RICE UNIVERSITY

Effect of Protein Thermostability on the Cooperative Function of Split Enzymes.
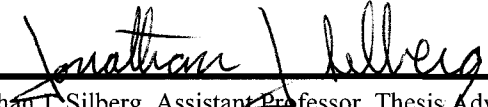
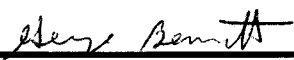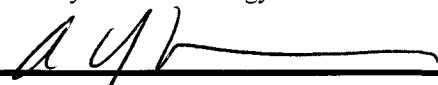by

Peter Q. Nguyen

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
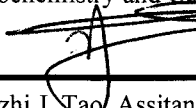
Doctor of Philosophy
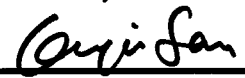
APPROVED, THESIS COMMITTEE:

Jonathan J. Silberg, Assistant Professor, Thesis Advisor
Biochemistry and Cell Biology

George N. Bennett, E. Dell Butcher Professor of Biochemistry
Biochemistry and Cell Biology

Yousif Shamoo, Associate Professor,
Biochemistry and Cell Biology

Yizhi J. Tao, Assitant Professor,
Biochemistry and Cell Biology

Ka-Yiu San, E.D. Butcher Professor of Bioengineering
Chemical Engineering, Bioengineering

HOUSTON, TEXAS
FEBRUARY, 2010

UMI Number: 3421397

UMI®

Dissertation Publishing

ProQuest®

Abstract

Although the effects of mutational events on protein function cannot yet be predicted *a priori*, molecular evolution studies have shown that the tolerance of protein structure and function to random mutation is positively correlated with the thermostability of the protein mutated. To test whether thermostability also influences protein function upon random fission, I have characterized the function of split *Bacillus subtilis* and *Thermotoga neapolitana* adenylate kinases (AK_{Bs} and AK_{Tn}, respectively), enzymes that are required to maintain adenine energy charge. Using libraries of split AK_{Bs} and AK_{Tn} variants, I show that mesophilic and thermophilic AK orthologs can be split at multiple sites into fragments that complement the growth of *Escherichia coli* with a temperature-sensitive AK at 40°C. However, I find that the fraction of split AK_{Tn} variants that function is ~7-fold higher than that observed for split AK_{Bs} variants. I also find that AK_{Tn} can be split within the AMP-binding and LID domains to create functional variants, whereas AK_{Bs} can only be split within the AMP-binding domain. Biochemical and biophysical analysis of one pair of homologous split AK variants reveal that polypeptide fragments derived from the more thermostable AK exhibit greater secondary structure and enzymatic activity, suggesting that residual structure of these fragments could account for their retention in function. In addition, complementation studies show that the association and cooperative function of AK_{Bs} fragments with little residual structure can be increased by fusing these peptides to interacting proteins. Similarly, the interaction of a split AK_{Tn} can be enhanced at a temperature (78°C) where the fragments are non-functional by fusion to proteins that interact. This split AK_{Tn}, which represents the first high-temperature protein fragment complementation assay (ht-PCA) for analyzing protein-protein interactions within a living thermophilic bacterium, is capable of detecting

predicted interactions among *Thermotoga maritima* chemotaxis proteins. These findings show that split proteins with varying functions can be rapidly discovered by fragmenting orthologs with a range of thermostabilities. Moreover, the novel ht-PCA described herein will aid in creating genome-wide maps of thermophilic protein-protein interactions, studying the effects of temperature on biomolecular interactions, and engineering oligomeric thermostable nanomaterials.

Acknowledgements

My mentor Joff Silberg has been a consistent force for scientific scrutiny and enthusiasm. He has given me to opportunity to spread my wings during my time at Rice, allowing me to play central roles in other research endeavors that are not covered in this work. For his tenacity, guidance, and friendship, I owe many, many thanks. I hope that my involement with the Silberg Lab will extend beyond the time I have spent as a graduate student.

My committee deserves the utmost praise, for guiding me through my evolution as a scientist and raising concerns that may have escaped me. Professor George Bennett has been inspirational from my first moments at Rice, always asking questions that strike right to the core of the issues at hand. Professor Yousif Shamoo deserves an enormous amount of gratitude from me, as his humor, levity, and sharp scientific criticism have substantially elevated my thinking. Thanks for keeping me sharp. Professor Jane Tao, the only original member of my committee, has been there to see me from the beginning and all your suggestions and insight into protein biophysics has been extremely helpful. Professor Ka-Yiu San has worked with me through the Rice iGEM team, and has provided not only kind assistance in terms of equipment, but fantastic suggestions that made us consider new avenues of thought. To Professor Pernilla Wittung-Stafshede, my former committee head, thanks for all your very insightful past and current strategic advice regarding navigating academic science. To my past committee member Professor Charles Stewart, I regret conditions that have precluded you from continuing as a member in my committee, but your questions and suggestions during your short time are greatly appreciated.

To those in the Silberg Lab, you've made these past 6 years whiz by. Shirley, you're the 'heart' of the Silberg Lab, always there to beam your smile and spread your upbeat attitude. Dr. Peng Zhai, we came to Rice together, joined the same lab together, and now we leave at nearly the same time. To previous members of the Silberg Lab Crystal Stanworth and Rui Li, thanks for keeping the office an absolute riot while you guys were here. I have too many undergrads to thank, so I'll keep it short. Thomas Segall-Shapiro has been working with me throughout the majority of my time as a graduate student. His sharp intellect and prowess in the lab has earned the utmost respect from me; I see him going on after his time here at Rice to be a great scientist. Tina Chen, thanks for keeping lab lively and for your consistent friendship. To David Kim and Jeremy Thompson, I appreciate all of the hard work you guys have done to move our projects forward. A deep thanks to Professor Kathleen Matthews and Professor Sarah Bondos, who have been there from the beginning. To everyone in Keck -- thanks for allowing me to either use your equipment, pick your brain, or for otherwise always being there for moral support. And finally to my family and friends, thank you for always being there to support me in my scientific pursuits.

Acronyms

Ap5A $= P^1P^5$-di(adenosine 5')-pentaphosphate
PIN = Protein Interaction Network
PCA = Protein-Fragment Complementation Assay
htPCA = High Temperature Protein-Fragment Complementation Assay
AK = Adenylate Kinase
Y2H = Yeast Two-Hybrid
MS = Mass Spectrometry
ORF = Open Reading Frame
AK_{Ec} = *Escherichia coli* Adenylate kinase.
AK_{Bs} = *Bacillus subtilis* Adenylate kinase.
AK_{Tn} = *Thermotoga neapolitana* Adenylate kinase.
EcAK-N = N-terminal fragment (1-76) of the *Escherichia coli* Adenylate Kinase.
EcAK-C = C-terminal fragment (77-214) of the *Escherichia coli* Adenylate Kinase.

Contents

Chapter 1

Introduction

1.1 Understanding Complexity in Biological Systems

1.1.a Systems Biology

Biological research has historically employed a reductionist approach to understand the complex nature of the molecular systems that underly living organisms. This affords investigators a tractable strategy to fully probe and characterize a biological molecule or process in a context that is removed from complex environment of the cell where the molecule of interest is subject to multiple forms of regulation. With the explosion of biological data from fully sequenced genomes (Blattner et al., 1997; Deckert et al., 1998; Goffeau et al., 1996; Gregory et al., 2006; Henne et al., 2004a; Kawarabayasi et al., 1998a; Kunst et al., 1997; Nelson et al., 1999a; Tomb et al., 1997; Waterston et al., 2002), we have ushered in a new era where the traditionally laborious characterization of biological entities is greatly accelerated using automation, computational tools, and high-throughput methods (Schena et al., 1995; Lashkari et al., 1997; Rigaut et al., 1999; Lesley et al., 2002; Pham et al., 2007; Fiebitz et al., 2008; Michael et al., 2008; Mayr and Bojanic, 2009; Lennon et al., 2010). Large datasets can now be routinely generated, and these are beginning to highlight the inadequacies of the traditional reductionist paradigm. As these datasets are compared with biochemical data, it is becoming increasingly apparent that the careful study of individual biomolecules fails to capture the complex dynamics of the larger-scale system of which they are a part. For example, one cannot accurately model geoclimate dynamics by considering only local changes in weather patterns or understand the fundamentals of an airplane from a catalog of parts. Similarly, knowledge of the biomolecular constituents of a

cell (or a handful of biomolecules) does not lend itself to understanding what makes a living organism 'living'. Instead, a systems-level approach must be taken towards understanding the structure and dynamics of biological systems. This new approach, aptly named "systems biology," aims at: i) understanding the complex structure of biological systems in terms of how individual components interact to form networks, ii) how these networks change dynamically with time and/or internal[1] or external perturbations, and iii) what emergent properties (properties only held by the system at a higher-level organization) can arise from these complex dynamic networks (Kitano, 2002; Fu and Panke, 2009; Longtin, 2005; Mazzocchi, 2008). Such an ambitious goal requires large datasets encompassing descriptions for all of the molecular entities in the cell, the acquisition of which leads to information in the form of the cellular genome, proteome (Pandey and Mann, 2000), metabolome[2], lipidome[3], and glycome[4]. Understanding the complex dynamics of these biological molecules is central for establishing a unified perspective of the cell on a molecular scale and decoding the function of cellular systems. In addition, a better understanding of biomolecules and their genome-wide interactions will aid in understanding the molecular basis of human diseases and ageing, assist in drug discovery, and help with the development of novel biotechnologies (Pazos et al., 2003; Butcher et al., 2004; Schadt and Lum, 2006; Delorenzo, 2008; Armengaud, 2009; de Groot et al., 2009; Gardy et al., 2009; Schadt et al., 2009; Villoslada et al., 2009).

---

[1] This may be due to evolved properties of the system, such as regulatory structures or network stability. In addition, biological networks can be affected by stochastic effects.
[2] The totality of metabolites present in an organism.
[3] The totality of lipid molecules present in an organism.
[4] The totality of carbohydrates present in an organism.

1.1.b Protein Interaction Networks

Advances in systems biology are being accelerated by proteomic studies that identify cellular proteins and map the interactions between these proteins. The resulting *protein interaction networks* (PINs) are proving to be useful for establishing information flow within biological systems, determining functional modules within cells, and assigning functions to unannotated open reading frames (ORFs) by virtue of association with proteins of known function (Legrain et al., 2001). PIN data sets can provide spatial information on the subcellular localization of the networks, as well as the temporal organization of protein-protein interactions (Kim et al., 2008; Monteiro et al., 2008; Albert and Wang, 2009; Wu et al., 2009; Yamada and Bork, 2009). However, a single data set describing PINs is not always sufficient to explain complex cellular behaviors. PINs must be determined under different conditions to understand the how changes in protein concentrations and post-translational modifications affect protein-protein interactions. When multiple PIN data sets are generated, the topology of these networks[5] can be compared to understand the emergent properties of these networks (Agrafioti et al., 2005; Cesareni et al., 2005; Gandhi et al., 2006; Zhu and Qin, 2005). To date, PINs have been generated for diverse model organisms, including *Escherichia coli, Saccharomyces cerevisiae,* and *Drosophila melanogaster* (Tarassov et al., 2008a; Middendorf et al., 2005; Butland et al., 2005). However, we cannot gauge how accurate and comprehensive the PINs are even in the best characterized model organisms, and PINs remain poorly defined for many multicellular organisms and unicellular organisms that live in unconventional environments.

---

[5] Topology as applied to biological networks derives its meaning from the mathematical field of graph theory, where interacting objects can be characterized as a collection of nodes (elements) and vertices (directed interactions between the elements). Abstract properties can be used to classify such network topologies based on the global structure.

1.2   Strategies for Mapping Protein Interaction Networks

1.2.a   *In vitro* Methods

Numerous methods have been developed for analyzing interactions among proteins within complex cell extracts (*e.g.*, protein affinity chromatography, co-immunoprecipitation, and protein crosslinking) and between purified proteins *in vitro* (*e.g.* isothermal titration calorimetry, surface plasmon resonance, and fluorescence polarization). *In vitro* methods are widely used to study the interactions of purified recombinant proteins because they are easy to implement and are frequently quite sensitive (Phizicky and Fields, 1995). However, they require homogeneous samples of proteins, which is not always possible. In addition, *in vitro* measurements are performed in highly controlled environments outside of a cellular context, and it is frequently unclear if the results from these analyses are physiologically relevant. In contrast, the *in vivo* methods listed above obtain information on protein-protein interactions within a cellular context. While these methods are quick and easy to implement, they are limited in their throughput and lack the ability for large-scale parallel implementation. To generate systems biology models for living systems, proteome-wide protein-protein interactions need to be mapped in a near-physiological context (in a highly similar, if not native cell type and subcellular localization) using methods that are high-throughput.

1.2.b   The Yeast-Two Hybrid Assay

A pioneering breakthrough in the high-throughput (HT) analysis of PINs within living cells was the development of the yeast two-hybrid (Y2H) assay (for review, see Young, 1998; Vidal and Legrain, 1999). This method allows rapid *in vivo* screening of protein-protein interactions within a cellular context without the need for purifying

recombinant proteins (Fields and Song, 1989; Chien et al., 1991). The original Y2H was built by splitting the yeast Gal4 transcriptional activator protein into two separate domains (the DNA binding and activation domains) that are not functional when coexpressed in yeast. However, when these proteins are fused to proteins that interact, the two Gal4 domains are brought into close proximity and capable of activating the transciption of a reporter gene product, typically a prototrophic marker or a color. The Y2H has the advantage of being easy to use and requires little specialized equipment. Furthermore, the Y2H is amenable to high-throughput analysis. Libraries can be screened against a single protein, called the 'bait', or against a second library (Chien et al., 1991; Finley and Brent, 1994). However, the Y2H has one major drawback. This strategy frequently results in false positives and negatives, as well as poor reproducibility, because the proteins being studied are expressed at non-physiological levels (Uetz and Hughes, 2000). Recent meta-analyses of multiple Y2H data sets estimated that the Y2H error rate can be greater than 50% (Ito et al., 2001; Deane et al., 2002a).

1.2.c   Mass Spectrometry Analysis of Protein Compelexes

PINs can also be mapped within living cells by coupling tandem affinity purification (TAP) to mass spectrometry (MS) (Rigaut et al., 1999). TAP-MS is being increasingly used to assess PINs because it overcomes some limitations of the Y2H, such as false positive and negative signals (Deane et al., 2002b). In TAP-MS, a single bait protein of interest is expressed with an affinity purification tag within the organism in which that protein evolved. Proteins that associate with the bait are identified by: i) recombining an affinity tag into the allele that encodes the protein bait in the organism of interest, ii) using affinity chromatography to isolate a complex composed of the bait and proteins bound to it, and iii)

determining the identity of the proteins associated with the bait using MS-based protein sequencing[6] (Ethier et al., 2006). TAP-MS has multiple advantages over the Y2H approach. The proteins of interest are not overexpressed from a plasmid, larger multi-protein complexes can be isolated directly from the host of interest, and the protein can be expressed within its normal subcellular location. However, false positives from the association of cellular proteins with the affinity tags and/or solid supports can be problematic. While stringent washing protocols can remove these false positives, they do it at the expense of weak interactions that are physiologically relevant (Ethier et al., 2006).

1.2.d   Protein Fragment Complementation Assays

Protein fragment complementation assays (PCAs) are a relatively recent advent in the analysis of protein-protein interactions within living cells, which are being increasingly used for proteomic studies (Pelletier et al., 1998; Michnick et al., 2000, 2007). As with Y2H assays, PCAs are created by splitting a protein into two fragments. In the case of a PCA, however, you split a single domain protein into two polypeptide fragments that lack function when mixed together but exhibit parent-like function when they are fused to associating proteins (see Figure 1.1, page 20). To date, PCAs have been created that generate outputs that can be screened (*e.g.*, fluorescence, colorimetric substrate production) or selected (*e.g.*, antibiotic resistance and auxotrophic rescue). This has been accomplished by splitting diverse single-domain proteins, including dihydrofolate reductase, β-lactamase, green fluorescent protein, and β-galactosidase (Rossi et al., 1997; Pelletier et al., 1998; Doi and Yanagawa, 1999; Galarneau et al., 2002). The results from the applications of these diverse PCA systems

---

[6] These are considered 'coupled-MS' methods, in which a purification step (e.g., liquid chromatography for small molecules or affinity purification for proteins) is used to resolve a complex mixture, which is then coupled to MS analysis.

have provided evidence that the key to building an effective PCA is the use of protein fragments that remain unfolded unless brought into close proximity by other interacting proteins (Michnick et al., 2000; Michnick, 2003).

PCAs, like the Y2H system, have the advantage of being facile and inexpensive to implement. However, unlike the Y2H assay, PCA strategies can be utilized in diverse organisms, tissues, cell types, and sub-cellular compartments to provide detection of truly context-relevant protein interactions (Michnick et al., 2000). In addition, PCAs are thought to be better at detecting weak protein-protein interactions than Y2H assays (Tarassov et al., 2008b), suggesting that this approach should be used when mapping PINs. PCAs have also found utility in other biotechnological applications. They have been used as screens for drug discovery and as screens to find artificial proteins that bind natural biomolecules (Villalobos et al., 2008; Vidi and Watts, 2009; Luker, Gupta, and Luker, 2009; Luker, Gupta, Steele, et al., 2009; Rossi et al., 1997; Michnick, 2003; Löfdahl et al., 2009; Michnick et al., 2007; Nyfeler and Hauri, 2007). The versatility of existing PCAs suggests that this strategy should be considered when constructing new assays for mapping biomolecular interactions, such as assays for monitoring interactions within organisms that live in extreme environments. The main shortcoming of the PCA approach is the difficulty in identifying split sites that yield protein fragments which functionally complement *only* when they are fused to interacting proteins.

1.2.e Novel Proteomic Assays are Needed to Study Extremophiles

All of the aforementioned methods for detecting protein-protein interactions have been developed for use in a subset of model organisms, as many of the relevant proteins

being analyzed are compatible with such systems. Unfortunately, these assays cannot be used within extremophilic organisms that have adapted their constituent biomolecules for survival in very harsh physical environments. Extremophiles have been found in environments that seem inhospitable to life, including those that are hyperacidic (pH of 2 or below), high molarity (> 2M salt), very cold (down to -20°C), and very hot (up to 121°C) (Závodszky et al., 1998; Cavicchioli, 2006; Kashefi and Lovley, 2003; Baker-Austin and Dopson, 2007; Pikuta et al., 2007; D'Amico et al., 2006). For many extremophiles, their environments do not affect their cellular milieu, as they have adapted mechanisms to protect their cellular contents from their environmental conditions. In the case of acidophiles, for example, ion transporters are used to generate a chemiosmotic gradient that maintains a near-neutral pH in their cytosol (Baker-Austin et al., 2007). Because the cytosolic physical environment of these organisms is similar to the cytosol of model organisms, existing protein-protein interaction assays (PCA, Y2H, TAP-MS) are likely to be useful for probing PINs within these extremophiles. In contrast, thermophiles and hyperthermophiles that thrive at extreme temperatures cannot protect their biomolecular contents from the harsh conditions of their environment, and all of the proteomic strategies developed for use within mesophilic organisms cannot be used to characterize protein-protein interactions within thermophiles and hyperthermophiles. Because protein-protein interactions cannot be measured within these organsisms, it remains unclear whether the existing high-throughput assays have captured all of the PINs that occur in these organisms.
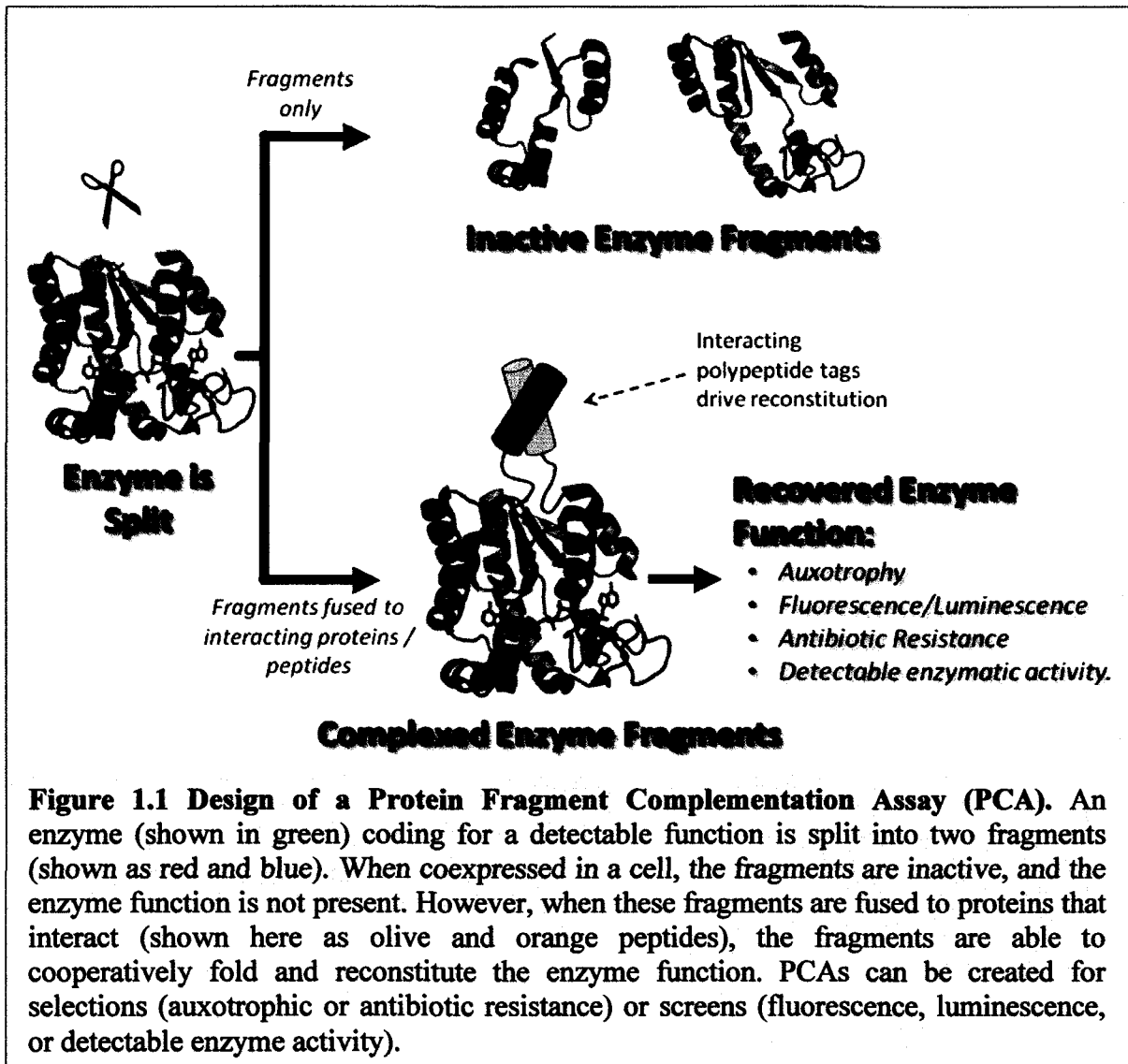
Figure 1.1 Design of a Protein Fragment Complementation Assay (PCA). An enzyme (shown in green) coding for a detectable function is split into two fragments (shown as red and blue). When coexpressed in a cell, the fragments are inactive, and the enzyme function is not present. However, when these fragments are fused to proteins that interact (shown here as olive and orange peptides), the fragments are able to cooperatively fold and reconstitute the enzyme function. PCAs can be created for selections (auxotrophic or antibiotic resistance) or screens (fluorescence, luminescence, or detectable enzyme activity).

1.3   Extremophile Interaction Maps Remain Poorly Defined

1.3.a   Thermotolerant Microbes: *Life on the Edge*

Thermophiles[7], bacteria, and archaebacteria which have adapted to thrive at temperatures exceeding 60°C, were first identified in the 1960s within hot springs at Yellowstone National Park by Thomas D. Brock (Brock and Freeze, 1969). Since their initial discovery, a diverse array of organisms have been found that live at extreme temperatures, and some people now differentiate thermophilies from hyperthermophiles, which grow optimally above 80°C (see Figure 1.2, page 23) (Ramaley and Bitzinger, 1975; Noll and Vargas, 1997; Pentecost, 1996; Pikuta et al., 2007). Nearly all thermophiles described are bacterial or archaeal in nature, although fungi have been discovered that live at temperatures as high as 62°C (Maheshwari et al., 2000). In contrast, all hyperthermophiles are either archaea or bacteria. Both thermophiles and hyperthermophiles are found in habitats where water is proximal to a heat source, *e.g.*, hot springs, fumaroles, hydrothermal vents, exothermic compost piles, and domestic water heaters (Figure 1.3, page 24) (Brock and Boylen, 1973; Ramaley et al., 1975; Huber et al., 2000; Thummes et al., 2007). In high-temperature deep ocean biotas, phototrophic organisms are non-existent, and thermophiles and hyperthermophiles form the foundation of the ecosystem (Perner et al., 2009; Rundqvist et al., 2009; Jeanthon, 2000; Reysenbach and Cady, 2001; Zhou et al., 2009; Segerer et al., 1993; Pentecost, 1996; Krishnamurthy et al., 2009c; Petursdottir et al., 2009). To date, the highest recorded temperature for hyperthermophile survival was demonstrated for a Crenarcheatoa dubbed *Strain 121*, which was able to reproduce at the temperature used by autoclaves for sterilization (121°C) and survive heating up to 130°C (Kashefi et al., 2003). The ability of an organism to withstand such mind-boggling temperatures, which for our cells

---

[7] Greek for *thermo-* = heat and *-phile* = loving.

would cause irreversible biomolecular degradation and/or denaturation, raises the question about the temperature limits for life forms based on DNA, RNA, proteins, and lipids (Naylor, 2004). The study of the PINs in these thermotolerant microbes could provide insights into the evolution of life on Earth, the potential for life elsewhere in the Universe, and the limits of life that are based on DNA and protein systems.

Atomic Blast — $10^7\,^\circ C$

Intergalactic Matter — $10^6\,^\circ C$

Surface of Sun — $10^4\,^\circ C$

Magma — $10^3\,^\circ C$

Temperature Range of Life on Earth { $\sim 10^2\,^\circ C$ / $\sim 10^1\,^\circ C$

Deep Space — $10^{-2}\,^\circ C$

121°C ···· Strain 121 ····
P. horikoshii
P. furiosus
M. jannashii   **Hyperthermophiles**
T. neapolitana
T. maritima
80°C ···· T. thermophilus
T. aquaticus
B. stearothermophilus   Thermophiles
50°C ····
E. coli
H. sapiens
S. cerevisae
B. subtilis   Mesophiles
15°C ····
M. marinus
P. articus   Psychrophiles

Figure 1.2. Classification of Life by Growth Temperature. Life on Earth is known to exist within a narrow range when compared to the range of temperatures in the universe. Organisms can be classified as psychrophiles (up to 15°C), mesophiles (15-50°C), thermophiles (50-80°C), and hyperthermophiles (80°C and beyond).

Figure 1.3. Thermophile Habitats. Natural habitats consist of water heated by geothermal heat, such as terrestrial hot springs (The Grand Prismatic Spring at Yellowstone National Park is shown; image from the National Park Service), deep sea hydrothermal vents (shown is a 'black smoker' from the Juan de Fuca Ridge in the Pacific Ocean, image from the National Oceanic and Atmospheric Adminstration). Natural and artifical compost heaps generate heat and are also sources of thermophiles. Thermophiles have even been found in completely man-made biotas, such as commercial water heaters. All images copyright-free.

1.3.b   Thermophile Molecular Adaptations to Extreme Temperatures

The incubation of biomolecules that are the building blocks for life on Earth (*e.g.*, lipids, DNA, RNA, and proteins) at increasing temperatures will eventually destabilize them to a level that cannot be counteracted by increased biosynthesis. This suggests the existence of a temperature limit for life where the degradation of essential biomolecules exceeds their rate of synthesis. Biochemical and genetic studies with thermophiles (and hyperthermophiles) have yielded insights into the distinctive nature of their physiology that allows them to grow near the boiling point of water where many biomolecules undergo thermal degradation (Kelly and Adams, 1994; Noll et al., 1997; Tehei and Zaccai, 2005; Oshima, 2007; Averhoff, 2009; Perugino et al., 2009). Table 1.1 on page 28 lists some of these strategies. Thermophiles maintain their membrane integrity by using unique lipids and fatty acids with higher melting temperatures. This adaptation is thought to have occurred because mesophilic lipids lose functional integrity at higher temperatures due to increased lipid fluidity and permeability (Charlier and Droogmans, 2005). Some thermophile metabolites also appear to be protected from degradation by metabolic channeling, in which substrates are directly transferred from one enzyme to another without ever diffusing into the cytoplasm (Massant and Glansdorff, 2004; Massant et al., 2002a). Furthermore, there is evidence that thermophiles protect their genetic material from thermal degradation through the production of DNA-stabilizing small molecules, unique DNA repair systems, and the expression of a unique reverse gyrase enzyme that induces positive DNA supercoiling that is thermostabilizing (Kikuchi and Asai, 1984; Bouthier de la Tour et al., 1990; Marguet and Forterre, 1994). In addition, several researchers have proposed that thermophiles maintain the

structure and function of their transfer RNAs by altering their guanine and cytosine content (de la Tour et al., 1995; Lynn et al., 2002; Perugino et al., 2009).

Biophysical studies have shown that multiple strategies can be used to stabilize proteins against thermal denaturation, including increased compactness, contact order, ionic interactions, hydrogen bonding, helical content, subunit interactions, hydrophobic core packing, and oligomerization (Vogt et al., 1997; Kumar et al., 2000; Kannan and Vishveshwara, 2000; Chakravarty and Varadarajan, 2002; Dominy et al., 2004; Ge et al., 2008; Maugini et al., 2009; Baldasseroni and Pascarella, 2009). However, these strategies are not used uniformly by all thermophilic proteins to increase their thermostability compared with mesophilic proteins. Bioinformatics studies comparing the physicochemical properties of large numbers of thermophilic proteins have identified differences between protein orthologs in mesophiles and thermophiles. Thermophilic proteins are enriched in charged versus polar residues and also contain a higher density of salt bridges when compared to their mesophilic counterparts (Cambillau and Claverie, 2000; Suhre and Claverie, 2003b; Robinson-Rechavi et al., 2006a). In addition, structural genomics studies indicate that proteins from thermophiles have smaller surface loops than their mesophilic orthologs (Kumar et al., 2000).

