NOTE TO USERS

This reproduction is the best copy available.

UMI®
RICE UNIVERSITY

Single Gene Evolution: a Punctuated History

of Chance Events

Rafael Couñago

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Yousif Shamoo, Assistant Professor, Thesis Advisor
Biochemistry and Cell Biology

John S. Olson, Professor,
Biochemistry and Cell Biology

George N. Bennett, Professor,
Biochemistry and Cell Biology

Edward P. Nikonowicz, Professor,
Biochemistry and Cell Biology

David C. Queller, Professor,
Ecology and Evolutionary Biology
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
ABSTRACT

In the present work I have developed and characterized an in vivo selection system to investigate molecular evolution. The essential gene that codes for adenylate kinase (AK; EC 2.4.7.3) in the gram-positive moderate thermophile Geobacillus stearothermophilus was replaced through homologous recombination with its counterpart from the mesophilic Bacillus subtilis. PCR, DNA sequencing and Southern analysis confirmed proper gene replacement and preservation of neighboring genes in the recombinant strain. Recombinant cells (NUB3621-R:ThEV) displayed a temperature sensitive phenotype, with a highest growing temperature almost 20 °C lower than that of wild-type cells (56 °C vs 75 °C). The temperature sensitive phenotype in recombinant cells was linked to a disruption of adenylate homeostasis at high temperatures, secondary to B. subtilis AK heat inactivation, as shown by enzyme activity assays, temperature denaturation profiles and adenylate level measurements. Evolution of a single gene, B. subtilis adk, was investigated by steady-state growth of the recombinant strain from 55 to 70 °C in a turbidostat. The temporal characteristics of B. subtilis adk evolution were probed by DNA sequence analysis at various temperatures. The appearance of more fit strains from pre-existing genetic variation, rapid extinctions generated by clonal interference and selection coupled to a strong dependence on historical context and chance events resulted in a punctuated pattern of evolution. The organism fitness at different temperatures could be linked to its expressed AK variant and allowed
the investigation of the biochemical and structural basis of adaptation. Heat denaturation and enzyme activity studies showed that all isolated AK mutants were more stable than the wild-type protein and were responsible for the punctuated pattern of evolution detected during the population analyses. The atomic structure of one of the isolated AK variants (Q199R) revealed that its increased thermostability is due to unique electrostatic interactions absent in the wild-type structure. Taken together our population, biochemical and structural analysis suggests that natural selection observed at the molecular level is guided by the same principles that act at the organismal level.
Acknowledgements

"The road goes ever on and on". Without the presence of several people the ride would not have been so enjoyable.

I would like to thank Paula for her cheerful presence, unconditional companionship and love throughout this long road.

I thank my parents and sister for their love and support throughout my life and especially during the last 6 years that we have been apart.

The rest of my family and friends that I left behind in Brasil, I thank for the good times spent during my seasonal appearances every December.

I thank Gil and Hatice for all the great times together and the long arguments about life. That your fish tank never dries!

I thank my advisor, Dr. Shamoo, for support and insight. His Herculean endeavor to build a multidisciplinary lab from ground zero should not be forgotten. His personal and professional skills have created a pleasant place to work and make science.

I thank Kevin for his big brains and his endless desire to spread knowledge.

My committee also deserves many thanks. They have been both patient and helpful. My committee chair, Dr. Olson, deserves special thanks. Few people have his charisma and kind heart.

Thanks to the Shamoo crew for support and help with technical problems and patience with personal issues. The original twin brothers, John and Jeff,
made the Shamoo lab a very interesting place to be. The happy girls Siyang, Megan and Linda made the lab a very pleasant place to be. I also thank Marc, Stephen and Victoria for all their effort and help. I also thank Jim and Dr. Wu for assistance and useful discussions.

It would be unfair not to thank the many people in the third floor of Keck hall: Endah, Corey, Phil, “grandpa”, Ian, Todd, Ye, Jayashree, Susan and all the other for help with various equipments and for being such nice peers.

This work would have never been accomplished without the Keck ladies, especially Lisa, and the G.R. Brown crew: Dolores, Jo, Meredith, Minnie, Sandy, Merilyn and Juan.

This work was in part funded by the National Science Foundation and the W. M. Keck Center for Computational and Structural Biology.
# Table of Contents

Figures and Tables ........................................................................................................... viii

Chapter 1 - Introduction ............................................................................................... 1

Chapter 2 - Evolution .................................................................................................... 5

2.1 From Darwin to the Synthesis (1859 - 1942) ....................................................... 5
2.2 Modifying the synthesis; a chance for history and neutrality ......................... 8
2.3 Phylectic gradualism vs Punctuated equilibria ................................................. 9
2.3 Molecular evolution .............................................................................................. 12
2.4 Summary ................................................................................................................ 18

Chapter 3 - Weak Link Evolution .................................................................................. 20

3.1 The phosphotransfer network as a weak link .................................................. 22
3.2 Adenylate kinase (EC 3.4.2.7) in the phosphotransfer network .................... 24
3.3 B. subtilis AK ......................................................................................................... 25
3.4. Geobacillus stearothermophilus ..................................................................... 26
3.5 Summary ................................................................................................................ 28

Chapter 4 - Mutation .................................................................................................... 30

4.1 Patterns of mutations ......................................................................................... 30
4.2 The power of mutations and selection .............................................................. 33
4.4 Summary ................................................................................................................ 37

Chapter 5 - Protein folding and stability ..................................................................... 39

5.1 Hydrophobic interactions and conformational entropy ................................... 41
5.2 Electrostatic interactions ..................................................................................... 42
5.3 B. subtilis AK stability ......................................................................................... 44
5.4 Summary ................................................................................................................ 45

Chapter 6 - Methods ................................................................................................... 46

6.1 Bacterial cultures and Media ............................................................................. 46
6.1.1 Escherichia coli ................................................................................................. 46
6.1.2 Geobacillus stearothermophilus .................................................................. 48
6.2 Isolation and manipulation of Plasmid and Genomic DNA ............................ 49
6.3 Polymerase chain reactions .............................................................................. 49
6.4 Construction of G. stearothermophilus strain NUB3621-R:ThEV ................. 50
6.4.1 Construction of a homologous recombination vector ................................ 52
6.4.2 Transformation of G. stearothermophilus NUB3621-R with plasmid DNA 53
6.4.3 Isolation of recombinant NUB3621-R:ThEV cells .................................... 55
6.4.4 Characterization of NUB3621-R:ThEV cells ............................................. 55
6.5 Growth of NUB3621-R:ThEV in a turbidostat ................................................ 57
6.6 Isolation of adk mutants generated during turbidostat runs ............................ 59
6.7 Mutant adk population analysis ........................................................................ 60
6.8 Mutant adk cloning, expression and purification ............................................. 61
6.8.1 Site-directed mutagenesis .............................................................................. 61
6.8.2 AK expression in E. coli .................................................................................. 62
6.8.3 AK purification from cell lysates

6.8.3.1 Miscellaneous information on chromatography columns and protein handling

6.8.3.2 DE52 chromatography

6.8.3.3 FPLC MonoQ chromatography

6.8.3.4 Blue Sepharose chromatography

6.8.3.5 Size exclusion chromatography

6.9 Temperature denaturation of AK followed by Circular Dichroism

6.10 AK activity assays

6.11 Crystallization of B. subtilis AK mutants

6.12 AK mutant crystal analysis and data collection

6.13 X-ray data processing

6.14 Molecular Replacement

6.15 Solution refinement

6.16 Total Adenine Nucleotide Extraction

6.17 Total protein extraction

Chapter 7 - The Recombinant NUB3621-R:ThEV strain

7.1 NUB3621-R adk is part of the S10-spca cluster

7.2 Homologous recombination

7.3 Genetic characterization of NUB3621-R:ThEV cells

7.4 B. subtilis AK expression in recombinant NUB3621-R:ThEV cells

7.5 AK activity and cellular metabolism

Chapter 8 - Single gene evolution: a punctuated history of chance events

8.1 Alternative evolutionary strategies

8.2 Evolution of B. subtilis AK in NUB3621-R:ThEV

8.3 Evolution of NUB3621-R:ThEV was punctuated

8.4 Punctuated evolution under strong selective pressure

8.5 NUB3621-R evolution; first come first served

8.6 Random neutral evolution in NUB3621-R:ThEV

8.8 B. subtilis AK evolution was punctuated

8.9 History and chance dictated the evolution of B. subtilis AK

Chapter 9 - The biochemical and structural basis of adaptation

9.1 Expected patterns

9.2 Primary sequence constraints dictated AK evolution

9.3 Biochemical basis for AK stability

9.4 Structural basis of adaptation

Chapter 10 - Conclusions
Figures and Tables

Figure 2-1- Gradualism vs Punctuated equilibria. 10
Figure 3-1 AK is a highly flexible enzyme. 27
Table 4-1 - Pattern of nucleotide substitution in pseudogenes 31
Figure 4-1 - Selection greatly increases the rate of evolution. 35
Figure 5-1- The theoretical temperature stability curve for $\Delta G_{stab}$ vs T. 40
Table 6-1 Bacterial strains and DNA plasmids 47
Table 6-2 Oligonucleotide primers 51
Figure 6-1- The “evolution machine”. 58
Figure 6-2 - Purified AK 69
Figure 7-1- Genome location and sequence information for NUB3621-R adk gene. 80
Figure 7-2- The homologous recombination cassette and the p12ThEV construct. 82
Figure 7-3 - Temperature sensitive phenotype of NUB3621-R:ThEV cells. 84
Figure 7-4 - Genetic characterization of NUB3621-R:ThEV cells. 85
Table 7-1 – Purification parameters for AK from NUB3621-R and NUB3621-R:ThEV cells 87
Table 7-2 Energy Charge and total protein levels for NUB3621-R and NUB3621-R:ThEV at different growth conditions 92
Figure 8-1 - The two employed strategies to follow NUB3621-R:ThEV evolution in the turbidostat. 99
Figure 8-2 - B. subtilis AK population dynamics displayed a punctuated pattern. 101
Figure 8-3 - Temperatures where B. subtilis AK mutants first appear in the population. 105
Figure 8-4 - Evolved B. subtilis AK variants are more thermostable than the wild-type enzyme. 112
Figure 8-5 - Single point mutations had different impacts on Q199R background. 115
Table 8-1 - Stability gains and mutational patterns for B. subtilis AK variants 118
Figure 9-1 - B. subtilis AK evolution was constrained by primary and tertiary structures. 122
Figure 9-2 - Protein crystals for the B. subtilis AK Q199R mutant. 127
Table 9-1 Crystallographic data statistics 128
Figure 9-3- Electron density details of B. subtilis Q199R AK atomic structure at 1.8 Å resolution. 129
Figure 9-4- B. subtilis Q199R AK mutant displayed an identical overall atomic structure to the wild-type protein. 131
Figure 9-5- The mutant Arg199 promoters better ionic interactions than the wild-type Gln199 residue. 133
Figure 9-6 - Putative structural basis for the higher thermostability of mutant Q16L. 135
Figure 9-7- Putative structural basis for the higher thermostability of mutant A193V. 136
Figure 9-8 - Putative structural basis for the higher thermostability of mutant T179I. 138
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>adk</td>
<td>adenylate kinase gene</td>
</tr>
<tr>
<td>AK</td>
<td>Adenylate Kinase</td>
</tr>
<tr>
<td>Ap5A</td>
<td>P¹, P⁵-di(adenosine-5') pentaphosphate</td>
</tr>
<tr>
<td>cat</td>
<td>chloramphenicol acetyl-transferase gene</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyl-transferase protein</td>
</tr>
<tr>
<td>EC</td>
<td>Adenylate Energy Charge</td>
</tr>
<tr>
<td>infA</td>
<td>initiation factor IF-1</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kiloDaltons</td>
</tr>
<tr>
<td>RBS</td>
<td>ribosomal binding site</td>
</tr>
<tr>
<td>M</td>
<td>Malthusian coefficient</td>
</tr>
<tr>
<td>map</td>
<td>methionine aminopeptidase</td>
</tr>
<tr>
<td>secY</td>
<td>pre-protein translocase secY subunit</td>
</tr>
<tr>
<td>s</td>
<td>selection coefficient</td>
</tr>
<tr>
<td>W</td>
<td>Darwinian fitness</td>
</tr>
</tbody>
</table>
Chapter 1 - Introduction

“Call me morbid, call me pale
I've spent six years on your trail
Six full years of my life on your trail
And if you have five seconds to spare
Then I'll tell you the story of my life.”

The Smiths - Half a person

Populations are not static entities. Darwin was the first to realize that phenotypical variations observed for single species could be used to explain all of life’s diversity on Earth. Darwin called the process by which populations change over time evolution. He also identified natural selection as a mechanism through which evolution could take place.

We now know that the presence of different phenotypes in populations is due to their genetic heterogeneity. Identification and characterization of the mechanisms behind changes in the genetic content of a population over time is the major concern of evolutionary theory. Evolutionary mechanisms have been identified through the use of quantitative approaches and careful experimentation. These studies were mainly concerned with speciation in higher taxa and have established the importance of selection, chance and history in guiding the fate of genetic variability in a population over time.
Studies with microbial populations have started to probe the genetic basis of evolution. Thus far, these studies have been concerned with the adaptation of whole genomes during microbial evolution and have greatly contributed to our understanding of evolution at the molecular level. The present study aims to identify and characterize the evolutionary mechanisms working at the single gene level and to investigate the biochemical and structural basis for adaptation.

Single gene evolution will be studied by a weak link approach. In our weak link experiment an organism’s growth at non-permissive conditions is made dependent on the modification of a key gene. Through genetic manipulation, I have made the proper functioning of the energy metabolism of a thermophile, G. stearothermophilus, dependent on the thermostabilization of one of its key components, adenylate kinase. The incorporation of the weak link, B. subtilis adk, into the phosphotransfer network of a thermophile ensured that the evolutionary process followed in this work was of physiological relevance.

The recombinant organism was cultivated under steady-state conditions in a small turbidostat. Growth temperature was increased from permissive, 55°C, to non-permissive, 70°C, at a rate of 0.5°C per day. Thermostabilization of AK was expected to happen through the accumulation of point mutations in the adk gene and, therefore, to be restricted by DNA chemistry and the host’s DNA replication and repair machinery. More importantly, the founder gene, B. subtilis adk, possesses a previous evolutionary history and was also expected to impose constraints on the evolution of the adk gene.
The temporal characteristics of *B. subtilis adk* evolution were investigated by withdrawing culture samples and sequencing the *adk* gene every 1°C. Coexistence of different *adk* mutants in the turbidostat population was expected to lead to Malthusian competition, where better fit variants experience greater reproductive success and, therefore, have a better chance to be represented in the following generations. The importance of chance, history and selection during the evolution of the weak link was investigated by head to head (Malthusian) competition between founder and evolved cell lines.

Protein thermostability is achieved by a set of limited thermodynamic rules that limits the fitness landscape available for the protein. Therefore, in addition to the mutational landscape evolution of *B. subtilis* AK is further constrained by the thermodynamic landscape. As the host's growth is directly dictated by the physicochemical characteristics of its AK enzyme, the thermodynamic basis of adaptation could be investigated. *B. subtilis* AK variants were cloned, expressed and purified from a heterologous *E. coli* system. Temperature stability of mutant AKs were probed by heat denaturation and enzyme activities. These studies investigated the different thermodynamic strategies employed for the various AKs. Protein X-ray crystallography was employed to investigate the unique structural differences responsible for the increased thermostability of the selected AK enzymes.

The following chapters will introduce the concepts that are going to be employed in analyzing and discussing this work's findings. Chapter 2 is concerned with our current evolutionary theory. Chapter 3 introduces the weak
link approach and the rationale behind the utilized model system. Chapter 4 deals with the accumulation of genetic changes in a population through mutations and how selection can affect the outcome of evolution. Chapter 5 introduces basic notions in protein thermodynamics and how it affects protein thermostability. Chapter 6 describes the methods employed during this study and Chapter 7, 8 and 9 describe and analyze its findings. In Chapter 10 a summary of this work’s achievements can be found as well as its future directions.
Chapter 2 - Evolution

"There is grandeur in this view of life, with its several powers, having been originally breathed by the Creator into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved." Darwin, C.

The main goal of the present work was to investigate evolution at the molecular level in a bacterial population under selective pressure. In this chapter the vast and fascinating literature of our current evolutionary theory will be briefly presented. Special attention will be observed to findings at the molecular level and to the use of microorganisms to probe evolution.

2.1 From Darwin to the Synthesis (1859 - 1942)

Charles Darwin (1809 - 1882) was the first to propose a mechanism to explain the astonishing diversity of life on earth. Darwin's field observations made him realize three of life's attributes: overproduction of offspring, natural variation and heritability. Darwin's greatest insight was the addition of a new ingredient to life's recipe, natural selection; a mechanism by which naturally occurring, better-adapted variants among overproduced offspring experienced differential
reproductive success. Better-adapted variants would, in turn, pass their characters to their offspring through heritability (Darwin, 1998). His theory on the origin of the species through natural selection is, to this day, called Darwinism.

Darwin’s ideas relied on countless observations and examples gathered by Darwin himself, mainly during his voyage in the HMS Beagle. The impossibility of applying quantitative methods to Darwin’s theory fueled critics and created divergent opinions among supporters. Particular concerns were focused on the origin of diversity and natural selection’s attributes and role during the origin of the species. Moreover, the heritability model at Darwin’s time was incorrect and did not predict the maintenance of variability over time in a population.

Rediscovery of Mendel’s work solved the heritability problem and, in time, was used to answer most critiques to Darwin’s theory. Incorporation of Mendel’s particulate model onto Darwin’s theory resulted in a “synthetic” theory of evolution (Fisher, 1930). The resulting modern synthesis furnished evolutionists with a powerful framework to design experiments and test hypotheses in evolutionary theory. In time, it dismissed all alternative explanations to evolution in favor of Darwinism.

The modern synthesis had a profound impact in our current understanding of evolution. Its findings and principles will be heavily employed throughout the present work. The synthetic theory defined evolution as changes in the genetic makeup of a population over time. Mutations (Chapter 4) were established as the source of genetic variation (Wright, 1932). Theoretical (Haldane, 1924) and
experimental studies (Dobzhansky and Pavlovsky, 1957) led to the identification of new mechanisms by which evolution could work. Factors such as migration and genetic drift were added to Darwin’s natural selection as evolutionary forces behind changes in the genetic makeup of a population over time (Dobzhansky and Pavlovsky, 1957). These factors, working continuously on a population, would, in time, lead to the appearance of new biological species.

The synthetic theory’s success in explaining evolution through population genetics consolidated Darwinism as the only explanation for evolution and speciation. Darwin’s theory stated, and the synthetic theory agreed, that variants arose by the accumulation of gradual changes and were fixed in the population mainly through natural selection. A caveat of identifying natural selection as the chief force behind evolution is that all observed changes in a population must be adaptive (Cain, 1964). The modern synthesis defined fitness as the adaptive worth of an organism and a direct measure of how a genotype will be represented in the next generation of a population. The modern synthesis stated that modifications that are not followed by an increment in fitness will be eliminated from the population. Neutrality, history and chance were considered not important for the accumulation of diversity and the eventual production of new biological varieties.
2.2 Modifying the synthesis; a chance for history and neutrality

The availability of molecular data led to the discovery of non-Darwinian (neutral) evolution. Analysis of sequence information from various proteins from different sources led Kimura (Kimura, 1968) and others (King and Jukes, 1969) to propose that most changes at the molecular level are neutral (non-adaptive) and, therefore, invisible to natural selection. Kimura showed that the chance of a neutral mutation becoming fixed in a population was only dependent on its initial frequency (Kimura, 1968). Thus, contrary to modern synthesis thinking, evolution could occur in the absence of selection and changes needed not to be adaptive. Kimura's neutral evolution acts as a generator of potentially selectable diversity because changes that are selectively neutral when they first appear in a population may be advantageous under different environmental conditions.

Another attack to the synthesis' primacy of natural selection in evolution was delivered by Gould and Lewontin (Gould and Lewontin, 1979). They argued that morphological changes are restricted by historical constraints (phyletic heritage). The organism's basic architecture imposes limitations on the possible derived modifications. These morphological constraints then become more important in establishing the pathways of possible changes than the selective pressure that may facilitate the fixation of such changes in the population. The concept of historical contingencies is also important at the molecular level and will be relevant for the current work. The nucleotide sequence of any gene imposes serious constraints to the encoded protein's attainable adaptive
landscape. Furthermore, the fixation of an allele in the population automatically restricts the starting point for the next generation of mutations in the same gene.

2.3 Phyletic gradualism vs Punctuated equilibria

Phyletic gradualism and punctuated equilibria are two theories concerned with speciation events. Both theories comply with strict Darwinian evolution (overproduction of offspring, variability and differential reproductive success). Phyletic gradualism states that derived species arise through gradual modifications in ancestral forms and that the rate of evolution is kept constant before and after speciation events (Figure 2.1). A consequence of phyletic gradualism is that the fossil record should display an unbreakable and directional series of slightly modified forms ranging from the ancestral to the descendant species. Unfortunately, examples of phyletic gradualism recorded in fossil samples are rare, an observation often attributed to imperfections in the preservation of fossil fauna and flora.

Eldredge and Gould (Eldredge and Gould, 1972) have argued that discontinuities of the fossil record reflect an important aspect of the speciation event and are not mere imperfections of the available data. They have reinterpreted the fossil record to suggest that the data it holds is more in accord with a punctuated pattern of evolution than with a single gradual line of changes from ancestral to derived forms. Under punctuated equilibrium, evolution
Figure 2-1- Gradualism vs Punctuated equilibria. Fossil records for speciation events are often found to be discontinued and usually only reveal ancestral (species 1) and fully derived (species 2) forms. A - phyletic gradualism model where small changes in morphology accumulate steadily over time. The intermediate forms are not represented in the fossil record due to the non-preservation of particular strata (dotted lines=not preserved). B - punctuated equilibrium model where speciation events occur quickly and is preceded and followed by long periods of stasis. Adapted from the University of California Museum of Paleontology web site; Evolution101: (http://evolution.berkeley.edu/evosite/evo101/VIIA1aHypotheses.shtml)
happens fast during speciation events and remains constant during long periods of stasis (Figure 2.1).

