The Pennsylvania State University

The Graduate School

Eberly College of Science

STRUCTURAL BASIS FOR THERMOSTABILITY AND THERMAL DEPENDENCE

OF ACTIVITY IN ALPHA/BETA BARREL GLYCOSYL HYDROLASES

A Thesis in

Biochemistry, Microbiology, and Molecular Biology

by

Nicholas Panasik Jr.

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2002

We approve the thesis of Nicholas Panasik Jr.

Date of Signature

_____   _____

Jean E. Brenchley
Professor of Microbiology/Biotechnology
Thesis Advisor
Chair of Committee

_____   _____

J. Gregory Ferry
Professor of Anaerobic Microbiology;
Stanley Person Professor of Biochemistry
and Molecular Biology

_____   _____

B. Tracy Nixon
Professor of Biochemistry

_____   _____

Allen T. Phillips
Professor of Biochemistry

_____   _____

Richard Koerner
Assistant Professor of Chemistry

_____   _____

Robert A. Schlegel
Professor of Biochemistry and Molecular Biology
Head of the Department of Biochemistry and
Molecular Biology

ABSTRACT

A fundamental biochemical observation is that most enzymes have a limited temperature range for activity and stability. Many attempts have been made to understand the mechanisms that determine this range for a specific enzyme. Studies comparing the structural variations between mesophilic and thermophilic enzymes have led to speculation that different structural factors such as the number of ion pairs, the number of hydrogen bonds, the amount and type of solvent exposed surface area, and the amino acid composition are important. In order to determine whether any of these factors varied significantly between mesophilic enzymes alone, I performed a statistical analysis of the distributions of these factors in all mesophilic and thermophilic α/β barrel glycosyl hydrolase structures. These data provide a spectrum of values against which researchers can compare findings from pairwise studies of naturally occurring homologues to evaluate whether the differences they observe are significant and worth further study. Only two significant differences were observed between the two groups studied: the number of glycine residues in turns at the bottom of the barrel was higher in mesophilic structures and the B factors were consistently lower in the thermophilic structures. These observations suggest differences in enzyme flexibility may play a key role in temperature adaptation. For the factors most commonly conjectured as imparting thermostability (ion pairs, hydrogen bonds, solvent accessible surface area, and amino acid composition) the degree of fluctuation within one temperature regime (mesophiles) was equal to the variability between the two regimes (mesophiles vs. thermophiles). Consequently, such comparisons of gross structural features between naturally occurring extremophilic homologs were found not to contain sufficient resolution to determine the molecular basis for thermostability.

I examined several approaches that would more specifically address the question of what structural changes, if any, could alter the temperature range for enzyme activity. I selected a directed evolution approach in which I used random mutagenesis to increase low temperature activity of one thermophilic and one psychrophilic family 42 β-galactosidase.

In the first case, a low temperature enrichment strategy was developed to select for increased activity at low temperatures in the psychophilic enzyme, and the existence of a selective pressure was validated. Two mutations, K206G and L423F, were found to double the specific activity of the β-galactosidase toward ONPG substrate. The effects of these two mutations on enzyme kinetics were examined with lactose as a substrate. The K_m values decreased by a factor of two while the k_{cat} value remained unchanged.

In directed evolution of the thermophilic enzyme, the low temperature limit of the mutant enzyme was extended 25°C below that of its parent while maintaining stability at high temperatures (60°C). The individual mutations responsible for this phenotype were identified as V3E, L4I, F187I, and F258S. The effects of these four mutations on enzyme kinetics were examined. The K_m value remained unchanged while the k_{cat} value was increased. The effects of the individual mutations were found to be additive. Titration experiments using dithiobisnitrobenoic acid (DTNB) to determine the number of cysteine residues accessible in the soluble enzyme indicated that molecular flexibility is increased. This work demonstrates that the temperature range of enzyme activity may be broadened without the loss of thermostability.

Although there is not sufficient data to identify the location of mutations F187I and F258S in relation to structure or the active site, the localization of the other two mutations (V3E and L4I) to the N-terminus of the protein and their resultant effects on enzyme flexibility

suggests that the N-terminus of an $(\alpha/\beta)_8$ barrel strongly influences the flexibility and thermostability of these molecules and allowed us to propose this as a general mechanism for adaptation to low temperatures. Subsequent saturation mutagenesis in the N-terminal region of the gene for the thermophilic enzyme led to a high percentage (over 50%) of mutant enzymes that were significantly altered in their thermal characteristics. To determine if the same region was important in the thermal adaptation of a related psychrophilic enzyme, the N-terminal region of the psychrophilic gene was subjected to saturation mutagenesis and two variants were obtained that exhibited increased activity at low temperatures. A statistical comparison of the crystallographic structures of other thermophilic and mesophilic $(\alpha/\beta)_8$ barrel glycosyl hydrolases was then performed, and results suggest this mechanism may be common to many $(\alpha/\beta)_8$ barrels. The unique structural role that the N-terminus plays in $(\alpha/\beta)_8$ barrel architecture further suggests that some mechanisms leading to low temperature activity and thermostability are protein fold dependent.

Taken as a whole, this work shows that while the commonly undertaken comparisons of naturally occurring extremophilic homologues that are based on differences in the gross amounts of structural features may not reveal mechanisms of thermostability or thermal dependency of activity, strategies may be developed that can lead to the identification of general mechanisms that control an enzyme's thermostat.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

.

ACKNOWLEDGEMENTS

gracious help in so many experiments and her great efforts and long days with secondary

structural analysis, Dr. Vanya Miteva for her kind encouragement and guidance, Dr. Hemmant

Yennawar for his patient mentoring and tutelage in crystallographic techniques, and to Jonna

Coombs for being a wonderful colleague. Special thanks to Jennifer Biddle and Dr. Mike

Frodyma for making lab fun and enjoyable every day.

I would especially like to thank the National Science Foundation for the opportunity to be

a part of the research training grant and gain an education in two major fields. But most

importantly, I am grateful to the NSF for giving me the opportunity to have more than one

mentor. The experience has been exhilarating.

To my family and friends, know that your support has made this all possible.

Lastly, I would like to give special thanks and acknowledgement to my fiancé Dr. Celena

E. Kusch for help in editing this work and for her warm and loving support thought graduate

school.

To all of you, thank you again so much.

Chapter 1

BIOCHEMISTRY OF THERMAL ADAPTATION

1.1 DIVERSITY OF MICROORGANISMS. EXTREMOPHILES, AND THE TEMPERATURE GRADIENT

Microorganisms have been isolated from nearly every environment found on Earth, and researchers have found them even in the most extreme habitats [1]. The remoteness of some of these extreme environments, and difficulties in cultivating many of the organisms may have initially impeded the study of extremophiles, but now with the increased availability of specialized transportation and advances in culture techniques, the field of extremophilic research is growing rapidly.

The range in temperature across environments from which microorganisms have been isolated is very large. High temperature environs rich in prokaryotic life include hydrothermal vents at the ocean floor where high pressures increase the boiling point of water and allow liquid temperatures in excess of 130°C [2]. Terrestrial hot springs, which can be found on six of the seven continents, support temperatures of 60°C and higher and have been a rich source of thermophilic species [3]. Natural environments, however, represent only one area where thermophilic and hyperthermophilic organisms can be found. Hot water pipes, thermal exhaust vents, and sewage treatment plants [4] are but a few man-made environs where organisms are found to grow and thrive.

Low temperature environments are far more plentiful. 85% of Earth's biosphere experiences temperatures of less than 5°C during the year [5], and over 75% is permanently cold. The oceans, which have an average temperature of 4°C, are a vast source of psychrophilic species [6, 7] as are the permanently cold environments of underground caves [8], cold springs

[9], arctic basins [10], the Antarctic continent [11], and even glaciers [12]. Psychrophiles, or "cold-loving" organisms, have also been isolated in regions which experience low temperatures only transiently. These areas are as diverse as they are common, and there are now reports of the isolation of cold-loving organisms from places like seasonally frozen lakes [13] or Pennsylvania farm lands in winter [14]. Man-made cold environments such as refrigerated meats [15] and cooling systems have also been found to contain psychrophiles.

Overall, microorganisms can have optimal growth conditions at temperatures from as low as -4°C to as high as 130°C, with different species growing optimally at temperatures distributed throughout that gradient.

Thermophiles are perhaps the most studied group of extremophilic organisms. Many of these are archeal, but not all. Both eubacterial [16] and eukaryotic [17] species have been found. Studies of the ribosomal RNA sequences of several of the hyperthermophilic species have revealed these are phylogenetically "deep branching" organisms on the tree of life, causing some researchers to postulate that early life on our planet evolved in a hot environment. *Pyrococcus furiosus*, *Thermotoga maritima*, *Thermus aquaticus* and *Thermus thermophilus* are but a few well known species.

Psychrophiles, while perhaps less studied, are no less prolific. The definition of a psychrophile given by Neidhardt and used in this work is an organism which can grow at 5°C and has an optimal growth temperature which is below 37°C [18]. A common misconception about these organisms is that they are usually dormant and merely psychrotolerant, but growth rate doubling times as high as 4 hours at 4°C have been observed, indicating that the catalytic efficiencies of their metabolic enzymes are on par with many mesophilic counterparts [19].

These psychrophilic prokaryotic organisms are mainly Bacteria [20], but both Archaea [21] and Eucarya [22] have been isolated as well.

As a whole, extremophilic organisms on both ends of the temperature gradient are overwhelmingly unicellular and for the most part prokaryotic. Still, representatives of the three domains of life have been found in each of these environments, suggesting that the enzymes common to and required by all organisms can be adapted to such conditions.

1.2 CELLULAR ADAPTATIONS TO TEMPERATURE

Optimal growth requires that all essential metabolic functions are occurring at well-regulated and efficient rates and that macromolecular structures, such as cellular membranes, remain fluid. Extremophiles employ a variety of strategies to maintain these processes at the temperatures in which they grow. To compensate for the negative effects of high or low temperatures on the lipid bilayer, these organisms demonstrate homeoviscous adaptation. Psychrophiles increase their membrane fluidity by decreasing the degree of saturation of fatty acids in their membrane phospholipids; conversely, thermophiles increase the degree of saturation in the fatty acids of membrane phospholipids to decrease membrane fluidity. In some psychrophiles, membrane fluidity is also achieved by a higher presence of branched-chain fatty acids [23], Some thermophiles likewise have novel macrocyclic membrane lipids which help to maintain membrane stability [24], and some psychrophiles even produce 'ice structuring proteins' that work like anti-freeze to inhibit the formation of ice crystals in the surrounding environment that might damage the cellular membrane [25].

Within the cell, molecular chaperones aid in the folding of proteins at both high and low temperatures. Chaperonins, which help proteins overcome steps in the folding pathway that may have high activation energies, may be especially important in the folding of proteins at low temperatures. They are also known to help refold enzyme molecules as they begin to unfold [26], a property perhaps quite useful at high temperatures.

Adaptation occurs in the treatment and character of nucleic acids as well. Proteins homologous with CspA, a cold shock protein found in *Escherechia coli,* facilitates transcription at low temperatures by preventing formation of secondary structures in RNA molecules. Both thermophiles and psychrophiles utilize post-transcriptional modification of tRNA molecules. Higher levels of nucleoside modification with dihydrouridine are thought to play a role in maintenance of conformational flexibility of RNA in some psychrophiles [27], while ribose-methylated nucleosides, 2'-O-methyladenosine, N(2),2'-O-dimethylguanosine and N(2),N(2),2'-O-trimethylguanosine, play a role in stabilizing the tRNA of hyperthermophiles [28]. Interestingly, the temperature stability of mRNA does not increase in thermophiles: for many of these organisms, the half-life of the mRNA decreases as the growth temperature increases [29]. The stability index (half-life of mRNA/doubling time of cells), however, is remarkably constant regardless of the growth temperature, suggesting that kinetic considerations play a significant role in the explanation of mRNA thermophily [29].

1.3 MOLECULAR ADPATATIONS TO TEMPERATURE IN THE STRUCTURE OF PROTEINS

Thermodynamic challenges at high temperatures

While adaptations of the cell inventory are important, of no less concern is the molecular adaptation of the enzymes themselves. There are important and difficult thermodynamic challenges for proteins to overcome in order to remain functional and stable at extremely high or low temperatures. At either extreme, enzymes must remain folded in the correct three-dimensional conformation while maintaining enough flexibility to carry out the molecular motions required for catalysis. In thermodynamic terms, maintenance of a folded state requires that the loss in chain entropy upon folding must be compensated by the sum of intermolecular interactions (enthalpic) and the positive solvation entropy from segregating non-polar residues from water (entropic). The value of free energy of stabilization of a folded protein represents small differences between very large and competing forces and typically does not exceed 50-100 kJ mol^{-1}. This means that, on balance, stabilization is based on the equivalent of a few hydrogen bonds, ion pairs, or hydrophobic interactions [30]. As temperatures increase, the destabilizing forces of greater molecular motions compete against the stabilizing forces of hydrogen bonds, ion pairs, and hydrophobic interactions. Additionally, at high temperatures, some amino acids, such as glutamine, undergo deamination or other modifications [31]. Proposed mechanisms for the molecular basis of thermostability will be presented in Chapter 2.

The low temperature thermodynamic barrier

At lower temperatures there are equally challenging albeit different concerns. According to the Boltzman equation, rates of enzymatic reactions increase with higher temperature as more reactants contain the necessary energy of activation. Conversely, reaction rates decrease as temperature is lowered, following the same distribution. Table 1.1 illustrates the effect of temperature on reaction rate considering the range of activation energies typical for biochemical reactions.

Table 1.1.Relative reaction rates based on free energy and temperature
(Rates at 20°C are normalized to 1) [32]

Temperature	$\Delta G^{\ddagger}$/kJ mol^{-1}=20	$\Delta G^{\ddagger}$/kJ mol^{-1}=40	$\Delta G^{\ddagger}$/kJ mol^{-1}=60	$\Delta G^{\ddagger}$/kJ mol^{-1}=80
20 C	1	1	1	1
0 C	1.8	3.6	6.7	11.4
-20 C	4	14	46	120

Thus, in order for cold-active enzymes to maintain catalytic efficiencies commensurate with their mesophilic counterparts, there must be structural changes which enhance their turnover rates and/or decrease their K_m values. These structural changes may be unrelated to those involved in the maintenance of stability. This will be discussed further in Chapter 4.

The higher thermodynamic barrier at low temperatures may affect catalysis in other ways as well. It has been observed that many mesophilic enzymes undergo small protein motions while carrying out catalysis and that this "breathing" is actually required for catalysis [33]. At lower temperatures the free energy barriers to these kinetic motions become greater, and motion and breathing become restricted, possibly interfering with catalysis. It is believed that cold-active

enzymes would need to be more flexible than their mesophilic counterparts to overcome this restriction [34]. While the concept of flexibility provides a useful framework, it has not yet been fully demonstrated.

1.4 CURRENT STRUCTURAL MODELS OF LOW TEMPERATURE ENZYME ACTIVITY

Despite a lack of specific answers, there are some general models for explaining cold-activity. Many researchers believe the key to maintaining activity at low temperatures is a change in the relative flexibility of the enzyme [34-36]. Various factors have been proposed which would give a structural basis for an increased conformational flexibility in cold-active enzymes at low temperatures. I divide these adaptations into two classes. The first class, which is chiefly enthalpic, includes such proposed mechanisms as a decrease in the number of salt bridges [37-39]; a decrease in the number of hydrogen bonds [40]; a decrease in the number of arginine residues relative to lysine residues (and thereby a decrease in the number of hydrogen bonds by one in each case) [41]; amino acid substitutions [42] that decrease inter-domain interactions in multimeric proteins [43]; and/or a decrease in the number of ions bound or their dissociation constants [44]. Some ions, such as calcium, exhibit a strong electric field which can have effects distantly throughout an enzyme's structure. It is proposed that this field acts to hold the molecule together more tightly by attracting oppositely charged residues further away in the molecule [45]. The second class, chiefly entropic, includes a reduction in the hydrophobicity in the core of the enzyme; the replacement of residues to allow greater rotational freedom in the polypeptide chain (which could include a decrease in the number of proline residues or increase in the number of

glycine residues); and/or an increase in the number of surface charges leading to increased interactions with solvent.

Significantly, these mechanisms for low temperature activity appear to act conversely to mechanisms proposed for thermostability at high temperatures. By lessening the number of enthalpic interactions within the protein or increasing the conformational entropy and/or the number of interactions with solvent, the propensity to enter the unfolded state will be increased and thus the molecule will be held together less tightly and therefore have greater conformational flexibility. That these mechanisms appear converse in principle and effect is one reason why the properties of low temperature activity and thermostability at high temperatures are sometimes believed to be mutually exclusive.

If it is true that the free energy of stabilization of the folded state is on the order of magnitude of only three or four hydrogen bonds or ion pairs, then any changes in hydrogen bonds to increase the conformational flexibility to provide cold-activity would have drastic (negative) effects on thermostability. Thus, it would superficially appear that the changes that make an enzyme cold-active would necessarily make it thermolabile. As a result, one could conclude that the conformational flexibility required for low temperature catalysis is mutually exclusive to conformation rigidity required for thermostability at high temperatures. This has led to speculation that enzymes can remain active over only a limited range. If one considers the possibility of local flexibility, however, this conclusion is challenged. A change in one hydrogen bond or ion pair may be sufficient to increase the local flexibility in a key region of the molecule where a molecular motion is required for catalysis without necessarily sacrificing the overall stability of the molecule. If this is true, rather than conceiving of either thermostability or low

temperature activity as enzyme-wide properties, the importance of local determinants of both thermostability and enzyme flexibility should be given greater consideration.

Identification of a single local mutation is all but impossible when doing pairwise (side by side) comparisons of distantly related homologs where the number of amino acid differences can be in the hundreds. This may explain why some conflicting conclusions regarding the molecular mechanisms of thermostability and low temperature activity have come from comparisons of naturally occurring, but distantly related thermophilic, mesophilic, and psychrophilic homologs. These local mutations may also explain why we have been able to increase the low temperature range of an enzyme without sacrificing thermostability, as described in Chapter 4.

To test any of these hypotheses, it would be helpful to have structures of cold-active enzymes. Unfortunately, only four structures of cold-active enzymes have been solved to date. They include a psychrophilic α-amylase from *Alteromonas haloplanctis* [46], a cold-active citrate synthase from an Antarctic bacterium [47], a cold-active malate dehydrogenase from *Aquaspillium arcticum* [48], and a triose phosphate isomerase from the psychrophilic *Vibrio marinus* [49].

The superposition of the psychrophilic α-amylase structure (AhA) with a mesophilic counterpart from porcine pancreas (PPA) showed that the 24 residues forming the five binding sites and having direct or water-mediated contacts with inhibitor are strictly conserved [50]. No changes were found in the catalytic center, indicating that the mechanism of cold adaptation occurs elsewhere in this molecule. This is a common trend seen in many studies of enzyme adaptation to temperature [51-53]. In these cases, structural effects may be mediated through

movements in regions of secondary structure [22]. For example, an additional disulfide bridge from domain A to the end of domain B is present in the mesophilic structures but absent in the cold-active enzyme. The loss of this domain tethering may provide this enzyme with a greater degree of freedom and added overall conformational flexibility that may be key to its low temperature adaptation as seen with an *E. coli* acid phosphatase [54]. Metal binding can also have long range effects on protein structure. The cold-active *A. haloplanctis* α-amylase has a lower affinity for calcium binding when compared to its mesophilic counterpart from porcine pancreas, perhaps explaining some of its thermolability. Information on the overall number of hydrogen bonds, ion pairs, and hydrophobic packing arrangement and other features commonly attributed to changes in flexibility have yet to be published.

In a different study, the molecular structures of a cold-active citrate synthase (DsCS) from Antarctic eubacterial isolate DS2-3R was compared to a highly thermostable counterpart (PfCS) from *P. furiosus* [55]. It was observed that both the catalytic residues and those that play dominant roles in the active site were completely conserved. No significant differences were found in the number of intra- or inter-subunit hydrogen bonds. There was, however, a reduction in the number of residues that bind substrate, resulting in a more accessible active site.

Contrary to expectations, DsCS had far more intermolecular ion pairs than PfCS: 42 and 27, respectively. This has been used to explain the cold-active citrate synthase's resistance to low temperature denaturation [55]. However, the increased number of ion pairs in the DsCS enzyme casts doubt onto the proposal that enzymes from psychrophiles would generally have fewer ion pairs than their counterparts from mesophiles and thermophiles. The thermostable enzyme does have more ion pairs located in networks between subunits. Ion pairs participating in networks are

thought to be a possible stabilizing factor, since fewer water molecules are required to solvate the charged areas. The cold-active enzyme lacks the complex ion pair networks, isoleucine clusters, and tyrosine clusters that exist in the monomer association regions in the thermophilic enzyme. In PfCS, this may provide dimeric integrity at higher temperatures.

Another feature thought to promote structural stability are glycine residues. Glycines in alpha helices are thought to be destabilizing while glycine residues in loop regions are believed to promote increased structural flexibility. DsCS has 22 glycine residues (7 in helices) and PfCS has 29 glycine residues (7 in helices). While both have the same number within helices, the thermostable enzyme actually appears to have more glycines contributing to flexibility of loop regions than DsCS. Conversely, an absence of proline residues in loop regions (stabilizing) and the presence of three proline residues in α-helices (destabilizing) in the cold-active DsCS enzyme may increase the overall flexibility of the protein relative to PfCS.

Similar to the comparison of DsCS and PfCS, a comparison of the cold-active malate dehydrogenase from *Aquaspirillium arcticum* (Aa MDH) and the enzyme from the thermophile *Thermus flavus* (Tf MDH) revealed that the cold-active enzyme had the same number of intra-subunit ion pairs and ion pair networks, but only half the number of inter-subunit ion pairs and fewer residues participating in networks at subunit interfaces [48]. Similarly, there were fewer inter-subunit hydrogen bonds found in the cold-active enzyme but more intra-subunit hydrogen bonds. Additionally, *Aa*MDH had a smaller buried surface area and a smaller portion of charged residues in the dimeric interface. The decreased number of interactions in the dimer association regions may contribute to the cold-active enzyme's greater thermolability. Regional B-factor analysis suggested that the active site is more flexible in the cold-active enzyme even though the

molecule's overall B-factors remain lower than those from the thermophilic enzyme. Catalytic efficiency of the cold-active enzyme may also be increased by greater positive potentials and lower negative potentials around the substrate binding sites as both substrates are negatively charged.

Research on the cold-active triose phosphate isomerase from *Vibrio marinus* [49] did not include information on intermolecular hydrogen bonds, ion pairs, or amino acid distribution. However, a single point mutation in the active site (alanine to serine) of the cold-active enzyme appears to have increased thermostability with only a slight decrease in activity.

To date, there is no consensus for a common paradigm of structural changes that confer either cold-activity or thermostability. In some cases, the structural features that match predicted molecular mechanisms were observed, and in others, observations ran counter to expectations. It does appear that structural changes that affect either of these properties are often far from the active site. This implies that the key adaptations to temperature involve changes in the dynamics of the molecule.

Unfortunately, discussion of molecular adaptation in these enzyme structures has been chiefly based on the total numbers of these structural elements. Additionally, the homologs used in the comparisons were usually distantly related, and even in the cases of the highest homology, still represent sequence differences in the tens to hundreds of amino acids. When comparing distantly related homologs, genetic drift alone may be enough to explain the differences observed in distributions of the number of ion pairs, hydrogen bonds, and other structural elements. There is not a sufficient number of solved structures to determine if the distribution of these features follows a consistent pattern from mesophile to psychrophile. In addition, some if the observed

differences between mesophilic and thermophilic enzymes may actually reflect the naturally occurring variation between different mesophilic enzymes. The variation between mesophilic enzymes will be examined in Chapter 2.

Another large complication in this methodology for identifying molecular mechanisms of thermostability and low temperature activity is that different enzymatic structural folds are being used to test these hypotheses. The measurements may seem incoherent because the yardstick keeps changing. There is no obvious reason that molecular mechanisms which may reasonably be conjectured to deal with *local* flexibility, would not be *fold specific*.

Untangling the mess. A strategy for comparison.

How vast is the variation in structural features commonly thought to play a role in temperature adaptation among mesophilic proteins? The natural range in the number of ion pairs and other similar features of a mesophilic group of proteins must be determined to see if the range is large or small. If the range is too large, i.e. if the number of ion pairs in mesophilic proteins, as a group, varies from mesophilic protein to mesophilic protein as much or more than it varies between mesophile and thermophile (or psychrophile) then it should be clear that the current methodology for determining a molecular mechanism for thermostability is flawed.

Similarly, the range in the number of hydrogen bonds within a protein, the amount of solvent accessible surface area, and the amino acid composition should be determined for a mesophilic group of proteins. These distributions should be quantified and compared to either psychrophilic and/or thermophilic proteins to determine if the net number of these structural differences that have been used in structural studies is a valid measure for comparison.

Such a study should select a single enzymatic fold, with a sufficient number of solved crystallographic structures. The members of this structural family that are used in the study should catalyze similar chemical reactions so that differences observed are not likely to be due to chemical mechanism. In Chapter 2 I begin to undertake such a study on eight stranded α/β barrel glycosyl hydrolases.

1.5 EIGHT STRANDED ALPHA/BETA BARRELS

General architecture

The $(\alpha/\beta)_8$ barrel fold was chosen for analysis because it is the most common structural motif found in enzymes. 10% of all known protein structures are members of this family [56]. This fold is defined by a motif (β sheet–loop-α helix–loop) that is repeated eight times, yielding a structure that has eight parallel β strands on the interior of the protein and eight α helices on the exterior of the protein (Figure 1.1). The helices are usually amphipathic, with hydrophobic residues on the side of the helix that is packed against the β sheets and more polar residues on the side of the helix facing the solvent. Strand β1 makes hydrogen bonds with strands β2 and β8, and this pattern is circularly permuted throughout the interior. The cross section of the barrel however is not usually circular. Most proteins are elliptical and the major points of the ellipse can be at different β stands. The height of the barrel can also vary widely.
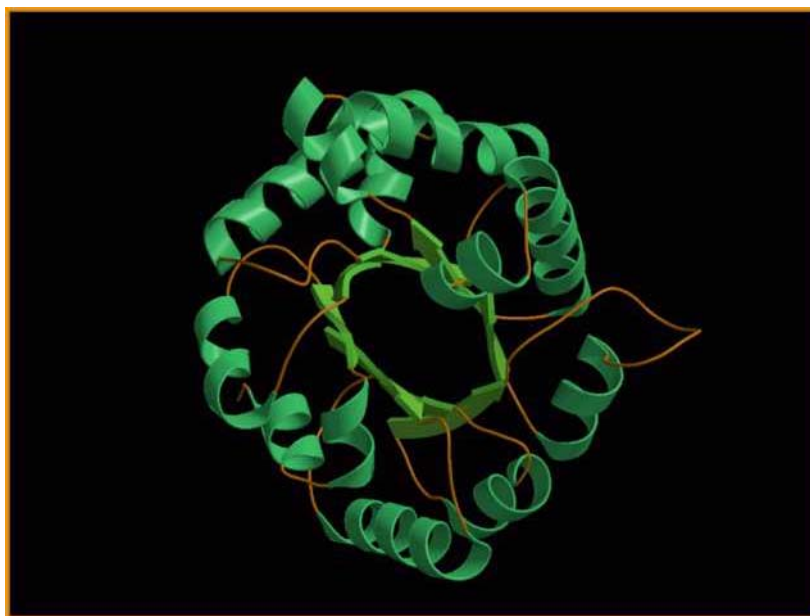
Figure 1.1 Eight stranded α/β barrel of triose phosphate isomerase

The top and bottom faces of the barrel are also uniquely different. The bottom or N-terminal face, located at the N-terminal side of the β sheets, is often characterized by short loops with tight turns. No additional secondary structural elements will be found inserted here. The top face, at the C-terminal side of the β sheets often has large loops and secondary structural elements inserted. It is on this top face that the active site residues (usually found in the loop regions) and the substrate recognition site residues (usually located at the C-terminal of the β sheet) may be found.

Evolution and relatedness of (α/β)₈ barrels

A striking characteristic of the $(\alpha/\beta)_8$ barrel family of proteins is the high variability in the sequences of its members. In fact, the amino acid sequences of the $(\alpha/\beta)_8$ barrels are so diverse that barrels cannot be identified based on sequence alone [57]. Many studies have been

undertaken to determine whether or not the different $(\alpha/\beta)_8$ barrel structures are a result of divergent or convergent evolution [56, 58, 59], and the issue has not been decided to date. Thus far, $(\alpha/\beta)_8$ barrels have been classified into families based on topological differences and hidden homologies. Since $(\alpha/\beta)_8$ barrels often have recognizable domains or secondary structures placed at the C-terminal end of the β sheets that make up the barrel, Reardon and Farber used these and other distinctive features, such as barrel ellipticity, height, and shape, to classify them into six families [56]. Other researchers have looked for "hidden homologies," localized regions of sequence that are highly conserved within a particular enzyme family that may be conserved among closely related families [60]. These studies are based on the idea that a small conserved sequence region in the $(\alpha/\beta)_8$ barrel enzyme should be more or less conserved in an equivalent region in other proteins, while the rest of the two structures being compared may not have large homology. This has been used with some degree of success. Bork et al. have found two sequence regions spanning the phosphate binding site in eight $(\alpha/\beta)_8$ barrels [58], while Janacek et al. have identified a strongly conserved sequence in β sheet 5 among the alpha amylases [61], The classification of glycosyl hydrolases has been greatly aided by this methodology with the discovery of a localization of functionally essential glutamate residues to strands β4 and β7 [62].

Hidden homologies also suggest that the homologous amino acid residues inherited from a primordial barrel structure can adopt different structural positions and/or functional roles. This can be seen with the identification of residues flanking the β2 strand in enzymes related to amylases, where glycine is positioned at the N-terminal end and a proline at its C-terminus even though the distance between them can vary from nine to fourteen residues in length throughout different structures [62].

Chemistry of (α/β)₈ barrels

This well-studied family is mainly found in the metabolic enzymes and extracellular glycosyl hydrolases [56]. They are known, however, to catalyze a plethora of different types of reactions including reductions, proton transfers, cycloisomerizations, phosphorylations, hydrolysis of glycosidic bonds, and the cleavage of aldols and carbon-carbon bonds [56]. Glycosyl hydrolases comprise the largest subclass of this protein fold, with over 40 different crystallographic structures solved to date.

1.6 GLYCOSYL HYDROLASES

Reaction Mechanism

O-Glycoside hydrolases (EC 3.2.1.-) catalyze the water mediated cleavage of glycosidic bonds. Enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires two critical residues: a proton donor and a nucleophile/base. This hydrolysis occurs via two major mechanisms, giving rise to an overall retention or inversion of configuration about the anomeric carbon.

In retaining enzymes, the reaction proceeds through a double displacement mechanism beginning with the general acid catalyzed attack of the nucleophilic carboxylate side chain on the C1 of the sugar substrate. The reducing end of the sugar is released leaving a glycosyl enzyme intermediate. The second step is hydrolysis of the glycosyl enzyme intermediate by a water molecule that is activated by a general base, yielding a product with retention of configuration about the anomeric carbon (i.e. equitorial to equitorial, etc). The transition states have high oxocarbenium ion character (i.e. positive charge on sugar).

In the case of inverting enzymes, protonation of the glycosidic oxygen occurs simultaneously with aglycon departure and attack by a base activated water molecule in a single displacement reaction. The water molecule is initially held in position between the base and the sugar. Thus, in inverting enzymes, the distance separating the catalytic residues is much larger (~10 Å) than that found separating the two bases in retaining enzymes (~5.5 Å). The proton is within hydrogen bonding distance of the glycosidic oxygen in both the retaining and the inverting mechanisms.

Active-site topology

The active site topologies of all glycosyl hydrolases whose structures are known fall into three categories: the pocket or crater, the cleft or groove, or the tunnel. The pocket, found in β-galactosidases, is optimal for recognition of saccharide. The cleft or groove, seen in α-amylases, is optimized for the binding of polysaccharides and endo-acting catalysis. The tunnel topology, found only in the cellobiohydrolases, allows for processive catalysis by remaining bound to the polysaccharide chain while product is released by sliding down the substrate.

Substrates (structures & linkages)

Glycosyl hydrolases can catalyze the cleavage of the glycosidic bonds between carbohydrates or between a carbohydrate and a non-carbohydrate moiety. Glycosyl hydrolases are notoriously promiscuous in their reactivity to a variety of carbohydrate substrates. While substrate specificity may be somewhat lax, the hydrolases are usually more specific in the linkages, such as β (1-4), on which they act. Glycosyl hydrolases have developed various ways to

lower the energy barrier of the hydrolysis reaction, such as substrate distortion into a half-chair conformation [62].

Families

The glycosyl hydrolases have been divided by Henrissat into over 87 families, based upon hydrophobic cluster analysis and other sequence differences [62]. These have been further grouped into 12 clans based on amino acid similarity and proposed enzymatic fold. For the 35 glycosyl hydrolase families for which a three-dimensional structure has been determined, 18 are of the $(\alpha/\beta)_8$ barrel type. Most of these families have been grouped into the GH-A superfamily, also known as the 4/7 superfamily, by virtue of the fact that the catalytic residues are found on the strand $\beta 4$ and $\beta 7$ of the barrel. All of the members of the 4/7 superfamily are retaining enzymes as well (meaning the catalytic residues will be separated by an average distance of 5.5 Å).

Thesis organization

The material in this thesis is organized into six chapters. The second chapter was accepted for publication in the journal *Biochimica Et Biophysica Acta*, and printed in the November 2000 issue. In this material, I analyzed the solved structures of 22 mesophilic and five thermophilic $(\alpha/\beta)_8$ barrels glycosyl hydrolases to determine the natural distributions of the structural features most commonly proposed to affect thermostability and adaptation to temperature. The distribution is expressed as an average value with a range given by the standard deviation and it is used to show that, for most of these commonly proposed mechanisms, the

distribution among the mesophilic groups is sufficiently broad as to call into question any conclusions based on differences in the gross numbers of structural features between naturally occurring homologues.

In Chapter 3, I discuss the reasons why these distributions may be so broad and what strategies one may employ to overcome these limitations to determine mechanisms that confer adaptation to extreme temperatures. Recent applications of and the discoveries made with some of these techniques will be reviewed.

In Chapter 4, I present the use of directed evolution techniques to increase the temperature range of an enzyme by extending the low temperature limit of a thermophilic β-galactosidase. The mutations conferring increased activity at low temperature are systematically identified and their effects on enzyme activity, stability, flexibility, and kinetics studied. Analysis of the location of some of these mutations leads to the proposal of a new mechanism for thermal adaptation in $(\alpha/\beta)_8$ barrel enzymes. This theory is tested by saturation mutagenesis of the 'hotspot' region suggested by this mechanism and the effects of this mutagenesis on the frequency of cold-activating mutations analyzed.

In Chapter 5, the paradigm of temperature adaptation proposed in Chapter 4 is extended to a related psychrophilic enzyme through saturation mutagenesis of the equivalent region. In order to have a sufficient selection for variants with increased activity at low temperatures in a psychrophilic enzyme, a novel strategy that employs enrichments is presented and the existence of adequate selective pressure verified. The variants derived form the saturation mutagenesis are characterized by their effects on growth rate and provide further evidence of the proposed structural mechanism.

Chapter 6 describes the random mutagenesis of the entire psychrophilic β-galactosidase gene in attempts to find more mutations that lead to increased catalytic activity at low temperatures. The enrichment strategy from Chapter 5 is employed to select for variants with increased activity on lactose at low temperature in this psychrophilic enzyme. The mutant β-galactosidases derived from some of these variants are purified and their enzymatic properties characterized.

1.7 REFERENCES

1.   Wynn-Williams, D.D., *Ecological aspects of Antarctic microbiology*, in *Advances in Microbial Ecology*, K.C. Marshall, Editor. 1990, Plenum Press: New York. p. 71-146.

2.   Takai, K. and K. Horikoshi, *Genetic diversity of archaea in deep-sea hydrothermal vent environments.* Genetics, 1999. **152**(4): p. 1285-1297.

3.   Ruff-Roberts, A.L., J.G. Kuenen, and D.M. Ward, *Distribution of cultivated and uncultivated cyanobacteria and Chloroflexus-like bacteria in hot spring microbial mats.* Appl. Environ. Microbiol., 1994. **60**(2): p. 697-704.