Proteins in thermophiles are also thought to be stabilized by adaptation mechanisms at the organismal level, such as the production of stabilizing solutes known as osmolytes[8] (Neelon et al., 2005; Empadinhas and da Costa, 2008, 2006a). In mesophiles, increases in solute concentrations cause volume-exclusion effects that stabilize proteins, alter biochemical equilibria, and increase macromolecular association constants by several orders of magnitude

---

[8] Osmolytes are organic molecules that influence osmotic balance and have been demonstrated to influence protein folding and stability through crowding effects.

(Eggers and Valentine, 2001; Neelon et al., 2005; Empadinhas et al., 2008). The yeast *Saccharomyces cerevisiae*, for example, has been shown to accumulate glycerol intracellularly during its temperature adaptation response (Siderius et al., 2000). Solutes produced in thermophiles are predicted to have similar effects. However, thermophiles produce solutes unique from those found in mesophiles (Martins et al., 1997; Empadinhas and da Costa, 2006b), which may use additional mechanisms to promote thermostablization (Empadinhas et al., 2006a). Thermophile-specific solutes are thought to include diglycerol phosphate, dimyoinositol-1-1'-phosphate, mannosylglycerate, dimannoyl-dimyoinositol-1,1'-phosphate, and mannosylglyceramide (Santos and da Costa, 2001). Currently, the extent to which thermophile solutes influence protein interactions is not known. Furthermore, it remains unclear whether these compounds influence macromolecular interactions solely through volume exclusion effects, or if they bind and specifically modulate some interactions.

Table 1.1. Thermophile Adaptations to High Temperatures

Nucleic acids	• Higher G+C Content in RNA molecules (Galtier and Lobry, 1997). • Production of DNA-stabilizing polycations and polyamines (Hensel and König, 1988; Terui et al., 2005). • Exclusive presence of reverse gyrase enzyme, which promotes positive supercoiling of DNA (Bouthier de la Tour et al., 1990). • Highly efficient DNA repair mechanisms (Wood et al., 1997).
Lipids	• Unique ether-linked membrane lipid molecules (van de Vossenberg et al., 1998; Xu and Glansdorff, 2002). • Some lipids also contain internal cyclopentane rings, allowing an increase in lipid rigidity (Gliozzi et al., 1983). • Secondary membrane ion transport systems to combat increased proton permeability at high temperatures (Albers et al., 2001).
Proteins	• Greater number of charged residues and salt bridges (Robinson-Rechavi et al., 2006b; Missimer et al., 2007). • Decrease in surface-exposed hydrophobic groups (Vogt et al., 1997). • More efficient packing of the protein hydrophobic core (Fang et al., 2008). • Stabilizing charge-neutral hydrogen bonding networks (Tanner et al., 1996; Vogt et al., 1997). • Stabilizing disulfide bridges (Cacciapuoti et al., 1994). • Shortening of external loops (Thompson and Eisenberg, 1999). • Increased metal cofactor binding (Teplyakov et al., 1990; Kojoh et al., 1999; Smith et al., 1999). • Aromatic clusters (Kannan et al., 2000; Trebbi et al., 2005). • Higher oligomeric forms and intersubunit interactions (Shima et al., 1998; Thoma et al., 2000; Robinson-Rechavi et al., 2006b). • Metabolic chanelling of substrates (Massant et al., 2002b). • Protein stabilization by production of compatible solutes (Martins et al., 1997; Kumar et al., 2000; Borges et al., 2002).

1.3.c Thermophiles as a Biological Resource

The explosion of molecular biology in the past few decades is attributed to the discovery of thermophilic DNA polymerases, which led to the development of the polymerase chain reaction (PCR) for DNA synthesis. The key discovery that changed DNA synthesis from a slow method to a rapid and convenient process was the replacement of a mesophilic DNA polymerase with a homolog from a thermophilic bacterium (Saiki et al., 1988). A variety of thermophilic enzymes have since been harnessed for biotechnology applications, such as DNA ligation (ligases), DNA winding (helicases), DNA degradation (nucleases), as well as a reverse transcription (by RNA-dependent DNA polymerases) (A. Pantazaki, 2002). Proteins discovered in thermophiles have not only revolutionized molecular biology, but they have also begun to transform industrial processes. Xylanases have made paper processing greener, and oligosaccharide-modifying enzymes have been harnessed for corn syrup production (Egorova and Antranikian, 2005; Mathrani and Ahring, 1992). While there is interest in further harnessing thermotolerant microbes and their enzymes for other industrial processes such as biomass conversion to bioethanol or biohydrogen, many proteins of a hyperthermophilic origin are still annotated as hypothetical proteins, limiting the full biotechnological utilization of these unique organisms (Sommer et al., 2004; Koskinen et al., 2008; Taylor et al., 2009).

1.3.d The Thermophile Proteomics Bottleneck

Although the genomes of many of thermophiles and hyperthermophiles have been sequenced, our understanding of their PINs is still limited (Kawarabayasi et al., 2001, 1998b; Smith et al., 1997; Nakamura et al., 2002; Nelson et al., 1999b; Chen et al., 2005; Henne et

al., 2004b). To date, only two proteome analyses have been performed on thermophilic organisms that seek to determine their genome-wide PINs, and both of these have used conventional methods for probing protein-protein interactions. In the first study, a two-hybrid strategy was used to screen for interactions among proteins from the hyperthermophile *Pyrococcus horikoshii* (optimal growth 98°C; González et al., 1998). In the first study, only 56 hetero-interactions were discovered, even though the pairwise interactions of 960 soluible proteins were probed, *i.e.*, approximately one million protein-protein interactions (Usui et al., 2005). In the second study, the large scale fractionation of the native proteome of a related hyperthermophile (*Pyrococcus furiosus*) was performed under non-denaturing conditions to isolate multi-protein complexes, and the sequence of the proteins within these complexes were identified by nano LC-ESI-MS/MS (Menon et al., 2009). This strategy identified approximately 2-fold more protein complexes than the two hybrid study.

The small number of PINs discovered in high-throughput studies of thermophilic proteomes is rather surprising given that protein interactome studies in mesophilic microbes with similar numbers of open reading frames (ORFs) have yielded PINs with more interactions. For example, a high-throughput screen of 285 *Heliobacter pylori* proteins against a library of genome-encoded polypeptides identified over 1,200 interactions; far more than observed with *P. horikoshii* (Rain et al., 2001). Several explanations could account for the paucity of thermophilic protein-protein interactions discovered in high-throughput screens. First, thermophilic microbes could simply have fewer protein-protein interactions. Second, some protein-protein interactions could be hard to detect with existing assays. Screens for interacting *P. horikoshii* pairs were performed at a temperature (37°C) significantly lower than the physiological temperature optima (98°C) of the organism, and

the *P. furiosis* protein fractionation studies involved multiple complex protein purification steps at non-physiological temperatures (some as low as 18°C). Under these conditions, the relative affinity of some thermophilic protein-protein interactions could be decreased. This could arise because heat activation is required for many thermophilic proteins to achieve the proper native conformation for competent binding to a protein partner, as observed with recombinant *P. islandicum* glutamate dehydrogenase (Kujo and Ohshima, 1998). This could also arise in cases where thermophilic protein-protein interactions weaken as temperature is decreased from thermophilic levels. Calorimetric studies have shown that this occurs with a handful of thermophilic protein-protein interactions. In the case of the alpha- and beta-subunits of tryptophan synthase from *P. furiosus*, for example, the association constant increased by >10-fold upon a 25°C increase in assay temperature (Ogasahara et al., 2003). Another concern is that the compatible solutes produced by thermophiles (as discussed in section 1.3.b) influences thermophilic protein-protein interactions and therefore a similar *in vivo* environment would be necessary to identify these interactions. Taken together, these considerations suggest high-throughput screens for thermophilic protein-protein interactions *in vivo* at physiological temperatures would be best for generating thorough maps of their PINs.

1.4   Adenylate Kinase as a Model System for Studying Split Protein Function

PCAs have been successfully constructed by splitting a variety of single domain proteins (Rossi et al., 1997; Pelletier et al., 1998; Doi et al., 1999; Galarneau et al., 2002). However, there are no PCAs for mapping PINs within thermotolerant microbes, and it remains unclear how to engineer structurally-related PCAs that function over a range of

temperatures. In addtion, it remains unknown how the choice of different homologous proteins for fragmentation affects the likelihood that split protein variants will cooperatively function. The adenylate kinase (AK) protein family is ideal for addressing these issues. Extensive structural and functional data is available for AK orthologs from psychrophiles, mesophiles, thermophiles, and hyperthermophiles (Schulz, 1987; Haney et al., 1999; Criswell et al., 2003; Bae, 2004). In addition, their functions can be selected over a broad temperature range, unlike many proteins previously used in split protein studies, *e.g.*, GFP, beta-lactamases, and luciferases (Rossi et al., 1997; Doi et al., 1999; Villalobos et al., 2008).

1.4.a Essential Role as Phosphotransferases

Adenylate kinases are small, ubiquitous phosphotransferases that maintain adenine energy charge[9] in all cells by catalyzing the reversible interconversion of ATP and AMP to ADP as shown in Figure 1.4 on page 34 (Atkinson, 1969; Brune et al., 1985). These enzymes were first characterized in the 1950s, and in the decades hence, AK family members have been used as model systems for thousands of enzyme structure-function studies. The essential nature[10] of AKs is highlighted by the lack of a viable gene knockout. However, simple selections for AK function have been described in bacteria, including the mesophilic *E. coli* and thermophilic *Geobacillus stearothermophilus* (Glaser et al., 1975; Kishi et al., 1987; Rose et al., 1991; Kawai et al., 1992; Noma et al., 2001; Stolworthy and Black, 2001; Counago and Shamoo, 2005; Gigliobianco et al., 2008). The *E. coli* strain used to select for AK function was originally identified in a screen for bacteria that exhibit temperature-dependent defects in membrane phospholipid synthesis (Cronan and Godson, 1972). One of

---

[9] The adenine energy charge is a concept by Atkinson that describes how cells try to maintain a balance between pools of ATP and ADP + AMP. This ratio has been shown to change according to the growth phase of the cell.

[10] In the cell, adenylate kinases maintain ATP/ADP/AMP balance while ATP synthase generates ATP from ADP+P_i.

the strains discovered (CV2), which was able to grow at 30°C but was not viable at 40°C, was found to have a single mutation (P87S) within *E.coli* adenylate kinase (AK_{Ec}) (Gilles et al., 1986; Haase et al., 1989). This mutation results in an AK with a decreased catalytic rate, decreased substrate affinity, and increased susceptibility to proteolysis. In theory, similar AK mutants could be created within any organism by replacing their chromosomally-encoded AK gene (*adk*) with a homolog that is less thermostable, similar to a *G. stearothermophilus* mutant that has been reported (Counago and Shamoo, 2005).

Figure 1.4. Function and Structure of Adenylate kinases. A) Adenylate kinases are essential phosphotransferases that catalyze the reversible transfer of terminal phosphate groups between adenosine monophosphate and adenosine triphosphate. B) The sequence of AK proteins showing a division into three domains: the CORE domain (residues 1-30, 60-127, and 160-217; contains the majority of the structural and catalytic core), the AMP binding domain (residues 31-59; mediates highly specific binding of AMP), and the LID domain (residues 129-159; closes over site of phosphoryl transfer, may coordinate a zinc ion). The P-loop motif is underlined. Residues that mediate zinc binding are shown in orange, underlined. Secondary structure prediction from Jnet (Cuff and Barton, 2000). C) Crystal structure of the *Bacillus stearothermophilus* AK (PDB ID = 1ZIN), with the CORE domain shown in green, the LID domain shown in blue, and the AMP binding domain shown in red. The bound substrate analogue, P^1P^5-di(adenosine 5')-pentaphosphate (Ap5A) is shown as a magenta stick model. The zinc confactor bound in the LID domain is shown as a space-filling sphere, in orange. All structures were acquired using PyMOL (DeLano Scientific, LLC).

1.4.b Sequence, Structure, and Function Conservation within the AK Family

The structure and function of AK orthologs have been characterized for organisms that live over a range of temperatures, including homologs from hyperthermophiles (*Thermotoga neopolitana, Aquifex aeolicus, Sulfolobus acidocaldarius,* and *Methanococcus igneus*), thermophiles (*Geobacillus stearothermophilus, Methanococcus jannaschii,* and *Methanococcus thermolithotrophicus*), mesophiles (*Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae,* and *Methanococcus voltae*), and psychrophiles (*Bacillus globisporus and Marinibacillus marinus*) (Konrad, 1988; Haney et al., 1999; Song et al., 1996; Brune et al., 1985; Kladova, Gavel, Mukhopaadhyay, et al., 2009; Krishnamurthy et al., 2009a; Perrier, Burlacu-Miron, Boussac, et al., 1998; Saint Girons et al., 1987; Gilles et al., 1993; Davlieva and Shamoo, 2009; Bae, 2004; Criswell et al., 2003; Batra et al., 1989; Ferber et al., 1997; Vieille, Krishnamurthy, Hyung-Hwan, et al., 2003; Kladova, Gavel, Zhadan, et al., 2009). These studies have found that AK family members can be subdivided into two classes (short and long forms) which differ in length by ~30 residues. The long form is typically found in bacteria and archaea, whereas the short form of AKs is typically found in higher organisms (Fukami-Kobayashi et al., 1996). The long AK family members can be further divided into family members that bind to a zinc ion and those that do not (Glaser et al., 1992; Gilles et al., 1994).

Structural studies have revealed a highly conserved structure among AK homologs from diverse branches of life. Among all monomeric AK variants, their structure can be functionally divided into three domains, including the CORE, the AMP-binding, and the LID domain (Schlauderer et al., 1996). As previously shown in Figure 1.4.B-C (page 34), the AK CORE domain contains the bulk of the structure and forms the catalytic site of the enzyme.

The N-terminal AMP-binding domain contains a phosphate-binding 'P-loop' motif that is highly conserved among all AK proteins and is thought to be essential for proper folding (Kumar et al., 2001). The LID domain is thought to close during catalysis to protect the phosphate group, and it may have a role in thermostabilization, as this domain coordinates a structurally stabilizing zinc ion in some AK family members (Perrier et al., 1994; Glaser et al., 1992; Gilles et al., 1994; Schulz et al., 1986, 1990).

A number of studies have used AK homologs to investigate how proteins adapt to function at different temperatures (Haney et al., 1999; Vieille et al., 2003; Bae, 2004). These studies have demonstrated that thermostable AK family members are able to maintain high activities at temperatures lower than the physiological temperatures of their host organisms. For example, the hyperthermophilic *T. neapolitana* AK (AK_{Tn}) exhibits maximal activity at 80°C, but retains similar activity to *E. coli* AK at 30°C (Vieille, et al., 2003). This finding is not consistent with the predictions from the 'cooresponding state hypothesis,' which postulates that protein homologs from mesophiles and thermophiles exhibit comparable flexibilities and functions at their physiological temperatures but distinct flexibilities and functions when compared at the temperatures where mesophiles live (Somero, 1978). The findings from studies with thermophilic AK homologs suggest that this ability to maintain the structural dynamics necessary for enzyme activity at temperatures below their physiological range may arise because the regions required for flexibility (and catalysis) may be distinct from those that are important for thermostability (Krishnamurthy et al., 2009).

1.4.c Thermostability of Natural and Engineered Orthologs

The conservation of primary sequences within the AK protein family has led to a wealth of information on how their structures and functions have evolved to function within different temperature ranges. A comparison of AK homologs from a psychrophile (*Bacillus globisporus*), mesophile (*Bacillus subtilis*), and thermophile (*Geobacillus stearothermophilus*) revealed that AK melting temperatures and activity temperature optima correlate with the optimal growth temperatures of the source organisms (Bae, 2004). Structural analysis of these AK homologs also found that increased AK thermostability correlates with an enrichment of ionic interactions, structural flexibility, and buried hydrophobic regions. Bae and coworkers also made chimeras out of these AK family members and found that the thermostabilizing regions of these proteins are spatially separated from those that control activity (Bae and Phillips, 2006). A more recent study with the hyperthermophilic AK_{Tn} from *Thermotoga neapolitana*, which grows from 55-90°C and has an optimal growth temperature of 80°C, obtained a similar result. AK_{Tn} was found to have equivalent enzyme activity at mesophilic temperatures as AK homologs from *B. subtilis* and *E. coli*, organisms that live at mesophilic temperatures (Vieille, et al., 2003). AK_{Tn} is thought to exhibit high activity at mesophilic temperatures because itmaintains local rigidity in its AMP-binding and LID domains while retaining the flexibility in its hinge regions that connect these domains and are needed for catalysis (Krishnamurthy et al., 2009a). This can be contrasted with other hyperthermophilic proteins, which tend to be more structurally rigid at lower temperatures and exhibit decreased catalytic activity under these conditions (Jaenicke and Böhm, 1998; Kohen et al., 1999; Gershenson et al., 2000; Manco et al., 2000).

Biophysical studies with recombinant AK homologs of varying thermostability have also provided evidence that proteins from this family can tune their thermostability through metal binding. AK family members found in gram-positive bacteria frequently contain a conserved Cys-X_2-Cys-X_{16}-Cys-X_2-Cys/Asp motif within their LID domain that has seen shown to bind a thermostabilizing Zn^{2+}, as illustrated in Figure 1.5 on the next page (Gilles et al., 1994). Point mutations of the cysteines within this motif in *Bacillus subtilis* AK (AK_{Bs}) severely impact zinc binding and decrease the thermostability of this protein by 7°C (Perrier et al., 1994). Furthermore, this motif can be introduced into a gram-negative AK to increase thermostability. Perrier and colleagues introduced this zinc-binding motif into the AK_{Ec} LID domain and showed that this mutant binds zinc and exhibits an 11°C increase in thermostability (Perrier, Burlacu-Miron, Bourgeois, et al., 1998). This mutant exhibited parent-like catalytic activity, indicating that the zinc-binding motif simply stabilizes the protein. A structural study performed on this AK_{Ec} mutant provided evidence that zinc-coordination leads to global enhancement of protein stability by increasing the rigidity of the solvent-exposed LID domain and altering the unfolding kinetics of the protein (Burlacu-Miron et al., 1998).
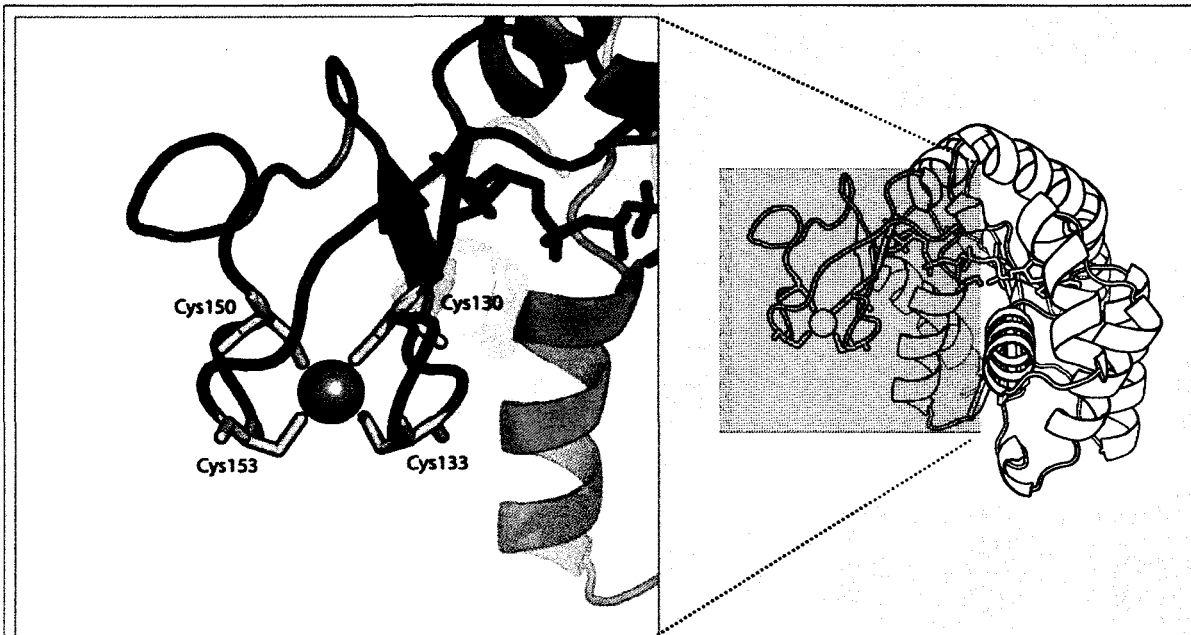
Figure 1.5 Zoomed view of the Zinc-binding motif in *B. stearothermophilus* AK. From PDB structure 1ZIN. In most gram-positive bacterium, four cysteines in the LID domain coordinate a zinc ion, the presence of which thermostabilizes the protein. Some organisms (such as *B. subtilis*) have an aspartic acid at position 150 instead of a cysteine (Perrier et al., 1994). Other bacterium (such as the gram-negative *E. coli*) do not contain a zinc-binding motif in their AKs (Gilles et al., 1994). All structures were acquired using PyMOL (DeLano Scientific, LLC).

1.4.d Identification of a Functional Split AK

Over twenty years ago, an *in vitro* study with purified recombinant AK_{Ec} found that this enzyme can be chemically cleaved into two fragments, [residues 1-76 (EcAK-N) and 77-214 (EcAK-C)] that are able to reassociate and regain some modicum of enzymatic activity (Saint Girons et al., 1987). When expressed separately, these fragments did not exhibit activity and could not bind to their nucleotide substrates. When mixed together in an equimolar ratio, in contrast, EcAK-N and EcAK-C were able to associate and function *in vitro* like full length AK_{Ec}. The function of this split protein differed from the full-length protein. When mixed together, EcAK-N and EcAK-C displayed catalytic activity that is ~8-fold lower than full-length AK_{Ec}, and the EcAK-N/EcAK-C complex displayed a ten-fold greater susceptibility to proteolytic degradation by trypsin compared to full-length AK, although the addition of ATP substrate was shown to increase resistance to this degradation (Saint Girons et al., 1987). A subsequent spectropolarimetry study of revealed that EcAK-N and EcAK-C exhibit increased secondary structure when mixed together (Monnot et al., 1987). While this study provided evidence that AK family members could be used to construct a PCA, there was no evidence prior to this thesis that split AK proteins could be used to report on protein-protein interactions within a living organism.

1.5   The Aims of the Works Presented Herein

Split proteins are increasingly being engineered as assays for probing molecular processes within living cells (for reviews, *see* (Michnick, 2003; Michnick et al., 2007; Nyfeler et al., 2007)). However, rules governing the design of split proteins with tailored functions are lacking. Herein, I investigate the link between protein stability and the

cooperative function of protein fragments created by randomly splitting a single domain protein, using the AK protein family as a model system. In pursuing this line of inquiry, I also construct the first protein-protein interaction assay that can be used within a living thermophilic bacterium to probe protein-protein interactions. This assay extends the temperature where a PCA can measure protein-protein interactions within a microbe from 40 to 78°C.

In Chapter 2, I describe studies that investigate whether protein thermostability affects AK fragment complementation by characterizing the function of split adenylate kinases from the mesophile *Bacillus subtilis* and the hyperthermophile *Thermotoga neapolitana*. For these studies, I generate split proteins that are homologous to the split *E. coli* AK that was previously found to be functional *in vitro* (Saint Girons et al., 1987). In addition, I characterize the *in vitro* functions of orthologous AK fragments (using far-UV circular dichroism, enzyme activity assays, and a zinc determination assay) to determine if there is a correlation between metal binding and residual structure and the cooperative function of split AK variants.

In Chapter 3, I further investigate the link between thermostability and split protein function by evaluating the functions of randomly created split *B. subtilis* and *T. neapolitana* AK variants. To achieve this, I use a transposase-based method to construct libraries of vectors that express randomly split AK variants, and I use *E. coli* CV2 to select for variants in each library that coexpress split AK variants that function as bimolecular complexes. To verify the findings from library selections, seven pairs of homologous *B. subtilis* and *T. neapolitana* AK variants were then generated, and their relative functions were assessed using bacterial complementation.

I investigate in Chapter 4 whether a split *T. neapolitana* AK that funtions in *E. coli* at 40°C also functions to support thermophile growth at at 78°C. To accomplish this, I created a *Thermus thermophilus* HB8 strain that is temperature-sensitive AK due to replacement of its chromosomal adk gene with a homolog from a less thermotolerant microbe. Using a *E. coli–T. thermophilus* shuttle vector for multi-gene expression, I show that this defect for this strain can be complemented by expression of full-length *T. neapolitana* AK. Furthermore, I use this novel complementation system to investigate the function of the split *T. neapolitana* AK system characterized in Chapter 2 in this background, and I show that rational mutation can be used optimize the activity of this split AK for use as a PCA at 78°C. Finally, I investigate whether this novel high-temperature PCA can successfully detect predicted interactions between chemotaxis proteins from the thermophile *T. maritima* MS8.

In the closing Chapter 5, I review the findings presented in this thesis work and describe how my experiments have increased our knowledge of split protein sytems and extended our understanding of the 'thermostability buffer' theory. In addition, novel technologies developed in the course of my work are highlighted and potential applications outlined. Finally, I end this chapter with a look at future paths others can take that would build upon the discoveries and technology presented herein.

Chapter 2

Thermostability Promotes the Cooperative Function of Split Adenylate kinases[11]

2.1   Abstract

Proteins can often be cleaved to create inactive polypeptides that spontaneously associate into functional complexes through non-covalent interactions, but little is known about what influences the cooperative function of the ensuing protein fragments. In this chapter, I examine whether protein thermostability affects protein fragment complementation by characterizing the function of split adenylate kinases from the mesophile *Bacillus subtilis* (AK_{Bs}) and the hyperthermophile *Thermotoga neapolitana* (AK_{Tn}). Using complementation studies, I show that the split AK_{Tn} supports the growth of *Escherichia coli* with a temperature-sensitive AK, but not the fragmented AK_{Bs}. However, I find that weak complementation occurrs when the AK_{Bs} fragments are fused to engineered or natural polypeptides that are known to associate, and this is enhanced by a Q16L mutation that thermostabilizes the full-length protein. To examine how the split AK homologs differ in structure and function, I also characterized their catalytic activity, zinc content, and circular dichroism spectra. I demonstrate that the reconstituted AK_{Tn} had higher levels of zinc, greater secondary structure, and $>10^3$-fold more activity than the AK_{Bs} pair, albeit 17-fold less active than full-length AK_{Tn}. These findings provide evidence that protein fragment complementation can be improved by choosing parents with the greatest thermostability, and they suggest that this arises because hyperthermophilic protein fragments exhibit greater residual structure compared to their mesophilic counterparts.

---

[11] The experiments and data in Chapter 2 have been previously published as a peer-reviewed article in:
Protein Engineering Design and Selection, May 2008; 21: 303-310.

2.2 Introduction

In nature, protein sequence diversity is created through a variety of processes, including gene fission. The frequency of fission has been quantified in several organisms (Snel et al., 2000; Pasek et al., 2006), but its effects on protein function cannot yet be predicted *a priori*, limiting our ability to design split proteins with tailored functions for applications in systems (Michnick, 2003) and synthetic (Giesecke et al., 2006) biology. Single-domain proteins having diverse topologies have been bisected into polypeptide fragments that cooperatively function, including adenylate kinase (Monnot, et al. 1987, Saint Girons, et al. 1987), aminoacyl-tRNA synthetase (Shiba and Schimmel, 1992), chymotrypsin inhibitor-2 (de Prat Gay et al., 1994), dihydrofolate reductase (Gegg et al., 1997), green fluorescent protein (Cabantous et al., 2005), imidazoleglycerol phosphate synthase (Hocker et al., 2001), and ribonuclease S (Kato and Anfinsen, 1969). These split proteins have been used to test questions about protein evolution, folding, solubility and structure-function relationships. In addition, studies with non-functional split proteins have revealed that fragment complementation can often be enhanced by fusing the inactive polypeptides to proteins that associate and promote oligomerization (Johnsson and Varshavsky, 1994; Remy and Michnick, 1999, 2006; Nyfeler et al., 2005; Magliery et al., 2005; Wilson et al., 2004). Such assisted protein fragment complementation has been useful for probing protein-protein interactions in cellular compartments (Spotts et al., 2002; Nyfeler et al., 2005) and organisms (Paulmurugan et al., 2002; Singh et al., 2006) inaccessible by the classical yeast-two hybrid assay (Fields and Song, 1989).

Combinatorial protein engineering studies have shown that for a given enzyme, a diverse array of cut points can yield polypeptides that associate and form functional

heterodimers (Ostermeier, et al. 1999). Non-disruptive breaks have been found within all types of structural elements (*i.e.*, in loops, alpha-helices, and beta-sheets) and between residues that constitute active sites (Ostermeier, et al. 1999). What is poorly understood is whether bisection of protein orthologs at the same site leads to polypeptide pairs with similar levels of parent-like structure and function. Directed evolution studies examining the effects of other mutational processes on the retention of protein function suggest that protein thermostability will influence the activity of split proteins. Among homologous proteins, those with the highest thermostability produce the largest fraction of functional variants upon random mutation (Besenmatter, et al. 2007; Bloom, et al. 2005; Bloom, et al. 2006). Furthermore, the fraction of functional variants in chimeric libraries created by recombination can be enhanced by incorporating mutations that are known to stabilize the structure of the proteins recombined (Meyer, et al. 2006).

To establish whether protein thermostability affects the cooperative function of fragments that arise when a polypeptide backbone undergoes fission, I evaluated the structure and function of protein fragments derived from protein orthologs having different thermostabilities (see Table 2.1 on the next page and Figure 2.1 on page 44). As my model system, I chose Adenylate kinases, ubiquitous phosphotransferases that maintain cellular adenine energy charge by catalyzing the reversible reaction of ATP + AMP $\leftrightarrow$ 2ADP.

Table 2.1 Characteristics of AK proteins used in Chapters 2-4.

