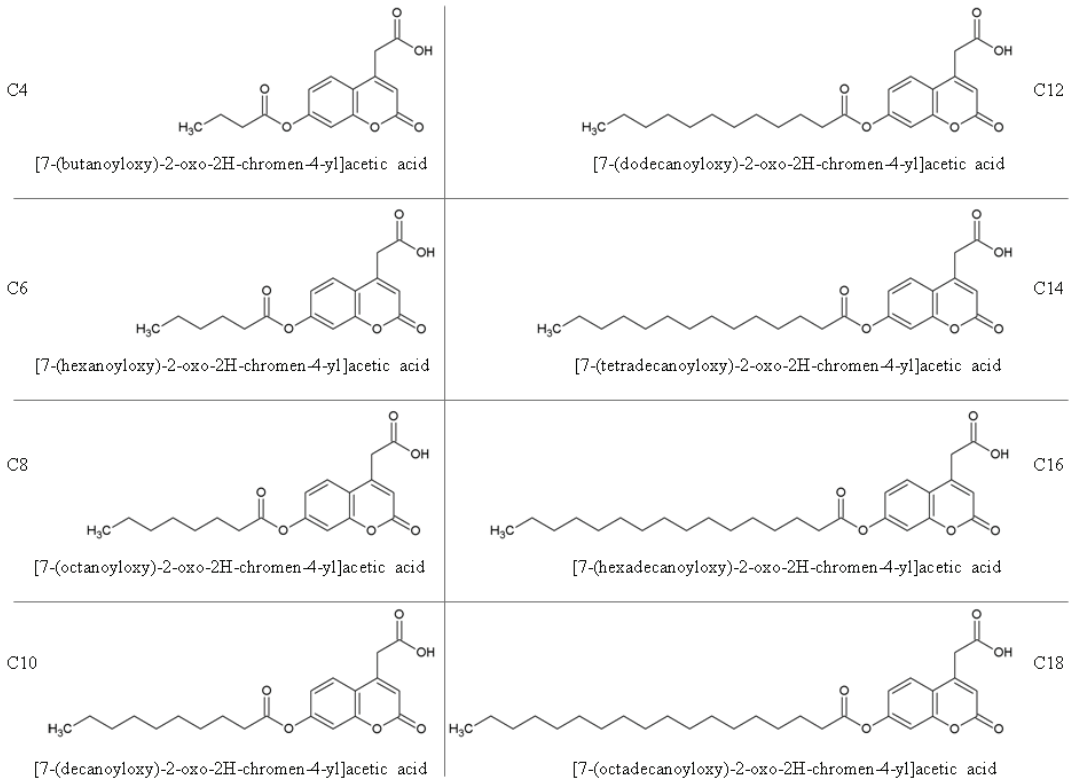


Figure 4.6 Custom synthesized substrates (chain length of 4 to 18) with carboxylic acid functional group.

4.5 Results and Discussion

In this section experimental results of the high throughput techniques developed in this thesis were presented and discussed.

4.5.1 Linear Cassette Transformation

We have designed different primer sets to amplify expression cassette parts. They were annotated as Part_A1: whole AOX promoter, Part_A2: a part of AOX promoter from SacI RE site which is generally used for cutting site to transform whole plasmid approach, Part_A3: exactly the half of the promoter for promoter part of expression cassette. Part_C: the whole zeocin containing part. Part_D1: the remaining 5' AOX

promoter which was not included in Part_A2, Part_D2: remaining part of 5' AOX promoter which was not included in Part_A2. We have first amplified the parts using the appropriate primers listed in Table 4–1, and the gel extraction procedure of each was performed before using the parts in OE-PCR. After obtaining parts A1, A2, A3, B, C, D1, and D2, A and B parts were amplified together to create A1-B, A2-B and A3-B. C-D1 and C-D2 were amplified and the gel extraction was done similarly (Figure 4.8).

Finally whole linear expression cassettes were amplified using these fragments and A1-B1-C, A1-B2-C, A2-B1-CD1, A2-B2-CD1, A3-B1-CD2, and A3-B2-CD2 were obtained (B1 and B2 were the purified and non-purified PCR products of gene) (Figure 4.9).

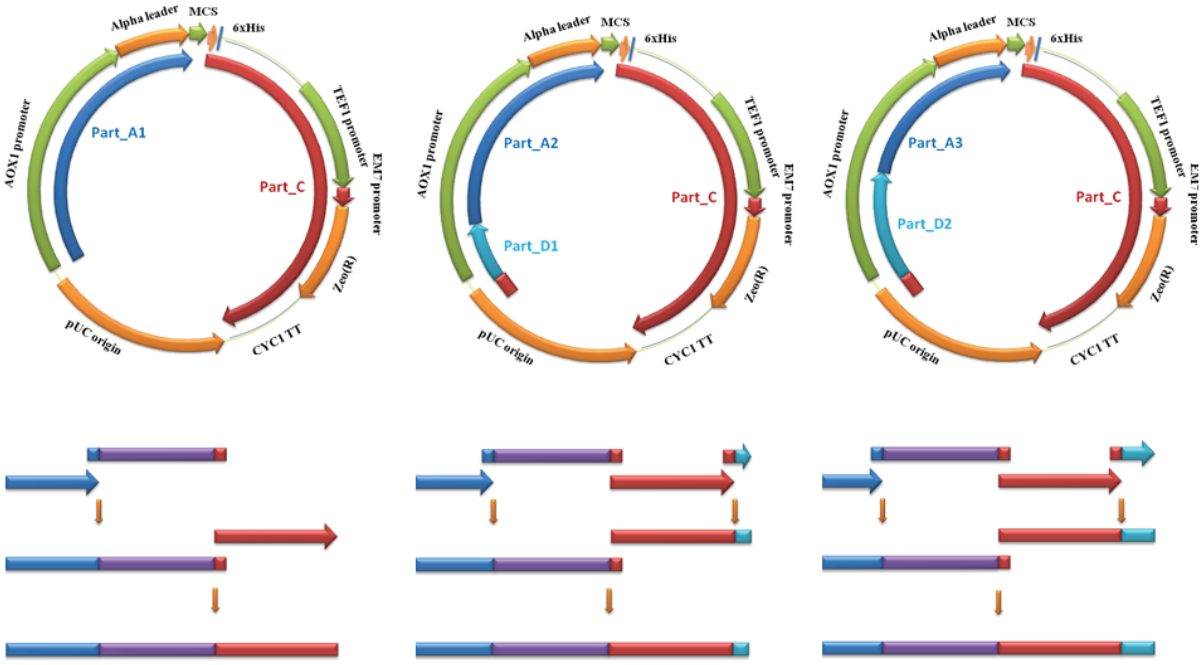


Figure 4.7 Expression cassettes generated from different parts of the pPICZαA vector.

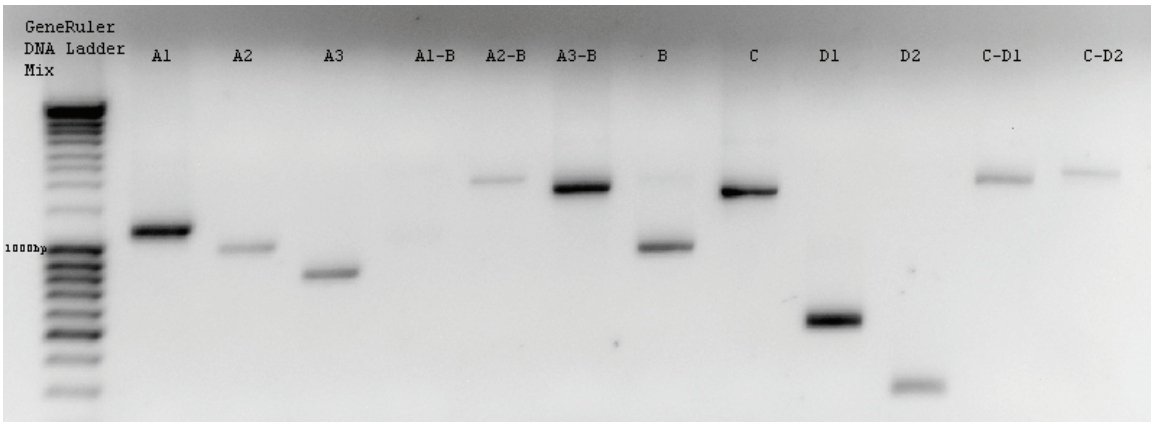


Figure 4.8 PCR amplification of the OE-PCR components (A1, A2, A3, A1-B, A2-B, A3-B, B, C, D1, D2, C-D1, C-D2).

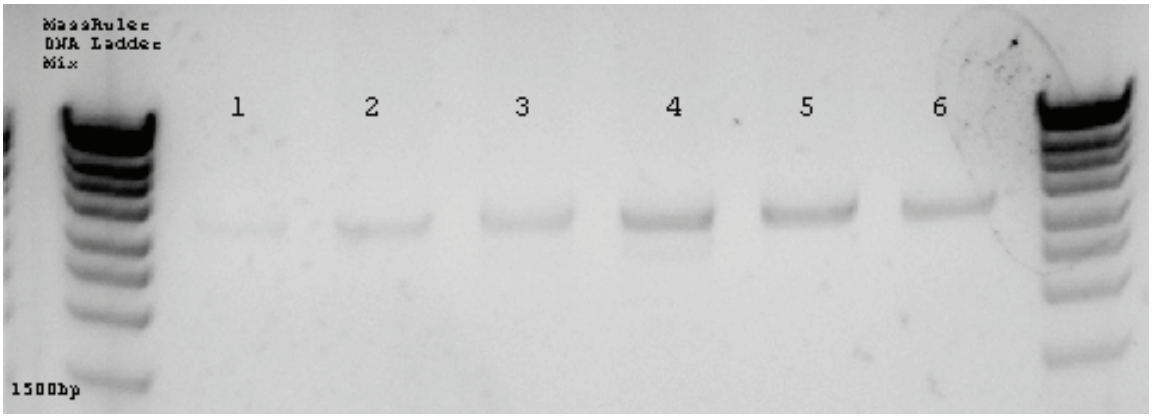


Figure 4.9 Overlap-extension PCR generated linear cassettes after purification. 1ul was loaded to agarose gel for conformation, whereas 5ul was used for transformation. The DNA fragments are as follows: 1. A1-B1-C, 2. A1-B2-C, 3. A2-B1-CD1, 4. A2-B2-CD1, 5. A3-B1-CD2, 6. A3-B2-CD2.

Transformation of different linear cassettes (Figure 4.9) into *P. pastoris* gave different number of colonies on YPD-Zeo plates (Table 4–3). Randomly selected colonies were transferred onto Rhodamine-MM plates to Transformants were analyzed using rhodamine plate assay (Figure 4.10). A2-B-CD1 colonies gave similar halo sizes, which is directly proportional to the copy number of expression cassette, comparing to the A3-B-CD2 and A3-B-CD2 transformants. Although copy number of expression cassette variations among these colonies has not been analyzed, our results clearly showed that probability of finding high protein expression colonies are increasing when full length of the AOX promoter sequence was used in linear expression cassette.

Unfortunately, we could not able to transform other linear expression cassettes (Figure 4.7).

Table 4–3 Transformation and plate-expression results of the OE-PCR generated cassettes.

	YPD-Zeo	Rhodamine-MM (37 randomly Selected Colonies)			
		High Secretion	Low Secretion	No Secretion	No growth
Plate 4	58	31	2	-	4
Plate 5	92	23	5	3	6
Plate 6	167	22	-	9	6

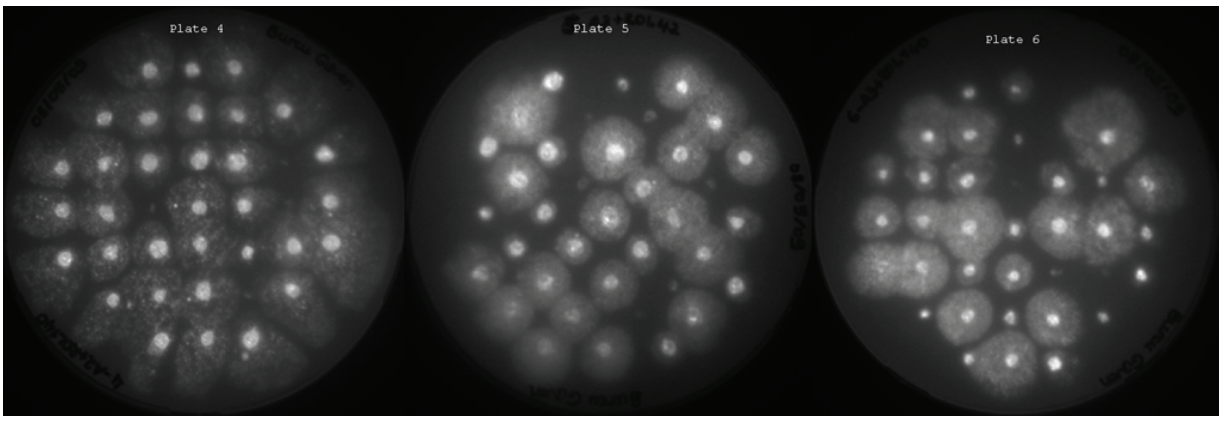


Figure 4.10 Rhodamine plate assay for the linear cassette transformed *P. pastoris* cells. Random colonies chosen from the transformation plates 4 (A2-B-CD1), 5 (A3-B-CD2), 6 (A3-B-CD2) on Rhodamine-MM assay plates visualized under UV exposure. Fluorescent halo formation (represented as white color) around colonies (intense white colored) shows lipase activity.

My results on linear cassette transformation clearly showed that overlap-extension PCR could be used to generate *P. pastoris* transformants and this method is compatible with high throughput transformation.

Besides being easy and efficient the transformation method can be coupled with the PCR generated mutant libraries, such as, random mutagenesis or site saturation mutagenesis. Even this methodology can be used with the mega-primer generated site directed mutagenesis to produce mutant proteins.

Copy number variation of integrated cassettes in *Pichia pastoris* genome can vary, therefore each candidate colony should be sequenced to determine the exact sequence of expressed protein. In case of mutagenesis driven libraries, sequencing can give exact mutation points in each candidate colony. Although it seems possible to obtain more than one different mutations in the same colony, sequencing and re-cloning will be much easier compared to the job has been performed by *P. pastoris* after screening procedure.