4.   Chen, M., *Adaptation of mesophilic anaerobic sewage fermentor populations to thermophilic temperatures.* Appl Environ Microbiol, 1983. **45**(4): p. 1271-6.

5.   Baross, J.A. and R.Y. Morita, *Microbial life at low temperatures: ecological aspects*, in *Microbial life in extreme environments*, D.J. Kushner, Editor. 1978, Academic press: New York. p. 9-71.

6.   Dalton, R., *Drillers dig deep for microbes under the sea floor.* Nature, 2002. **415**(6872): p. 566.

7.   Beja, O., et al., *Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces.* Appl Environ Microbiol, 2002. **68**(1): p. 335-45.

8.   Schabereiter-Gurtner, C., et al., *Altamira cave Paleolithic paintings harbor partly unknown bacterial communities.* FEMS Microbiol Lett, 2002. **211**(1): p. 7-11.

9.   Rudolph, C., G. Wanner, and R. Huber, *Natural communities of novel archaea and bacteria growing in cold sulfurous springs with a string-of-pearls-like morphology.* Appl Environ Microbiol, 2001. **67**(5): p. 2336-44.

10.   Deming, J.W., *Psychrophiles and polar regions.* Curr Opin Microbiol, 2002. **5**(3): p. 301-9.

11.   Mountfor, D.O., et al., *Psychromonas antarcticus gen. nov., sp. nov., a new aerotolerant anaerobic, halophilic psychrophile isolated from pond sediment of the McMurdo Ice Shelf, Antarctica.* Arch. Microbiol., 1998. **169**: p. 231-238.

12.   Price, P.B., *A habitat for psychrophiles in deep Antarctic ice.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1247-51.

13. Ferroni, G.D. and J.S. Kaminski, *Psychrophiles, psychrotrophs, and mesophiles in an environment which experiences seasonal temperature fluctuations.* Can J Microbiol, 1980. **26**(10): p. 1184-90.

14. Loveland, J., et al., *Characterization of psychrotrophic microorganisms producing β-galactosidase activities.* Appl. Environ. Microbiol., 1994. **60**(1): p. 12-18.

15. Ivanova, S., *[Contamination of butchered poultry with psychrophilic and psychrotrophic microorganisms].* Vet Med Nauki, 1984. **21**(10): p. 51-8.

16. Ueda, K., et al., *Distribution and diversity of symbiotic thermophiles, Symbiobacterium thermophilum and related bacteria, in natural environments.* Appl Environ Microbiol, 2001. **67**(9): p. 3779-84.

17. Sutherland, L.A., W.M. Wong, and R.N. Nazar, *Sequence and putative regulatory elements in a 5S rRNA gene from a eukaryotic thermophile, Thermomyces lanuginosus.* Nucleic Acids Res, 1989. **17**(24): p. 10504.

18. Neidhardt, F.C., J.L. Ingraham, and M. Schaechter, eds. *Physiology of the bacterial cell: a molecular approach.* 1990, Sinauer Associates, Inc.: Sunderland.

19. Zecchinon, L., et al., *Did psychrophilic enzymes really win the challenge?* Extremophiles, 2001. **5**(5): p. 313-21.

20. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya.* Proc. Natl. Acad. Sci. USA, 1990. **87**: p. 4576-4579.

21. Nozhevnikova, A.N., et al., *Temperature characteristics of methanogenic archaea and acetogenic bacteria isolated from cold environments.* Water Sci Technol, 2001. **44**(8): p. 41-8.

22. Gianese, G., P. Argos, and S. Pascarella, *Structural adaptation of enzymes to low temperatures.* Protein Eng, 2001. **14**(3): p. 141-8.

23. Reizer, J., et al., *The effect of growth temperature on the thermotropic behavior of the membranes of a thermophilic Bacillus. Composition-structure-function relationships. Glycolipids from some extreme thermophilic bacteria belonging to the genus Thermus.* Biochim Biophys Acta, 1985. **815**(2): p. 268-80.

24. Kaneshiro, S.M. and D.S. Clark, *Pressure effects on the composition and thermal behavior of lipids from the deep-sea thermophile Methanococcus jannaschii.* J Bacteriol, 1995. **177**(13): p. 3668-72.

25.   Clarke, C.J., S.L. Buckley, and N. Lindner, *Ice structuring proteins - a new name for antifreeze proteins.* Cryo Letters, 2002. **23**(2): p. 89-92.

26.   Veinger, L., et al., *The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network.* J Biol Chem, 1998. **273**(18): p. 11032-7.

27.   Dalluge, J.J., et al., *Posttranscriptional modification of tRNA in psychrophilic bacteria.* J Bacteriol, 1997. **179**(6): p. 1918-23.

28.   Kowalak, J.A., et al., *The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles.* Biochemistry, 1994. **33**(25): p. 7869-76.

29.   Stenesh, J. and J.B. Madison, *Stability of bacterial messenger RNA in mesophiles and thermophiles.* Biochim Biophys Acta, 1979. **565**(1): p. 154-60.

30.   Fersht, A.R., et al., *Pathway and stability of protein folding.* Philos Trans R Soc Lond B Biol Sci, 1991. **332**(1263): p. 171-6.

31.   Arii, K., T. Kai, and Y. Kokuba, *Degradation kinetics of L-alanyl-L-glutamine and its derivatives in aqueous solution.* Eur J Pharm Sci, 1999. **7**(2): p. 107-12.

32.   Stroz, G. and R. Hengge-Aronis, *Bacterial stress response.* 20th Ed. ed. 2000, Washington D. C.: ASM Press.

33.   Rasmussen, B.F., et al., *Crystallin ribonuclease A loses function below the dynamical transition at 220 K.* Nature, 1992. **357**: p. 423-424.

34.   Gerday, C., et al., *Psychrophilic enzymes: A thermodynamics challenge.* Biochim. Biophys. Acta., 1997. **1342**: p. 119-131.

35.   Russell, N.J., *Toward a molecular understanding of cold activity of enzymes from psychrophiles.* Extremophiles, 2000. **4**(2): p. 83-90.

36.   Jaenicke, R., *How do proteins acquire their three-dimensional structure and stability?* Naturwissenschaften, 1996. **83**(12): p. 544-54.

37.   Perutz, M.F. and H. Raidt, *Stereochemical basis of heat stability in bacterial ferredoxins and in hemoglobin A2.* Nature, 1975. **255**: p. 256-259.

38.   Yip, K.S., et al., *The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures.* Structure, 1995. **3**: p. 1147-1158.

39.   Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs.* Journal of Molecular Biology, 1997. **269**: p. 631-643.

40.   Shortle, D., *Mutational studies of protein structures and their stabilities.* Quarterly Review of Biophysics, 1992. **25**: p. 205-250.

41.   Feller, G., et al., *Enzymes from cold-adapted microorganisms. The class C beta-lactamase from the antarctic psychrophile Psychrobacter immobilis A5.* Eur J Biochem, 1997. **244**(1): p. 186-91.

42.   Gianese, G., F. Bossa, and S. Pascarella, *Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes.* Proteins, 2002. **47**(2): p. 236-49.

43.   Wallon, G., et al., *Crystal structures of Escherichia coli and Salmonella typhimurium 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from Thermus thermophilus.* Journal of Molecular Biology, 1997. **266**: p. 1013-1016.

44.   Almog, O., et al., *Structural basis of thermostability: analysis of stabilizing mutations in subtilisin BPN'.* J Biol Chem, 2002.

45.   Lin, S.J., et al., *Weakly bound calcium ions involved in the thermostability of aqualysin I, a heat-stable subtilisin-type protease of Thermus aquaticus YT-1.* Biochim Biophys Acta, 1999. **1433**(1-2): p. 132-8.

46.   Feller, G., D. d'Amico, and C. Gerday, *Thermodynamic stability of a cold-active α-amylase from the Antarctic bacterium Alteromonas haloplanctis.* Biochemistry, 1999. **38**: p. 4613-4619.

47.   Miyazaki, K., et al., *Directed evolution study of temperature adaptation in a psychrophilic enzyme.* Journal of Molecular Biology, 2000. **297**: p. 1015-1026.

48.   Kim, S.Y., et al., *Structural basis for cold adaptation.* Journal of Biological Chemistry, 1999. **274**: p. 11761-11767.

49.   Alvarez, M., et al., *Triose-phosphate isomerase (TIM) of the psychrophilic bacterium Vibrio marinus.* J. Biol. Chem., 1998. **273**(4): p. 2199-2206.

50.   Aghajari, N., et al., *Crystal structures of the psychropilic alpha-amaylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor.* Protein Science, 1998. **1998**: p. 564-572.

51.   Folcarelli, S., et al., *Effect of Lys--Arg mutation on the thermalstability of Cu,Zn superoxide dismutase: influence on the monomer-dimer equilibrium.* Protein Engineering, 1996. **9**: p. 323-325.

52.   Fields, P.A. and G.N. Somero, *Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic nototothenoid fishes.* Proceedings of the National Academy of Sciences, 1998. **95**: p. 11476-11481.

53.   Giver, L., et al., *Directed evolution of a thermostable esterase.* Proc. Natl. Acad. Sci. USA, 1998. **95**: p. 12809-12813.

54.   Rodriguez, E., et al., *Site-directed mutagenesis improves catalytic efficiency and thermostability of Escherichia coli pH 2.5 acid phosphatase/phytase expressed in Pichia pastoris.* Arch Biochem Biophys, 2000. **382**(1): p. 105-12.

55.   Russell, R.J.M., et al., *Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium.* Structure, 1998. **6**: p. 351-361.

56.   Reardon, D. and G.K. Farber, *The structure and evolution of α/β barrel proteins.* FASEB Journal, 1995. **9**: p. 497-503.

57.   Nagano, N., E.G. Hutchinson, and J.M. Thornton, *Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels.* Protein Sci, 1999. **8**(10): p. 2072-84.

58.   Bork, P., et al., *Divergent evolution of a β/α-barrel subclass: detection of numerous phosphate-binding sites by motif search.* Protein Science, 1995. **4**: p. 268-274.

59.   Janecek, S., *Close evolutionary relatedness among functionally distantly related members of the (α/β)$_8$-barrel hydrolases suggested by the similarity of their fifth conserved sequence region.* FEBS Letters, 1995. **377**: p. 6-8.

60.   Janecek, S., *Invariant glycines and prolines flanking in loops the strand beta 2 of various (alpha/beta)8-barrel enzymes: a hidden homology?* Protein Sci, 1996. **5**(6): p. 1136-43.

61.   Janecek, S., *Sequence similarities in (α/β)$_8$-barrel enzymes revealed by conserved regions of α-amylase.* FEBS Letters, 1993. **316**(1): p. 23-26.

62.   Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases.* Structure, 1995. **3**: p. 853-859.

Chapter 2

DISTRIBUTIONS OF THE STRUCTURAL FEATURES CONTRIBUTING TO THERMOSTABILITY IN MESOPHILIC AND THERMOPHILIC (α/β)₈ BARREL GLYCOSYL HYDROLASES

Nicholas Panasik Jr., Jean E. Brenchley, Gregory K. Farber *

Department of Biochemistry and Molecular Biology, 108 Althouse Laboratory, Pennsylvania

State University, University Park,

PA 16802, USA

2. 1 ABSTRACT

Analysis of the structural basis for thermostability in proteins has come mainly from pairwise comparisons of mesophilic and thermophilic structures and has often yielded conflicting results. Interpretation of these results would be enhanced by knowing the normal range of features found for mesophilic proteins. In order to provide the average and distribution values of structural features among similar mesophilic proteins, we compared the amino acid composition, solvent accessible surface area, hydrogen bonds, number of ion pairs, and thermal factors of 22 structures of $(\alpha/\beta)_8$ barrel glycosyl hydrolases. These distributions are then compared to values from seven $(\alpha/\beta)_8$ barrel glycosyl hydrolases from thermophilic organisms. We find that the distribution of each structural feature is broad within the mesophilic proteins and illustrates the difficulty of making pairwise comparisons of mesophiles to thermophiles where differences for individual proteins may be within the normal range for the group. In comparing mesophiles to thermophiles as a group, we find that thermophilic structures have fewer glycines in a particular region of the structure and higher thermal factors at room temperature. These results suggest the basis for thermostability may be related to protein motion rather than to static features of protein structure.

Keywords: Thermostability; Alpha/beta barrel; Ion pair; Glycosyl hydrolase; Thermal factor

2.2 INTRODUCTION

In recent years, organisms have been isolated from some of the most extreme environments on the planet. Some of these organisms are capable of growing at temperatures above the boiling point of water or below its freezing point. To date, attempts to explain the structural basis of thermostability have come from pairwise comparisons of thermophilic enzymes to their mesophilic counterparts. These pairwise comparisons have led to many different suggestions for the basis of thermostability [1]. The explanations have included the number of ion pairs [2-4], the number and type of hydrogen bonds per residue [5], the amount and type of solvent accessible surface area [6], the amino acid composition [7], the hydrophobic core size [8], the cavity size [9], and the multimer association regions of the protein [10].

Sometimes these suggested causes for thermostability have contradicted each other. In one case, an increased number of ion pairs appeared to be the only identifiable difference that could lead to thermostability in the high temperature citrate synthase enzyme [11], whereas in psychrophilic (cold-active) citrate synthase, an increased number of ion pairs was postulated as the reason behind resistance to low temperature denaturation [12]. While many pairwise comparisons have suggested ion pairs play a crucial role, several researchers have argued against it [13-16].

In another case, a pairwise comparison proposed that hyperthermostability in glutamate dehydrogenase is caused by larger ion pair networks [3]. Unfortunately, studies on this protein at pH values lower than 2.0 show that the protein is still stable after most ion pairs have been neutralized [17].

Conflicting results have also been found in amino acid composition. Lysine replacement by arginine and the loss of amino acids that can undergo chemical modification at high temperatures, such as glutamine, have been proposed as global mechanisms for proteins to adapt to high temperature [18]. In some cases, proteins from thermophiles meet this criterion; in other examples, they do not [7]. Small internal cavity sizes have been observed in some thermostable enzymes [19], but even smaller cavity size ratios have been found in less thermostable mesophilic proteins [10]. In each case, where a structural basis for thermostability is proposed, other cases can be found where the mechanism does not hold.

This wide array of conflicting explanations raises the question as to whether the observed differences are the cause of thermostability or whether they are due to the limitations of pairwise comparisons.

One possibility for these differing results is that each protein takes its own structural and evolutionary route to thermostability and further analysis to find global trends will be fruitless. Another possibility is that pairwise comparisons may not contain enough information to reveal the causes of thermostability. Structural arguments for the basis of thermostability are often given based on observed differences in the number and type of a particular structural feature. If a single comparison shows a difference in the number of ion pairs, then this might be interpreted as a cause for thermostability. It may be that among the mesophilic proteins alone there are variations of many more ion pairs with no concurrent increase or decrease in thermostability. Any interpretations based only on pairwise comparisons are subject to misinterpretation or overrepresentation. It must be determined whether the mesophilic structures and their

thermophilic counterparts chosen for these studies are representative for their temperature optimum.

A second complication with pairwise comparisons in determining general rules for thermostability is that the thermostability literature uses examples from many different protein folds. Structural mechanisms that evolution has employed to achieve thermostability could vary from one type of protein fold to another. If so, extrapolating information from one protein fold to another may be misleading. An analysis looking at the structural features possibly related to thermostability focusing on one protein fold would solve this problem. With the recent accumulation of many high resolution structures, it has now become possible for such an analysis.

A summary of the values found for structural features within a class of mesophilic proteins would be useful. For example, having these data readily available would serve as a reference to help determine whether a difference observed in the number of ion pairs or hydrogen bonds between a thermophilic protein and its mesophilic counterpart is significant. Considering the high level of interest in discovering the molecular basis of thermostability and in genetically engineering changes in thermostability, having an average analysis of each feature would serve as a guide for future experiments.

In order to provide a reference point for examining the significance of different values, we first compared 22 mesophilic structures all from the same family: the eight stranded α/β barrel glycosyl hydrolases. Eight stranded α/β barrels form the largest group of protein structures solved to date and comprise perhaps 10% of all known proteins [20]. The glycosyl hydrolases are structurally and mechanistically well characterized [21, 22]. For this group, we examine the

secondary structure, the number and type of hydrogen bonds, the number of ion pairs, the solvent accessible surface area, the amino acid composition, the turn length, and the B factors to find an average and a distribution among the mesophiles. These data provide a spectrum of values found for mesophilic enzymes. We were also interested in determining whether a similar analysis of seven thermophilic enzyme structures of the same fold and family showed any marked differences from our results for the mesophilic structures. We use this analysis of both groups to determine if any of these features are sufficiently different to identify an enzyme as thermostable as well as to validate or invalidate the use of pairwise assessments of structural differences causing thermostability.

2.3 EXPERIMENTAL PROCEDURES

Criterion for selecting structures

All $(\alpha/\beta)_8$ barrel glycosyl hydrolases in the Brookhaven Protein Database as of March 2nd, 2000 were used. Only those X-ray structures that had resolution better than 2.8 Å were selected. All but five structures in this survey have better than a 2.4 Å resolution. Candidate structures were assessed for good stereochemistry based on torsional angles, bond angles, and bond lengths. Native structures of enzymes were chosen over structures solved with substrate or inhibitors bound when possible. PDB entries 1bya, 1exp, and 2myr had substrate or inhibitor present. Table 2.1 shows the enzymes, their host organisms, and the optimal growth temperature for that organism.

Table 2.1 Structures Analyzed

	E.C. Number	Source Species	Environmental Temperature	PDB Code	Resolution
Mesophilic Enzyme					
α-Amylase (taka amylase)	3.2.1.1	*Aspergillus oryzae*	36[a]	6taa	2.1
Acid α-amylase	3.2.1.1	*Aspergillus niger*	26[a]	2aaa	2.1
Human pancreatic α-amylase	3.2.1.1	*Homo sapiens*	37[b]	1hny	1.8
Human salivary amylase	3.2.1.1	*Homo sapiens*	37[b]	1smd	1.6
α-1,4 glycan-4-glucanohydrolase	3.2.1.1	*Hordeum vulgare*	20-37[h]	1amy	2.8
Pig α-amylase	3.2.1.1	*Sus scrofa*	37	1pif	2.3
Soybean β-amylase	3.2.1.2	*Glycine max*	20-37[h]	1bya	2.2
β-1,4-D-glycanase CEX-CD	3.2.1.8	*Cellulomonas fimi*	30[c]	1exp	1.8
Endo-1,4-β-xylanase 3.2.1.91	3.2.1.8	*Penicillium simplicissium*	24[a]	1bg4	1.75
Hevamine A	3.2.1.14	*Hevea brasiliensis*	20-37[h]	2hvm	1.8
Chitinase A	3.2.1.14	*Serratia marcescens*	40	1ctn	2.3
β-Glucosidase A	3.2.1.21	*Bacillus polymyxa*	40	1bga	2.4
Cyanogenic β-glucosidase	3.2.1.21	*Trifolium repens*	20-37[h]	1cbg	2.15
1,3-β-Glucanase	3.2.1.39	*Hordeum vulgare*	20-37[h]	1ghs	2.3
1,4-β-D-glucan maltotetrahydrolase	3.2.1.60	*Pseudomonas stutzeri*	35	2amg	2.2
1,3-1,4-β-Glucanase	3.2.1.73	*Hordeum vulgare*	20-37[h]	1ghr	2.2
1,3-1,4-β-Glucanase	3.2.1.73	*Hordeum vulgare*	20-37[h]	1aq0	2.0
6-Phospho-β-D-galactosidase	3.2.1.85	*Lactococcus lactis*	40[d]	1pbg	2.3
Endo-β-*n*-acetyl glucosaminidase F1	3.2.1.96	*Flavobacterium meningosepticum*	37[d]	2ebn	2.0
Endo-β-*n*-acetylglucosaminidase H	3.2.1.96	*Streptomyces plicatus* ATCC 10739	50-55[d]	1edt	1.9
Thioglucoside glucohydrolase	3.2.3.1	*Sinapis alba*	20-37[h]	2myr	1.6
Cyclodextrin glucanotransferase	2.4.1.19	*Bacillus* sp 1011	37[e]	1pam	1.8
Endo-1,4-β-D-glucanase	3.2.1.4	*Acidothermus cellulolyticus*	60[f]	1ece	2.4
Thermophilic enzyme					
Endo-1,4-β-glucanase CELC	3.2.1.4	*Clostridium thermocellum*	60	1cec	2.15
Endo-1,4-β-D-glucanase	3.2.1.4	*Thermomonospora fusca*	80	1tml	1.8
β –Glycosidase	3.2.1.23	*Sulfolobus solfataricus*	87	1gow	2.6
β –Glycosidase*	3.2.1.23	*Thermosphaera aggregans*	80	1qvb	2.4
α-Amylase*	3.2.1.1	*Bacillus licheniformis*	70	1bli	1.7
Cyclodextrin glycosyltransferase	2.4.1.19	*Thermoanaerobacterium thermosulfurigenes* EM1	70[g]	1ciu	2.3

All environmental temperatures were taken from Bergey's Manual [37] except where noted. [a]From Jong and Gantt [38]. [b]Grays Anatomy [39]. [c]White et al. [40]. [d]Barlows, 1992 [41]. [e]Haga et al. 1994 [42]. [f]Eppard et al. [43]. [g]Knegtel et al. [44]. [h]Plant species are normally considered to grow optimally in the mesophilic range.

Hydrogen bonds

The WHATIF program [23] was used to identify all hydrogen bonds in the structures. Hydrogen bonds had to satisfy the following criteria: a maximal donor to acceptor distance of 3.5 Å, a maximal hydrogen to acceptor distance of 2.5 Å, a maximal donor to hydrogen to acceptor angular error of 60.00°, and a maximal hydrogen to acceptor to the atom covalently bound to the acceptor atom angular error of 90.00°. Sulfurs participating as hydrogen bond donors were considered uncharged.

Ion pairs

A program written by the authors was used to calculate the distance between all charged atoms in the molecules. The program listed all interactions less than 4 Å [24] between oxygen or nitrogen atoms of Lys, His, Arg, Asp, Glu, the carboxy-terminal oxygens and amino-terminal nitrogen. The net number of ion pairs per residue was defined as the number of interactions between oppositely charged side chains minus the number of interactions between side chains of like charge, divided by the total number of residues.

Solvent accessible surface area

Solvent accessible surface area for each structure was determined using the ASC program [25, 26] with a probe radius of 1.4 Å. In the coordinate file of 1ghs, there were 25 side chain atoms with an occupancy of zero. The occupancy of these atoms were set to one for the purposes of this calculation as the main chain atoms are defined and the disordered side chain atoms should contribute to solvent accessible surface area.

Assignment of secondary structure

Secondary structure was assigned based on the Phi and Psi and angles of each residue within the structures using PROCHECK [27]. α helices and 3/10 helices were included as helices. The helices and β strands which formed the walls of the barrel were assigned by visual inspection. In structures 1hny, 1amy, 2ebn, 1edt, 1ece, and 1tml, the residues that would form helices 5 or 6 were not listed as helices by PROCHECK because of interactions with a second domain. In those cases, glycines were assigned as being in either the carboxy-terminal or at the amino-terminal side of the barrel by visual inspection. (In 1hny, glycines 238, 239, 259, 263, and 271 were assigned at the carboxy-terminal side, while glycines 251, 281, 283, and 285 were assigned at the amino-terminal side. In 1amy, glycines 250 and 258 were assigned to the carboxy-terminal side, while glycines 268, 272, and 275 were assigned to the amino terminal side. In 2ebn, glycines 200 and 201 were assigned to the carboxy-terminal side, while glycine 216 was assigned to the amino-terminal side. In 1edt, glycines 171 and 197 were assigned to the carboxy-terminal side, while glycines 180 and 181 were assigned to the amino-terminal side. In 1ece, glycines 208, 214 and 215 were assigned to the carboxy-terminal side. Finally, in1tml, glycines 223, 225, 228 and 236 were assigned to the carboxy-terminal side, while glycine 248 was assigned to the amino-terminal side).

Thermal factors

The average thermal factors for main chain and side chain protein atoms were then calculated using the BAVERAGE program of the CCP4 program suite [28]. All hetatoms and hydrogens were removed prior to this calculation using PDBSET of the CCP4 package.

2.4 RESULTS

Hydrogen bonds

Hydrogen bonds were divided into four categories: hydrogen bonds between two charged atoms, hydrogen bonds between a charged atom and an uncharged atom, hydrogen bonds between two uncharged atoms and hydrogen bonds between two protein backbone atoms (Table 2.2). The mesophilic structures show a large distribution in the number of hydrogen bonds in each of the four categories. The average number of these bonds in thermophiles falls within one standard deviation of the mean found for the mesophiles.

Ion pairs

Ion pairs examined in each structure were divided into three categories: number of oppositely charged interactions, number of like (repelling) charge interactions, and the net sum of these two (unlike-like) (Table 2.3). The average number of ion pairs per residue was found to be 0.047, similar to the 0.04 ion pairs per residue found in an earlier study of 38 high resolution protein structures by Barlow and Thornton [24]. However, the number of ion pairs per residue shows a large distribution within the mesophilic structures varying from as few as 0.028 ion pairs per residue to as many as 0.096. The average number of ion pairs per residue for the thermophilic structures is found to be within one standard deviation of the mean for the mesophilic structures. Three of the thermophilic structures have fewer ion pairs per residue than the average for mesophiles and four are above.

Table 2.2 Hydrogen Bond Summary

File	Res	Total	#H/res	%C-C	%C-U	%U-U	%B-B	#C-C	#C-U	#U-U	#B-B
Mesophiles											
6taa	476	403	0.847	4.5	17.4	20.8	57.3	18	70	84	231
2aaa	484	416	0.860	5.1	16.6	19.2	59.1	21	69	80	246
1hny	496	489	0.986	5.7	20.9	21.9	51.5	28	102	107	252
1smd	496	504	1.016	6.7	23.0	21.0	49.2	34	116	106	248
1amy	403	367	0.910	7.4	20.4	11.2	61.0	27	75	41	224
1pif	496	466	0.940	5.6	20.4	20.8	53.2	26	95	97	248
1bya	495	437	0.883	5.5	14.4	16.9	63.2	24	63	74	276
1exp	312	310	0.994	5.8	19.4	14.2	60.7	18	60	44	188
1bg4	302	313	1.036	6.4	19.8	14.7	59.1	20	62	46	185
2hvm	273	239	0.875	2.5	13.8	21.8	61.9	6	33	52	148
1ctn	561	479	0.854	5.6	16.1	17.8	60.5	27	77	85	290
1bga-A	447	450	1.007	7.1	12.2	17.3	63.3	32	55	78	285
1bga-B	447	446	0.998	7.2	12.3	17.0	63.5	32	55	76	283
1bga-C	447	447	1.000	7.2	12.5	17.0	63.3	32	56	76	283
1bga-D	447	448	1.002	7.1	12.5	17.0	63.4	32	56	76	284
1cbg	490	459	0.937	7.8	16.8	13.7	61.7	36	77	63	283
1ghs-A	306	257	0.840	3.5	12.8	17.1	66.5	9	33	44	171
1ghs-B	306	273	0.892	3.7	11.4	20.5	64.5	10	31	56	176
2amg	418	389	0.931	9.8	20.3	13.4	56.6	38	79	52	220
1ghr	306	247	0.807	1.2	7.3	20.2	71.3	3	18	50	176
1aq0-A	306	255	0.833	2.0	8.2	21.6	68.2	5	21	55	174
1aq0-B	306	255	0.833	1.6	9.0	22.0	67.5	4	23	56	172
1pbg-A	468	445	0.951	7.6	16.0	12.6	63.8	34	71	56	284
1pbg-B	468	443	0.947	8.1	15.1	14.0	62.8	36	67	61	279
2ebn	289	248	0.858	5.2	14.9	14.5	65.3	13	37	36	162
1edt	270	230	0.852	2.2	14.8	18.3	64.8	5	34	42	149
2myr	501	496	0.990	5.6	15.7	18.8	59.9	28	78	93	297
1pam	686	661	0.963	4.8	19.9	21.5	54.4	33	131	135	357
Average			0.924	5.4	15.6	17.9	61.1				
S. D.			0.068	2.1	4.1	3.3	5.2				
Thermophiles											
1ece	358	354	0.989	4.5	19.2	21.7	54.6	16	68	76	194
1cec	343	323	0.942	9.0	19.5	11.8	59.8	29	63	38	193
1tml	286	260	0.909	3.1	19.2	15.0	62.7	8	50	39	163
1gow	489	495	1.012	8.3	17.8	16.4	57.6	41	88	81	285
1qvb	481	433	0.900	7.6	15.7	13.4	63.5	33	68	57	275
1bli	483	524	1.084	8.4	24.8	17.6	49.1	44	130	93	257
1ciu	683	640	0.937	5.0	19.8	21.6	53.6	32	127	138	343
Average			0.961	6.2	18.4	16.6	58.8				
S. D.			0.059	1.9	2.8	3.3	4.7				

Res, number of residues in structure; Total, total number of hydrogen bonds calculated for structure; #H/res, number of hydrogen bonds per residue found in the entire polypeptide chain; %C-C, percent of hydrogen bonds between two charged atoms, #C-C is the actual number of these bonds; %C-U, percent of hydrogen bonds between charged and uncharged atoms, #C-U is the actual number of these bonds; %U-U, percent of hydrogen bonds between two uncharged atoms, #U-U is the actual number of these bonds; %B-B, percent of hydrogen bonds between backbone atoms, #B-B is the actual number of these bonds.

Table 2.3 Ion Pair Analysis Summary

File:	+/-	-/-	+/+	Net	Opp/res	Like/res	Net/res
Mesophiles							
6taa	23	2	1	20	0.048	0.006	0.042
2aaa	21	9	1	11	0.043	0.021	0.023
1hny	35	2	5	28	0.071	0.014	0.057
1smd	39	2	3	34	0.079	0.010	0.069
1amy	31	7	1	23	0.077	0.020	0.057
1pif	32	4	2	26	0.065	0.012	0.052
1bya	26	5	3	18	0.053	0.016	0.040
1exp	21	2	1	18	0.067	0.010	0.058
1bg4	19	0	2	17	0.063	0.007	0.056
2hvm	9	2	0	7	0.033	0.007	0.026
1ctn	32	3	3	26	0.057	0.011	0.046
1bga[a]	33	1	5	27	0.074	0.013	0.060
1cbg	30	4	4	22	0.061	0.016	0.045
1ghs[a]	11	0	0	11	0.034	0	0.036
2amg	38	5	0	33	0.091	0.012	0.079
1ghr	5	0	0	5	0.016	0	0.016
1aq0[a]	6	0	1	5	0.020	0.003	0.016
1pbg[a]	47	2	7	38	0.100	0.019	0.081
2ebn	11	2	0	9	0.038	0.007	0.031
1edt	7	1	0	6	0.026	0.004	0.022
2myr	26	2	3	21	0.052	0.010	0.042
1pam[a]	38	2	3	33	0.055	0.007	0.048
Average	25.6	2.2	2.5	20.8	0.057	0.010	0.047
S. D.	12.6	2.2	2.2	10.4	0.023	0.006	0.019
Thermophiles							
1ece[a]	22	4	3	15	0.062	0.020	0.042
1cec	33	2	2	29	0.096	0.012	0.085
1tml	11	3	0	8	0.039	0.011	0.028
1gow[a]	54	1	6	47	0.110	0.014	0.096
1bli	38	10	3	25	0.079	0.027	0.052
1qvb	34	1	5	28	0.070	0.012	0.058
1ciu	36	5	1	30	0.053	0.009	0.044
Average	33.3	2.7	3.0	27.6	0.073	0.015	0.058
S. D.					0.025	0.006	0.024

+/-, Number of oppositely charged atoms within 4 angstroms; -/-, Number of like charged (negative) atoms within 4 angstroms; +/+, number of like charged (positive) atoms within 4 angstroms; net, number of oppositely charged interactions - number of like charged interactions; opp/res, number of oppositely charged ion pairs per residue in polypeptide chain; like/res, number of similarly charged interactions per residue in polypeptide chain; net/res, net ion pairs per residue of polypeptide chain (like - unlike).
[a]Average values for multiple chains.

Solvent accessible surface area

Solvent accessible surface area was calculated for all 29 structures (Table 2.4). Surface area was divided into three categories: percent non-polar, percent polar, and percent charged. The average solvent accessible surface area was found to be 57% non-polar, 32% polar, and 11% charged. These values differed slightly from the 57% non-polar, 24% polar, and 19% charged found in an earlier survey of high resolution structures [29]. The distribution for solvent accessible surface area was found to be small for both the mesophiles and thermophiles. In all three categories, the amounts of surface area for the thermophiles came within one standard deviation of the mean for the mesophilic structures implying there are no significant differences in this area that can account for thermostability.