AK Enzyme	Organism Max. Growth Temp.	Optimal Enzyme Temp.	Enzyme Melting Temp.	Refs.
B. subtilis	52°C	45°C	51°C	(Bae and Phillips, 2004)
B. subtilis$^{Q16L, Q199R}$	52°C	60°C	69°C	(Counago et al., 2006)
T. neapolitana	90°C	80°C	99°C	(Vieille, Krishnamurthy, Hyung-Hwan, et al., 2003)

Figure 2.1. AKs used in Chapter 2. A) Multiple sequence alignments are shown for the N-terminal fragments (red outline) and C-terminal fragments (blue outline) of the protein sequences for AKs from *E.coli* (not used in this study), *B. subtilis*, and *T. neapolitana*, using ClustalW and BOXSHADE. B) Splitting of the AK protein orthologues between residues D76 and C77 (shown as circle, using AK$_{Bs}$ numbering) produces an N-terminal AK fragment shown in red, containing the AMP binding site, and a C-terminal AK fragment shown in blue containing the majority of the CORE domain and the zinc-binding LID domain. The N- and C-termini of the full-length protein are labeled. For the full-length AK, the Ap5A substrate is shown in transparent magenta as sticks. For the full-length and C-terminal structures, the bound zinc ion is shown in orange as a sphere. From PDB 1ZIN, using PyMol (Delano Scientific, LLC).

2.3   Materials and Methods.

2.3.a   Reagents and Molecular Biology Materials.

E. coli XL1-Blue, and Rosetta 2 cells were from Stratagene and EMD Biosciences, respectively. Enzymes for DNA manipulation were obtained from Roche Biochemical, New England Biolabs, and Promega. Synthetic oligonucleotides were obtained from Fischer Scientific, and pET vectors were from EMD Biosciences. Bacterial growth media components were from BD Biosciences, and all other reagents were from Sigma-Aldrich.

2.3.b   Plasmids Constructed.

The gene encoding AK_{Bs} was amplified from genomic DNA using PCR with VENT DNA polymerase and cloned into pET21d using NcoI and HindIII restriction sites to create pAK_{Bs}. Previously described pET vectors for expressing AK_{Tn} (pAK_{Tn2}::Km) and AK_{Bs} with Q16L and Q199R mutations (designated $pAK_{Bs}^{Q16L/Q199R}$ herein) were used for studies involving full-length proteins (Counago et al., 2006; Vieille, Krishnamurthy, Hyun, et al., 2003). Vectors for producing AK_{Bs} fragments with associating polypeptide tags fused to their N-terminus were generated by chemically synthesizing genes encoding the IAAL-E3 and IAAL-K3 peptides (Litowski and Hodges, 2002) followed by a flexible linker of sequence GASGGGSSGGHM and ligating these to PCR amplified gene fragments encoding AK_{Bs} residues 1-76 (AK-N) and 77-217 (AK-C), respectively. In each of these gene fragments, three unique restriction sites (*nheI*, *xhoI*, and *ndeI*) were incorporated within codons of the linker. The IAAL-E3–AK-N and IAAL-K3–AK-C gene fusions were cloned into pET21d and pET24d using *ncoI* and *hindIII* cloning sites to generate pB-E3-N and pB-K3C, respectively. Vectors for producing AK_{Bs} fragments alone (pB-N and pB-C) were

created by converting the *ncoI* restriction sites in pB-E3N and pB-K3C to *ndeI* using Quikchange mutagenesis, removing the gene fragments encoding the IAAL-E3 and IAAL-K3 peptides and linker by *ndeI* digestion, and circularizing the vector by ligation. Q16L and Q199R mutations were introduced into pB-N, pB-E3N, pB-C, and pB-K3C by Quikchange mutagenesis to create pB-N^{Q16L}, pB-E3N^{Q16L}, pB-C^{Q199R}, and pB-K3C^{Q199R}, respectively. AK$_{Tn}$ gene fragments corresponding to residues 1-79 (AK-N) and 80-220 (AK-C) were amplified from pAK$_{Tn2}$::Km using VENT DNA polymerase, and cloned in place of the genes encoding structurally-related *B. subtilis* fragments in pB-N, and pB-C using *ndeI* and *hindIII* to create pTn-N and pTn-C, respectively. AK$_{Tn}$ gene fragments were also cloned in place of AK$_{Bs}$ fragments in pB-E3N and pB-K3C using *ncoI* and *hindIII* to create pT-E3N and pT-K3C, respectively. Vectors for expressing *B. subtilis* (pB-His-N and pB-His-C) and *T. neapolitana* (pT-His-N and pT-His-C) AK fragments with N-terminal (His)$_6$ tags were generated by subcloning AK gene fragments from pET21d (pB-N and pT-N) and pET24d (pB-C and pT-C) derived vectors into pET28b using *ndeI* and *hindIII*. All plasmids used this aim are presented in Table 2.3 on the next page.

Table 2.3 Plasmids used in Chapter 2.

Plasmid Name	Plasmid Backbone	Insert	Comment
pAK$_{Bs}$	pET21d	AK$_{Bs}$	Complementation positive control of the mesophilic AK$_{Bs}$.
pAK$_{Bs}$$^{Q16L/Q199R}$	pET11d	AK$_{Bs}$$^{Q16L/Q199R}$	Complementation positive control of the thermophilic AK$_{Bs}$$^{Q16L/Q199R}$.
pAK$_{Tn2}$::Km	pET21b	AK$_{Tn}$	Complementation positive control of the hyperthermophilic AK$_{Tn}$.
pB-N	pET21d	AK$_{Bs}$$^{1-76}$ fragment	Co-expression complementation plasmid, BsN fragment.
pB-C	pET24d	AK$_{Bs}$$^{77-217}$ fragment	Co-expression complementation plasmid, BsC fragment.
pB-E3-N	pET21d	E3 tag::AK$_{Bs}$$^{1-76}$ fragment	Co-expression complementation plasmid, E3::BsN fusion.
pB-K3-C	pET24d	K3 tag::AK$_{Bs}$$^{77-217}$ fragment	Co-expression complementation plasmid, K3::BsC fusion.
pB-N^{Q16L}	pET21d	AK$_{Bs}$$^{1-76,Q16L}$ fragment	Co-expression complementation plasmid, BsN^{Q16L} fragment.
pB-C^{Q199R}	pET24d	AK$_{Bs}$$^{77-217,Q199R}$ fragment	Co-expression complementation plasmid, BsC^{Q199R} fragment.
pB-E3-N^{Q16L}	pET21d	E3 tag::AK$_{Bs}$$^{1-76,Q16L}$ fragment	Co-expression complementation plasmid, E3::BsN^{Q16L} fusion.
pB-K3-C^{Q199R}	pET24d	K3 tag::AK$_{Bs}$$^{77-217,Q199R}$ fragment	Co-expression complementation plasmid, K3::BsC^{Q199R} fusion.
pT-N	pET21d	AK$_{Tn}$$^{1-79}$ fragment	Co-expression complementation plasmid, TnN fragment.
pT-C	pET24d	AK$_{Tn}$$^{80-220}$ fragment	Co-expression complementation plasmid, TnC fragment.
pT-E3-N	pET21d	E3 tag::AK$_{Tn}$$^{1-79}$ fragment	Co-expression complementation plasmid, E3::TnN fusion.
pT-K3-C	pET24d	K3 tag:: AK$_{Tn}$$^{80-220}$ fragment	Co-expression complementation plasmid, K3::TnC fusion.
pB-CheA^{P1P2}-N^{Q16L}	pET21d	TmCheA^{P1P2} tag::AK$_{Bs}$$^{1-76,Q16L}$ fragment	Co-expression complementation plasmid, TmCheA^{P1P2}::TnN fusion.
pB-CheY-C^{Q199R}	pET24d	TmCheY tag::AK$_{Bs}$$^{77-217,Q199R}$ fragment	Co-expression complementation plasmid, TmCheY::TnC fusion.
pB-His-N	pET30a	6xHis::AK$_{Bs}$$^{1-76}$ fragment	His-tagged BsN fragment for *in vitro* studies.
pB-His-C	pET28b	6xHis::AK$_{Bs}$$^{77-217}$ fragment	His-tagged BsC fragment for *in vitro* studies.
pT-His-N	pET30a	6xHis::AK$_{Tn}$$^{1-79}$ fragment	His-tagged TnN fragment for *in vitro* studies.
pT-His-C	pET28b	6xHis::AK$_{Tn}$$^{80-220}$ fragment	His-tagged TnC fragment for *in vitro* studies.

2.3.c The CV2 Complementation Assay.

The *E. coli* CV2 strain (*fhuA22, ΔphoA8, adk-2*(ts), *adkompF627*(T2R), *ompF627, ompF627*(T2R), *fadL701*(T2R), *fadL701, relA1, glpR2(glp^c^), glpD3, pitA10, spoT1, rrnB-2, mcrB1, creC510)* was used for all growth selections (Haase et al., 1989). Electrocompetent cells (designated CV2_T7) were generated that harbor pTara, a vector that produces T7 RNA polymerase under control of the araC promoter (Wycuff and Matthews, 2000). This vector constitutively produces T7 RNA polymerase when cells are grown on LB medium, thereby driving low level expression of transcripts controlled by T7 promoters. CV2_T7 exhibited a reversion frequency of <1 in 10^7 cfu.

To assess complementation, CV2_T7 were transformed using electroporation with plasmids encoding each AK fragment pair, and the transformed cells were grown at 30°C for 48 hours on LB-agar plates containing 17 μg/mL chloramphenicol, 50 μg/mL ampicillin, and 50 μg/mL kanamycin. Studies with full-length AK homologs were performed similarly with only those antibiotics required to select for maintenance of pTara and the AK-containing plasmid. Colonies obtained from initial plates were streaked onto LB-agar plates lacking antibiotic and incubated at 40°C to assess AK complementation. Colonies obtained from growth at 30°C were also used to inoculate 500 μL LB cultures in 96-well plates containing 17 μg/mL chloramphenicol, 50 μg/mL ampicillin, and 50 μg/mL kanamycin. These plates were grown for 24 hours at 30°C while shaking at 250 rpm. A 96-pin replicator was used to transfer 1 μL of cells from each well into plates containing fresh LB medium. These plates were grown at 40°C in a humidified incubator shaking at 250 rpm, and the cellular density was evaluated by measuring absorbance at 600 nm.

2.3.d Protein Expression and Purification.

Rosetta 2 *E. coli* transformed with pB-His-N, pB-His-C, pT-His-N and pT-His-C were grown in LB at 37°C, induced with 1 mM IPTG at $A_{600} \approx 1$ and grown for 6 to 8 hours at 25°C to allow for protein expression. Cells were harvested by centrifugation, resuspended in TND buffer (50 mM Tris pH 8.0, 300 mM NaCl, and 1 mM DTT) containing 25 mM imidazole, 0.1 mM PMSF, 1 mM $MgCl_2$, 300 µg/mL lysozyme, 2 U/mL DNase, and a complete protease inhibitor cocktail tablet (Roche Applied Science). Resuspened cells were frozen at −80°C for ≥2 hours, thawed, and centrifuged at 40k x g to remove cell debris. The cleared lysate was bound to NTA resin and washed with TND buffer containing 25 mM imidazole, 5 mM MgATP, and 12.5 µg/mL of denatured *E. coli* lysate to remove any ATP-dependent chaperones bound to the AK fragments (Rial and Ceccarelli, 2002). The denatured lysate was prepared from Rosetta 2 cells by boiling cleared cell lysate at 100°C for 10 minutes and then centrifugation at 14k rpm to remove all aggregates. AK fragments were eluted using a linear gradient of 0 to 250 mM imidazole in TND buffer and further purified using a combination of ion-exchange and gel filtration chromatography. The purified proteins were concentrated in TND buffer and stored at −70°C. SDS-PAGE analysis of each purified protein revealed that they were all ≥95% homogeneous.

Rosetta 2 *E. coli* transformed with pAK_{Tn2}:Km was grown in LB at 37°C, induced with 0.1 mM IPTG at $A_{600} \approx 1$ and grown for 18 h at 37°C to allow expression. Full length AK_{Tn} was purified using a combination of ion exchange, reverse phase, and size exclusion chromatography. AK_{Tn} was concentrated, dialyzed against TED buffer (50 mM Tris pH 8.0, 0.5 mM EDTA, and 1 mM DTT) and stored at −70°C.

2.3.e   The Adenylate Kinase Coupled Activity Assay.

AK activity was determined using an endpoint assay that measures the amount of ADP formed in a reaction containing ATP and AMP (Vieille, Krishnamurthy, Hyun, et al., 2003). Reactions (350 μL) containing HKM buffer (50 mM HEPES pH 7.4, 250mM KCl, and 4mM MgCl$_2$) and 0.0005 to 1 μM protein were incubated at 40°C for 5 min prior to initiating the reaction by adding AMP and ATP to final concentrations of 1 mM. After 2 to 5 minutes, reactions were quenched by placing them on ice and adding P^1,P^5-di(adenosine-5) pentaphosphate (55 μL) to a final concentration of 0.25 mM. A mixture (150 μL) containing 1.85 mM phosphoenolpyruvate, 0.925 mM NADH, and 2.5 units of lactate dehydrogenase in HKM buffer was added to the quenched reaction, and NADH consumption was monitored by following increases in absorbance at 340 nm after the addition of 2.5 units of pyruvate kinase. The amount of NADH consumed was calculated using an ε_{280} of 6,220 M^{-1}cm^{-1}, and the concentration of ATP consumed in the original reaction mixture was calculated as 0.793x[NADH]$_{consumed}$.

2.3.f   Far-UV Circular Dichroism to Measure Secondary Structure.

Protein concentrations were assessed using the Bradford method with BSA as a standard. UV-Vis absorbance was measured using SpectraMax M2 microplate and Cary 50 spectrophotometers. Far-UV circular dichroism (CD) spectra were recorded with a Jasco J-810 spectrapolarimeter using a pathlength of 0.1 cm. All CD spectra were acquired with proteins that were dialyzed into TD buffer (20 mM Tris pH 8.0 and 0.5 mM DTT). Thermal denaturation curves were obtained by monitoring changes in ellipticity at 222 nm. Protein unfolding was irreversible for all fragments.

2.3.g Determination of Zinc Content using the MMTS/PAR Assay.

The relative zinc content of AK fragments was determined by incubating 3 to 6 μM of each fragment with 100 μM PAR[12] in 50 mM HEPES pH 7.5 and evaluating the release of zinc upon addition of 200 μM MMTS[13]. The reaction was incubated at room temperature for 60 minutes, centrifuged to remove any precipitated protein, and changes in absorbance at 500 nm were monitored. The concentration of zinc was calculated using an extinction coefficient for the PAR-Zn complex (ϵ_{500} = 60,423 $M^{-1}cm^{-1}$) that was obtained in HEPES pH 7.5 using $ZnCl_2$ as a standard. The data reported represents the average of three replicates, each corrected for zinc levels in the buffer, with error corresponding to one standard error.

2.4   Results.

2.4.a   AK Protein Fragment Design.

To test whether thermostability affects protein fragment complementation, I have split AK orthologs from *Bacillus subtilis* and *Thermotoga neapolitana* that exhibit a range of thermostabilities and compared their structure and function at 40°C. AK_{Bs} and AK_{Tn} are monomeric zinc-binding enzymes that are functional at 30°C (Vieille, et al. 2003; Bae and Phillips 2004). However, they differ in the temperature profiles for their activity. AK_{Bs} exhibits maximal activity at ~45°C and is largely inactive at 60°C (Bae and Phillips 2004), whereas AK_{Tn} is maximally active at ~80°C and retains almost 50% of its activity at 100°C (Vieille, et al. 2003). An AK_{Bs} mutant (Q16L/Q199R) with intermediate thermostability was also used in these studies. This protein is maximally active at ~60°C and exhibits a midpoint

---

[12] 4-(2-Pyridylazo)resorcinol, reacts with free zinc to form a colorimetric complex.

[13] Methyl methanethiosulfonate, which alkylates cysteine residues; in this case freeing the bound zinc.

for thermal denaturation that is 14°C greater than AK_{Bs} (Counago, et al. 2006) and 30°C less than AK_{Tn} (Vieille, et al. 2003).

The AK_{Bs} structure shown previously in Figure 2.1 (page 47) illustrates the site of bisection and the polypeptide fragments used for complementation studies. All AK orthologs have been split to produce N-terminal fragments (designated AK-N) that contain the AMP binding site and C-terminal fragments (designated AK-C) that contain all of the residues involved in zinc coordination and the mobile LID that mediates substrate entry into the active site (Schulz, et al. 1990; Muller et al., 1996). Both fragments contain residues from the hydrophobic core of the protein. This fragmentation site was chosen based on a previous study that showed that AK_{Ec} can be similarly cleaved at this site *in vitro* to create polypeptides that recover partial catalytic activity when reconstituted at 23°C (Saint Girons, et al. 1987). However, AK_{Ec} was avoided herein because functional complementation was assayed in an *E. coli* strain with a temperature-sensitive AK (Haase, et al. 1989). Complementation assays using non-native AK homologs are less likely than AK_{Ec} to lead to chromosomal recombination and reversion of the temperature-sensitive phenotype.

2.4.b   AK Fragment Complementation

To initially compare the activities of split AK homologs, I examined whether they could complement the growth of *E. coli* CV2, a strain with a temperature-sensitive AK (Pro87Ser) that cannot support growth at 40°C (Haase, et al. 1989). Genes encoding each AK-N and AK-C fragment were cloned into pET vectors with different antibiotic selection markers. In pET vectors, AK fragment expression is controlled by the T7 promoter, and all studies involving *E. coli* CV2 used competent cells that harbor pTara (designated $CV2_{T7}$), a

plasmid that constitutively produces T7 RNA polymerase when cells are grown on LB medium (Wycuff and Matthews, 2000).

To evaluate the relative activities of split AK homologs, $CV2_{T7}$ were transformed with vectors for expressing three full-length AK homologs (pAK_{Bs}, $pAK_{Bs}^{Q16L/Q199R}$, and $pAK_{Tn2}::Km$) and their corresponding fragment pairs (pB-N/pB-C, $pB-N^{Q16L}/pB-C^{Q199R}$, and pT-N/pT-C). Cells were selected for growth at 30°C on LB-agar plates containing antibiotics that maintained the desired sets of plasmids, colonies obtained from selections were streaked on LB-agar plates lacking antibiotics, and $CV2_{T7}$ growth was evaluated under selective (40°C) and non-selective (30°C) conditions.

Figures 2.2 (page 54) and 2.3 (page 55) shows the complementation of $CV2_{T7}$ by full-length and fragmented AK homologs, respectively. Complementation occurred at 40°C after 24 hours with $CV2_{T7}$ harboring plasmids for producing *T. neapolitana* AK-N and AK-C fragments, similar to that obtained with $CV2_{T7}$ expressing full-length AK_{Tn}, AK_{Bs}, and $AK_{Bs}^{Q16L/Q199R}$. In addition, cells harboring the $AK_{Bs}^{Q16L/Q199R}$ fragment pair derived from the AK with intermediate thermostability appeared at a low density after 72 hours. However, no growth occurred after 72 hours with $CV2_{T7}$ alone or $CV2_{T7}$ expressing the split AK_{Bs}, although these cells were viable at 30°C.

Figure 2.2 Complementation of the *E.coli* CV2$_{T7}$ strain by full-length AK orthologs. Each plate is streaked with CV2$_{T7}$ cells containing (clockwise from the top of the plate) no plasmid, *B. subtilis* AK, thermostable *B. subtilis* AK mutant with Q16L and Q199R mutations, and *T. neapolitana* AK. Duplicate streaked plates were grown at 30°C and 40°C for 24 hours.
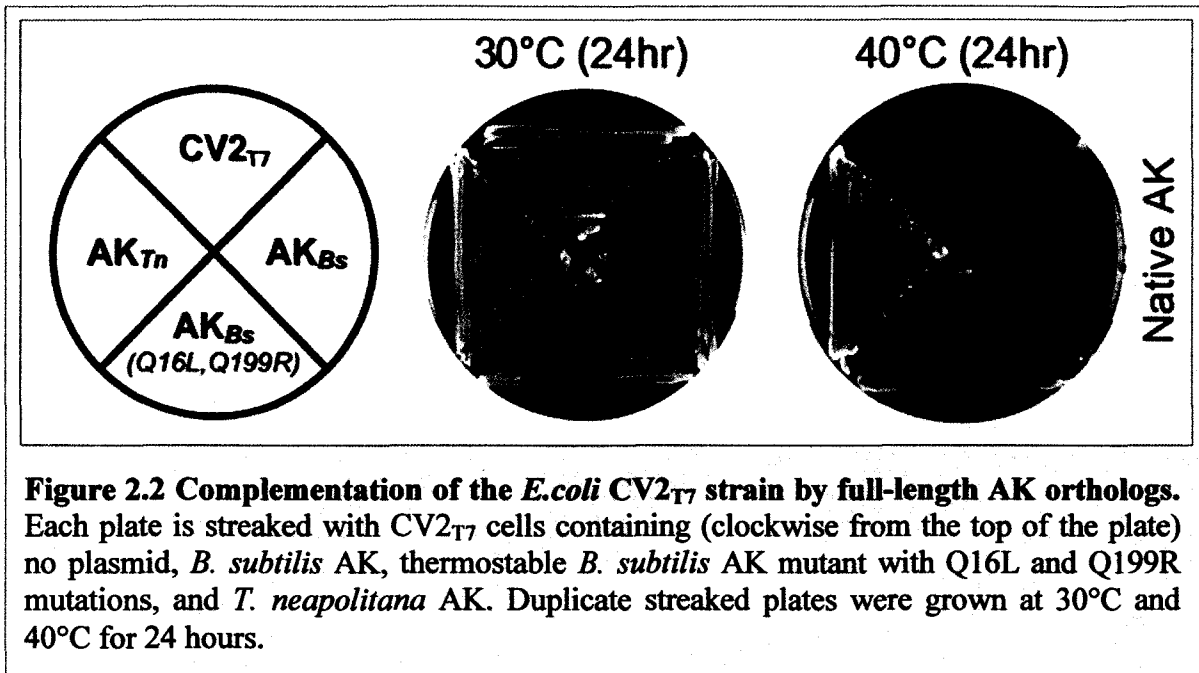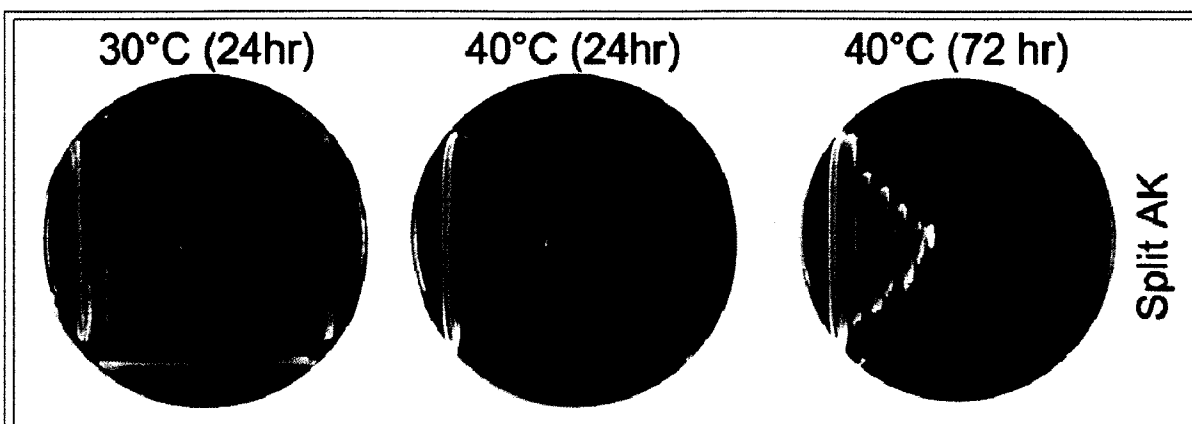
Figure 2.3 Complementation of the *E.coli* CV2$_{T7}$ strain by split AK orthologs. Each plate is streaked with transformed CV2$_{T7}$ cells containing (clockwise from the top) untransformed cells, plasmids coexpressing split fragments from the mesophilic *B. subtilis* AK, split fragments from the thermophilic *B. subtilis* AK mutant with Q16L and Q199R mutations, and split fragments from the hyperthermophile *T. neapolitana* AK (same plate outlay as that used for Figure 2.2, except with split AKs). Each AK was split at homologous sites cooresponding to residues 76 and 77, using the *B. subtilis* AK numbering convention. Duplicate streaked plates were grown at 30°C for 24 hours and 40°C for 24 and 48 hours.

2.4.c Assisted Protein Fragment Complementation.

Fragment complementation can frequently be improved by fusing the inactive fragments to proteins that associate (Johnsson and Varshavsky 1994; Remy and Michnick 1999; Pelletier, et al. 1998; Wilson, et al. 2004; Nyfeler, et al. 2005; Magliery, et al. 2005; Remy and Michnick 2006). To test whether this strategy can promote AK-fragment function, I fused the AK_{Bs} fragments to the IAAL-E3 and IAAL-K3 peptides that strongly associate ($K_D = 70$ nM) to form a heterodimeric coiled-coil (Litowski and Hodges 2002) and evaluated whether these tagged AK fragments exhibit improved $CV2_{T7}$ complementation. The pET vectors created for expressing these polypeptide fusions were designed so that each AK fragment has the associating polypeptides fused to their N-terminus through a twelve-residue linker that is predicted to be flexible (Linding, et al. 2003).

Figure 2.4 on the following page shows the effect of IAAL-E3 and IAAL-K3 on the function of split AK_{Bs}. When fused to these associating tags, $AK-N^{Q16L}$ and $AK-C^{Q199R}$ complemented $CV2_{T7}$ growth at 40°C after 24 hours, as did $AK-N^{Q16L}$ and AK-C. In contrast, $CV2_{T7}$ expressing the native AK-N and AK-C fusions exhibited weak complementation after 48 hours, although these cells grew readily after 24 hours at 30°C under non-selective conditions. A similar weak growth phenotype was observed at 40°C for $CV2_{T7}$ harboring plasmids for producing AK-N and $AK-C^{Q199R}$ fused to associating tags.
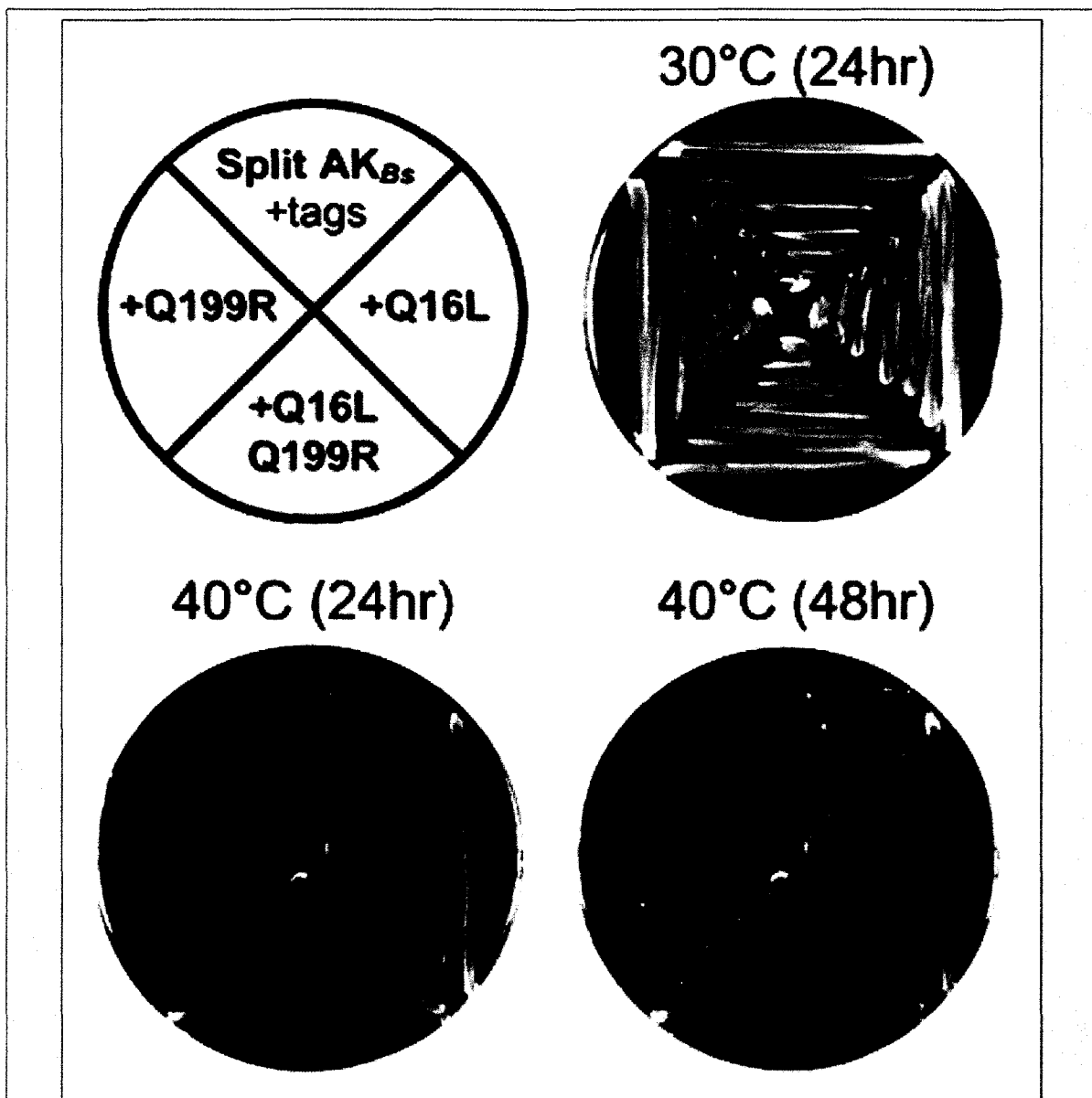
Figure 2.4 Complementation of the *E.coli* CV2$_{T7}$ strain by split AK orthologs fused to interacting E3 and K3 coiled-coils. Each plate is streaked with cotransformed CV2$_{T7}$ cells containing (clockwise from the top) plasmids coexpressing: E3-N terminal and K3-C terminal AK fusion proteins from *B. subtilis* AK, thermostable *B. subtilis* AK mutant with Q16L and Q199R mutations, and *T. neapolitana* AK. Duplicate streaked plates were grown at 30°C for 24 hours and 40°C for 24 and 48 hours.

