THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

EURYTHERMALISM OF A DEEP-SEA SYMBIOSIS SYSTEM FROM AN ENZYMOLOGICAL ASPECT

A thesis submitted in partial fulfillment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Charles Kai-Wu Lee

Department of Biological Sciences, University of Waikato

2007

Abstract

The recently proposed and experimentally validated Equilibrium Model provides the most detailed description of temperature's effect on enzyme catalytic activity to date. By introducing an equilibrium between E_{act}, the active form of enzyme, and E_{inact}, a reversibly inactivated form of enzyme, the Equilibrium Model explains apparent enzyme activity loss at high temperatures that cannot be accounted for by irreversible thermal denaturation. The Equilibrium Model describes enzyme behavior in the presence of substrates and under assay conditions; thus its associated parameters, ΔH_{eq} and T_{eq}, may have physiological significance.

The Equilibrium Model parameters have been determined for twenty-one enzymes of diverse origins. The results demonstrated the wide applicability of the Equilibrium Model to enzymes of different types and temperature affinity. The study has also established ΔH_{eq} as the first quantitative measure of enzyme eurythermalism and demonstrated the relationship between T_{eq} and optimal growth temperature of organisms. The Equilibrium Model is therefore a useful tool for studying enzyme temperature adaptation and its role in adaptations to thermophily and eurythermalism. Moreover, it potentially enables a description of the originating environment from the properties of the enzymes.

The Equilibrium Model has been employed to characterize enzymes isolated from bacterial episymbionts of *Alvinella pompejana*. *A. pompejana* inhabits one of the most extreme environments known to science and has been proposed as an extremely eurythermal organism. A metagenomic study of the *A. pompejana*

episymbionts has unveiled new information related to the adaptive and metabolic properties of the bacterial consortium; the availability of metagenomic sequences has also enabled targeted retrieval and heterologous expression of *A. pompejana* episymbiont genes. By inspecting enzymes derived from the unique episymbiotic microbial consortium intimately associated with *A. pompejana*, the study has shed light on temperature adaptations in this unique symbiotic relationship. The findings suggested that eurythermal enzymes are one of the mechanisms used by the microbial consortium to achieve its adaptations.

By combining metagenomic and enzymological studies, the research described in this thesis has lead to insights on the eurythermalism of a complex microbial system from an enzymological aspect. The findings have enhanced our knowledge on how life adapts to extreme environments, and the validation of the Equilibrium Model as a tool for studying enzyme temperature adaptation paves the way for future studies.

Preface

The major components of the research, which was centered on enzyme eurythermalism and temperature adaptation of the bacterial episymbionts of the uniquely eurythermal, vent-dwelling worm, *Alvinella pompejana*, are as follows:

1) Development and validation of the Equilibrium Model as a tool for characterizing and comparing enzyme eurythermal properties, described in Chapter Three. This work formed the basis of this project and has been published in the FASEB Journal.
2) Bioinformatic analysis of *A. pompejana* episymbiont metagenome and construction and testing of the corresponding metagenome database, described in Chapter Four. I contributed to this international collaborative project as both a bioinformatician and an end-user of the metagenome database. A manuscript containing results from this work is being prepared for submission to Nature.
3) Identification and isolation of suitable genes from the *A. pompejana* episymbiont metagenome using the metagenome database, and characterization of enzymes derived from these genes using the Equilibrium Model to understand their potential eurythermal properties, described in Chapter Four and Chapter Five. This work has revealed new and useful information about this unique symbiotic system and has been submitted to Applied and Environmental Microbiology for publication.

Acknowledgement

I would like to thank the following people:

my family, for their love and support, despite my inability to show them exact what it is that I do;

my advisors, Prof. Roy Daniel and Prof. Craig Cary, for being great mentors and friends;

my friend and colleague, Dr. Michelle Peterson, for tirelessly reading and commenting on this thesis;

my lab mates, Dr. Tom Niederberger, Dr. Andreas Ruckert, Colin Monk, Lynne Parker, David Clement, Murielle Lopez, and more… for helping me along the way in and outside the lab;

Christina, for her endless love and (mostly) endless patience;

the land of New Zealand and its delightful people, for making me feel so welcomed and comfortable in my newly adopted homeland;

The University of Waikato and the United States National Science Foundation, for making possible all the research presented in this thesis.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

Abbreviations

APS	*Alvinella pompejana Proteobacteria* Episymbionts
ATCC	American Type Culture Collection
BLAST	The Basic Local Alignment Search Tool
COG	Clusters of Orthologous Groups of proteins
cDNA	Complementary DNA
dNTP	Deoxynucleoside triphosphate
EC number	Enzyme Commission
GOI	Genes of Interest
HMM	Hidden Markov Model
hsp	Heat Shock Proteins
KEGG	Kyoto Encyclopedia of Genes and Genomes
K_m	The substrate concentration at which an enzyme exhibits exactly half of its maximal activity
OGT	Optimal Growth Temperature
ORF	Open Reading Frame
PCR	Polymerase chain reaction™
PRIAM	*PRofils pour l'Identification Automatique du Métabolisme* (enzyme-specific profiles for metabolic pathway prediction)
qPCR	Real-time quantitative PCR
RT-qPCR	Reverse transcription-qPCR
RTS	Roche™ Rapid Translation System®
rDNA	ribosomal DNA

SNP	Single Nucleotide Polymorphism
WGS	Whole Genome Shotgun (sequencing)
aa	amino acids
bp	(DNA/RNA) base pairs
Kb	Kilobases
Mb	Megabases
Gb	Gigabases

Chapter One

– Introduction and Literature Reviews

The first section of this chapter is a brief introduction to enzyme temperature kinetics studies and how the development and validation of the Equilibrium Model has broadened scientists' view of this field. This part is largely based on findings described in [1, 2] and includes material from [3] and [4].

The second section is an overview of temperature adaptation studies with special emphases on eurythermalism and enzymes isolated from extremophiles. It provides a historical perspective on eurythermalism studies, a critical analysis on the findings so far, and how they may be complemented by studies of temperature adaptation at the level of individual enzymes. This section also looks at some of the findings on temperature adaptation at the enzymic level from studying enzymes isolated from extremophiles.

The last section of the chapter describes the deep-sea hydrothermal vent ecosystem, the polychaetous annelid *Alvinella pompejana* (The *Pompeii worm*), and its associated symbiotic bacteria, which are relevant to Chapter Four and Chapter Five. Since the *A. pompejana* episymbiont metagenome project was a critical component of this thesis, a brief description of metagenomics is also included.

We have to remember that what we observe is not nature in itself but nature exposed to our method of questioning.

– Werner Heisenberg (1901-1976)

I. A Review of the 'Equilibrium Model'

a. State of the Field prior to the Equilibrium Model

Temperature is one of the pivotal factors that facilitate life; an increasingly evident fact as we enter into an era in which global warming starts to cause significant impacts to life on Earth. Temperature affects various aspects of life, and the most crucial one may be the chemical reactions that constitute the basis of life. As a result, changes in environmental temperature can have significant effects on an organism's well-being [5]; for example, fluctuations of a homeotherm's core temperature, such as what one experiences from a fever, can seriously affect its viability. Even a slight change in temperature (~1ºC) can have significant effects on life, as research on coral reef bleaching has shown [6].

Other facets of life are also affected or challenged by temperature, such as the maintenance of cellular components stability at high temperatures and the retention of water in its liquid state at sub-freezing temperatures [5]. Yet, in the past few decades, scientists have been finding organisms that not only survive, but thrive, in environments of extreme conditions that seem to defy what we had thought of as the temperature tolerance of life [7]. These extreme-loving organisms were aptly, although somewhat anthropocentrically, termed *extremophiles* [8]. The study of extremophiles is likely to reward scientists with not only the knowledge on how organisms adapt to different thermal environments, but also whether life is possible in an environment more hostile to life than Earth. In particular, how life meets

the enigmas created by different temperature ranges and the volatility within those ranges is one area of research that has been significantly rewarding both in terms of fundamental science [7] and biotechnology [9].

Life achieves temperature adaptation through a variety of means [7]. At the high end of temperature, an organism faces issues such as maintaining cellular component stability and the relatively low gas solubility in water, while at low temperatures, the formation and thawing of intracellular ice crystal and the decrease in membrane fluidity may lead to cell damage. Evolution responds to those problems by adjusting membrane composition, altering protein flexibility and stability, manipulating intracellular ionic strength and pH to stabilize DNA, lowering freezing temperature with *antifreeze* proteins, etc. In addition to the challenges described above, one major hurdle remains: how to maintain and control the chemical reactions that enable a cell to grow, perform repairs, and reproduce. Enzymes, as the facilitator of all biochemical reactions, may be the most important target for studying this problem.

The way an enzyme's catalytic activity responds to change in temperature can be described with its catalytic rate constant, k_{cat}:

Equation 1-1: $$k_{cat} = \frac{k_B T}{h} \cdot e^{\frac{-\Delta G^{\ddagger}_{cat}}{RT}}$$

(k_B is Boltzmann's constant; T is the temperature in kelvin; h is Planck's constant; $\Delta G^{\ddagger}_{cat}$ is the activation energy of the catalytic reaction; R is the gas constant)

Equation 1-1 describe the temperature dependence of enzyme-catalyzed

reactions, but enzyme activities that follow its description can only be observed within a relatively narrow and low temperature range [10]. This is because enzymes denature and lose activity at sufficiently high temperatures, thus limiting scientists' ability to observe and characterize temperature-dependent enzyme catalytic activity at high temperatures.

- **Enzyme "Temperature Optima"**

The quest to find "temperature optima" for enzymes has been a long, but unfruitful one. The implications of a *bona fide* temperature optimum reach beyond academic interests, as any user of cold-water laundry detergent can attest. A brief search of the term in PubMed returns thousands of entries, most of them from journals concerned with biotechnological application of enzymes and enzymes isolated from extremophiles, but only a misinformed researcher would consider those numbers valid and intrinsic parameters of enzymes. A product of an unknown mixture of enzyme thermal stability, catalytic rate constant (k_{cat}), and assay duration, the idea of temperature optimum as an intrinsic property of enzymes has long been discredited [1, 10, 11]. As temperature increases, observed enzyme activity also increases as defined by k_{cat} of the enzyme until the point at which thermal denaturation of the enzyme becomes significant, leading to the illusion of a temperature optimum. Since thermal denaturation is dependent on both time and temperature, the apparent temperature optimum changes with assay length, as described by the 'Classical Model' [1].

Representing a long-standing implicit assumption of the temperature dependence of enzyme activity, the Classical Model comprises of only the effects of catalytic rate constant in the form of $\Delta G^{\ddagger}_{cat}$, and irreversible enzyme thermal denaturation in the form of $\Delta G^{\ddagger}_{inact}$, the activation energy of the thermal denaturation process. It is defined by the following equations (assuming that the enzyme remains saturated with substrate[s] throughout the duration of the assay):

Equation 1-2: $V_{max} = k_{cat} \cdot E_0 \cdot e^{-k_{inact} \cdot t}$

V_{max} is the maximal enzyme velocity; E_0 is total enzyme concentration; t is assay duration; and the thermal inactivation rate constant, k_{inact}, is defined by the following equation:

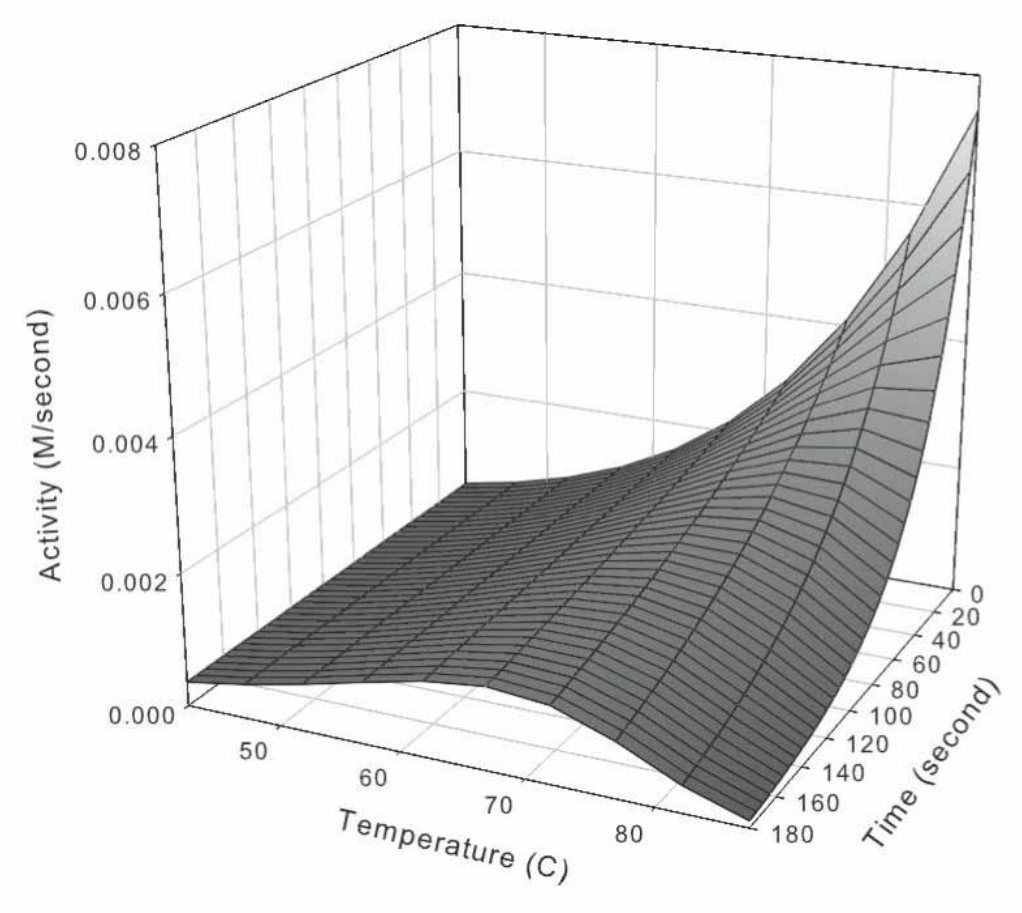
Equation 1-3: $k_{inact} = \frac{k_B T}{h} \cdot e^{\frac{-\Delta G^{\ddagger}_{inact}}{RT}}$

($\Delta G^{\ddagger}_{inact}$ is the activation energy of the thermal denaturation process)

A three dimensional plot of the Classical Model, as defined by Equation 1-1 and Equation 1-3, can be seen in Figure 1-1, in which it is evident that the apparent temperature optimum decreases with increasing assay duration. Note, however, that according to the Classical Model, V_{max} at zero time should increase indefinitely with temperature since there is no denaturation when $t = 0$. This may be the case for inorganic catalysts, but it is not plausible for a biological molecule, whose catalytic activity arises from various intermolecular forces that are significantly

affected by temperature. Moreover, there have been reports of enzymes that appear less active above a certain temperature than can be accounted for by thermal denaturation [12-15], suggesting that a new mechanism is required to explain such phenomena.

Figure 1-1 – Enzyme activity-temperature relationship as described by the Classical Model.



The following parameter values were used to generate this plot: $\Delta G^{\ddagger}_{cat}$ = 55 kJ · mol^{-1}, $\Delta G^{\ddagger}_{inact}$ = 100 kJ · mol^{-1}, and total enzyme concentration = 100 nM.

- **Does Stability Equate to Activity?**

The Classical Model specifies that thermal stability dictates the observable enzyme activity at high temperatures. In other words, when two enzymes with a similar intrinsic activity level (k_{cat}) are assayed, the more stable one is likely to give an impression of elevated activity at high temperatures due to the lack of activity loss caused by thermal denaturation. This may in some cases lead to a false perception that a greater stability is connected to greater activity, an idea that has been shown repeatedly to be incorrect, particularly by studies of psychrophilic enzymes [14, 16-18]. For example, a comprehensive survey of twenty-two lactate dehydrogenases from porcelain crabs suggested that structural stability and functional properties of proteins can evolve independently [18]; chimeric mutants of citrate synthases from a hyperthermophile and a moderate thermophile revealed that stability and activity depend on different regions of the protein and are thus individually manipulable [14].

To complete our understanding of how enzyme activity is affected by temperature, it is necessary to investigate both protein thermal stability and enzyme catalytic properties, since these factors collectively determine how an enzyme behaves *in vivo*. However, a parameter that results from a mixture of both factors, such as the enzyme "temperature optimum", is clearly inadequate for any comparative studies.

b. The Foundation of the Equilibrium Model

Prompted by anomalies that cannot be explained by the Classical Model [12-14], a new model describing enzyme-temperature relationship, named the 'Equilibrium Model', was formulated and reported by Daniel, Danson and Eisenthal [1]. The Equilibrium Model offers a more complete description of enzyme-temperature relationship through the inclusion of a third, intrinsic, temperature-dependent parameter of enzymes and has been experimentally verified with several enzymes [2, 19, 20].

- **The Active-Inactive Equilibrium**

The main characteristic of the Equilibrium Model is the inclusion of an inactive form of enzyme in addition to the two forms (active and thermally-denatured) described by the Classical Model. In the Equilibrium Model, the active form of enzyme (E_{act}) is in equilibrium with the inactive form (E_{inact}), and only the inactive form can undergo the irreversible thermal inactivation to the thermally denatured state (X):

Equation 1-4: $$E_{act} \underset{}{\overset{K_{eq}}{\rightleftharpoons}} E_{inact} \xrightarrow{k_{inact}} X$$

The equilibrium constant K_{eq} describes the equilibrium between E_{act} and E_{inact}, which plays a role in determining enzyme activity because only E_{act} is capable of catalysis. As a result, V_{max} is influenced by the position of the E_{act}–E_{inact} equilibrium, as described by the following equations:

Equation 1-5: $$V_{max} = k_{cat} \cdot [E_{act}]$$

Equation 1-6: $$[\mathrm{E}_{\mathrm{act}}] = \frac{E_0 - [X]}{1 + K_{\mathrm{eq}}}$$

Intrinsic to enzymes, K_{eq} joins k_{cat} and k_{inact} as the third temperature-dependent property of enzymes; its variation with temperature is defined by this equation:

Equation 1-7: $$\ln(K_{\mathrm{eq}}) = \frac{\Delta H_{\mathrm{eq}}}{\mathrm{R}}\left(\frac{1}{T_{\mathrm{eq}}} - \frac{1}{T}\right)$$

(ΔH_{eq} is the enthalpic change associated with the conversion of an enzyme from $\mathrm{E}_{\mathrm{act}}$ to $\mathrm{E}_{\mathrm{inact}}$; T_{eq} is the temperature at the midpoint of the transition between the two forms)

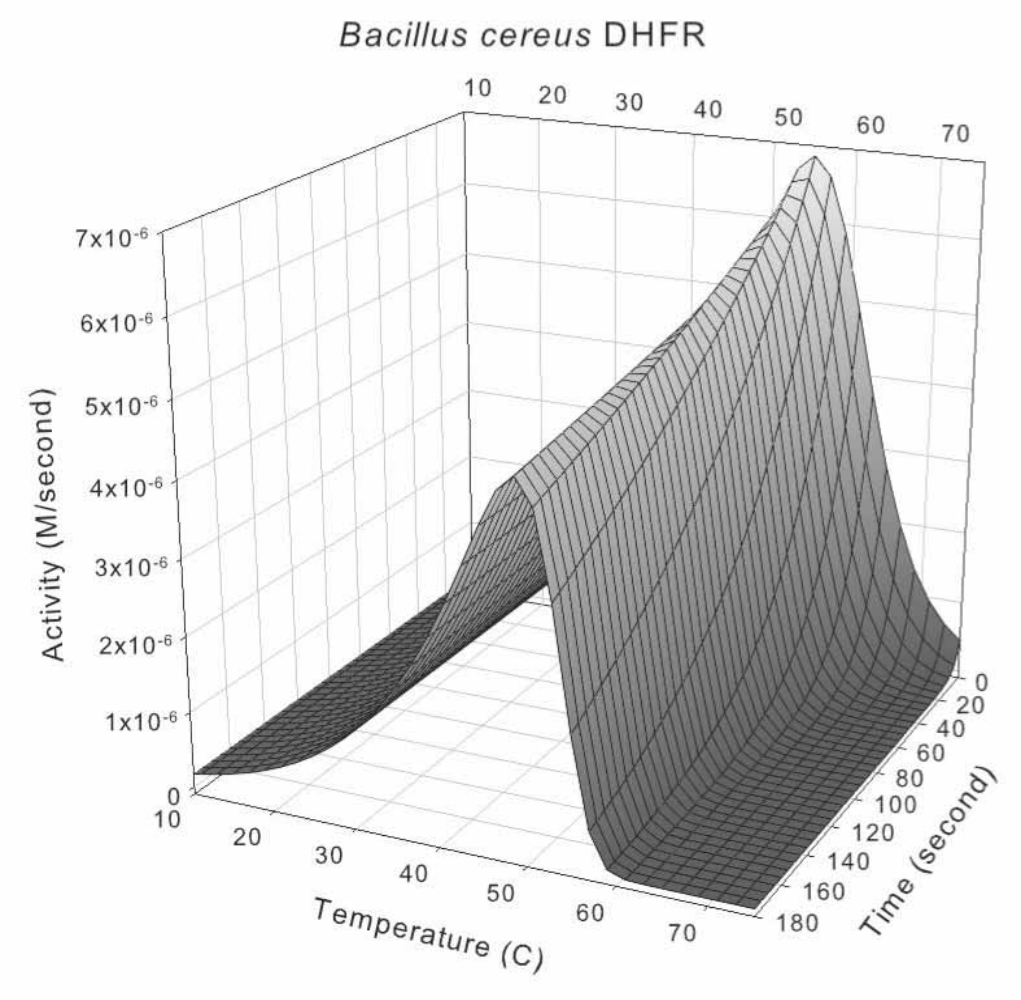
Since T_{eq} arises from K_{eq}, it is independent of enzyme thermostability. In other words, T_{eq} is an independent and intrinsic property of enzymes, albeit one that is not directly measurable with experiments, due to its dependence on ΔH_{eq}, which is also not directly measurable. Therefore, one has to consolidate the equations listed above into one that depends on experimentally determinable parameters (i.e., V_{max}, t, and T), then fit the said equation to experimentally determined enzyme reaction data acquired over an appropriate temperature range to estimate T_{eq} and ΔH_{eq}. The overall dependence of enzyme activity on temperature with time is defined by the following equation [2], and an example of enzyme-temperature relationship as described by the Equilibrium Model can be seen in Figure 1-2.

Equation 1-8:

$$V_{\max} = \frac{k_B T \cdot e^{\left(-\frac{\Delta G^{\ddagger}_{cat}}{RT}\right)} \cdot E_0 \cdot e^{\left(-\frac{k_B T \cdot e^{\left(-\frac{\Delta G^{\ddagger}_{inact}}{RT}\right)} \cdot e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{T_{eq}}-\frac{1}{T}\right)}{R}\right)} \cdot t}{h\left(1+e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{T_{eq}}-\frac{1}{T}\right)}{R}\right)}\right)}\right)}}{h\left(1+e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{T_{eq}}-\frac{1}{T}\right)}{R}\right)}\right)}$$

For obvious reasons, had the Equilibrium Model been developed fifty years ago it would have been extremely difficult to verify using the methodology described above and the technology available to the general scientific community at the time. Fortunately, modern technology is to scientists' rescue—through the availability of cheap computing power, one can apply a data set consisting of thousands of data points to the above equation and derive the parameters of interest (i.e., $\Delta G^{\ddagger}_{\text{cat}}$, $\Delta G^{\ddagger}_{\text{inact}}$, ΔH_{eq}, and T_{eq}) using a reiterative, nonlinear fitting process which yields a minimal sum of squared residuals. This process is described in detail in Chapter Two.

Figure 1-2 – Activity-temperature relationship of *Bacillus cereus* dihydrofolate reductase based on the Equilibrium Model.



The following parameter values were used to generate this plot: $\Delta G^{\ddagger}_{cat}$ = 66 kJ·mol^{-1}, $\Delta G^{\ddagger}_{inact}$ = 90 kJ·mol^{-1}, ΔH_{eq} = 261 kJ·mol^{-1}, T_{eq} = 331.5 K and total enzyme concentration = 39.5 nM.

---

c. The Structural Basis of the E_{act}–E_{inact} Equilibrium

Irreversible enzyme thermal denaturation and the inactivation process described by the Equilibrium Model may be distinct in both timescale and the

extent of structural changes. Although there is currently no direct evidence related to the molecular basis of the equilibrium between the active and the inactive states, it has been shown that the equilibration process is a rapid one, possibly at least two orders of magnitude faster than the irreversible denaturation process [2, 19]. Collecting data for use with the Equilibrium Model requires a rapid addition of enzyme held at 0ºC to a substrate-buffer mixture pre-equilibrated at the desired temperature, and data collection begins immediately after the enzyme was added and mixed. Yet, the effect of the E_{act}–E_{inact} equilibrium is evident from within the first few seconds of measurement, indicating a sub-second timescale [2]. On the other hand, total unfolding occurs on timescales ranging up to minutes, as demonstrated in Figure 1-2 and by a survey of the Protein Folding Database [21].

Meanwhile, some information is available for the effects of substrate, cofactor, and additives on the E_{act}–E_{inact} equilibrium [19] and provides additional insights into the nature of the phenomenon. The physical structure of substrates is likely to affect the flexibility of the enzyme active site, and comparative studies using different substrates of ß-lactamase revealed that structurally distinct substrates led to markedly different T_{eq} and ΔH_{eq} values; enzyme cofactors were found to have a similar influence on the Equilibrium Model parameters. Stabilizing and destabilizing agents were found to have little effect on $\Delta G^{\ddagger}_{cat}$, ΔH_{eq}, and T_{eq}, whereas $\Delta G^{\ddagger}_{inact}$ and consequently enzyme stability was found to be significantly modified by the additives. In summary, compounds that interact directly with the active site tend to result in more pronounced effects on the E_{act}–E_{inact} equilibrium than

those that don't, such as stabilizing or destabilizing agents, suggesting that the equilibrium described by the Equilibrium Model may be a local phenomenon around the active site. The relatively small effects of stabilizing and destabilizing agents on $\Delta G^{\ddagger}_{cat}$, ΔH_{eq}, and T_{eq} support the hypothesis that enzyme activity is effectively detached from stability [22].

Consistent with the theory of an active site-based E_{act}–E_{inact} equilibrium is the evidence that a local unfolding of the active site can precede the irreversible denaturation process [23, 24], potentially leading to a reduction in catalytic rate that cannot be accounted for by irreversible thermal inactivation. This is supported by evidence demonstrating that catalytic residues within an enzyme are optimized for catalytic function rather than stability [25], resulting in an active site that is more flexible than other parts of the enzyme and more susceptible to the effects of temperature. In brief, enzyme conformational change during thermal denaturation occurs in two steps, the first one is the local and potentially reversible unfolding of the active site, the second is an overall denaturation of the protein. While the exact molecular basis of the E_{act}–E_{inact} transition is still unclear, it shares features with the first, local and reversible, unfolding step of protein thermal denaturation.

d. The Implications of Parameters Derived from the Equilibrium Model

T_{eq} is the temperature at which [E_{act}] equals [E_{inact}] and the enzyme exhibits 50% of maximal activity, thus it can be considered as a *thermal equivalent* of K_m. An enzyme's T_{eq} partially determines its optimal working

temperature, an association that hints at potential evolutionary and adaptive significance for T_{eq}. Additionally, since T_{eq} is intrinsic to each individual enzyme, it can be used for comparative purposes, allowing scientists to directly compare thermal characteristics of unrelated enzymes. T_{eq} also adds a new dimension to enzyme engineering, which has traditionally been focused on altering thermal stability; a detailed discussion on the biotechnological implications of the Equilibrium Model has been published [22].

Evidence supporting a connection between $\Delta G^{\ddagger}_{cat}$ and an enzyme's thermal environment has been largely limited to enzymes from psychrophiles [26-28], and thermal stability is known to be affected by factors unrelated to thermal environment (e.g., protease resistance [29], intracellular milieu conditions). Novel parameters derived from the Equilibrium Model may provide additional insights into an enzyme's intrinsic properties and its native thermal environment. A detailed study of the connections between enzymes' thermal environments and their Equilibrium Model parameters is described in Chapter Three.

II. Temperature Adaptation and Eurythermalism

Eurythermalism refers to an organism's ability to tolerate a wide range of temperatures; its antonym is stenothermalism. Temperature variation, be it circadian, seasonal, or climatic, is a matter of fact for most ecosystems.

Although some species react to harsh temperature changes by entering into hibernation or other states of inactivity, most other organisms do not possess this ability and have to remain reasonably active within a range of temperatures. Therefore, the study of eurythermalism is an important part of understanding how life achieves temperature adaptation.

The following is an attempt at providing a brief overview on the history of temperature adaptation studies from an enzymological perspective and how the discovery of ecosystems in extreme environments has added a new dimension to it.

a. The Challenges of Temperature Adaptation

Eurythermalism is a term that only applies to ectotherms, in other words, animals whose body heat comes from and is affected by the surrounding environment. One of the most significant challenges ectothermic organisms faces in terms of temperature adaptation is the enzyme catalytic rate constant [30], which, as indicated in Equation 1-1, is dictated by temperature. For most biological systems, a decrease of 30ºC in body temperature translates to a 3 to 8-fold reduction in chemical reaction rate (assuming a Q_{10}* of 1.5-2). Yet, despite this problem, there's no shortage of evidence of ectotherms adapted to various temperature conditions. Without considering the more exotic examples, the fact that most ectotherms are acclimated to daily temperature fluctuations and season climate changes has been fascinating

* Temperature coefficient, the measure of the rate of change in catalytic activity as a consequence of increasing the temperature by 10ºC.

scientists for decades [31]. However, scientists had very little idea about how all this was achieved at the enzymic level [31] until the 1960s, when Hochachka brought together the two fields of biochemistry and comparative physiology and began researching adaptations for eurythermality in a biochemical context [32, 33].

- **Three Ways to Skin a Cat**

In his 1965 paper on lactate dehydrogenase isozymes in different tissues of gold fish [34], Hochachka proposed three mechanisms through which enzymes achieve adaptation to wide ranges of temperature [30]: (1) adaptive changes in gene expression pattern and (consequently) enzyme concentrations, (2) genetically based differences in the kinetic and structural properties of enzymes (e.g., K_m, k_{cat}, and disulfide bonds), and (3) post-translation modifications and modulations of enzymes (e.g., change in protein covalent bonds, altering pH, and low molecular weight stabilizers). These mechanisms not only provided potential explanations for eurythermality observed in nature, but also proposed that multiple levels of cellular functions are involved in the process, significantly broadening the scope of temperature adaptation studies.

Following the establishment of the three mechanisms, scientists have identified common themes for each of the mechanisms in relatively eurythermal species: (1) the up-regulation of enzyme-coding genes to compensate for lower reaction rates at a lower temperature [35, 36] and high protein turnover rates to compensate for thermal degradation at high

temperatures [37, 38]; (2) having multiple allelomorphic variants of enzymes (isozymes/allozymes) that are kinetically or thermally adapted to a given thermal condition, and through the regulation of gene expression or varied frequencies of genes coding for different allozymes in the genome, resulting in the expression of a particular population of isozymes/allozymes to respond to specific thermal conditions [39-41]; (3) regulation of pH and small compounds that modify protein stability.

The first two schemes may be expensive to the organism, the former in terms of energy and resources, and the latter in terms of maintaining multiple gene copies and gene expression regulation, making the second mechanism impractical at the individual level [40], particularly for prokaryotes, which have relatively compact genomes. Therefore, these two mechanisms are probably better suited for daily or seasonal reversible temperature acclimation of eukaryotes or even tissue-specific adaptation rather than the long-term, organism-wide, temperature adaptation.

In general, the third mechanism is manifested in two distinct processes: the regulation of pH and the presence of protein-stabilizing organic solutes and heat shock proteins (*hsp*). While a temperature-dependent change of intracellular pH can have profound effects on the activity of enzymes [42], it is in general limited to metazoans due to their unique intracellular compartmentalization [43]. The influence of organic solutes, on the other hand, is largely limited to protein structural stability. Heat shock response by *hsp* is a common

theme in many species; most known *hsp*s belong to the chaperone family, whose main function is to assist with the (re)folding of proteins [40, 44]. Heat shock response is largely a short-term tolerance mechanism, and a high level of *hsp* under normal circumstances might actually be detrimental to the cell [45].

- **One More Way to Skin a Cat**

We propose here an additional fourth mechanism for adapting to wide ranges of temperatures exists, which is to have a single eurythermal enzyme that is kinetically and thermally adapted over a wide temperature range; this mechanism is the most energetically favorable and perhaps the most effective mechanism for achieving eurythermalism [40].

While there have been reports of enzymes that work over an exceptionally wide range of temperatures [46-48], scientists have generally relied on k_{cat}/K_m as the main measure of enzyme adaptation in this regard [27]. Not only is k_{cat}/K_m specific to the temperature it was measured at, but both k_{cat} and K_m are affected by temperature. k_{cat}/K_m has also been shown mathematically to be an inadequate parameter for comparing homologous enzymes, let alone for organism-oriented comparative studies [49]. Therefore, a strong need exists for a more generally applicable measure of enzyme eurythermality to facilitate comparative enzyme temperature adaptation investigations.

b. Enzymes from Extremophiles

From the isolation of *Strain 121* from hydrothermal vents at Juan de Fuca Ridge that raised the known upper temperature limit for life to 121ºC [50], to the discovery of the strictly psychrophilic bacterium *Psychromonas ingrahamii*, which is capable of growing at subfreezing temperatures in Arctic sea ice core [51], it almost seems that the only limit to life is the imagination of the people trying to find it. The study of enzymes from some of these *extremophiles*, many of which are extraordinary in their own ways, have richly rewarded scientists/biotechnologists.

- **Psychrophiles**

Since liquid water is the *sine qua non* of life, in order to survive in temperatures significantly below 0ºC, an organism must have certain mechanisms to prevent water from freezing/crystallizing. The only known exception to this is the nematode *Panagrolaimus davidi*, which can withstand freezing of intracellular water [52]. In other organisms, protections against freezing of cellular fluids include elevated unsaturated fatty acids content in membranes, cold-shock proteins, and antifreeze proteins [27, 53]. Ultimately, however, enzymes in psychrophiles must be kinetically adapted to low temperatures in order to sustain an acceptable level of catalytic activity and consequently cellular functions for extended periods [53].

One of the earliest studies of psychrophilic enzymes was Somero's research on the properties of succinate dehydrogenase from the Antarctic

notothenioid fish *Trematomus bernacchii* [54]. It was among the earliest evidence to show that a link existed between cold adaptation and a lower $\Delta G^{\ddagger}_{cat}$, the latter allowing enzyme catalytic activity to remain relatively high at lower temperatures. Since then, various independent studies have found similar results with enzymes isolated from cold-adapted organisms [26, 28, 55], and the reduction of $\Delta G^{\ddagger}_{cat}$ has been considered a major adaptive parameter of cold-adapted enzymes [27].

- **Hyperthermophiles**

Hyperthermophiles are defined as organisms that thrive in unusually hot environments (> 60ºC) and have optimal growth temperatures (OGT) above 80ºC. To acclimatize to their environments, which are often extreme in ways other than temperature, hyperthermophiles often possess a wide array of adaptations, including abilities to cope with high salinity, high acidity, high pressure, and high heavy metal concentration, etc. Moreover, there is one common challenge that all hyperthermophiles face: maintaining stability of biomolecules, particularly enzymes, at high temperatures.

Ever since the first thermophile was isolated [56], it was clear that enhanced thermal stability, even *in vitro*, is a trait of almost all enzymes isolated from thermophilic organisms, a notion supported by a still growing body of experimental data [27, 57-59]. For thermostable enzymes in general, stability is closely linked to reduction of protein molecular movements compared with mesophilic enzymes at the same

temperature [60]. The activity of thermophilic enzymes are adapted to the native temperature of host organisms, and as a result typically exhibit relatively low activity at mesophilic temperatures compared to their mesophilic counterparts.

However, a trade-off between stability at high temperatures and activity at low temperatures is not a universal trait of enzymes isolated from thermophiles, as suggested by reports of enzymes that possess both high overall stability and relatively high activity compared to their thermophilic or even mesophilic counterparts at lower temperatures [46, 47, 61-64]. While these enzymes' relatively high levels of activity at both low and high temperatures have been suggested as a way to *protect* thermolabile substrates from rapid degradation at high temperatures [46, 65], more importantly it implies that stability and activity can be selected separately by evolution, and that enzyme activity, which is widely regarded as connected to the configuration of the active site, cannot be unequivocally inferred from stability. This echoes the growing consensus that stability can be detached from activity in enzymes, as reviewed earlier in this introduction. Therefore, enzyme stability data obtained using non-assay-based techniques are likely to be a poor predictor of the temperature dependence of enzyme activity.

By gaining fundamental understandings for enzyme temperature adaptation, scientists can better comprehend the effects of temperature changes on terrestrial

ecosystems and the organisms that constitute them. Proper tools must be developed for facilitate such studies, and the enzyme thermodynamic parameters derived from the Equilibrium Mode parameters, being intrinsic and quantitative, may be useful. Through investigating the universality and the usefulness of the Equilibrium Model parameters, this thesis describes an attempt at validating the Equilibrium Model as a generally applicable and comprehensive tool for investigating enzyme temperature adaptation.

III. *Alvinella pompejana* and Its Episymbionts

The discovery of the Mid-Atlantic Ridge by Marie Tharp and Bruce Heezen onboard the original R/V *Atlantis* in 1955 was instrumental in validating the theory of plate tectonics. The existence of mid-ocean implied that there was likely to be active, concentrated geothermal activities on the deep seafloor. Evidence for undersea geothermal activities emerged in 1977, when a scientific team onboard the submersible DSRV *Alvin*, lead by Dr. Robert Ballard, made the first discovery of submarine hydrothermal vents along the Galapagos Rift, part of the East Pacific Rise at a depth of over 2,000 meters. The hydrothermal vents are characterized by their high core temperature (often >400ºC), extreme pressure (over 150 bars), and toxic geochemistry, so it was indeed a surprising moment for the scientists when a thriving ecosystem was discovered around the hydrothermal vents, a discovery that would fundamentally change our understanding of life.

a. The Environment: Its Chemistry and Fauna

Deep-sea hydrothermal vents are fissures on the seafloor where geothermally heated water is release into the ocean, and the environments surrounding the vents are distinguished for their extreme chemical complexity. When seawater reacts with volcanic rocks at around 400ºC at the hydrothermal vents, the cations from seawater react with the volcanic rocks and form hydroxyl-bearing alteration minerals while simultaneously releasing hydrogen ions, resulting in a hot, mineral-rich, and acidic solution. From the volcanic rocks, the hot and low-pH seawater releases metals which complex with reduced sulfur (H_2S) [66]; these reduced metal-sulfur complexes, along with the poor availability of oxygen, characterize the extreme chemical nature of the hydrothermal vents environment. It is this unique chemical signature that provides the energy to the deep-sea hydrothermal vents to support ecosystems much larger (in terms of biomass) than anywhere else in the deep ocean [66].

The chemical complexity associated with the hydrothermal vent also plays a role in shaping the surrounding environment. *Chimneys* are the result of deposits of crystallized minerals contained in the vent fluid, and while the majority of the superheated vent fluid is channeled through chimneys into the open ocean, a significant volume of vent fluid passes through the chimney structures into the surrounding environment, resulting in cooler (below 100ºC) *diffuse flow regions* around the chimneys [67, 68].

At the hydrothermal vents, common deep-sea phyla such as *Porifera* (sponges), *Ctenophora* (comb jellies), *Echinodermata* (echinoderms) are

absent. It has now been determined that over 95% of the vent fauna are endemic [68]. Organisms found in vent systems in different oceans are also more closely related to each other than to those of nearby habitats [68]. Some of the notable members of this diverse ecosystem in the Pacific Ocean are *Riftia pachyptila*, a type of tubeworm inhabiting moderate temperature areas somewhat away from the vents and typically forms the major biomass at the vents; *Bathymodiolus thermophilus*, the vent mussel; *Calyptogena magnifica*, the vent clam; *Alvinella pompejana*, a polychaetous annelid colonizing the hydrothermal vent chimneys; and several vent crabs and shrimps.

In terms of microbial life at deep-sea hydrothermal vents, four distinct but sometimes overlapping communities are currently known: 1. free-living microorganisms associated with fluids discharged from the vents; 2. free-living microorganisms in microbial mats on exposed surfaces around the vents; 3. bacteria in an endo- or ectosymbiotic relationship with metazoans in the vent system; 4. microbes associated with the inside of hydrothermal vent plumes [69]. The majority of these microbes are mesophilic bacteria, and although a number of thermophilic archaea have been found, far fewer thermophilic bacteria have been isolated [69]. Symbiotic bacteria in the vent system fall predominantly into epsilon and gamma subdivisions of class *Proteobacteria*, as do the majority of free-living bacteria in the ecosystem [69, 70]. A considerable number of thermophilic and hyperthermophilic archaea have been found associated with hydrothermal fluid as well, indicating the existence of a high temperature habitat under the seafloor.

Hydrogen sulfide (H_2S) is highly toxic to aerobic life forms, but for some microbes in the hydrothermal vent environment found free-living or in symbiotic relationship, it serves as the primary energy source for chemosynthesis, the primary production method in the ecosystem. During this process, the symbiotic chemoautotrophic bacteria are thought to detoxify the sulfide for their host metazoan in the vent system [71], although some detoxification mechanisms (e.g., sulfide-binding hemoglobin, oxidation of sulfide to thiosulfate) are often found in the host metazoan as well [66].

The hydrothermal fluid, being highly reduced, is typically anoxic but enriched in CO_2 and free sulfide (H_2S/HS^-). Aerobes in the system counter the problem of oxygen availability through a variety of means. For example, the tubeworm *R. pachyptila* has a respiratory plume which extends into the oxygen-rich bottom water [66], while *A. pompejana* adapts to highly variable oxygen availability through an internal gas transfer system which consists of one vascular hemoglobin and one coelomic hemoglobin [72, 73].

The chemical conditions in different areas of the hydrothermal vent system are strongly correlated with the distribution of the fauna [74]. For example, the availability of free sulfide and high pH (> 5.7) are essential for the survival of *R. pachyptila,* vent mussels, and vent clams, while *A. pompejana* is absent from such environments and generally colonizes high iron diffuse flow areas where sulfide is almost totally complexed and therefore biologically inert [75].

b. The Pompeii Worm

The polychaetous annelid *Alvinella pompejana* belongs to a group of tube forming worms that colonize active chimneys at hydrothermal vents of the East Pacific Rise at depths of around 2500 meters [76]. The name *Pompeii worm* came from that fact that it lives under a constant rain of mineral particles [77]. As its colonies are located directly on the sides of hydrothermal vent chimneys and appear to be flushed by high temperature fluids, it is generally considered the most thermotolerant vent invertebrate known to science [76, 78]. The harsh environment inhabited by *A. pompejana* is characterized by an extraordinary temperature gradient (2ºC to over 80ºC), flushing toxic chemical compounds (free sulfides and heavy metals), and is both temporally and spatially volatile.

The Pompeii worm feeds on free-living chemoautotrophic bacteria found on surfaces around the colony, which it shares with several other species of annelids, including *Alvinella caudata*, *Paralvinella grasslei*, and *Hesiolyra bergi* [77]. It has also been observed making regular excursions of up to 1 m to feed on free-living bacteria growing on the outer surfaces of the colony, where it is more likely to fall prey to larger invertebrates, amphipods and crustaceans [67]. Therefore, its native environment may provide the worm protection against predators in addition to other benefits. *A. pompejana* is a gonochoric, sexually dimorphic species that exercises pseudocopulation, a behavior that potentially prevents dispersion of the gametes. *A. pompejana* reaches new colonies through larval dispersal in the water column; as a result, the *founder effect* can be observed in *A. pompejana* colonies, in which the

number of worms colonizing a new site is far below the average size of a local population [77].

A. pompejana has been established as one of the most eurythermal and thermotolerant metazoan known to science [67]. It may be regularly exposed to environmental temperatures up to 80ºC, and a thermal gradient up to 60ºC or more spans the length of its body [67, 75]. Although there have been suggestions that *A. pompejana* may have an internal thermal limit of 50ºC or less [77, 78], the only experimental evidence directly in support of this speculation was derived from *Paralvinella sulfincola*, which is from shallower water (1500–1800 m), and the experiments were performed in a uniformly heated pressurized vessel with the worms' mucus tubes removed, which may provide thermoprotective effects [79]. On the other hand, *A. pompejana*'s rDNA has one of the highest melting temperatures found in a marine invertebrate (88ºC) [77, 80], suggesting that it is adapted to an atypically high temperature vent environment.

Due to the difficulties in performing any precise *in situ* experiments, much of the information obtained on the worm to date is based on *in vitro* investigations. Although it is generally agreed that *in vivo* experiments are needed to obtain indisputable evidence, to date there hasn't been any report of *A. pompejana* being successfully maintained alive outside its native environment [70]. Also, even if this was accomplished, heat tolerance experiment by basking the whole worm in heated water may be an inadequate method since *in situ* only part of the worm may be exposed to high temperature fluid, and then only for limited periods of time.

c. The Episymbionts

Through the use of 16S rDNA sequence analysis and other molecular techniques, it has been established that the bacterial assemblage on the dorsal side of *A. pompejana* is predominantly composed of filamentous bacteria belonging to the epsilon subdivision of class *Proteobacteria* (ε-*Proteobacteria*). The community is characterized by both its high genetic diversity and the prominence of a few phylotypes [81, 82]; it is also metabolically diverse, containing sulfide oxidizers, sulfate reducers, nitrifiers, nitrate respirers, denitrifiers, and nitrogen fixers [81, 83].

Members of the ε-*Proteobacteria* subdivision in general share some common properties, such as adaptation to low oxygen environment, sulfide oxidation or sulfate reduction capability, and close association with eukaryotes [82]. ε-*Proteobacteria* have also been found to be associated with vent invertebrates such as the vent shrimp *Rimicaris exoculata*, and in a microbial mat at Pele's Vent on Loihi Seamount, Hawaii, suggesting that they play an important role in the ecology of the hydrothermal vent systems [81, 84, 85]. Furthermore, bacterial symbionts, both ε-*Proteobacteria* and γ-*Proteobacteria*, are often found intimately linked with their marine invertebrate hosts. For example, *Riftia pachyptila* endosymbionts (γ-*Proteobacteria*) oxidize hydrogen sulfide (H_2S), thereby producing the energy required to fix carbon from CO_2 and provide sugars and amino acids (predominantly as glutamate) that nourish the tubeworms [86, 87].

It has been suggested that the epibionts may be involved in the nutrition and/or detoxification of the host, but direct *in vivo* evidence is still missing,

since the dominant phylotypes have so far eluded all isolation attempts [71, 77, 88]. Using molecular biological techniques, however, it has been shown that members of *Alvinella* episymbiont community, as well as those of free-living microbial mat, express genes that facilitate the reverse tricarboxylic acid (rTCA) cycle, allowing them to contribute carbon fixed from CO_2 to other organisms in the vent community [89]. It is worth noting that the rTCA cycle has been hypothesized to be an ancient metabolic cycle, a fact potentially relevant to the evolutionary history of the vent ecosystem [89]. All ε-*Proteobacteria* isolated to date are capable of sulfate reduction or sulfide oxidation, or both [75, 83]. This fact, coupled with the abundance of electron acceptors (e.g., nitrate) in the environment, suggest that sulfate respiration is one, if not the most important respiration mechanism. This hypothesis is supported by the identification of a variety of dissimilatory bisulfite reductases, an enzyme crucial to sulfate respiration [90], in the episymbiont community.

The episymbionts have been implicated in sulfide detoxification for the host. Various heterotrophic, non-filamentous bacteria have been successfully cultured from *A. pompejana*'s dorsal integument (all non-dominant phylotypes), and many of them display resistance to heavy metals [91]. The role of the episymbionts in sulfide detoxification is less clear since it has been shown that at the vents, high temperature enables Fe^{2+} to interact with free sulfide to form FeS, which lowers pH and detoxifies sulfide by making it less biologically available [74, 75]. Overall, however, the vent-dwelling ε-*Proteobacteria* communities, both symbiotic and

free-living are likely to have a significant role in shaping the chemical conditions of the vent environment they inhabit [75].

The symbiotic relationship between *A. pompejana* and its epibionts may be unilaterally obligatory for the host [77]. Even the smallest worms have the dorsal bacterial filaments, and the filaments have been found to originate from within the intersegmentary spaces of the worm [92]. Yet *in situ* hybridization using episymbiont-specific DNA probes showed that one of the two dominant phylotypes of the epibionts could be found on surfaces around the colony, while another phylotype seemed to be exclusively associated with *Alvinella* spp. [81]. It should be noted that the dominant filamentous phylotypes have been determined to be non-autotrophic under experimental conditions and are at least facultative heterotrophs or mixotrophs [77].

It has been proposed that the fleece of epibiotic bacteria on *A. pompejana*'s dorsal surface may have important roles in shielding the worm from the extreme temperature fluctuations [67, 75, 77], but to date this hypothesis remains untested. While the host *A. pompejana* is capable of behavioral adaptations, the ε-*Proteobacteria* episymbionts, which face the same thermal challenges as their host, must be intrinsically adapted to high temperatures spikes and a large temperature gradient.

ε-*Proteobacteria* have also been shown to be the predominant early colonizers of a potential *A. pompejana* colony [70]. They possess the ability to colonize new environments with highly variable physicochemical conditions, and may be involved in altering the environment to allow *A. pompejana* settlement. It has also been observed that the *A. pompejana*

worms appear to feed on the epibionts of other worms (personally communications, S. Craig Cary), suggesting that apart from the symbiotic roles, the ε-*Proteobacteria* episymbionts may be a food source to *A. pompejana* also.

d. Metagenomics

Understanding the metabolism and physiology of the *Alvinella pompejana* ε-*Proteobacteria* episymbionts is crucial to further elucidating how it adapts to its harsh environment, and will likely provide insights to extremophily in general. Given the lack of success in isolating or separating the dominant phylotypes of those bacteria, it is fortunate that recent genetic technological advances now grant us the ability to investigate the community *en masse* using a metagenomic approach. Metagenomics is the application of modern genomics techniques, particularly bioinformatic analyses, to the study of communities of microbial organisms taken directly from their natural environments, bypassing the need for isolation and lab cultivation of individual species [93]. The use of metagenomic has in many cases provided useful and detailed information about microbial communities that would otherwise be nearly impossible to obtain [94-98].

The National Science Foundation Biocomplexity project [99], of which the studies presented in this thesis are a part, is one such attempt. A more detailed introduction to metagenomics and an overview of the Biocomplexity project can be found in Chapter Four.

The Pompeii worm is one of the most conspicuous and extreme organisms known to science, and its associated bacterial episymbionts are likely to play significant roles in how the worm adapts to its extreme environment. Although metagenomics is likely to yield valuable data such as genes related to metal detoxification, biosynthetic capabilities, etc, direct information on temperature adaptation needs to be gather using other methods, particularly thermodynamic characterization of enzymes originated from the episymbionts. Chapter Five is a description of the application of the Equilibrium Model to enzymes derived from the *A. pompejana* ε-*Proteobacteria* episymbiont metagenome, and how it may further our understandings of temperature adaptation in this unique ecosystem.

Chapter Two

– Methods for Determining the Equilibrium Model Parameters

An assortment of issues and criteria must be considered when employing the Equilibrium Model to describe enzyme behavior; some of them concern data collection and are crucial for obtaining valid temperature profiles of enzymes, while others are related to the application of experimental data to the Equilibrium Model formulae. Although far from being comprehensive, this chapter is a detailed and practical account of issues and processes critical to rigorous and successful applications of the Equilibrium Model, and can be considered an extension to previous publications [2, 20].

My Contributions

Materials which I contributed to wholly in this chapter include the study on the effects of K_m correction on final parameters and the verification and application of an initial reaction rates estimation method.

An expert is a person who has made all the mistakes that can be made in a very narrow field.

– Niels Bohr (1885-1962)

I. Key Findings Prior to This Work

Much of this chapter is intended to extend and add to existing publications, particularly [20]. Below is a summary of the most important findings and requirements related to applying the Equilibrium Model from reference [20].

- For spectrophotometric assays, quartz cuvettes must be used to avoid background absorbance by the cuvette, and to achieve quick temperature equilibration. Reaction vessel must also be capped at higher temperatures to avoid error due to evaporation [20].
- Buffer pH must be consistent across the assay temperature range [20].
- Enzyme velocity must be maintained at V_{max}, or the collected assay data must be adjusted accordingly and adequately. In particular, substrate concentration must be maintained at 10 times K_m. [20].
- Errors in enzyme concentration have little effect upon Equilibrium Model parameter determination, except in the case of $\Delta G^{\ddagger}_{cat}$, which is directly influenced by enzyme concentration [20].
- Equilibrium Model parameters can be accurately determined using a few as ten data points for each assay progress curve (e.g., sampling every 150 seconds in a 25-minute assay) [20].
- At least eight temperature points, including two points above T_{opt} (the apparent zero-time temperature optimum) and showing a clear downward trend, are required to yield reasonable estimates of Equilibrium Model parameters [20].

All enzyme data described in this thesis were collected under conditions satisfying the above criteria.

II. Obtaining Enzyme Temperature Profiles

In order to describe how an enzyme's activity responds to temperature, one needs to apply experimental data of enzyme activity (i.e., its temperature profile) to formulae derived from the Equilibrium Model to calculate a set of parameters that best describes the behavior of the enzyme according to the Equilibrium Model; in other words, fitting the experimental data to the Equilibrium Model. An enzyme temperature profile consists of data from a number of robust time-course enzyme assays performed in replicates at designated temperature points.

a. Performing Valid and Accurate Enzyme Assays

All enzyme assays performed for this study were continuous spectrophotometric assays measured using a Thermospectronic™ Helios γ-spectrophotometer equipped with a Thermospectronic™ single cell peltier-effect cuvette holder. Absorbance data were collected using Vision™ (version 1.25, Thermo Spectronic Inc.) on a Windows PC connected to the Helios. This setup allows accurate data collection with a precision of 0.001 absorbance units and up to eight data points per second. The temperature control unit is accurate to 0.1ºC and can provide temperature control from 2ºC to 96ºC (in conjunction with a heated and refrigerated water bath). The spectrophotometer also accepts cuvettes of various light path lengths (0.2 mm to 10 cm), allowing greater flexibility in substrate and enzyme concentrations.

As the famous saying goes, garbage in, garbage out. The result

generated from applying the Equilibrium Model is only as good as the temperature profile provided in the first place. However, the quality of the input data isn't always apparent from the parameters that result from it, and precautions must be taken to ensure the validity of the temperature profile [2, 19, 20]. These include the following:

- the enzyme being investigated must have an assay that allows the collection of comparable, quantitative, and progressive data across the intended range of temperatures.
- V_{max} must be maintained throughout the assays (see below for details).
- precise temperature control [20].
- a consistent and optimized enzyme concentration in assays.
- an adequate temperature range and a sufficient number of data points [20].
- the absence of enzyme/assay issues that lead to inherent bias or cause non-linearity in assays, for example;
    - substrate depletion
    - spectrophotometer detection limit
    - enzyme inhibition by reaction substrate or product
    - non-enzymic activities [10] such as spontaneous reverse reaction or thermal degradation of substrate or product
    - temperature-related changes in buffer pH, particularly buffers with high $\Delta pK_a / \Delta t$ values [10]

For details on identifying and rectifying these issues, see Appendix I.

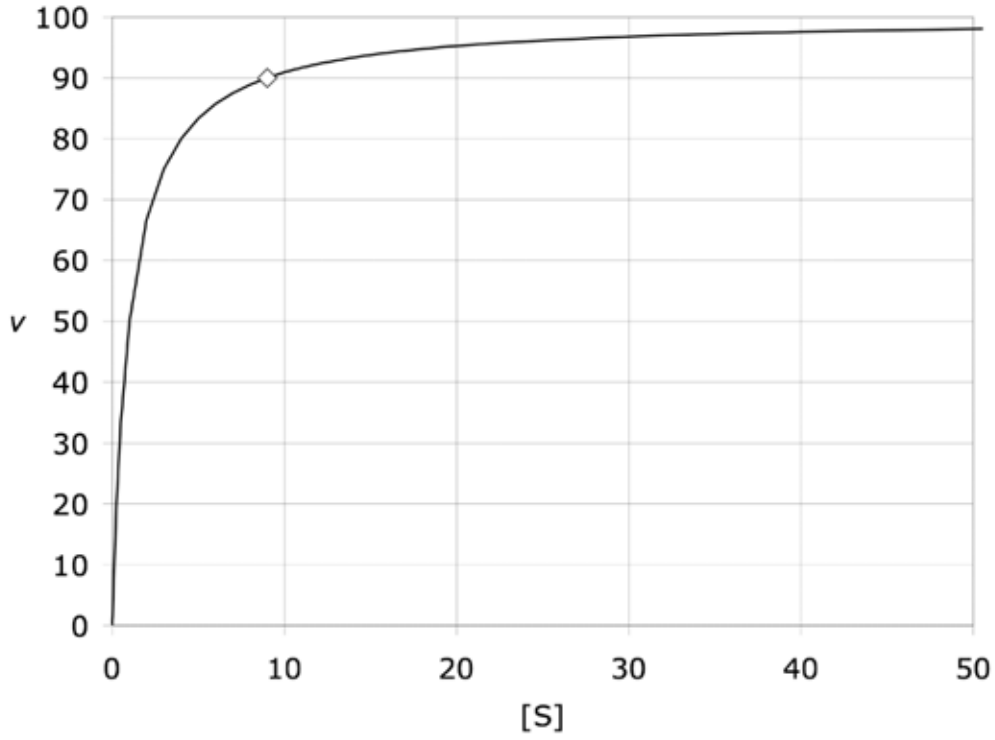
- **Maintaining V_{max}.**

The Equilibrium Model assumes that the enzyme operates at V_{max} at all times, and it is crucial that this assumption be taken into consideration when designing enzyme assays [20]. The theoretical maximal reaction rate (V_{max}) of an enzyme is achieved when it is saturated with all substrates and free of any inhibition, and the relationship between V_{max} and substrate concentration ([S]) can be described with this equation:

Equation 2-1: $$v = \frac{V_{max} \cdot [S]}{K_m + [S]}$$ (v is the actual reaction rate)

Since it is mathematically impossible to have $v = V_{max}$, an approximation of $v > 0.9 \times V_{max}$ is accepted as a practical compromise, and this can be achieved when $[S] > 9 \times K_m$ (see Figure 2-1). Therefore, substrate concentrations of ten times K_m at all assay temperatures are generally considered the minimum for obtaining temperature profiles. Should this be difficult due to various reasons (e.g., low substrate solubility, substrate inhibition), compensations can be made at the data processing step, although K_m determinations must first be performed at all relevant temperatures (see Appendix I for more details). It must be noted that the relationship between v and [S] is not a linear one (see Figure 2-1), and such compensation must be made using Equation 2-1. However, this procedure potentially introduces additional errors and should only be used if all other methods for increasing [S] have been exhausted.