Amino acid composition

The amino acid composition of each structure is presented (Table 2.5). Arginine had been proposed [30] to replace lysine in thermostable proteins based on its ability to maintain charge and provide an additional hydrogen bond. These preferences were not observed for the thermophilic α/β barrels. Decreases in potentially thermolabile amino acids, like glutamine, were also not found in the thermophilic structures. Amino acid analysis of all 29 structures found only one potential difference between thermophiles and mesophiles. The largest difference in amino acid composition occurs for glycine content (Table 2.5). Glycine appears to be less prevalent in thermophilic structures by a factor of almost one standard deviation. It is known that the loops linking β strands to α helices at either end of the $(\alpha/\beta)_8$ barrel are quite different. Turns at the

Table 2.4 Solvent Accessible Surface Areas

File	% Nonpolar	% Polar	% Charged
Mesophiles			
6taa	57.44	31.58	10.97
2aaa	57.53	33.38	9.09
1hny	56.55	32.28	11.17
1smd	57.55	30.79	11.65
1amy	59.00	26.62	14.38
1pif	57.36	31.86	10.78
1bya	57.64	27.81	14.55
1exp	56.35	30.66	12.99
1bg4	57.84	33.15	9.01
2hvm	56.01	35.18	8.81
1ctn	58.44	30.64	10.92
1bga	57.45	30.68	11.87
1cbg	59.39	26.23	14.38
1ghs	54.92	36.29	8.79
2amg	53.35	36.64	10.00
1ghr	59.06	34.24	6.70
1aq0	58.65	34.81	6.54
1pbg	59.40	24.66	15.94
2ebn	54.82	34.55	10.63
1edt	53.57	33.79	12.64
2myr	56.74	29.68	13.58
1pam	55.84	35.90	8.26
Average	57.04	31.88	11.07
S. D.	1.74	3.36	2.57
Thermophiles			
1ece	58.01	34.25	7.73
1cec	58.62	24.92	16.46
1tml	60.43	32.10	7.47
1gow	58.07	27.78	14.14
1bli	57.40	30.49	12.11
1qvb	59.41	24.95	15.65
1ciu	55.88	38.18	5.95
Average	57.71	31.36	10.93
S. D.	2.06	5.33	4.15

Table 2.5 Amino Acid Composition

File	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
6taa	7.77	1.89	8.82	2.52	2.94	8.82	1.47	5.88	4.20	6.93	1.89	5.46	4.41	3.99	2.10	7.14	8.40	6.09	2.10	7.14
2aaa	7.35	1.89	8.61	3.36	2.94	7.98	1.47	6.09	2.52	7.56	1.68	5.25	3.99	2.52	2.10	11.13	7.77	6.30	2.31	7.14
1hny	5.45	2.22	6.87	3.84	5.25	10.10	2.22	5.66	4.44	5.05	1.82	8.48	4.65	2.63	5.66	6.46	4.24	7.27	3.43	4.24
1smd	5.25	2.22	7.07	4.04	5.25	10.30	2.42	5.45	4.65	4.85	2.02	8.28	4.44	2.22	5.66	6.67	4.44	7.07	3.43	4.24
1amy	8.68	0.74	8.93	4.22	4.22	11.41	3.72	6.45	5.96	6.95	1.49	3.97	4.71	2.98	4.22	3.23	4.22	6.20	3.97	3.72
1pif	6.06	2.42	6.06	4.24	4.85	10.71	1.82	4.85	3.84	5.05	1.82	7.68	4.24	2.83	5.66	7.27	4.44	8.48	3.84	3.84
1bya	6.31	1.22	6.52	5.91	4.28	7.74	2.24	5.50	6.11	9.57	2.24	6.11	6.11	4.28	3.87	4.89	3.46	6.52	2.24	4.89
1exp	14.42	1.28	8.33	4.49	5.13	7.69	1.60	2.56	5.77	6.09	1.60	4.49	3.21	4.49	4.49	5.45	4.81	8.65	2.24	3.21
1bg4	10.96	0.66	5.98	2.66	2.99	8.64	1.99	6.98	6.64	7.31	1.00	7.97	2.99	2.33	2.33	8.64	6.31	6.98	2.33	4.32
2hvm	7.33	2.20	5.13	1.47	3.30	11.72	0.73	7.33	5.13	8.42	0.73	7.69	5.13	4.03	2.56	9.16	4.76	4.40	2.56	6.23
1ctn	9.29	0.74	6.88	3.53	4.65	11.15	1.12	4.46	7.43	7.43	1.86	6.13	4.28	4.46	2.04	5.58	6.32	6.32	2.60	3.72
1bga	5.59	1.12	6.26	6.04	5.15	8.72	3.80	6.71	1.79	6.94	2.68	5.59	4.03	5.37	5.59	4.70	4.92	6.49	3.36	5.15
1cbg	7.35	1.02	6.33	4.90	6.33	7.55	2.65	4.69	6.53	8.98	1.84	5.71	5.31	1.84	5.31	6.12	4.08	4.29	2.45	6.73
1ghs	14.05	0.33	3.92	2.61	4.25	9.80	0.33	5.56	3.27	7.19	1.63	8.82	4.90	3.59	4.25	7.84	4.90	6.86	0.65	5.23
2amg	8.67	0.96	8.43	2.65	4.58	11.57	2.65	3.61	2.41	4.58	1.45	6.51	4.82	5.30	6.27	7.23	2.65	6.75	4.58	4.34
1ghr	14.05	0.65	1.96	1.96	4.58	10.78	2.29	3.59	2.94	5.56	3.59	7.84	5.88	3.27	2.61	7.52	4.90	9.80	1.31	4.90
1aq0	14.05	0.65	1.96	1.96	4.58	10.78	2.29	3.59	2.94	5.56	3.59	7.84	5.88	3.27	2.61	7.52	4.90	9.80	1.31	4.90
1pbg	7.25	0.44	8.35	7.91	5.49	7.69	4.07	6.15	7.03	6.15	1.76	4.62	4.40	2.20	3.08	2.97	4.18	6.15	1.98	8.13
2ebn	9.47	0.35	5.96	2.46	4.56	6.32	1.40	3.51	6.32	9.12	2.11	9.82	4.91	3.16	2.81	7.72	6.67	6.67	0.35	6.32
1edt	12.08	0.00	7.17	3.02	3.77	9.43	1.13	3.40	3.40	7.55	0.75	7.17	3.77	4.91	4.15	6.42	6.04	9.06	0.75	6.04
2myr	4.61	1.60	7.41	3.81	5.41	9.22	2.40	6.41	5.01	7.01	1.60	6.41	5.21	3.21	3.61	6.81	6.81	4.01	2.00	7.41
1pam	8.02	0.29	6.41	2.77	4.81	9.77	1.75	6.27	3.50	5.25	2.04	9.33	4.37	3.79	3.35	5.54	8.89	7.00	1.90	4.96
Average	8.57	1.15	6.73	3.73	4.51	9.39	2.06	5.29	4.71	6.84	1.79	6.83	4.56	3.50	3.89	6.59	5.39	6.73	2.40	5.33
S. D.	2.98	0.74	1.69	1.53	0.91	1.54	0.97	1.34	1.66	1.45	0.62	1.66	0.76	1.05	1.39	1.88	1.63	1.46	1.10	1.41
1ece	5.87	1.12	7.26	1.96	3.35	9.22	1.68	5.31	2.51	8.38	1.40	6.70	5.87	5.59	3.63	7.26	6.15	6.42	5.03	5.31
1cec	4.83	1.51	7.25	9.97	6.04	4.83	2.72	7.85	8.16	8.46	2.11	6.65	3.32	2.11	4.23	4.53	2.42	4.53	2.42	6.04
1tml	15.03	1.40	5.24	3.15	2.10	8.39	2.80	6.64	2.10	4.20	2.80	6.29	6.99	4.55	4.20	9.09	4.20	4.90	2.80	3.15
1gow	5.73	0.20	6.54	6.13	4.70	8.18	2.66	4.09	4.70	7.36	2.25	6.13	5.32	2.25	6.54	6.54	4.09	6.13	3.48	6.95
1ciu	5.42	0.15	6.30	2.20	4.83	9.22	1.61	6.30	3.37	5.42	2.05	8.93	3.66	3.37	2.93	8.20	10.69	7.17	1.90	6.30
1bli	7.04	0.00	7.66	4.97	4.35	9.52	4.97	4.14	5.80	6.00	1.45	4.55	3.11	3.93	4.35	5.59	5.59	6.13	3.52	6.42
1qvb	6.03	0.42	6.03	6.86	4.57	7.28	2.49	4.78	6.03	8.73	2.29	5.61	6.24	1.87	4.99	5.61	2.29	8.52	3.53	5.82
Average	7.14	0.69	6.61	5.03	4.28	8.09	2.70	5.59	4.67	6.94	2.05	6.41	4.93	3.38	4.41	6.69	5.06	6.26	3.24	5.71
S. D.	3.54	0.64	0.84	2.88	1.24	1.63	1.11	1.40	2.17	1.75	0.49	1.33	1.56	1.40	1.14	1.60	2.87	1.34	1.01	1.24
Signif.[a]	-0.40	-0.36	-0.12	0.62	-0.34	-0.92	0.24	0.56	-0.33	-0.06	0.53	0.07	0.62	0.07	0.30	0.28	0.07	-0.62	0.66	0.16

Mesophiles (rows 6taa through 1pam); *Thermophiles* (rows 1ece through 1qvb).

[a]Signif., significance (thermophilic average minus mesophile average)/standard deviation for mesophile.

carboxy-terminal side of the barrel tend to be quite large. The active site is always located at this side of the barrel. The amino-terminal side of the barrel is characterized by smaller loops [20]. Since glycine is often found in tight two-residue turns [31], it might be enriched at the amino-terminal side of the barrel in mesophiles. Further analysis was done to assess glycine content in turns in loop regions in general, and the length of turns at the amino-terminal side of the barrel.

Glycine content of loops in general was found to be similar in both mesophiles and thermophiles (Table 2.6). There were no statistical differences in glycine content of loops at the carboxy-terminal side of the barrel between mesophiles and thermophiles. However, mesophiles have far more glycines located in the short turns at the amino-terminal side of the barrel (Table 2.6). This difference was not caused by differences in the size of the loops found in the thermophilic and mesophilic proteins (Table 2.7).

Table 2.6 Glycine Composition by Location

File	Total Glys	% in loops	% in turn at N-terminus	%in turn at C-terminus
Mesophiles				
6taa	28	75.0	14.3	60.7
2aaa	28	71.4	10.7	60.7
1hny	38	84.2	23.7	60.5
1smd	39	79.5	17.9	61.5
1amy	39	76.9	20.5	56.4
1pif	39	79.5	17.9	61.5
1bya	37	73.0	13.5	59.5
1exp	22	63.6	13.6	50.0
1bg4	25	60.0	20.0	40.0
2hvm	31	66.7	20.0	46.6
1ctn	44	61.4	4.5	56.8
1bga	39	76.9	17.9	59.0
1cbg	36	72.2	13.9	58.3
1ghs	29	65.5	20.7	44.8
2amg	38	68.4	7.9	60.5
1ghr	32	75.0	28.1	46.9
1aq0	32	71.9	28.1	43.8
1pbg	35	74.3	22.9	51.4
2ebn	18	72.2	38.9	33.3
1edt	24	75.0	33.3	41.7
2myr	44	68.1	11.4	56.8
1pam	33	84.8	24.2	60.6
Average		72.5	19.3	53.2
S. D.		6.6	8.2	8.5
Thermophiles				
1ece	29	62.1	13.8	48.3
1cec	16	43.8	0.0	43.8
1tml	23	60.9	4.4	56.5
1gow	39	79.5	10.3	69.2
1bli	32	79.4	9.4	70.0
1qvb	34	94.7	14.7	75.0
1ciu	34	73.5	8.8	64.7
Average		70.5	8.7	61.1
S. D.		16.4	5.2	11.8

Table 2.7 Amino Acid Lengths of Turns at N-terminal Side of $(\alpha/\beta)_8$ Barrels

File	Turn 1	Turn 2	Turn 3	Turn 4	Turn 5	Turn 6	Turn 7
Mesophiles							
6taa	3	2	2	1	7	6	2
2aaa	4	3	2	1	7	6	2
1hny	7	2	2	16	2	12	3
1smd	7	2	2	15	3	6	3
1amy	3	2	4	3	3	15	2
1pif	7	3	2	15	3	6	3
1bya	2	2	2	5	3	2	15
1exp	1	1	4	3	5	3	3
1bg4	1	2	5	3	5	4	3
2hvm	4	2	16	3	4	3	4
1ctn	1	3	4	3	4	5	2
1bga	3	2	5	4	4	7	3
1cbg	3	2	4	1	38	4	3
1ghs	1	4	7	4	2	4	13
2amg	2	2	1	3	2	8	2
1ghr	2	4	6	5	2	7	12
1aq0	2	4	6	5	2	7	13
1pbg	3	1	4	4	40	5	3
2ebn	13	2	2	3	11	9	2
1edt	13	2	2	3	11	7	3
2myr	2	2	4	5	31	3	1
1pam	9	2	3	2	2	6	2
Average	4.2	2.3	4.1	4.9	8.6	6.1	4.5
S. D.	3.6	0.8	3.1	4.4	11.6	3.0	4.3
Thermophiles							
1ece	3	2	5	3	10	4	11
1cec	2	1	6	4	7	1	2
1tml[a]	3	6	3	1	4	11	-[a]
1gow	3	1	4	3	20	2	3
1qvb	2	2	5	3	15	6	5
1bli	3	2	4	3	5	4	4
1ciu	2	2	4	4	2	6	2
Average	2.6	2.4	4.4	3.0	8.6	4.8	4.5

[a]Sheet 8 not present.

Thermal factors

Thermal factors for main chain and side chain atoms of each of the structures are tabulated (Table 2.8). The average main chain and side chain thermal factors for the thermophilic structures are found to be consistently higher than for mesophiles (Table 2.8). For the well-refined models considered in this paper, a high thermal factor can be due to three different causes: static disorder due to multiple conformations within the crystal lattice, dynamic disorder due to protein motion, and computational or experimental considerations, such as using a flash frozen crystal or comparing thermal factors from several different refinement programs. The third component makes comparing structures solved by different researchers difficult. To control for the effect of crystal freezing and refinement programs, the average thermal factors of structures solved from unfrozen crystals and refined with the same program (XPLOR) were examined. In this subset of structures, the difference in thermal factors between mesophiles and thermophiles is still obvious. The average main chain thermal factor for the mesophilic structures solved in XPLOR at 294°K was 18.8 with a standard deviation of 4.2 as compared to 23.8 for the thermophilic structures solved at the same temperature and with the same method. Concurrently, the average side chain thermal factor for the mesophilic structures was 20.0 with a standard deviation of 4.4 as compared to 25.8 found for the thermophilic structures.

Table 2.8 Average Thermal Factors for Main Chain and Side Chain Atoms

File	Main Chain	Side Chain	Exp.T	Refinement Program
Mesophiles				
6taa	13.8	15.2	294 K	XPLOR
2aaa	16.6	18.3	294 K	XPLOR
1hny	20.0	22.1	294 K	PROLSQ
1smd	24.0	24.9	100 K	XPLOR
1amy	18.9	19.8	294 K	XPLOR
1pif	32.8	33.7	275 K	XPLOR
1bya	23.6	25.9	294 K	XPLOR
1exp	24.1	25.9	294 K	XPLOR
1bg4	9.7	10.6	176 K	XPLOR
2hvm	11.4	17.9	279 K	XPLOR
1ctn	21.8	26.2	100 K	XPLOR
1bga	17.9	18.9	278 K	XPLOR
1cbg	10.3	11.0	100 K	PROLSQ
1ghs	14.7	15.3	294 K	XPLOR
2amg	20.2	20.9	294 K	XPLOR
1ghr	22.4	22.5	294 K	XPLOR
1aq0	13.1	15.6	294 K	XPLOR
1pbg	16.5	21.1	294 K	XPLOR
2ebn	16.7	19.1	294 K	XPLOR
1edt	24.8	28.3	294 K	XPLOR
2myr	11.4	14.1	100 K	XPLOR
1pam	19.5	21.3	294 K	PROLSQ
Average	18.0	19.9		
S. D.	5.2	5.2		
Thermophiles				
1ece	28.7	29.0	294 K	XPLOR
1cec	27.8	33.3	294 K	XPLOR
1tml	11.5	13.7	294 K	XPLOR
1gow	32.5	34.0	100 K	XPLOR
1bli	19.5	22.7	294 K	XPLOR
1qvb	28.9	30.4	294 K	XPLOR
1ciu	18.9	22.3	294 K	TNT
Average	24.0	26.5		
S. D.	6.8	6.7		

2.5 DISCUSSION

Our comparison shows that the distribution of hydrogen bonds, ion pairs, and amino acid composition of mesophilic α/β barrels covers a large range. The range in surface area is much narrower. The distribution of these structural features does not differ significantly between the mesophilic and thermophilic proteins. Thus, for α/β barrel glycosyl hydrolases, it is not possible to find a structural feature that differentiates thermophiles from mesophiles.

This suggests pairwise comparisons of thermophilic structures to mesophilic structures do not contain useful information in uncovering the structural basis of thermostability. These results concur with the findings that the number of ion pairs and hydrogen bonds do not appear to play a role in thermostability between thermophilic and mesophilic 8 stranded α/β barrel family 10 xylanases [32].

From our comparison of a large group of proteins from the same family, we observed two potentially useful explanations for the difference between thermophilic and mesophilic protein stability: the number of glycines in loops at the amino side of the barrel and thermal factors. Both of these observations can be related to molecular flexibility. The increase in glycine in the turns at the N-terminal face of mesophilic α/β barrels may allow access to conformations that are destabilizing for thermophilic proteins at high temperatures. The fewer glycyl residues in the thermophilic structures may serve to decrease the flexibility of this region and thus provide greater stability at the higher temperatures. Localization of temperature effects to "hinge regions" between elements of secondary structure is also suggested in the work of Fields and Somero [33]. They concluded that some lactate dehydrogenases have adapted to cold temperatures by increasing the flexibility of small areas of the molecule that affect the mobility

of adjacent active-site secondary structural elements. They also observed that the thermostable lactate dehydrogenases have smaller numbers of glycines in the loop regions between helices. For eight stranded α/β barrels, it has been shown earlier that the loops at the amino-terminal face of the barrel are more important for thermostability than those at the other face [34]. While the concept of rigidity as a means of adaptation to high temperatures has been explored by other researchers [35-37], its role as a general mechanism has not been proven experimentally.

A second result from this study also suggests that mobility is important in understanding the difference between mesophilic and thermophilic structures. The average main chain and side chain thermal factors for the thermophilic structures are found to be consistently higher than for mesophiles. This change in thermal factors correlates with the two published structures of a psychrophilic proteins, citrate synthase [12], and malate dehydrogenase [38]. For both citrate synthase and malate dehydrogenase, the thermophilic counterpart had significantly larger thermal factors than the psychrophile. It should be noted that in the case of citrate synthase, both structures were solved in the same laboratory using the same programs [12]. Higher thermal factors for thermostable enzymes as compared to mesophilic enzymes have been seen in other cases as well [3].

One interpretation of these observations is that the thermal factors are dominated by the dynamic component and that thermophilic enzymes are more flexible than mesophilic enzymes at room temperature where structures where determined. This explanation, however, is contradicted by experimental work which demonstrates that protein flexibility decreases as the temperature decreases [36].

There is a second explanation that must be considered. The dynamic component of thermal factors represents the degree of conformational flexibility, while the static component would reflect different protein conformations in each unit cell of one crystal. A recent publication by Zavodszky et al. [39] suggests that thermophilic proteins are indeed more conformationally rigid than mesophilic proteins at the same temperature. Thus, our observations could be explained by multiple conformations for thermophilic proteins at mesophilic temperatures. If this is true, the high thermal factors observed for thermophiles are due to static rather than dynamic disorder. At room temperatures, thermophilic enzymes may be frozen into a number of slightly different conformations. The existence of multiple conformations of the enzyme could also explain why thermophilic enzymes are not functionally active at mesophilic temperatures.

In the comparison of malate dehydrogenases from psychrophile and thermophile mentioned above, researchers also compared relative thermal factors. Here, thermal factors of each atom were divided by the thermal factors for the entire molecule and compared regionally. In this case, the cold-active enzyme appeared to have two-fold higher relative thermal factors indicating increased flexibility of the low temperature enzyme. This observation supports both the idea that flexibility decreases with increasing thermostability and that the static disorder of the crystal may be increased in thermophilic structures solved at room temperature. To confirm this, static disorder should be observable in very high resolution structures of thermophilic proteins. Taken together, these results suggest that there are large differences in protein dynamics between thermophiles and mesophiles. These differences may significantly contribute to enzyme thermostability.

2.6 REFERENCES

1.   Gupta, M.N.E., *Thermostability of enzymes*. 1993, Berlin: Springer.

2.   Perutz, M.F. and H. Raidt, *Stereochemical basis of heat stability in bacterial ferredoxins and in hemoglobin A2*. Nature, 1975. **255**: p. 256-259.

3.   Yip, K.S., et al., *The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures*. Structure, 1995. **3**: p. 1147-1158.

4.   Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs*. Journal of Molecular Biology, 1997. **269**: p. 631-643.

5.   Shortle, D., *Mutational studies of protein structures and their stabilities*. Quarterly Review of Biophysics, 1992. **25**: p. 205-250.

6.   Chan, M.K., et al., *Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferrodozin oxidoreductase*. Science, 1995. **267**: p. 1463-1469.

7.   Fleming, T. and J. Littlechild, *Sequence and structural comparison of thermophilic phophoglycerate kinases with a mesophilic equivalent*. Comparative Biochemical Physiology, 1997. **118A**: p. 439-451.

8.   Eriksson, A.E., et al., *Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic eject*. Science, 1992. **255**: p. 178-183.

9.   Eriksson, A.E., et al., *A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene*. Nature, 1992. **355**: p. 371-373.

10.   Wallon, G., et al., *Crystal structures of Escherichia coli and Salmonella typhimurium 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from Thermus thermophilus*. Journal of Molecular Biology, 1997. **266**: p. 1013-1016.

11.   Russel, R.J.M., et al., *The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiousus at 1.9 Å resolution*. Biochemistry, 1997. **36**: p. 9983-9994.

12.   Russell, R.J.M., et al., *Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium*. Structure, 1998. **6**: p. 351-361.

13.   Dao-pin, S., et al., *Contributions of engineered salt bridges to the stability of T4 lysozyme determined by site directed mutagenesis*. Biochemistry, 1991. **30**: p. 7142-7153.

14.	Lumb, K.J. and P.S. Kim, *Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper.* Science, 1995. **268**: p. 436-439.

15.	Waldburger, C.D., J.F. Schildbach, and R.T. Sauer, *Are buried salt bridges important for protein stability and conformational specificity?* Structural Biology, 1995. **2**(2): p. 122-128.

16.	Stark, W., et al., *The structure of a neutral protease from Bacilus cereus at 0.2 nm resolution.* European Journal of Biochemistry, 1992. **207**: p. 781-791.

17.	Chiaraluce, R., et al., *Acid-induced disassembly of glutamate dehydrogenase from hyperthermophilic archaeon Pyrococcus furiosus occurs below pH 2.0.* European Journal of Biochemistry, 1997. **247**: p. 224-230.

18.	Argos, P., et al., *Thermal stability and protein structure.* Biochemistry, 1979. **18**: p. 5698-5703.

19.	Adams, M.W.W. and R.M. Kelly, *Finding and using hyperthermophilic enzymes.* Trends in Biotechnology, 1998. **16**: p. 329-332.

20.	Reardon, D. and G.K. Farber, *The structure and evolution of α/β barrel proteins.* FASEB Journal, 1995. **9**: p. 497-503.

21.	Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases.* Structure, 1995. **3**: p. 853-859.

22.	McCarter, J.D. and S.G. Withers, *Mechanisms of enzymatic glycoside hydrolysis.* Current Opinions in Structural Biology, 1994. **4**: p. 885-892.

23.	Vriend, G., *A molecular modeling and drug design program.* Journal of Molecular Graph., 1990. **8**: p. 52-56.

24.	Barlow, D.J. and J.M. Thornton, *Ion pairs in proteins.* Journal of Molecular Biology, 1983. **168**: p. 857-885.

25.	Eisenhaber, F., et al., *The double cube lattice methode:efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular systems: handling of singularities and computational efficiency.* Journal of Computational Chemistry, 1995. **16**: p. 273-284.

26.	Eisenhaber, F. and P. Argos, *Improved strategy in analytic surface calculation for molecular systems: handling of singularities and computational efficiency.* Journal of Computational Chemistry, 1993. **14**: p. 1272-1280.

27.   Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structure.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

28.   Collaborative Computational Project, N., *The CCP4 suite: programs for protein crystallography.* Acta Crystallography, 1994. **D50**: p. 760-763.

29.   Miller, S., et al., *Interior and surface of monomeric proteins.* Journal of Molecular Biology, 1987. **196**: p. 641-656.

30.   Folcarelli, S., et al., *Effect of Lys--Arg mutation on the thermalstability of Cu,Zn superoxide dismutase: influence on the monomer-dimer equilibrium.* Protein Engineering, 1996. **9**: p. 323-325.

31.   Richardson, J.S., *The anatomy and taxonomy of protein structure.* Advanced Protein Chemistry, 1981. **34**: p. 167-339.

32.   LoLeggio, L., et al., *High resolution structure and sequence of T. aurantiacus xylanase I implications for the evolution of thermostability in family 10 xylanases and enzymes with (beta) alpha-barrel architecture.* Proteins, 1999. **36**: p. 295-306.

33.   Fields, P.A. and G.N. Somero, *Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenoid fishes.* Proceedings of the National Academy of Sciences, 1998. **95**: p. 11476-11481.

34.   Urfer, R. and K. Kirschner, *The importance of surface loops for stabilizing an eightfold α/β barrel protein.* Protein Science, 1992. **1**: p. 31-45.

35.   Scandurra, R., et al., *Protein thermostability in extremophiles.* Biochimie, 1998. **80**: p. 933-941.

36.   Cao, W., et al., *Cloning and thermostability of TaqI endonuclease isoschizomers from thermus species SM32 and Thermus filiformis TOK-GA1.* Biochemistry Journal, 1998. **333**: p. 425-431.

37.   Nakamura, S., et al., *Improving the thermostability of Bacillus stearothermophilus neutral protease by introducing proline into the active site helix.* Protein Engineering, 1997. **10**: p. 1263-1269.

38.   Kim, S.Y., et al., *Structural basis for cold adaptation.* Journal of Biological Chemistry, 1999. **274**: p. 11761-11767.

39.   Závodszky, P., et al., *Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins.* Proceedings of the National Academy of Sciences, 1998. **95**: p. 7406-7411.

Chapter 3

APPROACHES FOR DECIPHERING THE STRUCTURAL

BASIS OF LOW TEMPERATURE ENZYME ACTIVITY

3.1 ABSTRACT

I will consider why researchers have been unable to substantiate a general paradigm that explains how enzymes can be active at low temperature and discuss how the techniques of directed evolution may yield insights into this question. I will then examine what has been learned from recent directed evolution studies of temperature adaptation. The study of the thermal characteristics of enzymes in this way may lead to a deeper understanding of the interplay between enzyme structure, function, primary sequence, and temperature, as well as open up valuable biotechnological applications.

3.2 WHY IS THERE NO GENERAL PARADIGM FOR CONFERRING

THERMOSTABILITY OR LOW TEMPERATURE ACTIVITY?

As described in previous chapters, there are almost as many proposed mechanisms for achieving thermostability at high temperature and/or low temperature activity as there are examples of extremophilic enzymes. Why does there appear to be no general mechanism? One possibility is that there may be numerous independent mechanisms which could confer these properties. However, it is also possible that the proposed mechanisms result from the interpretation of limited data. For example, my results in Chapter 2 indicated that some of the proposed differences between mesophilic and thermophilic enzymes actually fit within the standard deviation of mesophilic proteins. Thus, the natural variation of many of the proposed structural interactions that might confer these properties is as large or larger than the difference seen when phenotypically different enzymes have been compared to each other.

This conflict may be the result of the methodologies used in these comparisons. If one examines the way comparisons between cold-active enzymes and mesophilic and/or thermophilic homologues have been performed, a noticeable trend appears. The enzyme sequences used in the comparisons are often from organisms found in extremely divergent evolutionary lineages [1]. The organisms used in the comparisons are sometimes from entirely different Domains on the Tree of Life [1].

Even when the enzymes compared are from organisms within one Domain, the taxa are often distantly related [2]. For example, a comparison using taxa from different lineages within the Eubacteria, such as a Gram positive and a Gram negative bacterium, may be equivalent to

using taxa from different Eukaryotic kingdoms, such as a plant and an animal [1, 3]. And even when closely-related enzymes are selected, selective pressures on each enzyme may not be equivalent within the different lineages over time [4]). There could be numerous selective pressures on the enzyme throughout its history that are completely unrelated to temperature adaptation. The enzymes may have experienced selection at the level of the enzyme subunit interface to adapt to evolutionary changes made in another subunit of a hetero-dimer, at the secondary structural level through insertions and/or deletions, or at the DNA level because of constraints due to codon usage or base substitution rates [4, 5]. All these factors are unknowns when dealing with enzymes from distant lineages, and they influence the protein sequence in ways that cannot be predicted.

The fact that the free energy of stabilization of the folded state is in the range of 50-100 kJ/mole, which represents the gain or loss of only a few hydrogen bonds or ion pairs, suggests that the number of sequence changes which are important in cold-activity or thermostability will be small. It would thus be nearly impossible to ascertain which of these changes confer cold-activity or thermostability when the number of sequence differences between enzymes are in the tens or hundreds. Because we can only examine the enzymes from extant organisms and not those from organisms in the intermediate steps in adaptation, an *in vitro* approach may be necessary.

Various combinations of proposed mechanisms for cold-activity and/or thermostability are usually seen in each enzyme family as well. Whether or not all these factors actually contribute to that phenomenon in all enzyme families has not been determined. While it is entirely possible that the basis of either cold-activity or thermostability is the result of multiple

factors working in concert, some of the mechanisms identified may likely be unique to the particular enzyme fold. Fold specific mechanisms would further complicate analysis. For example, the comparison of proteases with amylases may discount the proposal of mechanisms that confer cold-activity because they are not present in both enzyme families.

3.3 SITE-DIRECTED MUTAGENESIS STUDIES OF THE MECHANISM(S) OF LOW TEMPERATURE ACTIVITY AND THERMOSTABILITY

To avoid the problems inherent in comparing natural homologs with a high number of sequence differences, researchers have begun to manipulate DNA *in vitro* to investigate the proposed mechanisms controlling thermal dependence of activity. *In vitro* mutagenesis goes a step beyond pairwise sequence or structural comparisons of cold-active or thermostable enzymes obtained from environmental isolates. Site-directed mutagenesis has been used to test temperature adaptation hypotheses derived from sequence and structural studies. This method allows one to target amino acid residues believed to be involved in ion pairs, hydrophobic interactions, etc.

Increased thermolability in high-temperature enzymes using site-directed mutagenesis has been achieved by targeting ionic networks and salt bridges [6, 7], changing residues at highly conserved positions in the protein [8], altering residues at the subunit interfaces [9, 10], and changing helix-capping residues at the N-terminus [11]. In these cases, the thermal stability of the mutant enzymes has been reduced, sometimes with partial or complete loss of enzyme

activity. All of these studies, however, represent a loss of function of an enzyme property and not a gain of function [12, 13].

There are a few reports of site-directed mutations which have successfully enhanced the thermostability of mesophilic enzymes without loss of function. Studies with *Apr*P, an alkaline protease isolated from *Pseudomonas* sp. KFCC 10818, used site-directed mutagenesis to create an additional disulfide bond. The resultant variant, G199C/F236C, formed a disulfide bond and exhibited improved thermostability (2 to 4.8 times longer half-life) as well as improved kinetic parameters for hydrolysis. No significant differences in the half-life of the variant were observed under reducing conditions, suggesting that the new disulfide bond did contribute to the enhanced stability [14].

In studies of the mesophilic xylose isomerase from *Streptomyces diastaticus* No.7, strain M1033, site-directed mutagenesis was used to introduce prolines separately (G138P, G247D) and in tandem (G138P/G247D) in the turn of a random coil. The G138P variant showed a significant increase in thermostability, and G247D showed an increased catalytic activity with respect to the wild-type enzyme. Significantly, the double mutant, G138P/G247D, displayed even higher activity than G247D and better heat stability than G138P alone, indicating that the advantages of two single mutants can be combined effectively. The half-life of the double mutant was about 2.5-fold greater than the wild-type enzyme. Molecular modeling suggested that the introduction of these proline residues in the turns of a random coil reduced the backbone flexibility. This change in thermostability may, therefore, be explained based upon changes in the molecular rigidity of the molecule [15].

The following cases have used site-directed mutagenesis with psychrophilic enzymes to increase their thermostability. Studies with subtilisin, α-amylase, and triose phosphate isomerase (TIM), have been conducted focusing on the residues involved in ion pairs and hydrophobic interactions [16], co-factor binding [17] and those participating in a hydrogen bonding network [18] respectively.

Site-directed mutagenesis was used to increase the stability of a subtilisin from a psychrophilic *Bacillus* species. A change which increased calcium affinity at the cofactor binding site caused a marked increase in protein stability, in addition to an increase in specific activity. Incorporation of an additional salt bridge, a disulfide bridge, and an aromatic interaction, however, all resulted in mutant enzymes which were less stable or equally stable as compared to the wild-type enzyme while low temperature activity was unaffected [16].

Mutagenesis of cofactor binding sites using an α-amylase from *A. haloplanctis* has also been conducted [19]. Two separate changes of amino acids believed to be stronger ion binding residues were introduced into the protein. Both of these mutations were found to lower cofactor affinity and decrease the mutant enzyme's overall activity.

Site-directed mutagenesis was performed on the psychrophilic triose phosphate isomerase from *Vibrio marinus*. vTIM is highly thermolabile compared to its mesophilic counterparts (half-life of only 10 min at 25 °C) and has an alanine residue in loop 8 (at position 238), whereas all other TIM sequences known to date have a serine in that loop. This serine is believed to function in a hydrogen bonding network. Alanine 258 in vTIM was replaced with serine, and the mutation caused an increase in the enzyme's melting temperature by 5°C. It also caused a slight decrease in the enzyme's specific activity. This may be an example where, due to the loss of a selective

pressure to maintain stability at higher temperature (the optimal growth temperature of *V marinus* is 15℃), mutations were acquired through a process similar to evolutionary drift.

Significantly, some site-directed mutagenesis studies have yielded enzymes with increased stability and an increase in activity. Yet in other cases, increases in stability have merely maintained or decreased activity. Reports indicate that for each success, there have been many more failures. From this and previous examples, it is clear that the current thermostability rules for stabilizing cold-active enzymes are not adequate.

3.4 DIRECTED EVOLUTION

Site-directed mutagenesis efforts for examining thermoactivity have met with limited success. This is not because of poor design, but because the large number of possible single mutations and their combinations prevents the analysis of all possibilities using directed mutagenesis. Recently, a new approach has been developed which holds great potential for determining which residues are involved in conferring low temperature activity upon enzymes. Directed evolution has been used to create a large population of randomly altered enzymes, some of which have increased cold-activity or thermostability. A biological selection for the desired phenotype can be made to isolate the enzyme variant of interest. Through controlling the mutation rate, the number of amino acid changes in the variant can be kept to a minimum, allowing for a practical identification of the non-neutral mutations. This may be the fastest and most effective method for exploring changes which confer temperature adaptation.

The techniques of directed evolution can be used to make mesophilic enzymes cold-active (or cold-active enzymes even more cold-active) and to make cold-active enzymes into

mesophilic or even thermophilic enzymes. There are benefits to both approaches. Making

enzymes more cold-active will show which changes increase low temperature activity. Making a

cold-active enzyme more mesophilic may help uncover the molecular mechanisms leading to

greater thermostability and/or mesophilic temperature activity. Such studies would further

address questions as to whether the mechanisms conferring changes in an enzyme's temperature

of activity are directionally dependent—in other words, if the mechanisms conferring cold-

activity are the reverse of those that confer mesophilic activity, or if there are separate

mechanisms that are directionally specific to the temperature toward which the enzyme is being

evolved.

Directed evolution studies can use many methods. Larger changes made at the DNA level

can be produced through recombination of closely related genes or a collection of mutants made

from the same source [20]. These recombinant genes are often produced by DNAse I treatment

followed by re-assembly through ligation or by PCR [21, 22]. Random priming PCR may be

used to produce recombinant enzymes from closely related genes [23], or through incremental

truncation followed by gene fusion [24]. Oligonucleotide cassette mutagenesis can be used to

randomly alter sites in specific parts of proteins (such as the active site), which can then be

screened for activity [25]. Directed evolution may also be carried out by random mutagenesis at

a low mutation rate, followed by screening for a desired activity under specific, controlled

conditions [26]. The individual mutations resulting from these mutagenesis experiments can then

be recombined *in vitro* using staggered extension PCR (StEP) [27].

There are only a few examples of directed evolution being used to alter cold-active

enzymes. A p-nitrobenzyl esterase from the mesophile in *Bacillus subtilis* has been used as a

starting point for directed evolution experiments involving DNA shuffling [28, 29]. After six

generations of sequential directed evolution, one of the resulting mutant enzymes demonstrated a

14°C increase in melting temperature without loss of activity at lower temperatures. In addition,

at the new optimum temperature, the activity of the enzyme increased by nearly a factor of 30.

Analysis of a thermally adapted esterase, mutant 8g8, indicated that the mutations resulting

from directed evolution served to reorganize the active site and stabilized loops at the surface of the

molecule [30].

Mutations enhancing the thermostability of β-glucosidase A of *Bacillus polymyxa*, a

family 1 glycosyl hydrolase, were derived through random mutagenesis [31]. β-Glucosidase

activity remained after a thermal treatment that inactivated the wild-type enzyme. Replacement

of glutamate at position 96 by lysine and of methionine at position 416 by isoleucine showed

increased resistance to denaturing agents, such as pH and urea, while their kinetic parameters did

not change. The CD spectra indicated that the E96K replacement caused an increase in alpha-

helix content. The authors identified the M416I mutation, but they do not suggest any role that

mutation would play in a mechanism for low temperature activity, other than to say that it is

consistent with the lower amounts of methionine found in some psychrophiles.

Another enzyme that has been explored using directed evolution is a loss-of-function

subtilisin BPN'. An evolved form of the enzyme was produced through random mutagenesis and

screened for restoration of function [32]. The mutant protein had a k_{cat}/K_m value which was

100% higher than wild type when measured at 10°C, and it exhibited three amino acid changes

relative to the wild type enzyme. Single-site mutants of these three amino acid changes showed

that one was negative and two were positive with respect to low temperature activity. The triple

mutant was also less stable than the wild-type at 25°C, though the two positive-acting single mutants had comparable stability to wild-type.

A different subtilisin gene, subtilisin E, was used for DNA shuffling experiments [33]. Ten different mutations within a single gene were analyzed in order to determine their contributions to the protein's thermal stability. Of the six synonymous (mutations in nucleic acid sequence that do not lead to amino acid changes) and four non-synonymous mutations, eight were found to be neutral, and two were found to be the cause of increased thermal stability based on segregation of activity in a subsequent DNA shuffling experiment. A psychrophilic protease, subtilisin S41, was also subjected to directed evolution, by random point mutagenesis followed by saturation mutagenesis [34]. Four of the most thermostable mutants resulting from these experiments contained changes which incorporated hydrophobic residues within an extended mobile loop of the protein. These changes, which were made possible by multiple base substitutions at a single codon site, would not likely have been observed with a simple point mutation library [34]. The finding of mutations in unexpected regions of the protein illustrates the power of random mutagenesis over site-directed mutagenesis where it would have been impossible to predict which interactions to target. In addition, the analysis of phenotypically different enzymes with limited amino acid changes, made possible with directed evolution, allows for the dissection of structural effects due to these mutations without the complications of genetic drift that are observed in natural evolution.