2.4.d Complementation by Hybrid AK Fragments.

To examine which AK fragments cooperatively function in liquid cultures, I evaluated $CV2_{T7}$ complementation in 96-well microtiter plates for all possible AK-N and AK-C combinations. For these assays, colonies selected to grow at 30°C on LB-agar plates were used to inoculate 200 µL LB cultures in microtiter plates containing ampicillin, chloramphenicol, and kanamycin. After 24 hours of growth at 30°C, these cultures were replicated into fresh microtiter plates lacking antibiotic and grown for 72 hours in a humidified shaking incubator at 40°C. Complementation was evaluated at various times by measuring optical density at 600 nm.

Figure 2.5.A on page 60 shows the growth of $CV2_{T7}$ transformed with all possible combinations of AK-N and AK-C fragments. As with the selections on LB-agar medium, complementation was observed when cells contained vectors for expressing *T. neapolitana* AK-N and AK-C. However, *B. subtilis* AK fragments containing the Q16L and Q199R mutations did not weakly complement growth in liquid medium as was observed on LB-agar plates. This indicates that the liquid growth selections are more stringent than those on agar plates.

Figure 2.5.B on page 60 shows the results from complementation studies that used AK fragments fused to the IAAL-E3 and IAAL-K3 associating polypeptides. Again, the AK_{Tn} fragments complemented $CV2_{T7}$, with maximal growth occurring within 24 hours. In addition, complementation was observed for cells coexpressing *B. subtilis* AK-N^{Q16L} and AK-C, *B. subtilis* AK-N^{Q16L} and AK-C^{Q199R}, and *B. subtilis* AK-N^{Q16L} and *T. neapolitana* AK-C. These findings show that only AK_{Bs} fragments with the Q16L mutation can form a functional split protein when selection are performed in liquid medium, and they demonstrate

that only AK_{Bs} fragments with this mutation can cooperatively function with an AK_{Tn} fragment.

Figure 2.5 Liquid culture analysis of AK fragment complementation. The growth of $CV2_{T7}$ cells transformed with combinations of AK-N and AK-C fragments was assessed in the absence (A) and presence (B) of associating IAAL-E3 and IAAL-K3 polypeptide tags. Growth was monitored in LB medium at 40°C in a humidified shaker by measuring absorbance at 600 nm. Each experiment was performed eight times, and corrected for the absorbance of the LB medium. Error is shown as standard error.

2.4.e Demonstrating the AK$_{Bs}$-LR PCA system Using Natural Proteins.

In the previous sections, I used artificially designed heterodimeric polypeptide tags for benchmarking of the complementation ability of the various AK fragments. The most promising candidate pair was the BsN^{Q16L} and BsC^{Q199R} fragments which exhibited a strong signal when fused to the interacting tags and a very low background complementation when coexpressed as fragments only. To demonstrate the utility of the BsN^{Q16L} and BsC^{Q199R} system as a PCA system using naturally occurring proteins, I replaced the artificial IAAL-E3 and IAAL-K3 tags with the interacting chemotaxis proteins from *Thermotoga martima*, TmCheA^{P1P2} and TmCheY. As shown in Figure 2.6 on the next page, CV2$_{T7}$ cells transformed with these constructs and streaked onto LB-agar plates displayed strong complementation only when the fragments were fused to the interacting proteins. This is the first published PCA system that has been engineered from AK proteins.
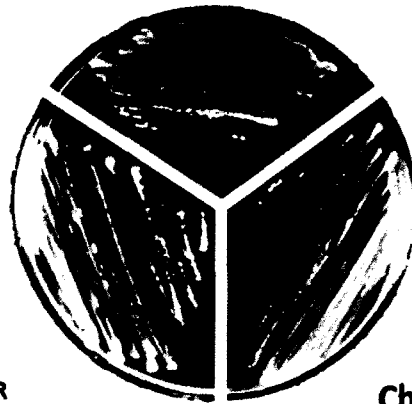
Figure 2.6 Demonstration of the AK-based PCA system using the interacting proteins, CheA^{P1P2} and CheY from *T. maritima*. Both plates are streaked with (clockwise from top) untransformed CV2$_{T7}$ cells, CV2 $_{T7}$ cells cotransformed with CheY / CheA^{P1P2} fusions to the thermostable AK$_{Bs}$ mutant fragments, and CV2 $_{T7}$ cells cotransformed the thermostable AK$_{Bs}$ mutant fragments without fusion tags. Ducplicate plates are shown for incubation at 30°C for 24hours (top) and at 40°C for 24hours (bottom).

2.4.f Protein Expression and Purification of the AK_{Bs} and AK_{Tn} Fragments.

To obtain AK fragments for *in vitro* studies, I created IPTG-controlled expression plasmids that produce AK-N (pB-His-N and pT-His-N) and AK-C (pB-His-C and pT-His-C) fragments with a $(His)_6$ affinity tag at their N-terminus. The addition of a His tag to AK_{Tn} fragments did not affect their ability to complement $CV2_{T7}$ *E. coli*. Cells (10^8 cfu) cotransformed with pT-His-N and pT-His-C yielded a large number of colonies (10^5 cfu) at 40°C after 48 hours.

AK-N and AK-C fragments were overexpressed in *E. coli* growing in LB medium with 1 mM IPTG and purified using a similar combination of NTA-affinity, anion-exchange, and size-exclusion chromatography in the absence of zinc chelators like EDTA. All fragments were soluble. Whereas the yields for the purified AK-C fragments were ≤1.5 mg/L culture medium, the yields for AK-N fragments were ~5-fold lower (≤0.3 mg/L). The final purified protein was determined by SDS-PAGE to be >95% pure (Figure 2.7, next page).
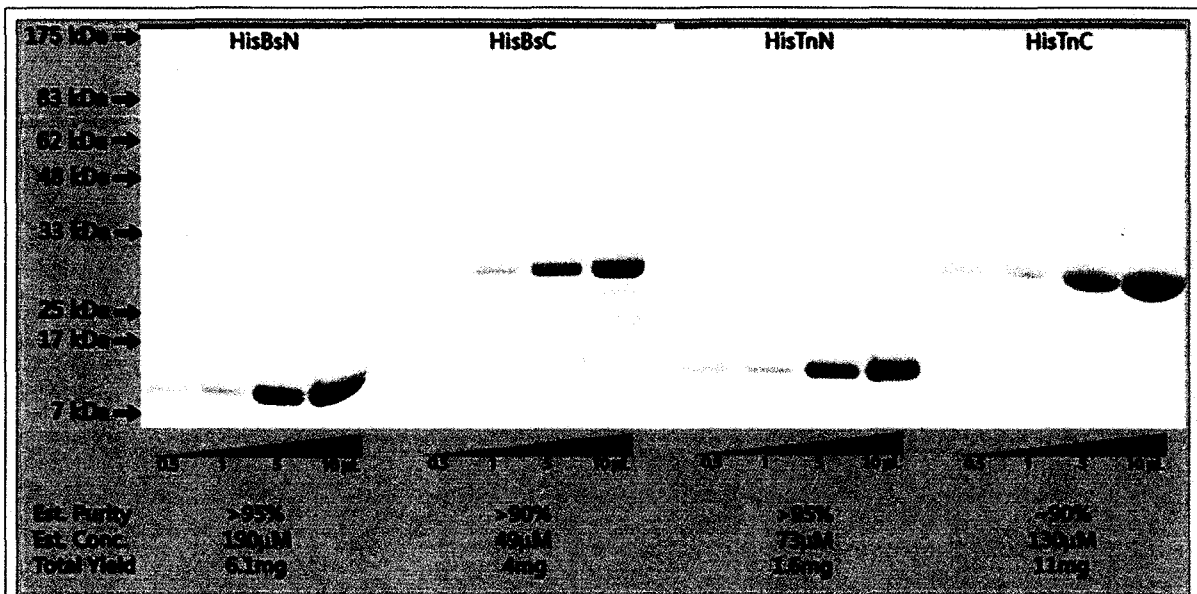
Figure 2.7 His-tagged AK$_{Bs}$ and AK$_{Tn}$ fragments purified by FPLC. Purified His-tagged (from left to right) *B.subtilis* N-terminal AK fragment (1-76), *B.subtilis* C-terminal AK fragment (77-217), *T.neapolitana* N-terminal AK fragment (1-79), and *T.neapolitana* C-terminal AK fragment (78-220). All samples were analyzed with 0.5, 1, 5, and 10 μL of the final sample to estimate purity. Concentrations were estimated using extinction coefficients determined from the protein sequence.

2.4.g Catalytic Activity of the AK fragments.

To examine how fragmentation affects AK function, I compared the catalytic activity of the split and full-length AK homologs. Adenylate kinase activity is typically measured using a coupled end-point assay (Vieille, Krishnamurthy, Hyun, et al., 2003), which I employed for these measurements. The reverse reaction (2ADP $\rightarrow$ ATP + AMP) is detected using lactate dehydrogenase and pyruvate kinase, which produces the spectophotometrically active NADH (Abs_{max} = 340 nm). Figure 2.8 on the next page shows the first-order rates obtained at 40°C in the presence of AMP and ATP concentrations (1 mM) that that are >10-fold higher than the K_m values reported for the full-length proteins. At this temperature, full-length AK_{Tn} and AK_{Bs} displayed similar rates of 1.5 x 10^4 ± 0.1 min^{-1} and 1.4 x 10^4 ± 0.2 min^{-1}, respectively. The *T. neapolitana* AK-N and AK-C fragments also had detectable activity (0.09 x10^4 ± 0.01 min^{-1}) when mixed together. The rate observed for this fragment mixture was ~17-fold lower than that of AK_{Tn} and ~300-fold greater than the sum of the rates obtained for either of the fragments alone, indicating that these fragments cooperatively function *in vitro*. In contrast, a mixture containing equimolar levels of the *B. subtilis* AK fragments exhibited a rate (0.8 ± 0.2 min^{-1}) that was <4-fold greater than the sum of the rates obtained for the fragments and >10,000-fold lower than either of the full-length AK homologs. These findings are consistent with those from bacterial selections showing that the split AK_{Bs} is not functional *in vivo* at 40°C.
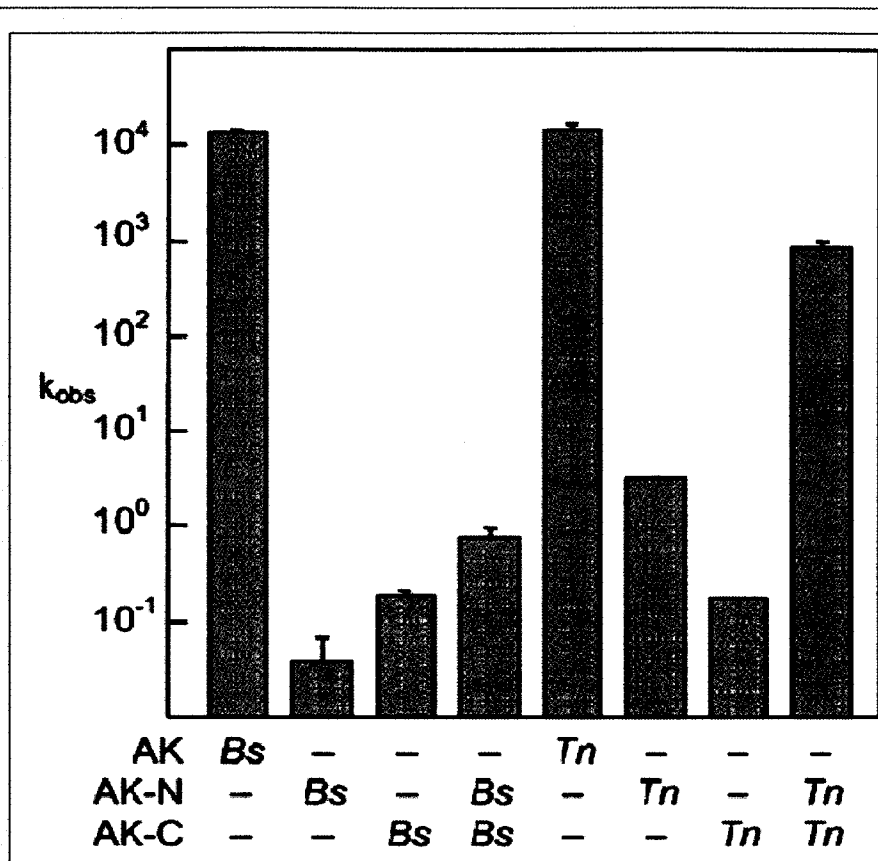
Fig. 2.8 AK and split AK enzymatic activities. Specific activities of *B. subtilis* and *T. neapolitana* AK, AK-N, AK-C, and AK-N/AK-C mixtures measured at 40°C in HKM buffer containing 1 mM AMP and ATP. Protein concentrations in reactions were varied so that each reaction consumed no more than 5% of the substrates over the time course of the assay. The concentrations were: 0.5 nM AK_{Bs}, 0.5 nM AK_{Tn}, 10nM *T. neapolitana* AK-N/AK-C, 1 mM *B. subtilis* AK-N/ AK-C, and 1 mM for measurements involving any individual AK fragment. Each experiment was performed in triplicate and corrected for background levels of NADH degradation. Error shown as standard error.

2.4.h Secondary Structure of the AK Fragments by far-UV CD.

The disparity in the functions of the AK fragments could arise because the *T. neapolitana* fragments retain greater secondary structure and more readily self-associate. To establish the relative secondary structure for each AK fragment, far-UV circular dichroism spectra were obtained. Figure 2.9.A on page 69 shows the mean residue ellipticity for each AK-N over a range of wavelengths. Comparison of these spectra to a reference set of proteins using the K2D algorithm predicts that *B. subtilis* and *T. neapolitana* AK-N both contain low levels of α helix, ~16 and 12%, respectively. This is much lower than the level of α helix (69%) observed in the AK$_{Bs}$ crystal structure (PDB structure 1P3J), suggesting that these fragments are largely unstructured. Figure 2.9.B on page 69 shows the mean residue ellipticity for each AK-C fragment over a range of wavelengths. *B. subtilis* AK-C displays a strong negative peak at 203 nm, suggesting that this polypeptide is largely disordered. *T. neapolitana* AK-C, in contrast, has clear peaks at 208 and 222 nm, consistent with the presence of greater helical content. However, a comparison of this spectrum to a reference set of proteins using the K2D algorithm suggests that its overall α helix content (21%) is still significantly lower than the 50% α helix predicted for this fragment from the structure of other AK homologs. Nonetheless, the far-CD analysis of the individual fragments clearly demonstrates a substantial difference in the folded state of the *B. subtilis* and *T. neapolitana* AK-C fragments that may indicate differences in the *in vivo* steady-state levels of the protein. To determine whether the higher catalytic activity of the *T. neapolitana* fragment pair arises because these fragments associate more readily than the *B. subtilis* fragment pair, I investigated whether the CD spectra of these fragment pairs change upon mixing. Figure 2.10.A on page 70 shows the spectrum obtained at 40°C for a mixture containing *B. subtilis*

AK-N and AK-C (10 µM) and the spectrum calculated for a mixture using the spectra acquired with the individual fragments. These spectra are similar indicating that the AK_{Bs} fragments do not acquire increased parent-like structure when mixed at this concentration. In contrast, the spectrum obtained under similar conditions for a mixture of *T. neapolitana* fragments (10 µM) at 40°C exhibits a 17% decrease in ellipticity at 222 nm compared with the spectrum calculated from measurements of the individual fragments (see Figure 2.10.B on page 70). This indicates that the *T. neapolitana* fragments self associate to some degree upon mixing and acquire increased levels of parent-like secondary structure. In addition, they suggest that these fragments exhibit a higher binding affinity than the *B. subtilis* fragment pair. However, the ellipticity observed for the mixture is less than that of AK_{Tn} at 222 nm (-15.2 x10^{-3} deg·cm^2/dmol), indicating that the split enzyme has less overall helical content than the full-length AK in the absence of substrates. Thermal denaturation profiles by CD analysis of the AK fragments show enhanced stability when the *T. neapolitana* fragments are mixed, in comparison to the fragments alone. However, the unfolding curve was shown to be irreversible, preventing clear assignment of the changes in CD signal to the folding/unfolding process rather than off-pathway events such as aggregation.
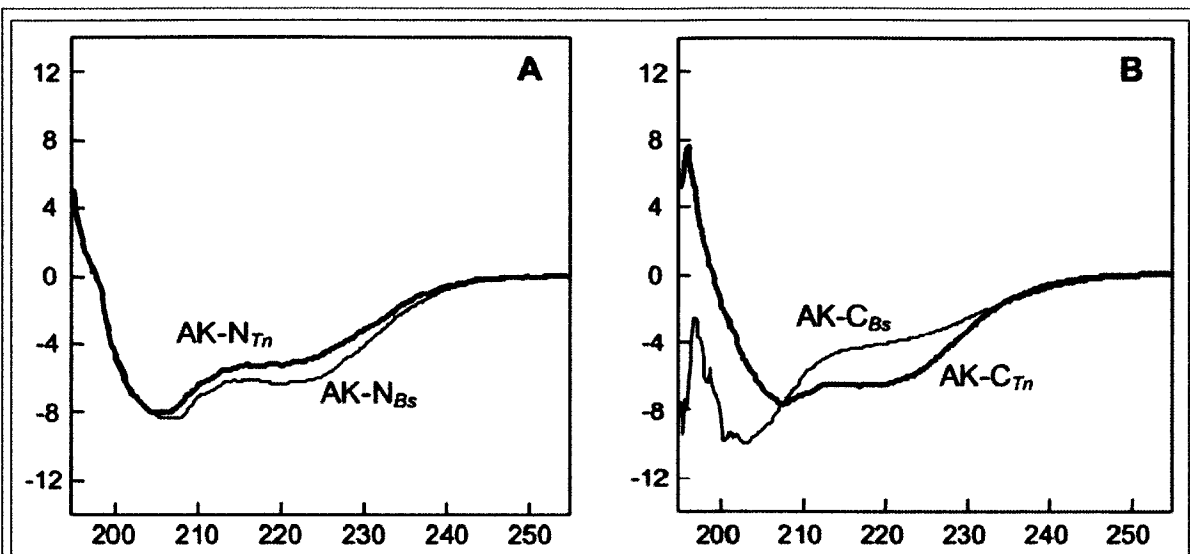
Figure 2.9 Far-UV circular dichroism measurement of secondary structure for N- and C-terminal fragments. A) A comparison of CD spectra for AK_{Tn} and AK_{Bs} N-terminal fragments. B) A comparison of CD spectra for AK_{Tn} and AK_{Bs} C-terminal fragments. All graphs are shown with wavelength (λ) on the abcissa and mean-residue ellipticity ($\times 10^{-3}$ deg•cm^2/dmol) on the ordinate.
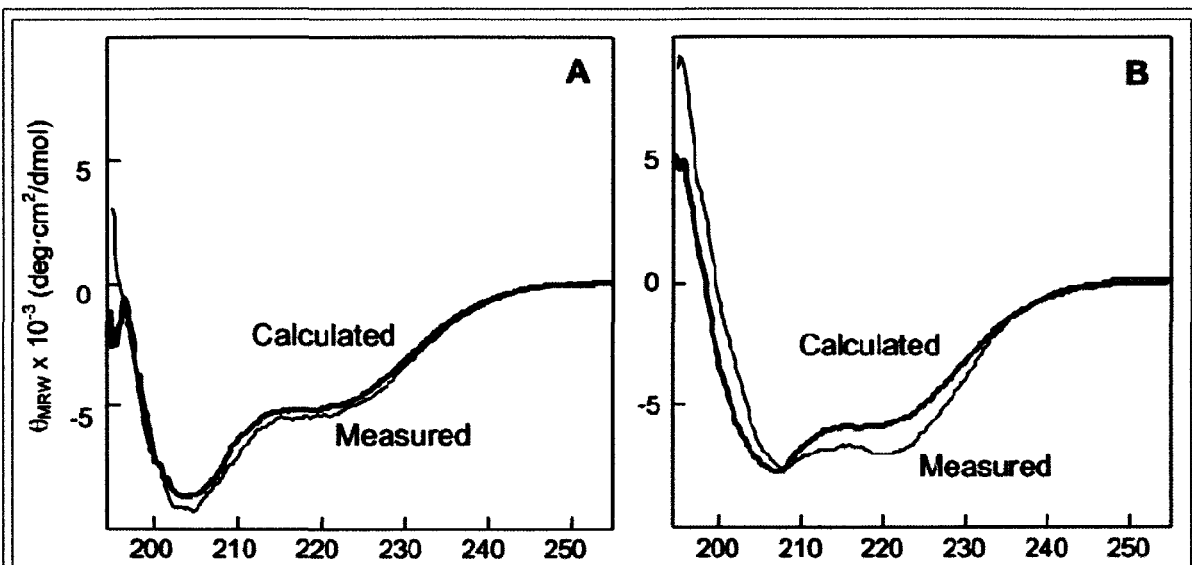
Fig. 2.10 Structural changes upon mixing of AK fragments. Far-UV CD spectra for mixtures (–) of *B. subtilis* (A) and *T. neapolitana* (B) AK-N and AK-C fragments were compared with the spectrum calculated (—) by averaging ellipticity values from measurements with the individual fragments. Spectra were obtained using 10 mM of each polypeptide in buffer containing 20 mM Tris pH 8.0 and 0.5 mM DTT. All graphs are shown with wavelength (λ) on the abcissa and mean-residue ellipticity ($\times 10^{-3}$ deg•cm^2/dmol) on the ordinate.

2.4.i Zinc Content of the AK Fragments.

AK_{Bs} and AK_{Tn} both coordinate a zinc ion and exhibit increased thermostability when zinc is bound, suggesting that the differences in split protein structure and function could arise from changes in zinc stoichiometry. Whereas AK_{Bs} uses three cysteines and an aspartate residue in its AK-C fragment to coordinate zinc, AK_{Tn} uses four cysteines found at structurally-homolgous positions. To quantify the relative zinc levels in each AK-C fragment, the purified polypeptides were mixed with a 40-fold excess of MMTS, a reagent that reacts specifically with sulfhydryl groups, and zinc release was monitored using PAR, a zinc chelator that exhibits increased absorbance at 500 nm upon zinc binding. When MMTS was added to the *T. neapolitana* AK-C in the presence of excess PAR, the absorbance increased to a level that corresponded with 0.84 ± 0.02 equivalents of zinc released per AK-C. Similar experiments with *B. subtilis* AK-C revealed that this protein contains a reduced number of equivalents of zinc (0.62 ± 0.01). However, the presence of >50% bound zinc in this fragment suggests that the reduced activity of the split AK_{Bs} does not arise from differences in zinc binding.

2.5 Discussion

In this chapter I present evidence that the level of AK fragment complementation is directly proportional to the stability of the parents bisected: $AK_{Tn} > AK_{Bs}^{Q16L/Q199R} > AK_{Bs}$. *E. coli* $CV2_{T7}$ growth at 40°C was only complemented by AK fragments derived from the hyperthermophilic AK_{Tn}, which exhibits a mid-point for thermal denaturation (T_m) that is near the boiling point of water (Vieille, Krishnamurthy, Hyun, et al., 2003). In contrast, polypeptides derived from the mesophilic *B. subtilis* homolog with the lowest thermostability

(T_m = 51-55°C) were non-functional under similar conditions, even though full-length AK_{Bs} displayed similar steady-state activity as AK_{Tn} at the temperature where growth selections were performed. Weak AK_{Bs} fragment complementation occurred on LB-agar plates when the Q16L and Q199R mutations were incorporated that increase the T_m for the full-length protein by 14°C (Counago et al., 2006). The growth of $CV2_{T7}$ containing these fragments on LB-agar was slower than that of cells containing split AK_{Tn}, and they could not complement growth in liquid medium like AK_{Tn}.

The enhanced complementation observed when AK_{Bs} fragments were fused to associating IAAL-E3 and IAAL-K3 peptides suggests that split AK proteins can be used as an assay to report on protein-protein interactions in organisms where AK selections are available (Counago et al., 2005). In addition, these findings provide further evidence that parental thermostability influences the cooperative function of protein fragments. *E. coli* $CV2_{T7}$ growth was poorest on LB-agar plates when cells harbored tagged AK_{Bs} fragment pairs that lacked mutations and those containing the Q199R mutation, which only increases the T_m for AK_{Bs} <2°C (Counago et al., 2006). In contrast, *E. coli* $CV2_{T7}$ growth was strongest for tagged AK_{Bs} fragments containing the Q16L mutation, which is thought to be more thermostabilizing than Q199R (Counago et al., 2006).

My findings are consistent with those from a study examining the folding and function of AK chimeras created by recombining mesophilic and thermophilic homologs (Bae et al., 2006). With these chimeras, those having only their counterweight loops derived from residues in the mesophilic AK exhibited the smallest reduction in thermostability (Bae et al., 2006). This is similar to the finding that peptide backbone cleavage within this region can yield a split AK_{Tn} that is functional like the full-length protein. In addition, protein

thermostability was most strongly correlated with the fraction of the core residues inherited from the thermophilic parent in these chimeras (Bae et al., 2006). These residues are found in both the AK-N and AK-C fragments, and of the AK_{Bs} fragments assayed, only the polypeptide containing the most thermostabilizing mutation (Q16L) functioned cooperatively with a *T. neapolitana* AK fragment.

Some hyperthermophilic enzymes are thought to achieve their thermostability through increased residual structure in their unfolded state (Robic et al., 2003). My results provide evidence that this occurs when AK homologs having different thermostability are bisected at a single structurally-related position. Whereas the N-terminal fragments from AK_{Bs} and AK_{Tn} both had similar low levels of secondary structure, the C-terminal fragment derived from *T. neapolitana* AK exhibited greater parent-like helical content compared with the *B. subtilis* fragment. The AK_{Tn} C-terminal fragment also contained a higher fraction of bound zinc. Previous studies examining the role that zinc plays in regulating AK thermostability have shown that zinc removal reduces the T_m of *Geobacillus stearothermophilus* and *T. neapolitana* AK by 7.5°C and 6.2°C, respectively . This suggests that the reduced activity of the split AK_{Bs} at 40°C compared with the split AK_{Tn} could arise in part because AK_{Bs} has a lower amount of bound zinc. However, I was unable to directly assess the influence of zinc removal on the activity of the split AK homologs because the purified C-terminal fragments aggregated upon displacement of the zinc by MMTS.

Previous complementation studies with fibronectin type III domain and β-lactamases provided anecdotal evidence that protein thermostability and fragment complementation are correlated in single domain proteins. With the fibronectin type III domain, stabilizing mutations were identified using a yeast-two hybrid assay that improved complementation

among fragments derived from a destabilized mutant (Dutta et al., 2005). When these mutations were incorporated into the full-length domain containing the same mutation, they all increased protein thermostability. In addition, the incorporation of a mutation shown to thermostabilize β-lactamases by 2.7 kcal/mol (Wang et al., 2002) led to a more sensitive protein fragment complementation assay (Galarneau et al., 2002). When lactamase fragments were fused to interacting proteins, they had increased catalytic activity. The results presented herein are the first to show that the cooperative function of protein fragments can be tuned by choosing naturally-occurring protein orthologs with different thermostabilities.

Protein evolution is driven by a variety of processes, including random mutation, intragenic recombination, domain insertion, and domain fission. Thermostability is known to influence a protein's tolerance to amino acid substitutions created randomly and by recombination (Besenmatter et al., 2007; Bloom, Labthavikul, et al., 2006; Bloom et al., 2005; Meyer et al., 2003). Thermodynamic models have proposed that this arises because random substitutions produce similar average effects on the free energy of folding for proteins with the same topology (Bloom et al., 2005). Thus, when protein homologs incur similar mutations, the homolog with the lowest free energy of folding (i.e., greatest thermostability) is least likely to have its structure destabilized and its function altered. These results suggest that a similar rationale can be used to explain the effects of backbone bisection on protein function. I hypothesize that upon fission of homologous proteins at structurally-related sites, fragmentation has similar destabilizing effects on protein homologs, and proteins with increased thermostability will on average be more likely to retain parent-like structure and function. Therefore, a quick way to create protein fragments with a range

of cooperative functions is to select structurally-related parent proteins with a range of thermostabilities. Such an approach is expected to accelerate the directed evolution of protein fragments that cooperatively function (Ostermeier et al., 1999) for uses in proteomics (Michnick, 2003), drug discovery (Michnick et al., 2007), and synthetic biology (Giesecke et al., 2006).

One question raised by these observations is whether the split AK_{Tn} can complement an extremophile with a temperature-sensitive AK (Counago et al., 2005). Ongoing studies are evaluating how temperature affects AK fragment structure and function and examining whether a split AK could be used to generate protein fragment complementation assays for thermophiles (60 to 80°C) and hyperthermophiles ($\geq$80°C) (Rothschild and Mancinelli, 2001). The development of such assays should facilitate the discovery of protein-protein interactions in microbes where classical two-hybrid systems have had limited success mapping genome-wide macromolecular interactions (Usui et al., 2005).