Figure 2-1 – A plot of *v* versus [*S*] demonstrating the approximation of V_{max} relative to K_m.



Reaction rate *v* is expressed as percentage of V_{max}, and substrate concentration [*S*] is expressed as the number of times K_m. Note that while *v* approaches V_{max}, it can never be equal to V_{max}. When $[S] > 10 \times K_m$, enzyme reaction rate is an accepted approximation of V_{max}.

- **The Effect of K_m Adjustments.**

For temperature profiles that contain reactions performed at sub-optimal substrate concentrations (i.e., less than ten times K_m), adjustments are made to the raw data to compensate for deviations from V_{max} before they were imported into Scientist™. To investigate the

potential effects of sub-optimal substrate concentrations, several temperature profile were fitted in both their K_m-adjusted and unadjusted forms. The results can be seen in Table 2-1. With the exception of one enzyme (i.e., *B. taurus* γ-glutamyl transferase), the adjustments did not have significant effects on the final parameters, even when substantial levels of adjustment were applied (e.g., *P. dulcis* ß-glucosidase and *G. stearothermophilus* α-glucosidase). This may be due to the fact that the Equilibrium Model assumes enzyme activity to be near V_{max}, and since the final parameters describes a set of behavior that conforms to the Equilibrium Model, the effects of sub-optimal substrate concentrations were by default reduced or removed.

Enzyme K_m values determined at relatively high temperatures are often associated with significant errors due to rapid enzyme denaturation at those temperatures. The relatively insignificant effects of K_m adjustments on final parameters suggest that errors in K_m estimations do not lead to pronounced errors in the final parameters.

Table 2-1 – The effects of sub-saturating substrate concentration on applying the Equilibrium Model: thermodynamics parameters derived from K_m-adjusted and unadjusted experimental data.

Enzyme	PF_AAA		PD_ßGLU		GS_αGLU		SS_FUM		BT_ γGTP	
Adjustment Applied	16%		56%		66%		21%		26%	
	A	U	A	U	A	U	A	U	A	U
$\Delta G^{\ddagger}_{inact}$ ($kJ \cdot mol^{-1}$)	92	92	95	95	96	96	92	93	98	98
$\Delta G^{\ddagger}_{cat}$ ($kJ \cdot mol^{-1}$)	74	74	63	64	64	64	60	60	63	62
ΔH_{eq} ($kJ \cdot mol^{-1}$)	115	122	100	111	225	219	378	349	109	89
T_{eq} (ºC)	36	36	56	57	68	66	59	58	52	43

PF_AAA, *P. fluorescens* aryl acylamidase; PD_ßGLU, *P. dulcis* ß-glucosidase; GS_αGLU, *G. stearothermophilus* α-glucosidase; SS_FUM, *S. scrofa* fumarase; BT_γGTP, *B. taurus* γ-glutamyl transferase; **"A"**, adjusted; **"U"**, unadjusted. Adjustment applied refers to the highest amount applied to one or more temperature points to the data set. With the except of one data set, K_m adjustment does not resulting in significant changes in estimated Equilibrium Model parameters.

---

b. Data Processing

To compile the assay data into a temperature profile ready for Equilibrium Model application, absorbance data from Vision™ are first exported as text files and imported into Excel (version 11 for Windows/Mac, Microsoft Corp.) where they are converted to progress curves of product concentration (M) versus time (second) using this formula:

Equation 2-2: $$[P] = \frac{\left|\left(\text{change in absorbance at } t\right) - \left(\text{background at } t\right)\right|}{\varepsilon \cdot a}$$

where [*P*] is product concentration, ε is the molar extinction coefficient ($\text{absorbance unit}/\text{M} \cdot \text{cm}$) of the substrate/product being monitored, *a* is the light path length of the cuvette (cm), and *t* is the length of time (second) from the beginning of the assay for the data point being processed. At this point, an assay is described by a progress curve consisting of three columns of data: time (second), product concentration (M), and temperature (K). Since the measurements are performed in triplicate, there are three progress curves for each temperature point; if necessary, the progress curves are adjusted for deviation from V_{max} with adjustment factors calculated using Equation 2-1. The temperature profile is now ready for fitting to the Equilibrium Model.

Before proceeding with the fitting process, the enzyme molar concentration in the assay (E_0) needs to be calculated using this formula:

Equation 2-3:

$$E_0 = \frac{\left[\text{enzyme stock}\right] \times \left(\text{volume of enzyme in assay}\right)}{\left(\text{dilution factor for enzyme stock}\right) \times \left(\text{enzyme Mw}\right) \times \left(\text{assay volume}\right)}$$

where enzyme stock concentration is in g/l, volume of enzyme in assay and assay volume are in liter, enzyme molecular weight is in g/mol, and E_0 is in M. Note that it has been show that moderate overestimations (less than ten-fold) of protein concentration determination do not significantly affect final Equilibrium Model parameters (ΔH_{eq} and T_{eq}) [20], suggesting that a low level of contaminating enzyme activity does not compromise the robustness of the parameters.

c. Initial Rates Estimation

In the topics discussed so far, reaction initial rates (v_0) play a very significant role. Both T_{opt} estimation and K_m determination rely on v_0 as the input data; initial rates are also the input data for the zero-time Equilibrium Model and can be used for validation of Equilibrium Model fitting results.

- **Initial Rates by Visual Estimation**

The most convenient method for obtaining initial rates is to do so visually from the reaction curves. One can perform this operation by importing the reaction curves into a computer software capable of basic measurement of curves (e.g., Vision^{TM}), performing a linear search or best fit on a portion of the curve that appears to be suitable, then inspecting the resulting fit to decide whether it is likely to reflect the initial rate of the reaction. For long-period assays such as those used during enzyme initial characterization, it is best to derive v_0 from a linear section close to the beginning of the assay (using Vision's *linear search* function, for example) because such assays tend to maintain a stable initial rate for some length of time, and occasionally mixing artifacts can be seen near the very beginning of those assays, making readings from that area less reliable. This should not be treated as a mechanical process, however, and one must take into consideration the overall shape of the reaction curve when evaluating the obtained v_0. If reliable and consistent initial rates are needed, a mathematically based method described below should be considered. On the other hand, visual

estimation usually works quite well for high activity level, short (e.g., 30 seconds) assays with high enzyme concentration such as those of K_m determination.

- **Initial Rates from Regressive Fitting of Progress Curves**

This approach utilizes a logarithmic regression method developed for estimating initial rates from nonlinear progress curves of enzyme reactions [100]. The method is based on an approximation of the Michaelis-Menton equation at near zero-time developed by Lu and Fei [100]:

Equation 2-4: $$y = y_0 + b\ln\left(1 + \frac{t}{t_0}\right)$$

where y is the amount of product (i.e., the amount of absorbance change), y_0 represent the scale of background signal, $b > 0$ is the shape parameter for the progress curve, t is time, and $t_0 > 0$ is the scale of the progress curve. The mathematical basis of the equation is not within this scope of this discussion, but the goal is to recursively fit the reaction curves, which are first converted from raw absorbance data to progress curves of product concentration, to Equation 2-4 to obtain the best fit from which the initial rate is calculated. Any mathematical software suite capable of logarithmic regression can be used to perform the analysis (e.g., MATLAB®, Mathematica®, Prism®).

A set of initial values for the fitting process has been determined empirically: t_0 = the *minimum* of y, $b = 1$, and $t_0 = 1000$. As all data are

in triplicates, R-squared values (R^2, the goodness-of-fit) should be calculated from the mean value of y from triplicates at the same temperature. The value for t depends on assay duration and the type of assay anomaly encountered, but since 30 seconds is the typical length of an initial rate assay, t = 15 is a good start. Perform a fit using these initial conditions and examine the R-squared value of the fit. Adjust t to maximize R^2, and at the same time compare the fitted curves to the actual progress curves to see if portions of the curves affected by product inhibition or equilibrium shift are still included in the fit.

Once a satisfactory fit is obtained, calculate initial rates with this formula:

Equation 2-5: $$v_0 = \frac{b}{t_0 \cdot \varepsilon}$$

After initial rates were obtained, they should be corrected for background activity, if there is any. The initial rates can now be used to evaluate potential T_{opt}, determine K_m, or be fitted to the zero-time Equilibrium Model. Although v_0 is not a mandatory element of a normal temperature profile, it can be useful in several ways, so it is a good idea to visually estimate v_0 of each individual assay while collecting data for temperature profiles.

III. Fitting Data to the Equilibrium Model

Since the temperature profiles described above are composed of progress curves of product concentration [*P*], Equation 1-8, which is based on enzyme activity, cannot be directly used and is integrated to give one based on [*P*] [20]:

Equation 2-6:

$$[P] = -\frac{e^{\left(\frac{\Delta G^{\ddagger}_{cat}}{RT}\right)} \cdot E_0 \cdot e^{\left(\frac{-\frac{k_B T \cdot e^{\left(-\frac{\Delta G^{\ddagger}_{inact}}{RT}\right)} \cdot e^{\left(-\frac{\Delta H_{eq}\left(\frac{1}{T}-\frac{1}{T_{eq}}\right)}{R}\right)} \cdot t}{h\left(1+e^{\left(\frac{\Delta H_{eq}\left(\frac{1}{T}-\frac{1}{T_{eq}}\right)}{R}\right)}\right)} \right)}}{e^{\left(-\frac{\Delta G^{\ddagger}_{inact}}{RT}\right)} \cdot e^{\left(-\frac{\Delta H_{eq}\left(\frac{1}{T}-\frac{1}{T_{eq}}\right)}{R}\right)}} + \frac{e^{\left(-\frac{\Delta G^{\ddagger}_{cat}}{RT}\right)} E_0}{e^{\left(-\frac{\Delta G^{\ddagger}_{inact}}{RT}\right)} \cdot e^{\left(-\frac{\Delta H_{eq}\left(\frac{1}{T}-\frac{1}{T_{eq}}\right)}{R}\right)}}$$

By fitting a 3D landscape depicting the enzyme activity–time–temperature relationship, such as those depicted in Figure 2-2, to the Equilibrium Model, a set of parameters are generated, and these parameters collectively depict a set of enzyme-temperature relationships most similar to that of the original temperature profile while conforming to the enzyme behavior described by the Equilibrium Model. In addition, because the Equilibrium Model has three dimensions, various issues can arise during the fitting process that could skew, or even invalidate the results.

Although a standalone MATLAB® (Version 7.1.0.246 (R14) Service Pack 3, The MathWorks, Inc.) application is now available from Prof. Roy M. Daniel

(r.daniel@waikato.ac.nz) for the determination of Equilibrium Model parameters, all the Equilibrium Model work described in this work was performed using Scientist™ (version 2.01, MicroMath Inc.) as it allows greater control over the fitting process. Fitting results from the two software packages agree well (unpublished data, Peterson, M. E.) and can be considered comparable, although MATLAB® tends to perform considerably better with very large data sets.

a. The Fitting Process

Raw data of a typical temperature profile generated using methods described above are imported into Scientist™ along with E_0. A set of *common estimates* [2] of the parameters ($\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq}) is used as initial values, which are then optimized using a *Simplex* fit until they stabilize (i.e., additional iterations of fitting do not alter the values). The *optimized* parameter estimates are then used as initial values to reiteratively *Least Square* fit the data set to the Equilibrium Model to generate the final stable parameter values. Basic statistics are calculated for each parameter and for the fit itself using Scientist™'s built-in statistics function, and the parameters then used to generate a 3D activity plot (rate versus temperature versus time) of the fitted profile. The parameters and E_0 (except $\Delta G^{\ddagger}_{inact}$) are also used to generate a 2D zero-time activity plot using the zero-time Equilibrium Model to allow visual determination of T_{opt}.

- **Zero-Time**

Fitting initial rates to the two-dimensional zero-time Equilibrium

Model is identical in concept to the normal fitting process. The main differences are the absence of $\Delta G^{\ddagger}_{inact}$ from the parameter set and the use of reaction rates (M/second) rather than [*P*]. Initial rates are calculated using methods described above, but any reliable initial rates, be they derived from discontinuous spectrophotometric or other types of assays, can be used. It should be note that parameters derived from the zero-time Equilibrium Model tend to have larger statistical errors. This is due to the inherent small size of the data set, so parameters derived from the zero-time model must be carefully verified against raw data.

b. Verifying the Fitting Results and Resolving Fitting Anomalies

The fitting process used to generate Equilibrium Model parameters is a least square regression analysis aimed at finding a *global* minimum of the sum of the squares of the ordinal differences between the original data from the temperature profile and those predicted by the Equilibrium Model. Occasionally the algorithm fails to identify the *global* minimum and becomes stuck in a *local* minimum, leading to *anomalous* results, or even goes off on a tangent and produce *implausible* results. While implausible results are easy to identify (e.g., a T_{eq} of 500 K), not all anomalous results can be spotted without comparing the original raw data with the behavior predicted by the final parameters.

- **Verifying the Fitting Results**

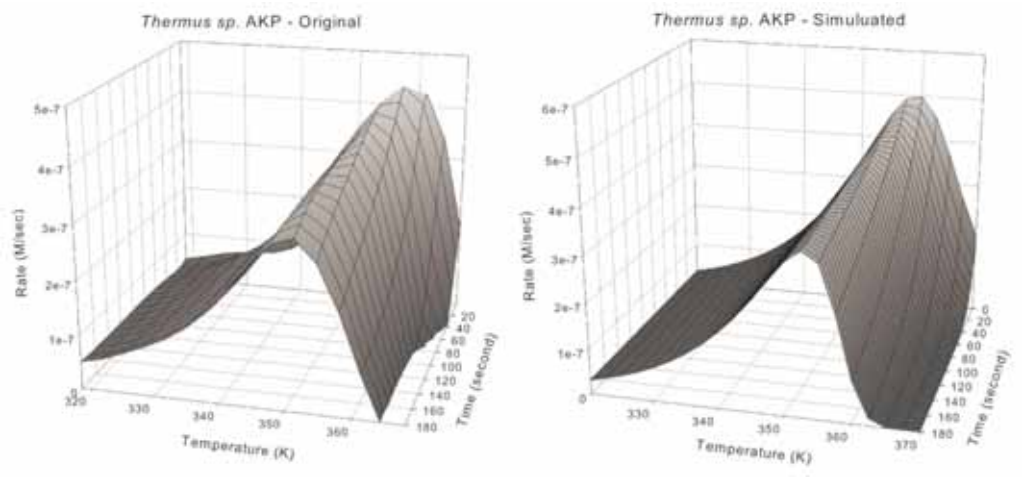
The verification of the Equilibrium Model parameters derived from the fitting process can be summarized into three steps, as detailed below:

(1) Comparison of 3D plots. To facilitate comparisons between two very different types of data (assay absorbance data versus the Equilibrium Model parameters), they need to be rendered graphically. Because enzymologist brains are more attuned to interpreting enzyme activity curves than progress curves of product concentration, the parameters are rendered with a version of the Equilibrium Model that yields enzyme activity curves (see Figure 2-2). This is performed by importing the final parameters to an alternate version of the model and simulating a 3D plot using temperature and time ranges similar to those of the original data. For the original experimentally determined progress curves, in addition to being converted to activity curves, they also need to be smoothed to improve the readability of the graphs. The conversion is performed using this formula:

Equation 2-7:

$$(\text{enzyme activity at } t) = \frac{|[\text{change in absorbance from } (t - 1) \text{ to } t] - (\text{background activity})|}{\varepsilon \cdot a}$$

Figure 2-2 – 3D plots (enzyme activity versus temperature versus time) of *Thermus sp.* alkaline phosphatase: comparing experimental data to Equilibrium Model fit.



'Original' represents smoothed activity curves derived from raw data as assay progress curves, and 'Simulated' is a plot of activity curves generated using the Equilibrium Model thermodynamic parameters derived from the same raw data, applied to an activity-based Equilibrium Model formula. Both plots are generated using SigmaPlot for easy comparison. In the given example, the shapes of the plots are very similar, and the activity axis indicates that the maximal activity levels are very close.

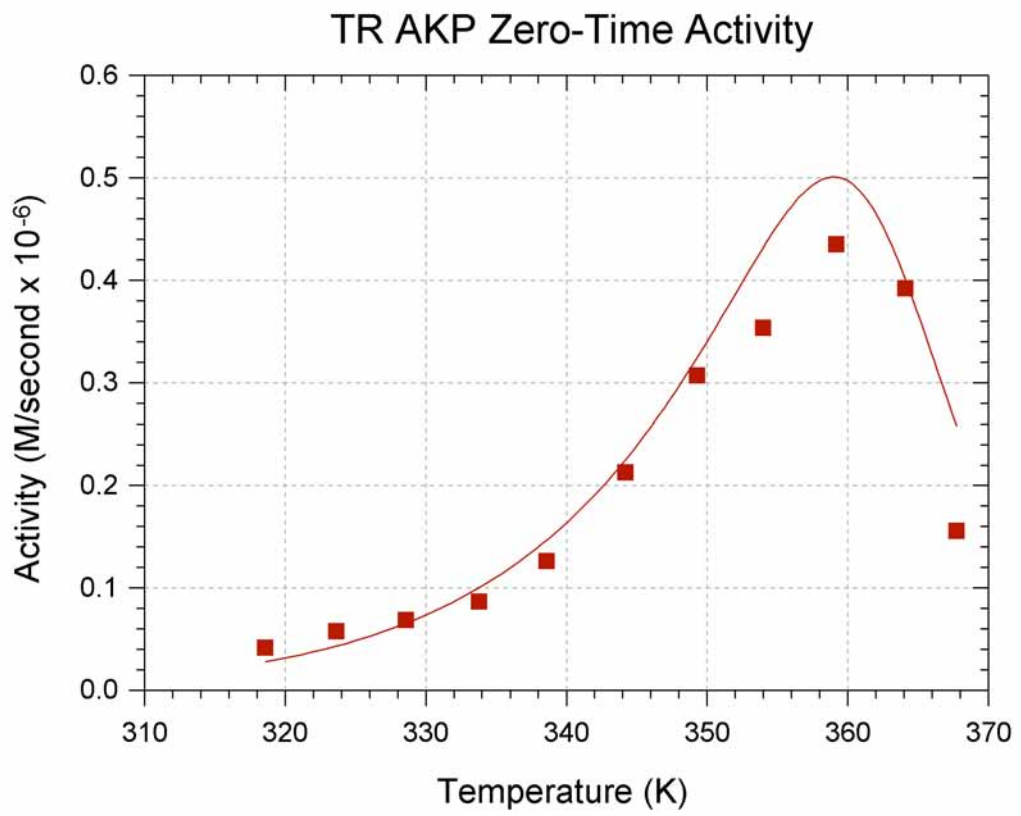
The resulting activity curves are imported into SigmaPlot® (Version 7, SysStat Software, Inc.) and smoothed with 3D *Loess* smoothing. For an example of such comparisons, see Figure 2-2. Several features of the 3D plots should be inspected: the activity scales (should be within 20% of each other), the overall shape of the plots, and where significant

denaturation started to occur (see Figure 2-2).

(2) Generating the zero-time plot. Once a set of Equilibrium Model parameters has been visually verified with the 3D plot, its should be applied to the zero-time Equilibrium Model to generate a zero-time plot, which can then be used to visually estimate T_{opt} (see Figure 2-3). At this point, one should also make sure that there are temperature points over T_{opt} in the original temperature profile, as well as that K_m has been determined for at least one point over T_{opt}. If either of these criteria were not satisfied, the validity of the parameters may be in question, and additional assays or K_m determinations should be performed to remove any doubt.

(3) Verification of the zero-time plot. One should also plot the manually estimated initial rates on the zero-time plot for additional confirmation (see Figure 2-3 for an example); this is also the way to verify parameters generated from fitting initial rates to the zero-time Equilibrium Model. The zero-time verification plot is very useful in assessing the validity of $\Delta G^{\ddagger}_{cat}$ and ΔH_{eq}, which are represented by how activity points below T_{opt} correspond to the plot and by the overall outline of the activity points, respectively (see Figure 2-3).

Figure 2-3 – Zero-time plots (enzyme activity versus temperature) of *Thermus sp.* alkaline phosphatase: comparing experimental data to Equilibrium Model fit.



The solid dots represent the average initial rates at the temperature points included in the original temperature profile. The line represents zero-time enzyme activity versus temperature as predicted using the Equilibrium Model parameters derived from the average initial rates.

- **Resolving Fitting Anomalies**

  Should the fitting results appear anomalous or implausible, a variety of strategies can be employed to rectify the fitting process. The strategies described below mostly apply to Scientist™, but some of them

can be used with MATLAB®. While the MATLAB® package is easier to use and requires little user intervention, if a data set consistently results in irrational parameters, it may be worthwhile to attempt fitting using Scientist™.

The most common problem is the algorithm being stuck in a local minimum during the Simplex fit optimization process. One can manually change the parameters to a more reasonable value and perform more Simplex fits until the parameters settle as more sensible values (this particular trick can be used during the Least Square fit process, albeit very moderately), or if Simplex fit keeps becoming trapped in a local minimum, simply stop the optimization process at a set of rational parameters. Since the recommended initial parameter values are merely *common estimates* [2], it is perfectly justifiable to change the initial parameters to values more rational to the enzyme being investigated.

Scientist™ allows the user to set arbitrary limits on one or more of the parameters, which can be very helpful if the Simplex algorithm keeps going off on a tangent. For example, if Simplex fits consistently lead to a T_{eq} value of 100 K, one can safely set the lower limit for T_{eq} to 250 K, knowing that an enzyme is unlikely to have a T_{eq} below the freezing point. The same can be applied to other parameters, but such limits must be very broad lest the algorithm becomes artificially destined to produce a certain set of parameters.

If all the above measures fail, one can simply leapfrog the Simplex optimization and use parameters obtained from fitting the initial rates to

the zero-time model as initial values. However, this technique does depend on a successful fit using the zero-time Equilibrium Model.

All the measures described above can be used in conjunction with others, but a keen common sense must be applied when using some of the techniques to avoid bias in the resulting parameters.

The Equilibrium Model is a powerful tool that offers useful description of enzymes' temperature-related behaviors. Its application relies on accurate and comprehensive temperature profiles, and the successful and objective use of regression analysis software. Based on experiences gained from the present and previous studies [2, 19, 20], satisfactory results can be reliably obtained using the Equilibrium Model if guidelines and rules described here are followed.

Chapter Three

– Eurythermalism and the Temperature Dependence of Enzyme Activity

Being quantitative and potentially physiologically based, the Equilibrium Model has been hypothesized to be relevant for studying enzyme temperature adaptation. Enzymes from a wide variety of sources were characterized using the Equilibrium Model, and the resulting data were analyzed to identify statistical relationships among enzyme thermodynamic parameters. The findings have not only validated the Equilibrium Model's wide applicability to enzymes of different types and temperature affinity, but also its usefulness for enzyme temperature adaptation studies.

This chapter was prepared as a standalone research article and has been published in the June 2007 issue (volume 21, issue 8) of the FASEB Journal (page 1934-1941). An extended discussion of this chapter can be found in Appendix III.

My Contributions

Some of the enzyme data presented in the survey were provided by colleagues and collaborators, specifically Michelle E. Peterson (*B. taurus* alkaline phosphatase, *P. fluorescens* aryl acylamidase, and wheat germ acid phosphatase) and Charis Shepherd (all isopropylmalate dehydrogenases). I performed the majority of experimental work with assistance from others as noted in acknowledgements, and designed the statistical analyses, which were confirmed to be valid and robust by Dr. Ray Littler. I also wrote the paper with assistance from Prof. Daniel and valuable feedbacks from co-authors of the paper.

Eurythermalism and the Temperature Dependence of Enzyme Activity

Charles K. Lee[*], Roy M. Daniel[*, *], Charis Shepherd[†], David Saul[†], S. Craig Cary[*, ‡], Michael J. Danson[§], Robert Eisenthal[¶], and Michelle E. Peterson[*]

[*]Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton, New Zealand;

[†]School of Biological Sciences, University of Auckland, Auckland, New Zealand;

[‡]College of Marine and Earth Studies, University of Delaware, Lewis, Delaware 19958, USA;

[§]Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK;

[¶]Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

Running Title: The Temperature Dependence of Enzyme Activity

[*]Correspondence: Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. Email: r.daniel@waikato.ac.nz Telephone: +64-7-838-4213 Fax: +64-7-8384324

Abbreviations: GDH, glutamate dehydrogenase; MDH, malate dehydrogenase; DHFR, dihydrofolate reductase; ASB, Antarctic sea bacterium; AKP, alkaline phosphatase; AAA, aryl acylamidase; ACP, acid phosphatase; GLU, glucosidase; FUM, fumarate hydratase; GGTP, γ-glutamyl transferase; PAL, phenylalanine ammonia-lyase; IPMDH, isopropylmalate dehydrogenase

Definitions: $\Delta G^{\ddagger}_{cat}$, the free energy of activation of the catalytic reaction; $\Delta G^{\ddagger}_{inact}$, the free energy of activation of the thermal denaturation process; ΔH_{eq}, the enthalpic change for the transition between active and inactive forms of the enzyme; T_{eq}, the temperature for the mid-point of active-inactive transition; T_{opt}, the temperature at which enzyme activity at zero-time is highest; Half Width at Half Maximum (HWHM), the width of the zero-time temperature/activity plot between T_{opt} and the upper temperature at which 50% of maximum activity is exhibited (see Figure 3-1).

I. Abstract

The 'Equilibrium Model' has provided new tools for describing and investigating enzyme thermal adaptation. It has shown that the effect of temperature upon enzyme activity is not only governed by $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$, but also by two new intrinsic parameters, ΔH_{eq} and T_{eq}, which describe the enthalpy and mid-point, respectively, of a reversible equilibrium between active and inactive (but not denatured) forms of enzyme.

Twenty-one enzymes from organisms with a wide range of growth temperatures were characterized using the Equilibrium Model. Statistical analysis indicates that T_{eq} is a better predictor of growth temperature than enzyme stability ($\Delta G^{\ddagger}_{inact}$). As expected from the 'Equilibrium Model', ΔH_{eq} correlates with catalytic temperature tolerance of enzymes and thus can be declared the first intrinsic and quantitative measure of enzyme eurythermalism. Other findings shed light on the evolution of psychrophilic and thermophilic enzymes. The findings suggest that the Equilibrium Model's description of the effect of temperature on enzyme activity applies to all enzymes regardless of their temperature origins; and that its associated parameters, ΔH_{eq} and T_{eq}, are intrinsic and necessary parameters for characterizing the thermal properties of enzymes, and their temperature adaptation and evolution.

Key Words: enzyme temperature optimum • Equilibrium Model • growth temperature • protein stability • temperature adaptation

II. Introduction

One of the most important environmental factors for life is temperature. It affects diffusion, membrane fluidity, nucleic acid stability, salt and gas solubility, and, significantly, the behavior of enzymes. Until recently, two parameters have been used to characterize the effect of temperature on enzymes: firstly, the temperature-related reaction kinetics of the chemical reaction catalyzed by the enzyme, expressible as $\Delta G^{\ddagger}_{cat}$, the free energy of activation of the catalytic reaction, and secondly thermal stability, which can be described using $\Delta G^{\ddagger}_{inact}$, the free energy of activation of the thermal denaturation process.

Some functional studies of enzyme temperature-linked properties have defined a so-called enzyme "temperature optimum" which arises from this two-parameter 'Classical Model' [1]. However, this "temperature optimum" is not an intrinsic enzyme property, being derived from a complex mixture of both activity and thermal stability effects [1] and dependent on assay duration. Consequently, it is of limited value for measuring enzyme temperature adaptation. Evidence of anomalously low enzyme activity at high temperatures [12, 16] also suggests that the Classical Model is insufficient to describe fully the effect of temperature on enzyme activity. In addition, there has not been any quantitative tool for characterizing an enzyme's eurythermalism. Traditionally, eurythermalism has been used to describe an organism's ability to withstand a wide range of temperatures. To date, there is limited information on how eurythermalism is achieved in eurythermal organisms, and an important possibility is that special characteristics of enzymes in these organisms play an important role.

Enzyme thermal stability clearly has a general correlation with growth temperature; enzymes from thermophiles are typically more thermally stable than those from mesophiles, and those from mesophiles generally more stable than their psychrophilic counterparts. However, there are exceptions, such as RNase [101, 102], mammalian adenylate kinase [103, 104], and catalase [105-109], which are much more stable than might be expected from their environmental temperatures. Factors other than a high temperature environment may in some cases be responsible for high thermal stability. Resistance to proteolysis [29] and organic solvents [110], for example, are known to be associated with high thermal stability. In addition, almost all thermal stability determinations are done in the absence of substrate, which often affects stability.

Two new, intrinsic, temperature-dependent parameters of enzymes, T_{eq} and ΔH_{eq}, have recently been described [1] and validated [2]. These arise from the 'Equilibrium Model', which proposes that the active form of enzyme (E_{act}) is in reversible equilibrium with an inactive form (E_{inact}), and that it is the inactive form that undergoes irreversible thermal inactivation to the thermally denatured state (X), as described in Equation 3-1.

Equation 3-1: $$E_{act} \underset{}{\overset{K_{eq}}{\rightleftharpoons}} E_{inact} \xrightarrow{k_{inact}} X$$

The behavior of the model is described by k_{cat}, the rate constant of the enzyme-catalyzed reaction, k_{inact}, the rate constant for the irreversible inactivation reaction, and K_{eq}, the equilibrium constant governing the ratio E_{inact}/E_{act} in the reversible interconversion. The dependence of catalytic rate on temperature is

obtained from these constants expressed in terms of the thermodynamic parameters, $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq} as follows:

Equation 3-2: $$k_{cat} = \frac{k_B T}{h} \cdot e^{-(\frac{\Delta G^{\ddagger}_{cat}}{RT})}$$

Equation 3-3: $$k_{inact} = \frac{k_B T}{h} \cdot e^{-(\frac{\Delta G^{\ddagger}_{inact}}{RT})}$$

Equation 3-4: $$K_{eq} = e^{\frac{\Delta H_{eq}}{R} \cdot (\frac{1}{T_{eq}} - \frac{1}{T})}$$

where R is the gas constant, k_B is Boltzmann's constant and h is Planck's constant.

T_{eq} is the temperature representing the mid-point of transition between the active and inactive (but undenatured) forms of the enzyme. In other words, T_{eq} is the temperature at which $K_{eq} = 1$ and $\Delta G^{\ddagger}_{eq} = 0$ (we have previously [1] used the term T_m to designate this temperature, but now prefer the term T_{eq} [2] since it is now clear that T_{eq} is entirely separated from and unrelated to thermal denaturation). Of more than 40 enzymes examined in our laboratories [2, 20, 22], including those presented here, all followed the temperature-dependent behavior predicted by the Equilibrium Model, and none followed that expected from the Classical Model. Furthermore, T_{eq} is independent of $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$. Preliminary evidence suggests that the active-inactive transition described by the Equilibrium Model arises from a conformational transition [2] at or near the active site, since the active-inactive equilibrium is established in a timescale at least two orders of magnitude faster than that of thermal denaturation [2]; stabilizing and destabilizing agents have no effect on T_{eq} [19, 22]; and different substrates can affect T_{eq} without altering stability [19, 22].

Not only does the Equilibrium Model provide a description of enzyme temperature-linked behavior, it does so under assay conditions, so that the derived constants such as T_{eq} and ΔH_{eq} are potentially physiologically based. Since enzyme temperature adaptation is the functional basis of metabolic and organism adaptation studies, the Equilibrium Model is not only useful, but necessary for conducting such studies.

In this work, the intrinsic temperature-dependent properties of twenty-one enzymes, including several isoactive sets of enzymes derived from a wide range of growth temperature sources, were determined. Parameters generated from model fitting were analyzed to identify any associations with each other or with growth temperature optima.

III. Material and Methods

Enzymes and reagents. For detailed descriptions of enzyme sources, see Table 3-5 in the Supplemental Data section of this chapter. All chemicals used in this study were purchased from Sigma-Aldrich Inc. and Wako Chemical Company, and are of analytical grade or higher (Table 3-4 in Supplemental Data).

Enzyme assays and data collection. In general terms, these were carried out as previously described [20]. All enzyme assays performed in this work were continuous photometric assays measured using a Thermospectronic™ Helios γ-spectrophotometer equipped with a Thermospectronic™ single cell peltier-effect cuvette holder. All reactions were started by the rapid addition and mixing of a few microliters of enzyme solution at zero degrees into 400-3000 µL

of temperature-equilibrated reaction mixture to enable data collection to begin within the first few seconds; the temperature of the reaction mixture remains within ± 1ºC of desired temperature during data collection. Absorbance data were collected at one-second intervals for three minutes using Vision^{TM} (version 1.25, Thermo Spectronic Inc.) on a Windows PC connected to the Helios, within a range of temperatures depending upon the source of the enzyme (see Material and Methods in Supplemental Data for assay details). In all cases, precautions were taken so that no decrease in reaction rates was due to substrate depletion (see Material and Methods in Supplemental Data). Where possible, substrate concentrations were maintained at over ten times K_m at all temperatures, otherwise appropriate adjustments were made to compensate for the effect of increases in K_m. No signs of substrate/product inhibition were observed under the experimental conditions described. Blank rates were measured at all temperatures and used to correct reaction rates if they were significant. See Table 3-4 in Supplemental Data for assay conditions for individual enzymes.

Based on the variation between the individual triplicate rates from which the parameters are derived for all the enzymes we have assayed thus far, we find that the experimental errors in the determination of $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$ and T_{eq}, are less than 0.5%, and less than 6% in the determination of ΔH_{eq} [20].

Data analysis. The method for processing experimental data is as described in detail in [20]. Absorbance data from Vision^{TM} were first converted in Excel (version 11 for Windows, Microsoft Corp.) to progress curves of product concentration (M). The data were subsequently imported into Scientist^{TM} (version 2.01, MicroMath Inc.), where a set of common estimates of

thermodynamic parameters (see Supplemental Data) was optimized by *Simplex* searching through the complete data set (rate versus time versus temperature). The *optimized* estimates of parameters were then used as initial values to *Least Square* fit the complete data set to the Equilibrium Model, which generated the final values of the parameters; $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq}. The generated parameters were then used in the 'Zero-time Model' [20] to draw a zero-time activity versus temperature plot. This plot was used to visually estimate T_{opt} and HWHM. Initial rates estimated from assay data were also plotted on the zero-time plot to verify the validity of the fittings.

A standalone MATLAB® (Version 7.1.0.246 (R14) Service Pack 3, The MathWorks, Inc.) application, enabling the facile derivation of the Equilibrium Model parameters from a Microsoft® Office Excel file of experimental progress curves (product concentration versus time), is available on CD from the corresponding author. This application is suitable for computers running Microsoft® Windows XP, and is for non-commercial research purposes only.

Protein determination. Protein determinations were performed using the Bradford assay [111].

Statistical analysis. All parameters were imported into STATISTICA [112] for further analysis. Correlation analysis was done for nine parameters, i.e., growth temperature, $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, T_{eq}, T_{opt}, HWHM, the difference between T_{eq} and growth temperature, and the difference between T_{eq} and T_{opt}. The non-parametric Goodman-Kruskal Gamma test was used, and Gamma statistics and the associated two-tailed P values were obtained. Gamma statistic is equivalent to Kendall *Tau* in terms of the underlying assumptions and

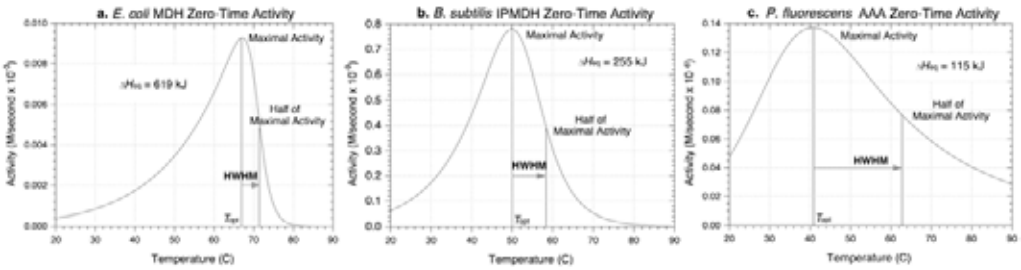
interpretation, but takes tied ranks (e.g., growth temperature) into account. Two-tailed statistical power values were calculated with these conditions: N = 21, *Rho*0 = 0, Alpha = 0.05, and *Exact* algorithm. Best-subset regression analysis was done using growth temperature as the response (dependent) factor, and $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, T_{eq}, and T_{opt} as predictors. A maximum step of 100 and a probability of 0.05 were used for stepwise selection; and best subsets measures were calculated using R squared values. Non-parametric *t* test was carried out using Mann-Whitney U test, and a two-tailed P value was calculated. Scatter plots were drawn using MiniTab (R14 for Windows, Minitab Inc.).

IV. Results and Discussion

All enzymes, including those from extremophiles, obeyed the Equilibrium Model and were found to display zero-time activity optima with temperature, characteristic of the model (e.g., Figure 3-1 & Figure 3-2). It should be noted that given the *fast-start* of the enzyme assays (see Materials and Methods), the decline in activity at zero time at temperatures above T_{opt} is evidence that equilibration of the active-inactive forms of the enzyme (E_{act} and E_{inact}) takes place over timescales significantly shorter than we can observe, i.e., less than 1-3 seconds, and that the equilibration is about two orders of magnitude or more faster than the rate of irreversible activity loss as shown by the activity versus time plots for each temperature in the 3D graphs in Figure 3-2 (although it should be noted that the very slow irreversible denaturation for the ASB HK47 psychrophilic alkaline phosphatase above T_{opt} is not typical of the enzymes examined here).

The parameters of all enzymes are listed in Table 3-1. They are derived by fitting assay data to the Equilibrium Model and thus relate to active enzymes in the presence of substrate and cofactor. The exception is the growth temperature optimum of the source organism, which is cited from various sources of reference (see Supplemental Data). Growth temperatures of the source organisms range from 2ºC to 75ºC. The difference between the source growth temperature and T_{eq} ranges from -6.1 to 52.6ºC.

Figure 3-1 – Zero-time activity plot of enzymes with different eurythermal properties.



a. *Escherichia coli* MDH, b. *Bacillus subtilis* IPMDH, c. *Moritella profunda* DHFR.

Table 3-1 – Enzyme thermodynamic parameters.

Organism	Enzyme	Growth Temp. (°C)	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (°C)	T_{opt} (°C)	HWHM (degree)	T_{eq} minus Growth Temperature	T_{eq} minus T_{opt}
C.utilis	GDH	31	57	94	409	59.5	55.5	6	28.5	4
B.taurus	MDH	39	53	85	826	67.4	64.5	4.5	28.4	2.9
E.coli	MDH	40	55	96	619	70.7	67	4.5	30.7	3.7
M.profunda	DHFR	2	67	93	104	54.6	60.5	32	52.6	-5.9
B.cereus	DHFR	30	66	90	261	58.5	55	8.5	28.5	3.5
G.stearothermophilus	DHFR	55	67	97	96	53.9	63	40.5	-1.1	-9.1
ASB HK47	AKP	15	55	90	149	17.5	15	12	2.5	2.5
B.taurus	AKP	39	57	97	86	59.6	69	44	20.6	-9.4
Thermus sp. RT41a	AKP	75	72	99	305	90	86	9	15	4
P.fluorescens	AAA	25	74	92	115	35.8	41	24	10.8	-5.2
Wheat germ	ACP	20	79	95	142	63.9	66	20	43.9	-2.1
S.cerevisiae	A-GLU	28	71	90	272	39.4	36.5	7.5	11.4	2.9
G.stearothermophilus	A-GLU	55	64	96	225	67.8	64	10.5	12.8	3.8
P.dulcis	ß-GLU	20	63	95	99.8	55.9	61.5	33.5	35.9	-5.6
C.saccharolyticus	ß-GLU	70	81	98	154	73.9	75	18.5	3.9	-1.1

Organism	Enzyme	Growth Temp. (°C)	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (°C)	T_{opt} (°C)	HWHM (degree)	T_{eq} minus Growth Temperature	T_{eq} minus T_{opt}
S.scrofa	FUM	39	60	92	378	59.1	55	6.5	20.1	4.1
B.taurus	GGTP	39	63	98	109	52	55.5	26	13	-3.5
R.glutinis	PAL	24	80	97	181	56.5	56	13	32.5	0.5
B.psychrophilus	IPMDH	20	72	100	123	56.7	60	24	36.7	-3.3
B.subtilis	IPMDH	30	73	94	255	52.8	50	8.5	22.8	2.8
B.caldovelox	IPMDH	70	76	101	573	63.9	61	4.5	-6.1	2.9

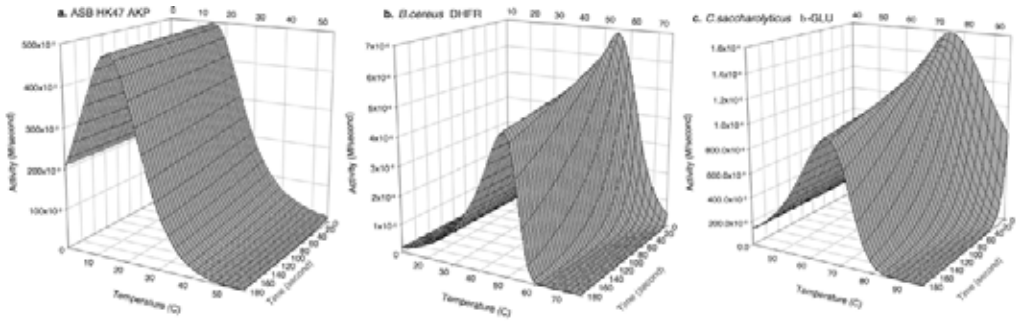
$\Delta G^{\ddagger}_{cat}$ is the free energy of the catalytic reaction, and $\Delta G^{\ddagger}_{inact}$ is that of the thermal denaturation process. ΔH_{eq} is the enthalpic difference between active (E_{act}) and inactive forms (E_{inact}) of the enzyme. T_{eq} is the temperature at which the concentration of E_{act} equals that of E_{inact}. T_{opt} and HWHM are both graphically determined from the zero-time temperature versus activity plots (see Figure 3-1). See Supplemental Data for complete names of the organisms.

T_{opt} is the graphical optimum temperature of the enzyme at time zero (see Figure 3-1). In most cases, T_{eq} is within a few degrees of T_{opt}, and for almost all enzymes T_{eq} (and T_{opt}) are higher than the respective source growth temperature. In the case of T_{eq}, the mean difference is 21.1 ± 15.2ºC. This is not surprising, since at T_{eq} half of the enzyme activity is unavailable, while at 20ºC below T_{eq} most of the enzyme will be in the active form. HWHM is the upper temperature half width of the zero-time temperature/activity peak at 50% activity (the upper temperature half width has been used because the low temperature half of this peak is dominated by $\Delta G^{\ddagger}_{cat}$), and is included to show its relationship to ΔH_{eq} (see Figure 3-1); a large ΔH_{eq} leads to a sharp decline in activity with increasing temperature above T_{opt} (and thus a small HWHM), since ΔH_{eq} is the enthalpic change associated with the reversible conversion of the active to the inactive form of enzyme. There are wide variations in both HWHM (from 4.5 to 44ºC) and ΔH_{eq} (from 86 to 826 $kJ \cdot mol^{-1}$), and thus considerable variations in the sensitivity of enzyme activity to temperature above T_{eq} (and T_{opt}). $\Delta G^{\ddagger}_{inact}$ describes the stability of the enzyme under reaction conditions; as expected, the average $\Delta G^{\ddagger}_{inact}$ of thermophilic enzymes tends to be higher, and the average $\Delta G^{\ddagger}_{inact}$ of psychrophilic/psychrotrophic enzymes lower, than that of mesophilic enzymes.

3D plots of one thermophilic, one mesophilic, and one psychrophilic enzyme are shown in Figure 3-2a-c. Although the plots are very similar in form, the temperature axes show that the three enzymes have very distinct responses to temperature. It is interesting to note that for the ASB HK47 alkaline phosphatase (Figure 3-2a), $\Delta G^{\ddagger}_{inact}$ has relatively little effect on activity over time at temperatures somewhat above T_{opt}, and physiological activity is therefore very

dependent upon T_{eq}, but for the *Bacillus cereus* DHFR (Figure 3-2b), activity over time above T_{opt} depends largely on $\Delta G^{\ddagger}_{inact}$.

Figure 3-2 – 3D plots of enzymes from different thermal origins.



a. Antarctic sea bacterium HK47 – psychrotrophic; b. *Bacillus cereus* – mesophilic; c. *Caldicelulosiruptor sacchrolyticus* – thermophilic. The plots are generated from parameters derived directly from the raw data.

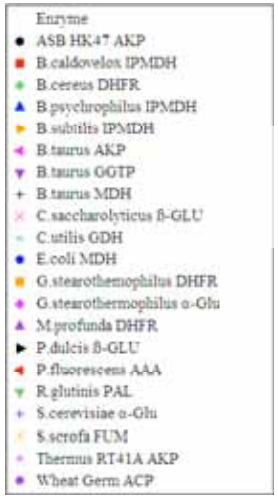
Correlation analysis. A correlation analysis of the parameters in Table 3-1 is shown in Table 3-2. Two sets of significant correlations are evident. The strongest correlation set (in red) involves ΔH_{eq}, HWHM (negatively), and the difference between T_{eq} and T_{opt}. ΔH_{eq} is the enthalpic change associated with the reversible, temperature driven, interconversion of an enzyme between its active and inactive state. It can be considered as a measure of the sensitivity of an enzyme's catalytic activity to temperature. ΔH_{eq} will therefore influence the broadness of zero-time activity versus temperature plots; a small ΔH_{eq} will lead to a broader zero-time activity versus temperature peak, thus giving rise to a strong negative correlation between ΔH_{eq} and HWHM. ΔH_{eq} can thus be considered to be a quantitative measure of an enzyme's ability to function over a temperature

range; a small ΔH_{eq} indicating that the enzyme will function at relatively high activity over a wide range of temperatures, i.e., behave in a eurythermal manner. Conversely, a large ΔH_{eq} indicates stenothermal behavior. The positive correlation between ΔH_{eq} and the difference between T_{eq} and T_{opt} can be explained on the basis that a low ΔH_{eq} will confer a broader time-zero graphical peak, and thus T_{opt} is likely to be higher than T_{eq}.

Compared to the correlation set above, other correlations involving ΔH_{eq}, HWHM, or the difference between T_{eq} and T_{opt} are much weaker. It may be that ΔH_{eq}, and hence an enzyme's ability to function over a wide range of temperatures (e.g., in an environment with large temperature fluctuations) is more likely to be determined by the temperature range or temperature variability of the environment(s) than it is by other thermal properties, although this hypothesis has not been tested here.

Table 3-2 – Correlations of thermodynamic parameters to each other and scatter plots of relevant correlations.

Parameters	Growth Temperature								
Gamma	0.2929	ΔH_{eq}							
p-value	0.0712								
Power	0.2549								
Gamma	-0.0202	-0.1048	$\Delta G^{\ddagger}_{cat}$						
p-value	0.9010	0.5065							
Power	0.1417	0.0733							
Gamma	0.3535	-0.1333	0.3048	$\Delta G^{\ddagger}_{inact}$					
p-value	0.0295	0.3978	0.0533						
Power	0.3582	0.0883	0.2733						
Gamma	**0.5253**	0.3143	-0.0286	0.3048	T_{eq}				
p-value	**0.0012**	0.0463	0.8562	0.0533					
Power	**0.7186**	0.2890	0.0517	0.2733					
Gamma	0.3878	-0.0288	0.0288	0.4327	**0.6442**	T_{opt}			
p-value	0.0176	0.8555	0.8555	0.0063	**<0.0001**				
Power	0.4252	0.0517	0.0517	0.5196	**0.9146**				
Gamma	**-0.4646**	-0.0286	-0.0667	-0.0571	0.0667	0.1346	T_{eq} minus Growth Temperature		
p-value	**0.0042**	0.8562	0.6725	0.7171	0.6725	0.3956			
Power	**0.5890**	0.0517	0.0593	0.0568	0.0593	0.0890			
Gamma	-0.2680	**-0.9512**	0.1122	0.1707	-0.2585	0.0936	0.0537	Half Width at Half Maximum	
p-value	0.1017	**<0.0001**	0.4821	0.2848	0.1053	0.5596	0.7367		
Power	0.2188	**1.0000**	0.0768	0.1141	0.2061	0.0685	0.0560		
Gamma	0.3232	**0.7333**	-0.1048	-0.0952	0.2952	-0.0673	-0.0667	**-0.7366**	T_{eq} minus T_{opt}
p-value	0.0466	**<0.0001**	0.5065	0.5459	0.0612	0.6710	0.6725	**<0.0001**	
Power	0.3039	**0.9818**	0.0733	0.0692	0.2584	0.0595	0.0593	**0.9831**	



Strong correlations are in color, while weaker ones are in normal typeface. Correlation analysis was done using the non-parametric Gamma test (where Gamma is equivalent to the correlation coefficient on the scale of -1 to +1). P value is the measure of the correlation by chance, while Power value is the probability of rejecting a false null hypothesis. A low Power value (i.e., lower than 0.8) indicates that a correlation is subject to type II errors (false negative); in other words, the sample size is too small to successfully detect a weak correlation between the two parameters. For the scatter plots, vertical axes correspond to the parameter along the horizontal plane of the table, and horizontal axes correspond to the parameter below. Least squares regression curves in the scatter plots were calculated using a linear model with fit intercept. The regression curves are not a mathematically correct representation of non-parametric correlation and are only used to indicate general trends in the data. See main text for a detailed description of statistical analyses used.

The second strong correlation set (in green) is that between T_{eq} and T_{opt}, and T_{eq} and growth temperature; and moderate or weak correlations between all three of growth temperature, T_{opt}, and $\Delta G^{\ddagger}_{inact}$, and between T_{eq} and $\Delta G^{\ddagger}_{inact}$. Since T_{opt} can theoretically be calculated from T_{eq} [2], the correlation between these two is not surprising, and the weak correlation of these two with protein stability ($\Delta G^{\ddagger}_{inact}$) simply confirms the well-established general tendency of thermophiles to have more stable proteins than mesophiles. Of the other correlations in this set, the correlation between T_{eq} and optimal growth temperature is particularly significant. It indicates that in a broad sense, T_{eq} is a better indication of an enzyme's temperature origin than either its stability ($\Delta G^{\ddagger}_{inact}$) or the graphically determined T_{opt}. This is supported by the result from best-subset regression analysis, which identifies T_{eq} as the single best parameter for predicting growth temperature (see Table 3-3).

The moderate negative correlation (in blue) between growth temperature and the difference between T_{eq} and growth temperature is interesting. From Table 3-1, it is evident that at high growth temperatures (growth temperature $\geq$ 55ºC), i.e., in organisms that are clearly thermophilic, there tend to be relatively small differences between T_{eq} and growth temperature (7.3 ± 6.6ºC); however, for psychrophiles, psychrotrophs, and low-temperature mesophiles (growth temperature $\leq$ 31ºC), they are larger (27.8 ± 15.1º). A non-parametric T-test between the T_{eq} and growth temperature differences of thermophilic and non-thermophilic enzymes reveals that the two groups are significantly different (P value = 0.008) in this respect.

Table 3-3 – Results of best-subset regression analysis of individual Equilibrium Model parameters.

Parameter	Ranking	R squared value
T_{eq}	1	0.6556
T_{opt}	2	0.6025
$\Delta G^{\ddagger}_{inact}$	3	0.5278
$\Delta G^{\ddagger}_{cat}$	4	0.2992
ΔH_{eq}	5	not determined

Equilibrium Model parameters were analyzed individually. The predictive power of parameter combinations were not examined since the implication of the results would not be clear for the purpose of this study.

A previous study using k_{cat} as a measure of temperature optima also suggested that enzymes tend to have optimal activity at temperatures higher than their hosts' living temperature, and the higher the growth temperature of the organism the narrower the gap becomes [27]. There are several possible reasons for this. There may be a poor match between the environmental temperature of thermophiles and their laboratory growth temperature optima; thermophiles are often isolated from sources that are appreciably hotter than their determined growth temperatures. Alternatively, it might be argued that the origin of the current mesophilic microbial life is thermophilic, i.e., that microbial evolution has proceeded down-temperature from a (hyper)thermophilic last common ancestor [113-115], and that the evolution of T_{eq} lags behind optimal growth temperature. Furthermore, depending on how crucial their roles may be in the cell, different

enzymes might evolve at different speeds to adapt their T_{eq} to the environment, thereby resulting in the variation in T_{eq} among enzymes from one organism (e.g., for *Bos taurus*, Table 3-1).

Conclusions based on weak correlations must be speculative. Although its p-value is slightly over the adopted threshold of 0.05, the correlation between $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$ may suggest that $\Delta G^{\ddagger}_{cat}$ is higher for more stable proteins and thus support various studies [26, 27, 116] that suggested that a change in activation energy ($\Delta G^{\ddagger}_{cat}$) might be important to the temperature-adaptation of psychrophilic enzymes. To investigate this idea, correlation analysis was performed with psychrophilic and psychrotrophic enzymes (growth temperature ≤ 20ºC, not including plant enzymes) excluded; the correlation between $\Delta G^{\ddagger}_{inact}$ and $\Delta G^{\ddagger}_{cat}$ was found to deteriorate significantly (P value > 0.1), tending to confirm that a lower $\Delta G^{\ddagger}_{cat}$ might indeed be one of the main adaptations for psychrophilic enzymes, while non-psychrophilic enzymes take on other means of adaptation.

The scatter plots in Table 3-2 are consistent with the correlations described.

Growth temperature. In general, any correlations with growth temperature must be regarded as tentative since determining environmental growth temperature from laboratory-determined optimal growth temperatures of bacteria is problematic, partly because the temperature of most microbial environments varies with time, including that of *Escherichia coli*, for example [117, 118], and partly because the laboratory environment on which these are usually based may not represent environmental conditions. In the absence of better measures, however, we have generally adopted here the published data for organisms involved in this study (see Supplemental Data).

Individual enzymes. The DHFR from the strict psychrophile *Moritella profunda*, isolated from deep Atlantic sediments [119, 120], is a surprisingly stable enzyme with a concomitantly high T_{eq}, while its ΔH_{eq} (104 kJ·mol^{-1}) is significantly lower than the 25th percentile (115 kJ·mol^{-1}), suggesting eurythermal behavior. We speculate that its low temperature activity has been achieved by evolving a lower ΔH_{eq} into an enzyme with an incidentally high T_{eq}, thus conferring eurythermalism. Part of the lower half (below T_{opt}) of the enzyme activity curve is influenced by ΔH_{eq}, but the effect of $\Delta G^{\ddagger}_{cat}$ may dominate, as stated above.

The alkaline phosphatase from isolate ASB HK47 (Figure 3-2a) gives another illustration of the relative effects of the four parameters on enzyme activity over a range of temperatures. For many enzymes (e.g., *B. cereus* DHFR, Figure 3-2b), the denaturation rate is such that it will play the major role in controlling enzyme activity over timescales important to the organism (over minutes rather than seconds) at temperatures above T_{eq}. However, for the ASB HK47 alkaline phosphatase it is evident that the lower rate of denaturation above T_{eq} leads to a situation where activity will be dominated by T_{eq} and ΔH_{eq} over significant timescales.

To achieve temperature adaptation, evolution imposes a strong selection force for thermophilic enzymes to have both high thermal stability and a high T_{eq}. High thermal stability has been thought to be achieved at the expense of flexibility to explain why thermophilic enzymes generally exhibit relatively little activity at lower temperatures [62]. However, the rigidity required for stability is global, while the flexibility required for catalysis is likely to be local [121], as

demonstrated by some exceptions [63, 122]. It has been suggested that a combination of local flexibility at the active site and high overall stability [60, 61], is behind such phenomena. In other words, enzymes active at low temperatures don't necessarily have to be thermally unstable [17, 48]. Recent reports also suggested that the higher flexibility of psychrophilic enzymes is local and on the microsecond to millisecond timescale [123], while their global flexibility may not be necessarily higher [17]. In addition, random and directed mutagenesis studies [48, 124, 125] have shown that it is possible to engineer an enzyme so that it retains low-temperature activity while gaining thermostability.

V. Conclusion

The correlations observed here do not necessarily denote a causative relationship. The finding that T_{eq} correlates best with growth temperature does not necessarily mean that it is the parameter most directly selectable by evolution, although it does provide insight into the effect of temperature on the evolution of enzymes. It should also be noted that the strong correlation between T_{eq} and growth temperature means that the correlation is likely to be reproducible with any selection of enzymes of similar number and diversity to this study.

ΔH_{eq} is a new quantitative measure to characterize eurythermalism at an enzymic level. The new parameters (i.e., T_{eq} and ΔH_{eq}), combined with established parameters (i.e., $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$), define fully how temperature influences enzyme activity, while the former provide new tools to describe and investigate this influence.

Much of the data presented here comes from bacterial sources, although results from correlation analysis remain similar when data from eukaryotic sources are excluded (data not shown). Although the twenty-one sets of data from discrete enzymes presented here are a relatively small data set, the main correlations are robust and generally consistent with findings or proposals based on other data; they also provide new targets and perspectives for molecular and structural studies.

The development and verification of the Equilibrium Model has provided additional and quantitative measures of the thermal behavior of enzymes [1, 2] that are essential for describing the effect of temperature on enzyme activity, and useful parameters for measuring the temperature adaptation of enzymes. Additionally, the results here show the Equilibrium Model's potential usefulness as a tool to investigate the relationship between enzyme thermal properties and the influence of temperature on the physiology and evolution of the host organism.

VI. Acknowledgements

This work was partly supported by a grant from the National Science Foundation (Biocomplexity 0120648) and a grant from the New Zealand Marsden fund. We thank Y. Xu at Free University of Brussels for *M. profunda* DHFR, Colin Monk for providing purified *Caldicellulosiruptor saccharolyticus* ß-GLU, David Clement for providing purified *Geobacillus stearothermophilus* DHFR, and Judith Klinman at UC Berkeley for the gift of the *G. stearothermophilus* DHFR

clone. We also thank Dr. Ray Littler at the University of Waikato for some assistance with the statistical analysis. We are grateful to Ophélie Planckaert and An-Chi Tsuei for their help with data collection.

VII. Supplemental Data

The following material provides additional experimental details and the rationale for assigning optimal growth temperature to host organisms. Information related to enzymes and their respective assays is also included.

a. Material and Methods

Enzyme Assays. For general methods relating to these assays, see [20]. Assays with NAD(P)H as substrate were monitored at 365 nm ($_{e}$NAD(P)H = 3,400 $M^{-1}cm^{-1}$) while the release of *p*-nitrophenyl from *p*-nitrophenyl phosphate, *p*-nitrophenyl α-D-glucopyranoside, and *p*-nitrophenyl ß-D-glucopyranoside were measured at 405 nm ($_{e}$*p*NP, pH 9.5/10 = 18,000 $M^{-1}cm^{-1}$, $_{e}$*p*NP, pH 8 = 15,800 $M^{-1}cm^{-1}$). The release of *p*-nitroanilide from L-γ-glutamyl-p-nitroanilide was monitored at 410 nm ($_{\varepsilon}$*p*NA = 8800 $M^{-1}cm^{-1}$), while the formation of fumarate was measured at 240 nm ($_{\varepsilon}$fumarate = 2440 $M^{-1}cm^{-1}$). The formation of trans-cinnamate from L-phenylalanine was monitored at 270 nm ($_{\varepsilon}$trans-cinnamate = 19,730 $M^{-1}cm^{-1}$).

In order to prevent under-estimation of reaction rates due to substrate depletion and to maintain V_{max}, substrates concentrations were set to at least ten times of K_m at all temperatures where possible, which theoretically allows

the enzyme to operate at over 90% of V_{max}. In cases where it is technically difficult to do so (e.g., low substrate solubility, high background/initial absorbance, substrate inhibition, etc), reaction curves were adjusted accordingly at each temperature point to compensate for the effect of K_m increases. In cases where the enzyme is active at temperatures where the determination of K_m is difficult, the K_m value at the highest possible temperature is used for all temperature points above it.

Cloning of *Bacillus cereus* DHFR. An ORF homologous to *Bacillus subtilis* DHFR was found in the *B. cereus* ATCC14579 complete genome using GenBank's BLAST services. The putative DHFR sequence (listed below) translates to a polypeptide 162 amino acids in length, which aligned very well with DHFR sequences from other bacilli. PCR primers were designed based on the sequence with additional Enterokinase cleavage site at 5' end and restriction cut sites (*Nco*I & *Sma*I) on either end. The resulting PCR product was then ligated into the Roche RTS pIVEX2.4d expression vector, which added His-tag to 5' end of the gene. The ligation mix was transformed into JM109 competent cells using electroporation and plated on selective medium (100 μg/ml ampicillin). A clone with the correct size insert was identified using PCR screening and used for plasmid purification. The purified plasmid was sequenced for validation and matched perfectly with predicted *B. cereus* ATCC14579 DHFR sequence.

>*B. cereus* DHFR

ATGATTGTTTCATTTATGGTCGCTATGGACGAAAATAGAGTAATTGGTAAAGATAATAATTT
ACCTTGGCGTTTACCGAGTGAATTACAATACGTGAAGAAAACAACGATGGGTCACCCGCTTA
TTATGGGAAGAAAGAACTATGAAGCGATTGGTAGACCACTGCCTGGAAGACGTAATATTATT

GTAACTCGTAATGAAGGGTATCATGTTGAAGGTTGTGAAGTTGTTCATTCTGTAGAAGAAGT
GTTTGAACTATGTAAAAATGAAGAAGAGATTTTTATTTTTGGTGGAGCGCAAATTTACGACC
TCTTTTTACCTTATGTAGACAAGTTATATATAACAAAAATCCATCATGCATTTGAAGGAGAT
ACATTCTTCCCAGAAATAGATATGACAAATTGGAAAGAGATTTTTGTAGAAAAAGGTTTAAC
GGATGAAAAAAACCCATATACGTATTATTACCATGTATATGAGAAACAACAATAA

Over-expression and Ni-NTA purification of *B. cereus* DHFR. Constructed pIVEX2.4d vector with *B. cereus* DHFR insert was transformed into Invitrogen® BL21 Star™ (DE3) pLysS One Shot® Chemically competent cells, which has an IPTG-inducible T7 expression system. Cells were grown at 37ºC and induced with 0.5 mM IPTG. Harvested cell pellet was resuspended in non-denaturing lysis buffer from the QIAGEN Ni-NTA Spin Kit and incubated on ice for 30 mins with 1 mg/ml lysozyme. The lysate was then sonicated using (Sonicator Ultrasonic Processor XL, Misonix Inc.) with micro tip at 25% power for a total of 3 minutes at half minutes intervals and spun down. Protein with DHFR activity was purified from the supernatant following QIAGEN Ni-NTA Spin Kit non-denaturing protocols. The protein appeared as a single band on SDS-PAGE and native protein gel with an estimated molecular weight around 23 kDa, which corresponds well to the predicted molecular weight of *B. cereus* DHFR (22.7 kDa).

Enzyme Purity Information. The enzymes used in this study can be classified into three categories based on how they were obtained: 1. Purchased from a commercial supplier; 2. Cloned from native organism with a 6xHis-tag attached and purified using a metal affinity column (i.e., Ni-NTA resin); 3. Cloned from native organism without a 6xHis-tag or purified directly from native organism using various chromatography techniques.

The final concentration of enzyme in the reaction mixture is sub-μM and sometimes sub-nM. Detailed information about enzymes used in this study and their purity can be found in Table 3-5.

b. Growth Temperature Assignments

Assigning environmental growth temperature optima to microorganisms is problematic, since the incubating environments in the lab are in many cases not necessarily those experienced by the microorganisms in nature, and in any event, environmental temperatures usually fluctuate. However, the non-parametric Goodman-Kruskal Gamma correlation analysis relies on the ranks of the data rather than the actual values, so minor errors in growth temperatures for the organisms are unlikely to influence the analysis significantly. We provide here the basis on which the growth temperature optima were chosen, and some rationale in cases of disagreement between data sources.

Candida utilis (a.k.a. Torula yeast) has a reported optimal growth temperature of 31ºC [178].

Escherichia coli is one of the most ubiquitous and well-known intestinal bacteria in large mammals. The optimal growth temperature of *Escherichia coli* in rich complex medium is generally considered to be 39ºC [179], the commonly known body temperature of ***Bos taurus***. Recent articles [180, 181], however, suggested a higher temperature range (40-41ºC) for growth optima, which corresponds to the physiological environment of *E. coli* better, since intestinal temperature is expected to be higher than rectally measured

body temperature. Therefore, 40ºC is used as the optimal growth temperature of *E. coil* in this study.

Moritella profunda was originally isolated at a depth of 2815 m in the eastern tropical Atlantic ocean [120], where the environmental temperature is 2ºC. Under atmospheric pressure, *M. profunda* exhibited maximum growth rate at 2ºC. Although the optimal growth temperature wasn't determined under pressurized conditions, it grew significantly better at 6ºC than at 10ºC under pressure, suggesting a preference for lower temperature. 2ºC has been used as the optimal growth temperature for *M. profunda*.

According to American Type Culture Collection catalog, the suggested growth temperature for ***Bacillus cereus*** ATCC14579 is 30ºC. *B. cereus* (*senso stricto*) has long been considered as a standard soil bacterium, as have most members of the genus *Bacillus* [182]. Recent discoveries, however, have suggested that its *Arthromitus* stage inside insect guts is probably its main proliferation stage [183, 184], as sporulation occurs quickly in open-air environments [184]. Although insects are by no means endothermic, evidence suggests that they are excellent in thermoregulation, especially at preventing overheating. In other words, *B. cereus* probably enjoys a relatively low temperature living environment with rare exposures to high temperatures in its non-spore form. Therefore, there is unlikely to be significant evolutionary pressure for *B. cereus* enzymes to be stable at high temperatures. This is consistent with the low $\Delta G^{\ddagger}_{inact}$ of *B. cereus* DHFR.

The growth optimum of ***Geobacillus stearothermophilus*** DSM13240 is unclear. The German Collection of Microorganisms and Cell Cultures

(DSMZ) suggests cultivation of *G. stearothermophilus* (previously named *Bacillus stearothermophilus* strain 10 [185]) at 65ºC, whereas all other members of *Geobacillus stearothermophilus* have a suggested growth temperature of 55ºC, which is the growth temperature suggested by the original isolator [186]. The same strain listed in Collection de I'Institut Pasteur (CIP106956) also has a suggested growth temperature at 55ºC. 55ºC was adopted here as the optimal growth temperature for this organism.

Antarctic sea bacterium HK47 (ASB HK47) has a determined growth optimum at 15ºC [187], and the majority of strains isolated in the same study were also psychrotrophs.

***Thermus* sp. RT41a** is a thermophilic eubacterium isolated from pool RT4 of Kuirau Park, Rotorua, New Zealand in 1983, but the organism has not yet been validly named [188]. The organism grows between 55ºC and 75ºC, with the optimal growth temperature being 75ºC [189].

Growth temperature optima for ***Pseudomonas fluorescens*** and **wheat germ** are taken from [2]. The *P .fluorescens* strain that aryl-acylamidase was isolated from was deposited to ATCC as *Pseudomonas putida* (Trevisan) Migula ATCC39004 [190], whose optimal growth temperature was listed as 25ºC. Wheat germ germinates in a range of temperature (15-25ºC), and a mid-point of 20ºC was adopted as its growth temperature.

Saccharomyces cerevisiae. The documented optimal growth temperature for baker's yeast is 28ºC [191].

Prunus dulcis. For almonds, the optimal temperature for pollen tube growth and the subsequent germination is usually considered to be in the

range of 20-25ºC, although pollination can still occur at temperature as low as 12ºC at a much slower pace [192]. 20ºC was adopted as the growth temperature for almond here.

Caldicellulosiruptor saccharolyticus ATCC43494 has a reported growth optimum of 70ºC [193-195], which corresponds with its ATCC record.

***Sus scrofa*.** The body temperature of the domestic pig, defined by rectal temperature measurements, is 39ºC, which has been adopted in this study as its optimal growth temperature.

Rhodotorula glutinis ATCC10788 is listed in ATCC as its teleomorph, *Rhodosporidium toruloides*. Its documented optimal growth temperature is 24ºC.

Bacillus psychrophilus is listed in DSM as *Sporosarcina psychrophila*. It has a minimal and maximal growth temperature of -4.5ºC and 30ºC, respectively, and is considered psychrotrophic [196]. The growth optimum of 20ºC suggested in DSM is adopted in this study.

A wide range of suggested optimal growth temperature for ***Bacillus subtilis*** can be found in ATCC, ranging from 26ºC to 37ºC. However, all of its substrains are listed in DSM as having a growth optimum of 30ºC, which is adopted in this study.

Bacillus caldovelox YT-G, listed as *Bacillus* sp. DSM411, has a documented optimal growth temperature of 60-70ºC and a maximal growth temperature of 76ºC [197]. This corresponds well with the optimal growth temperature of 70ºC in DSM's record.

Table 3-4 – Enzyme assay conditions.

Enzyme	Substrate	Cofactor	Reaction Buffer	Enzyme in assay
C. utilis GDH	50 mM 2-ketoglutarate*	2.5 mM NADPH*	100 mM Na^+ phosphate buffer with 150 mM NH_4Cl, pH 7.4	2.2 nM
B. taurus MDH	12 mM oxaloacetate*	1.25 mM NADH*	100 mM Na^+ phosphate buffer, pH 8	0.36 nM
E. coli MDH	10 mM oxaloacetate*	3 mM NADH*	100 mM Na^+ phosphate buffer, pH 8	0.45 nM
M. profunda DHFR [119]	0.5 mM dihydrofolate*	0.5 mM NADPH*	100 mM Na^+ phosphate buffer, pH 8	36.4 nM
B. cereus DHFR	1.5 mM dihydrofolate*	1.6 mM NADPH*	100 mM Na^+ phosphate buffer, pH 8	39.5 nM
G. stearothermophilus DHFR	0.5 mM dihydrofolate*	0.8 mM NADPH*	100 mM Na^+ phosphate buffer, pH 8	30.3 nM
ASB HK47 AKP	2mM *p*-nitrophenol phosphate*	None	100mM diethanolamine buffer with 5mM $CaCl_2$, pH 9.5 (temperature adjusted)	1.2 nM
B. taurus AKP† [2]	10 mM *p*-nitrophenol phosphate*	None	100mM diethanolamine buffer with 0.5mM $MgCl_2$, pH 8.5 (temperature adjusted)	0.07 nM
Thermus sp. RT41a AKP [188]	4 mM *p*-nitrophenol phosphate*	None	50mM borate buffer, pH 10	2.5 nM
P. fluorescence AAA† [2]	0.75 mM *p*-nitroacetanilide*	None	100 mM Tris buffer, pH 8.6 (temperature adjusted)	143 nM
Wheat germ ACP† [2]	10 mM *p*-nitrophenol phosphate*	None	100 mM Na^+ acetate buffer, pH 5	450 nM
S. cerevisiae α-GLU	10 mM *p*-nitrophenol α-D-glucopyranoside*	None	100 mM Na^+ phosphate buffer, pH 8	94.4 nM

* Purchased from Sigma-Aldrich Inc.
† Characterized by M.E. Peterson

Enzyme	Substrate	Cofactor	Reaction Buffer	Enzyme in assay
G. stearothermophilus α-GLU	10 mM *p*-nitrophenol α-D-glucopyranoside*	None	100 mM Na^+ phosphate buffer, pH 8	1.1 nM
T. aquaticus α-GLU	20 mM *p*-nitrophenol α-D-glucopyranoside*	None	100 mM Na^+ phosphate buffer, pH 8	1.44 μM
P. dulcis ß-GLU	40 mM *p*-nitrophenol ß-D-glucopyranoside*	None	100 mM Na^+ phosphate buffer, pH 8	1.48 nM
C. saccharolyticus *ß*-GLU	40 mM *p*-nitrophenol ß-D-glucopyranoside*	None	100 mM Na^+ phosphate buffer, pH 8	686 nM
S. scrofa FUM	100 mM L-malic acid*	None	100 mM K^+ phosphate buffer, pH 7.6	1.07 nM
B. taurus GGTP	20 mM L-γ-glutamyl p-nitroanilide*	500 mM Glycylglycine*	100 mM Na^+ phosphate buffer, pH 8	4.09 nM
R. glutinis PAL	150 mM L-phenylalanine*	None	50 mM Tris-HCl buffer, pH 8.5 (temperature adjusted)	46.8 nM
B. psychrophilus IPMDH‡	1 mM *threo*-D,L-3-isopropylmalate§	1 mM NAD^+ *	20 mM K^+ carbonate buffer with 0.3M KCl and 0.2 mM $MnCl_2$, pH 9	47.2 nM
B. subtilis IPMDH ‡	1 mM *threo*-D,L-3-isopropylmalate§	1 mM NAD^+ *	20 mM K^+ carbonate buffer with 0.3M KCl and 0.2 mM $MnCl_2$, pH 9.5	0.12 nM
B. caldovelox IPMDH ‡	1 mM *threo*-D,L-3-isopropylmalate§	1 mM NAD^+ *	20 mM K^+ phosphate buffer with 0.3M KCl and 0.2 mM $MnCl_2$, pH 8	44.4 nM

‡ Characterized by C. Shepherd and M.E. Peterson. (C. Shepherd and D. Saul, unpublished data)
§ Purchased from Wako Chemical Company

Table 3-5 – Enzyme purity information.

Enzyme	Supplier and Product ID (Lot #) / Source	Purity Information
Candida utilis GDH	Sigma-Aldrich Inc., G5027 (91H02135)	Protein >99%.
Bos taurus MDH	Sigma-Aldrich Inc., M9004 (063K7040)	Impurities: 0.003% Lactic dehydrogenase, 0.004% Glutamic-Oxalacetic Transaminase, 0.001% Glutamic-Pyruvic Transaminase.
Escherichia coli MDH	Megazyme Ltd, E-DMDHEC (30801)	Purity verified with SDS-PAGE and IEF gel (single band)
Moritella profunda DHFR	Courtesy of Y. Xu, J. M. Wiame Research Institute, Microbiology, Free University of Brussels (VUB), B-1070 Brussels.	This enzyme was originally prepared by D. Clement for neutron scattering work and therefore is of very high purity. Purity verified with SDS-PAGE and HPLC.
Bacillus cereus DHFR	Cloned and purified by C. K. Lee	6xHis-tagged enzyme purified using QIAGEN Ni-NTA Spin kit. Purity and molecular weight verified with SDS-PAGE and native protein gel (single band).
Geobacillus stearothermophilus DHFR	Purified by D. Clement in R. M. Daniel's lab using a clone provided by J. Klinman at UC Berkeley, USA.	This enzyme was originally prepared by D. Clement for neutron scattering work and therefore is of very high purity. Purity verified with SDS-PAGE, HPLC and neutron scattering data.
ASB HK47 AKP	Epicentre Biotechnologies, HK^{TM} Thermolabile Phosphatase.	Molecular biology grade.

Enzyme	Supplier and Product ID (Lot #) / Source	Purity Information
Bos taurus AKP	Roche Applied Science, Catalog # 1 097 075 (84731327)	Molecular biology grade. Contaminating activities (from 1000 units of enzyme solution): less than 50 units of DNase 1 & 2 activity, less than 50 units of nicking activity, less than 50 units of exonuclease activity, Less than 50 units of RNase activity.
Thermus sp. RT41a AKP	Purified by J. M. Bragger from native organism in R. M. Daniel's lab	This enzyme was purified with affinity and hydroxyapatite (HAP) chromatography. The preparation was then dialyzed to improve purity. The protein was visualized on a SDS-PAGE gel as a dominant band compared to contaminants.
Pseudomonas fluorescence AAA	Sigma-Aldrich Inc., A4951 (034H6844)	2.0% protein by Biuret. Product contains Tris buffer salts and stabilizer.
Wheat germ ACP	Serva Electrophoresis GmbH, 27971 (08098)	The enzyme is a crude preparation and is essentially the 0.35-0.55 saturated ammonium sulfate fraction from wheat germ described by Singer [198]. According to the manufacturer, the preparation has 0.24 U/mg phosphatase activity and 0.08 U/mg lipase activity (EC 3.1.1.3). According to BRENDA and a more recent study of wheat germ lipase [199, 200], lipase does not exhibit any activity towards the substrate (pNP-Pi) used in this study.
Sacchromyces cerevisiae α-GLU	Sigma-Aldrich Inc., G5003 (081K7415)	Impurities: less than 0.1% of ß-glucosidase, less than 0.1% of α-galactosidase, less than 0.1% of ß-galactosidase.
G. stearothermophilus α-GLU	Sigma-Aldrich Inc., G3651 (115K1451)	Impurities: 0.023% of ß-glucosidase, less than 0.001% of α-galactosidase, less than 0.003% of ß-galactosidase.
Prunus dulcis ß-GLU	Sigma-Aldrich Inc., Fluka 49290 (1198506)	Biochemika grade.
Caldicellulosiruptor saccharolyticus ß-GLU	Purified by C. Monk from a non-His-tagged clone constitutively expressed in *E. coli* in R. M. Daniel's lab	The *E. coli* cell extract containing the *C. sacchrolyticus* ß-GLU was heat-treated during purification, practically destroying any host contaminating enzyme. Since the targeted enzyme is exceptionally stable, the resulting enzyme preparation should be free of any contaminating activity.

Enzyme	Supplier and Product ID (Lot #) / Source	Purity Information
Sus scrofa FUM	Sigma-Aldrich Inc., F1757 (125K7435)	Purified using affinity chromatography. Less than 0.1% malic dehydrogenase.
Bos taurus GGTP	Sigma-Aldrich Inc., G4756 (65H9550)	Impurities: 0.2% Creatine phosphokinase, 0.1% Glutamic-Oxalacetic Transaminase, 0.1% Glutamic-Pyruvic Transaminase.
Rhodotorula glutinis PAL	Sigma-Aldrich Inc., P1016 (n/a)	Sigma Grade I.
Bacillus psychrophilus	Cloned and purified by C. Shepherd	6xHis-tagged enzyme purified using QIAGEN Ni-NTA His-Bind® Superflow kit. Purity verified with SDS-PAGE (single band).
Bacillus subtilis	Cloned and purified by C. Shepherd	6xHis-tagged enzyme purified using QIAGEN Ni-NTA His-Bind® Superflow kit. Purity verified with SDS-PAGE (single band).
Bacillus caldovelox	Cloned and purified by C. Shepherd	6xHis-tagged enzyme purified using QIAGEN Ni-NTA His-Bind® Superflow kit. Purity and molecular weight verified with SDS-PAGE (single band).

Chapter Four

– The *Alvinella pompejana* Episymbiont Metagenome Project

Functional Metagenomics, Gene-Mining, and Proteomic Investigations

This chapter aims to provide a general introduction to metagenomics, a description of the National Science Foundation (USA) Biocomplexity project (title: A Meta-Genome Level Analysis of An Extreme Microbial Symbiosis, NSF proposal number 0120648) [99], and my contributions to the project. Lead by Prof. Cary, the Biocomplexity project is an international collaboration between University of Delaware, Desert Research Institute, SymBio Corp., and University of Waikato. Information regarding the Biocomplexity project presented in this chapter is mostly compiled from research findings of individual participants of the project; a manuscript based on these findings is being prepared for publication.

This chapter begins by reviewing the development of metagenomics, an emerging culture-independent tool for studying microbial ecology, as both a concept and a discipline. The current state of the field of metagenomics and various other approaches that have been employed are discussed, followed by a brief description of the *Alvinella pompejana* episymbiont metagenome project. The final section contains a description of the data-mining protocol which I developed for this thesis. The cloning and overexpression of genes identified in the metagenome database are described, along with related results and discussion.

One of the challenges I encountered during the gene-mining process is the relatively short sequencing reads, which was in part the result of a deliberate decision by the senior investigators on the project to employ highly stringent

quality trimming criteria to maximize the quality of sequence being annotated and analyzed. Although the gene identification step relied on assembled sequence data and was not significantly affected, the shorter sequences made the processes of designing PCR cloning primers and identifying genotypic variations rather time consuming and labor intensive. However, techniques which I developed to circumvent these problems (see chapter content) were sufficiently effective, and several gene candidates were successfully identified and isolated.

The phylogenetic structure of the *A. pompejana* episymbiont community, which is composed of two dominant phylotypes and hundreds of other minor members, meant that isolating alternate genotypes of target genes (i.e., PCR amplicons carrying rare single nucleotide polymorphisms) may be difficult. This is due to PCR's inherent tendency to bias toward dominant genotypes and our limited ability to screen clones carrying isolated genes. No genotypic variants were identified for the successfully expressed gene candidates, and although more than five genotypic variants were identified for one gene candidate, none of them resulted in functional enzyme product, despite repeated attempts.

My Contributions

See section IV of this chapter for a detailed synopsis of my participation and contributions to the Biocomplexity project.

We are drowning in information, but starved for knowledge.

– John Naisbitt, American writer (1929-)

I. A Historical Overview

The field of microbiology was established in 1676 when van Leeuwenhoek, aided by his microscopes, officially set off his microbiologist career with his observations of *animalcules*. Like all microbiologists for the next two hundred years, he observed all forms of life that fitted under his microscope, be they free-living or parasitic, heterotrophic or autotrophic. Unfortunately, observation was also virtually the only method of research for early microbiologists; limited by technology, early microbiologists were confined to describing the morphology and basic physiology of microbial life [126].

The field of microbiology underwent a major transformation in the late 19th century, when Robert Koch established culturing as a valid tool for studying microorganisms. Suddenly, it was possible to characterize all aspects of a prokaryote with great detail and confidence. The study of readily-cultured model organisms stimulated an explosion of knowledge in microbial physiology and genetics in the 20th century [126]. Although microbiologists were aware of the *great plate count anomaly* (i.e., the discrepancy between the sizes of populations estimated by dilution plating and by microscopy) [127], the *catch-22* situation of the 'unculturable' microbes left them no choice but to ignore those inaccessible organisms.

By the mid 1980s, however, it became apparent that uncultured microorganisms have significant effects on ecosystems [128] and human wellbeing [129, 130], and could no longer be overlooked. However, without a fundamental change in methodology, microbiologists could do little except to

acknowledge the existence of uncultured microbial life. Indeed, the field of microbial ecology was in desperate need for a radical change in methodology.

A paradigm shift in microbiology took place in 1977, when Woese, using information derived from 16S/18S ribosomal RNA (rRNA), established the Archaea branch of life and drastically redefined biologists' view of the biosphere [131]. 16S/18S rRNA allows the classification of microbial life to be based on genetics rather than morphology; its ubiquitous nature, functional consistency, and relative ease of isolation have made it a useful and common molecular chronometer and identifier with a strong evolutionary significance [132].

Another paradigm shift happened in 1985, when several papers describing a novel experimental approach to microbial ecology were published by Pace and colleagues, in which they described direct analyses of 5S and 16S rRNA sequences from the environment [95, 133]. This new technique enabled scientists to describe the diversity of microorganisms in an environment without the need to culture and opened a new page in microbiology that would eventually lead to direct investigation of environmental genomes, or *metagenomics*.

II. Decipher the Undecipherable

Although Pace et al. demonstrated the feasibility of working directly with uncultured environmental samples, the experimental work was technically demanding, requiring direct sequencing of rRNA or its reverse-transcribed DNA counterpart [95, 133]. Therefore, the approach was not widely adopted until a

major technical breakthrough: the invention of the polymerase chain reaction (PCR), which enabled rapid and targeted amplification of DNA, and the identification of PCR amplification primers capable of amplifying almost the entire 16S ribosomal DNA (rDNA) gene [134]. Unaffected by the culturing bias, PCR-based phylogenetic studies (e.g., RFLP*, DGGE† [135]) revealed that the uncultured majority of a microbial community is highly diverse and quite often phylogenetically distant from the cultured minority [126].

The use of 16S rDNA phylogenetic analysis has greatly enhanced our understanding of the microbial world. In 2003, 52 phyla had been established within the *Bacteria* domain alone, a remarkable increase from 12 phyla in 1987 [136]. Also significant was the fact that only 26 phyla out of the 52 had cultured representatives [136], highlighting the main problem with 16S rDNA phylogenetic analysis; namely, the lack of any direct information on the physiology and biochemical capabilities of the targeted organism(s). Although phylogenetic information may in some cases contribute to the culturing efforts (e.g., quantitative assessment of enrichment efforts using organism-specific DNA probes), it is difficult to accurately predict whether an organism of interest may have special nutritional requirements or occupy an ecological niche, etc based on its 16S rDNA information.

Although phenotypic information cannot be inferred from phylogeny, much can be learned about an organism's lifestyle and metabolic/biosynthetic capabilities from its complete genomic sequence [137]. Yet, until very recently

* Restriction Fragment Length Polymorphism
† Denaturing Gradient Gel Electrophoresis

[138], one of the prerequisites for sequencing a genome was the isolation and culturing of the organism, precluding access to unculturable microbes. Furthermore, to understand the environmental roles and capabilities of complex microbial communities, a collective and simultaneous approach is required.

Metagenomics is by far the most promising method for comprehensively profiling microbial communities and has lead to some very significant findings [97, 98]. Coined by Handelsman and colleagues, the term 'metagenomics' captures the idea of collectively and simultaneously analyzing a community of microorganisms using their genetic information [126, 139]. Depending on the approach taken, metagenomics can be used to investigate specific traits of a community (e.g., enzyme activity [140, 141], antibiotics [142, 143]) or to produce an overview of the functional roles of each community member [97, 98]. After relying on culturing techniques to study microorganisms for more than a century, microbiologists now have a tool that allows them to not only observe, but also characterize and analyze heterotrophs, autotrophs, and obligate parasites alike.

III. The Metagenomic Approach…in Its Present State

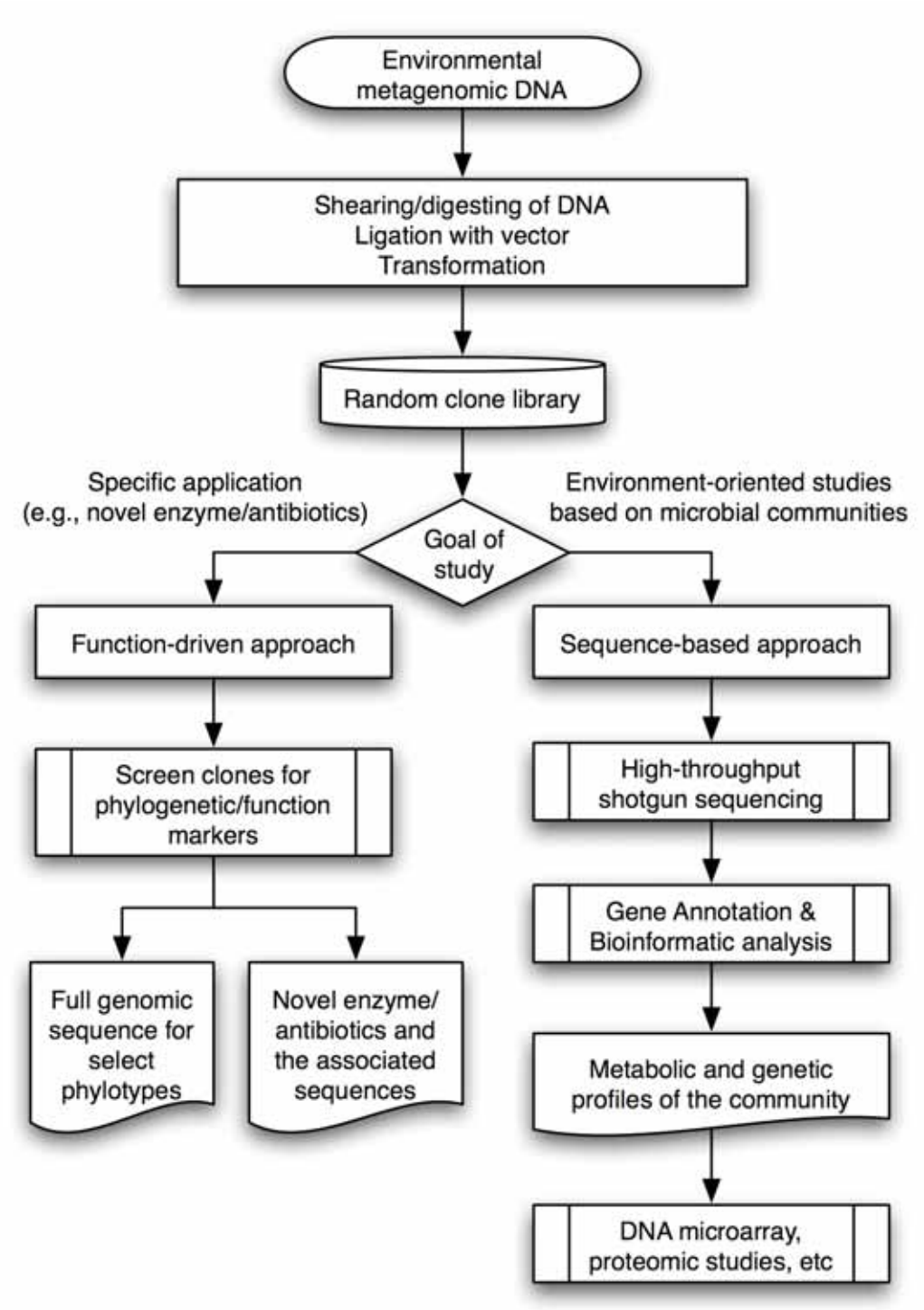
Broadly speaking, a metagenomic analysis begins with the isolation of DNA from the environment, cloning the metagenomic DNA into a suitable vector (e.g., BAC[‡]), and transforming the clones into a host organism (e.g., *E. coli*) to construct shotgun libraries of clones carrying fragments of environmental

[‡] Bacterial artificial plasmid

genomic DNA. From this point on, however, experimental approaches diverge and, depending on the objectives of the investigation, fall into two categories: function-driven and sequence-based metagenomics (see Figure 4-1).

Function-driven Analysis. An metagenomic approach often used for specific applications is to first screen the clone library for the presence of certain genes or activity markers using DNA hybridization, multiplex PCR [144], or biological activity (e.g., enzyme) assays [140-143]; the selected clones are then subjected to complete sequencing in order to obtain the gene(s) of interest and/or to establish the phylogeny of the clones [126] (see Figure 4-1). As this approach is invariably linked to certain functions or genes, it can be generalized as the 'function-driven approach'. This approach has several advantages. If a multiple-copy phylogenetic marker (e.g., the rDNA operon) can be identified in the selected clones, it may be relatively easy to identify and sequence clones belonging to the same species and obtain its complete genomic sequence [126] (provided that the target organism is relatively abundant and phylogenetically distant from other members of the community, see Figure 4-1). Secondly, by using enzyme activity assays as a screening tool, it is possible to identify clones with novel enzymes/antibiotics without any prior knowledge of their sequences [140, 141]. The technique is not without its limitations, particularly in the challenge of achieving heterologous expression [126]. In addition, the likelihood of shotgun clones expressing activity of interest can be diminishing low, making this technique economically impractical in many cases [145]. Nevertheless, functional screening of metagenomic clone libraries is arguably the most efficient way to assign definitive roles to 'hypothetical proteins' [126].

Figure 4-1 – Different approaches of metagenomics: experimental methods and potential applications.



Metagenomic investigations can be assigned to one of the categories: function-driven approaches or sequence-based approaches, each with their perspective strength and applications.

Sequenced-based Analysis. As the cost of DNA sequencing decreases and our ability to handle vast quantities of genomic data improves, a new approach to metagenomics has emerged. Metagenome-wide sequence-based metagenomic analysis (aka whole-genome shotgun sequencing, WGS sequencing), which utilizes parallel shotgun sequencing of entire clone libraries, is potentially the most informative approach [96-98, 146]. Not only does it provide information on the functional capabilities and genetic diversity of a community, it also sheds light on the distribution and redundancy of functions within the community. Additionally, it is the most effective method for uncovering genes/pathways of novel functions [98]. In cases where the community structure is relatively simple and phylogenetically distinct, WGS sequencing might even allow genome reconstruction of the entire community *in silico* [97]. Due to the wealth of information available with such approaches, significant hypotheses can often be inferred from the analysis, which can then be tested using genetic, proteomic, or even geochemical methods [94, 97] (see Figure 4-1).

Processing Metagenomic Sequence Data. The amount of DNA sequences generated for a typical sequence-based metagenomic study ranges from 25 Mb [96] to over 1.5 Gb [98], and is likely to increase substantially with the availability of novel sequencing technologies [147]. Therefore, one of the greatest challenges associated with WGS sequencing is that of bioinformatics. Apart from traditional approaches such as open reading frame (ORF) identification, genome assembly, and phylogenetic analysis, some new problems need to be solved with fresh approaches; for example, how to separate sequences from different organisms *in silico*, how to model population structures, and how

to compare different communities [93]. Since most existing programs have been designed for single genome applications, *ad hoc* heuristics must also be employed to adapt them for metagenomic studies [93]. Furthermore, programs designed to process metagenomic data must face additional challenges such as elevated intraspecies polymorphism (due to the potential presence of multiple strains of the same species in the metagenome), highly conserved genes across the community, truncated ORFs, low sequence coverage, etc [93].

Other Approaches. Additional experimental approaches complementing metagenomics have been examined (see Figure 4-1), such as the use of DNA microarray [148], proteomic investigations [94], and functional characterization of enzymes obtained from a WGS sequencing project (described in Chapter Five). The use of DNA microarray for metagenomic profiling has so far been limited to high-throughput taxonomic classification of shotgun clone libraries [148], but this approach may have great potentials at revealing *in situ* gene expression information for environmental studies (see section III). A 'shotgun' mass spectrometry-based proteomic study conducted in connection to the acid mine drainage metagenome project provided interesting information on gene expression and adaptations of the community [94, 97]. As the WGS sequencing approach becomes more widely adopted, new methods have been devised to take advantage of the wealth of sequence data characteristic of such projects. The latter part of this chapter describes one such method, which was designed and validated for this thesis.

IV. The National Science Foundation Biocomplexity Project

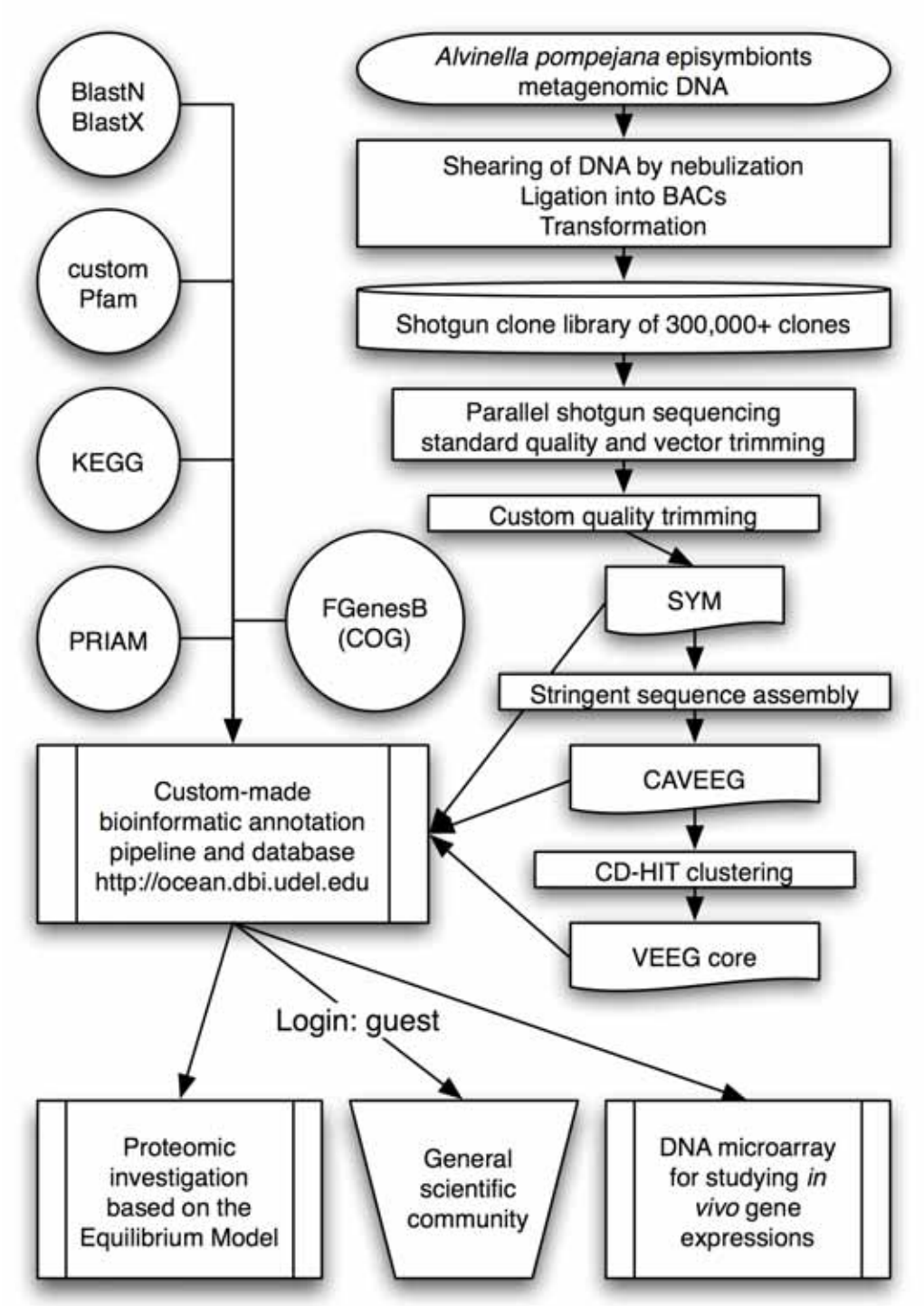
The Target Organisms. The National Science Foundation Biocomplexity project is a metagenomic scale analysis of the complex microbial community intimately associated with the dorsal surface of the deep-sea hydrothermal vent polychaete, *Alvinella pompejana*. It has been established that this microbial community (*A. pompejana Proteobacteria* episymbionts, APS) is predominantly composed of filamentous bacteria belonging to the epsilon subdivision of class *Proteobacteria* (ε-*Proteobacteria*) [81, 82]. This complex microbial community of ε-*Proteobacteria* is unique in that all the community members are closely related while at the same time being highly diverse [81, 82]. The symbiosis between *A. pompejana* and its episymbiotic bacterial consortium takes place at active chimneys at deep-sea hydrothermal vents along with East Pacific Rise, an ecological niche known for its high thermal and chemical complexity [67, 74]. As a result, *A. pompejana* and its episymbionts may be exposed to one of the most thermally extreme environment known to science [67]. A metagenomic investigation of this microbial community provides a unique opportunity to study various environmental adaptations associated with this ecosystem.

Hypotheses. The project was founded on five hypotheses: 1. under the intense constraints imposed by the geochemical environment, the episymbiotic consortium employs a congruous basic metabolic strategy (core metabolism); 2. the episymbiotic association can be defined as a mutualistic partnership enabling survival in this hostile environment; 3. the episymbiont community employs an extensive range of sensing capabilities to change and respond to their dynamic

environment; 4. the episymbiont community detoxifies the harsh chemical environment experienced by the worm; 5. protein eurythermalism is a common adaptation of the *A. pompejana* episymbionts.

Experimental Approaches. To investigate these hypotheses, various experimental approaches have been undertaken (see Figure 4-2). Small-insert shotgun clone libraries of episymbiont genomic DNA were first constructed, followed by high throughput parallel sequencing of these clones by SymBio Corp (Menlo Park, CA, USA). The resulting sequences were then quality-screened and pooled into a metagenomic sequence database. A custom-made bioinformatic pipeline (see Figure 4-2) was used to annotate and organize the database to facilitate further analyses, including defining a core metabolism, community phylogenetics, comparative genomic studies, functional gene distribution analysis, evaluation of the community's sensing capabilities, assessment of the community's detoxification mechanism, etc. For a proteomic investigation, suitable genes were identified in the metagenome database, cloned, and overexpressed to produce functional enzymes (see Figure 4-2). Those enzymes were then characterized using the Equilibrium Model in order to assess their eurythermal potentials (described in Chapter Five). Furthermore, a DNA microarray-based analysis based on information obtained from the bioinformatic pipeline is currently underway (see Figure 4-2).

Figure 4-2 – Information flow and various components of the NSF Biocomplexity project.



The Biocomplexity metagenomics project incorporates multiple experimental approaches, including DNA microarrays and proteomic investigations.

Metagenomic Application of DNA Microarray. DNA microarrays based on identified ORFs of interest from the database will be hybridized with cDNA preparations generated from total RNA extracted from APS samples to examine *in vivo* expression levels of individual genes and pathways (see Figure 4-2). In contrast to a similar study based on proteomics [94], DNA microarrays have the ability to detect low-level expressions and allow comparative studies using total RNA extracted from different sections of the worm or worms from different colonies, etc with relative ease.

Sample Collection and Preparation. *A. pompejana* specimens were collected from colonies at 9ºN and 13ºN of the East Pacific Rise during the annual Extreme oceanographic cruises between 2001 and 2004 and some earlier cruises. Geochemical samples and cinematographic data were also collected during sampling. Total genomic DNA samples were collected from the episymbiotic biomass of worms (in 1995 and 2001) and a single worm (AP201, aka 'Wayne') collected at Bio9 vent (9ºN East Pacific Rise) during the Extreme 2002 cruise, and made into several large clone libraries. For details on *A. pompejana* sampling and genomic DNA extraction, see Appendix IV. The genomic DNA samples were confirmed to be free of eukaryotic contamination from *A. pompejana* by PCR-screening of 18S rDNA and qPCR-screening of ß-actin, both eukaryote-specific genes. The genomic DNA samples were then nebulized, size selected by pulse field gel electrophoresis (see Figure 4-2), and cloned into a high-copy vector (pPCR4-TOPO™ from Invitrogen® or pSmart™ from Lucigen®).

a. The *A. pompejana* Episymbiont Metagenome

Basic Statistics. 96-well (from pools of worm) and 384-well (from Wayne) clone libraries were constructed from extracted metagenomic DNA, resulting in over 300,000 shotgun clones containing 1.5-3 Kb inserts. Libraries derived from Wayne accounted for more than 90% of the clones. The libraries were picked, and 180,000 clones from the clone libraries were sequenced in both forward and reverse directions, resulting in over 377,326 quality-filtered sequences (see Table 4-1 and Figure 4-2). A basic description of the metagenome can be found in Table 4-1. The sequences then went through another round of stringent quality trimming[§], and 301,361 sequences were selected, adding up to a total of 134.4 Mb of high quality, vector-trimmed, sequence data (around 66% of the original data) (see Table 4-1). These sequences were collectively referred to as the 'SYM' data set (see Figure 4-2). The average length of sequence reads is 446 bp, and the overall GC content of the metagenome is 35% (Table 4-1), suggesting that individual members of APS are likely to have AT-rich genomes.

Sequence Assembly. Stringent assembly (95% nucleotide match) of the sequences yielded 26,339 contigs and 102,831 unassembled singletons (see Table 4-1). The contigs had an average length of 1,463 bp, smaller than the average length of inserts in the shotgun clone library; this in turn implied that there were too many gaps for large scaffolds to form, and that the metagenome would have to be assessed using clustering and redundancy

[§] First, a 100-bp cut-off was applied to the sequences, then sequences were examined through a moving 11-base 'window' and discarded if a pre-determined quality threshold was not met. Only the longest high-quality contiguous portion of a sequence was kept (personal communications, S. C. Cary).

analysis. This assembly was named the 'Complete Assembled Vent Epibionts Environmental Gene' data set (CAVEEG) (see Figure 4-2).

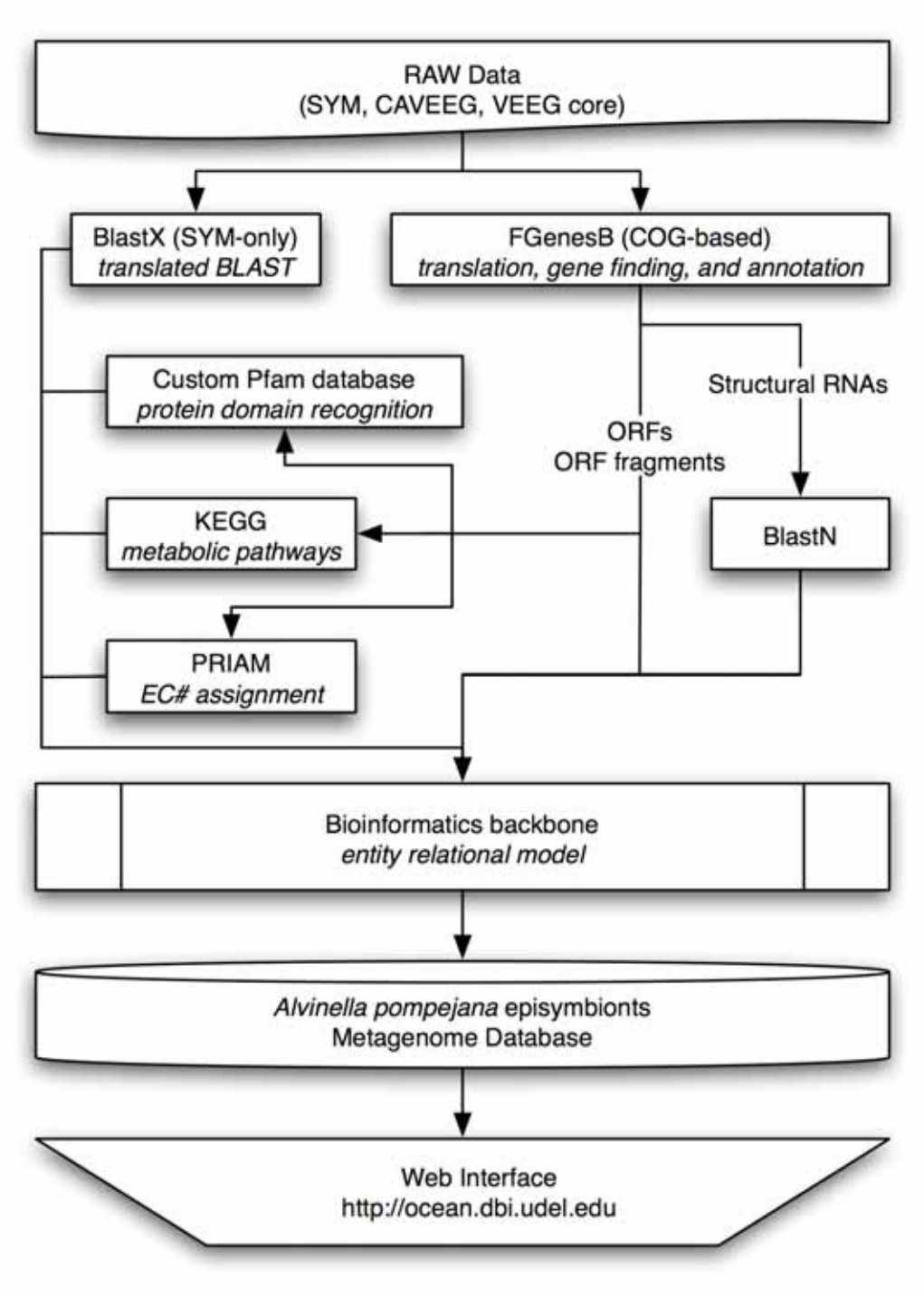
Table 4-1 – Basic statistics for the Biocomplexity metagenomics project.

Unassembled Data	
Total sequences	377,326
Sequences retained after custom quality trimming	301,361
Total bases after quality and vector trimming	134.4 Mb
Genome equivalents retained (2.5 Mb)	54 at 1X
Average trimmed sequence length	446 bp
Metagenome GC content	35%
Assembled Data	
Total sequences assembled into contigs	198,534
Number of contigs	26,339
Singletons (unassembled sequences)	102,831
Average contig length	1,463
Total bases assembled	85 Mb
Predicted genes	103,371
SSU ribosomal DNA genes (16S)	484
LSU ribosomal DNA genes (5S and 23S)	566
tRNA genes	918

Special Challenges. A custom-made annotation pipeline was developed for the Biocomplexity project, and Figure 4-3 shows a flow diagram of the pipeline. The bioinformatic component had to be customized for the Biocomplexity project due to the uniqueness of the community under study, namely a high level of diversity exists among very closely related community members. The similarity in genomic sequences among distinct community members implies that complete genomes of the dominant members are unlikely to be obtained, and that the metagenome must be analyzed collectively. In addition, since the community is dominated by a few phylotypes, sequences derived from them are likely to make up the bulk of the metagenome data, and special attention must be paid to identify the core metabolism potentially associated with these dominant phylotypes.

Annotation Pipeline. The annotation pipeline was based on the FGenesB package (SoftBerry) for ORF finding and COG-based gene annotation (Clusters of Orthologous Groups [149]) (see Figure 4-2 and Figure 4-3). Coding genes were also matched to several custom-made Pfam databases [150] for further protein domains-based annotation, queried against PRIAM [151] for EC number assignments, and mapped onto KEGG metabolic pathways [152] for easy access (see Figure 4-3). All sequences were also BLAST searched against GenBank databases for additional information. For a brief overview of various databases, see Table 4-2.

Figure 4-3 – The annotation and bioinformatic analysis pipeline.



The sequence annotation pipeline of the Biocomplexity project utilizes the strength of multiple reference databases to maximize annotation coverage. Parallel analysis of multiple data sets also allows comparisons and additional insight on the structure of the episymbiont community.

Table 4-2 – Overview of various bioinformatic reference databases.

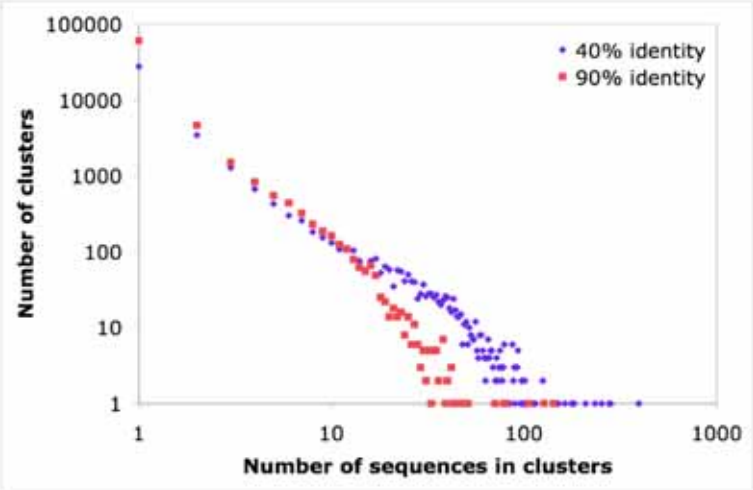
Name	Full Name	Description
COG	Clusters of Orthologous Groups	COG is a database of protein found in complete genome sequences organized based on functional categories and taxonomic information. It is gene-oriented rather than enzyme-oriented, and thus cannot be directly related to EC numbers.
KEGG	Kyoto Encyclopedia of Genes and Genomes	A comprehensive database of genes, enzymes, metabolic pathways, and organisms. KEGG is built upon various metabolic pathways found in organisms and provides a visual representation of all components involved in a metabolic process. The components themselves are also described in details at both gene and enzyme level, including information such as gene name, COG entry, catalyzed reaction, structure, and links to other databases. The pathways can also be limited to a specific organism, allowing easy examination of the status of the pathway in that organism.
Pfam	Pfam	Pfam contains information about protein domains and families. The information is based on multiple sequence alignments and Hidden Markov Models of known proteins.
PRIAM	PRofils pour l'Identification Automatique du Métabolisme (Enzyme-specific profiles for metabolic pathway prediction)	PRIAM is a database of position-specific score matrices (PSSMs) calculated from characterized homologous proteins from SWISS-PROT. It was specifically designed to automate the enzyme annotation process for genome sequencing projects. Its annotation is based on EC numbers.
KGB	Keyword General Browser	A comprehensive tool built into the Biocomplexity metagenome database website to enable keyword-based searches against all annotation fields included in the database.

Clustering Analysis. To evaluate the diversity of predicted proteins in the metagenome, CAVEEG was clustered at 40% amino acid identity[**] using CD-HIT [153]; this resulted in 36,221 clusters, of which 27,583 were singletons (see Figure 4-2 and Figure 4-4). The five largest clusters made up more than 1% of the predicted ORFs (see Figure 4-4), and they all appeared to be plasmid replication initiation proteins. This suggested a very high level of functional diversity in the data set. Clusters containing more than five sequences were grouped together, and the resulting non-redundant subset of CAVEEG was referred to as the 'VEEG core' data set (see Figure 4-2), containing 2,814 clusters of predicted genes. The purpose of this data set was to identify the 'core metabolism' of the microbial consortium.

[**] Amino acid sequences that cluster at 40% identity are considered to be functionally homologous.

Figure 4-4 – Gene clustering analysis of the 'Complete Assembled Vent Epibionts Environmental Gene' data set (CAVEEG).

(*Credit: Dr. J. J. Grzymski, Desert Research Institute*)

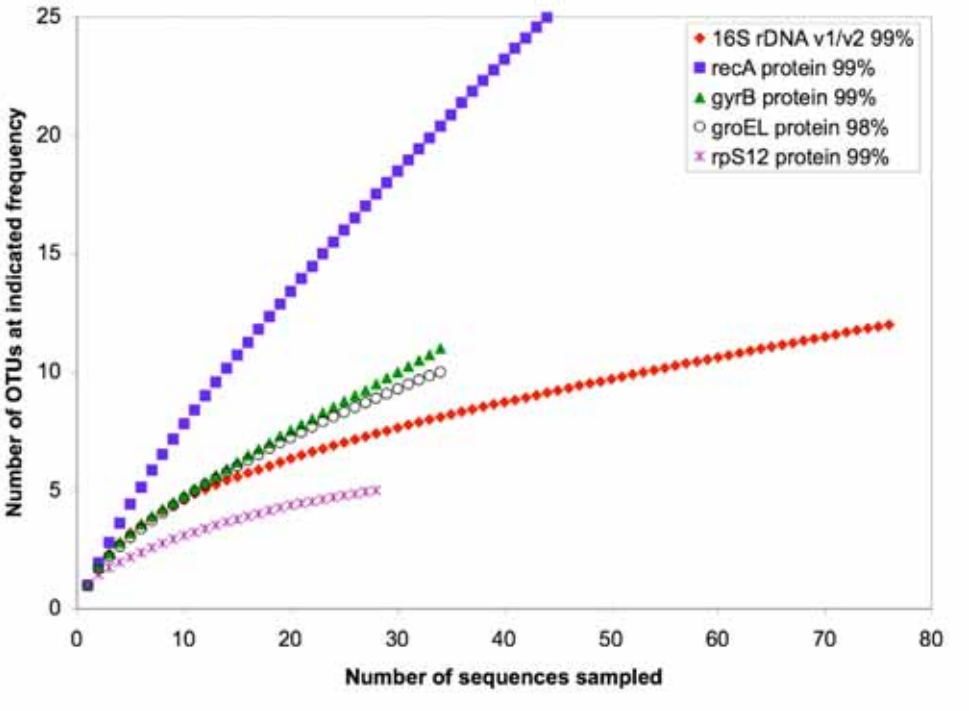


CD-HIT was used to cluster translated gene sequences based on the level of overall amino acid identify. 36,221 clusters were formed at 40% aa identity. 60,494 clusters were formed at 90% aa identify. Clusters containing more than five sequences were pooled to form the 'VEEG core' data set, which was used to identify the 'core metabolism' of the episymbiotic community.

Phylogenetic Diversity. In order to assess the phylogenetic diversity of the metagenome, seventy-six 16S rDNA sequences containing the hypervariable v1/v2 regions were retrieved from the SYM data set and analyzed with DOTUR, a species richness estimation program based on rarefaction analysis [154]. At 99% sequence identity, at least ten distinct ε-*Proteobacteria* phylotypes were found (see Figure 4-5), which largely corresponded to similar analyses based on the amino acid sequences of *groEL* and *gyrB* (see Figure 4-5). However, this result reflected only the diversity in the sequenced metagenome, and the true diversity of the community is likely to be higher, which may be partly due to short sequence reads resulting in a small sampling region, since PCR libraries of 16S rDNA from the episymbionts hinted at higher diversities [81, 82]. Since the community is dominated by a few phylotypes [81, 82], it would require substantial additional sequencing efforts to uncover the minor phylotypes in the metagenome.

Figure 4-5 – Rarefaction analysis of the v1/v2 region of metagenomic 16S rDNA sequences compared with several ubiquitous housekeeping genes.

(***Credit: Dr. B. J. Campbell, University of Delaware***)



OTU: operational taxonomic units. The analysis was performed using data from CAVEEG, based on alignments/distance matrix taking one representative clone from each contig. DOTUR was used to generate the rarefaction curves at the indicated % similarity , which were estimated from results of similar studies.

b. Community Characteristics and Comparative Analysis

This subsection describes unpublished findings from the Biocomplexity metagenome project that have been compiled from results provided by various participants of the project. Some of these results may be relevant to interpreting data presented in the thesis, and some were directly derived from my contributions to the bioinformatic analysis of the metagenome data (see the next subsection for details).

Comparative Metagenomics. In order to conduct a comparative analysis, a combined whale fall metagenomic sequence data set [96] was retrieved and put through the custom annotation pipeline. A combined ε-*Proteobacteria* data set was also constructed using COG assignment data of seven ε-*Proteobacteria* genomes from GenBank (*Helicobacter pylori* J99 and 26695, *Helicobacter hepaticus*, *Campylobacter jejuni* RM1221 and NCTC11168, *Wolinella succinogenes*, and *Thiomicrospira denitrificans*) and analyzed with the custom annotation pipeline.

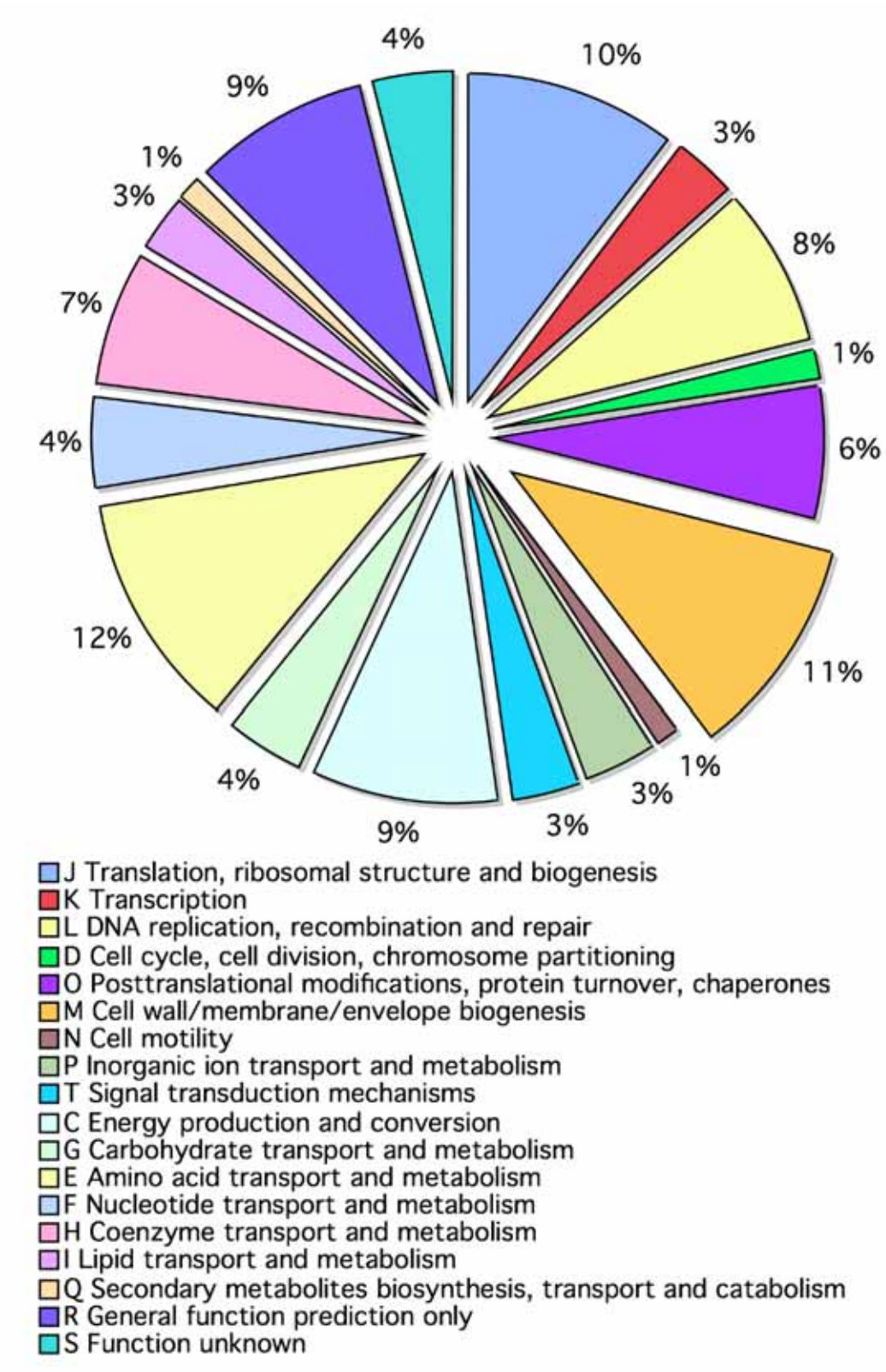
41% of genes in CAVEEG were assigned to at least one COG functional category, compared to 63% of VEEG core; both are higher in the percentage of ORFs not assigned to any COG functional category than the combined whale fall metagenome data, for which 55% of the predicted genes fell into one or more COG functional category. One example of the similarities between VEEG core and the combined ε-*Proteobacteria* data set is the underrepresentation of transcriptional genes (1.5% in VEEG core, 2.6% in combined ε-*Proteobacteria* data set, and 4.9% in all available bacterial

genomes). On the other hand, the combined ε-*Proteobacteria* genomes have relatively few genes assigned to cell wall, membrane, envelope biosynthesis proteins, and far more genes involved in cell mobility, for example. In addition, the episymbionts have the unusual ability to synthesize vitamin B12 and other essential vitamins both aerobically and anaerobically. In fact, many pathways in the metagenome were found to exist in both aerobic and anaerobic forms (e.g., components of both the regular TCA cycle and the reverse TCA cycle were found). This is consistent with the intermittent availability of various electron acceptors and donors in the vent environment [74, 75, 77].

COG-based Functional Analysis. Findings from the annotated data set correspond to known physiological and biochemical properties of the community [71, 75, 77, 81]. Based on the COG functional category assignments, the CAVEEG data set is rich in cell wall, membrane, and envelope biosynthesis proteins (see Figure 4-6). This is not surprising since in their native habitat the episymbionts are protected by a complex sheath of biofilm-like structure, which is likely to have been synthesized by the bacterial community [75, 77]. Many genes involved in metal detoxification are well represented in the metagenome, such as *Czc* proton/cation antiporter (Cd^{2+}, Zn^{2+}, Co^{2+} and Ni^{2+}) and the *Ars* chemiosmotic efflux pump (arsenate/arsenite). All cytochromes required for oxidative phosphorylation are present, with a few in high abundance, suggesting that oxidative phosphorylation is a prominent function of the episymbiont community. All enzymes involved in respiratory nitrate reduction are also present in very

high abundance, suggesting that in addition to oxygen, nitrate and/or nitrite are likely to act as terminal electron acceptors; the presence of dissimilatory bisulfite reductases among the symbionts further hints that the community is at least partially autotrophic. Furthermore, mandatory genes for the reverse TCA cycle (ATP citrate lyase and 2-oxoglutarate synthase) can be found in high abundance, lending credence to the ε-*Proteobacteria* episymbionts' proposed role as primary producers in the ecosystem [89].

Figure 4-6 – Clusters of Orthologous Groups (COG) functional group distribution of assigned sequences in the 'Complete Assembled Vent Epibionts Environmental Gene' data set (CAVEEG).



c. Proteomic Investigation of the Metagenome

Traditionally, the term 'functional metagenomics' refers to the functional analysis of heterologously expressed genes in metagenomic random clone libraries [126]. It relies upon the successful heterologous transcription and translation of gene(s) of interest, and has commonly been used to identify pre-selected genes with well-defined functions (e.g., antibiotic resistance genes and degradative enzymes). Although this approach has seen some interesting discoveries, it is often plagued by low positive response rate and largely limited to targets that can be effectively screened [126]. Therefore, functional metagenomics, as it was traditionally defined, has not been widely adopted by microbiologists as a tool for understanding the metabolic framework of a microbial community.

While insights into the cellular machineries of a microbial community can be gained from examining its metagenome, the gathered information is not necessarily a reflection of *in vivo* activity. An effort was made to corroborate acid mine drainage metagenomic data with those obtained through mass spectrometry-based proteomic methods [94]. The study was able to identify highly expressed components of the community and correlate the findings with metagenomic data, highlighting the community features needed for adapting to the special chemical environment.

The ε-*Proteobacteria* episymbiont community associated with *A. pompejana* puts forward another layer of complexity, namely the extreme thermal environment surrounding the organisms. Although metagenomic studies have revealed elements of the community related to its thermal

adaptation (e.g., biofilm formation), the evidence so far is indirect. Since temperature adaptation of an organism ultimately depends on the adaptation of its enzymes, proteomic investigations are essential for understanding the unique thermal adaptations of the episymbiont community. While there have been suggestions of thermophile-specific amino acid usage patterns, the studies have mostly been conducted on a genome-wide basis and focused on identifying overall trends in the genomes rather than predicting thermal properties of individual proteins [155-157]. Furthermore, the APS community is distinguished for its potential eurythermal features, and to date the only way to quantitatively describe and compare enzyme eurythermalism is with the Equilibrium Model [158].

Although the idea of isolating enzyme-coding genes directly from metagenomic libraries is not new, the screening process has so far been limited to performing PCR with primers based on highly conserved regions of known homologous genes [142]. Clearly, the requirement of gene conservation at the nucleotide level is the greatest shortcoming of this approach and severely limits its usefulness for poorly cultivated branches of microbial life. With the availability of metagenomic sequence data, it is now possible to perform *in silico* screening of genes of interest and design cloning primers based on sequence data readily available. This eliminates the guesswork involved in the selection of target genes and allows researchers to focus on genes relevant to the community and/or other interests. This newfound ability, coupled with the provision of a pan-community metabolic pathway analysis such as that of the Biocomplexity metagenome

database, potentially allows researchers to identify important pathways of the community and quickly access the sequences of all the genes involved. Surveys of this type, which in the past had taken decades [159], can now be conducted at the community level in a drastically reduced timeframe. Although a similar approach can be and has been directly applied to complete genomes of individual organisms [160, 161], the widely discussed unculturability of most microorganisms somewhat diminishes the ecological significance of such studies.

d. A Synopsis of My Contributions

In addition to the gene-mining studies described in this chapter (see section V), I have been involved in and have contributed to various aspects of the Biocomplexity project. Below is a brief account of my participation in the project and specific contributions.

In April 2004, I participated in the Biocomplexity project meeting held at Desert Research Institute in Reno, Nevada, USA. For the two following months, I worked at Prof. Cary's laboratory in College of Marine Studies, University of Delaware. During this period I performed some analyses on the metagenome and tested the metagenome web interface; some preliminary work on gene mining and cloning was also conducted.

From November to December 2004, I took part in the Extreme 2004 expedition (http://www.ocean.udel.edu/extreme2004/) during which additional *A. pompejana* specimens and geochemical data were collected.

Since the entire metagenome project team was onboard, various discussions and analyses were also part of the activities during the cruise.

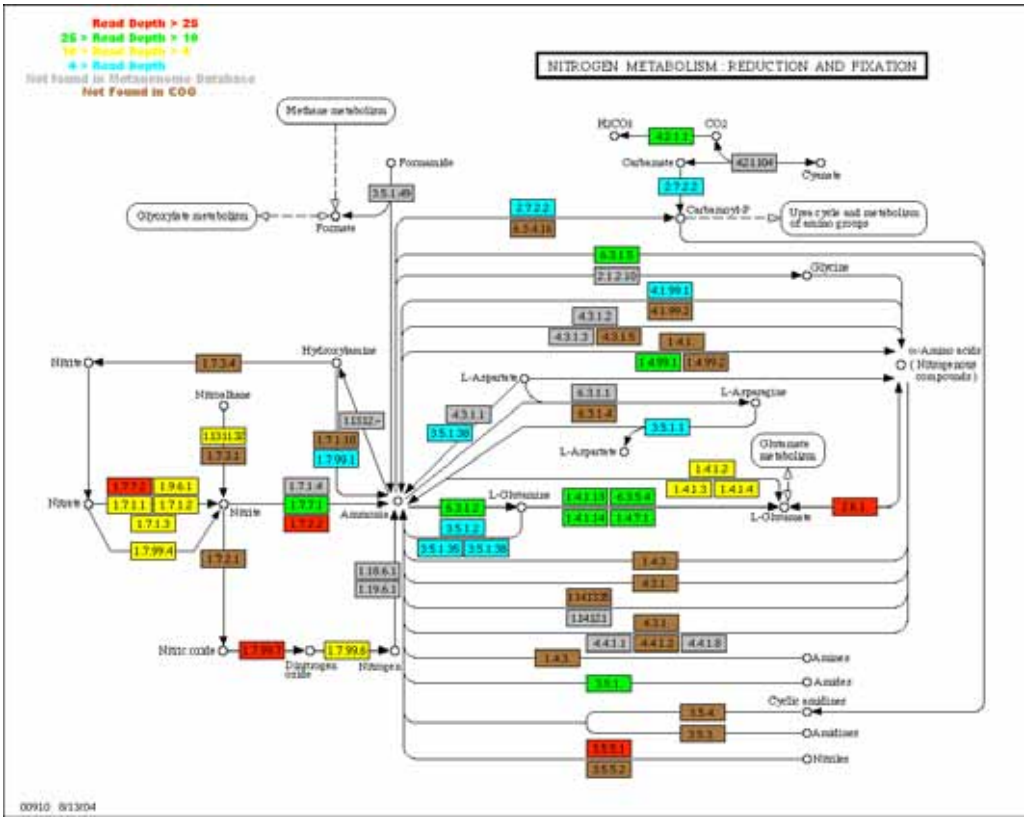
Since coming onboard the Biocomplexity project in November 2003, I have participated in numerous conference phone calls and email communications with the rest of the Biocomplexity project members.

From April to June 2005, I worked close to full time on testing the metagenome database and took active roles in designing and carrying out various analyses. I designed and constructed a bioinformatic pipeline for extracting and compiling COG-related data from clustered CAVEEG and VEEG core data sets. As a direct result of this bioinformatic pipeline, we were able to assign relative abundance information to individual COG entries and use the information to analyze the status of various KEGG metabolic pathways in the metagenome. I personally carried out such analyses on the following pathways: nitrogen metabolism, oxidative phosphorylation, two component system/quorum sensing, cell mobility, and detoxification. The results of these analyses (see Figure 4-7 for an example) have been presented by Dr. Alison Murray at the 105th general meeting of the American Society for Microbiology in 2005, and by Prof. Cary at various conferences and meetings. In addition, findings from these analyses have been consolidated and will contribute both directly and indirectly to the research article reporting the metagenome.

From November 2005 to January 2006, I worked full time on investigating genes of interest using the metagenome database using a protocol designed by myself, and during this process I provided significant

feedbacks on the metagenome database itself and the web interface. For details of this investigation, see section V of this chapter.

Figure 4-7 – An example of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping.



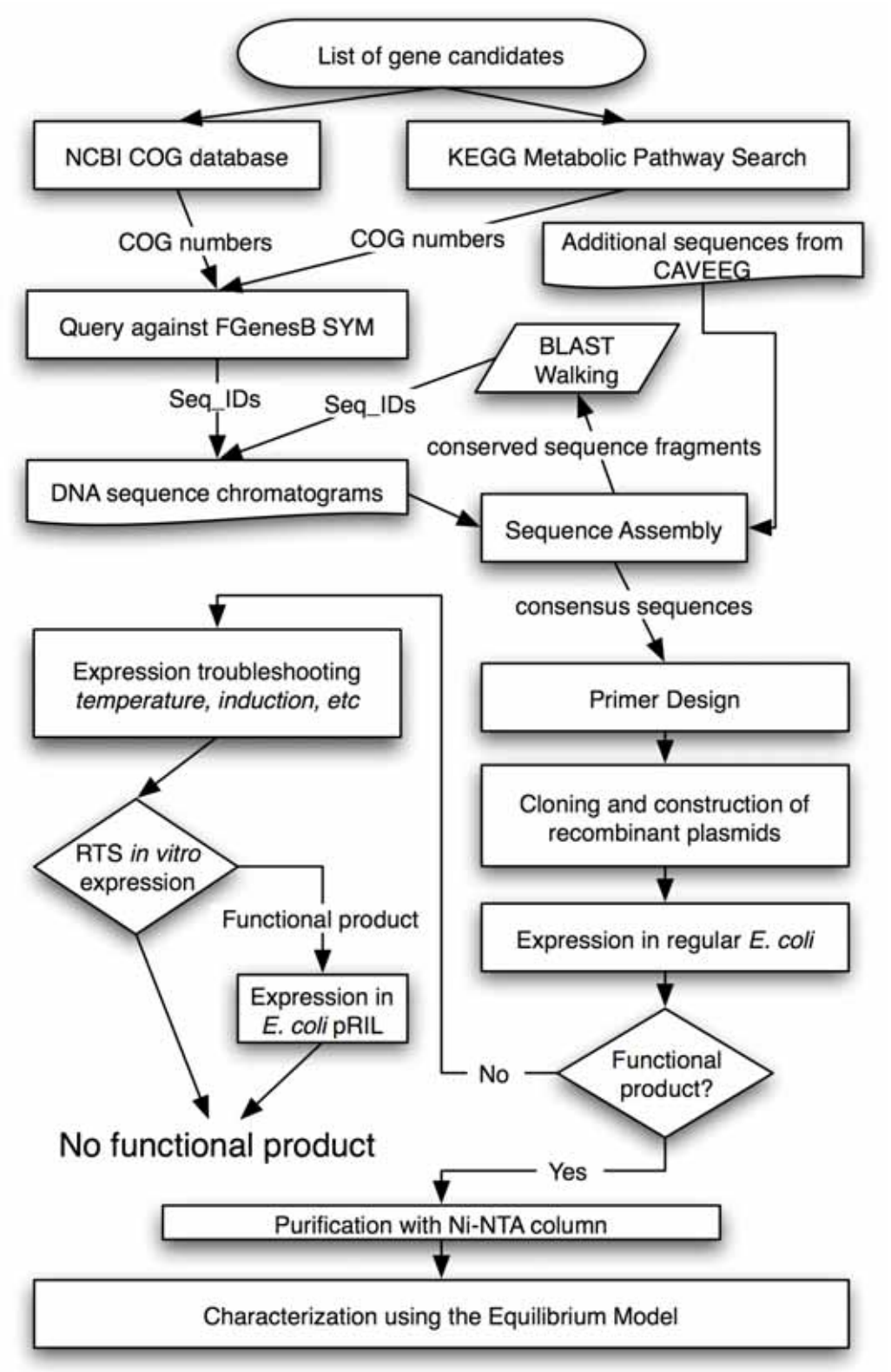
Manually annotated genes were color-coded based on their status in the metagenome. An automated implementation of this kind of pathway presentation has been developed for the Biocomplexity metagenome database, and findings derived from similar graphs that I produced had been presented in various conferences.

V. Mining the Metagenome Database

The main goal of this thesis within the Biocomplexity project is to characterize enzymes originated from the metagenome using the Equilibrium Model [1, 2]. The metagenome data set contains a staggering number of potential genes, many of which code for enzymes crucial to the functioning and survival of the community. However, due to the practical limitations of the Equilibrium Model (see Chapter Two), only a subset of those enzymes was considered for this work. Emphasis was also placed on identifying genes with noisy single nucleotide polymorphisms (SNPs), since proteins derived from those genes would have point mutations in their amino acid sequences, thus providing additional insights on the adaptive strategies of the community.

The general approach of the study is as follows (also see Figure 4-8). Sequences associated with target genes were identified using the metagenome database web interface and various bioinformatic tools, and PCR cloning primers were designed based on the consensus sequences (see Figure 4-8). PCR amplicons containing those genes were then cloned into a special vector designed for *in vitro* cell-free expression (pIVEX2.4d – Roche Applied Science), which incorporates a N-terminus His-tag for easy purification using Ni-NTA columns. The cloned genes were screened for activity in the cell-free expression system (RTS – Roche Applied Science) if necessary, or directly overexpressed in *E. coli* (see Figure 4-8). The expression products were then purified and characterized using the Equilibrium Model (see Figure 4-8).

Figure 4-8 – Data-flow pipeline map of a proteomic investigation of the *A. pompejana* episymbiont metagenome.



a. Gene-Mining Strategy and Results

A list of gene candidates can be found in Table 4-3. The first and foremost criterion for selecting target genes was the availability of a continuous direct photometric assay, since it gives more accurate results than discontinuous assays [20]. Subunit configuration of enzymes is an important factor when selecting target genes for this work. An enzyme with heterogeneous subunits requires significantly more efforts since all individual components have to be cloned and expressed, and *in vitro* assembly of heterogeneous subunits into functional enzymes often has unpredictable outcomes. Therefore, monomeric or homogeneous multimeric enzymes were the primary targets for this work. Physiological roles of the gene were also taken into consideration.

Table 4-3 – List of candidate genes, the associated Kyoto Encyclopedia of Genes and Genomes (KEGG) and Clusters of Orthologous Groups (COG) entries, and characterization status.

Enzyme Name	EC Number	Associated KEGG Ortholog	Associated COG Entry(s)	Number of hits in SYM	Status
Malate dehydrogenase	1.1.1.37	K00026	COG0039	181	*Cloned*
Isopropylmalate dehydrogenase	1.1.1.85	K00052	COG0473	125	**Cloned and Characterized**
Alanine dehydrogenase	1.4.1.1	K00259	COG0686	4	Insufficient number of sequences
Glutamate dehydrogenase	1.4.1.4	K00262	COG0334	48	**Cloned and Characterized**

Enzyme Name	EC Number	Associated KEGG Ortholog	Associated COG Entry(s)	Number of hits in SYM	Status
Dihydrofolate reductase	1.5.1.3	K00287	COG0262	1	Insufficient number of sequences
γ-Glutamyl transpeptidase	2.3.2.2	K00681	COG0405	4	Insufficient number of sequences
Lipase	3.1.1.3	K01046	COG1075	none	Not found in MGDB
Alkaline phosphatase	3.1.3.1	K01017	COG1785	1	Insufficient number of sequences
Acid phosphatase	3.1.3.2	K01078	COG0496	53	*Cloned*
α -amylase	3.2.1.1	K07405	COG1449	13	Insufficient number of sequences
α -Glucosidase	3.2.1.20	K01187	COG0366 COG1501	3 4	Insufficient number of sequences
β-Glucosidase	3.2.1.21	K01188 K05349 K05350	COG1472 COG2723	60 none	*Cloned*
β-Lactamase	3.5.2.6	K01467	COG2367 (Class A) COG1236 (Class B) COG1237 (Class B) COG1680 (Class C) COG2602 (Class D)	none 4 3 5 none	Insufficient number of sequences
Adenosine deaminase	3.5.4.4	K01488	COG0590 COG1816 COG2169	1 6 none	Insufficient number of sequences
Fumarase	4.2.1.2	K01679	COG0114	34	No significant ORF found

Enzymes in this list were selected on the basis of the availability of direct spectrophotometric assays and subunit configuration (no heterogeneous multimers).

- **An Overview of Gene-Mining Protocol**

In the current release of the metagenome database, the recommended method for identifying sequences associated with a particular gene of interest (GOI) is to use the *Keyword General Browser Services* (*KGB Service*), which allows direct searches using COG number, EC number, etc and returns results based on FGenesB, Pfam and BLAST annotation data (see Table 4-2 and Figure 4-3). Although both the *PRIAM portal* and the *KEGG Metabolic Pathway Search* allow genes to be rapidly located according to their metabolic functions, they are not relevant for this work since there is a predefined list of GOIs. Queries based on EC numbers consistently returned poor results in the KGB Service, and since the main annotation program, FGenesB, is based on the COG database (see Figure 4-3), COG entries were used as the primary query for gene mining.

Figure 4-8 is a flowchart for the gene-mining work described here and below. Using the corresponding EC numbers, GOIs were first located in KEGG Orthology [152], which in most cases provided a link to the corresponding NCBI COG entry. Searches were also performed in the NCBI COG database using gene names as queries to identify relevant COG entries not listed in KEGG Orthology. The COG numbers of the identified COG entries were then used to query the FGenesB SYM data set (see Figure 4-3). The FGenesB SYM data set (unassembled), rather than the FGenesB CAVEGG data set (redundant assembled), was used because the SYM data set provided direct

identification of all related sequence IDs (Seq_ID), which were required to retrieve DNA sequence chromatograms. Sequence chromatograms for each gene were assembled separately at high stringency using SeqMan (Version 5.51, DNASTAR Inc., Madison, WI, USA), and PCR cloning primers for each gene were designed based on the consensus sequences of the identified ORF(s).

- **Gene-Mining Methods and Results**

Not every GOIs returned a significant number of hits from the FGenesB SYM data set (see Table 4-3). Only GOIs with more than thirty retrieved DNA sequence chromatograms were chosen for PCR cloning primer design; these genes hypothetically encode the following enzymes: malate dehydrogenase (MDH, EC 1.1.1.37), 3-isopropylmalate dehydrogenase (IPMDH, EC 1.1.1.85), glutamate dehydrogenase (GDH, EC 1.4.1.4), acid phosphatase (ACP, EC 3.1.3.1), and β-glucosidase (ß-GLU, EC 3.2.1.21). Detailed descriptions of sequence assembly, editing, and primer sequences for these genes can be found in Appendix IV. Discussed below in the main text are general observations and issues encountered during the course of working with the assemblies, how they were overcome, and their implications.

Overall, there were several major issues related to the assemblies: 1. high sequence diversity (and consequently low depths for prominent contigs and large numbers of singletons in the assemblies); 2. gaps in contigs due to short sequence lengths; 3. since raw sequence

chromatograms files were used for assembly, in some cases errors in the unscreened sequences lead to poor sequence assembly and difficulties in obtaining a consensus sequence. The most important technique for overcoming these problems was 'BLAST-walking', which was developed for this study.

The concept of BLAST-walking is similar to that of chromosome walking. By taking a short, conserved region of the gene in question (the 'primer') and BLAST-searching it against a shotgun sequence metagenome, one essentially hybridizes the 'primer' to genomic DNA *in silico* (as opposed to *in vitro* hybridization of PCR primer to DNA in chromosome walking), allowing the identification of additional sequences that may either bridge an existing gap in the assembly or increase assembly depth over critical regions (see Figure 4-8). Nucleotide-nucleotide BLAST (blastn) is appropriate for this purpose, although it may be necessary to adjust the parameters of the program to optimize the match. For example, to BLAST-walk with a relatively short sequence, one might find it desirable to reduce the penalty for opening a gap but raise the penalty for extending a gap, or to reduce the word size for the BLAST query.

The idea, or indeed that term 'BLAST-walking' is not new [162], but the application described here is novel. Previously, BLAST-walking had been used to identify poorly conserved gene homologs in a complete or partially complete genomic sequence *in vitro*, whereas here it describes a new approach to *mining* metagenomic data *in*

silico. Although the technique requires a high level of human input, it was instrumental in the mining of the Biocomplexity metagenome database and should be considered for similar work in the future.

Short, highly diverse sequences typical of metagenome shotgun sequencing projects generally result in poor assembly performance. One way to deal with this problem was to assemble using a slightly less stringent match threshold (96% or even 91% as opposed to 99%, which is normally used for protein-coding genes from closely-related species) and a shorter overlap (50 bp rather than 150 bp or 200 bp as commonly adopted). One other method that can be considered is to reassemble the sequences after some sequences have been edited using information available in the original assembly. Quite often, this can significantly increase the coverage and depth of the prominent contig as many sequences that may have previously been rejected from the contig due to errors in existing sequences in the contig now join the prominent contig.

The corresponding consensus sequence(s) of a gene in CAVEEG can also be a helpful tool. In this gene-mining procedure, only Seq_IDs that have been annotated by FGenesB were included in the assembly, whereas CAVEEG was assembled from the entire SYM data set and is based solely on nucleotide sequence similarity (see Figure 4-2). Sequences that contain frameshift-causing errors may have been translated incorrectly by FGenesB and therefore not annotated, but such sequences would still be assembled with other annotated sequences in CAVEEG and can provide additional depth or coverage for the assembly

(see Figure 4-8). In addition, sequences of gene homologs from other ε-*Proteobacteria* were retrieved from GenBank and used to aid the assembly editing (identifying frameshift, incorrect start or stop codon, etc) of some of the GOIs (see Appendix IV).

b. Cloning and Overexpression of Genes of Interest

The cloning and overexpression strategy used in this study is similar to the strategy used to obtain *B. cereus* DHFR in Chapter Three (also see Figure 4-8), and details for the cloning, verification, overexpression, and purification of each gene can be found in Appendix IV and Appendix V. All five GOIs described in the previous subsection were successfully amplified, cloned, and verified (with the exception of ß-GLU, see below and Appendix IV), but only two of the GOIs (IPMDH and GDH) were successfully overexpressed and resulted in functional enzymes (see below and Appendix V for details).

- **Construction of Clones Containing GOIs**

All five GOIs were PCR amplified from a genomic DNA sample AP201 (Wayne) collected during the Extreme 2002 cruise; the same genomic DNA preparation was also the source of the majority of metagenomic sequences (see section IV). For details on sampling of *A. pompejana* and genomic DNA extraction, see Appendix IV.

The recombinant gene constructs were based on the Roche pIVEX2.4d expression vector, which is compatible with both a cell-free

in vitro expression system (Roche Rapid Translation System, RTS) and *in vivo* heterologous expression using an *E. coli* host carrying the DE3 genotype. The pIVEX2.4d vector also adds a N-terminus His-tag to the overexpressed product to enable rapid purification with Ni-NTA columns. For detailed descriptions of the pIVEX2.4d vector and RTS, see Appendix IV and Appendix V.

The GOIs were PCR amplified using PCR cloning primers under the conditions described in Appendix IV. The amplicons were analyzed with gel electrophoresis and confirmed to be the correct sizes. After the PCR amplicons were cleaned up (see Appendix IV), they were restriction digested with *Nco*I and *Sma*I following protocols described in Appendix IV. The digested fragments were then ligated to pIVEX2.4d vector digested with the same set of restriction endonucleases. The identities of the PCR amplicons were verified by sequencing, as described below.

The ligation mix was then transformed into Invitrogen™ One Shot® TOP10 competent *E. coli* (Invitrogen Corp, Carlsbad, CA, USA), and transformants containing the correct size inserts were selected with PCR-screening (see Appendix IV). Plasmids containing GOIs were then prepared using the selected transformants, and purified plasmids were verified again using the corresponding PCR cloning primers and sequenced.

With the exception of ß-GLU (possibly due to strong secondary structures in the recombinant constructs), all constructed clones were sequenced in forward and reverse directions to verify the identity of the

recombinant genes and to examine the presence of SNPs. The verified clone sequences were also BLAST-searched (blastx) against GenBank's protein sequence database to confirm their functions. Lastly, codon usage patterns of the clones were checked against that of *E. coli* with Graphical Codon Usage Analyser (http://gcua.schoedl.de/index.html). For details, see Appendix IV.

- **Expression of GOIs**

Expression candidates were chosen based on verified sequences of the clones (see Appendix V) and transformed into Invitrogen™ One Shot® BL21 Star™ (DE3)pLysS chemically competent *E. coli* (Invitrogen Corp, Carlsbad, CA, USA) for heterologous expression (see Figure 4-8). Functional products for two GOIs (IPMDH and GDH) were obtained and purified as described in Appendix V. The purity of these proteins was verified using protein gel electrophoresis (see Appendix V). Repeated attempts at overexpressing the remaining clones by altering incubating temperature, induction time and length, and inducer concentration did not result in any detectable expression product (see Appendix V).

Cell cultures were checked for signs of inclusion body formation; inclusion bodies were found inside cells carrying some of the ACP clones, but putative expression product could not be released from the inclusion bodies even under harsh denaturing conditions. Post-induction cultures were verified to have retained the recombinant

plasmids. Finally, clones of the remaining GOIs were tested with RTS and *E. coli* BL21(DE3)pRIL, a codon bias-correcting strain. No significant activity or signs of overexpression were observed following these attempts. For details on individual GOIs, see Appendix V.

c. Gene-Mining and Expression Discussion

One major hurdle during the gene-mining process is the short sequence length due to very stringent quality trimming; the average sequence length in SYM is 446 bp, shorter than sequences obtained in similar projects. For example, the average length for the acid mine drainage metagenome is 737 bp [97], and around 800 bp for the Sargasso Sea metagenome [98]. Apart from the challenges this problem created during the gene-mining process, it may also have had undesirable effects on the annotation process since HMM-based gene finders (e.g., FGenesB) have been found to perform considerably less well with truncated genes [93]. The accuracy of HMM-based gene finders may not be uniform over the length of the genome [93], so annotation-independent approaches (e.g., BLAST-walking) should be employed when mining the Biocomplexity metagenome.

The metagenome's high sequence diversity resulted in very few full-length, ORF-containing contigs in the sequence assemblies. Of the genes examined, none had more than one recognizably prominent type in their respective assemblies. This is unsurprising, since the episymbiont consortium is dominated by a handful of closely related, yet distinct,

ε-*Proteobacteria* phylotypes.

Although large numbers of SNPs were observed in some of the prominent contigs, most of them were silent and did not alter amino acid sequences. Of the five genes examined, only the ACP clones exhibited significant diversity (see Appendix V). Since silent SNPs were present in all genes with more than one clone, the lack of amino acid mutation is unlikely to be the result of PCR or other bias during the screening process.

Many of the SNPs predicted from the assemblies disappeared in the final sequences of the clones. Also, the examination of all eight ACP clones revealed SNPs at positions not predicted by the original assembly, suggesting that the raw metagenomic sequence chromatograms contained a considerable number of sequencing errors.

- **Expression Issues**

  Of the five clones, only the overexpression of IPMDH and GDH resulted in functional proteins. Although a wide variety of overexpression techniques were tested (e.g., altering incubation and induction conditions, the use of an *in vitro* cell-free expression system), no functional products were obtained from the remaining clones. Some potential remedies (e.g., changing the expression vector, salvaging the inclusion bodies) had not been trialed because the amount of resources required for carrying out these tests outweighed their potential usefulness.

  The AT-rich nature of the episymbiont genomes implies that the

codon usage of genes isolated from the community may be instrumental in determining their eventual successful overexpression in *E. coli*, for which tRNAs for Arg (AGG, AGA), Leu (CUA), Ile (AUA) and Pro (CCC) are known to be underrepresented [163]. Codon usage pattern analysis revealed that some rare codons (relative to *E. coli*) are widely used by the *A. pompejana* episymbionts: Leu (CUA) was abundant in the cloned genes, as was, to a less extent, Arg (AGG) (see Appendix V). It should be noted that clones without any rare codon in their first fifty amino acids (GDH and IPMDH) were also the only ones successfully overexpressed in a normal *E. coli* host, even though rare codons exist further down their gene sequences. However, codon bias-neutral expression systems such as RTS and a tRNA-compensated *E. coli* host strain did not produce any functional product from the other clones. Some of the gene products may be toxic to the host (e.g., ACP), however, the BL21 Star™ (DE3)pLysS strain used for initial expression trial is specifically designed for expressing toxic genes. Furthermore, expression products obtained with the *in vitro* RTS system did not result in significant levels of ACP activity.

In summary, the following potential remedies and technical issues were examined and excluded as causes for the lack of expression:

- The potential expression products, both from *in vivo* and *in vitro* expressions, were assayed under a wide range of conditions within reason, and no activity was observed.
- Whole *E. coli* cells, cell extract, and reaction mixture from

RTS runs were analyzed using SDS-PAGE and compared to negative control culture/run. No bands attributable to overexpression could be identified from the gels.

- Induced *E. coli* cells were microscopically examined, and there was no conclusive evidence pointing to the existence of inclusion bodies.
- Various combinations of induction temperature, time length, and inducer concentration were tested for *in vivo* expression, and no activity was observed from the resulting cell extracts. Post-induction cultures were confirmed to carry the recombinant plasmids using PCR.
- Induction in *E. coli* BL21(DE3)pRIL, a special strain carrying extra tRNA genes to reduce the effects of rare codons, did not result in observable expression.
- The clones were examined several times by sequencing and were confirmed to be in frame and intact (containing correct start and stop codons).

In the light of these results, it was judged that it was unlikely that issues specific to ε-*Proteobacteria* were the cause of unsuccessful expression for the following reasons:

- Two candidate genes were successfully isolated, cloned, and expressed both *in vivo* using *E. coli* as host organism and *in vitro* using the Roche RTS expression system.
- Several key enzymes related to reverse TCA cycle had been

successfully isolated from *Nautilia profundicola* [89], a minor, culturable member of the *A. pompejana* episymbiotic community, and the resulting recombinant proteins have been functionally characterized (Campbell, B. J., personal communication).

Therefore, it remains unclear why three out of five of the isolated and cloned gene candidates fail to result in heterologous expression.

Metagenomics is currently the most productive and comprehensive method for investigating complex microbial communities, and functional studies based on metagenomics allow scientists to tap into a new, rich source of information about environmental adaptations. The use of functional metagenomics in conjunction with the Equilibrium Model may provide insights on the enzymic temperature adaptations of a complex microbial system in a universal and quantitative manner.

Several suitable genes have been identified and isolated from the *A. pompejana* episymbiont metagenome database, and functional enzymes have been derived from two of those genes. Chapter Five describes the characterization of these two enzymes using the Equilibrium Model and an attempt at linking the resulting enzyme thermodynamic parameters to *in vivo* data.

Chapter Five

– Enzyme Eurythermalism and Environmental Temperature Variability

Chapter Three describes the validation and application of the Equilibrium Model for describing enzyme temperature-related behavior and enzyme temperature adaptation. The work described in this chapter is an attempt to apply the Equilibrium Model to enzymes derived from genes isolated from an environment that is extreme in many aspects, including temperature range and variation. The results presented in this chapter shed light on the role of enzyme eurythermalism in the thermal adaptation of *Alvinella pompejana* episymbionts and provided additional information related to the controversy surrounding the thermal adaptation of the host, *A. pompejana*.

This chapter was prepared as a self-contained research article and has been submitted to AEM (invertebrate microbiology) for publication. An extended discussion of this chapter can be found in Appendix VI.

My Contributions

I performed all the experimental work presented in this chapter with help from others as described in the Acknowledgements. I also wrote the paper with inputs from Prof. Daniel and Prof. Cary.

Working Title:

The Eurythermalism of *Alvinella pompejana* and its episymbionts – An Enzymic Approach

Charles K. Lee[*], S. Craig Cary[*, †], and Roy M. Daniel[*, ‡]

[*]Department of Biological Sciences, University of Waikato, Hamilton, New Zealand; [†]College of Marine and Earth Studies, University of Delaware, Lewis, Delaware, USA

[‡] Corresponding author. Mailing address: Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. Phone: +64-7-8384466-4213. E-mail: r.daniel@waikato.ac.nz.

I. Abstract

The Equilibrium Model, which describes the influence of temperature on enzyme activity, has been established as a valid and useful tool for characterizing enzyme eurythermalism and thermophily. By introducing K_{eq}, a temperature-dependent equilibrium constant for the interconversion between E_{act}, the active form of enzyme, and E_{inact}, a reversibly inactive form of enzyme, the Equilibrium Model provides currently the most complete description of enzyme-temperature relationship; its derived parameters are intrinsic and apparently universal to enzymes and, since they are derived under reaction conditions, potentially have physiological significance.

The vent-dwelling annelid *Alvinella pompejana* has been reported as an extremely eurythermal organism, and the symbiotic complex microbial community associated with its dorsal surface is likely to experience similar environmental thermal conditions. The *A. pompejana* episymbiont community, which is predominantly composed of filamentous ε-*Proteobacteria*, has recently been analyzed using metagenomics, which also enabled direct and targeted retrieval of genes coding for enzymes suitable for Equilibrium Model applications.

Two such genes have been isolated from the *A. pompejana* episymbiont community and heterologously expressed. The genes have been shown by RT-qPCR to be actively expressed in the community. The expression products were characterized using the Equilibrium Model, and their Equilibrium Model parameters suggested that enzyme eurythermalism constitutes part of the overall thermal adaptation strategy employed by the *A. pompejana* episymbiont

community. Moreover, the thermal characteristics of the enzymes are ostensibly in accord with the enzymes' predicted physiological roles and the abundance and expression of the corresponding genes.

II. Introduction

Temperature variation is an intrinsic property for almost all ecosystems, and many environments feature large temporal and/or spatial temperature gradients. Organisms adapted to such wide ranges of temperatures are termed *eurythermal*. While eurythermal poikilotherms can achieve adaptation through behavioral means, strict ectotherms and microbes have to be metabolically and structurally adapted. Therefore, enzymes of a eurythermal ectotherm or microbe must be adapted accordingly to facilitate cellular functions. The question of how this is achieved has challenged scientists for decades [31], and important discoveries have been made during the last few decades on how cellular mechanisms and enzymes of eurythermal organisms collectively contribute to such adaptations [30].

However, one aspect of enzyme temperature adaptation had not been satisfactorily addressed: how to assess and compare enzymes for their eurythermal qualities. Traditionally, biochemists have used enzyme catalytic efficiency (k_{cat}/K_m) as a measure of how well an enzyme operates [27]. Catalytic efficiency has been shown to be an inadequate and even misleading parameter for comparing enzymes, and is unsatisfactory as a measure of enzyme efficiency [49]. Furthermore, an enzyme's *in vivo* activity depends on its stability as well as

catalytic qualities [25], but stability cannot be directly measured in the process of obtaining an enzyme's catalytic efficiency and is often assessed in the absence of substrates, making the interpretation of these data difficult and potentially less physiologically significant.

While amino acid usage has been suggested to associate with enzyme thermophily, the associations tend to lack predictive value for individual proteins [155, 156]. A clear demonstration of this comes from a study of 26 glyceraldehyde phosphate dehydrogenase from organisms with OGT between 20ºC to 102ºC [164], in which findings from individual thermophilic enzymes conflicted with the "predictions" from the overall data in several aspects.

The recently proposed and experimentally validated Equilibrium Model [1, 2, 20] has provided a solution to the above problem. By incorporating a reversibly inactive form of enzyme (E_{inact}) in addition to the active form (E_{act}) and the irreversibly denatured form (X), the Equilibrium Model proposes that an equilibrium exists between the active and the inactive forms of enzyme, as expressed in the following equation:

Equation 5-1: $E_{act} \overset{K_{eq}}{\rightleftharpoons} E_{inact} \xrightarrow{k_{inact}} X$

K_{eq} is the equilibrium constant describing the E_{inact}/E_{act} ratio in the reversible E_{act}–E_{inact} interconversion, and k_{inact} is the rate constant for irreversible thermal denaturation. K_{eq} gives rise to two new, intrinsic, temperature-dependent parameters of enzymes, ΔH_{eq} and T_{eq}, which are the change in enthalpy and temperature midpoint, respectively, associated with the E_{act}–E_{inact} equilibrium. These two parameters, along with two previously

established parameters, $\Delta G^{\ddagger}_{cat}$ (the free energy of activation of the catalytic reaction) and $\Delta G^{\ddagger}_{inact}$ (the free energy of activation of the thermal denaturation process), provide the most complete quantitative description of how temperature influences enzyme activity and stability to date. Enzyme thermodynamic parameters derived from the Equilibrium Model, being intrinsic and universal, can be used to quantitatively compare enzymes of diverse origins and nature, and since ΔH_{eq} and T_{eq} are obtained under assay conditions, they potentially reflect *in vivo* conditions.

The Equilibrium Model's usefulness in studying enzyme temperature adaptation has been demonstrated through a survey of 21 enzymes from a wide variety of backgrounds [158]. The survey uncovered a strong link between T_{eq} and host optimal growth temperature (OGT) and showed that T_{eq} correlated with OGT better than enzyme stability ($\Delta G^{\ddagger}_{inact}$) did. The study also established ΔH_{eq} as the first intrinsic and quantitative measure of enzyme eurythermalism, with findings from the survey supporting the theory that ΔH_{eq} and T_{eq} are independent of $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$ [2]. Overall, the findings established the Equilibrium Model as a valid tool for assessing and comparing enzyme eurythermalism and thermophily [158].

Although eurythermalism is a common theme in many ecosystems, examples of highly eurythermal organisms are predominantly metazoans adapted to relatively gradual temperature shifts [165]. The vent-dwelling annelid *Alvinella pompejana* and its episymbionts, meanwhile, may be adapted to an extremely steep temperature gradient, with a measured temperature difference of up to 60ºC or more along its body length [67]. *Alvinella pompejana* is a polychaetous

annelid colonizing active chimneys at deep-sea hydrothermal vents along the East Pacific Rise [76]; it resides in the diffuse flow regions around the vents, one of the most extreme and thermally dynamic environment known to science. Apart from being possibly the most eurythermal eukaryote discovered to date, *A. pompejana* is also likely to be highly thermotolerant, since frequent temperature spikes of 80ºC or more have been observed inside tubes in which it resides [67].

The dorsal surface of a typical *A. pompejana* specimen is covered by an attached fleece of filamentous ε-*Proteobacteria* [81, 82, 166]. The bacterial consortium appears to maintain a unilateral non-obligatory symbiotic relationship with the host since *A. pompejana* has never been found without the episymbionts, and members of the consortium can be found on surfaces around *A. pompejana* colonies [81, 82]. The bacterial episymbionts associated with *A. pompejana* are collectively referred to as the *A. pompejana Proteobacteria* episymbionts (APS) and are dominated by a few phylotypes [81, 82]. APS have been suggested to be involved in the nutrition [71, 89] and detoxification [71, 91] of the host, and the bacteria and their associated structures (i.e., biofilm) may play an important role in shielding the host from rapid high temperature fluctuations [67, 75, 77].

There have been some controversies surrounding the thermal adaptedness of *A. pompejana*. While several studies reported temperatures spikes of 80ºC or more in immediate vicinities of live *A. pompejana* worms [67, 75], others have argued that *A. pompejana* itself lacks the adaptations necessary for coping with such temperatures [77, 167, 168]. Answers to whether *A. pompejana* experiences such high temperatures may be provided by examining the thermophilic and eurythermal features of APS enzymes, because although *A.*

pompejana may employ physiological, behavioral, and multiple levels of cellular adaptations to achieve this unique eurythermal adaptation, its bacterial episymbionts are much more limited in terms of available adaptation mechanisms while facing conditions just as, if not more, severe as those experienced by the host.

An international collaborative effort, funded by the National Science Foundation, is underway to study APS using a metagenomic approach [99]. Over 300,000 sequences have been obtained through community-wide shotgun sequencing and put through a custom-designed annotation pipeline (Project Metagenome, http://ocean.dbi.udel.edu/, log in with "guest"). Not only can the wealth of information available in the metagenome reveal valuable information about the consortium and its relationship with *A. pompejana*, it also provides access to sequence information for APS enzymes, which would otherwise be very difficult to obtain.

By utilizing the Equilibrium Model, this study aimed to characterize and assess the eurythermal qualities of enzymes derived from genes identified in the APS metagenome database in the hope of uncovering new information related to the potential thermal protective roles of APS, and consequently the relationship between *A. pompejana* and its episymbionts. In this study, suitable genes were identified using bioinformatics tools in the metagenome, cloned into a suitable expression vector, and overexpressed; the functional products were then characterized using the Equilibrium Model. To further relate the results to *in vivo* data, the expression levels of target genes were also examined by reverse transcription quantitative PCR (RT-qPCR) of RNA extracted from APS samples.

Through the development and validation of the Equilibrium Model, it is now possible to quantitatively and objectively assess an enzyme's eurythermal qualities [1, 2, 158]. A general relationship between enzyme intrinsic property and host optimal growth temperature has also been observed based on parameters derived from the Equilibrium Model, potentially allowing prediction of host habitat properties from those of its enzymes [158]. By applying the Equilibrium Model to enzymes originated from *A. pompejana* episymbionts, one may gather new information on the microbial consortium and its potential contribution to the thermal adaptation of *A. pompejana*. Such information may be instrumental in resolving longstanding controversies surrounding the eurythermalism and thermal tolerance of *A. pompejana*. To the authors' knowledge, this is the first attempt at elucidating the eurythermalism of a complex microbial system by quantitatively characterizing enzymes isolated from its metagenome.

III. Material and Methods

Genomic DNA and RNA. APS genomic DNA sample 'AP201' was extracted from the episymbiont biomass of a single *A. pompejana* specimen collected at Bio9 vent (9ºN East Pacific Rise) during dive #3836 of the Extreme 2002 cruise, and APS genomic DNA sample 'AP243' originated from an *A. pompejana* specimen collected at Q vent (9ºN East Pacific Rise) during dive #3837 of the same cruise.

To extract APS metagenomic DNA, the worms were slightly thawed on an iced aluminum block, and the episymbiont biomass was removed with sterile

forceps and placed in Tris-SDS-proteinase K lysis buffer for one hour. The nucleic acids were recovered with high salt/CTAB and phenol/chloroform extractions and subsequently precipitated with isopropanol [169]. RNA was removed with RNase Cocktail™ (Ambion® Inc., Austin, TX, USA), and DNA concentration and integrity were measured by UV spectrophotometry and agarose gel electrophoresis.

For RNA extraction, *A. pompejana* specimens were immersed in RNA*later*® (Ambion® Inc., Austin, TX, USA) immediately after collection on the seafloor to fully preserve their RNA. APS total RNA sample '#4064' was extracted from the episymbiont biomass of an *A. pompejana* specimen collected during the Extreme 2004 cruise using a Total RNA Isolation Kit (Ambion® Inc., Austin, TX, USA). The RNA sample was quantified (~200 ng/µl) using a NanoDrop™ (NanoDrop Technologies, Wilmington, DE, USA), and it integrity was verified using agarose gel electrophoresis.

Metagenomics. APS genomic DNA samples were restriction digested with 4-cutters, size selected by pulse field gel electrophoresis, and cloned into high-copy clone vectors (pPCR4-TOPO from Invitrogen Corp, Carlsbad, CA, USA; or pSmart from Lucigen Corp, Middleton, WI, USA). Over 300,000 shotgun clones with 1.5 Kb to 3 Kb inserts were generated; with clones derived from genomic DNA sample AP201 accounting for more than 90% of them. Over 180,000 clones were sequenced in forward and reverse directions, totaling around 134 Mb of high quality sequence data.

The metagenomic sequences were annotated using a custom-made annotation pipeline based on the FGenesB package (SoftBerry Inc, Mount Kisco, NY, USA).

The sequences were assembled with stringent criteria and resulted in ~26,000 contigs and ~100,000 singletons; this data set was designated the 'complete assembled Vent Epibionts Environmental Gene' (CAVEEG) data set and was also annotated. The sequence and annotation data were compiled into a metagenome database and made available through WWW access.

Bioinformatics. Since the FGenesB package was based on the NCBI Clusters of Orthologous Groups (COG) service [149], COG entries related to genes of interest (GOIs) were identified and used to retrieve chromatograms of metagenomic sequences related to the GOIs. The sequence chromatograms were assembled at high stringency (91% or 96%) using SeqMan (Version 5.51, DNASTAR Inc., Madison, WI, USA); the assemblies were edited to identify consensus sequences and SNPs. Additional sequence chromatograms were retrieved based on sequence homology to assist assembly editing, as did consensus sequences from CAVEEG and gene homologs from other ε-*Proteobacteria*.

The identified GOI consensus sequences were checked for the presence of *Nco*I and *Sma*I cut sites and used for designing PCR cloning primers. The sequences of constructed clones were compared to the original sequence assemblies, and their codon usage patterns were analyzed using Graphical Codon Usage Analyser (http://gcua.schoedl.de/index.html).

Cloning primer design. PCR cloning primers were designed for *aps_leuB* and *aps_gdhA* based on their consensus sequences. The PCR cloning primers are as follows: *aps_leuB* forward, 5'-ATGGTTAAGAGATATAAGATTGGTATT; *aps_leuB* reverse,

5’-TTAGCTTCTGCTAACCAAACTAG; *aps_gdhA* forward, 5’-ATGAGTGCAAGAGAGTATGTTAATAGTA; and *aps_gdhA* reverse: 5’-TTAGTTAGATTAAAGAGCCAAGAGG. Restriction endonuclease cut sites for *Nco*I and *Sma*I were added to the forward and reverse primers, respectively, along with extra overhangs (ATCGAT) at the 5’ end of primers to improve digesting efficiency. The primers were confirmed to be in frame with the expression vector and have similar T_m.

Cloning of GOIs. The GOIs were PCR amplified using the designed PCR cloning primers with AP201 genomic DNA as template. PCRs were carried out with a final volume of 25 µl, containing 1 µl of AP201 genomic DNA template, 100 nM of each respective primer, 100 µM of dNTPs, 3 mM of $MgCl_2$, 2.5 µl of 10X PCR buffer, and 1 U of *Taq* DNA polymerase (Roche Molecular Diagnostics, Pleasanton, CA, USA). PCRs were performed under the following conditions: an initial denaturation at 94ºC for 2 minutes; followed by 45 cycles of denaturation at 94ºC for 30 seconds, annealing at 58ºC for 30 seconds, and extension at 72ºC for 30 seconds; with a final extension of 5 minutes at 72ºC. The PCR amplicons were size verified with agarose gel electrophoresis, restriction digested with *Nco*I and *Sma*I, and ligated into pIVEX2.4d expression vector (Cat. No. 3269019, Roche Molecular Diagnostics, Pleasanton, CA, USA), which places the recombinant gene under the control of a T7 promoter and includes a N-terminus His-tag for easy purification. The ligation mixes were transformed into Invitrogen™ One Shot® TOP10 competent *E. coli* (Invitrogen Corp, Carlsbad, CA, USA). The transformation culture was spread on LB agar plates containing 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol, and

incubated overnight.

The transformants were screened with PCR using T7 primers, and cells picked from transformant colonies were added to PCR reactions as templates. Colonies which lead to PCR amplicons of the correct size (by agarose gel electrophoresis) were chosen for further inoculation and used for plasmid preparation with GenScript QuickClean 5M Miniprep Kit (GenScript Corp, Piscataway, NJ, USA). Multiple colonies were selected for both genes, and the resulting plasmids were sequenced in both directions.