4.5.2 Effect of Carbon source for HT Expression in *P. pastoris*

4.5.2.1 Glucose Depletion

As glucose is a repressor for AOX promoter, we have measured glucose concentration in growth medium (BMGlu), to determine the induction time without changing growth media before expression. For this purpose we have used DNS colorimetric glucose detection method. We have followed the glucose concentration in growth medium which was inoculated with different starting ODs. After 20 hour, all glucose in 2 ml growth medium has been consumed (Figure 4.11, Figure 4.12, Figure 4.13) for all three different clones. Following these results, all induction experiments were started after 18 hour initial growth phase.

The concentration of the glucose was adjusted to 2% in the initial medium. The DNS method was used to measure the glucose amount in the medium as described in Section 4.4.7. A standard curve was initially generated to assign the upper and lower limits of the measurable glucose concentration for ELISA reader. The three clones, BTGL, ROL, and CALIPA were cultivated in the same 48-deep well plate with 4 different initial OD600 values; 0.05, 0.1, 0.2, and 0.4. After inoculation all of the wells were subsequently sampled for DNS method. 100 µl of inoculums from each well were added 100 µl of DNS reagent and incubated 5 minutes at 95°C. 100 µl of reaction mixture was measured in a high throughput manner in ELISA plate reader at 590 nm. For 3 different time points same plate was sampled and DNS method was carried out in the same manner. For all the samples measured for the glucose concentration the optical density at 590 nm was also measured in the ELISA reader with the required dilutions. The values that are found to be in between the limits were then converted to glucose

concentration values using DNS standard curve. The values obtained from the cell density measurements and the glucose concentration values of the same wells were used to analyze the glucose depletion time line graphically in addition to the particular cell density at the depleted glucose time. These experiments were repeated 3 times.

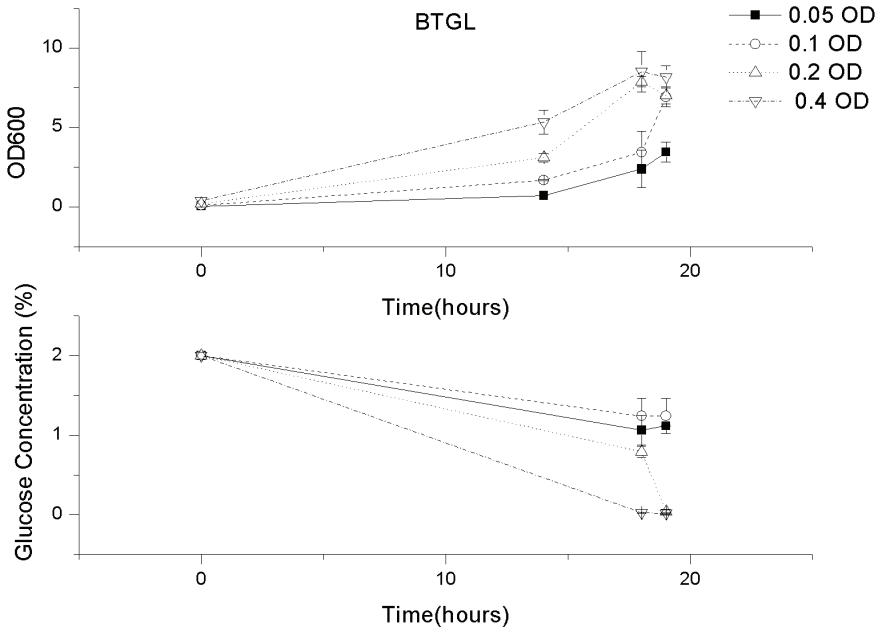


Figure 4.11 Glucose consumption and cell growth for BTGL (*Bacillus thermocatenulatus* Lipase) clone, Section 4.4.1.

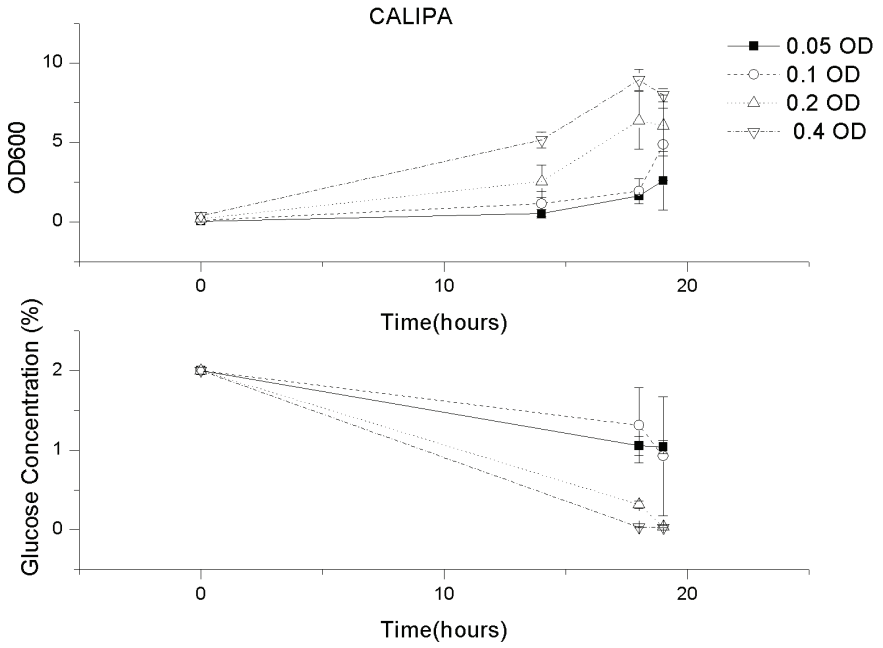


Figure 4.12 Glucose consumption and cell growth for CALIPA (*Candida antarctica* Lipase) clone, Section 4.4.1.

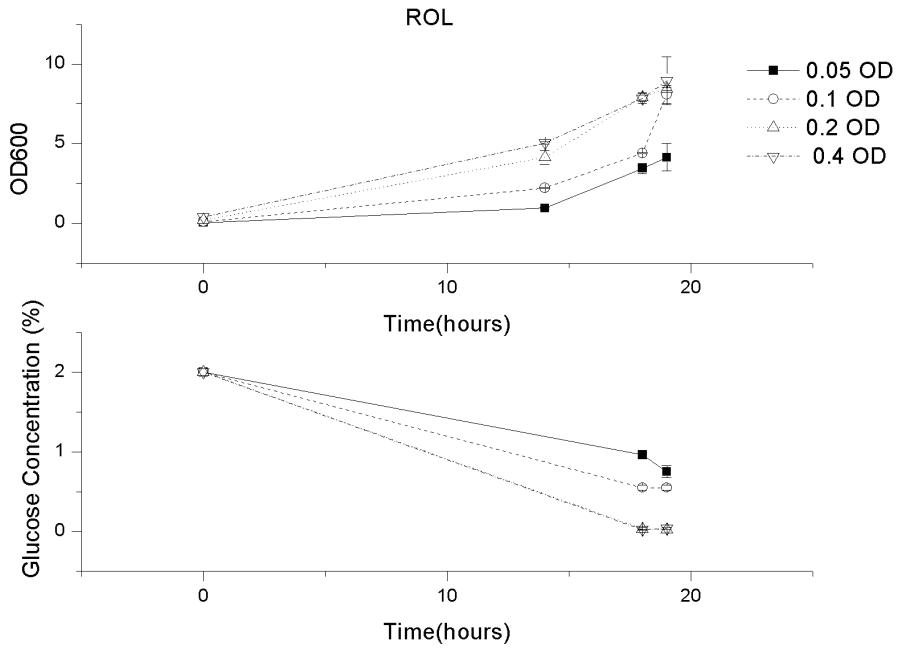


Figure 4.13 Glucose consumption and cell growth for ROL (*Rhizopus oryzae* Lipase) clone, Section 4.4.1.

4.5.2.2 Growth variations

In order to determine the effect of 48-well plate on growth characteristics of different clones, we have performed a series of experiments using the same overnight starter culture for each clone in every well of plate. We have measured the OD values at 18th hour of initial growth phase, as we have found that all glucose was consumed at that time. We have found that nearly all clones reached the same OD values after the initial growth phase, so that variations in expression levels could be minimized using this strategy (Figure 4.14).

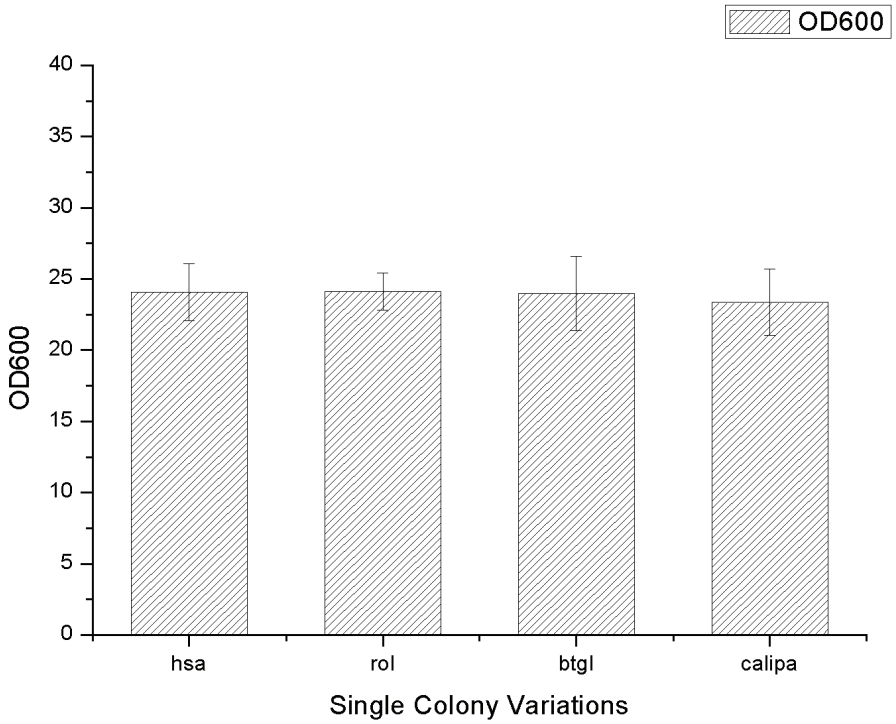


Figure 4.14 Variations in OD values after the initial growth phase (20th hour). Human Serum Albumin (Invitrogen control plasmid), ROL (*Rhizopus oryzae* Lipase), BTGL (*Bacillus thermocatenulus* lipase), and CALIPA (*Candida antarctica* Lipase A) containing *P. pastoris* clones were used for all studies.

P. pastoris KM71H cells harboring clones ROL, CALIPA and BTGL were grown on zeocin containing YPD plates. 12 single colonies were selected for each of them and inoculated into 2 rows of the 48-deep well plate (Abgene) filled with 2 ml of buffered minimal media without methanol with 2% glucose. The media in the first 2 rows were

either sterile or inoculated with *P. pastoris* harboring HSA clone to test the intrinsic hydrolysis of the secretion medium. After the colonies were resuspended in the media the initial optical density of the all wells were measured at 600 nm and recorded in a high throughput manner using ELISA reader (Biorad). The plate was covered with a breathable lid and allowed to shake at 260 rpm at 30°C. For 3 different time points 50 µl of samples were taken, diluted sufficiently not to go below or above the best line plotted for the ELISA measurement against Spectrophotometric measurement (Section 4.4.7). The values were then converted to cell density values using the standard curve. The variation in the cell density for each clone was calculated. These experiments were repeated 2 times.

4.5.2.3 Effect of non-repressing carbon source on expression

As our methodology was based on high throughput expression, we have investigated the effect of non-repressing carbon sources on expression of three different lipases in 48-well plates. Cells were grown in 2 % Glucose supplemented BMGlu medium for 18 hours which was determined as growth phase as all glucose was consumed. Our aim on the other hand was allowing the cells to consume all the repressing carbon source glucose and then starting an induction with methanol. During this induction apart from abandoning the culture without any carbon source we have targeted a variety of carbon sources that would not repress the AOX promoter but will enhance the growth. Therefore we have narrowed down our search from the previous work by Inan et.al. (Inan and Meagher, 2001). 5 different carbon sources; glucose as repressing, sorbitol, mannitol- sugar alcohols, trehalose-disaccharide and Alanine as non-repressing were investigated. 48-deep well plate was inoculated with single colonies selected from three different *P. pastoris* transformants; BTGL, ROL and CALIPA. The cultivation conditions were carried out as described in the previous two sections (Sections 4.5.2.1 and 4.5.2.2). The glucose was supplied in 2% for whole plate. After 18 hours the sampling and induction was initiated. After each sampling the OD600 was measured for the whole plate (Figure 4.16). The methanol and the carbon source were fixed to 1% for whole plate. The lipase activities of the supernatants of the whole samples were quantified using 4MU-Caprylate as substrate (Figure 4.15).

The clone ROL showed increased activity with the 1% sorbitol and with 1% trehalose. Since the other two cloned did not show any increase in the 1% trehalose the carbon sources apart from sorbitol were concluded as ineffective. Moreover for the clones BTGL and CALIPA the culture supplied with sorbitol at the concentration of 1% was monitored to show same total activity of that of no carbon source added culture. Nevertheless the finding at first sight would be thought to be a conflict with the prior findings of Inan et.al. but we have used *P. pastoris* KM71H strain in other words Mut^{s}. The prior findings however were on Mut^{-}. Our strain of interest is capable of methanol utilization but slow. Therefore our strain is able to utilize methanol as carbon source via its slow AOX2 promoter as an inducer via by expressing our clones with AOX1. The sorbitol is sufficient enough to assist expression as a non-repressing source for the particular promoter. Also it is known that some over expressed enzymes has toxic effect on cells. On the other hand, I have found that all carbon sources increased the cell density during the induction phase (Figure 4.16). Although this finding was not directly correlated with expression level, Methanol concentration might be the rate limiting step in protein expression. Since our aim was to develop a methodology for multitier plate *P. pastoris* expression supplemented with non-repressing carbon source, we have selected sorbitol as a non-repressing carbon sources for further studies.

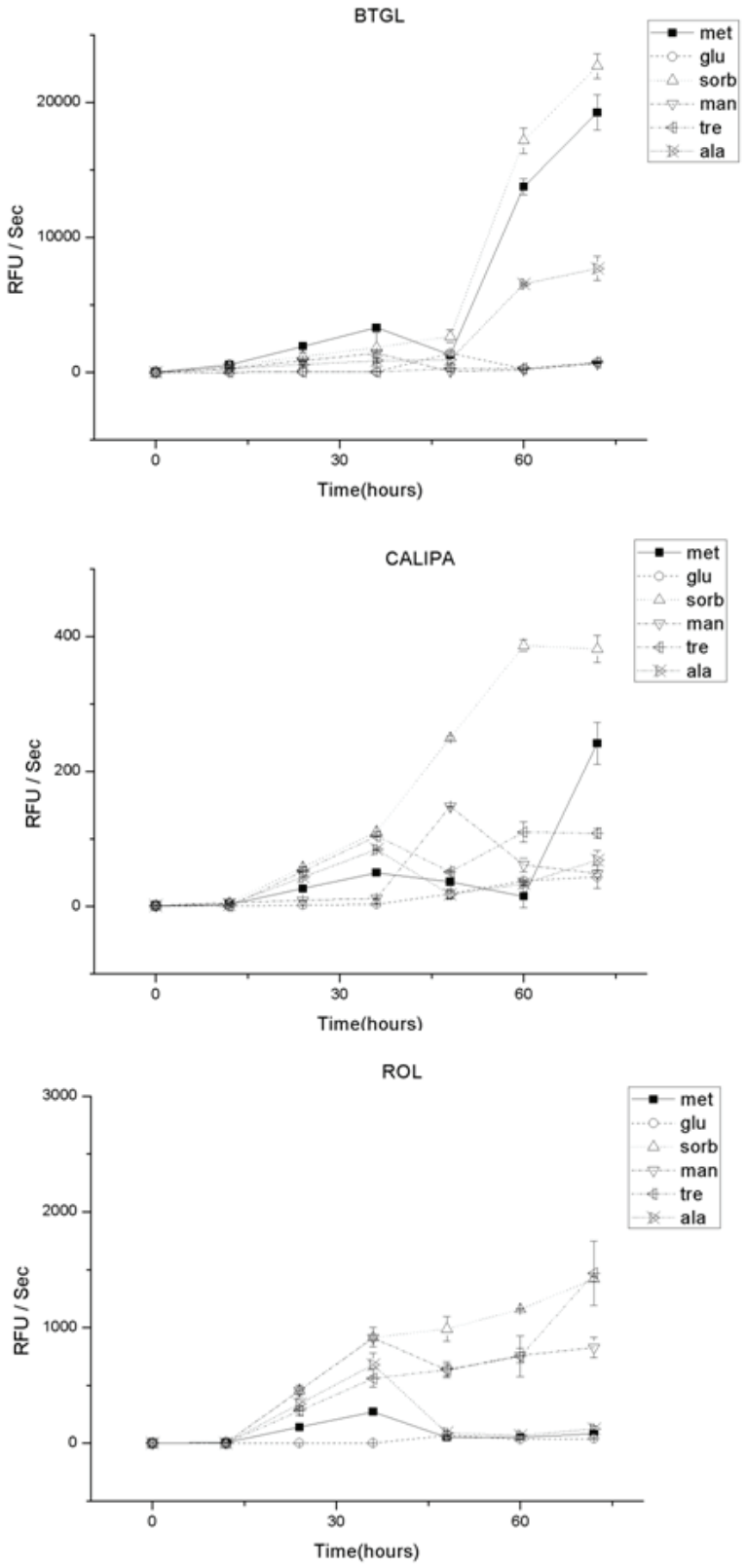


Figure 4.15 Effect of the non-repressing carbon sources on the enzyme activity. Methanol (met), Glucose (glu), Sorbitol (sorb), Mannitol (man), Trehalose (tre), and Alanine (ala) were used as carbon sources at 1 % concentration in BMM (1 % Methanol) medium.

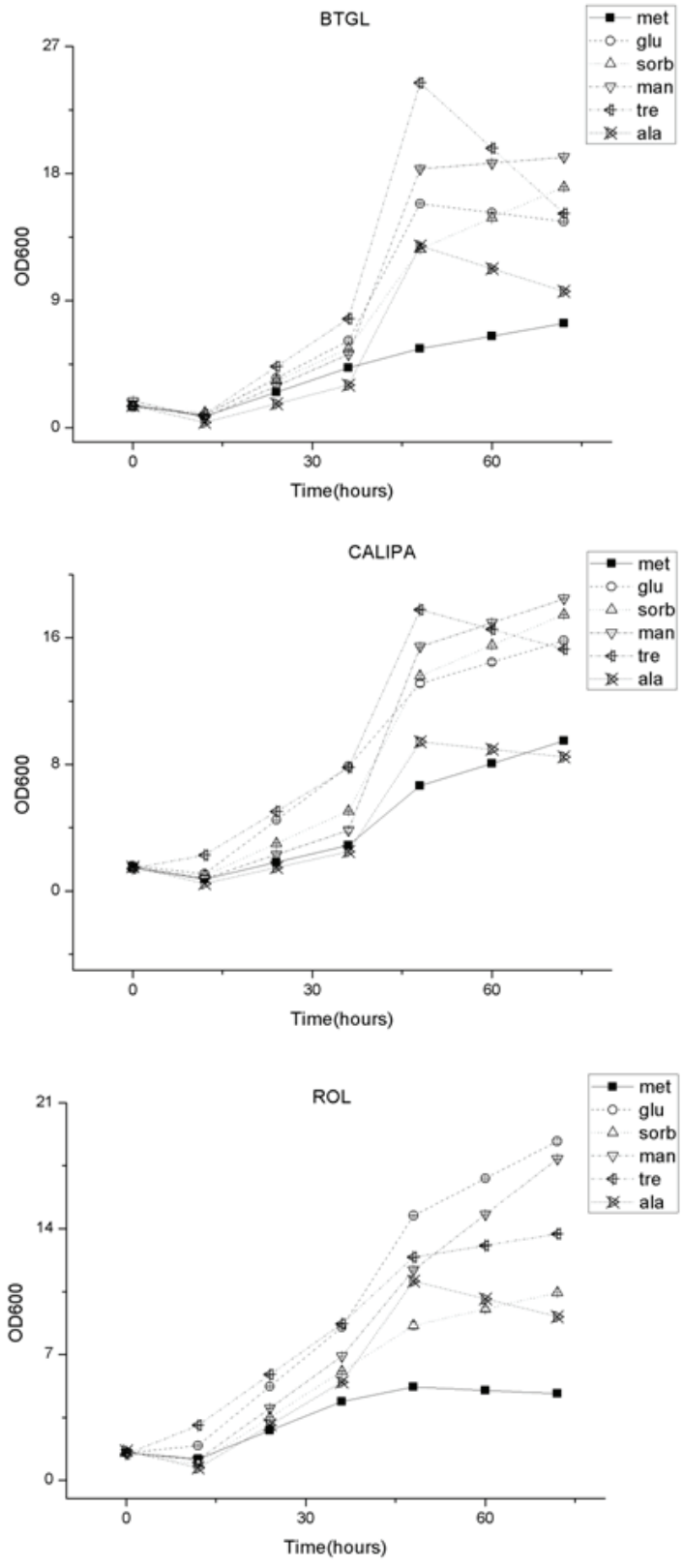


Figure 4.16 Effect of the non-repressing carbon sources on the cell growth. Methanol (met), Glucose (glu), Sorbitol (sorb), Mannitol (man), Trehalose (tre), and Alanine (ala) were used as carbon sources at 1 % concentration in BMM (1 % Methanol) medium.

4.5.2.4 Effective Sorbitol Concentration for HT Expression

We have shown that all non-repressing carbon sources are not effective in terms of lipase activity during induction phase except sorbitol. Sorbitol on the other hand increased the total enzyme production by 50% by comparing activity levels of other carbon sources in the BMM medium. Since the first section of the described work presents a detailed optimization procedure for the high throughput system based on our approach. The sorbitol concentration needs to be optimized as well. The sorbitol has chosen from the 5 different carbon sources, 4 of which non-repressing from the previous section was solely used to monitor expression of the all three clones. The cultivation approach and sampling was performed in the same manner. The OD600 values are shown in the Figure 4.17. The sorbitol was used in increasing concentrations starting from 0 % to 2.5 %. The methanol was again fixed to 1%. Yet again the total activities of the clones were different (Figure 4.18). The clone ROL showed an increase with an increase in the sorbitol concentration. Over the concentrations of 0.5% regardless of the sorbitol amount the total activity was increased to a maximum value and fixed. The presence of sorbitol is enough to increase the total activity obtained in a high throughput manner. For the other clones the sorbitol concentration 0.5 % is the most effective concentration. The sorbitol supply higher than 0.5% is not better than the bare methanol fed culture. As described by Granot et.al. (Granot et al., 2003) glucose presence may harm and force them to undergo apoptosis a similar mechanism induced by excess sorbitol concentration may harm the cell metabolism and thereby affecting the expression.

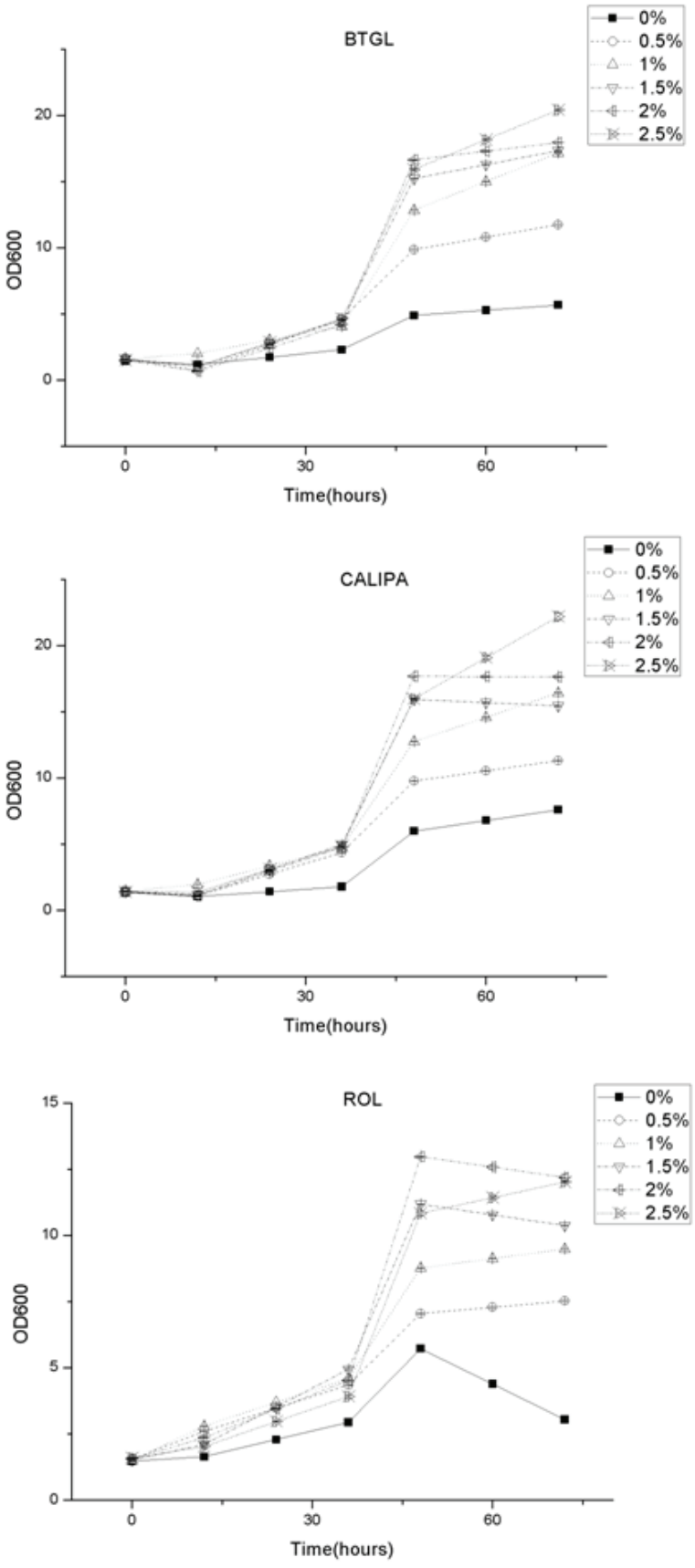


Figure 4.17 Effect of different Sorbitol concentration on the cell growth

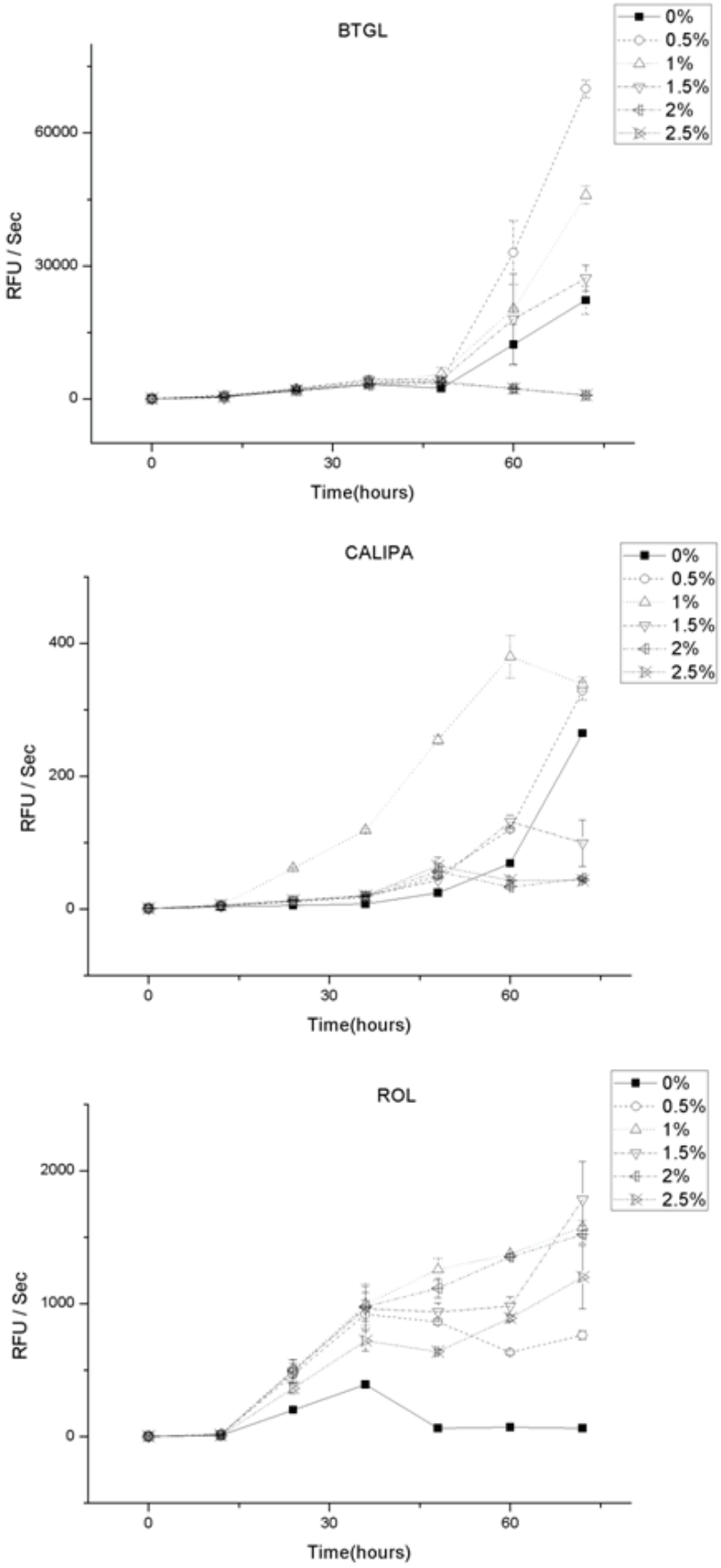


Figure 4.18 Effect of different Sorbitol concentration on the lipase activity

4.5.2.5 Effective Methanol Concentration for HT Expression

Unless otherwise stated the expression was optimized using fixed 1 % methanol concentration. The amount of the methanol supply is another critical factor for the expression quantity. Therefore we aimed to investigate the effective methanol concentration. Having used Mut^s strain we were aware of the importance of the methanol for the expression. According to the our findings for the clones BTGL and CALIPA the methanol when fixed to 1% the sorbitol supply of 1% does not affect much the total activity. One plausible reason would be excessive methanol supply. Sorbitol would be ineffective to assist expression quality. Moreover due to the ability of the strain of our study to utilize methanol in a slow rate the excess supply of methanol itself might promote the expression. Therefore we have investigated 4 different methanol concentrations 0 %, 0.5 %, 1.0% and 1.5 % against increasing sorbitol concentration for the clone BTGL only (Figure 4.20). Interestingly we have encountered a corroborative result to our first hypothesis. The methanol was seemed to be excess for the clones BTGL since the highest expression was monitored at the methanol 0.5% and sorbitol 0.5%. The cultivation plate was contained the BTGL clone only. The methanol concentration changing from row to row and sorbitol concentration from column to column enabled us to work with duplicates (Figure 4.19). The total activity where there is no methanol addition we could observe non-specific hydrolysis of the 4-MU substrate. Sorbitol when supplied in 0.5% regardless of the methanol concentration the lipase activity showed an increasing fashion. Interestingly as the methanol concentration increasing to 1% and 1.5% the sorbitol effect on expression behave same. As the concentration of the carbon source increases to 1.5% and higher values the activities have been monitored to be descending. The aforementioned decline is more explicit in the methanol concentration of 1.5%. Methanol concentration and sorbitol concentration

though being a non-repressive carbon source have been affecting one another. Higher supply of sorbitol than 1% would lead to the undesired loss of total yield in case of higher methanol concentrations. Nevertheless the higher methanol and sorbitol concentration when encounter in medium would affect the protein yield we have also observed silent effects of sorbitol concentration on methanol 1% for the clone ROL. The methanol concentration is another critical factor that has been shown to be affected by the media size and the presence of other carbon sources as well as the expression by-product. The employment of the Mut^s strain reasoned the critical importance of the methanol. The strain at the same time utilizes the methanol as a energy sources and recruits to initiates the transcription. The high density growth and over-expression are both well known features of *P. pastoris*. Therefore there should be fine tuning of the methanol employment as a carbon source and/or a transcription initiator.

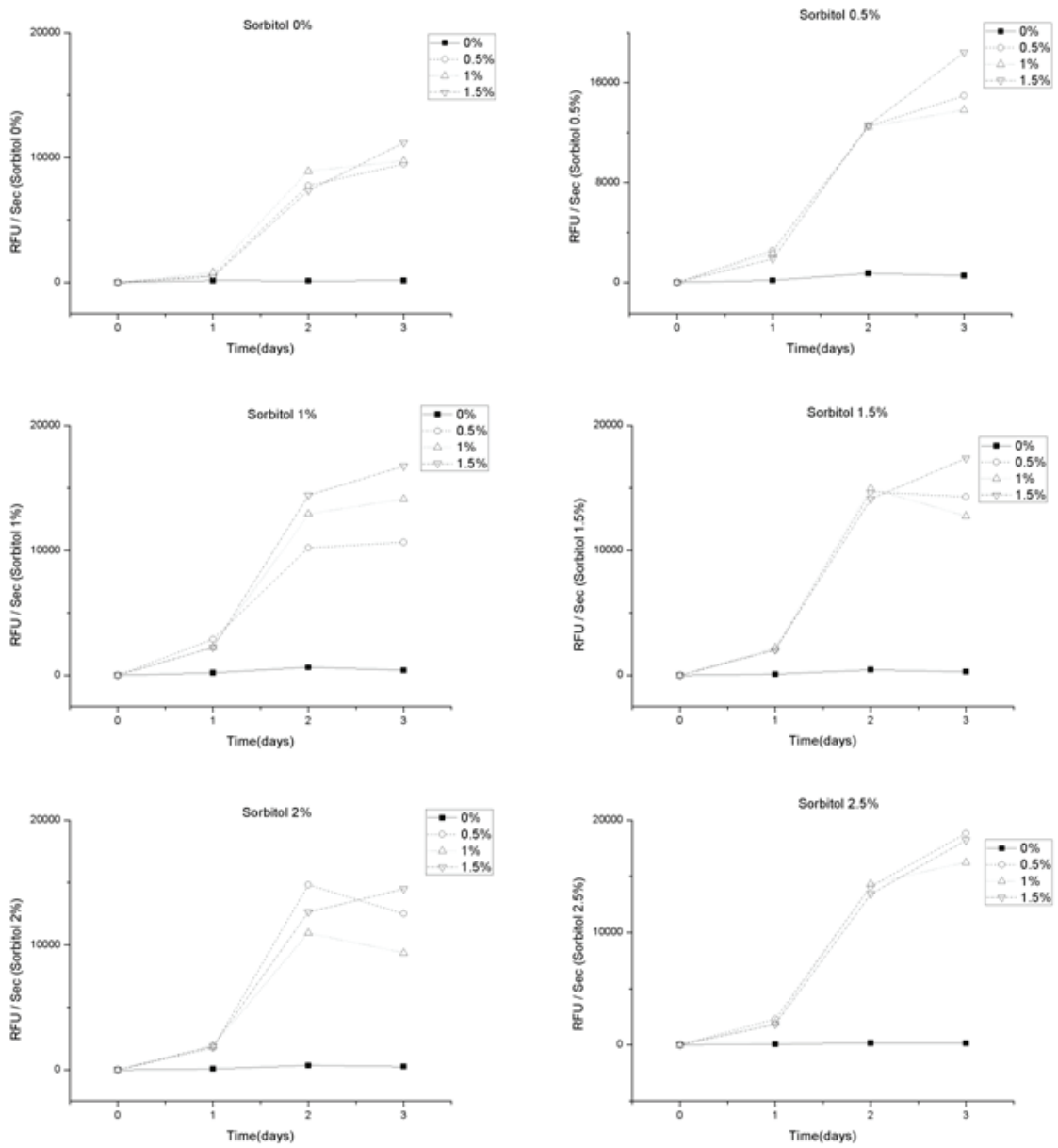


Figure 4.19 Effect of the Methanol concentration on different Sorbitol feeding

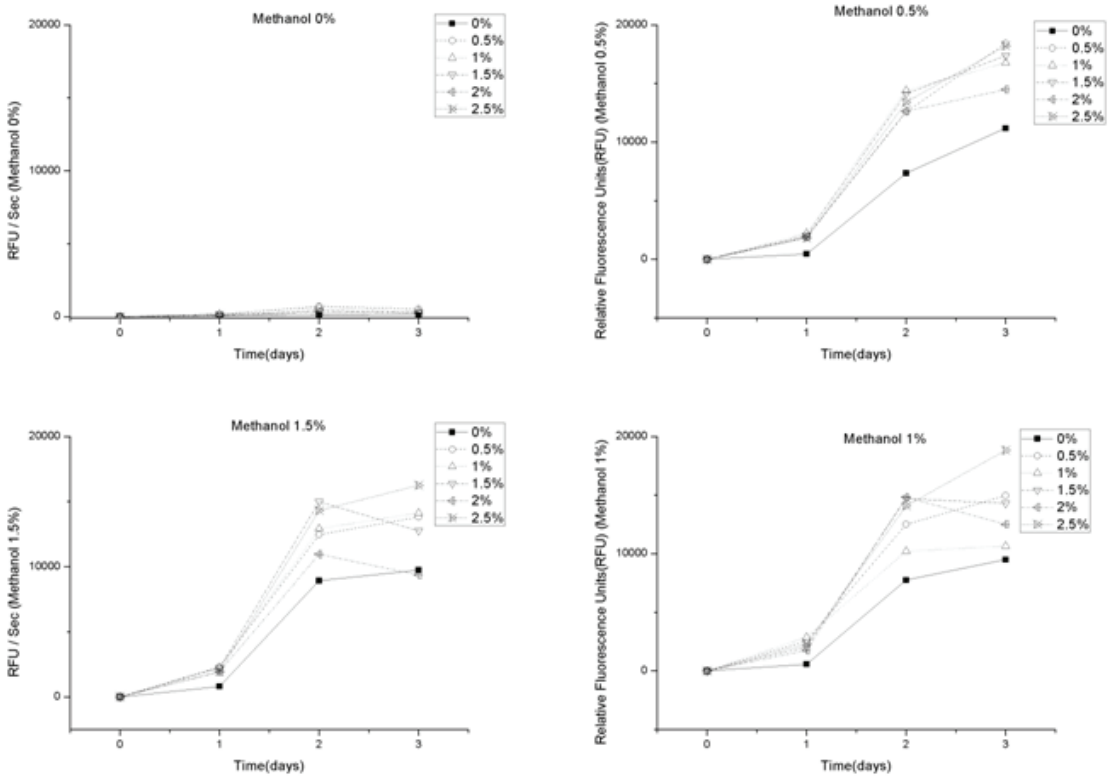


Figure 4.20 Effect of the Sorbitol feeding on different Methanol induction conditions

4.5.2.6 Purification of His-tagged BTGL enzyme

After establishing a high throughput expression system, we have purified secreted enzymes using the metal affinity chromatography in 96 well-plate format. As we stated in the introduction part of this thesis, one of the aims in this study was to establish a high throughput experimental procedure that can be applied to variety of experiments.

We have used Swellgel Nickel Chelated Discs in 96-well kit from Pierce for purification directly from the expression medium. The purity of expressed proteins after metal affinity chromatography is good enough for many applications, such as enzyme assays. Activity analysis of the purified proteins with the lipase zymogram can be seen in Figure 4.21. Similarly, activity graphs are also added to the zymogram graph as RFU versus time pictures.

SDS-PAGE gels were run for purified proteins. In Figure 4.22, purity of eluted proteins can be seen.

High throughput protein expression system using sorbitol feeding strategy was successful for increased amount of enzyme production and purification was easily applied to the 96-well plate without the buffer exchange procedures. We have just added 8X washing buffer and 1 M Imidazole into each well and directly loaded onto purification resin.

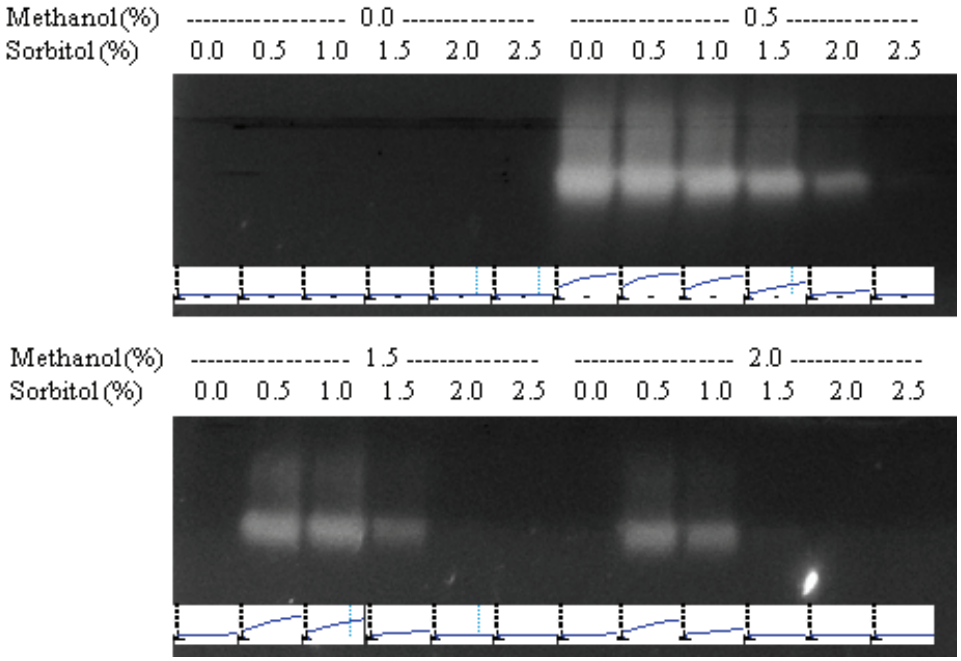


Figure 4.21 Zymogram and activity analysis of purified proteins. Activity assay results are given as RFU vs time graphs under zymogram bands. Each assay graph shows the activity of corresponding band.

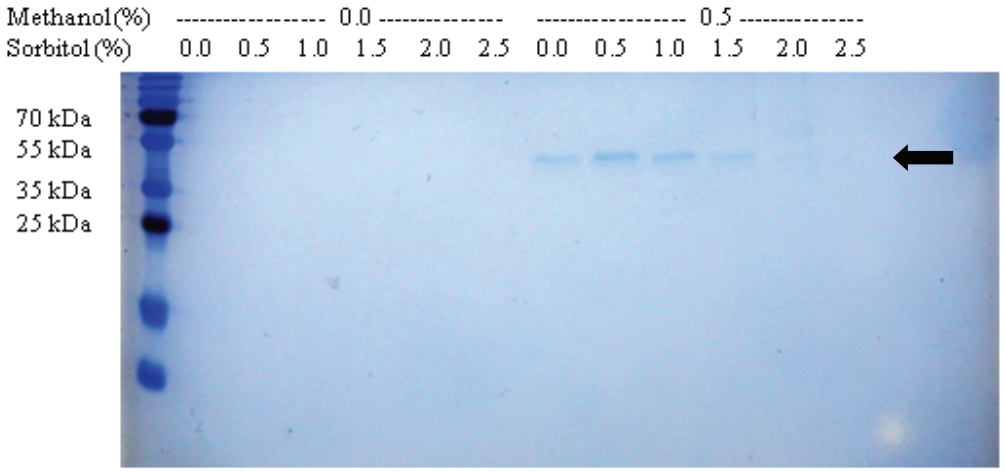


Figure 4.22 SDS-PAGE analysis of purified BTGL (43 kDa) protein. 20 ul of the eluted proteins were loaded into each well. Arrow shows band which corresponds to purified BTGL protein.

4.5.3 Enzyme Microarrays

4.5.3.1 TLC Array Fabrication and uniform coating

Uniform coating of the given substrates onto TLC slides were investigated by spotting same enzyme solutions on different parts of slides. Uniform coating was achieved by a simple coating procedure (Figure 4.23 – A). Signal intensities from all spots were plotted and a linear line was obtained (Figure 4.23 – B). Although we have used one enzyme at defined concentration for uniformity analysis, we can clearly but limitedly comment on substrate coating onto silica gel coated glass slides. Unfortunately, microarraying system has its own drawbacks. Enzyme concentrations, buffer compositions, TLC coating quality, glass quality, printing and washing cycles should be considered as a source of variation in this kind of assay systems.

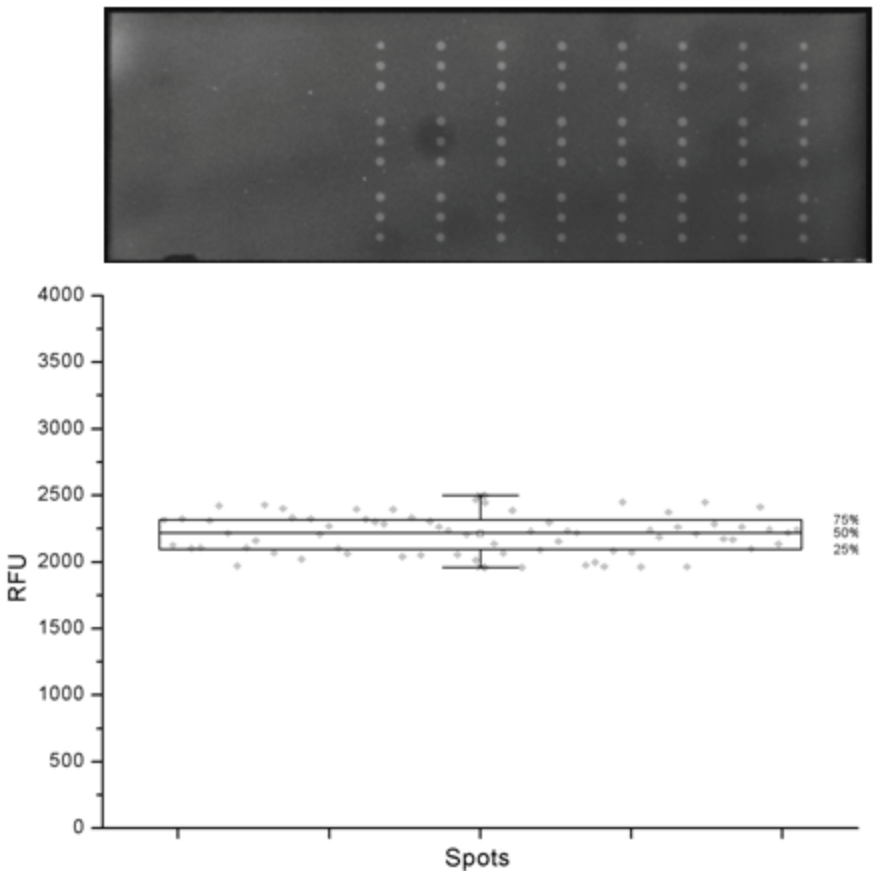


Figure 4.23 Uniform Slide Coating. Same enzyme solution (CRL) was spotted onto different parts of TLC array coated with 4MU-Caprylate to show uniform substrate deposition.

4.5.3.2 Spot shape and size analysis

Analysis of spot shapes and corresponding intensity values were analyzed using ArrayWorx Software. As seen in Figure 4.24, I have obtained uniform spot shape and good correlated fluorescent signal increase with respect to enzyme concentration (Figure 4.29). On the other hand, I have observed some donut-shape spots (Figure 4.33) when we have used the metal (Fe and Cu) ions in the spotting buffer at final concentration of 10 and 25 mM. Buffer composition is very important for TLC based microarray assays, since fluorescent substrates are deposited onto the surface and if they have dissolved in spotting buffer eventually substrates will move towards outer parts of the spots.

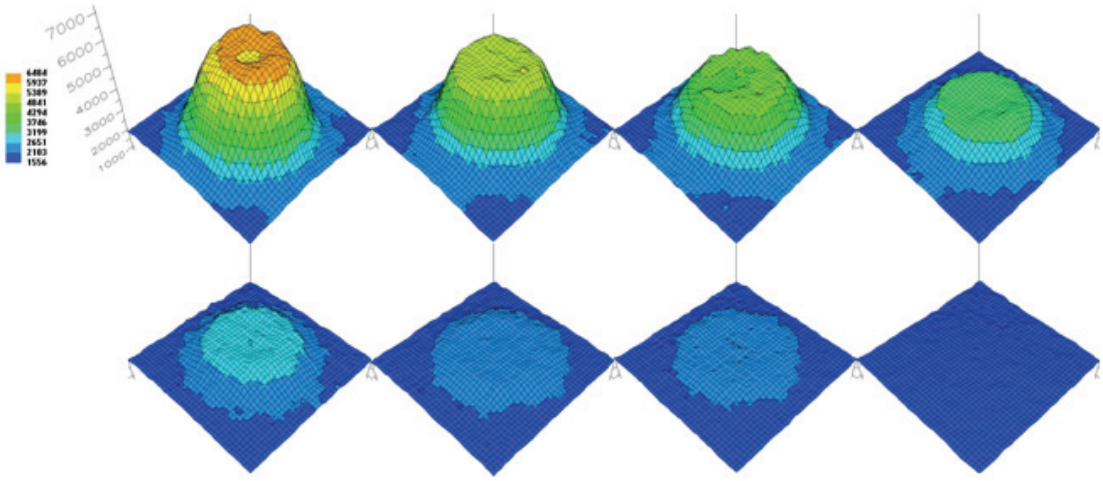


Figure 4.24 Spot shape analysis for the microarray scale lipase assay. Color bar shows signal intensities from blue (1000 RFU) to Orange (7000).

4.5.3.3 Enzyme Profiling using TLC Array

Enzyme solutions prepared as a 10 fold dilution series starting from 1 mg/ml and directly spotted onto TLC microarray. CALA, CALB, CRL and HPL were diluted in molecular biology grade water and directly used on TLC microarray (Figure 4.25). Detailed analysis of each substrate microarray and printing layout can be found in Figure 4.26. Custom synthesized substrates with 4, 8, 12, and 16 carbon chain lengths were used for coating TLC microarray (Figure 4.6).

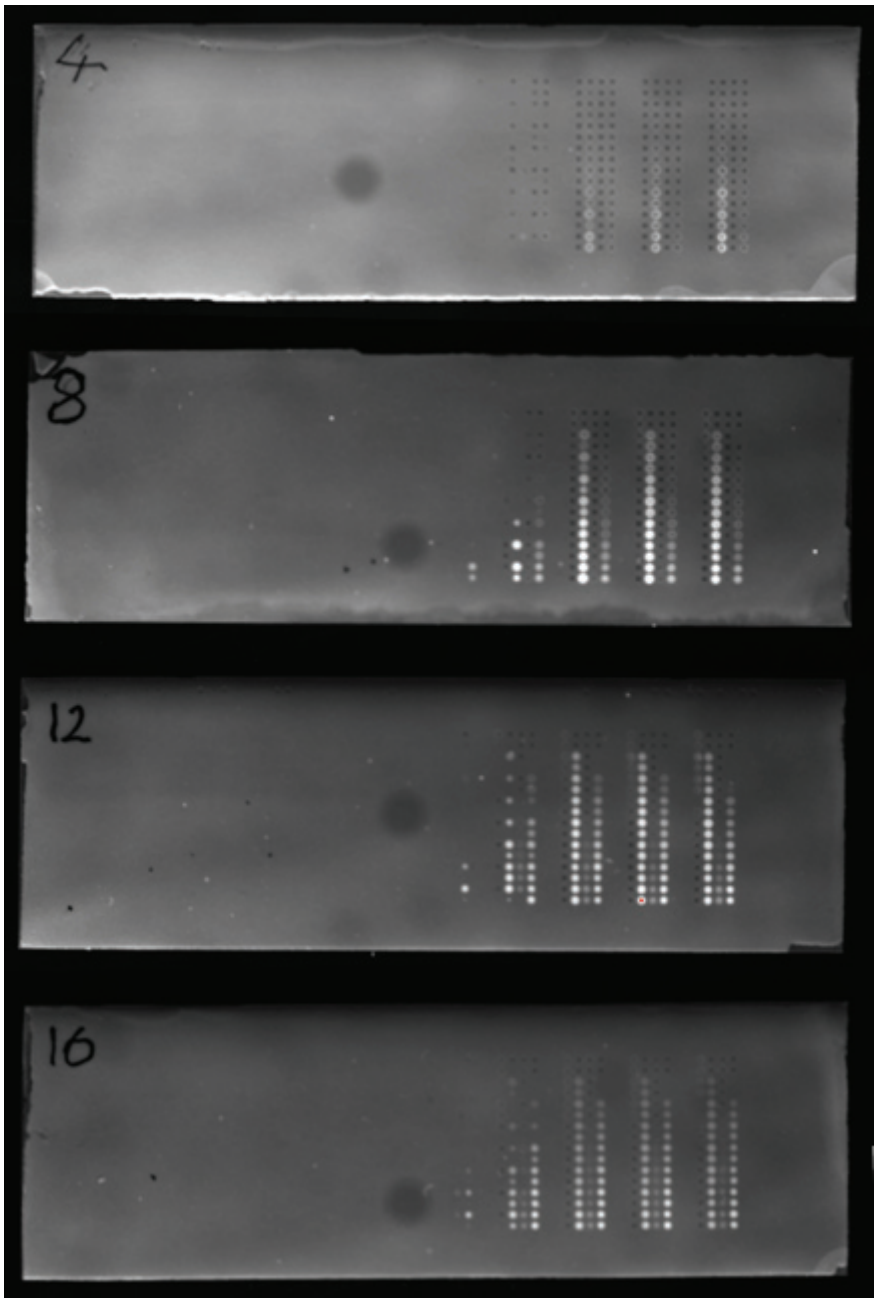


Figure 4.25 CALA, CALB, CRL and HPL (Table 4–2) enzymes were spotted onto the substrates with different chain lengths (4, 8, 12 and 16). Eight sub-arrays which are exact replicates of each other, were spotted. Only first three sub-arrays (numbered from right to left) showed enzyme activity.

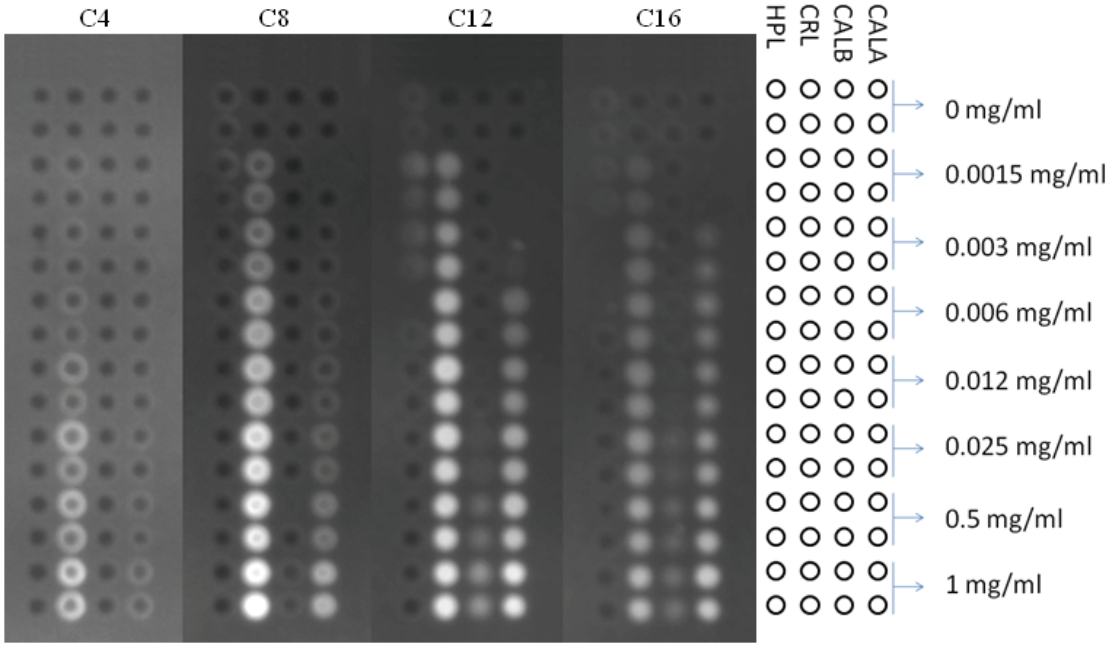


Figure 4.26 Lipase Assay on TLC Array. Slides were prepared according to the layout on the right of the array pictures. Printing was started from the upper right spots and finished at the bottom left. The right part of graph shows how to read the microarray scale enzyme assay. Each empty circle represent one spot on silica coated glass slide.

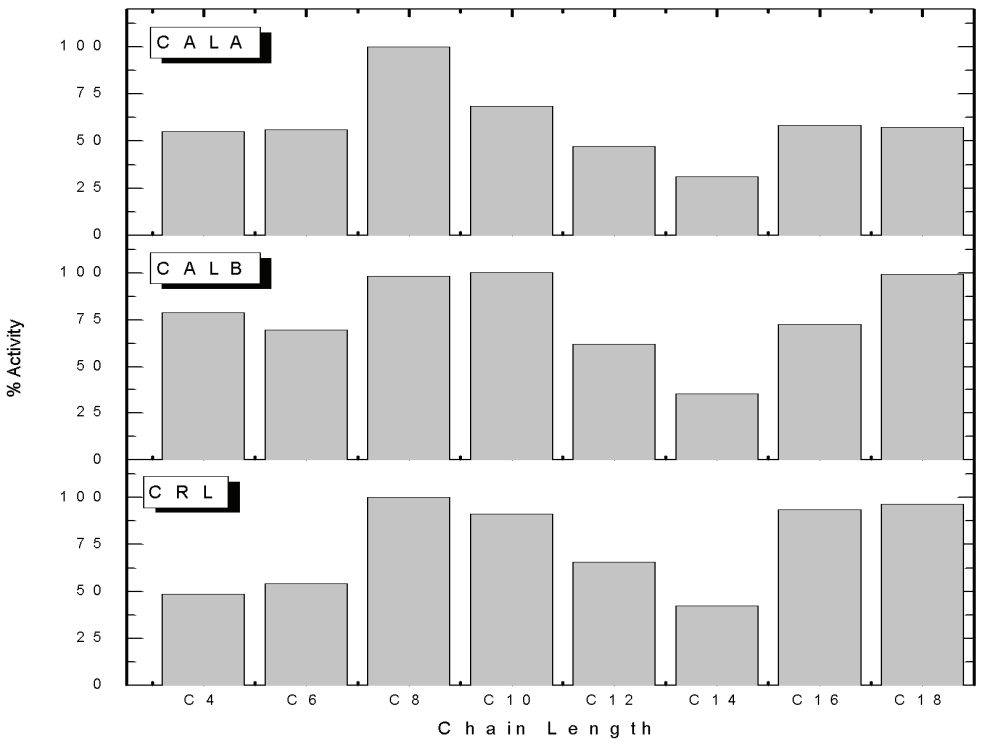


Figure 4.27 Solution phase substrate profiling for CALA, CALB, and CRL enzymes using custom synthesized substrates. X-axis shows the activity in percentage scale and Y-axis shows different substrates with varying chain lengths from 4 to 18.

4.5.3.4 Concentration Based Assay on TLC Slide

I performed a series of experiments to test the ability of TLC microarray system to differentiate different enzyme concentrations. I used CRL and CALA enzymes, 2 fold serially diluted in water starting from 10 mg/ml concentration (Figure 4.28). I observed linearity in signal intensity values in the range of 10 mg/ml and 1 mg/ml. 3 point logarithm analysis, Equation (4.1) was used for data fitting and R^2 values of 0.99948 for CRL and 0.9989 for CALA have been obtained.

$$y = a - \ln(x + c) \tag{4.1}$$

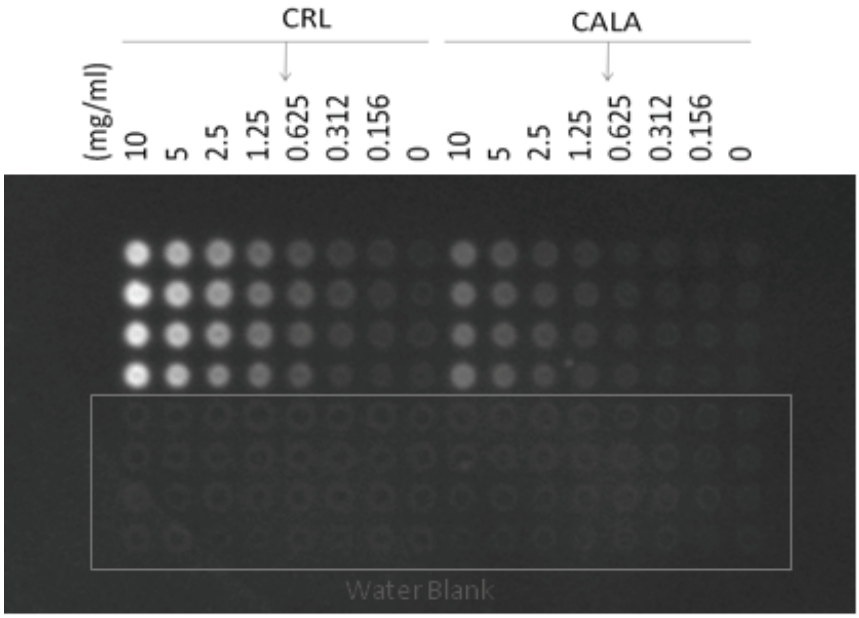


Figure 4.28 CRL and CALA enzymes were spotted onto TLC slide

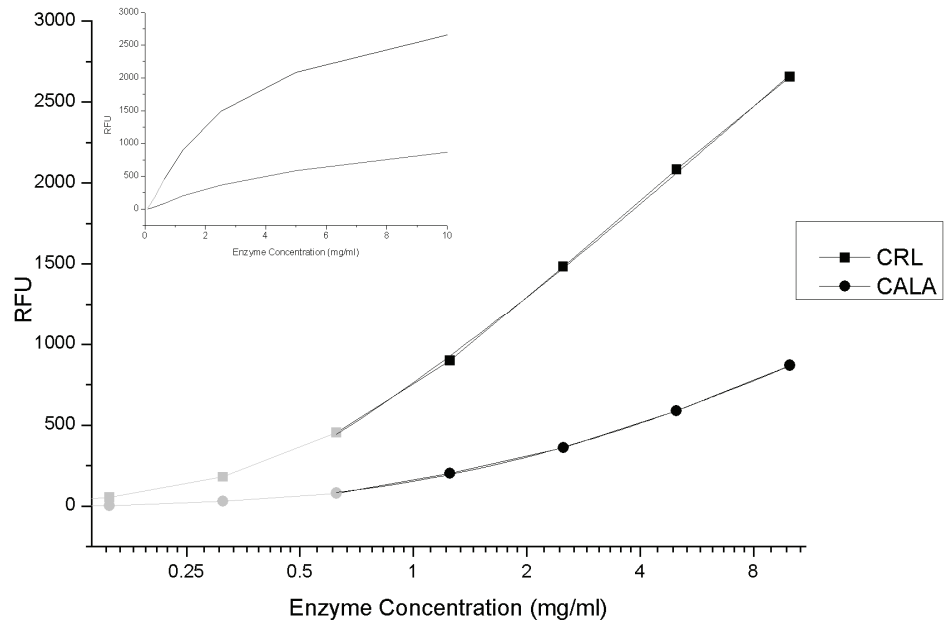


Figure 4.29 Enzyme activity assay on TLC slide. X-axis shows Relative fluorescence unit (RFU) and Y-axis shows enzyme concentration in mg/ml unit. Small graph represents the normal scale for enzyme concentrations and the large graph shows the log2 scale of enzyme concentrations.

The enzymes RAL, HPL, PFL, CAL, PCL, AL, CCL, MML, RNL, CALA, CALB, and CRL were spotted onto TLC slides which were previously subjected to 0.5

mM 4-MU caprylate in DCM (Figure 4.31). All of the enzymes were 10 mg/ml and their 1:10 dilution up to respective 7 fold 1:10 dilution with a ddH_2O blank was used for spotting. The enzyme CRL-1 and 1:10 dilution was chosen for the initial trial. Although commercial enzyme preparations (RAL, HPL, PFL, CAL, PCL, AL, CCL, MML, and RNL) were not pure (Figure 4.30) and specific activities were different, I have successfully analyzed these enzymes using TLC-microarray system.

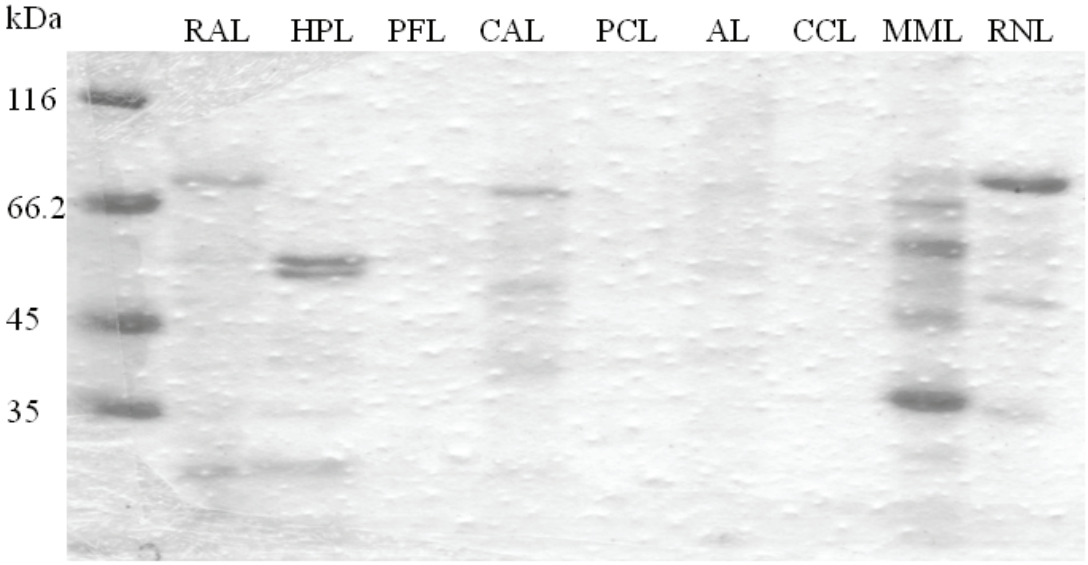


Figure 4.30 SDS-PAGE analysis of the lipases used in this study. Complete list of lipases are given in Table 4–2.

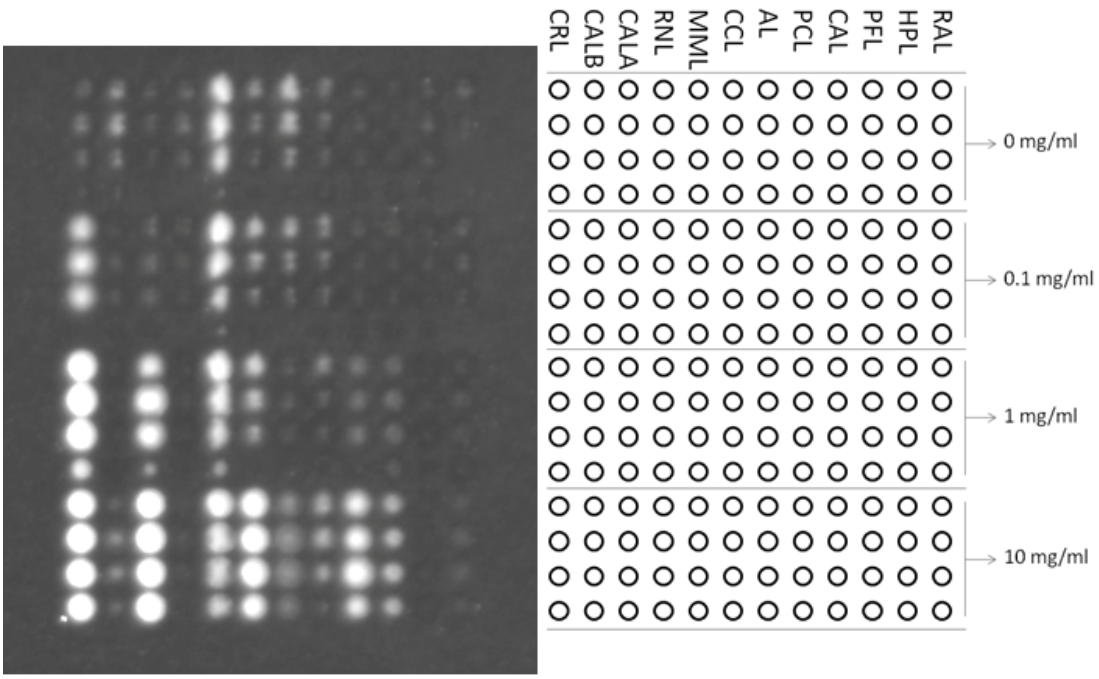


Figure 4.31 TLC array analysis of all the enzymes used in this study. 4MU-Caprylate coated silica gel plate was used as a microarray surface.

4.5.3.5 Metal Inhibition Screening Using Microarrays

I studied the effect of different metal ion concentrations on the selected enzymes (CALA, CALB, CRL, and BTGL). Enzymes and sources can be found in Table 4–2. As seen in Figure 4.32, I obtained the enzyme inhibition profile for each enzyme. The inhibitory effect of some metal ions on the lipolytic activity was determined by carrying out the same assay in 96-well plate format. The metals used in the assay were; Al, Ca, Cd, Cu, Fe, Li, K, Mg, Mn, Na, and Zn respectively. The concentrations of each metal solution provided in the final reaction mixture were 1mM, 5mM, 10 mM and 25mM. The lipases CALA, CALB, CRL and purified BTGL were used in the metal inhibition assays. The enzymes were incubated with metals for 5 minutes. The assay reactions were then set to determine the rate of the 4-MU-caprylate hydrolysis using fluorescent measurements (Section 3.4.17). The reactions were performed in duplicates.

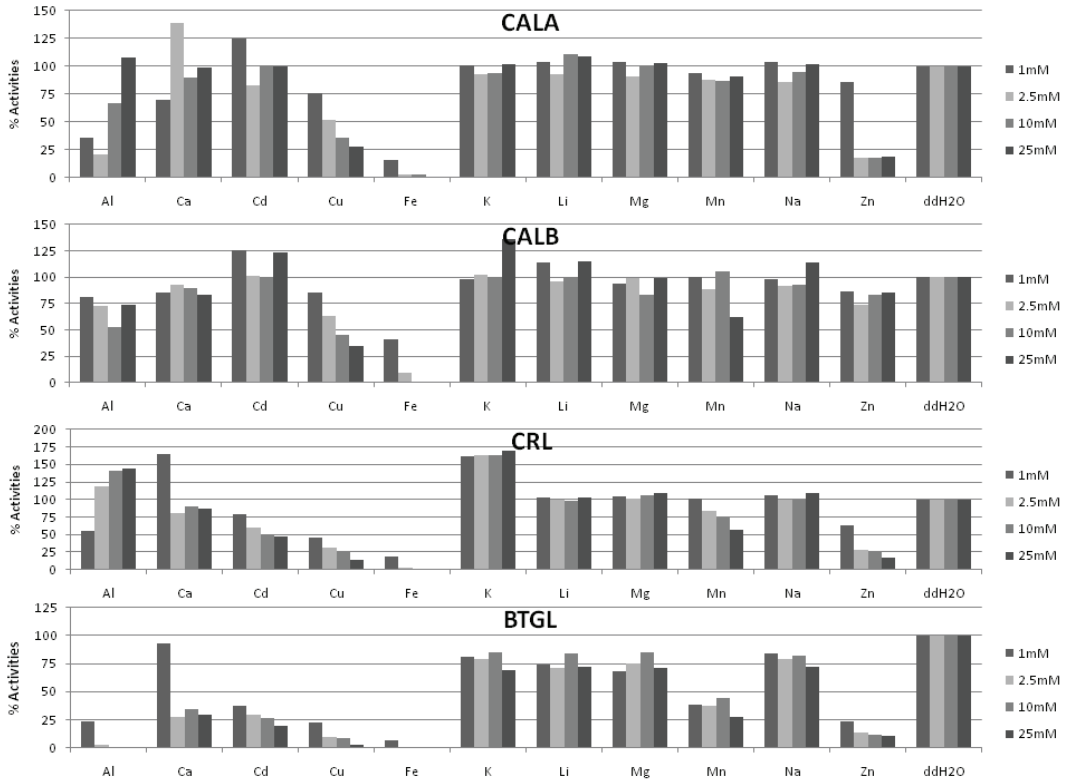


Figure 4.32 Effect of different metal ions on CALA, CALB, CRL and BTGL enzymes. Enzymes were incubated 5 min. at room temperature and then assayed against 4MU-Caprylate substrate.

The metal inhibition assay carried out in the array format was developed. The enzymes CRL and CALA were used in the same slide. The concentrations were 1 mg/ml and 3 mg/ml, respectively. The metals used in the analysis were Al, Cu and Fe. The selected inhibitory concentrations of the metal ions were 1 mM, 10 mM and 25 mM. The enzymes CRL and CALA were left in ddH_2O, 1 mM Al, 10 mM Al and 25 mM Al for 2 minutes. The reactions were set same for the metal Cu and Fe. The spotting started from (right-left) CRL-1 and its Al concentration gradient followed by Cu and Fe, next to respective Cal-A printing. In order to clean the pin head after each enzyme printing 0.1 M HCl was spotted horizontally. 4 replicas were spotted for each metal concentration.

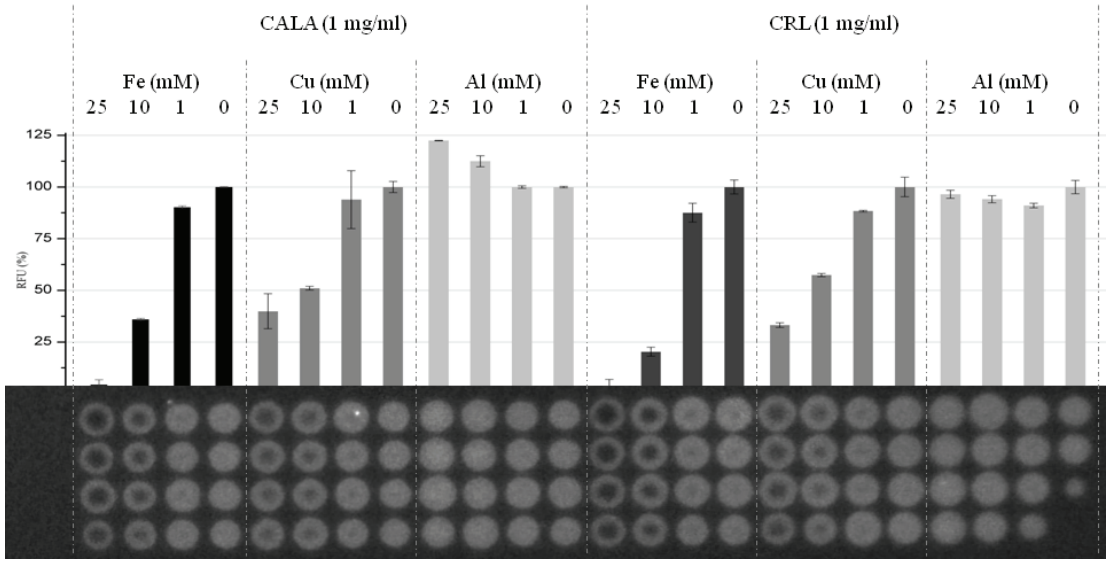


Figure 4.33 Effect of different metal concentration on enzyme activity analyzed using microarray technology. CALA and CRL enzymes were used at 1 mM concentration. Fe, Cu, and Al were mixed at 25 mM, 10 mM and 1 mM final concentration.

4.6 Conclusion

In this chapter, results of the experiments in order to investigate the feasibility of *P. pastoris* in high throughput methods are given. First, I established a cloning strategy for easy and fast transformation of PCR products without *E. coli* transformation and selection procedures. As a second part of this work, I established a microtiter plate expression protocol for KM71H cells using non-repressing carbon source. I showed that Sorbitol can increase enzyme production in small scale and expressed proteins can be purified directly from expression medium using metal affinity chromatography. Purified proteins can be used for further analysis.

I also showed that, enzyme assays can be performed on microarray scale on TLC slides. I complemented my molecular biology work with high throughput cloning and enzyme expression and purification with microarray scale enzyme assays. In the microarray scale enzyme assays part of this thesis, I established a solid strategy for analysis of the metal ion effect on the enzyme activity. In addition, I showed that the substrate selectivity of lipase enzymes can be analyzed on the microarray surfaces with different substrates.

I showed that the enzyme assays can be investigated using the microarray technology. In this thesis I analyzed lipase chain length selectivity and metal inhibition using microarray technology. Although previous research showed that it is possible to perform enzyme assays on TLC plates, I further improved the usage of TLC coating and performed microarray scale enzymatic assays. In my experimental setup, I used SMP3 pins which spot 0.7 nl of solution onto solid surface.

In general, mutant libraries do only require high transformation efficiencies other than multi-well formats, because the DNA solution already has many different mutations for screening and single but high efficiency transformation is enough for the purpose. On the other hand, developments in the rational design approaches using computational methodologies can generate more target positions for site directed mutagenesis that is not feasible to be handled in single tube based methods. Investigation such kind of problems in a high throughput manner may be improve the outcomes. Although I could not performed high throughput transformation in *P. pastoris*, available methods for *S. cerevisiae* or other yeasts can be adapted. Recently, multi-well electoporation systems have been launched. *P. pastoris* responds very well for electroporation based transformation.

4.7 References

Arnold, F. H. and G. Georgiou (2003). Directed enzyme evolution : screening and selection methods. Totowa, New Jersey, Humana Press.

Aslanidis, C. and P. J. de Jong. "Ligation-independent cloning of PCR products (LIC-PCR)."(1990) Nucl. Acids Res. **18**(20): 6069-6074.

Babiak, P. and J. L. Reymond. "A high-throughput, low-volume enzyme assay on solid support."(2005) Anal Chem **77**(2): 373-377.

Baneyx, F. "Recombinant protein expression in Escherichia coli."(1999) Current Opinion in Biotechnology **10**(5): 411-421.

Boettner, M., B. Prinz, C. Holz, U. Stahl and C. Lang. "High-throughput screening for expression of heterologous proteins in the yeast Pichia pastoris."(2002) J Biotechnol **99**(1): 51-62.

Bottner, M. and C. Lang. "High-throughput expression in microplate format in Pichia pastoris."(2004) Methods Mol Biol **267**: 277-286.

Brakmann, D. S. and B. F. Lindemann (2005). Generation of Mutant Libraries Using Random Mutagenesis. Evolutionary Methods in Biotechnology. D. A. S. Dr. Susanne Brakmann**:** 5-11.

Cereghino, J. L. and J. M. Cregg. "Heterologous protein expression in the methylotrophic yeast Pichia pastoris."(2000) FEMS Microbiol Rev **24**(1): 45-66.

Cha-aim, K., T. Fukunaga, H. Hoshida and R. Akada. "Reliable fusion PCR mediated by GC-rich overlap sequences."(2009) Gene **434**(1-2): 43-49.

Cregg, J., J. Cereghino, J. Shi and D. Higgins. "Recombinant protein expression in Pichia pastoris."(2000) Molecular Biotechnology **16**(1): 23-52.

Dubendorf, J. W. and F. W. Studier. "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor."(1991) Journal of Molecular Biology **219**(1): 45-59.

Farinas, E. T., T. Bulter and F. H. Arnold. "Directed enzyme evolution."(2001) Curr Opin Biotechnol **12**(6): 545-551.

Goddard, J.-P. and J.-L. Reymond. "Enzyme Activity Fingerprinting with Substrate Cocktails."(2004) Journal of the American Chemical Society **126**(36): 11116-11117.

Granot, D., A. Levine and E. Dor-Hefetz. "Sugar-induced apoptosis in yeast cells."(2003) FEMS Yeast Res **4**(1): 7-13.

Gräslund, S., P. Nordlund, J. Weigelt and B. M. Hallberg. "Protein production and purification."(2008) Nat Meth **5**(2): 135-146.

Grognux, J. and J.-L. Reymond. "A red-fluorescent substrate microarray for lipase fingerprinting."(2006) Molecular BioSystems **2**(10): 492-498.

Gul, O., E. Calay, U. Sezerman, H. Basaga and Y. Gurbuz. "Sandwich-type, antibody microarrays for the detection and quantification of cardiovascular risk markers."(2007) Sensors and Actuators B: Chemical **125**(2): 581-588.

Han, M. K., Y. H. Oh, J. Kang, Y. P. Kim, S. Seo, J. Kim, K. Park and H. S. Kim. "Protein profiling in human sera for identification of potential lung cancer biomarkers using antibody microarray."(2009) Proteomics **9**(24): 5544-5552.

Hartley, J. L., G. F. Temple and M. A. Brasch. "DNA cloning using in vitro site-specific recombination."(2000) Genome Res **10**(11): 1788-1795.

Haun, R. S., I. M. Serventi and J. Moss. "Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors."(1992) Biotechniques **13**(4): 515-518.

Heckman, K. L. and L. R. Pease. "Gene splicing and mutagenesis by PCR-driven overlap extension."(2007) Nat. Protocols **2**(4): 924-932.

Higgins, D. R. and J. M. Cregg (1998). Pichia protocols. Totowa, N.J., Humana Press.

Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen and L. R. Pease. "Site-directed mutagenesis by overlap extension using the polymerase chain reaction."(1989) Gene **77**(1): 51-59.

Horton, R. M., H. D. Hunt, S. N. Ho, J. K. Pullen and L. R. Pease. "Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension."(1989) Gene **77**(1): 61-68.

Huang, R. P., R. Huang, Y. Fan and Y. Lin. "Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system."(2001) Anal Biochem **294**(1): 55-62.

Inan, M. and M. M. Meagher. "Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris."(2001) Journal of Bioscience and Bioengineering **92**(6): 585-589.

Jeya, M., S. Thiagarajan, J. K. Lee and P. Gunasekaran. "Cloning and expression of GH11 xylanase gene from Aspergillus fumigatus MKU1 in Pichia pastoris."(2009) J Biosci Bioeng **108**(1): 24-29.

Kononen, J., L. Bubendorf, A. Kallionimeni, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M. J. Mihatsch, G. Sauter and O.-P. Kallionimeni. "Tissue microarrays for high-throughput molecular profiling of tumor specimens."(1998) Nat Med **4**(7): 844-847.

Kuruvilla, F. G., A. F. Shamji, S. M. Sternson, P. J. Hergenrother and S. L. Schreiber. "Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays."(2002) Nature **416**(6881): 653-657.

Lo, M. C., A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas and G. Ellestad. "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery."(2004) Anal Biochem **332**(1): 153-159.

MacBeath, G., A. N. Koehler and S. L. Schreiber. "Printing Small Molecules as Microarrays and Detecting Protein−Ligand Interactions en Masse."(1999) Journal of the American Chemical Society **121**(34): 7967-7968.

Minning, S., C. Schmidt-Dannert and R. D. Schmid. "Functional expression of Rhizopus oryzae lipase in Pichia pastoris: high-level production and some properties."(1998) Journal of Biotechnology **66**(2-3): 147-156.

Niesen, F. H., H. Berglund and M. Vedadi. "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability."(2007) Nat Protoc **2**(9): 2212-2221.

Olsen, M., B. Iverson and G. Georgiou. "High-throughput screening of enzyme libraries."(2000) Curr Opin Biotechnol **11**(4): 331-337.

Orchekowski, R., D. Hamelinck, L. Li, E. Gliwa, M. vanBrocklin, J. A. Marrero, G. F. Vande Woude, Z. Feng, R. Brand and B. B. Haab. "Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer."(2005) Cancer Res **65**(23): 11193-11202.

Pantoliano, M. W., E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane and F. R. Salemme. "High-density miniaturized thermal shift assays as a general strategy for drug discovery."(2001) J Biomol Screen **6**(6): 429-440.

Pfeffer, J., M. Rusnak, C.-E. Hansen, R. B. Rhlid, R. D. Schmid and S. C. Maurer. "Functional expression of lipase A from Candida antarctica in Escherichia coli--A prerequisite for high-throughput screening and directed evolution."(2007) Journal of Molecular Catalysis B: Enzymatic **45**(1-2): 62-67.

Ramón, R., P. Ferrer and F. Valero. "Sorbitol co-feeding reduces metabolic burden caused by the overexpression of a Rhizopus oryzae lipase in Pichia pastoris."(2007) Journal of Biotechnology **130**(1): 39-46.

Schena, M., D. Shalon, R. W. Davis and P. O. Brown. "Quantitative monitoring of gene expression patterns with a complementary DNA microarray."(1995) Science **270**(5235): 467-470.

Simionatto, S., S. B. Marchioro, V. Galli, T. D. Luerce, D. D. Hartwig, A. N. Moreira and O. A. Dellagostin. "Efficient site-directed mutagenesis using an overlap extension-PCR method for expressing Mycoplasma hyopneumoniae genes in Escherichia coli."(2009) J Microbiol Methods **79**(1): 101-105.

Studier, F. W., A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff. "Use of T7 RNA polymerase to direct expression of cloned genes."(1990) Methods Enzymol **185**: 60-89.

Tian, J., H. Gong, N. Sheng, X. Zhou, E. Gulari, X. Gao and G. Church. "Accurate multiplex gene synthesis from programmable DNA microchips."(2004) Nature **432**(7020): 1050-1054.

Torres, F. E., P. Kuhn, D. De Bruyker, A. G. Bell, M. V. Wolkin, E. Peeters, J. R. Williamson, G. B. Anderson, G. P. Schmitz, M. I. Recht, S. Schweizer, L. G. Scott, J. H. Ho, S. A. Elrod, P. G. Schultz, R. A. Lerner and R. H. Bruce. "Enthalpy arrays."(2004) Proc Natl Acad Sci U S A **101**(26): 9517-9522.

Weis, R., R. Luiten, W. Skranc, H. Schwab, M. Wubbolts and A. Glieder. "Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena."(2004) FEMS Yeast Res **5**(2): 179-189.

Xiao, Z., R. Storms and A. Tsang. "Microplate-based carboxymethylcellulose assay for endoglucanase activity."(2005) Anal Biochem **342**(1): 176-178.

Xie, J., Q. Zhou, P. Du, R. Gan and Q. Ye. "Use of different carbon sources in cultivation of recombinant Pichia pastoris for angiostatin production."(2005) Enzyme and Microbial Technology **36**(2-3): 210-216.

Zeder-Lutz, G., N. Cherouati, C. Reinhart, F. Pattus and R. Wagner. "Dot-blot immunodetection as a versatile and high-throughput assay to evaluate recombinant GPCRs produced in the yeast Pichia pastoris."(2006) Protein Expression and Purification **50**(1): 118-127.

Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein and M. Snyder. "Global analysis of protein activities using proteome chips."(2001) Science **293**(5537): 2101-2105.

Zhu, Q., M. Uttamchandani, D. Li, M. L. Lesaicherre and S. Q. Yao. "Enzymatic Profiling System in a Small-Molecule Microarray."(2003) Organic Letters **5**(8): 1257-1260.

Chapter 5

5 CONCLUSION

We have established a methodology for thermal stability studies by combining bioinformatics, molecular dynamics and molecular biology. We have tested the flexibility concept and its applicability for increased stability using computational and molecular biology. Flexibility is an important factor for determining protein stability. Increasing the flexibility of residues around active site of enzymes without changing the overall flexibility dramatically could be beneficial towards increased stability. In this thesis we have hypnotized that flexibility can be important factor for thermal stability by increasing the strength of functional domains. We have testes our mutant proteins using molecular dynamics at elevated simulation temperatures in 10 ns time period. Although computationally heavy work, rational design of enzymes can greatly reduce the search space and produce more powerful enzyme.

Although we could not purify mutant enzymes to homogeneity, secreted mutant enzymes from *P. pastoris* were enough to investigate thermal stability assays. We have used expressed proteins directly for enzymatic analysis. Finally we have found that P195G mutation has positive impact on stability and they have increased thermal stability relatively to the native ROL enzyme. We have found that the predicted mutations showed similar thermostability patterns in correlation with computational work. As a future work, our methodology can be applied to other lipases or even for other enzymes or proteins.

During the course of this thesis, we have tried to improve protein engineering techniques in terms of combining computational studies with experimental investigations. We have shown that Molecular Dynamics studies can be used as a part of protein engineering approaches and tried to generate a guideline for enzyme thermostability. We have found that mutation prediction methodology used in this thesis for increasing the thermostability of proteins can be beneficial, and this methodological approach can find broad application areas.

During this thesis work, we have also investigated the feasibility of microarray scale enzyme assays complemented by high throughput protein expression in *P. pastoris*. We have successfully increased the amount of secreted protein using sorbitol as a non-repressing carbon source in micro scale protein expression studies using Mut^s strain of *Pichia pastoris*. Linear cassette transformation methodology was developed for direct transformation of PCR products into *P. pastoris* in order to get benefit of high throughput expression system explained in this thesis. We have successfully implemented the overlap extension PCR methodology to *P. pastoris* system.

Moreover, we have established straight forward protocols for enzyme assay on microarray scale complemented with high throughput cloning, expression and purification system. Not only randomly generated libraries but also rationally designed libraries can be analyzed using this methodology. *Pichia pastoris* system was a good candidate for our purposes, because of its secretion mechanism. For enzyme analysis on microarray scale, secreted proteins are preferred since native cells do not secrete too much protein into expression medium. Although secreted proteins are diluted by nature, sensitivity of microarray systems compensates this dilution, gaining relatively pure protein content.

Appendix A

Instruments and Software

Gemini XS Spectrofluorometer	Molecular Devices
OmniGrid Microarray	Genemachines
OmniGrid Software	Genemachines
ArrayWorx Scanner	Applied Precision
ArrayWorx Software	Applied Precision
Geneamp 9700 PCR System	Applied Biosystems
Thermomixer R 1.5	Eppendorf
Mini centrifuges	Eppendorf
Gel Documentation System	Bio-Rad
Power supply	Bio-Rad
Vertical Gel Apparatus	Bio-Rad
Horizontal Gel Apparatus	Labnet
Incubated Shakers	New Brunswick
Autoclave	Hamalaya
FinchTV viewer	http://www.geospiza.com/Products/finchtv.shtml
VectorNTI	Invitrogen
Softmax Pro 4.8	Molecular Devices

Appendix B

Chemicals

Chemical Name	Supplier	Catalog Number
Acrylamide/Bis-acrylamide	Amresco, USA	0254
Agar	Merck, Germany	101614
Agarose Low EO	Applichem, Germany	A2114
Ammonium persulphate	Sigma, Germany	
Ampicillin	Sigma, Germany	A9518
Biotin	Calbiochem, Germany	2031
Coomassie Brilliant Blue	Merck, Germany	115444
EDTA (Ethylene diamine tetraacetic acid)	Riedel-de Haen, Germany	27248
Ethanol	Riedel-de Haen, Germany	32221
Ethidium Bromide	Merck, Germany	OCO28942
D-(+)-glucose	Sigma, Germany	G-7021
Glycerol	Riedel-de Haen, Germany	15523
HCl	Merck, Germany	100314
Isopropanol	Riedel-de Haen, Germany	24137
Kanamycin	Sigma, Germany	K4000.102
KH2PO4	Riedel-de Haen, Germany	04243
K2HPO4	Merck, Germany	105099
KOH	Riedel-de Haen, Germany	06005
Liquid Nitrogen	Karbogaz, Turkey	
Lithium Chloride	Fluka, Switzerland	62478
Luria Agar	Sigma, Germany	L-3147
Luria Broth	Sigma, Germany	L-3022
2-Mercaptoethanol	Aldrich, Germany	M370-1
Methanol	Riedel-de Haen, Germany	24229
NaCl	Riedel-de Haen, Germany	13423
NaOH	Merck, Germany	106462
PageBlue Protein Stain	Fermentas, Lithuania	R0571

Peptone	Merck, Germany	107213
Phenol/Chloroform/ Isoamylalcohol	Applichem, Germany	A0889
Sodium Dodecyl Sulphate	Sigma, Germany	L-4390
Sodium Acetate Trihydrate	Riedel-de Haen, Germany	25022
D (-) Sorbitol	Applichem, Germany	A2222
TEMED	Sigma, Germany	T-7029
Triton X-100	Applichem, Germany	A1388
Tris	Fluka, Switzerland	93349
Tween 20	Merck, Germany	822184
Yeast Extract	Applichem, Germany	A1552
Yeast Nitrogen Base (with ammonium sulphate without amino acids)	Invitrogen, Germany	Q300-07
Zeocin	Invitrogen, Germany	R250
Glass Slides	Cel-1	
48-well plates	Abgene	AB-0988

Appendix C

Enzymes

EcoRI	Fermentas
XbaI	Fermentas
XhoI	Fermentas
NdeI	Fermentas
SalI	Fermentas
NotI	Fermentas
Pfu Polymerase	Fermentas

Appendix D

Kits

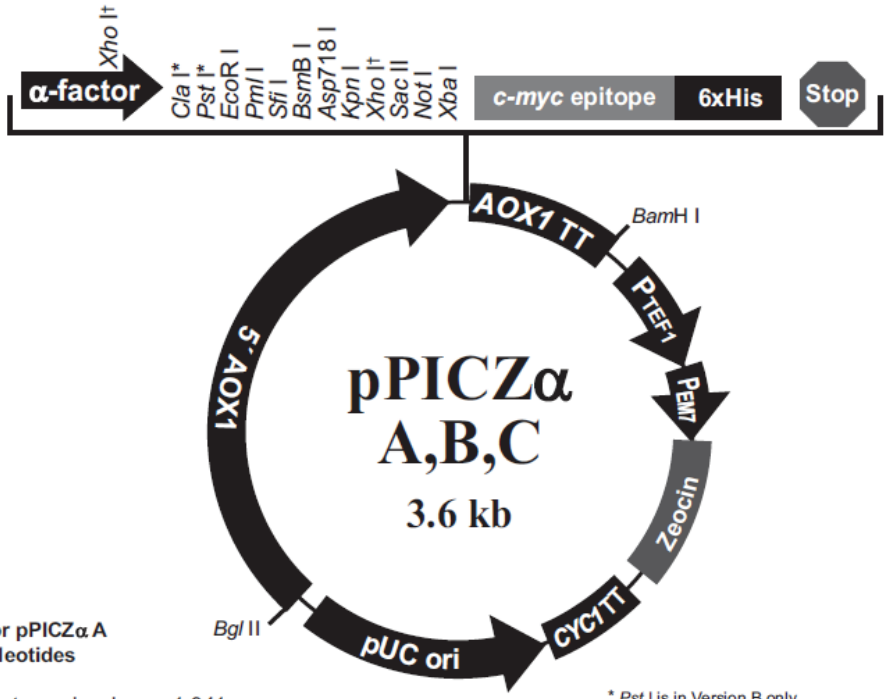
Quick ligation kit	Fermentas
In-fusion Cloning kit	BD bioscience
Silver Staining Kit	Fermentas
QiaQuick Gel Extraction Kit	Qiagen
QiaPrep miniprep Kit	Qiagen
Maxi prep kit	Qiagen
Quick Change SDM kit	Stratagene
Pichia Easy Select Kit	Invitrogen

Appendix E

Vector Maps

pPICZαA,B,C vector map (http://www.invitrogen.com)

Product Catalog number: **V195-20**



Comments for pPICZα A
3593 nucleotides

5´ *AOX1* promoter region: bases 1-941
5´ *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
α-factor priming site: bases 1144-1164
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
3´ *AOX1* priming site: bases 1423-1443
AOX1 transcription termination region: bases 1341-1682
TEF1 promoter: bases 1683-2093
EM7 promoter: bases 2095-2162
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)

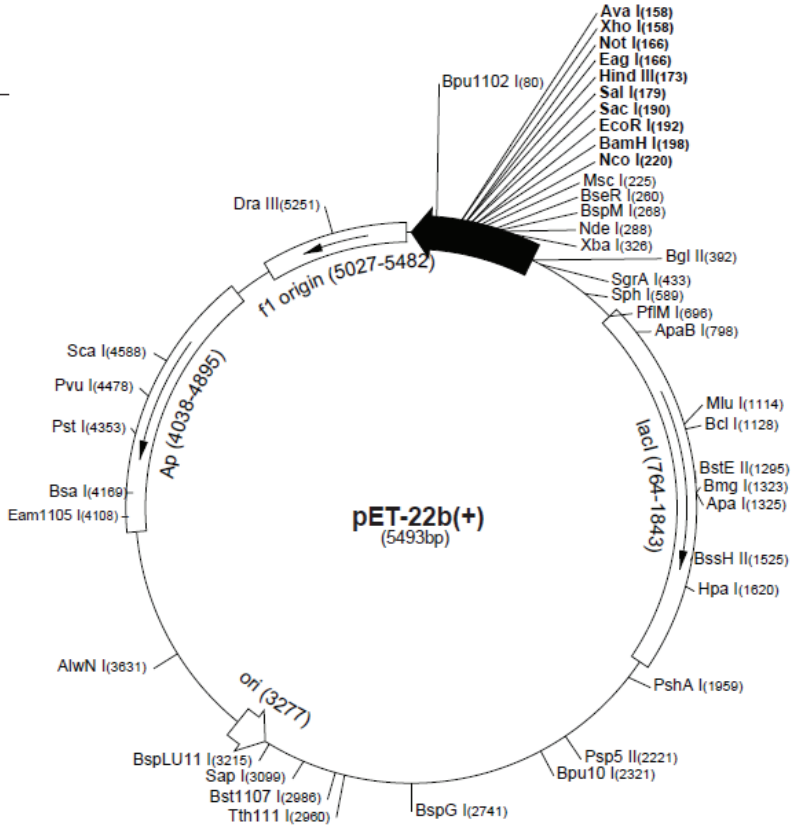
* *Pst* I is in Version B only
Cla I is in Version C only

†The two *Xho* I sites in the vector allow the user to clone their gene in frame with the Kex2 cleavage site, resulting in expression of their native gene without additional amino acids at the N-terminus.

pET22b(+) vector map (http://www.emdchemicals.com)

Product Catalog number: **69744**

pET-22b(+) sequence landmarks	
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482



pET-22b(+) cloning/expression region