Chapter 3

Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

3.1 Abstract

The tolerance of protein structure and function to random amino acid substitutions is positively correlated with its thermostability. However, it remains unclear whether proteins with enhanced thermostability are also more tolerant to *random* fission events on average, similar to that observed for a single fission site in Chapter 2. To test this, I created libraries of *Bacillus subtilis* and *Thermotoga neapolitana* adenylate kinase (AK_{Bs} and AK_{Tn}, respectively) expression vectors that have *notI* randomly inserted within each *adk* gene using the MuA transposase. I then subcloned a regulatory element (stop codon, promoter, and ribosomal binding site fusion) into the *notI* sites to create libraries of vectors that coexpress the AK polypeptides encoded before and after the *notI* site, and evaluated the function of split-protein variants in each library. I show that the library created using the more thermostable AK_{Tn} contains 7-fold more variants that function at 40°C than library creating using AK_{Bs}. Sequencing of the vectors from both libraries encoding functional variants identified fission sites within the AMP binding and LID domains of AK_{Tn}, but only in the AMP binding domain of AK_{Bs}. In addition, a comparison of seven homologous AK_{Bs} and AK_{Tn} split proteins further revealed that split proteins derived from the more thermostable homolog exhibit greater activity at 40°C. These findings show that the discovery of polypeptide fragments that cooperatively function can be accelerated by splitting thermophilic proteins, and they suggest that non-disruptive fission sites discovered within thermophilic proteins could be used in less thermostable protein homologs to create protein fragment complementation assays.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

76

3.2   Introduction

In protein engineering, libraries of protein variants are frequently screened to find mutants that have acquired novel functions (Arnold, 1998; Crameri et al., 1998; Giver et al., 1998; Neylon, 2004). In contrast to rational approaches, this 'directed evolution' methodology identifies desirable protein variants through highly parallel and combinatorial processes. Previous research has demonstrated that processes that generate diversity, such as random mutations and recombination, often have deleterious effects on the protein structure and function (Besenmatter et al., 2007; Bloom et al., 2005; Bloom, Labthavikul, et al., 2006; Meyer et al., 2006). This is due to the fact that many residues in a protein of interest have been evolutionarily fixed to confer a stability that is optimal for the fitness of an organism; random changes to the residues in this context will typically generate globally unstable variants (Bloom, et al., 2005). Hence, on average, random mutation results in unfolded and nonfunctional proteins, with increasing multiple substitution events leading to an exponential decline in functional proteins (Daugherty et al., 2000).

Protein thermostability has recently been found to play an important role in determining the fraction of functional variants in libraries created by random mutation and recombination (Bloom, et al., 2006). In directed evolution experiments that assessed the tolerance of protein homologs with differing thermostability to random mutations, Bloom and coworkers showed that variants derived from the more thermostable homolog could tolerate more random amino acid substitutions without losing function than the less stable protein (Bloom, et al., 2006). These and other findings have led to the idea that increased protein stability can at times act as a buffer against deleterious mutations, for example, a thermostable mutant of TEM1 beta-lactamase was found to have a higher fraction functional

when subjected to random mutations, in comparison to the wildtype enzyme. Thus, in generating libraries for directed evolution, proteins which are more thermostable are predicted to have a greater evolutionary capacity (Bloom, Raval, et al., 2006; Bloom et al., 2005, 2007) and anticipated to yield a higher fraction of folded and potentially useful proteins. Consistent with this prediction, recent studies have shown that thermostability increases the fraction of functional protein chimeras in libraries that are generated by recombination, an evolutionary process that is more disruptive than random mutation on a per mutation basis (Meyer et al., 2006). Currently, it is not known how thermostability affects other types of mutational processes, such as random fission events.

In Chapter 2, I have shown that protein thermostability influences the cooperative reassembly of AK orthologs split at one homologous site. I found that a split AK_{Tn} derived from a hyperthermophile is functional at 40°C and able to complement an AK deficiency in *E. coli*. In contrast, homologous AK_{Bs} fragments derived from a mesophile were not able to complement this *E. coli* deficiency unless they were fused to proteins that drove their association (Figures 2.4-2.6). The strength of this complementation also appeared to depend on thermostability, as the incorporation of a thermostabilizing mutation into these fragments further enhanced complementation. *In vitro* biophysical experiments using purified AK_{Bs} and AK_{Tn} fragments further revealed that the individual peptide fragments from the hyperthermophilic AK exhibit more secondary structure and activity than their corresponding mesophilic orthologs. This finding suggests that hyperthermophilic proteins may in general be more tolerant to fission events than their mesophilic homologs, due to their increased thermostability. Although the studies in Chapter 2 provide evidence for this hypothesis, I only examined the function of AK homologs having a single split site. Whether thermostable

protein homologs can be split at more sites than mesophilic homologs to create protein fragments that cooperative function was not determined.

To test whether thermostability affects the fraction of fission sites that are non-disruptive to function, I have generated libraries of vectors that express different split variants of AK_{Bs} and AK_{Tn} and identified functional split AK variants in each library through selections in *E. coli* $CV2_{T7}$ at 40°C. To create well-defined libraries of split AK_{Bs} and AK_{Tn} variants, I developed a three-step design strategy that uses a variant of the transposase MuA to randomly bisect each AK gene. This strategy was used because previous studies have shown that MuA inserts transposons into target DNA in a random fashion (Butterfield et al., 2002).

3.3 Materials and Methods

3.3.a Materials

E. coli XL1-Blue cells used for plasmid amplification and cloning were from Stratagene, whereas *E. coli* $CV2_{T7}$ was acquired from *E. coli* Genetic Resources at the Yale Coli Genetic Stock Center[14]. HyperMu™, an engineered mutant MuA transposase, was from EPICENTRE Biotechnologies (Cat. no. THM03210), and the MuA transposon used in the insertion reactions were the chloramphenicol acetyltransferase-containing (M1-CamR) entrancesposon from Finnzymes (Cat. no. F-760). The pTara plasmid that expresses the T7 RNA Polymerase (T7-RNAP) was a kind gift from K.S. Matthews (Wycuff et al., 2000), pET vectors were from EMD Biosciences, and the pSB1A2 vector was from the 2008 iGEM distribution. Synthetic oligonucleotides were from Eurofins MWG Operon, kits for plasmid purification were from Zymo Research and Qiagen, and all other enzymes were from New

---

[14] http://cgsc.biology.yale.edu/

England Biolabs, Fermentas, and Finnzymes (*i.e.*, restriction enzymes, Vent$_R$ polymerase, Phusion® High-Fidelity DNA Polymerase, and T4 DNA ligase).

3.3.b Transposon Insertion Reactions

The AK$_{Bs}$ and AK$_{Tn}$ genes were cloned behind the T7 promoter from a pET vector for high levels of constitutive expression. The promoter and gene were then transferred to the smaller pUC-18 vector to reduce the plasmid size for the library generation procedure. The *ndeI* restriction site was removed from pUC18 by QuikChange mutagenesis to produce pUC18-Δ*ndeI*, the *adk* genes were excised from pAK$_{Bs}$ and pAK$_{Tn2}$::Km (from experiments described in Chapter 2) using BglII and HindIII, and the *adk* genes were ligated into pUC18-Δ*ndeI* that had been digested with BamHI and HindIII to produce p18T7-*ncoI*-AK$_{Bs}$ and p18T7-AK$_{Tn}$, respectively. The *ncoI* site in p18T7-*ncoI*-AK$_{Bs}$ was converted to a *ndeI* site by QuikChange mutagenesis to yield p18T7-AK$_{Bs}$. M1-CamR was randomly inserted into p18T7-AK$_{Bs}$ and p18T7-AK$_{Tn}$ by incubating 20 μL reactions (1X HyperMu buffer, 300 ng of target plasmid, 100 ng entrancesposon, and 1 U of HyperMu MuA transposase) at 37°C for 16 hours. Transposase reactions were terminated by adding 2 μL of the HyperMu 10x Stop Solution, gently mixing, and incubating each reaction at 70°C for 10 minutes. After storage at -20°C, multiple electroporations were performed in parallel in which 1 μL of each transposase reaction was electroporated into 40 μL XL1-Blue cells (~10^9 cfu/mL). Small volumes (1 μL) of the ligation were electroporated in parallel to maximize the amount of transformants obtained. After electroporation, bacteria were allowed to recover for ~1 hr at 37°C in SOC medium, cells were plated onto LB-agar plates containing chloramphenicol (15 μg/mL), and plates were incubated for 48 hr at 37°C. Colonies from each transformation

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

80

were harvested by adding 2 mL of LB to each plate, scraping the cells into the LB using a sterile spreader, and pooling the slurry. After pelleting the pooled cells, the resulting vector libraries of p18T7-AK$_{Bs}$ and p18T7-AK$_{Tn}$ harboring randomly inserted M1-CamR entrancesposon were obtained using the Qiagen Midiprep Kit and concentrated with the DNA Clean & Concentrator™ Kit (Zymo Research). Preps of the AK$_{Bs}$ cells and the AK$_{Tn}$ cells yielded ~ 9.35 and ~ 12.1 µg of DNA, respectively.

3.3.c   Creation of Size-Selected Libraries

Each plasmid library (~200 ng) was digested with NdeI (20 U) and HindIII (20 U) for 16 hrs at 37°C, and the *adk* genes harboring a single M1-CamR transposon (~1,900 base pairs) were separated from *adk* genes lacking a transposon (~650 base pairs) using agarose gel electrophoresis. Gene fragments having a size consistent with the presence of an inserted transposon (~1900 base pairs) were gel purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research) and ligated into the vector backbone of p18T7 using NdeI and HindIII restriction sites. This digested plasmid, which contains the T7 promoter in the pUC18-ΔndeI background, was obtained by digesting p18T7-AK$_{Tn}$ with NdeI and HindIII and gel-purifying the resulting vector backbone. Prior to electroporation, the DNA in the ligation reactions was purified and concentrated from 20 to 10 µl using the DNA Clean & Concentrator™ Kit (Zymo Research). This DNA was split into two equal (5 µL each) aliquots and electroporated into XL1-Blue cells (40 µL x 2). After electroporation, bacteria were allowed to recover for ~1 hr at 37°C in SOC medium, cells were plated onto LB-agar plates containing chloramphenicol (15 µg/mL), and the plates incubated for 48 hours at 37°C. Colonies from each transformation were harvested by adding 2 mL of LB to each plate, scraping the cells

into the LB using a sterile spreader, and pooling the slurry. After pelleting the pooled cells containing size-selected (SS) vector libraries harboring M1-CamR minitransposons inserted within *B. subtilis adk* (Bs-SS library) and *T. neapolitana adk* (Tn-SS) genes were purified using a Qiagen midiprep.

3.3.d Creation of Fragment Co-expression Libraries

A kanamycin resistance cassette was PCR amplified from pET24d using primers to yield a product with stop codons in all three frames upstream to the resistance cassette. This product was used as a template in subsequent PCR reactions with primers that incorporate a T7 promoter, ribosomal binding site (RBS), and a start codon downstream of the resistance cassette as well as flanking *notI* sites. This regulatory element (designated F1-KanR) was subcloned into the *notI* site of pSB1A2 to create pSB1A2::F1-Kan^R and sequence verified (*see* Supplemental Data). The size-selected Bs-SS and Tn-SS libraries were treated with NotI to remove the M1-CamR transposon, and the ensuing vectors lacking the transposon were separated by agarose gel electrophoresis and gel purified. These vectors were ligated to F1-KanR that had been gel purified from NotI digested pSB1A2::F1-Kan^R to create libraries of vectors that contain two promoters for coexpressing AK fragments, herein referred to as 'F1-KanR' libraries.

3.3.e Functional Selections

The F1-Kan^R ligation reactions were transformed using electroporation into *E. coli* CV2$_{T7}$ cells already harboring pTara. Cells were then allowed to recover for 1 hr at 30°C in SOC liquid medium, plated onto LB-agar medium lacking antibiotics and LB-agar medium

containing ampicillin (25 µg/mL), kanamycin (25 µg/mL), and chloramphenicol (15 µg/mL). Plates lacking antibiotic were incubated 40°C to select for vectors that express functional AK mutants, and the plates containing antibiotic were incubated at 30°C to establish the fraction of cells that contain p18T7-derived vectors that contain the F1-Kan^R insert (and pTara). The fraction of functional variants for each library was calculated using the relative titers of cells at 30 and 40°C after 72 hours of growth. Untransformed CV2$_{T7}$ cells were plated at titers equivalent to that for the transformed cells LB-agar medium and incubated at 40°C yielded no viable colonies.

Colonies from the F1-Kan^R-inserted Bs-SS and Tn-SS libraries selected at 40°C and 72 hours were randomly picked and used to inoculate 96-deep well plates containing 1 mL of LB with ampicillin (25 µg/mL) and kanamycin (25 µg/mL). A total of 192 Bs-SS+F1-Kan^R and 192 Tn-SS+F1-Kan^R clones were grown at 30°C for 48 hours with vigorous shaking. Glycerol was added to all wells to a final concentration of 25% and these master plates were stored at -20°C. Each of these 384 clones was used to inoculate a 5 mL LB culture with ampicillin (25 µg/mL) and kanamycin (25 µg/mL), grown for 48 hours at 30°C, and hand-miniprepped.

3.3.f   Library Analysis and Verification

Colonies from the split protein F1-Kan^R-Bs and -Tn libraries selected at 40°C and 72 hours were randomly picked and used to inoculate 96-deep well plates containing 1 mL of LB, ampicillin (25 µg/mL), and kanamycin (25 µg/mL). A total of 192 F1-Kan^R-Bs and 192 F1-Kan^R-Tn colonies were grown at 30°C for 48 hours at 300 rpm. Glycerol was added to a final concentration of 25% in each well, and these master plates were stored at -20°C. Liquid cultures of each colony (5 mL) were grown in the presence of antibiotic (25 µg/mL

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

83

ampicillin and 25 μg/mL kanamycin) for 48 hours at 30°C, and the DNA vectors in each colony were purified from these cultures using a Qiagen miniprep kit. Each vector was sequenced by SeqWright, Inc. (Houston, TX) using a custom primer that hybridizes 168 base pairs from the 3' end of the F1-Kan^R insert. Of the 384 plasmids sequenced, 275 (~72%) yielded some *adk* sequence and 233 (~61%) were high enough quality to assign a fission site.

To assess the complementation strength of each *unique* sequenced variant (53 AK$_{Tn}$ and 38 AK$_{Bs}$), sequenced vectors were transformed into CV2 cells, plated onto LB medium with kanamycin (25 μg/mL) and chloramphenicol (15 μg/mL), and incubated at 40°C for 24 hrs. Single colonies obtained from these transformations were used to inoculate 5 mL cultures of LB containins kanamycin (50 μg/mL) and chloramphenicol (15 μg/mL). These liquid cultures were grown at 30°C overnight. After measuring the optical density of each culture (OD$_{600}$), cells were pelleted and resuspended in 25% glycerol to an OD$_{600}$ = 2. This was used to spot four 10x serial dilutions of resuspended cells onto LB plates at a volume of 10 μL per spot. Replicate spotted plates were incubated at 30°C and 40°C, and imaged after 24 hours of incubation using a FluorChem 5500 imager (Alpha-Innotech). Cell growth was categorized as strong (observable growth at all 4 titers analyzed), medium (growth observed at the 3 spots having the highest titer), or weak (growth only observed at the one spot having the highest titer).

3.3.g   Construction of Vectors for Expressing Homologous Split AK Variants.

Four vectors were created for expressing split AK$_{Bs}$ variants homologous to functional split AK$_{Tn}$ proteins discovered in selections, whereas three vectors were created for expressing split AK$_{Tn}$ variants homologous to functional split AK$_{Bs}$ proteins obtained

from library selections. To create these vectors, Quikchange mutagenesis was used to introduce *notI* and a five base-pair duplication into the different *adk* genes to mimic the sequence changes that would happen if this clone was created through a MuA-mediated transposition reaction (Mizuuchi, 1983). To create vectors that coexpress the AK peptides encoded before and after the inserted *notI* site, the F1-KanR regulatory element was subcloned into the *notI* site of each vector. All vectors were verified by sequencing.

3.3.h Calculations of Split Library Coverage

The approximate coverage of our split protein libraries were generated by using simulation approach. Library generation was simulated N times using a Python script[15] in the following manner. First, to simulate the MuA transposition step, X random numbers were generated between *1* and *A*. Each random number between *1* and *B* was stored in a list and the number of unique entries in this list returned as the outcome of the MuA insertion step. Second, to simulate the size selection step, Y elements of the list generated in the first step were randomly picked. Again, each of these values was stored in a list, and the number of unique values was returned as the outcome from the size selection step. Third, to simulate the final subcloning step, Z elements were randomly picked from the list from step 2. Each of these elements had a 50% chance of being added to a list, and the number of unique elements in this list was returned as the result from the subcloning step. A 50% probability was used here because the F1-Kan^R insert has a similar probability of being inserted with its promoter in the same orientation as the promoter before the N-terminal AK fragments. The returned values for each of the N runs were analyzed to find the mean and 95% confidence interval. Note: N = the number of simulations to run, A = the size of the plasmid used, B =

---

[15] This simulation algorithm was written by Thomas Segall-Shapiro.

the size of the gene of interest, X = the number of unique clones experimentally observed in the MuA insertion step, Y = the number of unique clones experimentally observed in the size selection step, and Z = the number of unique clones observed in the final subcloning step. For the AK_{Tn} simulation, N = 100,000, A =660 , B=3430 , X=9750 , Y=2097 , Z=2221 . For the AK_{Bs} simulation, N = 100,000, A =651 , B=3421 , X=6250 , Y=3236 , Z=3656 .

3.4   Results

3.4.a   Adenylate kinases as a Model System for Library Comparisons

Since the goal of this study is to extend the results obtained in Chapter 2, we have used the adenylate kinases (AK_{Tn} and AK_{Bs}) previously characterized in those experiments. These proteins make an optimal model system to use for a number of reasons. First, they have similar sequences (~50% identity) and structure, as well as similar activities at mesophile growth temperatures (Vieille, Krishnamurthy, Hyung-Hwan, et al., 2003). Second, a simple bacterial selection can be used to compare their functions at 40°C, as demonstrated in Chapter 2. Third, AK_{Tn} and AK_{Bs} are only 220 and 217 amino acids long, respectively, suggesting that they can be fragmented at a large fraction of the total possible sites using a transposase-based strategy.

3.4.b   Library Construction

To create libraries that express split proteins arising from the different possible fission sites, I turned to transposon-mediated strategy, which has been widely used for HTP sequencing, random protein insertions, and protein truncation studies. In my library design strategy (Figure 3.1, Step 1, page 89), I used HyperMuA to randomly insert a mini-transposon containing a selectable marker into AK-expression plasmids (p18T7-AK_{Bs} and

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

86

p18T7-AK$_{Tn}$). The efficiency of this first step was assessed by restriction digest analysis of libraries acquired from cells transformed with the transposase reaction. As shown in Figure 3.2.A, page 90 (lanes 1 and 3), digests to drop out the AK gene show an insertion that results in a DNA piece that is equivalent in size to the sum of the AK gene (650 base pairs) and the mini-transposon (1254 base pairs). Another digest that cleaves at the transposon integration sites (Figure 3.2.A, lanes 2 and 4, page 90) exhibits a single band at ~1254 base pairs corresponding to the transposon and a smear beginning at ~650 bp to corresponds to the randomly split AK gene fragments. To remove the plasmids in each library that have a minitransposon integrated outside of the *adk* gene, the *adk* genes were excised from each library using restriction endonucleases that cleave 5' and 3' to the AK genes, and *adk*-minitransposon fusions were purified away from parental *adk* genes using agarose electrophoresis (Figure 3.1, Step 2, page 89). After religating this ensemble of *adk*-minitransposon gene fusions into a new plasmid, the minitransposon was replaced with a regulatory element to create a library of vectors with two promoters, one for each AK fragment (Figure 3.1, Step 3, page 89). This regulatory element, F1-Kan^R, contains a fusion of three stop codons (in each frame), a kanamycin selectable marker, a T7-RNAP promoter, and a start codon.

The efficiencies for each of the steps shown in Figure 3.1 were estimated using a Python script that numerically analyzes the loss in library coverage at each step, based on the number of clones actually acquired at each step (Figure 3.2.B, page 90). The number of transformants (number of colonies divided by $2^{(\text{number of doubling times})}$) was calculated for each step. In addition, the fraction of the possible fission sites sampled within each *adk* was calculated using the transformation efficiencies and assuming a random distribution of

insertion sites. The final expression libraries are expected to contain vectors that express more than half of the possible split AK_{Bs} and AK_{Tn} variants (Figure 3.2.B, Step 3, page 90).

In this library design strategy, only a subset of vectors in each library are expected to coexpress the AK peptides that are separated by a regulatory element. Since the insert harboring the second promoter is cloned into a single *notI* site, only half of the variants in each library are predicted to contain the second promoter in an orientation that is productive for split protein expression, *i.e.*, in the same direction as the promoter before the *adk* gene. The value shown for the final library in Figure 3.2.B (page 90) represents the number calculated to be in the productive orientation. All of the vectors with the regulatory elements inserted in a productive orientation are expected to terminate expression of each N-terminal AK fragment, due to the presence of stop codons in each frame. However, only one third of the constructs with the regulatory element in a productive orientation are expected to express the AK fragments following the insertion site, because the regulatory element must be in frame to express the AK peptide that follows the *notI* insertion site. In addition, this library design strategy adds amino acids to each AK fragment proximal to the fission site due to the nature of HyperMuA-mediated insertion (Figure 3.3, page 91). N-terminal fragments contain an extra 4 to 6 residues, depending on the site and frame of each minitransposon insertion. C-terminal fragments begin with a 5 residue peptide that depends in part on the minitransposon insertion site, except in cases where these fragments are expressed from an alternative start codon.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate Kinase Fragments

88
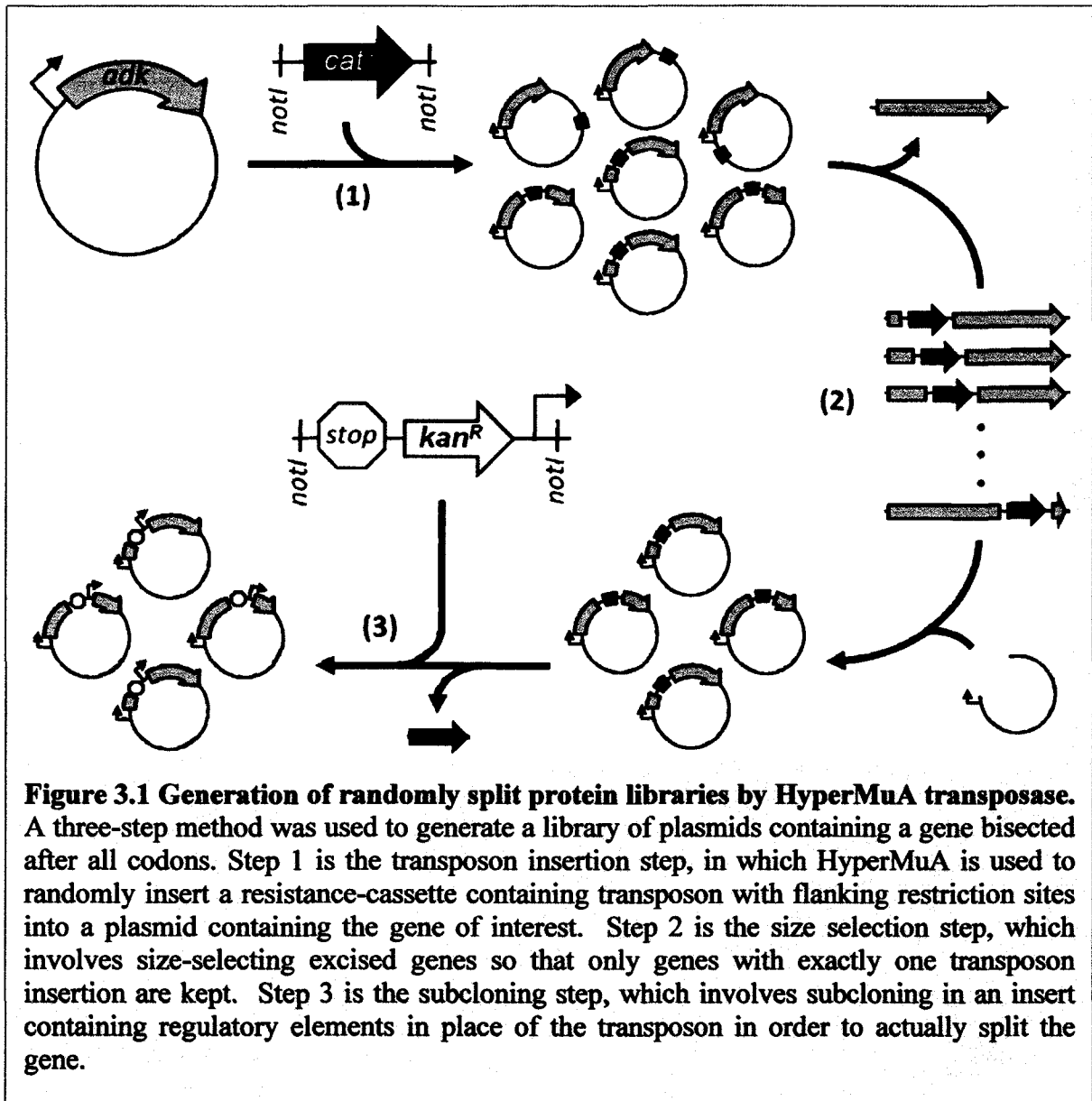
Figure 3.1 Generation of randomly split protein libraries by HyperMuA transposase. A three-step method was used to generate a library of plasmids containing a gene bisected after all codons. Step 1 is the transposon insertion step, in which HyperMuA is used to randomly insert a resistance-cassette containing transposon with flanking restriction sites into a plasmid containing the gene of interest. Step 2 is the size selection step, which involves size-selecting excised genes so that only genes with exactly one transposon insertion are kept. Step 3 is the subcloning step, which involves subcloning in an insert containing regulatory elements in place of the transposon in order to actually split the gene.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

89
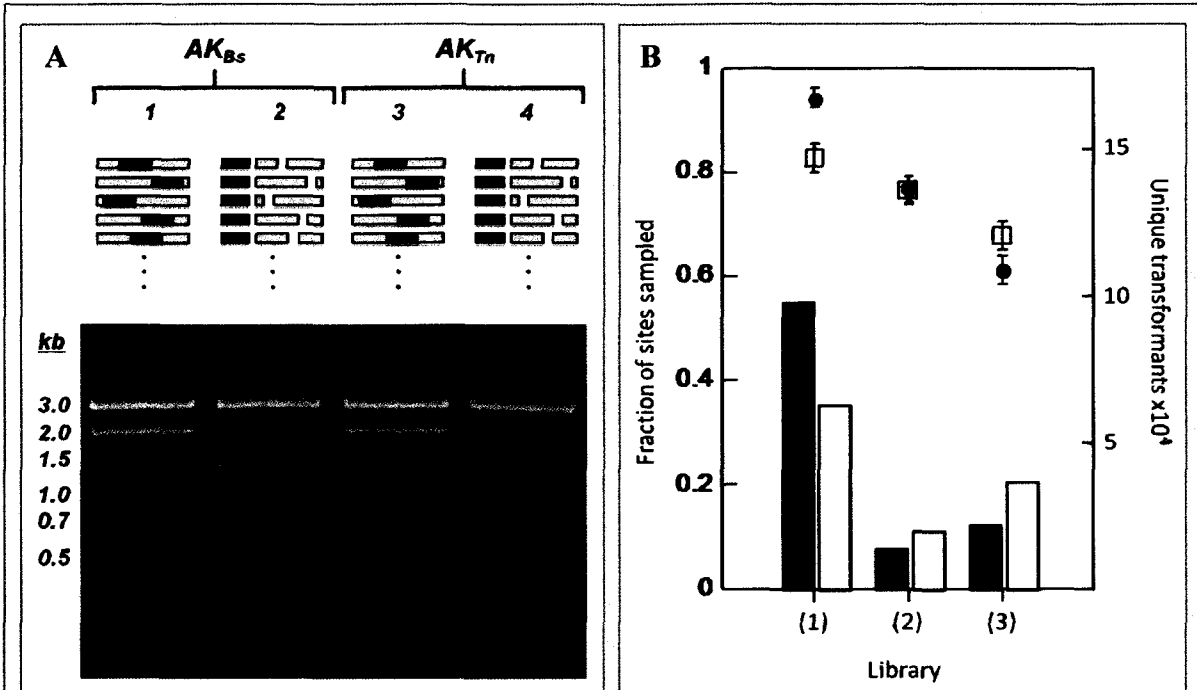
Figure 3.2 Library Construction and Estimated Efficiency of Random Insertion. A) Size-selected AK_Bs and AK_Tn libraries were analyzed by DNA restriction mapping to ensure the random nature of the transposon insertion. Lanes 1 and 3 are NdeI and HindIII-digested samples of the AK_Bs and AK_Tn libraries, respectively. Lanes 2 and 4 are NdeI, HindIII, and NotI-digested samples of the AK_Bs and AK_Tn libraries, respectively. B) Steps 1 and 2 of the library construction process were completed using the plasmid pUC18Δ-ndeI. The AK_Bs library is shown in white and the AK_Tn libraries are shown in black. The bars for the first two steps show the number of transformants (number of colonies divided by 2^number of doubling times) produced at each step. The points on the lines for these steps show what fraction of the possible split points this number of transformants should cover, assuming random distribution. These values were calculated by running a brute-force simulation that randomly places fission sites in the plasmid and counts what fraction of the possible ones are represented.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

90
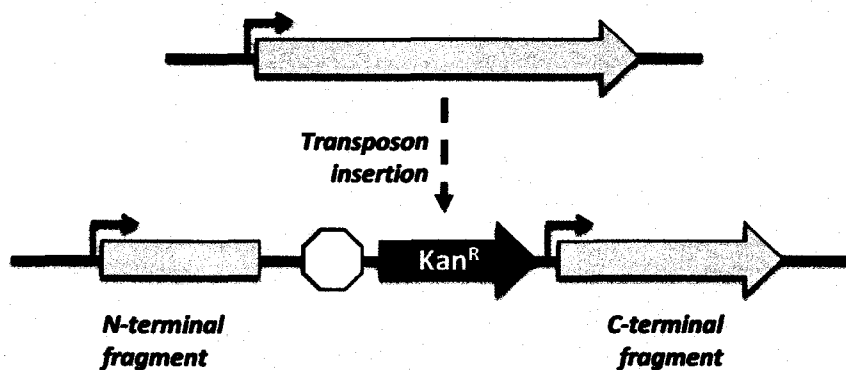
Frame	End of N-terminal fragment	
0	-NNN-tgc-ggc-cgc-tag x C G R stop	x = AK$_{Bs}$/AK$_{Tn}$ sequence
-1	-NNt-gcg-gcc-gct-agc-tag x A S S S stop	x = A, D, N, C, G, H, L, P, R, I, F, S, T, Y, or V
-2	-Ntg-cgg-ccg-cta gct-agc-taa x R P L A S stop	x = L, M, or V

Frame	Beginning of C-terminal fragment	
0	Atg-gcg-gcc-gca-NNN- M A A S X	X = AK$_{Bs}$/AK$_{Tn}$ sequence
-1/-2	Not expressed, or truncated peptide with alternative start	

Figure 3.3 Possible Sequences of AK fragments produced by the transposase insertion method. Depending on the site of transposase insertion, the N-terminal AK fragments will end with a 4-6 residue insertion with the identities predicted as above. For the C-terminal AK frame, only in-frame transposon insertions will lead to a translated protein product, starting with the sequence MAASX-.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

91

3.4.c   Selection of Functional Split AK_{Bs} and AK_{Tn} Variants

The $CV2_{T7}$ strain of *E. coli*, which contains a temperature-sensitive AK, was used to select both libraries at 40°C for AK variants that are functional. *E. coli* $CV2_{T7}$ cells transformed with F1-Kan^R library ligations were plated onto multiple LB-agar medium lacking antibiotic. For selections, 19 plates for each library transformation were incubated at either 40°C for 24 hours. Similarly, 5 plates each of the same library transformations was incubated at 30°C for 24 hours to obtain transformation counts at non-selection conditions. The mean number of transformants per plate obtained from the non-selection temperature plates (30°C) for the AK_{Bs} and AK_{Tn} split libraries were 577±40 and 350±17 colony forming units (CFUs), respectively. For $CV2_{T7}$ transformant libraries selected at 40°C, for an equivalent number of plates and plating volume, a total of 682 and 2938 complementing colonies were found for the split libraries created using AK_{Bs} and AK_{Tn}, respectively. A comparison of the average number of complementing colonies per plate to the number of total cells per plate ($cfu^{40°C}/cfu^{30°C}$) yields a fraction functional of 6.2% and 44.1% for the libraries created using AK_{Bs} and AK_{Tn}, respectively.