Overexpression of GOIs. Clones of GOIs were transformed into Invitrogen™ One Shot® BL21 Star™ (DE3)pLysS chemically competent *E. coli* (Invitrogen Corp, Carlsbad, CA, USA) and screened using PCR as described above to verify successful transformation. The transformants were added to LB media containing 100 μg/ml of Ampicillin and 34 μg/ml of Chloramphenicol and incubated at 37ºC. 1mM IPTG was added to the cultures at an O.D. of 0.6 to induce overexpression for 4-6 hours. Cells were harvested by centrifugation and cell lysates were purified following instructions from the QIAGEN Ni-NTA Spin Kit (QIAGEN, Hilden, Germany). Purified expression products, APS_IPMDH and APS_GDH, were analyzed by SDS-PAGE and native protein gel electrophoresis.

Enzyme assay. In general terms, these were carried out as described in [20]. All enzyme assays performed in this work were continuous spectrophotometric assays measured using a Thermospectronic™ Helios γ-spectrophotometer equipped with a Thermospectronic™ single cell peltier-effect cuvette holder. Assays were performed across a suitable temperature range for the enzyme, and

absorbance data were collected at one-second intervals for three minutes using Vision^{TM} (version 1.25, Thermo Spectronic Inc.) on a Windows PC connected to the Helios. The formation of NADH and the oxidation of NADPH were both monitored at 365 nm ($_{\varepsilon}$NAD(P)H = 3,400 $M^{-1}cm^{-1}$). A quartz cuvette with 5 mm light path length was used for all assays.

All reactions were started by the rapid addition and mixing of micro-liter quantities of enzyme solution held at zero degrees into 400 or 500 µL of temperature-equilibrated reaction mixture. This allowed data collection to begin within the first few seconds, and the temperature of the reaction mixture remains within ± 1ºC of desired temperature during data collection. Substrate concentrations were set to at least ten times of K_m at all temperatures where possible to keep the enzyme operating at close to V_{max}: otherwise K_m values were determined and reaction curves were adjusted accordingly at each temperature point to compensate for the effect of temperature-related K_m increases. No signs of substrate/product inhibition were observed under the experimental conditions described. Blank rates were measured at all temperatures and used to correct reaction rates if significant. All the criteria set here and in [20] for accurate determination of Equilibrium Model parameters were met.

Due to its low $\Delta pK_a / \Delta t$ value, potassium phosphate buffer was used and was pH-optimized for both enzymes. APS_IPMDH was assayed with the following reaction mixture: 7.5 mM of NAD^+ and 2–6 mM of isopropylmalate in 100 mM K^+ phosphate buffer, pH 8 with 1 mM of $MgCl_2$. 1.06 µg of enzyme (10 µl of working stock) was added to the assay mixture to give a total volume of

400 µl, and the final enzyme concentration was 35.4 nM. APS_GDH was assayed with the following reaction mixture: 1 mM of NADPH, 35 mM of α-ketoglutarate, and 600 mM of NH_4Cl in 100 mM K^+ phosphate buffer (pH 8). 77 ng of APS_GDH (1 µl of working stock) was added to the assay mixture to give a total volume of 500 µl, and the final enzyme concentration was 430 pM.

Enzyme data analysis. The method for processing experimental data is as described in detail in [20]. Briefly, absorbance data from Vision™ were first converted to progress curves of product concentration (M) versus time (seconds) in Excel (version 11 for Windows, Microsoft Corp.). The data were subsequently imported into Scientist™ (version 2.01, MicroMath Inc.), where a set of initial estimates of thermodynamic parameters was optimized by *Simplex* searches through the complete data set (product concentration versus time versus temperature). The *optimized* initial estimates of parameters were then used to perform *Least Square* fits of the complete data set to the Equilibrium Model to generate the final values of the parameters; $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq}. 3D plots derived from the resulting parameters were compared to 3D plots generated from smoothed raw activity data to verify the fittings. The generated parameters were then used in the 'Zero-time Model' [20] to draw a zero-time activity versus temperature plot. Based on the variation between the individual triplicate rates from which the parameters are derived, for all the enzymes we have assayed thus far, we find that the experimental errors in the determination of $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$ and T_{eq}, are less than 0.5%, and less than 6% in the determination of ΔH_{eq} [20].

An alternate version of the Equilibrium Model, presented as a standalone application based on MATLAB® (The MathWorks, Inc.) is available on CD from

the corresponding author. The application enables the facile derivation of the Equilibrium Model parameters from a Microsoft® Office Excel file of experimental progress curves (product concentration versus time) and is suitable for computers running Microsoft® Windows XP or Vista, and is for non-commercial research purposes only.

Enzyme half-lives were calculated from enzyme assay data using Prism® (GraphPad Software, Inc.).

cDNA generation. APS RNA sample (#4064) was treated with 7 µg/ml ethidium monoazide bromide (EMA, Catalog #40015, Biotium Inc, Hayward, CA, USA) to deactivate any residual genomic DNA as described in [170]. cDNA preparations were generated from the RNA sample using a RETROscript® Kit (Ambion® Inc., Austin, TX, USA) with both gene-specific primers (GDH_qR and IPMDH_qR, cDNA preparation '4064GS') and random decamers provided with the RETROscript® Kit (cDNA preparation '4064RD'). Reverse transcription was performed according to the manufacturer's instruction (except 2 µl of MMLV reverse transcriptase was used), and 1 µg of EMA-treated total APS RNA was used as template.

Quantitative PCR. To examine *in vivo* expression levels of *aps_leuB* and *aps_gdhA* in the APS consortium, reverse transcription quantitative PCR (RT-qPCR) was used to amplify internal fragments of the said genes from the APS cDNA pool generated as described above. RT-qPCR primers based on verified clone sequences were designed for *aps_leuB* and *aps_gdhA* using Primer3 [171]. The primers are as follows:

GDH_qL, 5'-AAAGCCATCCAGAAGCTACC;

GDH_qR, 5'-TTTGCACCCTCATTAACCAT;

IPMDH_qL, 5'-TGCGACAATCTCTAGCCTCT;

IPMDH_qR, 5'-CCCATCTCAGTAGAGGAGCA.

Product sizes are 150 bp for *aps_gdhA* and 165 bp for *aps_leuB*, and T_m for all primers are 58 ± 0.3ºC.

qPCRs were carried out with a final volume of 25 µl, containing 1 µl of genomic DNA/cDNA/plasmid template, 100 nM of each respective primer, 100 µM of dNTPs, 5 mM of $MgCl_2$, 2.5 µl of 10X PCR buffer, 1 U of *Taq* DNA polymerase (Roche Molecular Diagnostics, Pleasanton, CA, USA), and SybrGreen (Invitrogen Corp, Carlsbad, CA, USA). For *aps_leuB*, 1 µg/ml of PCR-grade bovine serum albumin (BSA) was also added to the reaction mixture to reduce the effects of primer dimer formation. qPCR for *aps_leuB* was performed under the following conditions: an initial denaturation at 94ºC for 2 minutes; followed by 45 cycles of denaturation at 94ºC for 7 seconds, annealing at 58ºC for 10 seconds and extension at 68ºC for 10 seconds; with no final extension. qPCR for *aps_gdhA* was performed under the following conditions: an initial denaturation at 94ºC for 2 minutes; followed by 45 cycles of denaturation at 94ºC for 7 seconds and annealing/extension at 60ºC for 20 seconds; with no final extension. All amplifications were performed using a Rotor-Gene™ 6000 Real-Time Rotary Analyzer (Corbett Life Science, Sydney, Australia), which automatically determines the C_T values (cycle threshold values, the cycle number where the target signal level reached a pre-set threshold during qPCR) for the reactions. Fluorescence readings were acquired immediately after the extension step on the green channel with maximal gain.

For unknown samples, 1 µl of undiluted cDNA/genomic DNA preparations were used. Standard curves were created separately for *aps_leuB* and *aps_gdhA* using 10-fold dilution series (100 ng to 1 fg) of plasmids containing the respective genes. All unknown samples were analyzed in triplicates.

IV. Results

Candidate Genes. Genes of interest (GOIs) were selected based on assay availability, potential physiological significance, and sequence representation in the metagenome data set. Two genes, *leuB* and *gdhA*, which respectively code for 3-isopropylmalate dehydrogenase (IPMDH, EC 1.1.1.85) and glutamate dehydrogenase (GDH, EC 1.4.1.4), were chosen, and sequences related to these genes were identified in the metagenome database using annotation data (COG number) and sequence homology.

DNA sequence chromatograms associated with the genes were retrieved and assembled locally, and the sequence assemblies were edited to reveal the consensus sequences of each gene and to examine the level of single nucleotide polymorphism (SNP) present in the genes. Only one consensus sequence was identified for each GOI, although both sequence assemblies contained large numbers of singletons that cannot be incorporated into the prominent contigs even when assembled at a lower stringency (80% similarity). The candidate genes are hereafter referred to as *aps_leuB* and *aps_gdhA*, and PCR cloning primers were designed based on their consensus sequences. Although SNPs were identified in both *aps_leuB* and *aps_gdhA*, none of the SNPs in *aps_gdhA* was noisy (resulting

in an amino acid change), whereas *aps_leuB* contained one noisy SNP.

Cloning and expression of GOIs. Both *aps_leuB* and *aps_gdhA* were successfully amplified from genomic DNA sample AP201. The use of AP201 genomic DNA sample promises a definitive correspondence between the sequences in the metagenome database and clones constructed for this study, and all genomic DNA samples used in this study were confirmed to contain insignificant amounts of eukaryotic DNA. PCR amplicons of the GOIs were cloned into the pIVEX2.4d vector, and constructed clones were sequenced in forward and reverse directions. The sequence of the *aps_gdhA* clone corresponded to the predicted consensus sequence perfectly, and those of the *aps_leuB* clones fell into two types distinguishable by one noisy SNP at residue 314 (Lys or Arg). Codon usage analysis of clone sequences showed that *aps_leuB* and *aps_gdhA* clones contained fewer than ten amino acids below the 10% relative adaptiveness threshold for *E. coli*, and none of which was in the first fifty amino acids of the genes.

APS_IPMDH and APS_GDH, the respective overexpression products of *aps_leuB* and *aps_gdhA* clones, were purified using a Ni-NTA column and analyzed with SDS-PAGE and native protein gel electrophoresis, which confirmed the enzyme preparations to be essentially pure. However, the overexpression product of *aps_leuB* clone with Arg at residue 314 was non-functional, thus APS_IPMDH referred to only the overexpression product of *aps_leuB* clone with Lys at residue 314.

cDNA generation. Typically, genomic DNA contamination in RNA extractions is removed with DNase I treatment. For the APS RNA sample,

however, a standard DNase I treatment could not completely remove genomic DNA contamination, and PCR amplification of DNase I-treated RNA samples using RT-qPCR primers produced significant amounts of amplicon. Ethidium monoazide bromide (EMA) is a photoreactive analogue of ethidium bromide that selectively eliminates double-stranded DNA from PCR amplifications by covalently cross-linking with nucleic acid after light exposure [170]. By treating RNA samples with EMA, all double-stranded genomic DNA was eliminated from further reactions while the single-stranded RNA remained unaffected. PCR amplifications of the EMA-treated APS RNA sample yielded no amplicon detectable by agarose gel electrophoresis, suggesting that the EMA treatment had effectively rendered all genomic DNA in the sample unavailable for amplification. The genomic DNA-free RNA samples were then reverse transcribed with both random decamers (4064RD) and equal amounts of gene-specific reverse primers for both *aps_leuB* and *aps_gdhA* (4064GS). The resulting samples both tested positive for amplifiable cDNA using PCR as described above.

Quantitative PCR. To verify that the GOIs were actively expressed *in vivo*, qPCR studies of the GOIs were carried out using cDNA samples reverse transcribed from APS total RNA (4064RD & 4064GS) as well as APS genomic DNA samples (AP201 & AP243) as templates; quantified clone plasmid DNA was used as standards. Table 5-1 lists the qPCR results, including C_T values and calculated copy numbers of each gene in all samples, and reaction statistics presented in Table 5-2 demonstrated the robustness of both assays. Compared to the qPCR assay for *aps_gdhA*, the qPCR assay for *aps_leuB* was more prone to primer dimer formation and consequently loss of linearity at low template

concentrations. Therefore, the *aps_leuB* qPCR assay had a relatively high background signal and a C_T value for the negative control substantially lower than that of the *aps_gdhA* qPCR assay (Figure 5-1). However, both standard curves were extensive enough to cover all unknown samples, and signal levels of all unknown samples were significantly higher than those of the negative controls (Figure 5-1).

The ratios between *aps_leuB* and *aps_gdhA* in AP201 and AP243 are nearly identical, suggesting that *aps_leuB* is more abundant than *aps_gdhA* in the collective APS metagenome. Meanwhile, the ratios between *aps_leuB* and *aps_gdhA* in 4064RD and 4064GS cDNA samples showed some disparities, which may have arisen from fundamental differences between the ways the two cDNA samples were generated. Nevertheless, results from both cDNA preparations confirmed that both genes are expressed *in vivo*, with *aps_leuB* potentially at a collectively higher level than *aps_gdhA* in the ε-*Proteobacteria* consortium.

Table 5-1 – Cycle threshold (C_T) values of RT-qPCR assays and copy numbers of *aps_gdhA* and *aps_leuB* in episymbiont genomic DNA and cDNA samples.

Sample	C_T (mean ± S_E)		Copies per µl (×1000) (mean ± S_E)		Ratio
	aps_gdhA	*Aps_leuB*	*aps_gdhA*	*aps_leuB*	
AP201 genomic DNA	19 ± 0.1	18.8 ± 0.2	69.6 ± 4.2	1746.9 ± 233.1	25.1
AP243 genomic DNA	19.6 ± 0.2	19.4 ± 0.2	47.4 ± 7	1158.3 ± 114.8	24.4
4064GS cDNA	12.4 ± 0.1	15.8 ± 0.2	6563.7 ± 438.9	14091.6 ± 158.8	2.2
4064RD cDNA	23 ± 0.1	23.9 ± 0.1	4.3 ± 0.4	50 ± 4.4	11.6

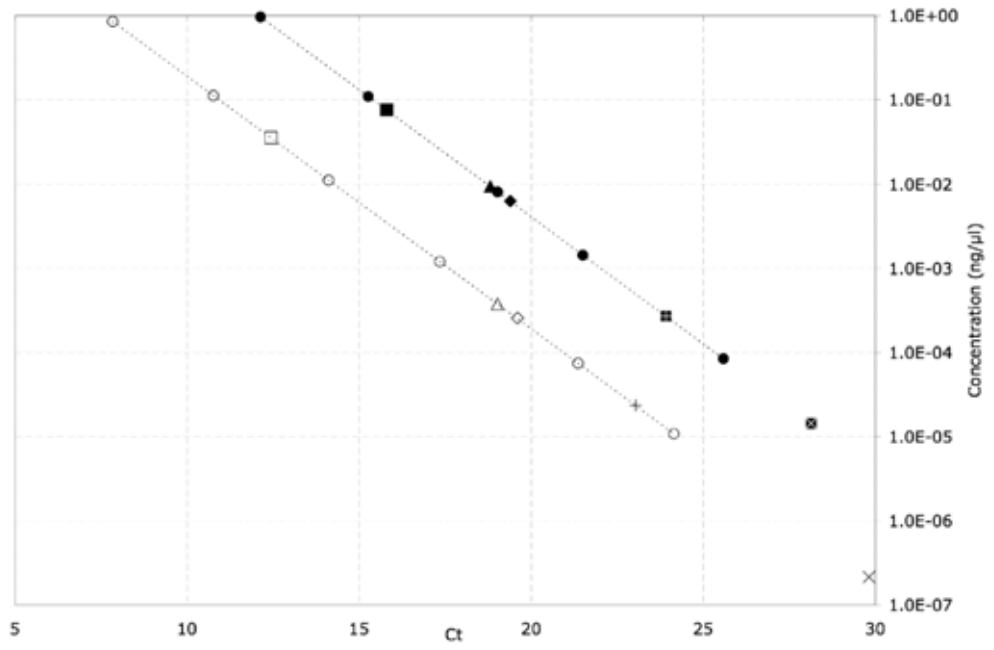
C_T value is the cycle number where the target signal level reached a pre-set threshold during qPCR. The ratios between *aps_gdhA* and *aps_leuB* were calculated from average copy numbers. The results suggest that *aps_leuB* exists in greater abundance than *aps_gdhA* both in the metagenome and the metatranscriptome (the pool of collective community mRNA), but in terms of expression per copy of gene, *aps_gdhA* appears to be expressed at a higher level.

Table 5-2 – Correlation coefficients, slopes, and efficiencies of standard curves of RT-qPCR assays.

Assay	r^2	Slope	Efficiency
aps_gdhA	0.99807	-3.33024	0.99654
aps_leuB	0.99594	-3.31424	1.00322

Both RT-qPCR assays appear highly robust and linear. The slight difference in efficiency may be attributed to different primer hybridization characteristics.

Figure 5-1 – Cycle threshold (C_T) value and copy number of individual samples plotted against *aps_gdhA* and *aps_leuB* RT-qPCR assay standard curves.



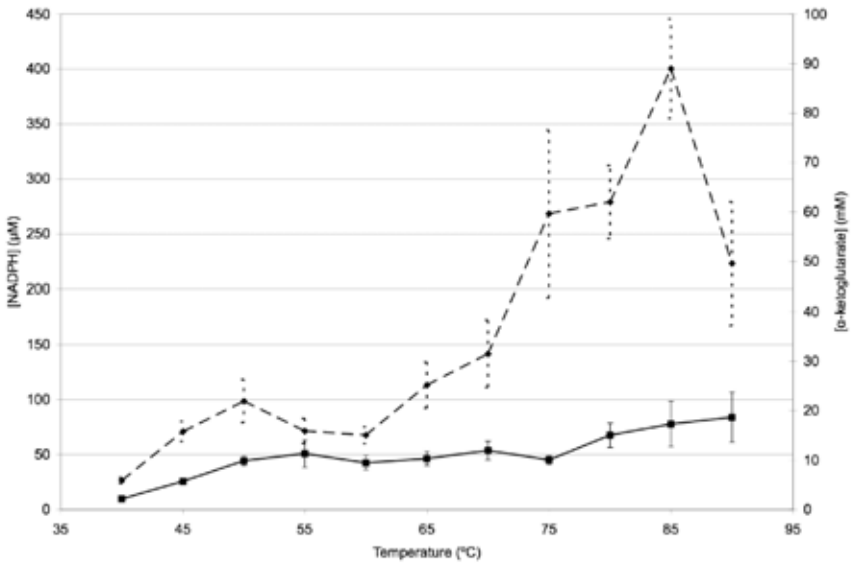
Hollow circles: *aps_gdhA* plasmid standards (1 ng to 10 fg). Solid circles: *aps_leuB* plasmid standards (1 ng to 100 fg). Hollow square: *aps_gdhA* 4064GS cDNA. Solid square: *aps_leuB* 4064GS cDNA. Cross: *aps_gdhA* 4064RD cDNA. Inverted cross: *aps_leuB* 4064RD cDNA. Hollow triangle: *aps_gdhA* AP201 genomic DNA. Solid triangle: *aps_leuB* AP201 genomic DNA. Hollow rhombus: *aps_gdhA* AP243 genomic DNA. Solid rhombus: *aps_leuB* AP243 genomic DNA. Saltire: *aps_gdhA* negative control. Inverted saltire: *aps_leuB* negative control. Signal levels of both negative controls are well below those of the lower concentration samples of the respective assays.

[THE UNDERLINED SENTENCES WILL NOT APPEAR IN THE

MANUSCRIPT SUBMITTED TO AEM]

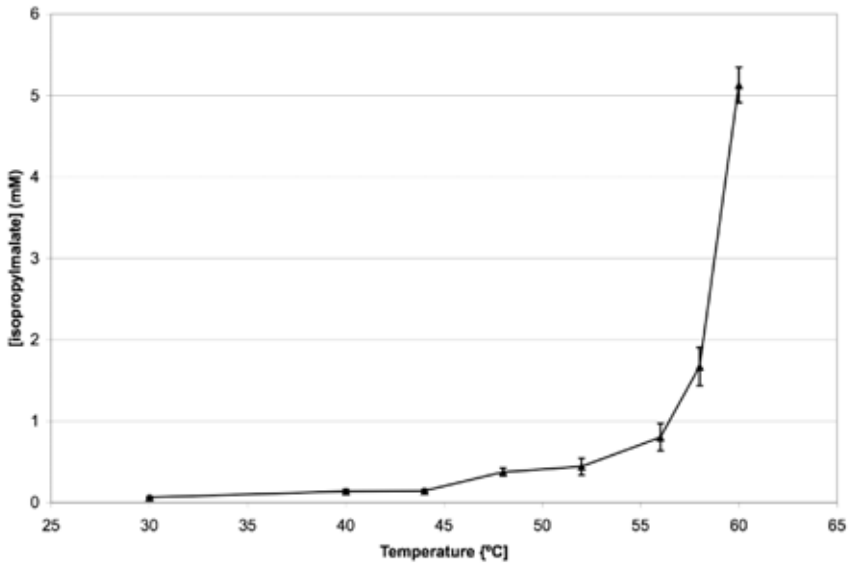
Enzyme properties. The purified APS_GDH and APS_IPMDH enzymes were fully characterized to obtain their kinetic parameters. For APS_GDH, K_m values for its substrates were determined at up to 90ºC (see Figure 5-2), and for APS_IPMDH the highest temperature for K_m determination was 60ºC (see Figure 5-3). APS_GDH is a highly stable (high $\Delta G^{\ddagger}_{inact}$) enzyme. More than 90% of its activity remained after 24 hours of incubation at room temperature (data not shown). Meanwhile, APS_IPMDH is significantly less stable since it denatured significantly at temperatures above 60ºC and completely and irreversibly lost activity at 70ºC within a minute.

Figure 5-2 – Substrates K_m plot for APS_GDH.



Solid rhombi: [NADPH]. Solid squares: [α-ketoglutarate]. NH_4Cl not shown since [NH_4Cl] is over 10X K_m at all temperatures. Over a temperature range of 50 degrees, K_m for α-ketoglutarate increased over ten fold. [THIS FIGURE WILL NOT APPEAR IN THE MANUSCRIPT SUBMITTED TO AEM]

Figure 5-3 – Substrate K_m plot for APS_IPMDH.



Solid triangles: [isopropylmalate]. NAD^+ not shown since [NAD^+] is over 10X K_m at all temperatures. K_m for isopropylmalate increased over fifty fold from 55ºC to 60ºC. [THIS FIGURE WILL NOT APPEAR IN THE MANUSCRIPT SUBMITTED TO AEM]

Equilibrium Model parameters and plots. Temperature profiles of both APS_GDH and APS_IPMDH were obtained and fitted to the Equilibrium Model, and the resulting Equilibrium Model parameters, which were derived directly from raw data, are listed in Table 5-3. The free energy of activation of the catalytic reaction ($\Delta G^{\ddagger}_{cat}$) for APS_IPMDH is 70 $kJ \cdot mol^{-1}$, and 58 $kJ \cdot mol^{-1}$ for APS_GDH. Simulated 3D plots of both enzymes generated using the parameters listed in Table 5-3 can be seen in Figure 5-4. The contrast in the widths of the

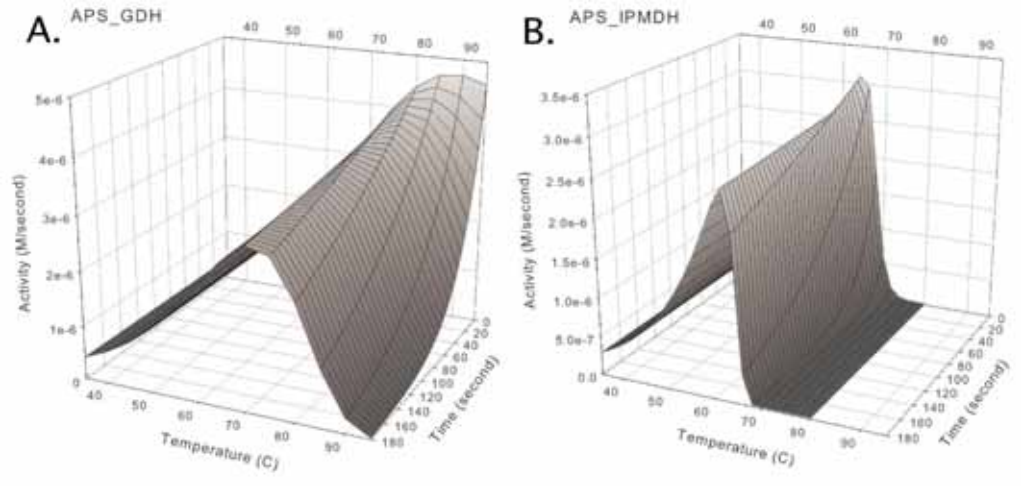
peaks between the two plots highlights the different ranges of temperatures over which these enzymes work optimally, and the positions of the activity peaks suggest that both enzymes are adapted for working at relatively high temperatures. Zero-time activity plots of the enzymes can be seen in Figure 5-5; they are the zero-time intersections of the plot in Figure 5-4, but are shown here to further demonstrate the difference in temperature adaptation of the two enzymes.

Table 5-3 – Equilibrium Model parameters of *A. pompejana* episymbiont enzymes compared to parameters of mesophilic enzymes from [158].

Description / Enzyme	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (ºC)	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})
APS_GDH	101	93	81	58
APS_IPMDH	92	793	67	70
25% Percentile*	91	112	53	n/a
Median*	94	218	58	n/a
75% Percentile*	97	394	62	n/a
Mean ± S_D*	94 ± 4	277 ± 218	56 ± 9	n/a

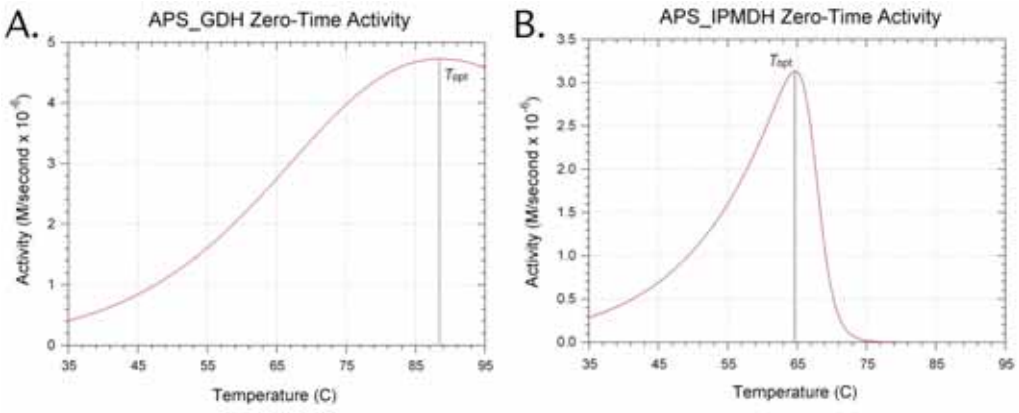
$\Delta G^{\ddagger}_{inact}$ is the free energy of activation of the thermal denaturation process. ΔH_{eq} is the enthalpic difference between active (E_{act}) and inactive forms (E_{inact}) of the enzyme. T_{eq} is the temperature at which the concentration of E_{act} equals that of E_{inact}. It is clear that the T_{eq} of both enzyme are above the normal range of enzymes of mesophilic origin, and that the two enzymes' ΔH_{eq} are on the extreme ends of the spectrum. * Calculated from data from [158].

Figure 5-4 – 3D plots (enzyme activity versus temperature versus time) of APS_GDH and APS_IPMDH.



The plots were generated from Equilibrium Model parameters of the enzymes, which were derived directly from raw data, and are the *fittest* depictions of the enzymes' actual temperature response by the Equilibrium Model. Note the sharp contrast in the widths of the *active* portions of the two plots.

Figure 5-5 – Zero-time activity plots of APS_GDH and APS_IPMDH.



The plots were generated by putting the determined Equilibrium Model parameters of enzymes into a version of the Equilibrium Model formulae specific for time zero.

V. Discussion

Two genes have been identified and isolated from the *A. pompejana* episymbiont metagenome, and heterologous overexpression of the isolated genes has resulted in functional enzymes. The product of *aps_leuB*, APS_IPMDH, catalyzes the oxidative decarboxylation of 3-isopropylmalate (IPM) to 2-oxoisocaproate, the penultimate step in leucine biosynthesis. Although the role of IPMDH is confined to a peripheral pathway, it is widely found in all types of organisms and is normally a homodimer [172, 173]. APS_GDH is the expressed product of *aps_gdhA* and plays central roles in nitrogen, amino acid, and glutamine metabolism and is a key link between catabolic and metabolic pathways. This homohexamer is a housekeeping gene ubiquitous in most organisms [174].

APS_IPMDH and APS_GDH are both metabolically essential enzymes, but the central roles of APS_GDH suggest that *aps_gdhA* is likely to be a constitutively expressed *housekeeping* gene whereas the same may not necessarily be true for *aps_leuB*. Evidence from the RT-qPCR assays suggests that mRNA derived from *aps_leuB* is present at a collectively higher level than that derived from *aps_gdhA* in the *A. pompejana* episymbionts mRNA pool. More importantly, the fact that *aps_leuB* and *aps_gdhA* are both actively transcribed in the APS consortium lends relevance to discussions of the thermodynamic characteristics of APS_IPMDH and APS_GDH in a physiological context. APS_GDH appears to be an enzyme geared toward working over wider range of temperatures (eurythermal) than APS_IPMDH, and both enzymes appear to be adapted for activity at relatively high temperatures. Additionally, the facts that

only one dominant consensus sequence could be identified for *leuB* and *gdhA* and that the consensus sequences themselves lack heterogeneity at the amino acid level potentially reflect the structure of the APS community, which is dominated by a few closely related phylotypes.

qPCR results listed in Table 5-1 suggest that the total copy number of *aps_leuB* is much higher than that of *aps_gdhA* in the metagenome, and the ratios between *aps_leuB* and *aps_gdhA* in two individual genomic DNA samples agree remarkably well. Although RT-qPCR results from 4064GS and 4064RD cDNA preparations both suggested that *aps_leuB* mRNA is more abundant in the collective community, results from the two data sets do not agree on the ratio between *aps_leuB* and *aps_gdhA*; this difference may have come from the different ways in which the cDNA samples were prepared. The 4064GS cDNA sample was generated using gene-specific reverse (antisense) qPCR primers for both *aps_leuB* and *aps_gdhA* during the reverse transcription step. cDNA generated using this approach may be more adequate for metagenomic applications since the resulting target cDNA concentrations are much higher (see Table 5-1) compared to the 4064RD cDNA sample, which was generated using random decamer nucleotides as reverse transcription primers. In 4064GS cDNA, the final ratio of *aps_gdhA* cDNA versus *aps_leuB* cDNA should in theory closely reflect the ratio of two mRNA populations in the total RNA sample since reverse transcription is a linear reaction performed at temperatures far below the melting temperatures of both primers and thus relatively insensitive to different primer efficiencies.

APS_IPMDH and APS_GDH were characterized using the Equilibrium

Model, and their thermodynamic parameters ($\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq}) were compared with reference data compiled from parameters of fourteen mesophilic enzymes (OGT between 20ºC and 40ºC) surveyed in an earlier study [158] (see Table 5-3). APS_GDH and APS_IPMDH are very different enzymes in terms of both temperature-activity response and stability. The most significant difference between the two enzymes is perhaps that between their ΔH_{eq} values (see Table 5-3). ΔH_{eq} is the enthalpic change associated with the reversible, temperature driven, interconversion between E_{act} and E_{inact}, and it has been established as a quantitative measure of eurythermalism at the enzymic level [158]: a large ΔH_{eq} infers stenothermal behavior (i.e., sharp decline in activity with increasing temperature), and vice versa. The small ΔH_{eq} of APS_GDH predicts eurythermal behavior, and the large ΔH_{eq} of APS_IPMDH implies a very narrow operating temperature range for the enzyme. However, the physiological significance of ΔH_{eq} cannot be elucidated without considering $\Delta G^{\ddagger}_{inact}$, since *in situ* the usefulness of an enzyme depends on its stability as much as its activity.

The exceptionally high $\Delta G^{\ddagger}_{inact}$ of APS_GDH clearly indicates that it is a thermally stable enzyme, as suggested by its long half-life at high temperatures (half-life at T_{eq} > 10 minutes, half-life at 70ºC > 1 hour) and resistance to activity loss at the room temperature. This combination of low ΔH_{eq} and high $\Delta G^{\ddagger}_{inact}$ suggests that APS_GDH is a highly stable enzyme capable of working across a relatively wide range of temperatures (i.e., eurythermal), as demonstrated in Figure 5-4A and Figure 5-5A. This corresponds ostensibly to the eurythermal and thermotolerant nature of APS.

APS_IPMDH appears to work optimally within a relatively narrow range of

temperatures, as inferred from its high ΔH_{eq}. Its $\Delta G^{\ddagger}_{inact}$ is also significantly below the average $\Delta G^{\ddagger}_{inact}$ of enzymes isolated from mesophiles in [158] (94 $kJ \cdot mol^{-1}$), suggesting that other mechanisms of adaptation (e.g., high turnover, temperature-linked gene expression) may be employed by APS to adapt functions associated with APS_IPMDH to the high temperature fluctuations it regularly experiences. The half-life of APS_IPMDH is around 30 seconds at T_{eq}, around 3 minutes at 55ºC, and over an hour at 45ºC. While low stability can in some cases be detrimental to cells due to short half-lives at high temperatures, high stability may hinder routine enzyme turnover in cells since stability is generally correlated to resistance to proteolysis. Overall, the findings are in harmony with the general consensus that a fine balance between stability and flexibility (i.e., thermolability) is a consistent feature of protein evolution in virtually all proteins [40], and that temperature adaptations occur at multiple levels of cellular functions [30].

T_{eq} is the temperature at which $[E_{act}]$ is equal to $[E_{inact}]$, and it may be helpful to conceptualize it as the thermal equivalent of K_m for the E_{act}–E_{inact} equilibrium. Although a difference of 14 degrees exists between T_{eq} values of the two enzymes, such discrepancies (i.e., variation in T_{eq} among enzymes from the same organism) have been previously observed [158] and may be common. Therefore, predictions based on T_{eq} are tentative.

Since the average T_{eq} of the reference data is 56 ± 9ºC, T_{eq} of both APS_GDH and APS_IPMDH are above the normal T_{eq} range of mesophilic enzymes. This implies that overall the *A. pompejana* episymbionts are adapted to a relatively thermophilic lifestyle. As described previously [158], the mean difference

between T_{eq} and OGT is 21 ± 15ºC, and this gap closes with increasing OGT. Considering that APS_IPMDH is stable at temperatures up to 60ºC (see Figure 5-4B) and that its apparent zero-time activity at 50ºC is less than one third of the maximum (see Figure 5-5B), APS_IPMDH is likely to work mostly in the temperature range of 50-60ºC, and the operational temperature range of APS_GDH is potentially higher and wider. This potentially reflects the nature of APS_GDH being a housekeeping enzyme and APS_IPMDH being part of a peripheral biosynthesis pathway that doesn't have to be constitutively expressed.

Stability is usually, but not necessarily, linked to temperature adaptation, and it has been suggested that $\Delta G^{\ddagger}_{cat}$ has a weak positive correlation with stability [158]. The fact that APS_GDH has a $\Delta G^{\ddagger}_{cat}$ higher than that of *Candida utilis* GDH, and that the average $\Delta G^{\ddagger}_{cat}$ of IPMDHs in the earlier survey [158] is higher than that of APS_IPMDH, appear to support previous observations.

[THE FOLLOWING PARAGRAPH WILL NOT APPEAR IN THE MANUSCRIPT SUBMITTED TO AEM]

From 40ºC to 90ºC, APS_GDH exhibits a ten-fold increase in its K_m for both α-ketoglutarate and NADPH (see Figure 5-2), whereas APS_IPMDH shows a more than 75-fold increase in its K_m for isopropylmalate from 30ºC to 60ºC, with the most dramatic increase taking place between 55ºC and 60ºC (see Figure 5-3). While the increase of K_m with temperature is common among enzymes [12, 175], APS_IPMDH is somewhat unusual in that most of the changes in K_m occurred within a space of 5ºC. Considering APS_IPMDH's extraordinarily high ΔH_{eq}, a hypothesis can be formed regarding a relationship between the two phenomena

since the shift in K_m and possibly the $E_{act} \rightleftharpoons E_{inact}$ transition both concern changes at or around the active site [2], but this theory was not investigated here. The free energy of activation of the catalytic reaction ($\Delta G^{\ddagger}_{cat}$) for APS_IPMDH is 70 kJ · mol^{-1}, close to the average value of three IPMDHs (74 kJ · mol^{-1}) described the earlier survey [158], and APS_GDH has a $\Delta G^{\ddagger}_{cat}$ of 59 kJ · mol^{-1}, close to that of *C. utilis* GDH (57 kJ · mol^{-1}) described previously [158].

It may be relevant to interpret RT-qPCR results in the context of the enzymes' properties. The presence of *aps_gdhA* transcript in the APS mRNA pool is expected due to it being a housekeeping gene. The eurythermal and thermophilic nature of APS_GDH also agrees with the high environmental temperature variability and occasional temperature spikes likely experienced by APS. On the other hand, the relatively high level of *aps_leuB* mRNA can perhaps be explained by the high turnover necessary to compensate for APS_IPMDH's low stability at high temperatures. The observation that *aps_leuB* exists in 25 times higher abundance than *aps_gdhA* in the APS metagenome implies that the APS consortium as a whole may be capable of expressing *aps_leuB* at a high level rapidly under suitable circumstances, such as when the host *A. pompejana* resides in a lower temperature area for an extended period of time. The low thermal stability of APS_IPMDH may also support this hypothesis since it allows quick degradation of APS_IPMDH, in response to changes in environmental conditions. While mRNA levels do not 100% reflect *in vivo* protein levels [94, 176], they do provide a glimpse into intracellular protein expressions.

The controversy between the high temperature characteristics of *A.*

pompejana's natural habitat and the absence of evidence for thermophilic and eurythermal adaptation in the worm itself so far may lie in the thermal adaptations of its episymbionts. During a temperature spike in the vent flow, the environment around the host worm equilibrates less rapidly than the milieu surrounding its episymbionts, therefore the episymbionts may be exposed to more direct and severe temperature changes. Although there is currently little information on the physical protection that APS and the associated biofilm may provide to the host, much can be learned about the thermal tolerance of the *A. pompejana* episymbionts from the temperature adaptation of APS enzymes. The extraordinarily stable and eurythermal APS_GDH enzyme may indicate that APS consortium is capable of withstanding a steep temperature gradient and the frequent high temperature spikes characteristic of the habitat. However, it must also be pointed out that host behavior and other acclimatization mechanisms are also likely to be important since APS_IPMDH is unlikely to sustainably function in the same manner as APS_GDH. The high pressure of APS's native environment may also contribute to the adaptation of some of its proteins.

One of the goals of the APS metagenome project was to investigate whether protein eurythermalism is a common adaptation of APS. Evidence presented in this study partially supports this hypothesis, but host behavior and temperature-related regulation of gene expression are also likely to play important roles in the temperature adaptation of APS enzymes in addition to intrinsically eurythermal enzyme.

The identification, isolation, overexpression, and characterization of novel enzymes from the *A. pompejana* ε-*Proteobacteria* episymbionts have revealed

relevant and useful information about the episymbionts. The use of the Equilibrium Model has provided insights on enzyme properties that were previously unavailable, and RT-qPCR analysis of the genes of interest has not only provided information on *in vivo* expression of the GOIs, but also lead to a hypothesis on intricate relationships between *A. pompejana* and its episymbionts. Furthermore, this study has shown how insights on the temperature adaptation of a complex microbial community can be gained by retrieving and expressing genes of interest from its metagenomic sequences through bioinformatic investigations and characterizing the resulting enzymes using the Equilibrium Model.

VI. Acknowledgements

This work was partly supported by grants from the National Science Foundation and the New Zealand Marsden fund. We're grateful for the support and suggestions from other members of the Biocomplexity project, specifically Dr. J. Grzymski and Dr. A. E. Murray of Desert Research Institute for the APS total RNA samples; Dr. B. J. Campbell of University of Delaware for the APS total genomic DNA samples; and Dr. M. Kaplarevic of University of Delaware for assistance with the database. We thank Dr. A. Rueckert of University of Waikato for assistance with qPCR assays and J. D. Steemson of University of Waikato for helpful comments on heterologous protein expression. Our gratitude goes to the following people for their critical reading of the manuscript: Dr. Ian R. MacDonald and Dr. Michelle E. Peterson, University of Waikato.

Chapter Six

– Additional Results and Conclusion

This chapter is an assemblage of expansions on findings from previous chapters, additional observations, and suggestions for future research directions. This chapter also contains a survey of all Equilibrium Model parameter sets available as of the writing of this thesis, including experimental results provided by Colin Monk (subtilisin, alcohol dehydrogenase, lipoxygenase, and some of the ß-glucosidase data), Michelle E. Peterson (ß-lactamase data and several other data sets), and Charis Shepherd (isopropylmalate dehydrogenases from various *Bacilli*).

My Contributions

In addition to the enzyme survey, I performed all the experimental work, designed the analyses, and interpreted the results presented in this chapter.

If I have not seen as far as others, it is because giants were standing on my shoulders.

– Hal Abelson

I. Consolidation and Expansion of Findings So Far

Enzyme Survey. The survey of 21 enzymes of diverse origins described in Chapter Three has validated the wide applicability of the Equilibrium Model and its usefulness for studying enzyme temperature adaptation by establishing ΔH_{eq} as the first intrinsic and quantitative measure of eurythermalism at the enzymic level.

Additional data are now available for enzymes that were not included in the previous survey. These enzyme' thermodynamic parameters and other characteristics (e.g., host optimal growth temperature, subunit configuration and size) are compiled here into a single expanded table of enzyme properties (Table 6-1), which now contains 45 data sets. This larger collection of enzyme parameters may provide new information, and some observations not examined in Chapter Three are briefly discussed below.

Table 6-1 – Expanded table of enzyme thermodynamic parameters.

Organism	Enzyme	Temperature Regulation	OGT (°C)	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (°C)	T_{eq} – OGT (°C)	Total Mw (kDa)	Subunit Mw (kDa)	No. of Subunits
P. fluorescens	AAA	Ectotherm	25	74	92	115	35.8	10.8	52	52	1
Wheat germ	ACP	Ectotherm	20	79	95	142	63.9	43.9	55	55	1
B. taurus	ADA	Homeotherm	39	65	99	101	55.6	16.6	33	33	1
G. stearothermophilus	ADH	Ectotherm	70	45	N/A	183	79.3	9.3	158	37	4
Thermus sp. RT41a	AKP	Ectotherm	75	72	99	305	90.0	15.0	97	45	2
ASB HK47	AKP	Ectotherm	15	55	90	149	17.5	2.5	67	67	1
B. taurus	AKP	Homeotherm	39	57	97	86	59.6	20.6	125	65	2
E. coli	AKP	Ectotherm	40	68	99	121	69.8	29.8	87	47	2
G. stearothermophilus	DHFR	Ectotherm	55	67	97	96	53.9	-1.1	19	19	1
B. cereus	DHFR	Ectotherm	30	66	90	261	58.5	28.5	19	19	1
M. profunda	DHFR	Ectotherm	2	67	93	104	54.6	52.6	18	18	1
E. coli	DHFR	Ectotherm	40	72	93	134	53.6	13.6	18	18	1
S. scrofa	FUM	Homeotherm	39	60	92	378	59.1	20.1	194	49	4
B. taurus	GDH 1.4.1.3 (NADH)	Homeotherm	39	66	92	295	53.0	14.0	330	56	6

Organism	Enzyme	Temperature Regulation	OGT (°C)	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (°C)	T_{eq} – OGT (°C)	Total Mw (kDa)	Subunit Mw (kDa)	No. of Subunits
B. taurus	GDH 1.4.1.3 (NADPH)	Homeotherm	39	66	91	181	48.0	9.0	330	56	6
C. utilis	GDH 1.4.1.4	Ectotherm	31	57	94	409	59.5	28.5	276	47	6
A. pompejana Symbionts	GDH 1.4.1.4	Ectotherm	72	58	99	133	86.0	14.0	320	60	6
B. taurus	GGTP	Homeotherm	39	63	98	109	52.0	13.0	88	48	2
B. caldovelox	IPMDH	Ectotherm	70	76	101	573	63.9	-6.1	80	40	2
B. psychrophilus	IPMDH	Ectotherm	20	72	100	123	56.7	36.7	80	40	2
B. subtilis	IPMDH	Ectotherm	30	73	94	255	52.8	22.8	80	40	2
B. psychrosaccharolyticus	IPMDH	Ectotherm	20	72	86	138	41.4	21.4	80	40	2
B. thurigiensis	IPMDH	Ectotherm	30	74	99	269	64.7	34.7	76	38	2
A. pompejana Symbionts	IPMDH	Ectotherm	72	70	92	793	67.5	-4.5	85	43	2
G. max	LOX	Ectotherm	25	63	89	86	17.2	-7.8	94	94	1
B. taurus	MDH	Homeotherm	39	53	85	826	67.4	28.4	70	33	2
E. coli	MDH	Ectotherm	40	55	96	619	70.7	30.7	117	35	4
B. cereus	MDH	Ectotherm	30	51	86	243	40.4	10.4	140	45	4
R. glutinis	PAL	Ectotherm	24	80	97	181	56.5	32.5	275	83	4
P. dulcis	ß-GLU	Ectotherm	20	63	95	100	55.9	35.9	135	135	1
C. saccharolyticus	ß-GLU (CKL, glucoside)	Ectotherm	70	81	98	154	73.9	3.9	54	54	1

Organism	Enzyme	Temperature Regulation	OGT (°C)	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (°C)	T_{eq} – OGT (°C)	Total Mw (kDa)	Subunit Mw (kDa)	No. of Subunits
C. saccharolyticus	ß-GLU (fucoside)	Ectotherm	70	77	101	166	79.1	9.1	54	54	1
C. saccharolyticus	ß-GLU (galactoside)	Ectotherm	70	77	103	147	76.2	6.2	54	54	1
C. saccharolyticus	ß-GLU (xyloside)	Ectotherm	70	83	102	83	64.9	-5.1	54	54	1
B. cereus	ß-LAC (ampicillin)	Ectotherm	30	65	95	114	31.4	1.4	30	30	1
B. cereus	ß-LAC (cefazolin)	Ectotherm	30	76	99	152	57.9	27.9	30	30	1
B. cereus	ß-LAC (nitrocefin)	Ectotherm	30	69	94	146	53.1	23.1	30	30	1
B. cereus	ß-LAC (penicillin)	Ectotherm	30	66	95	111	31.0	1.0	30	30	1
S. carlsbergensis	SUB (AAPA)	Ectotherm	30	72	102	111	63.6	33.6	27	27	1
S. carlsbergensis	SUB (AAPL)	Ectotherm	30	61	99	93	60.3	30.3	27	27	1
S. carlsbergensis	SUB (AAPN)	Ectotherm	30	62	98	86	57.8	27.8	27	27	1
S. carlsbergensis	SUB (AAPV)	Ectotherm	30	78	95	137	68.7	38.7	27	27	1
T. aquaticus	α-GLU	Ectotherm	70	88	103	149	74.4	4.4	61	61	1
S. cerevisiae	α-GLU	Ectotherm	28	71	90	272	39.4	11.4	62	62	1
G. stearothermophilus	α-GLU	Ectotherm	55	64	96	225	67.8	12.8	62	62	1

The full organism and enzyme names are as follows: *Pseudomonas fluorescens*, *Bos taurus*, *Geobacillus stearothermophilus*, *Escherichia coli*,

Bacillus cereus, *Moritella profunda*, *Sus scrofa*, *Candida utilis*, *Bacillus caldovelox*, *Bacillus psychrophilus*, *Bacillus subtilis*, *Bacillus psychrosaccharolyticus*, *Bacillus thurgiensis*, *Glycine max*, *Rhodotorula glutinis*, *Prunus dulcis*, *Caldicellulosiruptor saccharolyticus*, *Saccharomyces carlsbergensis*, *Thermus aquaticus*, *Saccharomyces cerevisiae*. AAA, aryl acylamidase; ACP, acid phosphatase; ADA, adenosine deaminase; ADH, alcohol dehydrogenase; AKP, alkaline phosphatase; DHFR, dihydrofolate reductase; FUM, fumarate hydratase; GDH, glutamate dehydrogenase; GGTP, γ-glutamyl transferase; IPMDH, isopropylmalate dehydrogenase; LOX, lipoxygenase; MDH, malate dehydrogenase; PAL, phenylalanine ammonia-lyase; GLU, glucosidase; LAC, lactamase; SUB, subtilisin.

a. Examining Known Correlations in the Expanded Table

The expanded table (Table 6-1) was analyzed as described in Chapter Three, and previously identified correlations involving parameters directly derived from the Equilibrium Model and the optimal growth temperature (OGT) were re-examined (see Table 6-2). T_{opt} was not examined since it is a peripheral parameter; and since correlations involved in establishing ΔH_{eq} as a measure of enzyme eurythermalism were extremely robust (Table 3-2), Half Width at Half Maximum and T_{eq} minus T_{opt} were also not examined.

Table 6-2 – Comparison of correlation analysis results from Chapter Three and the expanded analysis.

Correlation	Gamma (Chapter 3)	Gamma (Table 6-1)	*P* value (Chapter 3)	*P* value (Table 6-1)
OGT & T_{eq}	0.5253	0.5581	0.0012	<0.0001
OGT & $\Delta G^{\ddagger}_{inact}$	0.3535	0.3636	0.0295	0.0014
OGT & (T_{eq}-OGT)	-0.4646	-0.4171	0.0042	0.0002
$\Delta G^{\ddagger}_{cat}$ & $\Delta G^{\ddagger}_{inact}$	0.3048	0.3188	0.0533	0.0035
T_{eq} & ΔH_{eq}	0.3143	0.1507	0.0463	0.146

In Table 3-2, T_{eq} has been identified as a better predictor of OGT than stability ($\Delta G^{\ddagger}_{inact}$), an observation that remains true for the expanded table (Table 6-1). The Gamma for the correlation between T_{eq} and OGT is now 0.56 (versus 0.53 in Table 3-2), and the *P* value is lower than 0.0001. Meanwhile, the larger data set has also lead to an improved correlation

between OGT and $\Delta G^{\ddagger}_{inact}$, for which the Gamma is now 0.36 (versus 0.35 in Table 3-2), and the *P* value is 0.0014. Nevertheless, T_{eq} remains a better predictor of OGT than $\Delta G^{\ddagger}_{inact}$.

A previously identified negative correlation (Table 3-2, Gamma = -0.46, *P* value = 0.0042) between OGT and the difference between T_{eq} and OGT remains significant in the expanded table (Table 6-2, Gamma = -0.42, *P* value = 0.0002). This correlation may be an additional piece of evidence supporting the theory of a hyperthermophilic last common ancestor for the present mesophilic microbial life, although it also suggests that host OGT cannot be unequivocally predicted from T_{eq}.

The weak correlation between $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$ identified in Table 3-2 appears to support the general assertion that a heat stable (rigid) enzyme is less optimized for catalytic activity. This correlation is stronger in Table 6-2.

A previously undiscussed correlation is that between T_{eq} and ΔH_{eq}. This correlation was unstable because its statistical significance appeared to be dependent on several outliers with high ΔH_{eq} values (see the corresponding scatter plot in Table 3-2). This correlation is statistically insignificant with the larger sample size of the expanded table (Table 6-2, Gamma = 0.15, *P* value = 0.146), supporting the idea that ΔH_{eq}, or enzyme eurythermalism, may be determined more by the variability or temperature range of the organism's thermal environment than by other thermal properties. Overall, results from the substantially expanded table corroborate findings from Chapter Three.

b. Enzyme Subunit Configuration and Enzyme Eurythermalism

One intriguing trend newly identified in the expanded table is the potential correlation between ΔH_{eq} and the molecular weight or number of subunits of the enzyme. Examination of correlations between ΔH_{eq} and other properties in the expanded table revealed that ΔH_{eq} is significantly correlated to the total molecular weight of the protein (Gamma = 0.41, *P* value = 0.0049), but not to the size of individual subunits, nor to the molecular weight of monomeric enzymes. This suggests that ΔH_{eq} may in fact be linked to the subunit configuration of enzyme rather than the total molecular weight.

The data sets were then grouped into monomeric and multimeric enzymes, and a *t* test based on ΔH_{eq} values was performed as described in Chapter Three. The results suggest that the ΔH_{eq} values of these two groups significantly differ (*P* value = 0.0055). It is unclear what this correlation implies, but given that all multimeric enzymes in Table 6-1 are composed of subunits that contain a complete active site, it is consistent with a simple multiplication of ΔH_{eq} according to the number of active sites. However, no significant difference was found between ΔH_{eq} values of dimers and tetramers or those of dimers and hexamers.

c. The Influence of Artificial Substrates

Many of the enzymes analyzed so far were assayed using artificial substrates, which may cast doubt on the direct physiological significance of

the resulting parameters. To answer this question, correlation analysis (same as described above and in Chapter Three) was limited to only 21 enzymes assayed with physiological substrates in the expanded table, namely *Bos taurus* adenosine deaminase, assayed with adenosine as the substrate; various dihydrofolate reductases, assayed with dihydrofolate and NADPH as substrates; *Sus scrofa* fumarate hydratase, assayed with malic acid as the substrates; various glutamate dehydrogenases, assayed with α-ketoglutarate and NAD(P)H as substrates; various isopropylmalate dehydrogenases, assayed with isopropylmalate and NAD^+ as substrates; various malate dehydrogenases, assayed with oxaloacetate and NADH as substrates; *Glycine max* lipoxygenase, assayed with octadecadienoic acid as the substrate; and *Rhodotorula glutinis* phenylalanine ammonia-lyase, assayed with phenylalanine as the substrate. The results are listed in Table 6-3.

With the exception of the correlation between OGT and $\Delta G^{\ddagger}_{inact}$, all relevant correlations remained significant. The *P* value of the correlation between OGT and T_{eq} is higher relative to that described in Table 3-2, as expected from a smaller and less diverse data set. When the same criterion was applied to data from Table 3-1, only 11 data sets were available for analysis. Despite the small sample size, the correlation between ΔH_{eq} and HWHM remained extremely robust (data not shown). Overall, the results suggest that the use of artificial substrates does not reduce the significance of the resulting Equilibrium Model parameters.

Table 6-3 – Correlation analysis of only enzymes assayed with physiological substrates.

Correlation	Gamma (Chapter 3)	Gamma (physiological substrate)	*P* value (Chapter 3)	*P* value (physiological substrate)
OGT & T_{eq}	0.5253	0.4031	0.0012	0.0148
OGT & $\Delta G^{\ddagger}_{inact}$	0.3535	0.224	0.0295	0.1845
OGT & (T_{eq}-OGT)	-0.4646	-0.5158	0.0042	0.0019
$\Delta G^{\ddagger}_{cat}$ & $\Delta G^{\ddagger}_{inact}$	0.3048	0.3196	0.0533	0.0516

Enzyme Temperature Adaptation. An analysis of enzymes derived from the episymbiotic ε-*Proteobacteria* associated with *Alvinella pompejana* using the Equilibrium Model, described in Chapter Five, has provided insights into the thermal adaptation of the *Pompeii worm* and its episymbionts. One enzyme, APS_GDH, possesses traits associated with a eurythermal and thermophilic lifestyle, while properties of APS_IPMDH suggest that additional mechanisms may be employed within the microbial community to adapt to a rapidly changing and occasionally sweltering thermal environment (see Table 5-3). Both enzymes are actively expressed in the episymbiont community (see Table 5-1), and the lack of genetic diversity for APS_GDH, a housekeeping gene, hints at the existence of a core metabolism. Overall, these findings indicate to temperature adaptation mechanisms that span across multiple levels of cellular functions.

The study presented in Chapter Five is an example of targeted functional metagenomics, combining bioinformatics, recombinant protein overexpression, examination of *in vivo* gene expressions, and enzyme temperature adaptation studies based on the Equilibrium Model, and therefore serves as a proof-of-concept for future studies of similar nature. Moreover, the study has lead to some interesting observations relevant to understanding the molecular basis of the Equilibrium Model.

d. The Relationship between K_m Shift and ΔH_{eq}

In Figure 5-3, it was noted that APS_IPMDH exhibited a marked increase in its K_m for isopropylmalate within a narrow range of temperatures close to its T_{eq}, hinting at a potential link between K_m shift and the E_{act}–E_{inact} transition; this hypothesis is briefly explored here.

To visualize the E_{act}–E_{inact} transition, one can calculate the concentration of E_{act} relative to E_0 at time zero with the following equation, which is derived from Equation 1-6 and Equation 1-7 ([X] = 0 at time zero):

Equation 6-1: $$[\mathrm{E_{act}}] = \frac{E_0}{1 + e^{\frac{\Delta H_{eq}}{R}\left(\frac{1}{T_{eq}} - \frac{1}{T}\right)}}$$

One can then calculate the temperatures at which [E_{act}] = 10% of E_0 (T_{10}) and [E_{act}] = 90% of E_0 (T_{90}) with these equations derived from Equation 6-1:

Equation 6-2: $$T_{10} = \frac{1}{\frac{1}{T_{eq}} - \left(\ln(9) \cdot \frac{R}{\Delta H_{eq}}\right)}$$

Equation 6-3: $$T_{90} = \frac{1}{\frac{1}{T_{\text{eq}}} - \left(\ln(\frac{1}{9}) \cdot \frac{R}{\Delta H_{eq}} \right)}$$

The respective T_{10}, T_{90}, and T_{eq} of APS_GDH and APS_IPMDH are listed in Table 6-4, and the transition phases are mapped on to the respective projected zero-time activity plots of each enzyme, along with their K_{m} shifts with temperature (only the substrate/cofactor with the most significant changes in K_{m}) in Figure 6-1 and Figure 6-2.

Table 6-4 – T_{10}, T_{90}, and T_{eq} of APS_GDH, APS_IPMDH, and several other enzymes.

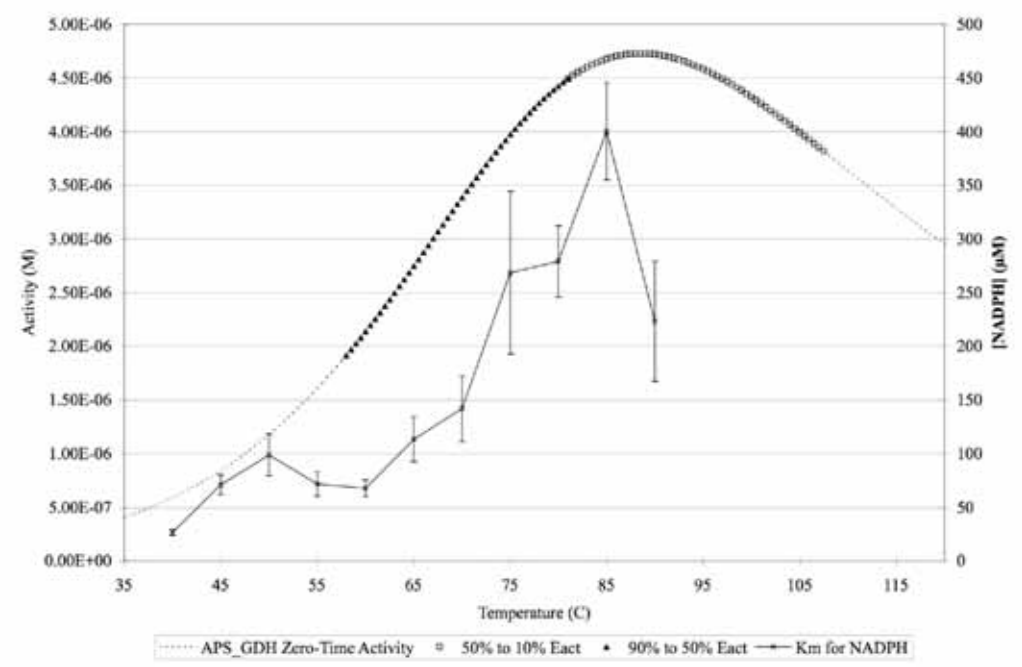
Enzyme	ΔH_{eq} (kJ · mol^{-1})	T_{90} (ºC)	T_{eq} (ºC)	T_{10} (ºC)
APS_GDH	93	58	81	107.5
APS_IPMDH	793	64.5	67	70
P. dulcis ß-GLU	100	37.2	55.9	77
G. stearothermophilus α-GLU	225	58.6	67.8	77.5
E. coli MDH	619	67.2	70.7	74.2

At T_{10}, the overall activity level of enzyme appears as if 10% of the enzyme population exist as E_{act}. Likewise, at T_{90}, the overall activity level of the enzyme appears as if 90% of the enzyme population exist as E_{act}.

From the plots, it is clear that the E_{act}–E_{inact} transition is a gradual process for APS_GDH (ΔH_{eq} = 93 kJ · mol^{-1}), spanning over almost 50ºC, whereas the shift from 90% of the enzyme being in the E_{act} form to just 10% happens in a space of just 5.5ºC for APS_IPMDH (ΔH_{eq} = 793 kJ · mol^{-1}). This corresponds to the contrasting ΔH_{eq} values of APS_GDH and APS_IPMDH

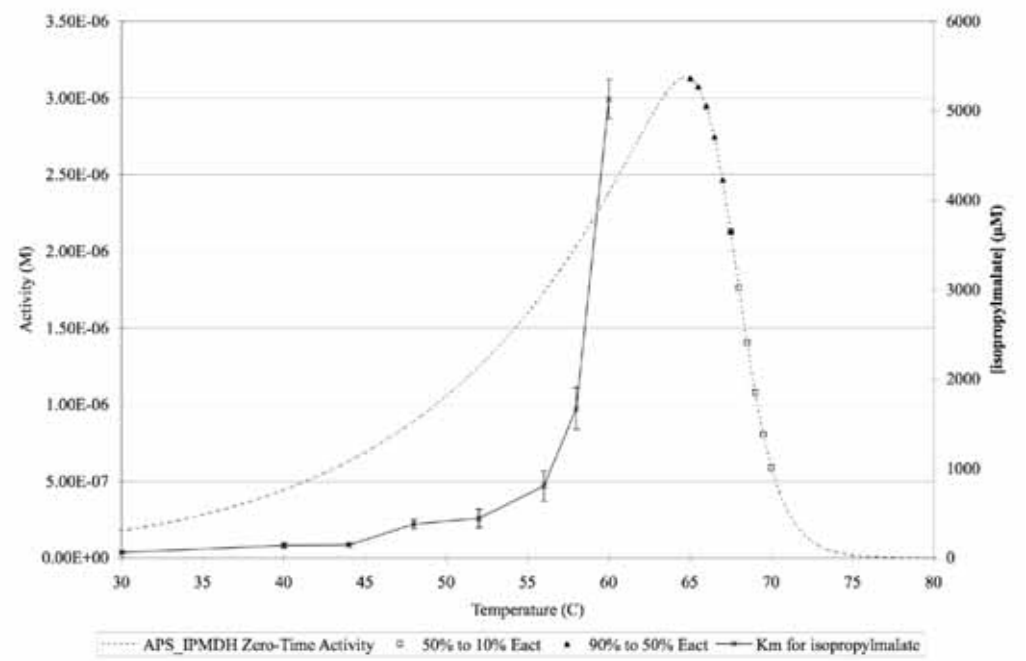
(see Table 5-3) and the conclusion from Chapter Three and Table 3-2 that ΔH_{eq} is a quantitative measure of enzyme eurythermalism. Intriguingly, for APS_GDH the E_{act}–E_{inact} transition started before any significantly shift in the enzyme's K_m for NADPH occurred, while the abrupt E_{act}–E_{inact} transition of APS_IPMDH took place at temperatures higher than the sharp increase in its K_m for isopropylmalate.

Figure 6-1 – Zero-time activity plot of APS_GDH.



ΔH_{eq} for APS_GDH is 93 kJ·mol^{-1}. Squares represent the transition from 90% E_{act} (T_{90}) to 50% E_{act} (T_{eq}). Triangles represent the transition from 50% E_{act} (T_{eq}) to 10% E_{act} (T_{10}). Saltires represent K_m values for NADPH.

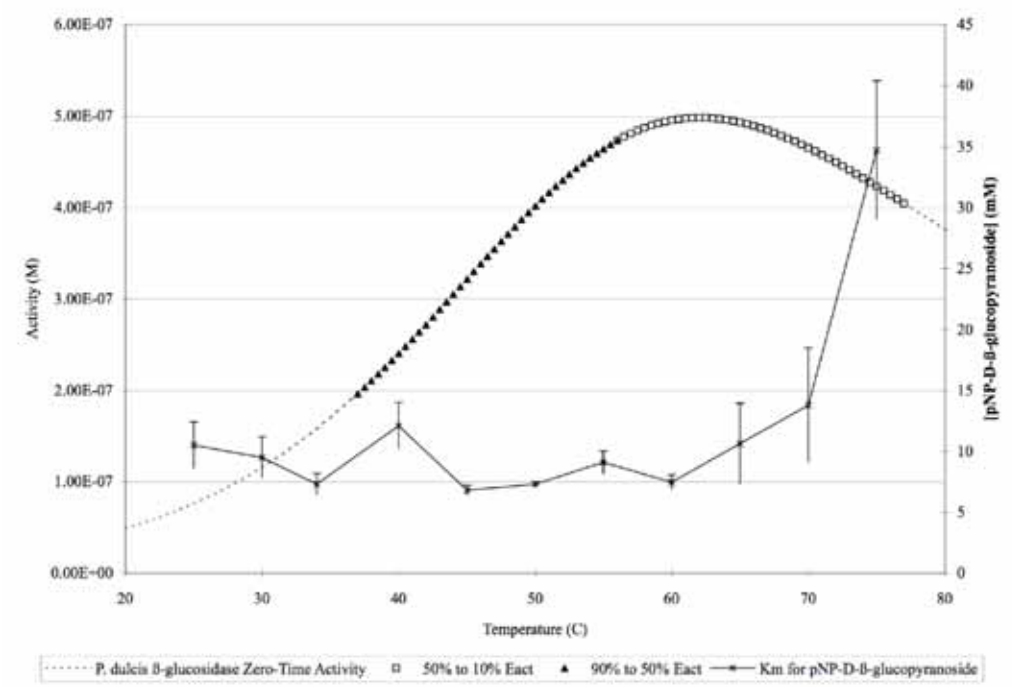
Figure 6-2 – Zero-time activity plot of APS_IPMDH.



ΔH_{eq} for APS_IPMDH is 793 kJ·mol^{-1}. Squares represent the transition from 90% E_{act} (T_{90}) to 50% E_{act} (T_{eq}). Triangles represent the transition from 50% E_{act} (T_{eq}) to 10% E_{act} (T_{10}). Saltires represent K_m values for isopropylmalate.

Similar trends can be observed in zero-time activity plots of *Prunus dulcis* ß-glucosidase (ΔH_{eq} = 100 kJ·mol^{-1}), *Geobacillus stearothermophilus* α-glucosidase (ΔH_{eq} = 225 kJ·mol^{-1}), and *Escherichia coli* malate dehydrogenase (ΔH_{eq} = 619 kJ·mol^{-1}) in Figure 6-3, Figure 6-4, and Figure 6-5, respectively. It must be stressed here that the V_{max} was maintained at all temperatures in the temperature profiles as specified in Chapter Two, so the observed effects were not due to rate drops caused by changes in K_m.

Figure 6-3 – Zero-time activity plot of *P. dulcis* ß-glucosidase.

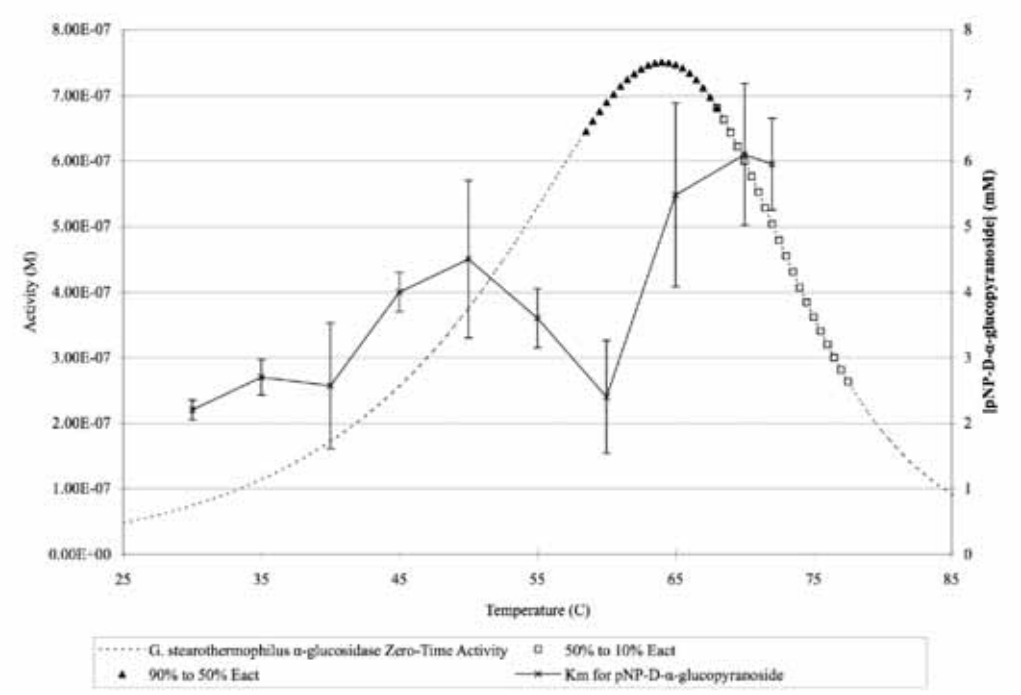


ΔH_{eq} for *P. dulcis* ß-glucosidase is 100 kJ·mol^{-1}. Squares represent the transition from 90% E_{act} (T_{90}) to 50% E_{act} (T_{eq}). Triangles represent the transition from 50% E_{act} (T_{eq}) to 10% E_{act} (T_{10}). Saltires represent K_m values for pNP-D-ß-glucopyranoside.

This phenomenon explains the positive correlation between ΔH_{eq} and the difference between T_{eq} and T_{opt} observed in Table 3-2. For example, the E_{act}–E_{inact} transition of an enzyme with a low ΔH_{eq} value starts at a lower temperature, leading to a T_{eq} lower than its T_{opt}. It also explains why, for enzymes with low ΔH_{eq} values (i.e., APS_GDH, *P. dulcis* ß-glucosidase), the lower left halves of their zero-time activity curves bear less resemblance to

the ideal Arrhenius equation curve (as shown in Figure 6-6) than those of enzymes with high ΔH_{eq} values, particularly near the peak of the curves.

Figure 6-4 – Zero-time activity plot of *G. stearothermophilus* α-glucosidase.

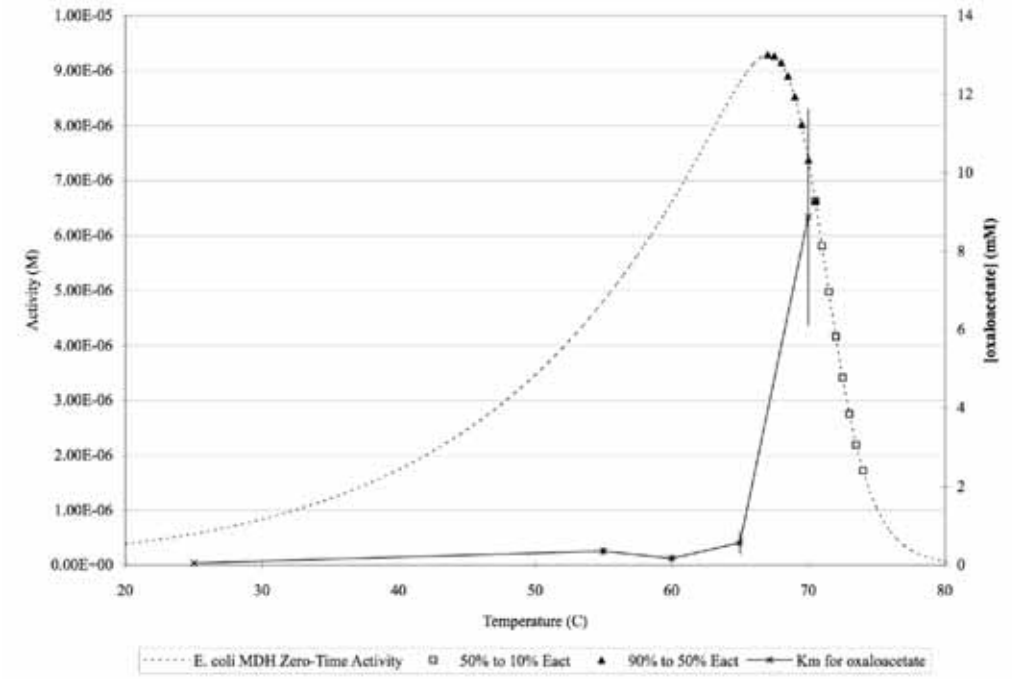


ΔH_{eq} for *G. stearothermophilus* α-glucosidase is 225 $kJ \cdot mol^{-1}$. Squares represent the transition from 90% E_{act} (T_{90}) to 50% E_{act} (T_{eq}). Triangles represent the transition from 50% E_{act} (T_{eq}) to 10% E_{act} (T_{10}). Saltires represent K_m values for pNP-D-α-glucopyranoside.

This occurs because for enzymes with low ΔH_{eq} values, the E_{act}–E_{inact} transition starts to occur at a temperature lower than T_{opt}, and the loss of activity due to increasing [E_{inact}] causes the curve to deviate from the ideal

curve shown in Figure 6-6 more significantly.

Figure 6-5 – Zero-time activity plot of *E. coli* malate dehydrogenase.



ΔH_{eq} for *E. coli* malate dehydrogenase is 619 kJ·mol^{-1}. Squares represent the transition from 90% E_{act} (T_{90}) to 50% E_{act} (T_{eq}). Triangles represent the transition from 50% E_{act} (T_{eq}) to 10% E_{act} (T_{10}). Saltires represent K_m values for oxaloacetate.

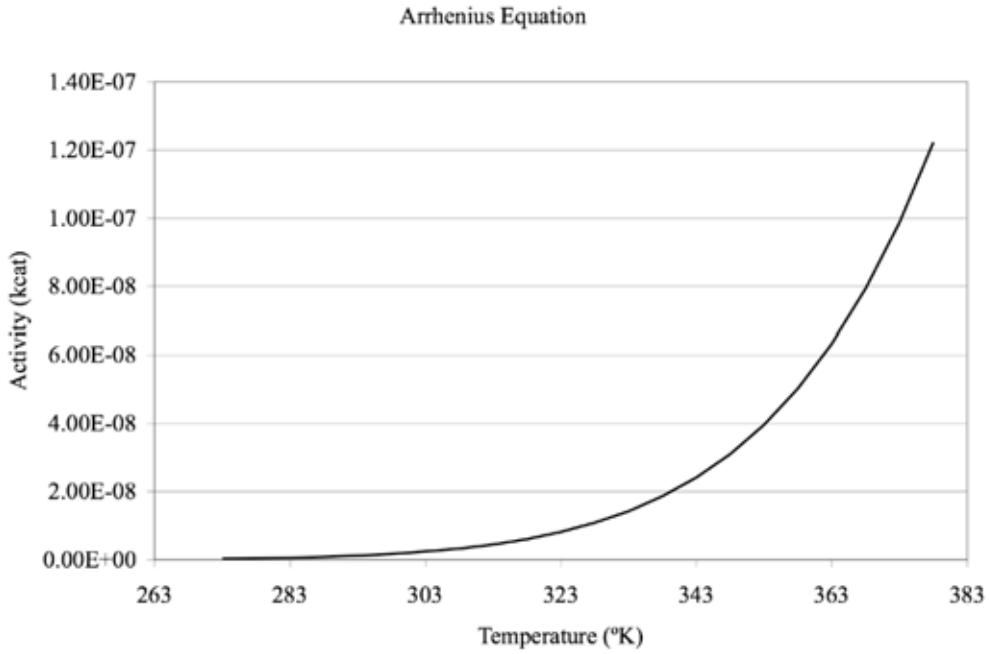
K_m is related to K_s, the dissociation constant for the enzyme–substrate complex (ES), by the following equation:

Equation 6-4: $$K_m = \frac{k_{-1}}{k_1} + \frac{k_2}{k_1} = K_s + \frac{k_2}{k_1}$$

(k_1 is the rate constant for the formation of ES from enzyme and substrate; k_{-1} is the rate constant for the breakdown of ES to enzyme

and substrate; k_2 is the rate constant for the formation of product from ES)

Figure 6-6 – Enzyme activity-temperature relationship as described by the Arrhenius equation.



K_s is inversely related to substrate affinity of the active site, and under certain circumstances (i.e., $k_{-1} >> k_2$), K_m can also be considered as a measure of an active site's affinity for substrate(s): a higher K_m indicates weaker interactions between the active site and the substrate(s). A sharp increase in K_m, such as that observed for APS_IPMDH, potentially means that the active site rapidly loses affinity for substrate(s) as temperature increases, which is likely to be related to structural changes at or around the active site.

Whether an active site loses affinity for substrate(s) rapidly with increasing temperature may depend on the composition of the molecular forces that make up the bindings between the two [27], and if a general correlation between sharp K_m shifts and high ΔH_{eq} values exists, it may provide answers to these molecular forces' involvement in the E_{act}–E_{inact} equilibrium. Thus, the findings presented here may provide an additional piece of evidence regarding the proposed active site-based nature of the E_{act}–E_{inact} equilibrium, and potentially provide further insights and a future research direction into the molecular basis of the Equilibrium Model.

II. Speculations and Future Research Directions

During the course of conducting studies presented in this thesis, some observations were made that are outside the main scope of this thesis. Detailed below are some speculations based on these observations, and suggestions are included for further studies of the Equilibrium Model that may lead to its wider application.

a. The Influence of ΔH_{eq} on $\Delta G^{\ddagger}_{cat}$ determination

Since it has been established that ΔH_{eq} influences apparent enzyme activity (see above and Chapter Three), $\Delta G^{\ddagger}_{cat}$ values determined using the conventional method (see below) based on the Arrhenius activation energy alone and without considering ΔH_{eq} (i.e., under the assumptions of the

Classical Model) may be systematically biased. To investigate this possibility, two different approaches were tested with temperature profiles of select enzymes with extreme ΔH_{eq} values (*Bos taurus* malate dehydrogenase, ΔH_{eq} = 826 kJ·mol^{-1}; and *Geobacillus stearothermophilus* dihydrofolate reductase, ΔH_{eq} = 96 kJ·mol^{-1}).

At temperatures above T_{eq}, most of the enzyme is in the E_{inact} form, which results in a drop of apparent activity that, according to the Classical Model, can only be explain by irreversible thermal denaturation. In the first approach, the temperature profiles were truncated to remove all temperature points above T_{eq}. These points are thus likely to have been excluded from analysis for $\Delta G^{\ddagger}_{cat}$ determinations performed using the conventional method.

The truncated temperature profiles were fitted to the Classical Model, which in theory produces $\Delta G^{\ddagger}_{cat}$ values that are fundamentally the same as $\Delta G^{\ddagger}_{cat}$ values calculated using the conventional method (i.e., using Equation 6-5 in conjunction with an Eyring plot of $\ln(k_{cat}/T)$ on the *y*-axis against T^{-1} on the *x*-axis).

Equation 6-5: $\Delta G_{cat} = \Delta H_{cat} - T \cdot \Delta S_{cat}$

(in an Eyring plot, the intercept of the *y*-axis is equal to $\ln(\frac{k_B}{h}) + \frac{\Delta S_{cat}}{R}$, and the slope is equal to $-\frac{\Delta H_{cat}}{R}$)

Fitting the truncated temperature profiles to the Classical Model should result in $\Delta G^{\ddagger}_{cat}$ values different from those determined by fitting the original temperature profiles to the Equilibrium Model since the E_{act}–E_{inact} equilibrium is not considered in the Classical Model. The results are listed

in the upper half of Table 6-5.

Table 6-5 – $\Delta G^{\ddagger}_{cat}$ and k_{cat} values derived from manipulated *B. taurus* malate dehydrogenase and *G. stearothermophilius* dihydrofolate reductase data sets fitted to the Classical Model and the Equilibrium Model.

Data Description	$\Delta G^{\ddagger}_{cat}$ ($kJ \cdot mol^{-1}$)	k_{cat} (S^{-1})
B. taurus MDH original data fitted to Equilibrium Model	53.0	34,400 (at 60ºC)
B. taurus MDH truncated data fitted to Classical Model	52.7	37,800 (at 60ºC)
G. stearothermophilus DHFR original data fitted to Equilibrium Model	66.6	162 (at 54ºC)
G. stearothermophilus DHFR truncated data fitted to Classical Model	67.2	128 (at 54ºC)
B. taurus MDH original data fitted to Classical Model	53.1	50,100 (at 67ºC)
B. taurus MDH adjusted data fitted to Classical Model	52.4	63,100 (at 67ºC)
G. stearothermophilus DHFR original data fitted to Classical Model	70.6	37 (at 54ºC)
G. stearothermophilus DHFR adjusted data fitted to Classical Model	68.6	76 (at 54ºC)

Standard deviations are only shown where significant. All presented numbers are rounded, but k_{cat} was calculated form raw numbers. k_{cat} values were determined at the respective T_{eq} of the enzymes except *B. taurus* MDH using the first approach, for which experimental data at T_{eq} had been truncated, and k_{cat} was calculated at 60ºC instead.

The second approach required calculating K_{eq} for all temperature points

in the *complete* temperature profiles and using the calculated K_{eq} values to adjust all reaction curves for any reduction in activity due to the E_{act}–E_{inact} equilibrium. In other words, the adjusted temperature profiles were now of theoretical enzymes that follow behaviors described by the Classical Model. The adjusted temperature profiles, along with the originals, were fitted to the Classical Model (there is no point in fitting the adjusted temperature profiles to the Equilibrium Model since they will not fit), and the final parameters for the adjusted and unadjusted temperature profiles are compared in the lower half of Table 6-5.

From Table 6-5, it appears that $\Delta G^{\ddagger}_{cat}$ values derived from fitting truncated temperature profiles to the Classical Model are marginally different from the *correct* one derived from the Equilibrium Model, and the difference is more pronounced for *G. stearothermophilus* DHFR, an enzyme with relatively small ΔH_{eq} value. This is expected, since from a comparison between Figure 6-1 and Figure 6-2, it is clear that the shift from E_{act} to E_{inact} (i.e., deviation from behaviors predicted by the Arrhenius equation) starts to occur at a relatively low temperature (and consequently a greater influence by the E_{act}–E_{inact} equilibrium on the determination of $\Delta G^{\ddagger}_{cat}$ using the conventional method) for enzymes with low ΔH_{eq} values (i.e., eurythermal) than for those with high ΔH_{eq} values. Since $\Delta G^{\ddagger}_{cat}$ is a logarithmic number, a small difference in $\Delta G^{\ddagger}_{cat}$ in fact represents significant changes in enzyme catalytic activity, as reflected by the more significantly different k_{cat} values.

A comparison of parameters obtained using the second method largely reflect findings from the first approach, but the differences in parameters are

more significant. This may be due to the use of complete, rather than truncated, temperature profiles and consequently the inclusion of additional temperature points that are more significantly affected by the E_{act}–E_{inact} equilibrium.

In summary, the findings above suggest that determining $\Delta G^{\ddagger}_{cat}$ values without considering the effects of the E_{act}–E_{inact} equilibrium potentially leads to overestimation of $\Delta G^{\ddagger}_{cat}$ values, and that the bias may be more pronounced for eurythermal, low ΔH_{eq}, enzymes.

b. Other Factors Affecting the Equilibrium Model Parameters

As the Equilibrium Model parameters are activity-based, assay condition variations are likely to affect the resulting parameters to some extent; two such factors that have so far not been investigated are buffer pH and reaction direction, although the latter is essentially a special case of substrate variation. The ionic strength and pH of the reaction buffer can have profound effects on enzyme activity and stability, but reality constrains enzymologists to study enzymes under relatively arbitrary conditions. Many enzymes are capable of catalyzing chemical reactions in both directions, which are not always equally important for cellular functions and may be subjected to different selective pressures; since findings from Chapter Three suggested that the parameters are potentially linked to the physiology and the evolutionary history of the organism, the Equilibrium Model parameters of the two directions may vary.

Table 6-6 – Thermodynamic parameters of *B. taurus* glutamate dehydrogenase and *E. coli* malate dehydrogenase under different assay conditions.

Enzyme	Description	$\Delta G^{\ddagger}_{cat}$ (kJ · mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ · mol^{-1})	ΔH_{eq} (kJ · mol^{-1})	T_{eq} (°C)
B. taurus GDH	Oxidizing NADPH	52.6	91.3	140 ± 11	34 ± 1
	Oxidizing NADPH (zero-time)	52.1	N/A	156 ± 19	33 ± 2
	Reducing $NADP^+$	70.4	90	86 ± 1	41 ± 1
	Reducing $NADP^+$ (zero-time)	71.1	N/A	90 ± 14	43 ± 5
E. coli MDH	Oxidizing NADH (pH 8)	55.3	96.4	619 ± 21	71
	Reducing NAD^+ (pH 8, zero-time)	63.8	N/A	556 ± 241	69 ± 1
	Reducing NAD^+ (pH 9, zero-time)	64.1	N/A	788 ± 292	65

Standard errors are listed where significant. Parameters for *B. taurus* GDH were derived from both full assay data and initial rates to detect potential bias caused by spontaneous reverse reactions. Initial rates were calculated using the regression fit method described in Chapter Two.

To assess the extent of influence these factors may have on the Equilibrium Model, tests were conducted using *Bos taurus* glutamate dehydrogenase (GDH) (assayed in both directions, analyzed with both full assay data and initial rates) and *Escherichia coli* malate dehydrogenase (MDH) (assayed in both directions, pH and buffer variations for one

direction). The parameters are listed in Table 6-6, and their corresponding assay conditions are listed in Table 6-7. Although these parameters were not presented in Table 3-1 due to difficulties in obtaining K_M for those enzymes, they appear reasonable and have indicative value.

Table 6-7 – Assay conditions of parameters listed in Table 6-6.

Enzyme	Description	Substrate	Cofactor(s)	Reaction Buffer	Enzyme Conc.
B. taurus GDH	Oxidizing NADPH	5 mM 2-ketoglutarate	0.75 mM NADPH	100 mM imidazole buffer with 150 mM NH_4Cl and 100 mM NaCl, pH 7	0.4 nM
	Reducing $NADP^+$	12.5 mM glutamate	1 mM $NADP^+$	100 mM K^+ phosphate buffer, pH 7.5	56.2 nM
E. coli MDH	Oxidizing NADH (pH 8)	10 mM oxaloacetate	3 mM NADH	100 mM Na+ phosphate buffer, pH 8	0.45 nM
	Reducing NAD^+ (pH 8, initial rates only)	20 mM malate	20 mM NAD^+, 10 mM semicarbazide	100 mM K+ phosphate buffer, pH 8	6.4 nM
	Reducing NAD^+ (pH 9, initial rates only)	20 mM malate	20 mM NAD^+, 10 mM semicarbazide	50 mM Na+ pyrophosphate buffer, pH 9	12.8 nM

Due to a spontaneous reverse reaction, NAD^+-reducing *E. coli* MDH assays were carried out with an assay length of 30 seconds, from which the initial reaction rates were calculated.

As expected, the $\Delta G^{\ddagger}_{inact}$ values of *B. taurus* GDH assayed in both directions are virtually identical, while the $\Delta G^{\ddagger}_{cat}$ values varied significantly, reflecting the two reaction directions' different catalytic efficiencies (over

140 times more enzyme had to be used for the NADP-reducing reaction to achieve an activity level similar to that of the NADPH-oxidizing reaction). ΔH_{eq} and T_{eq} are both somewhat different between the two reactions, but the differences are difficult to rationalize without further understandings for the underlying mechanism of the E_{act}–E_{inact} equilibrium. Parameters derived from initial rates and the zero-time model are largely similar to those derived from full assay data, albeit with larger errors as expected [20].

At pH 8, the *E. coli* MDH exhibits eight times higher activity in 100 mM K^+ phosphate buffer than in 50 mM Na^+ pyrophosphate buffer at pH 9 (data not shown, and monovalent cations have no effect on *E. coli* MDH activity), likely accounting for the small difference in $\Delta G^{\ddagger}_{cat}$ values of the two pH points ($\Delta G^{\ddagger}_{cat}$ of NADH-oxidizing reaction differed from these two, as expected). When standard deviations are considered, the three ΔH_{eq} values of *E. coli* MDH are essentially identical, and the differences in T_{eq} are small.

It is worth noting that *in vivo*, as part of the aerobic TCA cycle, *E. coli* MDH catalyzes the oxidation of malate to oxaloacetate [177], a reaction with a highly unfavorable standard free energy (~ +7 kCal/mol), in accordance with experimental observations (i.e., malate as a substrate is highly favored by the chemical equilibrium). As a result, *in vivo* the oxidation of malate essentially depends on the removal of oxaloacetate by citrate synthase (oxaloacetate + Acetyl-CoA → citrate, standard free energy = ~ -8 kCal/mol). Therefore, MDH may only be subjected to selective pressure in one reaction direction (reduction of oxaloacetate), potentially explaining the relative similarity of $\Delta G^{\ddagger}_{cat}$, ΔH_{eq}, and T_{eq} between the two reaction directions.

However, such hypotheses are highly speculative and difficult to verify using the Equilibrium Model parameters alone, or even with structural studies.

c. ΔS_{eq} – A Peek Hole into the E_{act}–E_{inact} Equilibrium?

One aspect of the Equilibrium Model that has not received much attention is ΔS_{eq}, the entropy of the reversible enzyme active-inactive state transition. The relationship between ΔH_{eq} and ΔS_{eq} is described by the following equation::

Equation 6-6: $\Delta G_{eq} = \Delta H_{eq} - T \cdot \Delta S_{eq}$

At equilibrium (i.e., $T = T_{eq}$ and $\Delta G_{eq} = 0$), ΔS_{eq} can be calculated using this equation:

Equation 6-7: $\Delta S_{eq} = \Delta H_{eq} / T_{eq}$

For an active site with greater molecular movements (i.e., relatively flexible), its E_{act} form may exist in relatively disordered conformations (i.e., greater number of alternative conformations occupying a broader distribution of conformational states). Also, its relatively flexible structure may entail a smaller ΔH_{eq} since the amount of energy required for conformational changes (change in enthalpy) is less. However, this also means a smaller, less favorable ΔS_{eq} for the E_{act}–E_{inact} inactivation process, which in turn implies that it must undergo a relatively large increase in order. This possibly suggests an additional ordered transitional state between E_{act} and E_{inact} or that

a flexible active site somehow has a more ordered conformation as E_{inact}.

Although ΔH_{eq} has been demonstrated as a measure of enzyme eurythermalism, in reality it may be ΔS_{eq} that dictates how an active site responds to change in temperature. A larger ΔS_{eq} (larger ΔH_{eq}) means that the conversion of E_{act} to E_{inact} is favorable, and so as temperature increases, $[E_{act}]$ decreases relatively more significantly and results in a sharper peak. It is worth noting that, in some cases, it has been shown that protein stability can be achieved through an increase in the entropy of the native state [61].

Speaking from a statistical thermodynamics point of view, if one thinks in terms of entropy and order rather than in enthalpy, then a very large ΔH_{eq} is not particularly remarkable. A rigid active site is likely to have a very ordered conformation (fewer alternative conformations) in its E_{act} state, and during the E_{act}–E_{inact} transition it may undergo a massive increase in ΔS_{eq}, implying that in its E_{inact} state it has many more alternative conformations, some of which involving a large increase in enthalpy.

d. Future Research Directions

To further understand the nature of the E_{act}–E_{inact} equilibrium, protein structural studies are necessary. Since the timescale of the equilibrium may be sub-second [2], NMR studies of the active site coupled with crystal structure of the protein may provide clues on the molecular movements associated with the E_{act}–E_{inact} transition. Once specific types of molecular interactions or amino acid residues are identified, modification of these

intramolecular bonds or residues can be performed, and its effects on the final Equilibrium Model parameters potentially identified.

Alternatively, a directed evolution approach may be the quickest way to facilitate large-scale comparative studies of homologous proteins and also generate enzymes with practical applications under the assumptions of the Equilibrium Model.

Furthermore, with the rapid advancements in computational technology, it may be possible in the near future for individual researchers to have access to one petaFLOPS of computational power (approximately that of the fastest supercomputer in existence as of mid-2007). Should this be the case, it may then be plausible to perform complete *in silico* folding and comprehensive, real-life timescale, dynamic simulation of proteins. This is possibly the Holy Grail for the study and application of the Equilibrium Model since one can then examine effects of all kinds of weak molecular forces within the protein in any timescale, and perform *in silico* modification of proteins based on the assumptions of the Equilibrium Model. Such capabilities would make current enzyme engineering efforts look rudimentary at best. However, understandings for the molecular nature of the Equilibrium Model may be a prerequisite for such *in silico* operations, thus structural studies remain the most important next step for research on the Equilibrium Model.

III. Final Conclusion

By applying the Equilibrium Model to a wide selection of enzymes, studies presented in this thesis have established the Equilibrium Model as a valid tool for studying enzyme temperature adaptation and its parameters as intrinsic and universal properties of enzymes. The application of Equilibrium Model to enzymes derived from a microbial consortium situated at one of the most unique habitats on Earth has enhanced our understandings of thermal adaptation not only for the organisms in question but also extreme environments in general. Furthermore, the studies presented in this thesis have offered additional insights regarding the underlying mechanisms and nature of the Equilibrium Model.

References

1. Daniel, R.M., Danson, M.J., and Eisenthal, R. (2001). The temperature optima of enzymes: a new perspective on an old phenomenon. Trends Biochem Sci *26*, 223-225.
2. Peterson, M.E., Eisenthal, R., Danson, M.J., Spence, A., and Daniel, R.M. (2004). A new intrinsic thermal parameter for enzymes reveals true temperature optima. J Biol Chem *279*, 20717-20722.
3. Daniel, R.M., Danson, M.J., Eisenthal, R., Lee, C.K., and Peterson, M.E. (2007). The effect of temperature on enzyme activity: new insights and their implications. Extremophiles *In Press*.
4. Daniel, R.M. (2007). Protein adaptation to high temperature/Proteins from extreme thermophiles. In Anatomy and Physiology of Extremophilic Proteins (in press), K.S. Siddiqui and T. Thomas, eds. (Hauppauge, New York, USA: Nova Science Publishers).
5. Clarke, A. (2003). Costs and consequences of evolutionary temperature adaptation. Trends in Ecology & Evolution *18*, 573-581.
6. Hughes, T.P., Baird, A.H., Bellwood, D.R., Card, M., Connolly, S.R., Folke, C., Grosberg, R., Hoegh-Guldberg, O., Jackson, J.B., Kleypas, J., et al. (2003). Climate change, human impacts, and the resilience of coral reefs. Science *301*, 929-933.
7. Rothschild, L.J., and Mancinelli, R.L. (2001). Life in extreme environments. Nature *409*, 1092-1101.
8. Macelroy, R.D. (1974). Some comments on the evolution of extremophiles. Biosystems *6*, 74-75.
9. Schiraldi, C., and De Rosa, M. (2002). The production of biocatalysts and biomolecules from extremophiles. Trends in biotechnology *20*, 515-521.
10. Daniel, R.M., and Danson, M.J. (2001). Assaying activity and assessing thermostability of hyperthermophilic enzymes. In Methods in Enzymology, Volume 334, M.W.W. Adams and R.M. Kelly, eds. (Academic Press), pp. 283-293.
11. Cornish-Bowden, A. (1995). In Fundamentals of Enzyme Kinetics. (London, UK: Portland Press Ltd), pp. 193-197.
12. Thomas, T.M., and Scopes, R.K. (1998). The effects of temperature on the

kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. Biochem J *330 (Pt 3)*, 1087-1095.

13. Buchanan, C.L., Connaris, H., Danson, M.J., Reeve, C.D., and Hough, D.W. (1999). An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. Biochem J *343 Pt 3*, 563-570.
14. Arnott, M.A., Michael, R.A., Thompson, C.R., Hough, D.W., and Danson, M.J. (2000). Thermostability and thermoactivity of citrate synthases from the thermophilic and hyperthermophilic archaea, *Thermoplasma acidophilum* and *Pyrococcus furiosus*. Journal of molecular biology *304*, 657-668.
15. Medina, D.C., Hanna, E., MacRae, I.J., Fisher, A.J., and Segel, I.H. (2001). Temperature effects on the allosteric transition of ATP sulfurylase from *Penicillium chrysogenum*. Archives of biochemistry and biophysics *393*, 51-60.
16. Gerike, U., Danson, M.J., Russell, N.J., and Hough, D.W. (1997). Sequencing and expression of the gene encoding a cold-active citrate synthase from an Antarctic bacterium, strain DS2-3R. Eur J Biochem *248*, 49-57.
17. Svingor, A., Kardos, J., Hajdu, I., Nemeth, A., and Zavodszky, P. (2001). A better enzyme to cope with cold. Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic IPMDHs. J Biol Chem *276*, 28121-28125.
18. Stillman, J.H., and Somero, G.N. (2001). A comparative analysis of the evolutionary patterning and mechanistic bases of lactate dehydrogenase thermal stability in porcelain crabs, genus *Petrolisthes*. The Journal of experimental biology *204*, 767-776.
19. Peterson, M.E. (2005). PhD Thesis: Evidence for a third thermal parameter of enzymes, (Hamilton, New Zealand: University of Waikato).
20. Peterson, M.E., Daniel, R. M., Danson, M. J., and Eisenthal, R. (2007). The dependence of enzyme activity on temperature: determination and validation of parameters. Biochem J *402*, 331-337.
21. Fulton, K.F., Devlin, G.L., Jodun, R.A., Silvestri, L., Bottomley, S.P., Fersht, A.R., and Buckle, A.M. (2005). PFD: a database for the investigation of protein folding kinetics and stability. Nucleic acids

research *33*, D279-283.

22. Eisenthal, R., Peterson, M. E., Daniel, R. M., Danson, M. J. (2006). The thermal behaviour of enzymes: implications for biotechnology. Trends in biotechnology *24*, 289-292.
23. Edwards, R.A., Jacobson, A.L., and Huber, R.E. (1990). Thermal denaturation of beta-galactosidase and of two site-specific mutants. Biochemistry *29*, 11001-11008.
24. Zhang, Y.L., Zhou, J.M., and Tsou, C.L. (1993). Inactivation precedes conformation change during thermal denaturation of adenylate kinase. Biochimica et biophysica acta *1164*, 61-67.
25. Shoichet, B.K., Baase, W.A., Kuroki, R., and Matthews, B.W. (1995). A relationship between protein stability and protein function. Proc Natl Acad Sci USA *92*, 452-456.
26. D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M.A., Feller, G., and Gerday, C. (2002). Molecular basis of cold adaptation. Philos Trans R Soc Lond B Biol Sci *357*, 917-925.
27. Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J.C., Sonan, G., Feller, G., and Gerday, C. (2004). Some like it cold: biocatalysis at low temperatures. FEMS Microbiol Rev *28*, 25-42.
28. Lonhienne, T., Gerday, C., and Feller, G. (2000). Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. Biochimica et biophysica acta *1543*, 1-10.
29. Daniel, R.M., Cowan, D.A., Morgan, H.W., and Curran, M.P. (1982). A correlation between protein thermostability and resistance to proteolysis. Biochem J *207*, 641-644.
30. Somero, G.N. (2004). Adaptation of enzymes to temperature: searching for basic "strategies". Comparative biochemistry and physiology *139*, 321-333.
31. Bullock, T.H. (1955). Compensation for temperature in the metabolism and activity of poikilotherms. Biol Rev *30*, 311-342.
32. Hochachka, P.W., Hayes, F.R. (1962). The effect of temperature acclimation on pathways of glucose metabolism in the trout. Can. J. Zool. *40*, 261-270.
33. Somero, G.N., and Suarez, R.K. (2005). Peter Hochachka: adventures in biochemical adaptation. Annual review of physiology *67*, 25-37.

34. Hochachka, P.W. (1965). Isoenzymes in metabolic adaptation of a poikilotherm: subunit relationships in lactic dehydrogenases of goldfish. Archives of biochemistry and biophysics *111*, 96-103.
35. Crawford, D.L., and Powers, D.A. (1989). Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. Proceedings of the National Academy of Sciences of the United States of America *86*, 9365-9369.
36. Crawford, D.L., and Powers, D.A. (1992). Evolutionary adaptation to different thermal environments via transcriptional regulation. Molecular biology and evolution *9*, 806-813.
37. Allen, M.B. (1953). The thermophilic aerobic sporeforming bacteria. Bacteriological reviews *17*, 125-173.
38. Bubela, B., and Holdsworth, E.S. (1966). Protein synthesis in *Bacillus stearothermophilus*. Biochimica et biophysica acta *123*, 376-389.
39. Hochachka, P.W.S., G.N. (1984). Temperature Adaptation. In Biochemical Adaptations, P.W.S. Hochachka, G.N., ed. (Princeton, NJ: Princeton University Press), pp. 355-449.
40. Somero, G.N. (1995). Proteins and temperature. Annual review of physiology *57*, 43-68.
41. Kengen, S.W., Tuininga, J.E., de Bok, F.A., Stams, A.J., and de Vos, W.M. (1995). Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Biol Chem *270*, 30453-30457.
42. Butler, P.J., and Day, N. (1993). The relationship between intracellular pH and seasonal temperature in the brown trout *Salmo trutta*. The Journal of experimental biology *177*, 293-297.
43. Burton, R.F. (2002). Temperature and acid-base balance in ectothermic vertebrates: the imidazole alphastat hypotheses and beyond. The Journal of experimental biology *205*, 3587-3600.
44. Fenton, W.A., and Horwich, A.L. (2003). Chaperonin-mediated protein folding: fate of substrate polypeptide. Quarterly reviews of biophysics *36*, 229-256.
45. Feder, J.H., Rossi, J.M., Solomon, J., Solomon, N., and Lindquist, S. (1992). The consequences of expressing hsp70 in *Drosophila* cells at normal temperatures. Genes & development *6*, 1402-1413.

46. Sterner, R., Kleemann, G.R., Szadkowski, H., Lustig, A., Hennig, M., and Kirschner, K. (1996). Phosphoribosyl anthranilate isomerase from *Thermotoga maritima* is an extremely stable and active homodimer. Protein Sci *5*, 2000-2008.
47. Ogasahara, K., Khechinashvili, N.N., Nakamura, M., Yoshimoto, T., and Yutani, K. (2001). Thermal stability of pyrrolidone carboxyl peptidases from the hyperthermophilic Archaeon, *Pyrococcus furiosus*. Eur J Biochem *268*, 3233-3242.
48. Van den Burg, B., Vriend, G., Veltman, O.R., Venema, G., and Eijsink, V.G. (1998). Engineering an enzyme to resist boiling. Proceedings of the National Academy of Sciences of the United States of America *95*, 2056-2060.
49. Eisenthal, R., Danson, M.J., and Hough, D.W. (2007). Catalytic efficiency and k(cat)/K(M): a useful comparator? Trends in biotechnology.
50. Kashefi, K., and Lovley, D.R. (2003). Extending the upper temperature limit for life. Science *301*, 934.
51. Auman, A.J., Breezee, J.L., Gosink, J.J., Kampfer, P., and Staley, J.T. (2006). *Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. Int J Syst Evol Microbiol *56*, 1001-1007.
52. Wharton, D., and Ferns, D. (1995). Survival of intracellular freezing by the Antarctic nematode *Panagrolaimus davidi*. The Journal of experimental biology *198*, 1381-1387.
53. Marshall, C.J. (1997). Cold-adapted enzymes. Trends in biotechnology *15*, 359-364.
54. Somero, G.N., Giese, A.C., Wohlschlag, D.E. (1968). Cold adaptation of the Antarctic fish *Trematomus bernacchii*. Comp. Biochem. Physiol. *26*, 223-233.
55. Sharpe, M., Love, C., and Marshall, C. (2001). Lactate dehydrogenase from the Antarctic eelpout, *Lycodichthys dearborni*. Polar Biology *24*, 258-269.
56. Brock, T.D., and Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. Journal of bacteriology *98*, 289-297.
57. Glasemacher, J., Bock, A.K., Schmid, R., and Schonheit, P. (1997).

Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. Eur J Biochem *244*, 561-567.

58. Bock, A.K., Glasemacher, J., Schmidt, R., and Schonheit, P. (1999). Purification and characterization of two extremely thermostable enzymes, phosphate acetyltransferase and acetate kinase, from the hyperthermophilic eubacterium *Thermotoga maritima*. Journal of bacteriology *181*, 1861-1867.
59. Musfeldt, M., and Schonheit, P. (2002). Novel type of ADP-forming acetyl coenzyme A synthetase in hyperthermophilic archaea: heterologous expression and characterization of isoenzymes from the sulfate reducer *Archaeoglobus fulgidus* and the methanogen *Methanococcus jannaschii*. Journal of bacteriology *184*, 636-644.
60. Daniel, R.M., Dines, M., and Petach, H.H. (1996). The denaturation and degradation of stable enzymes at high temperatures. Biochem J *317 (Pt 1)*, 1-11.
61. Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L.H. (1997). Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. Journal of molecular biology *271*, 789-802.
62. Ichikawa, J.K., and Clarke, S. (1998). A highly active protein repair enzyme from an extreme thermophile: the L-isoaspartyl methyltransferase from *Thermotoga maritima*. Archives of biochemistry and biophysics *358*, 222-231.
63. Merz, A., Knochel, T., Jansonius, J.N., and Kirschner, K. (1999). The hyperthermostable indoleglycerol phosphate synthase from *Thermotoga maritima* is destabilized by mutational disruption of two solvent-exposed salt bridges. Journal of molecular biology *288*, 753-763.
64. Jaenicke, R. (1996). Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. FASEB J *10*, 84-92.
65. Daniel, R.M., and Cowan, D.A. (2000). Biomolecular stability and life at high temperatures. Cell Mol Life Sci *57*, 250-264.
66. Zierenberg, R.A., Adams, M.W.W., and Arp, A.J. (2000). Life in Extreme Environments: Hydrothermal Vents. Proceedings of the National Academy of Sciences of the United States of America *97*, 12961-12962.

67. Cary, S.C., Shank, T., and Stein, J. (1998). Worms bask in extreme temperatures. Nature *391*, 545-546.
68. Tunnicliffe, V. (1992). The Nature and Origin of the Modern Hydrothermal Vent Fauna. PALAIOS *7*, 338-350.
69. Jeanthon, C. (2000). Molecular ecology of hydrothermal vent microbial communities. Antonie van Leeuwenhoek *77*, 117-133.
70. Alain, K., Zbinden, M., Le Bris, N., Lesongeur, F., Querellou, J., Gaill, F., and Cambon-Bonavita, M.A. (2004). Early steps in microbial colonization processes at deep-sea hydrothermal vents. Environmental microbiology *6*, 227-241.
71. Alayse-Danet, A.M., Desbruyeres, D., and Gaill, F. (1987). The possible nutritional or detoxification role of the epibiotic bacteria of *Alvinellid* polychaetes: review of current data. Symbiosis (Philadelphia, PA) *4*, 51-61.
72. Hourdez, S., Lallier, F.H., De Cian, M.C., Green, B.N., Weber, R.E., and Toulmond, A. (2000). Gas transfer system in *Alvinella pompejana* (*Annelida polychaeta, Terebellida*): functional properties of intracellular and extracellular hemoglobins. Physiol Biochem Zool *73*, 365-373.
73. Terwilliger, N.B., Terwilliger, R. C. (1984). Hemoglobin from the «Pompeii worm», *Alvinella pompejana*, an annelid from a deep sea hot hydrothermal vent environment. Marine Biology Letters *5*, 191-201.
74. Luther, G.W., 3rd, Rozan, T.F., Taillefert, M., Nuzzio, D.B., Di Meo, C., Shank, T.M., Lutz, R.A., and Cary, S.C. (2001). Chemical speciation drives hydrothermal vent ecology. Nature *410*, 813-816.
75. Di Meo-Savoie, C.A., Luther, G.W., and Cary, S.C. (2004). Physicochemical characterization of the microhabitat of the epibionts associated with *Alvinella pompejana*, a hydrothermal vent annelid. Geochim. Cosmochim. Acta *68*, 2055-2066.
76. Chevaldonné, P., Desbruyéres, D., and Childress, J.J. (1992). Some like it hot. . . and some even hotter. Nature *359*, 593-594.
77. Desbruyères, D., Chevaldonné, P., Alayse, A.M., Jollivet, D., Lallier, F.H., Jouin-Toulmond, C., Zal, F., Sarradin, P.M., Cosson, R., and Caprais, J.C. (1998). Biology and ecology of the "Pompeii worm" (*Alvinella pompejana* Desbruyeres and Laubier), a normal dweller of an extreme deep-sea environment: A synthesis of current knowledge and recent developments.

Deep Sea Research Part II: Topical Studies in Oceanography *45*, 383-422.

78. Chevaldonné, P., Fisher, C.R., Childress, J.J., Desbruyères, D., Jollivet, D., Zal, F., and Toulmond, A. (2000). Thermotolerance and the 'Pompeii worms'. Mar. Ecol. Prog. Ser *208*, 293–295.

79. Lee, R.W. (2003). Thermal Tolerances of Deep-Sea Hydrothermal Vent Animals from the Northeast Pacific. Biological Bulletin *205*, 98-101.

80. Dixon, D.R., Simpson-White, R., Dixon, L. R. J. (1992). Evidence for thermal stability of ribosomal DNA sequences in hydrothermal-vent organisms. Journal of the Marine Biological Association of the United Kingdom *72*, 519-527.

81. Cary, S.C., Cottrell, M.T., Stein, J.L., Camacho, F., and Desbruyeres, D. (1997). Molecular Identification and Localization of Filamentous Symbiotic Bacteria Associated with the Hydrothermal Vent Annelid *Alvinella pompejana*. Applied and environmental microbiology *63*, 1124-1130.

82. Haddad, A., Camacho, F., Durand, P., and Cary, S.C. (1995). Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. Applied and environmental microbiology *61*, 1679-1687.

83. Prieur, D., Chamroux, S., Durand, P., Erauso, G., Ph, F., Jeanthon, C., Borgne, L.L., Mével, G., and Vincent, P. (1990). Metabolic diversity in epibiotic microflora associated with the Pompeii worms *Alvinella pompejana* and *A. caudata* (Polychaetae: Annelida) from deep-sea hydrothermal vents. Marine Biology *106*, 361-367.

84. Moyer, C.L., Dobbs, F.C., and Karl, D.M. (1995). Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Applied and environmental microbiology *61*, 1555-1562.

85. Polz, M.F., and Cavanaugh, C.M. (1995). Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. Proceedings of the National Academy of Sciences of the United States of America *92*, 7232-7236.

86. Felbeck, H., and Jarchow, J. (1998). Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. Physiological zoology *71*, 294-302.

87. Laue, B.E., and Nelson, D.C. (1994). Characterization of the gene encoding the autotrophic ATP sulfurylase from the bacterial endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. Journal of bacteriology *176*, 3723-3729.
88. Gaill, F., Desbruyères, D., and Laubier, L. (1988). Relationships between the Pompeii worms and their epibiotic bacteria. Oceanol. Acta *8*, 147–154.
89. Campbell, B.J., Stein, J.L., and Cary, S.C. (2003). Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaete. Applied and environmental microbiology *69*, 5070-5078.
90. Cottrell, M.T., and Cary, S.C. (1999). Diversity of dissimilatory bisulfite reductase genes of bacteria associated with the deep-sea hydrothermal vent polychaete annelid *Alvinella pompejana*. Applied and environmental microbiology *65*, 1127-1132.
91. Jeanthon, C., and Prieur, D. (1990). Susceptibility to Heavy Metals and Characterization of Heterotrophic Bacteria Isolated from Two Hydrothermal Vent Polychaete Annelids, *Alvinella pompejana* and *Alvinella caudata*. Applied and environmental microbiology *56*, 3308-3314.
92. Gaill, F., Desbruyères, D., and Prieur, D. (1987). Bacterial communities associated with "Pompei worms" from the East Pacific rise hydrothermal vents: SEM, TEM observations. Microbial Ecology *V13*, 129-139.
93. Chen, K., and Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS computational biology *1*, 106-112.
94. Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., 2nd, Shah, M., Hettich, R.L., and Banfield, J.F. (2005). Community proteomics of a natural microbial biofilm. Science *308*, 1915-1920.
95. Stahl, D.A., Lane, D.J., Olsen, G.J., and Pace, N.R. (1985). Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. Applied and environmental microbiology *49*, 1379-1384.
96. Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., et al.

(2005). Comparative metagenomics of microbial communities. Science *308*, 554-557.

97. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature *428*, 37-43.

98. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science *304*, 66-74.

99. Cary, S.C. (2002). Meta-genome analysis of extreme microbial symbioses.

100. Lu, W.P., and Fei, L. (2003). A logarithmic approximation to initial rates of enzyme reactions. Analytical biochemistry *316*, 58-65.

101. Notomista, E., Catanzano, F., Graziano, G., Dal Piaz, F., Barone, G., D'Alessio, G., and Di Donato, A. (2000). Onconase: an unusually stable protein. Biochemistry *39*, 8711-8718.

102. Padmanabhan, S., Zhou, K., Chu, C.Y., Lim, R.W., and Lim, L.W. (2001). Overexpression, biophysical characterization, and crystallization of ribonuclease I from *Escherichia coli*, a broad-specificity enzyme in the RNase T2 family. Archives of biochemistry and biophysics *390*, 42-50.

103. Noda, L., and Kuby, S.A. (1957). Adenosine triphosphate-adenosine monophosphate transphosphorylase (myokinase). I. Isolation of the crystalline enzyme from rabbit skeletal muscle. J Biol Chem *226*, 541-549.

104. Tsuboi, K.K., and Chervenka, C.H. (1975). Adenylate kinase of human erythrocyte. Isolation and properties of the predominant inherited form. J Biol Chem *250*, 132-140.

105. Calera, J.A., Sanchez-Weatherby, J., Lopez-Medrano, R., and Leal, F. (2000). Distinctive properties of the catalase B of *Aspergillus nidulans*. FEBS Lett *475*, 117-120.

106. Diaz, A., Rangel, P., Montes de Oca, Y., Lledias, F., and Hansberg, W. (2001). Molecular and kinetic study of catalase-1, a durable large catalase of *Neurospora crassa*. Free Radic Biol Med *31*, 1323-1333.

107. Lopez-Medrano, R., Ovejero, M.C., Calera, J.A., Puente, P., and Leal, F.

(1995). An immunodominant 90-kilodalton *Aspergillus fumigatus* antigen is the subunit of a catalase. Infect Immun *63*, 4774-4780.

108. Switala, J., O'Neil, J.O., and Loewen, P.C. (1999). Catalase HPII from *Escherichia coli* exhibits enhanced resistance to denaturation. Biochemistry *38*, 3895-3901.

109. Wasserman, B.P., and Hultin, H.O. (1981). Effect of deglycosylation on the stability of *Aspergillus niger* catalase. Archives of biochemistry and biophysics *212*, 385-392.

110. Owusu, R.K., and Cowan, D.A. (1989). Correlation between microbial protein thermostability and resistance to denaturation in aqueous: organic solvent two-phase systems. Enzyme and Microbial Technology *11*, 568.

111. Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry *72*, 248-254.

112. StatSoft Inc (2005). STATISTICA, version 7.1. (Tulsa, OK 74104 USA). www.statsoft.com.

113. Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science *276*, 734-740.

114. Schwartzman, D.W., and Lineweaver, C.H. (2004). The hyperthermophilic origin of life revisited. Biochem Soc Trans *32*, 168-171.

115. Wachtershauser, G. (1998). In Thermophiles: the Keys to Molecular Evolution and the Origin of Life?, K. Wiegel, Adams, M., ed. (Philadelphia: Taylor and Francis), pp. 47-57.

116. Johns, G.C., and Somero, G.N. (2004). Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydrogenases of Pacific damselfishes (*Chromis* spp.). Molecular biology and evolution *21*, 314-320.

117. Flint, K.P. (1987). The long-term survival of *Escherichia coli* in river water. J Appl Bacteriol *63*, 261-270.

118. Hartl, D.L., and Dykhuizen, D.E. (1984). The population genetics of *Escherichia coli*. Annu Rev Genet *18*, 31-68.

119. Xu, Y., Feller, G., Gerday, C., and Glansdorff, N. (2003). *Moritella* cold-active dihydrofolate reductase: are there natural limits to optimization of catalytic efficiency at low temperature? Journal of bacteriology *185*,

5519-5526.

120. Xu, Y., Nogi, Y., Kato, C., Liang, Z., Ruger, H.J., De Kegel, D., and Glansdorff, N. (2003). *Moritella profunda* sp. nov. and *Moritella abyssi* sp. nov., two psychropiezophilic organisms isolated from deep Atlantic sediments. Int J Syst Evol Microbiol *53*, 533-538.
121. Daniel, R.M., Dunn, R.V., Finney, J.L., and Smith, J.C. (2003). The role of dynamics in enzyme activity. Annu Rev Biophys Biomol Struct *32*, 69-92.
122. Pappenberger, G., Schurig, H., and Jaenicke, R. (1997). 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 *274*, 676.
123. Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V.N., Gerday, C., and Feller, G. (2003). Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J Biol Chem *278*, 37015-37023.
124. Gershenson, A., Schauerte, J.A., Giver, L., and Arnold, F.H. (2000). Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. Biochemistry *39*, 4658-4665.
125. Miyazaki, K., Wintrode, P.L., Grayling, R.A., Rubingh, D.N., and Arnold, F.H. (2000). Directed evolution study of temperature adaptation in a psychrophilic enzyme. Journal of molecular biology *297*, 1015-1026.
126. Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev *68*, 669-685.
127. Staley, J.T., and Konopka, A. (1985). Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual review of microbiology *39*, 321-346.
128. Brock, T.D., and Brock, M.L. (1968). Measurement of steady-state growth rates of a thermophilic alga directly in nature. Journal of bacteriology *95*, 811-815.
129. Colwell, R.R., and Grims, D.J. ed. (2000). Nonculturable microorganisms in the environment. (Washington, D.C.: ASM Press).
130. Marshall, B.J., McGechie, D.B., Rogers, P.A., and Glancy, R.J. (1985). *Pyloric Campylobacter* infection and gastroduodenal disease. The Medical journal of Australia *142*, 439-444.

131. Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences of the United States of America *74*, 5088-5090.
132. Woese, C.R. (1987). Bacterial evolution. Microbiological reviews *51*, 221-271.
133. Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences of the United States of America *82*, 6955-6959.
134. Neefs, J.M., Van de Peer, Y., Hendriks, L., and De Wachter, R. (1990). Compilation of small ribosomal subunit RNA sequences. Nucleic acids research *18 Suppl*, 2237-2317.
135. Muyzer, G., and Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie van Leeuwenhoek *73*, 127-141.
136. Rappe, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annual review of microbiology *57*, 369-394.
137. DeLong, E.F. (2000). Extreme genomes. Genome biology *1*, REVIEWS1029.
138. Hawkins, T.L., Detter, J.C., and Richardson, P.M. (2002). Whole genome amplification--applications and advances. Current opinion in biotechnology *13*, 65-67.
139. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & biology *5*, R245-249.
140. Knietsch, A., Bowien, S., Whited, G., Gottschalk, G., and Daniel, R. (2003). Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Applied and environmental microbiology *69*, 3048-3060.
141. Majernik, A., Gottschalk, G., and Daniel, R. (2001). Screening of environmental DNA libraries for the presence of genes conferring Na(+)(Li(+))/H(+) antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. Journal of

bacteriology *183*, 6645-6653.

142. Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., et al. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Applied and environmental microbiology *69*, 49-55.

143. MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Phillips, T., Narula, S., Sundaramoorthi, R., et al. (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. Journal of molecular microbiology and biotechnology *3*, 301-308.

144. Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. Journal of bacteriology *178*, 591-599.

145. Riesenfeld, C.S., Goodman, R.M., and Handelsman, J. (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environmental microbiology *6*, 981-989.

146. Schloss, P.D., and Handelsman, J. (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome biology *6*, 229.

147. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.

148. Sebat, J.L., Colwell, F.S., and Crawford, R.L. (2003). Metagenomic profiling: microarray analysis of an environmental genomic library. Applied and environmental microbiology *69*, 4927-4934.

149. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003). The COG database: an updated version includes eukaryotes. BMC bioinformatics *4*, 41.

150. Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., et al. (2006). Pfam: clans, web tools and services. Nucleic acids research *34*,

D247-251.

151. Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D. (2003). Enzyme-specific profiles for genome annotation: PRIAM. Nucleic acids research *31*, 6633-6639.

152. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. Nucleic acids research *34*, D354-357.

153. Li, W., Jaroszewski, L., and Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics (Oxford, England) *17*, 282-283.

154. Schloss, P.D., and Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Applied and environmental microbiology *71*, 1501-1506.

155. Ponnuswamy, P., Muthusamy, R., and Manavalan, P (1982). Amino acid composition and thermal stability of globular proteins. Int J Biol Macromol *4*, 186-190.

156. Kreil, D.P., and Ouzounis, C.A. (2001). Identification of thermophilic species by the amino acid compositions deduced from their genomes. Nucleic acids research *29*, 1608-1615.

157. Zeldovich, K.B., Berezovsky, I.N., and Shakhnovich, E.I. (2007). Protein and DNA Sequence Determinants of Thermophilic Adaptation. PLoS computational biology *3*, e5.

158. Lee, C.K., Daniel, R.M., Shepherd, C., Saul, D., Cary, S.C., Danson, M.J., Eisenthal, R., and Peterson, M.E. (2007). Eurythermalism and the temperature dependence of enzyme activity. FASEB J *21*, 1934-1941.

159. Siebers, B., and Schonheit, P. (2005). Unusual pathways and enzymes of central carbohydrate metabolism in *Archaea*. Current opinion in microbiology *8*, 695-705.

160. Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y., and Tomita, M. (2006). GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. BMC bioinformatics *7*, 168.

161. Ward, N., Larsen, O., Sakwa, J., Bruseth, L., Khouri, H., Durkin, A.S., Dimitrov, G., Jiang, L., Scanlan, D., Kang, K.H., et al. (2004). Genomic insights into methanotrophy: the complete genome sequence of

Methylococcus capsulatus (Bath). PLoS biology *2*, e303.

162. Muresu, R., Sulas, L., Polone, E., and Squartini, A. (2005). PCR primers based on different portions of insertion elements can assist genetic relatedness studies, strain fingerprinting and species identification in rhizobia. FEMS microbiology ecology *54*, 445-453.
163. Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Current opinion in biotechnology *6*, 494-500.
164. Bohm, G., and Jaenicke, R. (1994). Relevance of sequence statistics for the properties of extremophilic proteins. International journal of peptide and protein research *43*, 97-106.
165. Somero, G.N. (2005). Linking biogeography to physiology: Evolutionary and acclimatory adjustments of thermal limits. Frontiers in Zoology *2*, 1.
166. Campbell, B.J., Jeanthon, C., Kostka, J.E., Luther, G.W., 3rd, and Cary, S.C. (2001). Growth and phylogenetic properties of novel bacteria belonging to the epsilon subdivision of the *Proteobacteria* enriched from *Alvinella pompejana* and deep-sea hydrothermal vents. Applied and environmental microbiology *67*, 4566-4572.
167. Le Bris, N., Zbinden, M., and Gaill, F. (2005). Processes controlling the physico-chemical micro-environments associated with Pompeii worms. Deep Sea Research Part I: Oceanographic Research Papers *52*, 1071-1083.
168. Dahlhoff, E.P., J. O'Brien, G. N. Somero and R. D. Vetter (1991). Temperature effects on mitochondria from hydrothermal vent invertebrates: Evidence for adaptation to elevated and variable habitat temperatures. Physiological zoology *64*, 1490-1508.
169. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1992). Preparation of Genomic DNA from Bacteria. In Current Protocols in Molecular Biology., E.W. Harkins, ed. (New York City: Greene Publishing Associates and Wiley-Interscience.).
170. Rueckert, A., and Morgan, H.W. (2006). Removal of contaminating DNA from polymerase chain reaction using ethidium monoazide. J Microbiol Methods.
171. Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology *132*, 365-386.

172. Nurachman, Z., Akanuma, S., Sato, T., Oshima, T., and Tanaka, N. (2000). Crystal structures of 3-isopropylmalate dehydrogenases with mutations at the C-terminus: crystallographic analyses of structure-stability relationships. Protein engineering *13*, 253-258.

173. Wallon, G., Lovett, S.T., Magyar, C., Svingor, A., Szilagyi, A., Zavodszky, P., Ringe, D., and Petsko, G.A. (1997). Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium *Vibrio* sp. I5 suggest reasons for thermal instability. Protein engineering *10*, 665-672.

174. Hudson, R.C., and Daniel, R.M. (1993). L-glutamate dehydrogenases: distribution, properties and mechanism. Comp Biochem Physiol B *106*, 767-792.

175. Liang, Z.X., Tsigos, I., Bouriotis, V., and Klinman, J.P. (2004). Impact of protein flexibility on hydride-transfer parameters in thermophilic and psychrophilic alcohol dehydrogenases. Journal of the American Chemical Society *126*, 9500-9501.

176. De Vriendt, K., Theunissen, S., Carpentier, W., De Smet, L., Devreese, B., and Van Beeumen, J. (2005). Proteomics of *Shewanella oneidensis* MR-1 biofilm reveals differentially expressed proteins, including AggA and RibB. Proteomics *5*, 1308-1316.

177. Park, S.J., Cotter, P.A., and Gunsalus, R.P. (1995). Regulation of malate dehydrogenase (mdh) gene expression in *Escherichia coli* in response to oxygen, carbon, and heme availability. Journal of bacteriology *177*, 6652-6656.

178. Pozmogova, I.N., Shkop, I., Tsvid, E.E., Riabchuk, V.A., and Vinarov, A. (1982). Chemostat cultivation of *Candida utilis* yeasts under nonstationary temperature conditions. Mikrobiologiia *51*, 236-239.

179. Madigan, M.T., Martinko, J.M., Parker, J. (2000). In Brock Biology of Microorganisms, Ninth Edition, P.F. Corey, ed. (Upper Saddle River, New Jersey: Prentice-Hall, Inc.), p. 148.

180. Gonthier, A., Guerin-Faublee, V., Tilly, B., and Delignette-Muller, M.L. (2001). Optimal growth temperature of O157 and non-O157 *Escherichia coli* strains. Lett Appl Microbiol *33*, 352-356.

181. Rosso, L., Lobry, J.R., and Flandrois, J.P. (1993). An unexpected correlation between cardinal temperatures of microbial growth highlighted

by a new model. J Theor Biol *162*, 447-463.

182. Slepecky, R.A., Hemphill, H. E. (1992). In The Prokaryotes, Volume 2, 2 Edition, A. Balows, Truper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H., ed. (New York: Springer), pp. 1663-1696.

183. Jensen, G.B., Hansen, B.M., Eilenberg, J., and Mahillon, J. (2003). The hidden lifestyles of *Bacillus cereus* and relatives. Environmental microbiology *5*, 631-640.

184. Margulis, L., Jorgensen, J.Z., Dolan, S., Kolchinsky, R., Rainey, F.A., and Lo, S.C. (1998). The Arthromitus stage of *Bacillus cereus*: intestinal symbionts of animals. Proceedings of the National Academy of Sciences of the United States of America *95*, 1236-1241.

185. Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V.V., Osipov, G.A., Belyaev, S.S., et al. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. Int J Syst Evol Microbiol *51*, 433-446.

186. Stenesh, J., and Schechter, N. (1969). Cell-free amino acid-incorporating systems from *Bacillus licheniformis* and *Bacillus stearothermophilus* 10. Journal of bacteriology *98*, 1258-1262.

187. Kobori, H., Sullivan, C.W., and Shizuya, H. (1984). Heat-Labile Alkaline Phosphatase from Antarctic Bacteria: Rapid 5' End-Labeling of Nucleic Acids. PNAS *81*, 6691-6695.

188. Hartog, A.T., and Daniel, R.M. (1992). An alkaline phosphatase from *Thermus sp* strain Rt41A. International Journal of Biochemistry *24*, 1657.

189. Lim, S.H. (1983). Master Thesis: A protease from a *Thermus* spp., Volume Master Thesis, (Hamilton, New Zealand: University of Waikato).

190. Hammond, P.M., Price, C.P., and Scawen, M.D. (1983). Purification and properties of aryl acylamidase from *Pseudomonas fluorescens* ATCC 39004. Eur J Biochem *132*, 651-655.

191. Chevillotte-Brivet, P., and Meunier-Lemesle, D. (1980). Cytochrome b-565 in *Saccharomyces cerevisiae*: use of mutants in the cob- box region of the mitochondrial DNA to study the functional role of this spectral species of cytochrome b. 2. Relationship between energetic data and cytochrome b-565 content. Eur J Biochem *111*, 161-169.

192. Malcolm, P. (2004). Effects of Root Temperatures and Genotype on the Growth of Peach Rootstocks and Their Scions. In Centre for Horticulture and Plant Science, Volume Doctor of Philosophy. (Richmond, Australia: University of Western Sydney (Hawkesbury)), p. 301.

193. Rainey, F.A., Donnison, A.M., Janssen, P.H., Saul, D., Rodrigo, A., Bergquist, P.L., Daniel, R.M., Stackebrandt, E., and Morgan, H.W. (1994). Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiol Lett *120*, 263-266.

194. Rainey, F.A., Ward, N.L., Morgan, H.W., Toalster, R., and Stackebrandt, E. (1993). Phylogenetic analysis of anaerobic thermophilic bacteria: aid for their reclassification. Journal of bacteriology *175*, 4772-4779.

195. Sissons, C.H., Sharrock, K.R., Daniel, D.M., and Morgan H.W. (1987). Isolation of Cellulolytic Anaerobic Extreme Thermophiles from New Zealand Thermal Sites. Appl. Environ. Microbiol. *53*, 832-838.

196. Fritze, D. (2002). Bacillus Identification - Traditional Approaches. In Applications and Systematics of Bacillus and Relatives, R. Berkeley, ed. (Oxford, UK: Blackwell Publishing Ltd.).

197. Heinen, U.J., and Heinen, W. (1972). Characteristics and properties of a caldo-active bacterium producing extracellular enzymes and two related strains. Arch Mikrobiol *82*, 1-23.

198. Singer, T.P. (1947). On The Mechanism of Enzyme Inhibition by Sulfhdryl Reagents. J Biol Chem *147*, 11-21.

199. Kapranchikov, V.S., Zherebtsov, N.A., and Popova, T.N. (2004). Purification and Characterization of Lipase from Wheat (*Triticum aestivum* L.) Germ. Applied Biochemistry and Microbiology *40*, 84-88.

200. Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. Nucleic acids research *32*, D431-433.

Appendices

Appendix I – Practical Considerations for Applying the Equilibrium Model

Assay consistency. The enzyme assay being used must be consistent across the intended temperature range, or at least the effect of temperature on the assay must be assessable and can consequently be used to adjust collected data (e.g., temperature-related substrate degradation). Generally speaking, direct continuous spectrophotometric assays produce ideal data sets for using with the Equilibrium Model. The high precision and continuous sampling rates of a modern spectrophotometer result in large quantities of high quality assay data points, and the convenience of direct continuous spectrophotometric assays not only allows quick data collection, but also has lead to a large number of assays being available.

Discontinuous spectrophotometric assays (i.e., stopped assays), while being more labor-intensive and provide fewer data points, has been shown to be a viable and valid method for obtaining temperature profiles [20]. Enzyme-coupled assays can also be used, as long as the stability and activity of the coupled enzyme is such that the coupled reaction is consistent across the entire temperature range and that the reaction catalyzed by the coupled enzyme is not the rate-limiting step. Provided that the above criteria are satisfied, other types of assay can be used in conjunction with the Equilibrium Model as well.

Activity ceiling. Apart from saturating the enzyme with substrates, other checks should also be performed to ensure assay linearity. The collective result

of these checks is an 'activity ceiling'. A valid activity ceiling takes into consideration the following factors:

- the upper detection limit of the spectrophotometer
- the intended assay length
- substrate saturation

In the absence of enzyme thermal denaturation, if an assay is within the limit of the its activity ceiling, its recorded absorbance curve should be completely linear. The activity ceiling should be determined at the highest non-denaturing temperature and applied to all assays in the temperature profile. Furthermore, one needs to adjust the enzyme concentration in assays so that the highest *overall* reaction rate (versus the highest *initial* reaction rate) of all assays in the temperature profile is around 80% of the activity ceiling. This allows sensitivity to be maximized without introducing any bias due to loss of linearity caused by non-ideal assay conditions.

If the *overall* reaction rate of any single assay in the temperature profile exceeds the activity ceiling, the linearity of that particular assay is then compromised. As a result, the data set is prone to a bias that may lead to incorrect estimation of $\Delta G^{\ddagger}_{inact}$ during the fitting process. Because the bias can only be corrected by adding less enzyme, the entire data set must be discarded since the concentration of enzyme must remain constant for a temperature profile.

Assay anomalies. Assay anomalies can undermine the validity of temperature profiles, and proper measures should be taken to detect and, if possible, rectify those anomalies. Several common anomalies are discussed here: substrate inhibition, product inhibition, spontaneous reverse reaction, and

non-enzymic activity.

Substrate inhibition is present if reaction curves recorded at non-denaturing temperatures deviate from straight line immediately after the assays began. One must perform enzyme assays at lower than the inhibiting substrate concentration and determine the enzyme's K_m at all assay temperatures to verify whether V_{max} was maintained at all temperatures and make adjustments during data processing if necessary.

Signs of product inhibition are similar to those of substrate depletion or enzyme thermal denaturation, namely reaction curves deviating from straight line toward the end of assays. However, product inhibition differs from these two in that it cannot be remedied by adding more substrate, or delayed by adding more enzyme. One can possibly work around it by restricting the final product concentration (i.e., having a very low activity ceiling) at the cost of assay sensitivity, but other measures such as shifting the buffer pH, chemically removing the reaction product during assay, etc should be first employed if available. If everything fails, however, one can resort to using initial reaction rates and the zero-time version of the Equilibrium Model [20], which can also be used for extremely unstable enzymes.

Effects of a spontaneous reverse reaction also appear similar to those of product inhibition, except the reaction curves actually tend to be U-shaped if the assay length is sufficiently long. This anomaly is difficult to overcome, but some potential measures exist: using an extremely high concentration of substrate (assuming that the assay monitors an increase in absorbance and no substrate inhibition is in place) and modifying assay conditions. The zero-time

Equilibrium Model is also an effective workaround for this anomaly.

Some non-enzymic issues may also affect assay consistency. Certain combinations of reaction and assay components (e.g., Mn^{2+} and phosphate buffer) may result in complex formation in the assay mixture. To check for this problem, simply examine the clarity of the reaction mixture. Substrate and product stability must also be taken into consideration. Blank rates of spontaneous substrate degradation must be obtained at temperatures where substrate stability is an issue, although certain cofactors (i.e., NAD[P]H) can be stabilized by a high buffer pH. For assays monitoring the formation of catalytic products, measures must be employed to minimize spontaneous product loss through the choice of buffer, use of additives, or even stop assays.

Buffer pH control. Apart from choosing a suitable buffer for the enzyme and determining its pH optimum, it is important to consider the stability of buffer constituents. Simple buffers such as phosphate, acetate, or borate buffers are often preferred over complex buffers like HEPES, MOPS, or PIPES buffer for their stability.

One also needs to consider temperature's effect on buffer pH ($\Delta pK_a/\Delta t$). A buffer with a large $\Delta pK_a/\Delta t$ (e.g., -0.028 $pH/^{\circ}C$ for Tris buffer) can exhibit a significant pH shift over a moderate range of temperatures (e.g., 1.12 pH units over 40ºC for Tris buffer), whereas buffers with sufficiently low $\Delta pK_a/\Delta t$ (e.g., phosphate buffer) can be used without adjustments for the entire temperature profile. Therefore, provided that activity levels are close, buffers with low

$\Delta pK_a/\Delta t$ values are often preferred. If a buffer with high $\Delta pK_a/\Delta t$ value has to be used, it must be adjusted to the desired pH value for each temperature point in the temperature profile. Additionally, if a known natural activator exists for the enzyme, it should also be included in the reaction mixture to maximize enzyme activity.

Temperature-related K_m changes. It is not uncommon for the K_m of an enzyme to increase with temperature [12, 175]. The magnitude of temperature-related K_m escalation differs from enzyme to enzyme, and since this relationship results from a complex mix of weak interactions between the enzyme and the substrate [27], it has to be determined experimentally using the same condition intended for the temperature profile assays. Determining K_m at all the necessary temperatures potentially involves tens, even hundreds of assays using significant amounts of enzyme, so this procedure must be well planned if enzyme supply is limited.

K_m determination should first be performed at the highest non-denaturing temperature, the highest temperature possible (i.e., clearly determinable initial reaction rate), and the lowest temperature point intended to be included in the temperature profile. At temperatures where enzyme thermal denaturation is significant, a larger amount of enzyme may be used to obtain reliable initial rates.

If K_m is higher than one tenth of the intended substrate concentration ([S]) at a certain temperature, K_m determination should be performed at all assay temperatures above this temperature, and successively at temperature points below the said temperature until K_m value is below one tenth of [S]. One can

then use Equation 2-1 to calculate how much observed rates theoretically deviate from V_{max}, and the adjustments can be made during data processing.

Temperature-related issues. Stringent temperature control is crucial to a valid temperature profile [2, 19, 20]. A digital thermometer accurate to ± 0.1ºC should be used to record assay temperature before and after measurement, and there should not be a detectable temperature gradient within the reaction vessel. At high temperatures, the reaction should be capped to avoid evaporation, and below the ambient temperature, condensation on surfaces in the light path can be avoided by blowing compressed gas across the surfaces.

Enzyme stability and addition. An ideal enzyme concentration in assay should be calculated from the activity ceiling, since a low enzyme concentration results in both low activity level at lower temperatures and rapid disappearance of enzyme activity at high temperatures, jeopardizing the overall robustness of the temperature profile.

Adding a large volume of enzyme solution to temperature-equilibrated reaction mixtures can have adverse effects on the precision of temperature control, but pipetting very small volumes of viscous enzyme solution is also problematic, thus it is desirable to add a moderate volume of relatively dilute enzyme stock. However, at low enzyme concentrations, a significantly portion of protein can attach to the walls of plastic containers. Glass containers significantly mitigate this problem, and a very low concentration of detergent (e.g., 0.0001% SDS) can also be used to solubilize protein. Very dilute solutions of less stable enzymes may loss activity on ice during the course of an experiment, so it is important to perform regular checks of the enzyme stock by repeating assays at a reference

temperature throughout the experiment. If an enzyme stock exhibits activity loss with time when stored on ice, one should prepare it in advance, aliquot it into small portions (e.g., enough for ten assays), and freeze the aliquots at -80ºC; fresh aliquots can then be thawed when necessary during the experiment. If the latter method is used, the experimenter should also assess potential enzyme activity loss during storage.

Enzyme solution stored on ice should be rapidly added to the reaction mixture to initiate the assay in order to minimize loss of activity due to denaturation and to maximize chances of observing activity loss due to the E_{inact}–E_{act} transition. Also, immediately following the addition of enzyme, the reaction must be sufficiently mixed without causing air bubbles in the vessel that may interfere data collection.

Temperature points. The influence of the number of temperature points on the final Equilibrium Model parameters has been examined and published [20]; below are a summary of the findings and guideline for estimating suitable temperature points and range.

A temperature profile must contain at least eight temperature points with at least two temperature points above T_{opt} (showing a clear downward trend) to validate the use of the Equilibrium Model, although additional temperature points, particularly at the high end of temperature, tend to have positive effects on the accuracy of the fitting process and the final parameters. Note that T_{opt} can only be accurately estimated after a temperature profile has been generated and fitted to the Equilibrium Model, thus obtaining a usable temperature profile can be a logical conundrum—one cannot get a good temperature profile without first using

parameters obtained from fitting the temperature profile to the Equilibrium Model.

A usable temperature profile consists of at least 24 assays (triplicates at eight temperature points), and data for continuous spectrophotometric assays contain up to thousands of data points [20]. A preliminary assessment can be performed by first assaying the enzyme at a conjectured T_{opt}, and then at two flanking temperature points, using assay length intended for the temperature profile. The *initial* rates of these measurements are then compared to see if the T_{opt} estimation was reasonable, and more measurements are performed at higher/lower temperatures if necessary. Once a reasonable T_{opt} estimation has been obtained, measurements should be made at several temperature points above the estimated T_{opt}, with at least one of them exhibiting a significant reduction in its initial rate. For some enzymes, the reduction of activity may be rapid with increasing temperature above T_{opt}, and it may be necessary to reduce the temperature increment for temperature points above the estimated T_{opt}.

The *estimated* T_{opt} isn't always close to the final T_{opt} derived from the Equilibrium Model, which should be compared against the temperature profile used. If none of the measurements were actually performed at temperatures above the final T_{opt} value and showing a clear downward trend, then additional temperature points should be obtained (assuming that an identical enzyme concentration in assay can be achieved) and the entire temperature profile refitted.

Appendix II – Eurythermalism and the Temperature Dependence of Enzyme Activity

This article has been published in the June 2007 issue (volume 21, issue 8) of the FASEB Journal (page 1934-1941, PubMed ID: 17341686)

Eurythermalism and the temperature dependence of enzyme activity

Charles K. Lee,* Roy M. Daniel,*,[1] Charis Shepherd,[†] David Saul,[†] S. Craig Cary,*,[‡] Michael J. Danson,[§] Robert Eisenthal,[‖] and Michelle E. Peterson*

*Department of Biological Sciences, University of Waikato, Hamilton, New Zealand; [†]School of Biological Sciences, University of Auckland, Auckland, New Zealand; [‡]College of Marine and Earth Studies, University of Delaware, Lewis, Delaware, USA; [§]Centre for Extremophile Research, Department of Biology and Biochemistry, and [‖]Department of Biology and Biochemistry, University of Bath, Bath, UK

ABSTRACT **The "Equilibrium Model" has provided new tools for describing and investigating enzyme thermal adaptation. It has been shown that the effect of temperature on enzyme activity is not only governed by $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$ but also by two new intrinsic parameters, ΔH_{eq} and T_{eq}, which describe the enthalpy and midpoint, respectively, of a reversible equilibrium between active and inactive (but not denatured) forms of enzyme. Twenty-one enzymes from organisms with a wide range of growth temperatures were characterized using the Equilibrium Model. Statistical analysis indicates that T_{eq} is a better predictor of growth temperature than enzyme stability ($\Delta G^{\ddagger}_{inact}$). As expected from the Equilibrium Model, ΔH_{eq} correlates with catalytic temperature tolerance of enzymes and thus can be declared the first intrinsic and quantitative measure of enzyme eurythermalism. Other findings shed light on the evolution of psychrophilic and thermophilic enzymes. The findings suggest that the description of the Equilibrium Model of the effect of temperature on enzyme activity applies to all enzymes regardless of their temperature origins and that its associated parameters, ΔH_{eq} and T_{eq}, are intrinsic and necessary parameters for characterizing the thermal properties of enzymes and their temperature adaptation and evolution.—Lee, C. K., Daniel, R. M., Shepherd, C., Saul, D., Cary, S. C., Danson, M. J., Eisenthal, R., Peterson, M. E. Eurythermalism and the temperature dependence of enzyme activity *FASEB J.* 21, 1934–1941 (2007)**

Key Words: enzyme temperature optimum · Equilibrium Model · growth temperature · protein stability · temperature adaptation

One of the most important environmental factors for life is temperature. It affects diffusion, membrane fluidity, nucleic acid stability, salt and gas solubility, and, significantly, the behavior of enzymes. Until recently, two parameters have been used to characterize the effect of temperature on enzymes: firstly, the temperature-related reaction kinetics of the chemical reaction catalyzed by the enzyme, expressible as $\Delta G^{\ddagger}_{cat}$, the free energy of activation of the catalytic reaction, and secondly thermal stability, which can be described using $\Delta G^{\ddagger}_{inact}$, the free energy of activation of the thermal denaturation process.

Some functional studies of enzyme temperature-linked properties have defined a so-called enzyme "temperature optimum" that arises from this two-parameter "Classical Model" (1). However, this "temperature optimum" is not an intrinsic enzyme property, being derived from a complex mixture of both activity and thermal stability effects (1) and dependent on assay duration. Consequently, it is of limited value for measuring enzyme temperature adaptation. Evidence of anomalously low enzyme activity at high temperatures (2, 3) also suggests that the Classical Model is insufficient to fully describe the effect of temperature on enzyme activity. In addition, there has not been any quantitative tool for characterizing an enzyme's eurythermalism. Traditionally, eurythermalism has been used to describe an organism's ability to withstand a wide range of temperatures. To date, there is limited information on how eurythermalism is achieved in eurythermal organisms, and an important possibility is that special characteristics of enzymes in these organisms play an important role.

Enzyme thermal stability clearly has a general correlation with growth temperature; enzymes from thermophiles are typically more thermally stable than those from mesophiles, and those from mesophiles are generally more stable than their psychrophilic counterparts. However, there are exceptions, such as RNase (4, 5), mammalian adenylate kinase (6, 7), and catalase (8–12), which are much more stable than might be expected from their environmental temperatures. Factors other than a high temperature environment may in some cases be responsible for high thermal stability. Resistance to proteolysis (13) and organic solvents (14), for example, are known to be associated with high thermal stability. In addition, almost all thermal stabil-

[1] Correspondence: Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. E-mail: r.daniel@waikato.ac.nz

doi: 10.1096/fj.06-7265com

0892-6638/07/0021-1934 © FASEB

ity determinations are done in the absence of substrate, which often affects stability.

Two new, intrinsic, temperature-dependent parameters of enzymes, T_{eq} and ΔH_{eq}, have recently been described (1) and validated (15). These arise from the Equilibrium Model, which proposes that the active form of enzyme (E_{act}) is in reversible equilibrium with an inactive form (E_{inact}) and that it is the inactive form that undergoes irreversible thermal inactivation to the thermally denatured state (X), as described in (1).

$$E_{act} \underset{}{\overset{K_{eq}}{\rightleftharpoons}} E_{inact} \xrightarrow{k_{inact}} X \qquad [1]$$

The behavior of the model is described by k_{cat}, the rate constant of the enzyme-catalyzed reaction, k_{inact}, the rate constant for the irreversible inactivation reaction, and K_{eq}, the equilibrium constant governing the ratio E_{inact}/E_{act} in the reversible interconversion. The dependence of catalytic rate on temperature is obtained from these constants expressed in terms of the thermodynamic parameters, $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq} as follows:

$$k_{cat} = \frac{k_B T}{h} \cdot e^{-\left(\frac{\Delta G^{\ddagger}_{cat}}{RT}\right)} \qquad [2]$$

$$k_{inact} = \frac{k_B T}{h} \cdot e^{-\left(\frac{\Delta G^{\ddagger}_{inact}}{RT}\right)} \qquad [3]$$

$$K_{eq} = e^{\frac{\Delta H_{eq}}{R} \cdot \left(\frac{1}{T_{eq}} - \frac{1}{T}\right)} \qquad [4]$$

where R is the gas constant, k_B is Boltzmann's constant, and h is Planck's constant.

T_{eq} is the temperature representing the midpoint of transition between the active and inactive (but undenatured) forms of the enzyme. In other words, T_{eq} is the temperature at which $K_{eq} = 1$ and $\Delta G^{\ddagger}_{eq} = 0$ [we have previously (1) used the term T_m to designate this temperature but now prefer the term T_{eq} (15) since it is now clear that T_{eq} is entirely separated from and unrelated to thermal denaturation]. Of more than 40 enzymes examined in our laboratories (15–17), including those presented here, all followed the temperature-dependent behavior predicted by the Equilibrium Model, and none followed that expected from the Classical Model. Furthermore, T_{eq} is independent of $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$. Preliminary evidence suggests that the active-inactive transition described by the Equilibrium Model arises from a conformational transition (15) at or near the active site, since the active-inactive equilibrium is established in a timescale at least two orders of magnitude faster than that of thermal denaturation (15); stabilizing and destabilizing agents have no effect on T_{eq} (16, 18); and different substrates can affect T_{eq} without altering stability (16, 18).

Not only does the Equilibrium Model provide a description of enzyme temperature-linked behavior, it does so under assay conditions, so that the derived constants such as T_{eq} and ΔH_{eq} are potentially physiologically based. Since enzyme temperature adaptation is the functional basis of metabolic and organism adaptation studies, the Equilibrium Model is not only useful but necessary for conducting such studies.

In this work, the intrinsic temperature-dependent properties of 21 enzymes, including several isoactive sets of enzymes derived from a wide range of growth temperature sources, were determined. Parameters generated from model fitting were analyzed to identify any associations with each other or with growth temperature optima.

MATERIALS AND METHODS

Enzymes and reagents

For detailed descriptions of enzyme sources, see Supplemental Table 4. All chemicals used in this study were purchased from Sigma-Aldrich Inc. and Wako Chemical Company and are of analytical grade or higher (see Supplemental Table 3).

Enzyme assays and data collection

In general terms, these were carried out as described previously (17). All enzyme assays performed in this work were continuous photometric assays measured using a Thermospectronic Helios γ-spectrophotometer equipped with a Thermospectronic single cell peltier-effect cuvette holder. All reactions were started by the rapid addition and mixing of a few microliters of enzyme solution at zero degrees into 400-3000 μl of temperature-equilibrated reaction mixture to enable data collection to begin within the first few seconds; the temperature of the reaction mixture remains within ±1°C of desired temperature during data collection. Absorbance data were collected at one-second intervals for three minutes using Vision (version 1.25, Thermo Spectronic) on a Windows PC connected to the Helios, within a range of temperatures depending on the source of the enzyme (see Material and Methods in Supplemental Data for assay details). In all cases, precautions were taken so that no decrease in reaction rates was due to substrate depletion (see Material and Methods in Supplemental Data). Where possible, substrate concentrations were maintained at over ten times K_m at all temperatures, otherwise appropriate adjustments were made to compensate for the effect of increases in K_m. No signs of substrate/product inhibition were observed under the experimental conditions described. Blank rates were measured at all temperatures and used to correct reaction rates if they were significant (see Table 3 in Supplemental Data for assay conditions for individual enzymes).

Based on the variation between the individual triplicate rates from which the parameters are derived for all the enzymes we have assayed thus far, we find that the experimental errors in the determination of $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, and T_{eq}, are <0.5% and <6% in the determination of ΔH_{eq} (17).

Data analysis

The method for processing experimental data is as described previously in detail (17). Absorbance data from

Vision were first converted in Excel (version 11 for Windows, Microsoft) to progress curves of product concentration (M). The data were subsequently imported into Scientist (version 2.01, MicroMath), where a set of common estimates of thermodynamic parameters (see Supplemental Data) was optimized by "Simplex" searching through the complete data set (product concentration *vs.* time *vs.* temperature). The *optimized* estimates of parameters were then used as initial values to "Least Square" fit the complete data set to the Equilibrium Model, which generated the final values of the parameters; $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, and T_{eq}. The generated parameters were then used in the "Zero-time Model" (17) to draw a zero-time activity *vs.* temperature plot. This plot was used to visually estimate T_{opt} and HWHM. Initial rates estimated from assay data were also plotted on the zero-time plot to verify the validity of the fittings.

A standalone MATLAB (Version 7.1.0.246 (R14) Service Pack 3, The MathWorks, Inc.) application, enabling the facile derivation of the Equilibrium Model parameters from a Microsoft Office Excel file of experimental progress curves (product concentration *vs.* time), is available on CD from the corresponding author. This application is suitable for computers running Microsoft Windows XP and is for noncommercial research purposes only.

Protein determination

Protein determinations were performed using the Bradford assay (19).

Statistical analysis

All parameters were imported into STATISTICA (20) for further analysis. Correlation analysis was done for nine parameters, *i.e.,* growth temperature, $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, T_{eq}, T_{opt}, HWHM, the difference between T_{eq} and growth temperature, and the difference between T_{eq} and T_{opt}. The nonparametric Goodman-Kruskal Gamma test was used, and Gamma statistics and the associated two-tailed *P* values were obtained. Gamma statistic is equivalent to Kendall *Tau* in terms of the underlying assumptions and interpretation, but takes tied ranks (*e.g.*, growth temperature) into account. Two-tailed statistical power values were calculated with these conditions: $n = 21$, *Rho*0 = 0, Alpha = 0.05, and "Exact" algorithm. Best-subset regression analysis was done using growth temperature as the response (dependent) factor and $\Delta G^{\ddagger}_{cat}$, $\Delta G^{\ddagger}_{inact}$, ΔH_{eq}, T_{eq}, and T_{opt} as predictors. A maximum step of 100 and a probability of 0.05 were used for stepwise selection, and best subsets measures were calculated using R squared values. Nonparametric *t* test was carried out using Mann-Whitney *U* test, and a two-tailed *P* value was calculated. Scatter plots were drawn using MiniTab (R14 for Windows, Minitab).

RESULTS AND DISCUSSION

All enzymes, including those from extremophiles, obeyed the Equilibrium Model and were found to display zero-time activity optima with temperature, characteristic of the model (*e.g.*, **Fig. 1** and **Fig. 2**). It should be noted that given the "fast-start" of the enzyme assays (see Materials and Methods), the decline in activity at zero time at temperatures above T_{opt} is evidence that equilibration of the active-inactive forms of the enzyme (E_{act} and E_{inact}) takes place over timescales significantly shorter than we can observe, *i.e.*, $<$1–3 s and that the equilibration is about two orders of magnitude or more faster than the rate of irreversible activity loss as shown by the activity *vs.* time plots for each temperature in the 3D graphs in Fig. 2 (although it should be noted that the very slow irreversible denaturation for the ASB HK47 psychrophilic alkaline phosphatase above T_{opt} is not typical of the enzymes examined here). The parameters of all enzymes are listed in **Table 1**. They are derived by fitting assay data to the Equilibrium Model and thus relate to active enzymes in the presence of substrate and cofactor. The exception is the growth temperature optimum of the source organism, which is cited from various sources of reference (see Supplemental Data). Growth temperatures of the source organisms range from 2°C to 75°C. The difference between the source growth temperature and T_{eq} ranges from -6.1 to 52.6°C.

T_{opt} is the graphical optimum temperature of the enzyme at time zero (see Fig. 1). In most cases, T_{eq} is within a few degrees of T_{opt}, and for almost all enzymes T_{eq}, (and T_{opt}) are higher than the respective source growth temperature. In the case of T_{eq}, the mean difference is 21.1 ± 15.2°C. This is not surprising, since

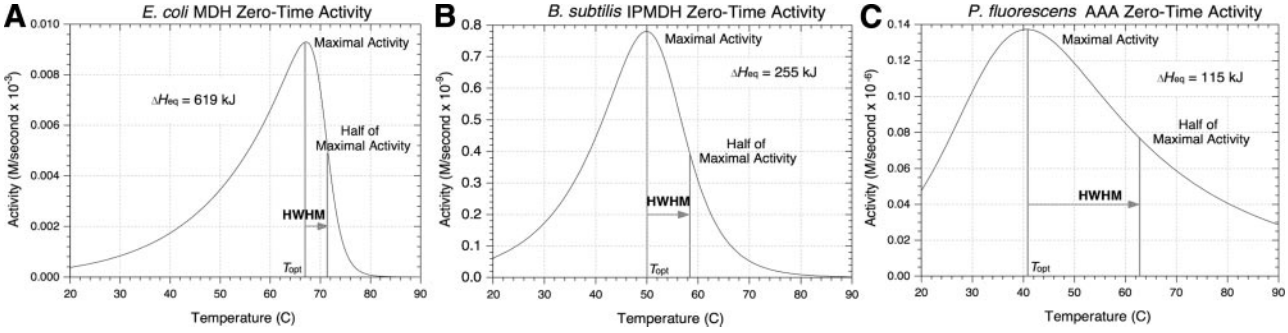


Figure 1. Zero-time activity plot of enzymes with different eurythermal properties [*Escherichia. coli* MDH (*A*); *Bacillus subtilis* IPMDH (*B*); *Moritella profunda* DHFR (*C*)]. ΔH_{eq} is the enthalpic change for the transition between active and inactive forms of the enzyme; T_{opt} is the temperature at which enzyme activity at zero-time is highest; Half Width at Half Maximum (HWHM) is the width of the zero-time temperature/activity plot between T_{opt} and the upper temperature at which 50% of maximum activity is exhibited. For enzyme name abbreviations, see Table 1 legend.

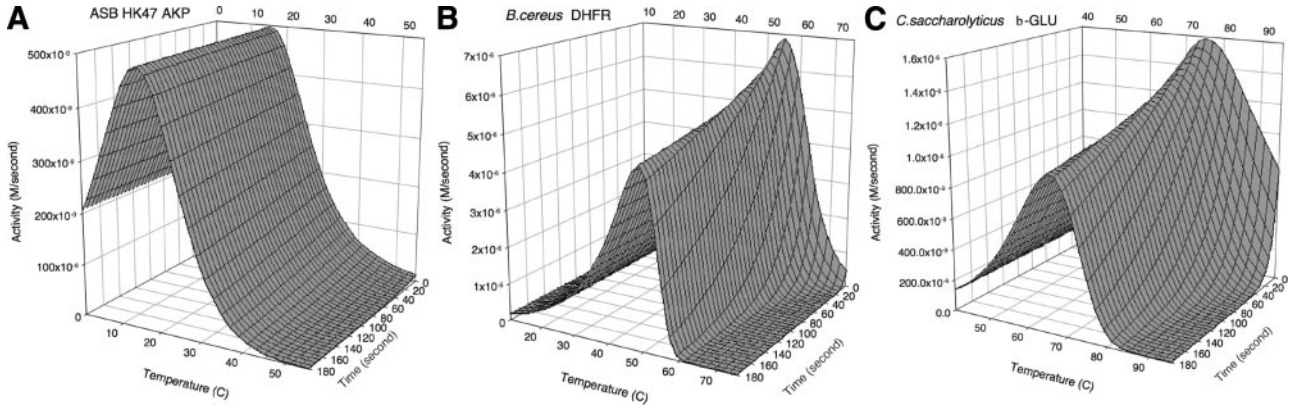


Figure 2. 3D plots of enzymes from different thermal origins [Antarctic sea bacterium HK47 – psychrotrophic (*A*); *Bacillus cereus* – mesophilic (*B*); *Caldicelulosiruptor sacchrolyticus* – thermophilic (*C*)]. The plots are generated from parameters derived directly from the raw data.

at T_{eq} half of the enzyme activity is unavailable, while at 20°C below T_{eq} most of the enzyme will be in the active form. HWHM is the upper temperature half width of the zero-time temperature/activity peak at 50% activity (the upper temperature half width has been used because the low temperature half of this peak is dominated by $\Delta G^{\ddagger}_{cat}$) and is included to show its relationship to ΔH_{eq} (see Fig. 1); a large ΔH_{eq} leads to a sharp decline in activity with increasing temperature above T_{opt} (and thus a small HWHM), since ΔH_{eq} is the enthalpic change associated with the reversible conversion of the active to the inactive form of enzyme. There are wide variations in both HWHM (from 4.5 to 44°C) and ΔH_{eq} (from 86 to 826 kJ·mol^{-1}), and thus considerable variations in the sensitivity of enzyme activity to temperature above T_{eq} (and T_{opt}). $\Delta G^{\ddagger}_{inact}$ describes the stability of the enzyme under reaction conditions; as expected, the average $\Delta G^{\ddagger}_{inact}$ of thermophilic enzymes tends to be higher, and the average $\Delta G^{\ddagger}_{inact}$ of psychrophilic/psychrotrophic enzymes lower, than that of mesophilic enzymes.

3D plots of one thermophilic, one mesophilic, and one psychrophilic enzyme are shown in Fig. 2*A–C*. Although the plots are very similar in form, the temperature axes show that the three enzymes have very distinct responses to temperature. Note that for the ASB HK47 alkaline phosphatase (Fig. 2*A*), $\Delta G^{\ddagger}_{inact}$ has relatively little effect on activity over time at tempera-

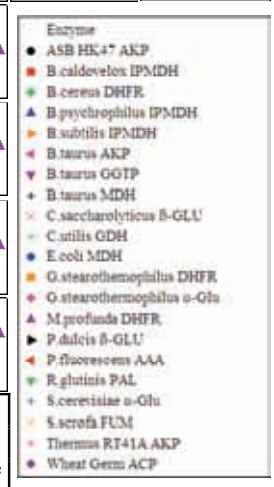
TABLE 1.

Organism	Enzyme	Growth Temp. (°C)	$\Delta G^{\ddagger}_{cat}$ (kJ·mol^{-1})	$\Delta G^{\ddagger}_{inact}$ (kJ·mol^{-1})	ΔH_{eq} (kJ·mol^{-1})	T_{eq} (°C)	T_{opt} (°C)	HWHM (degree)	T_{eq} minus Growth Temperature	T_{eq} minus T_{opt}
C.utilis	GDH	31	57	94	409	59.5	55.5	6	28.5	4
B.taurus	MDH	39	53	85	826	67.4	64.5	4.5	28.4	2.9
E.coli	MDH	40	55	96	619	70.7	67	4.5	30.7	3.7
M.profunda	DHFR	2	67	93	104	54.6	60.5	32	52.6	-5.9
B.cereus	DHFR	30	66	90	261	58.5	55	8.5	28.5	3.5
G.stearothermophilus	DHFR	55	67	97	96	53.9	63	40.5	-1.1	-9.1
ASB HK47	AKP	15	55	90	149	17.5	15	12	2.5	2.5
B.taurus	AKP	39	57	97	86	59.6	69	44	20.6	-9.4
Thermus sp. RT41a	AKP	75	72	99	305	90	86	9	15	4
P.fluorescens	AAA	25	74	92	115	35.8	41	24	10.8	-5.2
Wheat germ	ACP	20	79	95	142	63.9	66	20	43.9	-2.1
S.cerevisiae	α-GLU	28	71	90	272	39.4	36.5	7.5	11.4	2.9
G.stearothermophilus	α-GLU	55	64	96	225	67.8	64	10.5	12.8	3.8
P.dulcis	ß-GLU	20	63	95	99.8	55.9	61.5	33.5	35.9	-5.6
C.saccharolyticus	ß-GLU	70	81	98	154	73.9	75	18.5	3.9	-1.1
S.scrofa	FUM	39	60	92	378	59.1	55	6.5	20.1	4.1
B.taurus	GGTP	39	63	98	109	52	55.5	26	13	-3.5
R.glutinis	PAL	24	80	97	181	56.5	56	13	32.5	0.5
B.psychrophilus	IPMDH	20	72	100	123	56.7	60	24	36.7	-3.3
B.subtilis	IPMDH	30	73	94	255	52.8	50	8.5	22.8	2.8
B.caldovelox	IPMDH	70	76	101	573	63.9	61	4.5	-6.1	2.9

$\Delta G^{\ddagger}_{cat}$ is the free energy of activation of the catalytic reaction, and $\Delta G^{\ddagger}_{inact}$ is that of the thermal denaturation process. ΔH_{eq} is the enthalpic difference between active (E_{act}) and inactive forms (E_{inact}) of the enzyme. T_{eq} is the temperature at which the concentration of E_{act} equals that of E_{inact}. T_{opt} and HWHM are both graphically determined from zero-time temperature versus activity plots (see Fig. 1). See Supplemental Data for complete names of organisms. GDH, glutamate dehydrogenase; MDH, malate dehydrogenase; DHFR, dihydrofolate reductase; ASB, Antarctic sea bacterium, AKP, alkaline phosphatase; AAA, aryl acylamidase; ACP, acid phosphatase; GLU, glucosidase; FUM, fumarate hydratase; GGTP, γ-glutamyl transferase; PAL, phenylalanine ammonia-lyase; IPMDH, isopropylmalate dehydrogenase.

TABLE 2.

Parameters	Growth Temperature	ΔH_{eq}	$\Delta G^{\ddagger}_{cat}$	$\Delta G^{\ddagger}_{inact}$	T_{eq}	T_{opt}	T_{eq} minus Growth Temperature	Half Width at Half Maximum	T_{eq} minus T_{opt}
Gamma	0.2929								
P Value	0.0712								
Power	0.2549								
Gamma	-0.0202	-0.1048							
P Value	0.9010	0.5065							
Power	0.1417	0.0733							
Gamma	0.3535	-0.1333	0.3048						
P Value	0.0295	0.3978	0.0533						
Power	0.3582	0.0883	0.2733						
Gamma	**0.5253**	0.3143	-0.0286	0.3048					
P Value	**0.0012**	0.0463	0.8562	0.0533					
Power	**0.7186**	0.2890	0.0517	0.2733					
Gamma	0.3878	-0.0288	0.0288	0.4327	**0.6442**				
P Value	0.0176	0.8555	0.8555	0.0063	**<0.0001**				
Power	0.4252	0.0517	0.0517	0.5196	**0.9146**				
Gamma	**-0.4646**	-0.0286	-0.0667	-0.0571	0.0667	0.1346			
P Value	**0.0042**	0.8562	0.6725	0.7171	0.6725	0.3956			
Power	**0.5890**	0.0517	0.0593	0.0568	0.0593	0.0890			
Gamma	-0.2680	**-0.9512**	0.1122	0.1707	-0.2585	0.0936	0.0537		
P Value	0.1017	**<0.0001**	0.4821	0.2848	0.1053	0.5596	0.7367		
Power	0.2188	**1.0000**	0.0768	0.1141	0.2061	0.0685	0.0560		
Gamma	0.3232	**0.7333**	-0.1048	-0.0952	0.2952	-0.0673	-0.0667	**-0.7366**	
P Value	0.0466	**<0.0001**	0.5065	0.5459	0.0612	0.6710	0.6725	**<0.0001**	
Power	0.3039	**0.9818**	0.0733	0.0692	0.2584	0.0595	0.0593	**0.9831**	



Strong correlations are in color, while weaker ones are in normal typeface. Correlation analysis was done using the nonparametric Gamma test (where Gamma is equivalent to correlation coefficient on scale of −1 to +1). *P* value is the measure of correlation by chance, while Power value is the probability of rejecting a false null hypothesis. A low Power value (*i.e.,* lower than 0.8) indicates that a correlation is subject to type II errors (false negative); in other words, the sample size is too small to successfully detect a weak correlation between the 2 parameters. For scatter plots, vertical axes correspond to the parameter along the horizontal plane of the table, and horizontal axes correspond to the parameter below. Least squares regression curves in scatter plots were calculated using a linear model with fit intercept. The regression curves are not a mathematically correct representation of nonparametric correlation and are only used to indicate general trends in the data. See main text for a detailed description of statistical analyses used.

tures somewhat above T_{opt}, and physiological activity is therefore very dependent on T_{eq}, but for the *Bacillus cereus* DHFR (Fig. 2*B*), activity over time above T_{opt} depends largely on $\Delta G^{\ddagger}_{inact}$.

Correlation analysis

A correlation analysis of the parameters in Table 1 is shown in **Table 2**. Two sets of significant correlations are evident. The strongest correlation set (in red) involves ΔH_{eq}, HWHM (negatively), and the difference between T_{eq} and T_{opt}. ΔH_{eq} is the enthalpic change associated with the reversible, temperature driven, interconversion of an enzyme between its active and inactive state. It can be considered as a measure of the sensitivity of an enzyme's catalytic activity to temperature. ΔH_{eq} will therefore influence the broadness of zero-time activity *vs.* temperature plots; a small ΔH_{eq} will lead to a broader zero-time activity *vs.* temperature peak, thus giving rise to a strong negative correlation between ΔH_{eq} and HWHM. ΔH_{eq} can thus be considered to be a quantitative measure of an enzyme's ability to function over a temperature range: a small ΔH_{eq} indicating that the enzyme will function at relatively high activity over a wide range of temperatures, *i.e.*, behave in an eurythermal manner. Conversely, a large ΔH_{eq} indicates stenothermal behavior. The positive correlation between ΔH_{eq} and the difference between T_{eq} and T_{opt} can be explained on the basis that a low ΔH_{eq} will confer a broader time-zero graphical peak, and thus T_{opt} is likely to be higher than T_{eq}.

Compared to the correlation set above, other correlations involving ΔH_{eq}, HWHM, or the difference between T_{eq} and T_{opt} are much weaker. It may be that ΔH_{eq}, and hence an enzyme's ability to function over a wide range of temperatures (*e.g.,* in an environment with large temperature fluctuations), is more likely to be determined by the temperature range or temperature variability of the environment(s) than it is by other thermal properties, although this hypothesis has not been tested here.

The second strong correlation set (in green) is that between T_{eq} and T_{opt}, and T_{eq} and growth temperature; and moderate or weak correlations between all three of

growth temperature, T_{opt}, and $\Delta G^{\ddagger}_{inact}$, and between T_{eq} and $\Delta G^{\ddagger}_{inact}$. Since T_{opt} can theoretically be calculated from T_{eq} (15), the correlation between these two is not surprising, and the weak correlation of these two with protein stability ($\Delta G^{\ddagger}_{inact}$) simply confirms the well-established general tendency of thermophiles to have more stable proteins than mesophiles. Of the other correlations in this set, the correlation between T_{eq} and optimal growth temperature is particularly significant. It indicates that in a broad sense, T_{eq} is a better indication of an enzyme's temperature origin than either its stability ($\Delta G^{\ddagger}_{inact}$) or the graphically determined T_{opt}. This is supported by the result from best-subset regression analysis, which identifies T_{eq} as the single best parameter for predicting growth temperature (data not shown).

The moderate negative correlation (in blue) between growth temperature and the difference between T_{eq} and growth temperature is interesting. From Table 1, it is evident that at high growth temperatures (growth temperature $\geq$55°C), *i.e.,* in organisms that are clearly thermophilic, there tend to be relatively small differences between T_{eq} and growth temperature (7.3±6.6°C); however, for psychrophiles, psychrotrophs, and low-temperature mesophiles (growth temperature $\leq$31°C), they are larger (27.8±15.1°). A nonparametric *t* test between the T_{eq} and growth temperature differences of thermophilic and nonthermophilic enzymes reveals that the two groups are significantly different (*P* value=0.008) in this respect. A previous study using k_{cat} as a measure of temperature optima also suggested that enzymes tend to have optimal activity at temperatures higher than their hosts' living temperature, and the higher the growth temperature of the organism the narrower the gap becomes (21). There are several possible reasons for this. There may be a poor match between the environmental temperature of thermophiles and their laboratory growth temperature optima; thermophiles are often isolated from sources that are appreciably hotter than their determined growth temperatures. Alternatively, it might be argued that the origin of the current mesophilic microbial life is thermophilic, *i.e.,* that microbial evolution has proceeded down-temperature from a (hyper)thermophilic last common ancestor (22–24), and that the evolution of T_{eq} lags behind optimal growth temperature. Furthermore, depending on how crucial their roles may be in the cell, different enzymes might evolve at different speeds to adapt their T_{eq} to the environment, thereby resulting in the variation in T_{eq} among enzymes from one organism (*e.g.*, for *Bos taurus*, Table 1).

Conclusions based on weak correlations must be speculative. Although its *P* value is slightly over the adopted threshold of 0.05, the correlation between $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$ may suggest that $\Delta G^{\ddagger}_{cat}$ is higher for more stable proteins and thus support various studies (21, 25, 26) that suggested that a change in activation energy ($\Delta G^{\ddagger}_{cat}$) might be important to the temperature-adaptation of psychrophilic enzymes. To investigate this idea, correlation analysis was performed with psychrophilic and psychrotrophic enzymes (growth temperature$\leq$20°C, not including plant enzymes) excluded; the correlation between $\Delta G^{\ddagger}_{inact}$ and $\Delta G^{\ddagger}_{cat}$ was found to deteriorate significantly (*P* value>0.1), tending to confirm that a lower $\Delta G^{\ddagger}_{cat}$ might indeed be one of the main adaptations for psychrophilic enzymes, while nonpsychrophilic enzymes take on other means of adaptation.