Relatedness of mechanisms leading to thermostability and low temperature

Directed evolution could also be a powerful tool to determine if thermolability and activity at low temperatures can be uncoupled. Most natural enzymes that are cold-active have also been found to be thermally labile, suggesting that the two properties are linked. Psychrophilic enzymes are presumably selected for because of their ability to carry out reactions quickly at that the temperatures that the organism usually encounters. At these low temperatures, there is no direct selective pressure to maintain thermal stability. An organism with enzymes that have lost stability at high temperature is probably at no selective disadvantage so long as their environment remains cold. As a result, the observation that psychrophilic enzymes are usually thermolabile is likely more a result of genetic drift and how selections are done in natural environments than due to requirements of structure. Interestingly, this rational might explain why some mesophilic proteins are less thermolabile and can remain stable through a broader range of temperatures, as many experience selection in environments that fluctuate more widely in temperature. This will be discussed further in Chapter 4.

3.5 CONCLUSIONS

While there are an increasing number of examples of cold-active enzymes reported in the literature, little progress has been made in discerning the specific mechanisms enabling these enzymes to maintain high catalytic activities at low temperatures. Only one molecular mechanism leading to thermostability appeared to work broadly—specifically, an increase in calcium binding which did appear to increase thermostability in a number of cases while low temperature activity may or may not have been affected. Interestingly, the mutations leading to increased affinity for the metal have often been located far away from the metal binding site.

The suite of techniques known as directed evolution offers a powerful way to probe the functional effects of mutations. With only a few mutations resulting in phenotypically different variants, it becomes possible to study the structural effects that these mutations cause and understand the role they play in changing function. In the next three chapters, I will describe the application of directed evolution techniques to better understand the mechanisms of adaptation to low temperature. Because directed evolution produced fewer sequence changes than would have been found if comparing natural homologs, I was able to examine the resulting mutations alone and in combination to understand their functional effects. I will show that by discretely controlling selective pressures it is possible to uncouple thermostability from low temperature activity and show that they are not mutually exclusive. Additionally, new mechanisms of adaptation to low temperature will be proposed and tested.

3.6 REFERENCES

1. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya.* Proc. Natl. Acad. Sci. USA, 1990. **87**: p. 4576-4579.

2. Sheridan, P.P., et al., *Approaches for deciphering the structural basis of low temperature enzyme activity.* Biochimica et Biophysica Acta, 2000. **1543**: p. 417-433.

3. Wheelis, M.L., O. Kandler, and C.R. Woese, *On the nature of global classification.* Proc. Natl. Acad. Sci. USA, 1992. **89**: p. 2930-2934.

4. Li, W.-H., *Molecular evolution*, ed. A.D. Sinauer. 1997, Sunderland, MA 01375: Sinauer Associated, Inc. 487.

5. Kim, S.Y., et al., *Structural basis for cold adaptation.* Journal of Biological Chemistry, 1999. **274**: p. 11761-11767.

6. Pappenberger, G., H. Schurig, and R. Jaenicke, *Disruption of an ionic network leads to accelerated thermal denaturation of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium Thermotoga maritima.* Journal of Molecular Biology, 1997. **274**: p. 676-683.

7. Merz, A., et al., *The hyperthermostable inoleglycerol phosphate synthase from Thermotoga maritima is destabilized by mutational disruption of two solvent-exposed salt bridges.* Journal of Molecular Biology, 1999. **288**: p. 753-763.

8. Satoh, T., et al., *Primary structure, expression, and site-directed mutagenesis of inorganic pyrophosphatase from Bacillus stearothermophilus.* J. Biochem., 1999. **125**: p. 48-57.

9. Shinoda, H., et al., *Hydrophobic interactions of Val75 are critical for oligomeric thermostability of inorganic pyrophosphatase from Bacillus stearothermophilus.* J. Biochem., 1999. **125**: p. 58-63.

10. Kirino, H., et al., *Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, Thermus thermophilus.* European Journal of Biochemistry, 1994. **220**: p. 275-281.

11. Bogin, O., et al., *Enhanced thermal stability of Clostridium beijerinckii alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic Thermoanaerobacter brockii alcohol dehydrogenase.* Protein Science, 1998. **7**: p. 1156-1163.

12.   Cao, W., et al., *Cloning and thermostability of TaqI endonuclease isoschizomers from thermus species SM32 and Thermus filiformis TOK-GA1.* Biochemistry Journal, 1998. **333**: p. 425-431.

13.   Rahman, R.N., et al., *Ion pairs involved in maintaining a thermostable structure of glutamate dehydrogenase from a hyperthermophilic archaeon.* Biochem Biophys Res Commun, 1998. **248**: p. 920-926.

14.   Ko, J.H., et al., *Enhancement of thermostability and catalytic efficiency of AprP, an alkaline protease from Pseudomonas sp., by the introduction of a disulfide bond.* Biochem Biophys Res Commun, 1996. **221**(3): p. 631-5.

15.   Zhu, G.P., et al., *Increasing the thermostability of D-xylose isomerase by introduction of a proline into the turn of a random coil.* Protein Eng, 1999. **12**(8): p. 635-8.

16.   Narinx, E., E. Baise, and C. Gerday, *Subtilisin from psychrophilic antarctic bacteria: Characterization and site-directed mutagenesis of residues possibly involved in the adaptation to cold.* Protein Eng., 1997. **10**(11): p. 1271-1279.

17.   Aghajari, N., et al., *Crystal structures of the psychropilic alpha-amaylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor.* Protein Science, 1998. **1998**: p. 564-572.

18.   Alvarez, M., et al., *Triose-phosphate isomerase (TIM) of the psychrophilic bacterium Vibrio marinus.* J. Biol. Chem., 1998. **273**(4): p. 2199-2206.

19.   Chessa, J.-P., G. Feller, and C. Gerday, *Purification and characterization of the heat-labile α-amylase secreted by the psychrophilic bacterium TAC 240B.* Can. J. Microbiol., 1999. **45**: p. 452-457.

20.   Crameri, A., et al., *DNA shuffling of a family of genes from diverse species accelerates directed evolution.* Nature, 1998. **391**: p. 288-291.

21.   Stemmer, W.P.C., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. **370**: p. 389-391.

22.   Stemmer, W.P.C., *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution.* Proc. Natl. Acad. Sci. USA, 1994. **91**: p. 10747-10751.

23.   Shao, Z., et al., *Random priming in vitro recombination: an effective tool for directed evolution.* Nucleic Acids Research, 1998. **26**: p. 681-683.

24.   Ostermeier, M., A.E. Nixon, and S.J. Benkovic, *Incremental truncation as a strategy in the engineering of novel biocatalysts.* Bioorg. Med. Chem., 1999. **7**: p. 2139-2144.

25.   Palzkill, T. and D. Botstein, *Identification of amino acid substitutions that alter the substrate specificity of TEM-1 beta-lactamase.* Journal of Bacteriology, 1992. **174**: p. 5237-5243.

26.   Chen, K. and F.H. Arnold, *Turning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide.* Proc. Natl. Acad. Sci. USA, 1993. **90**: p. 5618-5622.

27.   Zhao, H., et al., *Molecular evolution by staggered extension process (StEP) in vitro recombination.* Nature Biotechnology, 1998. **16**: p. 258-261.

28.   Arnold, F.H., *Design by directed evolution.* Acc. Chem. Res., 1998. **31**(3): p. 125-131.

29.   Giver, L., et al., *Directed evolution of a thermostable esterase.* Proc. Natl. Acad. Sci. USA, 1998. **95**: p. 12809-12813.

30.   Spiller, B., et al., *A structural view of evolutionary divergence.* Proc. Natl. Acad. Sci. USA, 1999. **96**: p. 12305-12310.

31.   Lopez-Camacho, C., et al., *Amino acid substitutions enhancing thermostability of Bacillus polymyxa beta-glucosidase A.* Biochem J, 1996. **314 (Pt 3)**: p. 833-8.

32.   Almog, O., et al., *Structural basis of thermostability: analysis of stabilizing mutations in subtilisin BPN'.* J Biol Chem, 2002.

33.   Zhao, H. and F.H. Arnold, *Functional and nonfunctional mutations distinguished by random recombination of homologous genes.* Proc. Natl. Acad. Sci. USA, 1997. **94**: p. 7997-8000.

34.   Miyazaki, K. and F.H. Arnold, *Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function.* Journal of Molecular Evolution, 1999. **49**: p. 716-720.

Chapter 4

DIRECTED EVOLUTION OF A THERMOPHILIC β-GALACTOSIDASE FOR INCREASED ACTIVITY AT LOW TEMPERATURES

4.1 ABSTRACT

Recent research using directed evolution has suggested that differences in enzyme activity may be achieved with relatively few amino acid changes. With so few mutations, it becomes possible to ascertain potential mechanisms for low temperature activity. Therefore, I used random mutagenesis methods to increase low temperature activity of a thermophilic family 42 β-galactosidase. The low temperature limit of the mutant enzyme has been extended 25°C below that of its parent while maintaining stability at high temperatures (60°C). The individual mutations responsible for this phenotype were identified as V3E, L4I, F187I, and F258S, and the effects of these four mutations on enzyme kinetics were examined. The K_m values remained unchanged while the k_{cat} value was increased. Titration experiments using dithiobisnitrobenzoic acid (DTNB) to determine the number of cysteine residues accessible in the soluble enzyme suggested that molecular flexibility is also increased. Furthermore, the effects of the individual mutations were found to be additive. This work clearly demonstrates that the temperature range of enzyme activity may be broadened without the loss of thermostability.

Although there is not sufficient data to identify the location of mutations F187I and F258S in relation to structure or the active site, the localization of the other two mutations (V3E and L4I) to the N-terminus of the protein and their resultant effects on enzyme flexibility suggests that the N-terminus of this $(\alpha/\beta)_8$ barrel strongly influences the flexibility and thermostability of these molecules and allows us to propose this as a general mechanism for adaptation to low temperatures. Subsequent saturation mutagenesis in the N-terminal region of the gene for the thermophilic enzyme led to a high percentage (over 50%) of mutant enzymes

that were significantly altered in their thermal characteristics. A statistical comparison of the crystallographic structures of other thermophilic and mesophilic $(\alpha/\beta)_8$ barrel glycosyl hydrolases was then performed, and the results suggested this mechanism may be common to many $(\alpha/\beta)_8$ barrels. The unique structural role that the N-terminus plays in $(\alpha/\beta)_8$ barrel architecture further suggests that some mechanisms leading to low temperature activity and thermostability are protein fold dependent.

4.2 INTRODUCTION

Currently, there are no generally accepted structural models for thermal adaptation [1]. Structural elements such as hydrogen bonds, ion pairs, and buried solvent accessible surface area have all been suggested as factors for determining thermostability, but incorporation of these elements into proteins through "rational design" and/or site-directed mutagenesis has often failed. While structural flexibility is believed to play a role in the temperature dependence of enzyme activity, no consistent identification of areas responsible for the control of this property have been identified.

The reasons for this may be many. Structural comparisons have usually come from natural homologs which have many differences in sequence. Recently, however, researchers have begun to employ directed evolution to create enzyme variants with phenotypic differences caused by only a few amino acid substitutions. This has made it possible to explore more efficiently the determinants of temperature adaptation.

Moreover, these researchers have shown that only a few amino acid substitutions are necessary for significant changes in catalytic efficiency at low temperatures. In one study, only four amino acid changes were responsible for a 9.6 fold increase in enzyme activity at 10°C [2]. This may explain why it has been so difficult to deduce mechanisms of enzyme temperature adaptation from comparisons of naturally occurring homologs that may have upwards of 100 sequence differences.

Discrete control of mutational rates afforded by directed evolution and control over selective pressures gives researchers the potential to create enzyme variants with only a few mutations, but significantly different thermal properties. Recent studies have begun to explore

mechanisms of temperature adaptation by attempting to increase an enzyme's activity at low temperature. In a few cases, mesophilic enzymes have been successfully converted into cold active variants. In even fewer instances, thermophilic enzymes gained function at low temperature. Typical increases in activity have been between two and ten fold at a given lower temperature [2-4]. The low temperature limit (lowest temperature at which the enzyme retains activity) of these enzymes has been extended downwards to between zero and ten degrees in some cases [2, 5], while in others the limit remained unchanged. The location of amino acid substitutions found in these variants range throughout the molecule. In some cases, a mutation could be mapped to the subunit interface [6] while studies on β-glucosidase CelB [5], β-glucoronidase [7] and ornithine carbamoyltransferase [8] identified changes near the substrate binding site. Many mutations in these evolved variants occurred in otherwise unremarkable regions; therefore, no apparent mechanism could be proposed to explain the mutation's effects.

As research reveals new possibilities for the mechanisms of temperature adaptation, it has also begun to challenge the notion that cold-adapted enzymes must be necessarily thermolabile. Random mutagenesis followed by selection has shown that the stability of a psychrophilic protease subtilisin S41 could be enhanced without sacrificing thermostability. The half life of the seven mutation variant at 60°C was 500 times higher than that of wild type [4].

Here, I describe the directed evolution of low temperature activity in a thermophilic β-galactosidase, bgaB, from the hyperthermophilic *Geobacillus stearothermophilus* [9]. BgaB, a thermophilic β-galactosidase, has been classified as a family 42 glycosyl hydrolase based on sequence similarity and hydrophobic cluster analysis according to the classifications of Henrissat

[10]. Family 42 is a member of the 4/7 superfamily, which is characterized as having the $(\alpha/\beta)_8$ barrel enzymatic fold.

The BgaB enzyme has an optimal activity at 70°C and retains 5 percent activity at temperatures as low as 40°C. No activity for this enzyme can be observed at 35°C. Given the low lactose concentrations in hot springs and other environments where *Geobacillus* is found, lactose is probably not this enzyme's natural substrate. This enzyme does, however, hydrolyze lactose as well as the traditional chromogenic substrates X-gal and ONPG. The monomeric form of the enzyme has a mass of 78 kilodaltons but is most likely active as either a dimer or trimer.

In this study, a screening strategy was designed to require an increase in low temperature activity while allowing stability to vary. The gene for this enzyme was subjected to random mutagenesis at mutagenic rates designed to produce between 2 and 3 amino acid changes per gene product. This mutagenic rate is designed to explore the sequence landscape as fully as possible while still making it practical to identify the mutations leading to the phenotypic differences. The mutagenized genes were then ligated into screening vectors and transformed into *E. coli* host cells lacking the *lacZ* β-galactosidase. The resultant transformant libraries were screened at different temperatures for X-gal hydrolysis. Those transformants selected on the basis of a changed phenotype are termed "variants." The genes from these variants were purified and subjected to additional rounds of mutagenesis and screening, termed "generations," until suitably cold-active variants of the thermophilic gene were obtained. In this way, the strategy allowed us to use directed evolution to probe enzyme structure for the amino acid changes that lead to low temperature activity. By characterizing several of these mutations, it was possible to postulate a structural mechanism for low temperature activity that may be general to this fold.

4.3 EXPERIMENTAL PROCEDURES

Incorporation of *Nde*1 and *Hind*III restriction sites into pΔα18

An *Nde*1 restriction site was incorporated into the pΔα18 vector [11] using the Quickchange Site Directed Mutagenesis kit (Stratagene) and self complimentary primers (5'-GATAACAATTTCACACAGGAAACACATATGATTACGAATTCGAGCTCGG forward sequence). The altered plasmids were transformed into *Escherichia coli* DH5α cells and incubated overnight at 37⁰C on Luria-Bertani (LB) plates with 100 μg/mL ampicillin (Fisher Biotech). An isolated colony was selected with a sterile toothpick and inoculated into 3 mL LB media with 100 μg/mL ampicillin and the culture was incubated for 8 hours at 37⁰C. Plasmid was isolated using the Wizard Miniprep System (Promega). The resulting plasmid, designated pΔαN, was then subjected to Quickchange Site Directed Mutagenesis with self complimentary primers (5'-TGCAGGCATGCAAGCTTTGGTGCACTCTCAGTACAATCTGC-3' forward sequence) to restore the *Hind*III restriction site in the multiple cloning region that was deleted in the original creation of the pΔα18 vector [11]. The altered plasmids were transformed into *E. coli* DH5α cells and incubated overnight at 37⁰C on LB plates with 100 μg/mL ampicillin. An isolated colony was selected with a sterile toothpick and inoculated into 3 mL LB media with 100 μg/mL ampicillin and the culture was incubated for 8 hours at 37⁰C. Plasmid DNA was isolated using the Wizard Miniprep System and the resulting plasmid was designated pΔαH.

Construction of a *bga*B screening vector

The p*bga*B plasmid [12] was obtained from Wolfgang Schumann (University of Bayreuth, Bayreuth Germany). An *Nde*1 restriction site was incorporated at the start codon of the

*bga*B gene using the Quickchange Site Directed Mutagenesis kit (Stratagene) and self

complimentary primers (forward sequence 5'-AGCTAGGGGGGAATACATATGAA

TGTGTTATCCTCAATTTGTTACGGAGG). The altered plasmids were transformed into *E.*

coli DH5α cells and incubated overnight at 37°C on LB plates with 100 µg/mL ampicillin (Fisher

Biotech), and 100 µg/mL 8-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (United States

Biological). An isolated colony was selected with a sterile toothpick and inoculated into 3 mL

LB media with 100 µg/mL X-gal and the culture was incubated for 8 hours at 37°C. Plasmid was

isolated using the Wizard Miniprep System (Promega). This plasmid was designated p*bga*b*Nde*.

The *bga*B gene was excised by simultaneous restriction digests of p*bga*b*Nde* with *Nde*1

(Promega) and *Xba*1 (Promega) in Buffer D (Promega). The 2 kb gene fragment was then

purified from a 0.7% gel using the QIAquick Gel Extraction Kit (Qiagen), eluted in water and

then ligated (Epicentre Fast Link ligase, Epicentre Technologies) into a phosphatase treated

pΔαH plasmid vector (Shrimp alkaline phosphatase, Amersham Life Sciences, Arlington

Heights, IL). Recombinant plasmids were transformed into *E. coli* DH5α cells and incubated on

LB plates with 100 µg/mL ampicillin, and 100 µg/mL X-gal. Plasmid was isolated using the

Wizard Miniprep System (Promega). This plasmid was designated pΔαH/*bga*B.

Construction of expression host strain

In order to create a host strain that has no endogenous β-galactosidase production but is

capable of lactose transport, *E. coli* ER2585 cells (genotype *fhu*A2 *gln*V44 e14- *rfb*D1 *rel*A1

*end*A1 *spo*T1 *thi*-1 Δ(*mcr*C-*mrr*)114:IS10 Δ(*lac*I-*lac*A)200) were conjugated with *Salmonella*

typhimurium CH3 F' cells (genotype *gal*E496 *gal*K *met*A22, *met*E55, rpsL102 *xyl*404 H1B *nml*

H2-*enx ilz hsdl*6 HSDS-A29 F' 1020 Δz(h138), Tn:*pro*AB81 *lac*I- *lac*Y+). Isolated colonies of both strains were inoculated into 3 mL LB broth and incubated until an O.D.(600) of 0.6. Cultures were combined and incubated with minimal shaking for 1 hr at 37°C. 100 µL of a 10^{-7} dilution was spread plated onto minimal M9 plates with 10 µg/mL thiamine, 0.2% glucose, and 100 µg/mL tetracycline and incubated overnight at 37°C. The resulting tetracycline resistant colonies that were able to grow with only a thiamine requirement were designated *E. coli* ER2585 F'.

Random mutagenesis

Error prone PCR was carried out using the Diversify PCR Random Mutagenesis Kit (Clontech). Linearized template was prepared by *Sca*1 (Promega) restriction digest of the pΔαH/*bga*b plasmid, purification with the PCR Purification kit (Qiagen) and elution in 10 mM Tris (pH 8.0). Primers DE2-F (5'- TCGCCGCAGCCG-3') and DE2-R (5'- TGTAACCCACTCGTGC-3') were used to allow random mutagenesis of the entire *bga*B coding sequence. Reactions contained 40 mM tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl$_2$, 160 µM MnSO$_4$ (rounds 1 and 2) or 480 µM MnSO$_4$ (rounds 3 and 4), 40 µM dGTP, 1 mM each of dATP, dCTP, and dTTP, 10 µM of each primer, 5 ng of template plasmid, and 10 units of Titanium *Taq* DNA polymerase in a total volume of 25 µL. PCR was carried out on an Eppendorf Master Gradient Thermal Cycler (Eppendorf Scientific) at 95°C for 2 min, with 25 cycles of 95°C for 30 s, 60°C for 30 s, and 68°C for 2 min 45 s, followed by 7 min at 68°C. These conditions should generate an error frequency of approximately 2.3 substitutions per 1,000 base pairs or approximately 1-2 amino acid substitutions per gene copy in rounds one and two

and 3.5 substitutions per 1000 base pairs or approximately 2-3 amino acid substitutions per gene copy in rounds three and four [13].

Library construction

Products from the mutagenic PCR reaction were purified using a Qiagen PCR purification kit and the insert fragment was cut simultaneously with *Sal*1 (New England Biolabs) and *Nde*1 (New England Biolabs) in 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, and 0.1 M NaCl at 37°C for 4 hrs. The resulting DNA was then subjected to electrophoresis on a 0.7% agarose gel and the 2 kb fragment was excised from the gel and extracted using a QIAquick kit (Qiagen). Purified DNA was eluted in water. Ligation reactions were performed using Fastlink ligase (Epicenter Technologies). Vector DNA (the entire pΔαH excluding the region between *Nde*1 and *Sal*1), insert DNA, 10 x ligation buffer, 1 mM ATP, water and 1 unit of enzyme were combined and incubated at 10°C overnight. Resulting plasmids were transformed into *E. coli* ER2585 F' competent cells that were prepared using the Z competent cell procedure kit (Zymor Research) and that had transformation efficiencies of 10^8 cfu per mL. A typical transformation contained 5 μl of ligation reaction and 200 μl of Z competent cells that were incubated at 0°C for 1 hr. These transformed cells were plated onto LB agar plates containing 100 μg/ml ampicillin, X-gal, and 1 mM isopropyl β-D thiogalactoside (IPTG).

Screening and plasmid isolation

The above transformant plates were incubated overnight at 37°C (first generational variants), 24°C (second generational variants), or 18°C (third generation variants). Isolated

colonies that hydrolyzed X-gal were selected with a sterile toothpick, inoculated into 3 mL of LB broth containing 100 µg/ml ampicillin, and incubated for 8 hours at 37°C. Plasmid was isolated using the Wizard Miniprep System (Promega). Complete double stranded sequence of insert regions was obtained by primer walking (Nucleic Acid Facility, Pennsylvania State University).

Construction of over expression host MC1061 (DE3)

In order to create an over expression host with no native β-galactosidase expression, the gene for T7 RNA polymerase under *lac*UV7 promoter control was incorporated into the chromosome of *E. coli* MC1061 cells (genotype F- *hsd*R *mcr*B *ara*D139 Δ(*ara*ARC-*leu*)7679 Δ*lac*X74) using a lambda DE3 lysogenation kit (Novagen).

Construction of over expression vector

In order to express enzyme in quantities suitable for purification and biochemical characterization, over expression vectors were created. Plasmid DNA isolated from evolved mutants was cleaved simultaneously with the restriction endonucleases *Nde*1 (New England Biolabs) and *Sal*1 (New England Biolabs) in Buffer D (Promega) at 37°C for 10 hrs. The resulting DNA was then subjected to electrophoresis on a 0.7% agarose gel and the 2 kb fragment was excised from the gel and extracted using a QIAquick kit (Qiagen). Purified DNA was eluted in water. Ligation reactions were performed using Fastlink ligase (Epicenter Technologies). Vector DNA (the entire pET28a excluding the region between *Nde*1 and *Sal*1), insert DNA, 10x ligation buffer, 1 mM ATP, water and 1 unit of enzyme were combined and incubated at 10°C overnight. Resulting plasmids were transformed into *E. coli* MC1061 (DE3)

competent cells. These cells had been made competent using the Z competent cell procedure kit (Zymor Research) and demonstrated transformation efficiencies of 10^8 cfu per mL. A typical transformation contained 5 µL of ligation reaction and 200 µL of Z competent cells that were incubated at 0°C for 1 hr. These transformed cells were plated onto LB agar plates containing 25 µg/mL kanamycin and 1 mM X-gal.

Over expression and purification

A single isolated colony of wild type or variant was selected with a sterile toothpick and inoculated into 3 mL of LB containing 25 µg/mL kanamycin. The cultures were incubated at 37°C for 6 hours and then inoculated into 500 mL of Terrific Broth containing 25 µg/mL kanamycin in a 2 liter baffled flask. This culture was incubated at 37°C until O.D. (600 nm) reached 0.6. To induce cold shock proteins and bring the culture to lower temperature, the flask was cooled suddenly by swirling under cold (4°C) water for 3 minutes and then incubated at 18°C for 20 minutes. Protein over expression was induced by addition of IPTG (0.2 mM final concentration). The culture was incubated at 18°C for 16 hours. Cells were then harvested by centrifugation for 15 minutes at 6,000 x g in a Beckman centrifuge at 4°C. Supernatant was decanted and cells resuspended in 3 mL of cell resuspension buffer per gram of cell pellet. The cell resuspension buffer contained 50 mM Na_xPO_4 (pH 7.3), 0.3 M NaCl, 10 mM imidazole and recommended amounts of EDTA-Free Complete Protease Inhibitor Cocktail (Roche). Cells were lysed by two passes through a french pressure cell at 20,000 psi. Cell debris was removed through centrifugation for 30 minutes at 15,000 x g. The crude cell lysate was applied to a sepharose imino diacetic acid (Sigma) column (20 mL, 2.5 cm ID x 4 cm) that was pre-

equilibrated with 80 mL of 0.3 M NiCl followed by 100 mL of equilibration buffer containing 50 mM Na_xPO_4 (pH 7.3), 10 mM KCl, 1 mM $MgSO_4$, 0.3 M NaCl, and 10 mM imidazole. The column was washed with 100 mL equilibration buffer followed by 300 mL of 50 mM Na_xPO_4 (pH 7.3), 10 mM KCl, 1 mM $MgSO_4$, 0.3 M NaCl, and 50 mM imidazole. Protein was eluted with 50 mM Na_xPO_4 (pH 7.3), 10 mM KCl, 1 mM $MgSO_4$, 0.3 M NaCl, and 300 mM imidazole. Fractions were analyzed by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and those with greater than 98% purity were combined and dialyzed overnight against 50 mM Na_xPO_4 (pH 7.3) at room temperature. Protein concentration was determined using the Bio-Rad protein assay dye reagent concentrate (Biorad) using bovine serum albumin as a standard.

Purification of non-his tagged enzyme

Crude lysate of non-his tagged enzyme was incubated at 60°C for 20 minutes, centrifuged at 15,000 x g for 25 minutes and then dialyzed against equilibration buffer (50 mM Na_xPO_4 (pH 7.3), 10 mM KCl, 1 mM $MgSO_4$). Protein was applied to a DEAE Sepharose (Pharmacia) column (50 mL, 2.5 cm ID, x 10 cm), washed with 200 mL equilibration buffer. An ionic gradient from 0 to 0.5 M NaCl in equilibration buffer over 1 L was applied and protein eluted between 0.3 – 0.35 M NaCl. The material was then applied to a Sephadex G-75 gel filtration column (Sigma). Column fractions with enzyme activity were combined, concentrated in a Centriprep concentrator (Amicon), and dialyzed against 100 mM Na_xPO_4 (pH 7.3), 10 mM KCl, 1 mM $MgSO_4$.

Thermal dependence of enzyme activity

All assays were performed in triplicate. A reaction contained 1.19 mL of 25 mM Na_xPO_4 (pH 6.4), 50 mM KCl, 1 mM $MgSO_4$, and 2.2 mM of the chromagen ONPG (o-nitrophenyl β-D-galactopyranoside). The reaction was allowed to equilibrate to temperature for 15 minutes in an Isotemp refrigerated circulating water bath (Fisher Scientific) with a thermal accuracy of ±0.05°C. The reaction was started with the addition of 10 µL of enzyme and was quenched after 5.0 minutes with 500 µL of 1 M sodium carbonate and the absorbance at 420 nm was measured against a blank without enzyme.

Enzyme thermostability

All assays were performed in triplicate. 1 mL of enzyme solution was equilibrated at the specified temperature for the specified number of minutes and then a 10 µL aliquot was removed and added to 1.19 mL of 25 mM Na_xPO_4 (pH 6.4), 50 mM KCl, 1 mM $MgSO_4$, and 2.2 mM ONPG that was pre-equilibrated to 55°C. The reaction was quenched after 5.0 minutes with addition of 500 µL sodium carbonate and the absorbance at 420 nm was measured.

Mutational analysis

In order to deduce which mutations were responsible for the increased catalysis at low temperature, each of the mutations were introduced individually or in selected combinations in wild type *bga*B sequence. For each mutation, two self complimentary primers were constructed that had melting temperatures of 69°C or above. These were used in Quickchange Mutagenesis (Stratagene) reactions that contained 1 ng pΔαH/*bga*B plasmid and 1 µg each of the self

complementary primers. The forward sequences of these primers are: for V3E 5'-GGAAACACATATGAATGAGTTATCCTCAATTTGTTACGG-3', for L4I 5'-GGAAACACATATGAATGTGATATCCTCAATTTGTTACGGAGG-3', V3EL4I 5'-GGAAACACATATGAATGAGATATCCTCAATTTGTTACGG-3', K47E 5'-GGGATTTTCAGTTGGAGCGAGATCGAACCG-3', N75S 5'-CCACGGTG TTTATATTAGCTTGGGGACGGCG-3', Y149N 5'- GTGGCATGTTAATAATG AGAATGCATGTCACGTTTCC-3', F187I 5'-CGTTGGGGTACAAACAT TTGGGGACAGCG-3', E197V 5'-GCGATACAATCATTGGGATGTAATTAATC CCCCTAG-3', A204T 5'-CCCCCTAGAAAGACACCAACTTTTATTAATCCATCCC-3', F258S 5'-CCGTTAAACTATTCTCAATGGGCTCAGCATGTAG-3', V309I 5'-GGAGCAGGTAACCTCACATATTAACTGGCGCG-3', E393V 5'-GGATCTAGAATCAAGGCAGTGGTCGCGATC-3', G475D 5'-CGGCAATTT GTTGCTAACGGTGACACTTTGATTGTCAG-3'. 7.5 µL of the mutagenic PCR reaction was transformed into 200 µL *E. coli* ER2585 F' cells and spread onto 4 LB plates with 100 µg/mL ampicillin and 100 µg/mL X-gal. Plates were incubated one each at 37$^\mathbf{o}$C, 30$^\mathbf{o}$C, 24$^\mathbf{o}$C, and 18$^\mathbf{o}$C. Colonies were characterized by variations in the amount of hydrolysis of the chromogenic substrate at each temperature and isolated colonies were selected with a sterile toothpick and inoculated into 3 mL LB media with 100µg/mL ampicillin. Cultures were incubated for 8 hours at 37$^\mathbf{o}$C and plasmid was isolated using the Wizard Miniprep System (Promega).

Steady state kinetics

Reactions contained 25 mM Na_xPO_4 (pH 6.4), 50 mM KCl, 1 mM $MgSO_4$ and ONPG at the following concentrations: 100, 250, 500, and 750 μM, and 1, 1.25, 2.5, 5.0, 7.5, 10, and 15 mM. Absorbance at 420 nm was followed in a water jacketed cuvette holder which was pre-equilibrated for 15 minutes with an Isotemp refrigerated circulating water bath (Fisher Scientific) with a thermal accuracy of ±0.05°C. Final temperature was verified with an electronic thermometer. The extinction coefficient for ONP at pH 6.4 was determined to be 0.8809 $mM^{-1}cm^{-1}$. Kinetic and error analysis was performed using the Nonlin package.

Cysteine titration

The number of reactive cysteine groups in soluble BgaB wild type enzyme and variants was determined by titration of 1.4 μM enzyme in 1 mM dithiobisnitrobenzoic acid (DTNB) (Sigma), 1 mM EDTA (Sigma), and 150 mM Na_xPO_4 (pH 7.3). Absorbance at 412 nm was followed for 1 hour. The number of reactive cysteine groups in denatured BgaB wild type enzyme and variants was determined by titration of 1.4 μM enzyme in same reaction plus 8 M urea. An extinction coefficient of 14.1 mM^{-1} cm^{-1} at 412 nm was used to determine concentration of reacted groups.

Sequencing and sequence alignment

Sequence alignments of family 42 glycosyl hydrolases were compiled using MEGALIGN (DNAStar, Inc) via Clustal W. All DNA sequencing was performed at the Nucleic Acid Facility (The Pennsylvania State University).

Structural comparisons

Structures for $(\alpha/\beta)_8$ barrel glycosyl hydrolases from thermophilic species were identified

in the CAZY database (http://afmb.cnrs-mrs.fr/CAZY/). All structures with N-terminal domains

were rejected. Candidate structures were assessed for good stereochemistry based on torsional

angles, bond angles, and bonds lengths. When possible, native structures of enzymes were

chosen over structures solved with substrate or inhibitors bound. For the purposes of determining

the beginning and end of the beta sheets comprising the alpha/beta barrel, the secondary

structural assignments in the PDB file were used. The WHATIF program [14] was used to assign

all hydrogen bonds using a maximal donor to acceptor distance of 3.5 Å, a maximal hydrogen to

acceptor distance of 2.5 Å, a maximal donor to hydrogen to acceptor angular error of 60.00°, and

a maximal hydrogen to acceptor to the atom covalently bound to the acceptor atom angular error

of 90.00°. Sulfurs participating as hydrogen bond donors were considered uncharged for the

purposes of categorization. A program was written to calculate the distance between all charged

atoms in the molecules. Interactions were defined as occurring between polar oxygen or nitrogen

atoms of potentially ionizable side chains (Lys, His, Arg, Asp, Glu) or C-terminal oxygens and

N-terminal nitrogens that are positioned ≤4 Å apart [15]. The number of ion pairs per residue

was defined as the number of interactions between oppositely charged side chains minus the

number of interactions between side chains of like charge, divided by the total number of

residues [16]. Solvent accessible surface area and surface area buried by the N-terminal region

was determined using the ASC program [17, 18] with a probe radius of 1.4 Å. The N-terminal β

sheet 1 surface was defined as all residues up to the last residue of β sheet 1. All structures with N-terminal domains were rejected.

Saturation mutagenesis of N-terminal region of *bga*B

The *bga*B gene was amplified by PCR using the DE2-R primer and a degenerate primer (5'-CACACAGGAAACACATATGAATNAGNTANCCNCAATTTGTTACGG-3') complimentary to the N-terminal region. A 50 μL reaction contained 1 μM of each primer, 10 ng of linearized pΔαH/*bga*B plasmid DNA and two Ready to Go$_{TM}$ PCR beads (Amersham Pharmacia Biotech). PCR was carried out at 95°C for 5 minutes, with 35 cycles of 95°C for 30 s, 58°C for 30 s, and 72°C for 2.75 minutes, followed by 7 minutes at 72°C. The PCR product was purified using the Qiagen PCR Purification kit, eluted in water, and the insert fragment was cut simultaneously with *Sal*1 (New England Biolabs) and *Nde*1 (New England Biolabs) in Buffer B (Promega). The insert was then purified using the Qiagen PCR Purification kit, eluted in water and ligated (Epicenter Fastlink ligase) into pΔαH overnight at 10°C. Ligation products were transformed into Z competent *E. coli* ER2585 cells and plated onto LB plates containing 100 μg/mL ampicillin, and 100 μg/mL X-gal. Plates were incubated at 18°C for 48 hrs, then successively shifted to 24°C, 30°C, and 37°C for 10 hrs each to monitor hydrolysis of X-gal.

4.4 RESULTS

Generation one variants

In order to obtain enzyme variants that had increased activity at low temperature. The *Bga*B gene was subjected to random mutagenesis by error prone PCR under conditions expected

to yield two to three base substitutions per gene or one to two amino acid substitutions per gene product. A total of 40,000 transformants were screened for hydrolysis of X-gal at 30°C. Of the six transformants showing increased activity, two, designated Gen1-a and Gen1-b were selected for further characterization based on the intensity of the chromogenic indicator (Figure 4.1.).

	18 C	24 C	30 C	37 C
Wild Type				
Gen1-a				
Gen1-b				

Figure 4.1. Colony phenotypes of variants Gen1-a, Gen1-b, and wild type.
Plates contain 100 µg/mL X-gal. These pictures were taken 18 hours after the first round of mutagenesis.

The plasmids from selected variants were purified, their inserts removed using restriction endonucleases and then ligated into pET over expression vectors to create N-terminal 6 x his tagged versions of the gene products. The ezymes were over expressed and purified. Assays of thermal dependency of activity revealed that the optimal temperature for activity decreased by 5°C in Gen-1a (Figure 4.2.) while thermostability assays indicate that Gen1-a lost thermostability (Figure 4.3.). Gene sequencing revealed that this variant had three amino acid substitutions (Figure 4.4.).
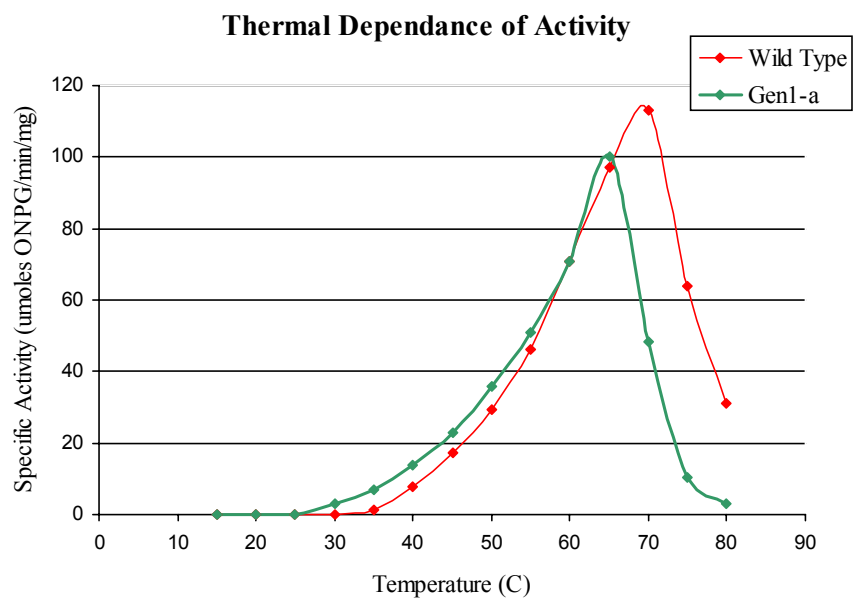
Figure 4.2. Thermal dependence of activity of Gen1-a and wild type.
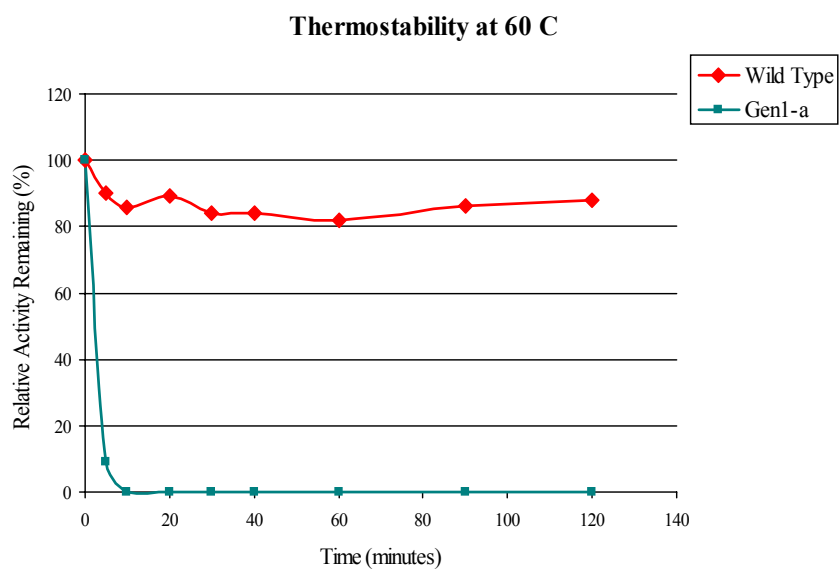
Figure 4.3. Thermostability of Gen1-a and wild type.

Figure 4.4. Amino acid substitutions in Gen1-a.

Generation two variants

Plasmids from all six transformants selected in round one were pooled and subjected to an additional round of mutagenesis with an error frequency similar to that of round one. A total of 41,000 transformants were screened for hydrolysis of X-gal at 24°C and two were selected for further characterization on the basis of increased hydrolysis of X-gal at lower temperatures (Figure 4.5). These were designated Gen2-a and Gen2-j. Characterization of purified his-tagged versions of these enzymes showed that they had a large increase, relative to wild type, in specific activity (Figure 4.6). Gen2-a had almost no change in thermostability properties while Gen2-j lost activity within 20 minutes at 60°C (Figure 4.7). Upon sequence analysis one mutation, V3E, was found to be common to both variants while a second mutation in Gen2-j, L4I, was located adjacent to this mutation (Figure 4.8). Interestingly, a mutation at the third position (V3A) had been found in the first generation variant Gen-1a which exhibited increased activity at low temperatures.
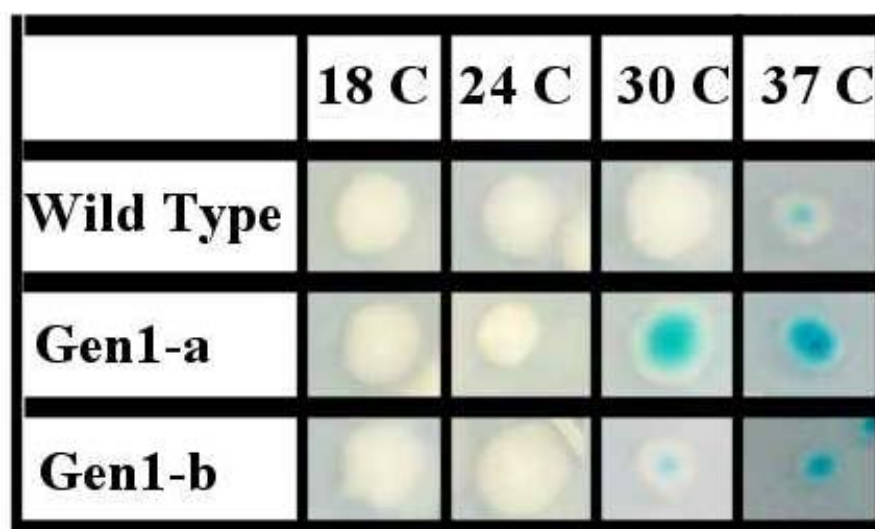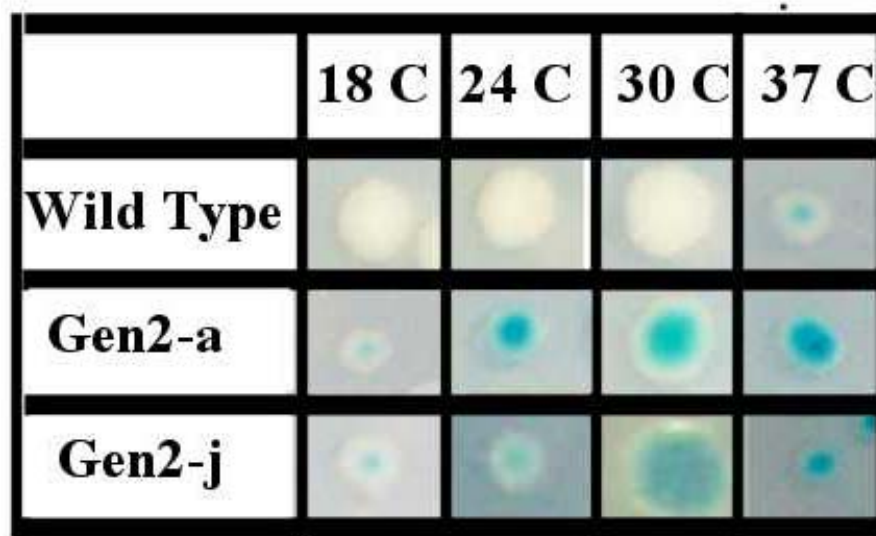
Figure 4.5. Colony phenotypes of variants Gen2-a, Gen2-j, and wild type.
Plates contain 100 µg/mL X-gal. These pictures were taken 18 hours after the first round of
mutagenesis.
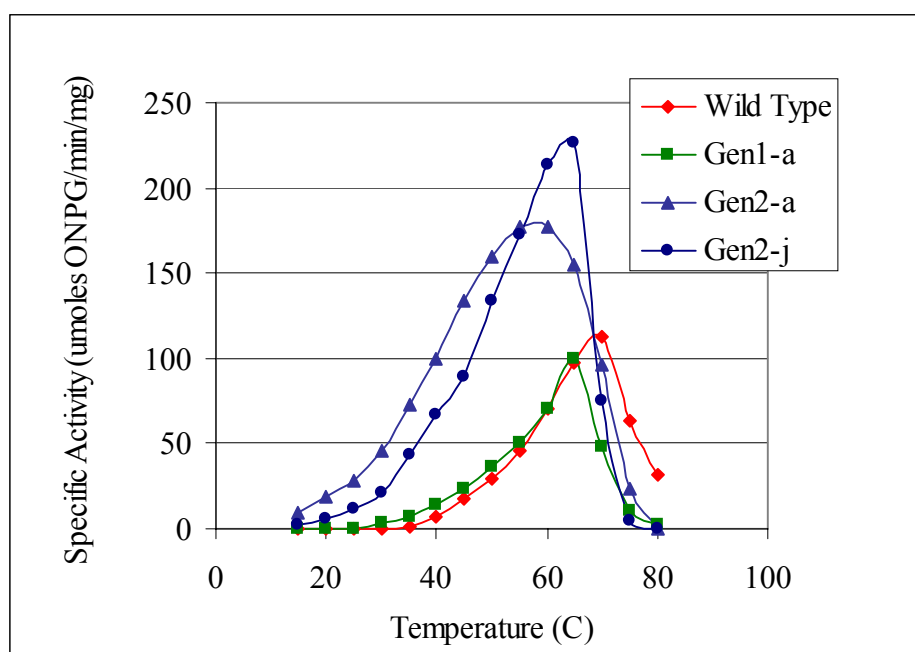
Figure 4.6. Thermal dependence of activity of Gen2-a, Gen2-j, and their parent.
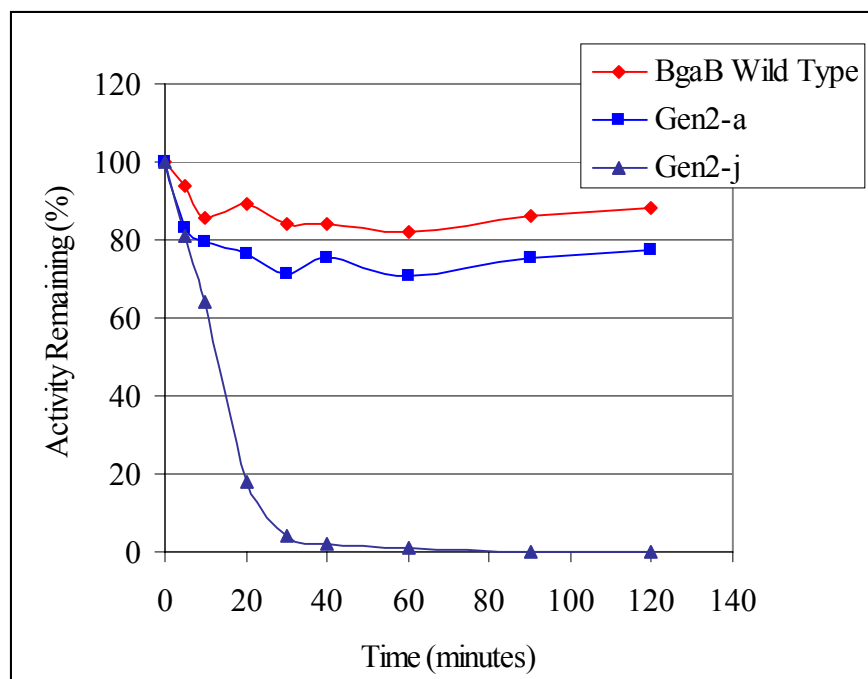
Figure 4.7. Thermostability at 60°C of second generation variants.

Amino Acid Substitutions in 2ⁿᵈ Generation
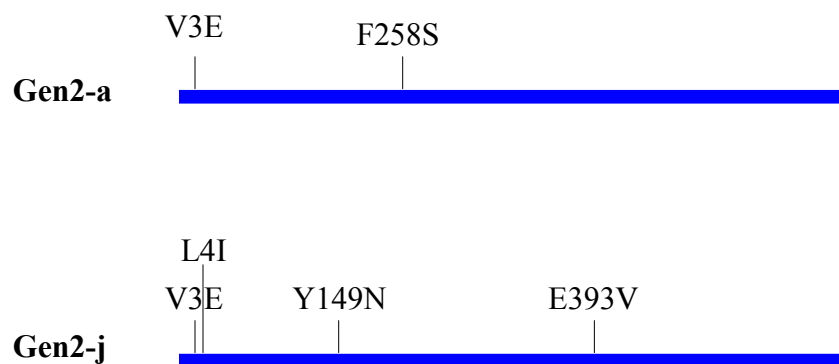
Figure 4.8. Amino acid substitutions in second generation.

Generation three variants

      The Gen2-a gene was used as template for the third round of random mutagenesis. A total

of 25,000 transformants were screened for hydrolysis of X-gal at 18°C and, of the 24 that

exhibited increased hydrolysis of X-gal at lower temperatures, one was selected for further

characterization (Figure 4.9). This variant, designated 2aGen3-e yielded a strong dark blue color

within 32 hours of plating at 18°C. After transfer into a pET28a over expression vector

incorporating a 6 x his tag, the enzyme was over expressed and purified. This variant exhibited

the greatest activity at low temperatures with 15% of activity remaining at 18°C (Figure 4.10).

This represents a 25°C extension in the low temperature range of the enzyme (wild type enzyme

has 15 % activity remaining at 43°C. A loss of some thermostability is observed (Figure 4.11)

but given the high number of mutations (Figure 4.12) it was deemed possible that by removing

excess mutations, the enzyme thermostability may be restored.
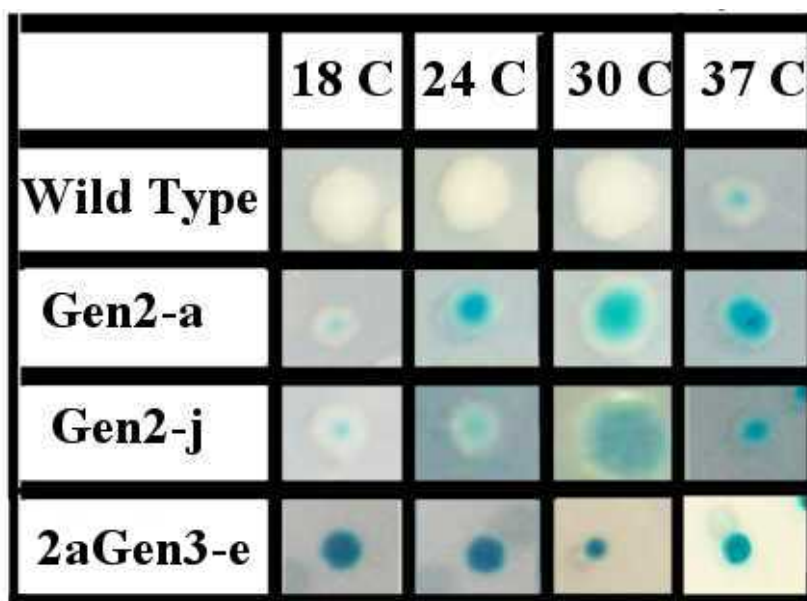
Figure 4.9. Colony phenotypes of second and third generation variants.
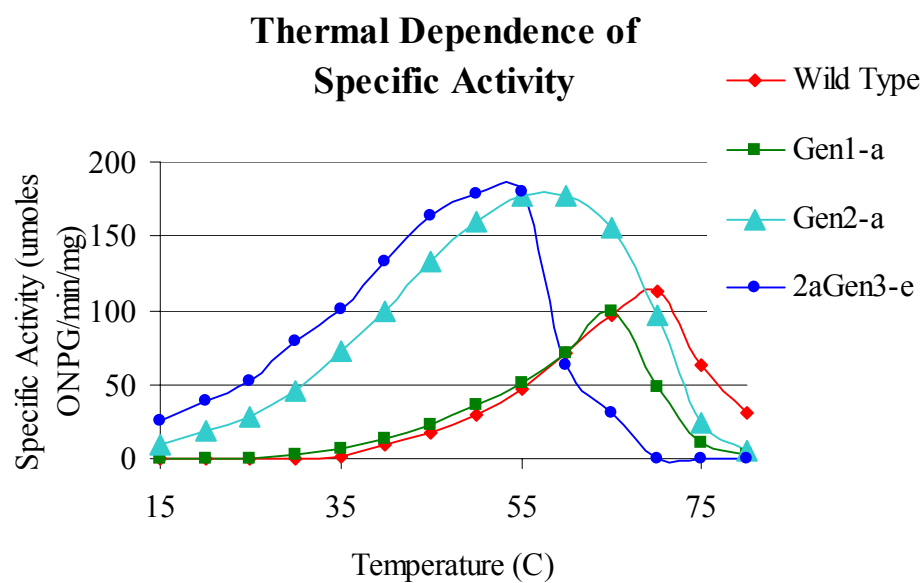
Figure 4.10. Thermal dependence of specific activity for directly evolved variants.
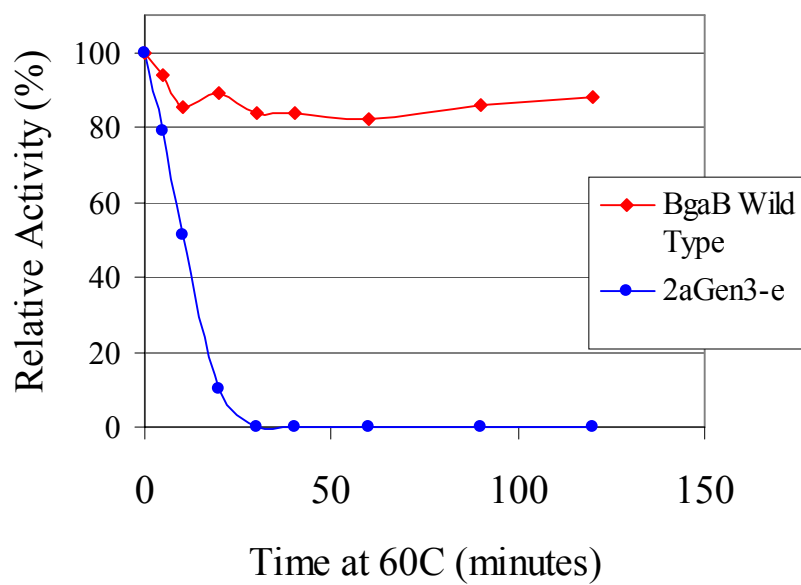
Figure 4.11 Thermostability of 2aGen3-e

3rd Generation Amino Acid Substitutions
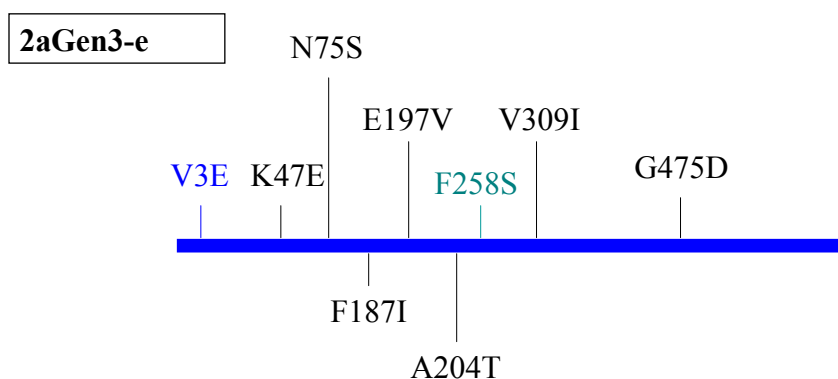
Figure 4.12. Amino acid substitutions found in 2aGen3-e variant.

Mutations shown in blue and green were inherited from Gen2-a

Mutational analysis

In order to ascertain the significance of each mutation and its structural effect on function, each mutation in Gen2-a, Gen 2-j, and 2aGen-3e was incorporated separately into wild-type enzyme. These variants were screened for hydrolysis of X-gal at 24°C, 30°C, and 37°C to assess their relative contribution (Figures 4.13, 4.14; Table 4.1). Mutations were also combined to assess if their effects were synergistic or dependent upon other mutations. V3E, L4I, F187I, and F258S were found to yield increased activity at low temperature, supporting hydrolysis at 30°C, 24°C, 30°C, and 24°C, respectively. When combined, their effects were additive and supported hydrolysis of chromogen at even lower temperatures. On chromogenic plates, these

mutations are sufficient to explain the low temperature activity characteristics found in their

respective parental variants.
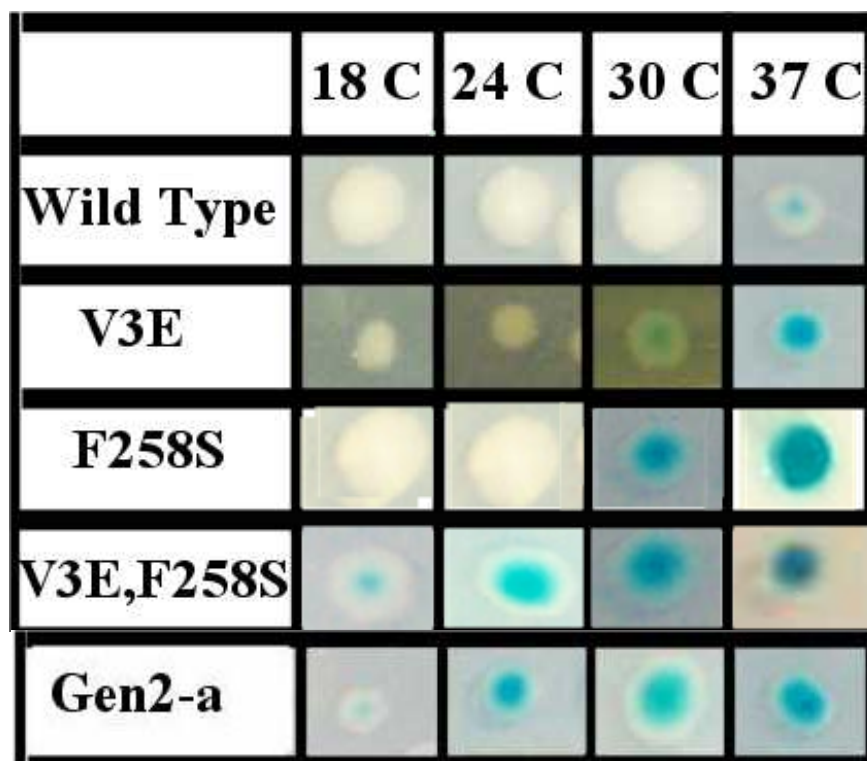
Figure 4.13. Analysis of mutations found in Gen2-a.
The activity of the parent, Gen2-a, at 18°C may be explained by the V3E F258S combination.
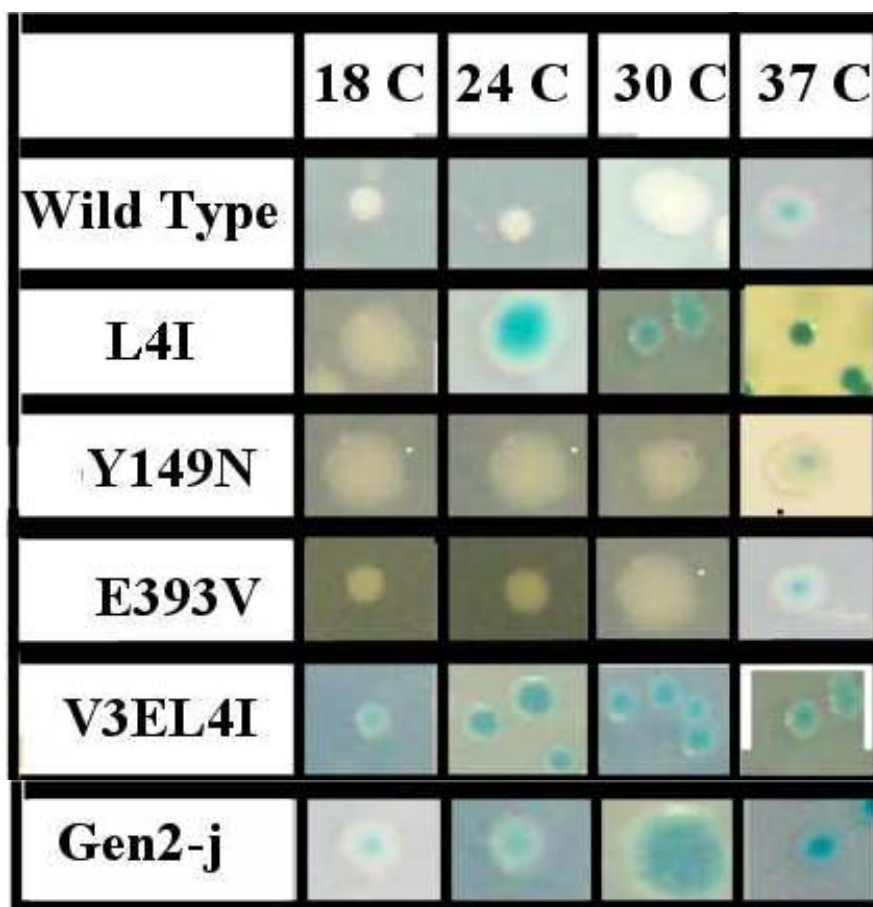
Figure 4.14. Analysis of mutations found in Gen2-j. The activity of Gen2-j at
18°C may be explained by the V3E L4I combination.

Table 4.1. Analysis of mutations found in 2aGen3-e.
The activity of 2aGen3-e at 18°C may be explained by the V3E F187I F258S combination.

Mutation	Colony color at	
	30 °C	37 °C
Wild type	white	Lt blue
K47E	white	Lt blue
N75S	white	Lt blue
Y149N	white	Lt blue
F187I	Lt blue	Dk blue
A204T	white	Lt blue
V309I	white	Lt blue
F258S	white	Lt blue
V3EF187I	Dk blue	Dk blue

In order to determine whether or not some mutations function only in tandem with others, each mutation that was found to yield no additional hydrolysis of X-gal was incorporated separately into each of the other single mutants (i.e. K47E was combined separately with V3E, N75S, Y149N, F187I, A204T, F258S, and V309I). In no case, was additional activity observed upon transformation and incubation on X-gal plates at 30°C or 24°C.

Thermal dependence of activity and thermostability of cold activating mutations

V3E, L4I, F187I, and F258S were found to yield increased cold-active properties (Figures 4.13, 4.14; Table 4.1). Genes containing these mutations singly and in combinations that reflected their appearance in directed evolution variants and their structural proximity were transferred into over expression vectors, over expressed with N-terminal 6 x his tags, and purified for enzymatic characterization. The specific activity of each of these enzymes was found to increase at lower temperatures (Figure 4.15) with no loss of thermostability (Figure 4.16). L4I had the greatest single overall effect and each of the mutational effects were additive. The enzyme that contained all four of these amino acid substitutions was the most active at lower temperatures.
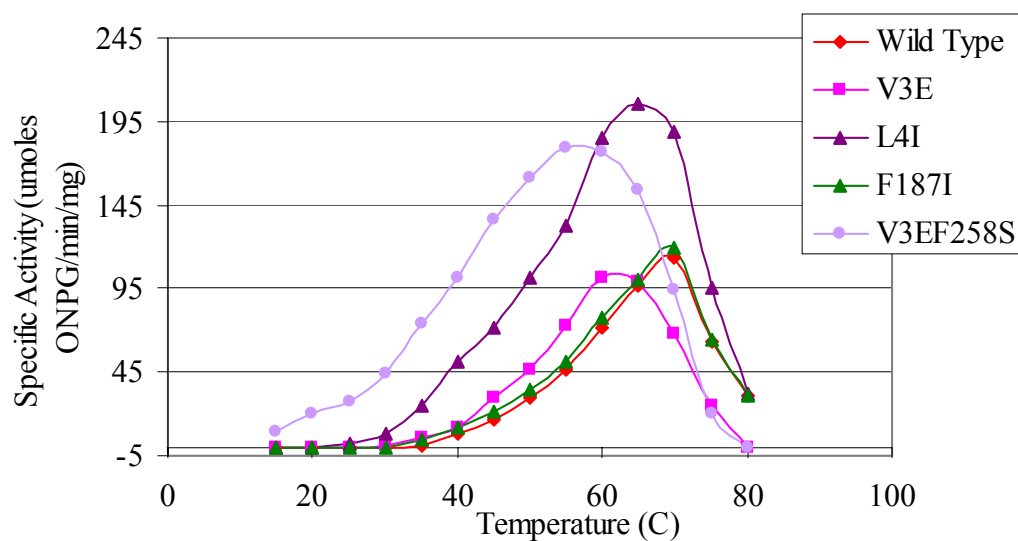
Figure 4.15. Thermal dependence of specific activity for single substitutions.
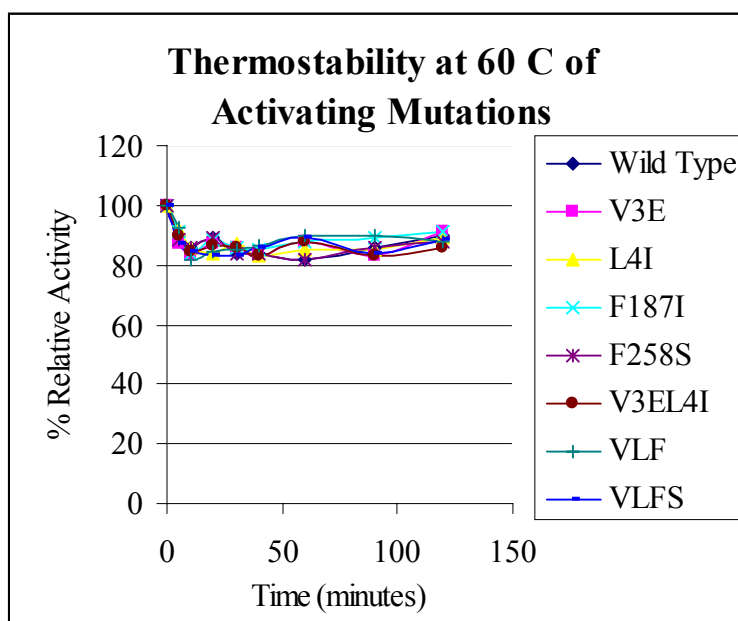
Figure 4.16. Thermostability at 60°C of all cold-activating substitutions.

Steady state kinetic analysis of activating mutations.

Kinetic characterization of the mutant enzymes were performed at 45, 55, and 65°C to determine the k_{cat} and K_m values for wild type, L4I, V3EL4I, and V3EL4IF187I enzymes. Significant increases in the k_{cat} value occurred for enzymes containing each of the mutations at each of the different temperatures while the K_m value remained relatively unchanged (Table 4.2).

Table 4.2. Steady state kinetics for wild type and mutants. All reactions were done in triplicate and NONLIN was used for nonlinear regression and error analysis.

Temp	Enzyme Source	k_{cat}	K_m	k_{cat}/K_m
65 °C	Wild Type	629	1.45	433
	V3E L4I F187I	2152	1.40	1537
	V3E L4I	1824	1.88	970
	L4I	1461	1.60	531
55 °C	Wild Type	351	1.00	356
	V3E L4I F187I	1453	1.20	1210
	V3E L4I	961	1.88	511
	L4I	430	1.27	339
45 °C	Wild Type	101	1.75	57
	V3E L4I F187I	521	1.10	473
	V3E L4I	199	1.41	141

Cysteine titration

Dithiobisnitrobenzoic acid (DTNB) was used to quantitate the number of sulfhydryl side chains that were accessible in wild type enzyme and the V3EL4I, and V3EL4IF187I variants under denaturing and non-denaturing conditions. Titration of available cysteine side chains is frequently used as a measure of flexibility in enzyme molecules [19]. The BgaB enzyme has 10 cysteines. Under denaturing conditions, using 14,120 as the extinction coefficient for the chromogenic TNB, an 11.2 µM concentration of free TNB released was determined for the wild-

type enzyme. Given a 1.48 µM concentration of wild-type enzyme, and thus a 14.8 µM total

concentration of thiols in the wild type reaction, this indicates that eight of the ten cysteines were

accessible to titration. Under non-denaturing conditions, only four cysteines or 40% could be

titrated in wild type enzyme after one hour. Similar calculations for the 1.35 µM solution of the

V3EL4IF187I variant indicate that seven cysteines (69.5%) could be titrated after one hour under

non-denaturing conditions (Table 4.3). This result indicates that structural changes induced by

these mutations give greater access to the DTNB molecule and perhaps indicate greater

flexibility in the variant molecule.

Table 4.3. Titration of accessible cysteines with DTNB.
Denaturing conditions contained 8 M urea while both conditions contained 1 mM EDTA.

	Absorbance at 412 nm			
	Denaturing conditions		Non-denaturing conditions	
Time (min)	Wild Type	V3EL4IF187	Wild Type	V3EL4IF187
1	0.087	0.114	0.061	0.080
5	0.122	0.140	0.063	0.094
25	0.157	0.144	0.064	0.100
55	0.158	0.144	0.064	0.100

Structural comparisons

The V3A, V3E, and L4I mutations are located in an extremely small section at the N-

terminal region of the enzyme. The high frequency of mutations arising in this small region

indicates that this area may play a strong role in controlling the thermal characteristics of this

enzyme. To assess if the N-terminal region and β sheet one play a critical role in thermal

characteristics of other $(\alpha/\beta)_8$ barrel enzymes, the N-terminus of the twenty-two mesophilic and

the five thermophilic structures analyzed in chapter two as well as five new thermophilic $(\alpha/\beta)_8$

barrel glycosyl hydrolase structures (Table 4.4) were examined. The N-terminus was defined as all residues up to and including the last amino acid in beta sheet one and all structures with N-terminal domains preceding the barrel were excluded. The residues in β sheet one were assigned based on the phi and psi angles of each residue within the structures using PROCHECK [20]. In both the thermophilic and mesophilic structures, this region was examined for the number of hydrogen bonds, the number of ion pairs, the amount of solvent accessible surface area, and the amount of surface area buried between the N-terminus (as defined here) and the rest of the molecule.

Hydrogen bonds

Hydrogen bonds were divided into four categories: hydrogen bonds between two charged atoms, hydrogen bonds between a charged atom and an uncharged atom, hydrogen bonds between two uncharged atoms and hydrogen bonds between two protein backbone atoms (Table 4.5). The average number of hydrogen bonds per residue was found to be similar in both mesophilic and thermophilic structures (Table 4.5).

Ion pairs

The average number of net ion pairs per residue for the mesophilic structures was 0.050 ± 0.018 while that for the thermophilic structures was 0.079 ± 0.029 (Table 4.6). These were similar to each other (each within one standard deviation) and also similar to the values found for the entire molecule, 0.047 and 0.062 net ion pairs per residue for the mesophilic and thermophilic structures, respectively.

Table 4.4 Structures Analyzed

	E.C. Number	Source Species	Environmental Temperature	PDB Code	Resolution
Mesophilic Enzyme					
α-Amylase	3.2.1.1	*Aspergillus oryzae*	36[a]	6taa	2.1
Acid α-amylase	3.2.1.1	*Aspergillus niger*	26[a]	2aaa	2.1
Pancreatic α-amylase	3.2.1.1	*Homo sapiens*	37[b]	1hny	1.8
Salivary amylase	3.2.1.1	*Homo sapiens*	37[b]	1smd	1.6
α-1,4 glycan-4-glucanohydrolase	3.2.1.1	*Hordeum vulgare*	20-37[h]	1amy	2.8
Pig α-amylase	3.2.1.1	*Sus scrofa*	37	1pif	2.3
Soybean β-amylase	3.2.1.2	*Glycine max*	20-37[h]	1bya	2.2
β-1,4-D-glycanase CEX-CD	3.2.1.8	*Cellulomonas fimi*	30[c]	1exp	1.8
Endo-1,4-β-xylanase 3.2.1.91	3.2.1.8	*Penicillium simplicissium*	24[a]	1bg4	1.75
Hevamine A	3.2.1.14	*Hevea brasiliensis*	20-37[h]	2hvm	1.8
Chitinase A	3.2.1.14	*Serratia marcescens*	40	1ctn	2.3
β-Glucosidase A	3.2.1.21	*Bacillus polymyxa*	40	1bga	2.4
Cyanogenic β-glucosidase	3.2.1.21	*Trifolium repens*	20-37[h]	1cbg	2.15
1,3-β-Glucanase	3.2.1.39	*Hordeum vulgare*	20-37[h]	1ghs	2.3
1,4-β-D-glucan maltotetrahydrolase	3.2.1.60	*Pseudomonas stutzeri*	35	2amg	2.2
1,3-1,4-β-Glucanase	3.2.1.73	*Hordeum vulgare*	20-37[h]	1ghr	2.2
1,3-1,4-β-Glucanase	3.2.1.73	*Hordeum vulgare*	20-37[h]	1aq0	2.0
6-Phospho-β-D-galactosidase	3.2.1.85	*Lactococcus lactis*	40[d]	1pbg	2.3
Endo-β-*n*-acetyl glucosaminidase F1	3.2.1.96	*Flavobacterium meningosepticum*	37[d]	2ebn	2.0
Endo-β-*n*-acetylglucosaminidase H	3.2.1.96	*Streptomyces plicatus* ATCC 10739	50-55[d]	1edt	1.9
Thioglucoside glucohydrolase	3.2.3.1	*Sinapis alba*	20-37[h]	2myr	1.6
Cyclodextrin glucanotransferase	2.4.1.19	*Bacillus* sp 1011	37[e]	1pam	1.8
Thermophilic enzyme					
Endo-1,4-β-D-glucanase	3.2.1.4	*Acidothermus cellulolyticus*	60[f]	1ece	2.4
Endo-1,4-β-glucanase CELC	3.2.1.4	*Clostridium thermocellum*	60	1cec	2.15
Endo-1,4-β-D-glucanase	3.2.1.4	*Thermomonospora fusca*	80	1tml	1.8
β –Glycosidase	3.2.1.23	*Sulfolobus solfataricus*	87	1gow	2.6
β –Glycosidase*	3.2.1.23	*Thermosphaera aggregans*	80	1qvb	2.4
α-Amylase*	3.2.1.1	*Bacillus licheniformis*	70	1bli	1.7
Cyclodextrin glycosyltransferase	2.4.1.19	*Thermoanaerobacterium thermosulfurigenes* EM1	70[g]	1ciu	2.3
β –mannanase	3.2.1.78	*Thermomonospora fusca*	80	1bqc	1.5
glycosyltrehalose trehydrolase	3.2.1.1	*Sulfalobus solfataricus*	87	1eh9	2.8
Malodextrin glycosyl transferase	2.4.1.25	*Thermatoga maritime*		1gjw	2.1
Cyclodextrin glucanotransferase	2.4.1.19	*Geobacillus stearothermophilus*		1cyg	2.5
α-amylase	3.2.1.1	*Geobacillus stearothermophilus*		1hvx	2.0

Table 4.5. Summary of Hydrogen Bonds in N-terminus and β1

File	Res in Nterm	Total	Hydrogen Bonds /Res	%C-C	%C-U	%U-U	%B-B
Mesophiles							
1aq0	5	7	1.4	0.00%	0.00%	0.00%	100.00%
6taa	14	21	1.50	4.76%	14.29%	14.29%	66.67%
2aaa	14	25	1.79	4.00%	16.00%	24.00%	56.00%
1hny	16	28	1.75	3.57%	21.43%	35.71%	39.29%
1smd	17	28	1.65	3.57%	21.43%	35.71%	39.29%
1amy	18	11	0.61	0.00%	9.09%	18.18%	72.73%
1pif	17	24	1.41	4.17%	20.83%	33.33%	41.67%
1bya	18	20	1.11	0.00%	15.00%	20.00%	65.00%
1exp	19	28	1.47	7.14%	17.86%	14.29%	60.71%
1bg4	22	36	1.64	2.78%	30.56%	19.44%	47.22%
2hvm	7	15	2.14	0.00%	6.67%	6.67%	86.67%
1bga	16	25	1.56	0.00%	8.00%	20.00%	72.00%
1cbg	29	27	0.93	7.41%	22.22%	18.52%	51.85%
1ghs	5	8	1.60	0.00%	0.00%	12.50%	87.50%
2amg	22	35	1.59	22.86%	22.86%	11.43%	42.86%
1ghr	5	7	1.40	0.00%	0.00%	0.00%	100.00%
1pbg	15	18	1.20	5.56%	0.00%	22.22%	72.22%
2ebn	18	26	1.44	3.85%	15.38%	15.38%	65.38%
1edt	17	24	1.41	4.17%	12.50%	16.67%	66.67%
2myr	35	35	1.00	0.00%	5.71%	40.00%	54.29%
1pam	20	38	1.90	10.53%	28.95%	23.68%	36.84%
Average	16.64	23.14	1.44	4.02	13.75	19.14	63.09
S. D.	7.56	9.4	0.347	5.26	9.46	10.73	19.19
Thermophiles							
1ece	28	38	1.35	5.26%	23.68%	21.05%	50.00%
1cec	10	16	1.6	0.00%	25.00%	25.00%	50.00%
1tml	42	44	1.05	4.55%	13.64%	22.73%	59.09%
1gow	14	19	1.35	0.00%	10.53%	21.05%	68.42%
1qvb	12	15	1.25	3.95%	15.51%	12.38%	69.38%
1bli	10	17	1.7	4.17%	12.50%	16.67%	66.67%
1ciu	20	33	1.65	9.09%	30.30%	24.24%	36.36%
1bqc	26	34	1.31	3.85%	15.38%	15.38%	65.38%
1eh9	17	15	0.88	4.17%	12.50%	16.67%	66.67%
1cyg	17	24	1.41	4.17%	12.50%	16.67%	66.67%
1hvx	11	18	1.64	0.00%	5.71%	40.00%	54.29%
Average	24	24.8	1.38	3.76	20.63	22.81	52.77
S. D.	24	10.51	0.26	3.86	8.26	1.81	11.93

Res, number of residues in structure; Total, total number of hydrogen bonds calculated for structure; #H/res, number of hydrogen bonds per residue found in the entire polypeptide chain; %C-C, percent of hydrogen bonds between two charged atoms, %C-U, percent of hydrogen bonds between charged and uncharged atoms; %U-U, percent of hydrogen bonds between two uncharged atoms; %B-B, percent of hydrogen bonds between backbone atoms, #B-B is the actual number of these bonds.

Table 4.6. Distribution of structural features in the N-terminus

File	Net ion pairs/residue	Net Ion Pairs	%Non-polar SASA	Buried Surface Area
Mesophiles				
6taa	0.07	1	62.02	2383
2aaa	0.07	1	60.25	2366
1hny	0	-1	55.32	2087
1smd	0	0	44.97	2761
1amy	0	0	55.6	1425
1pif	0	0	55.41	2757
1bya	0	0	63.68	2684
1exp	0	0	41.21	2294
1bg4	0.136	3	61.4	2414
2hvm	0.142	1	44.76	1616
1bga	0	0	58.23	2390
1cbg	0	0	66.89	2600
1ghs	0	0	55.57	1200
2amg	0.182	4	50.01	2616
1ghr	0	0	52.98	1185
1aq0	0	0	27.39	1183
1pbg	0.066	1	32.12	2053
2ebn	0	0	46.15	2525
1edt	0.059	1	54.36	2105
2myr	0.028	1	37	3105
1pam	0.05	1	69.49	3241
Average	0.050	0.035	53.56	2100
S. D.			10.95	530
Thermophiles				
1ece	0	0	51.9	3448
1cec	0.100	1	56.15	1765
1tml[a]	0.666	2	55.53	2908
1gow	0.047	2	53.12	2389
1qvb	0.286	4	54.23	2808
1bli	0.100	2	57.81	2811
1ciu	0.050	1	54.4	2908
1bqc	0.038	1	57.18	2825
1eh9	0	0	51.32	2490
1cyg	0.118	2	52.27	2944
1hvx	0	0	60.23	2790
Average	0.076	1.454	56.53	2729
S. D.			5.5	399

[a]Sheet 8 not present.

Ion pairs

The average number of net ion pairs per residue for the mesophilic structures was $0.050 \pm$ 0.018 while that for the thermophilic structures was 0.079 ± 0.029 (Table 4.6). These were similar to each other (each within one standard deviation) and also similar to the values found for the entire molecule, 0.047 and 0.062 net ion pairs per residue for the mesophilic and thermophilic structures, respectively.

Solvent accessible surface area

Surface area was divided into two categories: polar and non polar. The average solvent accessible surface area in the N-terminus and sheet one was found to be 46.4 ± 11 percent polar in the mesophilic structures and 43.5 ± 5.5 percent polar in the thermophilic structures (Table 4.6). These values do not differ from the 44 percent and 46 percent found through the entire mesophilic and thermophilic structures, respectively. The distribution for solvent accessible surface area was found to be small for both the mesophiles and thermophiles. In all three categories the amounts of surface area for the thermophiles came within one standard deviation of the mean for the mesophilic structures implying there are no significant differences in this area that can account for thermostability.

Surface area buried by N-terminus and β sheet one

The amount of surface area buried by the N-terminus and β sheet one residues was defined as the sum of the solvent accessible surface areas calculated for two fragments minus the solvent accessible surface area calculated for the entire molecule where fragment one is the N-

terminus plus all β sheet one residues treated as an independent object, and fragment two is the rest of the molecule (excluding those residues) treated as an independent object. This surface area buried due to docking was calculated for all 35 structures (Table 4.6). The average surface area buried due to docking was 2100 square angstroms for the mesophilic structures and 2,770 square angstroms for the thermophilic structures. These values are separated by more than the standard deviation for the mesophilic (±590) and the thermophilic (±345) groups.

Saturation mutagenesis of N-terminal residues

In order to target beta sheet one and the N-terminal region with a higher mutation rate, primers degenerate in the first base pair of codons three through six were used to amplify the wild type *bga*B gene. To determine the frequency of mutations leading to increased activity at lower temperatures, plates were incubated at 18°C for 48 hrs, then successively shifted to 24°C, 37°C, and 45°C for 6 hrs each to monitor hydrolysis of X-gal. Of the 4,000 clones that were screened, 16% demonstrated activity at 18°C, 26% demonstrated activity only when the temperature was raised to 24°C, a further 31% only demonstrated activity until the temperature was raised to 37°C, and 10% demonstrated activity only when the temperature was brought to 45°C (Table 4.7).

Table 4.7. Percent of transformants capable of hydrolyzing X-gal after saturation mutagenesis at N-terminus

Temperature	% of transformants hydrolyzing X-gal
18 °C	16
24 °C	26
37 °C	31
45 °C	10

4.5 DISCUSSION

In three rounds of directed evolution, the low temperature limit of the BgaB β-galactosidase was decreased by 25°C. Where wild type enzyme demonstrates 5% activity at 40°C, the directly evolved variant, 2aGen3-e, has 15% activity at 15°C. This represents a significant addition of function at low temperatures. The small number of amino acid substitutions leading to this new phenotype makes it possible to examine the role of each mutation and postulate structural mechanisms for achieving activity at low temperatures.

For each of the variants evolved in generation two, Gen2-a and Gen2-j, the individual mutations were examined individually and in combination for their ability to hydrolyze X-gal at different temperatures and by the thermostability and thermal dependence of activity of the purified enzyme. For Gen2-a, the complete thermal characteristics of the variant can only be fully reproduced in the presence of both mutations, V3E and F258S. Each mutation alone however, does confer additional activity at lower temperatures and in neither case is thermostability sacrificed (Figure 4.17).
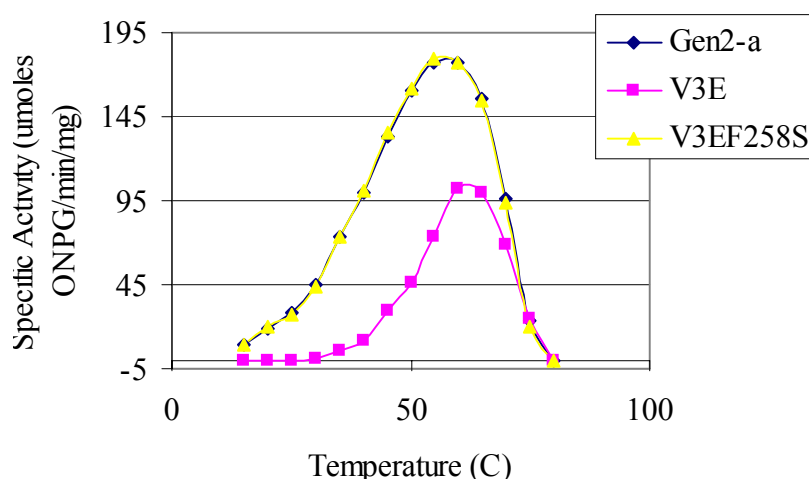
Figure 4.17. Thermal dependence of specific activity for Gen2-a and its component mutations

In the case of the Gen2-j (V3E L4I Y149N E393V) directly evolved variant, activity at low temperatures is increased but the activity decreased rapidly at temperatures over 55°C and the thermostability at 60°C decreased significantly. The rapid drop off of activity as temperatures increase above 55°C can be easily explained by the thermostability properties of this enzyme. In a five minute assay, when temperatures are above 55°C the enzyme will be undergoing rapid denaturation. Half way through the assay, over half of the enzyme will have been denatured and specific activity values will be lower. This loss of thermostability was a result of a mutation that did not have any effect on activity at lower temperatures. Through analysis of mutations individually, only the V3E and L4I mutations were shown to provide increased activity at low temperatures. Each of these mutations increased activity at low temperatures when incorporated individually, and in combination, reproduced the entire activity profile of the Gen2-j variant at low temperatures (Figure 4.18).
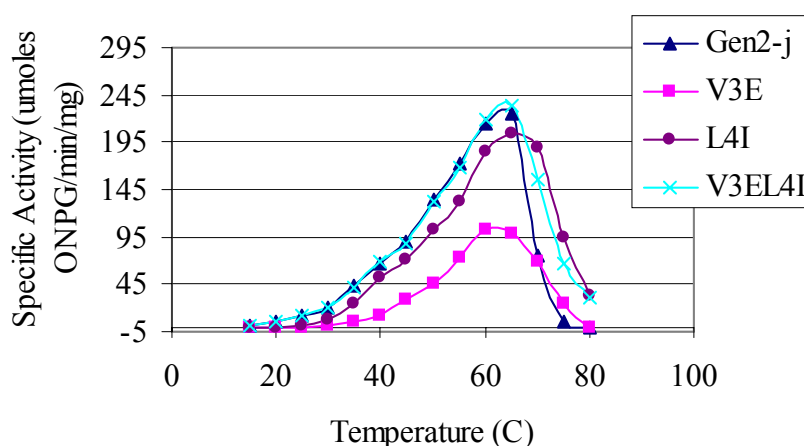
Figure 4.18. Thermal dependence of specific activity for Gen2-j and its component mutations

Interestingly, in the V3EL4I double mutant, there was no rapid drop off in activity at temperatures higher than 55°C, and thermostability at 60 °C was not sacrificed. This result indicates, that the loss of thermostability at 60°C and its resultant effects on activity at high temperatures were a result of one of the other mutations in Gen2-j. These results also indicate that activity at low temperatures can be increased without sacrificing stability at high temperatures as had been assumed by many researchers [4].

The third generation variant, 2aGen3-e (V3E K47E N75S Y149N F187I A204T F258S V309I), also demonstrated increased activity at low temperatures while both activity and stability at high temperatures in sacrificed. By incorporating mutations individually, V3E, F187I, and F258S were shown to increase activity at low temperatures. In combination, the effects of these mutations were additive and could completely reproduce the low temperature characteristics of the 2aGen3-e variant (Figure 4.19).
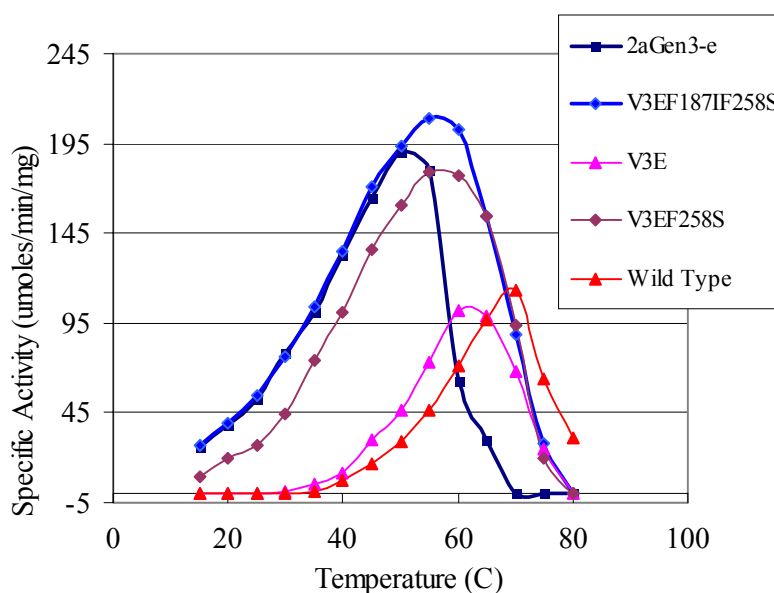
Figure 4.19. Thermal dependence of specific activity of 2aGen3-e variant and component mutations.

The triple mutant, V3EF187IF258S, did not demonstrate loss of stability at high temperatures nor a rapid drop off in activity at high temperatures. Again, this result shows that the structural changes required to increase activity at low temperatures do not necessitate decreases in stability. This finding is in agreement with the work on lactate dehydrogenase A4 and the subtilisin-like protease from *Bacillus sphaericus* in which no correlation could be found between adaptation temperature and stability at 50°C. [2].

Researchers have clearly shown that increases in the flexibility of an enzyme can increase the rate of product release [21]. To determine if these mutations resulted in an increase in flexibility of the enzyme, DTNB was used to titrate the accessible free sulfhydryl groups of cysteines within both the wild type and mutant molecules. Only four of the ten cysteines were accessible for titration with the wild type enzyme under non-denaturing conditions whereas seven were available for the V3E L4I and V3E L4I F187I mutants, suggesting that flexibility is indeed increased in the mutant molecules. Taken together, these results suggest that activity at low temperatures of the mutant enzymes is increased by amino acid substitutions which increase the flexibility of the enzyme, allowing for faster release of product and thus increased catalysis at low temperatures.

To determine the possible structural mechanism for the increase in flexibility provided by these mutations, the amino acid sequence of the BgaB enzyme from *G. stearothermophilus* was aligned with the amino acid sequences of the β-galactosidases from *Sulfolobus solfataricus* (SS-βgal) and *Thermus thermophila*,(A4-βgal) and the secondary structure of the SS-βgal enzyme, as found in its crystal structures [22].

was then matched against that sequence alignment (Appendix C, Figure 4.20). The SS-βgal bgaB

enzyme shares only a 20 % sequence identity with bgaB but it is the most closely related β-gal

for which a solved structure is published.

Figure 4.20 Sequence alignment of SS-βgal with bgaB at N-terminal region

Sequence alignment done in ClustalW. Secondary structure taken from

the structure of SS-βgal as found in 1GOW PDB file. A more extensive

alignment can be found in Appendix C.

The phenylalanine at position 187 in BgaB corresponds to phenylalanine 181 of the A4-

βgal aligned sequence. These residues may be located at the top of subdomain H and inserted

into an adjacent subunit. No functional role for subdomain H has yet been published.
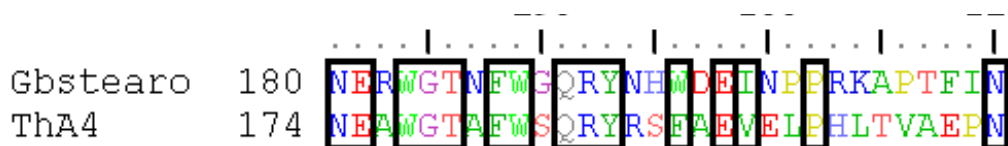
Figure 4.21. Alignment of *bga*B with A4 βgal.
Residues that are identical and/or similar are outlined.

Two residues downstream in the structure is a glutamine, which has been proposed to

make contacts with another subunit [23]. It is not possible to tell from the current study what role

the F187I mutation plays. The lack of significant changes in the K_m value for the ONPG

substrate (which contains a galactose moiety) suggests that substrate recognition is unaffected,

but this supposition is only true under the conditions that k1<<k-1 where the K_m value would be

a measure of substrate affinity. Alternately, F187I could play a role in subunit interaction by

changing the equilibrium constant between multimeric forms, but this is as yet untested.

The Phe258 in the BgaB sequence aligns with Phe252 in SS-βgal and thus may be

located in the middle of helix five of the $(\alpha/\beta)_8$ barrel. The role of the F258S mutation is unclear.

Because serine is capable of hydrogen bonding, it can be postulated that this mutation enables or

interferes with some hydrogen bonding network, but this is mere supposition. In the crystal

structure there are no nearby residues that appear to play critical roles in substrate recognition or

the multimer interface.

The mutations V3E, L4I, and V3A have all been observed in the variants screened for X-

gal hydrolysis at low temperature. While the V3A mutation, present in the first generation

mutant Gen1-a, was not characterized, the close proximity of all of these mutations at the N-

terminus of the enzyme cannot be ignored. By virtue of being located in the third and fourth

amino acid of the protein, there can be no other conclusion than that these residues are in sheet

β1 or in the N-terminus preceding that sheet. Sequence alignment with SS-bgal supports this conclusion (Figure 4.20).

Sheet one is unique in the $(\alpha/\beta)_8$ barrel architecture because, unlike the other seven β sheets that signify the fold, sheet one is not preceded by an alpha helix and turn. With the other sheets in the barrel, one can imagine that their mobility is stabilized by the anchoring provided by the preceding helix-turn as the preceding helix will be held in place by structural interactions with the rest of the molecule. Sheet one stability and mobility, however, can only be mediated by interactions with the adjacent sheets—two and eight—which comprise this domain. Any increase or decrease in the number or strength of interactions between sheet one and the surrounding environment would make that sheet more or less mobile and/or stable.

The contacts between sheet one and sheet eight of the barrel also have further significance. Whereas the barrel shape of the rest of the molecule is maintained by a set of contiguous peptide bonds, the portion of the ring shape between the first helix and the last helix of the barrel are not covalently connected and thus the closed barrel structure can only be maintained by contacts of sheet one with sheet eight and the contacts of alpha helix one with α helix eight. The significance of this aspect of the barrel architecture can be seen in denaturation studies of the $(\alpha/\beta)_8$ enzymes enolase and phosphoglycerate kinase [24]. In this study, researchers labeled the enzyme with deuterium in a time dependent fashion as the concentration of denaturants was increased. The labeling reaction was quenched, and the regions made accessible to label were identified through mass spectrometric analysis. In both enzymes, the initial step in the denaturation pathway was an opening of the barrel caused by separation of helix one from helix eight and strand one from strand eight (Figure 4.22).
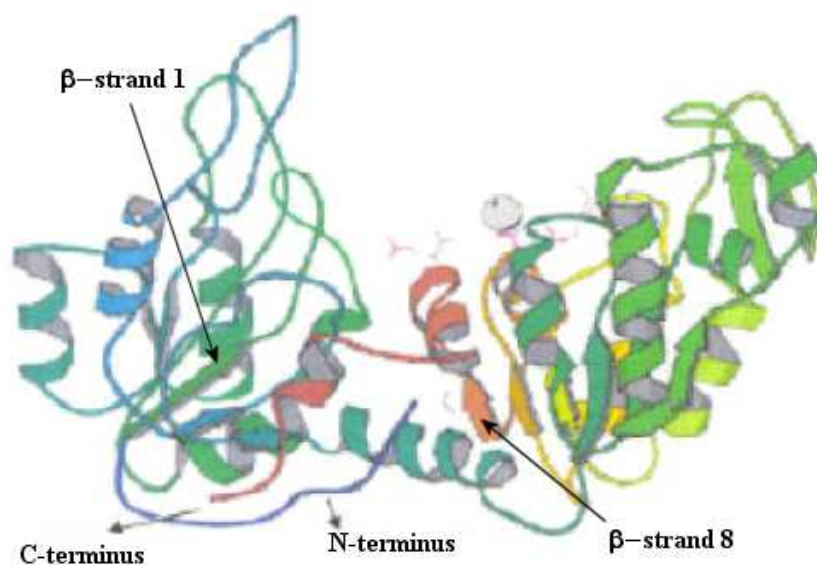
Figure 4.22. Phosphoglycerate kinase during initial phases of denaturation [24].

Several other (α/β)$_8$ barrels appear to follow this unfolding pathway as well [25]. The relative susceptibility of this region to denaturants emphasizes that this region may be a hot spot governing the stability and flexibility of the (α/β)$_8$ barrel.

Alignment of the *G. stearothermophilus* BgaB sequence with SS-bgal suggests that V3E and L4I exist in the region preceding sheet one (Figure 4.20). While it cannot be confirmed whether or not this region in bgaB precedes sheet one or is a part of it, it is instructive to note that in the structure of SS-bgal, this N-terminal sequence makes hydrogen bonds with helix eight of the barrel (Figure 4.23).
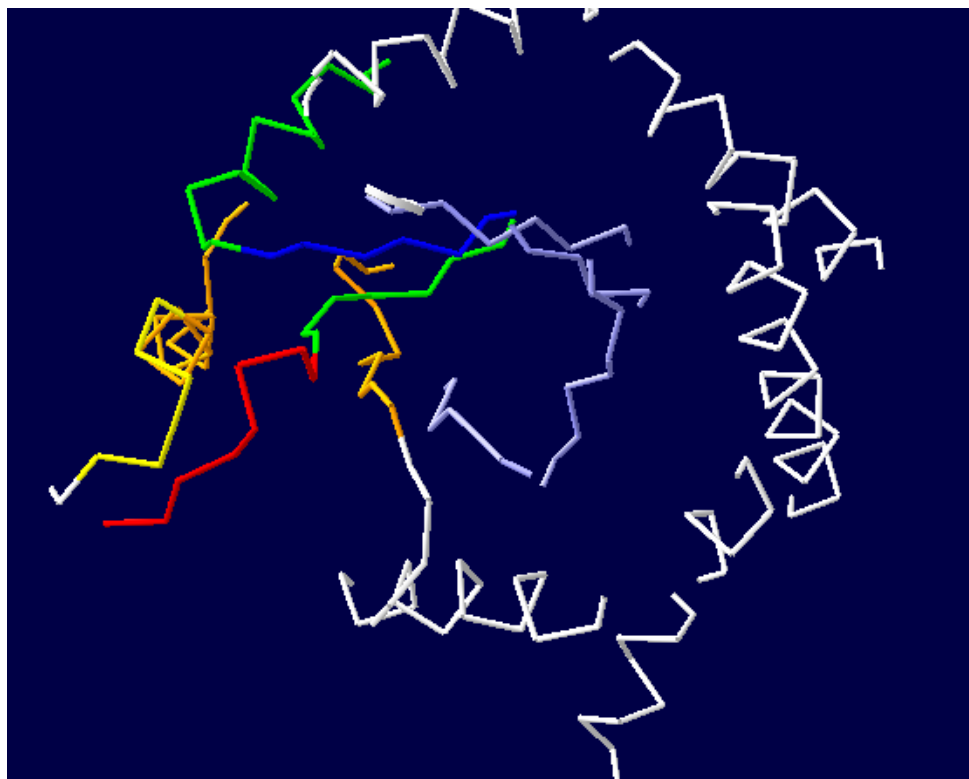
Figure 4.23. Structure of SS-βgal. Beta sheet one and helix one of the barrel are in green. Sheet eight and helix eight are in orange. The N-terminus, which is 'tethered' to the end of helix eight, is in red.

These N-terminal amino acids appear to act to tether sheet one to helix eight thereby keeping the ring of the barrel closed. As SS-βgal is a thermophilic enzyme, this tethering may prevent opening of the ring structure and provide a greater degree of stability for the molecule at high temperatures. Which interactions the V3E and L4I mutations alter cannot be precisely known but overall molecular flexibility does appear to be changed.

For the above structural reasons and in accordance with experimental data, I postulate that mutations at the N-terminus and in β sheet one of $(\alpha/\beta)_8$ barrels can alter the flexibility within the molecule. This theory predicts that this region would be a hot spot for mutations influencing stability and flexibility. Of course, any additional domain or sub-domain that spans

the region between helix/sheet one and helix/sheet eight of the barrel structure could provide

added stabilization and thus reduce the structural effects of this hotspot. Accordingly, this

prediction would only hold true for $(\alpha/\beta)_8$ barrels with no significant N-terminal domain and for

$(\alpha/\beta)_8$ barrels that do not contain a second domain that tethers these two boundaries together.

This prediction appears to be valid with the BgaB enzyme. When the N-terminus of the

BgaB gene was targeted for saturation mutagenesis by the use of an N-terminal primer that was

degenerate in the first base pair of each of the third through sixth codons, there was a high

number (over 50%) of transformants with altered temperature requirements for X-gal hydrolysis.

Both transformants that were able to hydrolyze X-gal at lower temperatures than wild type, and

transformants that could not hydrolyze X-gal until placed at higher than normal temperatures

($45^{o}C$) were observed. These results also indicate that mutations in this area can affect the

temperature requirements of the enzyme in either direction.

Other thermophilic $(\alpha/\beta)_8$ barrel glycosyl hydrolases appear to employ a similar

reinforcement of interactions in this region. When thermophilic structures were compared to

mesophilic structures, the thermophiles were found to have more surface area buried at the N-

terminus and sheet one region by a factor of one standard deviation, indicating that a clear trend

could be observed. An increase in the buried surface area indicates there is greater stabilization

in this region. This finding suggests that many thermophilic $(\alpha/\beta)_8$ barrel glycosyl hydrolases

increase their thermostability by stabilizing the interactions between the helix one/sheet one

interface with helix eight/sheet eight. While the number of hydrogen bonds and ion pairs per

residue and the amount of solvent accessible surface area were not statistically different between

the two groups, this result was not surprising as studies of secondary structural motifs indicate

that β sheet interactions that are internal to the molecular surface usually involve substantial hydrophobic interactions. Because the role of these mutations is unique to $(\alpha/\beta)_8$ barrel architecture, this is evidence that the molecular mechanisms altering the thermal dependence of activity can be fold specific.

Taken as a whole, these findings suggest that I have discovered a new structural mechanism for the control of thermostability and the thermal dependence of activity. Further evidence supporting this mechanism will be presented in the following chapters.

4.6 REFERENCES

1. Adams, M.W. and R.M. Kelly, *Thermostability and thermoactivity of enzymes from hyperthermophilic Archaea.* Bioorg Med Chem, 1994. **2**(7): p. 659-67.

2. Wintrode, P.L., K. Miyazaki, and F.H. Arnold, *Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution.* The Journal of Biological Chemistry, 2000. **275**(41): p. 31635-31640.

3. Giver, L., et al., *Directed evolution of a thermostable esterase.* Proc. Natl. Acad. Sci. USA, 1998. **95**: p. 12809-12813.

4. Miyazaki, K., et al., *Directed evolution study of temperature adaptation in a psychrophilic enzyme.* Journal of Molecular Biology, 2000. **297**: p. 1015-1026.

5. Lebbink, J.H., et al., *Improving low-temperature catalysis in the hyperthermostable Pyrococcus furiosus beta-glucosidase CelB by directed evolution.* Biochemistry, 2000. **39**(13): p. 3656-65.

6. Flores, H. and A.D. Ellington, *Increasing the thermal stability of an oligomeric protein, beta-glucuronidase.* J Mol Biol, 2002. **315**(3): p. 325-37.

7. Hirose, R., et al., *Crystal structures of mutants of Thermus thermophilus IPMDH adapted to low temperatures.* Protein Eng, 2001. **14**(2): p. 81-4.

8. Roovers, M., et al., *Experimental evolution of enzyme temperature activity profile: selection in vivo and characterization of low-temperature-adapted mutants of Pyrococcus furiosus ornithine carbamoyltransferase.* J Bacteriol, 2001. **183**(3): p. 1101-5.

9. Hirata, H., et al., *Structure of a β-galactosidase gene of Bacillus stearothermophilus.* Journal of Bacteriology, 1986. **166**(3): p. 722-727.

10. Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases.* Structure, 1995. **3**: p. 853-859.

11. Trimbur, D.E., et al., *Characterization of a psychrotrophic Arthrobacter gene and its cold-active β-galactosidase.* Appl. Environ. Microbiol., 1994. **60**(12): p. 4544-4552.

12. Mogk, A., R. Hayward, and W. Schumann, *Integrative vectors for constructing single-copy transcriptional fusions between Bacillus subtilis promoters and various reporter genes encoding heat-stable enzymes.* Gene, 1996. **182**: p. 33-36.

13. Leung, D.W., E. Chen, and D.V. Goeddel, *A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction.* Technique, 1989. **1**(1): p. 11-15.

14. Vriend, G., *A molecular modeling and drug design program.* Journal of Molecular Graph., 1990. **8**: p. 52-56.

15. Barlow, D.J. and J.M. Thornton, *Ion pairs in proteins.* Journal of Molecular Biology, 1983. **168**: p. 857-885.

16. Yip, K.S., et al., *The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures.* Structure, 1995. **3**: p. 1147-1158.

17. Eisenhaber, F. and P. Argos, *Improved strategy in analytic surface calculation for molecular systems: handling of singularities and computational efficiency.* Journal of Computational Chemistry, 1993. **14**: p. 1272-1280.

18. Eisenhaber, F., et al., *The double cube lattice methode:efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular systems: handling of singularities and computational efficiency.* Journal of Computational Chemistry, 1995. **16**: p. 273-284.

19. Huston, E.E., J.C. Grammer, and R.G. Yount, *Flexibility of the myosin heavy chain: direct evidence that the region containing SH1 and SH2 can move 10 A under the influence of nucleotide binding.* Biochemistry, 1988. **27**(25): p. 8945-52.

20. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structure.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

21. Caccuri, A.M., et al., *Human glutathione transferase T2-2 discloses some evolutionary strategies for optimization of the catalytic activity of glutathione transferases.* J Biol Chem, 2001. **276**(8): p. 5432-7.

22. Aguilar, C.F., et al., *Crystal structure of the beta-glycosidase from the hyperthermophilic archeon Sulfolobus solfataricus: resilience as a key factor in thermostability.* J Mol Biol, 1997. **271**(5): p. 789-802.

23. Hidaka, M., et. al. *Crystal structure of Thermus thermophilus A4 beta-galactosidase.* J. Mol Biol. 2002 (in press)

24. Raza A., Deng Y., and Smith D.L., *Identifying Unfolding Domains in Alpha-Beta Barrel Proteins Using Hydrogen Exchange and Mass Spectrometry.* ASMS. 2000. **45**(1): p. 1-21.

25. Zitzewitz, J.A. and C.R. Matthews, *Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein.* Biochemistry, 1999. **38**(31): p. 10205-14.

Chapter 5

EXTENDING THE PARADIGM

5.1 ABSTRACT

In this chapter, I apply the methodology of using directed evolution as a probe of protein structure as described in Chapter 4. By targeting the N-terminus and β1 residues for saturation mutagenesis and selection, I attempt to test the structural mechanisms affecting the thermal dependence of activity that were deduced in the previous chapter.

Here I present a unique scheme of selection and enrichment to select for variants that exhibit β-galactosidase activity at decreased temperatures. Cells containing mutations in the third through sixth amino acids of the N-terminus of a related $(\alpha/\beta)_8$ barrel glycosyl hydrolase, SOS β-galactosidase, were placed into enrichments at low temperatures to select for faster growth in minimal media with lactose as the sole carbon source. This scheme is first examined to show that a sufficient selective pressure exists under this format to expect the desired phenotypic changes; then the sensitivity to small increases in low temperature activity is quantified. Finally, directly evolved variants of the SOS β-galactosidase are characterized by their growth rates to show that targeting mutagenesis to the N-terminus region can produce enzymes with greater activity at lower temperatures in another enzyme of the $(\alpha/\beta)_8$ barrel.

5.2 INTRODUCTION

The possibility of identifying a common structural region that could control the internal flexibility and thus the thermal dependence of activity for $(\alpha/\beta)_8$ barrel enzymes is intriguing. The $(\alpha/\beta)_8$ barrel family is the largest family of protein folds, representing nearly 10 percent of all known soluble proteins [1]. Identification of a small region crucial for temperature adaptation would be of significant interest to industrial protein engineers eager to adapt their enzymes to function at the specific temperatures required for industrial processes. Moreover, this discovery would represent a significant advance in our understanding of the $(\alpha/\beta)_8$ barrel fold and its molecular mechanisms for determination of internal flexibility.

While much has been conjectured on the evolution of $(\alpha/\beta)_8$ barrels, much less is known about structural paradigms that govern internal flexibility, thermostability, or other physical properties. In one example, researchers suggested that $(\alpha/\beta)_8$ barrels have a preference for negatively charged substrates which is governed by the backbone contribution to a distinct electrostatic field that can act to steer negatively charged molecules to the active site [2]. Studies of the thermostability of certain $(\alpha/\beta)_8$ barrel proteins have pointed to increased hydrogen bonding networks or increases in the number of ion pairs as a molecular basis for their higher stability [3], but these findings have relied on comparisons of distantly related homologs and do not correlate with my studies, as detailed in Chapter 2, nor do they represent mechanisms that would be specific to the unique architecture of the $(\alpha/\beta)_8$ barrel. No general paradigms yet exist to explain mechanisms for controlling the mobility of internal structural elements, such as β strands, in $(\alpha/\beta)_8$ barrels. Nor has there been any identification of a "thermostat" region that

controls the internal enzyme flexibility that might be necessary to achieve changes in thermal dependence of activity.

The paradigm postulated in Chapter 4 suggests that flexibility in $(\alpha/\beta)_8$ barrels as a family can be altered by manipulating the interactions that hold $\beta 1$ in place. It is not known to what degree, if any, the sheets internal to the barrel are allowed to slide or move. It is known, however, that the substrate recognition sites in $(\alpha/\beta)_8$ barrels are usually located at the C-terminal end of these sheets. Furthermore, the active site is consistently found in the loops at the end of these sheets. Movement of a sheet containing or interacting with these regions could certainly affect catalysis, substrate recognition, or product release. The changes at the N-terminus in the BgaB V3E L4I mutants which appear to confer high flexibility and an increase in the rate of product release seem to support this idea.

The stability of the closed ring structure, and thus domain stability, may also be strongly influenced by changes at the N-terminus. The interface of $\beta 1/\alpha 1$ with $\beta 8/\alpha 8$ is the surface where the two ends of the ring meet. Interactions that stabilize this interface might inhibit initial steps of unfolding and enhance thermostability, or they may provide a measure of control over the amount of contortion or twist that the barrel can undergo during normal "breathing."

In order to provide the first test of this paradigm, I have attempted to extend it beyond the BgaB enzyme to the closely related SOS β-galactosidase by targeting the N-terminus and β1 residues for saturation mutagenesis and selection. The SOS Orange clone 5 encoding a family 42 β-galactosidase was obtained from a psychrophilic *Plannococcus* species collected in the dry valleys of Antarctica [4]. The enzyme was selected for use in this study on the basis of its high sequence identity (45%) with the BgaB enzyme from *G. stearothermophilus*. Unlike its

thermophilic counterpart, SOS has a temperature optimum of only 42°C and maintains 30% of its activity at 18°C. While the natural substrate for SOS is not known, it has a specific activity of 467 μmoles of ONPG/min/mg at 42°C and is highly salt tolerant, capable of maintaining activity in salt concentrations as high as 5 molar. The SOS gene has been cloned and its nucleic acid and amino acid sequences deduced [4] (Appendix B).

Because the SOS β-galactosidase maintains 30% of its activity at 18°C, it was necessary to develop a selection scheme different from the one used with BgaB as described in Chapter 4. *E. coli* is unable to grow at temperatures much below 15°C, and at 15°C the specific activity on ONPG of the SOS β-galactosidase would be 116 μmoles/min/mg—33 times higher than the minimum 3.5 μmoles/min/mg of specific activity necessary to observe X-gal hydrolysis on plates as revealed in Chapter 4. In other words, colonies growing at the lowest selection temperature possible with *E. coli* and containing wild type SOS enzyme would hydrolyze X-gal in such large quantities as to making selection on the basis of X-gal hydrolysis impossible.

To overcome this limitation, a host strain was constructed that was devoid of background β-galactosidase activity but maintained constitutive production of lactose permease to afford transport of lactose across the cellular membrane. Mutagenized SOS gene was then ligated into an expression vector and transformed into host cells. Cells with functional β-galactosidases were selected for in minimal liquid media containing lactose as its sole carbon source. It was posited that those cells which contained a β-galactosidase variant that catalyzed the breakdown of lactose more efficiently for its carbon and energy source at the selection temperatures would grow faster and thus out-compete cells with a less efficient enzyme. Consequently, these cultures were maintained in enrichments. By maintaining exponential growth at low enrichment temperatures,

cell containing β-galactosidases with even small increases in low temperature catalytic efficiencies became the dominant members of the population and thus could be efficiently selected.

5.3 EXPERIMENTAL PROCEDURES

Construction of an SOS screening vector

An *Nde*1 restriction site was incorporated at the start codon of the *SOS* gene in the pUC18/SOS plasmid [5] using the Quickchange Site Directed Mutagenesis kit (Stratagene) and self complimentary primers (forward sequence 5'-AGAAGGGATGGATGGAATCATATGATTAACGATAAATTGCCGAGATTTGG-3'). The altered plasmids were transformed into *E. coli* DH5α cells and cells were incubated overnight at 37°C on LB plates with 100 μg/mL ampicillin (Fisher Biotech), and 100 μg/mL 8-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (United States Biological). An isolated colony was selected with a sterile toothpick and inoculated into 3 mL LB media with 100 μg/mL X-gal. The culture was incubated for 8 hours at 37°C. Plasmid was then isolated using the Wizard Miniprep System (Promega). This plasmid was designated p*SOSNde*. The *SOS* gene was excised by simultaneous restriction digests of p*SOSNde* with *Nde*1 (Promega) and *Xba*1 (Promega) in Buffer D (Promega). The 2 kb gene fragment was then purified from a 0.7% gels using the QIAquick Gel Extraction Kit (Qiagen), eluted in water, and then ligated (Epicentre Fast Link ligase, Epicentre Technologies) into a phosphatase treated pΔαH plasmid vector (Shrimp alkaline phosphatase, Amersham Life Sciences, Arlington Heights, IL). Recombinant plasmids were transformed into *E. coli* DH5α cells and cells were incubated on LB plates with 100 μg/ml

ampicillin, and 100 µg/mL X-gal. Plasmids were isolated using the Wizard Miniprep System (Promega). This plasmid construct was designated pΔαH/*SOS*.

Construction of host strain for selections and enrichments

In order to create a host strain that has no endogenous β-galactosidase production but is capable of lactose transport, *E. coli* ER2585 cells (genotype *fhu*A2 *gln*V44 e14- *rfb*D1 *rel*A1 *end*A1 *spo*T1 *thi*-1 Δ(*mcr*C-*mrr*)114:IS10 Δ(*lac*I-*lac*A)200) were conjugated with *Salmonella typhimurium* CH3 F' cells (genotype *gal*E496 *gal*K *met*A22, *met*E55, rpsL102 *xyl*404 H1B *nml* H2-*enx ilz hsdl*6 HSDS-A29 F' 1020 Δz(h138), Tn:*pro*AB81 *lac*I- *lac*Y+). Isolated colonies of both strains were inoculated into 3 mL LB broth and incubated until an O.D.(600) of 0.6. Cultures were combined and incubated with minimal shaking for 1 hr at 37**°**C. 100 µL of a 10^{-7} dilution was spread plated onto minimal M9 plates with 10 µg/mL thiamine, 0.2% glucose, and 100 µg/mL tetracycline and incubated overnight at 37**°**C. The resulting tetracycline resistant colonies that were able to grow with only a thiamine requirement were designated *E. coli* ER2585 F'. In order to verify that the host strain had the desired phenotype, cells that had been transformed with the β-galactosidase containing pΔαH/*SOS* plasmid were grown on minimal M9 plates containing 10 µg/mL thiamine and 0.2% lactose. As a control, untransformed cells were plated on the same media.

Saturation mutagenesis of N-terminal region of *SOS*

The *SOS* gene was amplified by PCR using the DE2-R primer and a degenerate primer (5'-AGAAGGGATGGATGGAATCATATGATTNACNATNAANTGCCGAG

ATTTGG-3') complimentary to the N-terminal region. A 50 µL reaction contained 1 µM of each

primer, 10 ng of linearized pΔαH/*SOS* plasmid DNA and two Ready to Go$_{TM}$ PCR beads

(Amersham Pharmacia Biotech). PCR was carried out in an Ependorff Master Gradient Thermal

Cycler (Eppendorf Scientific) at 95°C for 5 minutes, with 35 cycles of 95°C for 30 s, 58°C for 30

s, and 72°C for 2.75 minutes, followed by 7 minutes at 72°C.

Library production

Products from the mutagenic PCR reaction were purified using a Qiagen PCR

purification kit, and the insert fragment cleaved simultaneously with the restriction enzymes

*Xba*1 (New England Biolabs) and *Nde*1 (New England Biolabs) in 50 mM Tris-HCl (pH 7.5), 10

mM MgCl$_2$, and 0.1 M NaCl at 37°C for 4 hrs. The resulting DNA was then subjected to

electrophoresis on a 0.7% agarose gel, and the 2 kb fragment was excised from the gel and

extracted using a QIAquick kit (Qiagen). Purified DNA was eluted in water. Ligation reactions

were performed using Fastlink ligase (Epicenter Technologies). Vector DNA (the entire pΔαH

excluding the region between *Nde*1 and *Xba*1), insert DNA, 10 x ligation buffer, 1 mM ATP,

water and 1 unit of enzyme were combined and incubated at 10°C overnight. Resulting plasmids

were transformed into *E. coli* ER2585 F' competent cells that were prepared using the Z

competent cell procedure kit (Zymor Research) and that had transformation efficiencies of 10^8

cfu per mL. A typical transformation contained 5 µl of ligation reaction and 200 µl of Z

competent cells that were incubated at 0°C for 1 hr. These transformed cells were inoculated into

800 µL of SOC media and incubated at 37°C for 40 minutes. The cells were pelleted by

centrifugation at 2,000 x g, washed with 1 mL of M9 media, repelleted, and resuspended in 1 mL of M9 media.

Enrichments

The above washed transformants were inoculated into 35 mL of M9 media containing 0.2% lactose, 10 μg/mL thiamine, 100 μg/ml ampicillin with 1 percent vitamin solution (Sigma). These cultures were shaken at 18°C at 350 rpm in an Innova 4335 refrigerated incubator shaker (New Brunswick Scientific) or at 250 rpm in a Psychrotherm refrigerated shaker (New Brunswick Scientific). When an optical density at 600 nm of ~0.3 was reached, a 0.5 mL aliquot was removed and inoculated into a fresh culture (as listed above).

To determine the minimal growth rate increase that is detectable as a 99 percent population of cells in an enrichment of this type, the following equation was derived:

$$\Delta GR = \{ 1 + [1/(r_w * t)] * \ln [(0.99 * C) / (0.01 * D)] \} * 100$$

Equation 5.1

where ΔGR is the percent increase in growth rate, r_w is the doubling time of wild type cells expressed in hours, t is the time spent in exponential growth expressed in hours, and $C + D = 1$ where C is the starting population of wild type cells and D is the starting population of the faster growing cells.

The percent change in growth rate is based on the number of hours spent in exponential growth that would yield a population that contains 99 percent of the fastest growing cells. For the calculations in this work, a conservative estimate of 1 in 10,000 was made for the initial

population of faster growing cells (C = 0.9999, D = 0.0001). A larger initial population of cells that grow faster than wild type would reach the distribution of 99 out of 100 cells in less time. If the fastest growing cells have increases in doubling time that are smaller than the minimal growth rate increase detectable or if there are no cells that grow faster than wild type, then a single colony selected after the enrichment is not likely to have a growth rate faster than wild type.

Plasmid isolation and retransformation

Cells from the enrichments were streaked onto LB plates containing 100 μg/mL ampicillin and incubated overnight at 37°C. Isolated colonies were selected by sterile toothpick and inoculated into a 3 mL LB culture containing 100 μg/mL ampicillin. The plasmids were purified using the Wizard Miniprep System (Promega).

Growth rate determination

To remove any possible effect of host mutations which may have occurred during the long enrichment times, new *E. coli* ER2585 F' host cells were transformed and used in the growth rate determinations. Starter cultures were prepared by inoculation into M9 minimal media containing 100 μg/mL ampicillin, μg/mL thiamine, vitamins, and 0.02% glucose. These starter cultures were grown to a density of 100 klett units and inoculated into 30 mL M9 minimal media containing 100 μg/mL ampicillin, μg/mL thiamine, vitamins, and either 0.2% glucose or lactose in a 250 mL sidearm klett flask to a final density of 10 klett units. These cultures were shaken at 18°C at 350 rpm and growth was followed by klett meter.

Growth rates were calculated by creating a graph of the log of the cell density plotted against time. Only the data in the linear portion of the growth curve was fitted to a straight line and the doubling time determined.

5.4 RESULTS

Creation of a host strain for selection

For efficient library screening, it was necessary to have a host strain that is highly transformable, has no background production of β-galactosidase activity, and has the ability to transport lactose across the cellular membrane. For this reason, *E. coli* ER2585 cells carrying a deletion of the entire *lac* operon were conjugated with *Salmonella typhimurium* CH3 F' cells which carry the gene for lactose permease on its F' plasmid. Conjugated *E. coli* ER2585 F' cells were selected for their ability to grow on minimal media with glucose, thiamine, and tetracycline. (*S. typhimurium* CH3 cells require methionine for growth while ER2585 cells have only a thiamine requirement. Genes for tetracycline resistance are carried on the F' plasmid).

To confirm that the ER2585 F' cells had no endogenous β-galactosidase production and that they could transport lactose across their membranes, the cells were transformed with plasmid carrying the gene for the SOS β-galactosidase. Transformed and untransformed cells were plated on minimal medium containing lactose and thiamine. Only cells transformed with the SOS β-galactosidase were able to hydrolyze lactose and grow on this medium, indicating that ER2585 F' cells had no endogenous β-galactosidase production and that they could transport lactose across their membranes.

Verification of selective pressure in *E coli* ER2585 F'

To determine if there is a strong selective pressure for β-galactosidase activity when

cultures are grown in lactose at lower temperatures and to determine if one can discriminate

between β-galactosidases with different temperature-activity ranges, plasmids carrying genes for

three β-galactosidases (SBA, SOS, bgaB) with different temperature optima (15°C, 42°C, 65°C,

respectively) were transformed into ER2585 F' cells and their growth in M9 minimal media with

vitamins and 0.2% glucose or 0.2% lactose was followed using a klett meter to monitor (Table

5.1).

Table 5.1. Doubling times of *E. coli* ER2585 F'

			Doubling time (hours)		
Carbon source	Temperature	pΔαN	pΔαN/SBA	pΔαN/SOS	pΔαN/bgaB
Glucose	37 °C	2.1	2.0	2.5	2.1
Lactose	37 °C	0	0	4	17
Glucose	18 °C	6.1	6.2	6.5	6
Lactose	18 °C	0	7	12	0

Saturation mutagenesis of the N-terminus of SOS

In order to determine if the N-terminal region of the SOS enzymes was important for

increasing the activity at lower temperatures, the SOS gene was subjected to saturation

mutagenesis using PCR with primers that were degenerate in the first base pair of the third

through sixth codons of the N-terminus. The mutagenized genes were ligated into expression

vectors, transformed into ER2585 F' cells, and inoculated into minimal media containing lactose,

thiamine, and ampicillin. Samples of these transformations, plated onto LB plates containing X-

gal and ampicillin, indicated that over 5,000 transformants were inoculated into the liquid

cultures.

Enrichments

Two cultures as well as a wild type control culture were maintained in the exponential growth phase for 60 hours at 18°C by periodically removing aliquots and inoculating them into fresh media. After 60 hours, calculations indicated that if the growth rate was improved by 2% or more, then 99 out of any 100 cells isolated from the culture would be the faster grower (Equation 5.1).

Growth rate characteristics of retransformed cells

Cells from each culture were streaked onto plates, incubated at 37°C, and isolated colonies were selected. The enrichment products were designated as *sos*E1, *sos*E2, and *sos*WT. To correct for the effects of host mutations that may have arisen during the enrichment procedure, plasmid from these isolates was purified and retransformed into fresh host cells. The effects on growth characteristics were examined.

The growth of ER2585 F' cells containing *sos*E1, *sos*E2, and *sos*WT at 18°C was followed by klett measurement (Figure 5.1), and doubling times were determined (Table 5.2).

Growth Curves of cells containing sosE1, sosE2, and wild type enzymes
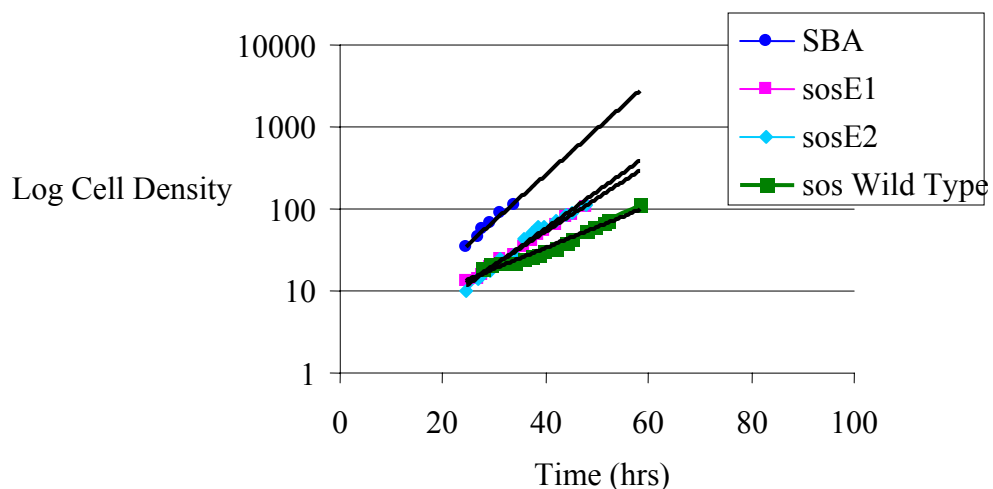
Figure 5.1. Growth curves of cells containing sosE1, sosE2, and wild type enzymes. Only the exponential phase of growth is included.

Table 5.2. Doubling Times

	Doubling time (hours)
SBA wild type	7.5
sosE1	8.5
sosE2	8.5
sosWild Type	13.2

5.5 DISCUSSION

The existence of a strong selective pressure is required for evolution both *in vivo* and *in vitro*. For directed evolution experiments, the selection pressure will define the number, distribution, and character of possible variant phenotypes. In the strategy used here, the existence of a selective pressure and the range in catalytic efficiencies that could be selected for was determined using the genes for three enzymes, each with different thermal properties.

SBA is a psychrophilic β-galactosidase that has an optimal temperature for activity of 15°C and loses all activity after 5 minutes at 37°C. At 18°C in lactose-containing media, cells containing this β-galactosidase grow fastest. Cells containing the SOS β-galactosidase which has a 42°C temperature optimum and only 20% activity at 18°C have doubling times that are nearly twice as high. Cells containing the BgaB β-galactosidase with a temperature optima of 65°C and no activity at 18°C do not grow at all. The higher doubling times of SOS and BgaB containing cells suggest that improvements in the catalytic efficiencies at 18°C of these two β-galactosidases would be detectable by changes in growth rates. Thus, under conditions that select for organisms based on growth rates, an adequate selective pressure exists for the evolution of β-galactosidases with higher catalytic efficiency at these temperatures. Growth in glucose-containing media indicate that the smallest doubling times achievable under these conditions irrespective of β-galactosidase activity is ~6 hrs at 18°C. Thus, β-galactosidase catalytic efficiencies that lead to smaller doubling times could not be selected. Because there may be additional limitations on growth rates due to the rate of lactose transport across the cellular membrane, this experiment does not reveal if the catalytic efficiencies can be improved beyond that of the SBA β-galactosidase.

Further evidence supporting the feasibility of this enrichment selection strategy is seen in the growth rates of the variants. In both cases where populations were enriched for β-galactosidase activity at low temperature, variants were isolated that exhibited faster growth characteristics. These characteristics are not likely due to spontaneous mutations that occur in the host strain during the enrichment. Transformation of fresh host strain cells with variant plasmid continue to carry the faster growth phenotype. While arguments could be made that conditions in

all of these minimal cultures may induce higher rates of mutation in host cells, this phenotype is not likely a result of such an occurrence when we consider that cultures of host strain cells with the wild type gene do not exhibit faster growth.

The increased growth rate of the variant isolated in the enrichments suggests that its β-galactosidase has improved catalytic properties. If improvements in catalytic efficiency can be equated with growth rate, as suggested above, then catalytic efficiency values for these β-galactosidases could provide useful information. The SBA β-galactosidase has been found to have a K_m value of 21 mM towards lactose and a k_{cat} value of 61.85 per second (Coker J.A., unpublished). The K_m and k_{cat} values for SOS β-galactosidase found in Chapter 6 are 15 mM and 1.5 per second, respectively. This ten fold difference in catalytic activity between SBA and SOS wild type enzymes may be responsible for the increase in doubling times from 7 hours to 12 hours, respectively. The observed 8.5 hour doubling time observed for both variants suggest the properties of the variant enzymes have significantly changed. It is not known if the relationship between catalytic efficiency of the β-galactosidase and growth rate of the organism is linear under these conditions. Purification and characterization of the variant enzymes may provide useful information about this relationship.

Also, great care must be given in making relations of this sort as observed growth rates may be highly dependent upon experimental conditions. Exact rpm speed of the shaker, temperature fluctuations, and the shape and type of flask used may each impact the growth rate determined over the long period of the experiment. Thus, it is best to make growth rate comparisons between cultures incubated in the same experiment.

The relative increased growth rates of cells isolated from enrichments also provide information about the structure of the SOS β-galactosidase. Because only the N-terminus of the protein was targeted for mutagenesis, it can be surmised that the N-terminus of SOS β-galactosidase contains determinants for catalysis. That this phenotype arose out of lactose selections at low temperature, suggests these determinants either improve specificity for lactose, increase catalysis at all temperatures, or are a specifically low temperature activating. While the exact identity and nature of these mutations has yet to be determined, they do signify that the N-terminal region of this family 42 $(\alpha/\beta)_8$ barrel glycosyl hydrolase has structural elements that can activate catalysis at low temperature.

These results provide the first evidence that this mechanism may reach beyond the BgaB enzyme. The next logical step will be to apply these same principles to other, more distantly related enzymes in family 42. Phylogenetic analysis of the 16s rRNA sequences of several family 42 members, indicate that there are other members in this family that are more distantly related to bgaB and SOS than they are to each other (Figure 5.2). Saturation mutagenesis or other mutagenic techniques could be applied to one of these more distantly related enzymes. Success in finding variants that gain the function of increased activity at lower temperatures or increased thermostability would indicate this mechanism is likely to apply to the entire family.
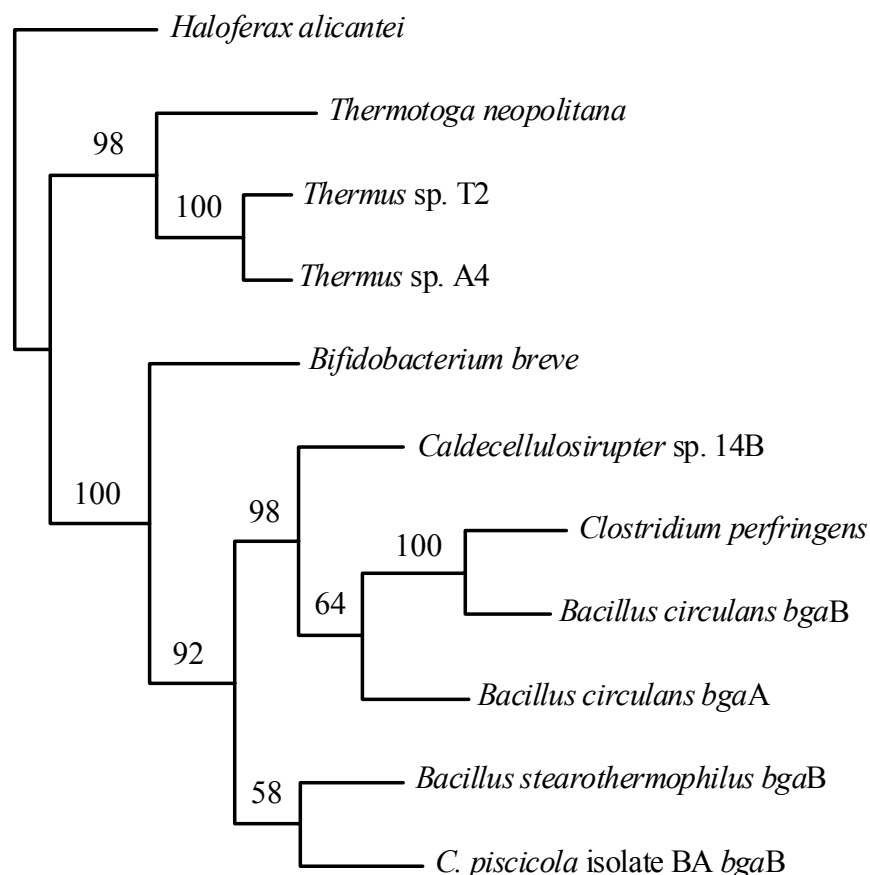
Figure 5.2. Phylogenetic tree for genes of selected family 42 enzymes

Perhaps a more exciting prospect however, is that this mechanism may be extended to enzymes outside of family 42. It is not known to what level such a principle could be extended. The 4/7 super family of glycosyl hydrolases, also known as the GH-A clan, is the next highest grouping of relatedness. These $(\alpha/\beta)_8$ barrels share similar topology and similar locations in the residues participating in catalysis. As mentioned in Chapter 1, in the 4/7 super family, the two conserved catalytic residues are located on the fourth and seventh α-helices of the barrel. Since these helices are located at the C-terminal end of the β strands integral to the barrel, this mechanism for temperature adaptation may function through interactions that support or restrict movement of β sheet(s). Effects of mutations that stabilize or destabilize the N-terminal region

may be communicated to the active site through the mobility of sheet $\beta 1$. It is not known to what extent the sheets internal to the barrel have mobility. Nor is it understood if the ellipticity of the barrel, which can vary among different enzymes, plays a role in any β sheet mobility. It is known, however, that β strands define the elliptical axis of the barrel and thus the barrel's ellipticity may be a factor to consider. Lastly, in $(\alpha/\beta)_8$ barrels, the β strands have a slight twist with respect to each other. The degree of this twist or skew varies between superfamilies of barrels [1]. All of these features - the mobility of β sheets, barrel ellipticity, and barrel twist - may play a role in the conformational flexibility of the $(\alpha/\beta)_8$ barrel motif. This flexibility, in turn, may translate changes at the N-terminus into changes in thermal properties.

To test if this paradigm of N-terminal modification for thermal adaptation is affected by such features it would be exciting to apply the principles of saturation mutagenesis to the N-terminus of $(\alpha/\beta)_8$ barrel enzymes beyond the glycosyl hydrolase group. Findings from such experiments could allow us to determine a new mechanism for temperature adaptation in all $(\alpha/\beta)_8$ barrel enzymes.

5.6 REFERENCES

1.   Reardon, D. and G.K. Farber, *The structure and evolution of α/β barrel proteins.* FASEB Journal, 1995. **9**: p. 497-503.

2.   Raychaudhuri, S., et al., *Backbone makes a significant contribution to the electrostatics of α/β-barrel proteins.* Protein Science, 1997. **6**: p. 1849-1857.

3.   Gupta, M.N.E., *Thermostability of enzymes.* 1993, Berlin: Springer.

4.   Sheridan, P.P. and J.E. Brenchley, *Characterization of a salt-tolerant family 42 β-Galactosidase from a psychrophilic antarctic Planococcus isolate.* Applied and Environmental Microbiology, 2000. **66**(6): p. 2438-2444.

5.   Trimbur, D.E., et al., *Characterization of a psychrotrophic Arthrobacter gene and its cold-active β-galactosidase.* Appl. Environ. Microbiol., 1994. **60**(12): p. 4544-4552.

Chapter 6

DIRECTED EVOLUTION OF ACTIVITY AT LOW

TEMPERATURES IN SOS ORANGE β-GALACTOSIDASE

6.1 ABSTRACT

Efforts to improve low temperature activity through directed evolution of enzymes already adapted to low temperatures are hampered by the complicated selection processes necessary to identify increased enzymatic efficiency in mesophilic host strains. The entire gene for a psychrophilic family 42 β-galactosidase from Antarctic isolate SOS Orange was subjected to random mutagenesis in order to find new mutations that lead to increased enzyme activity at low temperatures. Enrichments with lactose as the sole carbon source were used to select two variants with doubling times 30% faster. The β-galactosidases derived from these variants have 25% and 50% increases in catalytic efficiency towards lactose. These increases are primarily due to improvements in the k_{cat} value. The thermal dependence of activity however, remains unchanged.

6.2 INTRODUCTION

Creating a more psychrophilic enzyme from an enzyme that is already psychrophilic in nature is a difficult challenge. The window of selective pressure available to a researcher employing directed evolution becomes smaller and smaller as the activity becomes more extreme. A simple scheme, similar to the one used in Chapter 4, is limited by the growth characteristics of the host organism in which the selection is taking place. Using methods of enrichment however, where improvements in the catalytic efficiency of an enzyme can be selected for on the basis of marginal differences in growth rate, these limitations can partially be overcome, allowing us to explore the boundaries of temperature adaptation.

In this chapter, I use the enrichment strategy described in Chapter 5 to select for variants with increased β-galactosidase activity at low temperature in the psychrophilic SOS enzyme. In this case, however, mutagenesis is applied to the *entire* gene in order to determine new mutations that lead to temperature adaptation.

At least three classes of adaptive mutations are expected: increases in activity at low temperature, increased activity at all temperatures, and/or improvements in substrate specificity. The likelihood of finding any particular type of mutation is based on the degree of selective pressure available, the types of selective pressure, and the relative frequency of areas in the structure that contain determinants of the property of interest, in this case, the breakdown of lactose at low temperatures.

Because lactose is most likely not the native substrate of the SOS Orange enzyme (lactose is not usually found in high quantities in Antarctic environments) the use of lactose as a selective agent allows this research to have a broader impact. Up to 15 percent of Caucasians, 80

percent of blacks and Latinos, and nearly 100 percent of American Indians and Asians suffer from lactose intolerance [1]. Currently, processes to remove lactose from milk products involve the use of enzymes with higher temperature optima. These enzymes typically only function during the higher temperatures observed during the pasteurization process. At low temperatures, these enzymes cease to work and detectable amounts of lactose still remain present in milk. The evolution of an enzyme that could continue to catalyze the breakdown of lactose during the storage of milk at low temperatures would benefit the millions of people who remain sensitive to the lactose levels currently achieved in "lactose free" milk.

6.3 EXPERIMENTAL PROCEDURES

Random mutagenesis of the *SOS* β-galactosidase gene

The *SOS* gene was amplified by PCR using the DE2-F and DE2-R primer (sequence in Chapter 4). Reactions contained 1 μM of each primer, 10 ng of linearized pΔαH/*SOS* plasmid DNA, 7 mM $MgCl_2$, 50 mM KCl, 10 mM Tris (pH 8.3), 0.01% (wt/vol) gelatin, 200 μM dGTP, 200 μM dATP, 1 mM dCTP, 1 mM dTTP, 150 μM $MnCl_2$, and 0.05 units/μL of Taq DNA polymerase (Promega). A 500 μL master PCR mix was divided into 25 μL aliquots and PCR was carried out in an Ependorff Master Gradient Thermal Cycler (Eppendorf Scientific) at 95°C for 5 minutes, with 35 cycles of 95°C for 30 s, 62°C for 30 s, and 72°C for 2.75 minutes, followed by 7 minutes at 72°C. These conditions were expected to produce two to five base pair changes per gene sequence or one amino acid substitution per gene product [2].

DNA manipulation, enrichments, and growth rate determination

DNA manipulation, enrichments, and growth rate determination, over expression vector creation, expression, and purification were performed as described in Chapter 5.

Thermal dependence of enzyme activity

All assays were performed in triplicate. A reaction contained 1.19 mL of 25 mM sodium phosphate (pH 7.0), 10 mM KCl, 1 mM MgSO$_4$, and 2.2 mM of the chromogen O-nitrophenyl β-D-galactopyranoside (ONPG). The reaction was allowed to equilibrate to temperature for 15 minutes in an Isotemp refrigerated circulating water bath (Fisher Scientific) with a thermal accuracy of ±0.05°C. The reaction was started with the addition of 10 μL of enzyme and was quenched after 5.0 minutes with 500 μL of 1 M sodium carbonate and the absorbance at 420 nm was measured against a blank without enzyme in a Spectronic Genesis 2 Spectrophotometer (Spectronic Instruments).

Steady state kinetics with ONPG as substrate

Triplicate reactions contained 25 mM Na$_x$PO$_4$ (pH 7.0), 10 mM KCl, 1 mM MgSO$_4$ and ONPG at the following concentrations: 332, 442, and 664 μM and 1.3, 3.3, 6.6, and 13.3 mM. Absorbance at 420 nm was followed in a water jacketed cuvette holder which was pre-equilibrated for 15 minutes with an Isotemp refrigerated circular water bath (Fisher Scientific) with a thermal accuracy of ±0.05°C. Final temperature was verified with an electronic thermometer. The extinction coefficient for ONP at pH 7.0 was determined to be 2.2 mM^{-1}cm^{-1}. Kinetic and error analysis was performed using the Nonlin package.

Steady state kinetics with lactose as substrate

Triplicate reactions contained 25 mM sodium phosphate (pH 7.0), 10 mM KCl, 1 mM MgSO$_4$, 0.5 mM NAD(P)$^+$, 1 mM ATP, 0.01 units/µL hexokinase, 0.01 units/µL glucose-6-phosphate dehydrogenase, and lactose at the following concentrations: 0.1, 0.25, 0.5, 0.75, 1, 5, 10, 25, and 50 mM. Absorbance at 420 nm was followed in a water jacketed cuvette holder which was pre-equilibrated for 15 minutes to either 18°C or 37°C with an Isotemp refrigerated circulating water bath (Fisher Scientific) with a thermal accuracy of ±0.05^{o}C. Final temperature was verified with an electronic thermometer. Kinetic and error analysis was performed using the Nonlin package.

Sequencing and sequence alignment

Sequence alignment of family 42 glycosyl hydrolases were compiled using MEGALIGN (DNAStar, Inc) via Clustal W. All DNA sequencing was performed at the Nucleic Acid Facility (The Pennsylvania State University).

6.4 RESULTS

Random mutagenesis and enrichments

Mutagenic PCR was performed in ~70 independent 20 µL reactions in each of three rounds. Sequencing of 10 selected mutagenesis products revealed an average of 2 base pair changes per gene in the first two rounds and 3 base pair changes per gene in the third. The mutagenized genes produced in each round of mutagenesis were ligated into selection vectors

and transformed into *E. coli* ER2585 F' cells. Samples of each transformation were plated onto LB plates with ampicillin and X-gal in order to determine the approximate number of transformants that would be under selection in the enrichments. Approximately 5,000 transformants were under selection in each round.

Transformed cells were placed into a total of 92 enrichments at 18°C in minimal media plus vitamins and ampicillin with lactose as the sole carbon source (M9LVA). Enrichments proceeded for an average of 480 hours at which point cells were streaked out onto LB plates with ampicillin and a single colony was selected and its plasmid purified. Plasmids isolated from the 30 enrichments in round 1 were combined and used as template in the mutagenic PCR reactions of round 2. Only those plasmids isolated from round 2 enrichments that resulted in increased growth rates were used as template in round 3. Some variants were selected from each round for growth rate analysis.

Growth rate determinations

To remove any possible effect of host mutations which may have occurred during the long enrichment times, new host cells were retransformed prior to growth rate determination.