Of the final selected libraries, 192 clones were randomly picked from each library, grown overnight in LB liquid cultures containing 50μg/mL kanamycin, and individually miniprepped to acquire plasmids for sequencing. A single pass sequencing reaction yielded *adk* sequence data for ~70% of the plasmids. However, only 61% of the sequencing reactions yielded data that was of sufficient quality to identify the bisection site. A similar number of bisection sites were identified for the *T. neapolitana* (n = 124) and *B. subtilis* (n = 109) *adk* genes. The positions and frequencies of the bisection sites are shown in Figure 3.4, page 95, with all split sites numbered according to the last wild-type residue occurring before the

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

92

transposon insertion site. This figure shows that a large number of bisection sites occur proximal to the C-terminus each AK homolog, *i.e.*, at of after residues 185. In addition, many of these bisection sites contain the regulatory element in the non-productive orientation that does not allow for expression of C-terminal AK fragment. This finding suggests than many of the variants with a bisection site after residue 185 are functional because the C-terminally truncated fragments are sufficiently active to complement *E. coli* $CV2_{T7}$ at 40°C. Many of the other functional AK variants are thought to require both polypeptides to complement *E. coli* $CV2_{T7}$. For example, the bisection sites discovered between residues 75 to 175 in each AK are thought to represent split proteins that only function if both fragments are expressed and spontaneously reassemble. This region corresponds to part of the AK CORE domain, which contains residues that constitute the active site, as well as the entire LID domain, which protects the active site. Figure 3.4 (page 95) shows that functional split *T. neapolitana* AK variants were discovered with bisection sites throughout this region. In contrast, very few functional split *B. subtilis* AK variants were discovered with a fission site between residues 100 and 200, and most of these occurred proximal to the residue 100 and 200. These differences could arise AK_{Bs} cannot be fragmented in this region to create protein fragments that cooperatively function, or because of sequence biases in the libraries selected.

As noted above, the frame in which HyperMuA inserts the mini-transposon into each *adk* gene affects whether or not a full-length C-terminal AK fragment is expressed. In Figure 3.5 (page 96) all the frames of the C-terminal fragment for each split site are plotted, as well as all possible alternative start codons found in AK_{Bs} and AK_{Tn}. This figure shows that functional variants were obtained that have both in frame and out of frame insertions. In the

case of the out of frame insertions that are not expected to express a full C-terminal fragment, alternative start codons may be responsible for expressing a truncated C-terminal fragment. Figure 3.5 (page 96) shows that many of the out of frame insertions are proximal to a codon that can be used as an alternative start site for initiating the translation of the C-terminal fragments.
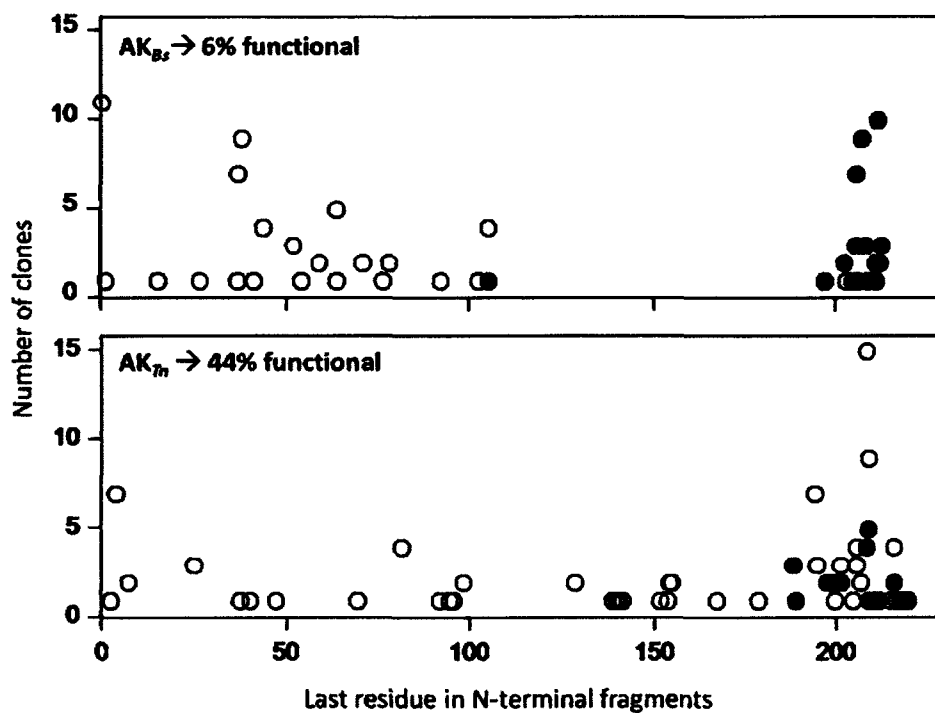
Figure 3.4 Incorporation of the internal promoter within the *adk* gene of functional clones. AK$_{Tn}$ and AK$_{Bs}$ split libraries were selected for functional complementation in CV2$_{T7}$ cells harboring. Complementing clones were randomly selected and sequenced. Shown is the frequency of the insertion site as a function of the location in the protein sequence, as determined by the sequencing of randomly selected clones. Black circles represent sequenced clones from in which the F1-Kan^R insert integrated in reverse orientation to that required for proper coexpression of the AK fragments. White circles represent sequenced clones from in which the F1-Kan^R insert integrated in the proper forward orientation, which would allow termination of the N-terminal fragment and transcription of the C-terminal fragment.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

95

AK$_{Bs}$

AK$_{Tn}$

ATG = dashed lines
GTG = dotted lines
TTG = not present

Last residue in N-terminal fragments

Number of base pairs out of frame
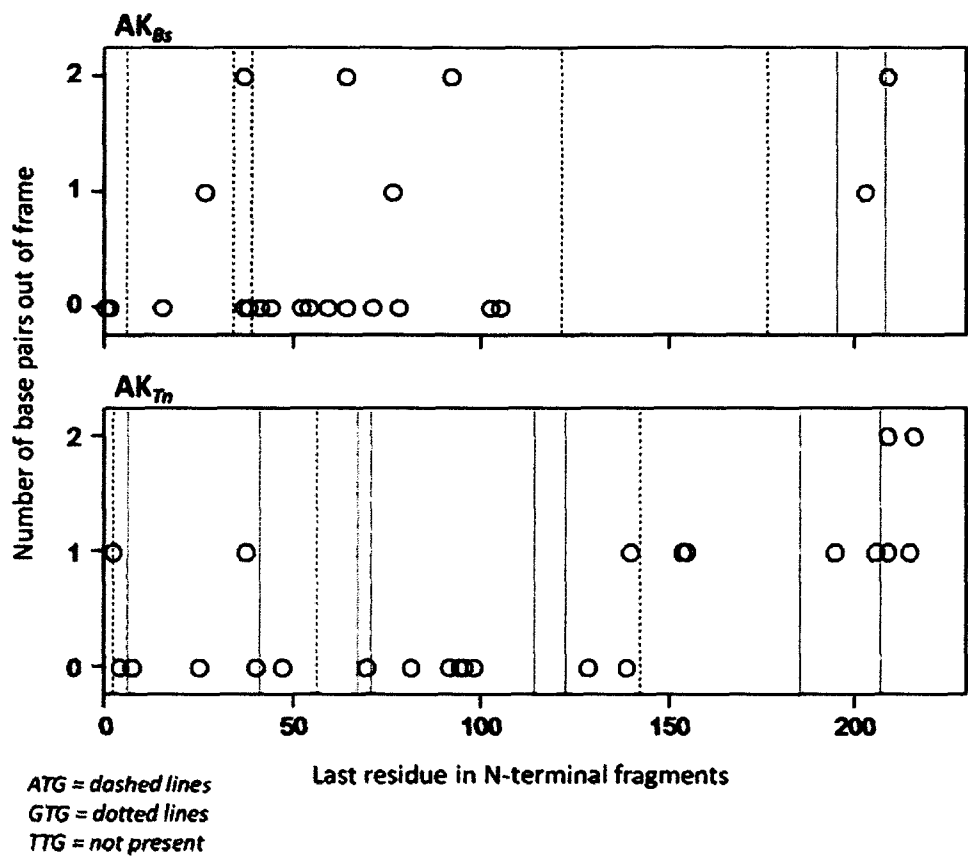
Figure 3.5 Frames of the C-terminal fragments relative to the alternative start codons. Split sites that lead to out of frame sequences for the C-terminal fragment shown as circles with values of 1 or 2 were mapped against alternate start sites, shown as vertical lines.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate Kinase Fragments

96

3.4.d Assessing the Complementation Strength of split AK_{Bs} and AK_{Tn} Homologs

In Chapter 2, I found that AK_{Bs} and AK_{Tn} protein fragments from a single homologous split site (between D76 and C77, by AK_{Bs} numbering) exhibited differences in complementation efficiency with *E. coli* $CV2_{T7}$ at 40°C. Whereas the split AK_{Tn} was able to support growth of $CV2_{T7}$ cells at 40°C, the split AK_{Bs} was only able to do when fused to interacting polypeptide tags or proteins (Figures 2.4-2.6). To see if this occurs with other split sites identified in the split AK_{Bs} and AK_{Tn} libraries, I assessed the relative complementation strength of each of the split sites identified by DNA sequencing. Plasmids corresponding to each of the 93 *unique* split fragment pairs were individually transformed into $CV2_{T7}$ cells, and spots representing multiple 10x serial dilutions of liquid cultures were analyzed at 40°C for 24 hours. As shown in Figure 3.6 on the following page, this analysis revealed that the proportion of strong to moderate to weak complementation for AK_{Bs} (21 to 39.5 to 31.6%) was different from that for AK_{Tn} variants (54.7 to 41.5 to 1.9%). This indicates that the AK_{Tn} variants are on average better at complementing an AK defect in *E. coli* at 40°C.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

97

Figure 3.6 Relative complementation strength of split protein clones in CV2 cells. Previously isolated split AK clones were re-transformed into CV2 cells containing pTara and individually assessed for complementation strength. Relative complementation strength is categorized on a scale of 1 to 3, with 1=weak (growth only observed at the one spot having the highest titer), 2=medium (growth observed at the 3 spots having the highest titer), and 3=strong (observable growth at all 4 titers analyzed).

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

98

3.4.e Comparison of split AK_{Bs} and AK_{Tn} Homolog Function

The difference in the frequency of fission sites observed between residues 100 and 200 in each selected library suggests that fragmentation of AK_{Bs} in this region may be more disruptive to AK structure and function. To test this, I created four constructs for expressing AK_{Bs} variants that are homologous to functional AK_{Tn} variants that were discovered in library selections. Figure 3.7.A (next page) shows the results from these selections. The two AK_{Bs} variants with fission sites within this region (125 and 175) were not able to complement $CV2_{T7}$ growth at 40°C like the homologous AK_{Tn} variants (128 and 178). In contrast, AK_{Bs} variants with fission sites before this region (66 and 95) complemented $CV2_{T7}$ growth at 40°C albeit to a lesser extent than homologous AK_{Tn} variants (69 and 98). To test whether fission sites discovered only in the functional selections of the *B. subtilis* AK library are also non-disruptive when incorporated into AK_{Tn}, I created and evaluated the function of three AK_{Tn} split variants (Figure 3.7.B, next page). Vectors encoding each of these AK_{Tn} variants (74, 81, and 108) complemented the growth of *E. coli* $CV2_{T7}$ at 40°C like their AK_{Bs} homologs (71, 78, and 105).

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

99

A Split AK$_{Tn}$ AK$_{Bs}$ homolog

69   *66*

98   *95*

128   *125*

178   *175*

B Split AK$_{Bs}$ AK$_{Tn}$ homolog

71   *74*

78   *81*

105   *108*
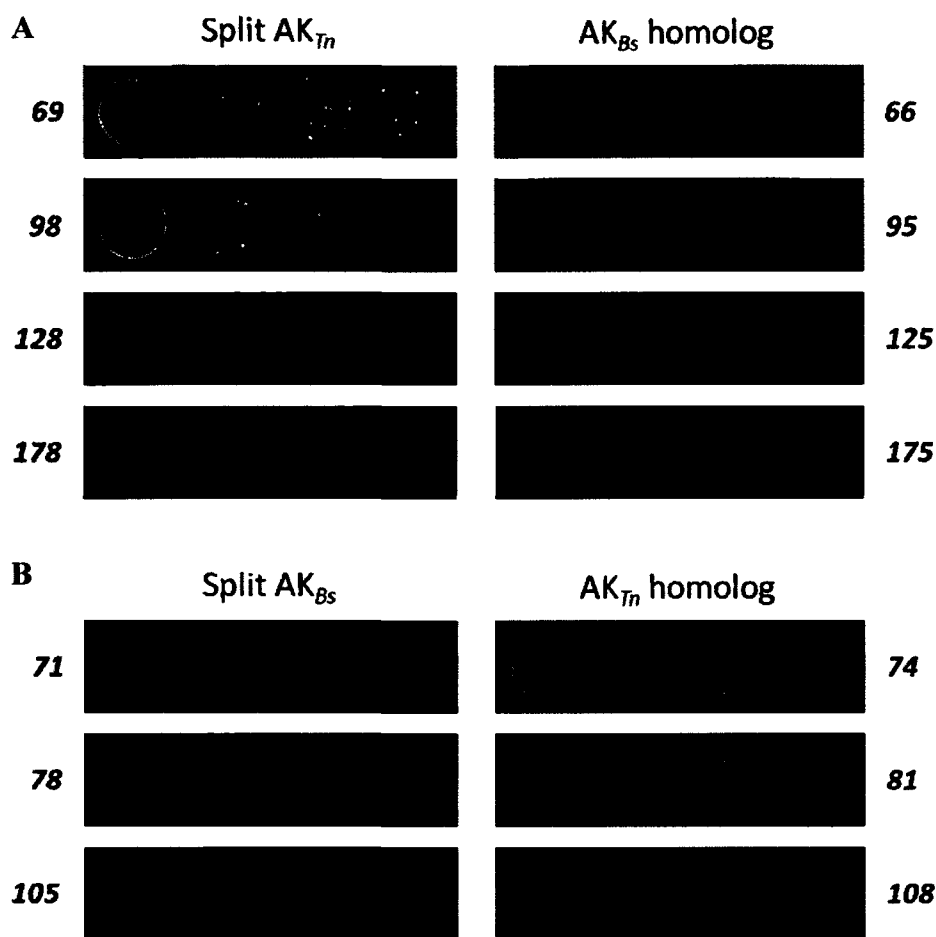
Figure 3.7 Comparison of homologous split sites for split AK$_{Tn}$ and AK$_{Bs}$. Complementation strength was compared by plating spots of cultures on LB plates, at 40°C for 24 hours. All spots shown are 10x dilutions, from left to right. A) AK$_{Tn}$ was engineered to have split sites at homologous sites as three split Bs clones identified in my selections. B) AK$_{Bs}$ was engineered to have split sites at homologous sites as four split Tn clones identified in my selections.

3.5 Discussion

In this chapter, I have designed, created, and analyzed libraries of randomly split AK$_{Bs}$ and AK$_{Tn}$ proteins using a novel transposase-based method. My hypothesis that thermostabile proteins are better able to withstand the destabilizing effects of protein fragmentation was supported by experiments in Chapter 2. However, those initial experiments only looked at a single split site among AK enzymes of differing thermostabilities. Analyzing a spectrum of split sites will corroborate the findings in Chapter 2 and provide further evidence that the correlation between thermostability and complementation efficiency is a general phenomena that occurs on average for split sites in AK proteins.

Comparison of the number of fraction functional split sites (cfu$^{40°C}$/cfu$^{30°C}$) for the split libraries show a striking difference. Of the split AK$_{Bs}$ library, only 6.2% were functional at 40°C whereas 44.1% of the split AK$_{Tn}$ library were found to be functionally complementing. The frequencies of all split sites identified with a high confidence are shown in Figure 3.4 (page 94). One predicted observation is that insertions that occur in the reverse orientations (where the T7 promoter in the F1-Kan^R insert is pointed away from the C-terminus) only occur in the C-terminal regions. This is because the N-terminal fragments from such integration events are still able to function as truncations. In contrast, N-terminal integrations of a F1-Kan^R regulatory insert in the non-productive orientation would result in the majority of the AK protein not being expressed. Focusing on regions in the middle (75 to 175) of the protein that are likely to be true split sites reveals more sites for the AK$_{Tn}$ (n=22; 17.7%) in comparison to the AK$_{Bs}$ protein (n=11; 10.0%).

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

101

To validate the complementation activity and to allow relative comparisons of the split sites, I retransformed library clones encoding for unique split sites identified in the library selections into $CV2_{T7}$ cells and assessed their relative complementation by analyzing diluted culture spots grown at selection temperature (40°C) for 24 hours. The results shown in Figure 3.6 (page 98) indicate that the AK_{Bs} fragments on average have a medium to low strength of complementation. In contrast, nearly all of the AK_{Tn} fragments displayed strong to medium complementation. In order to properly assess the differences between complementation efficiency for the thermophilic AK_{Tn} and mesophilic AK_{Bs}, I also compared split sites that occur at homologous positions between the two AKs (Figure 3.7, page 100). Split sites identified from the library selections were engineered into the corresponding ortholog as a homologous split sites, by Quikchange mutagenesis. Split sites originally identified in AK_{Tn}, when engineered into the less stable AK_{Bs} enzyme, were only able to complement weakly, if at all. To establish that this difference is due to the thermostability differences between the protein and not other contextual differences, I also analyzed split sites originally identified in AK_{Bs}, which were engineered into the thermostable AK_{Tn} enzyme. The engineered AK_{Tn} split sites were all able to complement as efficiently as the original AK_{Bs} fragment pairs.

An analysis of the split sites considering the distribution and complementation strength in the context of the AK protein sequence and structure is presented in Figure 3.8 on page 104. The heavy concentration of strongly complementing split sites near the C-terminus can be attributed to functional truncations of the AK protein, as I described above. Both the AK_{Bs} and AK_{Tn} split sites show similarities in the distribution of split sites in the N-terminal region, which contains the entire AMP binding domain and parts of the CORE domain

(although the AK$_{Tn}$ has stronger average complementation). In addition, the split sites for both orthologs mostly occur at solvent-exposed regions, and there is a lack of split sites in regions that make contact with the substrate or are close to the active site. However, one striking difference between the AK$_{Bs}$ and AK$_{Tn}$ split sites is the lack of any split sites for the mesophilic AK$_{Bs}$ protein between residues 105 – 191. This region of AK proteins contains the majority of the CORE domain as well as the entire LID domain (residues 129-159, AK$_{Bs}$ residue numbering), which protects the active site during catalysis (Schulz, 1987; Schulz et al., 1990; Liu et al., 2009). Also pertinent is the fact that the LID domain coordinates a zinc ion that has been shown to increase the thermostability of AK proteins by ~7-11°C (Perrier et al., 1994; Perrier et al., 1998). The lack of split sites in this domain for AK$_{Bs}$ may indicate that fragmentation in this region of the CORE domain and/or the LID domain results in fragments that are unfolded and cannot reassociate to form functional AK complexes. In contrast, the thermostable AK$_{Tn}$ can tolerate numerous split sites in this region, with 6 split sites occurring throughout the LID domain. Interestingly, 5 of these split sites in the LID domain would abolish the tetracysteine motif that binds to a thermostabilizing zinc atom by covalently splitting the motif in half. These results suggest that the AK$_{Tn}$ fragments produced by split sites in this region are able to retain sufficient partial folding at 40°C for functional complementation, whereas the AK$_{Bs}$ fragments cannot tolerate fission in these regions.

A

BsAK	--MNLVLMGLPGAGKG QGERIVEDYG PHISTGDMF A MK ET -LGLEAKS IDKGE	57
TnAK	MMA LV LGPPGAGKGTYAKRIQE TGIPHISTGDI RD VKK NDELGKKIKEIMEKGE	60
	:* ***** .:** *. *********:** :*:*. ** : *. ::***	

BsAK	L PDEV IGIVKE LGKDDC RGFLLDGFPRTV QAEALEEILE Y PIDYVINIEVDK	117
TnAK	LVPDELVNE VVKRRLSEKDC KGFILDGYP RTVAQAE LDSFLESQNKQLTAAVLFDVPE	120
	*****:. :**.**.:.***:**:***:******** *:.:**. .* : .: ::* :	

BsAK	DVLMERLTGRRICSVCGTTYHLVFNPPKTPGICDKDGGELYQRADDNEETVSKRLEVNMK	177
TnAK	DVVVQRL TSRRICPKCG RDY MISLPPKEDE LCD DCKVKLVQRDDD EETVRHRYKV LE	180
	::::*.****. ** *::: *** :**. :* ** **:**** :* :* ::	

BsAK	QTQPLLDFYSEKGYLANV GQQDIQD YA K L GGLKK	217
TnAK	KTQPVID YGKKG LK VD IGIDNVVAE LK IGWSD K	220
	:***::*:*.:** * .*:* .*::* *:* .::* .*	

■ Strong split-BsAK complementation   ■ Strong split-TnAK complementation

■ Medium split-BsAK complementation   ▨ Medium split-TnAK complementation

▦ Weak split-BsAK complementation   ☐ Weak split-TnAK complementation

B

C

Figure 3.7 Distribution of AK library split-sites. A) Split sites and their corresponding complementation strength, defined as the last wildtype residue of the N-terminal fragment, mapped onto the BsAK and TnAK sequences. The AMP-binding domain and LID domain are indicated by the brown and purple lines above the sequences, respectively. Complementation strength is color-coded for each AK ortholog. B) Split Bs sites mapped onto the BsAK structure (1P3J) using the same color coding. C) Split Tn sites mapped onto a TnAK structure acquired by protein threading, using the 1P3J structure as a template (SWISS-MODEL). All structures are shown with bound zinc as a space-filling model and the Ap5A substrate analog as a stick model.

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

104

The results presented in this chapter and in Chapter 2 provides further support for the 'thermostability buffer' theory, which postulates that proteins that are more thermodynamically stable are better able to withstand structurally destabilizing events than less stable homologs (Bloom et al., 2005). Although this has been previously demonstrated for random mutation and recombination (Besenmatter et al., 2007; Bloom et al., 2005; Bloom, Labthavikul, et al., 2006; Meyer et al., 2006), the work presented here is the first to establish that this phenomenon also occurs for protein fragments produced from the backbone cleavage of a protein. The evidence presented in here also elucidates a novel strategy for the design of PCA systems. To maximize the diversity of a split protein library, one can choose starting proteins of varying thermostabilities. Random fission of these protein homologs will generate protein fragment pairs with a wide range of complementation efficiency. Split proteins with the desired complementation strength can then be fine-tuned by random mutagenesis or rational design.

These PCA strategies we propose here can be implemented in the novel transposase method we have developed and validated. Previous methods employ multiple time-dependent enzymatic reactions, which could lead to significant inconsistencies and systemic error. Our method, which employs the ability of HyperMuA to randomly insert a mini-transposon at high efficiency, circumvents the current limitations in other methods. Generation of the randomly inserted library occurs in a single enzymatic step, the process is easy to implement (~5 minutes to set up the transposase reaction) the reagents are small in number, and thus error in the process is substantially reduced.

In Chapter 2 and this chapter I have demonstrated experimental evidence supporting the existence of a greater capacity for fragments from thermostable AK proteins to retain

Chapter 3. Functional Analysis of Libraries Containing Randomly Split Mesophilic and Thermophilic Adenylate kinase Fragments

105

thermodynamic stability. At higher temperatures closer to the temperature AK_{Tn} evolved in

(*T. neapolitana* optimally grows at ~90°C), one would expect that the AK_{Tn} split protein

fragments would be more destabilized and would be ideal to use as the basis for constructing

a PCA system that works at thermophilic temperatures. In the next chapter, I explore the

design and engineering of such a novel assay.

Chapter 4

Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus thermophilus*

4.1   Abstract

Many proteins can be split into fragments that exhibit enhanced function upon fusion to interacting proteins. While this strategy has been widely used to create protein-fragment complementation assays for studying natural and engineered proteins within mesophilic organisms, this strategy has not yet been applied within a thermotolerant microbe. In this chapter, I describe the development of a selection for protein-protein interactions within a temperature-sensitive *Thermus thermophilus* strain that is based upon growth complementation by fragments of *Thermotoga neapolitana* adenylate kinase (AK_{Tn}). Complementation studies with an engineered thermophile (PQN1) that is not viable above 75°C because its *adk* gene has been replaced by a *Geobacillus stearothermophilus* ortholog revealed that growth could be restored at 78°C by a vector that coexpresses polypeptides corresponding to residues 1-79 and 80-220 of AK_{Tn}. In contrast, PQN1 growth was not complemented by AK_{Tn} fragments harboring a C156A mutation in the zinc-binding tetracysteine motif unless these fragments were fused to *Thermotoga maritima* chemotaxis proteins that heterodimerize (CheA and CheY) or homodimerize (CheX). This enhanced complementation is interpreted as arising from chemotaxis protein-protein interactions, since AK_{Tn}-C156A fragments having only one polypeptide fused to a chemotaxis protein did not complement PQN1 to the same extent. This selection increases the maximum temperature where a protein-fragment complementation assay can be used to generate protein interaction maps for thermotolerant microbes and to engineer thermostable protein complexes.

Chapter 4. Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus Thermophilus*

107

4.2 Introduction.

Microbes that grow optimally above 60°C (*thermophiles*) and 80°C (*hyperthermophiles*) populate ecosystems where many biomolecules exhibit reduced stability (Rothschild & Mancinelli, 2001). To support cellular growth, their proteins have evolved distinct amino acid compositions (Zeldovich et al., 2007) and higher thermostability (Vieille and Zeikus, 2001) compared to orthologs found in mesophilic organisms that grow at lower temperatures. This latter feature has been exploited for a variety of biotechnological applications. Thermostable DNA polymerases have revolutionized molecular biology (Niehaus et al., 1999), xylanases have made paper processing greener (Bajpai et al., 2006), and oligosaccharide-modifying enzymes have been harnessed for corn syrup production (Crabb & Mitchinson, 1997). There is interest in further harnessing thermotolerant microbes and their proteins for other industrial processes such as biomass conversion to bioethanol (Shaw et al., 2008) or biohydrogen (Kongjan et al., 2009). However, no *in vivo* screens have been developed for creating protein complexes that help achieve these high-temperature metabolic engineering and synthetic biological goals (Blumer-Schuette et al., 2008; Chou et al., 2008).

A comparison of the findings from mesophile and hyperthermophile proteomic studies suggests that the lack of high-temperature protein-protein interactions screens may limit the discovery of some thermostable protein complexes. A high-throughput screen for pairwise protein-protein interactions among almost one thousand *Pyrococcus horikoshii* proteins found only 56 hetero-interactions using a two-hybrid assay implemented at a temperature (37°C) far below that of its optimal growth of 98°C (Usui et al., 2005). This finding can be contrasted with similar screens for protein complexes in bacteria and yeast

under near physiological conditions, which invariably find protein-protein interactions at a frequency that is more than an order of magnitude higher (Butland et al., 2005; Rain et al., 2001; Tarassov et al., 2008). The exact reason for paucity of thermophilic protein-protein interactions detected in these studies is unclear. Comparing thermophile PINs determined using parallel methods at both thermophilic temperatures and at mesophilic temperatures would allow one to pinpoint protein interactions that are temperature-dependent. The creation of an assay that can be used to assess protein complex formation at thermophile growth temperatures would have multiple advantages over available assays in studying natural and engineered proteins (Tarassov et al., 2008; Usui et al., 2005). High temperature assays are predicted to be superior at discovering interactions among proteins that require extreme temperatures to adopt their native conformation (Abd Rahman et al., 1997; Goda et al., 2005; Siddiqui et al., 1998) and among proteins whose interactions weaken as temperature is decreased from the levels where hyperthermophiles grow (Ogasahara et al., 2003).

To establish an assay for studying protein complex formation within a living thermophile, I split *Thermotoga neapolitana* adenylate kinase (AK_{Tn}) into fragments that can be used as a protein-fragment complementation assay (PCA) in *Thermus thermophilus*. AK_{Tn} has many characteristics that make it suitable for designing a split enzyme that reports on protein-protein interactions at extreme temperatures. AK_{Tn} is monomeric and extremely thermostable, exhibiting maximal phosphotransferase activity at 80°C (ATP + AMP $\leftrightarrow$ 2ADP) and having a melting temperature of 99.7°C (Vieille et al., 2003). Furthermore, AK_{Tn} can be split to generate fragments that spontaneously associate and cooperatively function within a mesophilic bacterium growing at 40°C. As demonstrated in Chapter 2, Polypeptides

corresponding to residues 1-79 (TnN) and 80-220 (TnC) complemented the growth defect of *Escherichia coli* CV2$_{T7}$, a strain that has a temperature-sensitive AK and pre-transformed with a plasmid encoding T7 RNA Polymerase. In addition, structurally-related fragments of a *Bacillus subtilis* AK with lower thermostability could not complement CV2$_{T7}$ growth unless they were fused to polypeptides that associate (Nguyen et al., 2008). This suggested that TnN and TnC may be used to report on protein-protein interactions at higher temperatures where the residual structures of these polypeptides are more destabilized, such as within thermophiles that have a temperature-sensitive AK (Counago & Shamoo, 2005).

4.3 Materials and Methods.

4.3.a Reagents and Molecular Biology Materials.

Agar was from Difco, Gelrite was from Research Products International, and all other bacterial growth media components were from BD Biosciences and Sigma-Aldrich. QuikChange mutagenesis reactions were performed using PfuTurbo DNA polymerase from Stratagene, genes were amplified for cloning using Vent$_R$ DNA Polymerase from New England Biolabs, and restriction endonucleases were obtained from Roche Biochemical, New England Biolabs, and Promega. Synthetic oligonucleotides were from Operon Biotechnology, pET vectors were from EMD Biosciences, *Escherichia coli* XL1-Blue cells used for plasmid amplification and cloning were from Stratagene, and kits for DNA purification were from Zymo Research and Qiagen.

4.3.b *T. thermophilus* Media, Strains, and Transformations.

Thermus thermophilus HB8 (ATCC #27634) was used as the parent strain for adenylate kinase (*adk*) gene replacement. Liquid growth was performed in Evian-Thermus Medium (EvTM) which contains 8 g tryptone, 4 g yeast extract, and 3 g NaCl per liter of Evian mineral water. EvTM-agar plates containing 3% agar (Difco) were used for growth at temperatures below 75°C, whereas EvTM-Gelrite plates containing 1.5% Gelrite were used for growth at higher temperatures to minimize desiccation. Electrocompetent HB8 cells were prepared using a protocol similar to that previously described (de Grado et al., 1999). Cells (500 mL) were grown to mid-logarithmic phase in EvTM medium (A$_{600}$ ≈ 0.7) at 65°C, concentrated by centrifugation at 5,000 rpm for 30 min, washed with 10% glycerol (250 mL) two times at room temperature, resuspended in 10% glycerol (2 mL), and frozen at -80°C in

aliquots (100 µL). Transformations were performed by mixing 3 µg of vector (100-800 µg/µL) and 100 µL of electrocompetent cells, incubating the mixture on ice for one hour, and electroporating the cells with a pulse of 12.5 kV/cm in 0.1 cm cuvettes using a BioRad MicroPulser. Cells were immediately transferred to EvTm medium (5 mL) that had been prewarmed to 60°C and incubated for 4 hrs at 60°C in a 50 mL flask shaking at 150 rpm. Cells transformed with pJJS-derived expression vectors (100 µL) were plated onto EvTM-agar plates containing 15 µg/mL bleocin and incubated at 65°C for 72 hours (Brouns et al., 2005). Cells transformed with the *adk* gene replacement vectors were plated onto EvTM-agar plates containing 250 µg/mL kanamycin and incubated at 60°C for 72 hours. Desiccation was minimized during incubation by placing agar plates into Ziploc bags, removing extra air from the bags, and sealing all but one inch of the bag. To avoid condensation on plates during incubation, freshly poured plates were incubated with the agar facing up at 37°C for 1-2 hours prior to use.

4.3.c   Constructing an AK Gene Replacement Vector for *T. thermophilus* HB8.

Figure 4.1 on page 114 illustrates how the vector used for *adk* gene replacement was built. A 1,600 bp segment of *T. thermophilus* HB8 genomic DNA including approximately 1,000 bp adjacent to the *adk* start codon, the *adk* gene, and 45 bp after the *adk* stop codon was PCR amplified from genomic DNA and cloned into pUC18 using *hindIII* and *xbaI*. In addition, approximately 1,000 bp of genomic DNA including the last 121 bp of the *adk* gene and sequence adjacent to the *adk* stop codon was PCR amplified from genomic DNA using and cloned into the pUC18-derived vector containing the first amplicon using *xbaI* and *ecoRI* to create pTT1. The gene encoding a highly thermostable kanamycin nucleotidyltransferase

(*htk*) (Hoseki et al., 1999) with only a RBS was PCR amplified from pMK-18 (de Grado et al., 1999) and cloned into pTT1 using *xbaI* to obtain pTT2. This plasmid was modified by two QuikChange mutagenesis reactions to create pTT2-BK, a plasmid that has a unique *bmtI* site 8 bp prior to the *adk* start codon and a *kpnI* site 147 bp before the *adk* stop codon. The *adk* gene from *Geobacillus stearothermophilus* was PCR amplified from genomic DNA and cloned into pTT2-BK using *bmtI* and *kpnI* to obtain pTT-GsteAK. Selection of HB8 transformed with pTT-GsteAK on EvTM-agar plates containing 250 µg/mL kanamycin yielded multiple colonies. After curing these strains of their plasmid by sequentially streaking them onto two EvTM-agar plates lacking antibiotic, PCR amplification with primers complementary to the HB8 *adk* gene revealed that the native *adk* had not been removed from the chromosome (*data not shown*), suggesting that the small amount of genomic DNA separating the *adk* and *htk* genes in pTT-GsteAK had facilitated off-pathway recombination. To remove this DNA, the *htk* resistance cassette was PCR amplified and cloned into the *kpnI* and *xbaI* sites in pTT-GsteAK to create pTTΔ200-GsteAK, which was the vector successfully used for gene replacement.

Figure 4.1 Creation of an *adk* gene replacement vector. HB8 chromosomal DNA was PCR amplified and cloned into pUC18 to create pTT1, a vector harboring *T. thermophilus adk* (white) and approximately one thousand base pairs of adjacent chromosomal DNA (hatched). The DNA cloned into pTT1 contains a tandem duplication of the last 121 bp of *adk*$_{Tt}$ and 45 bp of DNA adjacent to the *adk*$_{Tt}$ stop codon. The gene encoding a thermostable kanamycin nucleotidyltransferase (*htk*) fused to a ribosomal binding site was cloned into the *xbaI* site of pTT1 to create pTT2. Quikchange mutagenesis was used to introduce *bmtI* and *kpnI* sites into pTT2 to create pTT2-BK, and the gene encoding *G. stearothermophilus* adenylate kinase (*adk*$_{Gs}$) was cloned into these new restriction sites to create pTT-GsteAK. To prevent chromosomal recombination of *htk* without *adk*$_{Gs}$, *htk* was subcloned into pTT-GsteAK using the *kpnI* and *xbaI* sites to create pTTΔ200-GsteAK.

Chapter 4. Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus Thermophilus*

114

4.3.d *T. thermophilus* Strain Growth Rate Measurements.

Cultures of HB8 and PQN1 cells grown overnight at 60°C in EvTM were used to inoculate 50 mL flasks of prewarmed EvTM. Cells were grown over a range of temperatures (60 to 85°C), and the optical density at 600 nm was measured to acquire growth curves until the stationary phase. Each data trace was fit in Kaleidagraph to a modified Gompertz equation (Zwietering et al., 1990), $\ln(N/N_0) = A*(\exp(-\exp((\mu_m*(2.71828183)/A)*(\lambda\text{-}t)+1)))$, where N is the population size as determined by A_{600} measurements, N_0 is the initial A_{600} upon inoculation, A is the asymptotic value of the curve, λ is the lag period, and μ_m is the maximum specific growth rate. For each strain and temperature, I report the mean and standard error for the maximum specific growth rate calculated from three independent measurements.

4.3.e Protein Expression Vectors Constructed and Used.

The *E. coli—T. thermophilus* shuttle vector pWUR112 (Brouns et al., 2005) containing a thermostable bleomycin selectable marker (*shble*) was modified through QuikChange mutagenesis to remove the unique *ecoRI* and *xbaI* restriction sties and create pWUR112-ΔEX (Figure 4.2.A, next page). An ampicillin resistance cassette (*bla*) PCR amplified with primers that incorporated flanking *kpnI-ecoRI-notI-xbaI and speI-notI-pstI-kpnI* sites was cloned into the *kpnI* site of pWUR112-ΔEX to create pJJS. The strong constitutive *slpA* promoter (Faraldo et al., 1992) was PCR amplified from genomic DNA and subcloned into pJJS using *xbaI* and *pstI*, replacing the *bla* gene, to create pJJS-Pro (sequence shown in Figure 4,2.B, next page). Both pJJS and pJJS-Pro are compatible with the BioBrick assembly strategy (Shetty et al., 2008), which was used for construction of the majority of protein expression vectors (Tables 4.1 and 4.2 on pages 122 and 123).
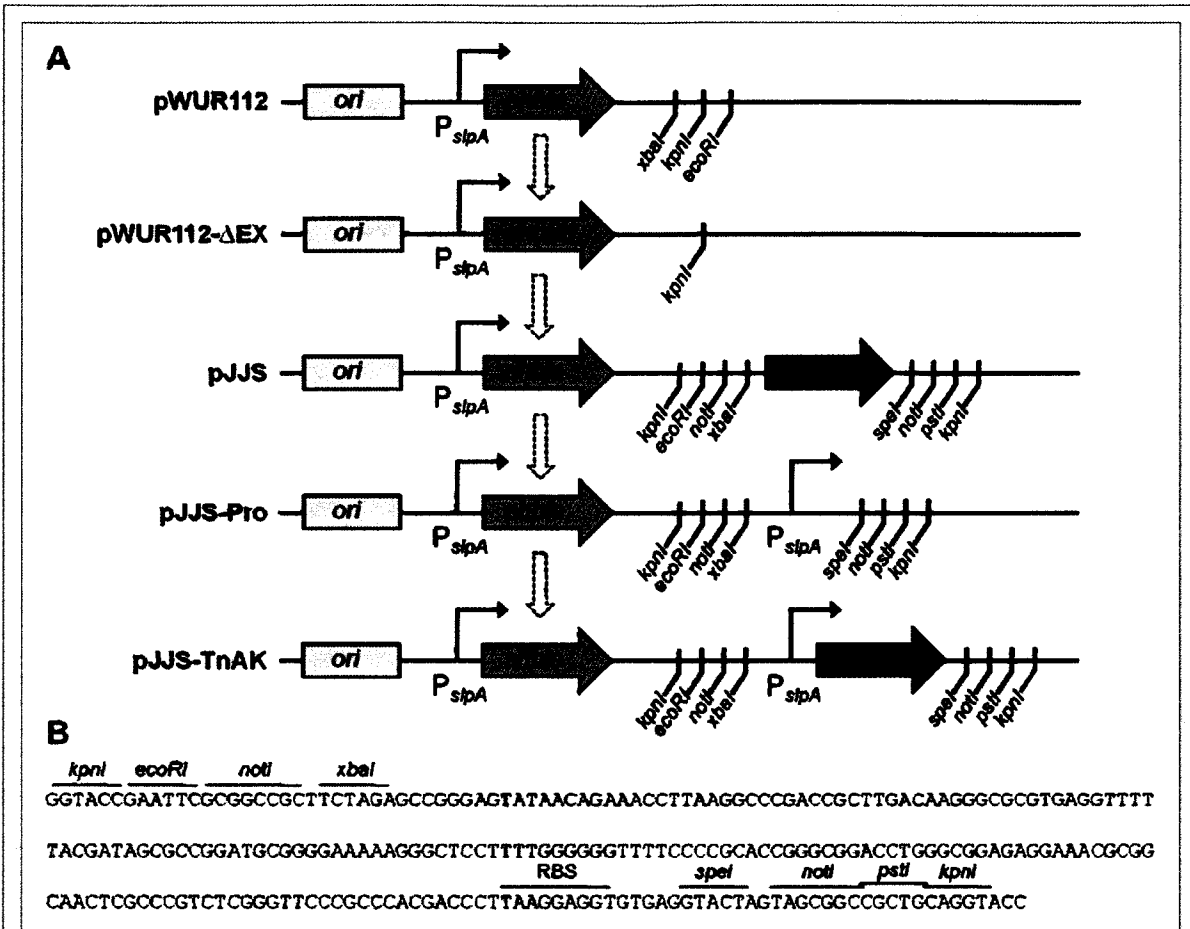
Figure 4.2 Creation of *E. coli–T. thermophilus* shuttle vectors for protein expression.
A) The *E. coli-T. thermophilus* shuttle vector pWUR112 containing a bleomycin selectable marker (*shble*) that has been evolved to function in hyperthmophiles was modified through Quikchange mutagenesis to remove the unique *ecoRI* and *xbaI* restriction sites and create pWUR112-ΔEX. An ampicillin resistance cassette (*bla*) amplified with primers that incorporated flanking *kpnI-ecoRI-xbaI* and *speI-pstI-kpnI* was cloned into the *kpnI* site of pWUR112-ΔEX to create pJJS. A PCR product containing the *T. thermophilus slpA* promoter with flanking *xbaI* and *speI-pstI* sites was cloned into pJJS using *xbaI* and *pstI* to create pJJS-Pro, and *adk_{Tn}* was PCR amplified with flanking *ecoRI-xbaI*, digested with XbaI and PstI, and clone into the *speI* and *pstI* sites of pJJS-Pro to create pJJS-TnAK. B) The multicloning site of pJJS-Pro contains a promoter and ribosomal binding site.

The *Thermotoga neapolitana adk* gene was PCR amplified from pAK$_{Tn2}$::Km (Nguyen et al., 2008) using primers that incorporate flanking *ecoRI-xbaI* and *speI-pstI* restriction sites, digested with *xbaI* and *pstI*, and cloned into pJJS-Pro that had been digested with *speI* and *pstI* to create pJJS-AK$_{Tn}$. In addition, fragments of the *T. neapolitana adk* gene that encode residues 1-79 (TnN) and 80-220 (TnC) were PCR amplified respectively from pET21d-TnN and pET24d-TnC, previously used for the experiments in chapter 2 of this work, using primers that incorporate flanking *ecoRI-notI-xbaI* and *speI-notI-pstI* restriction sites and cloned into pJJS-Pro using a similar protocol to create pJJS-TnN and pJJS-TnC, respectively. A vector for coexpressing TnN and TnC, pJJS-N+C, was created by digesting pJJS-TnC with *xbaI* and *pstI*, and subcloning the excised gene into pJJS-TnN that had been digested with *speI* and *pstI*. QuikChange mutagenesis was used to modify pJJS-TnC to create vectors that express TnC with C133A (pJJS-TnC133) and C156A (pJJS-TnC156) mutations. These plasmids were then used to create the coexpression vectors pJJS-N+C133 and pJJS-N+C156, using a cloning scheme similar to that described for pJJS-N+C.

The genes encoding the P1 and P2 domains of CheA (CheA^{P1P2}; residues 1-264), full-length CheX, and full-length CheY were PCR amplified from *Thermotoga maritima* MSB8 genomic DNA (ATTC #43589D-5) using primers that incorporate flanking *ncoI* and *nheI* restriction sites for the CheA^{P1P2} insert and *pciI* and *nheI* sites for the CheY insert. The PCR amplified chemotaxis genes were cloned in place of the genes encoding E3 and K3 peptides in the vectors pET21d-E3N and pET24d-K3C using the *ncoI* and *nheI* sites (Nguyen et al., 2008) to create vectors (p21d-AN, p21d-AC, p21d-YN, and p21d-YC) that produce TnN and TnC fragments fused at their N-terminus to CheA^{P1P2} and CheY through a linker

that is predicted to be flexible (GASGGGSSGGHM). The gene encoding this linker contains *nhel, xhol,* and *ndel* restriction sites.

The genes encoding chemotaxis proteins fused to AK fragments were PCR amplified from p21d-AN, p21d-AC, p21d-YN, and p21d-YC using primers that incorporate flanking *xbaI* and *speI-notI-pstI*, digested with *xbaI* and *pstI*, and cloned into pJJS-Pro digested with *speI* and *pstI* to yield plasmids that use the *slpA* promoter to constitutively express CheA^{P1P2}-TnN (pJJS-AN), CheA^{P1P2}-TnC (pJJS-AC), CheY-TnN (pJJS-YN), and CheY-TnC (pJJS-YC). The gene encoding *T. maritima* CheX was PCR amplified from genomic DNA using primers that remove its stop codon and incorporate flanking *ecoRI-notI-xbaI* and *nhel* sites and cloned into pJJS-AN using *ecoRI* and *nhel* to yield pJJS-noProXN, a vector that encodes CheX fused to the N-terminus of TnN through a GASGGGSSGGHM linker. The gene encoding TnC was subcloned from pJJS-AC into pJJS-noProXN using *nhel* and *pstI* to create pJJS-noProXC, a vector that encodes CheX fused to the N-terminus of TnC through the same linker as above. Vectors that constitutively express CheX-TnN (pJJS-XN) and CheX-TnC (pJJS-XN) were created by digesting pJJS-noProXN and pJJS-noProXC with *xbaI* and *pstI* and cloning the insert into *speI* and *pstI* digested pJJS-Pro. Vectors that express C156A mutants of CheA^{P1P2}-TnC (pJJS-AC156), CheX-TnC (pJJS-XC156), and CheY-TnC (pJJS-YC156) were created by introducing C156A mutations into the *adk* gene fragments in pJJS-AC, pJJS-YC, and pJJS-XC.
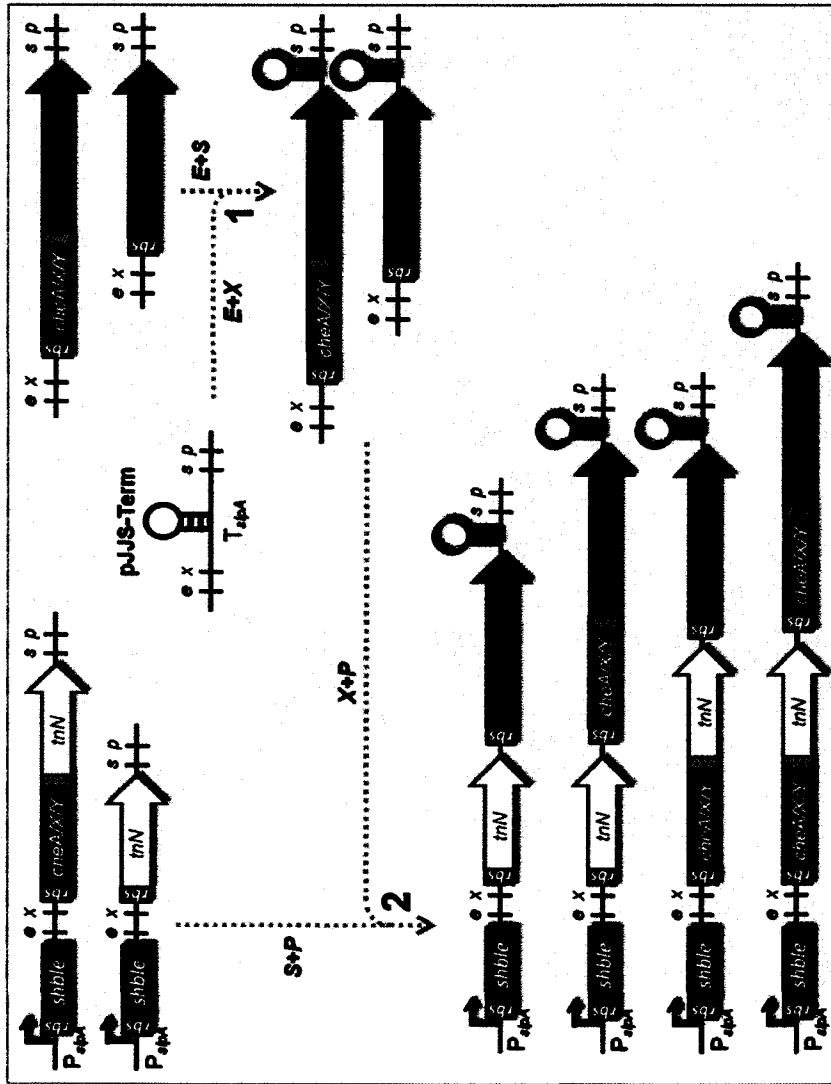
Vectors that use two promoters to coexpress the nine different combinations of TnN and TnC protein fusions were built from pJJS-Pro derived vectors using the BioBrick standardized gene-assembly strategy (Shetty et al., 2008). Promoter-gene fusions were obtained from pJJS-Pro derived vectors using EcoRI and SpeI (or XbaI and PstI) treatment

and subcloned into another pJJS plasmid harboring Promoter-gene fusions by EcoRI and

SpeI (or XbaI and PstI) subcloning to acquire the combinatorial coexpression plasmids. The

resulting vectors (pJJS-AN+AC156, pJJS-AN+XC156, pJJS-AN+YC156, pJJS-XN+AC156,

pJJS-XN+XC156, pJJS-XN+YC156, pJJS-YN+AC156, pJJS-YN+XC156, and pJJS-

YN+YC156) use two distinct $slpA$ promoters to drive the expression of TnN and TnC^{C156A}

protein fusions. Control vectors that use two $slpA$ promoters to coexpress the six possible

combinations of TnN and TnC fragments having only a single polypeptide fused to a

chemotaxis protein were created using a similar strategy (pJJS-AN+C156, pJJS-XN+C156,

pJJS-YN+C156, pJJS-N+AC156, pJJS-N+XC156, and pJJS-N+YC156).

As an alternative strategy for creating vectors that express TnN and TnC^{C156A}

protein fusions at lower levels, plasmids were generated that produce both fragments from a

single polycistronic transcript (Figure 4.3) on page 121. The $slpA$ transcriptional terminator

(Faraldo et al., 1992) was PCR amplified from $T.$ $thermophilus$ HB8 genomic DNA using

primers that incorporate flanking $ecoRI$-$notI$-$xbaI$ and $speI$-$notI$-$pstI$ sites, digested with

$ecoRI$ and $pstI$, and cloned into pJJS digested with $ecoRI$ and $pstI$ to yield pJJS-Term. The

genes encoding $CheA^{P1P2}$-TnC156 (and TnC156) fused to $T.$ $thermophilus$ ribosomal binding

sites were PCR amplified from pJJS-AC156 and pJJS-TnC156 using primers that incorporate

flanking $ecoRI$-$notI$-$xbaI$ and $speI$-$notI$-$pstI$ sites, digested with $ecoRI$ and $speI$, and cloned

into pJJS-Term to produce pJJS-rbsAC156-T$_{slpA}$ and pJJS-rbsTnC156-T$_{slpA}$. In addition, the

gene encoding CheY-TnN fused to the same ribosomal binding site was PCR amplified from

pJJS-YN using primers that incorporate flanking $ecoRI$-$notI$-$xbaI$ and $speI$-$notI$-$pstI$ sites,and

cloned into pJJS by ecoRI and $pstI$ to yield pJJS-rbsYN. Finally, pJJS-rbsAC156-T$_{slpA}$ and

pJJS-rbsTnC156-T$_{slpA}$ were digested with $xbaI$ and $pstI$, and the genes encoding $CheA^{P1P2}$-

TnC156 (and TnC156) fused to ribosomal binding sites and transcriptional terminators were cloned into SpeI and PstI digested pJJS-rbsYN to obtain pJJS-rbsYN+rbsC156-T$_{slpA}$ and pJJS-rbsYN+rbsAC156-T$_{slpA}$. All protein expression vectors are listed in Tables 4.1 and 4.2 on pages 122 and 123.

Chapter 4. Design and Construction of an Adenylate Kinase Protein Fragment Complementation System in *Thermus Thermophilus*

120

Figure 4.3 A two-step approach was used to create vectors that coexpress TnC and TnN protein fusions. Master pJJS plasmids containing fragments or fusion proteins with an N-terminal RBS sequence were acquired by simple subcloning from plasmids containing these genes under the control of the full *slpA* promoter. In step 1, genes encoding each TnC156 fragment alone or fused at their N-terminus to chemotaxis proteins through a flexible GASGGGSSGGHM linker were digested with EcoRI and SpeI and cloned into the EcoRI and XbaI digested pJJS-Term. This results in plasmids containing the genes in front of the *slpA* terminator. In step 2, the above vectors were digested with XbaI and PstI and cloned into SpeI and PstI digested pJJS plasmids containing genes encoding each TnN fragment alone or fused at their N-terminus to chemotaxis proteins through a flexible GASGGGSSGGHM linker. The final coexpression plasmids contain combinations of RBS-containing fragments or fusion genes downstream of the *shble* resistance gene. Coexpression of all three genes as a polycistronic transcript will be driven by the *slpA* promoter in front of the *shble* gene and terminated by the *slpA* terminator behind the TnC156-containing gene.

Chapter 4. Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus Thermophilus*

121

Table 4.1 Plasmids derived from pJJS that were used to construct protein expression vectors. The vector name, protein construct and any 5'and/ or 3' regulatory DNA is indicated.

Name	5' DNA	Protein coding region	3' DNA
pJJS	---	---	---
pJJS-Pro	P_{slpA}-RBS	---	---
pJJS-Term	---	---	T_{slpA}
pJJS-TnN	P_{slpA}-RBS	TnN	---
pJJS-TnC	P_{slpA}-RBS	TnC	---
pJJS-TnC133	P_{slpA}-RBS	TnC^{C133A}	---
pJJS-TnC156	P_{slpA}-RBS	TnC^{C156A}	---

Constructs used as templates for PCR.

Name	5' DNA	Protein coding region	3' DNA
pJJS-AN	P_{slpA}-RBS	CheA^{P1P2}–GASGGGSSGGHM–TnN	---
pJJS-XN	P_{slpA}-RBS	CheX–GASGGGSSGGHM–TnN	---
pJJS-YN	P_{slpA}-RBS	CheY–GASGGGSSGGHM–TnN	---
pJJS-AC156	P_{slpA}-RBS	CheA^{P1P2}–GASGGGSSGGHM-TnC^{C156A}	---
pJJS-XC156	P_{slpA}-RBS	CheX–GASGGGSSGGHM-TnC^{C156A}	---
pJJS-YC156	P_{slpA}-RBS	CheY–GASGGGSSGGHM–TnC^{C156A}	---

RBS Gene Fusion Constructs.

Name	5' DNA	Protein coding region	3' DNA
pJJS-rbsTnC156-T_{slpA}	RBS	TnC^{C156A}	T_{slpA}
pJJS-rbsAC156-T_{slpA}	RBS	CheA^{P1P2}–GASGGGSSGGHM–TnC^{C156A}	T_{slpA}
pJJS-rbsXC156-T_{slpA}	RBS	CheX–GASGGGSSGGHM–TnC^{C156A}	T_{slpA}
pJJS-rbsYC156-T_{slpA}	RBS	CheY–GASGGGSSGGHM–TnC^{C156A}	T_{slpA}
pJJS-rbsTnN	RBS	TnN	---
pJJS-rbsAN	RBS	CheA^{P1P2}–GASGGGSSGGHM–TnC^{C156A}	---
pJJS-rbsXN	RBS	CheX–GASGGGSSGGHM–TnC^{C156A}	---
pJJS-rbsYN	RBS	CheY–GASGGGSSGGHM–TnC^{C156A}	---

Chapter 4. Design and Construction of an Adenylate Kinase Protein Fragment Complementation System in *Thermus Thermophilus*

122

Table 4.2 Vectors used for complementation studies in *T. thermophilus* PQN1.

Name	Proteins expressed	Expression
pJJS-AK$_{Tn}$[‡]	AK$_{Tn}$	monocistronic
pJJS-N+C[‡]	TnN & TnC	monocistronic
pJJS-N+C133	TnN & TnC^{C133A}	monocistronic
pJJS-N+C156	TnN & TnC^{C156A}	monocistronic

Constructs for Y/A experiments

pJJS-N+C156	TnN & TnC^{C156A}	polycistronic
pJJS-YN+AC156	CheY–TnN & CheA^{P1P2}–TnC^{C156A}	polycistronic
pJJS-YN+XC156	CheY–TnN & CheX–TnC^{C156A}	polycistronic
pJJS-YN+YC156	CheY–TnN & CheY–TnC^{C156A}	polycistronic
pJJS-YN+C156	CheY–TnN & TnC^{C156A}	polycistronic
pJJS-N+AC156	TnN & CheA^{P1P2}–TnC^{C156A}	polycistronic

Constructs for X/X experiments

pJJS-XN+XC156	CheX–TnN & CheX–TnC^{C156A}	polycistronic
pJJS-N+XC156	TnN & CheX–TnC^{C156A}	polycistronic
pJJS-XN+C156	CheX–TnN & TnC^{C156A}	polycistronic
pJJS-XN+AC156	CheX–TnN & CheA^{P1P2}–TnC^{C156A}	polycistronic

4.3.f Complementation of the Temperature-sensitive PQN1 Strain.

Complementation analysis of the monocistronic constructs was assessed by streaking colonies obtained from transformations on solid medium in glass petri dishes, and incubating these plates at temperatures ($\geq 78°C$) where PQN1 cannot grow like parental HB8. Complementation studies involving the polycistronic constructs were performed by spotting a defined titer of cells grown at 65°C onto EvTM-Gelrite plates, and evaluating growth after 24 hours at 78°C. In the spotting experiments, 5 mL EvTM liquid cultures containing 5 µg/mL bleocin were inoculated with colonies obtained from transformations and cultured overnight at 65°C while shaking at 150 rpm. The optical density of each culture was measured after 24 hours, and 1 mL of each culture was diluted to an A_{600} of 0.5, pelleted by centrifugation, and resuspended in 250 mL of 25% glycerol. Serial dilutions (1x, 10x, 100x, and 1000x) of the resuspended cells (10 µL each) were spotted onto EvTM-Gelrite plates. After incubation at 78°C for 24 hours, growth at each spot was analyzed using a FluorChem 5500 imager (Alpha-Innotech), and the program ImageJ was used to quantify the relative intensities of growth at each spot (Abramoff et al., 2004).

4.3.g Protein Purification of *T. maritima* MSB8 Chemotaxis Proteins.