Both theories are seen as exclusive. But because of the immensely large time period required for a speciation event to happen, there is little hope that the dispute will ever be settled. Although a healthy argument, it is likely that both modes of speciation were present during the origin of new species.

Nevertheless, Raff and colleagues (2003) have gathered evidence of punctuated evolution during speciation in sea urchins. Species of sea urchins can display a feeding (indirect development; primitive condition) or a non-feeding (direct development; derived condition) larval stage. Cross fertilization experiments between direct and indirect developers from species closely (4 million years apart) and distantly (40 million years apart) related resulted in hybrids displaying a feeding larval stage. These results showed that a primitive (feeding larvae) condition could be restored in the hybrids and were used to argue that the developmental program leading to the primitive condition has been conserved for over 40 million years. On the other hand, changes that resulted in the derived condition (non-feeding larvae) happened relatively fast (less than 4 million years) (Raff et al., 2003). These results were used as an argument in favor of Gould’s punctuated model where long period of stasis are interrupted by sudden changes in morphology (in this case developmental programs).

Phyletic gradualism was championed by the modern synthesis and viewed as the norm for speciation events and, by extension, for any evolutionary process. Neutral evolution and the accumulation of mutations at a constant rate
(molecular clock) were viewed as supportive of phyletic gradualism at the molecular level. However, a punctuated pattern of evolution has been observed in the geographical distribution of vesicular stomatitis virus isolates. Sequencing a single gene from viruses collected at different localities revealed that nucleotide changes were correlated with regional zones and not to the divergence time between isolates. These findings indicated that single nucleotide substitutions in the viral population were not accumulated through a neutral molecular clock, but rather through a punctuated pattern (Nichol et al., 1993).

2.3 Molecular evolution

The modern synthesis advocated that a satisfactory theory of evolution must be quantitative (Haldane, 1924). The study of molecular evolution in microorganisms readily satisfies this prerogative. Microorganisms are well-suited to rigorously test hypothesis in evolution for several practical reasons. They are easy to handle, display short generation times and can sustain very high cell densities for extended periods of time. Most microorganisms employed in evolution experiments also display asexual reproduction and have a developed genetic system. These attributes allow evolutionists to investigate evolution through many generations in short time periods; they also facilitate experimental replication and robustness. Clonality, due to asexual replication, also makes it easier to trace lineage pedigrees through the introduction of genetic markers. Growth conditions can be readily and exquisitely controlled and thus evolutionists
can investigate the impact of single variables (resources, temperature, etc) during evolution experiments. As microorganisms can be viably preserved, difference in fitness between founder and evolved lineages can be investigated by head to head (Malthusian) competition experiments (Elena and Lenski, 2003).

The present work is concerned with the origin and fate of variation at the molecular level in a bacterial population. Therefore it is highly appropriate to briefly review the relevant findings in molecular evolution. Experiments with microorganisms were at the heart of seminal molecular evolution studies of the 1950's through the 1970's. It is quite surprising that evolutionary biologists did not employ these systems again until the 1990's (Bennett et al., 1990). Evolution experiments in microorganisms have, in general, employed two distinct approaches. First, phenotypical adaptation of _E. coli_ populations to certain environmental challenges (Bennett et al., 1990) and second, the use of phage systems to follow the evolution of an entire genome to specific growth conditions (Bull et al., 1997). Results from these studies indicated that the origin and fate of diversity in a population of microorganisms displayed general attributes found for the same event in higher taxa. More importantly, studies in microorganisms confirmed some previously untested aspects of Darwinism and also expanded our knowledge of how populations evolve.

A fundamental concept of evolutionary theory is that mutation must precede and not be directed by selection. It is a common observation that bacterial cultures can often adapt to new growth conditions. Manipulation of the culture media can produce variants with modified metabolic pathways. These
results are often interpreted as evidence that changes in the environment can lead to heritable changes in the cell's metabolism. In the 1930's Lamarckian inheritance was still a possibility for bacterial evolution. Indeed, results from bacterial evolution imposed such an embarrassment to evolutionists that microorganisms were not included in the modern synthesis (Sniegowski and Lenski, 1995).

Bacteria were reintroduced to the evolutionary theory in a series of elegant experiments. Luria and Delbrück (Luria and Delbruck, 1943) demonstrated for the first time that Darwin's concept of “variability precedes selection” was true. Their fluctuation test experiment showed that \textit{E. coli} variants displaying phage resistance pre-existed in cell cultures. Lederberg and Lederberg (Lederberg and Lederberg, 1952) and Cavalli-Sforza and Lederberg (Cavalli-Sforza and Lederberg, 1956) confirmed Luria and Delbrück's observations and demonstrated that the appearance of resistant variants in a cell population was not dependent to previous cell exposure to selective agents. Surprisingly, it is quite common to find reports of 'directed' or 'adaptive' mutations in bacteria (Cairns, 1988; Harris et al., 1994). Nevertheless careful analyses of these studies have identified experimental flaws or poor interpretation of the results as the causes of the observed phenomena (Sniegowski and Lenski, 1995).

Gould has suggested that the relative impact of adaptation, chance and history in evolution could be assessed by “replaying life's tape” (Gould, 1989). If selection is the major force behind the evolution of a trait, then “playing the evolutionary tape” multiple times should yield similar trajectories and final results.
On the other hand, if chance and history are important then the same starting point should lead to different outcomes. Lenski and colleagues have performed Gould’s thought experiment with 12 different *E. coli* populations grown under identical conditions for over 30,000 generations (over 10 years!) (Bennett et al., 1990; Travisano et al., 1995). Their results have shown that history, adaptation and chance contribute differently to the evolution of independent traits. Adaptation and history were more important to the evolution of traits directly correlated with the organism survival (fitness). On the other hand, a trait weakly linked to organism survival, cell size, was more influenced by history and chance (Travisano et al., 1995).

Lenski’s group also reported a punctuated pattern during microbial evolution. They observed that fitness increased rapidly for microorganisms grown under new environmental conditions (Bennett et al., 1990). Temporal analysis of fitness profiles for entire populations demonstrated that such fast adaptations were caused by a logarithmic increase in fitness. Surprisingly, sufficient temporal definition revealed that fitness gains occurred in a step wise manner (Elena et al., 1996), reminiscent of Gould’s punctuated equilibrium. Lenski proposed that the fitness profile observed for bacterial evolution could be explained by clonal interference (Gerrish and Lenski, 1998). New, strong mutations leading to fitness improvements in the population have to first outcompete preexisting variants. The waiting time required for the emergence of more fit variants is responsible for the stasis periods. After the new beneficial mutation reaches sufficiently large
numbers and replaces the original, less fit genotype, the entire population fitness will climb to the next step (Elena et al., 1996).

Findings from Lenski's group greatly inspired the current work. Their groundbreaking experimental design and fresh ideas on molecular evolution will be felt throughout this dissertation. Nevertheless, it is worth mentioning that their results were restricted to the observation of fitness fluctuations. The actual genetic traits behind the observed fitness gains were not determined. Therefore the genetic and biochemical basis of adaptation could not be investigated.

Lenski's findings at the phenotypical level were shown valid at the molecular level by evolution experiments in phage systems. Phages have small genomes that can be readily sequenced in their entirety. Two phage lineages cultivated under identical conditions accumulated point mutations that were correlated with fitness gains. Approximately half of the point mutations were observed in both lineages and were attributed to selective pressure. Chance and history were considered relevant to evolution at the molecular level because the two lineages accumulated different mutations and, more importantly, the order that identical mutations were accumulated varied. (Bull et al., 1997; Wichman et al., 1999). Although the biochemical mechanisms responsible for increase in fitness were not directly investigated, these experiments established that at the molecular level selection, chance and history also play a role in evolution.

The presence of a punctuated pattern in molecular evolution was first proposed by Perutz. His studies on hemoglobin physiology from different species led him to realize that a protein sequence and structure greatly constrains the
adaptive landscape. Hemoglobin can have many of its physiologically relevant attributes altered by few changes in key residues. Perutz employed Kimura’s neutral theory to propose that mutations accumulate at constant rates throughout the hemoglobin molecule and, for the most part, do not change the molecule’s physiological behavior. On the other hand, mutations at key residues have a chance to alter the physiological attributes of the protein in a way that would be beneficial to the host organism (ex.: increased oxygen affinity for high altitude dwelling mammals) (Perutz, 1983). Therefore, analysis of fitness (ex.: oxygen binding) gains over time for the various hemoglobin mutants would show a punctuated pattern. Large fitness increments would occur any time a beneficial mutation happened to hit a key residue.

During real time population genetic experiments, organisms experience controlled conditions that hardly reproduce the ones found in real environments. Therefore, findings based on laboratory experiments have always been met with reservations. Nevertheless, laboratory evolution has been shown to recapitulate what is observed for wild populations. Phage studies showed that nucleotide changes accumulated during laboratory evolution experiments could also be found in wild populations (Wichman et al., 2000). Directed evolution studies designed to investigate the structural basis of antibiotic resistance in bacteria were able to recapitulate previously observed changes for ampicilin resistance in wild isolates. More importantly, this study obtained a new β-lactamase variant that was later isolated from wild populations. This was the first time that such
laboratory evolution methods were shown to predict the emergence of new pathogenic bacteria variants (Orenza et al., 2001).

2.4 Summary

Darwin’s theory of evolution states that species are originated through the random and gradual accumulation of adaptive modifications over large periods of time. The synthetic theory of evolution used population genetics to dismiss all other alternatives to Darwinism. Introduction of quantitative methods allowed researchers to test hypotheses in evolutionary theory and resulted in a greater understanding of the forces behind changes in the genetic composition of a population over time. Major modifications were incorporated to the synthetic theory of evolution. The neutral theory showed that evolution can occur in the absence of selection and that changes need not to be adaptive. Organisms' life histories and chance are as important to evolution as selective forces. The punctuated equilibrium theory reinterpreted the fossil record to demonstrate that not all speciation events need to be gradual. The use of microorganisms generated new insights on how evolution works at the molecular level. These studies have confirmed Darwinism’s tenet that modifications precedes and are not directed by selective pressures. Evolution experiments performed in microbial (asexual) populations conform to the current evolutionary theory. It displays striking resemblance to evolution at higher taxonomic level, like the stepwise increment of fitness in microbial and molecular evolution, reminiscent of
punctuated equilibrium; the accumulation of neutral mutations during bacteriophage evolution; evolution of microbial organisms is subjected to history, chance and selection; and that microbial evolution in the lab recapitulates what is observed for wild isolates.
Chapter 3 - Weak Link Evolution

"The devil is very busy, and no one knows better than he, that nothing is stronger than its weakest part." Kingsley, C.

No organism is stronger than its weakest link. Microorganisms are amenable to genetic manipulation and therefore suitable for the generation of weak links. Following the evolution of a single weak link (gene) in a microbial population is a valid way to probe evolution. In fact, the microbial population studies by Lenski’s and Bull's groups discussed in the previous chapter can be seen as a large scale weak link experiment. In those experiments the weak link was the entire less fit genome of ancestral founders and can be seen as a chain with many weak links. In the current work, the weak link is a single gene. This approach greatly simplifies the interpretation of the results, is amenable to quantification and allows the investigation of the biochemical and structural basis of adaptation.

Organisms are not composed of individual parts that can be evolved separately. They are integrated entities with an evolutionary history that greatly constrains their adaptive landscape (Gould and Lewontin, 1979). It is an important point to make that at the single gene level the founder primary sequence for the weak link represents its previous evolutionary history and imposes constraints to its adaptive landscape. Therefore, the weak link gene is itself an evolutionary entity. Nevertheless, the introduction of the weak link gene
into a large, pre-existing chain further constrains its evolution in a physiologically relevant way. Therefore, it is desirable for the weak link gene to be part of a metabolic network, and not an autonomous entity, like most antibiotic resistance mechanisms.

The weak link evolution strategy has three ingredients: the “weak gene”, the host organism and the selective pressure. The “weak gene” must be essential to the host survival under the experimental conditions. The selective pressure must be amenable to exquisite control. The host organism must be easy to handle and amenable to genetic manipulations. In the present work I have employed *Bacillus subtilis adk* as the weak link, temperature as the selective pressure and the thermophile *Geobacillus stearothermophilus* as the host organism.

Temperature was chosen as selective pressure because it is relatively easy to control and enables the precise determination of the amount of selective pressure applied into the system. Temperature sensitive mutants often display a sharp phenotype associated with protein denaturation and consequent loss of function. Moreover, protein temperature stability is of major concern to biotechnology industries as well to academia.

No other metabolic pathway is more crucial to cell survival and growth than the phosphotransfer network. This metabolic pathway is a most satisfactory large chain into which a weak link gene can be introduced. Enzymes in this pathway must finely orchestrate their functions to ensure proper cellular metabolism. Adenylate kinase is a crucial enzyme in the phosphotransfer
network of prokaryotes. Gene-knockout studies have shown \textit{adk} to be essential for every organism in which it is found as a single copy gene, from \textit{E. coli} to the fission yeast \textit{Schizosaccharomyces pombe} (Cousin and Buttin, 1969; Glembotski et al., 1981; Haase et al., 1989; Konrad, 1993).

\textit{G. stearothermophilus} is a gram-positive, moderate thermophile. Its unusually wide growth temperature range (45 to 73 °C) (Streips and Welker, 1969) and robust genetic system (Wu and Welker, 1989) made it a suitable host for our evolution experiments.

The host organism was genetically modified in order to knock-out its endogenous \textit{adk} and replace it with the mesophilic \textit{B. subtilis adk}. The enzyme from \textit{B. subtilis} has a midpoint denaturation temperature (\textit{T}_m) of 50.7 °C (Glaser et al., 1992) and nicely matches the lowest growing temperature for the host organism, 45 °C. This system was made even more appropriate by the wealth of biochemical and structural information available for adenylate kinase.

### 3.1 The phosphotransfer network as a weak link

All life on Earth utilizes the same energy currency to link energy production and consumption, the chemical energy from phosphoanhydride bonds stored in nucleotides. ATP is the most common form of energy storage in the cell, although other nucleoside triphosphates can also be employed. Proton gradients across membranes are responsible for ATP formation during oxidative

The physical separation between energy production and consumption imposes a challenge to living cells. Energy transfer through simple diffusion would require significant nucleotide concentration gradients (Jacobus, 1985; Meyer et al., 1984). Moreover, accumulation of ATP at production sites or its hydrolysis products, ADP, Pi and H⁺, at consumption sites would lead to kinetic and thermodynamic hindrances of these reactions (Dzeja et al., 2000; Saks et al., 1994; van Deursen et al., 1993). Optimal cellular bioenergetics requires ATP to be produced and delivered at rates corresponding to the ATPase velocity and, at the same time, ATP hydrolysis products to be removed at equivalent speeds. This task is accomplished by a phosphotransfer network (Dzeja and Terzic, 2003).

Enzymes of the phosphotransfer network are at near-equilibrium conditions, ensuring that high energy phosphoryl groups, and not their vessel nucleotides, are transported from energy production to energy consumption sites (Dzeja and Terzic, 2003). Disturbance of the phosphotransfer network has serious consequences for cellular metabolism, ranging from reduced tissue energetics to severe genetic disorders (Bianchi et al., 1999; Black et al., 1979; Costa et al., 2005; Grabowska et al., 2004; Matsuura et al., 1989; Miwa et al., 1976).
3.2 Adenylate kinase (EC 3.4.2.7) in the phosphotransfer network

Adenylate kinase (AK) catalyzes the following reaction:

\[
\text{Mg}^{+2} \text{ATP} + \text{AMP} \leftrightarrow \text{Mg}^{+2} \text{ADP} + \text{ADP}
\]

AK’s unique capability of utilizing both high energy phosphoanhydride (β and γ) bonds found in ATP doubles the energetic potential of this molecule (Dzeja et al., 1998). AK is also the only enzyme capable of regenerating ADP from AMP (Noda, 1973). These two attributes make AK activity crucial for the phosphotransfer network.

In order for AK to perform its role in the phosphotransfer network it must carry the inter-conversion of adenylate species to near equilibrium rapidly. Indeed AKs from various organisms display comparable affinities for AMP, ADP or ATP (Glaser et al., 1992). AK activity maintains adenylate species near equilibrium and, therefore, the individual concentrations of the AMP, ADP and ATP are not appropriate to depict the energetic economy of the cell. Their ratios, given by the adenylate energy charge (EC), should be used instead (Atkinson, 1968):

\[
\text{EC} = ([\text{ATP}] + 0.5 [\text{ADP}] / ([\text{ATP}] + [\text{ADP}] + [\text{AMP}]).
\]

EC has been shown to directly measure the amount of energy stored in adenine nucleotide pools accessible for cellular metabolism. Cell viability and growth are influenced more by small reductions in EC values than large changes in the absolute concentrations of ATP. Furthermore, activities for main regulatory enzymes (AK included) have been shown to respond more readily to EC values.
than to absolute concentrations of adenylate species. AK inactivation has been shown to reduce EC values, leading to lower cellular metabolism and viability (Atkinson, 1968; Chapman et al., 1971; Swedes et al., 1975).

3.3  *B. subtilis* AK

AK isolated from *B. subtilis* belongs to the class of long eubacterial adenylate kinases. As all other AKs, the *B. subtilis* enzyme belongs to the ribonucleotide monophosphate kinase family (Leipe et al., 2003) (Schulz et al., 1986) and has four distinct domains, the P-loop, the AMP-binding, the LID and the CORE (Ferber et al., 1997; Glaser et al., 1992; Munier-Lehmann et al., 1999; Schulz et al., 1986).

The LID domain of long AK forms display high sequence identity and fold into similar anti-parallel β-sheets (Berry and Phillips, 1998; Muller and Schulz, 1992). In the enzymes from *B. subtilis* and *G. stearothermophilus* the LID domain binds a structural zinc atom through four highly conserved residues (Berry and Phillips, 1998; Glaser et al., 1992). The zinc atom does not play a role in substrate catalysis (Bae and Phillips, 2004; Glaser et al., 1992) nor is it the only factor implicated in protein thermostability (Perrier et al., 1994).

Adenylate kinase's three domains play distinct roles during the phosphotransfer reaction. Substrate shielding from surrounding solvent avoids ATP hydrolysis by water (Jenks, 1975; Schulz et al., 1986) and is accomplished by large domain movements of the LID and the AMP-binding domains (Berry et
al., 1994; Diederichs and Schulz, 1990; Schulz, 1991; Schulz, 1992). The CORE domain, on the other hand, stays relatively static during catalysis and serves as a scaffold for the molecular hinges movements performed by the other two domains (Diederichs and Schulz, 1990; Wolf-Watz et al., 2004)(Figure 3.1).

AK catalyzes the transfer of phosphoryl groups through a random Bi-Bi mechanism. In this model, the substrates, AMP and ATP, bind independently of each other (Rhoads and Lowenstein, 1968). The reaction’s rate-limiting step is the LID opening and closing movements (Wolf-Watz et al., 2004). AK can be inhibited by excess of substrate or their analogues (Ito et al., 1980; Rhoads and Lowenstein, 1968; Saint Girons et al., 1987). Diadenosine pentaphosphate (Ap5A) is one of the most powerful AK inhibitors. This compound has two adenosine rings linked by 5 covalently attached phosphates. The adenine rings bind to the AMP and ATP binding sites. The extra phosphate group found for Ap5A optimally mimics the spatial separation between the enzyme’s binding sites (Sinev et al., 1996). As AK is a highly flexible enzyme, Ap5A is routinely used to restrict protein movements and facilitate structural investigations.

3.4. **Geobacillus stearothermophilus**

*G. stearothermophilus* was chosen as the model system because it is amenable to genetic manipulations (Wu and Welker, 1989) and can grow over a wide range of temperatures (45 to 73°C) (Streips and Welker, 1969), allowing the weak link experiment to be performed over a large adaptive landscape.
Figure 3-1 AK is a highly flexible enzyme. The atomic structure for E. coli AK is shown in open (A) and closed (B) conformation. The LID domain closes over the substrate upon binding. Ap5A inhibitor is shown in B as a blue stick model. Open conformation is PDB entry: 2AK3; closed conformation PDB entry: 1AKE. All Figures depicting protein structures were fabricated employing PyMol (DeLano, 2002).
Since the late 1960's considerable efforts have been made to convert wild isolates of *G. stearothermophilus* into amenable genetic systems (Chen et al., 1986; Nakayama et al., 1992; Narumi et al., 1992; Streips and Welker, 1969; Vallier and Welker, 1990; Wu and Welker, 1989). Up to this day, three transformation protocols have been established for three different strains of *G. stearothermophilus* (Imanaka et al., 1982; Narumi et al., 1992; Wu and Welker, 1989). All protocols are strain specific and are not readily interchangeable, probably due to the large divergence time between different strains of this organism (Studholme et al., 1999). Another factor hindering the use of *G. stearothermophilus* as a genetic model for *in vivo* evolution is the lack of sequence information for strains with a developed transformation protocol.

Although *G. stearothermophilus* strain NUB3621-R is not a genetic system as well developed as *E. coli* and *B. subtilis* it is still the best characterized system for a moderate thermophilic bacteria, including a genetic map (Vallier and Welker, 1990) and a transformation protocol (Wu and Welker, 1989).

### 3.5 Summary

Evolution can be studied with a weak link approach. Single gene evolution is amenable to quantification and can probe the biochemical and structural basis for adaptation. Organisms are integrated entities with an evolutionary history. The weak link gene is an evolutionary entity. Nevertheless, it is desirable for the weak link evolution to be restricted by a physiological pathway. The weak link
strategy relies on the careful choice of selective pressure, host organism and "weak gene". Temperature is a good selective pressure because it can be easily controlled and quantifiable. The phosphotransfer network forms a large chain; its participating proteins are candidates for the weak link. \textit{adk} is a suitable weak link because it is an essential gene due to its pivotal role in the phosphotransfer network. \textit{G. stearothermophilus} is an appropriate host organism because of its wide temperature range and robust genetic system.
Chapter 4 - Mutation

The synthetic theory of evolution identified mutations as the source of genetic variability. In the current study, evolution of the weak link will rely exclusively on the spontaneous production and accumulation of point mutations. Point mutations are mainly caused by the modification of single DNA bases due to chemical instability or enzymatic activity. As most mutations are deleterious, cells have evolved very efficient DNA repair mechanisms.

4.1 Patterns of mutations

The direction of mutations in living organisms is clearly not random. Table 4.1 (Li, 1997) shows the observed patterns of nucleotide change in pseudogenes. As pseudogenes are no longer under selective constraints they are expected to provide a reliable estimate for the rate of spontaneous mutations. Some trends are easily identified from Table 4.1. First, transitions are more frequent than transversions. Second, C and G are the most mutable residues. The chemical properties of the DNA molecule are, in part, behind the skewed pattern of nucleotide substitution. DNA can be spontaneously modified by hydrolysis, oxidative damage or methylation. Moreover, enzymatic modification of
Table 4-1 - Pattern of nucleotide substitution in pseudogenes\textsuperscript{a}

<table>
<thead>
<tr>
<th>from</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>4.7</td>
<td>5.0</td>
<td>9.4</td>
<td>19.1</td>
</tr>
<tr>
<td>T</td>
<td>4.4</td>
<td>-</td>
<td>8.2</td>
<td>3.3</td>
<td>15.9</td>
</tr>
<tr>
<td>C</td>
<td>6.5</td>
<td>21</td>
<td>-</td>
<td>4.2</td>
<td>31.7</td>
</tr>
<tr>
<td>G</td>
<td>20.7</td>
<td>7.2</td>
<td>5.3</td>
<td>-</td>
<td>33.2</td>
</tr>
</tbody>
</table>

Column total | 31.6 | 32.9 | 18.5 | 16.9 |

\textsuperscript{a}values shown are percentages from 13 mammalian pseudogene sequences (Li, 1997)
DNA is also employed as a means to control gene expression and recognition in living systems (Lindahl, 1993; Lindahl and Wood, 1999).

DNA lesions are unavoidable and, in most cases, deleterious to the cell. Therefore, living organisms have developed very efficient DNA repair mechanisms. For example, deamination of cytosine and 5-methylcytosines (5-meC) is pronouncedly increased at high temperatures and is a threat to the genome integrity of thermophiles. In mesophiles, 5-meC is a hot spot for mutations because its deaminated form is less efficiently repaired (Lindahl, 1993). Thermophilic organisms display a more effective strategy to repair such lesions. In addition to their equivalent to the mesophilic uracyl-DNA glycosylases (UDG), thermophiles express a unique type of glycosylase that is capable of recognizing U:G and T:G mismatches and therefore countering the mutagenic action of deaminated cytosines and 5-meC (Sandigursky and Franklin, 1999; Yang et al., 2000).

The degeneracy of the genetic code also plays an important role in the outcome of point mutations. Changes in the second and third codon positions are more likely to result in conservative changes. On the other hand, changes in the first position of the codon are almost certain to produce missense mutations. Further restrictions to the adaptive landscape of the weak link through the accumulation of point mutations come from the limited number of changes that can be achieved by single nucleotide substitution. For example, Gln199 is coded by a CAG in *B. subtilis adk*. Through a single nucleotide change, this codon could only mutate to encode a Lys, Glu, Leu, Arg, Pro or His residue. If a
stabilizing mutation required the change in this position to any different residue, say a valine, no single mutation event would be able to satisfy this condition.

### 4.2 The power of mutations and selection

As a result of the very efficient DNA repair machinery of the cell, the rate of mutation ($\mu$) is extremely low ($5 \times 10^{-10}$ mutations per bp per replication for *E. coli* - (Drake, 1991)). At this point it would be very appropriate to perform a thought experiment to verify the power of selection and mutation in the weak link approach and to introduce some of the mathematical aspects of evolutionary theory.

First we shall impose some restrictions to the evolution of the weak link. *B. subtilis adk* is 651 bp. Considering that *B. subtilis* ($T_m = 50.7^\circ$C) and *G. stearothermophilus* ($T_m = 74.7^\circ$C) are 74% identical let's assume that only mutations that happen in non-conserved positions between the two enzymes will result in thermostabilization of the mesophilic enzyme. Another restriction is that only mutations that change the first codon positions will be considered, as they are the ones most likely to produce a change in amino acid. Using the $\mu$ value for *E. coli*, the rate of thermostabilizing mutations in the weak link gene ($\mu_{adk}$) will be $3.0 \times 10^{-8}$ mutations per replication ($5 \times 10^{-10}$ mutations per bp per replication $\times 0.26 \times 651 \times 1/3$).

In the absence of selection these beneficial mutations will be considered neutral. The fixation of a neutral allele in the population only depends on its initial
frequency (Kimura, 1968). Nevertheless, working in a large population regime ensures that very low probability events, such as $\mu_{adk}$, are actually possible. It is therefore essential for the success of the weak link approach that the population number (N) is somewhat larger than the rate of "beneficial" mutations. For a population with $10^6$ cells per ml growing in a 500 ml chemostat the number of cells carrying a “beneficial” mutation in the adk gene after one round of replication will be $N \times \mu_{adk} = 1.5 \times 10^4$.

The genetic content of the population in the fermentor can be described as the frequencies of wild-type (p) and mutant adks (q). At any time $p + q = 1$. Changes in the fermentor population as a function of time and mutational rate $\mu_{adk}$ can be described by (Lenski, 1991):

$$p_n = p_0 \times (1 - \mu_{adk})^n \text{ (equation 4.1)}$$

where $p_n$ is the frequency of wild-type genes at the $n^{th}$ generation; $p_0$ is the initial wild-type frequency and $\mu_{adk}$ is the rate of “beneficial” mutations accumulated in every generation.

In the absence of selection (i.e.: mutants are selectively neutral) it would take $\sim 9 \times 10^6$ generations for the wild-type gene to be reduced to 50% of the population and $\sim 3$ times longer for it to be reduced to 1% (Figure 4.1). If we consider an average time of 30 minutes for every generation it would take 513 years for the wild-type gene to be reduced to 50% of the population!
Figure 4-1- Selection greatly increases the rate of evolution. Curves depict the fate of the wild-type (blue) and mutant (red) genotypes in a population. Frequencies (f) for wild-type (p) and mutant (q=1-p) are dictated by the equations shown in A and B, respectively. $\mu_{adk}$ is the mutational rate for the adk gene. In A there is no selection and it would take approximately 500 years for both genotypes to be equally represented in the population assuming a generation time of 30 minutes. In B the mutant ($q_0 = \mu_{adk}$) is slightly more fit than the wild-type ($W=0.92$; $s=0.8$) and the waiting time to reach equal amounts of both genotypes would be reduced to approximately 4 days.
During the evolution of *B. subtilis* adk the spontaneously produced mutants are expected to competitively displace the wild-type cells. The relative fitness (*W*) for a particular mutant can be calculated from the ratio of its growth rate relative to that of the wild-type strain. If two strains, mutant and wild-type, are grown together for one day at the end of that period the Malthusian (*M*) parameter for the mutant strain will be (Lenski, 1991):

\[
M_{\text{mut}} = \ln(\text{mut}_t / \text{mut}_o) \quad (\text{equation 4.2})
\]

where \( \text{mut}_t \) is the number of mutant cells at time \( t \) and \( \text{mut}_o \) is the number of cells at the beginning of the experiment. The Malthusian parameter can be found for the wild-type cells in the same manner. The relative fitness for wild-type and mutant cells at the end of the first day will be (Lenski, 1991):

\[
W = \frac{M_{\text{wt}}}{M_{\text{mut}}} \quad (\text{equation 4.3})
\]

If both strains are equally fit at the experimental conditions \( W \) will be equal to 1. If mutant cells are more fit than wild-type than \( W \) will be < 1. Conversely, if wild-type cells are more fit than mutant cells than \( W > 1 \).

If the growth rate of wild-type cells differs from the one found for mutant cells, the frequency of wild-type cells over time will be given by (Lenski, 1991):

\[
p_n = \frac{p_o}{(1 - p_o)e^{\lambda n} + p_o} \quad (\text{equation 4.4})
\]
where \( p_n \) is the frequency of wild-type cells after \( n \) generations, \( p_0 \) is the initial frequency for wild-type cells and \( s \) is the selection coefficient defined as \( s=1-W \) (Lenski, 1991).

Selective pressure greatly increases the speed of evolution. Originally neutral mutations in \( B. \ subtilis \ adk \) are expected to increase their fitness values as selective pressure rises in the turbidostat. If the growth rate of wild-type cells is 90% of that found for the mutants (\( W=0.92; s=0.08 \)), the former will make 50% of the population after \( \sim 200 \) generations. After 60 more generations the mutants would make 99% of the population in the chemostat (Figure 8.2b). Considering the same doubling time as above (30 minutes), mutant cells would reach the 50% point after only 4 days.

These very simplified and reductionist models for the power of mutation and selection show that the weak link approach can work with a highly constrained rate of spontaneous "beneficial" mutations. In the real experiment, temperature increments will lead to the eventual occurrence of non-permissive growth conditions for the founder organism, and the consequent reduction of its fitness to 0.

### 4.4 Summary

Spontaneously occurring mutations are the fuel of evolution. The mutational pattern is not random, but dictated by the DNA chemical attributes and the cellular repair and replication machineries. Because of the low frequency
of spontaneous mutations, evolution of the weak link will be greatly facilitated by
the use of a large population size. Mutations in \textit{B. subtilis adk} are expected to be
neutral when they first appear. The fixation of a neutral mutation in the population
is only dependent on its initial frequency. However, in the absence of selection
changes in the gene frequency of a population would take a long time. Fitness is
defined as the adaptive worth of an organism. In the presence of selection,
different genotypes are expected to display different fitness. Relative fitness
values can be calculated from the growth rates of two cell lineages in a
Malthusian competition experiment. The presence of purifying selection greatly
increases the speed of evolution by removing less fit genotypes from the
population.
Chapter 5 - Protein folding and stability

"A dead frog placed in saturated urea solution becomes translucent and falls to pieces in a few hours." Ramsden, W. (1902)

Proteins are physicochemical entities and as such must obey the laws of thermodynamics. Protein stability is usually reported as the Gibbs free energy of stabilization $\Delta G_{\text{stab}}$, the difference between the free energies of the native ($\Delta G_{\text{fold}}$) and the unfolded ($\Delta G_{\text{unfold}}$) states for a particular protein:

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

$\Delta G_{\text{stab}}$ measures the difference between the enthalpic ($\Delta H_{\text{stab}}$) and entropic ($\Delta S_{\text{stab}}$) contributions to protein stability. The value of $\Delta G_{\text{stab}}$ for most proteins was found to fall in the range of 5 and 15 kcal/mol at 25°C; a very small number compared to the large values found for the enthalpic and entropic terms (Vieille and Zeikus, 2001).

The enthalpic and entropic terms for protein stability are fully described in its primary sequence. Single amino acid modifications can have dramatic effects on protein stability because their individual contributions to the protein’s enthalpic or entropic terms are usually larger than $\Delta G_{\text{stab}}$ values.

The protein stability curve is a convenient way to represent the $\Delta G$ dependence on temperature (Figure 5.1). In this curve, the folded state of the protein is favored at regions with $\Delta G_{\text{stab}} > 0$ and maximal at $T^*$. The two points where $\Delta G_{\text{unfold}} = 0$ represent the heat ($T_m$) and cold denaturation temperatures. It
Figure 5-1- The theoretical temperature stability curve for $\Delta G_{\text{stab}}$ vs $T$. The curve for a hypothetical mesophilic protein is shown (M). Curves A, B and C depict the possible strategies that can be found when stabilizing protein M into a thermophilic protein. A - the thermophilic protein is stabilized by a shift towards higher $T^*$. B - the thermophilic protein displays a flatter stability curve due to differences in the heat capacity of folded and unfolded states. C - the thermophilic protein displays a higher $\Delta G_{\text{stab}}$ value at the same $T^*$. Adapted from (Vieille and Zeikus, 2001).
follows from the protein stability curve that there are three ways for proteins to increase their stability. Shifting the curve to the right towards a higher $T^*$ (curve A in Figure 5.1); broadening the curve by changes in the heat capacity differences between the folded and unfolded states (an entropic term; curve B in Figure 5.1) and raising the curve by increasing the magnitude of $\Delta G_{\text{stab}}$ (an enthalpic term; curve C in Figure 5.1) (Becktel and Schellman, 1987). The study of thermostable proteins revealed that their increased thermostability is achieved by the combination of all of the above-mentioned strategies (Vieille and Zeikus, 2001).

Protein stability is thus a struggle between two opposing forces. The entropic cost of reducing the protein's virtually infinite conformations at the unfolded state to a set of defined native ensembles must be overcome by favorable enthalpic factors, such as the hydrophobic effect or electrostatic interactions (Bromberg and Dill, 1994; Chothia, 1975; Dill, 1990).

5.1 Hydrophobic interactions and conformational entropy

Theoretical and empirical studies have identified hydrophobicity as the main force in protein folding and stability. Almost half of the amino acids have aliphatic side-chains. The burial of these non-polar residues inside the protein core and away from the solvent increases protein stability and is also thought to drive the spontaneous folding of a protein and the formation of its secondary structures (Dill, 1985; Rackovsky and Scheraga, 1977).
Sequence and structural information from thermophilic proteins have confirmed the importance of the hydrophobic effect to protein thermostability. Thermophilic proteins favored amino acid substitutions that led to a tighter hydrophobic packing of the protein interior; mainly through the replacement of uncharged polar residues with hydrophobic ones (Haney et al., 1999). In addition, single amino acid substitutions (host-guest experiments) performed with model proteins have confirmed the importance of hydrophobic interactions and also shown that mutations that fill in solvent cavities also lead to higher protein thermostability (Matthews, 1995; Milla and Sauer, 1994).

Tighter packing of the protein core leads to a reduced flexibility in thermophilic proteins. Solvent exchange experiments have demonstrated that thermophilic proteins display lower flexibility at room temperature than their mesophilic counter-parts. The decrease flexibility is thought to be responsible for the reduced catalytic activity displayed by these proteins at temperatures removed from their physiological range (Zavodszky et al., 1998). Host-guest experiments have also confirmed that decrease flexibility leads to higher protein stability. Substitutions that decrease the conformational entropy of residues are correlated to increased protein stability (Matthews et al., 1987).

5.2 Electrostatic interactions

Electrostatic interactions are yet another stabilizing factor to protein structures. Their contributions to protein stability are not as well accepted in the
literature as the hydrophobic effect because their numbers in the folded and unfolded states are expected to be the same (Strop and Mayo, 2000). Nevertheless, many examples can be found where electrostatic interactions greatly contributed to protein thermostability; mainly through the stabilization of secondary structures (Bae and Phillips, 2004; Vogt et al., 1997). Moreover, thermophilic proteins also display a larger number of electrostatic interactions than their mesophilic counterparts, which also supports their importance to protein stability (Criswell et al., 2003; Vogt et al., 1997).

Electrostatic interactions arise due to the non-covalent attraction of opposite charges. They can be of two types: hydrogen bonds and salt bridges. The interacting energy of two opposite charges is given by Coulomb's law:

$$ \Delta E = \frac{Z_A Z_B \varepsilon^2}{D r_{AB}} \quad \text{(equation 5.1)} $$

where $Z_A$ and $Z_B$ are the charges for atoms A and B, $\varepsilon$ is the electron's charge, D is the dielectric constant for the media (80 for water) and $r_{AB}$ is the distance between atoms A and B.

Hydrogen bonds occur when two partially charged residues compete for the same hydrogen atom. Hydrogen bonds have a distance cut-off of 3.0 Å between the H-donor and the H-acceptor and should display a donor-hydrogen-acceptor angle bellow 90°. Formation of hydrogen bonds has been shown to be important for protein stability (Tanner et al., 1996) and, in at least one case, to be as important as hydrophobic interactions (Shirley et al., 1992).
Salt bridges occur between two fully charged residues of opposite characters. Although they do not have a cut-off distance, their interacting energy decays quickly with increasing distance. The strength of a salt bridge between two fully charged amino acids is determined by the number of opposite charges in close proximity. Therefore the stabilizing energy of a salt bridge is also somewhat restrained by geometry (Kumar and Nussinov, 1999; Kumar and Nussinov, 2002). A distance cut-off of 5.0 Å is usually employed when identifying salt bridges in protein structures.

Because of the large energy penalty that results from burying charged residues in the hydrophobic interior of a protein, salt bridges can be destabilizing and their importance for protein stability are commonly overlooked. Nevertheless, recent studies have shown that salt bridges with good geometries are likely to be stabilizing, regardless of their location in the protein structure (Kumar and Nussinov, 1999; Kumar and Nussinov, 2002). Moreover, thermophilic proteins also display a larger percentage of charged residues, often involved in the formation of ion pair networks (Yip et al., 1998).

5.3 B. subtilis AK stability

Comparison of the AK atomic structures obtained for the AK enzymes isolated from the mesophile B. subtilis, the psycrophile B. globisporus and the thermophile G. stearothermophilus suggested that AK thermostability can be accomplished by the use of different strategies. The thermophilic AK displayed a
reduced solvent exposed surface and an increased number of ion pairs and hydrogen bonds when compared to its mesophilic and psychrophilic counterparts. (Bae and Phillips, 2004). It is important to mention that all these stabilizing modifications originated from mutations at the protein’s primary sequence (Figure 9.1). As these three proteins display a high degree of homology, variants sites may represent “hot spots” for AK thermostability.

5.4 Summary

Proteins are subject to thermodynamic constraints. Proteins are only marginally stable at room temperature; their stability is given by the difference between entropic and enthalpic terms. Single amino acid substitutions may result in significant changes in protein stability. Hydrophobic interactions are thought to be the main contributor to protein stability and a major driving force during folding. Tighter packing of the hydrophobic core increases protein rigidity reduces entropic costs and is correlated to protein stability. Ionic interactions, hydrogen bonds and salt bridges, have been correlated with protein stability. AK structures from mesophilic and thermophilic enzymes revealed that higher protein stability can be accomplished by the use of different strategies.
Chapter 6 - Methods

6.1 Bacterial cultures and Media

6.1.1 Escherichia coli

Throughout this work two strains of *E. coli* were employed. Strain XL-1 Blue (New England BioLabs) was used for plasmid transformation and plasmid DNA isolation. Strain BL21DE3 (Promega) was utilized for protein expression. Chemically competent *E. coli* cells were transformed by the heat shock method (Hanahan et al., 1991). Relevant genotypes for utilized strains can be found on Table 6.1.

Liquid *E. coli* cultures were grown in LB medium (1 % tryptone, 1 % NaCl and 0.5 % yeast extract) under aeration at 37 °C for the specified times. Solid LB medium was obtained by adding agar to liquid LB medium (15 g/l) prior to sterilization. Plates were kept at 37 °C overnight in an incubator. All media was sterilized with an autoclave and supplemented with ampicillin (0.1 g/l) or carbenicillin (0.1 g / l) prior to use. *E. coli* cultures containing plasmids of interest were stored at -80 °C in 20 % glycerol (volume by volume).
<table>
<thead>
<tr>
<th>Strain / plasmid</th>
<th>Relevant genotype / phenotype&lt;sup&gt;a&lt;/sup&gt;</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. stearothermophilus NUB3621-R</td>
<td>Rif&lt;sup&gt;r&lt;/sup&gt;</td>
<td>Dr. N. Welker</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Northwestern U., Evanston, IL)</td>
</tr>
<tr>
<td>G. stearothermophilus NUB3621-R:Thev</td>
<td>Rif&lt;sup&gt;r&lt;/sup&gt; Cml&lt;sup&gt;r&lt;/sup&gt; ts</td>
<td>This work</td>
</tr>
<tr>
<td>E. coli XL1-Blue</td>
<td>lac&lt;sup&gt;1&lt;/sup&gt;Z&lt;sup&gt;Δ&lt;/sup&gt;M15</td>
<td>New England BioLabs</td>
</tr>
<tr>
<td>E. coli BI21(De3)</td>
<td>F&lt;sup&gt;−&lt;/sup&gt;omp&lt;sup&gt;T&lt;/sup&gt; hsdS (r&lt;sub&gt;6&lt;/sub&gt;m&lt;sub&gt;6&lt;/sub&gt;) gal dcm</td>
<td></td>
</tr>
<tr>
<td>pUC19</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt;</td>
<td>New England BioLabs</td>
</tr>
<tr>
<td>pPW15(7)</td>
<td>Cml&lt;sup&gt;r&lt;/sup&gt;</td>
<td>Wei and Stewart, 1993</td>
</tr>
<tr>
<td>pSTE12</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt; Tet&lt;sup&gt;r&lt;/sup&gt;</td>
<td>Riken (Japan)</td>
</tr>
<tr>
<td>pETadk&lt;sub&gt;Bsb&lt;/sub&gt;</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt; adk&lt;sub&gt;Bsb&lt;/sub&gt;</td>
<td>Dr. G.N. Phillipps, Jr. (U. Wisconsin, WI)</td>
</tr>
<tr>
<td>pThEV</td>
<td>Ap&lt;sup&gt;r&lt;/sup&gt; Tet&lt;sup&gt;r&lt;/sup&gt; adk&lt;sub&gt;Bsb&lt;/sub&gt; Cml&lt;sup&gt;r&lt;/sup&gt;</td>
<td>this work</td>
</tr>
</tbody>
</table>

<sup>a</sup> Abbreviations: Rif<sup>r</sup>, rifampin resistant; Cml<sup>r</sup>, chloramphenicol resistant; Ap<sup>r</sup>, ampicillin resistant; Tet<sup>r</sup>, tetracycline resistant; adk<sub>Bsb</sub>, <i>B. subtilis</i> adenylate kinase.
6.1.2 *Geobacillus stearothermophilus*

*G. stearothermophilus* NUB3621-R (kind gift of Dr. N. Welker, Northwestern University, Evanston, IL) cultures were grown in modified LB medium (mLB - LB medium supplemented with 0.6 mM MgSO4.7H2O, 0.9 mM CaCl2.2H2O, 0.04 mM FeSO4.7H2O and 1.1 mM Nitrilotriacetic acid) (Chen et al., 1986). Alternatively, *G. stearothermophilus* NUB3621-R cells were cultivated in TS medium (4% tryptone, 0.5% NaCl, supplemented with 0.6 mM MgSO4.7H2O, 0.9 mM CaCl2.2H2O, 0.04 mM FeSO4.7H2O and 1.1 mM Nitrilotriacetic acid) (personal communication, Dr. N. Welker, Northwestern U., Evanston, IL). Solid media was prepared by the addition of 15 g / l of agar (Sigma). Tryptone and yeast extracts were purchased from Difco. Liquid cultures were cultivated at the indicated temperatures in a water shaker or in a New Brunswick tabletop fermentor as indicated in the text. Solid cultures were kept at the indicated temperatures in a Fisherbrand Scientific Isotemp Forced Air Incubator 600 (± 0.1 °C).

All media was sterilized in autoclave and supplemented with the appropriate antibiotics (chloramphenicol - 0.07 g / l; rifampicin 0.05 g / l or tetracycline 0.05 g/l) prior to use. *G. stearothermophilus* cultures were stored at -80 °C in 10 % glycerol (volume by volume).

Transformation of *G. stearothermophilus* NUB3621-R protoplasts with plasmid DNA was performed according to the protocol developed by Wu and Welker (1989) (Wu and Welker, 1989).
6.2 Isolation and manipulation of Plasmid and Genomic DNA

All plasmid DNA preparations were performed using the Wizard plasmid purification kit (PROMEGA). *G. stearothermophilus* NUB3621-R genomic DNA was isolated using the Genomic-tip 20 / G kit (QIAGEN). Restriction enzymes employed in this work were purchased from Promega and New England BioLabs and utilized according to the manufacturer’s instructions. DNA ligation reactions were performed overnight at 20°C using T4 DNA Ligase (Promega). Electrophoretic separation of nucleic acids was performed in a BioRad system. Agarose gels were prepared in 1 x TAE buffer. Electrophoresis was performed in 1 x TAE at 100 volts. Gels were stained with Ethidium Bromide, DNA fragments and plasmids were visualized with the aid of UV light.

6.3 Polymerase chain reactions

All PCR amplifications were performed on a Primus PCR system (MWG Biotech). The Expand Long Template PCR System (Roche Applied Science) was routinely used for PCRs performed on genomic DNA. The reaction mixture containing genomic DNA (100 ng), primers (0.3 μM), dNTPs (25 mM) and enough water to 40 μl was incubated at 94 ºC for 5 minutes. To the reaction were added 5 μl of buffer 3 and 0.75 units of polymerase mixture from the Expand Long Template PCR System together with enough water to 50 μl. Reaction
conditions varied in accordance with the oligonucleotide primers utilized (Table 6.2). Elongation times were 1 minute per kb of expected product.

PCR amplifications performed on plasmid DNA employed enzyme, buffer and magnesium solutions from the MasterTaq PCR system (Eppendorf). The reaction mixture contained plasmid DNA (10 ng), oligonucleotide primers (0.3 \( \mu \text{M} \)), dNTPs (25 mM), magnesium acetate (3.0 mM), reaction buffer and 0.75 units of Taq polymerase.

We employed a PCR cleanup kit (Qiagen) to directly purify PCR products from the reaction mixture. PCR products containing more than one product or DNA fragments obtained by restriction enzyme digestions were separated in 0.8 % agarose gels prior to purification. The DNA fragment of interest was extracted from the gel using the gel cleanup kit (Qiagen).

6.4 Construction of \textit{G. stearothermophilus} strain NUB3621-R:ThEV

All plasmids utilized for the construction of the homologous recombination vector can be found on Table 6.1. Degenerate oligonucleotide primers were designed based on sequence information for \textit{G. stearothermophilus} strain S10 (\textit{Bacillus}) and \textit{B. subtilis} (Li et al., 1997). All oligonucleotide primers employed in this section can be found on Table 6.2, together with the relevant restriction enzyme sites they introduced in the amplicon and their PCR parameters.
<table>
<thead>
<tr>
<th>Oligonucleotide primers&lt;sup&gt;a&lt;/sup&gt;</th>
<th>sequence (5′ to 3′)</th>
<th>Restriction site</th>
<th>Annealing temperature (°C)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub_adk_seq_f</td>
<td>taggagcaacgcctcttgcc</td>
<td>none</td>
<td>56.0</td>
</tr>
<tr>
<td>sub_adk_seq_r</td>
<td>ccataaacctcaactcccttgcc</td>
<td>none</td>
<td>56.0</td>
</tr>
<tr>
<td>adk_f</td>
<td>ccgccgatcatgaactaggctctgtttgctagtgaatggagbagttgctaatgg</td>
<td>none</td>
<td>56.0</td>
</tr>
</tbody>
</table>
| adk_r                            | gtcctctgtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
6.4.1 Construction of a homologous recombination vector

Construction of the homologous recombination vector required first knowledge of the DNA sequence surrounding the wild-type NUB3621-R adk gene. Genomic DNA isolated from NUB3621-R cells was used as template for PCRs performed with degenerate primers for the ribosomal genes rplO (primer L15_deg_F) and rpmJ (primer L36_deg_R). The PCR product was sequenced by Lone Star Labs (Houston, TX – USA) using degenerate primers for secY (SecYDEgF) and map (MAPDEgR). The sequence information (genebank # AY729037) was used to design primers to amplify the last 1.0 kb from secY (NUB_LSecY_F_Sacl and Rec_SecY_R) and the first 1.0 kb from map (Rec_Map_F and NUB_LMap_R_SphI) genes from wild-type NUB3621-R genomic DNA.

Wild-type B. subtilis adk gene was amplified with primers adk_F and adk_R from plasmid pETADK (kind gift of Dr. G.N. Phillipps, Jr.; University of Wisconsin, WI, USA). Primers dCATf and RC6 were employed to amplify a promoterless copy of the cat gene, utilized as a genetic marker, from plasmid pPW15(7) (Wei and Stewart, 1993).

The PCR fragments generated for NUB3621-R wild-type secY and map genes were digested with the appropriate restriction enzymes (Table 6.2) and ligated into the pUC19 plasmid (Yanisch-Perron et al., 1985). The resulting construct and PCR amplicons for wild-type B. subtilis adk and the promoterless cat gene were treated with the proper restriction enzymes and ligated using T4
DNA Ligase (Promega), yielding the p19ThEV construct. The homologous recombination cassette, composed of NUB3621-R's 1.0 kb from secY, *B. subtilis* adk, promoterless cat and 1.0 kb from NUB3621-R map, was excised from p19ThEV using the restriction enzymes SaeI and SphI and ligated into the *E. coli* / *G. stearothermophilus* shuttle vector pSTE12 (Narumi et al., 1992) treated with the same enzymes. The final construct was dubbed pThEV and employed as a homologous recombination vector in the subsequent step.

**6.4.2 Transformation of *G. stearothermophilus* NUB3621-R with plasmid DNA**

*G. stearothermophilus* NUB3621-R cells were transformed with plasmid pThEV using Dr. Welker's protoplast protocol developed for NUB3621 cells (Wu and Welker, 1989). NUB3621-R cells were kept at 50 °C throughout the entire process. All media was pre-warmed at 50 °C prior to use. All centrifugation steps were performed at room temperature. mP medium is mL8 complemented with 10 % lactose (weight / volume). Solid R medium consisted of mP medium complemented with MgCl₂ and CaCl₂ and agar (Difco).

A NUB3621-R glycerol stock was used to inoculate solid TS medium complemented with rifampicin. Cells were kept at 50 °C for 16 hour and the resulting colonies used to inoculate 30 ml of liquid TS media (without antibiotics) in a 500 ml Erlenmeyer, yielding an OD₆₀₀ of ~ 0.01. The liquid culture was cultivated at 60 °C in a water shaker at 200 rpm until the OD₆₀₀ reached 1.0.
Cells from a 14 ml aliquot were harvested by centrifugation at room temperature (5,000 rpm, 5 minutes). Pelleted cells were brought back into solution with 1.4 ml of mP medium. Lysozyme (Sigma) was added to a final concentration of 0.01 mg / ml. Cells were incubated for 2 minutes in a water shaker (50 °C and 150 rpm). 4.0 ml of mP medium was added to the lysozyme treated cells. Protoplasts were harvested by centrifugation (3,000 rpm for 5 minutes) and gently brought back into solution with 1.4 ml of mP medium. NUB3621-R protoplasts, in 100 μl aliquots, were stocked at -80 °C.

Prior to transformation with plasmid DNA, protoplasts were thawed at room temperature. Up to 1 μg of plasmid DNA was added to a 100 μl protoplast aliquot. The resulting solution was then transferred to a 15 ml plastic tube containing 0.9 ml of sterile 40 % 1500 PEG in mP medium. The solution was gently homogenized and incubated for 2 minutes at 50 °C, 130 rpm. The PEG:protoplast mixture was diluted with 2.5 ml of mP medium. Protoplasts were recovered by centrifugation (5,000 rpm for 5 minutes), excess PEG was removed and cells brought back to solution with 100 μl of mP medium. Protoplasts were incubated for one hour at 50 °C at 130 rpm, diluted and spread on solid R medium complemented with tetracycline (5 mg / l). Cells were incubated for 24 to 48 hour at 50 °C.

Single colonies able to regenerate their cell wall were transferred to fresh selective solid R medium and grown overnight at 50 °C. Cells were then transferred to solid TS medium containing tetracycline (5 mg / l) and cultivated at 50 °C for one night. Liquid cultures were started from single colonies and were
cultivated at the indicated temperatures. PCR and restriction enzyme analyses were employed to confirm the identity of plasmid pThEV isolated from transformed protoplasts.

6.4.3 Isolation of recombinant NUB3621-R:ThEV cells

To screen for recombinant cells and to cure them for the plasmid pThEV, transformants were continuously grown in liquid TS cultures without antibiotics at 50 °C. Cell splits into fresh liquid TS media were performed twice every 24 hour. Cell culture samples were diluted and spread onto solid TS media containing chloramphenicol (7 mg / l). Plates were incubated 16 hour at 55 °C. Colonies growing on chloramphenicol were replica-plated to TS plates containing tetracycline (5 mg / l) and grown for 16 hour at 55 °C.

Screens for temperature-sensitive cells were performed with cultures grown at 55 °C on solid TS medium in the presence of chloramphenicol for 16 hour. Resulting colonies were replica-plated onto fresh solid TS medium plates complemented with chloramphenicol. One set of replicas was incubated at 55 °C and the other cultivated at 65 °C for 24 hour.

6.4.4 Characterization of NUB3621-R:ThEV cells

The correct insertion of the homologous recombination vector and the loss of plasmid pThEV were verified by PCR. Total DNA was utilized as template in
PCR with oligonucleotide primers WAC and RC6. WAC is complementary to a region in secY upstream from the homologous recombination cassette, while primer RC6 anneals at the 3’ end of the cat gene, present in the homologous recombination cassette. The presence of plasmid pThEV in cells was investigated through PCR using primers Tet_ATG_F_EcoRI and Tet_STOP_R_EcoRI. These oligonucleotide primers are complementary to the tet gene, found within plasmid pThEV.

The genomic organization of recombinant cells was probed by Southern blots. DNA probes for Southern blots for B. subtilis adk, pThEV tet and cat genes and G. stearothermophilus map and secY genes were obtained using the PCR DIG Probe synthesis kit (Roche Applied Science). DNA templates and oligonucleotide probes used in generating the southern blot probes can be found on Table 6.2.

Approximately 5.0 µg of total DNA isolated from NUB3621-R and NUB3621-R:ThEV cells were treated with 10 units of EcoRI and HindIII at 37 °C overnight. 2 ng of pThEV were similarly treated. The DNA samples were separated on a 0.6% agarose gel and transferred to a Byodine B membrane (Pall Life sciences), according to the manufacturer’s instructions. The same nylon membrane was used for hybridization with different probes. Hybridization and stripping procedures followed the membrane manufacturer’s protocol. Detection of DNA fragments was performed with the DIG Nucleic Acid Detection Kit (Roche Applied Science), according to the manufacturer’s instructions.
6.5 Growth of NUB3621-R:ThEV in a turbidostat

Evolution of *B. subtilis adk* in NUB3621-R:ThEV cells was performed in a modified New Brunswick Scientific BF3000 Benchtop fermentor system (Figure 6.1). The fermentor's original set up was modified by adding a Mettler Toledo turbidity monitoring system. The turbidity probe was used to follow the cell culture density and control the amount of fresh media input and old media outtake. This set up allowed us to run the fermentor as a turbidostat. The turbidity probe was calibrated using plain LB media as 0 % and an overnight *E. coli* culture diluted to an OD$_{600}$ of 2.0 as 100 %.

The vessel, tubing, probes and other parts in contact with the cell culture were sterilized by UV-light radiation for one night in a hood. All parts were autoclaved prior to use. The assembled fermentor was autoclaved with 1.0 l of water prior to the addition of sterile liquid media.

Fresh culture media was kept in autoclavable 20 l carboys (Nalgene). Tryptone (100 g), yeast extract (50 g) and NaCl (100 g) were weighed and transferred to a 20 L carboy. 10 liters of MilliQ water was added to the vessel together with 1.0 ml of a 1 % anti-foam solution. The media was autoclaved by a 1 hour cycle at 121 °C. Rifamycin (5 mg / l) and chloramphenicol (7 mg / l) were added to the media prior to use. Waste media was stocked in a 20 l carboy and treated with bleach prior to disposal.

The turbidostat was run with approximately 0.7 l of LB media. A 100 ml overnight culture of NUB3621-R:ThEV grown at 55 °C in liquid TS medium was
Figure 6-1- The "evolution machine". A New Brunswick Scientific BF3000 Benchtop fermentor was modified to be used in the evolution experiments described in this work. Relevant probes are shown.
used to inoculate the turbidostat. The turbidity probe controlled the peristaltic pumps for media intake and outtake. Both pumps were activated every time the laser probe reported a turbidity of 30 %, equivalent to an OD$_{600}$ of 0.6. pH and oxygen levels of the culture were monitored, but not controlled. Because the communication between the fermentor and the controlling software were found to "freeze" after prolonged periods of inactivity, the agitation speed was set to cycle between 150 and 200 rpm every 3 minutes. The temperature was set to increase 0.5 °C every 24 hour.

Cell culture samples (12 ml) were withdrawn from the turbidostat every 24 hour, just before the temperature was increased. Glycerol (10 % final concentration) was added to the cell samples prior to storage - 80 °C.

6.6 Isolation of adk mutants generated during turbidostat runs

A cell culture sample from the frozen glycerol stock was spread onto solid TS medium containing rifamycin and chloramphenicol. Cells were kept for 16 hour at the same temperature as they were originally isolated.

Individual colonies were grown in 5 ml liquid TS medium for one night and harvested by centrifugation (5,000 rpm for 10 minutes). Cells were brought back into solution with 0.5 ml of buffer B1 from the Genomic-tip 20 / G kit (QIAGEN) complemented with lysozyme and RNAse A and incubated for 30 minutes at 37 °C. 175 μl of buffer B2 (QIAGEN) were added to the lysed cells and the final mixture kept at 55 °C for 30 minutes. Total DNA was isolated with one phenol:
chloroform followed by two chloroform extractions. Total DNA was recovered by isopropanol precipitation (0.7 volumes) followed by a 70 % ethanol wash. DNA pellets were dried for one night at room temperature and brought back into solution with 100 µl of water.

Isolated genomic DNA was used as template for PCR with primers sub_adk_seq_F and sub_seq_R. PCR amplicons were purified using the Montage 96 system from Millipore. Purified PCR samples were subject to sequencing analysis (Lone Star Labs, Houston, TX, USA) with the same oligonucleotide primers.

6.7 Mutant adk population analysis

Alternatively to growing fermentor samples at the isolation temperatures, cells were also screened at temperatures higher than the one they were originally isolated. Glycerol stocks isolated from 55 to 57 °C were diluted and incubated at 60, 65 and 70 °C. Likewise, stocks from 60 to 63 °C were screened at 65 and 70 °C. Finally, cells isolated from temperatures 65 to 67 °C were grown at 70 °C. Isolated colonies had their adk sequence analysed as described above (Figure 8.1).
6.8 Mutant adk cloning, expression and purification

The biophysical and biochemical characterization of mutant AKs isolated during the turbidostat runs required large amounts of highly purified protein. The low protein expression levels observed for the genomic adk copy and the difficulty of growing G. stearothermophilus cells to high optical densities made us look for a different protein expression system. We chose to work with an E. coli system for cloning, expression and purification of mutant AKs because of its low maintenance cost and its high protein expression yields. Furthermore, heterologous expression of AKs from various organisms, including B. subtilis, had already been shown to be possible for this system (Glaser et al., 1992; Perrier et al., 1994). Our strategy to express the mutant AKs can be divided in three parts. First, the mutation of interest was introduced into a protein expression vector carrying the wild-type copy of B. subtilis (vector pETADK; kind gift of Dr. George Phillips, University of Wisconsin, MA, USA) through site-directed mutagenesis. Second, the mutant AKs were produced in high levels using E. coli cells grown in a 14 liter fermentor. Third, the mutant AKs were purified from whole cell extracts by a series of chromatography steps.

6.8.1 Site-directed mutagenesis

Throughout this work we have employed a commercially available kit, QuickChange (Stratagene), which employs specially designed oligonucleotide
primers and the PCR reaction to introduce point mutations at the desired positions.

The point mutations of interest were introduced into the template plasmid (pETADK) using mutagenic oligonucleotide primers and PCR reactions (Table 6.2). The mutagenized plasmids were used to transform competent XL-1 Blue cells (Stratagene) through the heat shock method (Hanahan et al., 1991). The correct introduction of the point mutation was confirmed through DNA sequencing (Lonestar Labs, Houston, TX) after plasmid preparation and purification, performed using the Wizard kit (PROMEGA). Once the adk sequence was confirmed, the plasmid was transformed into a protein expression cell line, E. coli BL21DE3.

6.8.2 AK expression in E. coli

Early cell cultivation experiments were performed in baffled Erlenmeyer flasks kept in a New Brunswick Innova 4300 shaker at 37°C and 250 rpm. Due to the poor protein yields obtained under these growth conditions we switched our cell cultivation protocol. The use of a 14 liter fermentor (Braun Biostat C) greatly increased the cell mass and the protein expression yields obtained. The experimental specifications are described bellow.

A 1 liter Erlenmeyer flask containing 100 ml of LB media and ampicillin (0.1 g / l) was inoculated with the glycerol stock from the desired AK mutant. Cells were grown for approximately 16 hour in a New Brunswick 4300 shaker at
37°C and 200 rpm to yield the starting culture. 14 l of 2 x YT media used for the fermentor run were autoclaved in place. Prior to sterilization, 1 ml of anti-foam solution was added to the media. The media was allowed to cool down to 37°C and supplemented with 0.4 l of 40 mM Potassium phosphate buffer pH 7.4 and carbinicillin (0.1 g / l). The fermentor run was inoculated with the entire starting culture. The culture's stirring and air flow were kept variable in order to maintain a constant PO₂ of 50 %. When the cell culture reached an OD₆₀₀ nm of 0.5 the AK expression was induced with IPTG at a final concentration of 1 mM. At this time glucose was also added to a final concentration of 0.01 M.

The fermentor run was stopped 8 hours after IPTG was added to the media. Typical cell densities at this point were between OD₆₀₀ nm 4-6. Cells were harvested by centrifugation in a News Brunswick JA series centrifuge (10,000 rpm at 4 °C for 10 minutes). The cell pellets were stored at - 80°C in 30 g aliquots. The total cell mass obtained from 14 l of media varied between 90 and 120 grams (wet weight).

6.8.3 AK purification from cell lysates

Approximately 30 g of frozen cell pellets were brought back into solution in 200 ml of ice cold breaking buffer (25 mM Tris pH 7.4 50 mM NaCl, 1 mM EDTA). After homogenized the cells were lysed by sonication in a Branson Sonifier 250 (3 x 2 minutes pulses at 70 % duty cycle and an output control of 7). The cell lysate was clarified by centrifugation in a News Brunswick JA series
centrifuge (15,000 rpm at 4 °C for 60 minutes). AK was then purified from the clarified lysate through 4 chromatography steps.

6.8.3.1 Miscellaneous information on chromatography columns and protein handling

Throughout the purification process protein samples were dialyzed in 8,000 Da molecular weight cut off dialysis tubes. Protein samples were quantified through UV spectroscopy at 280 nm, using Lambert-Beer's law:

\[ A = \varepsilon \times c \times l \]

where \( A \) is absorbance at the desired wavelength, \( \varepsilon \) is the theoretical extinction coefficient calculated for AK \((12345 \text{ M}^{-1}\text{cm}^{-1})\) and \( l \) is the pathlength in cm. The ratio of the absorbance values found for the sample at 280 and 260 nm \( (A_{280} : A_{260}) \) was used to monitor nucleic acids contamination of protein samples.

Protein samples were separated in 12.5 % discontinuous SDS PAGEs. Protein gels were stained with 0.1% Coomassie Brilliant Blue in 40% methanol, 40% acetic acid (volume by volume).

All buffers were filtered through a 0.22 \( \mu \)m filter (Millipore) and degassed before applied into FPLC chromatography systems. Sample separated in such systems were also filtered through 0.22 \( \mu \)m filters (Millipore). All chromatography columns were stored in 20 % ethanol and washed with water before equilibrated with the appropriate buffer.
6.8.3.2 DE52 chromatography

The De52 is a weak anion exchanger resin. The theoretical isoelectric point for *B. subtilis* AK is 5.6. Thus, at neutral pHs, AK is expected to bind to the DE52 resin. The resin was prepared and equilibrated as described below.

Approximately 40 grams of DE52 resin were suspended in 300 ml of 0.1 M Tris pH 7.4, 1.0 M NaCl. The resin was allowed to sediment for 15 minutes, the buffer discarded and the resin washed 2 more times with the same buffer. The resin was then washed 3 more times with buffer A (25 mM tris pH 7.4, 50 mM NaCl) supplemented with 1 mM EDTA and poured into a 20 x 5 glass column (blabla). The DE52 column was equilibrated with at least 200 ml of buffer A, passed through by gravity. The clarified lysate was passed through the column and non-bound proteins washed with at least 500 ml of buffer A. AK was eluted from the column with a 200 ml linear salt gradient (50 mM to 1.0 M NaCl in buffer A). Fractions were collected in approximately 5.0 ml with the aid of a fraction collector. Protein samples were separated in a 12.5% SDS-PAGE. Fractions rich in AK were pooled together and dialyzed overnight at 4 °C against 2 l of buffer A containing 0.3 mM DTT (prepared fresh). The dialyzed sample was used in the next chromatography step.
6.8.3.3 FPLC MonoQ chromatography

The DE52 eluate is highly contaminated with nucleic acids. We used a strong anion exchanger, MonoQ (Amersham) to further purify AK and also reduce the amount of nucleic acid contamination in the protein sample. The column was run using a FPLC system (Amersham).

After equilibration with buffer A the sample was applied onto the MonoQ column. Non-bound proteins were washed with 5 column volumes of buffer A. AK was eluted by a 15 column volumes of a linear salt gradient (25 mM to 350 mM NaCl in buffer A). Eluted proteins from the salt gradient were collected in 5.0 ml fractions using an automatic fraction collector. Fractions were run in a 12.5 % SDS PAGE and the ones containing AK were pooled together and dialyzed over night at 4 °C against 2.0 l of Buffer A containing 0.3 mM DTT and 1 mM MgCl₂. The dialyzed sample was utilized in the next chromatography step.

6.8.3.4 Blue Sepharose chromatography

The blue sepharose (Amersham) column contains a blue cibracon group attached to an agarose matrix. This resin has been shown to work as an affinity chromatography for nucleic acid binding proteins. Also it has been previously employed to purify AK from various organisms (Barzu and Michelson, 1983). The presence of nucleic acids in the clarified cell lysate and in the DE52 eluent seemed to greatly decrease the AK ability to bind to the blue sepharose resin
and prevented its application earlier in the purification protocol. The use of DNAases to circumvent such problem was deliberately avoided in order to prevent the contamination of shared use glassware and chromatography systems with such enzymes.

Approximately 50 ml of Blue sepharose (Amersham) resin slur were poured into a glass column according to the manufacturer’s guidelines. The column was run at room temperature by gravity. The dialyzed fraction from the MonoQ step was applied onto the column and non-bound proteins washed with 4 column volumes of buffer A supplemented with 1 mM MgCl₂. AK was eluted with a 200 ml linear salt gradient (from 25 mM to 2.0 M in buffer A with 1 mM MgCl₂). Fractions (approximately 5 ml) from the salt gradient were collected using a fraction collector and analysed in a 12.5 % SDS PAGE. Fractions containing AK were pooled together and concentrated using Vivaspin concentrators with a 10,000 Da molecular weight cut off to no more than 3.0 ml. The concentrated fraction from the affinity chromatography step was used directly for the size exclusion chromatography step.

6.8.3.5 Size exclusion chromatography

Size exclusion chromatography is usually used as the final step in protein purification strategies for protein crystallography. It ensures that the purified protein of interest is free of higher molecular weight protein aggregates that could prevent crystal growth.
We employed a preparative grade Superdex 200 (Amersham) column connected to a FPLC system as the last step in AK purification. The concentrated sample from the Blue Sepaharose chromatography was applied into the column previously equilibrated with 25 mM Tris pH 7.4 150 mM NaCl. The sample was run at 0.50 ml per minute. Protein fractions (3.0 ml) were collected as described for the MonoQ step. Protein samples were analysed in a 12.5 % SDS PAGE. Fractions containing AK were pooled together and dialyzed against 2.0 l of 10 mM HEPES pH 7.0 for one night at 4 °C. The dialyzed AK fraction was concentrated using Vivaspin concentrators (10,000 Da MW cut off) and stored at -80 °C.

Typical protein yields were between 2 and 4 mg of purified AK for every gram of lysed cell. Visual analysis of purified AK in SDS PAGE gels stained with Coomasie blue determined that AK was more than 99.0% pure. Typical $A_{280} : A_{260}$ ratios for purified AK were between 1.7 and 2.0.

6.9 Temperature denaturation of AK followed by Circular Dichroism

The secondary structure content of AK at various temperatures was investigated by circular dichroism (CD). Prior to CD experiments, samples were dialyzed overnight against 2.0 l of CD buffer (10 mM potassium phosphate buffer pH 7.4). All CD experiments were performed with 20 mM protein samples in CD buffer. Buffers and samples were filtered through 0.22 μm syringe filters.
Figure 6-2 - Purified AK. The quality of AK purification was assessed by 12.5% SDS-PAGE. The protein fraction from the last step of purification for *B. subtilis* wild-type AK is shown (B). Molecular weight standard is shown in (A). All AK mutants could be purified to comparable qualities.
(Millipore). All CD data was acquired using an AVIV CD spectrometer. Fraction of unfolded protein was calculated by following the changes in ellipticity at 220 nm (bandwidth 1.0 nm, averaging time 1.0 second) as a function of temperature. The rate of temperature increase was 1 °C per minute (temperature deadband 0.1 °C). CD signal at 220 nm was acquired every minute. CD signal acquired for protein samples was subtracted from the traces obtained for buffer alone prior to data analysis. The traces were fit to the Hill equation and smoothed using a local technique with polynomial regression and weights computed from the Gaussian density function (Sigma Plot). The midpoint denaturation temperature was estimated from the CD traces assuming a reversible single step transition from the folded to the unfolded states (F↔U).

6.10 AK activity assays

AK specific activities towards ADP production were determined at various temperatures in reaction buffer containing 25 mM potassium phosphate 7.2, 0.3 mM DTT, 25 mM magnesium chloride, 65 mM potassium chloride, 2 mM ATP and 2 mM AMP. Reactions were performed in a Grant-Boekel GD120L water circulator (± 0.1 °C). 1.0 ml of reaction buffer was incubated at the indicated temperatures for 5 minutes. 1.0 μl of a 1.0 mM enzyme stock solution in reaction buffer was added to the reaction mixture, making a final AK concentration of 1.0 μM. The reaction was allowed to proceed for 1 minute. The reaction was stopped with 1.0 μL of a 50 mM Ap5A solution (final concentration 50 μM). Reaction tubes
were kept in an ice bath until analysed by HPLC chromatography. All reactions were performed in triplicate. The background formation of ADP through temperature hydrolysis of ATP was measured by omitting AK from the reaction and treating it the same way as AK containing mixtures.

The total amount of ADP generated by different AK mutants were estimated by separating the reaction mixture on a Eclipse XDB-C8 reversed-phase chromatography column (Agilent) attached to a Shimadzu HPLC system (pump: LC10AT VP; UV detector: SPD10A VP; degasser: DGU14A). The column was equilibrated with 50 mM potassium phosphate pH 7.0, 10 mM (tetrabutyl-ammonium borate), 7 % acetonitrile. Samples were run at 1.0 ml per minute. Peak detection was performed by following the sample absorption at 327 nm. The area under the peaks were integrated and used to estimate the amount of ATP, ADP and AMP present in the samples. The amount of ADP generated in the AK containing samples was subtracted from the samples where AK was omitted. All experiments were performed in triplicate.

6.11 Crystallization of B. subtilis AK mutants

All solutions utilized for crystallization trials were filtered through 0.22 μm filters (Millipore) prior to use. Buffer, precipitant and salt solutions were kept at room temperature, protected from light.

Vapor diffusion in hanging drops was the protein crystallization method employed throughout in this work. 24 well pre-greased trays and siliconized
cover slides (Hampton Research) were used for protein crystallization trials. The wells were cleaned with a brief spray from an air can prior to addition of the individual reagents forming the mother liquor solution (final volume of 0.5 ml). The well solution was homogenized by agitation in a rocking plate for at least 15 minutes prior to use. Air can spray was used to remove dust from the covering slides before protein drops were set up.

Purified AK was removed from the -80 °C freezer and thawed on ice. Enough Ap5A and water was added to yield a final protein concentration of 18 mg / ml and a final Ap5A concentration of 3.5 mM. The protein-inhibitor solution was incubated on ice for at least 15 minutes before crystallization drops were set up.

A volume of 6 µl of protein-inhibitor solution was added to the center of a cover slide. The same volume of mother liquor from one of the wells was added on top of the protein Ap5A solution drop. The cover slide was inverted and used to seal the well from which the mother liquor drop was drawn. Crystal trays were kept at 20 °C and screened for protein crystal growth with the aid of a microscope every 48 hour for the first week and every seven days thereafter.

Crystals obtained for the A193V AK mutant protein were used to seed crystallization drops for the Q199R and Q199R / Q16L AK mutants. The first step in the seeding technique was to gently touch the seed crystal in its crystallization drop with a cat whisker. The crystal seeds were then transferred to equilibrated drops (48 hour after original set up) containing the protein of interest. Seeded drops were screened as described before.
6.12 AK mutant crystal analysis and data collection

Crystals obtained for mutant AKs were transferred from the crystallization drop to a 20 μl mother liquor drop placed on a cover slide with the aid of a small crystal loop (Hampton Research). It is common for protein crystallization conditions containing high concentrations of PEG to develop a film covering the protein crystal and the drop. Prior to X-ray diffraction analysis, the PEG film was carefully removed from crystals. Crystals were flash-frozen in a -170 °C nitrogen cold stream directly from their mother liquor solutions and without the use of cryoprotectant solutions.

The ultimate test for a crystal is its X-ray diffraction limit and quality. All crystal data collection in this work was performed in a home source X-ray generator (MSC RU-H3R) with a rotating copper anode (wavelength = 1.54Å) set at 50 kV and 100 mA. The X-ray beam was focused with Osmic Mirrors™ and collimated to 0.3 mm. A Rigaku R-Axis IV++ image plate detector system was employed to collect diffraction data. All crystal data was obtained under cryogenic conditions.

Protein crystals were screened by acquiring four 10 minutes exposure images, 90° apart from each other. Diffraction spots shape (round or slightly elliptical) and intensities over the background (θ, higher than 10 at 1.8 Å) were used as parameters for selecting crystals that were further characterized. The four images acquired for crystals that met the spot shape and intensity requirements were processed using the Crystal Clear suite of crystallographic
tools (Pflugrath, 1999). We employed the program’s crystallographic routines to
assign the crystal’s unit cell parameters (dimensions - a, b and c; and angles - α,
β and γ), space group, mosaicity and data collection strategy.

6.13 X-ray data processing

The crystal clear package was employed to merge the reflection images.
The program calculates the reflection Miller indexes (hkl) and intensities (I). Data
merging quality was assessed by the statistical $R_{\text{merge}}$:

\[ R_{\text{merge}} = \frac{\sum I_{hkl} - \bar{I}_{hkl}}{\sum \bar{I}_{hkl}} \]

where $I_{hkl}$ is the intensity of the reflection (hkl) and $\bar{I}_{hkl}$ is the mean intensity of all
symmetry-equivalent (hkl) reflections.

6.13 Molecular Replacement

Experimental phases were obtained through molecular replacement. *B.
subtilis* AK (PDB 1P3J) (Bae and Phillips, 2004) was used as the crystallographic
search model. All protein ligands (zinc, magnesium and Ap5A) were removed
from the PDB file prior to its use as search model for molecular replacement
using CNS cross-rotation and translational routines.
6.14 Solution refinement

Following the molecular replacement routine an iterative process of model refinement and manual rebuilding was performed. Throughout this process the quality of the protein model was judged by the statistical R-factor values obtained for the working set (R) and the test set (R_{free}). The test set was composed of 5% of the data chosen randomly from the working set by a routine in CNS. The definition of the statistical R values is:

$$R = \frac{\sum |F_o - F_c|}{\sum |F_o|}$$

Where F is the structure factor for the observed (F_o) and calculated (F_c) amplitudes. The structure factors can be calculated from the intensities:

$$I_{(hkl)} = |F_{(hkl)}|^2$$

The iterative process was generally composed of the following steps: simulated annealing, energy minimization, individual B-factor refinement, map calculation and manual rebuilding. When R_{free} values could no longer be lowered through these steps, solvent molecules were added to the model. Water molecules were added using a routine in CNS and visual inspection of the model. The B-factor cut-off for waters was set at 50 Å. The quality of the model was monitored by inspection of the statistical R and the R_{free} values. The final structure quality was evaluated using the Procheck package.
6.15 Total Adenine Nucleotide Extraction

Wild type and recombinant NUB3621-R cells were cultivated in liquid TS media containing rifamycin (5 mg / l) in a New Brunswick Scientific BF3000 Benchtop fermenter system. Cells were grown at 50°C up to an O.D.590 of 0.5. A 50 ml aliquot of cell culture was then transferred to a 250 ml Erlenmeyer and placed into a New Brunswick Innova 4230 shaker set at 65°C and 255 rpm for the specified amount of time. For the transient increase in temperature experiment, cells were kept in the fermenter throughout the duration of the experiment. Cell culture samples (1.0 ml, in triplicates) were transferred into 9.0 ml of Boiling Tris Buffer (100 mM Tris acetate pH 7.75, 4 mM EDTA). The samples were kept at 100 oC for two minutes. Samples were then chilled on ice, centrifuged (2 minutes at 5.000 rpm at 4°C) and the supernatant transferred to a new tube. Growth media samples were obtained by passing 10 ml of cell culture through a 0.2 μm filter (Millipore) and treated as described above. Total adenine nucleotide samples were kept on ice until assayed with the luciferase activity assay. Two acid extraction methods (Swedes et al. 1975; Kahru et al. 1990) were also employed.

6.16 Total protein extraction

Total protein was extracted as described previously (Swedes et al. 1975). 750 μl of cell culture was added to 750 μl of an ice cold 10 % HClO4 solution.
Samples were then centrifuged (10 minutes, 14,000 rpm) and the pellet was brought back into solution with 100 μl of 50 mM Phosphate Buffer pH 7.3. Protein concentrations were estimated using the Bio-Rad protein Assay (Bio-Rad).

6.17 Sample preparation and the Luciferase activity assay

Sample preparation for the determination of ATP, ADP and AMP concentrations was carried out as described by Chapman and colleagues (1971)(Chapman et al. 1971). ATP concentrations were determined by means of the luciferase reaction. Firefly (*Photinus pyralis*) luciferase (1.13.12.7) and luciferin (D-[-]-2-(6'-hydroxy-benzothiazolyl)-Δ2-thiazoline-4-carboxylic acid) were purchased from Roche Applied Science. Sample preparation was performed according to the manufacturer’s instructions. 50 μl of an ATP sample was placed into a Greiner 96 well plate. 50 μl of luciferase solution (50 mM Tris acetate pH 7.75, 2 mM EDTA, 60 mM DTT, 0.01 % -w/v- BSA, 10 mM magnesium acetate, 35 μM D-Luciferin and 0.05 μg of Luciferase) was added to the samples and the plate was immediately placed on an Alpha Innotech Fluorchem 5500. The signal was integrated from 1 to 60 seconds and the ATP concentrations estimated through comparison with an ATP standard curve. AMP and ADP were converted to ATP (Chapman et al. 1971) prior to the luciferase experiments and their concentrations estimated as follows: [AMP] = ([ATP] + [ADP] + [AMP]) − ([ATP] + [ADP]); and [ADP] = ([ATP] + [ADP]) − [ATP]. All measurements were performed in triplicate.
Chapter 7 - The Recombinant NUB3621-R:ThEV strain

Evolution is constrained by the organism’s phyletic history. In our weak link approach, the phyletic history of the model gene was represented not only by the gene’s primary sequence but also by its location in the genome. In the present work the previous evolutionary history of adk in G. stearothermophilus NUB3621-R was respected by employing homologous recombination to replace the endogenous gene with the B. subtilis adk.

In this chapter I will show that introduction of the B. subtilis adk into the host’s genome did not alter expression levels for the B. subtilis adk gene nor disturb the organization of neighboring genes. More importantly, I will show that heat inactivation of B. subtilis AK in recombinant cells disrupted the adenine nucleotide homeostasis of the host organism and was the reason for its temperature sensitive phenotype. Taken together these results indicated that the evolution of the weak link is mainly constrained by its physiological role in the organism.

7.1 NUB3621-R adk is part of the S10-spc-α cluster

The weak link gene was introduced into the host’s genome through homologous recombination. In order to be successful, it was necessary to determine the DNA sequence neighboring adk in the NUB3621-R genome.
Unfortunately, there was no sequence information available for NUB3621-R. Nevertheless, in all other gram-positive bacteria adk is part of the S10-spc-α cluster (Fig 7.1). This region harbors mostly ribosomal genes and is under control of promoter sequences found upstream the S10 gene (Li et al., 1997; Suh et al., 1996). Degenerate primers were designed to amplify the genomic region between two of the genes in this cluster; rplO (primer L15_deg_F) and rpmJ (primer L36_deg_R). DNA sequence analysis of this region revealed that NUB3621-R adk, as in other gram-positive bacteria, is found between secY (upstream) and map (downstream) genes. The sequence information obtained for NUB3621-R genes coding for SecY, AK, MAP and InfA proteins were deposited at genebank with the access number AY729037 (Figure 7.1).

7.2 Homologous recombination

Although homologous recombination in G. stearothermophilus NUB3621-R had never been reported in the literature, it has been very well characterized in other bacterial systems, including the gram-positive B. subtilis (Harris-Warrick and Lederberg, 1978a; Harris-Warrick and Lederberg, 1978b; Jorgensen et al., 1990). In B. subtilis, homologous recombination between genomic and plasmid DNA requires approximately 70 bp of homology (Khasanov et al., 1992). Sequence information obtained for the genomic region flanking NUB3621-R adk was used to design primers to amplify approximately 1.0 kb upstream (secY) and downstream (map and infA) of NUB3621-R adk. These fragments were used as
Figure 7.1: Genome location and sequence information for NUB3621-R *adk* gene. A - gene organization of the spc operon. Ribosomal genes are depicted as red arrows, *secY, adk* and *map* are depicted as blue, green and orange arrows, respectively. Coding sequence for *secY, adk* and *map* are highlighted in blue, green and orange, respectively.
homologous regions to promote the insertion of *B. subtilis adk* into the host's genome.

Because NUB3621-R protoplasts can only be transformed with circular, replicative plasmid DNA (Chen et al., 1986; Wu and Welker, 1989) the recombination cassette composed of the host's homologous DNA sequences, *B. subtilis adk* and a promoterless genetic marker were introduced into an *E. coli* / G. stearothermophilus shuttle vector, plasmid pSTE12. The antibiotic selective marker for plasmid pSTE12 in G. stearothermophilus is tetracycline resistance (Narumi et al., 1992). Introduction of the homologous recombination cassette into pSTE12 yielded construct p12ThEV (Figure 7.2).

*B. subtilis adk* expression in recombinant cells was kept under control of the same regulatory regions controlling expression of *G. stearothermophilus adk* in wild-type cells. All putative ribosomal binding sites (RBS) found in the intervening region between secY and adk were preserved in the homologous recombination vector. In addition to *B. subtilis adk*, the homologous recombination cassette also contained a promoterless copy of the chloramphenicol acetyl-transferase gene (*cat*), which was employed as a genetic marker. Once the homologous recombination cassette had been properly introduced into the host's genome, *B. subtilis adk* and *cat* expression should be constitutively driven by the S10 promoter sequence.

Following transformation of NUB3621-R protoplasts with plasmid p12ThEV, selection for recombinant cells took advantage of two attributes of the homologous recombination cassette. First, the conditional expression of the
Figure 7-2— The homologous recombination cassette and the p12ThEV construct. A — schematic representation of the homologous recombination cassette showing its gene organization: secY — pre-protein translocase secY subunit; adk^{B,ab} — B. subtilis adenylate kinase; cat — chloramphenicol acetyltransferase; map — methionine aminopeptidase; infA — initiation factor IF-1. Relevant restriction sites are shown. B — p12ThEV construct showing the homologous recombination cassette (shown in A) and genes conferring antibiotic resistance to ampicillin (amp) and tetracycline (tet). C — 1.0 % agarose gel showing PCR products amplified from p12ThEV plasmid with oligonucleotide primers for secY (lane 1), B. subtilis adk (lane 2), cat (lane 3) and map / infA (lane 4) fragments found in the homologous recombination cassette. Lane 5 — molecular weight marker (1.0 μg of λ DNA treated with Pst I). Genes are depicted as boxes and RBSs as black bars. Approximate molecular weights are also shown.
promoterless cat gene should enable NUB3621-R cells to grow in chloramphenicol containing media. Second, the low midpoint denaturation temperature (T_m) of B. subtilis AK (50.7°C) (Perrier et al., 1994) should obliterate growth of recombinant cells at temperatures higher than the protein's T_m. Moreover, loss of the recombinant vector should result in a tetracycline sensitive phenotype.

Cells displaying chloramphenicol resistance, tetracycline and temperature sensitivity could be isolated after extensive growth in non-selective liquid media and plating in selective solid media (Figure 7.3). The highest growth temperature for these cells was established to be 56°C. At this temperature NUB3621-R:ThEV cells failed to produce colonies after a 48 hour incubation period.

7.3 Genetic characterization of NUB3621-R:ThEV cells

Recombinant cells displaying the proper phenotype had their genome organization investigated by PCR analyses, Southern blot and DNA sequencing. In a genetically stable recombinant strain the gene replacement was expected to have occurred through a unique double homologous recombination event. It was also necessary to cure the cells of plasmid p12ThEV.

PCR reactions performed with specific oligonucleotide primers for map and secY genes confirmed the introduction of the homologous recombination cassette into the host genome (Figure 7.4 B and C). Southern blots performed with specific probes for adk, cat and tet showed that the homologous
<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>rifampicin</th>
<th>rifampicin + chloramphenicol</th>
<th>rifampicin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>55°C</td>
<td>55°C</td>
<td>65°C</td>
</tr>
</tbody>
</table>

NUB3621-R ThEV  

NUB3621-R  

Figure 7-3- Temperature sensitive phenotype of NUB3621-R:ThEV cells. Recombinant NUB3621-R:ThEV and wild-type NUB3621-R cells were streaked on TS solid media supplemented with the indicated antibiotics. Plates were kept 24 hour at the indicated temperatures. Wild type NUB3621-R cells are rifamycin resistant and grow in the temperature range of 45°C-75°C. Following gene replacement of its endogenous adk gene with plasmid p12hEV recombinant cells should be rifampicin and chloramphenicol resistant and display a temperature phenotype.
Figure 7-4: Genetic characterization of NUB3621-R:ThEV cells. A - Schematic representation of NUB3621-R:ThEV genome after gene replacement. Genes are depicted as boxes and RBSs as black bars. Wild-type HindIII and EcoRI restriction sites and the BamHI restriction site introduced by the homologous recombination cassette are shown. Arrows indicate the approximate annealing location for primers used in B. B - PCR reaction performed with primers specific for secY (red arrow) and map (blue arrow) for p12ThEV plasmid DNA (pl) and genomic DNA extracted from recombinant (rec) and wild-type (wt) cells. Recombinant DNA displayed a larger PCR fragment due to the additional cat gene found in the homologous recombination cassette. C - BamHI treatment of the PCR products from B generated the expected restriction pattern, only the recombinant fragment was recognized by the enzyme. D - Southern blot analysis of NUB3621-R:ThEV genomic DNA. Total DNA from NUB3621-R wild-type cells (wt); total DNA from NUB3621-R:ThEV cells (rec) and plasmid p12ThEV control DNA (pl) were treated with HindIII and EcoRI, separated on an 0.6% agarose gel and transferred to a nylon membrane. The membrane was then hybridized with DIG labeled probes for B. subtilis adk, cat or tet genes (depicted as boxes on A).
recombination cassette was introduced as a single copy in the host's genome and NUB3621-R:ThEV cells were cured for the homologous recombination vector, p12ThEV (Figure 7.4 D). Genomic DNA sequencing further confirmed the integrity of all genes introduced by the homologous recombination cassette.

7.4 *B. subtilis* AK expression in recombinant NUB3621-R:ThEV cells

The homologous recombination cassette was designed to maintain the *adk* expression in recombinant cells at comparable levels to wild-type cells. AK levels in wild-type and recombinant cells were investigated by purifying the enzyme from natural abundance. The amount of AK units per mg of total protein was found to be comparable between the two cell lines cultivated under identical conditions (Table 7.1). The *cat* gene was also employed to indirectly assess *B. subtilis* AK expression levels in recombinant cells. Because the *cat* and the *B. subtilis adk* genes are under control of the same upstream promoter they were expected to be expressed at similar levels. CAT activity was not induced by chloramphenicol and was stably maintained in different cell lineages (Figure 7.5 A). These results taken together indicated that *B. subtilis* AK expression levels are comparable to the ones found for *G. stearothermophilus* AK in wild-type cells.
<table>
<thead>
<tr>
<th>Fraction</th>
<th>NUB3621-R</th>
<th>NUB3621-R:ThEV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specific activity (U/mg)</td>
<td>Purification factor</td>
</tr>
<tr>
<td>Clarified lysate Blue Sepharose Superdex 200</td>
<td>338</td>
<td>1</td>
</tr>
</tbody>
</table>

*Enzyme activities were measured at 20 °C by the coupled reaction method.*
Figure 7-5– Expression and activity profiles for AK in NUB3621-R:ThEV cells. A - CAT expression profiles for three wild-type (w1 to w3) and three NUB3621-R:ThEV (r1 to r3) cell cultures. Cells were grown at 50°C in the presence (+) or absence (-) of chloramphenicol (Chl) until an OD₅₀₀ of 1.0. Cells were lysed and the total units of CAT calculated from the rate of acetyl group transfer from acetyl-CoA to chloramphenicol. B - Temperature profiles for the activity of adenylate kinase isolated from wild-type NUB3621-R (solid squares) and recombinant NUB3621-R:ThEV (open circles) cells. AK was incubated at indicated temperatures for 5 minutes. Substrates AMP and ATP were added to the enzyme mixture and the reaction allowed to proceed for 3 minutes at indicated temperatures. The reaction was stopped by the addition of Ap5A. ADP production was estimated by the degradation of NADH through a coupled assay containing phosphoenolpyruvate, pyruvate kinase and lactate dehydrogenase. Curves shown are the average of three independent experiments. Error bars are shown for the standard deviations.
7.5 AK activity and cellular metabolism

In a weak link experiment the host organism's growth must be directly correlated with the evolution of the target protein. In our system, heat inactivation of AK was expected to be the main reason behind the observed temperature sensitive phenotype displayed by recombinant cells.

In order to assign the temperature sensitive phenotype of recombinant cells to AK's physiological role in the phosphotransfer network, I first investigated the enzyme activity at different temperatures. AK purified from NUB3621-R:ThEV cells reached its maximum activity at 60°C and was dramatically reduced at higher temperatures. In contrast, AK isolated from wild-type cells showed an activity maximum at 75°C (Figure 7.5 B). Previous studies have determined that the midpoint temperature denaturation ($T_m$) for *B. subtilis* AK is 50.7 °C (Glaser et al., 1992), although in my hands the $T_m$ of the enzyme is 54.9°C (Figure 8.4). Moreover, recombinant cells can not grow at temperatures higher then 56°C. It was quite surprising that the maximum activity observed for NUB3621-R:ThEV AK was 5 to 10°C higher than its previously determined $T_m$s and 5°C in excess of the highest growth temperature for the host organism.

The observed discrepancies between *B. subtilis* AK highest temperature activity, its $T_m$s and NUB3621-R:ThEV phenotype can be explained by two previous observations. First, the $T_m$ determinations by Glasser and colleagues (Glaser et al., 1992) were performed in the absence of any ligands or salts and it has been shown that an AK inhibitor, Ap5A, could increase the $T_m$ for *B. subtilis*
AK by more than 20 °C (Bae and Phillips, 2004). Therefore it is possible that the enzyme may have been protected against heat denaturation during the kinetic experiments by the reaction mixture (2 mM AMP and ATP, 5 mM MgCl₂, 65 mM KCl and 1 mM DTT).

Keeping in mind that in vitro conditions are unlikely to precisely reproduce what occurs in vivo, the results from the phenotypic characterization and the enzyme kinetics were in good agreement and showed that the host organism could not grow at temperatures where AK was inactivated. As previously observed for E. coli cells, AK heat inactivation resulted in a sharp temperature sensitive phenotype (Glembotski et al., 1981; Haase et al., 1989).

Because AK promotes the homeostasis of adenylate species, its activity in living cells can be better assessed by probing the adenylate energy charge (EC) under different growth regimes. EC is commonly utilized to investigate the relative ratios of ATP, ADP and AMP in living cells. It is defined as (Atkinson, 1968):

\[
EC = ([ATP] + 0.5 [ADP]) / ([ATP] + [ADP] + [AMP])
\]

and directly measures the amount of energy stored in adenine nucleotide pools accessible for cellular metabolism (Atkinson, 1968; Chapman et al., 1971; Swedes et al., 1975). More relevant to this work, E. coli cells expressing a heat labile AK displayed reduced EC values when grown at restrictive temperatures (Glembotski et al., 1981).

Unfortunately, studies on the energy metabolism of thermophiles are quite scarce (Kahru et al., 1982) and, in fact, none is available for G.
*stearothermophilus*. Therefore, before analyzing EC values for recombinant cells I had to first establish that EC was an appropriate indication of the metabolic state of wild-type NUB3621-R cells and, more importantly, that it correlated with AK activity. EC investigation was accomplished by probing the total adenylate species found in cell cultures media under different conditions. Cellular adenylate levels were estimated from subtracting total values from the ones found for the cell culture filtrates. ATP levels were reported by a coupled luciferase reaction. AMP and ADP were enzymatically converted to ATP prior to the luciferase experiments and their concentrations estimated as follows: \([\text{AMP}] = ([\text{ATP}] + [\text{ADP}] + [\text{AMP}]) - ([\text{ATP}] + [\text{ADP}])\); and \([\text{ADP}] = ([\text{ATP}] + [\text{ADP}]) - [\text{ATP}]\) (Chapman et al., 1971).

Growth curves performed for wild-type cells at 50 °C and 65 °C revealed a dramatic decrease in doubling times at lower temperatures (45 vs 18 minutes). Moreover, total protein levels were also shown to be higher at 65°C than at 50°C (Table 7.2). EC values have been used before to reliably assess the metabolic state of cells in other systems (Chapman et al., 1971; Glembotski et al., 1981; Kahru et al., 1982; Swedes et al., 1975). Assuming growth rate and total cellular protein concentration are fair indicators of cellular metabolism I investigated the EC values for wild-type cells at high (65 °C) and low (50 °C) metabolic states.

The first experiment to investigate energy charge values in wild-type cells probed changes in EC when cells were transiently (20 minutes) transferred from a low metabolic state to a high one then returned to original low metabolic levels. As expected, EC values were shown to increase once cells were shifted from
<table>
<thead>
<tr>
<th>Growth condition</th>
<th>NUB3621-R</th>
<th>NUB3621-R:ThEV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EC</td>
<td>total protein (mg / ml of culture)</td>
</tr>
<tr>
<td>Time at 65°C (minutes)a</td>
<td>0</td>
<td>0.22 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.45 ± 0.015</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.41 ± 0.016</td>
</tr>
<tr>
<td>Temperature shift (°C)b</td>
<td>50</td>
<td>0.22 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>50 to 65</td>
<td>0.43 ± 0.011</td>
</tr>
<tr>
<td></td>
<td>65 to 50</td>
<td>0.20 ± 0.020</td>
</tr>
</tbody>
</table>

*a* Cells were grown at 50 °C in a benchtop fermenter, transferred to flasks and kept in a shaker at 65 °C for the indicated times.

*b* Transient increase in temperature experiment. Cells were kept in the fermenter for the entire experiment.

*c* ND – Not determined due to high media concentrations of adenylate species.
50°C to 65 °C and, more importantly, EC values decreased to initial low levels upon return to 50 °C (Table 7.2b). EC values were also shown to be increased after 30 minutes at 65 °C and to be kept at comparable high values even after 1 hour at this temperature (Table 7.2a). These results taken together showed that EC values closely matched the metabolic state of wild-type cells.

The impact of AK inactivation in the cellular EC was investigated by performing the same experiments with recombinant NUB3621-R:ThEV cells. At 50°C, doubling times for recombinant and wild-type cells are comparable and, as expected, recombinant cells displayed low EC values at 50°C, similar to the ones observed for wild-type cells under the same conditions (Table 2). As opposed to the wild-type enzyme, AK purified from NUB3621-R:ThEV cells cannot sustain its activity at high temperatures (Figure 7.4). Moreover, recombinant cells cannot grow at 65°C. EC values for recombinant cells were found to be drastically reduced after 20 minutes at 65°C and correlated with the observed temperature sensitive phenotype. Upon return to 50°C, EC values for recombinant cells could not be reliably calculated due to the high levels of AMP found in the culture media (Table 7.2a); the same was true after extended incubation times (30 and 60 minutes) at 65°C.

In both experiments, the presence of adenylate species in the media was not followed by dramatic changes in optical density or total protein content. Thus, cell lysis followed by cytosol leakage seemed unlikely. However, cell viability for NUB3621-R:ThEV was dramatically decreased following the transient
temperature challenge. Only 0.001 % of NUB3621-R:ThEV cells were able to grow on solid media at 50 °C following a 20 minutes incubation at 65 °C.

High media adenylate levels have been correlated with stressful growth conditions. For an adenine-requiring mutant of E. coli growing under adenine starvation, up to 35 % of the total adenylates were found in the culture media (Swedes et al., 1975). Presence of adenylate species in the media was thus considered an indication of cell stress. For both wild-type and recombinant cells at 50 °C, AMP and ADP present in the media added to approximately 15 % of the total cellular adenylate species. At 50 °C wild-type cells may be under stress due to low temperatures, 15°C removed from its optimal growth temperature. Indeed, media adenylate levels were reduced for wild-type cells at 65°C. For recombinant cells, on the other hand, media adenylate levels were greatly increased after 30 minutes at 65°C or upon return to 50°C following a 20 minutes incubation at 65°C. Under these conditions, AMP and ADP found in the culture media added to more than 50% of total adenylates and prevented reliable EC measures. Presence of adenylate in the media was taken as an indication that recombinant cells are under stress at 65 °C, an observation corroborated by their low viability at this temperature.

Although the EC values and growth rates are likely to reflect the activities of all enzymes in the phosphotransfer network, the fact that low EC values and cell viability coincided with heat inactivation of B. subtilis AK clearly suggested that this enzyme was the main reason behind the temperature sensitive phenotype in recombinant cells.
Chapter 8 - Single gene evolution: a punctuated history
of chance events

8.1 Alternative evolutionary strategies

The physiological studies performed on NUB3621-R:ThEV showed that B. subtilis AK function was essential to cell survival. Therefore cell growth at non-permissive temperatures was expected to be linked with restoration of AK function at elevated temperatures. Although the anticipated result was the accumulation of thermostabilizing point mutations in the adk gene, by no means was this the only possible outcome.

Perhaps the easiest way for recombinant NUB3621-R:ThEV cells to alleviate their AK deficiency would be to up-regulate gene expression of the B. subtilis adk. In fact, only few changes in the promoter sequence for β-galactosidase are sufficient to make this enzyme constitutively expressed (Calos, 1978; Calos and Miller, 1981), which has been observed for E. coli cultivated under lactose-limited conditions in a chemostat (Horiuchi et al., 1962). However, our weak link experiment was designed to make such a strategy unlikely. In order to achieve higher expression levels of a thermolabile AK the mutant would have to up-regulate the expression of all genes in the S10-spc-α cluster. This cluster harbors a large number of ribosomal and other essential genes and are thought to be under tight control (Breitling et al., 1994; Li et al., 1997; Suh et al., 1996).
Altered expression levels for these genes have been shown deleterious in other systems (Allen et al., 2004). Moreover, immediately upstream of adk lies the coding sequence for another essential gene, secY. The chance of altering secY's internal sequence into a strong promoter and at the same time keeping its coding capabilities is likely to be quite small.

Another possible scenario would be the up-regulation of chaperone proteins in recombinant cells. Many chaperones facilitate the refolding of heat-denatured proteins (Lund, 2001). Refolding of AK would then extend the in vivo half-life of the enzyme and perhaps allow the cells to grow at elevated temperatures. With the turbidostat temperature constantly, raising the percentage of unfolded AK would also be higher following every temperature increment. It would then become unlikely that chaperones, alone, to promote AK activity through refolding of the enzyme, although they may delay the onset of the phenotype.

Adenylate kinase is one of the many members of the ribonucleotide monophosphate kinases (rNMK) family of proteins. Enzymes in this family catalyze the phosphorylation of monophosphate nucleotides and include adenylate, uridylate, cytidylate, guanylate and thymidylate kinases (Leipe et al., 2003). It is therefore possible that a different rNMK protein could accumulate point mutations in its active site in order to convert it into a functional adenylate kinase. Indeed a double mutant of mouse guanylate kinase (GMK) has been shown to complement AK deficiency in yeast (Stolworthy and Black, 2001). Several reasons make such strategy impractical in our system. First, this double
mutant guanylate kinase showed very low affinity for adenylate and probably required high expression levels in order to complement AK deficiency in yeast. Second, the observed single mutations in GMK by themselves could not turn guanylate kinase into a functional AK (Stolworthy and Black, 2001). Therefore both mutations would have to occur at the same time, a very low probability event. More importantly, the double mutant enzyme was no longer a functional guanylate kinase and GMK function has been shown essential for other systems (Konrad, 1992).

Evolution of the weak link was likely to follow the easiest path in the adaptive landscape. The strong selective pressure and the experimental design therefore favored the accumulation of thermostabilizing point mutations in \textit{B. subtilis} AK. Indeed, that was the observed outcome. Although the above mentioned strategies were not directly tested, their chances of happening are quite small and it is unlikely that they would interfere with the observed results.

8.2 Evolution of \textit{B. subtilis} AK in NUB3621-R:ThEV

The current evolutionary theory has greatly benefited from studies at the molecular level. Experiments with microorganisms have been used to test hard to prove hypothesis, like the role of selection, chance and history during evolution (Elena and Lenski, 2003; Travisano et al., 1995; Wichman et al., 1999). They have also shed light into new mechanisms by which evolution can work, such as punctuated equilibrium through clonal interference (Elena et al., 1996).
The study of evolution can only be achieved through the use of populations and should be pursued through quantitative methods. The current work satisfies both prerogatives by cultivating the model microorganism in a fermentor. NUB3621-R:ThEV cells (Chapter 7) were grown to very high densities (~ 10^9 cells per ml) in a small fermentor (1.2 liters). Cells were kept at steady-state growth conditions in midlog throughout the duration of the experiment by monitoring their optical density (which turned the fermentor into a turbidostat).

Growth of the host organism in the turbidostat was started at a permissive temperature, 55°C, and ended at the non-permissive temperature of 70°C. In order to allow sufficient time for point mutations to accumulate, the temperature was incremented by 0.5°C every day. During the evolution of approximately 1,000 generations of NUB3621-R:ThEV, approximately 300 individuals were screened following the strategy showed on Figure 8.1.

8.3 Evolution of NUB3621-R:ThEV was punctuated

Darwinian evolution occurs by gradual and accumulative fitness improvements in an ancestral background. The Darwinian expectation for AK evolution in NUB3621-R:ThEV cells would be to observe the accumulation of many, slightly stabilizing point mutations in the parental *B. subtilis adk*. It was quite remarkable that the population analysis of NUB3621-R:ThEV population did not exhibit such pattern.
Figure 8-1 - The two employed strategies to follow NUB3621-R:ThEV evolution in the turbidostat. 12 ml culture samples were withdrawn every 24 hour and frozen as glycerol stocks (10% v/v) at -80°C. Stocks from the indicated temperatures were diluted in liquid TS media and spread onto solid media containing antibiotics (Rifampicin 5 g/l and Chloramphenicol 7 g/l). Cells were incubated one night at the "screened at" temperatures. Genomic DNA was isolated from individual colonies and the adk gene amplified through PCR with specific primers. PCR products were purified and submitted to DNA sequencing analysis.
Over a temperature range of 15°C and a time span encompassing more than 1,000 generations, *B. subtilis adk* accumulated at most two point mutations. The first single mutant, Q199R, became so predominant in the population that it was the parent sequence for all subsequently occurring mutations; resulting in five double mutants: Q199R/G214R, Q199R/G213E, Q199R/T179I, Q199R/A193V and Q199R/Q16L. Therefore evolution of *B. subtilis* AK favored the appearance of a second mutation in the Q199R background rather than the accumulation of a third change in one of the five observed double mutants.

Darwinian evolution of NUB3621-R:ThEV would predict gradual changes in the composition of the population. In striking contrast, the population dynamics of NUB3621-R:ThEV displayed two long periods of stasis interrupted by a sudden burst in diversity (Figure 8.2). The single mutant Q199R appeared early on the experiment and dominated the population until 61°C. At 62°C the diversity burst was first detected by the appearance of two double mutants, Q199R/G213E and Q199R/Q16L. The decline of the single mutant Q199R was evident at 63°C. By this temperature all five double mutants were present in the turbidostat. Diversity in the population was greatly reduced at 65°C, where only two mutants could be readily detected; the double mutants Q199R/A193V and Q199R/Q16L. By the end of the experiment, at 70°C, the prevailing mutant was Q199R/Q16L.

Sudden decreases in genetic diversity are usually attributed to population bottlenecks. These are often a concern in microbial evolution experiments because the next generation of the population following the
Figure 8-2 - B. subtilis AK population dynamics displayed a punctuated pattern. NUB3621-R:ThEV cells were grown at steady state conditions in a small turbidostat from 55°C to 70°C, temperature was raised 0.5°C per day. Cell population was kept at 10⁸ cells/ml. Culture samples were withdrawn every 24 hour and used to isolate single colonies for genomic DNA analyses. Total number of individual adk genes sequenced at each temperature is shown (n) and frequencies for each mutant are reported (f). Because there was no observable changes in the population from 56°C to 60°C this interval was omitted. The same color scheme will be used for all figures to identify individual B. subtilis adk mutants (x-axis).
bottleneck will be the product of a random genetic drift and not selection. In our experiment the population was kept at steady-state growth conditions for the most part of the experiment. When technical problems prevented the maintenance of steady-state conditions, the fermentor was restarted using 10 ml of a frozen glycerol stock obtained from the nearest temperature increment. Selection was only resumed once the reanimated population reached mid-log. Therefore the observed reduction in diversity at 65°C was unlikely to have been influenced by genetic drift. Indeed head to head competition experiments between different cell lineages and biochemical analysis of the mutant enzymes showed that the observed population distribution during NUB3621-R:ThEV evolution was the result of differences in fitness due to differences in AK stability (see below).

Alternate periods of stasis and diversity bursts are reminiscent of Gould’s punctuated equilibrium (Eldredge and Gould, 1972). Such phenomenon has been reported before in microbial populations and has been attributed to the rare appearance of more fit mutants and the time needed for them to accumulate in the population (Elena et al., 1996). Previous reports on punctuated microbial evolution were not supported by genetic or biochemical evidence. Therefore our system provided the first opportunity to test a molecular origin for punctuated equilibrium.

It is important to mention that the punctuated pattern was only observed because samples had enough time resolution. If samples were analyzed every 5°C the diversity burst would have been overlooked and the population was likely
to have been composed of a single and two double mutants what would have made the punctuated pattern less evident.

### 8.4 Punctuated evolution under strong selective pressure

In our experiment, NUB3621-R:ThEV cells were clearly under a strong selection regime. Because of the sharp temperature sensitive phenotype displayed by NUB3621-R:ThEV cells, variants that did not accumulate stabilizing mutations would quickly disappear from the population. Indeed attempts to increase the temperature increment to \(1^\circ\)C per day resulted in the complete collapse of the population at low temperatures (>60\(^\circ\)C).

If adaptation is the most important evolutionary force, “life’s tape” is expected to repeat itself (Gould, 1989). In our system the strong selective force clearly favored adaptive changes (see below). Indeed during an independent turbidostat run (55 to 65\(^\circ\), at 0.5\(^\circ\)C per day) some of the same mutants were observed. Although not extensively screened as the original run, this partial run revealed that the single mutant Q199R and the double mutants Q199R/A193V and Q199R/Q16L were present in the population.

Previous reports on punctuated evolution of microbial populations employed mild selective forces. *E. coli* cells were grown under limiting conditions and selection favored faster growers (Elena et al., 1996). Under a strong selective pressure regime the expectation is that the diversity in the population will be reduced. Therefore the observation of a diversity burst and a punctuated
pattern of evolution were not obvious outcomes at the onset of the experiment. Our results showed that a punctuated pattern can be observed even in the presence of a strong selective pressure.

8.5 NUB3621-R evolution; first come first served

The punctuated equilibrium observed in *E. coli* populations was thought to be caused by the waiting period between the appearance of strong beneficial mutations and their sweep through the population (Elena et al., 1996). Because these previous studies were performed with whole genomes the exact instant that a strong beneficial mutation was produced could not be determined. On the other hand, our system provided the opportunity to investigate when later favored genotypes first appeared in the population. We accomplished this by performing an “early” screen for thermostable strains. Glycerol stocks originally isolated from various time points during the evolution of NUB3621-R:ThEV were subsequently plated at 60, 65 or 70°C, temperatures 3°C to 15°C in excess of the original temperature of isolation (Figure 8.1).

Population analysis from the “early” screen revealed two striking findings (Figure 8.3). First, that at very low frequencies a Q199R single mutant could be found growing at 65°C. This finding contrasted with later observations that 65°C was a non-permissive temperature for this single mutant. Survival of Q199R at 65°C may be attributed in part to one of the alternative evolutionary strategies available for NUB3621-R:ThEV cells (section 8.1). Nevertheless this altered
Figure 8.3- Temperatures where *B. subtilis* AK mutants first appear in the population. Glycerol stocks from the highlighted temperatures were diluted and grown on solid media as indicated on Figure 8.1. Frequencies for individual mutants at the indicated temperatures are shown.
single mutant Q199R happened at extremely low frequencies (3.3 x 10^{-6}). The probability of this mutant to be used as parent genotype for the double mutations in the adk gene was likely to be quite low. More importantly, this mutant was less fit than the double mutants and was completely absent from our “early screen” at temperatures higher than 59°C.

The second important observation was that the “early” screen indicated Q199R/A193V and Q199R/G213E as the first double mutants to appear in the population at 59°C. Moreover, the “early” screen showed that from 59 to 62°C the double mutant Q199R/A193V was the most abundant variant that could grow at 65°C (n=38) (figure 8.3). Q199R/A193V frequency in the population at these temperatures rose from 10^{-6} at 59°C to 10^{-3} at 62°C. At such low frequencies, it was not surprising that Q199R/A193V was not observed during the regular screen of the turbidostat population at 62°C (Figure 8.2). On the other hand, it was quite intriguing that mutant Q199R/Q16L was completely absent from the “early” screen from 59°C to 62°C because the “regular” screen showed this mutant to be first present at 62°C and we know it is able to grow at 65°C (Figures 8.2 and 8.3). These contrasting scenarios between “early” and “regular” screens were likely to have emerged due to small sampling sizes. Therefore the frequencies obtained from both sequencing screens may display high standard deviations. Nevertheless, the overall composition of the population could be determined and was not likely to have been influenced by the small sampling size. The fact that no other mutations could be found from 55 to 61°C even after 65 adk genes were sequenced is a strong indication that Q199R was indeed the
dominant mutant at these temperatures. Likewise, the fact that at temperatures higher than 65°C (n=82) only the double mutants Q199R/A193V and Q199RQ16L could be isolated was used as a strong evidence that these mutants were the most abundant in this temperature range.

Our “early” screen determined that the double Q199R/A193V prevailed among the “early 65°C growers” from 59 to 62°C (n=38). This early start may have conferred to Q199R/A193 a substantial advantage over later arriving mutants and may explain its persistence in the population until the final struggle with Q199R/Q16L for domination over the turbidostat. Because the frequency of a neutral mutation in a population is only dictated by its initial frequency (Kimura, 1969), Q199R/A193V persistence can be credited to its early start when compared to other equally or more fit mutants. Indeed, Malthusian competition experiments supported this hypothesis.

8.6 Random neutral evolution in NUB3621-R:ThEV

Kimura’s neutral theory states that mutations are neutral or quasi-neutral when they first appear in the population (Kimura, 1968). The Darwinian fitness (W, Chapter 2) of individual cell lines was estimated by Malthusian competition (Lenski, 1991). Because the single mutant Q199R was the parent strain for all other mutations this cell lineage was chosen as the reference strain. W values lower than 1 indicates that the Q199R is less fit than the competing strain.
Conversely, if $W$ is higher than 1 then the competing strain is less fit than the reference strain.

Competition experiments were performed at 50 and 58°C because at these temperatures both the single and the double mutants can grow. Liquid cultures were started with equal amounts of the desired mutants. Cells were allowed to grow for 8 hours at the investigated temperatures, diluted and plated at permissive and non-permissive temperatures. Colony counts were used to calculate $W$ values. It is important to keep in mind that the fitness values of individual mutants were calculated at "static" selective pressures while in the turbidostat the temperature was continuously increasing.

As expected, the single Q199R mutant was found to be less fit than the double Q199R/Q16L at higher temperatures ($W=0.92 \pm 0.086$; $T=58^\circ$C) and more fit than the same double mutant at lower temperatures ($W=1.035 \pm 0.144$; $T=50^\circ$C). These results were in good agreement with Kimura’s expectation of quasi neutral mutations. The same Malthusian competition experiment was performed with two double mutants at 58°C. The reference strain for this experiment was Q199R/A193V and it was shown to be less fit than Q199R/Q16L at this temperature ($W=0.92 \pm 0.020$; $T=58^\circ$C). This experiment showed that at its originating temperature, 58°C (Figure 8.2), Q199R/A193V is already less fit than the Q199R/Q16L mutant. Therefore it is quite safe to say that the predominance of Q199R/A193V in the population at intermediate temperatures is due to its sheer numbers in the population and that mutations were selectively neutral when they first happened. Another important implication from the
competition experiments was that the punctuated pattern of evolution was caused by the waiting period for a more fit, but rarer mutation (see below) to sweep through the population, as proposed before (Elena et al., 1996).

The most fundamental tenet of evolutionary theory is that "mutation is random with respect to its adaptive consequences for individual organisms" (Sniegowski and Lenski, 1995). A simple experiment was performed to verify that mutations in our weak link approach happened before selection. The single mutant Q199R was cultivated in 100 ml of liquid media for one night at 50 °C. The liquid culture was diluted and plated at the non-permissive temperature of 65°C. After a 24 hour incubation period, cells growing at 65 °C could be observed with a frequency of $10^{-6}$. DNA sequence analysis revealed that the temperature resistant cells carried a second point mutation in the Q199R background, T179I. This double mutation had been isolated before and shown to confer AK stability (see below). Because Q199R cells were cultivated in liquid cultures at a permissive temperature this experiment showed that mutations in our weak link system are indeed random with respect to selection.

8.7 NUB3621-R fitness is due to AK stability

Fitness values for different genotypes ultimately dictated their frequencies in the population. Although NUB3621-R:ThEV's whole genome contributed for its overall fitness at different temperatures, the stability of its AK was expected to be the major player. Linking the organism fitness with its enzyme stability would
strongly suggest that the evolutionary events leading to the punctuated evolution of the organism were due to the same events behind evolution of the enzyme. The following sections determined that AK stability was indeed behind fitness differences between the strains.

One of the major advantages of our single gene system was that the biochemical and structural basis of temperature adaptation for *B. subtilis* AK could be investigated. All isolated AK mutants were cloned, expressed and purified from an *E. coli* system. The thermostability of isolated AK variants was assessed by heat denaturation curves followed by circular dichroism (CD). CD signal can be used to follow changes in the secondary structure content of a protein as a function of increasing temperature. The resulting temperature denaturation curve can be used to estimate the temperature value where 50% of the enzyme is in its native configuration (*T_m*) (Woody, 1995). The strongest CD signal for all mutant AKs was observed at 220 nm. This wavelength was thus employed throughout this work to monitor the secondary structure content of various AK enzymes. A solution of 20 μM enzyme in 10 mM potassium phosphate buffer pH 7.4 was heated at a rate of a 1°C / minute. Readings at 220 nm were taken at each temperature increment. The resulting denaturation curves obtained for three independent experiments were used to estimate the protein's *T_m*. 
As expected, all the selected AK variants were found more stable than the wild-type protein (Figure 8.4) and showed that AK stability correlated with the organism fitness. As described previously, Q199R is less fit than Q199R/Q16L at 58°C. At this temperature Q199R AK is approximately 75% denatured while the double mutant enzyme is less than 25% unfolded. The same trend could be observed for the competition experiment between the two double mutants. At 58°C Q199R/A193V is about 50% denatured while Q199R/Q16L is only marginally destabilized. Therefore the lower fraction of denatured AK could again be correlated with higher W values.

Another yet interesting result was that NUB3621-R:ThEV strains expressing less stable enzymes were shown more fit than Q199R/Q16L at 50°C. At this temperature all enzymes were only marginally denatured (Figure 8.4) and the slight advantage of the single mutant may be due differences in enzyme activity. A good indication that this may be the case is that thermostable enzymes are usually less active at lower temperatures, as shown by the kinetics experiments performed for the wild-type B. subtilis and G. stearothermophilus enzymes (Figure 7.5). Nevertheless, enzyme kinetic experiments are needed to confirm this hypothesis. Another important point to make is that AK heat denaturation was irreversible and prevented a strict thermodynamic treatment of the data. Chemical denaturation and protein folding experiments are on their way in order to obtain true thermodynamic constants.

Together the CD and the Malthusian competition results indicated that the fitness of the mutant strain was indeed correlated with the stability of its AK
Figure 8-4 - Evolved *B. subtilis* AK variants are more thermostable than the wild-type enzyme. Temperature denaturation profiles for *B. subtilis* wild-type and mutant AKs as followed by changes in CD signal at 220 nm as a function of temperature (raised at 1°C per minute). Curves were fitted assuming a single state transition (F→U) and smoothed using polynomial regression and weights computed from the Gaussian density function (Sigma Plot); $T_m$s (in parenthesis) were calculated from the fitted curves at 50% denaturation. CD traces were acquired from 20 uM protein samples in 10 mM potassium phosphate pH 7.2. Data shown is the average of three independent experiments.
enzyme. Therefore it is a valid assumption to associate the punctuated equilibrium found for NUB3621-R:ThEV cells to the same evolutionary forces that shaped evolution of AK.

8.8  *B. subtilis* AK evolution was punctuated

The CD results demonstrated that the Darwinian expectation of gradual changes in fitness was once again frustrated by changes in protein stability that were neither linear nor gradual. The first point mutant, Q199R, increased the wild-type T_m by 1.7°C. Mutations G214R, G213E, T179I and A193V in the Q199R background increased the single mutant’s stability by 5 to 8°C, while accumulation of Q16L increased the enzyme’s stability by almost 14°C (Table 8.1).

Presence of few, large jumps in adaptive space are reminiscent of Gould’s punctuated equilibrium (Gould and Lewontin, 1979). Perutz was the first to argue that evolution at the single protein level should display a punctuated pattern because although the chance of hitting any protein position is likely to be similar, the fitness increments provided by such changes are not (Perutz, 1983). Indeed jumps in AK stability were observed because not all sites are structurally equivalent.

Accumulation of two point mutations can result in additive or cooperative effects in protein stability. The impact of the second mutation in the Q199R
background was investigated by producing the single AK mutants A193V and Q16L in the wild-type background and testing their temperature stabilities (Figure 8.5). These experiments confirmed that different mutations contributed differently to *B. subtilis* AK stability.

Mutations in Q199R/Q16L AK were found to be slightly cooperative. The individual AK mutant Q16L increased the wild-type stability by 9.6°C, almost 8°C more stable than the Q199R single mutant (1.7°C) (Table 8.1). Alone Q16L is more stabilizing than any other double mutant. Added together, the individual contributions of Q199R and Q16L are slightly less stabilizing than contribution of the double mutant (11.3°C vs 13.8°C). The additional 2.5°C of stability resulted from the synergistic nature of these point mutations.

On the other hand, the single A193V mutant was found to be only slightly more stable than the wild-type protein and less stable than the Q199R protein. The algebraic contributions of the single mutations Q199R (1.7°C) and A193V (0.2°C) to the wild-type protein are almost 3°C short of the observed stabilization for the double mutant Q199R/A193V (1.9°C vs 5.0°C). As with Q199R/Q16L there is a positive cooperation between the two individual mutations, but for Q199R/A193V this synergy appears to be accentuated.
Figure 8-5 - Single point mutations had different impacts on Q199R background. Temperature denaturation profiles for *B. subtilis* wild-type and single mutant AKs Q199R, A193V, Q16L and the relevant double mutants Q199R/A193V and Q199R/Q16L. Experimental conditions were identical to the ones on Figure 8.4.
8.9 History and chance dictated the evolution of *B. subtilis* AK

The use of a single gene system provided a unique opportunity to contrast the roles of adaptation, history and chance in evolution. Because selective pressure was kept high throughout the experiment all mutations were clearly adaptive and led to variant AKs with higher $T_m$s (Figure 8.4). Indeed no silent or neutral mutations were observed.

The strong synergy between the Q199R and A193V individual point mutations evoked the importance of historical aspects during the weak link evolution. As opposed to the Q16L, the A193V mutation would not be expected to be present by itself because it required the accumulation of a previous change in order to effectively increase the stability of the enzyme.

If evolution of AK was solely driven by adaptive constraints, Q16L could have been expected to be the dominant mutation instead of Q199R. The fact that Q16L is not the first mutant is clearly a combination of historical constraints and chance. The previous history of codon 16 dictated that it could only be made into a stabilizing Leu through an A to T transversion, an event with lower probability than the A to G transition responsible for Q199R. Q16L was indeed the only observed transversion.

The Q16L mutation reinforced the importance of the non-random pattern of mutations during evolution. Transitions, especially C to T, are the most frequent mutational events. Indeed five out of the six observed mutations were produced by transitions; two were C to T, two were G to A and one was an A to
G. Although the host organism G. *stearothermophilus* NUB3621-R has been reported as a *dam* and *dcm* positive strain (Wu and Welker, 1989) none of the observed mutations occurred at possible methylation sites and therefore are unlikely to be hot spots for mutations. The fact that most mutations occurred at the 3′OH end of the gene was not attributed to a higher probability of mutation in that region, but to structural constraints (Chapter 9).

It is an important point to make that in five out of the six observed variants, the frequency of their mutational event could be inversely correlated with their fitness gains (Table 8.1). Double mutants conferring 5-8°C of protein stability were produced by more frequent mutational events. G213E and G214R accumulated in the Q199R background through a G to A transition and A193V and T179I were generated by a C to T transition. The first single mutant, Q199R, was produced by an A to G transition. The only transversion (A to T) observed produced the most stable double mutant Q199R/Q16L.

The results presented in this chapter indicated the presence of punctuated evolution in a population under strong selective pressure. Results from DNA sequencing analysis, enzyme stability and Malthusian competition experiments showed that the origin of the punctuated pattern was due to the different probabilities of accumulating beneficial mutations resulting in dissimilar fitness gains. Therefore the punctuated evolution of NUB3621-RThEV was caused by the chain of events leading to *B. subtilis adk* heat adaptation, which depended on the gene’s previous history and chance events. Because all evolutionary
processes are likely to be influenced by adaptation, chance and history, punctuated equilibrium may be a frequent trend during evolution.

Table 8-1 - Stability gains and mutational patterns for *B. subtilis* AK variants

<table>
<thead>
<tr>
<th>single mutant</th>
<th>mutation type</th>
<th>transition</th>
<th>transversion</th>
<th>ΔT_m (°C) (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q199R</td>
<td></td>
<td>X</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>Q16L (^b)</td>
<td></td>
<td></td>
<td>X</td>
<td>9.6</td>
</tr>
<tr>
<td>A193V (^b)</td>
<td></td>
<td>X</td>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td>Double mutant (^c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q199R / G213E</td>
<td></td>
<td>X</td>
<td></td>
<td>4.8</td>
</tr>
<tr>
<td>Q199R / A193V</td>
<td></td>
<td>X</td>
<td></td>
<td>5.0 (0.2)</td>
</tr>
<tr>
<td>Q199R / T179I</td>
<td></td>
<td>X</td>
<td></td>
<td>6.5</td>
</tr>
<tr>
<td>Q199R / G214R</td>
<td></td>
<td>X</td>
<td></td>
<td>7.9</td>
</tr>
<tr>
<td>Q199R / Q16L</td>
<td></td>
<td>X</td>
<td>X</td>
<td>13.8 (9.6)</td>
</tr>
</tbody>
</table>

\(^a\) difference in T_m between the mutant and wild-type *B. subtilis* enzyme. Numbers in parenthesis indicate the stability gains for the respective single mutants.

\(^b\) single mutants Q16L and A193V were not observed in the turbidostat population.

\(^c\) all double mutants accumulated in the Q199R background
Chapter 9 - The biochemical and structural basis of adaptation

As discussed previously *B. subtilis adk* evolution is constrained by two independent landscapes. The mutational landscape limits evolution because not all amino acid changes are equally attainable from single point mutations and the structural landscape restrains evolution because not all residues are functionally equivalent. Restrictions from both landscapes were present during *B. subtilis* AK evolution because the enzyme function was essential to the host organism and, as seen on Chapter 8, resulted in its punctuated evolution.

These constraints produced two striking patterns. First, that four out of the six mutations occurred at absolutely conserved positions between *B. subtilis* and *G. stearothermophilus* AKs. Second, five out of the six mutations were located at the C-terminal region of the enzyme.

9.1 Expected patterns

Primary sequence comparison between closely related mesophilic and thermophilic pair of enzymes is commonly employed to identify putative stabilizing positions. Mutating the variant site in the mesophilic protein into the thermophilic residue has been shown to be a valid strategy to increase protein thermostability (Perl and Schmid, 2001).
The obvious choice for *B. subtilis* AK primary sequence analysis is with the thermophilic enzyme from *G. stearothermophilus*. The two enzymes are highly identical (76%) and there is structural information available for both of them (Bae and Phillips, 2004; Berry and Phillips, 1998). Although the archaeal AK enzymes have been shown to be extremely thermostable they are not appropriate for this type of analysis due to their trimeric nature (Criswell et al., 2003) (Vonrhein et al., 1998) and low sequence identities with other AKs. In fact, the archaeal enzymes constitute an independent class of the AK family (Ferber et al., 1997).

Adenylate kinase is a relatively small protein with 217 residues. Approximately 35% of them are involved in substrate binding or catalysis and, not surprisingly, these residues are extremely well conserved among different AKs. Most of these invariant amino acids locate to the P-loop (residues 7-15); the AMP-binding (residues 35-59) or the LID domain (residues 128-159), but some can be found in the CORE domain, (arginines 88, 127, 160 and 171; Asp162; Asp163 and Gin92) (Figure 9.1). Together these residues ultimately bestow to AK its unique capability of maintaining adenylate homeostasis in living cells. Single point mutations in these regions were likely to alter AK’s finely-tuned kinetic mechanisms and, as expected, were not observed.

The remaining (147) residues of AK’s CORE domain are not involved in catalysis or ligand binding and may impose less stringent constraints during the evolution of *B. subtilis* AK. Protein alignment between *B. subtilis* and *G. stearothermophilus* revealed that there are 40 variant positions in the CORE
domain and that they are equally distributed between the N-terminal (17 changes from β1 to β4) and C-terminal region of the enzyme (23 changes from β4 through α9) (Figure 9.1A and 9.1B). Because *B. subtilis* and *G. stearothermophilus* belong to very distant phylogenetic groups (Studholme et al., 1999) it is important to keep in mind that some of their variant residues may be the product of genetic drift and not adaptation to elevated temperatures.

Although useful for protein stabilization studies, this type of analysis completely disregards the presence of mutational constraints during evolution. Considering only the structural landscape, these 40 variant structural residues would be the most likely positions chosen for changes during the *B. subtilis* AK evolution in NUB3621-R:ThEV cells. Moreover, if all variant positions were structurally equivalent, mutations were expected to accumulate in both the N-terminal and the C-terminal halves of the enzyme.

### 9.2 Primary sequence constraints dictated AK evolution

In our system, *B. subtilis adk* evolution relied solely on spontaneous mutational events, which favors transitions over transversions. Out of the forty variant sites found between *B. subtilis* and *G. stearothermophilus* AKs, fourteen would not be attainable by *B. subtilis adk* through single point mutations. Twenty-one of the remaining sites would require a transversion event and, therefore, had
Figure 9-1 - B. subtilis AK evolution was constrained by primary and tertiary structures. A - Protein sequence alignment between the weak link B. subtilis AK and the thermophilic enzyme from G. stearothermophilus (G. stearo) showing the highly conserved P-loop (salmon), AMP-binding (orange) and LID (slate) domains. The more variant CORE domain is shown in green. Conservative changes in the CORE domain are shown in yellow, non-conservative in red. Arrowheads indicate isolated mutations during the in vivo evolution of B. subtilis AK. Q199R is indicated in purple, G213E in blue, A193V in green, T179I in yellow, G214R in orange and Q16L in red. B - Atomic model of B. subtilis Q199R AK showing the domains and variant residues highlighted as in A; the zinc atom is shown in blue. C - Same model depicted in B, but with the location of the isolated mutations highlighted as the arrowheads in A.
a lower chance to be observed. Out of the five sites that required a more likely transition event, four introduced conservative changes and were less likely to be involved in protein stability. Therefore the stability landscape available to *B. subtilis* AK through single point mutations was likely to favor changes at conserved positions. Indeed four (Q16, Q199, G213 and G214) out of the six changes observed during *B. subtilis* evolution were at completely conserved sites between the mesophilic and thermophilic enzymes (Figure 9.1C).

Because *B. subtilis* AK evolution was constrained by structural and mutational landscapes, primary sequence alignments would only succeed at positions where both of these landscapes coincided. Position 179 is a *bona fide* example of such phenomenon. Residue 179 is a Met in the thermophilic enzyme. Previous structural comparisons between the two enzymes indicated that a hydrophobic residue at this position would be stabilizing (Bae and Phillips, 2004). In fact, the only hydrophobic residue that codon T179 can yield through a single nucleotide change is the observed T179I mutation.

Residue 193 is the other variant position where a mutation was observed to accumulate during the *in vivo* evolution of *B. subtilis adk*. In the thermophilic enzyme arginine is found at this position. No single nucleotide change in *B. subtilis* codon Ala193 can produce an arginine or any charged residue. This mutation could be considered as a suboptimal compromise between the two adaptive landscapes. But, as will be discussed later (section 9.4), there is structural evidence to justify the increased stability promoted by a hydrophobic residue at this position.
Together with pair wise sequence comparisons, site-directed mutagenesis is yet another method routinely employed to produce thermostable protein variants (Ahern et al., 1987). One could try and use this method to recreate the path that B. subtilis AK evolution followed in our system. Because site-directed mutagenesis is not constrained by a mutational landscape she or he would have the opportunity to explore the entire adaptive landscape. If this very brave graduate student decided to use only the 40 variant positions between B. subtilis and G. stearothermophilus AKs she or he would have to make 1,026 single point mutations (19 x 40). Because selection in an in vitro system is very much like a breeding process (you get to pick which parent to use for the next round of selection), she or he would probably choose the most stable out of the 1,026 mutants. And we know that this would not have been Q199R. Therefore the real evolutionary path followed by B. subtilis AK in our fermentor would have been broken at the very first step.

Knowledge of the complete adaptive landscape available for B. subtilis AK would be extremely useful, although certainly not worth the effort of our hypothetical (and miserable) graduate student. This information, together with the known mutational landscape for the gene, would enable us to build a powerful predictive model. Although impossible to be attainable in real life, our lab is building the complete adaptive landscape for B. subtilis AK in silico. Molecular dynamics simulations will be used to test the stability of all 1,026 mutants in relation to the wild-type protein.
9.3 Biochemical basis for AK stability

None of the observed mutations (Q199R and the five double mutants Q199R/G214R, Q199R/G213E, Q199R/T179I, Q199R/A193V and Q199R/Q16L) can be considered conservative, as far as protein stability is concerned. In summary, changes for the selected AK mutants favored higher residue volume (all) and hydrophobicity (Q16L, A193V and T179I); reduction in the number of uncharged polar residues (T179I, Q16L and Q199R) or an increased number of charged side chains (G213E, G214R and Q199R). All these changes are consistent with previous primary sequence comparisons between mesophilic and thermophilic homologue pairs (Haney et al., 1999).

It is an important point to make that although useful, primary sequence analyses can be misleading (Bohm and Jaenicke, 1994). For example, reduction in the number of uncharged polar residues, especially Gln and Asn, is often associated with protein stability (Tamakoshi et al., 2001; Yano et al., 2003). Nevertheless, the thermophilic AK from *G. stearothermophilus* displays a higher number of (Gln + Asn) residues than the enzyme from *B. subtilis*. Indeed, algorithms to assess protein thermostability from the ratio of uncharged polar to charged residues (Farias et al., 2004) would incorrectly predict that *B. subtilis* AK is more stable than its thermophilic counterpart from *G. stearothermophilus*. 
9.4 Structural basis of adaptation

Since temperature was employed as selective pressure, the final products of *B. subtilis* AK evolution were expected to be thermostabilized proteins. Therefore I determined the atomic structure of one of the isolated AK mutants, Q199R, via protein X-ray crystallography in order to establish the structural features behind the mutant’s increased stability. Protein crystals also were obtained for other mutants, but their atomic structure determination was hindered by poor diffraction quality. Extensive crystal optimization screens are being performed in order to investigate the remaining mutants at the atomic level.

Co-crystals of Q199R bound to the inhibitor Ap5A (Figure 9.2) were readily obtained using similar conditions to the ones previously employed to crystallize the wild-type protein (Bae and Phillips, 2004) and diffracted to high-resolution (1.8Å). Q199R atomic structure was determined by molecular replacement using as search model the wild-type enzyme from *B. subtilis* (PDB 1P3J) (Bae and Phillips, 2004). Relevant statistics for the final Q199R protein model solved at 1.8Å can be found on Table 9.1. Low R values, the absence of forbidden phi and psi angles or bond length violations indicated that the final model was of good quality (Table 9.1). Moreover, OMIT maps revealed electron density for all but the C-terminal residue. The zinc and the Ap5A ligands could be clearly defined on electron density maps even at high contours (I=2σ). More importantly, electron density could be found for the mutant residue Arg199 to permitted atomic analyses (Figure 9.3).
Figure 9-2- Protein crystals for the *B. subtilis* AK Q199R mutant. Crystals were obtained by vapor diffusion methods in 50 mM CHESS pH 9.0; 25 mM CaCl₂; 32.5% PEG 1500. 6 µl of a protein-inhibitor solution (18 g/l protein and 3.5 mM Ap5A in 10 mM HEPES pH 7.0) was diluted in an equal volume of mother liquor. Crystals were first visible after 3 days and stopped growing after 5 to 6 days. Diffraction quality crystals could only be obtained after seeding with previously obtained crystals. Drops were allowed to equilibrate for 48 hour prior to seeding.
<table>
<thead>
<tr>
<th></th>
<th>Q199R&lt;sup&gt;a&lt;/sup&gt;</th>
<th>B. subtilis&lt;sup&gt;b&lt;/sup&gt;</th>
<th>G. stearo&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P&lt;sub&gt;2&lt;/sub&gt;_1</td>
<td>P&lt;sub&gt;2&lt;/sub&gt;_1</td>
<td>P&lt;sub&gt;2&lt;/sub&gt;_1</td>
</tr>
<tr>
<td><strong>Cell constants</strong></td>
<td>34.7 Å x 75.1 Å x 77.4 Å, β=98.3°</td>
<td>32.8 Å x 73.2 Å, β=101.6°</td>
<td>41.2 Å x 62.3 Å, β=117.1°</td>
</tr>
<tr>
<td>Molecules per unit cell</td>
<td>2 (216 / 216)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1 (212)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1 (217)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>1.80 (1.86-1.80)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.90</td>
<td>1.96</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>35,105</td>
<td>13,641</td>
<td>12,454</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.0 (88.6)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>96.8</td>
<td>93.1</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)</td>
<td>3.6 (9.6)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.0</td>
<td>10.6</td>
</tr>
<tr>
<td>R (%)</td>
<td>19.00 (20.10)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>19.7</td>
<td>16.1</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>22.84 (25.13)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>24.6</td>
<td>22.8</td>
</tr>
<tr>
<td>RMSD from ideality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bonds (Å)</td>
<td>0.004</td>
<td>0.021</td>
<td>0.012</td>
</tr>
<tr>
<td>angles (°)</td>
<td>1.23</td>
<td>1.2</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Average B-factors (Å)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>main-chain</td>
<td>13.28 (12.37&lt;sup&gt;A&lt;/sup&gt;;14.19&lt;sup&gt;B&lt;/sup&gt;)</td>
<td>27.7</td>
<td>10.9</td>
</tr>
<tr>
<td>side-chain</td>
<td>14.79 (13.94&lt;sup&gt;A&lt;/sup&gt;;15.64&lt;sup&gt;B&lt;/sup&gt;)</td>
<td>36.2</td>
<td>19.4</td>
</tr>
<tr>
<td>solvent</td>
<td>24.1</td>
<td>37.5</td>
<td>28.0</td>
</tr>
<tr>
<td>Ap5A</td>
<td>10.75 (9.6&lt;sup&gt;A&lt;/sup&gt;;11.9&lt;sup&gt;B&lt;/sup&gt;)</td>
<td>21.3</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Ramachandran

<table>
<thead>
<tr>
<th>favored (%)</th>
<th>allowed (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.60</td>
<td>x</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

<sup>a</sup> B. subtilis AK Q199R mutant; values in parenthesis represent the highest resolution shell.
<sup>A</sup> values for molecule A
<sup>B</sup> values for molecule B
<sup>b</sup> values for wild-type B. subtilis AK (Bae and Phillips, 2004)
<sup>c</sup> values for G. stearothermophilus AK (Berry and Phillips, 1998)
<sup>d</sup> number of amino acids accounted for in the electron density
Figure 9-3- Electron density details of *B. subtilis* Q199R AK atomic structure at 1.8 Å resolution. A - detail showing the electron density for the mutant residue Arg199 and neighboring residues. B - detail showing the zinc binding residues found in the LID domain. Only the side-chains of relevant residues are shown. C - detail of the inhibitor Ap5A found in the active site. As seen before in other AK-inhibitor co-structures, the magnesium atom was found coordinated by four water molecules (yellow) and two phosphate atoms. Electron density maps shown in B and C were originated from omit maps at a density level 2 times higher than the background intensity (I=2σ). Electron density in A is from the same map, but at I=1σ. Color code: green for carbon, blue for nitrogen, pink for phosphate, red for oxygen, blue for zinc, purple for magnesium and yellow for waters.
The mutant enzyme crystallized with two protein molecules in its unit cell as opposed to the single molecule found in the wild-type crystal (Bae and Phillips, 2004). Nevertheless both molecules found in Q199R crystals are identical and show a root-mean-square deviation (rmsd) of only 0.46 Å between their Cα traces. This figure is even lower, 0.32 Å, if part of the highly mobile C-terminal helix and LID domains are removed (Figure 9.4). The crystal structure of Q199R is also virtually identical to the one found for the wild-type protein (Figure 9.4). The rmsd between wild-type and the mutant molecules are around 0.7 Å. This number is reduced to 0.44 Å if the entire LID domain is removed from the alignment.

It is interesting to mention that position 199 is a Gln in all mesophilic and thermophilic AK enzymes and therefore was not an obvious choice for a stabilizing mutation. In the wild-type structure for B. subtilis or G. stearothermophilus Gln199 does not seem to play a role in protein stabilization. The only ionic interaction that Gln199 could support in the wild-type protein would be a pair of hydrogen bonds with Asn196. However, the geometry of this interaction is clearly not favorable (Figure 9.5). Moreover, position 199 is exposed to the solvent and a suboptimal hydrogen bond between Gln199 and Asn196 would be prone to constant solvent screening and not be of much stabilizing power (Lockhart and Kim, 1993). Introduction of an arginine at position 199 was expected to promote protein stability via ionic interactions. Arginine's fully charged group is long and flexible enough to reach the fully charged group of Asp207 and facilitate the formation of a strong salt bridge. Moreover, arginines
Figure 9-4- *B. subtilis* Q199R AK mutant displayed an identical overall atomic structure to the wild-type protein. The two molecules found in the unit cell (shown as a blue frame) for Q199R crystals are depicted in green (molecule B) and yellow (molecule A). The *B. subtilis* wild-type enzyme is depicted in slate. Molecule A from the Q199R crystal and the wild-type enzyme were aligned and have a rmsd of 0.70 Å (0.44 Å using 188 out of the 212 residues and excluding the LID domain and the 4 C-terminal residues from the alignment). The rmsd between molecules A and B was found to be 0.46 Å (0.32 Å using 203 out of 216 residues and excluding part of the LID and the C-terminal helix from the alignment). For molecule B the different AK domains are shown in salmon (P-loop), slate (LID), orange (AMP-binding) and green (CORE). The ligands Ap5A (green for carbon, blue for nitrogen, pink for phosphate and red for oxygen) and zinc (blue) are also shown.
can serve as hydrogen donor in up to five hydrogen bonds (Shimoni and Glusker, 1995).

Indeed, the crystal structure of mutant Q199R showed that Arg199 is capable of making several ionic interactions with the surrounding negatively charged residues (Figure 9.5). Three hydrogen bonds with good geometries and distances between 2.59 and 2.76 Å can be formed between Asn196, Arg199 and Asp207. Arg199 N$^\text{II}$ makes two hydrogen bonds with Asp207 O$^\text{61}$ and the main-chain oxygen from Asn196. A third hydrogen bond can be found between Arg199 N$^\text{II}$ and Asp207 O$^\text{61}$. Besides the hydrogen bonds, all fully charged residues from Asp207 and Arg199 are within 5.0 Å of each other and therefore are capable of making reasonably strong salt bridges. Other electrostatic interactions may be present between the partially charged main-chain oxygen from Asn196 and Asp203 and the side chain nitrogen atoms from Arg199. The multiple charge-charge interactions made by Arg199 with other residues in the protein are likely to avoid the problem of solvent screening and therefore make this a stabilizing interaction (Kumar and Nussinov, 2002). More importantly, the above described ionic interactions could easily account for the 1.7°C increase in T$_m$ observed for the mutant protein during temperature denaturation studies described earlier. A unique salt bridge between His31 and Asp 70 has been shown to contribute 3-5 kcal/mol to $\Delta G_{\text{stab}}$ of T4 Lysozyme (Anderson et al., 1990).

Determining the atomic structure for Q199R allowed me to suggest a structural basis of adaptation by the double mutants. Mapping the remaining mutations into the Q199R structure showed that four out of the five observed
Figure 9-5- The mutant Arg199 promotes better ionic interactions than the wild-type Gln199 residue. **A** - location of residue 199 in the superimposed structures of wild-type and Q199R AKs is shown as a deep purple sphere. The protein region shown in details in **B**, **C** and **D** is highlighted in purple. **B** - detail of the highlighted region in **A** showing the side chains for relevant amino acids. **C** and **D** - hydrogen bonds (red) and charge-charge interactions (blue) supported by residue 199 in wild-type (**C**) and Q199R mutant (**D**) AKs. Hydrogen bonds were assigned using a cut-off distance of 3.0Å, charge-charge interactions with a cut-off distance of 5.0Å. In **C** N$^\varepsilon$ from Gln199 is at 2.9 Å from Asp207's O$^{\varepsilon1}$ and the same distance from Asn196's main-chain O. In **D** Arg199 N$^{\eta2}$ is at 2.59 Å from Asp207's O$^{\varepsilon1}$ and at 2.76 Å from Asn196's main-chain O. Arg199 N$^{\eta1}$ is at 2.76 Å from Asp207's O$^{\varepsilon1}$. Color key: Nitrogen: blue; oxygen: red; carbon: slate for wild-type and green for the mutant AK.
mutants could be implicated in stabilizing the C-terminal helix α9. Mutant Leu16 may fit into a hydrophobic pocket made by Ile113, Gly197, Val195 and Val208 and promote a better packing of α9's hydrophobic face (Figure 9.6). This hydrophobic pocket is also observed in G. stearothermophilus AK (Berry and Phillips, 1998).

Although an arginine in G. stearothermophilus AK, introduction of a hydrophobic residue at position 193 may be a more stabilizing mutation. A valine at position 193 may facilitate a better hydrophobic packing of α9 by interacting with Val110, Ile111, Leu211, and Tyr109 (Figure 9.7). This hydrophobic pocket is again conserved among thermophilic and mesophilic AKs (Berry and Phillips, 1998).

Mutations G213E and G214R are located at the very end of α9. Amino acids have different helix-forming capabilities. Due to entropic penalties, Gly is a poor helix forming residue. Large polar residues, like Arg and Glu, on the other hand, have been shown to have higher helical propensities (Blaber et al., 1993; Chou and Fasman, 1978). The stabilization of secondary structures by the introduction of residues with high helical propensities is a well documented strategy to achieve higher stability in proteins (Blaber et al., 1993).

It is interesting to note that although both mutations G213E and G214R are thought to stabilize the same secondary structure through similar means, the T_m's for the double mutants Q199R/G213E and Q199R/G214R displayed a 3°C difference in favor of the former. An explanation for this disparity in protein
Figure 9-6 - Putative structural basis for the higher thermostability of mutant Q16L. Mutant residue Leu16 is capable of occupying a hydrophobic pocket made by Ile113, Val208 and Val195 and Gly197 (not labeled) (B and D) that the wild-type Gln16 residue leaves exposed to the solvent (A and C). The van der Waals surface is depicted for relevant residues. C and D depict the hydrophobic pocket in detail. Atomic coordinates for the Q16L point mutation were produced in PyMol using data from the Q199R crystal structure. Domains and atoms follow the same color key employed for previous figures.
Figure 9-7- Putative structural basis for the higher thermostability of mutant A193V. Mutant residue Val193 is capable of better hydrophobic interactions with Tyr109, Val110, Ile111 and Leu213 (B and D) than the wild-type Ala193 (A and C). The van der Waals surface is depicted for relevant residues. C and D depict the hydrophobic pocket in detail. Atomic coordinates for the A193V point mutation were produced in PyMol using data from the Q199R crystal structure. Domains and atoms follow the same color key employed for previous figures.
stability will have to wait protein structure determination and folding studies to be resolved.

Although all the above mentioned may be implicated in stabilizing α9, there is no strong evidence that this protein region is particularly unstable. Nevertheless, because both the wild-type and the mutant enzymes have identical structures the stabilization promoted by Arg199 is likely to have its origins in a local effect. Position 199 locates to the connecting loop between α9 and β8. Solvent exchange experiments performed with the E. coli enzyme revealed that this loop is highly mobile (Wolf-Watz et al., 2004). Therefore it is possible that together with the other mutations Q199R aid in stabilizing the C-terminal α-helix and promote an enhanced thermostability.

Previous structural studies on AK proteins from B. subtilis and G. stearothermophilus have suggested residue 179 as important for protein stability (Bae and Phillips, 2004). In the atomic structure for G. stearothermophilus AK Met179 fits into a hydrophobic pocket made by Met6, Gly7 and Leu8. For the mesophilic protein, Thr179 cannot occupy the conserved hydrophobic pocket and a water molecule is found instead. The T179I substitution probably mimics the effect of the thermophilic Met 179 and occupies the conserved hydrophobic pocket (Figure 9.8).

It is satisfying to come full circle and realize that the punctuated evolution displayed by NUB3621-RThEV cells could be related to structural features of B. subtilis AK. As discussed above, not all sites contributed equally to protein stability. Investigation of the adaptive landscape available for B. subtilis AK in the
Figure 9-8 - Putative structural basis for the higher thermostability of mutant T179I. Mutant residue Ile193 is capable of filling in a hydrophobic pocket formed by Met6, Gly7 and Met8 (B and D) that the wild-type residue Thr109 leaves exposed to the solvent. C and D depict the hydrophobic pocket in detail. Atomic coordinates for the T179I point mutation were produced in PyMol using data from the Q199R crystal structure. Domains and atoms follow the same color key employed for previous figures.
light of the DNA sequence and protein structure information revealed why most changes in AK occurred at absolutely conserved sites between mesophilic and thermophilic enzymes. Structural analysis of mutant proteins also offered explanations for the different amounts of stability achieved by different mutants and observed during protein stability experiments.

Considering all factors that lead to AK stability it is not altogether surprising to have found a punctuated pattern of evolution in our system. Adaptation of NUB362-R to higher temperatures required the thermostabilization of its AK enzyme. Protein thermostability is dictated by physicochemical laws and primary and tertiary structural constraints that were unlikely to produce a slow and gradual pattern of changes. Moreover, strong selective pressure in the turbidostat did not allow the accumulation of variants conferring only slight $T_m$ increments and, therefore, small fitness gains. Our results taken together indeed support that all these mechanisms played an important role and lead to the punctuated evolution of *B. subtilis* AK.
CHAPTER 10

Conclusions

“They may reveal what happens to a hundred rats in the course of ten years under fixed and simple conditions, but not what happened to a billion rats in the course of ten million years under the fluctuating conditions of earth history. Obviously, the latter problem is much more important.” Simpson, G.G.

Gould has argued that to be a “unit of selection” biological objects must have five properties: birth points, death points, sufficiently stable existence, reproduction and inheritance of parental traits by offspring. Darwin’s biological object of concern was the organism. But, following Gould’s definition, an “evolutionary unit” can vary from a single gene to an entire clade. Because the units of selection are so dissimilar they need not to display the same modes or tempos of evolution (Gould, 1994).

Our results indicated a punctuated pattern of evolution that is more consistent with Gould’s view of evolution than the classic gradualism advocated by the modern synthesis. In our experiment, adaptation was the driving force behind AK evolution. As a consequence, more stable AK enzymes could be isolated. Moreover, repeating the same experiment a second time revealed the accumulation of some of the same AK mutants. Repetition of evolutionary trajectories is an indication that selective force is a critical player during evolution.
Nevertheless, adaptation was restrained by history and chance. Such combination of factors produced the punctuated pattern of evolution observed for *B. subtilis* AK.

Establishing the biochemical and structural basis of adaptation allowed us to confirm Perutz's expectation that because not all sites are structurally equivalent, evolution at the single gene level should be punctuated (Perutz, 1983). Indeed the stability increments observed for *B. subtilis* AK variants isolated from NUB3621-R:ThEV cells were not gradual (Figure 8.4). The punctuated pattern of increased protein stability was produced due to restraints to the possible adaptive landscape of the enzyme. Chance dictated which mutations were more likely to accumulate and structural constraints accepted or rejected these changes as adaptive (Figure 9.1). The history of the evolutionary process was also important for the punctuated pattern of stability because some mutations were shown to be more dependent on previously accumulated changes than others (Figure 8.5).

Physiological (Chapter 7) and Malthusian competition experiments (Chapter 8) demonstrated that the fitness of the organism was a direct result of its AK stability. Therefore, evolution of NUB3621-R cells was dictated by the stability increments of AK. Indeed the temporal distribution of different NUB3621-R:ThEV variants displayed a punctuated pattern (Figure 8.2). Lenski's clonal interference mechanism was at the heart of such a pattern. More frequent mutations (Q199R) had an early start in the population and were able to reach transient fixation, despite their lower fitness values. Subsequent mutations in the
Q199R background followed the same pattern. More frequent double mutants (Q199R/A193V) were able to persist in the population even after the appearance of a more fit, but rarer mutant (Q199R/Q16L).

Gould was the first to point out that evolution needed not to be gradual. His theory of punctuated equilibria reinterpreted the fossil record and argued that the data it holds is more in accord with a punctuated pattern of evolution than with a single gradual line of changes from ancestral to derived forms. Gould's punctuated pattern of evolution would be fueled by historical contingencies and chance events and not only by adaptation (Eldredge and Gould, ; Gould, 1989; Gould and Lewontin, 1979).

Gould's theory of punctuated equilibria was concerned with fossil records and evolutionary processes that led to speciation events. However, a punctuated pattern of evolution has been observed in many levels of complexity; ranging from speciation events in sea urchins (Byrne and Voltzow, 2004) to the geographical distribution of viruses (Nichol et al., 1993). Real time evolution experiments also detected a punctuated pattern of evolution. Investigation of fitness increments in microbial evolution experiments with enough temporal resolution enabled Lenski and co-workers to detect that fitness gains in the population were accomplished in a step-wise manner, akin to Gould's punctuated equilibrium. They have proposed a mechanism, clonal interference, to explain how a punctuated pattern of evolution can emerge in a microbial population (Elena et al., 1996). However, fitness is not an evolutionary unit as defined by
Gould (Gould, 1994) and their work did not establish the genetic and biochemical basis of the observed punctuated pattern.

Our results linked for the first time the punctuated pattern of evolution at the molecular level with a *bona fide* evolutionary unit, the *B. subtilis adk* gene. In our work the genetic and biochemical basis of adaptation at the single gene level were revealed. More strikingly, the punctuated pattern of evolution observed at the single gene level was matched by the same pattern of evolution in a different evolutionary unit, NUB3621-R:ThEV cells. In both units evolution was restrained by adaptation, history and chance; the same factors Gould implicated in the punctuated pattern of evolution observed at (much) higher evolutionary unit levels.
References:


Bacillus (Geobacillus) stearothermophilus Genome Sequencing Project. 
http://www.genome.ou.edu/bstearo.html.


reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg2+ and Mn2+. Proteins 32, 276-288.