The growth of 64 variants was followed in M9LVA media at 18°C (Figure 6.1). Of these, 37 had growth rates faster than that of host cells containing wild type enzyme (Table 6.1).
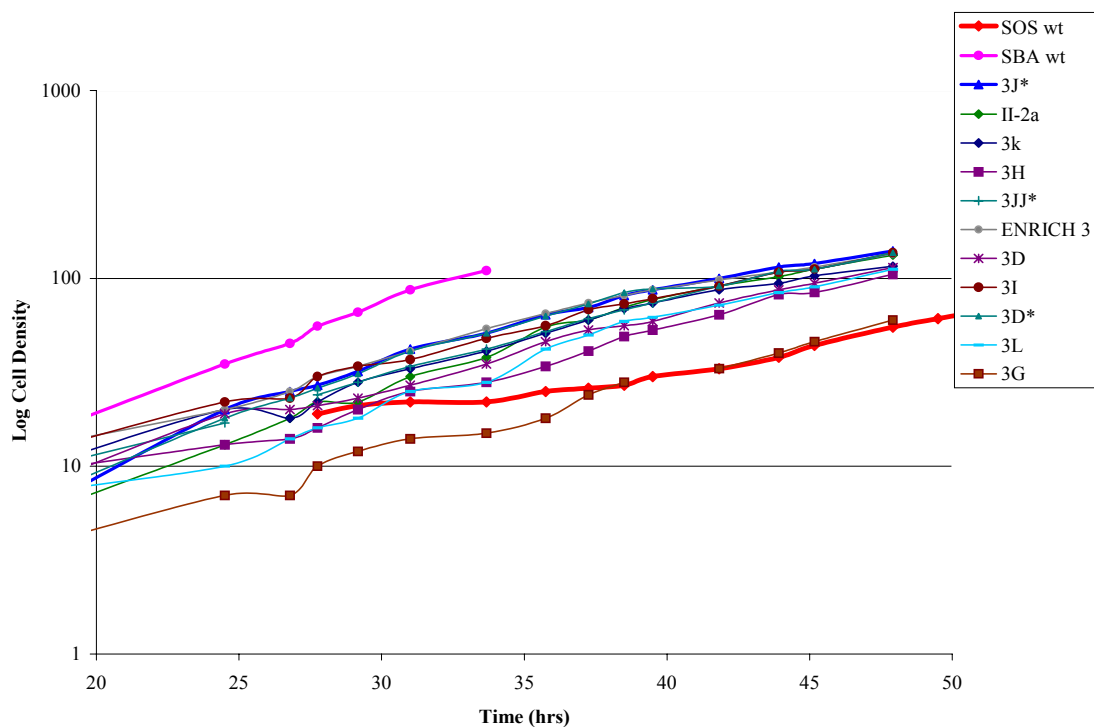
Figure 6.1. Growth curves of select variants

Enzyme characterization

The genes for seven of the variants exhibiting faster growth were cloned into over expression vectors and their enzymes over expressed and purified. Five had specific activities slightly greater than wild type while two, II-2a and 3J, had specific activities more than twice that of wild type (Table 6.1). The thermal dependence of activity was similar in all cases and their optimal temperatures were within 2$^{\mathbf{o}}$C of wild type enzyme.

Table 6.1. Characteristics of select variants.
Roman numeral designates the round of mutagenesis.
Specific activity is expressed in μmoles of ONPG/min/mg of enzyme.

Enzyme Studied	Doubling time at 18°C (hrs)	Specific activity at 39°C	Specific activity at 18°C	S.A. Ratio 18°/39°	$T_{optimum}$ (°C)
Wild type	10 ± 1.1	150 ± 12	57 ± 5	.380	43
SBA Wild Type	5.4 ± 0.8				
II-2A	7.0	270 ± 21	100 ± 8	0.370	42-44
III-3J	7.0	240 ± 15	95 ± 5	0.395	42-44
III-3U	6.5	211 ± 19	76 ± 6	0.361	42-44
III-3D	6.5	215 ± 18	83 ± 3	0.386	42-44
I-1A	7.5	191 ± 10	72 ± 2	0.378	42-44
I-3C	7.2	142 ± 14	54 ± 3	0.364	42-44
I-3I	8.5	127 ± 9	50 ± 3	0.393	42-44
III-3G	7.5				
III-3O	8				
III-3K	7.3				
III-3L	5.7				
III-3H	7.0				
III-3JJ	7.3				
II-3F	8.5				
I-4A	8.5				
I-4C	8.0				
I-4F	7.5				
II-2Q	9.0				
II-10	7.5				
III-5A	7.0				
III-7A	8.5				
II-3P	8.0				
II-3S	7.5				
II-2D	7.5				
II-2SS	7.75				
III-2R	8.5				
II-11	7.8				
III-13	7.4				

Kinetic characterization with ONPG as substrate

Steady state kinetic measurements were performed with the II-2a and wild type enzymes

to determine the k_{cat} and K_m values for the ONPG substrate (Table 6.2).

Table 6.2 Kinetic values for II-2a and III-3J variants on ONPG substrate

	k_{cat} (1/sec)	K_m (mM)	k_{cat}/ K_m
Wild Type	417 (+56 -58)	2.9 (+0.21 -0.21)	140
II-2a	858 (+79 -81)	3.1 (+0.28 -0.33)	276
III-3J	646 (+71 -63)	3.0 (+0.28 -0.33)	215

Kinetic analysis with lactose as substrate

In order to analyze the kinetic values for these enzymes with lactose as a substrate, a coupled assay using hexokinase and glucose-6-phosphate dehydrogenase was developed. In this coupled reaction, β-galactosidase hydrolyzes lactose to yield galactose and glucose. In the presence of ATP, hexokinase (HK) will catalyze the phosphorlyation of glucose to produce glucose-6-phosphate, ADP, and a proton. Glucose-6-phosphate dehydrogenase (G6PDH) can then catalyze the NAD(P) dependent oxidation of glucose-6-phosphate to produce 6-phosphogluconolactone, a proton, and the chromogenic NAD(P)H [3]. The production of NAD(P)H is followed at 340 nm.

To optimize the conditions necessary at the 18°C assay temperatures, the concentrations of HK and G6PDH were steadily increased independently for a given amount of β-galactosidase until the initial reaction rate remained unchanged. To verify that these concentrations were indeed sufficient to monitor β-galactosidase activity, the concentrations of SOS wild type and variant enzymes in the reaction were doubled, and a two-fold increase in the reaction rate was observed.

Kinetic assays of the wild type, II-2, and III-3J enzymes revealed that their K_m values on lactose substrate were essentially unchanged while their k_{cat} values and catalytic efficiencies were 50% and 20% higher, respectively (Table 6.3).

Table 6.3 Kinetic values for variants and wild type on lactose substrate at 18°C

	k_{cat} (1/sec)	K_m (mM)	k_{cat}/ K_m
Wild Type	1.5 (+0.09 -0.08)	14.95 (+0.83 -0.82)	0.10
II-2a	2.2 (+0.10 -0.12)	14.5 (+0.79 -0.81)	0.15
III-3J	1.8 (+0.13 -0.09)	14.6 (+1.2 -1.6)	0.12

Sequence analysis

Nucleic acid sequence analysis of the II-2a and III-3J variants revealed only one non-synonymous mutation, K206G and L423F respectively, in each enzyme.

6.5 DISCUSSION

50% and 20% increases in catalytic efficiency and k_{cat} values of the II-2a and III-3J variants appear to lead to a decrease in growth rate doubling time of 3 hours at 18 °C. Both II-2a and III-3J appear to have adapted to the selection in such a way as to increase their catalytic efficiencies. Analysis with ONPG substrate revealed similar, but less pronounced effects. Catalytic efficiency was doubled in the case of II-2a and increased by 50% in III-J each as a result of higher k_{cat} values. These results indicate that even minor increases in catalytic efficiency can be selected for on the basis of changes in growth rate in lactose selections.

6.5 REFERENCES

1. Swagerty, D.L., Jr., A.D. Walling, and R.M. Klein, *Lactose intolerance.* Am Fam Physician, 2002. **65**(9): p. 1845-50.

2. Miyazaki, K., et al., *Directed evolution study of temperature adaptation in a psychrophilic enzyme.* Journal of Molecular Biology, 2000. **297**: p. 1015-1026.

3. Tsai, C.S. and Q. Chen, *Purification and kinetic characterization of hexokinase and glucose-6-phosphate dehydrogenase from Schizosaccharomyces pombe.* Biochemistry and Cell Biology, 1998. **76**: p. 107-113.

Appendix A

bgaB DNA and Amino Acid Sequence

```
          10        20        30        40        50        60
....|....|....|....|....|....|....|....|....|....|....|....|
1   ATGAATGTGTTATCCTCAATTTGTTACGGAGGAGATTATAACCCAGAGCAATGGCCAGAG 60
     M  N  V  L  S  S  I  C  Y  G  G  D  Y  N  P  E  Q  W  P  E

          70        80        90       100       110       120
....|....|....|....|....|....|....|....|....|....|....|....|
61  GAAATTTGGTATGAAGATGCTAAGTTGATGCAAAAAGCGGGGGTGAATTTAGTATCTTTA 120
     E  I  W  Y  E  D  A  K  L  M  Q  K  A  G  V  N  L  V  S  L

         130       140       150       160       170       180
....|....|....|....|....|....|....|....|....|....|....|....|
121 GGGATTTTCAGTTGGAGCAAGATCGAACCGTCTGATGGAGTGTTCGACTTTGAATGGCTA 180
     G  I  F  S  W  S  K  I  E  P  S  D  G  V  F  D  F  E  W  L

         190       200       210       220       230       240
....|....|....|....|....|....|....|....|....|....|....|....|
181 GACAAGGTTATAGATATACTATATGACCACGGTGTTTATATTAACTTGGGGACGGCGACT 240
     D  K  V  I  D  I  L  Y  D  H  G  V  Y  I  N  L  G  T  A  T

         250       260       270       280       290       300
....|....|....|....|....|....|....|....|....|....|....|....|
241 GCAACTACTCCAGCTTGGTTTGTAAAAAAGTATCCAGATTCTTTGCCGATCGATGAAAGC 300
     A  T  T  P  A  W  F  V  K  K  Y  P  D  S  L  P  I  D  E  S

         310       320       330       340       350       360
....|....|....|....|....|....|....|....|....|....|....|....|
301 GGTGTCATTCTCTCGTTTGGCAGTAGACAACATTATTGTCCTAATCATCCTCAATTAATT 360
     G  V  I  L  S  F  G  S  R  Q  H  Y  C  P  N  H  P  Q  L  I

         370       380       390       400       410       420
....|....|....|....|....|....|....|....|....|....|....|....|
361 ACGCACATAAAGAGACTTGTGAGGGCTATAGCAGAACGGTATAAAAATCATCCGGCACTC 420
     T  H  I  K  R  L  V  R  A  I  A  E  R  Y  K  N  H  P  A  L

         430       440       450       460       470       480
....|....|....|....|....|....|....|....|....|....|....|....|
421 AAAATGTGGCATGTTAATAATGAGTATGCATGTCACGTTTCCAAGTGTTTTTGTGAGAAT 480
     K  M  W  H  V  N  N  E  Y  A  C  H  V  S  K  C  F  C  E  N

         490       500       510       520       530       540
....|....|....|....|....|....|....|....|....|....|....|....|
481 TGTGCTGTCGCGTTTAGAAAGTGGCTAAAGGAAAGATATAAAACAATCGATGAATTAAAT 540
     C  A  V  A  F  R  K  W  L  K  E  R  Y  K  T  I  D  E  L  N
```

```
              550         560         570         580         590         600
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
541  GAACGTTGGGGTACAAACTTTTGGGGACAGCGATACAATCATTGGGATGAAATTAATCCC 600
      E  R  W  G  T  N  F  W  G  Q  R  Y  N  H  W  D  E  I  N  P

              610         620         630         640         650         660
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
601  CCTAGAAAGGCACCAACTTTTATTAATCCATCCCAAGAACTTGATTACTACCGTTTTATG 660
      P  R  K  A  P  T  F  I  N  P  S  Q  E  L  D  Y  Y  R  F  M

              670         680         690         700         710         720
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
661  AATGACTCAATTCTCAAGTTGTTTTTAACAGAAAAGGAAATTTTACGTGAGGTAACACCA 720
      N  D  S  I  L  K  L  F  L  T  E  K  E  I  L  R  E  V  T  P

              730         740         750         760         770         780
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
721  GATATTCCAGTATCAACTAATTTCATGGGTTCATTCAAACCGTTAAACTATTTTCAATGG 780
      D  I  P  V  S  T  N  F  M  G  S  F  K  P  L  N  Y  F  Q  W

              790         800         810         820         830         840
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
781  GCTCAGCATGTAGATATTGTGACATGGGACTCATATCCTGATCCCAGAGAGGGCTTGCCA 840
      A  Q  H  V  D  I  V  T  W  D  S  Y  P  D  P  R  E  G  L  P

              850         860         870         880         890         900
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
841  ATTCAGCACGCCATGATGAATGACCTTATGCGTAGTTTAAGAAAAGGTCAACCGTTTATT 900
      I  Q  H  A  M  M  N  D  L  M  R  S  L  R  K  G  Q  P  F  I

              910         920         930         940         950         960
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
901  TTGATGGAGCAGGTAACCTCACATGTTAACTGGCGCGATATTAATGTTCCAAAACCGCCA 960
      L  M  E  Q  V  T  S  H  V  N  W  R  D  I  N  V  P  K  P  P

              970         980         990        1000        1010        1020
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
961  GGTGTAATGCGTCTATGGAGTTATGCAACTATTGCCCGTGGTGCAGATGGTATTATGTTT 1020
      G  V  M  R  L  W  S  Y  A  T  I  A  R  G  A  D  G  I  M  F

             1030        1040        1050        1060        1070        1080
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1021 TTCCAGTGGCGTCAAAGTAGAGCAGGAGCTGAAAAATTCCACGGTGCAATGGTGCCCCAC 1080
      F  Q  W  R  Q  S  R  A  G  A  E  K  F  H  G  A  M  V  P  H

             1090        1100        1110        1120        1130        1140
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1081 TTTTTGAACGAGAATAATAGAATTTATAGGGAAGTTACACAGTTAGGGCAAGAGCTGAAA 1140
      F  L  N  E  N  N  R  I  Y  R  E  V  T  Q  L  G  Q  E  L  K
```

```
              1150        1160        1170        1180        1190        1200
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1141 AAGTTAGATTGTTTGGTCGGATCTAGAATCAAGGCAGAGGTCGCGATCATTTTTGATTGG 1200
      K   L   D   C   L   V   G   S   R   I   K   A   E   V   A   I   I   F   D   W

              1210        1220        1230        1240        1250        1260
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1201 GAAAACTGGTGGGCTGTCGAACTAAGTTCCAAACCACATAATAAACTAAGATATATTCCT 1260
      E   N   W   A   V   E   L   S   S   K   P   H   N   K   L   R   Y   I   P

              1270        1280        1290        1300        1310        1320
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1261 ATAGTTGAAGCTTATTATAGGGAATTATATAAACGTAATATTGCTGTCGATTTTGTAAGG 1320
      I   V   E   A   Y   Y   R   E   L   Y   K   R   N   I   A   V   D   F   V   R

              1330        1340        1350        1360        1370        1380
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1321 CCATCTGATGATCTAACAAAATACAAAGTAGTTATTGCTCCAATGTTATATATGGTTAAA 1380
      P   S   D   D   L   T   K   Y   K   V   V   I   A   P   M   L   Y   M   V   K

              1390        1400        1410        1420        1430        1440
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1381 GAGGGAGAAGATGAAAACTTACGGCAATTTGTTGCTAACGGTGGCACTTTGATTGTCAGT 1440
      E   G   E   D   E   N   L   R   Q   F   V   A   N   G   G   T   L   I   V   S

              1450        1460        1470        1480        1490        1500
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1441 TTCTTCAGTGGCATTGTAGATGAAAATGACCGTGTACATCTAGGCGGATATCCTGGTCCT 1500
      F   F   S   G   I   V   D   E   N   D   R   V   H   L   G   G   Y   P   G   P

              1510        1520        1530        1540        1550        1560
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1501 CTGCGAGATATTTTGGGGATTTTTGTTGAGGAATTTGTACCATACCCAGAAACAAAGGTA 1560
      L   R   D   I   L   G   I   F   V   E   E   F   V   P   Y   P   E   T   K   V

              1570        1580        1590        1600        1610        1620
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1561 AACAAAATATATAGTAACGATGGGGAATATGATTGTACGACGTGGGCGGACATAATCCGA 1620
      N   K   I   Y   S   N   D   G   E   Y   D   C   T   T   W   A   D   I   I   R

              1630        1640        1650        1660        1670        1680
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1621 TTAGAAGGGGCAGAACCTCTAGCGACATTTAAGGGGGATTGGTATGCAGGACTTCCGGCG 1680
      L   E   G   A   E   P   L   A   T   F   K   G   D   W   Y   A   G   L   P   A
```

```
               1690        1700        1710        1720        1730        1740
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 1681 GTTACACGTAACTGCTACGGTAAAGGAGAGGGGATTTACGTCGGTACTTATCCAGATAGT 1740
        V  T  R  N  C  Y  G  K  G  E  G  I  Y  V  G  T  Y  P  D  S

               1750        1760        1770        1780        1790        1800
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 1741 AATTATTTAGGCAGGCTTTTAGAACAGGTTTTCGCTAAACATCATATTAATCCCATTCTT 1800
        N  Y  L  G  R  L  L  E  Q  V  F  A  K  H  H  I  N  P  I  L

               1810        1820        1830        1840        1850        1860
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 1801 GAAGTAGCTGAAAATGTAGAGGTGCAACAAAGAGAGACTGATGAATGGAAGTATTTGATT 1860
        E  V  A  E  N  V  E  V  Q  Q  R  E  T  D  E  W  K  Y  L  I

               1870        1880        1890        1900        1910        1920
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 1861 ATCATCAATCATAATGATTACGAAGTGACGCTGTCACTGCCAGAAGATAAGATATACCAG 1920
        I  I  N  H  N  D  Y  E  V  T  L  S  L  P  E  D  K  I  Y  Q

               1930        1940        1950        1960        1970        1980
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 1921 AATATGATTGATGGGAAATGTTTTCGAGGAGGTGAATTGAGGATTCAAGGGGTTGATGTA 1980
        N  M  I  D  G  K  C  F  R  G  G  E  L  R  I  Q  G  V  D  V

               1990        2000        2010        2020        2030        2040
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 1981 GCAGTGTTAAGAGAGCATGATGAAGCCGGGAAGGTTTAGAGAAGTCTCGTTCCGACAGTT 2040
        A  V  L  R  E  H  D  E  A  G  K  V  *  R  S  L  V  P  T  V

               2050        2060        2070        2080
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
 2041 GGCAACATAATATGCATAAGATGACAATGTCTATAAACATTGGATC 2086
        G  N  I  I  C  I  R  *  Q  C  L  *  T  L  D
```

Appendix B
SOS Orange β-Galactosidase DNA and Amino Acid Sequence

```
                10        20        30        40        50        60
     ....|....|....|....|....|....|....|....|....|....|....|....|
1    ATGATTAACGATAAATTGCCGAAGATTTGGCACGGTGGGGACTATAACCCCGAACAATGG  60
      M  I  N  D  K  L  P  K  I  W  H  G  G  D  Y  N  P  E  Q  W

                70        80        90       100       110       120
     ....|....|....|....|....|....|....|....|....|....|....|....|
61   GATTCGAAAGAAATTTGGGACGAAGACGTTCGGATGTTCAAACTGGCAGGAATTGACGTG  120
      D  S  K  E  I  W  D  E  D  V  R  M  F  K  L  A  G  I  D  V

               130       140       150       160       170       180
     ....|....|....|....|....|....|....|....|....|....|....|....|
121  GCCACTCTAAACGTTTTTTCATGGGCACTCAATCAGCCGAATGAAGATACCTACAACTTT  180
      A  T  L  N  V  F  S  W  A  L  N  Q  P  N  E  D  T  Y  N  F

               190       200       210       220       230       240
     ....|....|....|....|....|....|....|....|....|....|....|....|
181  GACTGGCTAGATGAAAAAATAAATCGCCTCTACGAAAACGGCATTTACACGTGCCTCGCA  240
      D  W  L  D  E  K  I  N  R  L  Y  E  N  G  I  Y  T  C  L  A

               250       260       270       280       290       300
     ....|....|....|....|....|....|....|....|....|....|....|....|
241  ACAAGTACGGCAGCGCATCCGGCATGGATGGCGAAGAAATACCCGGATGTCTTGCGTGTC  300
      T  S  T  A  A  H  P  A  W  M  A  K  K  Y  P  D  V  L  R  V

               310       320       330       340       350       360
     ....|....|....|....|....|....|....|....|....|....|....|....|
301  GACTTTTACGGCAGAAAACGCAAATTCGGCAGTCGTCACAACTCGTGTCCGAACAGCCCG  360
      D  F  Y  G  R  K  R  K  F  G  S  R  H  N  S  C  P  N  S  P

               370       380       390       400       410       420
     ....|....|....|....|....|....|....|....|....|....|....|....|
361  ACTTATCGCAAATACTCGGAGCGAATTGCTGAAACATTAGCAGAGCGCTACAAAGACCAC  420
      T  Y  R  K  Y  S  E  R  I  A  E  T  L  A  E  R  Y  K  D  H

               430       440       450       460       470       480
     ....|....|....|....|....|....|....|....|....|....|....|....|
421  CCAGCGGTCTTGATTTGGCACGTCTCCAACGAATACGGCGGCTACTGCTATTGCGACAAT  480
      P  A  V  L  I  W  H  V  S  N  E  Y  G  G  Y  C  Y  C  D  N

               490       500       510       520       530       540
     ....|....|....|....|....|....|....|....|....|....|....|....|
481  TGCCAAGACGCTTTCCGCAACTGGTTAAGCGACAAATACGGAACGCTTGAAAAGCTCAAC  540
      C  Q  D  A  F  R  N  W  L  S  D  K  Y  G  T  L  E  K  L  N
```

```
             550         560         570         580         590         600
        ....|....|....|....|....|....|....|....|....|....|....|....|
541     AAAGCCTGGAACACCGGATTCTGGGGCCATACGTTTTACGAATGGGACGAAATTGTTGCA 600
         K  A  W  N  T  G  F  W  G  H  T  F  Y  E  W  D  E  I  V  A

             610         620         630         640         650         660
        ....|....|....|....|....|....|....|....|....|....|....|....|
601     CCGAATATGCTAAGCGAAAAACGTGAAGACAATGTATCTGATTTCCAAGGCATTTCACTC 660
          P  N  M  L  S  E  K  R  E  D  N  V  S  D  F  Q  G  I  S  L

             670         680         690         700         710         720
        ....|....|....|....|....|....|....|....|....|....|....|....|
661     GATTACCGCCGTTTCCAGTCGGATCGTTTACTGGATTGCTACAAGCTGGAATACAACGCC 720
          D  Y  R  R  F  Q  S  D  R  L  L  D  C  Y  K  L  E  Y  N  A

             730         740         750         760         770         780
        ....|....|....|....|....|....|....|....|....|....|....|....|
721     ATCCGCAAGCACGTGCCAACCAGCATCCCAATCACTACTAACTTGATGGGTACATATCCG 780
          I  R  K  H  V  P  T  S  I  P  I  T  T  N  L  M  G  T  Y  P

             790         800         810         820         830         840
        ....|....|....|....|....|....|....|....|....|....|....|....|
781     ATGCTGGATTACTTCAAATGGGCAAAAGAAATGGACGTCGTGTCTTGGGACAATTACCCG 840
          M  L  D  Y  F  K  W  A  K  E  M  D  V  V  S  W  D  N  Y  P

             850         860         870         880         890         900
        ....|....|....|....|....|....|....|....|....|....|....|....|
841     TCAATCGATACGCCGTTTAGCTACACCGCCATGACCCACGATTTGATGCGTGGATTAAAA 900
          S  I  D  T  P  F  S  Y  T  A  M  T  H  D  L  M  R  G  L  K

             910         920         930         940         950         960
        ....|....|....|....|....|....|....|....|....|....|....|....|
901     GGCGGGAAGCCGTTCATGCTGATGGAGCAAACGCCGAGCCAGCAAAACTGGCAGCCGTAT 960
          G  G  K  P  F  M  L  M  E  Q  T  P  S  Q  Q  N  W  Q  P  Y

             970         980         990        1000        1010        1020
        ....|....|....|....|....|....|....|....|....|....|....|....|
961     AATTCCCTGAAGCGCCCGGGCGTTATGCGTTTATGGAGCTACCAGGCAATCGGCCGCGGA 1020
          N  S  L  K  R  P  G  V  M  R  L  W  S  Y  Q  A  I  G  R  G

            1030        1040        1050        1060        1070        1080
        ....|....|....|....|....|....|....|....|....|....|....|....|
1021    GCAGATACCATTTTGTACTTCCAGTTGCGTCGCTCAGTCGGCGCTTGTGAGAAATACCAC 1080
          A  D  T  I  L  Y  F  Q  L  R  R  S  V  G  A  C  E  K  Y  H

            1090        1100        1110        1120        1130        1140
        ....|....|....|....|....|....|....|....|....|....|....|....|
1081    GGAGCGGTCATCGAACACGTAGGTCACGAGCATACCCGTGTATTCAATGAAGTTGCTCAA 1140
```

```
          G   A   V   I   E   H   V   G   H   E   H   T   R   V   F   N   E   V   A   Q

               1150        1160        1170        1180        1190        1200
          ....|....|....|....|....|....|....|....|....|....|....|....|
1141      TTAGGGCAAGAATTAAATGGGTTGTCCGACACCTTGCTCGATGCACGGGTCAACGCCAAA      1200
            L   G   Q   E   L   N   G   L   S   D   T   L   L   D   A   R   V   N   A   K

               1210        1220        1230        1240        1250        1260
          ....|....|....|....|....|....|....|....|....|....|....|....|
1201      GTGGCGATCGTCTTTGACTGGGAAAATCGCTGGGCGACCGAACTGTCAAGCGGACCTTCT      1260
             V   A   I   V   F   D   W   E   N   R   W   A   T   E   L   S   S   G   P   S

               1270        1280        1290        1300        1310        1320
          ....|....|....|....|....|....|....|....|....|....|....|....|
1261      GTGTCGCTCGACTATGTCAACGAAGTCCATAAATATTACGACGCGCTGTACAAATTGAAT      1320
             V   S   L   D   Y   V   N   E   V   H   K   Y   Y   D   A   L   Y   K   L   N

               1330        1340        1350        1360        1370        1380
          ....|....|....|....|....|....|....|....|....|....|....|....|
1321      GTCCAAGTCGACATGATTGGCGTCGAAGAAGACTTGAGCAAGTACGACGTCGTCATTGCA      1380
             V   Q   V   D   M   I   G   V   E   E   D   L   S   K   Y   D   V   V   I   A

               1390        1400        1410        1420        1430        1440
          ....|....|....|....|....|....|....|....|....|....|....|....|
1381      CCGGTTCTGTACATGGTCAAAGAAGGCTATGCAGCGAAAGTCGAGAAATTCGTCGAAAAT      1440
             P   V   L   Y   M   V   K   E   G   Y   A   A   K   V   E   K   F   V   E   N

               1450        1460        1470        1480        1490        1500
          ....|....|....|....|....|....|....|....|....|....|....|....|
1441      GGCGGTACGTTCCTGACGACTTTCTTCAGCGGCATTGTCAATGAAACCGATATCGTGACA      1500
             G   G   T   F   L   T   T   F   F   S   G   I   V   N   E   T   D   I   V   T

               1510        1520        1530        1540        1550        1560
          ....|....|....|....|....|....|....|....|....|....|....|....|
1501      CTCGGTGGCTATCCAGGTGAACTGCGTAAGGTTCTTGGCATTTGGGCAGAAGAAATCGAT      1560
             L   G   G   Y   P   G   E   L   R   K   V   L   G   I   W   A   E   E   I   D

               1570        1580        1590        1600        1610        1620
          ....|....|....|....|....|....|....|....|....|....|....|....|
1561      GCACTGCATCCGGACGAAACCAATCAAATCGTTGTAAAAGGATCCCGCGGAATTTTGAGC      1620
             A   L   H   P   D   E   T   N   Q   I   V   V   K   G   S   R   G   I   L   S

               1630        1640        1650        1660        1670        1680
          ....|....|....|....|....|....|....|....|....|....|....|....|
1621      GGCAAGTATTCATGTAATTTGCTGTTTGATTTGATCCATACAGAAGGCGCAGAAGCCGTT      1680
             G   K   Y   S   C   N   L   L   F   D   L   I   H   T   E   G   A   E   A   V

               1690        1700        1710        1720        1730        1740
          ....|....|....|....|....|....|....|....|....|....|....|....|
```

```
1681 GCAGAATATGGCTCTGACTTTTACAAAGGCATGCCGGTCTTGACCGTCAACAAATTCGGA 1740
        A  E  Y  G  S  D  F  Y  K  G  M  P  V  L  T  V  N  K  F  G
```

```
              1750      1760      1770      1780      1790      1800
        ....|....|....|....|....|....|....|....|....|....|....|....|....|
1741 AAAGGAAAAGCGTGGTACGTGGCTTCAAGCCCGGACGCAGAATTCCTCGTCGATTTCCTC 1800
         K  G  K  A  W  Y  V  A  S  S  P  D  A  E  F  L  V  D  F  L
```

```
              1810      1820      1830      1840      1850      1860
        ....|....|....|....|....|....|....|....|....|....|....|....|....|
1801 CAGACAGTATGCGAGGAAGCGGGCGTCGAGCCATTGCTTGATGTACCAGCAGGCGTCGAA 1860
         Q  T  V  C  E  E  A  G  V  E  P  L  L  D  V  P  A  G  V  E
```

```
              1870      1880      1890      1900      1910      1920
        ....|....|....|....|....|....|....|....|....|....|....|....|....|
1861 ACAACCGAGCGCGTGAAAGACGGCCAAACGTATTTGTTCGTGTTGAACCACAACAACGAT 1920
         T  T  E  R  V  K  D  G  Q  T  Y  L  F  V  L  N  H  N  N  D
```

```
              1930      1940      1950      1960      1970      1980
        ....|....|....|....|....|....|....|....|....|....|....|....|....|
1921 GAAGTCACGATCGAGCTGCACGGCAGCCAGTATAGAGAAGTGCTGACGGATGAACAAGTG 1980
         E  V  T  I  E  L  H  G  S  Q  Y  R  E  V  L  T  D  E  Q  V
```

```
              1990      2000      2010      2020      2030
        ....|....|....|....|....|....|....|....|....|....|....|....
1981 AGCGGAAACCTGGTCTTGAAAGAAAAAGGCGTGTTGATTTTGGCGAAAGTGTAA 2034
         S  G  N  L  V  L  K  E  K  G  V  L  I  L  A  K  V  *
```

Appendix C

Alignment of bgaB with β-gal from *Sulfulobus solfacataricus*
Partial secondary structure of *S. solfacataricus* crystal structure 1GOW


```
                    10         20         30
         ....|....|....|....|....|....|....|....
Gbstearo  1  -MNVLSSICYGGDYNPEQWPEEIWYEDAKLMQKAG
SS-bGal   1  MYSFPNSRFG-------------WSQAGFQSEMG


                    50         60         70         80
         ....|....|....|....|....|....|....|....|....|
Gbstearo 40  LGIFSWSKIEPSDGVFDFEWLDKVIDILYDHGVYINLGTA
SS-bGal  23  ---P--GSEDP-----NTDWYKWVHDPE-------NMAAG


                    89         99        109        119
         ....|....|....|....|....|....|....|....|....|
Gbstearo 80  TATTPAWFVKKYPDSLPIDESGVILSFGSRQHYCPNHPQL
SS-bGal  47  LVSG----------DLPENGPG---YWGNYKTFHDNAQKM


                   130        140        150        160
         ....|....|....|....|....|....|....|....|....|
Gbstearo 120 ITHIKRLVRAIAERYKN-HPALKMWHVNNEYACHVSKCFC
SS-bGal   74 GLKIARLNSEWSRQFPNPLPRPQNFDESKQDVTEVEINEN


                   170        180        190        200
         ....|....|....|....|....|....|....|....|....|
Gbstearo 159 ENCAVAFRKWLKERYKTIDELNERWGTNFWGQRYNHWDEI
SS-bGal  114 E---------LKR----LDEYAN-------KDALNHYREI
```

```
                       210       220       230       240
              ....|....|....|....|....|....|....|....|....|....|
Gbstearo  199 NPPRKAPTFINPSQELDYYRFMNDSILKLFLTEKEILREV
SS-bGal   134 FKDLKS------RGLYFIQNMYHWPLPLWLHDPIRVRRG
```

```
                       250       260       270
              ....|....|....|....|....|....|....|....|....|..
Gbstearo  239 TPDIPVSTNFMGSFKPLNYFQWAQHVDIVTWDSYPDPREG
SS-bGal   167 DFTGPS-----GWLSTRTVYEFARFSAYTAWK--FD---D
```

```
                       290       300       310       320
              ....|....|....|....|....|....|....|....|....|....|
Gbstearo  279 LPIQHAMMNDLMRSLRKGQPFILMEQVTSHVNWRDINVPK
SS-bGal   197 LVDEYSTMNEPN--------------VVGGLGYVGVKSGF
```

```
                       330       340       350       360
              ....|....|....|....|....|....|....|....|....|....|
Gbstearo  319 PPGVMRLWSYATIARGADGIMFFQWRQSRAGAEKFHGAMV
SS-bGal   223 PPG----------------YLSFELSRR----------AM
```

```
                       370       380       390       400
              ....|....|....|....|....|....|....|....|....|....|
Gbstearo  359 PHFLNENNRIYREVTQLGQELKKLDCLVGSRIKAEVAIIF
SS-bGal   237 YNIIQAHARAYDGIKSVS---KKP----------VGIIY
```

```
                 410       420       430       440
          ....|....|....|....|....|....|....|....|
Gbstearo  399 DWENWWAVELSSKPHNKLRYIPIVEAYYRELYKRNIAVDF
SS-bGal   263 ANSSFQPLTDKDMEAVEMAENDRWWFFDAIIRGEITRGN


                 450       460       470       480
          ....|....|....|....|....|....|....|....|
Gbstearo  439 VRPSDDLTKYKVVIAPMLYMVKEGEDENLRQFVANGGTLI
SS-bGal   303 EKIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLGG---


                 490       500       510       520
          ....|....|....|....|....|....|....|....|
Gbstearo  479 VSFFSGIVDENDRVHLGGYPGPLRDILGIFVEEFVPYPET
SS-bGal   339 ----YGHGCERNSVSLAGLP------TSDFGWEFFPEGLY


                 530       540       550       560
          ....|....|....|....|....|....|....|....|
Gbstearo  519 KVNKIYSNDGEYDCTTWADIIRLEGAEPLATFKGDWYAGL
SS-bGal   370 DVLTKYWN--RYHLYMYVTEN---GIADDADYQRPYY--L


                 570       580       590       600
          ....|....|....|....|....|....|....|....|
Gbstearo  559 PAVTRNCYGKGEGIYVGTYPDSNYLGRLLEQVFAKHHINP
SS-bGal   402 ---VSHVYQVHRAINSGADVRG----------YLHWSLAD


                 610       620       630       640
          ....|....|....|....|....|....|....|....|
Gbstearo  599 ILEVAENVEVQQRETDEWKYLIIINHNDYEVTLSLPEDKI
SS-bGal   430 NYEWASGFSMRFG-------LLKVDYNTKRLYWR-PSALV


                 650       660       670
          ....|....|....|....|....|....|....|....
Gbstearo  639 YQNMIDGKCFRGGELRIQGVDVAVLREHDEAGKV 672
SS-bGal   462 YREIATNGAITDEIEHLNSVPPVKPLRH------ 489
```

CURRICULUM VITAE

NICHOLAS PANASIK
211 S. FREAR
DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PA 16802
814.865.3330
nxp7@psu.edu

EDUCATION

Ph.D.-Biochemistry and Molecular Biology   August 2002
The Pennsylvania State University
Dissertation: Structural Basis for thermostability and thermal dependence of activity in α/β barrel glycosyl hydrolases, directed by Professor Jean E. Brenchley

Bachelor of Arts-Biochemistry   December 1993
University of Wisconsin-Madison

RESEARCH EXPERIENCE

Research Assistant   1996-present
The Pennsylvania State University.
Funding: National Science Foundation Research Training Grant Fellowship
Advisors: Professor Jean E. Brenchley, Professor Gregory K. Farber

Assistant Scientist   1994-1996
Ophidian Pharmaceuticals
Madison, Wisconsin
Project: Organic Synthesis of novel therapeutic compounds.

Research Assistant   1991-1994
University of Wisconsin-Madison
Project: Organic synthesis of proline analogs and analyzing their inductive effects on protein folding using X-ray crystallography, NMR, & IR.
Advisor: Professor Ronald T. Raines

HONORS AND AWARDS

National Science Foundation Research Training Grant Fellowship	1997-2001
University Wide Graduate Assistant Award for Outstanding Teaching	1998
Departmental Award for Outstanding Teaching	1997