T. maritima MSB8 CheA^{P1P2}, CheX, and CheY were overexpressed in *E. coli* Rosetta cells (Novagen) using plasmids kindly provided by the Crane Lab that express these proteins with N-terminal (His)$_6$ tags (Park, Beel, et al., 2004). Cells transformed with these plasmids were grown at 37°C in LB supplemented with 25 µg/mL kanamycin, protein expression was induced at an A_{600} of ~0.5 by adding 1 mM isopropylthio-β-D-galactoside (IPTG), cells were harvested by centrifugation after 3 hours of growth, and cells were lysed by resuspending

them in PSI buffer (50 mM phosphate pH 7.5, 150 mM NaCl, 10 mM imidazole) containing

1 mM $MgCl_2$, 300 μg/mL lysozyme, and 2 U/mL DNase I. Cells were frozen at -80°C,

thawed, and centrifuged at 15k x g for 1 hr. Cleared lysate was filtered through a sterile 0.2

micron syringe filter, applied to a 2 mL nickel talon affinity column (Qiagen) equilibrated

with PSI buffer, washed with 10 column volumes of PSI buffer, and bound protein was

eluted using PSI containing 250 mM imidazole. The elution containing $CheA^{P1P2}$ (theoretical

pI = 4.6) was diluted 100 fold into 50 mM phosphate pH 7.5 buffer, applied to a 5 mL

HiTrap-Q HP anion exchange column (GE Healthcare), and eluted using a linear gradient

from 0 to 500 mM NaCl in 50 mM phosphate pH 7.5. All purified proteins were dialyzed

overnight into PS buffer (50 mM phosphate pH 7.5 and 150 mM NaCl), and stored at -80°C.

### 4.3.h Analytical methods.

An ÄKTA FPLC was used for all protein purification. Protein concentrations were

determined by measuring their absorbance with a Cary 50 spectrophotometer, using the

calculated extinction coefficients $\varepsilon_{280}(TmCheA^{P1P2}) = 5,120$ $M^{-1}cm^{-1}$, $\varepsilon_{280}(TmCheX) = 3,840$

$M^{-1}cm^{-1}$, and $\varepsilon_{280}(TmCheY) = 2,560$ $M^{-1}cm^{-1}$, as calculated from the primary sequence by the

PEPSTATS algorithm (Rice et al., 2000). Far-UV circular dichroism (CD) spectra were

recorded on a Jasco J-815 spectropolarimeter using samples that contained 5 μM protein in

PS buffer. Each spectrum was acquired in triplicate using a data pitch of 1 nm and a scan rate

of 100 nm/min, and temperature scans were performed at a rate of 1°C per minute. All

spectra are shown as mean residue ellipticity and corrected for the ellipticity of the buffer

alone. Dynamic light scattering was performed with a Malvern Instruments Zetasizer Nano

ZS using samples that contained 115 μM of each protein in PS buffer. These samples were

analyzed in triplicate using a 1 cm quartz cuvette at a backscattering angle of 173° after a 5 minute incubation at 78°C. Data processing was performed using the high resolution analysis model provided with Zetasizer software v6.01.

4.4 Results

4.4.a PQN1: Constructing a T. thermophilus HB8 Complementation Strain by Genomic Recombination of the *adk* gene.

To generate a *Thermus thermophilus* HB8 mutant for evaluating AK fragment complementation above 75°C, homologous recombination was used to replace its *adk* gene with the *Geobacillus stearothermophilus* ortholog *adk*$_{Gs}$ that encodes a protein that is not functional above 70°C *in vitro* (Bae et al., 2004). Gene replacement was performed by transforming HB8 cells with the pTTΔ200-GsteAK plasmid and selecting for growth at 60°C on EvTM-agar plates containing 250 μg/mL kanamycin as previously described (Cameron et al., 2004). Since HB8 cells cannot replicate this plasmid, they are only able to exhibit kanamycin resistance if the *htk* gene is integrated into the chromosome as illustrated in Figure 4.4.A on the next page. Kanamycin-resistant colonies were screened for growth at 80°C in EvTM-liquid culture, and strains identified as nonviable at this temperature were cured of their plasmid by restreaking them twice onto EvTM-agar plates lacking antibiotic at 60°C. PCR analysis of genomic DNA from one of the strains obtained (designated PQN1) confirmed integration of *adk*$_{Gs}$ into the targeted chromosome locus (Figure 4.4.B, next page), and DNA sequencing of these amplicons revealed that PQN1 expresses an AK$_{Gs}$ with an E70V mutation.
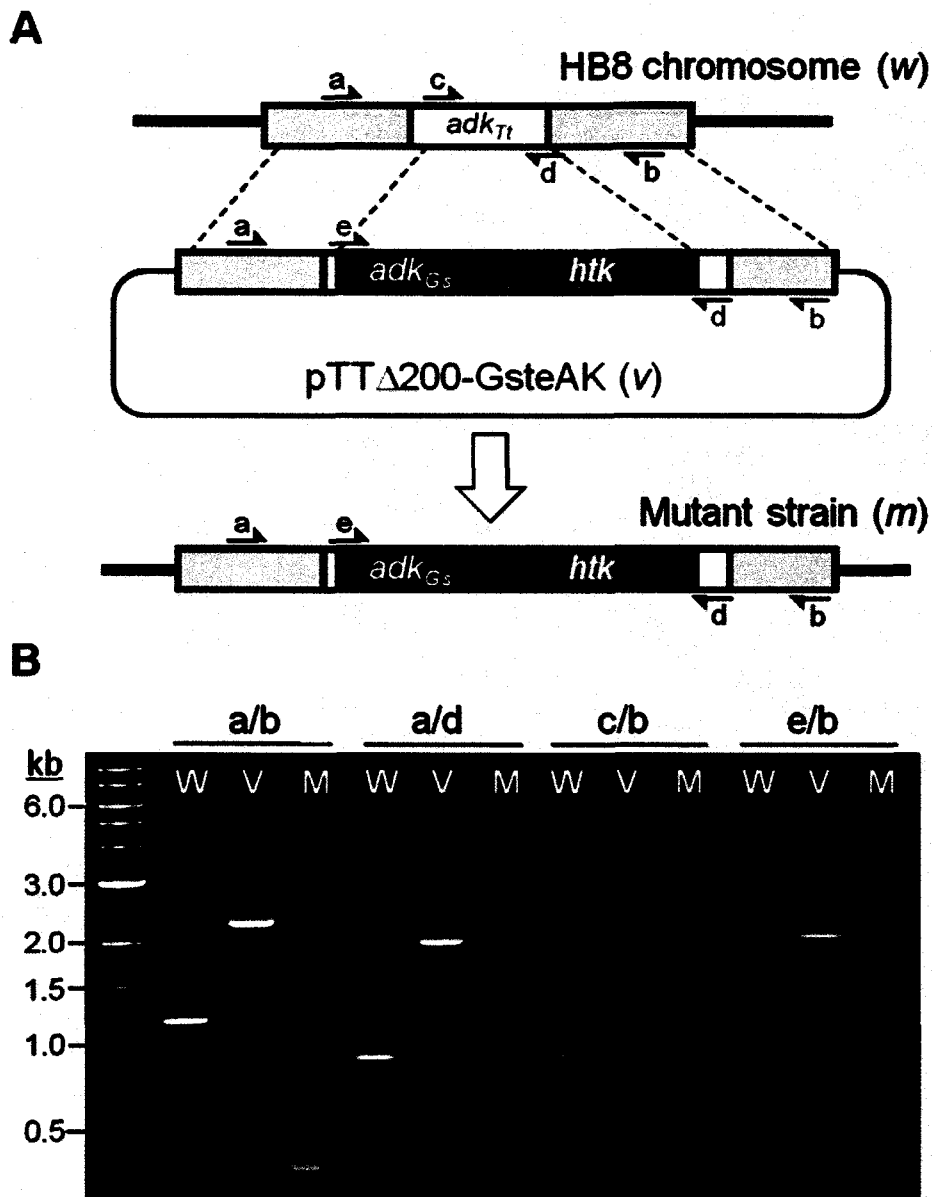
Figure 4.4 Creation and characterization of PQN1, a *T. thermophilus* mutant strain encoding an AK with decreased thermostability. A) Chromosomal replacement of *T. thermophilus adk~Tt~* was induced by transformation of HB8 with pTTΔ200-GsteAK, which contains a fusion of the *G. stearothermophilus adk* gene and the *htk* selectable marker flanked by 982 bp of genomic DNA that is upstream of the *adk~Tt~* start codon and 1,121 bp of genomic DNA that includes the last 121 bp of *adk~Tt~* and 1,000 bp following the gene. This region of *adk~Tt~* was included to ensure that the promoter that drives transcription of the predicted downstream methionine aminopeptidase gene was not disrupted upon recombination. B) Wildtype HB8 genomic DNA (*w*), the pTTΔ200-GsteAK vector (*v*), and PQN1 mutant (*m*) genomic DNA were used as template for PCR reactions involving pairs of primers that bind to genomic DNA flanking *adk~Tt~* (a and b) and sequences within the different *adk* genes (c, d, and e).

Chapter 4. Design and Construction of an Adenylate Kinase Protein Fragment Complementation System in *Thermus Thermophilus*

127

To better characterize the temperature-sensitive phenotype of PQN1, I compared its specific maximal growth rate in liquid medium to the parental strain used for recombination (Zwietering et al., 1990). The results shown in Figure 4.5.A (next page) demonstrate that *T. thermophilus* HB8 grows over a broad temperature range (60 to 85°C) as previously reported (Oshima and Imahori, 1974). In contrast, PQN1 only displayed detectable growth up to 75°C, ten degrees lower than the maximal growth temperature of the parental HB8 strain. I also investigated the effect of temperature on the growth of PQN1 on EvTM-Gelrite solid medium over a range of temperatures (Figure 4.5.B, next page). As in liquid medium, PQN1 did not exhibit significant growth above 75°C under these conditions. However, this temperature-sensitive phenotype could be rescued by transformation with a vector (pJJS-AK_{Tn}) that constitutively expresses the hyperthermophilic AK_{Tn} using the strong constitutive *slpA* promoter (Faraldo et al., 1992).
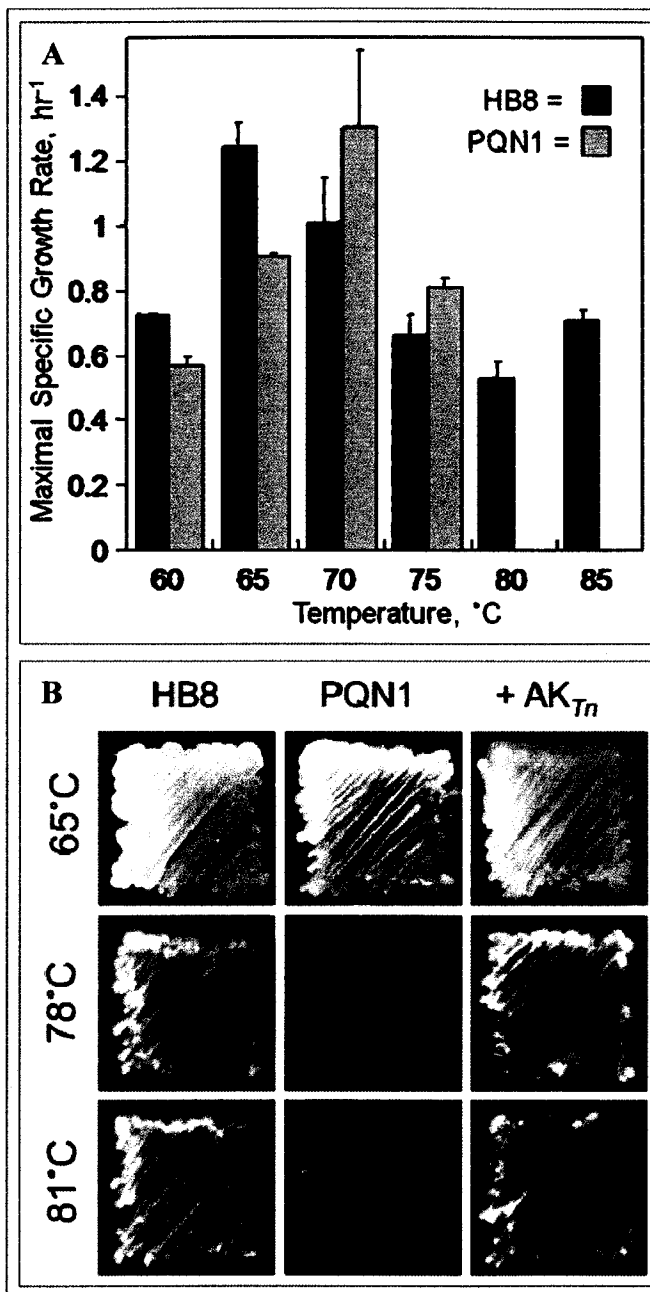
Figure 4.5 The PQN1 recombinant mutant is an AK temperature-sensitive strain. A) Temperature sensitivity of the PQN1 AK-recombinant strain in liquid culture. Single colonies of PQN1 and HB8 cells grown at 65°C on EvTM plates were used to inoculate 5 mL EvTM cultures. After an overnight incubation at 60°C while shaking at 150 rpm, 1 mL of each culture was used to inoculate fresh 50 mL EvTM cultures. These cultures were grown at the indicated temperatures, the change in optical density at 600 nm was measured until the stationary phase was reached (≥ 12 hours), and the growth rate was calculated. For each strain and temperature, the mean and standard error of the mean were calculated from three independent measurements. B) AK_{Tn} complementation of PQN1 on solid medium. Colonies of HB8, PQN1, and PQN1 transformed with a plasmid expressing full-length AK_{Tn} streaked on EvTM-Gelrite plates after incubation at 65, 78, and 81°C for 48 hours. Representative streaks are shown from selections that were performed in triplicate.

Chapter 4. Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus Thermophilus*

129

4.4.b   Testing the *T. neapolitana* Unassisted Adenylate Kinase Fragment Complementation.

To evaluate whether fragments from AK_{Tn}, TnN and TnC, can be used to detect protein complex formation within *T. thermophilus*, I evaluated the growth of PQN1 transformed with pJJS-N+C, a vector that constitutively expresses both polypeptides using *slpA* promoters. Figure 4.6 on the next page shows, surprisingly, that PQN1 growth was complemented by the TnN and TnC fragments above 75°C without the assistance of proteins that drive their association. I next examined whether point mutations within the tetracysteine motif (Glaser et al., 1992) that coordinates zinc abrogate the cooperative function of TnC and TnN sufficiently to allow them to be used as a PCA in PQN1. These cysteines were targeted for mutagenesis, since previous studies have shown zinc binding by this tetracysteine motif in the LID domain enhances the thermostability of AK homologs (Vieille, Krishnamurthy, Hyung-Hwan, et al., 2003; Perrier et al., 1994; Perrier, Burlacu-Miron, Bourgeois, et al., 1998). I found that AK_{Tn} fragments harboring either the Cys133Ala or Cys156Ala mutation could not complement PQN1 growth at 78°C (Figure 4.6, next page). This suggests that disruption of zinc binding in the TnC^{C133A} and TnC^{C156A} mutants causes a reduction in thermostability, as observed previously upon loss of zinc from full-length AK_{Tn} (Vieille, Krishnamurthy, Hyun, et al., 2003).
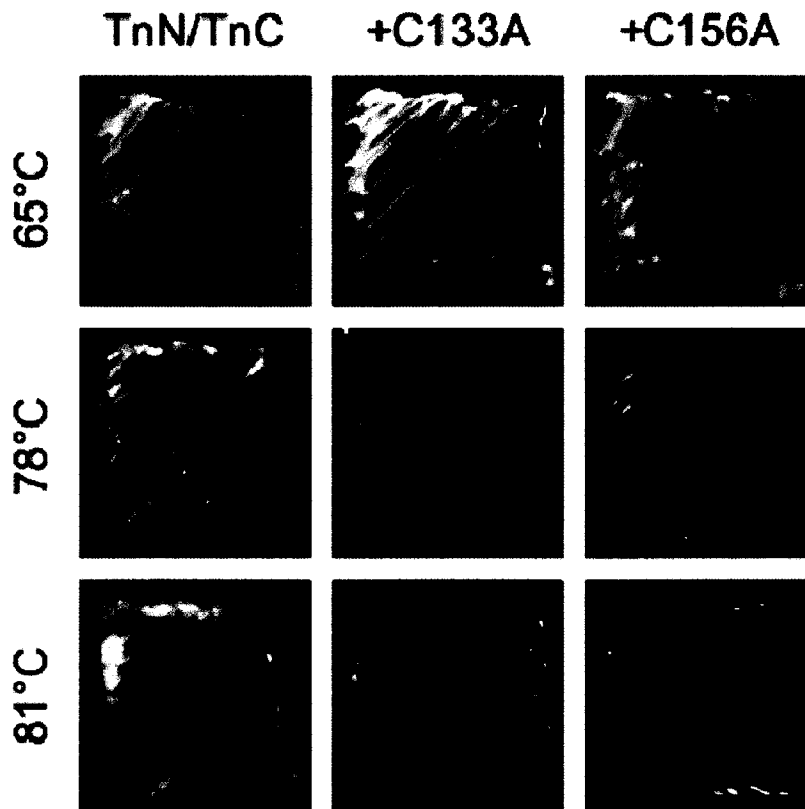
Figure 4.6 Effect of AK$_{Tn}$ fragmentation after residue D77 on PQN1 complementation. Colonies of PQN1 *T. thermophilus* transformed with plasmids that coexpress TnN and TnC, TnN and TnC^{C133A}, and TnN and TnC^{C156A} streaked on EvTM-Gelrite plates after incubation for 48 hours at 65, 78, and 81°C. Representative streaks are shown from selections that were performed in triplicate.

### 4.4.c Prototype ht-PCA: *T. neapolitana Assisted AK fragment complementation.*

To test whether TnN and TnC^{C156A} can be used with PQN1 as a protein-fragment complementation assay to report on protein-protein interactions above 75°C, I examined whether PQN1 growth could be complemented at 78°C by vectors that express these polypeptides fused to different combinations of the *Thermotoga maritima* MSB8 chemotaxis proteins (CheA^{P1P2}, CheX, and CheY). Shown in Figure 4.7 on the next page is a diagram of the chemotactic pathway and protein-protein interactions of these proteins. These proteins were chosen because biochemical and biophysical studies have provided evidence for interactions between CheA^{P1P2} and CheY (Park, Beel, et al., 2004), CheX and CheY (Park, Chao, et al., 2004), and CheX and CheX (Park, Chao, et al., 2004). Among these pairwise interactions, only the CheA^{P1P2} and CheY interaction has been studied at a temperature where thermophiles grow, although the maximum temperature where this interaction has previously been analyzed (70°C) is lower than the minimum temperature (78°C) where the ht-PCA assay can be performed (Park, Beel, et al., 2004).
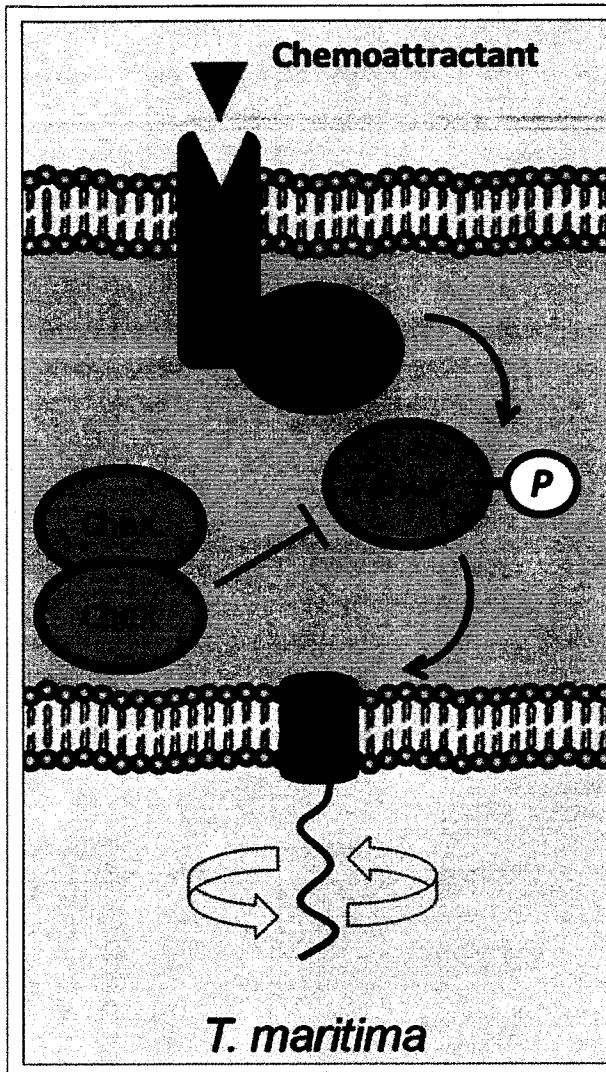
Figure 4.7 The *T. maritima* chemotaxis pathway. The presence of an extracelllar chemoattractant binds to the Methyl-accepting Chemotaxis Protein (MCP, shown as blue) receptor and initiates a phosphorylation cascade by activating CheA. The histidine kinase CheA (red), via its P2 domain, interacts with CheY (purple). CheX homodimers (green) bind to phosphorylated CheY to attenuate flagellar activation.
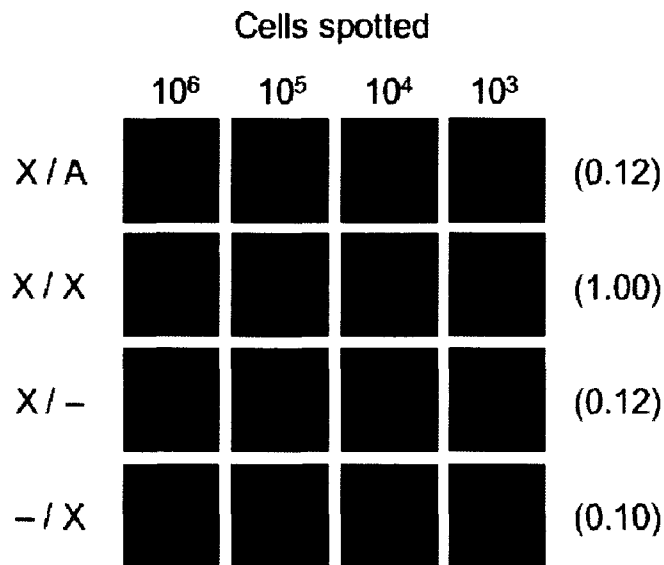
Chapter 4. Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus Thermophilus*

133

Figure 4.8 on page 136 compares the growth of PQN1 alone with cells coexpressing TnN and TnC^{C156A} fused to different combinations of chemotaxis proteins via a twelve amino acid linker. For these experiments, both TnN and TnC^{C156A} fragments were expressed from a single polycistronic transcript with the thermostable bleomycin selectable marker. I found that PQN1 coexpressing CheY-TnN and CheA^{P1P2}-TnC^{C156A} grew after 24 hours of incubation at 78°C at all titers analyzed, similar to that observed with PQN harboring a vector that expresses full-length AK$_{Tn}$. The average magnitude of the growth was calculated for the highest titers analyzed in each experiment by integrating the intensity of replicate spots using ImageJ. This analysis revealed that PQN1 coexpressing CheY-TnN and CheA^{P1P2}-TnC^{C156A} grows ~70% as dense as cells expressing full-length AK$_{Tn}$. In contrast, the growth detected with PQN1 expressing AK fragments fused to only CheY (CheY-TnN and TnC^{C156A}) or only CheA^{P1P2} (TnN and CheA^{P1P2}-TnC^{C156A}) is ≥4-fold lower than that of cells expressing AK fragments fused to CheY and CheA^{P1P2} at the highest titer and undetectable at the lowest titer analyzed. Similarly, little growth was observed after 24 hours with PQN1 harboring vectors that express TnN and TnC^{C156A} fused to nothing, CheY and CheX, and CheY and CheY. Taken together, these findings provide direct evidence that the complementation observed with PQN1 coexpressing CheY-TnN and CheA^{P1P2}-TnC^{C156A} arises because the association of these chemotaxis proteins drives the cooperative function of the AK$_{Tn}$ fragments at the physiological growth temperatures of *T. maritima* as predicted (Park, Beel, et al., 2004). CheY and CheX are also predicted to interact in *T. maritima* (Park, Chao, et al., 2004). However, the lack of growth with CheY and CheX protein fusions is consistent with studies which indicate that CheY must be phosphorylated in order to associate strongly with CheX (Muff et al., 2007). In contrast, CheA and CheY interact

regardless of CheA's phosphorylation state, as the P1 domain is responsible for autophosphorylation and the P2 domain drives CheY binding, and the two functions are independent (Stewart and Van Bruggen, 2004).
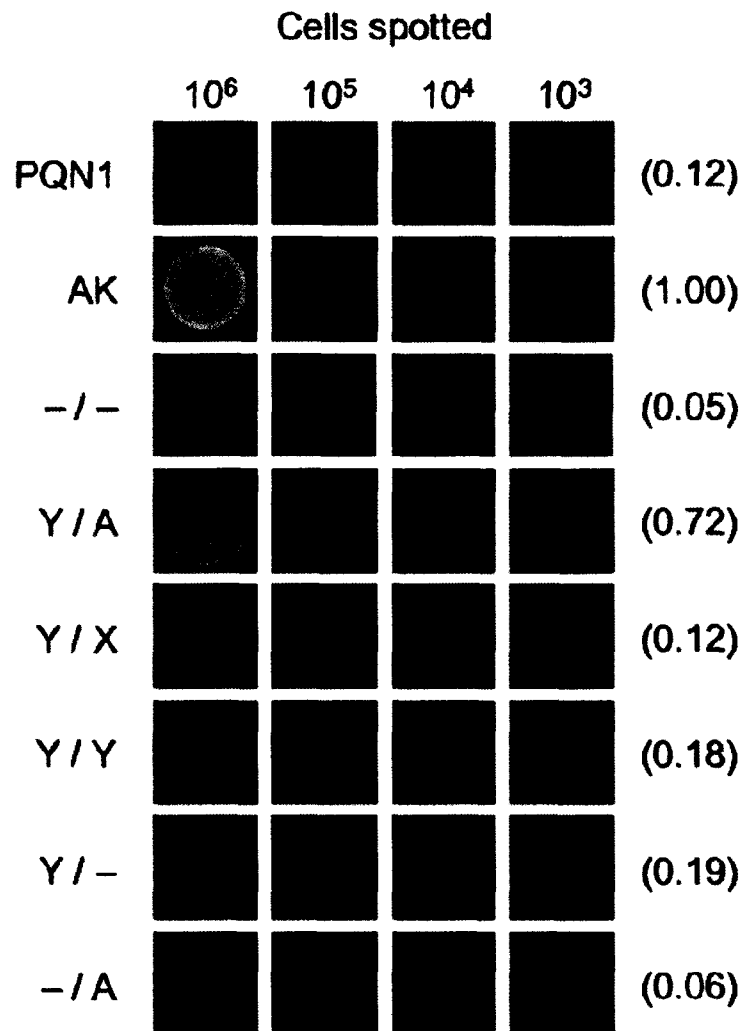
Figure 4.8 Analysis of CheA^{P1P2} and CheY binding using AK$_{Tn}$-fragment complementation. Relative growth after 24 hours at 78°C of PQN1 alone (PQN1) and PQN1 transformed with vectors that express TnN and TnC^{C156A} (–/–), CheY-TnN and CheA^{P1P2}-TnC^{C156A} (Y/A), CheY-TnN and CheX-TnC^{C156A} (Y/X), CheY-TnN and CheY-TnC^{C156A} (Y/Y), CheY-TnN and TnC^{C156A} (Y/–), and TnN and CheA^{P1P2}-TnC^{C156A} (–/A). The different spots represent a dilution series of cells obtained from a liquid culture grown at 65°C spotted onto EvTM-Gelrite plates. ImageJ (Abramoff et al., 2004) was used to quantify the density of cell growth relative to background levels for the spots corresponding to the highest titer, and the relative integrated densities are reported in parenthesis relative to that of PQN cells expressing AK$_{Tn}$. The number of cells spotted were determined by counting colonies obtained from plating cultures onto EvTM plates after 48 hours of incubation at 65°C.

Chapter 4. Design and Construction of an Adenylate Kinase Protein Fragment Complementation System in *Thermus Thermophilus*

136

Since CheX is a dimer in its crystal structure (Park, Chao, et al., 2004), I also investigated whether two CheX molecules could enhance AK_{Tn}-fragment complementation of PQN1 like $CheA^{P1P2}$ and CheY. Figure 4.9 on the next page shows that PQN1 cells coexpressing CheX-TnN and CheX-TnC^{C156A} display detectable growth after a 24 hour incubation at 78°C. At the highest titer analyzed, this growth is $\geq$5.5-fold higher than that of PQN1 cells harboring vectors that only express one of the two AK_{Tn} fragments as a fusion with CheX, *i.e.*, CheX-TnN and TnC^{C156A} or TnN and CheX-TnC^{C156A}. In addition, PQN1 cells coexpressing CheX-TnN and CheX-TnC^{C156A} grow to a density that is ~8-fold higher than that of cells expressing CheX-TnN and CheA-TnC^{C156A}, chemotaxis protein fusions that are not expected to interact and promote AK-fragment function. However, PQN1 were not complemented by CheX-TnN and CheX-TnC^{C156A} to the same extent as with CheY-TnN and CheA^{P1P2}-TnC^{C156A}. Growth was only observed with cells expressing CheX protein fusions at the highest titer of cells analyzed even after extended incubations (up to 48 hours), whereas growth could be detected with CheA^{P1P2} and CheY protein fusions at all titers analyzed after 24 hours.

Cells spotted

	10^6	10^5	10^4	10^3	
X / A	■	■	■	■	(0.12)
X / X	■	■	■	■	(1.00)
X / –	■	■	■	■	(0.12)
– / X	■	■	■	■	(0.10)

Figure 4.9 Analysis of CheX self-association using AK$_{Tn}$-fragment complementation. Relative growth after 24 hours at 78°C of PQN1 cells transformed with plasmids that express CheX-TnN and CheX-TnC^{C156A} (X/X), CheX-TnN and CheA^{P1P2}-TnC^{C156A} (X/A), CheX-TnN and TnC^{C156A} (X/–), and TnN and CheX-TnC^{C156A} (–/X). