

**Structural studies on thermostable variants of
Bacillus subtilis lipase evolved by *in vitro* methods**

THESIS

Submitted for the degree of

DOCTOR OF PHILOSOPHY

to

JAWAHARLAL NEHRU UNIVERSITY

New Delhi

by

SHOEB AHMAD

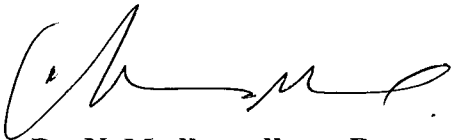


**Centre for Cellular and Molecular Biology,
(Council of Scientific and Industrial Research)
Hyderabad – 500 007.**

2008


CERTIFICATE

The research work embodied in this thesis has been carried out at the Centre for Cellular and Molecular Biology, Hyderabad. This work is original and has not been submitted in part or full for any other degree or diploma of any other university.



Dr. N. Madhusudhana Rao

PhD supervisor



Shoeb Ahma

PhD student

To My Parents

ACKNOWLEDGEMENTS

This thesis is the compilation of my work carried out during the PhD tenure. In this period, I have been encouraged and supported by many people. I express my gratitude to all of them.

First of all, I would like to thank my supervisor and guide, Dr. N. Madhusudhana Rao, who opened the doors of scientific research for me. As a student, I have always received utmost encouragement from him to learn and try new things in science. He gave me enough freedom and to work independently, along with valuable suggestions and constructive criticisms, which helped a lot in improving the shape of this work.

I would like to extend my gratitude to CCMB and especially to Dr. Lalji Singh, Director CCMB, for providing state-of-the art facilities and resources without which it would not have been possible to carry out this work. I am extremely thankful to the whole Instrumentation group of Shekhar, Asha, Vanaja, Lora, Mahesh, Dattatreya and many others for always maintaining the instruments in proper conditions. The cooperation extended by other staff; Nagesh for sequencing, Kishore for fine biochemicals, Giridharan for digital imaging, Prakash and his group for photocopying and Divakar for library, is greatly appreciated.

My seniors helped me a lot during the initial stages of PhD. I owe a lot to Priyamvada, Prasad and Satyalakshmi for their assistance and guidance which helped me during the entire duration of PhD. Long sessions of discussion with Satya, Amit, Saloni and Jobby were very fruitful.

I would like to thank Ramki, Raman, Srinivas and Manjula for valuable discussions and guidance from time to time. Assistance from Subbalakshmi and Krishnakumari on FPLC is deeply appreciated.

My past and present lab mates Vijaya Gopal, Vijayalakshmi, Zahid, Jennifer, Poornima, Viren, Bhanu, Gulzar, Asrar, Shiva and Subhadra always provided a cordial atmosphere in the lab and helped me in several ways. I am also thankful to Venugopal, Abdul Sayeed, Hanumanthu and (Late) Seyed Ehsan for providing all the necessary lab items and other support.

I had a pleasant and memorable time with my batch mates Tirumal, Aashish, Amit, Brijesh, Rupa, Ram, Vasanti, Praveen, Pallavi, Sumit, Rajkumar and Shweta. My friends Rupa, Gopalji, Hridesh, Hiten and Amit da helped me pull through a lot of difficult times.


I would like to thank all other members of west wing first floor including Dev, Bhairab, Aravind, Aftab, Prabhu, Abhishek, Saad, Abdullah and Rajeev. Working with them was a pleasure. I would especially like to thank Abhay for useful scientific and 'technical' discussions.

I am thankful to my friends Akif and Sheeba, with whom I have spent some memorable time in Hyderabad, for their support and motivation.

I am thankful to my closest friend Haseeb (NIN, Hyderabad) for listening to all my problems and coming out with solutions for them. His own hard work has been a driving force for me to handle large scale experiments during this study.

I thank the Council for Scientific and Industrial Research (CSIR) and CCMB for the financial assistance during my PhD tenure.

Finally, I am deeply indebted to my parents and all my family members for their constant support and encouragement.


(Shoeb Ahmad)

Contents

Abbreviations	i
Synopsis.....	iii

Chapter 1 Introduction

1.1 Introduction.....	1
1.2 Protein stability	3
1.2.1 Thermodynamic stability	3
1.2.2 Kinetic stability.....	4
1.3 Mechanisms of protein inactivation	5
1.3.1 Aggregation.....	5
1.3.2 Covalent modifications	6
1.4 Mechanism of protein thermostabilization.....	6
1.4.1 Amino acid composition.....	7
1.4.2 Hydrophobic interactions.....	8
1.4.3 Hydrogen bonds.....	9
1.4.4 Ion pairs	10
1.4.5 Entropy of unfolding, anchoring of loops and docking of N- and-C termini	11
1.4.6 Helix stabilization	12
1.4.7 Other factors.....	12
1.5 Engineering of protein thermostability.....	13
1.5.1 Rational design.....	13
1.5.2 Computational design	14
1.5.3 Methods of directed evolution	15
1.5.3.1 Generation of variants.....	17
1.5.3.2 Screening and selection procedures	18
1.5.3.3 Screening for enhanced protein stability.....	23
1.5.4 Semi-rational approaches.....	24
1.6 Protein thermostability, flexibility and catalytic activity.....	28
1.7 Lipases.....	29
1.7.1 Substrate specificity	30
1.7.2 Structure.....	31
1.7.3 Interfacial activation	33
1.7.4 Catalytic mechanism.....	33
1.7.5 Biotechnological application of lipases	34
1.7.6 <i>Bacillus subtilis</i> lipase.....	36
1.8 Focus of present thesis.....	38

Chapter 2 Establishment of cloning and expression system for *in vitro* evolution of lipase

2.1 Introduction	39
2.2 Materials and Methods	
2.2.1 Materials	41
2.2.2 Construction of vector pSA01	41
2.2.3 Subcloning of lipase gene in <i>E. coli</i> - <i>B. subtilis</i> shuttle vector pSA01	42
2.2.4 Transformation of <i>Bacillus subtilis</i> using electroporation.....	43
2.2.5 Natural transformation of <i>Bacillus subtilis</i>	44
2.2.6 Lipase assay	45
2.2.7 Error-prone PCR	46
2.2.8 Isolation of plasmid DNA and sequencing of lipase gene	46
2.3 Results	
2.3.1 Cloning and expression of lipase gene.....	47
2.3.2 Electro-transformation of <i>Bacillus subtilis</i>	48
2.3.3 Natural transformation of <i>Bacillus subtilis</i>	49
2.3.4 Solid phase lipase assay	49
2.3.5 Random mutagenesis of lipase gene	50
2.4 Discussion	52

Chapter 3 *In vitro* evolution of *Bacillus subtilis* lipase for thermostability

3.1 Introduction	56
3.2 Materials and Methods	
3.2.1 Materials	58
3.2.2 Mutagenesis	58
3.2.3 Screening.....	58
3.2.4 Recombination	60
3.2.5 Site-saturation mutagenesis	61
3.2.6 Plasmid DNA isolation and sequencing of lipase gene	62
3.2.7 Protein purification	62
3.2.8 Circular dichroism	63
3.2.9 Fluorescence	64
3.2.10 Thermal inactivation	64
3.2.11 Thermal unfolding	65
3.2.12 Static light scattering.....	65
3.2.13 Urea unfolding	66
3.2.14 Activity measurements.....	66
3.3. Results	
3.3.1 Generation of thermostable mutants	67
3.3.2 Characterization of lipase mutants.....	72
3.3.3 Thermal inactivation	74
3.3.4 Thermostability of mutants	77

3.3.5 Temperature induced aggregation of mutants.....	78
3.3.6 Temperature dependence of activity.....	79
3.3.7 Equilibrium unfolding in urea.....	80
3.3.8 Catalytic properties of lipase mutants.....	82
3.3.9 Structural basis of thermostability.....	82
3.3.10 Validation of mutational effect by site-saturation mutagenesis.....	86
3.4 Discussion.....	87

Chapter 4 Further improvement in thermostability of *Bacillus subtilis* lipase by site-saturation mutagenesis

4.1 Introduction.....	92
4.2 Materials and Methods	
4.2.1 Materials.....	94
4.2.2 Construction of vectors used for screening and protein purification.....	94
4.2.3 Site-saturation mutagenesis.....	96
4.2.4 Screening.....	97
4.2.5 Recombination.....	98
4.2.6 DNA sequencing.....	99
4.2.7 Protein purification.....	99
4.2.8 Circular dichroism.....	100
4.2.9 Fluorescence.....	100
4.2.10 Thermal inactivation and unfolding.....	100
4.2.11 GdmCl unfolding.....	100
4.2.12 Activity measurements.....	101
4.3 Results	
4.3.1 Periplasmic expression of lipase gene in <i>E. coli</i>	101
4.3.2 Generation of thermostable mutants.....	102
4.3.3 Characterization of lipase mutants.....	105
4.3.4 Thermal inactivation.....	107
4.3.5 Thermostability of mutants.....	109
4.3.6 Equilibrium unfolding of lipase mutants in GdmCl.....	110
4.3.7 Temperature dependence of activity.....	111
4.3.8 Catalytic properties of lipase mutants.....	113
4.3.9 Structural basis of thermostability.....	113
4.4 Discussion.....	115

Chapter 5 Thermal unfolding and refolding of *in vitro* evolved lipase variants

5.1 Introduction.....	120
5.2 Materials and Methods	
5.2.1 Materials.....	122
5.2.2 Thermal inactivation.....	122

5.2.3 Activity measurements.....	123
5.2.4 Thermal unfolding and refolding.....	123
5.2.5 Circular dichroism	123
5.2.6 Fluorescence	124
5.2.7 Static light scattering.....	124
5.2.8 Reversibility of thermal unfolding.....	124
5.2.9 Isoelectric focusing	125
5.3 Results	
5.3.1 Thermal inactivation	126
5.3.2 Thermal unfolding and refolding of lipase mutants.....	127
5.3.3 Reversibility of thermal unfolding of lipase mutants upon incubation at elevated temperatures	133
5.3.4 Aggregation of lipase mutants upon incubation at elevated temperatures	135
5.3.5 Bis-ANS binding to denatured mutants at elevated temperatures	137
5.3.6 Effect of ionic strength on the reversibility of thermal unfolding	138
5.3.7 Thermal unfolding and refolding of 6-B mutant.....	141
5.3.8 Structural basis of thermotolerance	143
5.4 Discussion.....	145
Conclusion and future perspective.....	151
References.....	154
Publications	171

ABBREVIATIONS

Amp	Ampicillin
BICINE	N,N-bis(2-hydroxyethyl)glycine
bis-ANS	1,1',-bis(4-anilino)naphthalene-5,5'-disulfonic acid
CD	Circular dichroism
CFU	Colony forming units
CHAPS	3--[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
dNTP	deoxy ribonucleotide tri phosphate
DTT	Dithiothreitol
ELISA	Enzyme linked immunosorbent assay
epPCR	Error-prone PCR
FACS	Fluorescent activated cell sorting
FPLC	Fast protein liquid chromatography
GdmCl	Guanidinium chloride
GSSM	Gene site-saturation mutagenesis
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
IPTG	Isopropyl- β -D-thiogalactopyranoside
Kan	Kanamycin
k_{cat}	Turnover number
kDa	kilo Dalton
K_m	Michaelis constant
LB	Luria Bertani
LBSP	Luria Bertani sucrose phosphate
mQ	Milli-Q
NMR	Nuclear magnetic rasonance
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PNPA	p-Nitrophenyl acetate
PNPB	p-Nitrophenyl butyrate
PNPO	p-Nitrophenyl oleate

RA	Residual activity
RP	Residual (soluble) protein in solution
SDS	Sodium dodecyl sulphate
SSM	Site-saturation mutagenesis
TB	Terrific broth

Synopsis

Synopsis

Structural solutions adopted by proteins to withstand extremes of temperature, have basic importance in the understanding of protein structure-function relations and also in industrial applications of enzymes. Understanding of principles and mechanisms employed by proteins, to maintain their functionally active form, at extremes of temperature, can be utilized to improve the stability of biocatalysts used in industrial context. Much of this understanding has come from comparative sequence and structural analysis of homologous proteins from organisms inhabited to extreme environments. These studies have revealed the importance of several structural features which play important role in conferring stability to proteins. However, these studies often lead to confounding interpretations on the relevance of a mutation to the property due to multiplicity of selection pressures and neutral drifts. In addition, stability of a protein translates into presence of few weak interactions. Despite several studies, understanding the structural basis of thermostability has proven elusive as there are no well defined rules to stabilize a protein at high temperature.

Recently, another strategy of improving protein property has been adopted, which has met with phenomenal success in recent past. Termed as “Methods of directed or *in vitro* evolution”, these methods mimic the natural evolution but are performed under controlled conditions on an accelerated time scale. It involves iterative rounds of *in vitro* generation of random mutants followed by screening of variants for the desired property. These methods are blind to the structural information in order to improve the property, rather, the information gathered can be used to understand sequence-structure-function relationship in a retrospective manner.

Bacillus subtilis lipase A is a small (19.3 kDa), monomeric secretory enzyme, with no cofactors and disulfides. It is a typical mesophilic enzyme, with optimum temperature of activity around 35 °C and shows broad substrate specificity, acting on both long as well as short chain esters and triglycerides, thus having potential industrial application. This makes it an important model protein, to improve its thermostability by directed evolution and to investigate the structural solutions a protein adopts to deal with heat.

General aspects of protein thermostability and the structural features that contribute to protein thermostability are discussed in **Chapter 1**. Strategies used in protein engineering, mainly for improving stability with emphasis on directed evolution are discussed. The structure, specificity, catalytic mechanism and application of lipases, in general, and *Bacillus subtilis* lipase, in particular, are also discussed.

Chapter 2 deals with the efforts made in the establishment of a platform for *in vitro* evolution of *Bacillus subtilis* lipase for thermostability. This involves cloning of lipase gene and development of an efficient expression and screening system. Lipase gene was cloned in a suitable *E. coli*-*B. subtilis* shuttle vector for expression and secretion of protein in *B. subtilis* host. Being gram positive, *B. subtilis* is very difficult to transform, thus, experimental procedures were optimized to get suitable transformation efficiencies. Solid- and liquid-phase lipase assays were standardized to develop a three-tier screening protocol for improving lipase thermostability. Finally, procedure of random mutagenesis by error-prone PCR was also optimized.

Using a partially stable triple mutant (TM) as parent, *in vitro* evolution of *Bacillus subtilis* lipase was performed, as described in **Chapter 3**. By performing two rounds of random mutagenesis and screening, six thermostabilizing mutations, three mutations per generation, were identified. Each of the six mutations, when evaluated individually, contributed to thermostability in an additive manner. These mutants were recombined in different orders to generate a series of variants with incremental increase in thermostability. The most thermostable variant, 4D3, having nine mutations (six identified in this study, A15S, F17S, A20E, N89Y, G111D, I157M ; and three previously identified L114P, A132D and N166Y) shows a remarkable 15 °C increase in melting temperature and a million fold reduction in rate of thermal inactivation. This is accompanied by 20 °C increase in optimum temperature of activity. Notably, increase in thermostability is achieved without any loss in activity at ambient temperature, rather, the specific activity of thermostable mutants has increased 2-5 fold in the temperature range of 25-65 °C. Structural analysis of the mutants revealed the importance of subtle non-covalent interactions in stabilization of unstructured regions (loops), in the vicinity of active site, in conferring stability to protein.

Random point mutagenesis followed by screening for thermostability, has the potential to identify “weak spots” in protein, although the substitution it provide might be sub-optimal. Optimizing of substitution at these sites by saturation mutagenesis may lead to phenomenal increase in thermostability. As described in **Chapter 4**, four positions, 134, 137, 158 and 163, which were identified to effect lipase thermostability, were optimized in 4D3 by site-saturation mutagenesis. All four mutations identified (M134E, M137P, G158D and S163P) significantly improved protein thermostability. The most thermostable variant, 6-D, having all the four mutations, shows remarkable shift of 7 degrees in melting temperature which is 22 degrees higher than wild-type protein. Notably, in addition to the significant increase in temperature optima by 10 degrees (30 degrees compared to wild-type enzyme), the catalytic efficiency has also increased by 3-6 fold in the temperature range of 25-65 °C. Characterization of mutants, having individual as well as in different combinations, by thermal inactivation and unfolding as well as by equilibrium unfolding in GdmCl revealed confounding results. Only two mutations, M137P and S163P, were found to increase conformational stability of the protein while the other two have marginal effects. These two mutants, M134E and G158D, however improve the ability of the protein to resist high temperature, without effecting conformational stability of the protein.

Chapter 5 deals with the detailed characterization of the mutants obtained by site-saturation mutagenesis at four positions in 4D3. All the four mutations phenomenally improve enzyme’s thermotolerance, with 25 °C shift in mid-point of thermal inactivation with little improvement in melting temperature. *Bacillus subtilis* lipase and all its thermostable variants till 4D3 undergo irreversible thermal unfolding and inactivation, primarily due to aggregation at elevated temperatures. Detailed characterization of mutants by monitoring thermal inactivation, unfolding and refolding by activity and several spectroscopic measurements has revealed that these mutations have altered the unfolding and refolding pathways. Parent 4D3 and two mutants partially unfold to intermediate forms around 75 °C, which are prone to aggregation. Introduction of these four mutations affects the generation of intermediate to different extent. While M137P has maximum effect which inhibits aggregation completely, M134E, G158D and S163P have partial effect. Upon further increase in temperature (at 85 °C), the denatured state acquired by all the mutants is highly similar, does not show any perceptible aggregation, but is still susceptible to chemical modifications. Thus, these four mutations

not only improve conformational stability of the protein but also increase its thermotolerance by improving the refolding yields after thermal unfolding.

Introduction

Chapter 1

1.1 Introduction

Proteins are important class of biomacromolecules which perform the most diverse functions inside the living cell. Enzymes are a subset of proteins with catalytic function. They catalyze almost all the metabolic reactions in the cell. Enzymes catalyze chemical reactions with remarkable specificity and enhance the rates by several orders ($10^6 - 10^{17}$), as compared to the uncatalyzed reaction. The remarkable catalytic power and versatility of enzymes has long been recognized and enzymes have proved to be useful workhorses outside the living system as well. However, in non-physiological conditions, the enzyme activity is compromised compared to their activities in natural milieu. Enzymes show optimum activity around the physiological temperature of the organism which tend to decrease with increase in temperature due to denaturation. Thermostable proteins maintain their native three dimensional structures even at elevated temperatures compared to less stable proteins, which loose their structure at high temperatures due to unfolding. The term protein thermostability refers to the preservation of the unique chemical and spatial structure of a polypeptide chain under extremes of temperature conditions (Jaenicke 1991; Jaenicke and Bohm 1998). In energy terms this refers to the extent to which the free energy of the native conformation differs from that of the (reversibly) unfolded form.

Life exists almost everywhere on the earth, from deep-sea hydrothermal vents to cold expanses of Antarctica. The organisms that inhabit and have adapted to these extreme and diverse environments are often classified by their altered habitat. Based on the temperature of growth, microorganisms are classified into four groups: psychrophiles (0 – 20 °C), mesophiles (20 – 50 °C), thermophiles (50 – 80 °C) and hyperthermophiles (80 – 120 °C). Proteins isolated from different groups vary significantly in their stabilities. Proteins from thermophiles and hyperthermophiles maintain their structure and remain active at higher temperatures compared to their mesophilic counterparts (Vieille et al. 1996; Vieille and Zeikus 2001). Identification and understanding of the forces contributing to the stability of proteins has been a long standing problem. The first high resolution crystal structure of thermolysin was reported way back in 1974 (Matthews et al. 1974), while Perutz and Raidt described the stereochemical basis of thermostability of ferredoxins and hemoglobin in 1975 (Perutz and Raidt 1975). Since then understanding the structural basis of thermostability of proteins has become an actively pursued area of research in biophysical chemistry, not only to understand the basic principles involved but

also to use those principles in designing better biocatalysts for industrial and therapeutic applications (Kumar et al. 2000; Szilagyi and Zavodszky 2000; Sterner, Liebl 2001; Kumar and Nussinov 2001; Razvi and Scholtz 2006).

Proteins perform important tasks in all biological systems, and they do so by maintaining a specific globular conformation. This functional state, called the native state, is marginally stabilized in a balancing act of stabilizing and destabilizing forces. The players in this balancing act have long been identified (Kauzmann 1959) though their relative contributions have been debated (Pace et al. 1996; Karshikoff and Ladenstein 2001). The major stabilizing forces include the hydrophobic effect and hydrogen bonding while conformational entropy favors the unfolded state. The forces stabilizing the native state outweigh the disruptive forces marginally in folded protein, in the range of 5 – 10 kcal mol⁻¹ (Pace 1975). This balance of forces is known as conformational stability of the protein and is defined thermodynamically as the free energy change, ΔG_{stab} , for the native to unfolded state transition. Studies on protein stability explore the sequence-structure-stability relationship, with stability being the measured thermodynamic quantity, since sequence defines the structure, whose interactions afford stability. Sequence is the variable that organisms alter as they evolve to adapt their proteins to the environments they inhabit.

Extensive work has been performed on the stability studies of thermophilic proteins from hyperthermophiles and thermophiles along with their mesophilic counterparts. Thermophilic proteins, which maintain their native structure and function at high temperatures, shows only marginal difference between free energy of stabilization $\Delta\Delta G_{\text{stab}}$, i.e. changes in ΔG_{stab} , in the range of 5 -20 kcal mol⁻¹, as compared to its mesophilic counterpart (Jaenicke and Bohm 1998; Vieille and Zeikus 2001). Stabilization may involve all levels of hierarchy of protein structure involving local packing of the polypeptide chain, secondary and tertiary structure elements, domains and subunits. Comparison of homologous thermophilic enzymes (thermozymes) with their mesophilic counterparts has revealed that, with the exception of phylogenetic variations, temperature range in which they are adapted is the only factor that can differentiate these enzymes. Otherwise, hyperthermophilic and mesophilic enzymes are highly similar: (i) the sequence similarity between hyperthermophilic and their mesophilic homologues are typically 40 to 85%; (ii) their three dimensional structures are superimposable; and (iii) they share the same catalytic mechanism (Vieille and Zeikus 2001). Recombinant thermozymes produced in mesophilic organisms retain the thermostability suggesting that this property

is intrinsic. Operationally, the adaptation of an enzyme to be a thermozyyme seems to be preservation of functionally important motions in such a way that the “functional state” of the enzyme is intact between a mesophilic and a thermophilic enzyme.

1.2 Protein Stability

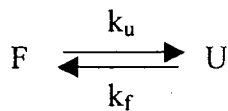
Proteins from thermophilic and hyperthermophilic organisms are generally described in the literature to be very (thermo) stable. However, the term “stability” is used with quite different meanings. The following sections deals with various definitions of protein stability.

1.2.1 Thermodynamic stability

Thermodynamic stability of the protein can be defined as the difference between the free energies of the folded (F) and unfolded (U) states of the protein, which is termed as ΔG_{FU} .

$$\Delta G_{FU} = G_u - G_f$$

ΔG_{FU} can be determined if the protein undergoes reversible two-state unfolding transition:



where, F is the native folded structure, and U is reversibly unfolded state; k_u and k_f are the rate constants for unfolding and refolding, respectively. The two-state model implies that (a) the unfolding from F to U is reversible, so that the system remains in equilibrium and (b) F and U are the only two states in protein is present, that is, significant amounts of folding intermediates are absent. From the equilibrium constant $K = k_u/k_f = [U]/[F]$, ΔG_{FU} can be calculated as follows:

$$\Delta G_{FU} = -RT \ln K$$

where, R is universal gas constant and T is temperature in Kelvin.

The only factors which affect the thermodynamic stability (ΔG_{FU}) are the relative free energies of the folded (G_f) and the unfolded (G_u) states. The larger and more positive ΔG_{FU} , the more stable is the protein to denaturation.

At the standard temperature of 25 °C, ΔG_{FU} values are generally determined by the analysis of protein unfolding induced by denaturants such as urea or guanidinium chloride

(Pace and Scholtz 1997). The degree of unfolding with respect to denaturant concentration [D] is measured by spectroscopic techniques such as fluorescence or CD. The normalized spectroscopic signals yield equilibrium constant K and thus ΔG_{FU} as a function of [D]. Assuming a linear relationship between ΔG_{FU} and [D], both ΔG_{FU} in the absence of denaturant ($\Delta G_{FU}(\text{H}_2\text{O})$) and the m-value, which is a measure of the dependence of ΔG_{FU} on [D] can be determined (Pace and Scholtz 1997).

$$\Delta G_{FU} = \Delta G_{FU}(\text{H}_2\text{O}) - m [D]$$

In order to obtain ΔG_{FU} over a broad range of temperatures, the enthalpy change (ΔH) and the heat capacity difference (ΔC_p) between U and F can be determined by analyzing the thermal unfolding process (Pace and Scholtz 1997). ΔG_{FU} as a function of temperature is calculated with the Gibbs-Helmholtz equation:

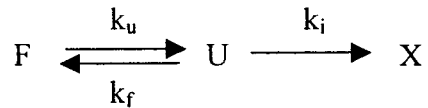
$$\Delta G_{FU}(T) = \Delta H_m(1-T/T_m) - \Delta C_p [(T_m-T) + T \ln(T/T_m)]$$

The curve of ΔG_{FU} vs. Temperature is skewed parabola and intersects the x-axis twice, indicating that unfolding occurs both at low and high temperatures. Other parameters of interest that can be calculated using modifications of the given equation include T_S and ΔG_S , where T_S is the temperature of maximum stability or temperature where the change in entropy between native and denatured states is zero and ΔG_S is the conformational stability at this temperature. Protein stability curves also allow the calculation of conformational stability at any temperature, including the habitat temperature of an organism (T_E).

1.2.2 Kinetic stability

Although many small mesophilic and thermophilic proteins show completely reversible unfolding transitions, thermal unfolding of most of mesophilic and thermophilic proteins is irreversible, mainly due to aggregation or other chemical modifications of the unfolded state. For these proteins, thermodynamic analyses could not be performed as the system does not achieve equilibrium. For these proteins the kinetic stability, or the rate of unfolding, is important. Kinetic stability is a measure of how rapidly a protein unfolds. A protein which is kinetically stable unfolds more slowly than a kinetically unstable protein. In kinetically stable protein, a large free energy barrier to unfolding is required and the factors affecting stability are the relative free energies of the folded (G_f) and the transition state (G_{ts}) for the first committed step on the unfolding pathway. It is the magnitude of this

difference (activation energy), that determines the rate of unfolding (and hence inactivation). A simple model for irreversible unfolding can be shown as:



where, F and U are the folded and unfolded states, which undergoes reversible transition, while X is the irreversibly unfolded state and k_i is the rate constant for the transition from U to X. In case of irreversible unfolding, the measured rate of the overall process from $F \rightarrow X$ is termed $k_{obs} = (k_u * k_i) / (k_f + k_i)$ and half-life ($t_{1/2}$) is given by $\ln 2 / k_{obs}$. k_{obs} at a given temperature is used frequently as an operational measure of stability. Sometimes, the half-inactivation temperature (T_{50}) is used, that is the temperature at which half of the protein is unfolded after a given time of incubation. According to above model, the upper limit of k_{obs} is set by k_u because only U (and not F) will undergo an irreversible transition. At high temperatures, when irreversible processes, such as aggregation and covalent modification of thermolabile amino acid residues, will be fast and therefore $k_i \gg k_f$, $k_{obs} \sim k_u$. Thus, a low k_u value is important to kinetically protect proteins at elevated temperatures (Plaza, I et al. 2000).

1.3 Mechanisms of protein inactivation

Native structure of proteins is held together by a delicate balance of noncovalent forces (e.g., H bonds, ion pairs, hydrophobic and Van der Waals interactions etc). When protein is subjected to high temperatures, these weak interactions are disrupted resulting in unfolding of the protein. Protein unfolding might be reversible or irreversible. A large number of proteins, including many small globular proteins, are found to unfold in simple two-state manner, where unfolding is completely reversible. The loss of secondary and tertiary structure is concomitant with the loss of activity upon unfolding at high temperature, while protein regains its native active conformation upon cooling. However, a large number of mesophilic as well as many thermophilic proteins unfold irreversibly in which protein unfolding transition is followed by another transition in which the unfolded protein undergoes additional interactions or modifications thus making the process irreversible.

1.3.1 Aggregation

Aggregation is the main cause of irreversibility of unfolding transition in majority of proteins. In the native structure of protein, most of the hydrophobic residues are buried

into the core of the molecule, thus minimizing their exposure to the polar solvent (water) as it is entropically unfavorable. Upon unfolding, these hydrophobic residues become exposed to the solvent, which then interact with hydrophobic residues of other unfolding protein molecules in order to minimize their exposure to solvent. This nonspecific intermolecular interaction between the unfolded protein molecules causes aggregation, often leading to precipitation (Tomazic and Klibanov 1988; Chi et al. 2003).

1.3.2 Covalent modifications

All proteins, irrespective of their origins, consist exclusively of the 20 canonical amino acids. Both the integrity of the natural amino acids and the proper folding of the polypeptide chain are important for protein function and stability. Ultimate limit of protein stability must come from the chemical stability of amino acid residues. A number of amino acid residues are known to undergo chemical modifications at elevated temperatures (80-120 °C) (Daniel et al. 1996). The most susceptible are asparagine and glutamine residues, which undergo deamidation leading to succinamide formation, which may follow peptide backbone cleavage. While both residues deamidates, asparagine shows much higher propensity to deamidate than glutamine residues. Beside these, a number of amino acid side chains undergo oxidation. This includes His, Met, Cys, Trp and Tyr residues (Daniel et al. 1996). Besides modification of amino acid side chains, hydrolysis of peptide bonds also occur which happens most often at the C-terminal side of Asp residues, with the Asp-Pro bond being the most labile of all. Other covalent changes which occur at elevated temperature include disruption of disulfide bonds, cis-trans isomerization of Xaa-Pro peptide bonds and racemization of some residues (mainly Asp and Ser) to their D-form (Vieille and Zeikus 2001).

The rates of degradative processes have been found to depend upon the geometry of the amino acid residue, the amino acids in its vicinity, steric and hydrogen bonding constraints and local flexibility. It is observed that in thermophilic proteins the deamidation rates of Asn and Gln are highly reduced in native folded structure, presumably due to steric constraints (Vieille and Zeikus 2001).

1.4 Mechanism of protein thermostabilization

Thermostable proteins maintain the functional state, or the so called “corresponding state” identical to their mesophilic counterpart at extreme temperatures.

This ability is acquired through a combination of many small structural modifications that are achieved with the exchange of some amino acids for others and the modulation of the canonical forces such as hydrogen bonds, ion pair interactions, hydrophobic interactions which are found in all the proteins (Zavodszky et al. 1998; Scandurra et al. 1998). Availability of genome sequence and structural information of several thermophilic and hyperthermophilic organisms have allowed sequence and structural comparison of proteins from such extremophiles against their mesophilic counterparts. These studies have provided important insights into our understanding of the features important in protein thermostability (Jaenicke and Bohm 1998; Kumar and Nussinov 2001; Razvi and Scholtz 2006; Luke et al. 2007). Although, site-directed mutagenesis studies based on these comparative analyses gave mixed results. A detailed comparative analysis of 18 thermophile-mesophile enzyme families revealed that firstly, protein stabilization strategies that are observed in an individual family may not show consistent trend across families and secondly, not all differences among thermophiles and mesophiles could be directly attributed to protein stability (Kumar et al. 2000). It is apparent that there is no single universal mechanism that governs protein thermostability rather different proteins have acquired thermostability by a variety of evolutionary devices. There are multiple molecular mechanisms that effect protein stability and different proteins adopt different mechanisms to improve their thermostability (Petsko 2001). Based on comparison among mesophiles, several factors have been identified including increased core hydrophobicity, increase in the number of hydrogen bonds, ion pairs, better packing efficiency, shortening and stabilization of loops, anchoring of protein termini, helix stabilization, reduction in number of amino acids prone to degradation. The following sections briefly discuss the features that are known to contribute to protein stability.

1.4.1 Amino acid composition

Comparative sequence and structural analyses of thermophilic and mesophilic proteins have shown the differences in the amino acid composition and their role in conferring thermostability (Zhou et al. 2008). Initial statistical analyses comparing amino acid compositions in mesophilic and thermophilic proteins indicated that the properties most correlated with the proteins of thermophiles include higher residue volume, higher residue hydrophobicity, more charged amino acids (especially Glu, Arg and Lys) and fewer uncharged polar residues (Ser, Thr, Asn and Gln) (Haney et al. 1999a). However,

later studies showed that the properties of side chain of amino acids were determinants of thermostability of protein, though thermophilic and mesophilic proteins have both similar polar and non polar contribution to the surface area and compactness (Sadeghi et al. 2006).

Compared to their mesophilic counterparts, thermophilic proteins (a) show marginally higher content of hydrophobic amino acids (Ala, Val, Leu and Ileu) (Chakravarty and Varadarajan 2000); (b) have fewer Gly (Panasik et al. 2000) and higher frequency of Pro (Sadeghi et al. 2006); (c) show less preference for thermolabile residues such as Met, Cys, Asn and Gln (Vieille and Zeikus 2001); (d) show strong preference for replacements (from meso to thermophile) Met→Ala, Cys→Ala, Trp→Tyr, Met→Leu, Cys→Val and Cys→Val (Gromiha et al. 1999b); (e) show preference for polar charged residues (including Arg, Lys, His, Asp and Glu), mostly at the expense of uncharged polar residues (in particular Gln), except His having lower frequency in thermophiles and (f) replacement of Lys with Arg (Chakravarty and Varadarajan 2000).

1.4.2 Hydrophobic interactions

Hydrophobic interactions are considered as main driving force for the folding of globular proteins (Dill 1990). They result in the burial of the hydrophobic residues in the core of the protein. Hydrophobicity is a quantitative estimation of hydrophobic interactions in the core of a protein molecule and it is expressed as the ratio of buried non-polar surface area to the total non-polar surface area of the molecule. There is a direct correlation between the total hydrophobicity of a protein and its thermostability. An estimated average increase in stability of 1.3 ± 0.5 kcal/mol is attained through each additional methyl group being buried in the molecule during protein folding (Pace 1992). In a mutagenesis study on barnase, 15 mutants were generated in which a hydrophobic interaction was deleted (V10A, V36A, V45A, I4A, I25A, I51A, I55A, I76A, I109A, I4V, I25V, I51V, I55V, I76V and I109V). This finding has strong correlation between the degree of destabilization (which ranges from 0.60 to 4.7 kcal/mol) and the number of methylene side chain groups deleted from the hydrophobic core ($r = 0.91$) (Serrano et al. 1992). Improving hydrophobic interactions in protein core, by filling the cavities, as in case of *E. coli* Ribonuclease H1 (Ishikawa et al. 1993), better packing of hydrophobic core, (Haney et al. 1999b; Xu et al. 2003) or improved inter-subunit hydrophobic interactions as in case of 3-isopropylmalate dehydrogenase from the thermophile *Thermus thermophilus* (Kirino et al. 1994) improves protein thermostability.

A recent comparative structural study on hyperthermophilic and thermophilic proteins with their mesophilic counterparts has shown that hyperthermophilic proteins underwent a significant increase in the hydrophobic contact area contributed by those residues composing the α -helices of the structurally conserved regions. The decreased flexibility of α -helices, which may be due to an increased number of buried methyl groups in the protein core and/or better packing of α -helices with the rest of the structure, act as a major factor contributing to the enhanced thermostability of a number of hyperthermophilic proteins (Paiardini et al. 2008).

1.4.3 Hydrogen bonds

A hydrogen bond is formed when two electronegative atoms interact with the same hydrogen atom. The hydrogen is covalently bonded to one of the atoms, which acts as donor, and interacts electrostatically with the other atom, the acceptor. In proteins hydrogen bonds involve mostly nitrogen and oxygen atoms. For two component system, the strongest hydrogen bonds are collinear (Creighton 1993). In protein structures, however, hydrogen bonds are most often found to deviate from linearity with 90% of N—H—O bonds lie between 140 and 180° and they are centered around 158°. For C=O—H, the range is between 90 and 160° and centered around 129° (Baker and Hubbard 1984). In the unfolded state, all the hydrogen bonding partners in polypeptide chain are satisfied by hydrogen bonding with water. Upon folding only some of these are replaced by intra-protein hydrogen bonds, many of which are sub optimal. So, from purely enthalpic point of view, hydrogen bonding contributes little to protein stabilization. However, taking into account the increase in entropy of solvent upon folding of protein, hydrogen bonding contributes favorably to the free energy of folding. Now it is generally accepted that hydrogen bonding contributes favorably to the protein thermostability (Pace et al. 1996).

Hydrogen bonds play a major role in the formation and stability of secondary structural elements such as helices and strands. Overall increase in secondary structural elements is strongly correlated with an increase in stability of the protein. The effect of hydrogen bonds on RNase T1 stability has been carefully studied by mutagenesis and it was found that individual hydrogen bonds contribute an average of 1.3 kcal/mol to the stabilization (Shirley et al. 1992). Thermostability of proteins is correlated with the number of hydrogen bonds and the polar surface that results in increased density of hydrogen bonding with surrounding water (Vogt et al. 1997). A study performed by

Tanner et al. showed a strong correlation between D-glyceraldehyde-3-phosphate dehydrogenase thermostability and the number of charged-neutral hydrogen bonds (Tanner et al. 1996). A structural comparison of *Bacillus licheniformis* α -amylase (BLA) and *Bacillus stearothermophilus* α -amylase (BSTA) explains the basis of higher stability of BLA. This is because of nine fewer hydrogen bonds that costs about 12 kCal/mol in BSTA, were among the main factors that reduced relative stability of BSTA (Suvd et al. 2001). Sequence analysis of putative soluble proteins from complete genomes of eight thermophiles and 12 mesophiles suggest that increase in number of residues involved in hydrogen bonding is one of the key factors responsible for high thermostability (Chakravarty and Varadarajan 2000).

1.4.4 Ion pairs (Salt bridges)

Typical charge-charge interactions between oppositely charged residues are known as salt bridges or ion pairs. They are usually found on the surface of protein but may also be buried. Salt bridges can be formed with the partner atoms in almost any orientation to each other, at distances near the van der Waals contact distance of about 3 Å. The contribution of salt bridges to protein thermostability has been a contentious issue. Thermophilic and hyperthermophilic proteins tend to show increased number of salt bridges as compared to their mesophilic counterparts (Tanner et al. 1996; Vogt et al. 1997), on the other hand mutational studies show that the contribution of salt bridges to stability is often very small. This might be due to the reason that the process of salt bridge formation involves desolvation and immobilization of two ions (Serrano et al. 1990; Waldburger et al. 1995). Salt bridge can stabilize protein structure in a number of manners which include surface or buried salt bridges, ion-pair networks and helix stabilization. Alsop et al. compared 127 mesophilic-thermophilic orthologous protein groups and showed that there was a clear preference for stabilizing acid-base pairs on the surface of thermophilic proteins (Alsop et al. 2003). Another study showed that the amino acid sequence between mesophilic and thermophilic alkaline endo-1,4- β -glucanase was very similar and presence of lysine residues at positions 137, 179 and 194 in the thermozyeme was responsible for thermal stabilization. Replacing the Glu137, Asn 179 and/or Asp 194 with lysine by site-directed mutagenesis made mesophilic enzyme more thermostable (Hakamada et al. 2001).

Optimized electrostatic interactions by increasing the number of salt bridges are clear contributors to enhance thermotolerance of proteins from hyperthermophilic organisms, which is less evident in proteins from thermophilic organisms (Karshikoff and Ladenstein 2001). In parallel with the increasing melting temperature of the protein, ion pairs tend to be organized in networks and are found on the surface of the protein or partially buried at domain or subunit interfaces. The largest ion-pair network is reported in glutamate dehydrogenase from hyperthermophile *Pyrococcus furiosus* (Yip et al. 1995). Another example is the one present at the dimer interface of disulfide oxidoreductase from *P. furiosus* (Ren et al. 1998). Ion pair networks are energetically more favourable than an equivalent number of isolated ion pairs, because for each new pair the burial cost is cut by half, thus only additional residue must be desolvated and immobilized (Yip et al. 1998; Karshikoff and Ladenstein 2001).

1.4.5 Entropy of unfolding, anchoring of loops and docking of N-and C- termini

Thermal unfolding of proteins at high temperatures is caused by a strong increase of entropy change that lowers Gibbs free energy of the unfolding transition (Pal and Chakrabarti 1999). A mutation that decreases the conformational freedom of the unfolded state is expected to increase the free energy of the unfolded state, thereby stabilizing the native state. In the protein's unfolded state glycine has the highest conformational entropy. On the other hand proline residue can adopt only a few conformations and has lowest conformational entropy. Thus, mutations Gly→Xaa or Xaa→Pro (Xaa = any amino acid) have been proposed to decrease the entropy of proteins unfolded state and stabilize the protein (Watanabe et al. 1994; Gaseidnes et al. 2003).

Loop stabilization is considered as a contributing factor for protein thermostability (Vieille and Zeikus 2001). Mechanism of loop stabilization in thermophilic proteins includes shortening or better anchoring to the rest of the molecule through non covalent interactions. Loop shortening can be a consequence of extension or creation of secondary structural elements. In the case of Phosphoglycerate kinase from *Thermotoga maritima*, a hyperthermophile, α 6-helix was extended by four residues (Auerbach et al. 1997). In *Aquifex pyrophilus* superoxide dismutase, loop 2 was extended and played a key role in forming a compact tetramer. This loop has low flexibility and it makes extensive contact with other subunits in the tetramer (Lim et al. 1997).

Loops and N and C termini are usually the regions with the highest thermal factors in a protein crystal structure, which suggest their high degree of flexibility in proteins. They are likely to unfold first during thermal denaturation. Stabilization of the N and C termini involves similar mechanisms to those in loop stabilization. In case of *Thermotoga maritima*, the N-terminus of ferredoxin is fixed to the protein core by three hydrogen bonds (edo-Ribeiro et al. 1996). Whereas in case of phosphoglycerate kinase (Auerbach et al. 1997) and phosphoribosyl anthranilate isomerase (Hennig et al. 1997) from same organism, N and C termini interact with each other for mutual stabilization.

1.4.6 Helix stabilization

α -Helix stabilization can be achieved by substituting residues with low helical propensity with those having a high helical propensity. Beta branched residues (Val, Ile, Thr) were found to destabilize α -helix. Comparative structural analysis of 13 thermophilic proteins with their mesophilic homologues performed by Facchiaano et al. showed that thermostable proteins lack β -branched residues in α -helix (Facchiano et al. 1998). The α -helix carries a considerable dipole moment, the effect of which can be approximated by placing 0.5-0.7 positive unit charge near the N-terminus and same unit negative charge near the C-terminus of the helix (Hol 1985). α -helix dipole can be stabilized by negatively charged residues (e.g. Glu) near their N-terminal end as well as by positively charged residues (e.g. Lys) near their C-terminal end. This kind of helix dipole stabilization has been found in several proteins from thermophiles and hyperthermophiles (Vieille and Zeikus 2001).

1.4.7 Other factors

Besides the factors described above, several other factors have been observed which are responsible for conferring stability to proteins. These include disulfide bonds, which are believed to stabilize protein by decreasing the entropy of the protein unfolded state (Matsumura et al. 1989); binding to metal ions, (mainly divalent metal cations including Mg^{2+} , Co^{2+} , Mn^{2+} Ca^{2+} etc.)(Vieille et al. 2001); reduction of conformational strain, inter-subunit interaction and increase in oligomerization, as has been found in case of many proteins of thermophiles and hyperthermophiles; reduction in solvent accessible hydrophobic surface with improved core packing and post translational modification, mostly, glycosylation (Vieille and, Zeikus 2001; Li et al. 2005). Given that a number of

factors are linked with the protein thermostability and considering the complexity of the protein structure, it is improbable that any one universal stabilization mechanism is responsible for thermostability. Thermostable proteins alter their sequence such that it optimizes the interactions holding their native structural conformations together at elevated temperatures.

1.5 Engineering of protein thermostability

Enzymes are adapted to their particular function in a living cell and are ill suited for industrial applications, due to extremes of pH, temperature and/or salinity. To broaden the industrial applicability of enzymes and to understand protein structure-function relationships, protein engineering is very actively pursued. Approaches to protein (thermostability) engineering can be broadly classified into four strategies. These include “rational designing”, which is based on the information available on sequence-structure-property relationship of the protein, based on which mutations are introduced, thus altering the structure and stability of the protein. Second approach is primarily computational, wherein increased knowledge on structure-stability relationships in proteins, the factors governing thermostability and increased computing powers are used to design mutations in order to generate proteins with increased thermostability. Third approach is “directed or *in vitro* evolution” which mimics the natural evolution, but performed in the lab on an enhanced rate under stringent conditions of screening and/or selection. Finally, the fourth approach is “semi-rational” approach, which combines the positive features of rest of the three approaches.

1.5.1 Rational design

Comparative studies of proteins from extremophiles as well as site-directed mutagenesis studies have revealed that there is no simple set of ‘rules’ for protein stabilization [see above] (Petsko 2001). Nevertheless, several rational approaches of protein stabilization do exist and have been applied successfully to improve thermostability of a variety of proteins. These approaches include improvement of hydrophobic core packing, the introduction of disulfide bridges, stabilization of α -helices, engineering of surface salt bridges, anchoring of loops and termini and introduction of mutations which reduces the entropy of unfolded state (Eijsink et al. 2004).

Entropic stabilization is one of the most promising strategies for improving thermostability, mainly by introduction of disulfides or Xaa→Pro and Gly→Xaa, where Xaa is any amino acid (Gaseidnes et al. 2003). However, a prerequisite is that the mutations should not introduce unfavorable strain in the folded state. Over the past few years, protein surface has become the focus of attention for improving thermostability, mainly by the introduction of salt bridges as well as ionic networks. Electrostatic calculations on proteins are still difficult and hence the prediction of mutation that could improve stability by means of electrostatic interactions is complicated (Sanchez-Ruiz and Makhatadze 2001). However, still it is a favored strategy.

Despite extensive efforts, rational designing of proteins for thermostability has resulted in mixed results. This is because prediction of mutations based on comparative study of sequences of proteins from extremophiles may not be easy as all the differences in sequence need not contribute to improved thermostability. Moreover, successful rational design of proteins is greatly dependent on sufficient knowledge of the mechanisms of thermal denaturation. Targeting specific regions whose unfolding is limiting in protein denaturation process can provide significant stabilization, however, identification of these regions and prediction of the kind of mutation is a difficult task.

1.5.2 Computational design

Computational design is one of the knowledge based approaches based on physico-chemical principles responsible for protein stability as well as on the three dimensional structure. Computer based algorithms are developed to predict protein rigidity and stability to design thermostabilizing mutations. Malakauskas and Mayo have developed an algorithm, which predicts the thermostabilizing mutations based on pair-wise interaction energies between amino acid side chains and between side chain and backbone (Malakauskas and Mayo 1998). Using this algorithm a hyperthermostable variant of the streptococcal protein G β 1 domain was designed. The thermostable variant shows increase in melting temperature from 81 to 100 °C. Structural analysis of the variant reveals that the main stabilizing mechanisms were the release of strained rotamer conformation, increased burial of hydrophobic area and higher helical propensity. Recently, Korkegian et al. have improved thermostability of a model protein 'yeast cytosine deaminase' by using a computational approach that increased the apparent melting temperature by 10 °C and resulted in a 30 fold increase in half life at 50 °C, with

no reduction in catalytic efficiency (Korkegian et al. 2005). Mutations were based on Rosetta design, which is one of the most popular algorithms for energy functions for evaluating the fitness of a particular sequence to a given fold. This program requires a backbone structure as an input and generates sequences that have the lowest energy for that fold (Kuhlman, Baker 2000). Nowadays, several programs such as WHAT IF (Vriend 1990), ProTherm (Gromiha et al. 1999a; Saraboji et al. 2006; Gromiha 2007) are available on various servers which can be used to predict the stabilizing effect of a mutation in protein, which can be efficiently utilized to predict thermostabilizing mutations in protein.

1.5.3 Methods of directed evolution

Although engineering proteins by “rational designing” is successful in number of cases, the success rate of this approach is limited. Rational designing is based on the thorough understanding of the sequence-property relationship, which is not clear in many cases. During the last two decades, another approach known as “directed evolution” or “*in vitro* evolution” has been developed to engineer biomacromolecules (mainly proteins), which does not rely on the prior knowledge of sequence and function relations. This approach has been successfully used to improve those properties of proteins, whose molecular basis is poorly understood, e.g., catalyzing the synthesis of synthetic polymer. The term “directed evolution” encompasses a number of experimental techniques that reproduce, the wheels of natural selection i.e., generation of diversity and selection (screening) of the fittest on an accelerated time scale in laboratory (Farinas et al. 2001; Bloom et al. 2005; Yuan et al. 2005; Bershtein and Tawfik 2008; Jackel et al. 2008).

Although the underlying principles of both processes are similar, there are some striking differences between these two processes. Directed evolution is very rapid compared to the natural selection, wherein typically variation in gene sequences and selection can be achieved in a matter of weeks. The fixation of a positive to a selection pressure is a chance event in case of natural evolution, while in case of directed evolution it is predetermined with well controlled parameters set by the investigator (Bloom et al. 2005).

The property or function of protein is defined by its three dimensional structure, information of which resides in its sequence or primary structure. Any change in the primary structure brings about changes in the tertiary structure, which will then ultimately reflect in the altered property or function. Thus, any alteration in the property requires

suitable change in the primary sequence of the protein. A protein can adopt innumerable sequence variations. Theoretically, the whole set of potential sequence variants, starting from a given primary sequence, are collectively referred to as its 'sequence space'. For a 100 amino acid long protein, the number of sequence variants possible is 20^{100} which is roughly equal to 10^{130} , an astronomical number (Saven 2002). However, all the structures arising out of it may not give rise to a functional protein. So, a small subset of this vast sequence space, which can give rise to a specific function, is termed as 'functional space'. Since, the sequence space is vast, it is presumed to have not one but many sub sets of functional space representing a diverse range of functions. Therefore, by changing the sequence of the protein, it is possible to hop from one functional space to other giving rise to the evolution of novel functions. Methods of directed evolution explore the sequence space (around the functional space) by making changes in the sequence without prior knowledge of the functional space and then the desired variants, occupants of the different functional space, are chosen from this library with appropriate screens.

Methods of directed evolution have advantages over rational methods as it explores additional sequence space and does not require prior structural information (Arnold and Volkov 1999). This becomes particularly important in case of complex properties in which the understanding of structure and function is uncertain. This strategy not only helps in improving proteins towards these non-natural complex properties but also helps in understanding these complex phenomenon by studying beneficial mutations and their contribution in improving the property in a retrospective analysis (Arnold et al. 2001; Wintrode and Arnold 2000). However, this technique has some limitations too. Given that the functional space is a small subset of sequence space and the active folded conformation of proteins are only marginally stable, random change in sequence mainly results in inactive variants compared to the active and "desired" variants (Arnold 1998). This imposes a severe experimental constraint on the screening strategies for identifying the improved active variants from the inactive ones. Hence, a major challenge in these methodologies lies in developing smart and efficient screening strategies (Cohen et al. 2001). Despite these limitations, directed evolution has emerged as a powerful tool in improving various properties of proteins (mainly enzymes) including stability (Eijsink et al. 2005), activity (Wong et al. 2004), substrate specificity (Iffland et al. 2001), enantioselectivity (Reetz 2004), pH compatibility (Bessler et al. 2003), stability towards

organic solvents (Moore and Arnold 1996) etc. Also, a number of proteins altered by directed evolution have found several commercial applications.

Any directed evolution protocol has two basic steps: generation of genetic diversity and screening of variants displaying improved property/function. Thus, the success of any directed evolution exercise to improve any property/function critically depends on two parameters: the method chosen to create genetic variations and the method chosen to select/screen the desired variants in the random population.

1.5.3.1 Generation of variants

Various methods for creating libraries of random mutants have been developed (Table 1.1). The choice of the method to generate genetic variation must take into account the factors of the throughput of selection procedure and the fitness distribution expected among the variants which in turn depends on the initial strategy chosen to create library diversity as well as the subsequent treatment met out to the improved variants such that only beneficial mutations are carried forward and the deleterious ones are eliminated in subsequent generations (Kurtzman et al. 2001).

The most widely used technique for diversity generation in one single gene is error prone PCR (epPCR) (Cadwell and Joyce 1992). Other techniques including the use of chemical mutagens as well as mutator strains have also been reported. Mutant generation by epPCR is convenient and well established where the frequency of mutation can be controlled easily by altering the concentration of PCR components, mainly $MnCl_2$, dNTPs ratio or by altering the initial amount of template taken for amplification. A low error rate of single amino acid substitution per sequence per generation accumulates mostly adaptive mutations, whereas higher error rates generate neutral and deleterious mutations which confounds analysis (Arnold 1998). The probability of improvement by a single mutation is small which decreases rapidly when multiple simultaneous mutations are made. However, recently it was reported that mutant libraries with 15 to 30 mutations per gene, on an average, have orders of magnitude higher variants that retain function than what was expected from low mutation rate trend (Drummond et al. 2005). Variants with improved or novel functions were isolated disproportionately from these high error rate libraries, indicating that high mutation rates unlock regions of sequence space enriched in positive coupled mutations. It was shown that while low mutation rates result in many functional

sequences, only a small number are unique. On the other hand, high mutational rates generate mostly unique sequences, but few retain function (Drummond et al. 2005).

Another approach of generating diversity employs the use of recombination techniques applied to gene pools derived from random mutagenesis methods and/or from natural variants. One of the key technology in directed evolution is ‘gene shuffling’, i.e., *in vitro* recombination of two or more genes encoding natural or engineered variants of the protein (Stemmer 1994). This method combines beneficial mutations while purging out neutral and deleterious ones. Thus, at any time point during an *in vitro* evolution procedure, genes may be crossed *in vitro*, leading to a much more efficient sampling of relevant sequence space (Cramer et al. 1998). Since Stemmer’s description of gene shuffling method, several other methods for random recombination *in vitro* have been described (Table 1.1). The original shuffling protocols depend on the degree of sequence similarity between genes in the gene pool (usually >70%). Moreover, these methods are not completely random in the sense that not all combinations of sequences/mutations are equally likely to occur (sequence and distance dependent). Recently, several methods of recombination of genes with low sequence similarity have been described (Sen et al. 2007). Some of these procedures make use of truncation and ligation of DNA fragments. Other procedures make use of specific primers or the insertion of tag sequences to induce specific similarity independent recombination events between sequences. These newer procedures often require careful experimental design and are less straightforward than standard shuffling.

Recently, focus has been shifted to generate “smart” libraries in contrast to random ones generated earlier. Several statistical and computational approaches have been utilized to generate focused libraries, which significantly reduce screening effort along with substantial improvement in properties (Wong et al. 2007; Fox and Huisman 2008).

1.5.3.2 Screening and selection procedures

Success of any directed evolution critically depends on the screening or selection strategy employed during the course of evolution. The outcome of the whole exercise depends on the selection pressure applied, which should be carefully designed and executed depending upon the final outcome required (Arnold et al. 2001). The basic rule of directed evolution is “you get what you have screened for”, highlighting the importance

Table 1.1 Examples of established methods used to create genetic variation in directed evolution.

Method	Description	Pros	Cons	Reference
Random mutagenesis (error-prone PCR)	Point mutations introduced by polymerase under altered reaction conditions	Both transitions and transversion are introduced; frameshift and deletions are rare	Limited sequence space sampled; bias due to polymerase and gene sequence	(Cadwell and Joyce 1992) (Chen and Arnold 1993)
Site-saturation mutagenesis (SSM)	Oligo based substitution of an amino acid to all other 19, by PCR	Complete sampling of sequence space of given location	Can be performed only at limited positions; screening of variants is required	(Miyazaki and Arnold 1999)
Gene site saturation Mutagenesis (GSSM)	Same as SSM but performed over entire length of gene	Same as SSM	Large number of oligos need to be synthesized; screening of large number of variants is required	(DeSantis et al. 2003) (Kretz et al. 2004)
Homology based methods				
DNA/Family Shuffling	Recombination of DNase I digested fragments by PCR	Combines beneficial mutation while purges out deleterious ones; multiple parents can be recombined	Bias can introduced due to sequence homology; all products are not equally represented	(Stemmer 1994) (Crameri et al. 1998)
Staggered extension process (StEP)	PCR induced recombination by template switching of partially extended products	One step process; no risk of DNase I digestion	Sequence homology based; requires standardization	(Zhao et al. 1998)
Random chimeragenesis on transient templates (RACHITT)	Using small oligo linkers, regions of low homology recombined	Regions of low homology are recombined	Creation of ss DNA templates by exonuclease digestion followed by phagemid cloning	(Coco et al. 2001) (Coco 2003)

Cont...

Method	Description	Pros	Cons	Reference
Combinatorial libraries enhanced by recombination in yeast (CLERY)	PCR and <i>in vivo</i> recombination based for recombining multiple parental genes in yeast	Useful for proteins that need to be expressed in yeast	Bias can introduced due to sequence homology	(Abecassis et al. 2000)
Multiplex PCR based Recombination (MUPREC)	Multiplex PCR based generation of fragments with mutation which were subsequently assembled to full length genes by PCR	Reduces introduction of novel point mutations; high frequency of recombination without wild type background	Requires synthesis of oligo for each mutation and need parental gene cloned in two separate vectors	(Eggert et al. 2005)
Non-homology based methods				
Exon-shuffling	Generates mosaic proteins by splicing together of various exons	High percentage of folded proteins, maintains structural domains	Restricted to intron containing (eukaryotic) genes	(Kolkman and Stemmer 2001)
Incremental truncation for the creation of hybrid enzymes (ITCHY)	Chimeric gene library by non-homologous recombination by nested gene deletions and fusions.	No sequence homology required. Full length gene products are selected	Each progeny has only one crossover point. May induce amino acid deletions or duplications at junctions	(Ostermeier et al. 1999)
Sequence homology independent protein recombination (SHIPREC)	Generates variants with single crossover without requiring sequence homology	Absolutely unrelated genes can be recombined; junctions are randomly distributed	Only one crossover per hybrid per round	(Sieber et al. 2001)
Nonhomologous random recombination (NRR)	Based on DNaseI digestion, blunt-end ligation and/or extension, and capping using two asymmetrical hairpins to stop the extension	Higher flexibility in modulating fragment size and crossover frequency and number of parental genes	The ligation and extension steps with the capping due to DNA hairpins might be tricky to execute	(Bittker et al. 2004)

of this step. The selection pressure applied should be such that all the variants falling within the desired functional space are sampled at the same time eliminating all the undesired variants falling outside it (Arnold et al. 2001). Selection pressure can be applied by performing “selection” or “screening” of desired mutants in the random population. While selection allows all the molecules in the pool to be tested for function at the same time with the active/desired ones being highlighted (by conferring some advantage to the survival of host under selection conditions), while during screening, each member of the pool is accessed individually and the best one is (or better ones are) selected on the basis of comparative analysis (Leemhuis et al. 2005; Boersma et al. 2007). While selection methods can assess large populations, it is not always easy to perform selection for most of the properties. Screening strategies, on the other hand, can be developed easily for many properties, but are more cumbersome to perform (Olsen et al. 2000).

The desired features of any selection or screening strategy include: a, should be preferably high throughput; b, should be sensitive to incremental changes in the desired function (to pick up weak signal); c, should be easily reproducible and, finally, should be robust (Cohen et al. 2001). With the advancement in the methods involved in the diversity generation, there has been improvement in the methods employed for screening or selection. The library sizes which can be easily assessed by present day strategies range from 10^3 - 10^{13} . Although this number is much smaller than the diversity which can be generated, these procedures have met with significant success in improving a number of properties. The most important factor governing the identification of desired variants apart from the size and nature of the library is the linkage of genotype to phenotype in the library (Leemhuis et al. 2005). In nature this is achieved by compartmentalizing genotype and phenotype within a cell. Based on this, many screening strategies are designed on *in vivo* models such as, bacteria (*E. coli*, *B. subtilis*), yeast or mammalian cell lines. The screening of the proteins can be performed either in intracellular or extra cellular milieu. However, several limitations are associated with this strategy such as inability to handle large populations, low transformation efficiency, protein refolding, protein toxicity etc. These limitations can be circumvented by the use of many *in vitro* strategies, such as many “display” technologies. These techniques not only handle much larger population but also allow screening under harsh *in vitro* conditions (Matsuura and Yomo 2006).

579-362
AR518
82



TH-16502

Conventional screening methodologies

Conventional approaches involve expression of random mutant library in a suitable expression system (*E. coli*, *B. subtilis*, yeast etc.) followed by screening using an easily identifiable phenotypic trait. These systems are based either on solid phase or liquid phase format of assays. Solid phase screening relies on product solubilization following an enzymatic reaction that gives rise to a zone of clearance (Song and Rhee 2001), a fluorescent product (Kouker and Jaeger 1987) or a strongly absorbing (chromogenic) product (Olsen et al. 2000). Toxic assays can be performed by transferring colonies followed by cell lysis on filter membranes (Matsumura et al. 1999). Moreover, quantitative estimation can be performed by using digital imaging techniques (Joo et al. 1999; Joern et al. 2001). However, many assays can not be implemented in solid phase format necessitating screening to be performed at the level of individual clones by growing them separately in microtitre plates. Libraries are screened in liquid phase format of assays following an enzymatic reaction leading to chromogenic or fluorescent product formation (Cedrone et al. 2005; Leroy et al. 2003). Although these assays are highly sensitive, they are more time consuming, limiting the size of the mutant population which can be screened. However, with the advent of robotic automation and colony picking technologies, these assays can screen a population of 10^3 - 10^6 on routine basis (Geddie et al. 2004; Pohn et al. 2007).

Advanced screening methodologies

Several *in vitro* strategies of screening have been developed which not only circumvent the limitations faced by conventional *in vivo* screening techniques, but are also capable of handling much larger population of mutants; in the range of 10^6 - 10^9 (Matsuura and Yomo 2006). Several such techniques have been collectively referred as 'display' technologies. In 'phage display' the protein of interest is fused with the phage coat proteins which then display the desired protein on their surface either in low or high copy number (Fernandez-Gacio et al. 2003). Another technique is 'cell surface display' in which the protein of interest is expressed and displayed over the microbial surface (*E. coli*, yeast) which is then coupled with flow cytometry based screening (Becker et al. 2004). Various other display technologies, based on protein synthesis machinery consisting of mRNA display (Wilson et al. 2001; Takahashi et al. 2003), ribosome display (Hanes et al. 2000; Pluckthun et al. 2000; Lipovsek and Pluckthun 2004; Yan and Xu 2006) have also

been reported. These techniques are particularly useful when screening for either activity or binding properties of enzymes.

Droplet based strategies (also known as *In vitro* Compartmentalization; IVC) are much more efficient as they are capable of screening $>10^9$ variants. It makes use of aqueous core of reverse micelle or water in oil emulsions as artificial compartments for *in vitro* protein synthesis and product capture. These *in vitro* compartments have all the machinery required for coupled transcription-translation as well screening of the product which can be screened for the desired variant using FACS (Bernath et al. 2004; Griffiths and Tawfik 2006; Bershtein and Tawfik 2008).

1.5.3.3 Screening for enhanced protein stability

Development of a screen for protein stability requires the knowledge of the kind of thermal unfolding mechanism the desired protein follows. While most of the proteins/enzymes of industrial importance undergo irreversible thermal unfolding, screening for high residual activity after thermal challenge is sufficient to improve thermostability. However, in case of reversible thermal unfolding, screening for high activities at elevated temperature is important (Bommarius et al. 2006). All these screenings are performed in 96- or 384- well plates, which severely limit the capacity of the screen. However, screening capacity can be increased by pooling the variants (Polizzi et al. 2006b). Screening protocols for stability, based on generating chemical denaturation profiles of the mutants, in 96-well plate format has been reported (Edgell et al. 2003; Aucamp et al. 2005). Another technique of measuring stability of proteins in high throughput format has been reported. Referred as SUPREX (stability of unpurified proteins from rates of H/D exchange) makes use of hydrogen-deuterium exchange in denaturant solutions coupled with mass spectroscopy to detect the extent of exchange (Ghaemmaghami et al. 2000).

The use of selection strategies for increased stabilities are still rare; however, thermophilic bacterium *Thermus thermophilus* has been successfully used as a host to perform selection at high temperature in few studies (Tamakoshi et al. 2001). Another useful screening strategy is 'phage display' using which stabilized protein variants can be selected after pretreatment designed to denature unstable variants (by using heat, guanidinium chloride or protease treatment). The approach, known as 'Proside' (protein stability increase by directed evolution) (Sieber et al. 1998), involves insertion of the

protein of interest into the multidomain capsid protein important for phage infectivity. Infectivity is lost when the domains are disconnected by proteolytic cleavage of unstable protein inserts. The method is best suited for small, monomeric proteins (Martin et al. 2001; Wunderlich and Schmid 2006), as it requires that the insertion does not abolish the assembly or infectivity of the phage.

Directed evolution has emerged as an important tool to engineer stability of a variety of proteins. It has been used not only to improve thermostability of a number of enzymes of industrial importance such as amylases (Kim et al. 2003), xylanases (Miyazaki et al. 2006; Xie et al. 2006), proteases (Miyazaki et al. 2000; Strausberg et al. 2005), lipase and esterases (Giver et al. 1998; Zhang et al. 2003; Acharya et al. 2004), polymerases (Ghadessy et al. 2001) etc. but have also been used to improve thermostability of a number of other proteins (Wunderlich et al. 2005a; Wunderlich and Schmid 2006). A list of few successful examples of improving protein thermostability by directed evolution is presented in Table 1.2.

1.5.4 Semi-rational approaches

Recently semi-rational approaches have been explored to improve the properties of proteins by combining the benefits of both rational designs with combinatorial design. To enhance success it is possible to reduce the sequence space and the number of variants to be screened by utilizing the information available from structure, sequence and the experimental data (Chica et al. 2005; Bommarius et al. 2006). Various strategies employed in semi-rational approaches are targeted randomization of defined residues based on structural knowledge, whole gene random mutagenesis to identify potentially advantageous positions followed by fine tuning by site saturation mutagenesis besides other computational and sequence homology based approaches.

Table 1.3 lists few selected examples in which protein stability was improved by different semi-rational approaches. Targeted saturation mutagenesis of selected residues along with Proside improved T_m of Streptococcal G β 1 protein by 24 °C (Wunderlich et al. 2005b). In another study, sequential randomization of 12 positions in a calcium free variant of subtilisin improved half-life of thermal inactivation at 75 °C by 1500 fold (Strausberg et al. 2005). Half-life of *B. subtilis* lipase was improved by 300 fold by

Table 1.2 Selected examples of stability improvement by directed evolution.

Protein/Enzyme	Strategy	Improvement	Reference
Xylanase (EvXyn11)	Gene site saturation mutagenesis, recombination	Increase in T_m by 25 °C Catalytic properties equivalent to parent	(Dumon et al. 2008)
<i>E. coli</i> phytase AppA2	EpPCR, screening in <i>S. cerevisiae</i>	Increase in T_m by 7 °C. Catalytic efficiency improved by 152% than wild type	(Kim and Lei 2008)
Fructosyl peptide oxidase	Random mutagenesis, then site-directed and cassette mutagenesis	Half-life at 50 °C was 79.8 fold higher; improved protease tolerance	(Hirokawa et al. 2008)
<i>B. subtilis</i> Xylanase A	Two rounds of EpPCR then one round of DNA shuffling	Increase in T_m by 20 °C	(Ruller et al. 2008)
Phosphite dehydrogenase	One round of random mutagenesis, site-saturation mutagenesis at 14 positions	Half-life at 45 °C improved by 23,000 fold	(McLachlan et al. 2008)
beta-glucuronidase	Four rounds of DNA shuffling and screening	Six mutations to be present concurrently, confers stability RA at 80 °C after 10 min, 75%	(Xiong et al. 2007)
Lactate oxidase	EpPCR and DNA shuffling	Increase in half-life at 70 °C by 36 fold compared to wt.	(Minagawa et al. 2007)
<i>B. subtilis</i> lipase	Iterative saturation mutagenesis at six locations	Increase in T_{50}^{60} by 45 °C	(Reetz et al. 2006)
<i>B. subtilis</i> GH11 Xylanase	Random point mutagenesis, saturation mutagenesis and recombination	Increase in T_{50}^{10} from 58 to 68 °C; increase in T_{opt} from 55 to 65 °C	(Miyazaki et al. 2006)
beta-Lactamase	Three round of RM, DNA shuffling and metabolic selection of truncated variants	Increase in T_{opt} from 35 to 50 °C with full retention of activity at low temperature	(Hecky and Muller 2005)
bleomycin binding protein	Random mutagenesis with selection at 77 °C in <i>Thermus thermophilus</i> HB27	Increased T_m by 17 °C, unfolds at 85 °C in the absence and 100 °C in the presence of bleomycin	(Brouns et al. 2005)

Cont...

Protein/Enzyme	Strategy	Improvement	Reference
<i>B. subtilis</i> lipase	Random mutagenesis followed by structure guided site-directed mutagenesis	Increase in half-life at 55 °C by 300 fold	(Acharya et al. 2004)
Xylanase (GH11 XYL7746)	Gene site saturation mutagenesis; screening of combinatorial variants; recombination by site-directed mutagenesis	Increase in T_m from 61 °C to 96 °C	(Palackal et al. 2004)
<i>Thermus</i> maltogenic amylase	Random mutagenesis by DNA shuffling and subsequent recombination	Shift in T_{opt} from 60 to 75 °C and T_m by 10.9 °C. Half-life at 80 °C increased by 170 fold	(Kim et al. 2003)
<i>Candida antarctica</i> Lipase B	Two rounds of EpPCR	Half-life at 70 °C increased by 20 fold with increase in catalytic efficiency	(Zhang et al. 2003)
N-Carbamyl-D-amino acid amidohydrolase	DNA shuffling	Increase in T_{50}^{30} from 61 to 73 °C	(Oh et al. 2002)
Taq DNA polymerase	Three rounds of compartmentalized self replication	Improvement in thermostability by 11 fold	(Ghadessy et al. 2001)
<i>B. subtilis</i> cold Shock protein	Site saturation mutagenesis and Proside	Increase in T_m by 28 °C	(Martin et al. 2001)
Prolyl endopeptidase	EpPCR, screening by active staining on membrane filters	Increased half-life at 60 °C by 60 fold	(Uchiyama et al. 2000)
Psychrophilic protease subtilisin S41	Random mutagenesis, saturation mutagenesis and DNA shuffling	Increase in half-life at 60 °C by 500 fold with enhancement in catalytic efficiency by 3 fold	(Miyazaki et al. 2000)
<i>B. subtilis</i> subtilisin	EpPCR, recombination	Increased half-life at 65 °C by 200 fold; shift in T_{opt} by 17 °C	(Zhao and Arnold 1999)
p-nitrobenzyl esterase	Six generation of random mutagenesis, screening and recombination.	Increase in T_m by 14 °C with no loss of activity at lower temperatures	(Giver et al. 1998)

EpPCR, Error-prone PCR; T_m , melting temperature; RA, residual activity; wt, wild-type; T_{50}^x , Temperature at which 50% of initial activity remains upon incubation for 'x' minutes; T_{opt} , temperature of optimal activity.

Table 1.3 Selected examples of improvement in protein stability obtained through semi-rational approaches.

Method	Protein	Improvement	Variants screened ^a	Reference
Error-prone PCR, DNA shuffling and selection of variants with truncated termini	β -Lactamase	T_{opt} shift from 35 °C to 65 °C	42 600	(Hecky and Muller 2005)
Targeted saturation mutagenesis and Proside	Streptococcal G β 1	Increased T_m by 24 °C	100	(Wunderlich et al. 2005b)
Successive saturation mutagenesis at selected residues	Calcium-free subtilisin	Half-life at 75 °C increased 1500 fold	3300	(Strausberg et al. 2005)
Random mutagenesis followed by site-saturation mutagenesis	Phosphite dehydrogenase	Half-life at 45 °C increased 23,000 fold	--	(McLachlan et al. 2008)
Error-prone PCR then site-directed mutagenesis	<i>B. subtilis</i> lipase	Half-life increased 300 fold	4000	(Acharya et al. 2004)
Gene-site saturation mutagenesis and Gene reassembly	XylanaseXYL7746	Increased T_m by 35 °C (to 96 °C)	72 500 + 4000	(Palackal et al. 2004)
Iterative saturation mutagenesis	<i>B. subtilis</i> lipase	Increase in T_{50} by 60 °C ^b	8700	(Reetz et al. 2006)
Calculation of optimal mutation load	<i>Candida antarctica</i> lipase B	Increased thermostability by 7.5 fold	10 000	(Chodorge et al. 2005)
Consensus method	Phytase	Increased T_m to over 33 °C	44	(Lehmann et al. 2002)
Combinatorial consensus mutagenesis	β -Lactamase	Increased T_m by 9.4 °C	90+352 ^c	(Amin et al. 2004)
Consensus method and structural Information	Penicillin G acylase	Increased half-life by 3 fold	21	(Polizzi et al. 2006a)

^a Number of variants generated or screened as reported in published data.

^b T_{50} : Temperature at which enzyme loses 50% activity upon incubation for 60 minutes.

^c Two rounds of evolution were employed.

identifying mutations by error-prone PCR, elimination of unfavorable mutations using structural information and the rules on amino acid compatibility (Acharya et al. 2004). Using consensus approach, which is based on the multiple sequence alignment of closely related proteins, thermostability of phytase was increased by substitution of an amino acid residue in a protein for the one that is predominant in the protein family. By synthesizing the consensus sequence determined from an alignment of 13 phytases, a more stable phytase was obtained (T_{opt} increased by 15-26 °C over parents) (Lehmann et al. 2002).

1.6 Protein thermostability, flexibility and catalytic activity

Apart from understanding the factors responsible for protein thermostability, the relationship between stability and activity is also an active area of investigation. Several studies show that thermophilic enzymes show much reduced activities at mesophilic temperature compared to their mesophilic counterparts (Vieille and Zeikus 2001). This is explained by the hypothesis that thermophilic enzymes are more rigid (less flexible) than their mesophilic homologues at mesophilic temperatures. This is supported by the studies measuring conformational flexibility of homologous thermophilic and mesophilic enzymes as probed by fluorescence quenching (Varley and Pain 1991), hydrogen deuterium (H/D) exchange (Zavodszky et al. 1998; Hernandez et al. 2000), molecular dynamics (MD) simulations (Lazaridis et al. 1997), and neutron scattering (Fitter and Heberle 2000). Many of these studies have found that the conformational flexibility of thermophilic enzymes at room temperature is considerably reduced compared to mesophilic enzymes, while both show comparable flexibility near their physiological temperatures (Zavodszky et al. 1998). This suggests that more stable proteins are less prone to have their structures perturbed by thermal fluctuations and therefore appear stable. Also, this reduced flexibility has functional consequences explaining the basis of reduced activity of thermophilic proteins at low temperatures. Conformational fluctuations are important for enzyme catalysis, but these fluctuations, if too large, can also lead to loss of structure and activity. Since the magnitude of fluctuations will depend on the available thermal energy ($k_B T$), evolution has modified the strength and number of stabilizing interactions in enzymes to achieve the optimal balance of stability and flexibility at a given temperature. Large increase in temperature increases flexibility thus causing loss of structure while large decrease in temperature makes enzyme too rigid, hence less active, affecting even those fluctuations which are necessary for catalysis.

Several studies, which are in marked contrast to those mentioned above, are also observed. Conformational flexibility of ruberodoxin from hyperthermophile *Pyrococcus furiosus* on millisecond time scale, as measured by NMR-monitored H/D exchange, was found to be comparable to that of its mesophilic homologue at room temperature (Hernandez et al. 2000). Room temperature dynamics of a pair of mesophilic and thermophilic α -amylases, found no significant difference in dynamics as monitored by hydrogen exchange, while increased mobility on picosecond timescale in the thermophilic protein, as measured by neutron scattering (Fitter and Heberle 2000). Proteins exhibit varied motions from 0.01 to 100 Å on distance scale and 10^{-15} to 10^3 seconds on time scale. Based on molecular dynamics simulations, it was argued that there is no single measure of protein flexibility, a protein can be rigid on a nanosecond scale but flexible on millisecond scale (Lazaridis et al. 1997). Adaptation of protein to high temperature results in a variety of alterations in dynamic behavior, including reduction in certain types of motions (which might cause protein unfolding) and increase in others (which enhances protein stability and activity) (Wintrode et al. 2003). Characterization of several hyperthermophilic enzymes, that are more active than their mesophilic counterparts at mesophilic temperatures (Sterner et al. 1996; Ichikawa, Clarke 1998) and generation of several laboratory evolved mesophilic enzymes (Giver et al. 1998; Zhao and Arnold 1999) which are more thermostable as well as more active at low temperature than their parents suggest that thermostability is not incompatible with high activity at moderate temperatures. The high catalytic activity at moderate temperatures suggests that these enzymes combine local flexibility in their active sites, which is responsible for their high activity at low temperatures, along with high overall rigidity, which is responsible for their thermostability (Wintrode and Arnold 2000).

For investigating the structure-stability-activity relations we have chosen lipase from *Bacillus subtilis* as a model protein in this study. To generate variance to study these aspects in lipase, we employed directed evolution approach.

1.7 Lipases

Lipases (triacylglycerolacylhydrolase, E.C. 3.1.1.3) are hydrolases, which act under aqueous conditions on the carboxyl ester bonds present in triacylglycerols to liberate fatty acids and glycerol (Figure 1.1). The natural substrates of lipases are long-chain

triacylglycerols, which have very low solubility in water, and the reaction is catalyzed at the lipid-water interface. Under micro-aqueous conditions, lipases also possess the unique ability to perform the reverse reaction i.e. esterification. Besides being lipolytic, lipases often also show esterase, phospholipase, cutinase and amidase activities. Thus, lipases show a diverse substrate range, although they are highly specific as chemo-, regio- and enantioselective catalysts (Jaeger et al. 1994; Jaeger and Reetz 1998; Jaeger and Eggert 2002; Gupta et al. 2004).

Lipases are ubiquitous in nature and are produced by various plants, animals and microorganisms. Lipases of microbial origin, mainly bacterial and fungal, represent one of the most widely used class of enzymes in biotechnological applications. Although a number of lipase-producing bacterial sources are available, only a few are commercially exploited as wild or recombinant strains (Jaeger and Eggert 2002; Gupta et al. 2004). Of these, the important ones are: *Achromobacter*, *Alcaligenes*, *Bacillus*, *Burkholderia*, *Chromobacterium* and *Pseudomonas*.

1.7.1 Substrate specificity

Based on substrate specificity, microbial lipases may be divided into three categories namely nonspecific, regiospecific and fatty acid specific. Non specific lipases act at random on the triglyceride molecule and result in the complete breakdown of triglyceride to fatty acids and glycerol. Examples of this group of lipases include those from *S. aureus*, *S. hyicus*, *Corynebacterium acnes* and *Chromobacterium viscosum* (Jaeger et al. 1994). In contrast, regiospecific lipases are 1,3-specific lipases which hydrolyze only primary ester bonds (i.e. ester bonds at C1 and C3 atoms of glycerol) and thus hydrolyze triglycerides to give free fatty acids, 1,2(2,3)-diacylglyceride and 2-monoacylglycerol. Examples of this group include lipases from *Bacillus* sp., *B. subtilis* 168, *Bacillus* sp. THL027, *Pseudomonas* sp. F-B-24, *P. aeruginosa* EF2 and *P. alcaligenes* 24 (Gupta et al.

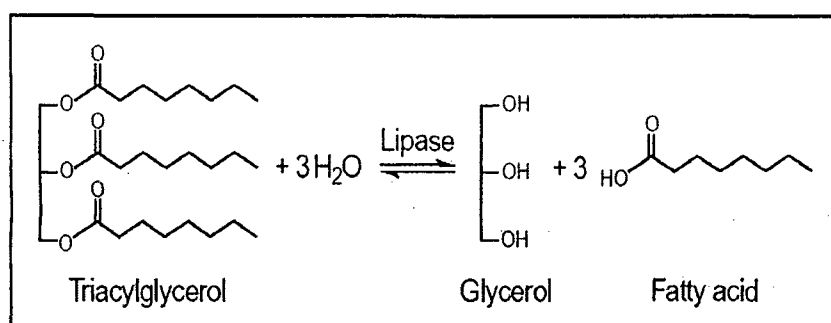


Figure 1.1 The catalytic action of lipases. A triglyceride can be hydrolyzed to form glycerol and fatty acids, or the reverse (synthesis) reaction can combine glycerol and fatty acids to form triglyceride.

2004). Lipase from *Penicillium camembertii* shows a different selectivity as it reacts only on mono- and diacylglycerols, whereas the one from *Bacillus sterothermophilus* reacts only on monoacyl glycerols. The third group comprises fatty acid-specific lipases, which exhibit a pronounced fatty acid preference. While lipases from *Bacillus* sp., *P. aicaligenes* EF2, and *P. alcaligenes* 24 show specificity for triglycerides with long-chain fatty acids, while lipases from *B. subtilis* 168, *Bacillus* sp. THL027, *P. fluorescens*, *C. viscosum* prefer small or medium chain triacylglycerols (Gupta et al. 2004). Lipase from *S. aureus* 226 shows a preference for unsaturated fatty acids (Muraoka et al. 1982).

Another important property of lipases is their enatio-/stereoselective nature, wherein they possess the ability to discriminate between the enantiomers of a racemic pair. Such enantiomerically pure or enriched organic compounds are steadily gaining importance in the chemistry of pharmaceutical, agricultural, synthetic organic and natural products (Reetz 2001). Most lipases from *Pseudomonas* family fall in this category. The stereospecificity of a lipase depends largely on the structure of the substrate, interaction of the active site and the reaction conditions (Reetz and Jaeger 1998).

1.7.2 Structure

The three dimensional structures of the fungal lipase *Rhizopus miehei* and the human pancreatic lipase, were first to be reported in 1990 (Brady et al. 1990; Winkler et al. 1990). Since then, structures of many lipases, both from fungal and bacterial sources, have been reported (Cygler and Schrag 1997; Cygler et al. 2008). These enzymes span a wide range of molecular weight, from ~19 kDa (cutinase) to ~60 kDa (*Candida rugosa* lipase). All of them, with the exception of pancreatic lipases, have only one domain. Comparative structural analysis of lipase with other enzymes as diverse as haloalkane dehalogenase, acetylcholinesterase and serine carboxypeptidase revealed that all these enzymes share a common folding pattern which was named as α/β hydrolase fold. The canonical α/β hydrolase fold consists of a central, mostly parallel β sheet of eight strands with the second strand antiparallel (Figure 1.2). The parallel strands β 3 to β 8 are connected by α helices, which pack on either side of the central β sheet. Excursions of the peptide chain at the C-terminal ends of strands in the C-terminal half of the β sheet form the binding subdomains of the α/β hydrolase fold proteins. They differ substantially in length and architecture, in agreement with the large substrate diversity of these enzymes. All bacterial

lipases structure have α/β hydrolase fold with some variations. The lipases from *B. glumae*, *B. cepacia*, and *C. viscosum* have six parallel β strands ($\beta 3$ to $\beta 8$) in the central β sheet of the α/β hydrolase fold (Lang et al. 1996). The *P. fluorescens* carboxylesterase contains seven strands ($\beta 2$ to $\beta 8$), while that of *S. exfoliates* lipase has the full canonical α/β hydrolase fold with one extra antiparallel β strand added as strand 9 (Kim et al. 1997a).

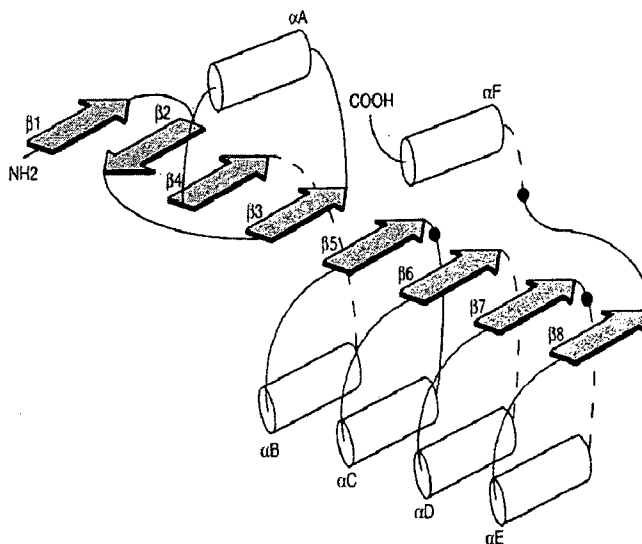


Figure 1.2 Canonical fold of α/β hydrolases.

The active site of the α/β hydrolase fold enzymes consist of three catalytic residues: a nucleophilic residue (serine, cysteines, or aspartate), a catalytic acid residue (aspartate or glutamate), and a histidine residue, always in this order in the amino acid sequence (Ollis et al. 1992). This order is different from that observed in any of the other proteins that contain catalytic triads (e.g. Trypsin, Papain, subtilisin). In lipases the nucleophile has so far always found to be serine, whereas the catalytic acids either an aspartate or glutamate. The nucleophilic serine is located in a highly conserved Gly-X-Ser-X-Gly pentapeptide (Ollis et al. 1992), which forms a sharp, γ -like turn between $\beta 5$ of α/β hydrolase central β sheet and the following α helix. This is the most conserved structural feature of α/β hydrolase fold. Because close contact exists between the residues, two positions before and after the nucleophile are occupied by glycine. However, in some cases they are substituted for other small residues such as alanine, valine, serine and threonine (Dartois et al. 1992; Franken et al. 1991; Lawson et al. 1994; Uppenberg et al. 1995). The strand-nucleophile-helix arrangement is termed as “nucleophilic elbow”, which positions the nucleophilic residue free of the active site surface and allows easy access on one side by the active site histidine and on the other side by substrate. The catalytic acid (Asp or Glu) occurs in a reverse turn after strand 7 of the central β sheet (Ollis et al. 1992). It is hydrogen bonded to active site histidine. The third catalytic residue in lipase is the

catalytic histidine, which is located in a loop after β strand 8 of α/β hydrolase fold. The length and conformation of this loop are variable (Ollis et al. 1992).

1.7.3 Interfacial activation

Lipolytic enzymes are characterized by their drastically increased activity when acting at the lipid-water interface of micellar or emulsified substrates (Jaeger et al. 1999), a phenomenon called ‘interfacial activation’. This increase in enzymatic activity is triggered by structural rearrangements of the lipase active-site region, as witnessed from crystal structures of lipases complexed by small transition state analogs (Derewenda et al. 1992; van et al. 1993). In the absence of lipid-water interfaces, the active site is covered by a so-called “lid.” However, in the presence of hydrophobic substances, the lid is open, making the catalytic residues accessible to substrate and exposing a large hydrophobic surface. This hydrophobic surface is presumed to interact with the lipid interface. The lid may consist of a single helix (Derewenda et al. 1992), or two helices (Kim et al. 1997b), or a loop region (Grochulski et al. 1994). However, not all lipases show this interfacial activation. Notable exceptions are lipase from *B. subtilis* (Lesuisse et al. 1993), cutinase from *Fusarium solani* (Martinez et al. 1992), and guinea pig pancreatic lipase (Hjorth et al. 1993). These lipases lack a lid that covers the active site in the absence of lipid-water interfaces. Moreover, lipases from *Pseudomonas aeruginosa*, *B. glumae* and *Candida antarctica* B do not show interfacial activation but nevertheless have an amphiphilic lid covering their active sites (Verger 1997).

1.7.4 Catalytic mechanism

Active site of lipase consists of a Ser-His-Asp/Glu catalytic triad. Hydrolysis of the substrate takes place in two steps as shown in Figure 1.3 (Jaeger et al. 1999). Reaction starts with an attack by the oxygen atom of the hydroxyl group of the nucleophilic serine residue on the activated carbonyl carbon of the susceptible lipid ester bond. A transient tetrahedral intermediate is formed, which is characterized by a negative charge on the carbonyl oxygen atom of the scissile ester bond and four atoms bonded to the carbonyl carbon atom arranged as a tetrahedron. The intermediate is stabilized by the helix macrodipole of helix C (Figure 1.2), and hydrogen bonds between the negatively charged carbonyl oxygen atom (the “oxyanion”) and at least two main-chain NH groups (the “oxyanion hole”). One of the NH groups is from the residue just behind the nucleophilic serine; the other one is from

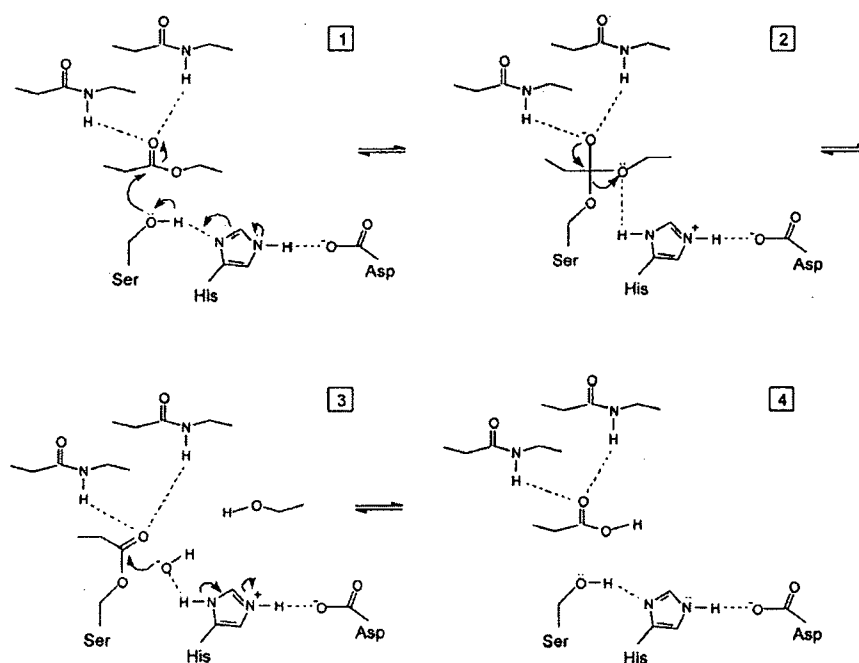


Figure 1.3 Catalytic mechanism of lipase.

the residue at the end of strand $\beta 3$. The nucleophilicity of the attacking serine is enhanced by the catalytic histidine, to which a proton from the serine hydroxyl group is transferred. This proton transfer is facilitated by the presence of the catalytic acid, which precisely orients the imidazole ring of the histidine and partly neutralizes the charge that develops on it. Subsequently, the proton is donated to the ester oxygen of the susceptible bond, which thus is cleaved. At this stage the acid component of the substrate is esterified to the nucleophilic serine (the “covalent intermediate”), whereas the alcohol component diffuses away (Figure 1.3). The next stage is the deacylation step, in which a water molecule hydrolyzes the covalent intermediate. The active-site histidine activates this water molecule by drawing a proton from it. The resulting OH^- ion attacks the carbonyl carbon atom of the acyl group covalently attached to the serine. Again, a transient negatively charged tetrahedral intermediate is formed, which is stabilized by interactions with the oxyanion hole. The histidine donates a proton to the oxygen atom of the active serine residue, which then releases the acyl component.

1.7.5 Biotechnological application of lipases

Lipases catalyze both the hydrolysis and the synthesis of esters formed from glycerol and long chain fatty acids, which usually proceed with high regio- and/or

enantioselectivity, making lipases an important group of biocatalysts in organic chemistry. The reasons for the enormous biotechnological potential of microbial lipases include the facts that they are stable in organic solvents, do not require cofactors, possess broad substrate specificity and exhibit high enantioselectivity. Lipases are used as hydrolases in detergent, food processing, paper and pulp industry and in leather processing. Besides these they are also used as synthetases in pharmaceutical industry, mainly in the synthesis of chiral drugs (Jaeger and Reetz 1998; Gupta et al. 2004).

The commercially most important application of lipases is their addition to detergents which are mainly used in household and industrial laundry. A number of lipase formulations such as LipolaseTM, which originated from fungus *T. thermophilus*, LumafastTM from *Pseudomonas mendocina* and LipomaxTM from *Pseudomonas alcaligenes* have been developed for use in detergents (Jaeger and Reetz 1998). Lipases also found application in modifying food ingredients and in making flavored compounds. An important example is the use of *Rhizomucor miehei* lipase in making cocoa butter substitute, which catalyzes transesterification reaction replacing palmitic acid by steric acid to provide the steric-oleic-stearic triglyceride with the desired melting point for use in chocolate making (Jaeger et al. 1999). Lipases are of particular use in flavor development of dairy products, mainly in the preparation of specially flavored cheese (Jaeger and Reetz 1998; Jaeger and Eggert 2002). Microbial lipases are also used to enrich polyunsaturated fatty acids (PUFA) from animal and plant lipids which are subsequently used to produce a variety of pharmaceuticals including anticholestoleemics, anti-inflammatories and thrombolytics (Gill and Valivety 1997). Another application of increasing importance is the use of lipase in removing the pitch from pulp produced in the paper industry. Lipase from *Candida rugosa* is extensively used for this purpose (Jaeger and Reetz 1998).

Lipases have been used for a long time to catalyze a wide variety of chemo-, regio- and stereoselective transformations (Jaeger and Reetz 1998; Jaeger et al. 1999; Gupta et al. 2004). This has been exploited in pharmaceutical industry for the synthesis of chiral drugs. The enantiomers of a chiral drug may exhibit marked differences in biological activity, toxicity, drug elimination and metabolism. Examples include production of β -blockers, containing an aryl-oxi-propanolaminic group, from which only the S-enantiomer is active, enantiomerically pure lipolytic hydrolysis of cyclopentendiol derivatives in the synthesis of prostaglandins, synthesis of chiral amines, catalyzed by the lipase from *Bukholderia plantarii* and the *Serratia marcescens* lipase based production of (2R,3S)-3-(4-

methoxyphenyl)methyl glycidate, which is used in the synthesis of the calcium antagonist DiltiazemTM.

The biotechnological potential of microbial lipases is steadily increasing. New lipases are being discovered and structures of many are becoming available. This along with the development of new molecular biological techniques provides novel tools for tailoring lipases for improved properties, thus expanding their potential for a variety of different applications.

1.7.6 *Bacillus subtilis* lipase

Bacillus subtilis is a gram-positive, aerobic, spore-forming bacterium found in soil and water, and in association with plants. This organism is of substantial commercial interest because of its highly efficient protein secretion system which can be used for the production of bulk quantities of enzymes such as proteases and amylases (Harwood 1992; Harwood and Cranenburgh 2008).

Extracellular lipolytic activity of *B. subtilis* was first observed in 1979 (Kennedy and Lennarz 1979), however, molecular research started in 1992 when a lipase gene, *lipA*, was cloned, sequenced and overexpressed, and the protein was characterized (Dartois et al. 1992; Lesuisse et al. 1993; Dartois et al. 1994). Later, a second gene, *lipB*, was found that is 68% identical with *lipA* at the nucleic acid level. This gene has been cloned and overexpressed and protein was purified and characterized (Eggert et al. 2000).

B. subtilis lipase LipA, with 181 amino acids and molecular mass of 19,348 Da, is one of the smallest lipases known. Also, it is one of the few examples of a lipase that does not show interfacial activation in the presence of oil-water interfaces (Dartois et al. 1992). Moreover, LipA is very tolerant to basic pH and has optimum activity at pH 10.0. It was classified as a lipase rather than an esterase, because of its ability to hydrolyse sn-1 and sn-3 glycerol esters with long fatty acid chains, showing optimal activity on glycerol esters with medium-length (C-8) fatty acid chains (Lesuisse et al. 1993).

Unlike other lipases *Bacillus subtilis* lipase does not show interfacial activation. This, together with its small size, suggests that it does not have a lid. This enzyme has an accessible active site, with an intact preformed oxyanion hole, which stabilizes the negatively charged reaction intermediates (van et al. 2001).

Bacterial lipases are classified into eight families according to their sequence similarities, conserved sequence motifs and biological properties (Arpigny and Jaeger

1999). The true lipases comprise family I, the largest family, which contains six subfamilies. *Bacillus* lipases have been placed in subfamilies 4 and 5. The common property between these two subfamilies is that alanine replaces the first glycine residue in the conserved G-X-S-X-G pentapeptide around the active site serine residue. Subfamily 4 consists of only three members, LipA and LipB from *B. subtilis*, and a lipase from *Bacillus pumilis*, which share 74-77% sequence identity. These are the smallest true lipases known and share very little sequence similarity (about 15 %) with the other, much larger, *Bacillus* lipases that constitute subfamily 5. There is no separate lid domain present, as for instance in the larger lipases.

The crystal structure of *B. subtilis* lipase (LipA) was reported in 2001 (van et al. 2001). Since then, several structure of same protein in different conditions (Kawasaki et al. 2002), and of its thermostable mutants have been reported (Acharya et al. 2004). The crystal structure reveals a globular shape with dimensions of 35Å×36Å×42Å (Figure 1.4). The structure shows a compact domain that consists of six β-strands in a parallel β-sheet, surrounded by five α-helices. Two α-helices are on one side of the β-sheet while three on the other side. The fold of *B. subtilis* lipase resembles that of the core of the α/β hydrolase fold enzymes. A comparison of the secondary structure elements of *B. subtilis* lipase and the canonical 1 α/β hydrolase fold is shown in Figure 1.5 (a) and (b). *B. subtilis* lipase



Figure 1.4 The *Bacillus subtilis* lipase. The catalytic residues Ser77, His156 and Asp 133 are labeled S, H and D respectively. The letters N and C indicate the N and C termini respectively. [von Pouderoen, G., Eggert, T., Jaeger, K.-E. & Dijkstra, B. W. (2001). The crystal structure of *Bacillus subtilis* lipase: a minimal α/β hydrolase fold enzyme. *J. Mol. Biol.* 309, 215–226.]

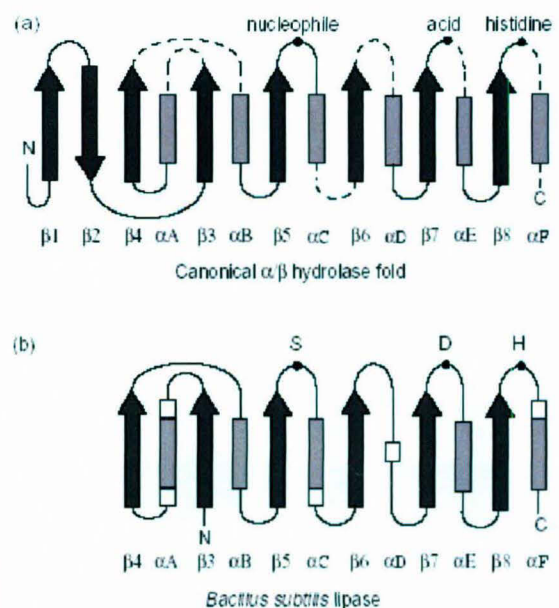


Figure 1.5 Canonical α/β hydrolase fold (a) and minimal α/β hydrolase fold of *Bacillus subtilis* lipase (b).

lacks the first two β -strands ($\beta 1$ and $\beta 2$) of the canonical fold and helix αD is replaced by a small 3_{10} helix. The helix αE is exceptionally small, with only one helical turn, and several α -helices start or terminate with 3_{10} helical turns. There is no separate lid domain present, as for instance in the larger lipases. Because of its small size and the absence of a lid domain, *B. subtilis* lipase is considered as a minimal α/β hydrolase fold enzyme.

1.8 Focus of the present thesis

Understanding the principles governing the stability in proteins has been a subject of great interest. This is not only to understand the sequence-structure-property relationship in proteins, but also to employ these rules in designing stable proteins for commercial applications. Much understanding on this topic has been provided by the comparative sequence and structural analysis of homologous proteins from organisms adapted to extreme environment. However, these kinds of studies lead to confounding interpretations due to a number of variations in homologous proteins, many of which are neutral or not at all related to the property of interest. However, in last one decade, with the advancement in molecular biological tools, it has become possible to ‘evolve proteins’ for a defined property in selected conditions in the laboratory on a highly accelerated time scales. This not only allows the exploration of the properties not found in nature but also presents a much clear picture as the changes observed are directly linked to the property for which the protein was evolved. When coupled with biophysical analysis and structure determination of improved variants and wild-type protein, it provides a better understanding of the structural basis of the phenomenon involved.

In the present work, *Bacillus subtilis* lipase has been used as a model protein to understand the basis of its thermostability. The protein is evolved by “*in vitro* evolution” to generate a set of variants differing in thermostability. Characterization of such variants of the same protein, with incremental increase in stability provides an opportunity for better understanding of the strategies employed by this protein in improving thermostability.

Establishment of cloning and
expression system for *in vitro*
evolution of lipase

Chapter 2

2.1. Introduction

The methods of “directed molecular evolution” have met with phenomenal success for engineering proteins in recent past (Farinas et al. 2001; Cherry, Fidantsef 2003; Otten and Quax 2005; Yuan et al. 2005; Johannes and Zhao 2006; Jackel et al. 2008). These methods rely on the principles of Darwinian evolution, and the process is highly accelerated on time scale. Similar to natural evolution, directed evolution involves generation of mutations and selection of mutants in the desired conditions to find the best variant. The conditions for creation of genetic diversity as well as screening of mutant population are carefully controlled, as the outcome depends on the screening conditions employed. Iterative cycles of mutagenesis followed by screening, enable identification and accumulation of beneficial mutations with gradual improvement in property. Since sequences of proteins are optimized for a specific function over duration, the chances of identifying positive mutations would be rare. Consequently, to isolate rare mutations the size of the mutant populations required to be screened would be large. The particular choice of the method to create genetic variation and selection depends on the protein and the property to be improved. For example, each one of the strategies, such as random mutagenesis of whole gene or targeted mutagenesis of specific regions of protein or recombination between natural variants of the gene could lead to new and improved variants. Therefore, the choice of a technique to generate genetic variation and the method to screen the variants generated are the two most important aspects of any directed evolution protocol (Kuchner and Arnold 1997).

The method chosen for generating genetic variations mainly depends upon the evolutionary distance between the existing properties and the desired ones and should be able to pave the evolutionary pathway from the former towards the latter. This means that longer the evolutionary distance between the existing property and the property to be evolved, larger would be the size of the variant library. Techniques which can create high fitness libraries with respect to size and diversity have to be used when evolving a property unrelated to the existing property. However, techniques which produce low fitness libraries can be used when fine tuning the existing property for better performance (Wintrode and Arnold 2000).

The most important as well as the ‘rate limiting step’ of any directed evolution algorithm is the selection/screening strategy employed to identify the positive variants.

Screening not only has to be specific towards the property to be improved, but also sensitive enough to monitor the small changes. Moreover, screening strategy has to be rapid to screen a large mutant population (Arnold et al. 2001). Selection allows all the mutants in the pool to be tested at the same time with the active/desired ones being highlighted. Although being very efficient, selection can not be used for many properties as it relies on the contributions of the improved property to the survival of the organism under selection pressure, not possible in many cases (Leemhuis et al. 2005; Boersma et al. 2007). Screening on other hand provides an opportunity to assess individual mutants under controlled conditions, often non-natural, and selects the best variant (Olsen et al. 2000).

Designing screening protocols in a high throughput format is relatively straightforward for enzymes by exploiting their catalytic properties. Both *in vivo* and *in vitro* assay formats can be used. While former offers the proper environment for folding and maturation of protein, the latter allows the use of harsh or non-physiological conditions during screening. Several colorimetric or fluorescence based assays are available for different class of enzymes which can be used in high throughput format upon miniaturization (Olsen et al. 2000; Cohen et al. 2001). This allows screening of a population of more than 10^4 mutants on a routine basis.

Bacillus subtilis lipase is a small protein of 19.3 kDa, without any co-factors and disulfides. This is one of the smallest, lidless lipase and does not show interfacial activation. *B. subtilis* lipase is a triglyceride hydrolase, with preference for C-8 fatty acid chain esters, but also hydrolyzes single chain esters (Dartois et al. 1992; Lesuisse et al. 1993). Activity of lipase could be performed by a variety of ways such as by monitoring the changes in the pH (using pH dyes or pH stat), hydrolysis of a substrate tagged with a fluorescent dye, changes in the physical properties of the substrates (surface pressure or light scatter). Many solid phase (plate based) assays, based on the hydrolysis of substrate, have been reported for lipases (Gupta et al. 2003). Having broad substrate specificity, *B. subtilis* lipase can perform hydrolysis of a number of substrate analogues whose products are colored or fluorescent (Schmidt and Bornscheuer 2005). Heterologous over-expression of *B. subtilis* lipase in *E. coli*, *B. subtilis*, yeast etc. have been shown to be possible (Sanchez et al. 2002). Thus, a high throughput screening protocol can be easily developed by using available solid phase as well as solution based lipase assays. Our earlier efforts indicated that screening of mutant population in *E. coli* sometimes identifies false positives (Acharya et al. 2004). In thermostability assays, heating steps in screening

protocols results in aggregation of proteins in the cell lysates, having high protein concentration, leading to confounding results. Hence, the cloning steps are performed in *E. coli* and the screening is performed in *B. subtilis*, where the protein is secreted out in the medium. *Bacillus* species are tough to transform (10^2 - 10^3 colonies per μg plasmid DNA) and to obtain large population sizes it is essential to have an efficient transformation protocol (Saunders and Saunders 1999). This chapter describes the generation of a suitable shuttle vector for propagation in *E. coli* and *B. subtilis*, subcloning of *Bacillus subtilis* lipase in the shuttle vector and expression in *B. subtilis*, efforts made to optimize transformation protocols for *B. subtilis* and standardization of mutagenic protocol.

2.2 Materials and Methods

2.2.1 Materials

Taq DNA polymerase and dNTPs were purchased from Invitrogen (Carlsbad, CA). Restriction enzymes were from New England Biolabs (Beverley, MA). Fast link™ DNA ligation kit was from EPICENTRE Biotechnologies (Madison, WI). Genemorph® Random mutagenesis kit was obtained from Stratagene (La Jolla, CA). QIAquick® gel extraction kit and QIAprep® spin Miniprep kits were from Qiagen (Hilden, Germany). Tributyrin, Gum Arabic, Lysozyme and Triton-X 100 were from Sigma Chemical Co. Oligonucleotides were purchased from BioServe technologies (Hyderabad, India). PNPO was synthesized as described earlier (Acharya and Rao 2002). All other reagents used were of analytical grade or higher.

2.2.2 Construction of vector pSA01

For efficient cloning and expression of lipase gene (*lip*) in *Bacillus subtilis*, vector pSA01 was constructed by cloning a strong gram positive promoter *Hpa* II (Zyprian and Matzura 1986) along with lipase signal sequence into the multiple cloning site of *E. coli*-*B. subtilis* shuttle vector pMK3 (Sullivan et al. 1984). As shown in Figure 2.1, the region coding for *Hpa* II promoter, ribosome binding site (rbs) and signal sequence of lipase gene was amplified from pLIP2031 (Dartois et al. 1994), using primers XmaFor (5'-ATACTACCTGTCCCGGGCTGATTTTA-3') and PstRev (5'-ACTGGATTGTGTTCTGCAGCTTTTGCTGA-3') (*Xma* I and *Pst* I sites underlined).

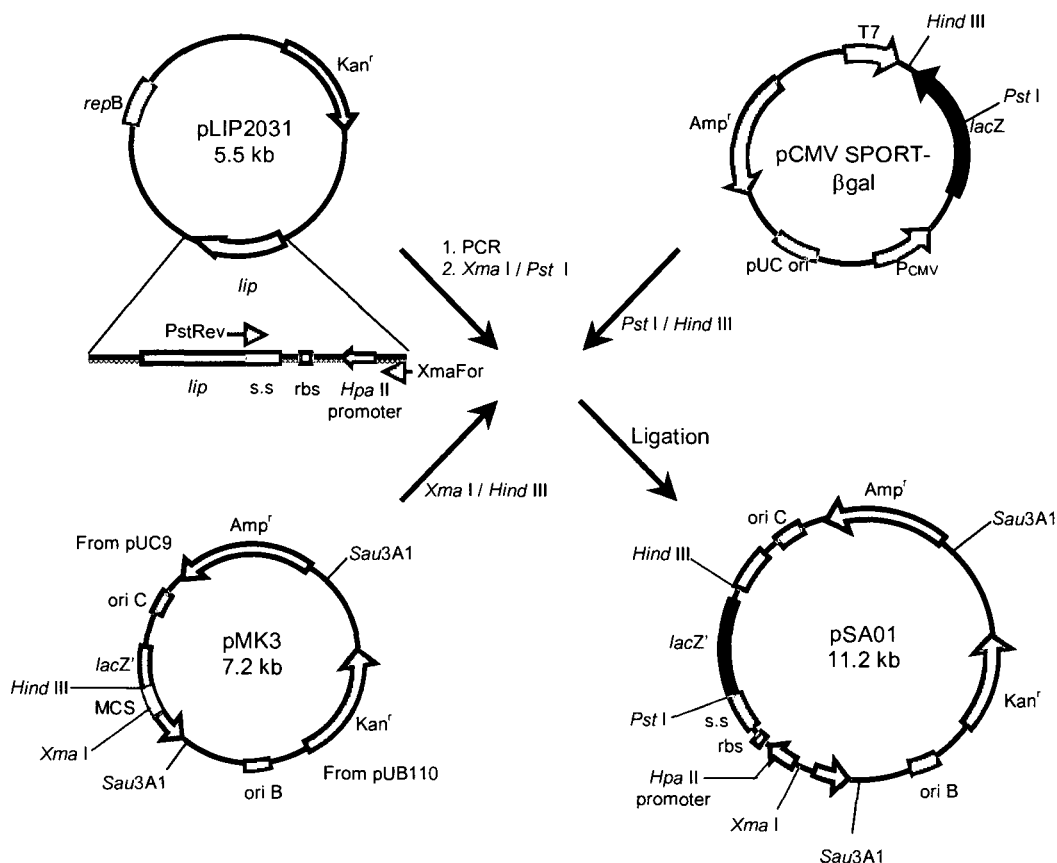


Figure 2.1 Construction of vector pSA01. A 500 bp fragment having *Hpa* II promoter and lipase signal sequence was amplified from pLIP2031 plasmid and cloned along with a 3.5 kb fragment from pCMV SPORT-βgal into the MCS of pMK3 *E. coli-B. subtilis* shuttle vector.

The amplified PCR product was gel purified and digested with *Xma* I and *Pst* I. In another step, a 3.5 kb fragment was isolated by digesting plasmid pCMV SPORT βgal (Invitrogen) with *Pst* I and *Hind* III. The digested and gel purified products were mixed and ligated into *Xma* I – *Hind* III digested pMK3 shuttle vector to get pSA01. This brings the regions coding for *Hpa* II promoter, ribosome binding site and the signal sequence of lipase in frame with *lacZ'* gene, which was cloned next to the signal sequence. A *Pst* I was also introduced at the junction of signal sequence and the subsequent gene by silent mutagenesis. This construct (pSA01) was used for the subcloning of wild-type and mutant lipase genes in the subsequent steps involved during screening.

2.2.3 Subcloning of lipase gene in *E. coli-B. subtilis* shuttle vector pSA01

For over-expression of lipase gene in *Bacillus subtilis*, cloning of lipase gene (*lip*) was performed in *E. coli-B. subtilis* shuttle vector pSA01 as shown in Figure 2.2. The structural genes encoding wild-type and triple mutant (TM), cloned in pET21b (Acharya et

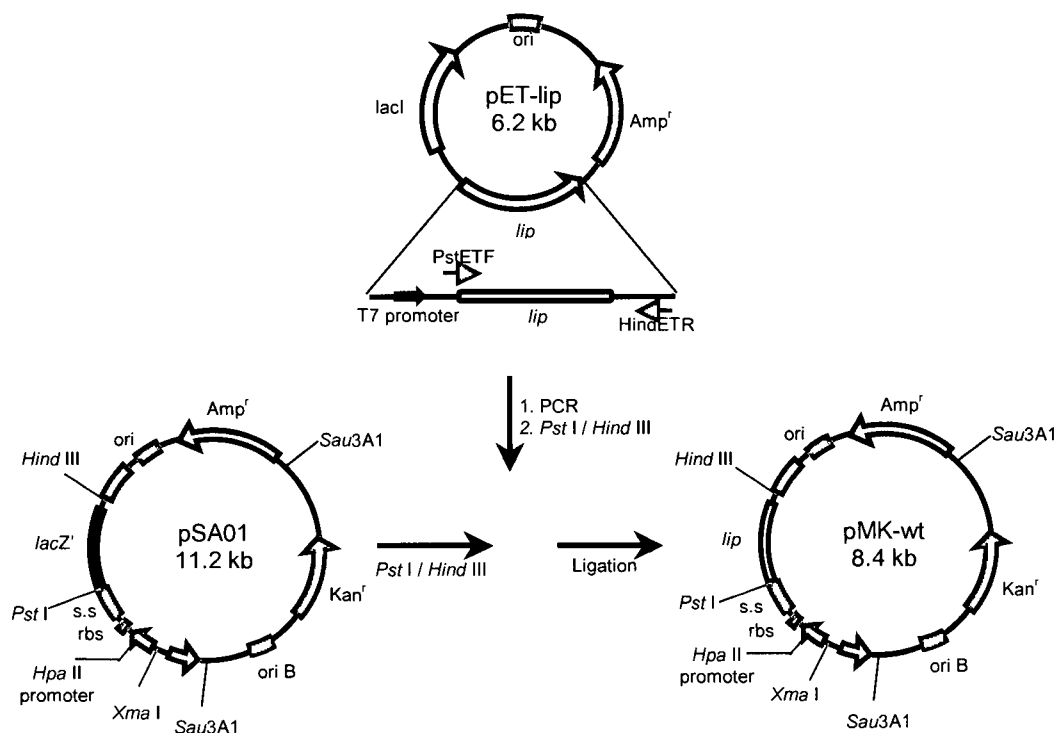


Figure 2.2 Subcloning of lipase gene (*lip*) in *E. coli*-*B. subtilis* shuttle vector pSA01. Lipase gene was cloned, along with its signal sequence, under gram positive *Hpa*II promoter for protein expression and secretion of lipase in *B. subtilis*.

al. 2004), were PCR amplified using primers PstETF (5'-GGAGATATACATACTGCAGAACACAATCCA-3') and HindETR (5'-GTTACCTGAAAGCTTGGGTGCATGTGT-3') (*Pst* I and *Hind* III sites underlined). The amplified products were gel purified and digested with *Pst* I and *Hind* III. The digested products were again gel purified and ligated with similarly digested pSA01, to get pMK-wt and pMK-TM. This cloning puts lipase gene under the control of strong gram positive *Hpa* II promoter, in frame with its signal sequence which allows high expression and secretion of the protein in the surroundings upon expression in *Bacillus subtilis*.

2.2.4 Transformation of *Bacillus subtilis* using electroporation

Preparation of cells

Electrocompetent *B. subtilis* (BCL 1050) cells (Dartois et al. 1994) were prepared as described earlier with little modifications (Vehmaanpera 1989). Briefly, 50 ml of LBSP medium (1% Tryptone, 0.5 % Yeast extract 0.5 % NaCl and 250 mM Sucrose in 50 mM Potassium phosphate buffer, pH 7.2) was inoculated with cells from a single fresh colony

and incubated at 37 °C with vigorous shaking (250 rpm) to late exponential phase (4.5 hours). Four flasks containing 200 ml of LBSP medium (in 1 L flasks; pre-warmed at 37 °C) were inoculated each with 10 ml of the grown culture. Cultures were grown at 37 °C with vigorous shaking until OD₆₀₀ reaches 1.0 (about 2 hours) and chilled immediately in ice-water bath for 10 minutes. Cells were harvested by centrifugation at 10,000 g for 5 min at 4 °C, using pre-cooled rotor and tubes. Supernatant was discarded. Cells were washed five times by resuspending in 200 ml of ice-cold SHMG (250 mM Sucrose, 1 mM Hepes, 1mM MgCl₂ and 10 % Glycerol; pH 7.0). Cells were recovered during each washing step by centrifugation at 10,000 g for 5 min at 4 °C. Finally, cells were resuspended in 6 ml of ice-cold SHMG, dispensed in 0.5 ml aliquots in microfuge tubes and stored directly at -80 °C for further use.

Electroporation

An aliquot (0.5 ml) of frozen electrocompetent cells was thawed by holding briefly in palm and then quickly transferring on ice. 1-10 µl of DNA (containing 0.1 to 1 µg of DNA in milli-Q water or 10 mM Tris-Cl buffer; pH 8.0) was added and mixed gently by pipetting. Mixture was transferred to a 2 mm pre-cooled electroporation cuvette (BioRad, Richmond, CA) and left on ice for 30 minutes. A single pulse of 1200 V (Electric field = 6 kV/cm) and 25 µF was applied using Gene Pulser II™ (BioRad). Immediately after the pulse, cells were diluted 10 times with LBSPG (LBSP having 10 % Glycerol) and transferred to sterile culture tubes. Cells were incubated at 37 °C with vigorous shaking for 60 minutes for the expression of antibiotic resistance markers before plating different aliquots/dilutions on suitable agar plates.

2.2.5 Natural transformation of *Bacillus subtilis*

Natural transformation of *Bacillus subtilis* (BCL 1050) was performed by following the procedure of Heierson et al. (Heierson et al. 1987) with few modifications as follows. Five ml of PM medium was inoculated with *Bacillus subtilis* (BCL 1050) cells from a single fresh colony and incubated overnight at 37 °C with vigorous shaking (250 rpm). The overnight grown culture was diluted 50 times in fresh PM (1 ml of overnight grown culture was added to 50 ml fresh PM) and incubated at 37 °C till OD₆₀₀ reaches ~1.0 (3 hours). The medium was divided in 0.5 ml aliquots in sterile tubes to which 10 µl of DNA (containing 0.1-1 µg of DNA in milli-Q water or 10 mM Tris-Cl buffer; pH 8.0)

was added. The tubes were further incubated at 37 °C with vigorous shaking for 4.5 hours with loose caps for proper aeration. Different aliquots/dilutions were plated on suitable agar plates for the selection of transformants.

2.2.6 Lipase Assay

Solid phase (Petri plate based) lipase assay

Bacillus subtilis BCL 1050 strain transformed

with appropriate plasmids (pMK3 derived plasmids or pLIP2031) were plated on LB Agar plates (1% Tryptone, 0.5% Yeast-extract, 1% NaCl and 1.5% Agar) having appropriate antibiotics (5 µg/ml Kanamycin for pMK3 derived plasmids while 50 µg/ml Kanamycin for pLIP2031) and incubated overnight at 37 °C. Overnight grown colonies were further incubated at 6-8 °C for 6-8 hours for lipase expression and secretion. Plates were brought back to room temperature and overlaid with 5 ml of soft-agar (0.5%) having 0.5% tributyrin emulsified in 0.4% gum arabic and 0.2% sodium azide. Plates were further incubated at 37 °C for 6-8 hours for the appearance of clear halos around the colonies.

Solution based lipase assay

Solution based lipase assay, using crude culture supernatant as lipase source, was performed by using p-nitrophenyl oleate (PNPO) as the substrate as described earlier (Acharya et al. 2004) with some modifications. Culture supernatant of lipase over-expressing *Bacillus subtilis* strain, grown either in 96-well plates or in culture tubes for sufficient period, was diluted 1:1 with 100 mM sodium phosphate buffer (pH 7.2). For microtitre plate based assay, activity measurement was performed by mixing 80 µl of diluted supernatant with 80 µl of 2 x PNPO-Triton X-100 substrate solution (2 mM PNPO micellized in 40 mM Triton-X100). Release of the product, p-nitrophenol, was monitored by measuring increase in absorbance at 405 nm, with time, in an ELISA reader (Spectramax 190, Molecular Devices, Sunnyvale, CA). For tube assays, 80 µl of diluted culture supernatant was mixed with 920 µl of 50 mM sodium phosphate buffer (pH 7.2)

Paris Medium (PM)

Sterile mQ	:	75 ml
10x PC Buffer	:	10 ml
20% Glucose	:	5 ml
2% Casamino acids	:	5 ml
1M Potassium Glutamate	:	2 ml
1 mg/ml L-Tryptophan	:	2 ml
1 M Magnesium Sulphate	:	300 µl
0.1% Ferric Ammonium Citrate	:	200 µl
10 mg/ml L-Histidine	:	500 µl

Phosphate Citrate Buffer (PC) 10x

K ₂ HPO ₄ (0.6 M)	:	107g/L
KH ₂ PO ₄ (0.4 M)	:	60 g/L
Na ₃ -citrate.5H ₂ O (30 mM)	:	10 g/L

having 0.2 mM PNPO micellized in 20 mM Triton-X100. Increase in absorbance at 405 nm, with time, was measured using spectrophotometer (Hitachi U-2000).

2.2.7 Error prone PCR of lipase gene

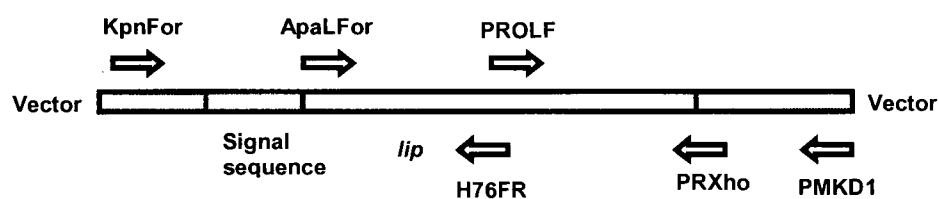
Lipase gene of triple mutant (TM) along with the signal sequence cloned in pMK-TM was mutagenized by error-prone PCR using GeneMorph[®] Random Mutagenesis kit from Stratagene according to the manufacturer's instructions. Primers Kpn1For (5'-GGGGTACCGTAATATAATTGGAGAA-3') and PMKD1 (5'-GAGCGCAACGCAATTAATGTGAGTT-3'), flank the gene beyond the *Pst* I and *Hind* III sites present on the plasmid, were used for amplification. A mutation condition which should generate low error frequency of ~ 3-5 mutations per 1000 bases was used. The reaction mixture contained 1× Mutazyme[®] reaction buffer, 0.2 mM of each dNTP, 0.5 pmoles/μl of each flanking primers, 90 pg of template plasmid and 2.5 units of Mutazyme[®] in a total volume of 50 μl. The reaction was heated at 94 °C for 5 min followed by 30 cycles of 94 °C for 1 min, 52 °C for 1 min and 72 °C for 1 min followed by a final extension of 10 min at 72 °C in a PCR thermal cycler (GeneAmp 9700, Applied Biosystems, Foster City, CA). Products were purified, digested with *Pst* I and *Hind* III and ligated in a similarly digested pSA01 plasmid. The mutagenic library thus obtained was used to transform *E. coli* DH5α by electroporation.

2.2.8 Isolation of plasmid DNA and sequencing of lipase gene

Plasmid DNA from *E. coli* DH5α or *Bacillus subtilis* BCL1050 were isolated using QIAprep[®] spin Miniprep kit according to the manufacturer's instructions. When isolation was done from *B. subtilis*, 1 mg/ml lysozyme was also added in buffer P1. Sequencing of lipase gene was performed by using six primers as listed in Table 2.1, (Figure 2.3). Sequencing reactions consisted of 50 ng of template DNA, 2.5 pmoles of each primer, sequencing buffer and BigDye[™] reagent (Applied Biosystem). Reactions were carried out for 30 cycles of 94 °C for 10 sec, 50 °C for 5 sec followed by 60 °C for 4 minutes in a PCR thermal cycler. Sequencing was done using ABI 3730 sequence analyzer (Applied Biosystems).

Table 2.1 Sequence of primers used in sequencing of lipase gene.

Primer	Sequence
KpnFor	5' GGGGTACCGTAATATAATTGGAGAA 3'
ApaLFor	5' CGTTATGGTGCACGGTATTGGA 3'
PROLF	5' GGCAAGGCGCCTCCGGGAACAGAT 3'
PMKD1	5' GAGCGCAACGCAATTAATGTGAGTT 3'
PRXho	5' GGTTTTGTTTTCTCGAGATTCGTATTCTGG 3'
H76FR	5' GGGGTACGACTTTCGCTGTTATAG 3'

**Figure 2.3** Schematic representation of various primer combinations used for sequencing of lipase gene.

2.3 Results

2.3.1 Cloning and expression of lipase gene

The lipase gene was cloned along with its signal sequence for export of enzyme into the medium. Cell free lipase was expected to provide cleaner signal compared to the intracellular protein. To screen thermostable mutants during “*in vitro* evolution”, *Bacillus subtilis* lipase was cloned and expressed in *Bacillus subtilis* BCL1050. The strain BCL1050 (*his nprR2 nprE18 aprΔ3, Δlip*) is not only deficient in two extra cellular neutral proteases and an alkaline protease but also lacks the chromosomal copy of the lipase gene (Dartois et al. 1994). Lipase being native to *B. subtilis* which has well defined secretory system can be used as an excellent host for the over-expression and secretion of lipase into the extra cellular milieu. To facilitate this, an *E. coli-B. subtilis* shuttle vector was made in which lipase gene (*lip*) was subcloned along with its signal sequence under the control of a strong gram positive promoter *Hpa* II (Zyprian and Matzura 1986) (Figure 2.1 and 2.2). To facilitate subcloning of lipase gene during the subsequent steps involved in screening, a *Pst* I site was also introduced at the junction of signal sequence and the structural gene, by silent mutagenesis. The vector construct pMK-wt (or pMK-TM, encoding a triple mutant of lipase (Acharya et al. 2004)), when transformed in *B. subtilis*

BCL1050, drives the constitutive expression and secretion of lipase into the culture supernatant. The enzymatic activity, when measured using late exponential phase culture supernatants, was comparable to the lipase over-expressing strain BCL1051 (harboring wild-type lipase under *Hpa* II promoter in plasmid pLIP2031) (Dartois et al. 1994).

2.3.2 Electro-transformation of *Bacillus subtilis*

Several procedures describing electroporation of *Bacillus subtilis* (Kusaoke et al. 1989; Masson et al. 1989; Ohse et al. 1995; Ohse et al. 1997; Xue et al. 1999) were tested to transform *B. subtilis* BCL1050 with variable success. Transformation efficiencies in the range of 10^0 - 10^2 (CFU/ μ g plasmid DNA) were obtained with varied reproducibility. However, the method as described by Vehmaanperä (Vehmaanpera 1989), yielded transformation efficiencies in the range of 10^3 - 10^4 CFU/ μ g plasmid DNA. The method was less cumbersome and the transformation efficiencies were highly reproducible. Growth parameters, like time of cell harvest, were optimized which had critical bearing on the transformation efficiency. The effect of electric field applied on the transformation efficiency and the rate of survival of cells is shown in Figure 2.4. The transformation efficiency was $\sim 5 \times 10^3$ / μ g of plasmid DNA (pMK3; 7.2 kb) at 5 kV/cm which increases to 1.5×10^4 at 6 kV/cm. The transformation efficiency decreases slightly on further increasing the electric field to 7-9 kV/cm but remains in the range of 10^4 transformants/ μ g of plasmid DNA. Viability of the cells decreases, as judged by colony forming units, with increase in electric field. While

$\sim 3\%$ of the viable cells ($\sim 4 \times 10^{10}$) survives the pulse at 5 kV/cm, barely 1.7% of the same remains viable after pulse at 8 kV/cm. Other electrical parameters like capacitance, resistance and number of pulses applied were also optimized. Capacitance of 25 μ F and resistance of 200 Ω in parallel gave the optimal results. Only a single pulse was used for

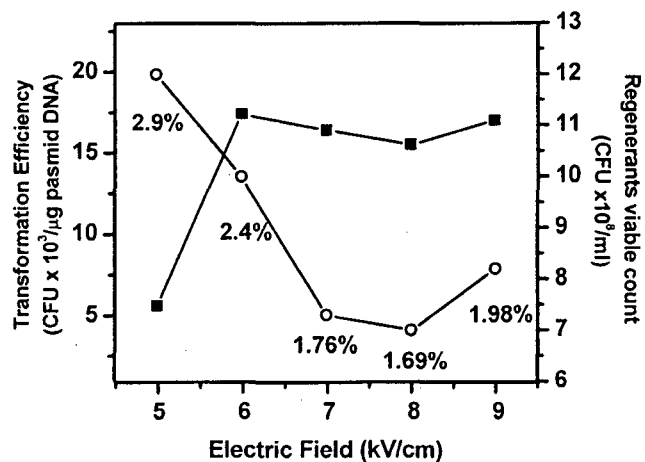


Figure 2.4 Electrotransformation of *B. subtilis*. Effect of electric field on the transformation (■) and survival (○) efficiencies of *B. subtilis* (BCL1050) upon electroporation using plasmid pMK3 (7.2 kb). Capacitance of 25 μ F and resistance of 200 Ω was used.

transformation.

2.3.3 Natural transformation of *Bacillus subtilis*

Natural transformation of *Bacillus subtilis* BCL 1050 was done as described earlier (Heierson et al. 1987). Several other protocols of transformation as described by Spizizen (Spizizen 1958) and later modified by others (Anagnostopoulos and Spizizen 1961; Dubnau and vidoff-Abelson 1971; Sadaie and Kada 1983; Matsuno et al. 1990), were tried but transformation efficiencies in the range of 10^0 - 10^2 (CFU/ μ g plasmid DNA) were obtained, however with protocol of Heierson et al. same was 10^2 - 10^3 (CFU/ μ g plasmid DNA) with supercoiled plasmid. We observed that the time of addition of plasmid DNA to the secondary culture has crucial effect on the efficiency of transformation. As shown in Figure 2.5, maximum efficiency was obtained when DNA was added after 4 hours of growth of secondary culture. It has been reported that higher transformation efficiencies were obtained with plasmid multimers as compared to supercoiled plasmid (Canosi et al. 1978). We obtained multimers of plasmid pMK3 by propagating it in *rec* positive *E. coli* strains GJ1885 (*ara zbh-900::Tn10dKan(Ts)llacZ4525::Tn10dKan*) (Reddy and Gowrishankar 1997) and GJ1885 *recD* (GJ1885 *recD1901::Tn10*) (Biek and Cohen 1986). Both strains yield multimers of plasmid but the yield was higher when isolated from latter. When transformed with multimers of pMK3, *B. subtilis* (BCL1050) yielded transformation efficiency in the range of 10^4 CFU/ μ g plasmid DNA, which was two orders of magnitude higher than what was obtained with supercoiled plasmid (Figure 2.5).

2.3.4 Solid phase lipase assay

Lipase assay on petriplates was performed by overlaying the overnight grown colonies of *Bacillus subtilis* with a layer of soft agar having substrate tributyrin emulsified in gum arabic. This makes a thin translucent

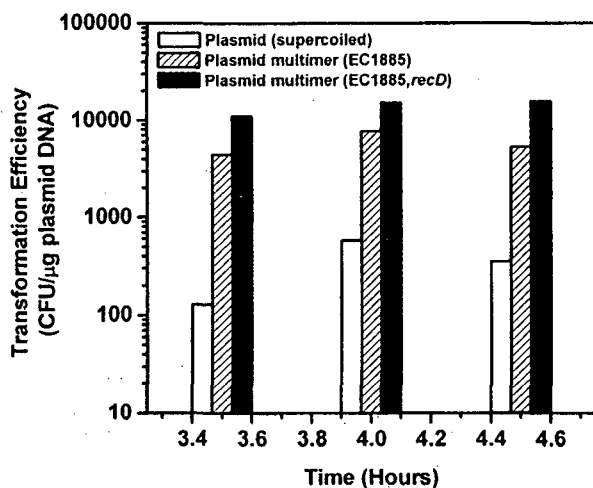


Figure 2.5 Natural transformation of *B. subtilis*. Effect of the time of plasmid DNA addition to the competent cells and the form of plasmid pMK3 (supercoiled or multimers) on the transformation efficiency of *B. subtilis* (BCL 1050).

layer (due to the emulsified substrate) over the colonies. Incubation of plates for 6-8 hours at 37 °C allows the enzyme to hydrolyze the substrate and makes a clear halo around lipase positive colonies (Figure 2.6). For residual activity measurement, 0.2 % of sodium azide was also added to the soft agar, which kills the bacteria in the colonies and so the activity (clearing due to hydrolysis of substrate) observed is only due to already secreted protein in the surrounding. Figure 2.6 shows the hydrolysis of tributyrin and clear haloes around the lipase positive clones *B. subtilis* (BCL1051) and BCL1050 transformed with pMK-wt (Figure 2.6 (b) and (d), respectively). Host *B. subtilis* BCL1050 alone and transformed with pMK3 were shown as controls (Figure 2.6 (a) and (c), respectively). It is important to note that lipase gene when cloned and expressed in pMK-wt, shows similar level of protein expression as that of lipase over-expressing clone *B. subtilis* BCL1051 (strain BCL1050 transformed with pLIP2031) when grown and incubated for the same period (Figure 2.6 (d) and (b), respectively).

2.3.5 Random mutagenesis of lipase gene

Random mutagenesis of lipase gene was performed by error-prone PCR, using Genemorph[®] Random Mutagenesis kit from Stratagene. The kit uses a DNA polymerase, Mutazyme[®], which has high intrinsic mutational rate. It randomly introduces error during elongation and has a unique mutational spectrum. Unlike *Taq* DNA polymerase which introduce error in the presence of Mn²⁺ and unbalanced dNTP ratio (Cadwell and Joyce 1992), Mutazyme[®], does not have such requirements. The mutational frequency was

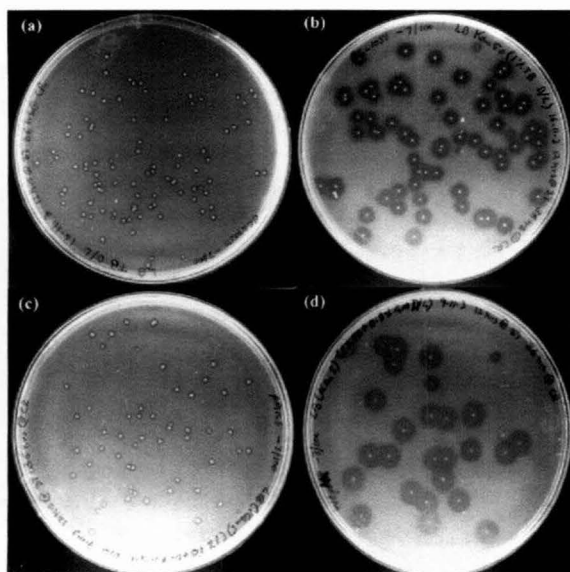


Figure 2.6 Solid phase (petriplate based) lipase assay. *Bacillus subtilis* (BCL 1050) colonies having suitable plasmids were overlaid with soft agar having substrate tributyrin emulsified in gum arabic. Lipase expressing clones shows clearing zones (halos) around the colonies as compared to lipase deficient ones. (a) Host strain, BCL 1050; (b) Lipase over-expressing strain, BCL1051 (having pLIP2031); (c) BCL 1050 transformed with pMK3 and (d) BCL1050 transformed with pMK-wt.

controlled by the initial amount of template taken for PCR and the number of cycles involved i.e. fold amplification of the target region occurring during the PCR reaction. Error-prone PCR of lipase gene was performed using KpnFor and PMKD1 primers (Table 2.1) and pMK-TM as template. The initial amount of template was taken such that 3-5 mutations occur per 1000 bases. The PCR product was cloned back in the pSA01 vector construct and used to transform *E. coli* DH5 α . In order to check the mutation frequency and the mutational spectrum, 24 clones were picked randomly, and their plasmids were isolated. The lipase gene in these plasmids was sequenced to assess the frequency and occurrence of mutations. It was observed that mutations among the clones are not clustered in any particular region of the gene but are well dispersed throughout the gene. The numbers of mutations observed per gene are listed in Table 2.2. Out of 24 clones, no mutation was observed in 7 clones while only one mutation was observed in a similar number of clones. We observed five clones with 2 mutations, 2 clones with 3 mutations and 3 clones with 4 mutations. More than 4 mutations per gene (690 bp) were not observed. This is well in accordance with the conditions used to achieve a low mutation frequency of one or two mutations per gene (690 bp).

The observed mutational spectrum of Mutazyme[®] and its comparison with that of *Taq* DNA polymerase is shown in Table 2.3. While Mutazyme[®] generates all kinds of mutations, its mutational spectrum was quite different from that of *Taq* DNA polymerase. Both enzymes show bias towards the kind of mutation they introduce (Shafikhani et al. 1997). The transition to transversion ratio (Ts/Tv) was roughly equal to 1 for both the enzymes, suggesting that both favor transitions over transversions. In the absence of any bias this ratio should be around 0.5. Moreover, the AT \rightarrow GC/GC \rightarrow AT ratio of Mutazyme[®] and *Taq* DNA polymerase is 0.35 and 1.9, respectively, indicating that Mutazyme[®] preferentially replaces Gs and Cs with As or Ts while *Taq* DNA polymerase shows a reverse preference.

Table 2.2 Number of mutations per gene obtained after random mutagenesis by error-prone PCR.

Number of mutations ^a	Number of clones ^b
0	7
1	7
2	5
3	2
4	3
5 or more	0

^a Total 38 mutations were observed including 2 insertions and a deletion.
^b Twenty four randomly picked clones were sequenced.

Table 2.3 Mutational spectrum of Mutazyme®.

Type(s) of mutations	Mutazyme® Number of mutations		DNA Polymerase Percent of total ^a	Taq DNA polymerase ^b
Transitions (Ts)				
A→G, T→C	4,	1	13.2%	27.6%
G→A, C→T	5,	8	34.2%	13.6%
Transversions (Tv)				
A→T, T→A	5,	2	18.4%	40.9%
A→C, T→G	1,	1	5.3%	7.3%
G→C, C→G	1,	0	2.6%	1.4%
G→T, C→A	4,	3	18.4%	4.5%
Insertions and Deletions				
Insertions	2		5.3%	0.3%
Deletions	1		2.6%	4.2%
Summary				
A→N, T→N	10,	4	36.8%	75.9%
G→N, C→N	10,	11	55.3%	19.6%
Bias Indicators				
Ts/Tv	1.1			0.8
AT→GC/GC→AT	0.35			1.9

^a Total 38 mutations were observed.

^b Taq DNA polymerase Mn²⁺containing buffer and unbalanced dNTP concentrations. [Shafikhani *et al.* (1997)]

To summarize, Mutazyme® DNA polymerase 1.6 times more likely replaces G→N and C→N than A→N and T→N while Taq DNA polymerase generates 3 times more A→N and T→N mutations than G→N and C→N.

2.4 Discussion

The success of a directed evolution experiment mainly depends on the screening protocol. Screening must be rapid to enable sampling of large mutant populations and also sensitive enough to differentiate between small changes in property that are expected to occur due to point mutations. This involves expression of protein in a suitable host system followed by screening of the mutant population upon exposure to the desired stress conditions. *Bacillus subtilis* lipase is a small (19.3 kDa) protein which can be expressed in many expression systems including *E. coli*, *B. subtilis* and *S. cerevisiae* (Sanchez *et al.* 2002; Becker *et al.* 2005; Mormeneo *et al.* 2008). In the present study, *Bacillus subtilis*

was chosen as the desired expression system as it offers many advantages over others. Being native to the host, lipase can be efficiently over-expressed and secreted out of the cell, thus avoiding the step of cell lysis and working with cell lysate for activity measurements. By using *B. subtilis* BCL1050 strain, which is deficient in three proteases and also lacks chromosomal copy of the lipase gene, lipase expressed from the plasmid does not have any contamination from the endogenous wild-type and also shows enhanced stability in the culture supernatant. Also, being native to the host, lipase is not toxic to *B. subtilis* as compared to other hosts such as *E. coli* (Krag and Lennarz 1975; Dartois et al. 1992; Dartois et al. 1994).

Being gram positive, *Bacillus subtilis* is difficult to transform as compared to other gram negative bacteria such as *E. coli* (Saunders and Saunders 1999). However, several methods of transformation of *Bacillus subtilis* are available which includes protoplast transformation (Chang and Cohen 1979), conjugation (Naglich and Andrews, Jr. 1988), transduction (Alonso et al. 1986), natural transformation (Spizizen 1958) and electroporation (Vehmaanpera 1989). Although transformation of protoplast gives highest yield of transformants, the method is cumbersome to perform. The other two methods can not be used for transforming mutant libraries as required in directed evolution protocol. Both natural transformation and electroporation can be used to transform mutant library in plasmid but generally gives less yield of transformants ($> 10^3$ transformants/ μg of supercoiled plasmid DNA). Transformation efficiency can be increased 10 times by using plasmid multimers for natural transformation and by fine tuning the growth and electrical parameters during electroporation. Transformation efficiency of 10^4 transformants/ μg of plasmid DNA was routinely achieved with both these methods, which was sufficient for the semi-highthroughput screening. However, this makes it mandatory to either use plasmid multimers for natural transformation or very high quality supercoiled plasmid for electroporation. To facilitate this, lipase gene along with its signal sequence was cloned under the control of a strong gram positive promoter, *Hpa* II, in an *E. coli*-*B. subtilis* shuttle vector, enabling all cloning steps to be performed in *E. coli* and expression and secretion in *B. subtilis*.

The natural substrates of *B. subtilis* lipase are triglycerides but can also perform hydrolysis of single chain esters (Lesuisse et al. 1993). Several solid phase assays are available to monitor lipase or esterase activity on petriplates (Kouker and Jaeger 1987; Beisson et al. 2000; Gupta et al. 2003). Using tributyrin as substrate, emulsified in gum

arabic, solid phase assay was developed which is based on the hydrolytic activity of the enzyme and development of a clearing zone (halo) around the colonies, secreting lipase. This assay is qualitative but can be used for quick, initial screening of the mutants in a population range of $\sim 10^4$. The positive clones identified in this coarse screening can then be subjected to next level of screening which is more stringent and quantitative. The fact that *B. subtilis* lipase can act on single chain esters, p-nitrophenyl esters were used for activity measurements. Moreover, since lipase is secreted out in the culture supernatant, this was used as the source of enzyme. Assays with p-nitrophenyl esters are highly sensitive because of the high extinction coefficient of the liberated p-nitrophenol ($\epsilon_{405\text{nm}}=18.3 \text{ mM}^{-1}\text{cm}^{-1}$) and can be easily adapted to 96-well plate format, allowing activity measurement in microplate reader. p-nitrophenyl oleate (PNPO), micellized in Triton-X 100, was used as the preferred substrate as it shows negligible background hydrolysis, compared to shorter esters of p-nitrophenol, with crude culture supernatant. This is a well-studied reaction system for lipase reactions as the micellized substrate is highly reproducible and easy to handle (Acharya and Rao 2002). Stringent and quantitative screening of mutants in a population range of 10^3 can be performed routinely in 96-well format. Positive clones identified in this screen will be further tested individually using tube assays. Again, culture supernatants can be used as the enzyme source, excluding the need of protein purification, thus cutting down time and thus allowing testing of large numbers of mutants, individually.

The second most important step of directed evolution protocol is generation of genetic variation. We used the most common and easy procedure of generating mutant population by using error-prone PCR. It is easy to perform and the mutation frequency can be controlled easily. Both, *Taq* DNA polymerase, which induces error in the presence of Mn^{2+} and unbalanced dNTPs concentrations, as well as other DNA polymerases, which has high intrinsic error rate, can be use. We used Mutazyme[®] DNA polymerase from Stratagene, as; a, it does not require unbalanced dNTPs concentrations; b, gives good yield of PCR product and c, mutational frequency can be easily controlled by altering the template concentration in PCR. Mutation frequency was optimized to get 1 or 2 mutations per gene (690 bp), which is important to get mainly single mutants at protein level. This is important for initial round of mutagenesis as occurrence of more than one mutation confounds the contribution of individual mutation to the property.

Thus, by standardizing all important steps involved in directed evolution protocol, including cloning of gene in a suitable expression system, optimization of transformation protocols for host, development of assays involved in screening and optimization of mutagenic protocols, the stage was ready for *in vitro* evolution of *Bacillus subtilis* lipase for thermostability.

In vitro evolution of *Bacillus subtilis*
lipase for thermostability

Chapter 3

3.1 Introduction

Stability of a protein is measured by the energy required for the disruption of its structure. The energies associated with stability are known to be small in proteins, since the difference between native and denatured state is a sum of only a few weak interactions (Jaenicke 1996; Jaenicke and Bohm 1998; Jaenicke 2000b). The small difference between native and denatured states of proteins has a critical bearing on protein turnover within a cell. Several lines of investigations including structural properties of homologous proteins from extremophiles and contributions of mutations on stability of proteins point to the fact that knowledge based design for stable proteins is still a challenge (Jaenicke and Bohm 1998; Kumar and Nussinov 2001). Site-directed approaches to improve protein stability have limited success since the prediction of weak but profound interactions in a protein are not trivial. The need to improve our understanding of stability of proteins is also important because biotechnological applications demand stabilized proteins (Arnold 1993; Renugopalakrishnan et al. 2005).

Simultaneous improvement of stability and activity of a protein was considered to be difficult since structural requirements for each of them are considered to be mutually exclusive. Stability in a protein implies rigidification of protein molecule whereas activity requires flexibility. Data on thermophilic proteins, which were less active at mesophilic temperatures, supported such observations (Vieille and Zeikus 2001). However, recent research has indicated that this activity-stability trade off is possibly an outcome of evolutionary pressures rather than of physico-chemical limitations within a protein. Few thermostable enzymes from natural sources (Vieille and Zeikus 2001), and others generated in the lab have shown the catalytic proficiencies comparable to their mesophilic counterparts (Giver et al. 1998; Zhao and Arnold 1999; Miyazaki et al. 2000; Wintrode and Arnold 2000).

Extensive work has been performed to understand the basis of thermostability by altering the protein structure by rational design (Lehmann and Wyss 2001; Eijsink et al. 2004). The most common method followed to predict stabilizing mutations is based on comparisons of mesophilic proteins to their thermophilic homologues (Vogt et al. 1997; Cambillau and Claverie 2000; Kumar et al. 2000; Sterner and Liebl 2001). Although there are a number of successful examples available in which thermostability of proteins was considerably improved, rational designing was not always productive. Homology based

approaches have provided valuable information but are limited in scope due to involvement of multiple selection pressures during the evolution of proteins. In the last decade, protocols based on *in vitro* screening of large population of protein variants, collectively called as directed evolution methods, have provided extraordinary success in altering the stability, affinity and selectivity of proteins (Petrounia and Arnold 2000; Arnold et al. 2001; Turner 2003; Eijsink et al. 2005). Since subtle variations in protein structure are responsible for altered stability, directed evolution methods are probably a better choice for improving protein thermostability as they sample a large sequence space in a shorter time. Using directed evolution thermostability of a number of proteins has been improved significantly such as p-nitrobenzyl esterase (Spiller et al. 1999), subtilisins (Zhao and Arnold 1999; Miyazaki et al. 2000), β -glucosidase (Gonzalez-Blasco et al. 2000), 3-isopropylmalate dehydrogenase (Akanuma et al. 1999), amylases (Kim et al. 2003; Machius et al. 2003), xylanases (Andrews et al. 2004; Miyazaki et al. 2006), and lipases (Zhang et al. 2003; Acharya et al. 2004) etc. Structural information on thermostable protein variants, developed from a screening method with temperature as the only variable, may provide unambiguous relation between observed mutations and thermostability.

To understand thermostability in proteins, several variants of a lipase from *Bacillus subtilis* were generated by *in vitro* evolution. Lipases are very important as industrial biocatalysts and in detergent industry (Jaeger and Reetz 1998; Jaeger et al. 1999; Jaeger and Eggert 2002). *Bacillus subtilis* lipase (LipA), a product of *lipA* gene, with a mass of 19.34 kDa, is a monomeric secretory protein without any bound ligand. LipA has broad substrate specificity but preferentially hydrolyzes C-8 fatty acid esters and does not show interfacial activation due to absence of lid on the active site. It is a typical mesophilic enzyme having temperature optima of activity at 37 °C (Dartois et al. 1992; Lesuisse et al. 1993). The crystal structures of wild-type LipA and two thermostable mutants have been reported earlier (van et al. 2001; Kawasaki et al. 2002; Rajakumara et al. 2004). In order to understand the structural basis of thermostability and stability-activity relationship, thermostability of LipA was improved, without affecting its activity at mesophilic temperature, using directed evolution. Using a moderately thermostable triple mutant (TM) of LipA as a starting molecule (Acharya et al. 2004), a set of mutants with single as well as multiple mutations, having graded thermostability, was generated. Thermostability of the protein increases with the introduction of individual mutation without any loss of

activity at low temperature. The most thermostable mutant with nine mutations (three previously identified and six newly identified in this study) shows a remarkable enhancement of 15 degrees in melting temperature and 20 degrees increase in temperature optima of activity as compared to wild-type protein. Thermostabilization of protein was achieved along with improvement in activity over a vast range of temperature. Detailed characterizations of mutants revealed the contribution of surface exposed residues present on non-regular secondary structure or loops in improving thermostability of the protein.

3.2 Materials and Methods

3.2.1 Materials

All molecular biology reagents, kits and enzymes were same as described earlier in section 2.2.1 (Chapter 2). Phenyl Sepharose 6 Fast Flow (High Sub) was from Amersham Biosciences. Bis-ANS, PNPA and PNPB, were from Sigma Chemical Co. Urea was purchased from USB Corporation (Cleveland, OH). All other reagents used were of analytical grade or higher.

3.2.2 Mutagenesis

Lipase gene of triple mutant (TM) (Acharya et al. 2004) along with the signal sequence cloned in pMK-TM was mutagenized by error-prone PCR using GeneMorph[®] Random Mutagenesis kit from Stratagene as described earlier in Section 2.2.6 (Chapter 2). A mutation condition which should generate low error frequency of ~ 2-3 mutations per 1000 bases was used. Mutagenic PCR product obtained was purified, digested with *Pst* I and *Hind* III and ligated in a similarly digested pSA01 plasmid. The mutagenic library thus obtained was used to transform *E. coli* DH5 α by electroporation.

3.2.3 Screening

Thermostability of mutants was assessed by following a three tier screening protocol. Initially coarse screening was performed on petriplates followed by a stringent screening in 96-well plate format. Positives were further confirmed by residual activity measurements in tube-assays (Figure 3.1).

Mutant library obtained by error-prone PCR was used to transform *E. coli* DH5 α . Cells were plated on LB (1% Tryptone, 0.5% Yeast extract and 1% NaCl) agar plates

having 75 $\mu\text{g/ml}$ ampicillin and incubated at 30 °C for 24 hours. All the colonies ($\sim 10^6$ per μg of ligated product) were scraped and plasmid was purified using QIAGEN® Plasmid Mini kit. The supercoiled plasmid thus obtained was used to transform *Bacillus subtilis* BCL1050 by

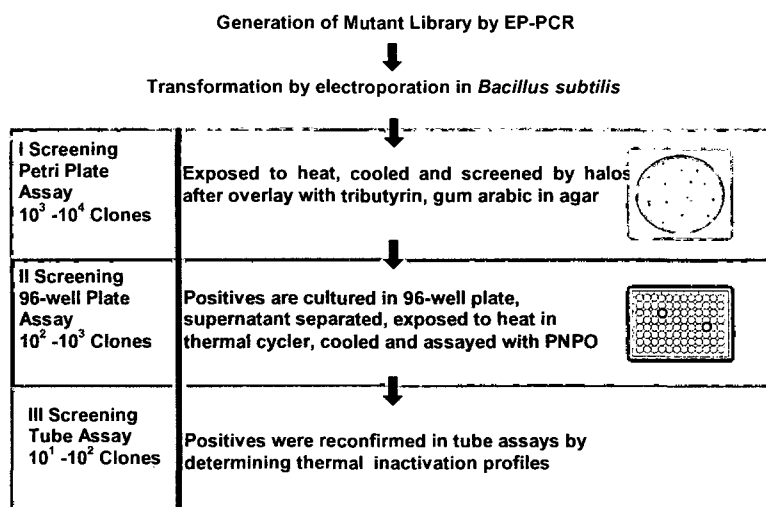


Figure 3.1 Schematic representation of the three-tier protocol employed to screen thermostable variants.

electroporation, as described in section 2.2.3 (Chapter 2). The transformants were patched on three similar LB agar plates having 5 $\mu\text{g/ml}$ kanamycin and incubated for 6 hours at 37 °C followed by 6 hours at 4-8 °C for enzyme secretion into the surroundings. Two set of plates were brought back to room temperature followed by incubation of one set at high temperature for 45 minutes followed by cooling back to room temperature. A thin layer of soft agar (0.5%) having 0.5 % tributyrin emulsified in 0.4 % gum arabic and 0.2 % sodium azide was poured over these two sets of plates followed by incubation at 37 °C for 6-8 hours. Appearance of clear haloes around the colonies upon incubation indicates the enzyme activity. Mutants showing clear haloes upon exposure to high temperatures were further screened for thermostability in 96-well plate format.

All the mutants showing high thermostability in petriplate assay were inoculated in individual wells of 96-well plates containing 200 μl of TB medium (1.2% Tryptone, 2.4% Yeast extract and 0.4% Glycerol) having 0.4 % gum arabic and 5 $\mu\text{g/ml}$ kanamycin. After 8 hours of growth at 37 °C, the cultures were used to inoculate fresh medium in an identical plate and incubated further for 8 hours. Cells were harvested by spinning the plates at 4000 rpm for 30 minutes at 4 °C in an Eppendorf centrifuge 5810 R. Hundred microliters of supernatant was diluted 1:1 with 100 mM sodium phosphate buffer (pH 7.2) and split into two identical 96-well PCR plates (Axygen Scientific, Union City, CA). One plate was incubated at high temperature for 20 minutes, cooled at 4 °C for 20 minutes followed by equilibration at 25 °C for 15 minutes in a PCR thermal cycler (GeneAmp 9700, Applied Biosystems, Foster City, CA). Other plate was incubated under identical

conditions except incubation at high temperature. Activity measurements were performed by mixing 80 μ l of diluted supernatant with 80 μ l of 2 \times PNPO-Triton X-100 substrate solution (2 mM PNPO micellized in 40 mM Triton X-100). Absorbance at 405 nm was recorded using ELISA reader (Spectramax 190, Molecular Devices, Sunnyvale, CA) at regular intervals while data was analyzed by using SoftMaxPro 4.7.1 software provided along with the instrument. Ratio of activity of each clone after incubation at higher temperature versus without incubation was taken as residual activity and was used to identify the positives.

Positives were further confirmed by performing tube assays. Ten milliliters of cell culture in TB medium was harvested after 16 hours of incubation at 37 °C and supernatant was diluted 1:1 with 100 mM Sodium Phosphate buffer (pH 7.2). Hundred microliters of diluted supernatant was incubated at various temperatures for 20 minutes followed by cooling at 4 °C for 20 minutes and final equilibration at 25 °C for 15 minutes in PCR thermal cycler. Activity was measured by adding 80 μ l of diluted supernatant to 920 μ l of 50 mM sodium phosphate buffer (pH 7.2) having 0.2 mM PNPO, emulsified in 20 mM Triton X-100, as substrate. Increase in absorbance at 405 nm with time was measured using spectrophotometer (Hitachi U-2000, Japan). Residual activity corresponding to each temperature of incubation was determined by taking the ratio of activities obtained upon incubation with that of without incubation. The temperature at which enzyme loses 50% of its activity after incubation for 20 minutes (T_{50}) was taken as the thermostability index for the selection of mutants. Plasmid was purified from mutants displaying highest T_{50} with respect to parent and gene was sequenced to identify the mutation occurred.

3.2.4 Recombination

Mutations obtained in generations 1 and 3 by error-prone PCR were recombined in generations 2 and 4 respectively, by restriction digestion and ligation using unique restriction sites within the structural gene (Figure 3.2). Two unique restriction sites corresponding to that of enzyme *Age* I, which lies between the codons of 52nd and 54th amino acid residues, while that of *Hae* II between the codons of 112th and 114th amino acid residues, were used for fragmentation of gene. Briefly, mutant 2D9 was created by ligation of fragments obtained by restriction digestion of clones 1-8D5, 1-14F5 and 1-17A4 by *Age* I and *Hae* II. The genes coding for the three proteins were amplified by PCR using primers Kpn1For and PMKD1. A four point ligation using *Pst* I-*Age* I fragment

from 1-8D5, *Age* I-*Hae* II fragment from 1-14F5, *Hae* II-*Hind* III fragment from 1-17A4 and pSA01, digested with *Pst* I and *Hind* III was performed and transformed in *E. coli* DH5 α for selection. Mutant 4D3 was made by combining mutation

A15S with A20E by site-directed mutagenesis on 3-11G1 template using mutagenic primers A15SF 5'-GGTATTGGAGGGTCATCATCCATTT-3' and A15SR 5'-CAA AATTGGATGATGACCCTCCAATA-3' (mismatch codon italicized) using standard procedures (Sambrook and Russel 2001). Forward and reverse mutagenic primers were used with PMKD1 and Kpn1For primers respectively to amplify the two fragments by PCR encoding first and second half of the structural gene having the desired mutation. In second round of PCR these two fragments were mixed, extended and used as template to get full length structural gene having the desired codon change. This fragment was digested with *Pst* I and *Age* I and ligated along with *Age* I-*Hind* III fragment from 3-18G4 in *Pst* I-*Hind* III digested pSA01 to recombine all the three mutations in 4D3.

3.2.5 Site-saturation mutagenesis

Site-saturation mutagenesis was carried out using a pair of oligonucleotide primers. The target amino acid position was coded by NNK (sense strand) and MNN (antisense strand), where N = A, G, C or T, K = G or T and M = A or C. The reaction was carried out using overlap extension method using two vector specific primers Kpn1For and PMKD1 along with position specific mutagenic primers (Table 3.1) using standard procedures (Sambrook and Russel 2001). Screening of mutant library was performed as described above.

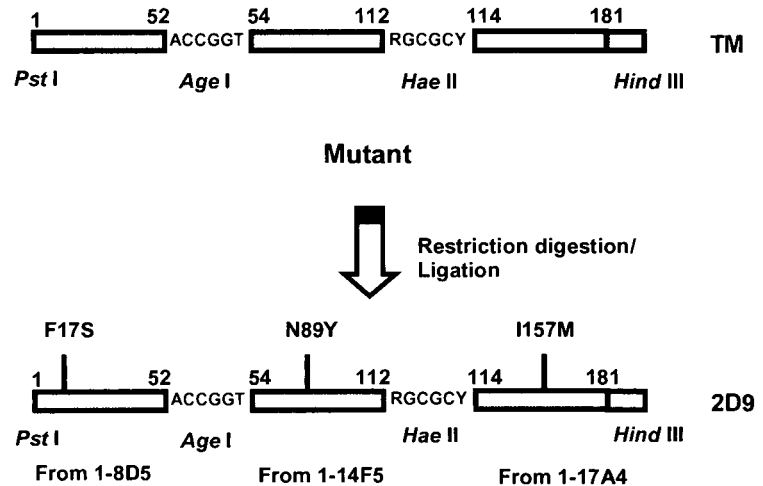


Figure 3.2 Schematic representation of the strategy used for recombination of mutations. Mutations were recombined by ligation of fragments, obtained from different mutants by restriction digestion, using two unique restriction sites within the structural gene.

Table 3.1. Oligonucleotide sequences used for site-saturation mutagenesis.

Primer	Position	Sequence ^{a,b}
S15X-U	15	5'-GGTATTGGAGGG <u>NNK</u> TCATCCAATTTTGAG-3'
S15X-D	15	5'-CTCAAATTTGGATG <u>MNN</u> CCCTCCAATACC-3'
S17X-U	17	5'-GGAGGGTCATC <u>ANNK</u> AATTTTGAGGG-3'
S17X-D	17	5'-CCCTCAAATTT <u>MNN</u> TGATGACCCTCC-3'
E20X-U	20	5'-GTCATCATCCAATTTT <u>NNK</u> GGAATTAAGAGCT-3'
E20X-D	20	5'-AGCTCTTAATTC <u>MNN</u> AAAATTGGATGATGAC-3'
Y89X-U	89	5'-ACTTTACTACATAAA <u>NNK</u> CTGGACGGCGGA-3'
Y89X-D	89	5'-TCCGCCGTCCAG <u>MNN</u> TTTTATGTAGTAAAGT-3'
D111X-U	111	5'-GTTTGACGAC <u>ANNK</u> AAGGCGCCTCCG-3'
D111X-D	111	5'-GGAGGCGCCTT <u>MNN</u> TGTCGTCAAACG-3'
M157X-U	157	5'-GGCGTTGGACAC <u>NNK</u> GGCCTTCTGTAC-3'
M157X-D	157	5'-GTACAGAAGGCC <u>MNN</u> GTGTCCAACGCC-3'

^a The target amino acid position was coded by degenerate codon NNK (sense strand) and MNN (antisense strand), where N = A, G, C or T, K = G or T and M = A or C.

^b Mutated codon underlined.

3.2.6 Plasmid DNA isolation and sequencing of lipase gene

Isolation of plasmid DNA from *E. coli* DH5 α or *Bacillus subtilis* BCL1050 and sequencing of lipase gene was performed as described earlier in Chapter 2 (Section 2.2.8).

3.2.7 Protein purification

For protein over-expression and purification, structural genes of wild-type and mutants, cloned in pMK3 shuttle vector, were PCR amplified using primers PMKNDE (5'-GCCGTCAGCACATATGGCAGAACACAA-3') and PMKBAM (5'-AGCGGATAACAGGATCCCACAGGAAACA-3') (*Nde* I and *Bam*H I sites underlined). The amplified products were digested with *Nde*I and *Bam*HI and ligated with similarly digested pET21b (Figure 3.3). The forward primer also introduced a start codon ATG which is part of the *Nde*I site. This would introduce an extra methionine residue, just before the N-terminal alanine of mature protein, upon expression in *E. coli*.

Wild-type and mutant proteins, cloned in pET21b, were purified upon expression in *E. coli* BL21 (DE3) as described earlier (Acharya et al. 2004). Since the mutants were very different in surface properties than wild-type protein, purification conditions for each mutant were optimized. All the mutants bind to phenyl sepharose and get eluted in the conditions similar to that of wild-type protein. However, in the subsequent ion-exchange chromatography step, mutants bind to different matrices depending on the isoelectric point

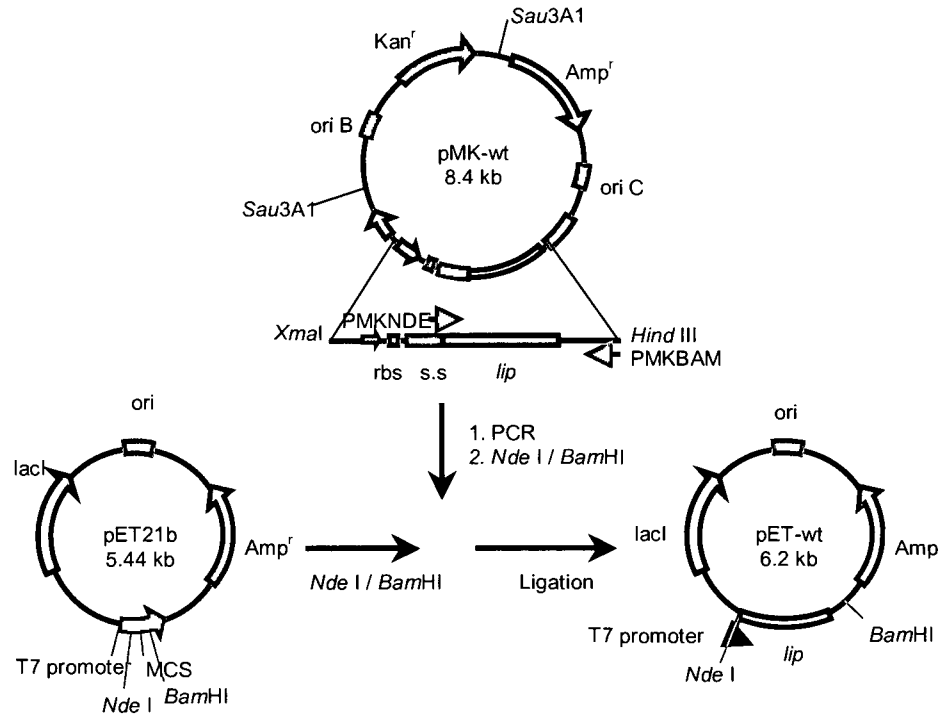


Figure 3.3 Sub cloning of lipase gene (*lip*) in expression vector pET21b. Lipase structural gene was cloned under the control of T7 promoter for protein over-expression in *E. coli*.

(pI) of the protein. Wild-type, TM, 1-8D5, 1-14F5, 1-17A4, 2D9 and 3-3A9 bind to cation exchanger Mono-S (Amersham Biosciences) in 50 mM BICINE-NaOH buffer (pH 7.7) while mutants 3-11G1, 3-18G4 and 4D3 bind to anion exchanger Mono-Q (Amersham Biosciences) in 20 mM Glycine-NaOH buffer (pH 10.5). Elution of the bound protein was performed by using linear gradient of NaCl from 0 to 1 M in the corresponding binding buffer. Active fractions were pooled, dialyzed overnight against 2 mM Glycine-NaOH (pH 10) at 4 °C and stored at -20 °C. Purity was checked by running SDS-PAGE and was found to be more than 95 percent. Protein quantitation was done by the modified Lowry method (Markwell et al. 1981).

3.2.8 Circular dichroism

Circular dichroism spectral measurements were performed using JASCO J-815 spectropolarimeter equipped with Jasco Peltier-type temperature controller (CDF-426S/15). Near-UV CD spectra were recorded in the range of 310-250 nm using 1 mg/ml protein solution in 50 mM sodium phosphate buffer (pH7.2) in 1 cm path length cuvette. Far-UV spectrum measurements were performed in the range of 250-185 nm range using same protein solution in 0.02 cm path length sandwich type cuvette. All spectra reported

are the average of three accumulations. Spectra were recorded in ellipticity mode at a scan speed of 50 nm/minute, response time of 2 seconds, band-width of 2 nm and data pitch of 0.2 nm. All spectra were corrected for buffer baseline by subtracting the respective blank spectra recorded identically without protein. Mean residue ellipticity was calculated using equation

$$[\Theta]_{MRW} = (M_r \times \Theta) / (10 \times l \times c)$$

where Θ is the ellipticity in degrees, l is the path length in cm and c is the concentration in g/ml. Mean residue weight (M_r) of 115 was used (Kelly and Price 1997).

3.2.9 Fluorescence

Fluorescence measurements were performed using Hitachi F-4500 fluorimeter. Protein concentration of 0.05 mg/ml, in 50 mM sodium phosphate buffer (pH 7.2) was used. Intrinsic tryptophan fluorescence was recorded between 310 and 380 nm by exciting the sample at 295 nm. The excitation and emission band passes were set at 5 nm while all spectra were recorded in corrected spectrum mode. All spectra were corrected for buffer baseline by subtracting the respective blank spectra, without protein, recorded under identical conditions.

Bis-ANS binding was monitored by adding bis-ANS to a final concentration of 10 μ M to 0.05 mg/ml protein in 50 mM sodium phosphate buffer (pH 7.2). A stock solution of bis-ANS was prepared in methanol and the concentration was determined by absorbance at 385 nm using extinction coefficient, $\epsilon_{385} = 16,790 \text{ M}^{-1}\text{cm}^{-1}$ (Sharma et al. 1998). Samples were excited at 390 nm while emission was recorded in the range of 400-600 nm. Both excitation and emission band passes were set at 5 nm while the scan speed was 120 nm/min.

3.2.10 Thermal inactivation

Heat treatment of purified protein was carried out by incubating the protein in 0.2 ml PCR tubes in a programmable thermal cycler for precise temperature control. Proteins (25 μ l of 0.05 mg/ml in 50 mM sodium phosphate buffer, pH 7.2) were heated at different temperatures for required time, cooled at 4 $^{\circ}$ C for 20 min followed by equilibration at 25 $^{\circ}$ C for 15 min. Samples were centrifuged to get rid of any aggregated protein before assaying for enzymatic activity. Thermal inactivation profiles were plotted by incubating

the enzymes at various temperatures (from 25 to 75°C) for 20 minutes followed by residual activity measurement at room temperature.

Kinetics of thermal inactivation was monitored by incubating the proteins at different temperatures, ranging from 51 to 70 °C, primarily at 55, 60 and 66 °C. At various time intervals, an aliquot was removed followed by residual activity measurement at room temperature. Typically, inactivation was followed until >80% of the activity was lost. The heat-treated protein sample (20 µl) was added to 1 ml of 50 mM sodium phosphate buffer (pH 7.2) having 2 mM PNPB, micellized in 20 mM Triton X-100, as substrate. Enzymatic activity was monitored by measuring the rate of increase in absorbance at 405 nm. $\ln(\text{Residual activity})$ was plotted against time of incubation. Inactivation rate constants (k_{inact}) were determined from the slope of the plots, while half-lives were calculated as $t_{1/2} = \ln 2/k_{\text{inact}}$.

3.2.11 Thermal unfolding

Thermal unfolding of lipase mutants was monitored by circular dichroism spectroscopy in a JASCO J-815 spectropolarimeter fitted with Jasco Peltier-type temperature controller (CDF-426S/15). The protein concentration used was 0.05 mg/ml in 50 mM sodium phosphate buffer (pH 7.2) with path length of 1 cm. Temperature dependent unfolding profiles were obtained by heating protein at a constant rate of 1 °C per minute from 25 to 85 °C and measuring the change in ellipticity at 222 nm.

3.2.12 Static light scattering

Static light scattering was performed on Fluorolog 3-22 fluorimeter, fitted with peltier based cuvette holder, controlled by LFI-3751 temperature controller (Jobin Yvon, USA). Both monochromators were set at 360 nm, while slit-widths at 2 nm each. Protein was added to 50 mM sodium phosphate buffer (pH 7.2), preheated at desired temperature, to a final concentration of 0.05 mg/ml. Data acquisition was started immediately after the addition of protein.

To monitor temperature dependent aggregation profiles of mutants, 0.05 mg/ml of purified protein in 50 mM sodium phosphate buffer (pH 7.2) was heated at constant rate of 1 °C/minute in Lambda-35 spectrophotometer, equipped with thermostatted cuvette holder attached to PTP-1 Peltier temperature programmer (Perkin Elmer). Increase in turbidity of the sample with increase in temperature was monitored at 360 nm.

3.2.13 Urea unfolding

Equilibrium unfolding in urea was monitored by using CD and fluorescence spectroscopy. Protein (0.5 mg/ml in 50 mM sodium phosphate buffer, pH 7.2) was incubated in varying concentration of urea for 12 hours at room temperature prior to measurements. Far-UV CD spectra were recorded in the 250-210 nm range using JASCO J-815 spectropolarimeter in 1 mm path-length cuvette. Scan speed of 100 nm per minute, response time of 2 s, bandwidth of 2 nm and data pitch of 0.2 nm was used for measurement. Each spectrum was an average of 3 accumulations. Same protein sample was diluted to 0.1 mg/ml concentration in equimolar denaturant solution and the fluorescence measurements were recorded using Hitachi F-4500 fluorimeter. The excitation wavelength was 295 nm and slit width of 5 nm was used for excitation and emission, while emission spectra were recorded from 310-400 nm in corrected spectrum mode. Unfolding profiles were determined by monitoring the change in ellipticity at 222 nm and shift in emission max (λ_{max}) of tryptophan fluorescence with increase in urea concentration. The concentration of stock urea solution was determined by measuring refractive index. The ΔG of the folding step was determined by performing the least-squares analysis of the unfolding, taking into account pre- and post-transition slopes (Pace 1986).

3.2.14 Activity measurements

All enzyme kinetic parameters were determined at 25 °C by measurements performed on thermostatted spectrophotometer (Lambda-35 attached with PTP-1 Peltier temperature programmer, Perkin Elmer) using PNPA as substrate. A concentrated substrate stock of water soluble PNPA was made in acetone (200 mM). Different volumes of substrate were added from stock to 1 ml sodium phosphate buffer (pH 7.2) to get varying concentration in the range of 0.05 – 2 mM, at which initial rate of substrate hydrolysis was monitored by monitoring increase in absorbance at 405 nm. The values for K_m and k_{cat} were derived from the corresponding Lineweaver-Burke plots.

Specific activities of lipase mutants at room temperature (25°C) were determined using 2mM PNPB, micellized in 20 mM TritonX-100 as the substrate. The use of micellized p-nitrophenyl esters as substrate has the advantage that the results are highly reproducible with very less sample to sample variations. However, this system can not be

used at higher temperatures because of the cloud-point of detergent. So, specific activities at elevated temperatures (from 25 to 70 °C) were monitored by using PNPB dissolved in acetonitrile. PNPB is a preferred substrate due to its high solubility in water as compared to other long chain substrate (p-nitrophenyl esters) and also due to very less spontaneous hydrolysis as compared to other substrates under identical conditions. Briefly, 970 µl of 50 mM sodium phosphate buffer (pH 7.2) was incubated at desired temperature for 5 minutes for temperature equilibration in 1 cm cuvette within the cuvette holder to which 1 µg of enzyme was added and incubated further for 1 minute. Reaction was started by adding 10 µl of PNPB from 100X stock (50 mM PNPB dissolved in acetonitrile) and rate of increase in absorbance at 405 nm was recorded. At each temperature, rate of spontaneous hydrolysis was also monitored for correction (Fourage et al. 1999).

3.3 Results

3.3.1 Generation of thermostable mutants

Mutant population of lipase was generated using error-prone PCR. The criteria for selecting a mutation depended on its contribution to thermostability in a three tier screening assay and also retention of activity at moderate temperatures comparable to parent. Only those mutants were selected which were showing high residual activities upon exposure to high temperature without affecting their catalytic efficiencies at low temperature. The lineage of mutants and the mutations incorporated during the course of evolution are depicted in the flowchart (Figure 3.4).

A moderately thermostable triple mutant (TM) (Acharya et al. 2004) was used to parent a mutagenic library generated by error-prone PCR. Approximately 7000 clones were screened for thermostability after exposure at 60 °C using a three tier screening protocol (Figure 3.1). After screening, 18 mutants, which were showing significantly higher thermostability at 60 °C compared to parent TM, were selected and sequenced to identify the mutations (Table 3.2). It was observed that 10 mutants out of 18 have either F17S or N89Y mutation in addition to other mutations in some clones. Importantly, mutant 1-8D5 and 1-14F5, which have only F17S and N89Y mutation, respectively, were displaying highest residual activities as compared to others. At position 17, another mutation F17Y was also found to be thermostabilizing, but the improvement it brought upon was significantly lesser than F17S. Another mutant 1-17A4 having only mutation,

I157M also showed high residual activity compared to parent. Other single mutants, having mutations V149I, and G158D were found to be moderately thermostabilizing (Table 3.2).

In second generation, in order to explore whether the three mutations F17S, N89Y and I157M act synergistically to improve thermostability, they were recombined in all possible combinations, using restriction digestion and ligation (Table 3.3). Thermal inactivation profiles of each mutant, having single as well as multiple mutations generated due to recombination, were made by monitoring residual activities in crude culture

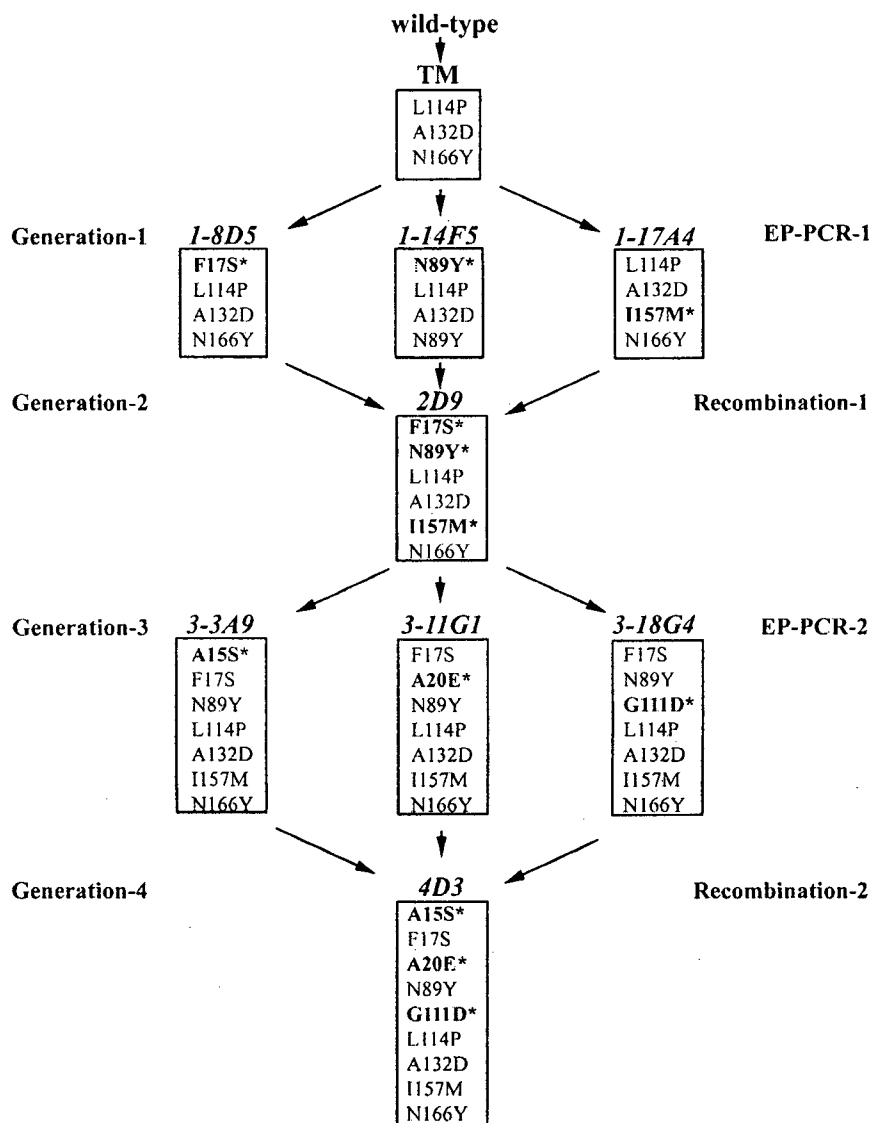


Figure 3.4. Lineage of thermostable variants of *Bacillus subtilis* lipase A. Amino acid substitutions accumulated through two rounds of error-prone PCR and recombination are shown. Newly introduced mutations in each generation are marked with asterisk.

Table 3.2. Thermostabilizing mutations isolated by error-prone PCR in generation 1.

Mutant	Codon change	Amino acid change	Residual activity ^a	Mutant	Codon change	Amino acid change	Residual activity ^a
TM	—	—	>15	1-12D11	GTA54GCA GTC74ATC	V54A V74I	25
1-3C10	TTC17TCC GTA27CTA	F17S V27L	30	1-13A5	GGC158GAC	G158D	30
1-3F5	AAT138AAC GTT149ATT	V149I	28	1-14F5	AAT89TAT	N89Y	45
1-5G6	CCG53CCA ATG137ACG GGC158GAC AGC167AGT	M137T G158D	37	1-14F9	AAT89TAT	N89Y	48
1-8D5	TTC17TCC	F17S	65	1-17A4	ATC157ATG AAC174AAT	I157M	47
1-9A11	TTC17TAC	F17Y	30	1-18B1	AAT89TAT CAA121CAG ACG180ACA	N89Y	44
1-10E6	TTC17TCC	F17S	63	1-22F2	GGC158GAC	G158D	32
1-10F1	AGC77AGT AAT89TAT	N89Y	47	1-23D1	GTC74GTT AAT89TAT ACA117ACT	N89Y	47
1-11C6	TTC17TAC CTG36TTG ACA47TCA CCG53CTG ATG134ACG	F17Y T47S P53L M134T	70	1-26D4	AAT89TAT ACG101ACT GTT149ATT ATC151ATA	N89Y V149I	48
				1-27A10	TTC17TCC	F17S	60

^a Monitored by incubating crude culture supernatant at 55 °C for 20 minutes, followed by activity measurement at room temperature.

supernatant upon incubation over a range of temperature. The temperature, at which 50% of activity is lost (T_{50}), shows that all the three mutations act in an additive manner to increase thermostability (Table 3.3). Thus mutant 2D9, which has all the three mutations, in addition to three of TM, (F17S, N89Y, L114P, A132D, I157M and N166Y), was most thermostable at 60 °C and used to parent next generation.

In the third generation, error-prone PCR was used again to generate mutant population using 2D9 as the parent. After screening 7000 clones at 65 °C, 16 mutants were selected, which were sequenced to identify mutation occurred (Table 3.4). Three single mutants 3-3A9, 3-11G1 and 3-18G4 that were displaying highest thermostability were found to have A15S, A20E and G111D mutations, respectively. In addition to that, out of 16 mutants selected, mutation A15S got selected in two clones while G111D in three clones. Mutation A20E was observed in only one mutant. Other single mutants, having mutations S16P, S24N, V149I, G158D and S163C, were found to bring modest improvement in thermostability (Table 3.4). In fourth generation, the three mutations, A15S, A20E and G111D, which were found to improve thermostability significantly, were recombined by restriction digestion ligation and site-directed mutagenesis in all possible combinations (Table 3.5). Thermostabilizing effect of each mutation was found to be additive as measured by thermal inactivation at different temperatures, (T_{50}), using crude culture supernatants. Combining all the mutations, mutant 4D3 was generated which shows highest thermostability at 65 °C and has nine mutations (A15S, F17S, A20E, N89Y, G111D, L114P, A132D, I157M and N166Y) as compared to wild-type protein.

Table 3.3 List of second generation mutants made after recombination of mutations identified in first generation.

Mutant	Mutations	T_{50} ^a	ΔT_{50} ^b
TM	—	51.9	—
1-8D5	F17S	54.5	2.6
1-14F5	N89Y	54.3	2.4
1-17A4	I157M	53.4	1.5
2A11	F17S, N89Y	56.7	4.8
2B12	N89Y, I157M	55.8	3.9
2C3	F17S, I157M	56.1	4.2
2D9	F17S, N89Y, I157M	58.6	6.7

^a Temperature at which enzyme loses half of its initial activity, upon 20 minutes of incubation (T_{50}), measured using crude culture supernatant.

^b Difference of T_{50} value of mutant and that of parent TM.

Table 3.4. Thermostabilizing mutations isolated by error-prone PCR in generation 3.

Mutant	Codon change	Amino acid change	Residual activity ^a	Mutant	Codon change	Amino acid change	Residual activity ^a
2D9	—	—	>15	3-13C6	GCA38GCT AGC163TGC	S163C	36
3-1D11	AGC24AAC	S24N	32	3-14C11	GTT149ATT	V149I	30
3-3A9	GCA15TCA	A15S	47	3-16F6	TCA16CCA CGG33CGA	S16P	40
3-4D1	CAC76CAT GGC158GAC	G158D	43	3-16G2	GGC111GAC ATT128ATA CTG160CTA	G111D	45
3-5D4	AGC24AAC AAG35AAA	S24N	30	3-18C11	AAC50ATC GGC111GAC	N50I G111D	47
3-5D8	CGA57CGG AGC163TGC	S163C	33	3-18G4	GGC111GAC TAC161TAT	G111D	45
3-6A3	GTT149ATT	V149I	35	3-18G5	GCA15GAA TCA16CCA	A15E S16P	50
3-11G1	GCG20GAG	A20E	46	3-25E8	GGC158GAC CTG160CTA	G158D	40
3-12D7	GCA15TCA GTT62GTC GGC111GGT	A15S	49				

^a Monitored by incubating crude culture supernatant at 60 °C for 20 minutes, followed by activity measurement at room temperature.

3.3.2 Characterization of lipase mutants

All the lipase mutants including wild-type were purified by standard protocols. The purity was checked by SDS-PAGE and was found to be more than 95% after staining with coomassie brilliant blue (Figure 3.5).

Spectroscopic characterization of lipase mutants was performed to probe any variation in the native state (Figure 3.6). Far-UV CD spectra of all the mutants completely overlap over each other, indicating invariant secondary structure (Figure 3.6 (a)). Tertiary structure as probed by near-UV CD shows that spectra corresponding to all mutants completely overlap except that of wild-type protein (Figure 3.6 (b)). The

characteristic peaks corresponding to that of aromatic amino acid residues were similar in case of wild-type and mutants except that ellipticity was higher in case of wild-type protein in the entire range of 250-295 nm. This might be due to the propensity of the wild-type protein to aggregate even at room temperature (Acharya and Rao 2003). This may cause wild-type protein to make soluble oligomers at room temperature, thus causing alteration in the chiral environment of the chromophores, causing shift in the entire spectrum. The mutants on the other hand do not show any aggregation at room temperature (data not shown), has very similar tertiary structure as shown by the complete overlap of near-UV CD spectra. Tertiary structure was further investigated by monitoring

Table 3.5 List of fourth generation mutants made after recombination of mutations identified in third generation.

Mutant	Mutations	T_{50}^a	ΔT_{50}^b
2D9	—	58.4	—
3-3A9	A15S	60.0	1.6
3-11G1	A20E	59.8	1.4
3-18G4	G111D	59.8	1.4
4A5	A15S, A20E	61.4	3.0
4B2	A20E, G111D	61.2	2.8
4C2	A15S, G111D	61.4	3.0
4D3	A15S, A20E, G111D	62.9	4.5

^a Temperature at which enzyme loses half of its initial activity, upon 20 minutes of incubation (T_{50}), measured using crude culture supernatant.

^b Difference of T_{50} value of mutant and that of parent 2D9.

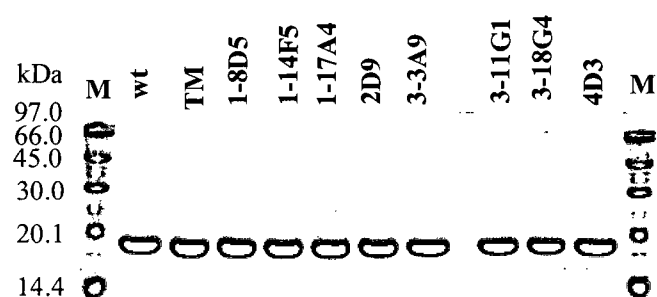


Figure 3.5 SDS-PAGE profiles of purified lipase mutants. M, Low molecular weight (LMW) marker (AmershamBiosciences).

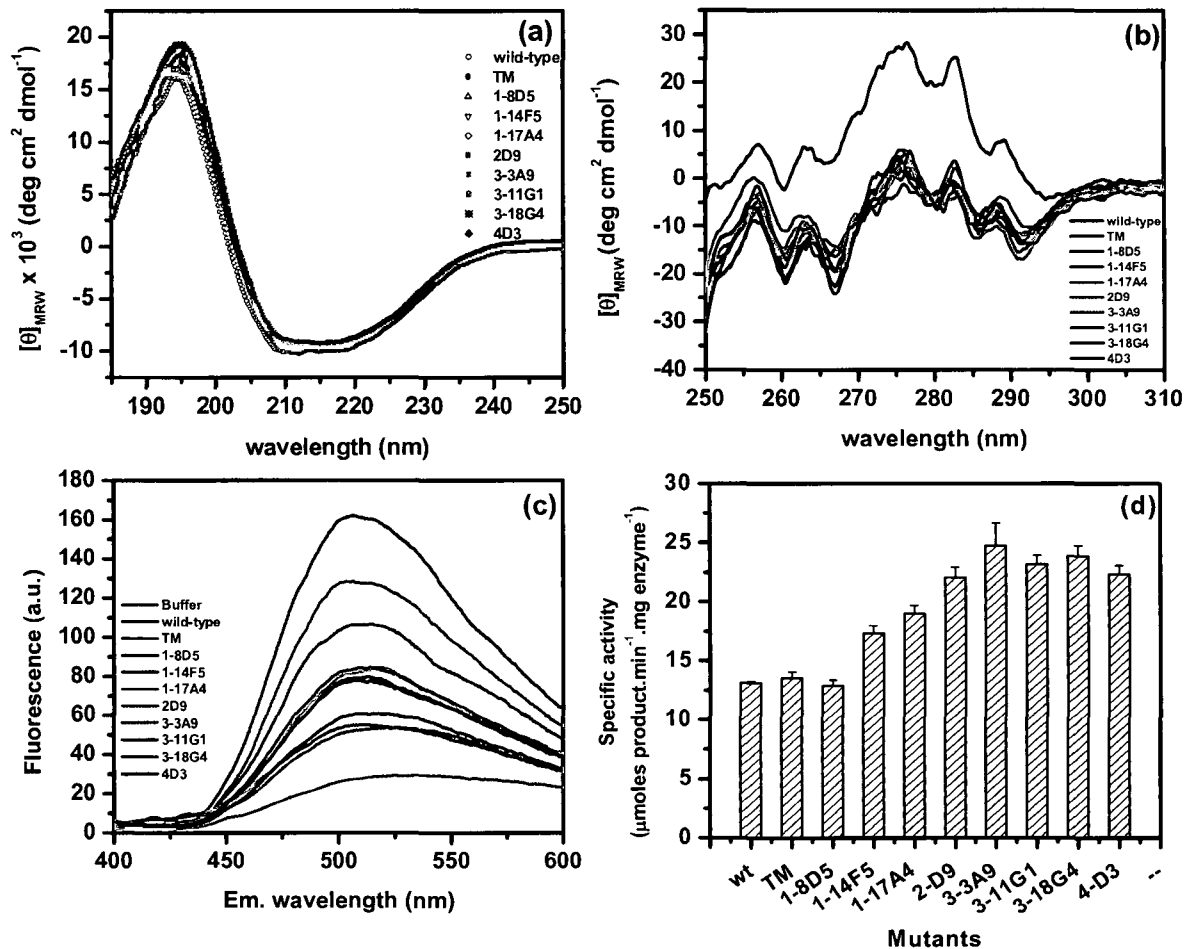


Figure 3.6 Spectral profiles and specific activities of lipase mutants. (a) Far-UV CD spectra, (b) Near-UV CD spectra and (c) bis-ANS binding spectra of lipase mutants. All spectra were recorded in 50 mM sodium phosphate buffer at 25 °C. (d) Specific activities of lipase mutants determined at 25 °C in 50 mM sodium phosphate buffer (pH 7.2), using 2 mM p-nitrophenyl butyrate (PNPB) micellized in 20 mM Triton-X 100 as substrate.

intrinsic tryptophan fluorescence and bis-ANS binding. The fluorescence emission spectra arising out of two tryptophan residues, (W31 and W42), overlaps with subtle variation in intensity in case of mutants including wild-type (data not shown). The emission maximum (λ_{max}) was found to be around 339 nm, indicating that tryptophans are partially buried. In the native state, tryptophan fluorescence is quenched due to close proximity of His 3 to Trp 31 (van et al. 2001). Little variation in fluorescence intensities among mutants may have arisen due to subtle change in the vicinity of Trp31. Tertiary structure of the lipase mutants was further probed by measuring their affinity to the hydrophobic dye bis-ANS. Bis-ANS is a valuable probe to investigate surface hydrophobicity of proteins. Upon

binding to hydrophobic pockets, the fluorescence of bis-ANS increases significantly, which can be used to titrate the surface hydrophobicity of the proteins (Sharma et al. 1998; Smoot et al. 2001). All mutants showed variable but significant binding to the dye (Figure 3.6 (c)). While wild-type protein shows maximum binding to bis-ANS, affinity of mutants decreases with accumulation of thermostabilizing mutations. While mutant TM and 2D9, having three and six mutations, respectively, shows intermediate binding to bis-ANS compared to wild-type protein, mutant 4D3, having all the nine mutations shows little binding. This suggests that wild-type protein has significantly higher exposed hydrophobic surface area as compared to mutants. *Bacillus subtilis* lipase is one of the few reported lipases which does not have lid over its active site (van et al. 2001). Since the substrates of this enzyme are hydrophobic triglycerides, the active site also has hydrophobic residues. Due to the absence of lid, wild-type protein displays a large patch of hydrophobic residues of the active site to the solvent which are responsible for the binding of bis-ANS. It is important to note that many mutations introduced during the course of evolution of this protein for thermostability occurred in the vicinity of the active site and most of them are hydrophobic to polar/charged substitutions. This might have reduced the hydrophobicity of the region, thus explaining reduced binding of thermostable mutants to hydrophobicity sensitive dye bis-ANS.

Specific activities of all the mutants, except TM and 1-8D5, was found to be superior than wild-type (Figure 3.6 (d)). While TM and 1-8D5 shows comparable specific activities than wild-type, other mutants show up to two fold increase at room temperature. First generation mutant 1-14F5 and 1-17A4, having mutations N89Y and I157M mutations respectively, displays significant increase in specific activity which was about 1.5 fold higher than parent TM and wild-type enzyme. Combining these two mutations with F17S of 1-8D5 in 2D9 further increases specific activity to about 2 fold than wild-type. However, mutations identified in third generation have little effect and as shown by similar specific activities of third and fourth generation mutants, all of which have ~ 2 fold higher specific activities than wild-type.

3.3.3 Thermal inactivation

To determine the resistance to irreversible thermal inactivation of wild-type lipase and mutants, enzymes were heated at different temperatures for 20 minutes followed by residual activity measurements at room temperature. For each mutant the transition was

sharp with equal width of transition. The temperature at which the enzyme loses half of its activity (T_{50}) increases with the accumulation of mutations (Figure 3.7, Table 3.7).

The rank order of T_{50} among the variants was wild-type < TM < (1-8D5, 1-14F5, 1-17A4) < 2D9 < (3-3A9, 3-11G1, 3-18G4) < 4D3. While wild-type shows T_{50} at 53 °C the same for TM was 58 °C. T_{50} of first

generation mutants, 1-8D5, 1-14F5 and 1-17A4 was 3, 1.7 and 2 degrees higher than the parent TM, respectively. Upon combining the three mutations in 2D9, T_{50} increases to 64 °C, which is 6 degrees higher than TM. Similarly, 3-3A9, 3-11G1 and 3-18G4 have T_{50} at 65 °C, one degree higher than parent 2D9. Upon combining the three mutations in 4D3, T_{50} increases by 4 degrees to 68 °C. These observations made by using purified proteins matches very well with the observations made earlier using crude culture supernatant (Table 3.3 and 3.5). All six mutations, when combined in all combinations, show additive nature of each mutation in increasing thermostability of the protein. Overall, the half-inactivation temperature (T_{50}) increases by 15 degrees upon accumulation of nine mutations in the wild-type protein.

Kinetics of irreversible thermal inactivation was monitored by incubating the enzymes at elevated temperatures followed by residual activity measurements at room temperature with respect to time of incubation (Figure 3.8). All the thermal inactivation traces were found to follow first order kinetics. Since the stability of the mutants differ widely, inactivation kinetics of all the mutants cannot be measured at same temperature. The half-life of inactivation of mutants, determined at three different temperatures, are compared in Table 3.7. At 55 °C the half-life of TM was 530 minutes while that of wild-type protein was just 2.8 minutes. At 60 °C, TM has a half-life of 4.4 minutes while that of 1-8D5, 1-14F5 and 1-17A4 was 47.5, 21.6 and 22.5 minutes respectively. Combining the three mutations in 2D9 increases the half-life at 60 °C to 1307 minutes, which is 300 fold

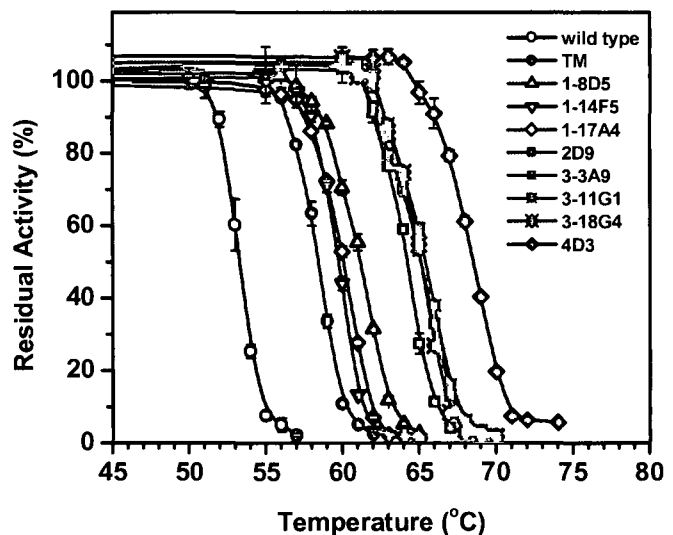


Figure 3.7 Thermal inactivation profiles of lipase variants. Enzymes (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2), were incubated at various temperatures for 20 minutes and assayed for residual activity at 25 °C.

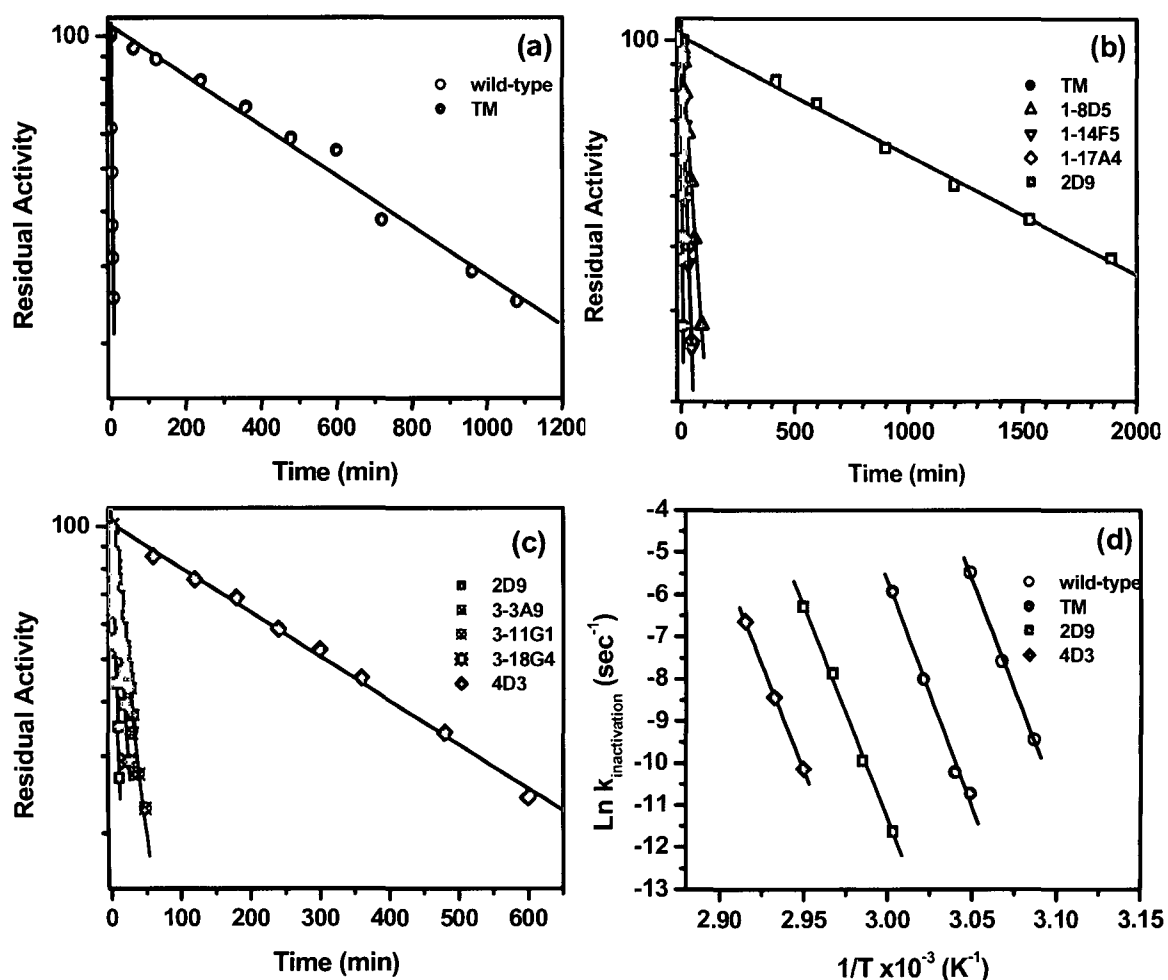


Figure 3.8 Thermal inactivation kinetics of lipase mutants. Time course of thermal stability was measured by calculating the residual activities at 25 °C after incubating the enzymes (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) at 55 °C (a), 60 °C (b) and 66 °C (c). Arrhenius plot of rate inactivation of mutants versus reciprocal of absolute temperature (d).

higher than TM. At 66 °C, 2D9 has half-life of 6.3 minutes while that of 3-3A9, 3-11G1 and 3-18G4 was 22.9, 22.8 and 14.7 minutes respectively. Combining these three mutations in 4D3 increases its half-life to 301 minutes at 66 °C, which is 50 fold higher than 2D9 at same temperature.

Since, stability of mutant 4D3 differs with wild-type so widely, direct comparison between the two can not be made at any temperature. For an indirect comparison, rates of inactivation wild-type along with mutants TM, 2D9 and 4D3 were monitored at different temperatures, ranging from 51 to 70 °C (Table 3.6). The rates of inactivation at different temperatures were used to draw Arrhenius plot of the mutants (Figure 3.8 (d)). The activation energies of the inactivation process does not show any significant difference

between wild-type and mutants as the corresponding values for wild-type, TM, 2D9 and 4D3 are 208.9, 214.8, 202.8 and 201.7 kcal/mol.

3.3.4 Thermostability of mutants

Thermostability of lipase variants was tested by monitoring temperature induced unfolding using circular dichroism (Figure 3.9, Table 3.7). The apparent melting temperature ($T_{m,app}$) of mutants increases in the same order as that of thermal inactivation, T_{50} (Table 3.7). Wild-type protein has $T_{m,app}$ of 56 °C while that of TM is 61 °C. Mutant 1-8D5, 1-14F5 and 1-17A4 are marginally better having $T_{m,app}$ of 64.4, 63.0 and 63.4 °C, respectively, which are 2-3 degree higher than TM. Combination of individual mutations in 2D9 increases $T_{m,app}$ to 67.4 °C, which is 6 degree higher than TM. Mutant 3-3A9, 3-11G1 and 3-18G4 melts at 68.7, 68.6 and 68.4 °C respectively. Combining all the mutations in 4D3 increases its melting temperature to 71.2 °C, which is approximately 4 degree higher than 2D9 and 15 degree

Table 3.6. Thermal inactivation parameters of wild-type lipase and its mutants.

Mutant	Temperature (°C)	Thermal inactivation k_{inact} (min^{-1})	Activation energy, E_a (kcal/mol)
Wild-type	51	0.0047	208.9
	53	0.030	
	55	0.25	
TM	55	0.0013	214.8
	56	0.0021	
	58	0.020	
	60	0.16	
2D9	60	0.00053	202.8
	62	0.0029	
	64	0.023	
	66	0.11	
4D3	66	0.0024	201.7
	68	0.013	
	70	0.078	

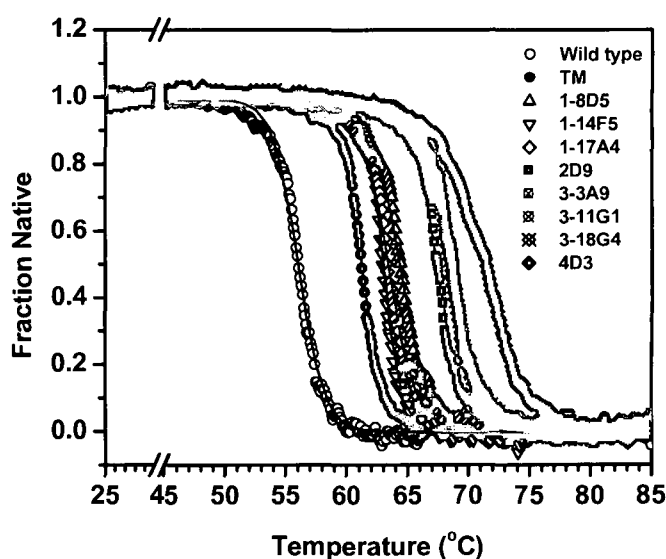


Figure 3.9 Thermal unfolding profiles of wild-type and mutant lipases as monitored by circular dichroism. Purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was heated in a cuvette of 1 cm path length at a constant rate of 1 °C/min. Change in ellipticity at 222 nm (Θ_{222}) with temperature was used to calculate unfolding transitions.

higher than wild-type protein. Collectively the thermal unfolding data is consistent with the observations made with thermal inactivation study.

3.3.5 Temperature induced aggregation of mutants

Temperature induced aggregation of mutants was monitored by static light scattering upon incubation of mutants at elevated temperatures (Figure 3.10). All mutants including wild-type were found to aggregate upon thermal unfolding, although data for only wild-type, TM 2D9 and 4D3 are shown in Figure 3.10 (a). All mutants, incubated at temperatures above which unfolding starts, show difference in maximum scatter values, which all the mutants appear to have achieved by 10 minutes of incubation. Wild-type shows maximum scatter values at 55 °C which was followed by TM, 2D9 and 4D3 at 60, 66 and 70 °C. This difference might be due to the difference in the amount of protein got aggregated at that particular temperature in the given time or it might be due to the difference in the size of the aggregate being formed.

In order to explore the aggregation trend of mutants compared to their unfolding, thermal unfolding profiles of mutants, as monitored by CD, were compared to their aggregation profiles, as monitored by static light scattering, upon heating proteins at an identical rate (Figure 3.10 (b)). It was observed that aggregation profiles of wild-type and all mutants, except 4D3, completely overlaps with their corresponding unfolding profiles.

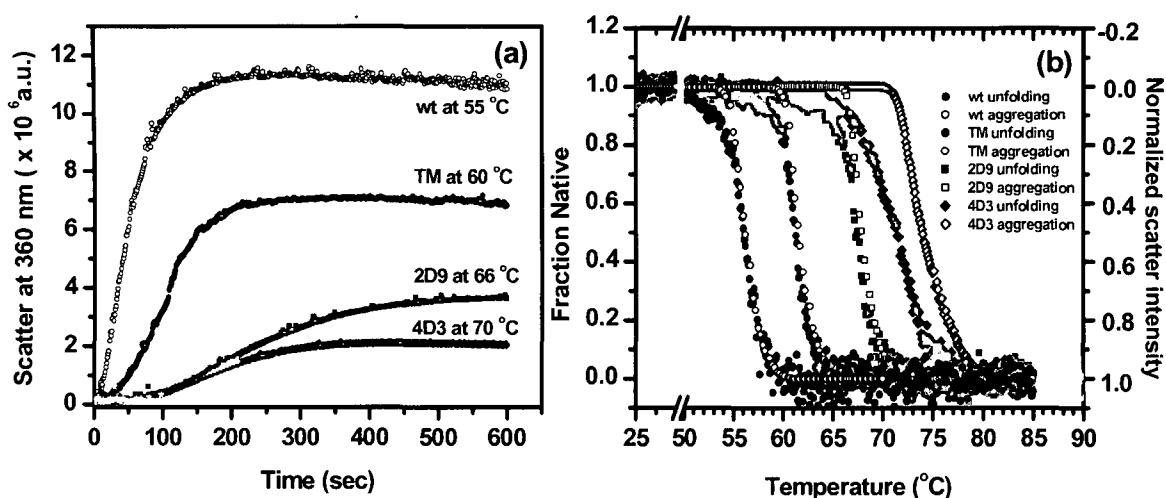


Figure 3.10 Aggregation profiles of wild-type lipase and its mutants. (a) Aggregation of wild-type lipase and its mutants upon incubation at elevated temperatures as monitored by static light scattering. (b) Aggregation profiles of wild-type lipase and its mutants, along with their corresponding unfolding profiles, as monitored by static light scattering and far-UV CD, respectively, upon heating the protein samples (0.05 mg/ml in 50 mM sodium phosphate buffer, pH 7.2) at a constant rate of 1 °C/min.

This suggests that the rate of aggregation of wild-type and that of mutants TM and 2D9 are comparable to that of their corresponding unfolding rates. However, in case of 4D3, the two profiles do not overlap but show that aggregation was significantly delayed compared to unfolding. This difference might be due to the thermostabilizing mutations introduced, thus affecting the denatured state of the protein, decreasing its propensity to aggregate.

3.3.6 Temperature dependence of the activity

The specific activity of wild-type, TM, 2D9 and 4D3 mutants, which differ significantly in thermostability, were determined over a range of temperature from 25-70 °C (Figure 3.11). The optimum temperature for wild-type enzyme was 35 °C which increases to 45, 50 and 55 °C in case of TM, 2D9 and 4D3, respectively (Table 3.7). Interestingly, with increase in temperature optima for activity, the activity per se in 4D3 has doubled at all temperatures compared to the wild-type. In addition, the activity vs. temperature profile was broader resulting in enhanced activities at all temperatures compared to the wild-type. This observation is in contrast to many reports on thermophilic proteins which show less activity at lower temperatures as compared to their mesophilic counterparts (Vieille and Zeikus 2001). This might be due to the double selection criteria employed during screening which picks only those mutants which gained enhanced thermostability without sacrificing activity at lower temperatures.

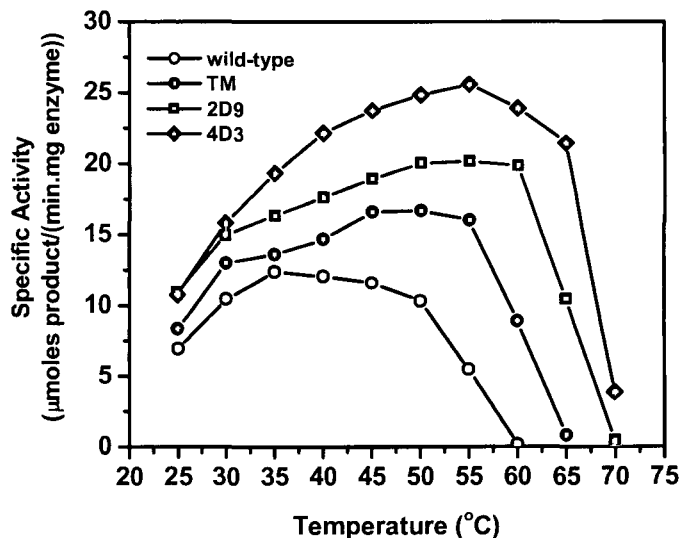


Figure 3.11. Temperature dependence of activities of wild-type lipase and its mutants. Enzymes (1 µg in 970 µl of 50 mM sodium phosphate buffer, pH 7.2) were incubated at various temperatures for 1 minute prior to activity measurement at the same temperature by the addition of substrate.

Table 3.7. Thermostability parameters of wild-type lipase and mutants.

Mutant	$t_{1/2}$ (min) ^a			$T_{(50)}$ ^b (°C)	$T_{m(app)}$ ^c (°C)	T_{opt} (°C)
	55 °C	60 °C	66 °C			
Wild-type	2.8	—	—	53.3	56.0	35.0
TM	530.0	4.4	—	58.3	61.2	45.0
1-8D5	—	47.5	—	61.1	64.4	—
1-14F5	—	21.6	—	59.7	63.0	—
1-17A4	—	22.5	—	60.0	63.4	—
2D9	—	1307.0	6.3	64.2	67.4	50.0
3-3A9	—	—	22.9	65.1	68.7	—
3-11G1	—	—	22.8	65.3	68.6	—
3-18G4	—	—	14.7	65.1	68.4	—
4D3	—	—	301.0	68.4	71.2	55.0

^a Half-life of thermal inactivation.

^b Temperature at which enzyme loses 50% activity upon incubation for 20 minutes.

^c Mid-point of thermal transition as calculated by circular dichroism spectroscopy.

3.3.7 Equilibrium unfolding in urea

Stability of wild-type and thermostable mutants was investigated by monitoring equilibrium unfolding behavior in the presence of urea using fluorescence and circular dichroism. Shift in tryptophan fluorescence emission (λ_{max}) and ellipticity at 222 nm (Θ_{222}) were monitored with increasing concentration of denaturant (Figure 3.12). The unfolding transitions in urea monitored by either fluorescence or circular dichroism were indistinguishable. Using the approach described by Pace (Pace 1986), the mid-point of unfolding transition and the ΔG values for urea induced unfolding were determined (Table 3.8). The mid-point of unfolding transition increases

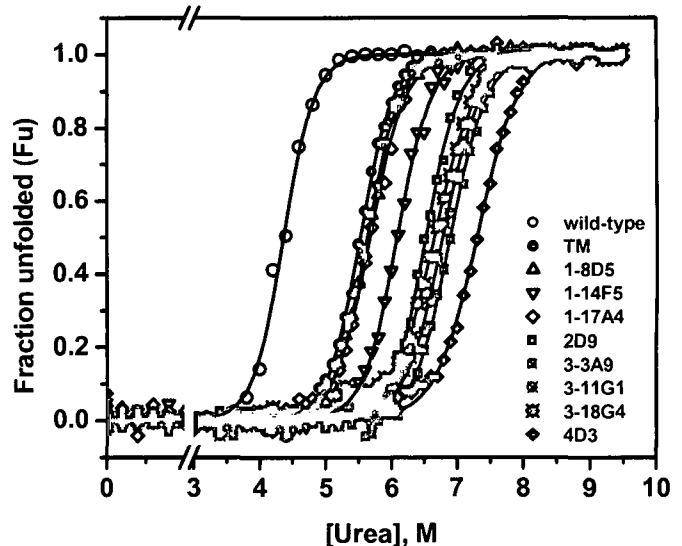


Figure 3.12. Equilibrium unfolding of wild-type lipase and its mutants in the presence of urea using CD. Fraction unfolded was calculated based on change in ellipticity at 222 nm (Θ_{222}) with urea concentration. (Identical transitions were obtained when fraction unfolded was calculated using shift in tryptophan fluorescence emission, λ_{max} , not shown here for the sake of clarity.)

with the accumulation of mutations. The concentration of urea in which 50% of wild-type, TM, 2D9 and 4D3 are unfolded (U_{50}) correspond to 4.4, 5.5, 6.5 and 7.3 M respectively. Also, the increment in U_{50} shown by mutants having combined mutations was found equal to the sum of increments shown by mutants having individual mutations. This again shows the additive nature of mutants in improving thermostability. The increase in concentration of denaturant required causing 50% denaturation is consistent with $T_{m,app}$ and thermal inactivation studies in showing that the mutant proteins are more stable than their progenitors.

The free energy of folding from the denatured state in the absence of denaturant ($\Delta G^{(H_2O)}$) was calculated by linear extrapolation of ΔG versus denaturant plot to 0 M Urea (Pace 1986) (Table 3.8). The increase in $\Delta G^{(H_2O)}$ for first and second generation mutants is consistent with their respective thermostability, however, same is not true with third and fourth generation mutants, which show less $\Delta G^{(H_2O)}$ as compared to their parents. This behavior is due to low m values displayed by third and fourth generation mutants, which may be due to altered interaction of the mutant proteins with the denaturant. Also these mutants, may fail to achieve fully unfolded state but have some residual structure in the denatured state as has been found for several proteins (Dill and Shortle 1991; Klein-Seetharaman et al. 2002).

Table 3.8. Equilibrium unfolding parameters of wild-type lipase and its mutants.

Mutant	$\Delta G^{(H_2O)^a}$ (kcal mol ⁻¹)	m^b (kcal mol ⁻¹ M ⁻¹)	$\Delta G/m$ (M)
Wild-type	11.39±0.72	2.61±0.16	4.36
TM	13.20±0.17	2.39±0.03	5.52
1-8D5	13.63±0.30	2.42±0.05	5.63
1-14F5	15.24±0.43	2.51±0.07	6.07
1-17A4	13.36±0.28	2.36±0.05	5.77
2D9	15.38±0.45	2.36±0.07	6.52
3-3A9	14.79±0.30	2.14±0.04	6.91
3-11G1	14.96±0.42	2.20±0.06	6.80
3-18G4	13.23±0.26	1.99±0.04	6.65
4D3	14.34±0.20	1.96±0.03	7.32

^a ΔG is the difference between the free energies of folded and unfolded states of protein.
^b m is rate of change in free energy as a function of denaturant concentration.

3.3.8 Catalytic properties of lipase mutants

Using p-nitrophenylacetate as substrate (PNPA), the kinetic parameters of wild-type lipase and its mutants were determined (Table 3.9). The kinetic parameters, k_{cat} and K_m , show small but significant variations in all the mutants. For 1-14F5 and 1-17A4, a marginal increase in both k_{cat} and K_m was observed, the overall catalytic efficiency, as shown by k_{cat}/K_m

either remained unaltered as for 1-14F5 or improved substantially in case of 1-17A4 as compared to wild-type enzyme. In the case of most thermostable mutant 4D3, the parameter k_{cat}/K_m has increased significantly. It is interesting to note that the introduction of stabilizing mutations enhanced the catalytic efficiencies of these enzymes. This observation is likely due to the strategy employed for screening, which identifies clones that exhibit activities comparable to that of parent enzyme at room temperature besides displaying high residual activities after exposure to higher temperatures.

3.3.9 Structural basis of thermostability

Crystal structure of wild-type protein at different resolutions and that of thermostable triple mutant (TM) have already been reported (van et al. 2001; Kawasaki et al. 2002; Rajakumara et al. 2004; Acharya et al. 2004). In order to explore the structural basis of thermostability shown by mutants, all the mutations were modeled on the structure of thermostable triple mutant, TM (PDB id: 1T2N) (Figure 3.13). These observations along with those made from the crystal structure of mutants 2D9 and 4D3, as reported by Ahmad *et al* (Ahmad et al. 2008) are reported here.

Bacillus subtilis lipase A is a single domain protein with a minimal α/β hydrolase fold (van et al. 2001). A six-stranded parallel β -sheet forms the central scaffold while helices and loops present on both the sides sandwich the sheet in-between. Mutation F17S

Table 3.9. Kinetic parameters of wild-type lipase and its mutants.

Mutant	K_m^a (mM)	$k_{cat} \times 10^2$ (min^{-1})	$k_{cat}/K_m \times 10^2$ ($\text{mM}^{-1} \text{min}^{-1}$)
Wild-type	0.98	2.2	2.2
TM	0.93	2.0	2.2
1-8D5	1.17	2.8	2.4
1-14F5	1.79	5.0	2.8
1-17A4	1.25	4.9	3.9
2D9	0.81	2.8	3.4
3-3A9	1.08	3.6	3.4
3-11G1	0.87	2.7	3.1
3-18G4	1.13	3.6	3.2
4D3	0.79	2.9	3.7

^a Kinetic parameters were estimated from assays conducted at 25 °C using PNPA as substrate.

is located in the loop connecting $\beta 3$ strand to αA helix (Figure 3.14 (a)). Stabilization by this mutation appears to be due to preference for a polar residue over a nonpolar residue on the protein surface. Phenylalanine, a bulky and planar residue, being completely exposed to the solvent keeps the protein native structure under constant strain of entropically unfavourable polar-nonpolar interaction. Its conversion to a small polar serine in mutants relieves this strain upon establishing favourable interactions with polar solvent.

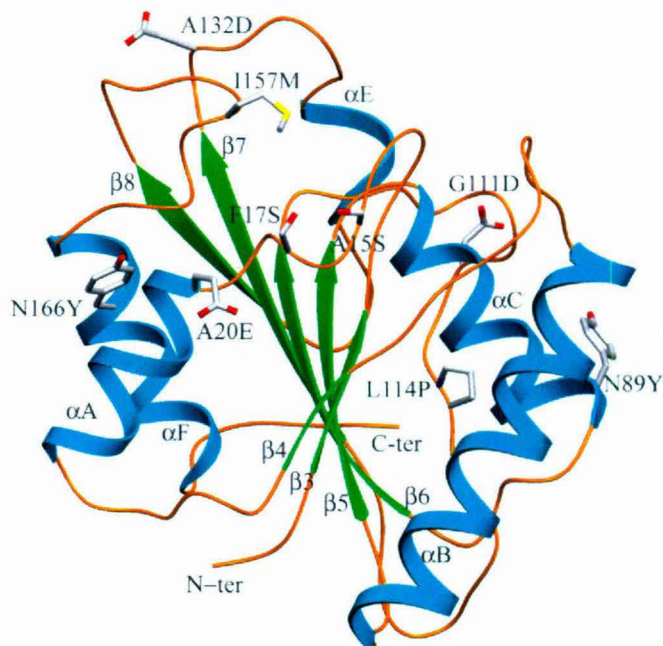


Figure 3.13. Ribbon diagram of the *Bacillus subtilis* lipase (LipA) indicating the relative position of each mutation.

Mutation N89Y occurred at the C-terminal residue of αC helix. The N-terminus of this helix is connected to $\beta 5$ strand through catalytic Ser77. C-terminal of this helix leads to $\beta 6$ strand through a short loop (Figure 3.14 (b)). Mapping of this mutation of the structure of TM does not provide any suitable explanation of the observed thermostability, however, crystal structure of mutant 2D9 shows that hydroxyl group of Tyr89, amino group of Lys88 and main chain carbonyl of Thr109 are all hydrogen bonded with a water molecule. While Tyr89 and Lys88 are ultimate and penultimate residues of αC helix, Thr109 is a part of longest loop of the molecule. Hence, this water molecule is bridging αC helix and the largest loop through hydrogen bonds resulting in a better anchoring of the loop to the core elements of the protein.

The third mutation I157M, which is surface-exposed, positioned at the first residue of G5, a five-residue long 3_{10} helix. In the wild-type lipase, CG2 methyl group of Ile157 is involved in van der Waals contact with Leu160 resulting in hydrophobic packing between the two residues (Figure 3.14(c)). Leu160 is a core residue. Replacement of Ile157 to methionine appears to have resulted in an increased van der Waals contact with Leu160 and consequently improves hydrophobic packing because of closer distance between

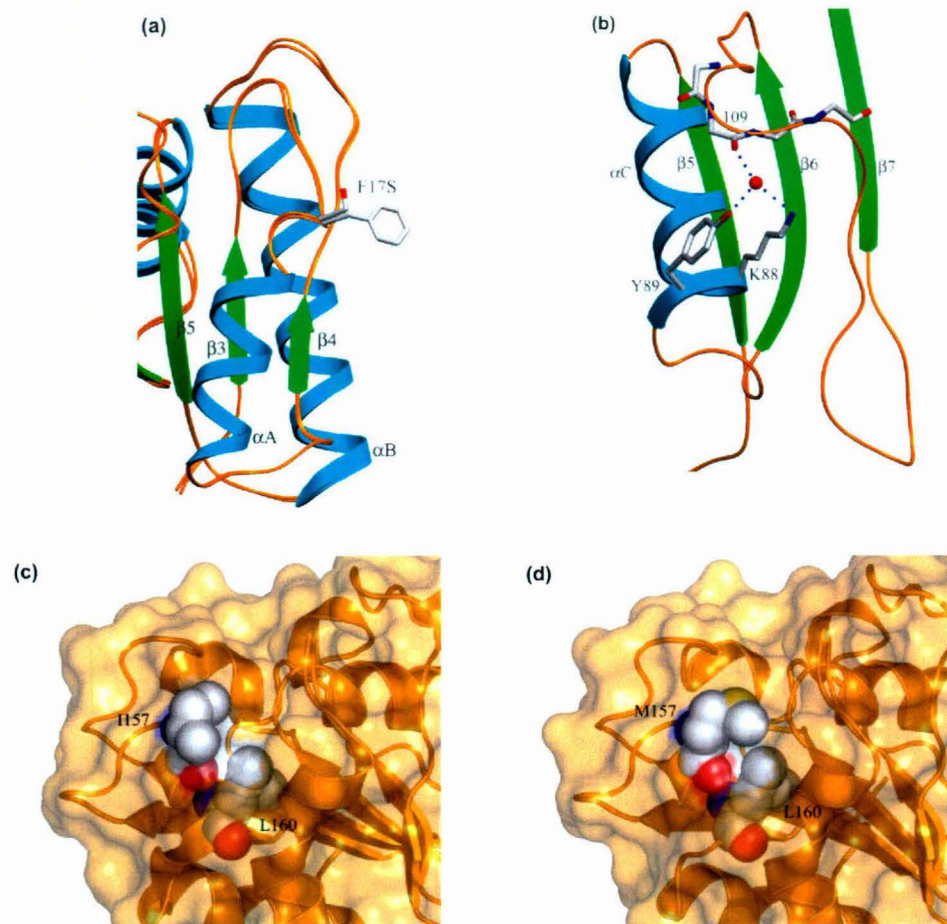


Figure 3.14. Structural changes in the regions around mutations identified in first generation. (a) Superimposed structure of mutant and wild-type protein around surface exposed mutation F17S. **(b)** A water molecule interacts with side chains of Lys88 and Tyr89 and also with the main chain carbonyl oxygen of Thr109. **(c)** Region around Ile157 shown in transparent, solvent accessible protein surface in wild-type. **(d)** Structural changes incurred due to I157M mutation are shown.

terminal methyl group of Met157 and Leu160 (Figure 3.14(d)). Interestingly, mutation I157M was independently identified as a thermostabilizing mutation in another study (Reetz and Carballeira 2007).

Mutation A15S is located in a loop that connects $\beta 3$ strand to αA continuous helix. Replacement of Ala15 to Ser, on one hand replaces solvent exposed hydrophobic side chain of alanine with polar serine, but also appears to establish a strong hydrogen bond with Ser 17, already present in 2D9 (Figure 3.15 (a)). This further improves stability to the loop with the introduction of an additional interaction.

Mutation A20E is present at the N-terminus of αA helix and is completely surface exposed. Presence of Glu20 at N-terminus of α -helix in mutants is a stabilizing change as it results in a favourable interaction between its negative charge and positive end of the

helix microdipole, thus improving the stability of the α -helix (Serrano and Fersht 1989). In addition to this, crystal structure of mutant 4D3 shows that side chain of Glu20 interacts with a water molecule, which in turn interacts with Ser24 (present on α A helix) and Arg33 (present on loop connecting α A to β 4) as shown in Figure 3.15 (b). In essence, this water molecule helps in holding two distinct secondary structural elements through hydrogen bonding.

The last mutation, G111D, is located in the longest loop of the protein. This loop is 14 residues long and connects preceding G4, a 3_{10} helix to β 7 strand. Presence of glycine in this structure could be presumed to be unfavourable due to intrinsic flexibility

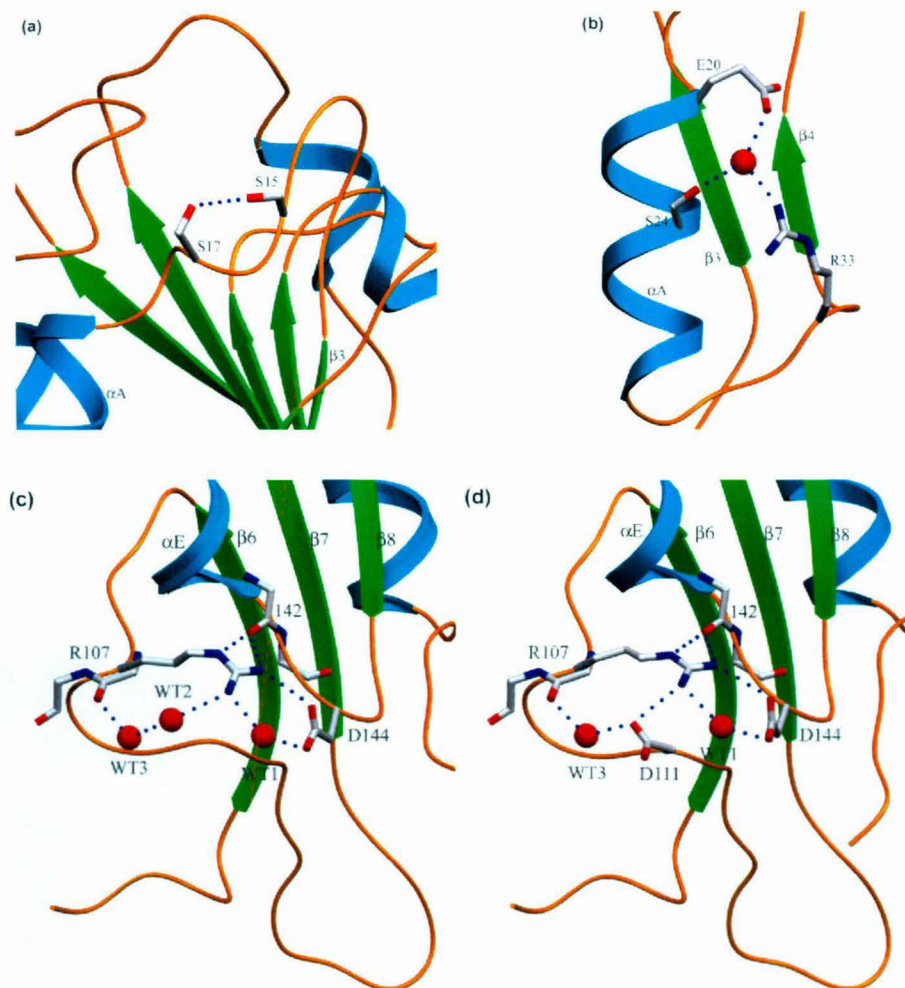


Figure 3.15. Structural changes in the regions around mutations identified in third generation. (a) A15S mutation in the background of F17S mutation makes hydrogen bond with the same. **(b)** Side chain of Glu20 making water mediated hydrogen bonds with side chains of residues Ser24 and Arg33. **(c)** Electrostatic network consisting of side chains of Arg107 & Asp144, main chain carbonyl groups of residues 107 & 142 and three water molecules named WT1, WT2 & WT3 in the wild-type protein. **(d)** G111D mutation replaces WT2 by carboxylate oxygen of Asp111 in mutant.

associated with this residue, replacement of which with aspartate stabilizes the loop. In crystal structure of wild-type protein, Gly111 is very close to an extensive electrostatic network comprising guanidino group of Arg107, carboxylate group of Asp144, main chain carbonyl groups of residues 107 & 142 and three water molecules (numbered WT1, WT2 and WT3) as shown in the Figure 3.15 (c). Crystal structure of 4D3 shows that mutation G111D replaces a centrally placed water molecule (WT2) that bridged side chain amino group of Arg107 and another water molecule (WT3) with its negatively charged OD oxygen atom (Figure 3.15(d)). Evidently, due to this substitution, relatively weak hydrogen bonding of Arg107 with water is replaced by stronger salt-bridge with Asp111, which also results in the appearance of a salt-bridge network comprising of residues Asp111, Arg107 and Asp144. Thus, stabilizing influence of G111D can be attributed to many factors viz., limiting the intrinsic flexibility associated by glycine residue, appearance of salt-bridge network (Karshikoff and Ladenstein 2001) and reducing the mobility associated with water molecules.

3.3.10 Validation of mutational effect by site-saturation mutagenesis

To verify the role of each mutation in conferring thermostability, site-saturation mutagenesis (Miyazaki and Arnold 1999) was performed at positions 15, 17, 20, 89, 111 and 157, identified during two rounds of random mutagenesis and screening. Mutant 4D3, which has all the nine mutations, was used as parent. For each position, ~ 400 mutants were screened after exposure to 70 °C for 20 minutes, a temperature at which 4D3 shows around 20% residual activity. Mutants showing higher residual activities were reconfirmed by calculating T_{50} using crude culture supernatant. Sequencing of clones displaying higher or comparable T_{50} to that of 4D3 revealed only two positions, 15 and 17 at which a substitution of previously identified mutation further increases the thermostability (Table 3.10). At position 15, substitution of serine (TCA) to aspartate (GAT) increases T_{50} by 1.8 degrees. Similarly, at position 17, substitution of serine (TCC) to glutamate (GAG) also improved T_{50} by 1.3 degrees. No other amino acid substitutions were identified at positions 20, 89, 111 and 157 which further improves thermostability of 4D3. At these positions, clones displaying comparable activities to that of 4D3 retained the mutations of the parent enzyme although the change in codon was observed. This corroborates the observation that at position 15 and 17, stabilization was brought upon by the substitution of surface exposed hydrophobic residues (Ala and Phe) with polar residues (Ser at both

Table 3.10 List of mutants obtained after site-saturation mutagenesis at six positions in 4D3.

Mutant	Position	Codon change	A.A. Change	EPCR/SSM ^a	T ₅₀ ^b	ΔT ₅₀ ^c
4D3	—	—	—	—	67.2	—
15-3B11	15	TCA→GAT	S15D	SSM	69.0	1.8
17-5C6	17	TCC→GAG	S17E	SSM	68.5	1.3
20-1A9	20	GAG→GAG	E20E	—	67.5	0.3
89-2F3	89	TAT→TAC	Y89Y	EPCR/SSM	67.0	-0.2
111-1G2	111	GAC→GAT	D111D	EPCR/SSM	67.8	0.6
157-5F3	157	ATG→ATG	M157M	—	67.5	0.3

^a Possibility of amino acid (A.A.) substitution obtained either by error-prone PCR (EPCR) or site-saturation mutagenesis (SSM).
^b Temperature at which enzyme loses half of its initial activity, upon 20 minutes of incubation (T₅₀), measured using crude culture supernatant.
^c Difference of T₅₀ value of mutant and that of parent 4D3.

positions), which further improves with increase in polarity as shown by the substitutions with charged residues at these two positions (Asp and Glu, respectively). On the other hand, no other substitution at rest of the four positions improves thermostability further, suggesting the crucial role played by each mutation, being involved in multiple interactions, to impart stabilization which is hard to replace by other substitutions.

3.4 Discussion

Sequence-property relationships derived from comparisons of proteins from organisms belonging to extreme phyla are often confounded by the multiple selection pressures experienced by the organism. Similar issues also arise in proteins evolved *in vitro*, especially when more than one mutation occurs per sequence in one round of mutagenesis. A ‘weightage’ assigned to each of the mutations would be essential for assessing the contribution of the mutation to the property (Yuen and Liu 2007). This issue was addressed in this study, to an extent, by controlling the mutational rate along with stringent screening and by evaluating the stability of mutants at protein level at each mutational cycle. A three tier screening protocol includes a coarse screening on petriplates which eliminates most of the inactive and less stable mutants followed by screening in 96 well plates under highly stringent conditions. Positives were further confirmed in tube assays and sequenced. While selecting a mutation for next generation, among several that occurred in one round, structural criteria of the substituted amino acids and also the

frequency of occurrence of mutation at a given site was used. Further, most importantly, by estimating the stability parameters (such as T_{50} , T_{mapp} and ΔG) on many mutant proteins with single mutations, the importance of a mutation to stability was assessed. Studies on each of the six mutant proteins along with the two combined mutant proteins (2D9 and 4D3) suggest that the contribution of each of the mutations on thermostability is additive. $\Delta G/m$ (U_{50}) clearly shows a progressive increase from the wild-type to 4D3, an increase by 3M of urea. Similarly, the mid point of transition, measured either by CD or residual activity and half-life of temperature denaturation has progressively increased with the addition of each mutation. Similar systematic evaluation of stability of each mutation was not attempted earlier, though several studies have reported improvement of stability after each round of mutation/screening cycles (Giver et al. 1998; Zhao and Arnold 1999). The information from solution studies indicates that these mutations alter local environment and their effects on structure were not synergistic but independent.

Spectroscopic characterization of the mutants showed little variation in the secondary and tertiary structure of the protein. This observation was further corroborated from the crystal structures of two mutants (Ahmad et al. 2008) which show no gross change in the overall structure compared to wild-type protein, except subtle variations in the vicinity of the mutations. This demonstrates that minimal changes in the structure are sufficient for enhancing thermostability (Dumon et al. 2008). Noticeable is the fact that all the nine thermostabilizing mutations so far created in this lipase are on the protein surface and most of them are present on non-regular secondary structure elements. Most of the mutations including A15S, F17S, G111D, L114P, A132D and I157M are present on various loops. Besides, other three mutations, which are present on three different α -helices, also bring their stabilizing influence through involvement of non-regular secondary structures. Further observation shows that two mutations, A20E and N89Y, happened to be one of the terminal residues of helices αA and αC respectively. Considering the prevalent thought that non-regular secondary structures are weak points in protein molecule from stability point of view and also that helix unfolding starts from their terminus, directed evolution approaches taken during thermostabilization of this protein have clearly identified weak points in protein molecule and stabilized them with additional interactions.

It is important to note that most of the mutations identified in this study (A20E, N89Y, G111D and I157M) stabilize the protein through weak interactions, such as charge-helix microdipole interaction (Serrano and Fersht 1989), hydrogen bonds (Vogt et al. 1997; Langhorst et al. 2000) and van der Waals contact. Cataloguing the stabilization mechanisms observed in studies, where directed evolution was combined with structure determination, reveal that largely the mechanisms involved stabilize loops, add salt bridges and hydrogen bonds or promote aromatic-aromatic interactions (Spiller et al. 1999; Wintrode and Arnold 2000). As mentioned above, most of the mutations are located on the protein surface and in the loop or turn region rather than in the elements of regular secondary structures such as helices and sheets. Lack of mutations in the core region would probably reflect the limited plasticity associated with the core region. Amino acid substitutions must be tolerated in a mutant protein and it is expected that the less ordered regions, loops and turns, would be more accommodative to the substitutions. In addition, disruption of packing in the core may interfere with the folding of the protein.

Further confirmation came from the validation of the role of each mutation identified in this study by site-saturation mutagenesis. Of the six positions, only at position 15 and 17, replacement of serine with aspartate and glutamate respectively, resulted in further improvement in thermostability. In the remaining four positions (20, 89, 111 and 157), the substitutions present in 4D3 are the best. This further supports the notion that replacement of solvent exposed non-polar side chains (Phe17 and Ala15) with highly polar charged residues on the surface seems to be highly favorable in increasing thermostability (Martin et al. 2001). On the other hand, none of the substitutions on the rest of the four positions could improve the lipase stability further. This suggests the specific role played by the side-chain of each residue (Glu20, Tyr89, Asp111 and Met157) in conferring thermostability which is hard to be substituted by other residues.

Mapping of all the mutations (total 19) identified during two rounds of random mutagenesis and screening on the structure shows that most of them are present on the loops or non-regular secondary structure followed by helices, while only two mutations occurred on β -sheets which forms the core of the molecule (Figure 3.16). Four mutations present on loops and two in helices were found to be truly thermostabilizing and characterized in detail. It is possible that other mutations, which were not selected for detailed study due to their modest contribution in improving thermostability, might also be important due to their location, while the amino acid change brought upon by point

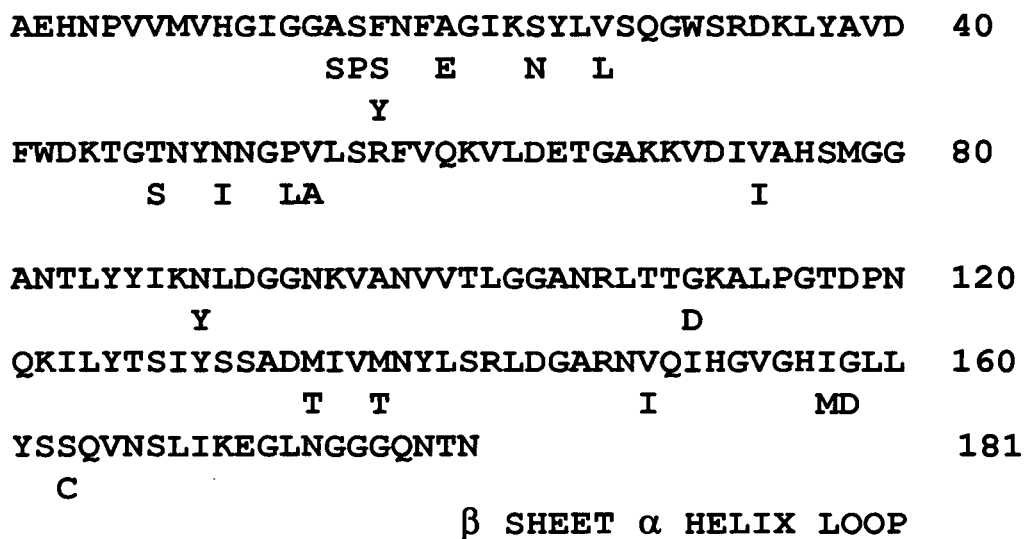


Figure 3.16 Relative locations of mutations, identified during two rounds of random mutagenesis and screening, on the primary structure of the protein. Out of 20 mutations identified, six mutations were characterized and found to be thermostabilizing, are marked in red while rest of 14 uncharacterized mutations are shown in black.

mutagenesis was not optimal (Miyazaki and Arnold 1999). Also, it is worthy to note that six out of nine mutations present in 4D3 are present on various loops which are part of the active site. This again suggests that this region in particular is more prone to thermal stress and mutations stabilizing this region of protein significantly improve thermostability of the whole protein. It is worthy to optimize all the locations, already identified by random mutagenesis and screening, particularly those present in loops in the vicinity of active site, by site-saturation mutagenesis to find the optimized substitution which improves the thermostability of the protein further.

During the course of *in vitro* evolution, LipA was evolved for thermostability without sacrificing its activity at lower temperature. It is interesting to observe that the catalytic efficiencies of all the mutants were not compromised while the thermostability was enhanced. Although the catalytic parameters k_{cat} and K_m show subtle but significant alterations upon introduction of mutations, the overall catalytic efficiency (k_{cat}/K_m) showed an improvement of 1.5-2 folds as compared to wild-type enzyme at room temperature. These mutants retain their native state at higher temperatures and perform catalysis till the onset of unfolding. Although we did not screen for the mutants which perform catalysis at high temperature, our screening criteria identified mutants that are thermoactive (Wintrode and Arnold 2000). Amide hydrogen / deuterium exchange, X-ray

analysis and molecular dynamic simulation studies on thermophilic and mesophilic proteins suggest that in thermophilic proteins the protein backbone dynamics are reduced compared to mesophilic proteins and the resultant conformational rigidity decreased the activity in thermostable proteins at lower temperatures (Jaenicke 2000a). Compared to the wild type, the mutants reported here resulted in improvement of thermostability and activity simultaneously, at all temperatures. The stabilizing influences of these mutations did not result in a loss of flexibility required for activity. It was observed that A132D and I157M are located at $<4 \text{ \AA}$ distance from Asp133 and His156, respectively of the active site, whereas the other mutations are further apart. The influence of the mutations on enhancement of activity in 4D3 at lower temperature requires further investigation. Effective screening and combination of mutations has resulted in a lipase mutant with both enhanced thermostability and activity at high temperatures. Such simultaneous improvement of stability and activity, reported uncommonly, provide us with important clues into the design of such mutations.

Further improvement in thermostability
of *Bacillus subtilis* lipase by site-
saturation mutagenesis

Chapter 4

4.1. Introduction

Though protein thermostability has been a subject of interest for long time, general methods for increasing protein stability are lacking (Jaenicke and Bohm 1998). Extensive work has been performed, to understand the rules that determine the protein stability, mainly by comparative genomics and structural analysis of thermophilic proteins with their mesophilic or psychrophilic counterparts (Kumar and Nussinov 2001; Vieille and Zeikus 2001). However, the sequence or structural differences observed in these studies are not confined only to the stability, since the proteins may have evolved in response to stresses other than temperature. Understanding sequence vs. stability relations become further complicated since proteins may have adapted multiple mechanisms to attain thermostability. More importantly, in proteins the free energy difference between native and its unfolded form is small (approx. 5-15 kCal/mol) which translates into few weak, hence non-covalent, interactions (Jaenicke 1996; Jaenicke 2000b). Predicting weak interactions is computationally difficult, consequently, knowledge based designing of proteins for improved stability is a challenging task (Jaenicke 2000b; Eijsink et al. 2004).

Directed or *in vitro* evolution methods to impart non-natural properties to proteins have met with phenomenal success in the recent past (Cherry and Fidantsef 2003; Chaput et al. 2008; Woycechowsky et al. 2007). This involves screening of genetic diversity created by random mutagenesis, under suitable constraints to evolve a property, such as stability, selectivity and affinity (Chirumamilla et al. 2001; Eijsink et al. 2005). The most common method used to generate genetic diversity is random mutagenesis using error-prone PCR (Leung et al. 1989; Chen and Arnold 1993), which like most of the other methods, introduces single base substitutions or point mutations. Although, the procedure is simple but has following limitations. On an average, with single base substitution, only 5.7 amino acid substitutions are accessible from any given amino acid residue (Miyazaki and Arnold 1999). A further consequence of the structure of the genetic code is that single base changes usually generates conservative amino acid substitutions, replacing one amino acid with others having similar physicochemical properties. As a result, a large sequence space of non conservative substitutions remains untapped. However, if the number of positions in the protein sequence is small, method of saturation mutagenesis could be applied to investigate the role of amino acids at those positions (Miyazaki and Arnold 1999; Parikh and Matsumura 2005).

Applying saturation mutagenesis at each site in a protein to substitute a residue with all the possible 19 residues is a daunting task, but the kind of information achieved can not be obtained otherwise. However, this method has been successfully performed to improve or optimize stability, selectivity and binding affinity of a number of proteins (Kretz et al. 2004). Alternatively, saturation mutagenesis can be used in combination with random mutagenesis to optimize the substitutions already identified by point mutagenesis (Geddie and Matsumura 2004; Miyazaki et al. 2006) or by some other rational approach (Reetz and Carballeira 2007). Site saturation mutagenesis at key locations, identified by random mutagenesis has been proved extremely useful in improving thermostability for a number of proteins such as protease (Miyazaki and Arnold 1999), xylanase (Palackal et al. 2004; Dumon et al. 2008), phosphite dehydrogenase (McLachlan et al. 2008), and lipase (Reetz et al. 2006). Using random mutagenesis during *in vitro* evolution of protein may accumulate conservative substitutions which may not provide best substitution, but it definitely identifies the key locations or “hot spots”, optimization of which by saturation mutagenesis may further improve the property of the protein. Drastic substitution with non conservative amino acids provides an opportunity to understand the ways involved in the improvement of property, in our case thermostability, which are seldom observed in nature.

To gain insight into the mechanisms employed by proteins to attain thermostability, lipase from *Bacillus subtilis* was evolved for thermostability by *in vitro* evolution (Chapter 3). Two rounds of random mutagenesis followed by screening were performed to generate a series of lipase variants with graded thermostability. The method employed for mutagenesis was error-prone PCR, which like most other methods incorporates point mutations. During each round of mutagenesis and screening, mutants displaying highest increase in thermostability, due to single mutation were selected for further investigations. However, a large number of single mutations, at varying locations, displaying modest increase in thermostability, were rejected. It was observed that since these positions were selected during screening, they might be the weak spots from stability point of view, however, the amino acid substitution incorporated by point mutation may not be optimal, which when optimized by saturation mutagenesis may increase the thermostability of the protein further.

The most thermostable variant 4D3, generated by *in vitro* evolution, has nine mutations and shows melting at temperature 15 degree higher as compared to wild-type

protein, along with 20 degree higher temperature optimum of activity. Most of these mutations were either located in the loops or unordered secondary structure region around the active site or if present in secondary structure element (helix) bring about their stabilizing effect by interacting with loops (Chapter 3). This suggests that the region around the active site, which is comprised of 6 loops, is the weak region in protein, stabilization of which can improve protein stability. Based on this assumption, four positions were selected, two each on two loops which are part of active site region. Two positions are present in a loop connecting $\beta 7$ strand and αE helix while the third one is present in the last loop of the protein, connecting $\beta 8$ strand with αF helix. The fourth position selected is the first residue of the last helix, αF , as the helix ends are also known to be weak spots, stabilization of which contributes to protein stabilization (Figure 3.13, Chapter 3). All these four positions were optimized in 4D3 by site-saturation mutagenesis, without affecting its activity at ambient temperatures. The most thermostable variant, which has total 12 mutations (nine previously identified in 4D3 and three newly identified in this study), shows additional enhancement of 7 degrees in melting temperature which is 22 degrees higher than wild-type protein. It also shows further improvement in temperature optima to 65 °C, which is 10 degrees higher than parent 4D3 and 30 degrees higher than wild-type enzyme. Importantly, improvement in thermostability is achieved with simultaneous improvement in catalytic efficiency over a vast range of temperature.

4.2. Materials and Methods

4.2.1. Materials

GdmCl was purchased from SERVA Electrophoresis GmbH (Heidelberg, Germany). All other reagents were of analytical grade or higher as described in previous chapters.

4.2.2. Construction of vectors used for screening and protein purification

The expression strategy used in this work is different from the methods describe earlier. By expressing the lipase in the periplasm of *E. coli*, the presence of cytoplasmic proteins in the assay can be avoided and the periplasmic protein also leaches out of periplasm into the medium. Mutant lipase (4D3) gene was cloned in pET22b expression vector (Novagen), under pelB signal sequence for periplasmic expression in *E. coli*. As

shown in Figure 4.1, structural gene, encoding 4D3 mutant lipase, cloned in pET21b, was PCR amplified using primers NcoFOR (5'-GAAGGAGATATACCCATGGCAGAACA-3') and T7-terminator (5'-GCTAGTTATTGCTCAGCGG-3') (*NcoI* site underlined). An internal *NcoI* site within the structural gene was knocked down by silent mutagenesis using two internal primers DNCOF (5'-GAAACGTTCAAATTCATGGCGTTGGA-3') and DNCOR (5'-GTCCAACGCCATGAATTTGAACGTTTCTA-3') (Sambrook and Russel 2001). The amplified product was digested with *NcoI* and *HindIII* followed by gel extraction. The digested and gel-purified product was ligated with similarly digested pET22b vector to get pET22-4D3. This construct was used for screening of thermostable mutants upon transformation of *E. coli* BL21 (DE3), which directs the expression of protein in periplasm upon induction with IPTG.

For protein over-expression and purification, structural genes of lipase mutants cloned in pET22b, were PCR amplified using primers PET22NDE (5'-

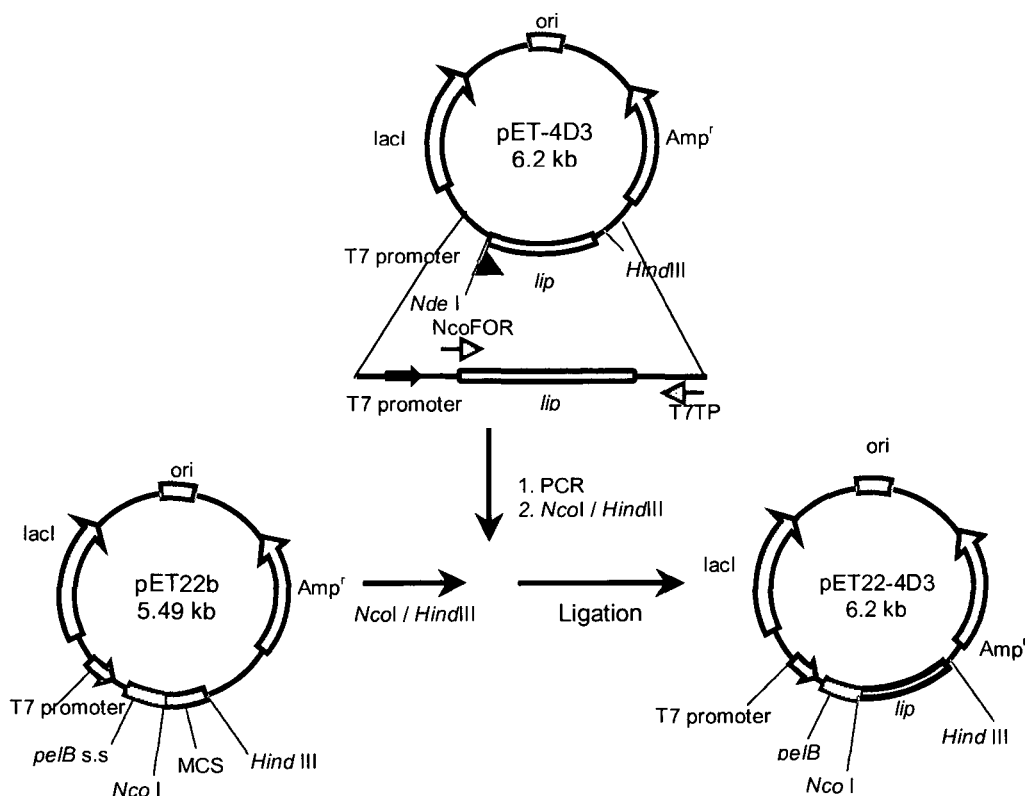


Figure 4.1 Cloning of lipase gene (*lip*) in expression vector pET22b. Lipase structural gene was cloned in-frame with *pelB* signal sequence, under the control of T7 promoter for periplasmic expression in *E. coli*.

GCCGGCGATGCATATGGCAGAACACA-3') and T7-terminator (*NdeI* site underlined). The amplified products were digested with *NdeI* and *HindIII* followed by gel extraction and ligated with similarly digested pET21b. The forward primer also introduced start codon ATG, which adds an extra methionine residue, before the N-terminal alanine of mature protein, upon expression in *E. coli*.

4.2.3. Site-saturation mutagenesis

Site-saturation mutagenesis at selected positions, in 4D3 mutant lipase gene, was performed by overlap extension method with little modifications (Sambrook and Russel 2001; Kretz et al. 2004). A schematic representation of the procedure is shown in Figure 4.2. A pair of internal primers, having the degenerate codons, was used to introduce mutations (Table 4.1). The target amino acid position was coded by NNK (sense strand) and MNN (antisense strand), where N = A, G, C or T, K = G or T and M = A or C. In the first round of PCR, position specific mutagenic primers were used along with vector specific primers T7-promoter (5'-TAATACGACTCACTATAGGG-3') and T7-terminator (5'-GCTAGTTATTGCTCAGCGG-3') to amplify first and second half of the structural gene, using pET22-4D3 as template. In the second round of PCR, these two fragments were mixed, extended and used as template to get full length structural gene having 32 kinds of codons, encoding all 20 kinds of amino acids, at the desired location. The full length amplified

products were digested with *NcoI* and *HindIII* followed by gel purification. Digested and purified products were ligated with similarly digested pET22b and used for transformation of *E. coli* BL21 (DE3) for gene expression.

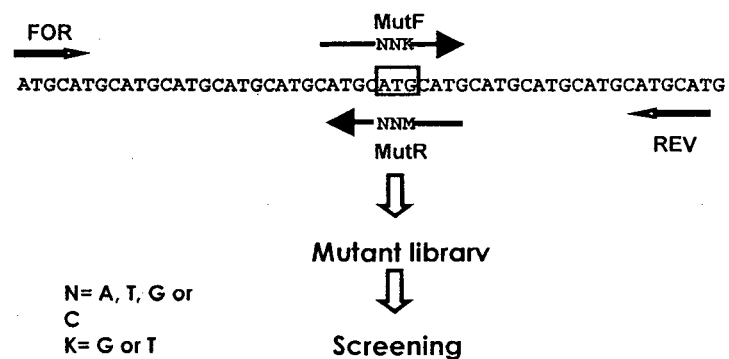


Figure 4.2 Schematic overview of site-saturation mutagenesis by overlap extension method. Two internal overlapping mutagenic primers having degenerate codons of target position are used in combination with vector specific forward and reverse primers. The two fragments generated were later used to amplify the whole gene.

Table 4.1. Oligonucleotide sequences used for site-saturation mutagenesis.

Primer	Position	Sequence ^{a,b}
M134X-U	134	5'-ACAGCAGTGACGAT <u>NNK</u> ATTGTCATAATTAC-3'
M134X-D	134	5'-GTAATTCATGACAAT <u>MNN</u> ATCGTCACTGCTGT-3'
M137X-U	137	5'-GACGATATGATTGTC <u>NNK</u> AATTACTTATCAAGAT-3'
M137X-D	137	5'-ATCTTGATAAGTAATT <u>MNN</u> GACAATCATATCGTC-3'
G158X-U	158	5'-GCGTTGGACACATG <u>NNK</u> CTTCTGTACAGCAGC-3'
G158X-D	158	5'-GCTGCTGTACAGAAG <u>MNN</u> CATGTGTCCAACGC-3'
S163X-U	163	5'-CCTTCTGTACAGC <u>NNK</u> CAAGTCTACAGC-3'
S163X-D	163	5'-GCTGTAGACTTGM <u>NNN</u> GCTGTACAGAAGG-3'

^a The target amino acid position was coded by degenerate codon NNK (sense strand) and MNN (antisense strand), where N = A, G, C or T, K = G or T and M = A or C.

^b Mutated codon underlined.

4.2.4. Screening

Thermostability of mutants was assessed by following a two tier screening protocol, similar to that described in chapter 3, except that the first step of coarse screening on petriplates was omitted. Initial screening was performed using a 96-well plate assay format followed by rigorous assessment of positives in tube assays.

The mutant library obtained after site-saturation mutagenesis was used to transform *E. coli* BL21 (DE3). Cells were plated on LB (1% tryptone, 0.5% yeast extract and 1% NaCl) agar plates having 100 µg/ml ampicillin and incubated at 37 °C for 12 hours. Transformants obtained were inoculated in individual wells of 96-well plates containing 200 µl of LB medium having 100 µg/ml ampicillin. Plates were incubated at 37 °C with shaking at 200 rpm for 6 hours. Grown cultures were used to inoculate fresh medium in identical plates and were incubated at 37 °C with shaking at 200 rpm. After 6 hours, cultures were induced by adding IPTG in each well to a final concentration of 0.5 mM and incubated further for 12 hours. Cells were harvested by spinning the plates at 4000 rpm for 30 minutes at 4 °C in an Eppendorf centrifuge, 5810 R. Supernatant (100 µl) was mixed 1:1 with 100 mM sodium phosphate buffer (pH 7.2) and split into two identical 96-well PCR plates (Axygen Scientific, Union City, CA). One plate was incubated at 70 °C for 20 minutes, cooled at 4 °C for 20 minutes followed by equilibration at 25 °C for 15 minutes in a PCR thermal cycler (GeneAmp 9700, Applied Biosystems, Foster City, CA). The other plate was incubated under identical conditions, except incubation at high temperature. Activity measurements were performed by mixing 80 µl of diluted supernatant with 80 µl of 2 x PNPB- Triton X-100 substrate solution (4 mM PNPB

micellized in 40 mM Triton X-100). Increase in absorbance at 405 nm with time was monitored using Spectramax 190 plate reader (Molecular Devices, Sunnyvale, CA) while data was analyzed by using SoftMaxPro 4.7.1 software provided with the instrument. The ratio of activity of each clone after incubation at higher temperature versus without incubation was taken as residual activity and was used to identify the positives.

Positives were further confirmed by performing tube assays. Ten milliliters of cell culture in LB medium, having 100 µg/ml ampicillin, was inoculated with 1% of overnight grown culture, followed by incubation at 37 °C with vigorous shaking (250 rpm). After 6 hours, culture was induced with 0.5 mM IPTG and incubated further for 12 hours after which cells were separated from culture supernatant by centrifugation at 18,000 g for 10 minutes at 4 °C. Culture supernatant was diluted 1:1 with 100 mM sodium phosphate buffer (pH 7.2) and incubated (100 µl) at various temperatures for 20 minutes followed by cooling at 4 °C for 20 minutes and final equilibration at 25 °C for 15 minutes in a PCR thermal cycler. Activity was measured by adding 80 µl of diluted supernatant to 920 µl of 50 mM sodium phosphate buffer (pH 7.2) containing 2 mM PNPB micellized in 20 mM Triton X-100 as substrate. Increase in absorbance at 405 nm with time was measured with a spectrophotometer (U-2000, Hitachi, Japan). Residual activity corresponding to each temperature of incubation was determined by taking the ratio of activities obtained upon incubation with that of without incubation. The temperature at which enzyme loses 50% of its activity after incubation for 20 minutes (T_{50}) was taken as the thermostability index for the selection of mutants. Plasmids were purified from mutants displaying higher T_{50} values with respect to parent and genes were sequenced to identify the mutation occurred.

4.2.5. Recombination

Mutations obtained by site-saturation mutagenesis at four positions, 134, 137, 158 and 163 were recombined by site-directed mutagenesis. Briefly, mutant 6-A was made by recombining mutations M134E and M137P using mutagenic primers DMF (5'-AGTGACGATGAGATTGTCCCGAATTA^{*CT*}ACTTATC-3') and DMR (5'-GATAAGTAATTCGGGACAATCTCATCGTCA^{*CT*}ACT-3') (mismatch codons of positions 134 and 137 are italicized). Using pET22-4D3 as template, forward and reverse mutagenic primers were used along with vector specific primers, T7-terminator and T7-promoter, respectively, to amplify the two fragments by PCR encoding the first and second half of structural gene with desired mutations. In second round of PCR, these two fragments were

mixed, extended and amplified to get full-length gene. Mutant 6-C, having three mutations recombined M134E, M137P and G158D, was made in a similar manner as mutant 6-A, except that the structural gene of mutant 5-C in pET22b (having mutation G158D in the background of 4D3) was used as template. Mutant 6-B, having mutations M134E, M137P and S163P, was generated by introducing mutation S163P, using previously amplified full length gene of mutant 6-A as template and mutagenic primers S163PF (5'-CTTCTGTACAGCCCGCAAGTCTACAGC-3') and S163PR (5'-GCTGTAGACTTGCGGGCTGTACAGAAG-3') (mismatch codon italicized). Two rounds of PCR was performed yielding first and second half of structural gene having desired mutation, which were mixed together, extended and amplified to get full length structural gene. Finally, mutant 6-D having all the four mutations M134E, M137P, G158D and S163P, was made by introducing S163P mutation in a similar manner as mutant 6-B, except that the full length structural gene of mutant 6-C was used as template. The full length amplified products were digested with *Nco*I and *Hind*III, purified and ligated with similarly digested pET22b to get mutant 6-A, having two mutations, 6-B and 6-C, having three mutations and 6-D, which has all the four mutations recombined in 4D3 background.

4.2.6. DNA sequencing

Plasmid DNA isolation from *E. coli* DH5 α or BL21 (DE3), and sequencing of lipase gene was performed as described earlier in section 2.2.8, Chapter 2. Two vector specific primers, T7-promoter and T7-terminator along with four gene specific primers as listed in Table 2.1 (Chapter 2) were used for sequencing lipase gene.

4.2.7. Protein purification

Lipase mutant proteins, genes of which cloned in pET21b, were purified upon expression in *E. coli* BL21 (DE3) in a similar manner as described in chapter 3 with some modifications. All mutant proteins bind to phenyl sepharose and were eluted under conditions similar to that of 4D3 mutant. In the subsequent ion-exchange chromatography step, all the mutants bind to anion exchanger Mono-Q (Amersham Biosciences) in 20 mM glycine-NaOH buffer (pH 10.5). Elution of bound protein was performed by using a linear gradient of NaCl, from 0 to 1 M, in binding buffer. Active fractions were pooled, dialyzed overnight against 2 mM glycine-NaOH buffer (pH 10) at 4 °C and stored at -20 °C. Purity

was checked by running SDS-PAGE and was found to be more than 95 percent. Protein quantitation was done by the modified Lowry method (Markwell et al. 1981).

4.2.8. Circular dichroism

Circular dichroism spectral measurements, both in Far and Near-UV range were performed as described earlier in section 3.2.8 of chapter 3.

4.2.9. Fluorescence

Fluorescence measurements, corresponding to intrinsic tryptophan fluorescence and bis-ANS binding were performed as described earlier in section 3.2.9, Chapter 3.

4.2.10. Thermal inactivation and unfolding

Thermal inactivation of lipase mutants was performed in a similar manner as described in section 3.2.10 (Chapter 3). Thermal inactivation profiles were generated by incubating the enzymes at various temperatures (from 60 to 100°C) for 20 minutes followed by residual activity measurement at room temperature. Kinetics of thermal inactivation was monitored by incubating the proteins at 75 °C as described earlier (section 3.2.10; Chapter 3).

Thermal unfolding of lipase mutants was monitored by circular dichroism spectroscopy in a JASCO J-815 spectropolarimeter fitted with Jasco Peltier-type temperature controller (CDF-426S/15). The protein concentration used was 0.05 mg/ml in 50 mM sodium phosphate buffer (pH 7.2) with path length of 1 cm. Temperature dependent unfolding profiles were obtained by heating protein at a constant rate of 1 °C per minute from 25 to 95 °C and measuring the change in ellipticity at 222 nm.

4.2.11. GdmCl unfolding

Equilibrium unfolding in the presence of GdmCl was monitored by using CD and Fluorescence spectroscopy. Protein (0.5 mg/ml in 50 mM sodium phosphate buffer, pH 7.2), was incubated in varying concentrations of GdmCl for 12 hours at room temperature prior to measurements. Far-UV CD and fluorescence measurements were performed as described earlier (section 3.2.13; Chapter 3). Unfolding profiles were determined by monitoring the change in ellipticity at 222 nm and shift in tryptophan emission maxima (λ_{\max}) with increasing concentration of GdmCl. The concentration of stock GdmCl

solution was determined by refractive index measurements. Data were analyzed as described by Pace et al. (Pace et al. 1989).

4.2.12. Activity measurements

All activity measurements including determination of enzyme kinetic parameters, specific activity measurements of lipase mutants at room temperature (25°C) and at elevated temperatures were performed as described earlier in section 3.2.14, Chapter 3.

4.3. Results

4.3.1. Periplasmic expression of lipase gene in *E. coli*.

In order to develop an efficient screening system, to screen for thermostable mutants of *Bacillus subtilis* lipase, the lipase structural gene was cloned downstream to a periplasmic targeting sequence for expression in periplasm of *E. coli*. The structural gene of lipase mutant was cloned in frame to pelB signal sequence in expression vector pET22b (Figure 4.1). The lipase gene was under the control of T7 promoter the expression of which can be induced in the presence of IPTG, when expressed in a suitable host such as BL21 (DE3). This allows over expression and targeting of the protein to periplasm. Cytoplasmic expression of lipase is toxic to *E. coli* (Dartois et al. 1992; Dartois et al. 1994), which gets minimized by targeting the protein to periplasm. Moreover, prolonged incubation of host after induction (more than 12 hours) causes leakage of the lipase from periplasm into the culture medium. This might be due to the associated toxicity of the protein, being lipase it may act on the outer membrane damaging it during prolonged incubation thus causing leakage into the culture medium or through an unknown mechanism (pET System Manual, 11th ed. Novagen). However, leakage offers an additional advantage as sufficient amount of enzyme got leaked out into the medium and activity measurements can be performed using culture supernatant. The enzymatic activity, when measured using culture supernatant harvested 12 hours after induction, was found comparable to the activity present in the late exponential phase culture supernatant of lipase over expressing *Bacillus subtilis* strain BCL1051 which actively secretes out protein into the medium (Chapter 2) (Dartois et al. 1994).

4.3.2. Generation of thermostable mutants

Thermostable mutants were generated by performing site-saturation mutagenesis at four positions 134, 137, 158 and 163, identified earlier by random mutagenesis and screening (Chapter 3). The criteria for selecting a mutation depended on its contribution to thermostability and retention of activity at moderate temperatures comparable to that of the parent. Only those mutants that showed high levels of residual activity upon exposure to high temperature without affecting its catalytic activity at ambient temperatures were selected. The lineage of mutants and the mutations incorporated are depicted in the flowchart (Figure 4.3).

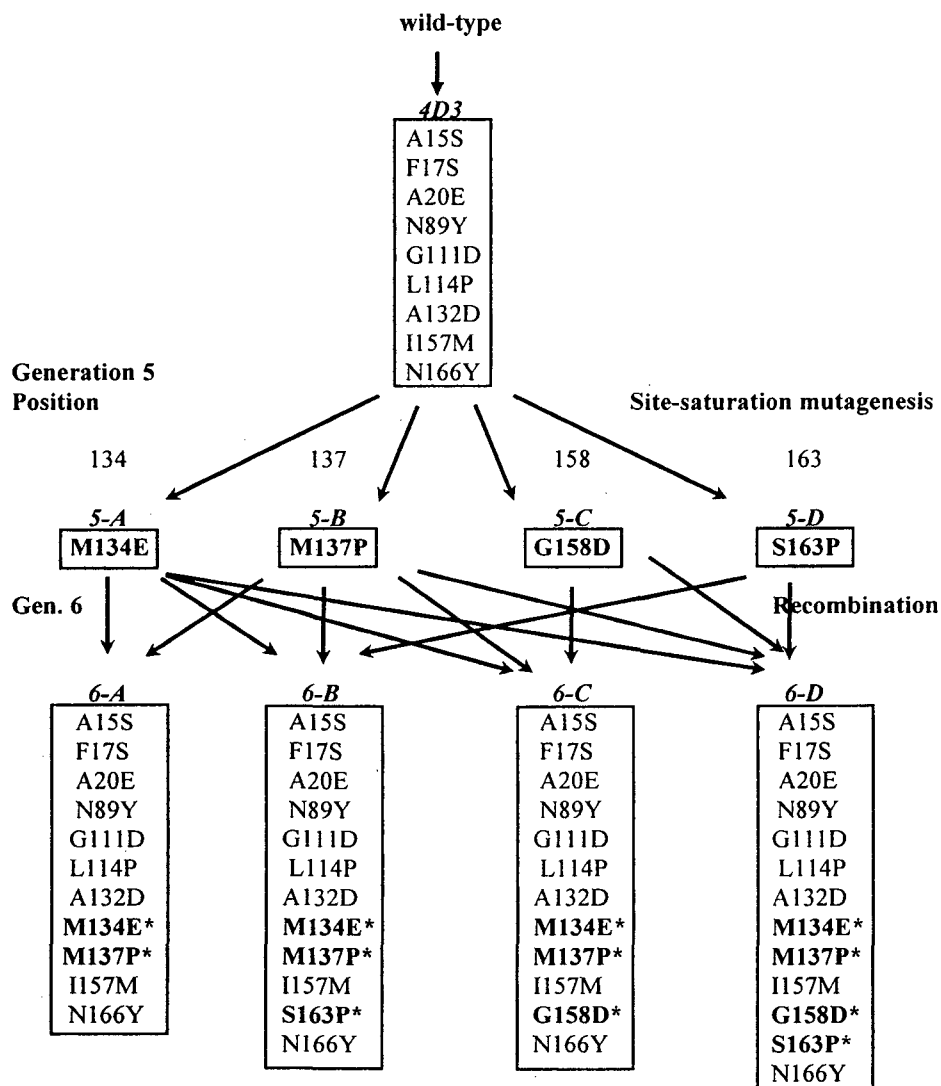


Figure 4.3 Lineage of thermostable variants of lipase. Amino acid substitutions accumulated during one round of site-saturation mutagenesis at four positions (134, 137, 158 and 163), followed by recombination are shown. Newly introduced mutations are marked with asterisk.

Thermostable mutant 4D3, which has nine mutations as compared to wild-type, was used as parent for mutagenesis (Chapter 3). The mutant population was generated by overlap-extension PCR using mutagenic primers encoding target position by degenerate codon NNK. The mutagenic library so obtained was used for periplasmic expression of protein in *E. coli*. Screening was performed using high-throughput assay format in 96-well plates, upon thermal challenge at elevated temperature. Approximately 2000 clones, 500 clones for each position, were screened for thermostability after exposure to 70 °C for 20 minutes, a temperature at which parent 4D3 shows less than 20 % residual activity. After screening, clones which were displaying significantly higher residual activities compared to parent (>50 %) were further tested in tube assays. Thermal inactivation profile of mutants was made by monitoring residual activities in culture supernatant upon incubation over a range of temperatures. The temperature at which 50 % of activity was lost (T_{50}) was used as selection criteria. Four mutants, two from each group, displaying highest and second highest increase in T_{50} were subjected to sequencing to determine the mutation occurred.

At position 134, substitution of methionine (ATG) with glutamate (GAG) increases T_{50} by 5.8 degrees (Table 4.2). The second best substitution at this position was aspartate (GAT), with an increment of 4.8 degrees. At position 137, replacement of methionine (ATG) with proline (CCG/CCT) increases T_{50} by 6.8 degrees while the second best substitution was aspartate (GAT), which shows increment of 3.6 degrees. At position 158, only two kinds of substitutions, both showing equivalent improvement in thermostability, were observed. Substitution of glycine (GGC) at position 158 by either aspartate (GAT) or glutamate (GAG) increases T_{50} by 1.6 degrees. Finally, at position 163, substitution of serine (AGC) with proline (CCG/CCT) increases T_{50} by 1.8 degrees followed by aspartate (GAT) which shows marginal improvement of 0.5 degrees. Thus the best substitutions identified at these four positions are M134E, M137P, G158D and S163P. The mutations identified at these four locations by error-prone PCR were M134T, M137T, G158D and S163C showing modest increase in thermostability, which when optimized by site-saturation mutagenesis improves thermostability to a significant extent. It is important to note that except for mutation G158D, none of the other three substitutions (M134E, M137P and S163) could be achieved by point mutagenesis as it requires change of more than one nucleotide of the codon (Table 4.2). All the mutants identified by site-saturation mutagenesis of 4D3 belong to fifth generation of evolution of lipase for thermostability.

Table 4.2 List of thermostable mutants got after site-saturation mutagenesis at four positions of 4D3.

Mutant	Position	Codon change	A.A. Change	EPCR/SSM ^a	T ₅₀ ^b	ΔT ₅₀ ^c
4D3	—	—	—	—	67.2	—
134-2B7	134	ATG→GAG	M134→E	SSM	73.0	5.8
134-3A3	134	ATG→GAG	M134→E	SSM	73.0	5.8
134-2D12	134	ATG→GAT	M134→D	SSM	72.0	4.8
134-4E11	134	ATG→GAT	M134→D	SSM	71.8	4.6
137-3C7	137	ATG→CCG	M137→P	SSM	74.0	6.8
137-4A4	137	ATG→CCT	M137→P	SSM	74.0	6.8
137-2D5	137	ATG→GAT	M134→D	SSM	70.8	3.6
137-3B2	137	ATG→GAT	M134→D	SSM	7.4	3.2
158-1B2	158	GGC→GAT	G158→D	EPCR/SSM	69.0	1.8
158-3F6	158	GGC→GAC	G158→D	EPCR/SSM	68.8	1.6
158-2D12	158	GGC→GAG	G158→E	SSM	68.8	1.6
158-3C3	158	GGC→GAG	G158→E	SSM	68.5	1.3
163-2D3	163	AGC→CCG	S163→P	SSM	69.0	1.8
163-4G2	163	AGC→CCT	S163→P	SSM	69.0	1.8
163-1G8	163	AGC→GAT	S163→D	SSM	67.5	0.3
163-4G7	163	AGC→GAT	S163→D	SSM	67.8	0.6

^a Possibility of amino acid (A.A.) substitution obtained either by error-prone PCR (EPCR) or site-saturation mutagenesis (SSM).

^b Temperature at which enzyme loses half of its initial activity, upon 20 minutes of incubation (T₅₀), measured by using crude culture supernatant.

^c Difference of T₅₀ value of mutant and that of parent 4D3.

In order to explore whether recombination of these three mutations further improves thermostability and how the mutations behave in improving thermostability, mutations M134E, M137P, G158D and S163P were recombined in 4D3 background, using site-directed mutagenesis (Figure 4.3). Four mutants, having above mutations in different combinations were made. Mutant 6-A, has two mutations M134E and M137P recombined. Mutant 6-B was made by introducing third mutation, S163P, into 6-A, thus recombining three mutations M134E, M137P, and S163P. Mutant 6-C was made by adding mutation G158D to 6-A, thus recombining M134E, M137P and G158D. Finally mutant 6-D, which has all the four mutations combined, M134E, M137P, G158D and S163P, was made by introducing mutation S163P into mutant 6-C. All these four mutations of the fifth generation, when recombined give a mutant (6-D), with thirteen mutations as compared to wild-type protein.

4.3.3 Characterization of lipase mutants

All the lipase mutants including wild-type and parent 4D3 were purified by standard protocols. The purity was checked by SDS-PAGE and was found to be more than 95% after staining with coomassie brilliant blue R-250 (data not shown). Specific activities of all mutants were determined at room temperature (25 °C) using PNPB as substrate (Figure 4.4). Specific activities of all the mutants including parent 4D3 was found to

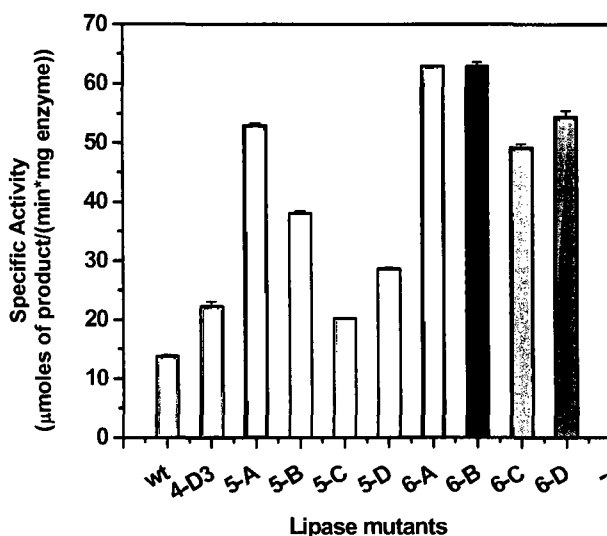


Figure 4.4 Specific activities of lipase mutants. Measurements were performed at 25 °C in 50 mM sodium phosphate buffer (pH 7.2), using 2 mM p-nitrophenyl butyrate (PNPB) micellized in 20 mM Triton-X 100 as substrate.

be superior than wild-type. Individual mutations at four positions had varied effect on the specific activity. The most profound effect was shown by the introduction of M134E, as mutant 5-A shows ~2.5 fold increase in specific activity compared to parent 4D3. Introduction of M137P in 6-B also increases specific activity by 1.8 fold. While mutation G158D in 6-C did not effect specific activity, introduction of mutation S163P in 6-D shows marginal increment. Interestingly, these mutations show additive behavior in increasing specific activity when combined together. Mutant 6-A, which has M134E and M137P recombined, and mutant 6-B, which has S163P in addition to the previous two, shows equivalent and highest increase in specific activities which was about 3 fold higher than parent 4D3 and more than 5 fold higher than wild-type. This shows the positive contribution of these three mutations in increasing specific activity of the enzyme. However, introduction of mutation G158D, either in the background of 6-A or 6-B, has a negative effect as shown by reduced specific activities of 6-C and 6-D, but still it was more than two fold higher than that of 4D3.

Spectroscopic characterization of lipase mutant was performed to probe any variation in the native state (Figure 4.5). Far-UV CD spectra of all the mutants completely overlap each other, indicating invariant secondary structure (Figure 4.5 (a)). Tertiary structure of all the mutants, as probed by near-UV CD, intrinsic tryptophan fluorescence and binding of hydrophobic dye bis-ANS was also found to be similar with little

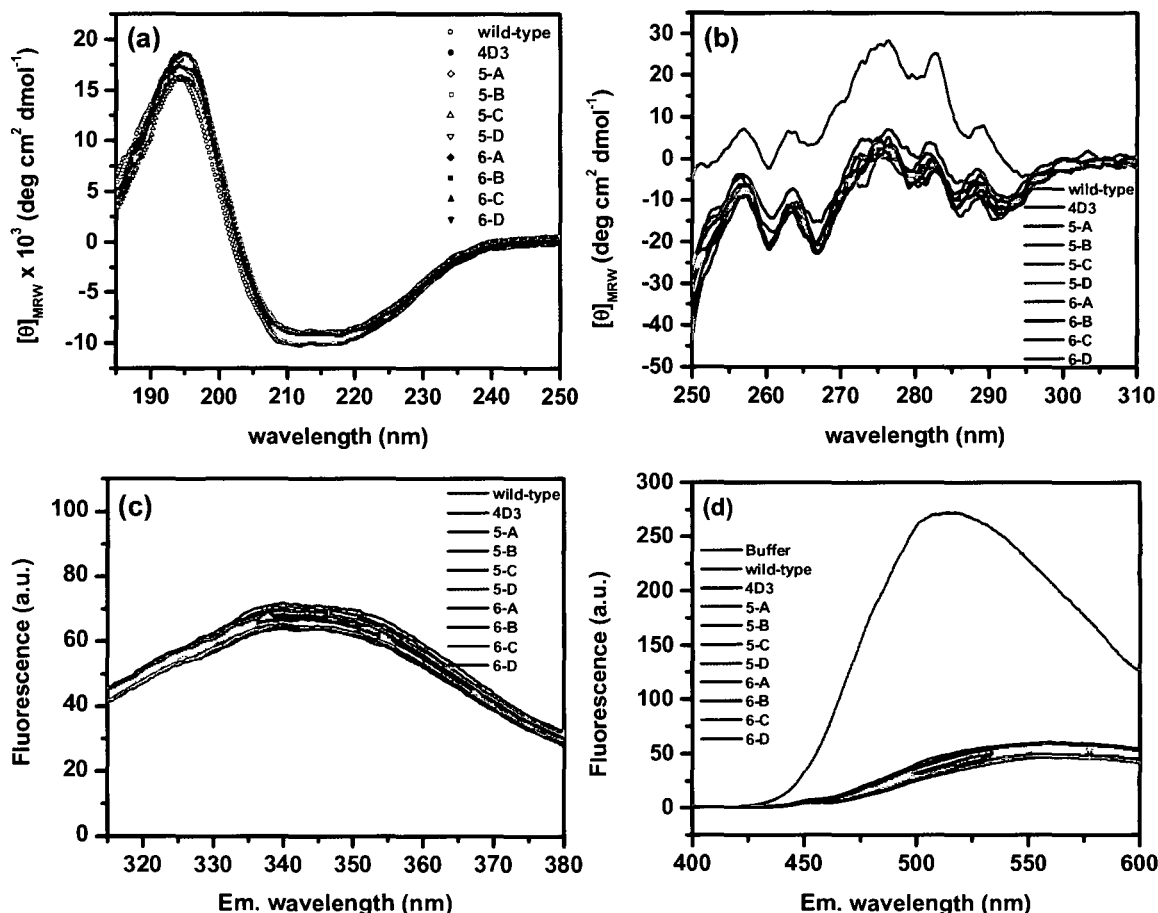


Figure 4.5 Spectral profiles of lipase mutants. (a) Far-UV CD spectra, (b) Near-UV CD spectra, (c) Intrinsic tryptophan fluorescence spectra and (d) bis-ANS binding spectra of lipase mutants. All spectra were recorded in 50 mM sodium phosphate buffer at 25 °C.

variations. Near-UV CD spectra of all mutants completely overlap except that of wild-type protein (Figure 4.5 (b)). The characteristic peaks corresponding to that of aromatic amino acid residues were similar in case of wild-type and mutants except that ellipticity was higher in case of wild-type protein in the entire range of 250-295 nm. This might be due to the propensity of the wild-type protein which may undergo limited aggregation even at room temperature, (data not shown) (Acharya and Rao 2003), which may cause wild-type protein to make soluble oligomers at room temperature, thus causing alteration in the chiral environment of the chromophores, thus causing shift in the entire spectrum. The fluorescence emission spectra arising out of two tryptophans, (W31 and W42), overlaps with subtle variation in intensity in case of mutants including wild-type (Figure 4.5 (c)). The emission maximum (λ_{max}) was found to be around 339 nm, indicating that tryptophans are partially buried. In the native state, tryptophan fluorescence is quenched

due to close proximity of His 3 to Trp 31 (Acharya and Rao 2003). Subtle variation in fluorescence intensities among mutants may have arisen due to subtle change in the vicinity of Trp 31. Tertiary structure of the lipase mutants was further probed by measuring their affinity to the hydrophobic dye bis-ANS. Bis-ANS is a valuable probe to investigate surface hydrophobicity of proteins. Upon binding to hydrophobic pockets, the fluorescence of bis-ANS increases significantly, which can be used to titrate the surface hydrophobicity of the proteins (Sharma et al. 1998). All mutants including parent 4D3 shows little binding of bis-ANS in contrast to wild-type which shows significant binding to the dye (Figure 4.5 (d)). This suggests that wild-type protein has significantly higher exposed hydrophobic surface area as compared to mutants. Due to absence of lid, *Bacillus subtilis* lipase has exposed active site which is hydrophobic in nature. This displays a large patch of hydrophobic residues of the active site to the solvent which are responsible for the binding of bis-ANS. It is important to note that many mutations introduced during the course of evolution of this protein for thermostability occurred in the vicinity of the active site and most of them are hydrophobic to polar/charged ones (Chapter 3). All the four positions reported here lies in the active site region with two mutations out of four resulted in the introduction of native charge. This might have reduced the hydrophobicity of the region, thus explaining reduced binding of mutants to hydrophobicity sensitive dye bis-ANS.

4.3.4. Thermal inactivation

To determine the resistance against irreversible thermal inactivation of lipase mutants, enzymes were heated at different temperatures for 20 minutes followed by cooling to ambient temperatures and residual activity measurements at room temperature. The thermal inactivation profiles of the mutants are shown in Figure 4.6. The inactivation transition of all the mutants are not similar, rather each shows a peculiar profile. Parent 4D3 inactivates at temperatures beyond 66 °C with a sharp transition which gets over by 75 °C. However, upon incubation at temperatures higher than 75 °C, residual activity increases, reaches a maximum by 90 - 92 °C and declines with further increase in temperature. Mutants 5-A, 5-C and 5-D, first show a decrease in residual activity in temperature range of 68 - 75 °C but increases with further increase in temperature till 85 °C. On the other hand, mutant 5-B and all the mutants of sixth generation (6-A, 6-B, 6-C and 6-D) show a gradual decrease in residual activity all along the temperature range of 75

to 100 °C. Inactivation profiles of all the mutants, except the parent 4D3, merge and follow the same trend beyond 85 °C. The temperature at which enzyme loses half of its activity (T_{50}) increases in the following order: 4D3 < (5-C, 5-D) < (5-A, 5-B, 6-A, 6-B, 6-C and 6-D) (Figure 4.6, Table 4.3). While 4D3 shows T_{50} around 68 °C, the same is 72 °C for 5-C and 5-D, which is 4 degree higher than parent. The half inactivation temperature of all other mutants, i.e. 5-A, 5-B, 6-A, 6-B, 6-C and 6-D is 93 °C, which is 25 degrees higher than parent 4D3. It is interesting to note that mutant 5-A and 5-B, which individually has only one additional mutation compared to parent 4D3, increases T_{50} to such an enormous extent. Moreover, combining these mutations in sixth generation does not increase T_{50} further, even marginally.

Kinetics of irreversible thermal inactivation was monitored by incubating the enzymes at elevated temperature followed by residual activity measurements at room temperature with respect to the time of incubation (Figure 4.7). All thermal inactivation traces were found to follow first-order kinetics. Half-life of inactivation of mutants ($t_{(1/2)}$), monitored at 75 °C are shown in Table 4.3. At 75 °C, the half-life of parent 4D3 was 4.4 minutes while that of

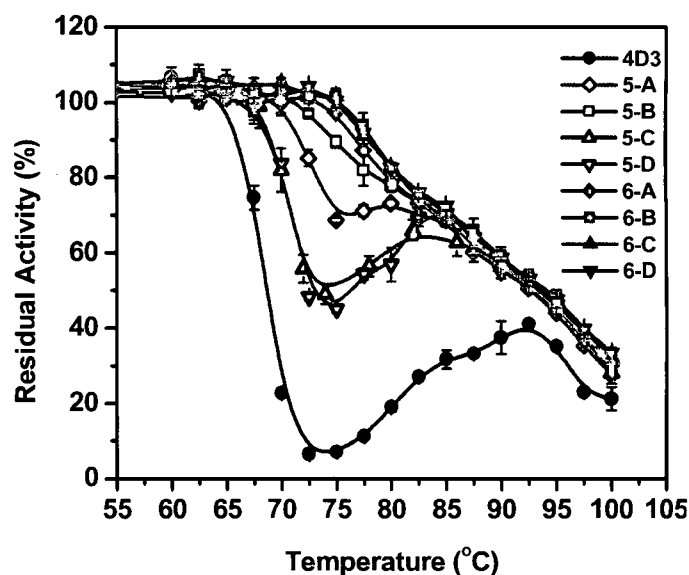


Figure 4.6 Thermal inactivation profiles of lipase variants. Enzymes (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2), were incubated at various temperatures for 20 minutes and assayed for residual activity at 25 °C.

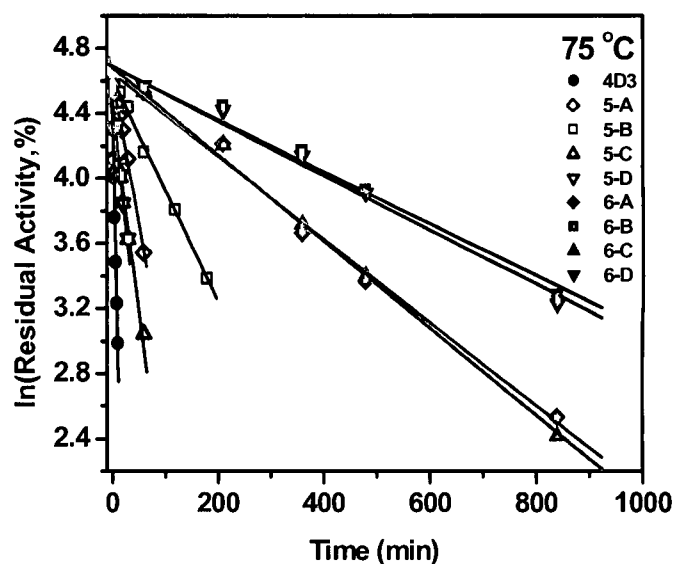


Figure 4.7 Thermal inactivation kinetics of lipase mutants. Time course of thermal stability was measured by calculating the residual activities at 25 °C after incubating the enzymes (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) at 75 °C.

mutants 5-A, 5-B, 5-C and 5-D, having one additional mutation than 4D3, was 38.8, 101.2, 28.5 and 22.5 minutes respectively. Combining individual mutations in sixth generation increases the half-life of inactivation further. Mutant 6-A, having two mutations M134E and M137P, has a half-life of 270.8 minutes which increases further upon adding the mutation S163P in 6-B to 430.5 minutes, which is 100 fold higher than parent 4D3 at the same temperature. Addition of mutation G158D in 6-A and 6-B background to get mutant 6-C and 6-D, does not increase half-life any further. This suggests that mutation G158D has little effect on thermal inactivation in the presence of other mutations. Unusual inactivation behavior of thermostable mutants, whopping increase in T_{50} values and non additive nature of individual thermostabilizing mutations suggests altered mechanism of inactivation.

4.3.5. Thermostability of mutants

The thermostability of lipase variants was tested by monitoring temperature induced unfolding using CD (Figure 4.8, Table 4.3). The apparent melting temperature ($T_{m,app}$) of mutants increases in the order of 4D3, 5-C < 5-D < 5-A < 5-B < (6-A, 6-C) < (6-B, 6-D). Parent 4D3 has an apparent melting temperature of 71.2 °C. Mutants 5-D and 5-A are marginally better with $T_{m,app}$ of 72.2 and 72.9 °C, respectively, which is 1 – 1.7 degree higher than that of 4D3. Mutant 5-B has $T_{m,app}$ of 74.1 °C which is 2.9 degrees higher than that of parent. However, mutant 5-C does not show any increase in $T_{m,app}$. Combining individual mutations in different order increases its $T_{m,app}$ further. Mutant 6-A, which has two mutations M134E and M137P recombined, shows $T_{m,app}$ of 76.9 °C, which is 5.7 degrees higher than that of parent 4D3 and ~ 21 degrees higher than wild-type protein. Incorporation of an

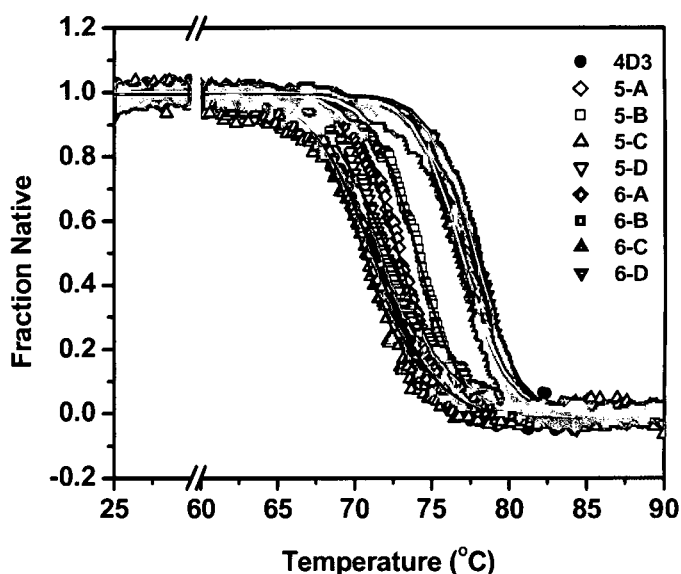


Figure 4.8 Thermal unfolding profiles of lipase mutants as monitored by circular dichroism. Purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was heated in cuvette of 1 cm path length at a constant rate of 1 °C/minute. Change in ellipticity at 222 nm (Θ_{222}) with temperature was used to calculate unfolding transitions.

additional mutation S163P in 6-B further increases $T_{m,app}$ to 77.8 °C, an increment of 1 degree. Addition of mutant G158D, either in 6-A or 6-B background does not effect $T_{m,app}$ as mutants 6-C and 6-D shows similar values as their progenitors 6-A and 6-B, respectively. This suggests that out of four mutations, only three mutations, M134E, M137P and S163P contributes in improving thermostability of the molecule and acts synergistically in improving the property while the fourth mutation G158D does not contribute in increasing thermostability.

4.3.6. Equilibrium unfolding of lipase mutants in GdmCl

The stability of lipase mutants was investigated by monitoring equilibrium unfolding in the presence of GdmCl by using fluorescence and CD spectroscopy. Since these mutants are very stable compared to wild-type protein, complete unfolding could not be achieved in urea, even at 9 M concentration. The shift in tryptophan fluorescence emission (λ_{max}) and ellipticity at 222 nm (Θ_{222}) were monitored with increasing concentration of denaturant (Figure 4.9). The unfolding transitions in GdmCl monitored either by fluorescence or far-UV CD were indistinguishable. Using the approach described by Pace et al. (Pace et al. 1989) data was normalized and mid point of unfolding transitions were determined (Table 4.3). The concentration of GdmCl required to cause 50% denaturation (U_{50}) of 4D3 was 2.90 M, which increases to 3.13 and 2.97 M in case of 5-B and 5-D mutants, respectively. Mutant 5-A and 5-C show lower conformational stability compared to parent 4D3, as shown by respective U_{50} values of 2.86 and 2.81 M. Upon combining these mutations in different order, conformational stability of the protein increases further. Mutant 6-A, which has

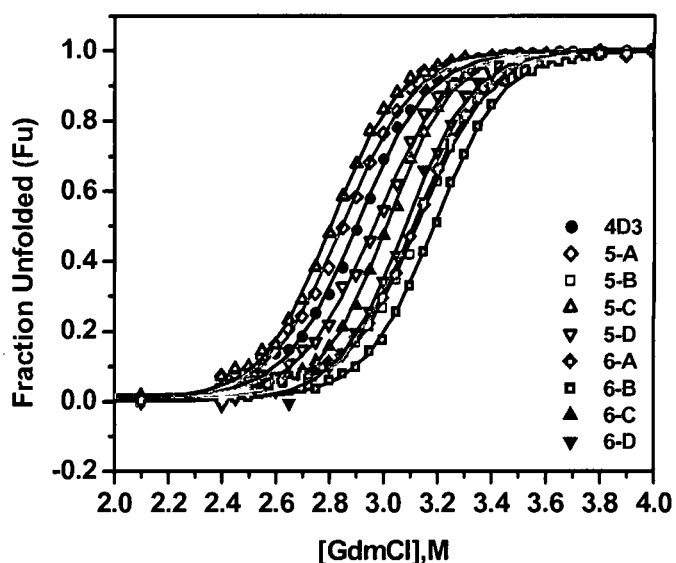


Figure 4.9 Equilibrium unfolding profiles of lipase mutants in the presence of GdmCl using CD. Fraction unfolded was calculated based on changes in ellipticity at 222 nm (Θ_{222}) with GdmCl concentration. (Identical transitions were obtained when fraction unfolded was calculated using shift in tryptophan fluorescence, λ_{max} , not shown for the sake of clarity).

two mutations, M134E and M137P combined, show U_{50} value of 3.11 M, which is equivalent to that of mutant 5-B having only mutation M137P. Upon adding mutation S163P with the two previous ones in mutant 6-B, U_{50} increases to 3.19 M, which is highest among displayed by all mutants. Incorporation of mutation G158D in the background of either 6-A or 6-B decreases the conformational stability of the protein as shown by the U_{50} values of mutant 6-C and 6-D at 3.01 and 3.08, respectively. These values are ~ 0.1 M less than that of their respective progenitors 6-A and 6-B. Mutant 6-B displays highest conformational stability, having U_{50} at 3.13 M, which is 0.3 M higher than parent 4D3 and 1.14 M higher than wild-type protein. The increment in U_{50} shown by fifth generation mutants, having individual mutations, and that by mutants of sixth generation, which has these mutations combined in different order, clearly shows that out of four mutations, only two mutations, M137P and S163P, contributes positively in increasing the conformational stability of the protein. The other two mutations, M134E and G158D decrease stability of the protein. Another significant observation is that all the four mutations, whether contributes positively or negatively, when combined together contributes in additive manner to the stability of the protein. The increment shown in U_{50} by mutants having recombined mutations is always equal to the sum of their individual contribution. Although two mutations out of four identified in the study are actually destabilizing in the presence of GdmCl, they show improvement in thermostability of the protein. This suggest different role played by each mutant in altering the mechanism involved in the stabilization of protein which needs to be investigated further.

4.3.7. Temperature dependence of the activity

The specific activity of wild-type, 4D3 and 6-B mutants, which differ significantly in thermostability, were determined over a range of temperature from 25 to 70 °C (Figure 4.10). The optimum temperature of activity (T_{opt}) for wild-type enzyme was 35 °C, which increases by 20 degrees to 55 °C in case of 4D3. In case of mutant 6-B, T_{opt} increases further to 65 °C, which is 10 degrees higher than parent 4D3 and 30 degrees higher than wild-type (Table 4.3). It is interesting to note that this increase in optimum temperature of activity in 6-B was accompanied by an increase in activity at all temperatures compared to wild-type and 4D3. The specific activity of 6-B was more than 5 and 3.5 folds higher as compared to wild-type and 4D3, respectively, at 25 °C. Moreover, the same was more than 12 and 6 fold higher at their respective temperature optima. This observation is in contrast

to many reports on thermophilic proteins, which show less activity at lower temperatures compared to their mesophilic counterparts (Sterner and Liebl 2001; Wintrode et al. 2001). This might be due to the conditions employed during screening, which picks only those mutants that gained enhanced thermostability without sacrificing activity at lower temperatures.

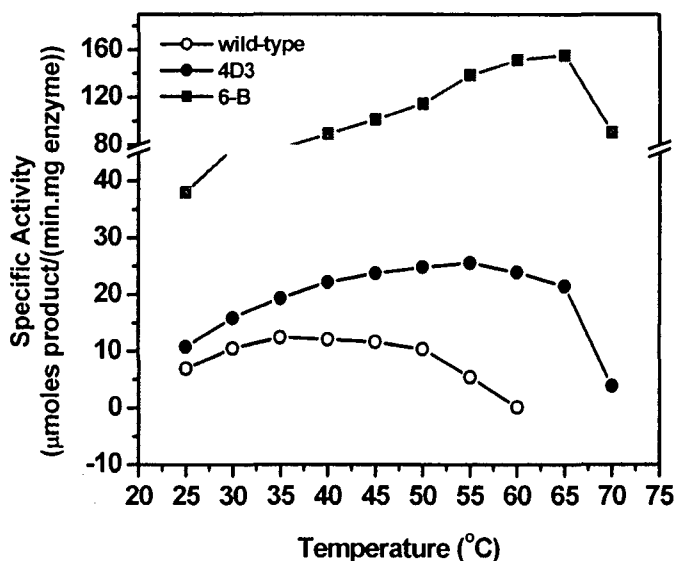


Figure 4.10 Temperature dependence of activities of wild-type lipase and its mutants. Enzymes (0.2-1 μg in 970 μl of 50 mM sodium phosphate buffer, pH 7.2) were incubated at various temperatures for 1 minute prior to activity measurement at the same temperature by the addition of substrate.

Table 4.3. Stability parameters of lipase mutants

Mutant	$t_{(1/2)}$ (min) ^a		$T_{(50)}$ ^b (°C)	$T_{m(app)}$ ^c (°C)	T_{opt} (°C)	U_{50} ^d (M)
	55 °C	75 °C				
wild-type	2.8	—	53.3	56.0	35.0	2.05
4D3	—	4.4	68.0	71.2	55.0	2.90
5-A	—	38.8	93.0	72.9	—	2.86
5-B	—	101.2	93.0	74.1	—	3.13
5-C	—	28.5	73.0	70.7	—	2.81
5-D	—	22.2	72.0	72.2	—	2.97
6-A	—	270.8	93.0	76.9	—	3.11
6-B	—	430.5	93.0	77.8	65.0	3.19
6-C	—	259.6	93.0	76.6	—	3.01
6-D	—	412.6	93.0	77.7	—	3.08

^a Half-life of thermal inactivation.

^b Temperature at which enzyme loses 50% activity upon incubation for 20 minutes.

^c Mid-point of thermal transition as calculated by CD.

^d Mid-point of unfolding transition in the presence of GdmCl.

4.3.8. Catalytic properties of lipase mutants

Using p-nitrophenyl acetate (PNPA) as substrate, kinetic parameters of lipase mutants were determined at 25 °C (Table 4.4). The kinetic parameters k_{cat} and K_m shows large variations in mutants as compared to parent 4D3 and wild-type enzyme. K_m of mutants are significantly lower, least for mutant 5-A at 0.28 mM, as compared to

wild-type enzyme which has around 1 mM. On the other hand k_{cat} values of all the mutants, except 5-C, are higher than wild-type. 5-D shows highest k_{cat} of 637 min^{-1} which is more than 2.5 fold higher than that of wild-type at 220 min^{-1} . The overall catalytic efficiency, as measured by k_{cat}/K_m , of mutants is 3 – 6 folds higher than wild-type enzyme. The mutants have achieved higher catalytic efficiencies in different ways. While mutant 5-B and 5-D display enhanced k_{cat} along with subtle decrease in K_m , 5-A, showing highest efficiency, has achieved it mainly due to a drastic reduction in K_m . The most thermostable mutant 6-B shows significant reduction in K_m along with improvement in k_{cat} , so that the overall catalytic efficiency (k_{cat}/K_m) has improved significantly. It is interesting to note that introduction of stabilizing mutations has enhanced the catalytic efficiencies of these enzymes along with increasing thermostability. This observation is likely due to the strategy employed during screening, which identifies clones that exhibit activities comparable or higher than that of parent enzyme at room temperature besides displaying high levels of residual activities after exposure to higher temperatures.

4.3.9. Structural basis of thermostability

All the four mutations identified in this study were modeled onto the structure of mutant 4D3 (PDB id: 3D2C, chain A) (Ahmad et al. 2008), in order to understand the structural basis of the observed thermostability (Figure 4.11). *Bacillus subtilis* lipase is a single domain protein with a minimal α/β hydrolase fold, having a six-stranded central β -sheet with helices and loops on either sides (van et al. 2001).

Table 4.4 Kinetic parameters of wild-type lipase and its mutants.

Mutant	K_m^a (mM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\text{mM}^{-1} \text{min}^{-1}$)
wild-type	0.98	220	224
4D3	0.79	290	370
5-A	0.28	380	1383
5-B	0.67	565	845
5-C	0.68	150	220
5-D	0.70	637	914
6-A	0.44	357	821
6-B	0.51	414	809
6-C	0.36	263	730
6-D	0.38	266	748

^a Kinetic parameters were calculated from assays conducted at 25 °C using PNPA as substrate.

Mutation M134E is located in a loop connecting $\beta 7$ strand and αE helix. Being hydrophobic and completely exposed to the solvent on the surface of the molecule, Met134 keeps the protein native structure under constant strain of entropically unfavourable polar-nonpolar interaction with solvent. Its conversion to charged Glu relieves this strain and establishes favourable interaction with polar solvent, thus stabilizing the native structure of protein. However, mutation M134E occurred adjacent to Asp133 and Asp132. This brings three negative charges in close vicinity, may cause repulsive strain

in the region, thus resulting in destabilization of the native state. It was indeed observed that mutation M134E, does not increase the conformational stability of the protein, as determined by equilibrium unfolding in GdmCl, but has a profound effect on the thermostability as determined by thermal unfolding and inactivation. The cluster of three consecutive negative charges might have blocked the aggregation of unfolded protein upon heating, thus increasing the half-life of inactivation and also shifting the equilibrium toward unfolded state, consequently increasing the apparent melting temperature.

The second mutation, M137P, is also located in the same loop in which mutation M134E is present. Met 137 is also present on the surface and side-chain is completely exposed to the solvent. Substitution of hydrophobic Met with Pro does not affect the polarity of the region but brings stability mainly due to the constraints imposed by the proline in the backbone conformation (Matthews et al. 1987; Hardy et al. 1993; Watanabe et al. 1994).

The third mutation, G158D, is located in a 3_{10} helix G5, and is completely surface exposed. Introduction of a charged residue on the surface may stabilize the protein by

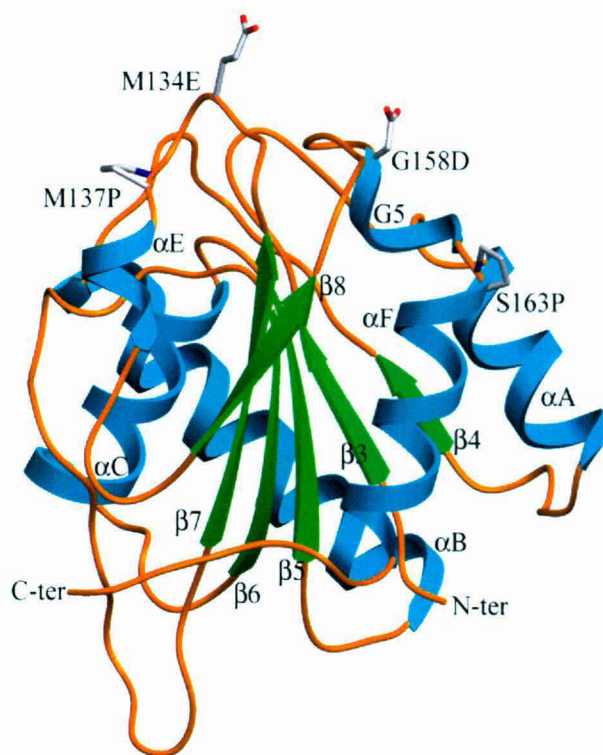


Figure 4.11 Ribbon diagram of 4D3 mutant lipase indicating the relative position of each mutation. Protein structure was made using the X-ray structure coordinates of mutant 4D3 (PDB id: 3D2C, chain A).

favourable polar interaction of the side chain with solvent. However, it was observed that this mutation actually destabilizes the protein as shown by equilibrium unfolding in GdmCl. This might be due to the replacement of glycine, which might be essential to maintain the flexibility of the region, which was compromised upon substitution with aspartate. Another reason for destabilization of the protein might be due to the close proximity (4.28 Å) of the side-chain carbonyl oxygen of Asp158 to that of carbonyl oxygen of main chain Gly155. Both being electronegative might be involved in a kind of electrostatic repulsion, thus keeping the region in strain. The effect of this mutation in increasing thermostability, as measured by thermal inactivation, might be due to the introduction of negative charge at the surface, which in turn causes reduced aggregation of protein upon thermal unfolding.

The fourth mutation, S163P, is located at the N-terminal of the helix α F, which is the last helix of the protein. Ser163 is the first residue of the helix and its side-chain is completely exposed to the solvent. Conversion of Ser to proline might have stabilized the protein by helix stabilization as proline is favored at the N-terminus of α -helix in several thermophilic proteins and also by decreasing the conformational entropy of the unfolded state (Matthews et al. 1987; Watanabe et al. 1996; Vieille, Zeikus 2001).

4.4. Discussion

Stability of protein is crucially dependent on the regions or positions prone to stress from which unfolding starts. Stabilization of such locations or regions by additional interactions may have a dramatic effect on the stability of the molecule (van den et al. 1998). However, prediction of these positions and the kind of interaction to be introduced is the most challenging task of any rational designing protocol (Eijsink et al. 2004). One approach relied on identifying the residues with high B-factor values, from crystallographic data, and mutate them to a substitute with lower B-factor value. This strategy was reported in one case (Reetz et al. 2006) but need to be proved in other cases also. On the other hand, directed or *in vitro* evolution has found phenomenal success in identifying and stabilization of weak spots in protein, increasing overall stability of the protein (Wintrode and Arnold 2000; Arnold et al. 2001; Eijsink et al. 2005). Use of random mutagenesis, mainly by error-prone PCR, followed by screening for stable variants is quite useful but suffers from the limitation that point mutation does not sample all possible substitutions at a give location. This limitation can be circumvented by the use

of site-saturation mutagenesis which screens all possible 19 amino acid substitution at a given location. By using this approach, we have further improved thermostability of a *Bacillus subtilis* lipase variant significantly. Four positions, 134, 137, 158 and 163, were identified as weak spots during an earlier round of random mutagenesis and screening. However, the mutations identified were showing only modest increase in thermostability (Chapter 3). Optimization of substitutions at just three locations increases the melting temperature by 7 degrees along with 25 degree increment in the half-inactivation temperature (T_{50}).

Examinations involving mutants having single mutation individually as compared to parent individually as well as mutant having all the mutations combined provides better insights into the role of each mutation have in improving the property (Yuen and Liu 2007). Mutation M137P has the most dramatic effect in improving thermostability followed by S163P, which show modest improvement in stability, as shown by increase in $T_{m,app}$ and U_{50} . On the other hand, mutation M134E and G158D does not increase protein stability, as shown equilibrium unfolding in the presence of GdmCl. It was observed that these mutations which in fact destabilize the protein, show positive effect on the thermostabilization of the protein as measured by thermal inactivation and unfolding. This positive effect is apparently due to the introduction of negative charge on the surface of the molecule. M134E not only replaces a solvent exposed hydrophobic residue with a charged one but also place a negatively charged residue adjacent to two negatively charged residues, Asp133 and Asp132. This generates a cluster of three negative charges, which is adjacent to the exposed hydrophobic region of active site, completely exposed to the solvent. This might have prevent the protein to aggregate upon thermal unfolding, thus having improved half-life of inactivation as well as increase in $T_{m,app}$, due to shifting of the equilibrium toward the unfolded state. Mutation G158D is in the middle of a stretch of hydrophobic amino acid residues $G_{153}VGHM(G158D)LL_{160}$, which might be involved in aggregation upon thermal denaturation. Introduction of a negative charge blocks the non specific hydrophobic interaction, thus reducing the rate of aggregation at elevated temperatures. It is interesting to observe that, although mutants were screened for improved thermostable variants, these mutations, which in fact decreases the conformational stability of the protein also got selected. This could be due to the conditions employed during screening. Direct measurement of enzymatic activity of large number of mutants is not possible at elevated temperatures in high throughput format,

although, this identifies mutants which retain their native folded structure at elevated temperature. On the other hand, a relatively simple method was employed by measuring residual activity of mutants, after exposure to elevated temperatures for a fixed period (Daniel and Danson 2001). However, this kind of screening can identify different class of mutants including those which maintains their structure at high temperatures (more thermostable; have higher melting temperature) as well as those which loose their structure at high temperature but refolds back to native state efficiently upon cooling to ambient temperatures (more thermotolerant). Mutation M137P and M163P belongs to the first category as both improves stability of the protein, thus allowing it to maintain its structure at higher temperature, while mutation M134E and G158D belongs to the second class, which do not increase the conformational stability of the protein, rather affects the irreversible step involved after unfolding (by reducing aggregation), thus allowing the unfolded protein to refold back efficiently upon cooling hence showing higher residual activities.

Combination of all the individual mutations in sixth generation shows additive nature of mutants in increasing stability of the protein as measured by equilibrium unfolding in GdmCl. The increment shown in mid-point of chemical unfolding transition (U_{50}) by mutants having recombined mutations is always equal to the sum of their individual contribution. This suggests that each mutation alter local environment therefore their effects on the structure were not synergistic but independent to each other.

It is important to note that all the four mutations, which have such a profound effect on the thermostability of protein, are in the vicinity of the active-site. Two mutants, M134E and M137P are located in same loop, while G158D is located in another loop, which are part of the binding region of active-site. This indicates that the region around the active site is prone to denaturation, from which unfolding starts. Stabilization of these loops has a significant effect on protein stability. It is important to note that out of four thermostabilizing mutations, only two mutations increase conformational stability of the protein. In both these mutations, stabilization was achieved by the introduction of proline residues. Proline residues are known to stabilize the folded state of proteins compared to the unfolded state due to the constraints they introduce in the backbone conformation, thus reducing the conformational entropy of the unfolded state (Watanabe et al. 1994; Sriprapundh et al. 2000). However, the other two mutations M134E and G158D seem to increase the thermostability of the protein in an indirect manner. Both the mutations

caused introduction of charged residue on the surface, which itself is a stabilizing change, but it appears that they bring about their stabilizing effect by reducing protein aggregation upon heating. Also, in contrast to other thermostable mutants which are described in chapter 3, thermal inactivation profiles of mutants reported here does not follow their thermal unfolding profiles. While all mutants show one step transition which gets over by 80 °C, inactivation profiles of mutants is not single step. In case of few mutants, it shows a sharp decrease which matches with protein unfolding, but shows increase with further increase in temperature, while in case of other mutants, it didn't matches with unfolding transition at all. Two out of four mutations display a positive effect on protein thermostability as measured by thermal inactivation, but found to be actually destabilizing in the presence of chemical denaturant GdmCl. This suggests that these mutations affect the mechanism involved in thermal unfolding and inactivation, which needs to be investigated further. All these discrepancies are explored in detail, which forms the matter of Chapter 5.

Another significant observation is the coevolution of thermostability with activity. During the course of evolution, mutants were selected for improved thermostability without sacrificing their activities at ambient temperatures. It is interesting to observe that the catalytic efficiencies (k_{cat}/K_m) of all the mutants have increased by at least 3.5 fold at room temperature along with the increase in thermostability. The most thermostable mutant 6-B, which melts at a temperature 7 degrees higher and has 10 degree increment in T_{opt} than parent 4D3, shows at least 3 and 6 fold higher specific activity, as compared to parent at 25 °C and their respective T_{opt} . Although screening for mutants that perform catalysis at higher temperature was not performed, the screening criteria employed identified mutants which are thermoactive. Measurements of protein flexibility by hydrogen-deuterium exchange and molecular dynamic simulation studies on thermophilic proteins and their mesophilic counterparts suggest that in thermophilic proteins, the backbone dynamics are reduced as compared to mesophilic proteins (Jaenicke 2000a; Vieille and Zeikus 2001). In addition, the resultant conformational rigidity decreased the activity in thermostable proteins at lower temperatures. Contrary to this, the mutants reported here resulted in improvement in thermostability along with simultaneous improvement in activity over a range of temperature. This observation is in agreement with other reports based on *in vitro* evolution of thermostability, which also suggest that

activity at ambient temperatures and thermostability are not inversely correlated and both can be improved simultaneously (Wintrode and Arnold 2000).

To conclude, *in vitro* evolution of a thermostable variant of *Bacillus subtilis* lipase was performed for further improving its thermostability, without sacrificing its activity at lower temperatures. By performing site-saturation mutagenesis at four key positions, 7 degree enhancement in melting temperature was obtained which was accompanied by enhanced catalytic efficiency at ambient temperatures and improvement in optimum temperature of activity. Certain discrepancies were observed in the mechanism of stabilization which are explored further in the next chapter.

Thermal unfolding and refolding of
in vitro evolved lipase variants

Chapter 5

5.1 Introduction

How do proteins cope with extreme denaturing conditions such as heat? Answer to this may lie in the thermodynamic stability of the protein, which is the measure of the difference between the free energies of unfolded and folded states of the protein. Greater this difference, more stable protein will be against denaturing conditions. However, this holds true for proteins undergoing fast, cooperative and completely reversible unfolding process, as shown by many small proteins (Sterner and Liebl 2001). On the other hand, it may also lie in the kinetic stability, as shown by many proteins undergoing slow or irreversible unfolding. In these cases, it is the difference between the free energies of the native protein and the transition state involved for the first committed step in the unfolding pathway. Larger the difference, more slowly the protein/enzyme unfolds or inactivates. Another important feature is the reversibility involved in the process of unfolding. A protein undergoing reversible unfolding may not show high thermodynamic stability, thus losing its structure upon heating but quickly refolds back to its native form upon cooling, thus regaining back structure as well as activity (in case of enzymes). On the other hand, another protein showing high conformational stability, undergoes irreversible unfolding, loses its structure at elevated temperatures than the former, upon unfolding but fails to refold back upon cooling. While former is the example of a protein which is able to tolerate high temperature, despite losing its structure, thus more “thermotolerant”, the latter is more “thermostable” as it resists loss of its structure at higher temperatures.

Reversibility of unfolding of protein upon thermal denaturation depends on a number of processes occurring along with or after unfolding. These reactions involve aggregation of partially unfolded species, which accumulate during the course of unfolding, chemical modifications of side chains of amino acid residues, such as deamidation of asparagines and glutamine residues, oxidation of methionine, histidine, tyrosine, tryptophan and cysteines residues as well as a number of other covalent modifications to the polypeptide chain that occur at elevated temperatures (> 80 °C) (Daniel et al. 1996). While in case of many proteins these reactions are imperceptible making the whole process of unfolding completely reversible, they are significant in case of large number of proteins, thus making the whole process irreversible (Tomazic and Klibanov 1988; Vieille and Zeikus 2001). In majority of cases, aggregation plays a major role in imparting irreversibility to the process of unfolding, other reactions involving

covalent modification of the protein are also involved. Besides occurring during or after unfolding, aggregation may occur even at physiological conditions at a temperature well below than that required for the melting of protein, a condition at which the native state is highly thermodynamically favoured (Chi et al. 2003). Stabilization of native state of protein may increase its conformational stability, allowing the protein to resist unfolding at higher temperatures, but this may have little effect on the reversibility of the unfolding process.

Since aggregation involves interaction between partly unfolded species accumulated during unfolding or the denatured state, prediction of these changes in native state, which abolish these interactions, is a difficult task (Shortle 1996). However, several studies based on *in vitro* evolution have shown improvement in thermostability of the protein accompanied by improved reversibility of unfolding (Gray et al. 2001; Zhang et al. 2003). On the other hand there are several reports in which proteins are evolved for improved resistance to aggregation (Famm and Winter 2006; Famm et al. 2008). The commonly used screen employed in these *in vitro* evolution protocols involves exposure of mutant population of protein at elevated temperature for a fixed period of time, followed by cooling to ambient temperatures and assessing residual activities or other properties associated with native state at lower temperatures. Since this involves heating and cooling of protein, thus allowing unfolding and refolding, this kind of screening can pick mutants which retain their structure at higher temperatures, thus more stable, along with mutants which might not be stable but have acquired better refolding efficiencies due to reduction in reactions involved in irreversible steps, such as aggregation and covalent modification of protein at elevated temperatures.

Bacillus subtilis lipase is a small monomeric protein of 19.3 kDa, which shows irreversible thermal unfolding at physiological pH, primarily due to aggregation (Acharya et al. 2004). However, at low pH (below 5), enzyme shows remarkable recovery of activity, even upon prolonged incubations at temperatures, well above its melting point (Rajakumara et al. 2007). This phenomenal recovery of enzymatic activity is mainly attributed to drastic reduction in aggregation behavior of protein upon unfolding at low pH. Being highly basic (pI = 9.9), enzyme has excessive positive charge at low pH, mainly due to the protonation of histidine residues, thus preventing aggregation of protein upon heat denaturation by intermolecular electrostatic repulsion. Since, aggregation is primarily

responsible for the loss of activity at neutral pH; abolishment of aggregation at low pH improves the reversibility of thermal unfolding.

In an attempt to improve thermostability of *B. subtilis* lipase at neutral pH by *in vitro* evolution, two rounds of random mutagenesis and screening were performed. This identifies nine mutations, which have profound effect on the thermostability of the protein, as the mutant having all the nine mutations combined show apparent melting temperature 15 degree higher and more than million fold reduction in the rate of inactivation, compared to wild-type protein. However, none of these mutations have any effect on the reversibility of unfolding as all mutants showed irreversible thermal unfolding due to aggregation (Chapter 3). Further optimization of four positions, identified during previous rounds of random mutagenesis, by site-saturation mutagenesis, identified four mutations which increased thermostability of the protein further (Chapter 4). Examination of mutants having individual mutation as well as in different combinations revealed that out of four, only two contribute in improving stability of the protein, while other two have relatively lesser contribution to stability. Further studies shows that all mutants affect the reversibility of protein upon thermal unfolding but to different extent. Reversibility effects in these mutants are brought about mainly due to suppression of aggregation of unfolded protein at elevated temperatures. However, all these mutants, some of which shows partial and other complete inhibition of aggregation, are still prone to chemical modification of protein at elevated temperatures, mainly due to deamidation of asparagine and glutamine residues, thus rendering the whole process of thermal unfolding partially reversible.

5.2 Materials and Methods

5.2.1 Materials

Bis-ANS, PNPB and Triton X-100 were purchased from Sigma Chemical Co. Immobilized pH gradient (IPG) strips were from BioRad (Richmond, CA). All other reagents used were of analytical grade or higher.

5.2.2 Thermal inactivation

Kinetics of thermal inactivation was monitored by incubating the proteins at 75 and 85 °C. The procedure followed was as described earlier. (Section 3.2.10, Chapter 3).

5.2.3 Activity measurements

Lipase activity measurements at room temperature were performed as described earlier (Section 3.2.14, Chapter 3). Briefly, 1 μ g enzyme was added to 1 ml of sodium phosphate buffer (50 mM, pH 7.2), having 2 mM PNPB micellized in 20 mM Triton X-100. Increase in absorbance at 405 nm with time was monitored in spectrophotometer (U-2000, Hitachi, Japan) and used to calculate enzymatic activity.

5.2.4 Thermal unfolding and refolding

Thermal unfolding of lipase mutants was monitored by circular dichroism spectroscopy in a JASCO J-815 spectropolarimeter fitted with Jasco Peltier-type temperature controller (CDF-426S/15). The protein concentration used was 0.05 mg/ml in 50 mM sodium phosphate buffer (pH 7.2) with path length of 1 cm. Temperature dependent unfolding profiles were obtained by heating protein at a constant rate of 1 °C per minute from 25 to 95 °C and measuring the change in ellipticity at 222 nm. Refolding was monitored by reversing the scan with an identical rate of cooling. Temperature wavelength scans were performed by heating the protein at a rate of 1 °C per minute, from 25 to 95 °C, with wavelength spectrum measurement, from 250 to 200 nm, at every 1 degree increment in temperature.

5.2.5 Circular dichroism

Far-UV CD spectra of protein were recorded in the 250-200 nm range using JASCO J-815 spectropolarimeter fitted with Jasco Peltier-type temperature controller (CDF-426S/15). The protein concentration used was 0.05 mg/ml in 50 mM sodium phosphate buffer (pH 7.2) with path length of 0.5 or 1 cm. All spectra reported, except that of temperature wavelength scans, were averaged from three accumulations. Scan speed of 100 nm/minute, response time of 2 seconds, bandwidth of 2 nm and data pitch of 0.2 nm was used for measurements. All spectra were corrected for buffer baseline by subtracting the respective blank spectra recorded identically without protein.

5.2.6 Fluorescence

Fluorescence measurements at room temperature were performed using Hitachi F-4500 fluorimeter. Intrinsic tryptophan fluorescence and bis-ANS fluorescence measurements were performed as described earlier in section 3.2.9, Chapter 3.

Bis-ANS fluorescence at elevated temperature was monitored using Fluorolog 3-22 fluorimeter, fitted with Peltier based cuvette holder, controlled by LFI-3751 temperature controller (Jobin Yvon, USA). Sample composition and measurement parameters were same as described above except that both excitation and emission band passes were set at 2 nm. Protein was added to preheated buffer at desired temperature and incubated for 5 minutes for temperature equilibration. Bis-ANS was added from stock and spectra were recorded after an additional incubation of 1 minute.

5.2.7 Static light scattering

Static light scattering was performed on Fluorolog 3-22 fluorimeter, fitted with Peltier based cuvette holder, controlled by LFI-3751 temperature controller (Jobin Yvon, USA). Both monochromators were set at 360 nm, while slit-widths at 2 nm each. Protein sample of 0.05 mg/ml in 50 mM sodium phosphate buffer (pH 7.2), was rapidly heated to desired temperature followed by incubation at elevated temperature for 10 minutes. Data acquisition was started as soon as the temperature of sample reaches the fixed temperature.

5.2.8 Reversibility of thermal unfolding

Thermal unfolding of lipase mutants was monitored using CD by heating purified protein (0.05 mg/ml in 50 mM sodium phosphate buffer, pH 7.2) from 25 to 95 °C at a constant rate of 1 °C/minute in 1 cm path length cuvette. Refolding was monitored by reversing the scan from 95 to 25 °C with identical rate of cooling. Change in ellipticity at 222 nm with temperature was observed to monitor unfolding and refolding transitions. Far-UV CD spectra of native protein and that of refolded protein were recorded before and after completion of heating and cooling scans. Extent of reversibility of thermal unfolding was also probed by measuring bis-ANS binding to the native and renatured protein. From a concentrated stock in methanol, bis-ANS was added to the protein sample (0.05 mg/ml) to a final concentration of 10 µM and fluorescence spectra was recorded after an incubation of 1 minute, as described above. Extent of refolding upon cooling and loss of protein due to aggregation was also monitored by measuring residual activity and the

fraction of soluble protein left in the solution. Protein samples were centrifuged at 20,000 x g for five minutes followed by passing through 0.22 µm filter (Millex-GV, Millipore, Japan) to remove any aggregated protein. Protein quantitation was done by the modified method of Lowry *et al*, using BSA as standard (Markwell *et al*. 1981).

Reversibility of thermal unfolding of lipase mutants, upon incubation at elevated temperatures for 20 minutes, was probed by Far-UV CD, intrinsic tryptophan fluorescence, bis-ANS binding and by estimating residual activity and the amount of protein left in the supernatant after removal of aggregated fraction. Three ml of protein solution (0.05 mg/ml in 50 mM sodium phosphate buffer, pH 7.2), was incubated at elevated temperatures for 20 minutes followed by rapid cooling and equilibration at 25 °C for 20 minutes in a 1 cm path length cuvette within the cuvette holder of CD spectropolarimeter. Far-UV CD spectra of native and heat treated protein was recorded as described above. One ml of the above sample was used to record intrinsic tryptophan and bis-ANS fluorescence spectra, which was also recorded for an identical sample without heat incubation. Rest of the sample was centrifuged at 20,000 g for five minutes followed by passing through Millex-GV, 0.22 µm filter (Millipore, Japan), to remove any aggregated protein. Enzymatic assay and protein estimation was performed to determine residual activity and the amount of protein left in the supernatant.

5.2.9 Isoelectric focusing

Isoelectric focusing was performed as described earlier with minor modifications (O'Farrell 1975). Thirty microgram of purified protein samples, incubated at elevated temperatures for 20 minutes, was mixed with 80 µl of rehydration buffer (8 M Urea, 50 mM DTT, 2% (w/v) CHAPS, 0.2% w/v Biolyte (3-10) (BioRad, Richmond, CA) with traces of Bromo Phenol Blue) and used to rehydrate Immobilized pH gradient (IPG) strips (pH 3-10) overnight. The strips were overlaid with mineral oil to prevent evaporation. After overnight rehydration, the IPG strip was loaded onto an Isoelectric Focusing (IEF) tray (Protean IEF tray, BioRad) with wet paper wicks over the electrode. The sample was overlaid with mineral oil. The focusing conditions were set up according to the manufacturer's instructions. Briefly, one-step ramping protocol was used as the pH range was broad and temperature was maintained at 20 °C. The second dimension SDS-PAGE was carried out according to standard procedure (Sambrook and Russel 2001). Prior to loading on gel, IPG strips were incubated in equilibration buffers I (6 M urea, 2% SDS,

2% DTT and 20% glycerol in 0.375 M Tris-HCl, pH 8) and II (buffer I, having 2.5% Iodoacetamide) for 20 minutes each, with gentle shaking. The strips were rinsed in 1 x Tris-glycine buffer and loaded on resolving gel followed by sealing with 1% agarose. Rest of the steps was identical to that of SDS-PAGE protocol.

5.3 Results

5.3.1 Thermal inactivation

Kinetics of irreversible thermal inactivation was monitored by incubating the enzymes at elevated temperatures followed by residual activity measurements at room temperature with respect to the time of incubation (Figure 5.1). All thermal inactivation traces were found to follow first-order kinetics. Half-life of inactivation of mutants ($t_{(1/2)}$), monitored at two different temperatures, 75 and 85 °C, along with other thermostability parameters are compared in Table 5.1. At 75 °C, the half-life of parent 4D3 was 4.4 minutes while that of mutants 5-A, 5-B, 5-C and 5-D, having one additional mutation than 4D3, was 38.8, 101.2, 28.5 and 22.5 minutes respectively (Figure 5.1 (a)). Combining individual mutations in sixth generation increases the half-life of inactivation further. Mutant 6-A, having two mutations M134E and M137P, has a half-life of 270.8 minutes which increases further upon adding the mutation S163P in 6-B to 430.5 minutes, which is 100 fold higher than parent 4D3 at the same temperature. Addition of mutation G158D in 6-A and 6-B background to get mutant 6-C and 6-D, does not increase half-life any

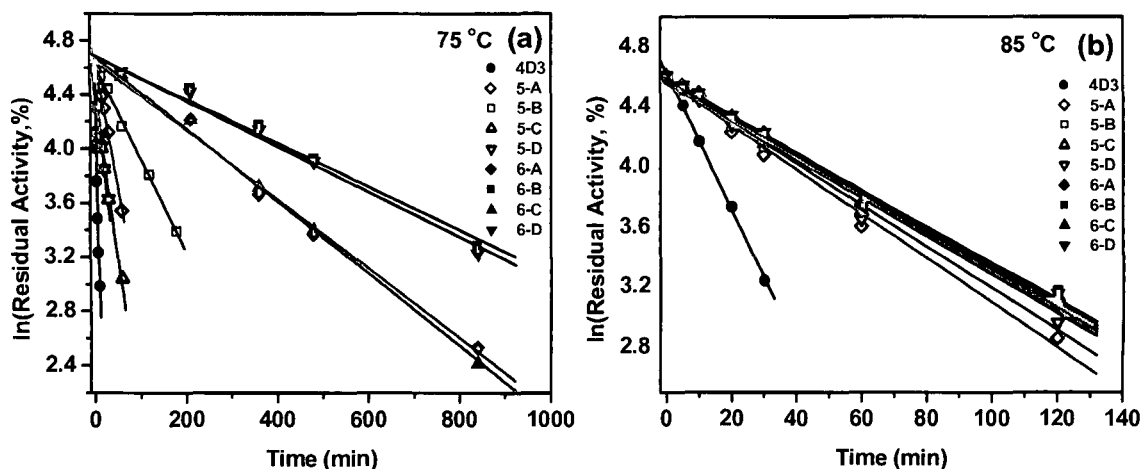


Figure 5.1 Thermal inactivation kinetics of lipase mutants. Time course of thermal stability was measured by calculating the residual activities at 25 °C after incubating the enzymes (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) at 75 °C (a) and 85 °C (b).

Table 5.1. Stability parameters of lipase mutants

Mutant	Mutation	$t_{(1/2)}$ (min) ^a			$T_{(50)}$ ^b (°C)	$T_{m(app)}$ ^c (°C)
		55 °C	75 °C	85 °C		
wild-type	—	2.8	—	—	53.3	56.0
4D3	9 mutations ^d	—	4.4	15.2	68.0	71.2
5-A	M134E	—	38.8	46.9	93.0	72.9
5-B	M137P	—	101.2	54.3	93.0	74.1
5-C	G158D	—	28.5	55.0	73.0	70.7
5-D	S163P	—	22.2	49.9	72.0	72.2
6-A	M134E,M137P	—	270.8	53.9	93.0	76.9
6-B	M134E,M137P,S163P	—	430.5	57.1	93.0	77.8
6-C	M134E,M137P,G158D	—	259.6	54.8	93.0	76.6
6-D	M134E,M137P,G158D,S163P	—	412.6	55.5	93.0	77.7

^a Half-life of thermal inactivation.

^b Temperature at which enzyme loses 50% activity upon incubation for 20 minutes (Figure 4.7, Chapter 4).

^c Mid-point of thermal transition as calculated by CD (Figure 4.9, Chapter 4).

^d 4D3 (A15S, F17S, A20E, N89Y, G111D, L114P, A132D, I157M, N166Y). Mutations listed are in the background of 4D3.

further. This suggests that mutation G158D has little effect on thermal inactivation in the presence of other mutations.

At 85 °C, half life of inactivation of 4D3 was 15.2 minutes which is more than 3 times higher to that at 75 °C (Figure 5.1 (b)). Also, the four individual mutants 5-A, 5-B, 5-C and 5-D show similar half-lives at 85 °C which are 46.9, 54.3, 55.0 and 49.9 minutes, respectively. Interestingly, combining individual mutations, in different combinations in sixth generation, does not improve half lives of mutants 6-A, 6-B, 6-C and 6-D at 85 °C, which remains between 54 to 57 minutes (Table 5.1). This suggests that mutants follow different mechanism of inactivation at these two temperatures. While individual mutations have varying effect on the rate of inactivation at 75 °C, they have little effect at 85 °C, suggesting an altered mechanism of inactivation, which is independent of these stabilizing mutations.

5.3.2 Thermal unfolding and refolding of lipase mutants

Thermal unfolding and refolding of lipase mutants were monitored using CD spectroscopy, by heating proteins at a rate of 1 °C per minute from 25 to 95 °C followed

by cooling at the same rate back to 25 °C. Ellipticity at 222 nm was monitored to follow unfolding and refolding transitions (Figure 5.2). Far-UV CD spectra of native proteins before the thermal scans and that of refolded proteins after heating and cooling scans were also recorded (Figure 5.3.(a)). It was observed that all mutants except 5-C, 5-D and parent 4D3 show reversible unfolding, although the unfolding and refolding transition does not overlap together. All these mutants, 5-A, 5-B, 6-B and 6-D show almost complete (> 95 %) recovery of signal at 222 nm upon cooling, which was also evident from far-UV CD spectra of native and refolded proteins (Figure 5.2 and 5.3 (a)). Parent 4D3 shows progressive loss in signal at 222 nm upon heating, which reaches almost to baseline upon completion of unfolding transition and does not show any recovery upon cooling (Figure 5.2 (a) and 5.3 (a)). Mutant 5-C and 5-D also display similar behavior, except that ellipticity at 222 nm does not reach baseline, but it does not show any recovery in signal upon cooling (Figure 5.2 (d) and (e)). Far-UV CD spectra of refolded 4D3 show complete

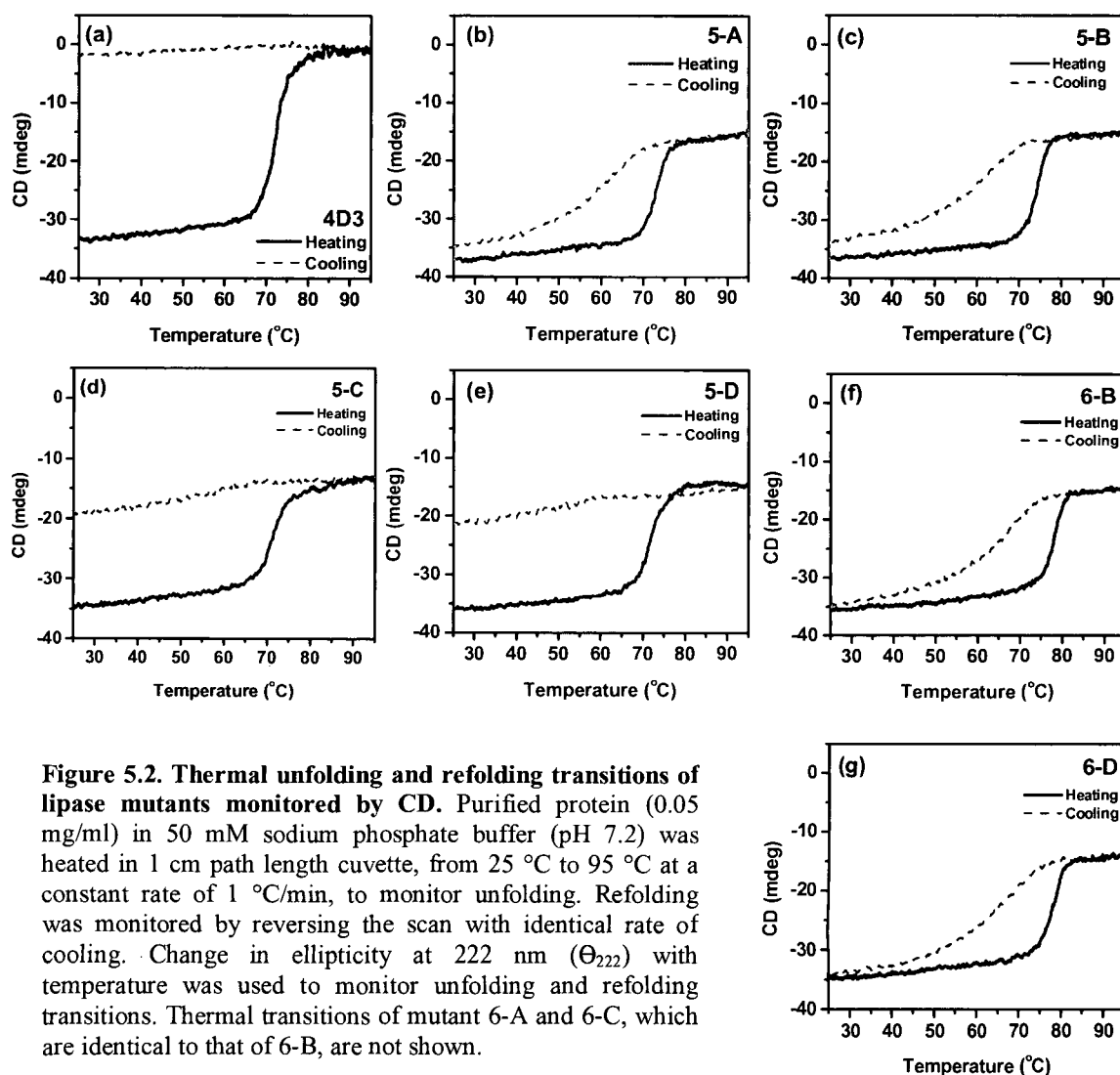


Figure 5.2. Thermal unfolding and refolding transitions of lipase mutants monitored by CD. Purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was heated in 1 cm path length cuvette, from 25 °C to 95 °C at a constant rate of 1 °C/min, to monitor unfolding. Refolding was monitored by reversing the scan with identical rate of cooling. Change in ellipticity at 222 nm (Θ_{222}) with temperature was used to monitor unfolding and refolding transitions. Thermal transitions of mutant 6-A and 6-C, which are identical to that of 6-B, are not shown.

loss of spectra along the entire wavelength while that of mutant 5-C and 5-D show partial loss on refolding (Figure 5.3(a)).

Residual activity measurements and quantitation of soluble protein left in solution after heating and cooling scans shows partial recover of activity, (40-55 %) for 5-A, 5-B, 6-B and 6-D mutants, however, the recovery of soluble protein was more than 84 % (Figure 5.3 (c); Table 5.2). Mutant 5-C and 5-D shows partial recovery of activity (36 and 30%, respectively) as well as soluble protein (53 and 43%, respectively) while 4D3 does not show any recovery of activity or protein. This shows that the loss of signal, as monitored by CD, during unfolding and refolding of 4D3, 5-C and 5-D mutants was due to loss of protein from the sample due to aggregation and precipitation of protein upon unfolding. No such loss of protein, due to aggregation, was observed in case of other mutants. Although, the other four mutants 5-A, 5-B, 6-B and 6-D display almost complete recovery of soluble protein (>85%), recovery of activity was partial (40-55%). To investigate this, tertiary structure of mutants were probed by bis-ANS binding (Figure 5.3 (b)). While all the mutants including

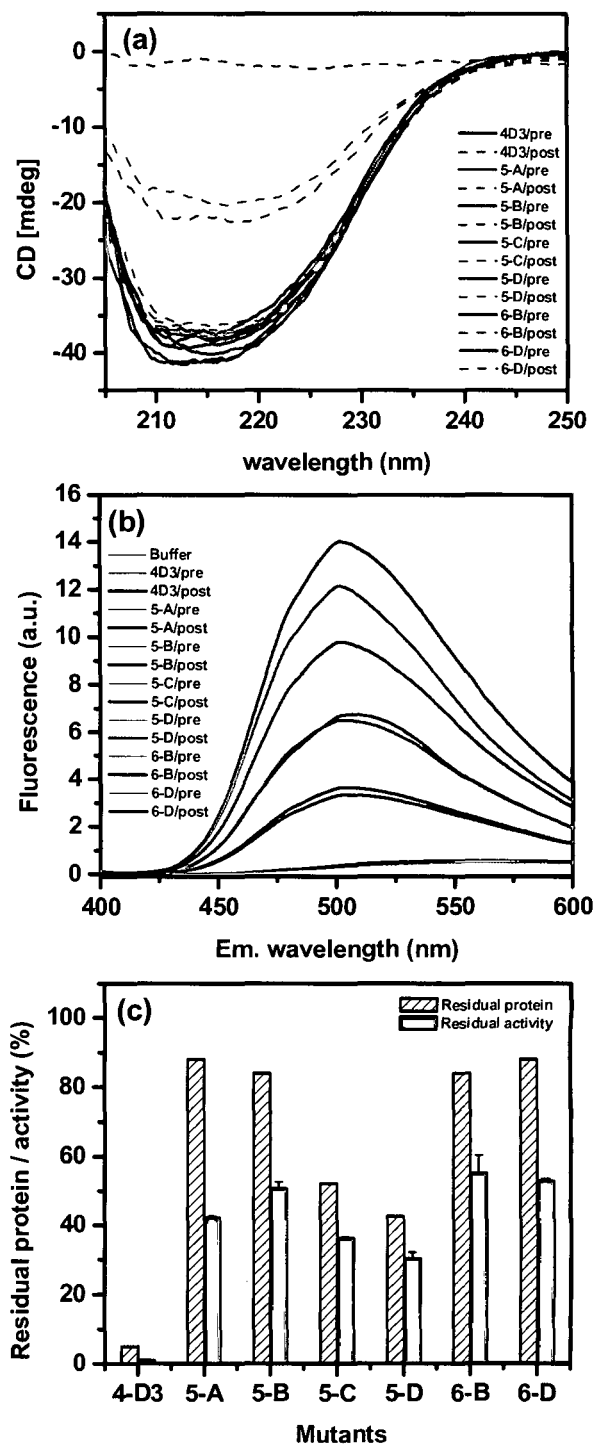


Figure 5.3. (a) Far-UV CD, (b) bis-ANS binding spectra of lipase mutants, before and after thermal unfolding and refolding transitions along with residual activity and soluble protein left in solution after completion of transitions (c). Purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was heated in 1 cm path length cuvette, from 25 °C to 95 °C at a constant rate of 1 °C/min, followed by cooling from 95 °C to 25 °C at an identical rate.

parent 4D3 do not bind to bis-ANS in native state, the refolded proteins bind to bis-ANS to a different extent. The order of bis-ANS binding to refolded mutants was 4D3>5-D>5-C> (5-A, 5-B) > (6-B, 6-D). Bis-ANS is a well established and highly sensitive probe to monitor exposed hydrophobic surfaces on proteins (Smoot et al. 2001). 4D3 does not refold back upon cooling and has lot of exposed hydrophobic surface available for bis-ANS binding. Same is true with mutant 5-C and 5-D, which show partial refolding. Mutants 5-A, 5-B, 6-B and 6-D, which show almost complete recovery of protein upon cooling also shows bis-ANS binding to a significant extent as compared to native protein. This suggests that the refolded protein of these mutants have a significant fraction which is not in the native state (has more exposed hydrophobic surface as that of native protein), which might be inactive. This is in agreement with the observation that although these mutants show complete recovery of protein, with no apparent loss due to aggregation, recovery of activity was only partial.

Monitoring of far-UV CD spectra of mutants during thermal unfolding provided further insights into the changes happening during unfolding (Figure 5.4). Mutants 5-A, 5-B, 6-B and 6-D follow similar pattern during unfolding, displaying spectra corresponding to that of native protein (α -helix along with β -sheet) before the onset of unfolding, and show progressive conversion to that of a denatured protein resembling random coil during the transition range. Progression of spectra corresponding to native state to that of denatured state was accompanied by a clear isosbestic point near 206 nm, which is a signature for reversible two-state model (folded versus unfolded) (Kelly and Price 2000) (Figure 5.4 (b, c, f and g)). Moreover, thermal unfolding in case of these mutants occurred without any increase in HT voltage, suggesting no aggregation of protein happening during the course of unfolding. However, in the case of parent 4D3, loss of signal along the entire spectrum was observed without any isosbestic point during the course of unfolding. The signal decreases along the entire spectrum and reaches almost to baseline upon completion of transition. The decrease in ellipticity was also accompanied with simultaneous increase in HT voltage, which was due to increase in turbidity of the solution due to aggregation of protein upon unfolding. This suggests that the observed loss of CD signal was actually due to the loss of protein from the solution due to aggregation (Figure 5.4 (a)). On the other hand, mutant 5-C and 5-D shows an intermediate trend (Figure 5.4 (d and e)). Both mutants show conversion of native state spectra to that of denatured state

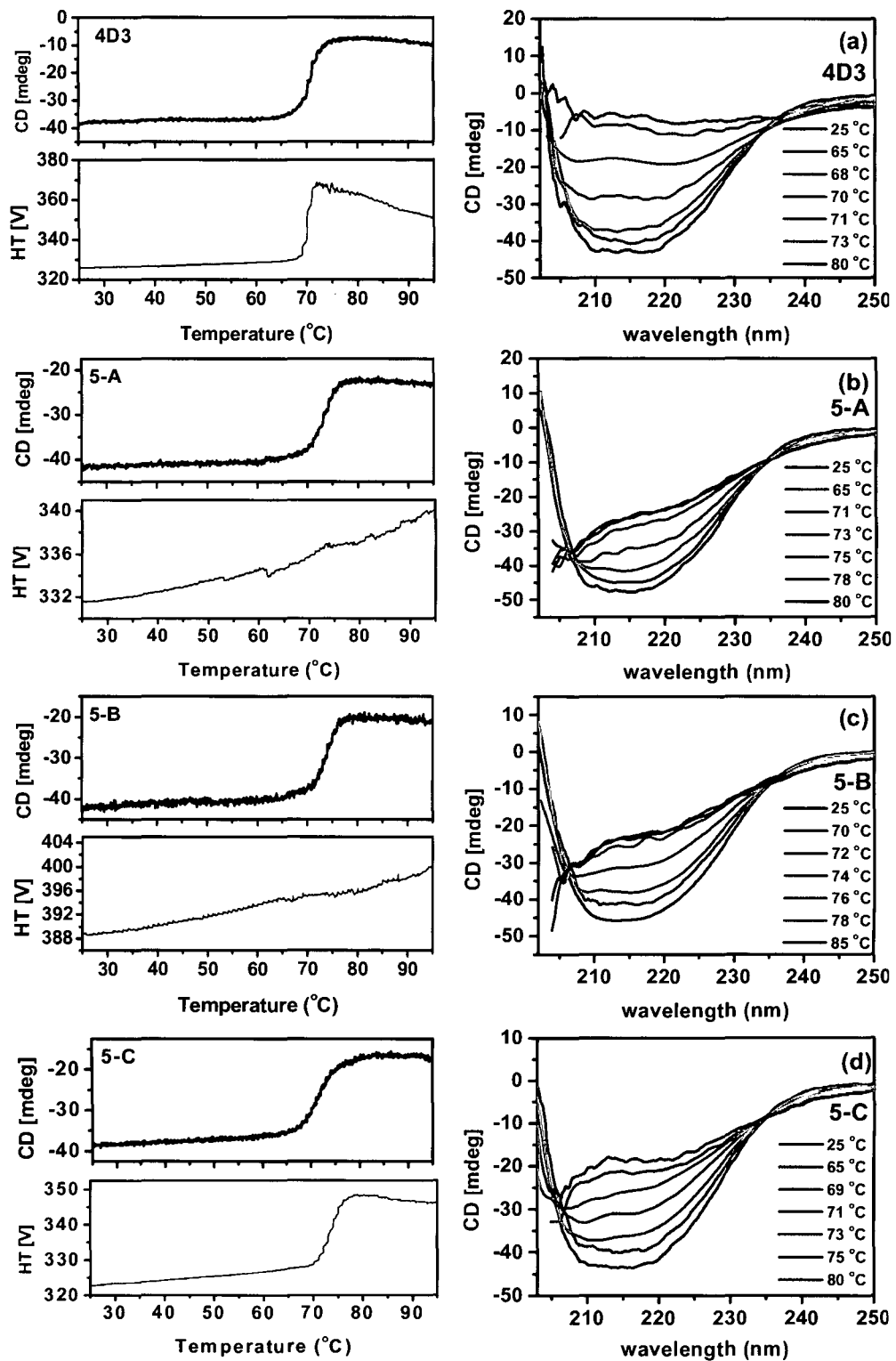


Figure 5.4. Temperature dependent far-UV CD spectra of lipase mutant 4D3 (a), 5-A (b), 5-B (c) and 5-C (d). *Figure is continued on next page.*

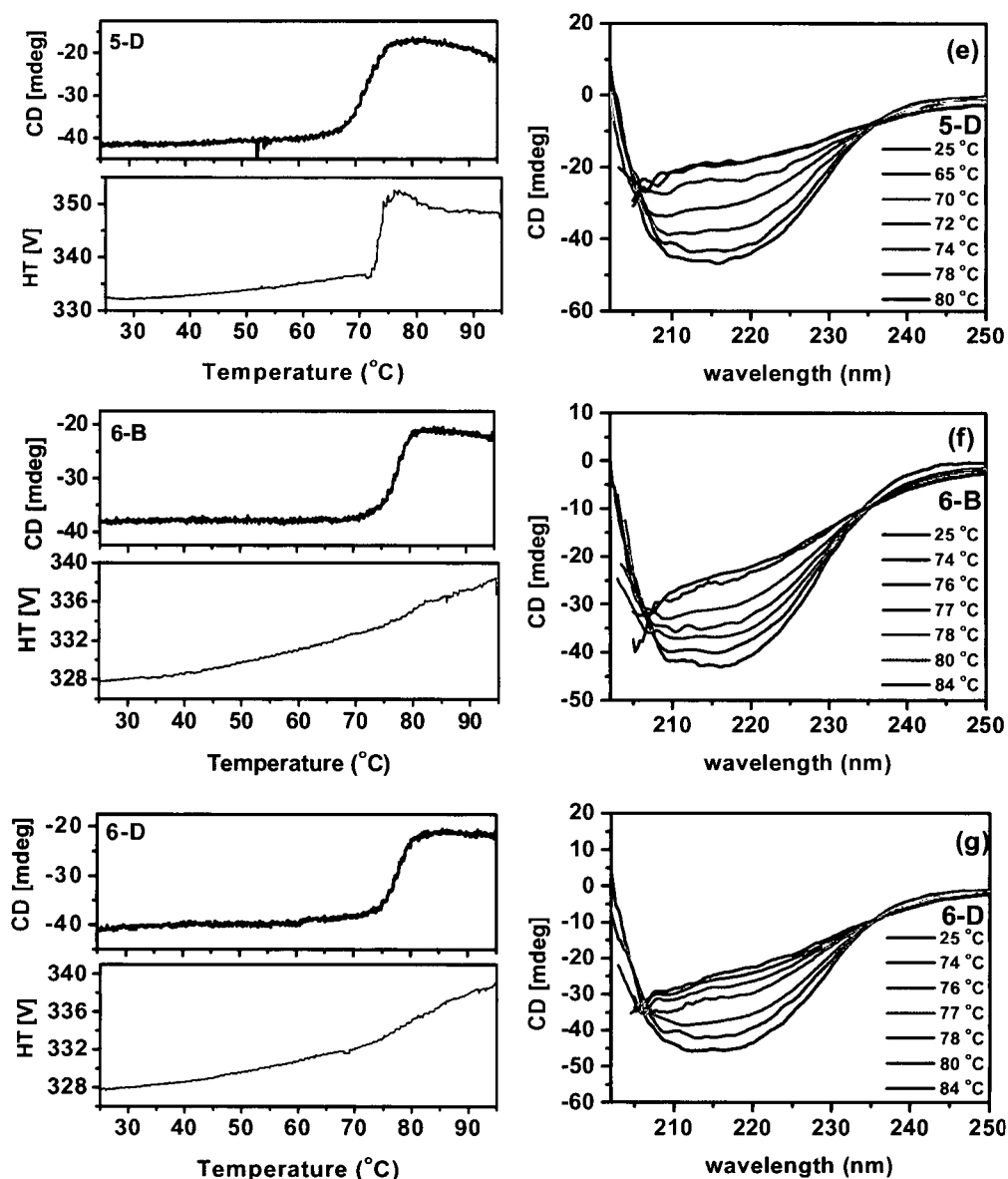


Figure 5.4. Temperature dependent far-UV CD spectra of lipase mutant 5-D (e), 6-B (f) and 6-D (g). Purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was heated in 1 cm path length cuvette, from 25 to 95 °C, at a constant rate of 1 °C/minute with wavelength spectrum measurement at every 1 degree increment in temperature.

resembling random coil upon unfolding along with some loss of signal over the entire spectrum with no clear isosbestic point. The loss in ellipticity at 222 nm was more when compared to other mutants, such as 5-B or 6-B, but did not reach to baseline as in the case of parent 4D3. The loss in ellipticity at 222 nm was also accompanied by increase in HT voltage, although the values do not reach to same level as in case of 4D3. Also unlike 4D3, in which increase in HT voltage was simultaneous with the loss in CD signal,

increase in HT voltage in case of these two mutants was delayed by 2-3 degrees when compared to transition monitored by CD. All these observations suggest that the thermal unfolding transitions of mutant 5-C and 5-D were not completely reversible. Partial loss of protein occurred due to aggregation, as monitored by increase in HT voltage due to increase in turbidity of the solution, which was responsible for the partial loss of ellipticity along the entire spectrum upon unfolding. Unfolding of parent 4D3 was completely irreversible as shown by almost complete loss of ellipticity along the entire spectrum upon completion of thermal transition due to loss of the protein from the solution due to aggregation. Moreover, increase in HT voltage and transition monitored by decrease in ellipticity at 222 nm was almost simultaneous which completely overlap with each other suggesting that protein aggregation was almost simultaneous to unfolding (Figure 5.4 (a)). Contrary to this, unfolding of mutant 5-C and 5-D was partially irreversible as both mutants show significant loss of ellipticity over the entire spectrum upon completion of unfolding transition. This was due to partial loss of protein from the solution due to aggregation as shown by increase in HT voltage upon unfolding. Moreover, since the increase in HT voltage was delayed compared to transition monitored by CD, it appears that aggregation was delayed in case of these mutants compared to unfolding (Figure 5.4 (d and e)). This was clear from the far-UV CD spectra of these mutants in post transition range which corresponds to that of denatured protein suggesting that significant amount of protein was still present in solution with non-regular secondary structure which may refold back to native state upon cooling. Other mutants, 5-A, 5-B, 6-B and 6-D, shows thermal transitions which clearly fits into reversible two state model with no apparent loss of protein due to aggregation.

5.3.3 Reversibility of thermal unfolding of lipase mutants upon incubation at elevated temperatures

The thermal inactivation profiles of lipase mutants 5-C, 5-D and parent 4D3, shows a sharp decrease in residual activity in the region of 68 to 75 °C after which it increases further and the profiles of all the mutants merge at 85 °C after which it shows gradual decrease with further increase in temperature (Figure 4.6, Chapter 4, page 91). In order to investigate it further, reversibility of thermal unfolding of all the mutants, including parent 4D3, was probed by heating the proteins at 75 and 85 °C for 20 minutes followed by cooling and equilibration at 25 °C for 20 minutes. Reversibility was probed

by measuring Far-UV CD, bis-ANS binding, residual activity and the fraction of soluble protein left in solution (Figure 5.5, Table 5.2). Upon incubation at 75 °C for 20 minutes, all mutants except 5-C, 5-D and parent 4D3, show complete recovery of secondary structure upon cooling, as shown by far-UV CD spectra of refolded proteins. While parent 4D3 shows complete loss of signal along the entire spectrum, mutant 5-C and 5-D show partial (~50 %) recovery upon refolding. On the other hand, upon incubation at 85 °C for 20 minutes, all mutants including 5-C and 5-D show complete recovery of secondary structure upon cooling. Also, parent 4D3 shows only partial loss of signal (~60 %) along the entire spectrum (Figure 5.5 (a) and (b)). Estimation of residual activity and the fraction

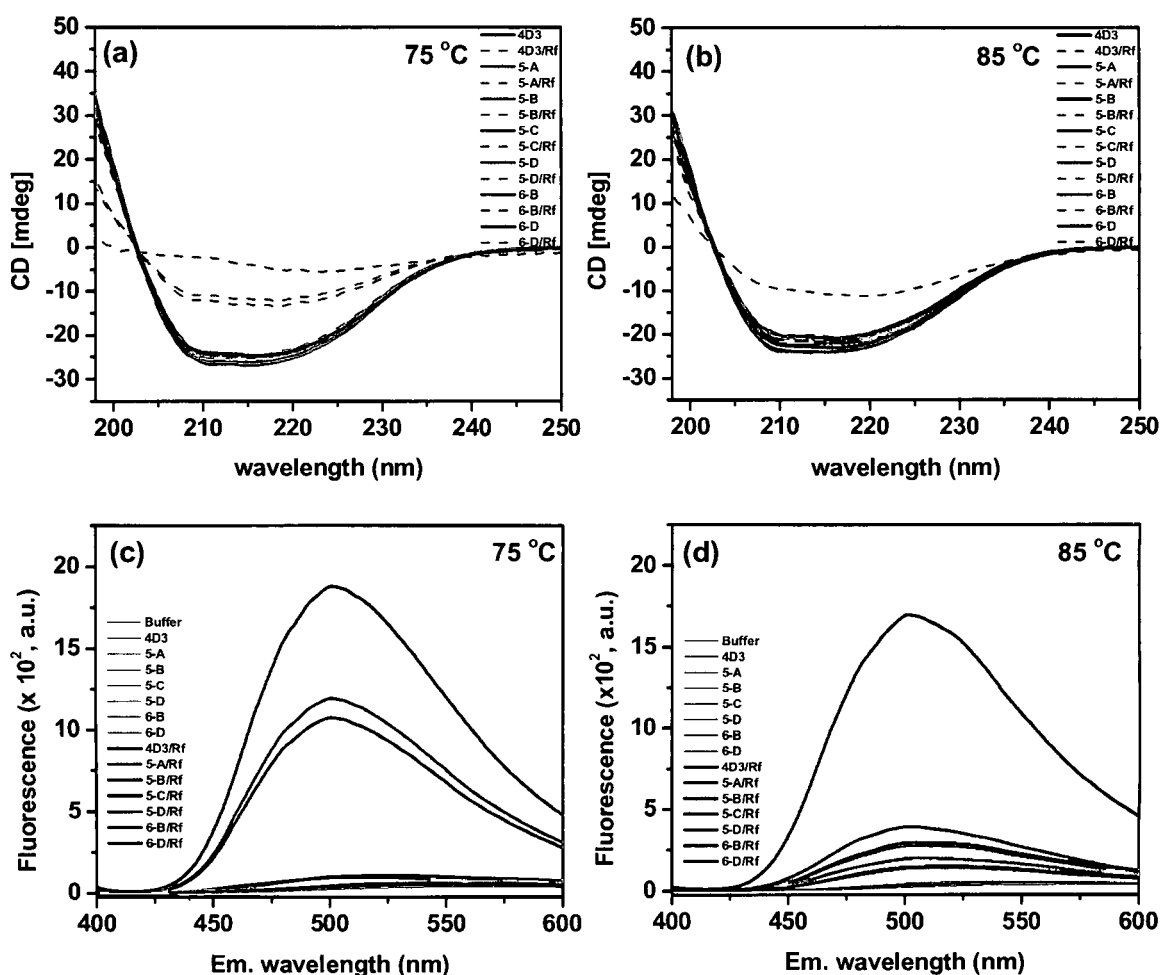


Figure 5.5. Far UV-CD spectra of lipase mutants before and after incubation at (a) 75 °C and (b) 85 °C. Bis-ANS binding to native and refolded lipase mutants after incubation at (c) 75 °C and (d) 85 °C. Far UV-CD spectra of purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was recorded using 0.5 cm cuvette before incubating at 75 or 85 °C for 20 minutes. Proteins were cooled to 25 °C and spectra of refolded proteins were recorded after an incubation of 20 minutes. To the same sample, bis-ANS was added and spectra of refolded proteins were recorded after 1 minute of addition. Bis-ANS binding to native proteins were recorded to a similar sample without heating.

of soluble protein left upon incubation of mutants at 75 °C for 20 minutes show almost complete recovery of protein in case of 5-B, 6-B and 6-D mutants, with little loss in residual activity (Table 5.2). However, in case of mutant 5-C and 5-D only partial recovery (~45-50%) of protein and residual activity was observed, while in case of 4D3, complete loss of soluble protein as well as residual activity was observed. On the other hand, upon incubation at 85 °C for 20 minutes, almost all the protein was recovered in soluble fraction but only partial residual activity (~70 %) for all the mutants including 5-C and 5-D. In addition to that, parent 4D3 shows approximately 30 % recovery of soluble protein as well as residual activity. It is interesting to observe that upon incubation at 75 °C, the residual activity was comparable to that of amount of soluble protein recovered for all the mutants suggesting whatever loss of activity occurs is due to aggregation, but upon incubation at 85 °C all the mutants showed similar loss in residual activity with no apparent loss of soluble protein (Table 5.2). This suggests different mechanism of inactivation involved at these two temperatures. Overall, these data are in agreement with the far-UV and half-life data. Tertiary structure of refolded mutant proteins, after incubation at 75 and 85 °C, were probed by bis-ANS binding (Figure 5.5 (c) and (d)). After incubating at 75 °C, refolded proteins 4D3 followed by 5-D and 5-C show large affinity for bis-ANS, while rest of the mutants show negligible binding. However, after incubating at 85 °C, binding of refolded 4D3 to bis-ANS was reduced compared to the binding after incubation at 75 °C. Refolded mutants 5-C and 5-D also show remarkable reduction in bis-ANS binding, while all other mutants show a significant increase compared to that refolded after incubation at 75 °C. This observation is in perfect agreement with quantitative estimation of soluble protein and residual activity measurements, which shows that upon exposure to 85 °C, although no loss of protein occurs due to aggregation, a significant fraction of protein fails to regain native tertiary structure and has exposed hydrophobic surfaces as probed by bis-ANS binding. This fraction does not have any activity and so is responsible for only partial recovery of activity.

5.3.4 Aggregation of mutants upon incubation at elevated temperatures

Aggregation of lipase mutants was monitored by static light scattering upon incubation at 75 and 85 °C (Figure 5.6). Upon incubation at 75 °C, parent 4D3 aggregates heavily as the scatter value reaches to maximum within 3 minutes after the onset of

aggregation (Figure 5.6 (a)). This was followed by mutants 5-D and 5-C which aggregates slowly when compared to 4D3. Both mutants show a lag phase of 1.5 minutes before the onset of aggregation. Mutant 5-D aggregates slightly more than 5-C, as shown by maximum scatter values, which both mutants achieve by 10 minutes. The maximum scatter values of mutants 5-C and 5-D, after an incubation of 10 minutes, correspond to 50 and 40 % to that of 4D3, respectively.

On the other hand, at 85 °C, only parent 4D3 aggregates while other mutants do not show significant aggregation even upon incubation for 10 minutes (Figure 5.6 (b)). The maximum scatter value of 4D3 at 85 °C was slightly lesser than that at 75 °C (85% of that at 75 °C), suggesting that 4D3 might aggregate to a slightly lesser extent at 85 °C. Overall these observations are in agreement with far-UV CD measurements, residual activity and quantitative soluble protein estimations after incubation of mutants at elevated temperatures which suggest that mutant 5-C and 5-D along with parent 4D3 aggregates upon incubation at 75 °C, which is responsible for the loss of protein as well as activity. However, upon incubation at 85 °C, all mutants except 4D3, do not show any aggregation thus allowing complete recovery of protein upon refolding.

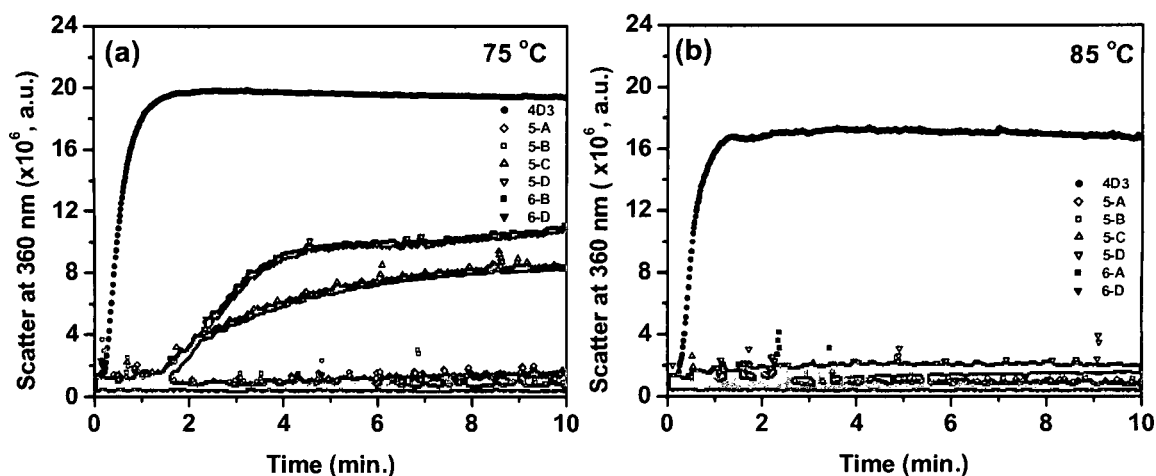


Figure 5.6. Aggregation of lipase mutants, as monitored by static light scattering, upon heating at 75 °C (a) and 85 °C (b). Three ml of protein solution (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was heated at desired temperature in 3 ml fluorescence cuvette for 10 minutes in Fluorolog 3-22 fluorimeter, fitted with peltier based cuvette holder. Both excitation and emission monochromators were set at 360 nm with 1 nm slit width. Data acquisition was started as soon as the sample acquires the desired temperature.

5.3.5 Bis-ANS binding to denatured mutants at elevated temperatures.

Denatured state of mutants at elevated temperatures were probed by their affinity to bis-ANS binding (Figure 5.7). All mutants show varying but significant binding to bis-ANS upon incubation at 75 °C (Figure 5.7 (a)). Parent 4D3 shows maximum binding followed by mutants which follow the order 4D3 > 5-D > 5-C > (5-A, 5-B) > (6-B, 6-D). All the mutants have little affinity for the hydrophobic dye in native state (Figure 5.5 (c); Figure 4.5 (d), Chapter 4). Moreover, upon complete unfolding in 6 M GdmCl, none of the mutant binds to bis-ANS (data not shown). This suggests that upon incubation at 75 °C all mutants (except 6-B and 6-D, which got unfolded at > 80 °C) were denatured due to unfolding but fail to acquire completely unfolded state and have significant residual structure. The partially unfolded state of mutants, accumulated at 75 °C, are different as evident by different extent of bis-ANS binding, suggesting that they differ in exposed hydrophobic surface area. Denatured mutant displaying higher bis-ANS binding has higher solvent exposed hydrophobic surface as compared to those which show reduced binding. This observation is in perfect agreement with the aggregation profiles of the mutants upon incubation at 75 °C. Parent 4D3, which shows highest bis-ANS binding at 75 °C, may probably display maximum hydrophobic surface, which was responsible for the aggregation of protein due to non-specific hydrophobic interactions. This was followed by mutant 5-D and 5-C, which show reduced bis-ANS binding as compared to 4D3, thus

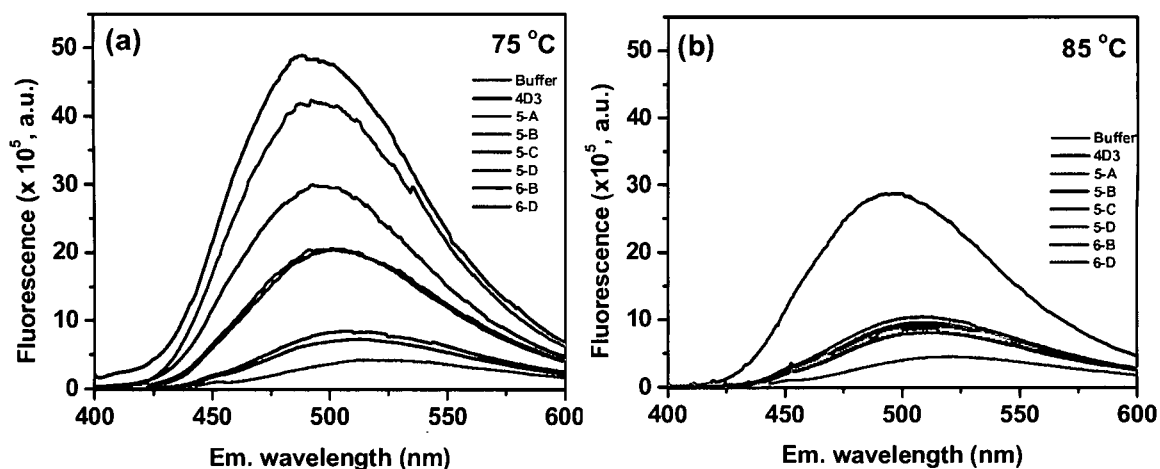


Figure 5.7. Bis-ANS binding to lipase mutants at 75 °C (a), and 85 °C (b). Three ml of 50 mM sodium phosphate buffer (pH 7.2) was heated at desired temperature in 3 ml fluorescence cuvette for 10 minutes. Purified protein was added to preheated buffer to get final concentration of 0.05 mg/ml and incubated for 5 minutes for temperature equilibration. Bis-ANS was added to a final concentration of 10 μM. Spectra was recorded after an additional incubation of 1 minute.

also have smaller exposed hydrophobic surface due to which these mutants aggregate to a lesser extent than 4D3. Although, mutants 5-A, 5-B, 6-B and 6-D show significant binding to hydrophobic dye at 75 °C and thus have some exposed hydrophobic surface, they do not show any perceptible aggregation (Figure 5.6 (a)). This might be possible either due to that the exposed hydrophobic region was not large enough for productive non-specific hydrophobic interaction to bring about aggregation or due to introduction of a negative charge which inhibits non-specific intermolecular hydrophobic interaction due to electrostatic repulsion.

Incubation of mutants at 85 °C shows that only parent 4D3 binds to bis-ANS to significant extent, while all other mutants show highly reduced binding compared to 4D3 (Figure 5.7 (b)). Moreover, spectra of all the mutants, except 4D3, completely overlap with each other suggesting that all the mutants have acquired denatured states which have similar exposed hydrophobic surfaces. Overall this observation is in agreement with the aggregation profiles of the mutants at 85 °C, which shows that only 4D3, showing significant binding to bis-ANS, aggregates while no significant aggregation was observed in case of other mutants (Figure 5.6 (b)).

5.3.6 Effect of Ionic strength on the reversibility of thermal unfolding

Effect of ionic strength on the reversibility of thermal unfolding of mutants was probed by monitoring aggregation, far-UV CD, residual activity measurements and quantitative estimation of soluble protein after incubation of mutants at elevated temperatures in the absence and presence of 1 M NaCl (Figure 5.8, Table 5.2). As shown in Figure 5.8 (a), in the absence of NaCl, only parent 4D3 and mutants 5-D followed by 5-C show significant aggregation upon incubation at 75 °C, as monitored by static light scattering. All other mutants including 5-A, 5-B, 6-B and 6-D do not show any perceptible aggregation under similar conditions (Mutant 6-B and 6-D were incubated at 80 °C). However, in the presence of 1M NaCl, there is a remarkable increase in the rate and extent of aggregation in case of mutant 5-D and 5-C, which follows the same behaviour as that of parent 4D3 (Figure 5.8 (b)). Moreover, mutant 5-A, which does not show any perceptible aggregation in the absence of salt, also shows significant aggregation in the presence of 1 M NaCl. However, other mutants including 5-B, 6-B and 6-D, which do not show any noticeable aggregation in the absence of salt, remain unaffected in the presence of 1 M NaCl. These observations were corroborated by far-UV CD measurements performed

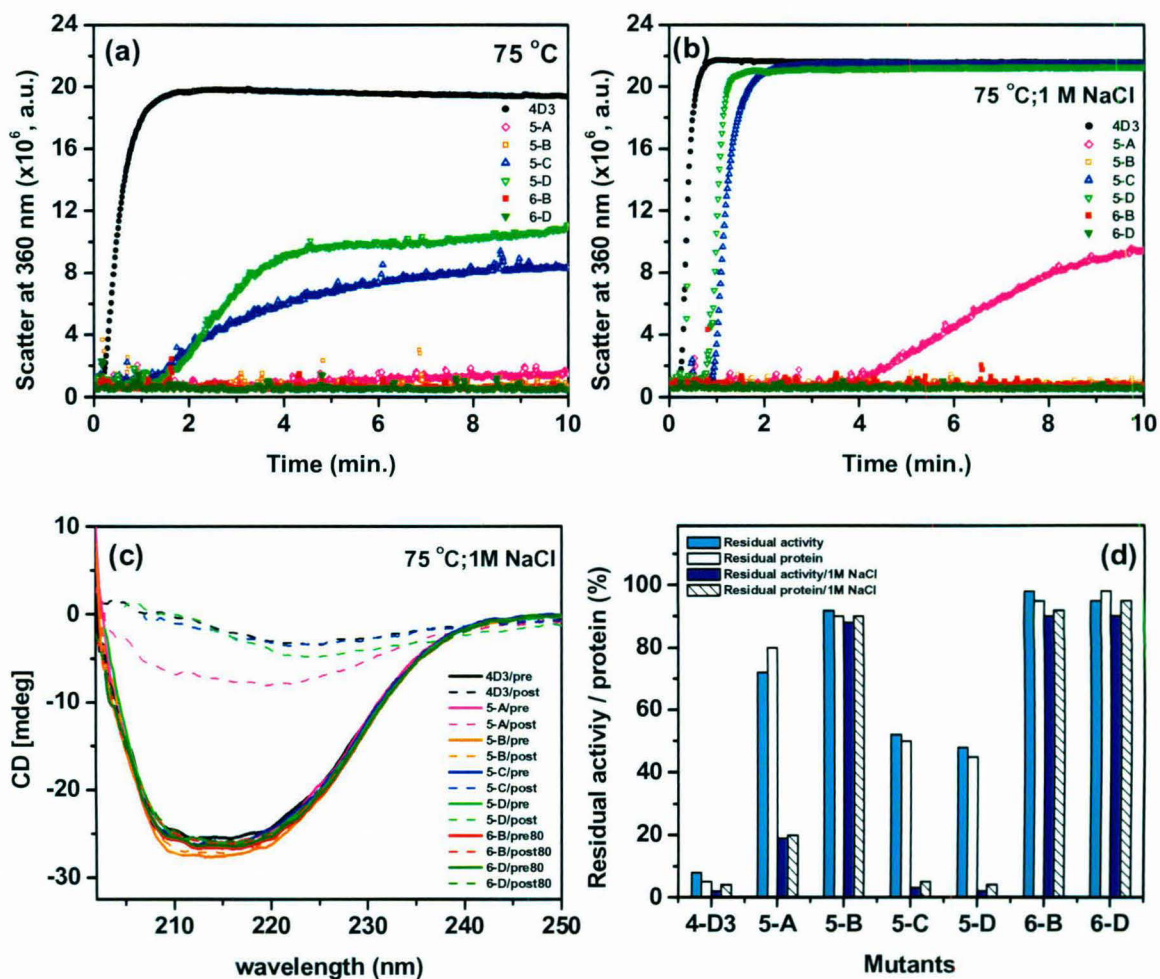


Figure 5.8. Effect of ionic strength on aggregation and reversibility upon thermal unfolding of lipase mutants. Aggregation of lipase mutants, as monitored by static light scattering, upon heating at 75 °C in the absence (a) and presence of 1 M NaCl (b). Far-UV CD spectra of lipase mutants before and after incubation at 75 °C for 20 minutes in the presence of 1 M NaCl (c). Residual activity (RA) and the amount of soluble protein recovered in the solution (RP) upon incubation of lipase mutants, at 75 °C, in the absence and presence of 1 M NaCl (d). Mutant 6-B and 6-D were incubated at 80 °C.

before and after incubating the mutants at high temperatures for 20 minutes, in the absence and presence of 1 M NaCl. It was observed that all mutants, except 4D3, 5-D and 5-C, regain almost all the secondary structure upon incubation at 75 °C for 20 minutes (Figure 5.5 (a)). While parent 4D3 shows no recovery, 5-D and 5-C show partial recovery (~50%) upon refolding. However, in the presence of 1 M NaCl, 4D3, along with mutant 5-D and 5-C, show complete loss of ellipticity along the entire spectrum. Moreover, mutant 5-A also shows about 80% loss in signal along the entire spectrum. On the other hand, mutant 5-B, 6-B and 6-D shows almost complete recovery of secondary structure as the spectra of native and refolded proteins completely overlaps with each other (Figure 5.8 (c)).

Table 5.2. Residual Activity and soluble protein recovered after incubation of lipase mutants at elevated temperatures.

Mutant	Thermal ^a transitions		75 °C ^b				85 °C ^b	
			1M NaCl				RA(%)	RP(%)
	RA (%)	RP (%)	RA (%)	RP (%)	RA (%)	RP (%)		
4D3	1	5	8	5	2	4	30	32
5-A	42	88	72	80	19	20	68	98
5-B	51	84	92	90	88	90	68	95
5-C	36	52	52	50	3	5	69	95
5-D	30	43	48	45	2	4	67	98
6-B	55	90	98	95	90	92	70	99
6-D	53	88	95	98	90	95	72	96

^a Residual Activity (RA) and Residual soluble protein (RP) left in solution after thermal unfolding and refolding scans, from 25 °C to 95 °C and back, at a rate of 1 °C/minute.

^b After incubating proteins at indicated temperatures for 20 minutes.

Residual activity measurements and quantitative estimation of soluble protein further confirms the above observations (Figure 5.8 (d)). All the mutants, except 4D3, show almost complete recovery (> 80 %) of protein as well activity upon incubation of mutants at 75 °C for 20 minutes in the absence of salt. While 4D3 shows no recovery (< 5%) of protein as well as activity, recovery of activity and protein was partial (~ 50 %) in case of 5-C and 5-D. However, in the presence of 1 M NaCl, 4D3 as well as 5-C and 5-D show little recovery of activity and protein, while other mutants, including 5-B, 6-B and 6-D shows complete recovery of protein as well as activity. On the other hand, mutant 5-A, which shows about 80% residual activity and protein in the absence of salt, shows less than 20% recovery of both in the presence of 1 M NaCl. This might be due to the fact that mutation M134E has introduced an additional negative charge in 5-A, which might have prevented the aggregation of mutant upon unfolding, due to electrostatic repulsion between denatured protein molecules. This electrostatic repulsion got abolished in the presence of NaCl due to shielding of charge thus facilitating non-specific hydrophobic interaction between denatured molecules, causing aggregation. On the other hand, mutant 5-B, which has M137P mutation, does not show aggregation neither in the absence or presence of salt. Aggregation was blocked in this mutant by the introduction of proline, which locks the denatured state of the mutants in such a conformation which was less prone to aggregation. Since it was not dependent on the electrostatic repulsion between denatured molecules, due to the introduction of any charged residue, aggregation of

mutant 5-B remains independent of the ionic strength of the medium as there is no charge involved, shielding of which may promote aggregation. Effect brought about by this mutation was so pronounced that it was sufficient to completely inhibit aggregation when present alone in 5-B or when present along with other mutations as in case of 6-B and 6-D.

5.3.7 Thermal unfolding and refolding of 6-B mutant

Thermal unfolding and refolding of mutant 6-B, which shows an atypical inactivation profile (Figure 4.6, Chapter 4), was investigated by incubating the protein at a range of temperature (60 – 100 °C) for 20 minutes and the extent of refolding was monitored upon cooling to 25 °C after 20 minutes using several probes (Figure 5.9). Far-UV CD spectra of native and protein refolded after incubation at 75 -95 °C overlaps over each other showing complete recovery of secondary structure (Figure 5.9 (a)). Residual activity measurement and estimation of soluble protein upon refolding shows gradual loss in residual activity upon incubation at temperature beyond 75 °C, with no loss of protein over entire range till 100 °C (Figure 5.9 (b)). Tertiary structure of 6-B mutant was probed by intrinsic tryptophan fluorescence and bis-ANS binding, before and after incubation at elevated temperatures. Fluorescence emission arising from two tryptophans remains quenched in native state, with emission maxima (λ_{max}) around 339 nm. Refolded mutant 6-B shows increase in tryptophan fluorescence along with red shift in emission maxima (345 nm) with increase in temperature of incubation (Figure 5.9 (c)). Similar observations were made by bis-ANS binding. Mutant 6-B does not bind to bis-ANS in the native state, but binding of refolded protein increases with increase in temperature of incubation (Figure 5.9 (d)). All these data suggest that mutant 6-B undergoes reversible thermal unfolding which is partial in nature. Although protein does not show any loss due to aggregation and regains complete secondary structure upon refolding, a fraction of it fails to regain tertiary structure of the native state. No aggregation was observed upon incubation at any temperature and the recovery of secondary structure was independent of the temperature of incubation but recovery of native state (tertiary structure) depends upon the temperature of incubation, as shown by decrease in residual activity, increase in tryptophan fluorescence and bis-ANS binding with increase in temperature.

Increase in failure to achieve native state with increase in the temperature of incubation may result due to the modification of amino acid side chains at elevated temperatures. Asparagine and glutamine residues are particularly prone, which deamidates

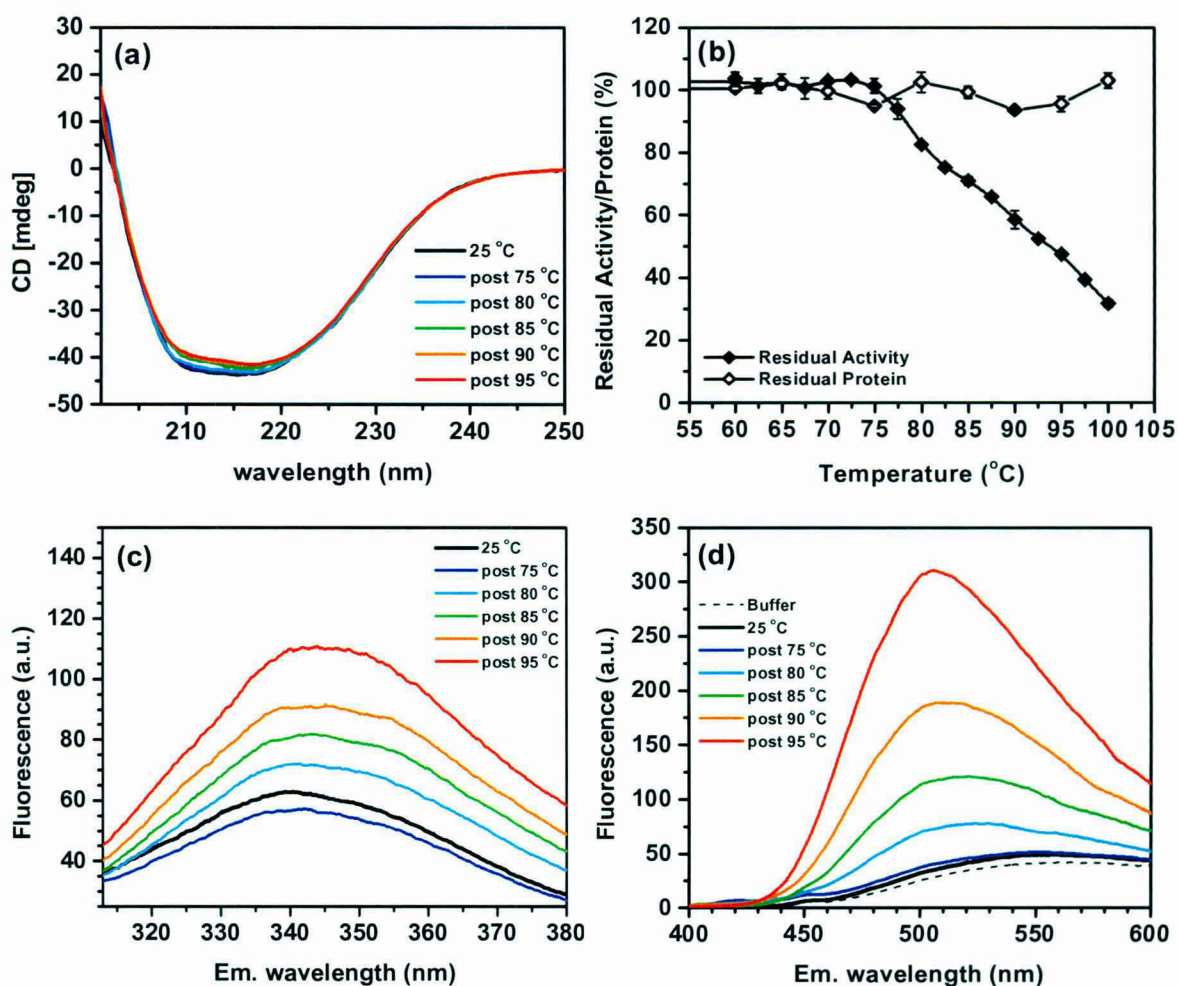


Figure 5.9. Reversibility of thermal unfolding of 6-B mutant upon incubation at a range of temperature, as monitored by (a) Far-UV CD, (b) Residual activity and protein left in supernatant, (c) Intrinsic Tryptophan Fluorescence and (d) bis-ANS binding. Purified protein (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2) was incubated at a range of temperature (75 – 100 °C) for 20 minutes followed by cooling to 25 °C. Residual activity, amount of protein left in supernatant and spectroscopic measurements on refolded protein were performed after an incubation of 20 minutes. All the measurements were performed with an identical sample of native protein without heating.

at high temperatures (> 80 °C), thus causing generation of aspartate and glutamate (Daniel et al. 1996). To check this possibility, isoelectric focusing (IEF) of 6-B mutant was performed after incubating the protein at 70 – 100 °C for 20 minutes (Figure 5.10). Isoelectric point (pI) of mutant 6-B is 8.2 compared to wild-type protein which has pI of 9.9 (Lesuisse et al. 1993). This is due to introduction of 4 negative charges during the course of *in vitro* evolution of wild-type to 6-B. Upon incubation of 6-B at 70 and 80 °C, only a single spot at pH value of 8.2, corresponding to that of native protein was observed. However, samples incubated at 90 and 100 °C shows an additional broad band in the range of pH 6-7. This clearly suggest that incubation of mutant 6-B at temperatures more than 80

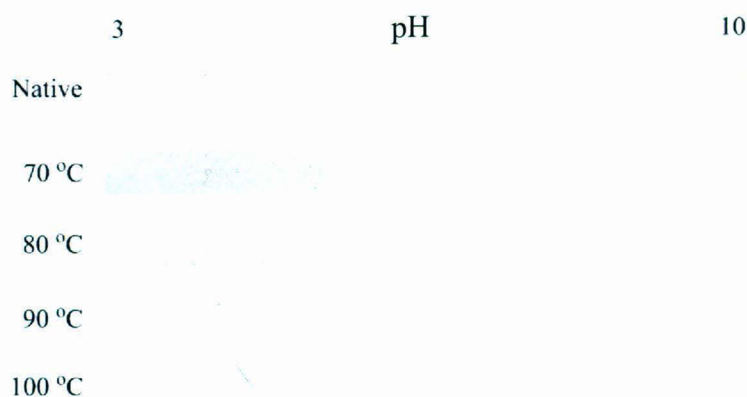


Figure 5.10. Isoelectric focusing profile of 6-B lipase mutant, after incubation at elevated temperatures (70 – 100 °C) for 20 minutes.

°C modifies the side chains of amino acid residues of the protein, mainly by deamidation of asparagines/glutamine residues, thus increasing the negative charge content of the protein which in turn have lower pI values than the native protein. This non-native fraction, which increases with the increase in temperature, has similar secondary structure content, but altered tertiary structure, making it inactive and was responsible for the loss of the activity.

5.3.8 Structural basis of thermotolerance

All the four mutations identified in this study were modeled onto the structure of mutant 4D3 (PDB id: 3D2C, chain A) (Ahmad et al. 2008), in order to understand the structural basis of the observed ability to resist thermal inactivation of mutants (Figure 5.11 (a)).

Mutation M134E is located in a loop connecting $\beta 7$ strand and αE helix. Substitution of surface exposed Met at position 134 with negatively charged Glu might have stabilized the protein by favoured polar interactions with solvent. Moreover, this mutation has dramatic effect on the reversibility of thermal unfolding. Mutation, M134E is located adjacent to Asp133 and Asp132 (Figure 5.11(b)). This brings three negative charges in close vicinity, thus making a cluster of three negative charges. This cluster of three consecutive negative charges is mainly responsible for blocking the aggregation of denatured protein due to intermolecular electrostatic repulsion during thermal unfolding.

The second mutation, M137P, is also located in the same loop in which mutation M134E is present (Figure 5.11 (b)). Met 137 is also present on the surface and side-chain

is completely exposed to the solvent. Substitution of hydrophobic Met with Pro does not affect the polarity of the region but brings stability due to the constraints imposed by the proline in the backbone conformation (Matthews et al. 1987). Besides improving stability of the protein, this mutation has the dramatic effect of the reversibility of thermal unfolding. Introduction of only this mutation in 4D3 completely blocks aggregation over entire range of incubation temperature. It is well known that proteins aggregate upon heating, due to the non-specific interactions among the hydrophobic surfaces, exposed in the partially unfolded/denatured species, which gets accumulated during unfolding (De Bernardez et al. 1999). Introduction of proline at this crucial position must have “locked” the denatured state in a conformation which is not prone to aggregation (Shortle 1996).

The third mutation, G158D, is located in a 3_{10} helix G5, and is completely surface

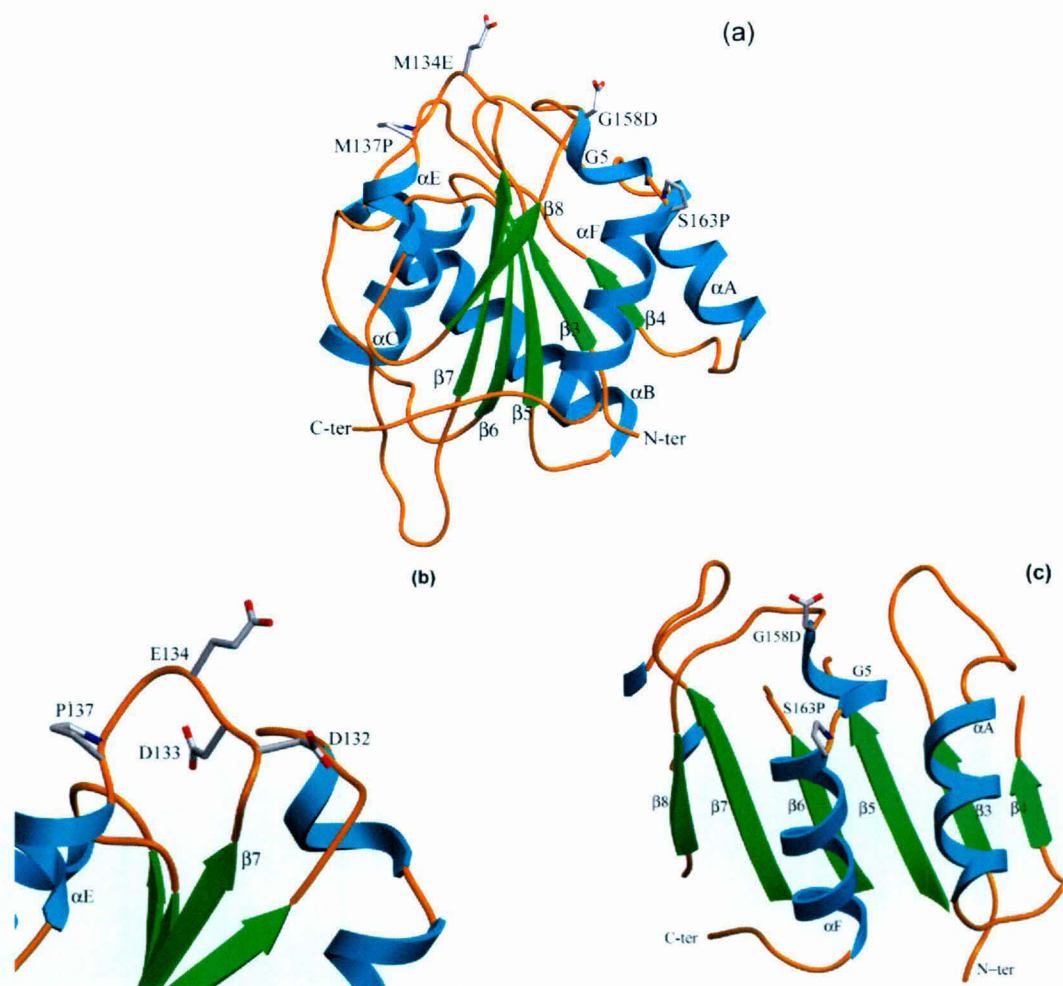


Figure 5.11. (a) Ribbon diagram of lipase mutant 4D3, indicating the relative position of four mutations. (b) Location of E134, which is adjacent to two negatively charged residues, D132 and D133, also showing P137 on the same loop. (c) Mutation G158D on 3_{10} -helix G5, followed by S163P, on N-term of last helix, αF.

exposed (Figure 5.11 (c)). Introduction of this mutation does not have any profound effect on the stability of the protein but have a remarkable effect on the reversibility after thermal unfolding. Mutation G158D occurred in the middle of a stretch of hydrophobic amino acid residues G₁₅₃VGHM(G158D)LL₁₆₀, which might be involved in aggregation upon thermal denaturation. Introduction of a negative charge blocks the non specific hydrophobic interaction, thus reducing the rate of aggregation at elevated temperatures.

The fourth mutation, S163P, is located at the N-terminal of the helix α F, which is the last helix of the protein. Ser163 is the first residue of the helix and its side-chain is completely exposed to the solvent (Figure 5.11 (c)). Mutation S163P affects the reversibility of thermal unfolding of the protein by blocking aggregation during unfolding in a similar manner as M137P mutation, but the effect was only partial. This might be due to that fact that position 163 which is comparatively outside of the protein making it less crucial than 137, which is centrally located, thus affecting the conformation of denatured state in a better manner.

All the four mutations reported effect the ability of the protein to resist high temperatures mainly by blocking aggregation. While two mutations bring about this effect due to the introduction of charged residues at crucial locations, two mutations performed the same action due to the introduction of proline, by locking the denatured states in such conformations, making them less prone to aggregation.

5.4 Discussion

B. subtilis lipase, like many other proteins, also undergoes irreversible thermal unfolding, primarily due to aggregation upon unfolding. In an attempt to increase the thermostability of this protein by *in vitro* evolution, 13 thermostabilizing mutations were identified during two rounds of random mutagenesis and screening followed by one round of site saturation mutagenesis at four positions identified in previous rounds (Chapter 3 and 4). Nine mutations out of thirteen have little effect on the reversibility of unfolding, despite having profound effect on protein thermostability. Combination of all these nine mutants in wild-type protein increases its apparent melting temperature by 15 °C along with more than a million fold reduction in inactivation rate. But the unfolding of all these mutants was found to be irreversible due to the aggregation of protein upon unfolding (Chapter 3). Site-saturation mutagenesis was performed on four additional positions identified during previous rounds of random mutagenesis. The variant, having all the four

optimized mutations, shows enhanced thermostability as shown by further increase in apparent melting temperature by 7 degrees (Chapter 4). Interestingly, all the four mutations, optimized by site-saturation mutagenesis, have profound effect on the reversibility of unfolding apart from increasing conformational stability of the protein. All the four mutations, M134E, M137P, G158D and S163P resist aggregation of the denatured species to different extent and also by different mechanisms apart from increasing thermostability. The main reasons for the identification of these mutations are the conditions employed during the screening of thermostable variants. Direct measurement of enzymatic activity of large number of mutants is not possible at elevated temperatures in high throughput format, due to number of complications involved (Daniel and Danson 2001). However, employing this kind of screening identifies mutants which retain their native folded structure at elevated temperatures. On the other hand, a relatively simple and more commonly employed way is to measure residual activity of mutants, after exposure to elevated temperatures for a fixed period. But, this kind of screening can identify different class of mutants including those which maintains their structure and remain active at high temperatures (more thermostable; have higher T_m), may or may not refold back efficiently after unfolding or those which loose their structure at high temperature but refold back to native state efficiently upon cooling to ambient temperatures (more thermotolerant). The mutations identified in this study belong to both the classes. All the nine mutations present in 4D3 compared to wild-type, belong to first category as they all increase the melting temperature but these mutants fail to refold back after unfolding while the four mutations identified by site-saturation mutagenesis belong to latter category as all of them increase the refolding efficiency by blocking aggregation. Among these four mutations, two mutations M137P and S163P, increase conformational stability of the protein along with improved refolding efficiency by blocking aggregation albeit to different extent. The other two mutations, M134E and G158D, have little effect on protein stability, but improve refolding efficiency of the protein by inhibiting aggregation.

Individually assessing, mutation M137P was found to have most dramatic effect in improving thermostability as well as the reversibility of unfolding. Mutation S163P also increases stability of the protein marginally and also improves the reversibility only partially. On the other hand, mutation M134E does not increase protein stability, but blocks aggregation drastically. Finally, mutation G158D, which was found to have little destabilizing effect of the conformational stability of the protein, also improves the

reversibility of unfolding partially. Combination of mutations in mutant 6-B and 6-D, improves the stability of the protein in additive manner, but hardly shows any effect on the reversibility of unfolding. Aggregation was completely blocked, over the entire range of temperature, by the introduction of only mutation M137P. This mutation appears to have a dominant role, remains unaffected in the presence of other mutations, which have only partial effect on aggregation of protein.

Thermal inactivation profile of parent 4D3, along with mutant 5-C and 5-D, shows sharp decrease in residual activity in the range of 68 – 75 °C, which closely follows their corresponding unfolding profiles, however, upon further increase in temperature, residual activity increases till 85 °C. Other mutants including 5-A, 5-B and 6-B, show gradual decrease in residual activity after unfolding with increase in temperature, which does not match with their unfolding transitions. Comparison of the inactivation profile with the fraction of soluble protein recovered gave better picture (Figure 5.12). Mutants 5-C and 5-D, along with parent 4D3, show loss of activity, primarily due to loss of protein due to aggregation after unfolding which reaches its maximum at 75 °C. Upon further increase in temperature, the loss of protein due to aggregation decreases accounting for the increase in residual activity. These mutants show loss of protein due to aggregation within the range of 70 – 85 °C with complete recovery of protein at temperatures beyond 85 °C. On the other hand, mutants 5-A, 5-B and 6-C show no loss of protein due to aggregation over entire range of temperature but show gradual decrease in residual activity upon incubation at temperatures beyond 80 °C, which is well above the melting temperature of these mutants. All these observations were corroborated with spectral and activity measurements performed upon incubation of

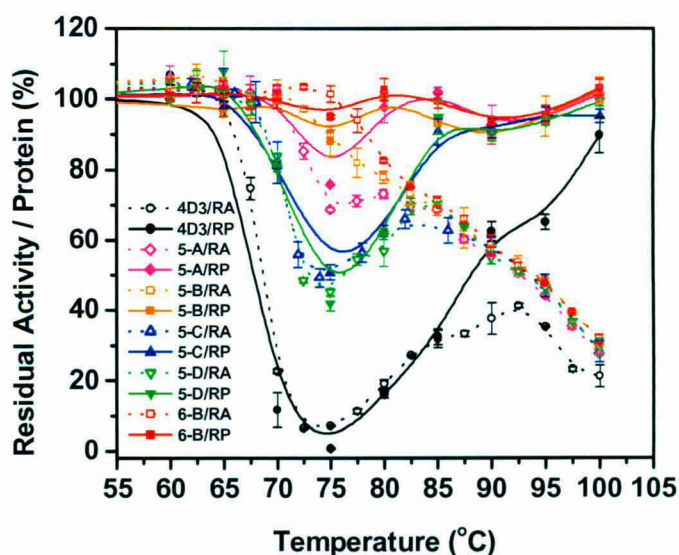
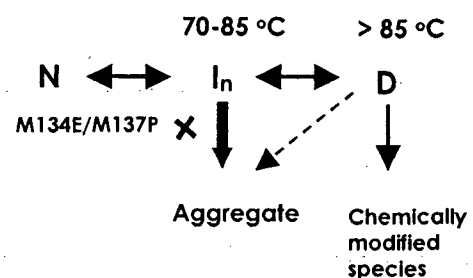


Figure 5.12. Thermal inactivation profiles of lipase mutants along with the fraction of protein recovered upon thermal challenge. Enzymes (0.05 mg/ml) in 50 mM sodium phosphate buffer (pH 7.2), were incubated at various temperatures for 20 minutes followed by equilibration at 25 °C for 15 minutes. Samples were centrifuged at 20×10^3 g for 10 minutes to remove any aggregated fraction prior to residual activity (RA) measurement at 25 °C and protein estimation to calculate residual protein (RP) left in solution

mutants at 75 and 85 °C. It shows that the partial recovery of unfolding, shown by few mutants upon incubation at 75 °C, was primarily due to the loss of the protein due to aggregation, which decreases with further increase in temperature. Moreover, all mutants, except 4D3, do not show any aggregation at 85 °C, accounting for complete recovery of protein. However, in case of mutants, which do not show any loss of protein upon unfolding due to aggregation, but show gradual loss in residual activity upon incubation at temperatures above their melting points must have involved altered mechanism of inactivation. It is well known that the denatured state of protein is not completely devoid of structure lacking all interactions present in the native state (Dill and Shortle 1991; Shortle 1996). Several proteins are known to unfold in multi steps and partially unfolded species or intermediates do populate during the course of unfolding. Also, the denatured state is not completely unfolded but has significant amount of residual structure present which protein loses progressively upon further increasing the denaturing conditions (Shortle 1996). Closer the denaturing condition/temperature to that required for unfolding transition, more the residual structure present, suggesting the accumulation of partially unfolded species or intermediates during or just after the transition. This clearly explains that partially unfolded species, having significant amount of structure and are prone to aggregation, get accumulated during the unfolding of the mutants 5-C and 5-D. However, with further increase in temperature, these species further lose their structure, making them less prone to aggregation. This was further confirmed by monitoring bis-ANS binding to the denatured states of individual mutants at elevated temperatures. All mutants bind to bis-ANS at elevated temperatures, well above their melting temperatures, to a significant extent suggesting that the denatured species accumulated at that temperature has significant structure with exposed hydrophobic surface. At 75 °C, mutants bind to bis-ANS in different order, suggesting the presence of an intermediate having significant residual structure, which differs in the extent of exposed hydrophobic surface. This order closely matches with the extent of aggregation shown by each mutant at this temperature. However, at 85 °C, only 4D3 shows bis-ANS binding to significant levels indicating the presence of significantly exposed hydrophobic surfaces, all other mutants show only marginal binding to bis-ANS. This suggests that all mutants are completely denatured by 85 °C, but still have some residual structure with little hydrophobic surfaces (as bis-ANS does not bind to completely unfolded mutants in 6 M GdmCl, data not shown), thus having little propensity to aggregate.

Mutants such as 6-B, do not show any loss of protein due to aggregation upon unfolding but show a gradual loss in residual activity upon incubation at temperatures above its melting point ($>80\text{ }^{\circ}\text{C}$). This loss was due to chemical modification of protein, mainly due to deamidation of asparagine and glutamine residues (Daniel et al. 1996; Volkin et al. 1997). Mutant 6-B has 15 asparagine and 6 glutamine residues, many of which might be prone to deamidation at high temperatures, besides other covalent changes which are also possible. Modified protein fraction fails to achieve native state and it increases with the temperature of incubation as shown by intrinsic tryptophan fluorescence and bis-ANS binding of the refolded protein. Clear evidence of deamidation was isoelectric focusing (IEF) profile of the protein, incubated at elevated temperature, indicating the generation of a fraction of protein having pI values much lower than the native protein (Gianazza 1995).

The unfolding and inactivation behaviour shown by different mutants can be explained by the pathways shown in scheme 1. Native protein (N) denatures and unfolds upon heating and partially unfolded species or intermediates (I_n) accumulate at elevated temperatures ($70 - 85\text{ }^{\circ}\text{C}$). These intermediates are prone to aggregation, have significant residual structure as probed by bis-ANS binding. At $75\text{ }^{\circ}\text{C}$, the intermediate states of mutants (I_{75}) are prone to aggregation to different extent, depending on the kind of mutation involved. While mutant 5-C and 5-D shows considerable aggregation, other mutants such as 5-A, 5-B and 6-B hardly aggregate. This irreversible step of aggregation affects the reversibility of unfolding of these mutants to different extent. All the four mutations affect this irreversible step by altering the structure of the intermediate generated upon unfolding. Further increasing the temperature to $85\text{ }^{\circ}\text{C}$ or more, all mutants further unfold to acquire states (D) which is similar in case of all the mutants. This denatured state (D) does not have propensity to aggregate but it still is prone to chemical modifications of amino acid residues. Once modified, protein fails to regain the native state, thus making the whole pathway irreversible. Thus, the cause of irreversibility at $75\text{ }^{\circ}\text{C}$ is primarily

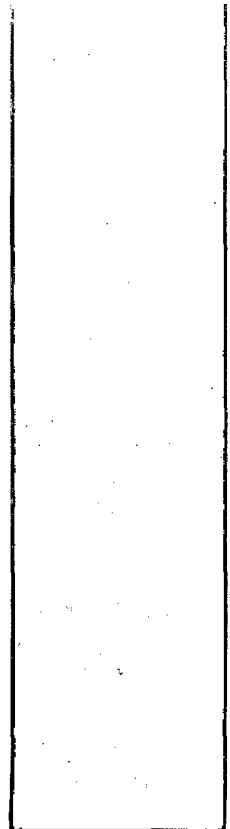


Scheme1: Scheme depicting the changes occurring to protein upon incubation at elevated temperatures.

aggregation while at temperatures more than 85 °C, it is the modification of the protein itself.

Bacillus subtilis lipase is only one of the two lipases known to have solvent exposed active-site due to the absence of lid (van et al. 2001). Being hydrophobic, this region might be involved in aggregation upon denaturation. Notably, all four mutations are in the vicinity of the active site. Two of the four mutations, which have such a profound effect on aggregation, M134E and M137P are located in the same loop, which is part of the binding region of active-site. Introduction of Glu 134 places a negative charge adjacent to Asp 133 and Asp 132, thus creating a cluster of three negative charges, which inhibits aggregation due to intermolecular electrostatic repulsion, which gets abolished in the presence of NaCl (1 M) due to shielding of charge. This observation was further confirmed by the identification of aspartate as the second best substitution at this position in site-saturation mutagenesis, suggesting that introduction of only negative charge brings about this effect by making a cluster of three negative charges and because of only that reason, no positive charge residue was selected as that would have neutralized the negative charge within its vicinity, thus reducing the overall charge of the region, required for effective intermolecular electrostatic repulsion of the denatured protein. On the other hand, effect of mutation M137P is independent of ionic strength of the medium, which blocks the aggregation by locking the conformation of denatured state, not allowing it to acquire conformation prone to aggregation. This indicates the importance of this loop and position 137 in particular in blocking the pathway of aggregation. Number of studies have been performed earlier in order to stabilize the shelf life of proteins both by *in vitro* evolution and knowledge based approaches by blocking the aggregation mainly by the introduction of charged residues in exposed, continuous hydrophobic stretches, which act as nucleation sites or by increasing the charge on the surface (Narhi et al. 2001; Strub et al. 2004). An extreme example is the engineering of super charged proteins having excess of either kind of charge on the protein surface (Lawrence et al. 2007). In contrast to these charge based approaches, inhibition of aggregation by other approaches, such as introduction of prolines have rarely been tried. Studies involving the identification of these substitutions will further help in engineering proteins of therapeutic and industrial importance with greater shelf-life.

Conclusion and future prospective



Conclusion and future perspective

Understanding the principles and mechanisms employed by proteins to improve thermostability has been a subject of great interest since long a time. This is not only to understand the basic principles involved, but also to utilize that knowledge for designing stable biocatalysts for industrial use. Much of our understanding on protein stability came from comparative sequence and structural analysis of homologous proteins from extremophiles, but interpretation of observations in those cases is often not straightforward. Directed or *in vitro* evolution of proteins, on the other hand, have provided a valuable tool to evolve a property of protein in carefully controlled manner and then to study the mechanisms employed by protein to achieve that improvement. Here, *in vitro* evolution techniques have been used to improve thermostability of a model protein, *Bacillus subtilis* lipase (LipA) to generate a set of mutants with 'graded' thermostability, in order to explore the structural solutions adopted by this protein to tolerate high temperature.

The strategy of *in vitro* evolution is based on repeated cycles of sequence randomization followed by screening for altered properties, thermostability in this case. A highly efficient and sensitive three-tier screening protocol, to identify thermostable variants was established. Lipase gene was expressed in *B. subtilis* to utilize its efficient expression and secretion system which makes the screening effort relatively easier. After two rounds of random mutagenesis and screening, nine thermostabilizing mutations were identified, which were recombined to generate a series of variants with incremental increase in thermostability. Solution studies to assess protein stability show that each mutation contributes positively in improving protein stability in an additive manner. With the incorporation of nine thermostabilizing mutations in mutant 4D3, melting temperature as well as temperature of half-inactivation got enhanced by more than 15 °C with a million fold reduction in the rate of inactivation. The improved thermostability was achieved with simultaneous improvement in catalytic efficiency at moderate temperatures. Compared to wild-type enzyme, all thermostable mutants show 3-5 fold increase in specific activity over a broad range of temperature along with shift in temperature optima by 20 °C. In order to understand the mechanism employed by protein and the structural changes responsible for increased thermostability, crystal structure of mutant 2D9 and 4D3 mutants was determined (Ahmad et al. 2008). Crystal structure of

thermostable mutants and wild-type protein was found to be highly similar, showing that only few additional interactions are responsible for the improved stability. The interactions identified to improve thermostability include replacement of solvent exposed hydrophobic side chain with polar group, helix stabilization by favorable interaction with dipole moment, additional hydrogen bonds (also water mediated) and better van der Waal interaction among amino acid residues on surface. Site-saturation mutagenesis on the six newly identified positions did not reveal any substitution which can further improve thermostability suggesting the crucial nature of the identified mutations which is hard to be substituted by any other change. All the six mutations identified in this study, as well as the three mutations of previous study, bring about their stabilization effect either by stabilizing loop regions by introducing additional interactions or by anchoring of loops with the rest of the molecule, suggesting these loops might be the potential sites from where unfolding starts at elevated temperatures, stabilization of which dramatically improved protein thermostability.

Random mutagenesis and screening for thermostability identified many positions at which mutation occurred but showed little improvement in stability. This might be due to the limitation of point mutagenesis. Moreover, the location of thermostabilizing mutations identified the loops around the active site region as the potential weak points. Based on these clues, four positions in two different loops in mutant 4D3 were selected and randomized by site-saturation mutagenesis. Mutations identified by site-saturation were found to be more thermostabilizing which upon recombination increases the melting temperature of the protein further by 7 °C. This was accompanied by further increase in temperature optima by 10 °C.

Bacillus subtilis lipase undergoes irreversible thermal inactivation due to aggregation of unfolded protein at elevated temperature. The four mutations identified in fifth generation by site-saturation mutagenesis not only improved protein thermostability but also improved reversibility of protein unfolding by blocking aggregation of unfolded protein at high temperature. Two mutations, M134E and M137P, were found to be very promising as the latter one alone is sufficient to completely block the aggregation of protein upon thermal unfolding.

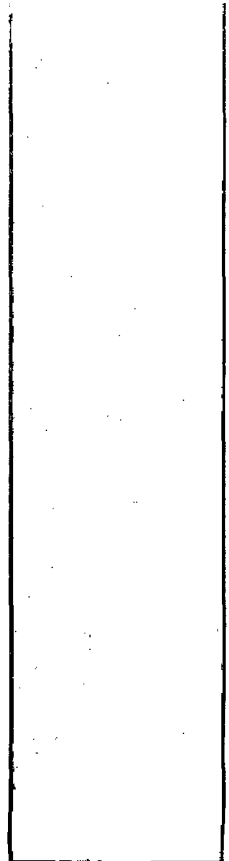
Random mutagenesis, mainly by error-prone PCR followed by screening for the desired trait, is the method of choice in most of the directed evolution exercises but this has serious limitations in its ability to search the available sequence space. Also, three rounds of random mutagenesis and screening performed in this study showed that most

of the positions at which a mutation positively contributes to thermostability lie on the loops. Based on this, site-saturation mutagenesis on all positions which are part of loop regions (90 positions) was performed. This exercise not only identified all the previous mutations identified through random mutagenesis, but also identified 15 additional positions, mutation at which the thermostability of the protein is improved. Although loops are the regions at which most of the thermostabilizing mutations were identified in this study (and also in a number of studies reported earlier), the contribution of secondary structure elements (both direct and indirect) in protein stability can not be neglected. In order to explore the role of these structural elements, site-saturation mutagenesis on rest of the positions could be performed.

Generation of a set of mutants of a model protein with graded improvement in thermostability is of great importance as it can be used to understand the mechanism adopted by the protein in achieving thermostability. Thermodynamic and protein folding kinetic analysis on this set of mutants can give important insights into the mechanisms adopted by protein for thermostabilization. Moreover, protein thermostability is also often related (positively or inversely) with other properties of interest, such as stability against other denaturants, organic solvents, protease susceptibility, cold adaptation etc. Study of this set of mutants for these properties may help us understand this correlation in a better manner.

The *in vitro* evolution strategy for protein engineering holds promise for the development of novel biocatalysts. The most important feature of this method is that it allows simultaneous improvement of multiple properties thus making it possible to make tailor-made enzymes for specific applications. Here, we have evolved *B. subtilis* lipase for thermostability along with improved catalytic efficiency demonstrating that these two properties are not inversely correlated. Since lipases are important biocatalysts, they can be evolved for multiple properties such as improved thermostability along with high activity in organic solvents, enhanced enantioselectivity, or improved activity towards a particular substrate.

References



- Abecassis,V., Pompon,D., and Truan,G. 2000. High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **28**:E88.
- Acharya,P. and Rao,N.M. 2002. Anomalous ester hydrolysis in mixed micelles of p-nitrophenyloleate-triton X-100 in the presence of guanidinium chloride: Implications in lipase assay. *Langmuir* **18**:3018-3026.
- Acharya,P. and Rao,N.M. 2003. Stability studies on a lipase from *Bacillus subtilis* in guanidinium chloride. *J Protein Chem.* **22**:51-60.
- Acharya,P., Rajakumara,E., Sankaranarayanan,R., and Rao,N.M. 2004. Structural basis of selection and thermostability of laboratory evolved *Bacillus subtilis* lipase. *J Mol Biol* **341**:1271-1281.
- Ahmad,S., Kamal,M.Z., Sankaranarayanan,R., and Rao,N.M. 2008. Thermostable *Bacillus subtilis* lipases: in vitro evolution and structural insight. *J Mol Biol* **381**:324-340.
- Akanuma,S., Yamagishi,A., Tanaka,N., and Oshima,T. 1999. Further improvement of the thermal stability of a partially stabilized *Bacillus subtilis* 3-isopropylmalate dehydrogenase variant by random and site-directed mutagenesis. *Eur. J Biochem.* **260**:499-504.
- Alonso,J.C., Luder,G., and Trautner,T.A. 1986. Requirements for the formation of plasmid-transducing particles of *Bacillus subtilis* bacteriophage SPP1. *EMBO J* **5**:3723-3728.
- Alsop,E., Silver,M., and Livesay,D.R. 2003. Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis. *Protein Eng* **16**:871-874.
- Amin,N., Liu,A.D., Ramer,S., Aehle,W., Meijer,D., Metin,M., Wong,S., Gualfetti,P., and Schellenberger,V. 2004. Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng Des Sel* **17**:787-793.
- Anagnostopoulos,C. and Spizizen,J. 1961. Requirements for transformation in *Bacillus subtilis*. *J Bacteriol.* **81**:741-746.
- Andrews,S.R., Taylor,E.J., Pell,G., Vincent,F., Ducros,V.M., Davies,G.J., Lakey,J.H., and Gilbert,H.J. 2004. The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca²⁺-independent stable xylanases. *J Biol Chem.* **279**:54369-54379.
- Arnold,F.H. 1993. Engineering proteins for nonnatural environments. *FASEB J* **7**:744-749.
- Arnold,F.H. 1998. When blind is better: protein design by evolution. *Nat Biotechnol.* **16**:617-618.
- Arnold,F.H. and Volkov,A.A. 1999. Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol* **3**:54-59.
- Arnold,F.H., Wintrode,P.L., Miyazaki,K., and Gershenson,A. 2001. How enzymes adapt: lessons from directed evolution. *Trends Biochem. Sci.* **26**:100-106.
- Arpigny,J.L. and Jaeger,K.E. 1999. Bacterial lipolytic enzymes: classification and properties. *Biochem. J* **343 Pt 1**:177-183.
- Aucamp,J.P., Cosme,A.M., Lye,G.J., and Dalby,P.A. 2005. High-throughput measurement of protein stability in microtiter plates. *Biotechnol. Bioeng.* **89**:599-607.
- Auerbach,G., Huber,R., Grattinger,M., Zaiss,K., Schurig,H., Jaenicke,R., and Jacob,U. 1997. Closed structure of phosphoglycerate kinase from *Thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability. *Structure.* **5**:1475-1483.
- Baker,E.N. and Hubbard,R.E. 1984. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol Biol* **44**:97-179.

- Becker,S., Schmoldt,H.U., Adams,T.M., Wilhelm,S., and Kolmar,H. 2004. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr. Opin. Biotechnol.* **15**:323-329.
- Becker,S., Theile,S., Heppeler,N., Michalczyk,A., Wentzel,A., Wilhelm,S., Jaeger,K.E., and Kolmar,H. 2005. A generic system for the *Escherichia coli* cell-surface display of lipolytic enzymes. *FEBS Lett.* **579**:1177-1182.
- Beisson,F., Tiss,A., Riviere,C., and Verger,R. 2000. Methods for lipase detection and assay: a critical review. *Eur. J. Lipid Sci. Technol.* 133-153.
- Bernath,K., Hai,M., Mastrobattista,E., Griffiths,A.D., Magdassi,S., and Tawfik,D.S. 2004. In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Anal. Biochem.* **325**:151-157.
- Bershtein,S. and Tawfik,D.S. 2008. Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol* **12**:151-158.
- Bessler,C., Schmitt,J., Maurer,K.H., and Schmid,R.D. 2003. Directed evolution of a bacterial alpha-amylase: toward enhanced pH-performance and higher specific activity. *Protein Sci.* **12**:2141-2149.
- Biek,D.P. and Cohen,S.N. 1986. Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *J Bacteriol.* **167**:594-603.
- Bittker,J.A., Le,B.V., Liu,J.M., and Liu,D.R. 2004. Directed evolution of protein enzymes using nonhomologous random recombination. *Proc. Natl. Acad. Sci. U. S. A* **101**:7011-7016.
- Bloom,J.D., Meyer,M.M., Meinhold,P., Otey,C.R., MacMillan,D., and Arnold,F.H. 2005. Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol* **15**:447-452.
- Boersma,Y.L., Droge,M.J., and Quax,W.J. 2007. Selection strategies for improved biocatalysts. *FEBS J* **274**:2181-2195.
- Bommarius,A.S., Broering,J.M., Chaparro-Riggers,J.F., and Polizzi,K.M. 2006. High-throughput screening for enhanced protein stability. *Curr. Opin. Biotechnol.* **17**:606-610.
- Brady,L., Brzozowski,A.M., Derewenda,Z.S., Dodson,E., Dodson,G., Tolley,S., Turkenburg,J.P., Christiansen,L., Huge-Jensen,B., Norskov,L., and . 1990. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Nature* **343**:767-770.
- Brouns,S.J., Wu,H., Akerboom,J., Turnbull,A.P., de Vos,W.M., and van der,O.J. 2005. Engineering a selectable marker for hyperthermophiles. *J Biol Chem.* **280**:11422-11431.
- Cadwell,R.C. and Joyce,G.F. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **2**:28-33.
- Cambillau,C. and Claverie,J.M. 2000. Structural and genomic correlates of hyperthermostability. *J Biol Chem.* **275**:32383-32386.
- Canosi,U., Morelli,G., and Trautner,T.A. 1978. The relationship between molecular structure and transformation efficiency of some *S. aureus* plasmids isolated from *B. subtilis*. *Mol Gen Genet* **166**:259-267.
- Cedrone,F., Bhatnagar,T., and Baratti,J.C. 2005. Colorimetric assays for quantitative analysis and screening of epoxide hydrolase activity. *Biotechnol. Lett.* **27**:1921-1927.
- Chakravarty,S. and Varadarajan,R. 2000. Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Lett.* **470**:65-69.

- Chang,S. and Cohen,S.N. 1979. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol Gen Genet* **168**:111-115.
- Chaput,J.C., Woodbury,N.W., Stearns,L.A., and Williams,B.A. 2008. Creating protein biocatalysts as tools for future industrial applications. *Expert. Opin. Biol Ther.* **8**:1087-1098.
- Chen,K. and Arnold,F.H. 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A* **90**:5618-5622.
- Cherry,J.R. and Fidantsef,A.L. 2003. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14**:438-443.
- Chi,E.Y., Krishnan,S., Randolph,T.W., and Carpenter,J.F. 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* **20**:1325-1336.
- Chica,R.A., Doucet,N., and Pelletier,J.N. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16**:378-384.
- Chirumamilla,R.R., Muralidhar,R., Marchant,R., and Nigam,P. 2001. Improving the quality of industrially important enzymes by directed evolution. *Mol Cell Biochem.* **224**:159-168.
- Chodorge,M., Fourage,L., Ullmann,C., Duvivier,V., Masson,J.M., and Lefevre,F. 2005. Rational strategies for directed evolution of biocatalysts-application to *Candida antarctica* lipase B (CALB). *Adv. Synth. Catal.* **347**:1022-1026.
- Coco,W.M. 2003. RACHITT: Gene family shuffling by Random Chimeragenesis on Transient Templates. *Methods Mol Biol* **231**:111-127.
- Coco,W.M., Levinson,W.E., Crist,M.J., Hektor,H.J., Darzins,A., Pienkos,P.T., Squires,C.H., and Monticello,D.J. 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol.* **19**:354-359.
- Cohen,N., Abramov,S., Dror,Y., and Freeman,A. 2001. In vitro enzyme evolution: the screening challenge of isolating the one in a million. *Trends Biotechnol.* **19**:507-510.
- Cramer,A., Raillard,S.A., Bermudez,E., and Stemmer,W.P. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**:288-291.
- Creighton,T.E. 1993. *Proteins:Structure and molecular properties*, 2nd ed. W. Freeman & Co.: New York.
- Cygler,M. and Schrag,J.D. 1997. Structure as basis for understanding interfacial properties of lipases. *Methods Enzymol.* **284**:3-27.
- Cygler,M., Hung,M.N., Wagner,J., and Matte,A. 2008. Bacterial structural genomics initiative: overview of methods and technologies applied to the process of structure determination. *Methods Mol Biol* **426**:537-559.
- Daniel,R.M. and Danson,M.J. 2001. Assaying activity and assessing thermostability of hyperthermophilic enzymes. *Methods Enzymol.* **334**:283-293.
- Daniel,R.M., Dines,M., and Petach,H.H. 1996. The denaturation and degradation of stable enzymes at high temperatures. *Biochem. J* **317 (Pt 1)**:1-11.
- Dartois,V., Baulard,A., Schanck,K., and Colson,C. 1992. Cloning, nucleotide sequence and expression in *Escherichia coli* of a lipase gene from *Bacillus subtilis* 168. *Biochim. Biophys. Acta* **1131**:253-260.
- Dartois,V., Coppee,J.Y., Colson,C., and Baulard,A. 1994. Genetic analysis and overexpression of lipolytic activity in *Bacillus subtilis*. *Appl Environ. Microbiol* **60**:1670-1673.

- De Bernardez,C.E., Schwarz,E., and Rudolph,R. 1999. Inhibition of aggregation side reactions during in vitro protein folding. *Methods Enzymol.* **309**:217-236.
- Derewenda,U., Brzozowski,A.M., Lawson,D.M., and Derewenda,Z.S. 1992. Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. *Biochemistry* **31**:1532-1541.
- DeSantis,G., Wong,K., Farwell,B., Chatman,K., Zhu,Z., Tomlinson,G., Huang,H., Tan,X., Bibbs,L., Chen,P., Kretz,K., and Burk,M.J. 2003. Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J Am. Chem. Soc.* **125**:11476-11477.
- Dill,K.A. 1990. Dominant forces in protein folding. *Biochemistry* **29**:7133-7155.
- Dill,K.A. and Shortle,D. 1991. Denatured states of proteins. *Annu. Rev Biochem.* **60**:795-825.
- Drummond,D.A., Iverson,B.L., Georgiou,G., and Arnold,F.H. 2005. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J Mol Biol* **350**:806-816.
- Dubnau,D. and vidoff-Abelson,R. 1971. Fate of transforming DNA following uptake by competent *Bacillus subtilis*. I. Formation and properties of the donor-recipient complex. *J Mol Biol* **56**:209-221.
- Dumon,C., Varvak,A., Wall,M.A., Flint,J.E., Lewis,R.J., Lakey,J.H., Morland,C., Luginbuhl,P., Healey,S., Todaro,T., DeSantis,G., Sun,M., Parra-Gessert,L., Tan,X., Weiner,D.P., and Gilbert,H.J. 2008. Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. *J Biol Chem.* **283**:22557-22564.
- Edgell,M.H., Sims,D.A., Pielak,G.J., and Yi,F. 2003. High-precision, high-throughput stability determinations facilitated by robotics and a semiautomated titrating fluorometer. *Biochemistry* **42**:7587-7593.
- edo-Ribeiro,S., Darimont,B., Sterner,R., and Huber,R. 1996. Small structural changes account for the high thermostability of [4Fe-4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. *Structure.* **4**:1291-1301.
- Eggert,T., Funke,S.A., Rao,N.M., Acharya,P., Krumm,H., Reetz,M.T., and Jaeger,K.E. 2005. Multiplex-PCR-based recombination as a novel high-fidelity method for directed evolution. *Chembiochem.* **6**:1062-1067.
- Eggert,T., Pencreac'h,G., Douchet,I., Verger,R., and Jaeger,K.E. 2000. A novel extracellular esterase from *Bacillus subtilis* and its conversion to a monoacylglycerol hydrolase. *Eur. J Biochem.* **267**:6459-6469.
- Eijsink,V.G., Bjork,A., Gaseidnes,S., Sirevag,R., Synstad,B., van den,B.B., and Vriend,G. 2004. Rational engineering of enzyme stability. *J Biotechnol.* **113**:105-120.
- Eijsink,V.G., Gaseidnes,S., Borchert,T.V., and van den,B.B. 2005. Directed evolution of enzyme stability. *Biomol. Eng* **22**:21-30.
- Facchiano,A.M., Colonna,G., and Ragone,R. 1998. Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. *Protein Eng* **11**:753-760.
- Famm,K. and Winter,G. 2006. Engineering aggregation-resistant proteins by directed evolution. *Protein Eng Des Sel* **19**:479-481.
- Famm,K., Hansen,L., Christ,D., and Winter,G. 2008. Thermodynamically stable aggregation-resistant antibody domains through directed evolution. *J Mol Biol* **376**:926-931.
- Farinas,E.T., Bulter,T., and Arnold,F.H. 2001. Directed enzyme evolution. *Curr. Opin. Biotechnol.* **12**:545-551.

- Fernandez-Gacio,A., Uguen,M., and Fastrez,J. 2003. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **21**:408-414.
- Fitter,J. and Heberle,J. 2000. Structural equilibrium fluctuations in mesophilic and thermophilic alpha-amylase. *Biophys. J* **79**:1629-1636.
- Fourage,L., Helbert,M., Nicolet,P., and Colas,B. 1999. Temperature dependence of the ultraviolet-visible spectra of ionized and un-ionized forms of nitrophenol: consequence for the determination of enzymatic activities using nitrophenyl derivatives-A warning. *Anal. Biochem.* **270**:184-185.
- Fox,R.J. and Huisman,G.W. 2008. Enzyme optimization: moving from blind evolution to statistical exploration of sequence-function space. *Trends Biotechnol.* **26**:132-138.
- Franken,S.M., Rozeboom,H.J., Kalk,K.H., and Dijkstra,B.W. 1991. Crystal structure of haloalkane dehalogenase: an enzyme to detoxify halogenated alkanes. *EMBO J* **10**:1297-1302.
- Gaseidnes,S., Synstad,B., Jia,X., Kjellesvik,H., Vriend,G., and Eijsink,V.G. 2003. Stabilization of a chitinase from *Serratia marcescens* by Gly-->Ala and Xxx-->Pro mutations. *Protein Eng* **16**:841-846.
- Geddie,M.L. and Matsumura,I. 2004. Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. *J Biol Chem.* **279**:26462-26468.
- Geddie,M.L., Rowe,L.A., Alexander,O.B., and Matsumura,I. 2004. High throughput microplate screens for directed protein evolution. *Methods Enzymol.* **388**:134-145.
- Ghadessy,F.J., Ong,J.L., and Holliger,P. 2001. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U. S. A* **98**:4552-4557.
- Ghaemmaghami,S., Fitzgerald,M.C., and Oas,T.G. 2000. A quantitative, high-throughput screen for protein stability. *Proc. Natl. Acad. Sci. U. S. A* **97**:8296-8301.
- Gianazza,E. 1995. Isoelectric focusing as a tool for the investigation of post-translational processing and chemical modifications of proteins. *J Chromatogr. A* **705**:67-87.
- Gill,I. and Valivety,R. 1997. Polyunsaturated fatty acids, Part 2: Biotransformations and biotechnological applications. *Trends Biotechnol.* **15**:470-478.
- Giver,L., Gershenson,A., Freskgard,P.O., and Arnold,F.H. 1998. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. U. S. A* **95**:12809-12813.
- Gonzalez-Blasco,G., Sanz-Aparicio,J., Gonzalez,B., Hermoso,J.A., and Polaina,J. 2000. Directed evolution of beta -glucosidase A from *Paenibacillus polymyxa* to thermal resistance. *J Biol Chem.* **275**:13708-13712.
- Gray,K.A., Richardson,T.H., Kretz,K., Short,J.M., Bartnek,F., Knowles,R., Kan,L., Swanson,P.E., and Robertson,D.E. 2001. Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.* **343**:607-617.
- Griffiths,A.D. and Tawfik,D.S. 2006. Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol.* **24**:395-402.
- Grochulski,P., Li,Y., Schrag,J.D., and Cygler,M. 1994. Two conformational states of *Candida rugosa* lipase. *Protein Sci.* **3**:82-91.
- Gromiha,M.M. 2007. Prediction of protein stability upon point mutations. *Biochem. Soc. Trans.* **35**:1569-1573.
- Gromiha,M.M., An,J., Kono,H., Oobatake,M., Uedaira,H., and Sarai,A. 1999a. ProTherm: Thermodynamic Database for Proteins and Mutants. *Nucleic Acids Res.* **27**:286-288.

- Gromiha,M.M., Oobatake,M., and Sarai,A. 1999b. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys. Chem.* **82**:51-67.
- Gupta,R., Gupta,N., and Rathi,P. 2004. Bacterial lipases: an overview of production, purification and biochemical properties. *Appl Microbiol Biotechnol.* **64**:763-781.
- Gupta,R., Rathi,P., Gupta,N., and Bradoo,S. 2003. Lipase assays for conventional and molecular screening: an overview. *Biotechnol. Appl Biochem.* **37**:63-71.
- Hakamada,Y., Hatada,Y., Ozawa,T., Ozaki,K., Kobayashi,T., and Ito,S. 2001. Identification of thermostabilizing residues in a *Bacillus* alkaline cellulase by construction of chimeras from mesophilic and thermostable enzymes and site-directed mutagenesis. *FEMS Microbiol Lett.* **195**:67-72.
- Hanes,J., Schaffitzel,C., Knappik,A., and Pluckthun,A. 2000. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat Biotechnol.* **18**:1287-1292.
- Haney,P.J., Badger,J.H., Buldak,G.L., Reich,C.I., Woese,C.R., and Olsen,G.J. 1999a. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci. U. S. A* **96**:3578-3583.
- Haney,P.J., Stees,M., and Konisky,J. 1999b. Analysis of thermal stabilizing interactions in mesophilic and thermophilic adenylate kinases from the genus *Methanococcus*. *J Biol Chem.* **274**:28453-28458.
- Hardy,F., Vriend,G., Veltman,O.R., van,d., V, Venema,G., and Eijsink,V.G. 1993. Stabilization of *Bacillus stearothermophilus* neutral protease by introduction of prolines. *FEBS Lett.* **317**:89-92.
- Harwood,C.R. 1992. *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol.* **10**:247-256.
- Harwood,C.R. and Cranenburgh,R. 2008. *Bacillus* protein secretion: an unfolding story. *Trends Microbiol* **16**:73-79.
- Hecky,J. and Muller,K.M. 2005. Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase. *Biochemistry* **44**:12640-12654.
- Heierson,A., Landen,R., Lovgren,A., Dalhammar,G., and Boman,H.G. 1987. Transformation of vegetative cells of *Bacillus thuringiensis* by plasmid DNA. *J Bacteriol.* **169**:1147-1152.
- Hennig,M., Sterner,R., Kirschner,K., and Jansonius,J.N. 1997. Crystal structure at 2.0 Å resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. *Biochemistry* **36**:6009-6016.
- Hernandez,G., Jenney,F.E., Jr., Adams,M.W., and LeMaster,D.M. 2000. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proc. Natl. Acad. Sci. U. S. A* **97**:3166-3170.
- Hirokawa,K., Ichianagi,A., and Kajiyama,N. 2008. Enhancement of thermostability of fungal deglycating enzymes by directed evolution. *Appl Microbiol Biotechnol.* **78**:775-781.
- Hjorth,A., Carriere,F., Cudrey,C., Woldike,H., Boel,E., Lawson,D.M., Ferrato,F., Cambillau,C., Dodson,G.G., Thim,L., and . 1993. A structural domain (the lid) found in pancreatic lipases is absent in the guinea pig (phospho)lipase. *Biochemistry* **32**:4702-4707.
- Hol,W.G. 1985. The role of the alpha-helix dipole in protein function and structure. *Prog. Biophys. Mol Biol* **45**:149-195.
- Ichikawa,J.K. and Clarke,S. 1998. A highly active protein repair enzyme from an extreme thermophile: the L-isospartyl methyltransferase from *Thermotoga maritima*. *Arch Biochem. Biophys.* **358**:222-231.

- Iffland,A., Gendreizig,S., Tafelmeyer,P., and Johnsson,K. 2001. Changing the substrate specificity of cytochrome c peroxidase using directed evolution. *Biochem. Biophys. Res. Commun.* **286**:126-132.
- Ishikawa,K., Nakamura,H., Morikawa,K., and Kanaya,S. 1993. Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* **32**:6171-6178.
- Jackel,C., Kast,P., and Hilvert,D. 2008. Protein design by directed evolution. *Annu. Rev Biophys.* **37**:153-173.
- Jaeger,K.E. and Eggert,T. 2002. Lipases for biotechnology. *Curr. Opin. Biotechnol.* **13**:390-397.
- Jaeger,K.E. and Reetz,M.T. 1998. Microbial lipases form versatile tools for biotechnology. *Trends Biotechnol.* **16**:396-403.
- Jaeger,K.E., Dijkstra,B.W., and Reetz,M.T. 1999. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. *Annu. Rev Microbiol* **53**:315-351.
- Jaeger,K.E., Ransac,S., Dijkstra,B.W., Colson,C., van,H.M., and Misset,O. 1994. Bacterial lipases. *FEMS Microbiol Rev* **15**:29-63.
- Jaenicke,R. 1991. Protein stability and molecular adaptation to extreme conditions. *Eur. J Biochem.* **202**:715-728.
- Jaenicke,R. 1996. How do proteins acquire their three-dimensional structure and stability? *Naturwissenschaften* **83**:544-554.
- Jaenicke,R. 2000a. Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci. U. S. A* **97**:2962-2964.
- Jaenicke,R. 2000b. Stability and stabilization of globular proteins in solution. *J Biotechnol.* **79**:193-203.
- Jaenicke,R. and Bohm,G. 1998. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol* **8**:738-748.
- Joern,J.M., Sakamoto,T., Arisawa,A., and Arnold,F.H. 2001. A versatile high throughput screen for dioxygenase activity using solid-phase digital imaging. *J Biomol. Screen.* **6**:219-223.
- Johannes,T.W. and Zhao,H. 2006. Directed evolution of enzymes and biosynthetic pathways. *Curr. Opin. Microbiol* **9**:261-267.
- Joo,H., Arisawa,A., Lin,Z., and Arnold,F.H. 1999. A high-throughput digital imaging screen for the discovery and directed evolution of oxygenases. *Chem. Biol* **6**:699-706.
- Karshikoff,A. and Ladenstein,R. 2001. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem. Sci.* **26**:550-556.
- Kauzmann,W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**:1-63.
- Kawasaki,K., Kondo,H., Suzuki,M., Ohgiya,S., and Tsuda,S. 2002. Alternate conformations observed in catalytic serine of *Bacillus subtilis* lipase determined at 1.3 Å resolution. *Acta Crystallogr. D. Biol Crystallogr.* **58**:1168-1174.
- Kelly,S.M. and Price,N.C. 1997. The application of circular dichroism to studies of protein folding and unfolding. *Biochim. Biophys. Acta* **1338**:161-185.
- Kelly,S.M. and Price,N.C. 2000. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **1**:349-384.

- Kennedy, M.B. and Lennarz, W.J. 1979. Characterization of the extracellular lipase of *Bacillus subtilis* and its relationship to a membrane-bound lipase found in a mutant strain. *J Biol Chem.* **254**:1080-1089.
- Kim, K.K., Song, H.K., Shin, D.H., Hwang, K.Y., and Suh, S.W. 1997b. The crystal structure of a triacylglycerol lipase from *Pseudomonas cepacia* reveals a highly open conformation in the absence of a bound inhibitor. *Structure.* **5**:173-185.
- Kim, K.K., Song, H.K., Shin, D.H., Hwang, K.Y., Choe, S., Yoo, O.J., and Suh, S.W. 1997a. Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an alpha/beta hydrolase with broad substrate specificity. *Structure.* **5**:1571-1584.
- Kim, M.S. and Lei, X.G. 2008. Enhancing thermostability of *Escherichia coli* phytase AppA2 by error-prone PCR. *Appl Microbiol Biotechnol.* **79**:69-75.
- Kim, Y.W., Choi, J.H., Kim, J.W., Park, C., Kim, J.W., Cha, H., Lee, S.B., Oh, B.H., Moon, T.W., and Park, K.H. 2003. Directed evolution of *Thermus maltogenic* amylase toward enhanced thermal resistance. *Appl Environ Microbiol* **69**:4866-4874.
- Kirino, H., Aoki, M., Aoshima, M., Hayashi, Y., Ohba, M., Yamagishi, A., Wakagi, T., and Oshima, T. 1994. Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus*. *Eur. J Biochem.* **220**:275-281.
- Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M., and Schwalbe, H. 2002. Long-range interactions within a nonnative protein. *Science* **295**:1719-1722.
- Kolkman, J.A. and Stemmer, W.P. 2001. Directed evolution of proteins by exon shuffling. *Nat Biotechnol.* **19**:423-428.
- Korkegian, A., Black, M.E., Baker, D., and Stoddard, B.L. 2005. Computational thermostabilization of an enzyme. *Science* **308**:857-860.
- Kouker, G. and Jaeger, K.E. 1987. Specific and sensitive plate assay for bacterial lipases. *Appl Environ Microbiol* **53**:211-213.
- Krag, S.S. and Lennarz, W.J. 1975. Purification and characterization of an inhibitor of phospholipase A1 in *Bacillus subtilis*. *J Biol Chem.* **250**:2813-2822.
- Kretz, K.A., Richardson, T.H., Gray, K.A., Robertson, D.E., Tan, X., and Short, J.M. 2004. Gene site saturation mutagenesis: a comprehensive mutagenesis approach. *Methods Enzymol.* **388**:3-11.
- Kuchner, O. and Arnold, F.H. 1997. Directed evolution of enzyme catalysts. *Trends Biotechnol.* **15**:523-530.
- Kuhlman, B. and Baker, D. 2000. Native protein sequences are close to optimal for their structures. *Proc Natl. Acad. Sci. U. S. A* **97**:10383-10388.
- Kumar, S. and Nussinov, R. 2001. How do thermophilic proteins deal with heat? *Cell Mol Life Sci.* **58**:1216-1233.
- Kumar, S., Tsai, C.J., and Nussinov, R. 2000. Factors enhancing protein thermostability. *Protein Eng* **13**:179-191.
- Kurtzman, A.L., Govindarajan, S., Vahle, K., Jones, J.T., Heinrichs, V., and Patten, P.A. 2001. Advances in directed protein evolution by recursive genetic recombination: applications to therapeutic proteins. *Curr Opin. Biotechnol.* **12**:361-370.
- Kusaoke, H., Hayashi, Y., Kadowaki, Y., and Kimoto, H. 1989. Optimum conditions for electric pulse mediated gene transfer to *Bacillus subtilis* cells. *Agric. Biol. Chem.* **53**:2442-2446.

- Lang,D., Hofmann,B., Haalck,L., Hecht,H.J., Spener,F., Schmid,R.D., and Schomburg,D. 1996. Crystal structure of a bacterial lipase from *Chromobacterium viscosum* ATCC 6918 refined at 1.6 angstroms resolution. *J Mol Biol* **259**:704-717.
- Langhorst,U., Backmann,J., Loris,R., and Steyaert,J. 2000. Analysis of a water mediated protein-protein interactions within RNase T1. *Biochemistry* **39**:6586-6593.
- Lawrence,M.S., Phillips,K.J., and Liu,D.R. 2007. Supercharging proteins can impart unusual resilience. *J Am. Chem. Soc.* **129**:10110-10112.
- Lawson,D.M., Derewenda,U., Serre,L., Ferri,S., Szittner,R., Wei,Y., Meighen,E.A., and Derewenda,Z.S. 1994. Structure of a myristoyl-ACP-specific thioesterase from *Vibrio harveyi*. *Biochemistry* **33**:9382-9388.
- Lazaridis,T., Lee,I., and Karplus,M. 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **6**:2589-2605.
- Leemhuis,H., Stein,V., Griffiths,A.D., and Hollfelder,F. 2005. New genotype-phenotype linkages for directed evolution of functional proteins. *Curr. Opin. Struct. Biol* **15**:472-478.
- Lehmann,M. and Wyss,M. 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.* **12**:371-375.
- Lehmann,M., Loch,C., Middendorf,A., Studer,D., Lassen,S.F., Pasamontes,L., van Loon,A.P., and Wyss,M. 2002. The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng* **15**:403-411.
- Leroy,E., Bense,N., and Reymond,J.L. 2003. A low background high-throughput screening (HTS) fluorescence assay for lipases and esterases using acyloxymethylethers of umbelliferone. *Bioorg. Med. Chem. Lett.* **13**:2105-2108.
- Lesuisse,E., Schanck,K., and Colson,C. 1993. Purification and preliminary characterization of the extracellular lipase of *Bacillus subtilis* 168, an extremely basic pH-tolerant enzyme. *Eur. J Biochem.* **216**:155-160.
- Leung,D.W., Chen,E., and Goeddel,D.V. 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**:11-15.
- Li,W.F., Zhou,X.X., and Lu,P. 2005. Structural features of thermozyms. *Biotechnol. Adv.* **23**:271-281.
- Lim,J.H., Yu,Y.G., Han,Y.S., Cho,S., Ahn,B.Y., Kim,S.H., and Cho,Y. 1997. The crystal structure of an Fe-superoxide dismutase from the hyperthermophile *Aquifex pyrophilus* at 1.9 Å resolution: structural basis for thermostability. *J Mol Biol* **270**:259-274.
- Lipovsek,D. and Pluckthun,A. 2004. In-vitro protein evolution by ribosome display and mRNA display. *J Immunol. Methods* **290**:51-67.
- Luke,K.A., Higgins,C.L., and Wittung-Stafshede,P. 2007. Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J* **274**:4023-4033.
- Machius,M., Declerck,N., Huber,R., and Wiegand,G. 2003. Kinetic stabilization of *Bacillus licheniformis* alpha-amylase through introduction of hydrophobic residues at the surface. *J Biol Chem.* **278**:11546-11553.
- Malakauskas,S.M. and Mayo,S.L. 1998. Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct. Biol* **5**:470-475.
- Markwell,M.A.K., Hass,S.M., Tolbert,N.E., and Bieber,L.L. Protein determination in membrane and lipoprotein samples: Manual and automated procedures. *Methods Enzymol.* **72**, 296-303. 1981.

- Martin,A., Sieber,V., and Schmid,F.X. 2001. In-vitro selection of highly stabilized protein variants with optimized surface. *J Mol Biol* **309**:717-726.
- Martinez,C., de,G.P., Lauwereys,M., Matthysens,G., and Cambillau,C. 1992. *Fusarium solani* cutinase is a lipolytic enzyme with a catalytic serine accessible to solvent. *Nature* **356**:615-618.
- Masson,L., Prefontaine,G., and Brousseau,R. 1989. Transformation of *Bacillus thuringiensis* vegetative cells by electroporation. *FEMS Microbiol Lett.* **51**:273-277.
- Matsumura,I., Wallingford,J.B., Surana,N.K., Vize,P.D., and Ellington,A.D. 1999. Directed evolution of the surface chemistry of the reporter enzyme beta-glucuronidase. *Nat Biotechnol.* **17**:696-701.
- Matsumura,M., Signor,G., and Matthews,B.W. 1989. Substantial increase of protein stability by multiple disulphide bonds. *Nature* **342**:291-293.
- Matsuno,Y., Hiraoka,H., Ano,T., and Shoda,M. 1990. Plasmid transformation in *Bacillus subtilis* NB22, an antifungal-antibiotic iturin producer. *FEMS Microbiol Lett.* **55**:227-229.
- Matsuura,T. and Yomo,T. 2006. In vitro evolution of proteins. *J Biosci. Bioeng.* **101**:449-456.
- Matthews,B.W., Nicholson,H., and Becktel,W.J. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. U. S. A* **84**:6663-6667.
- Matthews,B.W., Weaver,L.H., and Kester,W.R. 1974. The conformation of thermolysin. *J Biol Chem.* **249**:8030-8044.
- McLachlan,M.J., Johannes,T.W., and Zhao,H. 2008. Further improvement of phosphite dehydrogenase thermostability by saturation mutagenesis. *Biotechnol. Bioeng.* **99**:268-274.
- Minagawa,H., Yoshida,Y., Kenmochi,N., Furuichi,M., Shimada,J., and Kaneko,H. 2007. Improving the thermal stability of lactate oxidase by directed evolution. *Cell Mol Life Sci.* **64**:77-81.
- Miyazaki,K. and Arnold,F.H. 1999. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol.* **49**:716-720.
- Miyazaki,K., Takenouchi,M., Kondo,H., Noro,N., Suzuki,M., and Tsuda,S. 2006. Thermal stabilization of *Bacillus subtilis* family-11 xylanase by directed evolution. *J Biol Chem.* **281**:10236-10242.
- Miyazaki,K., Wintrode,P.L., Grayling,R.A., Rubingh,D.N., and Arnold,F.H. 2000. Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* **297**:1015-1026.
- Moore,J.C. and Arnold,F.H. 1996. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat Biotechnol.* **14**:458-467.
- Mormeneo,M., Andres,I., Bofill,C., Diaz,P., and Zueco,J. 2008. Efficient secretion of *Bacillus subtilis* lipase A in *Saccharomyces cerevisiae* by translational fusion to the Pir4 cell wall protein. *Appl Microbiol Biotechnol.* **80**:437-445.
- Muraoka,T., Ando,T., and Okuda,H. 1982. Purification and properties of a novel lipase from *Staphylococcus aureus* 226. *J Biochem.* **92**:1933-1939.
- Naglich,J.G. and Andrews,R.E., Jr. 1988. Tn916-dependent conjugal transfer of pC194 and pUB110 from *Bacillus subtilis* into *Bacillus thuringiensis* subsp. israelensis. *Plasmid* **20**:113-126.
- Narhi,L.O., Arakawa,T., Aoki,K., Wen,J., Elliott,S., Boone,T., and Cheetham,J. 2001. Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of *Escherichia coli*-derived erythropoietin. *Protein Eng* **14**:135-140.

- O'Farrell,P.H. 1975. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem.* **250**:4007-4021.
- Oh,K.H., Nam,S.H., and Kim,H.S. 2002. Improvement of oxidative and thermostability of N-carbamyl-d-amino acid amidohydrolase by directed evolution. *Protein Eng* **15**:689-695.
- Ohse,M., Kawade,K., and Kusaoke,H. 1997. Effects of DNA topology on transformation efficiency of *Bacillus subtilis* ISW1214 by electroporation. *Biosci. Biotechnol. Biochem.* **61**:1019-1021.
- Ohse,M., Takahashi,K., Kadowaki,Y., and Kusaoke,H. 1995. Effects of plasmid DNA sizes and several other factors on transformation of *Bacillus subtilis* ISW1214 with plasmid DNA by electroporation. *Biosci. Biotechnol. Biochem.* **59**:1433-1437.
- Ollis,D.L., Cheah,E., Cygler,M., Dijkstra,B., Frolow,F., Franken,S.M., Harel,M., Remington,S.J., Silman,I., Schrag,J., and . 1992. The alpha/beta hydrolase fold. *Protein Eng* **5**:197-211.
- Olsen,M., Iverson,B., and Georgiou,G. 2000. High-throughput screening of enzyme libraries. *Curr. Opin. Biotechnol.* **11**:331-337.
- Ostermeier,M., Shim,J.H., and Benkovic,S.J. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat Biotechnol.* **17**:1205-1209.
- Otten,L.G. and Quax,W.J. 2005. Directed evolution: selecting today's biocatalysts. *Biomol. Eng* **22**:1-9.
- Pace,C.N. 1975. The stability of globular proteins. *CRC Crit Rev Biochem.* **3**:1-43.
- Pace,C.N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**:266-280.
- Pace,C.N. 1992. Contribution of the hydrophobic effect to globular protein stability. *J Mol Biol* **226**:29-35.
- Pace,C.N. and Scholtz,J.M. 1997. Measuring the conformational stability of a protein. In *Protein structure-a practical approach.* (ed. TE Creighton), pp 299-321. Oxford University Press,Oxford: New York.
- Pace,C.N., Shirley,B.A., and Thompson,J.A. Mechanism of protein folding. In *Protein Structure: A Practical Approach* (Creighton.,ed). 311-330. 1989. Oxford Press, Oxford.
- Pace,C.N., Shirley,B.A., McNutt,M., and Gajiwala,K. 1996. Forces contributing to the conformational stability of proteins. *FASEB J* **10**:75-83.
- Paiardini,A., Sali,R., Bossa,F., and Pascarella,S. 2008. "Hot cores" in proteins: comparative analysis of the apolar contact area in structures from hyper/thermophilic and mesophilic organisms. *BMC. Struct. Biol* **8**:14.
- Pal,D. and Chakrabarti,P. 1999. Estimates of the loss of main-chain conformational entropy of different residues on protein folding. *Proteins* **36**:332-339.
- Palackal,N., Brennan,Y., Callen,W.N., Dupree,P., Frey,G., Goubet,F., Hazlewood,G.P., Healey,S., Kang,Y.E., Kretz,K.A., Lee,E., Tan,X., Tomlinson,G.L., Verruto,J., Wong,V.W., Mathur,E.J., Short,J.M., Robertson,D.E., and Steer,B.A. 2004. An evolutionary route to xylanase process fitness. *Protein Sci.* **13**:494-503.
- Panasik,N., Brenchley,J.E., and Farber,G.K. 2000. Distributions of structural features contributing to thermostability in mesophilic and thermophilic alpha/beta barrel glycosyl hydrolases. *Biochim. Biophys. Acta* **1543**:189-201.
- Parikh,M.R. and Matsumura,I. 2005. Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of beta-fucosidase from beta-galactosidase. *J Mol Biol* **352**:621-628.

- Perutz,M.F. and Raidt,H. 1975. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**:256-259.
- Petrounia,I.P. and Arnold,F.H. 2000. Designed evolution of enzymatic properties. *Curr. Opin. Biotechnol.* **11**:325-330.
- Petsko,G.A. 2001. Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol.* **334**:469-478.
- Plaza,d.P., I, Ibarra-Molero,B., and Sanchez-Ruiz,J.M. 2000. Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins* **40**:58-70.
- Pluckthun,A., Schaffitzel,C., Hanes,J., and Jermutus,L. 2000. In vitro selection and evolution of proteins. *Adv. Protein Chem.* **55**:367-403.
- Pohn,B., Gerlach,J., Scheideler,M., Katz,H., Uray,M., Bischof,H., Klimant,I., and Schwab,H. 2007. Micro-colony array based high throughput platform for enzyme library screening. *J Biotechnol.* **129**:162-170.
- Polizzi,K.M., Chaparro-Riggers,J.F., Vazquez-Figueroa,E., and Bommarius,A.S. 2006a. Structure-guided consensus approach to create a more thermostable penicillin G acylase. *Biotechnol. J* **1**:531-536.
- Polizzi,K.M., Parikh,M., Spencer,C.U., Matsumura,I., Lee,J.H., Realff,M.J., and Bommarius,A.S. 2006b. Pooling for improved screening of combinatorial libraries for directed evolution. *Biotechnol. Prog.* **22**:961-967.
- Rajakumara,E., Acharya,P., Ahmad,S., Sankaranaryanan,R., and Rao,N.M. 2007. Structural basis for the remarkable stability of *Bacillus subtilis* lipase (Lip A) at low pH. *Biochim. Biophys. Acta.*
- Rajakumara,E., Acharya,P., Ahmad,S., Shanmugam,V.M., Rao,N.M., and Sankaranarayanan,R. 2004. Crystallization and preliminary X-ray crystallographic investigations on several thermostable forms of a *Bacillus subtilis* lipase. *Acta Crystallogr. D. Biol Crystallogr.* **60**:160-162.
- Razvi,A. and Scholtz,J.M. 2006. Lessons in stability from thermophilic proteins. *Protein Sci.* **15**:1569-1578.
- Reddy,M. and Gowrishankar,J. 1997. Identification and characterization of *ssb* and *uup* mutants with increased frequency of precise excision of transposon Tn10 derivatives: nucleotide sequence of *uup* in *Escherichia coli*. *J Bacteriol.* **179**:2892-2899.
- Reetz,M.T. 2001. Combinatorial and Evolution-Based Methods in the Creation of Enantioselective Catalysts. *Angew. Chem. Int. Ed Engl.* **40**:284-310.
- Reetz,M.T. 2004. Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. *Proc. Natl. Acad. Sci. U. S. A* **101**:5716-5722.
- Reetz,M.T. and Carballeira,J.D. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc.* **2**:891-903.
- Reetz,M.T. and Jaeger,K.E. 1998. Overexpression, immobilization and biotechnological application of *Pseudomonas* lipases. *Chem. Phys. Lipids* **93**:3-14.
- Reetz,M.T., Carballeira,J.D., and Vogel,A. 2006. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed Engl.* **45**:7745-7751.
- Ren,B., Tibbelin,G., de,P.D., Rossi,M., Bartolucci,S., and Ladenstein,R. 1998. A protein disulfide oxidoreductase from the archaeon *Pyrococcus furiosus* contains two thioredoxin fold units. *Nat Struct. Biol* **5**:602-611.

- Renugopalakrishnan,V., Garduno-Juarez,R., Narasimhan,G., Verma,C.S., Wei,X., and Li,P. 2005. Rational design of thermally stable proteins: relevance to bionanotechnology. *J Nanosci. Nanotechnol.* **5**:1759-1767.
- Ruller,R., Deliberto,L., Ferreira,T.L., and Ward,R.J. 2008. Thermostable variants of the recombinant xylanase A from *Bacillus subtilis* produced by directed evolution show reduced heat capacity changes. *Proteins* **70**:1280-1293.
- Sadaie,Y. and Kada,T. 1983. Formation of competent *Bacillus subtilis* cells. *J Bacteriol.* **153**:813-821.
- Sadeghi,M., Naderi-Manesh,H., Zarrabi,M., and Ranjbar,B. 2006. Effective factors in thermostability of thermophilic proteins. *Biophys. Chem.* **119**:256-270.
- Sambrook,J. and Russel,D.W. 2001. *Molecular Cloning: A Laboratory Manual*, 3 ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
- Sanchez,M., Prim,N., Randez-Gil,F., Pastor,F.I., and Diaz,P. 2002. Engineering of baker's yeasts, *E. coli* and *Bacillus* hosts for the production of *Bacillus subtilis* Lipase A. *Biotechnol. Bioeng.* **78**:339-345.
- Sanchez-Ruiz,J.M. and Makhatadze,G.I. 2001. To charge or not to charge? *Trends Biotechnol.* **19**:132-135.
- Saraboji,K., Gromiha,M.M., and Ponnuswamy,M.N. 2006. Average assignment method for predicting the stability of protein mutants. *Biopolymers* **82**:80-92.
- Saunders,J.R. and Saunders,V.A. 1999. Introduction of DNA into bacteria. In *Methods in Microbiology*. (eds. CMS Margeret and ES R), pp 3-52. Academic Press.
- Saven,J.G. 2002. Combinatorial protein design. *Curr. Opin. Struct. Biol* **12**:453-458.
- Scandurra,R., Consalvi,V., Chiaraluca,R., Politi,L., and Engel,P.C. 1998. Protein thermostability in extremophiles. *Biochimie* **80**:933-941.
- Schmidt,M. and Bornscheuer,U.T. 2005. High-throughput assays for lipases and esterases. *Biomol. Eng* **22**:51-56.
- Sen,S., Venkata,D., V, and Mandal,B. 2007. Developments in directed evolution for improving enzyme functions. *Appl Biochem. Biotechnol.* **143**:212-223.
- Serrano,L. and Fersht,A.R. 1989. Capping and alpha-helix stability. *Nature* **342**:296-299.
- Serrano,L., Horovitz,A., Avron,B., Bycroft,M., and Fersht,A.R. 1990. Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **29**:9343-9352.
- Serrano,L., Kellis,J.T., Jr., Cann,P., Matouschek,A., and Fersht,A.R. 1992. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* **224**:783-804.
- Shafikhani,S., Siegel,R.A., Ferrari,E., and Schellenberger,V. 1997. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques* **23**:304-310.
- Sharma,K.K., Kaur,H., Kumar,G.S., and Kester,K. 1998. Interaction of 1,1'-bi(4-anilino)naphthalene-5,5'-disulfonic acid with alpha-crystallin. *J Biol Chem.* **273**:8965-8970.
- Shirley,B.A., Stanssens,P., Hahn,U., and Pace,C.N. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31**:725-732.
- Shortle,D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* **10**:27-34.

- Sieber,V., Martinez,C.A., and Arnold,F.H. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat Biotechnol.* **19**:456-460.
- Sieber,V., Pluckthun,A., and Schmid,F.X. 1998. Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol.* **16**:955-960.
- Smoot,A.L., Panda,M., Brazil,B.T., Buckle,A.M., Fersht,A.R., and Horowitz,P.M. 2001. The binding of bis-ANS to the isolated GroEL apical domain fragment induces the formation of a folding intermediate with increased hydrophobic surface not observed in tetradecameric GroEL. *Biochemistry* **40**:4484-4492.
- Song,J.K. and Rhee,J.S. 2001. Enhancement of stability and activity of phospholipase A(1) in organic solvents by directed evolution. *Biochim. Biophys. Acta* **1547**:370-378.
- Spiller,B., Gershenson,A., Arnold,F.H., and Stevens,R.C. 1999. A structural view of evolutionary divergence. *Proc. Natl. Acad. Sci. U. S. A* **96**:12305-12310.
- Spizizen,J. 1958. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *Proc. Natl. Acad. Sci. U. S. A* **44**:1072-1078.
- Sriprapundh,D., Vieille,C., and Zeikus,J.G. 2000. Molecular determinants of xylose isomerase thermal stability and activity: analysis of thermozymes by site-directed mutagenesis. *Protein Eng* **13**:259-265.
- Stemmer,W.P. 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A* **91**:10747-10751.
- Stemmer,W.P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**:389-391.
- Sterner,R. and Liebl,W. 2001. Thermophilic adaptation of proteins. *Crit Rev Biochem. Mol Biol* **36**:39-106.
- Sterner,R., Kleemann,G.R., Szadkowski,H., Lustig,A., Hennig,M., and Kirschner,K. 1996. Phosphoribosyl anthranilate isomerase from *Thermotoga maritima* is an extremely stable and active homodimer. *Protein Sci.* **5**:2000-2008.
- Strausberg,S.L., Ruan,B., Fisher,K.E., Alexander,P.A., and Bryan,P.N. 2005. Directed coevolution of stability and catalytic activity in calcium-free subtilisin. *Biochemistry* **44**:3272-3279.
- Strub,C., Alies,C., Lougarre,A., Ladurantie,C., Czaplicki,J., and Fournier,D. 2004. Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC. Biochem.* **5**:9.
- Sullivan,M.A., Yasbin,R.E., and Young,F.E. 1984. New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments. *Gene* **29**:21-26.
- Suvd,D., Fujimoto,Z., Takase,K., Matsumura,M., and Mizuno,H. 2001. Crystal structure of *Bacillus stearothermophilus* alpha-amylase: possible factors determining the thermostability. *J Biochem.* **129**:461-468.
- Szilagy, A. and Zavodszky,P. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure.* **8**:493-504.
- Takahashi,T.T., Austin,R.J., and Roberts,R.W. 2003. mRNA display: ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.* **28**:159-165.
- Tamakoshi,M., Nakano,Y., Kakizawa,S., Yamagishi,A., and Oshima,T. 2001. Selection of stabilized 3-isopropylmalate dehydrogenase of *Saccharomyces cerevisiae* using the host-vector system of an extreme thermophile, *Thermus thermophilus*. *Extremophiles.* **5**:17-22.

- Tanner,J.J., Hecht,R.M., and Krause,K.L. 1996. Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Angstroms Resolution. *Biochemistry* **35**:2597-2609.
- Tomazic,S.J. and Klibanov,A.M. 1988. Mechanisms of irreversible thermal inactivation of *Bacillus* alpha-amylases. *J Biol Chem.* **263**:3086-3091.
- Turner,N.J. 2003. Directed evolution of enzymes for applied biocatalysis. *Trends Biotechnol.* **21**:474-478.
- Uchiyama,H., Inaoka,T., Ohkuma-Soyejima,T., Togame,H., Shibana,Y., Yoshimoto,T., and Kokubo,T. 2000. Directed evolution to improve the thermostability of prolyl endopeptidase. *J Biochem. (Tokyo)* **128**:441-447.
- Uppenberg,J., Ohrner,N., Norin,M., Hult,K., Kleywegt,G.J., Patkar,S., Waagen,V., Anthonsen,T., and Jones,T.A. 1995. Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. *Biochemistry* **34**:16838-16851.
- van den,B.B., Vriend,G., Veltman,O.R., Venema,G., and Eijsink,V.G. 1998. Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci. U. S. A* **95**:2056-2060.
- van,P.G., Eggert,T., Jaeger,K.E., and Dijkstra,B.W. 2001. The crystal structure of *Bacillus subtilis* lipase: a minimal alpha/beta hydrolase fold enzyme. *J Mol Biol* **309**:215-226.
- van,T.H., Egloff,M.P., Martinez,C., Rugani,N., Verger,R., and Cambillau,C. 1993. Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography. *Nature* **362**:814-820.
- Varley,P.G. and Pain,R.H. 1991. Relation between stability, dynamics and enzyme activity in 3-phosphoglycerate kinases from yeast and *Thermus thermophilus*. *J Mol Biol* **220**:531-538.
- Vehmaanpera,J. 1989. Transformation of *Bacillus amyloliquefaciens* by electroporation. *FEMS Microbiol Lett.* **52**:165-169.
- Verger,R. 1997. Interfacial activation' of lipases: facts and artifacts. *Trends Biotechnol.* **15**:32-38.
- Vieille,C. and Zeikus,G.J. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* **65**:1-43.
- Vieille,C., Burdette,D.S., and Zeikus,J.G. 1996. Thermozyms. *Biotechnol. Annu. Rev* **2**:1-83.
- Vieille,C., Epting,K.L., Kelly,R.M., and Zeikus,J.G. 2001. Bivalent cations and amino-acid composition contribute to the thermostability of *Bacillus licheniformis* xylose isomerase. *Eur. J Biochem.* **268**:6291-6301.
- Vogt,G., Woell,S., and Argos,P. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* **269**:631-643.
- Volkin,D.B., Mach,H., and Middaugh,C.R. 1997. Degradative covalent reactions important to protein stability. *Mol Biotechnol.* **8**:105-122.
- Vriend,G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph.* **8**:52-6, 29.
- Waldburger,C.D., Schildbach,J.F., and Sauer,R.T. 1995. Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct. Biol* **2**:122-128.
- Watanabe,K., Kitamura,K., and Suzuki,Y. 1996. Analysis of the critical sites for protein thermostabilization by proline substitution in oligo-1,6-glucosidase from *Bacillus coagulans* ATCC 7050 and the evolutionary consideration of proline residues. *Appl Environ. Microbiol* **62**:2066-2073.

- Watanabe,K., Masuda,T., Ohashi,H., Mihara,H., and Suzuki,Y. 1994. Multiple proline substitutions cumulatively thermostabilize *Bacillus cereus* ATCC7064 oligo-1,6-glycosidase. Irrefragable proof supporting the proline rule. *Eur. J Biochem.* **226**:277-283.
- Wilson,D.S., Keefe,A.D., and Szostak,J.W. 2001. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U. S. A* **98**:3750-3755.
- Winkler,F.K., D'Arcy,A., and Hunziker,W. 1990. Structure of human pancreatic lipase. *Nature* **343**:771-774.
- Wintrode,P.L. and Arnold,F.H. 2000. Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv. Protein Chem.* **55**:161-225.
- Wintrode,P.L., Miyazaki,K., and Arnold,F.H. 2001. Patterns of adaptation in a laboratory evolved thermophilic enzyme. *Biochim. Biophys. Acta* **1549**:1-8.
- Wintrode,P.L., Zhang,D., Vaidehi,N., Arnold,F.H., and Goddard,W.A., III 2003. Protein dynamics in a family of laboratory evolved thermophilic enzymes. *J Mol Biol* **327**:745-757.
- Wong,D.W., Batt,S.B., Lee,C.C., and Robertson,G.H. 2004. High-activity barley alpha-amylase by directed evolution. *Protein J* **23**:453-460.
- Wong,T.S., Roccatano,D., and Schwaneberg,U. 2007. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environ. Microbiol* **9**:2645-2659.
- Woycechowsky,K.J., Vamvaca,K., and Hilvert,D. 2007. Novel enzymes through design and evolution. *Adv. Enzymol. Relat Areas Mol Biol* **75**:241-94, xiii.
- Wunderlich,M. and Schmid,F.X. 2006. In vitro evolution of a hyperstable G-beta 1 variant. *J Mol Biol* **363**:545-557.
- Wunderlich,M., Martin,A., and Schmid,F.X. 2005a. Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of Coulombic interactions. *J Mol Biol* **347**:1063-1076.
- Wunderlich,M., Martin,A., Staab,C.A., and Schmid,F.X. 2005b. Evolutionary protein stabilization in comparison with computational design. *J Mol Biol* **351**:1160-1168.
- Xie,H., Flint,J., Vardakou,M., Lakey,J.H., Lewis,R.J., Gilbert,H.J., and Dumon,C. 2006. Probing the structural basis for the difference in thermostability displayed by family 10 xylanases. *J Mol Biol* **360**:157-167.
- Xiong,A.S., Peng,R.H., Cheng,Z.M., Li,Y., Liu,J.G., Zhuang,J., Gao,F., Xu,F., Qiao,Y.S., Zhang,Z., Chen,J.M., and Yao,Q.H. 2007. Concurrent mutations in six amino acids in beta-glucuronidase improve its thermostability. *Protein Eng Des Sel* **20**:319-325.
- Xu,Z., Liu,Y., Yang,Y., Jiang,W., Arnold,E., and Ding,J. 2003. Crystal structure of D-Hydantoinase from *Burkholderia pickettii* at a resolution of 2.7 Angstroms: insights into the molecular basis of enzyme thermostability. *J Bacteriol.* **185**:4038-4049.
- Xue,G.P., Jhonson,J.S., and Dalrymple,B.P. 1999. High osmolarity improves the electro-transformation efficiency of the gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis*. *J Microbiol Methods* **34**:183-191.
- Yan,X. and Xu,Z. 2006. Ribosome-display technology: applications for directed evolution of functional proteins. *Drug Discov. Today* **11**:911-916.
- Yip,K.S., Britton,K.L., Stillman,T.J., Lebbink,J., de Vos,W.M., Robb,F.T., Vetriani,C., Maeder,D., and Rice,D.W. 1998. Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur. J Biochem.* **255**:336-346.

Yip,K.S., Stillman,T.J., Britton,K.L., Artymiuk,P.J., Baker,P.J., Sedelnikova,S.E., Engel,P.C., Pasquo,A., Chiaraluce,R., and Consalvi,V. 1995. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*. **3**:1147-1158.

Yuan,L., Kurek,I., English,J., and Keenan,R. 2005. Laboratory-directed protein evolution. *Microbiol Mol Biol Rev* **69**:373-392.

Yuen,C.M. and Liu,D.R. 2007. Dissecting protein structure and function using directed evolution. *Nat Methods* **4**:995-997.

Zavodszky,P., Kardos,J., Svingor, and Petsko,G.A. 1998. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. U. S. A* **95**:7406-7411.

Zhang,N., Suen,W.C., Windsor,W., Xiao,L., Madison,V., and Zaks,A. 2003. Improving tolerance of *Candida antarctica* lipase B towards irreversible thermal inactivation through directed evolution. *Protein Eng* **16**:599-605.

Zhao,H. and Arnold,F.H. 1999. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng* **12**:47-53.

Zhao,H., Giver,L., Shao,Z., Affholter,J.A., and Arnold,F.H. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol*. **16**:258-261.

Zhou,X.X., Wang,Y.B., Pan,Y.J., and Li,W.F. 2008. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino. Acids* **34**:25-33.

Zyprian,E. and Matzura,H. 1986. Characterization of signals promoting gene expression on the *Staphylococcus aureus* plasmid pUB110 and development of a gram-positive expression vector system. *DNA* **5**:219-225.

Publications

Ahmad S., Kamal Z., Sankaranarayanan R. and Rao N.M. (2008), Thermostable *Bacillus subtilis* lipases: *in vitro* evolution and structural insights. *J Mol Biol.* Aug 29;381(2):324-40

Ahmad S. and Rao N.M. (2008), Mutational analysis of thermostability in lipase: role of residual structure on the efficiency of refolding. *Manuscript in preparation.*

Rajkumara E., Priyamvada A., **Ahmad S.**, Sankaranarayanan R. and Rao N.M. (2007), Structural basis for the remarkable stability of *Bacillus subtilis* lipase (Lip A) at low pH. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, Volume 1784, Issue 2*, February 2008, Pages 302-311

Rajkumara E., Priyamvada A., **Ahmad S.**, Shanmugam V.M., Sankaranarayanan R. and Rao N.M. (2004), *Acta Crystallogr D Biol Crystallogr*, **60**, 160-162.