The scatter plots in Table 2 are consistent with the correlations described.

Growth temperature

In general, any correlations with growth temperature must be regarded as tentative since determining environmental growth temperature from laboratory-determined optimal growth temperatures of bacteria is problematic, partly because the temperature of most microbial environments varies with time, including that of *Escherichia coli*, for example (27, 28), and partly because the laboratory environment on which these are usually based may not represent environmental conditions. In the absence of better measures, however, we have generally adopted here the published data for organisms involved in this study (see Supplemental Data).

Individual enzymes

The DHFR from the strict psychrophile *Moritella profunda*, isolated from deep Atlantic sediments (29, 30), is a surprisingly stable enzyme with a concomitantly high T_{eq}, while its ΔH_{eq} (104 kJ·mol^{-1}) is significantly lower than the 25th percentile (115 kJ·mol^{-1}), suggesting eurythermal behavior. We speculate that its low temperature activity has been achieved by evolving a lower ΔH_{eq} into an enzyme with an incidentally high T_{eq}, thus conferring eurythermalism. Part of the lower half (below T_{opt}) of the enzyme activity curve is influenced by ΔH_{eq}, but the effect of $\Delta G^{\ddagger}_{cat}$ may dominate, as stated above.

The alkaline phosphatase from isolate ASB HK47 (Fig. 2*A*) gives another illustration of the relative effects of the four parameters on enzyme activity over a range of temperatures. For many enzymes (*e.g.*, *B. cereus* DHFR, Fig. 2*B*), the denaturation rate is such that it will play the major role in controlling enzyme activity over timescales important to the organism (over minutes rather than seconds) at temperatures above T_{eq}. However, for the ASB HK47 alkaline phosphatase, it is evident that the lower rate of denaturation above T_{eq} leads to a situation where activity will be dominated by T_{eq} and ΔH_{eq} over significant timescales.

To achieve temperature adaptation, evolution imposes a strong selection force for thermophilic enzymes to have both high thermal stability and a high T_{eq}. High thermal stability has been thought to be achieved at the expense of flexibility to explain why thermophilic enzymes generally exhibit relatively little activity at lower temperatures (31). However, the rigidity required for

stability is global, while the flexibility required for catalysis is likely to be local (32), as demonstrated by some exceptions (33, 34). It has been suggested that a combination of local flexibility at the active site and high overall stability (35, 36), is behind such phenomena. In other words, enzymes active at low temperatures do not necessarily have to be thermally unstable (37, 38). Recent reports also suggested that the higher flexibility of psychrophilic enzymes is local and on the microsecond to millisecond timescale (39), while their global flexibility may not be necessarily higher (38). In addition, random and directed mutagenesis studies (37, 40, 41) have shown that it is possible to engineer an enzyme so that it retains low-temperature activity while gaining thermostability.

CONCLUSION

The correlations observed here do not necessarily denote a causative relationship. The finding that T_{eq} correlates best with growth temperature does not necessarily mean that it is the parameter most directly selectable by evolution, although it does provide insight into the effect of temperature on the evolution of enzymes. It should also be noted that the strong correlation between T_{eq} and growth temperature means that the correlation is likely to be reproducible with any selection of enzymes of similar number and diversity to this study.

ΔH_{eq} is a new quantitative measure to characterize eurythermalism at an enzymatic level. The new parameters (*i.e.,* T_{eq} and ΔH_{eq}), combined with established parameters (*i.e.,* $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$), define fully how temperature influences enzyme activity, while the former provide new tools to describe and investigate this influence.

Much of the data presented here comes from bacterial sources, although results from correlation analysis remain similar when data from eukaryotic sources are excluded (data not shown). Although the 21 sets of data from discrete enzymes presented here are a relatively small data set, the main correlations are robust and generally consistent with findings or proposals based on other data; they also provide new targets and perspectives for molecular and structural studies.

The development and verification of the Equilibrium Model have provided additional and quantitative measures of the thermal behavior of enzymes (1, 15) that are essential for describing the effect of temperature on enzyme activity and useful parameters for measuring the temperature adaptation of enzymes. Additionally, the results here show the Equilibrium Model's potential usefulness as a tool to investigate the relationship between enzyme thermal properties and the influence of temperature on the physiology and evolution of the host organism. FJ

This work was partly supported by a grant from the National Science Foundation (Biocomplexity 0120648) and a grant from the New Zealand Marsden Fund. We thank Y. Xu at Free University of Brussels for *M. profunda* DHFR, C. Monk for providing purified *Caldicellulosiruptor saccharolyticus* β-GLU, D. Clement for providing purified *Geobacillus stearothermophilus* DHFR, and J. Klinman at UC Berkeley for the gift of the *G. stearothermophilus* DHFR clone. We also thank Dr. Ray Littler at the University of Waikato for some assistance with the statistical analysis. We are grateful to O. Planckaert and A.-C. Tsuei for help with data collection.

REFERENCES

1. Daniel, R. M., Danson, M. J., and Eisenthal, R. (2001) The temperature optima of enzymes: a new perspective on an old phenomenon. *Trends Biochem. Sci.* **26,** 223–225
2. Thomas, T. M., and Scopes, R. K. (1998) The effects of temperature on the kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. *Biochem. J.* **330**, 1087–1095
3. Gerike, U., Danson, M. J., Russell, N. J., and Hough, D. W. (1997) Sequencing and expression of the gene encoding a cold-active citrate synthase from an Antarctic bacterium, strain DS2–3R. *Eur. J. Biochem.* **248,** 49–57
4. Notomista, E., Catanzano, F., Graziano, G., Dal Piaz, F., Barone, G., D'Alessio, G., and Di Donato, A. (2000) Onconase: an unusually stable protein. *Biochemistry* **39,** 8711–8718
5. Padmanabhan, S., Zhou, K., Chu, C. Y., Lim, R. W., and Lim, L. W. (2001) Overexpression, biophysical characterization, and crystallization of ribonuclease I from *Escherichia coli*, a broad-specificity enzyme in the RNase T2 family. *Arch. Biochem. Biophys.* **390,** 42–50
6. Noda, L., and Kuby, S. A. (1957) Adenosine triphosphate-adenosine monophosphate transphosphorylase (myokinase). I. Isolation of the crystalline enzyme from rabbit skeletal muscle. *J. Biol. Chem.* **226,** 541–549
7. Tsuboi, K. K., and Chervenka, C. H. (1975) Adenylate kinase of human erythrocyte. Isolation and properties of the predominant inherited form. *J. Biol. Chem.* **250,** 132–140
8. Diaz, A., Rangel, P., Montes de Oca, Y., Lledias, F., and Hansberg, W. (2001) Molecular and kinetic study of catalase-1, a durable large catalase of *Neurospora crassa*. *Free. Radic. Biol. Med.* **31,** 1323–1333
9. Calera, J. A., Sanchez-Weatherby, J., Lopez-Medrano, R., and Leal, F. (2000) Distinctive properties of the catalase B of *Aspergillus nidulans*. *FEBS Lett.* **475,** 117–120
10. Wasserman, B. P., and Hultin, H. O. (1981) Effect of deglycosylation on the stability of *Aspergillus niger* catalase. *Arch. Biochem. Biophys.* **212,** 385–392
11. Lopez-Medrano, R., Ovejero, M. C., Calera, J. A., Puente, P., and Leal, F. (1995) An immunodominant 90-kilodalton *Aspergillus fumigatus* antigen is the subunit of a catalase. *Infect. Immun.* **63,** 4774–4780
12. Switala, J., O'Neil, J. O., and Loewen, P. C. (1999) Catalase HPII from *Escherichia coli* exhibits enhanced resistance to denaturation. *Biochemistry* **38,** 3895–3901
13. Daniel, R. M., Cowan, D. A., Morgan, H. W., and Curran, M. P. (1982) A correlation between protein thermostability and resistance to proteolysis. *Biochem. J.* **207,** 641–644
14. Owusu, R. K., and Cowan, D. A. (1989) Correlation between microbial protein thermostability and resistance to denaturation in aqueous: organic solvent two-phase systems. *Enzyme Microbial Techn.* **11,** 568
15. Peterson, M. E., Eisenthal, R., Danson, M. J., Spence, A., and Daniel, R. M. (2004) A new intrinsic thermal parameter for enzymes reveals true temperature optima. *J. Biol. Chem.* **279,** 20717–20722
16. Eisenthal, R., Peterson, M. E., Daniel, R. M., Danson, M. J. (2006) The thermal behaviour of enzymes: implications for biotechnology. *Trends Bio/Tech.* **24,** 289–292
17. Peterson, M. E., Daniel, R. M., Danson, M. J., and Eisenthal, R. (2007) The dependence of enzyme activity on temperature: determination and validation of parameters. *Biochem. J.* **402,** 331–337.

18. Peterson, M. E. (2005) *Evidence for a third thermal parameter of enzymes*. PhD Thesis: University of Waikato, Hamilton, New Zealand
19. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72,** 248–254
20. StatSoft Inc (2005) STATISTICA, version 7.1. p. www.statsoft.com, Tulsa, OK 74104 U. S. A.
21. Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J. C., Sonan, G., Feller, G., and Gerday, C. (2004) Some like it cold: biocatalysis at low temperatures. *FEMS Microbiol. Rev.* **28,** 25–42
22. Schwartzman, D. W., and Lineweaver, C. H. (2004) The hyperthermophilic origin of life revisited. *Biochem. Soc. Trans.* **32,** 168–171
23. Pace, N. R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740
24. Wachtershauser, G. (1998) *Thermophiles: the Keys to Molecular Evolution and the Origin of Life?* (Wiegel, K., Adams, M., ed) pp. 47–57, Taylor and Francis, Philadelphia
25. D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M. A., Feller, G., and Gerday, C. (2002) Molecular basis of cold adaptation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **357,** 917–925
26. Johns, G. C., and Somero, G. N. (2004) Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydrogenases of Pacific damselfishes (*Chromis* spp.). *Mol. Biol. Evol.* **21,** 314–320
27. Flint, K. P. (1987) The long-term survival of *Escherichia coli* in river water. *J. Appl. Bacteriol.* **63,** 261–270
28. Hartl, D. L., and Dykhuizen, D. E. (1984) *The. population. genetics. of. Escherichia. coli. Annu. Rev. Genet.* **18,** 31–68
29. Xu, Y., Nogi, Y., Kato, C., Liang, Z., Ruger, H. J., De Kegel, D., and Glansdorff, N. (2003) Moritella profunda sp. nov., and *Moritella abyssi* sp. nov., two psychropiezophilic organisms isolated from deep Atlantic sediments. *Int. J. Syst. Evol. Microbiol.* **53,** 533–538
30. Xu, Y., Feller, G., Gerday, C., and Glansdorff, N. *Moritella* cold-active dihydrofolate reductase: are there natural limits to optimization of catalytic efficiency at low temperature? (2003) *J. Bacteriol.* **185,** 5519–5526
31. Ichikawa, J. K., and Clarke, S. (1998) A highly active protein repair enzyme from an extreme thermophile: the L-isoaspartyl methyltransferase from *Thermotoga maritima*. *Arch. Biochem. Biophys.* **358,** 222–231
32. Daniel, R. M., Dunn, R. V., Finney, J. L., and Smith, J. C. (2003) The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.* **32,** 69–92
33. Merz, A., Knochel, T., Jansonius, J. N., and Kirschner, K. (1999) The hyperthermostable indoleglycerol phosphate synthase from *Thermotoga maritima* is destabilized by mutational disruption of two solvent-exposed salt bridges. *J. Mol. Biol.* **288,** 753–763
34. Pappenberger, G., Schurig, H., and Jaenicke, R. (1997) Disruption of an ionic network leads to accelerated thermal denaturation of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*. *J. Mol. Biol.* **274,** 676
35. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* **271,** 789–802
36. Daniel, R. M., Dines, M., and Petach, H. H. (1996) The denaturation and degradation of stable enzymes at high temperatures. *Biochem. J.* **317,** 1–11
37. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G., and Eijsink, V. G. (1998) Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 2056–2060
38. Svingor, A., Kardos, J., Hajdu, I., Nemeth, A., and Zavodszky, P. (2001) A better enzyme to cope with cold. Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic IPMDHs. *J. Biol. Chem.* **276,** 28121–28125
39. Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C., and Feller, G. (2003) Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *J. Biol. Chem.* **278,** 37015–37023
40. Gershenson, A., Schauerte, J. A., Giver, L., and Arnold, F. H. (2000) Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. *Biochemistry* **39,** 4658–4665
41. Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., and Arnold, F. H. (2000) Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J. Mol. Biol.* **297,** 1015–1026

Received for publication September 12, 2006.
Accepted for publication January 18, 2007.

Appendix III – Extended Discussion of Chapter Three

By examining over twenty enzymes from a wide range of thermal backgrounds, it was possible to identify correlations both among Equilibrium Model parameters themselves and between Equilibrium Model parameters and other enzyme-specific traits. Apart from the major correlations that have been examined in the main text in Chapter Three, the minor correlations and, in some cases, the lack of correlation are also interesting.

The observation that $\Delta G^{\ddagger}_{cat}$ correlates weakly with $\Delta G^{\ddagger}_{inact}$, but not with optimal growth temperature suggest that the existing proposal of enzyme temperature adaptation through changes in $\Delta G^{\ddagger}_{cat}$ is potentially incomplete [26, 27, 116]. In other words, the lower $\Delta G^{\ddagger}_{cat}$ values of psychrophilic enzymes are very likely the indirect result of the general correlations between optimal growth temperature and $\Delta G^{\ddagger}_{inact}$, and between $\Delta G^{\ddagger}_{cat}$ and $\Delta G^{\ddagger}_{inact}$ (see Table 3-2), instead of a direct relationship between $\Delta G^{\ddagger}_{cat}$ and optimal growth temperature.

As discussed in Chapter Six, the reason behind the correlation between ΔH_{eq} and T_{eq} minus T_{opt} is inherent to the parameter itself, but the correlation between ΔH_{eq} and HWHM (see Table 3-2) can be related to experimental results since the 3D plots generated using Equilibrium Model parameters had been verified to resemble those derived directly from experimental data. Findings from Chapter Three and Chapter Six suggest that ΔH_{eq} doesn't correlation with any of the examined parameters except HWHM and T_{eq} minus T_{opt}. This partly corroborates the hypothesis that the E_{act}–E_{inact} equilibrium (and consequently ΔH_{eq} is intricately linked to enzyme activity and the active site, and not with global

stability ($\Delta G^{\ddagger}_{inact}$), which has long been known as an indicator of thermal adaptation. However, ΔH_{eq} doesn't appear to be correlated with $\Delta G^{\ddagger}_{cat}$ (see Table 3-2); this suggests that the active site components involved in determining ΔH_{eq} have relatively minor roles in the conformation of the active site, and that these components have limited influence over the enzyme's catalytic mechanism.

The correlation between optimal growth temperature and T_{eq} minus optimal growth temperature potentially has evolutionary implications (see Table 3-2). Specifically, it may suggest that the origin of life is thermophilic and the evolution of T_{eq} lags behind optimal growth temperature. Although it is not an uncontroversial issue, most genetic trees, along with other evidences, suggest that microbial evolution has proceeded down-temperature from a (hyper) thermophilic last common ancestor; it is reasonable to assume that if such down-temperature evolution did happen, it would still be happening today. To adapt to its living environment, an organism must first be able to grow and propagate at its environmental temperature. Once this is achieved, however, its evolutional pressure would be greatly reduced. Depending on how crucial their roles may be in the cell, different enzymes might evolve at different speeds to adapt their T_{eq} to the environment. This second theory is attractive because it explains also why T_{eq} can be tens of degrees higher than the body temperature for mammalian enzymes (see Table 3-1).

Finally, a logical step to extend the work described in Chapter Three would be a survey of enzymes of select organisms, particularly known stenothermal and eurythermal metazoans with well-defined temperature tolerance.

Appendix IV – Cloning and Verification of *A. pompejana* Episymbiont Genes

1. An Overview of the Methods

In order for the clones to be compatible with the Roche Rapid Translation System (RTS, Roche Molecular Diagnostics, Pleasanton, CA, USA), the Roche pIVEX2.4d expression vector was used (Cat. No. 3269019, Roche Molecular Diagnostics, Pleasanton, CA, USA). The pIVEX2.4d expression vector was specifically developed for *in vitro* cell-free expression using the *E. coli*-based RTS, although it also can be used for *in vivo* expression in an *E. coli* host carrying the DE3 genotype, since the vector does not contain the T7 RNA polymerase operon. The pIVEX2.4d vector places the cloned gene under T7 promoter control and adds a N-terminus His-tag for easy purification. Restriction enzymes *Nco*I and *Sma*I are recommended by the vendor for optimal asymmetrical cloning with pIVEX2.4d.

Before proceeding with primer design, all predicted gene sequences were checked for *Nco*I (CCATGG) and *Sma*I (CCCGGG) cut sites. These cut sites were absent in all sequences and subsequently added to the ends of the forward and reverse primers, respectively. Extra bases were added to the 5' end of forward primer sequences to ensure that the resulting clones were in frame with the expression vector, and random extra overhangs (ATCGAT) were added to the 5' end of primer sequences to enhance the cutting efficiency of restriction endonucleases, particularly the blunt-end *Sma*I.

A high level of SNP exists in the metagenomic sequences. Although SNPs can provide useful insights into the diversity of genes, for this work only SNPs that result in an amino acid change (noisy SNPs) were considered.

The eubacterial genetic code table (NCBI transl_table 11) was used during gene translation and codon usage analysis, since it is the genetic code table assigned to the class ε-*Proteobacteria* in NCBI taxonomy database.

2. 3-isopropylmalate dehydrogenase

125 sequence chromatograms were retrieved from the metagenome database. The ends of the sequences were trimmed with high quality cutoff, and the sequences were assembled at 96% and 50 bp overlap. 78 contigs were present in the assembly with a single prominent contig among them. This contig was manually edited to remove apparent errors and discrepancies, and an ORF with the length of 188 aa was identified, which matched the latter half of *Wolinella succinogenes* IPMDH with 70% similarity at the amino acid level. This contig was therefore incomplete. The sequence of the first 150 bp of this contig was then used to *BLAST-walk* the metagenome, and 17 Seq_IDs not identified by FGenesB were found. Those sequence chromatograms were retrieved, and all the sequences were reassembled at 96% and 50 bp overlap.

The new assembly resulted in two additional prominent contigs, one of which was found to contain an ORF of 149 aa. This ORF matched the first half of *W. succinogenes* IPMDH with 67% similarity. The two ORFs identified so far were aligned to the *W. succinogenes* IPMDH using MacGDE's (version 2.3,

http://www.msu.edu/~lintone/macgde/) TCOFFEE function (default parameters), and a 19 aa gap was found which wasn't covered by either ORF. An IPMDH consensus sequence was retrieved from CAVEEG and added to the assembly, which confirmed that the two ORFs did belong to the same gene. The resulting ORF was 1,068 bp long and had one noisy SNP.

The remaining prominent contig was manually edited, but the poor sequence quality of its 3' end makes the stop codon difficult to identify. Again, a short section of this assembly was used to BLAST-walk the metagenome, and the resulting additional chromatograms allowed the 3' end of the gene to be resolved with confidence. The resulting ORF had the same amino acid sequence (sans the noisy SNP) as the above IPMDH sequence.

The primer sequences are

Forward: `ATGGTTAAGAGATATAAGATTGGTATT` (T_m = 55ºC)

Reverse: `TTAGCTTCTGCTAACCAAACTAG` (T_m = 54.3ºC)

The final primer sequences (with extra overhang and restriction cut sites) are

Forward: `ATCGATCCATGGTTAAGAGATATAAGATTGGTATT`

Reverse: `ATCGATCCCGGGTTAGCTTCTGCTAACCAAACTAG`

Using the designed primers, a PCR amplicon of roughly 1,100 bp was amplified from Wayne genomic DNA. The PCR product was digested using *Nco*I and *Sma*I and ligated to pIVEX2.4d vector cut with the same set of restriction enzymes. The resulting ligation mix was transformed into TOP10 competent cells. Fourteen colonies were PCR-screened for the presence of the IPMDH amplicon, and six of which were selected for plasmid preparation (C3, C4, C6, C10, C13 and C14). The resulting plasmids were verified again using

cloning primers and sequenced.

Twelve sequences were obtained by sequencing all six clones forward and reverse using primers corresponding to the T7 promoter and the T7 terminator, respectively. The sequences were assembled at 80% similarity to force them into a single contig; then the contig was trimmed to remove vector sequences. In summary, C3 and C4 appeared to be identical to C10, C6 seemed to the same as C14, while C13 looked to be unique. However, in this assembly C10, C13 and C14 each contained unique noisy SNPs that did not correspond to the noisy SNP predicted by the sequence assembly. Therefore, clones C10, C13, and C14 were sequenced again, and the resulting sequences were added to the assembly.

With the additional sequences, the consensus sequences of C13 and C14 now appeared to be identical, while C10 had one noisy SNP at position 314, which changed the corresponding amino acid to Arginine from Lysine.

A translated BLAST (blastx) search of the verified sequence (C13/C14) against the GenBank protein database returned 3-isopropylmalate dehydrogenase from *Wolinella succinogenes* as the closest hit. All top ten hits were IPMDHs.

The verified IPMDH sequence was checked for codon usage using GCUA against *E. coli* codon usage pattern. Ten amino acid residues were found to be below the 10% relative adaptiveness threshold for *E. coli*, but none of which was in the first fifty amino acids of the gene.

3. Glutamate dehydrogenase

48 sequence chromatograms were retrieved from the metagenome database

and assembled at 96% and 50 bp overlap. The deepest contig only had 3 sequences, and while GDHs from other ε-*Proteobacteria* were around 450 aa long, all contigs were shorter than 1200 bp.

The data set was reassembled at 91%, 50 bp overlap, and a complete GDH-coding ORF (1362 bp) was found in the most prominent contig (containing 6 sequences), which didn't have any noisy SNP. No other contigs contained a recognizable ORF.

The primer sequences are

Forward: `ATGAGTGCAAGAGAGTATGTTAATAGTA` (T_m = 56.4ºC)

Reverse: `TTAGTTAGATTAAAGAGCCAAGAGG` (T_m = 56.7ºC)

The final primer sequences (with extra overhang and restriction cut sites) are

Forward: `ATCGATCCATGGTAATGAGTGCAAGAGAGTATGTTA`
`ATAGTA`

Reverse: `ATCGATCCCGGGTTAGTTAGATTAAAGAGCCAAGAGG`

Using the designed primers, a PCR amplicon of roughly 1,400 bp was amplified from Wayne genomic DNA. The PCR product was digested using *Nco*I and *Sma*I and ligated to pIVEX2.4d vector cut with the same set of restriction enzymes. The resulting ligation mix was transformed into TOP10 competent cells. Although no noisy SNP was predicted from the sequence assembly, only one (F3) of more than eighty colonies PCR-screened contained the GDH amplicon, possibly as a result of the strong secondary structure of the GDH amplicon. The resulting plasmid was verified again using cloning primers and sequenced.

The forward and reverse sequences of the only GDH clone were retrieved and

added to the metagenomic sequence assembly. The alignment was sound apart from a small gap between the two sequences from the clone; hence no further sequencing was performed. The clone sequences did not have any noisy SNPs compared to the predicted consensus sequence from the metagenomic sequence assembly.

A translated BLAST (blastx) search of the verified sequence against the GenBank protein database returned glutamate dehydrogenase from *Vibrio alginolyticus* 12G01 as the closest hit. All top ten hits were NADP-specific GDHs.

The verified GDH sequence was checked for codon usage using GCUA against *E. coli* codon usage pattern. Seven amino acid residues were found to be below the 10% relative adaptiveness threshold for *E. coli*, but none of which was in the first fifty amino acids of the gene.

4. Acid Phosphatase

53 sequence chromatograms were retrieved from the metagenome database. Assembly at 91% and 50 bp overlap resulted in 4 prominent contigs and some minor ones. The deepest and longest contig was edited manually, and a 261 aa-long ORF was identified. One noisy SNPs can be found in this ORF. Two other ORF encoding nearly identical proteins (260 aa) were also found in two of the remaining prominent contigs. These two ORFs are different between themselves and to the first ORF at more than 10 amino acid residues, and they shared a slightly different reverse primer from the first ORF. Therefore, the

reverse primer for ACPs contains one ambiguity.

The primer sequences are

Forward: ATGAAAAAGATACTAATAACAAATGATG (T_m = 56.2ºC)

Reverse: MTAAATCCAACTCTTTAAACTCTCA (T_m = 55.3ºC)

The final primer sequences (with extra overhang and restriction cut sites) are

Forward: ATCGATCCATGGTAATGAAAAAGATACTAATAACAA
ATGATG

Reverse: ATCGATCCCGGGMTAAATCCAACTCTTTAAACTCTCA

Using the designed primers, a PCR amplicon of roughly 800 bp was amplified from Wayne genomic DNA. The PCR product was digested using *Nco*I and *Sma*I and ligated to pIVEX2.4d vector cut with the same set of restriction enzymes. The resulting ligation mix was transformed into TOP10 competent cells. Twenty-eight colonies were PCR-screened for the presence of the ACP amplicon, and eight of which were selected for plasmid preparation (D3, D7, D12, D13, D17, D18, D19 and D24). The resulting plasmids were verified again using cloning primers and sequenced.

Due to poor sequence quality, the eight ACP clones were each sequenced at least twice in order to obtain usable DNA sequences. After some editing, the retrieved sequences were added to the metagenomic sequence assembly. As predicted by the original assembly, many noisy SNP hotspots were found (seven versus ten in the original assembly), although many of them were in fact not at the same sites. In short, apart from D17 and D19, which appeared to be identical, all other clones are distinct at one or more amino acids.

A translated BLAST (blastx) search of the verified sequence (consensus with

ambiguities) against the GenBank protein database returned acid phosphatase from *Wolinella succinogenes* DSM 1740 as the closest hit; both ends of the query aligned extremely well with the search subject. All top hits were listed as SurE protein or acid phosphatase.

The verified sequences of ACP clones D3 and D24 were checked for codon usage using GCUA against *E. coli* codon usage pattern. For both clones, ten amino acid residues were below the 10% relative adaptiveness threshold for *E. coli*, and five of which were in the first fifty amino acids of the gene.

5. Malate dehydrogenase

181 sequence chromatograms were retrieved from the metagenome database. Assembly at 96% and 50 bp overlap resulted in three prominent contigs and numerous minor ones. After some editing, two ORFs of 954 bp were identified in two of the prominent contigs, which contained no noisy SNPs and therefore identical amino acid sequences. The remaining prominent contig only contained a partial gene.

The primer sequences are

Forward: `ATGGCAAAAGGAAAAAAGGT` (T_m = 56ºC)

Reverse: `TTACTTATCAAAAAAGTTTAATTCGTAT` (T_m = 56ºC)

The final primer sequences (with extra overhang and restriction cut sites) are

Forward: `ATCGATCCATGGCAAAAGGAAAAAAGGT`

Reverse: `ATCGATCCCGGGTTACTTATCAAAAAAGTTTAATTCGTAT`

Using the designed primers, a PCR amplicon of roughly 1,000 bp was

amplified from Wayne genomic DNA. The PCR product was digested using *Nco*I and *Sma*I and ligated to pIVEX2.4d vector cut with the same set of restriction enzymes. The resulting ligation mix was transformed into TOP10 competent cells. Since no noisy SNP was predicted from the sequence assembly, only four colonies were PCR-screened for the presence of the MDH amplicon, and two of which were selected for plasmid preparation (E1 and E4). The resulting plasmids were verified again using cloning primers and sequenced.

Both MDH clones were sequenced forward and reverse, and the sequences were added to the metagenomic sequence assembly. The alignment of the clone sequences to the original assembly showed a perfect match. None of the SNPs identified in the clone sequences was noisy, so for overexpression purposes the two MDH clones could be considered identical.

A translated BLAST (blastx) search of the verified sequence against the GenBank protein database returned malate dehydrogenase from *Thiomicrospira denitrificans* ATCC 33889 as the closest hit; both ends of the query aligned extremely well with the search subject. All hits listed by the BLAST output page were MDHs.

The verified MDH sequence was checked for codon usage using Graphical Codon Usage Analyser (GCUA) against *E. coli* codon usage pattern. Ten amino acid residues were found to be below the 10% relative adaptiveness threshold for *E. coli*, and two of which were in the first fifty amino acids of the gene.

6. β-Glucosidase

60 sequence chromatograms were retrieved using COG1472 as the query, while COG2723 returned no hits at all. The sequences were assembled at 91% and 50 bp overlap, and two prominent contigs were identified. It was apparent that the most prominent contig was missing the 5' end of the gene, so 10 additional sequences were identified using BLAST-walking and added to the assembly. Another BLAST-walking was performed using the 5' end sequence of the new contig, and 6 more sequences were retrieved. However, these 6 new sequences didn't assemble into the existing contig until after significant editing. The resulting 1188 bp-long ORF was translated and compared to 12 β-GLU genes from other organisms. It was found that both N-terminus and C-terminus of β-GLU were highly variable, potentially explaining why the 16 sequences were not annotated by FGenesB initially. The ORF contained no noisy SNPs.

The other prominent contig was also missing its 5' end, but BLAST-walking did not identify enough additional sequences to allow the determination of its 5' end sequence. The rest of the ORF, however, is identical to the first one.

The primer sequences are

Forward: `ATGGCAAAAAAAATTCTAATATTAG` (T_m = 55ºC)

Reverse: `TTAAATTCTACTCTTGAGGTCATCTAC` (T_m = 55.8ºC)

The final primer sequences (with extra overhang and restriction cut sites) are

Forward: `ATCGATCCATGGCAAAAAAAATTCTAATATTAG`

Reverse: `ATCGATCCCGGGTTAAATTCTACTCTTGAGGTCATCTAC`

Using the designed primers, a PCR amplicon of roughly 1,200 bp was amplified from Wayne genomic DNA. The PCR product was digested using *Nco*I and *Sma*I and ligated to pIVEX2.4d vector cut with the same set of

restriction enzymes. The resulting ligation mix was transformed into TOP10 competent cells. Although no noisy SNP was predicted from the sequence assembly, only one (G56) of more than eighty colonies PCR-screened contained the β-GLU amplicon, possibly as a result of the strong secondary structure of the β-GLU amplicon. The resulting plasmid was verified again using cloning primers and sequenced.

Despite numerous attempts by me and staff at the Waikato DNA Sequencing Facility, no usable sequences could be obtained for the β-GLU clone. Again, this may be partly due to strong secondary structure.

A translated BLAST (blastx) search of the predicted sequence against the GenBank protein database returned β -N-acetylhexosaminidase from *Thiomicrospira denitrificans* ATCC 33889 (EC 3.2.1.52) as the closest hit. There were some β-GLU listed among the outputs, but the function of this gene isn't entirely clear.

Since the clone sequence could not be obtained, codon analysis was not performed for the β-GLU clone.

7. Protocols

Genomic DNA Source. All recombinant DNA were generated from the same genomic DNA sample that was used to generate most of the metagenomic libraries that were eventually sequenced. The said genomic DNA sample was obtained from the episymbiont consortium of an *A. pompejana* specimen nicknamed 'Wayne' collected at Bio9 vent (9ºN East Pacific Rise) during dive

#3836 of the Extreme 2002 cruise, and was designated genomic DNA sample 'AP201'. Sipper data collected along with the specimen showed that the tube from which Wayne was collected had highly dynamic chemical and thermal fluctuations, with temperatures ranging from 38 to 51ºC, pH ranging from 2 to over 5, and high [Fe^{2+}] (180 to 350 mM) within a window of ten minutes. The designations 'AP201 genomic DNA' and 'Wayne genomic DNA' are used interchangeably throughout the description.

Genomic DNA extraction. Alvinella worms were slightly thawed on an iced aluminum block, and the symbiont biomass was removed with sterile forceps and placed in a Tris-SDS-proteinase K lysis buffer for 1 hour; the nucleic acids were recovered with high salt/CTAB and phenol/chloroform extractions and subsequently precipitated with isopropanol [169]. RNA was removed with RNAse Cocktail™ (Ambion® Inc., Austin, TX, USA), and the DNA concentration and integrity measured by UV spectrophotometry and agarose gel electrophoresis. The genomic DNA samples were verified using PCR and metagenomic sequencing to contain insignificant amounts of contaminating eukaryotic DNA.

PCR assays. PCRs were carried out with a final volume of 50 µl, containing 2 µl of AP201 genomic DNA as the template, 100 nM of both primers, 200 µM of dNTPs, 3 mM of $MgCl_2$, 5 µl of 10X PCR buffer, and 2 U of *Taq* DNA polymerase (Roche Molecular Diagnostics, Pleasanton, CA, USA). PCR was performed under the following conditions: an initial denaturation at 94ºC for 2 minutes; followed by 30 cycles of denaturation at 94ºC for 30 seconds, annealing at a suitable melting temperature for each primer set for 30 seconds,

and extension at 72ºC for 60 seconds; with a final extension of 5 minutes at 72ºC. The PCR amplicons were checked for correct sizes by agarose gel electrophoresis before being purified with a GenScript QuickClean 5M PCR Purification Kit (GenScript Corp, Piscataway, NJ, USA) to remove primers and proteins.

DNA restriction and ligation assays. The purified PCR amplicons were restriction digested using the following protocol: 3 µl of Invitrogen™ REact 4 10X buffer, 0.25 µl of *Nco*I and *Sma*I (Invitrogen Corp, Carlsbad, CA, USA), 0.6 µg of DNA, and deionized water to bring the final volume up to 30 µl. The reaction mix was incubated for 2 hours at 30ºC, followed by 2 hours at 37ºC, and finally 10 minutes at 65ºC to heat-inactivate the restriction enzymes. The DNA ligation mix contained the following components: 4 ul of 5X ligation buffer, 1 µl of T4 DNA ligase (Invitrogen Corp, Carlsbad, CA, USA), 180 fmol of digested DNA, and 60 fmol of digested pIVEX2.4d plasmid DNA. The mix was incubated at 16ºC overnight. Low ligation efficiency and a high level of vector self-ligation was observed for some clones, possibly caused by strong secondary structure of the amplicons. In retrospect, this could have been partially mitigated by treating the linearized vector with alkaline phosphatase.

Transformation of competent cells. The ligation mixes were transformed into Invitrogen™ One Shot® TOP10 competent *E. coli* (Invitrogen Corp, Carlsbad, CA, USA) following the protocol provided by the manufacturer. 2.5 µl of ligation mix (containing less than 5 ng of DNA) was added to 25 µl of competent cells, incubated on ice for 30 minutes, heat-shocked for 30 seconds in a 42ºC water bath, and placed back on ice. 250 µl of pre-warmed SOC medium was added to the cell mix, and the entire reaction was shake incubated at 37ºC and 225

rpm for 1 hour before being spread on a LB agar plate containing 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol and incubated at 37ºC overnight.

PCR-screening. Depending on the number of SNPs predicted by the consensus sequence, an appropriate number of colonies were picked and PCR-screened for each gene. The colonies were picked using 10 µl pipette tips and quickly dipped in aliquoted PCR mixes before being transferred into universal bottles containing 2.5 ml of LB medium with 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol. The universal bottles were shake incubated at 37ºC and 150 rpm overnight. PCRs were carried out with a final volume of 10 µl, containing 100 nM of both T7 primers, 200 µM of dNTPs, 3 mM of $MgCl_2$, 1 µl of 10X PCR buffer, and 0.4 U of *Taq* DNA polymerase (Roche Molecular Diagnostics, Pleasanton, CA, USA). PCR was performed under the following conditions: an initial denaturation at 94ºC for 2 minutes; followed by 30 cycles of denaturation at 94ºC for 30 seconds, annealing at 52.5ºC for 30 seconds, and extension at 72ºC for 60 seconds; with a final extension of 5 minutes at 72ºC. The PCR amplicons were checked for correct sizes by agarose gel electrophoresis, and only universal bottle cultures containing the correct sized insert were used for plasmid preparation.

Plasmid purification. A GenScript QuickClean 5M Miniprep Kit (GenScript Corp, Piscataway, NJ, USA) was used for plasmid purification. 2.5 ml of overnight culture containing the correct plasmid was used following instructions provided by the manufacturer.

Appendix V – Screening, Overexpression and Purification of *A. pompejana* Episymbiont Enzymes

1. An Overview of the Methods

The Roche Rapid Translation System (RTS, Roche Molecular Diagnostics, Pleasanton, CA, USA) is a cell-free protein expression system based on *E. coil* or wheat germ lysates. For the purpose of this study, it provides the following advantages: 1) rapid screening of potential clones for functional product; 2) no codon bias since all tRNA types exist in excess in the reaction mix; 3) the ability to overexpress genes that potentially cause loss of growth or cell lysis in commonly used host organisms; 4) the compatibility of pIVEX2.4d, a vector developed specifically for RTS, with any T7-based *E. coil* expression system, provided that the host strain contains the DE3 genotype. pIVEX2.4d also adds a N-terminus His-tag to the overexpressed product, allowing simple protein purification using Ni-NTA resins or spin columns.

Due to these potential benefits, all recombinant gene overexpression work described in this thesis was performed using pIVEX2.4d to ensure compatibility with the RTS system. Although RTS was only used for functional screening in this study, it could easily be scaled up to mass produce recombinant proteins should the need arise (e.g., products that are toxic to *E. coli*), albeit at a significantly higher cost compared to *in vivo* expression systems.

Before the overexpression work began, all clones were sequenced in both

directions and checked for rare codons, as described in Appendix IV. All clones were also confirmed to be in the correct frame and free of unexpected stop codons.

Initial functional screening of the clones through induced overexpression in BL21 Star™ (DE3)pLysS showed that only GDH (F3) and IPMDH (C13/C14) clones produced functional products with this expression system. BL21 Star™ (DE3)pLysS was the first choice of expression system due to its high performance and low level of basal expression due to very tight regulation of the chromosomal T7 polymerase.

Next, the RTS was used to troubleshoot overexpression since it is neutral to the effect of codon bias and gene toxicity, as stated above. The MDH (E1) and ß-GLU (G56) clones remained non-functional when screened with the RTS, as did most of the ACP clones. However, two ACP clones (D17 and D19) exhibited elevated activity compared to other ACP clones, so these two ACP clones (which are in fact identical), along with the MDH and ß-GLU clones, were tested in the BL21(DE3)pRIL expression system, which in theory should mitigate any negative effects rare codons in those clones may have on the production of recombinant proteins. However, despite repeated attempts (i.e., different inducer concentrations, lower growth temperatures, different induction points [O.D.], different length of induction), no functional products were obtained from these clones.

Although the N-terminus His-tags were not removed from the enzymes, it is unlikely to be the cause for the lack of overexpression since all cell extracts were purified using Ni-NTA spin columns and no signs of overexpression could be

found in the purification products of non-functional clones when examined with SDS-PAGE.

2. 3-isopropylmalate dehydrogenase

Since one noisy SNP was predicted from the metagenomic sequence assembly (see Appendix IV), a total of 6 colonies was isolated from the PCR-screening process: C3, C4, C6, C10, C13, and C14. The plasmid sequences suggested that C3, C4, and C10 were identical clones, and with additional sequencing, C6, C13, and C14 were discovered to be identical at nucleotide level as well. As a result, C10, C13, and C14 were initially chosen to be overexpressed.

C10, C13, and C14 plasmids were transformed into BL21 Star™ (DE3)pLysS, and for each clone, a colony containing the correct insert size was PCR selected, grown overnight at 37ºC, and used to inoculate LB medium with 100 μg/ml of Ampicillin and 34 μg/ml of Chloramphenicol (incubation temperature = 37ºC). At an O.D. of 0.6, 1 mM of IPTG was added to the cultures to induce overexpression. The cells were harvested and lysed, and significant amounts of exogenous IPMDH activity were found in both the cell extracts and the Ni-NTA column purified products of C13 and C14, but not C10. At this point, C13 and C14 were confirmed to be identical clones, so all C13/C14 samples were pooled together, and the clones from this point on are referred to collectively as APS_IPMDH.

The purity of the APS_IPMDH preparation was verified using gel

electrophoresis. The purified APS_IPMDH sample was analyzed with gel electrophoresis, and a dominant band corresponding to the predicted size of APS_IPMDH was found on both native gel and SDS-PAGE. Based on predictions from its amino acid sequence, the subunit configuration of its homologs, and experimental evidence described above, APS_IPMDH is likely to be a homodimer with an overall size of 85 kDa.

The APS_IPMDH clone was tested with RTS and produced functional products, therefore serving as a positive control.

3. Glutamate dehydrogenase

Since no SNPs were predicted from the metagenomic sequence assembly (see Appendix IV) and only one colony (F3) was isolated from the PCR-screening process, F3 was used for all subsequent work and from this point on referred to as APS_GDH.

The APS_GDH plasmid was transformed into BL21 Star™ (DE3)pLysS, and a colony containing the correct insert size was PCR selected, grown overnight at 30ºC, and used to inoculate LB medium with 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol (incubation temperature = 30ºC). At an O.D. of 0.4, 1 mM of IPTG was added to the culture to induce overexpression. The cells were harvested and lysed, and significant amounts of exogenous GDH activity were found in both the cell extract and the Ni-NTA column purified product. The purity of the APS_GDH preparation was verified using gel electrophoresis. The purified APS_GDH sample was analyzed with gel electrophoresis, and a

dominant band corresponding to the predicted size of APS_GDH was found on both native gel and SDS-PAGE. Based on predictions from its amino acid sequence, the subunit configuration of its homologs, and experimental evidence described above, APS_GDH is likely to be a homohexamer with an overall size of 320 kDa.

The overexpression of recombinant APS_GDH was found be equally successful when incubated and induced at 10ºC, and additional APS_GDH was purified.

The APS_GDH clone was tested with RTS and produced functional products, therefore serving as a positive control.

4. Acid Phosphatase

Since more than ten noisy SNPs were predicted from the metagenomic sequence assembly (see Appendix IV), a total of 8 colonies were isolated from the PCR-screening process: D3, D7, D12, D13, D17, D18, D19, and D24. The plasmid sequences suggested that while D17 and D18 are identical at nucleotide level, the remaining clones all contained one or two amino acid variations. Based on the uniqueness of their sequences, ACP clones D3 and D24 were picked for overexpression trials.

D3 and D24 plasmids were transformed into BL21 Star^{TM} (DE3)pLysS, and for each clone a colony containing the correct insert size was PCR selected, grown overnight at 30ºC, and used to inoculate LB medium with 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol (incubation temperature = 30ºC).

At an O.D. of 0.6, 1 mM of IPTG was added to the cultures to induce overexpression. The cells were harvested and lysed. The cell extracts did not exhibit any ACP activity attributable to exogenous recombinant protein; neither did the Ni-NTA column purified products. The same experiment was carried out several more times with at least one of the following variations: 2 mM or 5 mM of IPTG instead of 1 mM, earlier induction (between O.D. values of 0.3 and 0.4), longer induction period (which did not result in a higher final O.D. value), and incubation at 37ºC and 10ºC. No signs of functional recombinant protein were observed in any of the overexpression attempts from D3 or D24.

Cells from the induced cultures were examine under a microscope, and bright spots could be found inside most of the cells, hinting at the presence of inclusion bodies. SDS-PAGE of the D3 and D24 cell extracts did not show any signs of overexpression. Whole cell samples of D3 and D24 were also lysed under harsh denaturing conditions with the hope of releasing the protein inside inclusion bodies, but SDS-PAGE of these samples did not reveal any additional bands. This suggested that the potential recombinant products inside the inclusion bodies were unlikely to be recovered under conditions that allow simple refolding of the protein.

The D3 and D24 clones were tested with RTS, and while the GFP positive control, APS_IPMDH, and APS_GDH all produced functional products, no ACP activity could be detected under the assay condition used. Later on, all eight ACP clones were expressed using RTS, and the resulting products were assayed at 50ºC using a wide selection of buffers, including 100 mM Na^+ acetate buffer, pH 5 & pH 5.5, 100 mM MES buffer, pH 6, 50 mM citrate buffer, pH 5 & pH 5.5,

and 50 mM glycine-HCl buffer, pH 2.4. Minute amounts of ACP activity were observed when 100 mM MES buffer, pH 6 was used as the reaction buffer, and overnight assays suggested that D17 and D19 exhibited somewhat elevated activity (50%-200% higher compared to other ACP clones), while D7, D12, and D24 had ACP activity at the same level as the negative control (APS_GDH clone). This suggested that D17 and D19 were potentially producing recombinant products at very low concentrations, and all overexpression attempts from this point on were performed using these two clones. This experiment demonstrated the potential of RTS in rapid functional screening, and that the low expression level in RTS may have precluded the formation of inclusion bodies or protein aggregates, thus the observable ACP activity.

D17 and D19 clones were transformed into BL21(DE3)pRIL for functional screening, but very little recombinant ACP activity could be detected in the cell extracts, and activity from the Ni-NTA column purified products was indistinguishable from blank reactions. The lack of overexpression was confirmed by SDS-PAGE of the cell extract.

5. Malate dehydrogenase

Since no SNPs were predicted from the metagenomic sequence assembly (see Appendix IV), only four colonies were PCR-screened and two (E1 and E4) were eventually isolated. Based on the sequences of the plasmids, the two clones were found to be identical at nucleotide level, thus only E1 was used for subsequent work and from this point on referred to as APS_MDH.

The APS_MDH plasmid was transformed into BL21 Star™ (DE3)pLysS, and a colony containing the correct insert size was PCR selected, grown overnight at 30ºC, and used to inoculate LB medium with 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol. It should be noted that at this point the APS_MDH overnight culture is significantly more dilute than overnight cultures of other clones being overexpressed in parallel, and the culture did not reach an O.D. of 0.6 until 11 hours after inoculation whereas other cultures reached an O.D. of 0.6 after around 4 hours (incubation temperature = 30ºC). At an O.D. of 0.5, 1 mM of IPTG was added to the culture to induce overexpression. The cells were harvested and lysed. The cell extract did not exhibit any MDH activity attributable to exogenous recombinant protein; neither did the Ni-NTA column purified product. The same experiment was carried out several more times with a least one of the following variations: 2 mM or 5 mM of IPTG instead of 1 mM, earlier induction (between O.D. values of 0.3 and 0.4), longer induction period (which did not result in a higher final O.D. value), and incubation at 37ºC and 10ºC. No signs of functional recombinant protein were observed in any of the overexpression attempts.

Cells from the induced culture were examined under a microscope; the cell morphology was excellent, and no signs of inclusion body were observed. SDS-PAGE of the APS_MDH cell extracts did not show any signs of overexpression. The presence of APS_MDH plasmid in the post-induction culture was verified using PCR.

The APS_MDH clone was tested with RTS, and while the GFP positive control, APS_IPMDH, and APS_GDH all produced functional products, no MDH

activity could be detected under the assay condition used.

The APS_MDH clone was also transformed into BL21(DE3)pRIL for functional screening, and only *E. coli* constitutive MDH activity was detected. The lack of overexpression was confirmed by SDS-PAGE of the cell extract.

6. β-Glucosidase

Since no SNPs were predicted from the metagenomic sequence assembly (see Appendix IV) and only one colony (G56) was isolated from the PCR-screening process, G56 was used for subsequent work and from this point on referred to as APS_GLU.

The APS_GLU plasmid was transformed into BL21 Star™ (DE3)pLysS, and a colony containing the correct insert size was PCR selected, grown overnight at 37ºC, and used to inoculate LB medium with 100 µg/ml of Ampicillin and 34 µg/ml of Chloramphenicol (incubation temperature = 37ºC). At an O.D. of 0.6, 1 mM of IPTG was added to the culture to induce overexpression. The cells were harvested and lysed. The cell extract did not exhibit any activity (see Protocols below) attributable to exogenous recombinant protein; neither did the Ni-NTA column purified product. Although high levels of ß-galactocosidase activity was observed, this was most likely due to the native *E. coli* ß-galactocosidase being induced by IPTG. The same experiment was carried out several more times with at least one of the following variations: 2 mM or 5 mM of IPTG instead of 1 mM, earlier induction (between O.D. values of 0.3 and 0.4), longer induction period (which did not result in a higher final O.D. value), and

incubation at 30ºC and 10ºC. No signs of functional recombinant protein were observed in any of the overexpression attempts.

Cells from the induced culture were examined under a microscope; the cell morphology was excellent, and no signs of inclusion body were observed. SDS-PAGE of the APS_GLU cell extracts did not show any signs of overexpression.

The APS_GLU clone was tested with RTS, and while the GFP positive control, APS_IPMDH, and APS_GDH all produced functional products, no enzyme activity could be detected under the assay condition used.

The APS_GLU clone was also transformed into BL21(DE3)pRIL for functional screening, and no recombinant protein activity could be detected. The lack of overexpression was confirmed by SDS-PAGE of the cell extract.

7. Protocols

Transformation and electroporation of competent cells. The protocol for transforming BL21 Star™ (DE3)pLysS chemically competent *E. coli* was identical to the transformation of TOP10 competent *E. coli* described in Appendix IV. The electroporation of BL21(DE3)pRIL electrocompetent *E. coli* was performed with a 0.1 cm electroporation cuvette. 2.5 µl of plasmid DNA was mixed with 75 µl of competent cells and electroporated following instructions accompanying the electroporator. Immediately after electroporation, 1 ml of pre-warmed SOC medium was added to the electroporation mix, and the entire reaction was shake incubated at 37ºC and 225 rpm for 1 hour before being spread

on a LB agar plate containing 100 μg/ml of Ampicillin and 34 μg/ml of Chloramphenicol and incubated at 37ºC overnight.

Expression in *E. coli*. A typical recombinant protein overexpression operation in *E. coli* is as follows: a pilot culture containing around 60 ml of LB medium with 100 μg/ml of Ampicillin and 34 μg/ml of Chloramphenicol is inoculated with 1 ml of overnight culture from a single colony confirmed to carry the desired plasmid by PCR-screening. For screening purposes, this pilot culture is incubated at the desired temperature (e.g., 37ºC) until its has reached the designated O.D. value before a suitable amount IPTG is added to induce overexpression. The culture is then incubated at the desired temperature for an appropriate period, depending on the induction temperature (4-6 hours for 30ºC and 37ºC, 24 hours for 10ºC), before the culture is harvested by centrifugation. For large scale overexpressions, 50 ml of the pilot culture is used to inoculate 1 liter of pre-warmed LB medium containing 100 μg/ml of Ampicillin and 34 μg/ml of Chloramphenicol, and the remaining induction process is similar to that described above. Note that appropriately sized incubation vessels should always be used to allow sufficient oxygen exchange (i.e., large surface area) and mixing. For large-scale overexpressions, the remaining pilot culture can be used to prepare a glycerol stock. The culture is spun down and resuspended in 1 ml of 1:1 LB medium (containing the appropriate antibiotics) and glycerol, then incubated at room temperature for 15 minutes before being stored at -80ºC.

Cell lysis and protein purification. Harvested cell pellets were thawed and resuspended in a suitable amount of lysis buffer (roughly 1 ml of buffer for one gram of wet cells) as specified in the QIAGEN Ni-NTA Spin Kit. Lysozyme

was then added to the lysis mix to a final concentration of 2 mg/ml, and the mix was incubated on ice for one hour. The suspension was sonicated at the highest power output permitted for a microtip for six times, 15 seconds each. Using a refrigerated centrifuge at 4ºC, the sonicated cells were then spun down in 2 ml Eppendorf tubes at maximal speed for 30 minutes. The resulting supernatant (i.e., cell extract) was then assayed for exogenous enzyme activity. Where necessary, the cell extracts were purified using QIAGEN Ni-NTA Spin Kit following procedures as described in the manufacturer's instructions.

Functional screening of expression products. Functional screening of expression products was performed with the corresponding enzyme activity assays performed at 30ºC or 50ºC. For APS_IPMDH, the assay mix contained up to 5 mM of NAD and isopropylmalate, 1 mM of $MgCl_2$ (an activator for the enzyme), and 100 mM K^+ phosphate buffer at pH 8. The formation of NADH was monitored at 365 nm. For APS_GDH, the assay mix contained 0.6 mM of NADPH, 7.5 mM of α-ketoglutarate, 150 mM of NH_4Cl, and 100 mM K^+ phosphate buffer, pH 7.5. The oxidation of NADPH was monitored at 365 nm. For the ACP clones, initially the assay was perform with the follow condition: 5 mM of *p*NP-phosphate in 100 mM Na^+ acetate buffer, pH 5. The release *p*-nitrophenol phosphate (*p*NP) was measured at 410 nm. After buffer optimization (see above), the assay condition was changed to 5 mM of *p*NP-phosphate in 100 mM MEP buffer, pH 6. For APS_MDH, the assay mix contained 0.8 mM of NADH, 2 mM of OAA, and 100 mM K^+ phosphate buffer, ph 8. The oxidation of NADH was monitored at 365 nm. For APS_GLU, multiple substrates were used for the first few overexpression attempts. The

assays were performed under the following conditions: 1 mM of *p*NP-α-glucopyranoside, *p*NP-ß-glucopyranoside, *p*NP-ß-galactopyranoside, or *p*NP-ß-xyloside in 100 mM K^+ phosphate buffer, pH 8. The release of *p*NP was measured at 410 nm. In later screening, only *p*NP-α-glucopyranoside and *p*NP-ß-glucopyranoside were used.

RTS expression. *In vitro* expression using the RTS system was performed following the manufacturer's instructions. Around 500 ng of plasmid DNA was used for each 25 µl reaction, which was incubated at 30ºC and 600 rpm for 5 hours. The expression products were immediately assayed using conditions described above and examined using SDS-PAGE, which did not reveal any obvious signs of overexpression, even in the positive controls (APS_GDH and APS_IPMDH). The GFP positive control was incubated at 4ºC for 24 hours, and the presence of GFP fluorescence was confirmed using a UV transilluminator.

Preparation of *E. coli* BL21(DE3)pRIL competent cells. A single colony of BL21(DE3)pRIL was used to inoculate 5 ml of LB medium containing 50 µg/ml Chloramphenicol, which was incubated at 37ºC overnight. The culture was spun down in 2 ml Eppendorf tubes the next day and resuspended in 1 ml of LB medium containing 50 µg/ml Chloramphenicol, which was then used to inoculate 500 ml of pre-warmed LB medium containing 50 µg/ml Chloramphenicol in a 2 liter flask. This culture was incubated at 37ºC until its O.D. reached 0.6, then spun down, and resuspended in 500 ml of 10% sterilized ice-cold glycerol. The cells were then spun down again and resuspended in 250 ml of 10% sterilized ice-cold glycerol. The process was repeated two more times, and the cells were resuspended in 20 ml, then 2 ml of 10% sterilized ice

cold glycerol after each centrifugation step. During this process, the cell suspension was kept chill (i.e., on ice) at all times, and all containers were sterilized by autoclaving in advance. The final suspension of 2 ml was then made into 75 µl aliquots in 0.5 ml Eppendorf tubes and flash frozen in liquid nitrogen. The prepared BL21(DE3)pRIL electrocompetent cells were stored at -80ºC.

Appendix VI – Extended Discussion of Chapter Five

The research results presented in Chapter Five, while not containing sufficient evidence to identify clear trends of temperature adaptation in the episymbiotic bacteria of *Alvinella pompejana*, serves as a demonstration that by combining metagenomics and the Equilibrium Model, it is possible to examine temperature adaptation of otherwise inaccessible organisms through characterizing the thermal adaptation of their enzymes.

From Table 5-1, it is clear that *aps_gdhA* has a higher expression per copy of gene than *aps_leuB* does. This potentially corresponds to the housekeeping role of *aps_gdhA*'s expression product, APS_GDH. On the other hand, *aps_leuB* may be a much greater potential for elevated expression under certain conditions, as described in Chapter Five. However, it is difficult to investigate the direct significance of these findings since no information is available regarding the temperature condition when the worm was collected.

Although the difficulty in drawing conclusions from the results presented in Chapter Five is largely due to the small number of enzyme examined, additional factors further complicates their interpretation. The fact that a complex and unculturable microbial community was the target of investigation makes the interpretation of RT-qPCR results difficult and any potential conclusion vague, and the reported behavioral adaptation by the host, *Alvinella pompejana*, adds yet another layer of uncertainty to the thermal stress imposed on the episymbionts.

To adequately and confidently examine the role of enzyme temperature adaptation among multiple layers of cellular adaptations that can be employed by

the organism, two potential approaches may be taken. The first potential approach is to perform a comprehensive survey of both housekeeping and non-critical enzymes of a strictly stenothermal and immobile organism. Although a stenothermal lifestyle does not necessarily exclude the presence of eurythermal enzymes, the general trend of enzyme thermal adaptation in such organisms may differ from those living in more thermally dynamic environments. The second approach is to use a known eurythermal and immobile organism and target its housekeeping enzymes for characterization using the Equilibrium Model. However, this approach must be coupled with gene expression analysis using RT-qPCR or microarray of the organism growing at different temperatures since the organism will likely respond metabolically to different temperatures. For example, a bacterium may be grown at five different temperatures, and active cultures can be either snap-frozen or treated with RNA preserving material (e.g., RNAlater) to obtain a snapshot of its gene expression pattern at a particular temperature. Both approaches require the use of immobile organisms or at least organisms that cannot react to temperature changes by seeking refuge in areas with more favorable thermal conditions in order to avoid complications in interpreting the findings due to host behavioral adaptation.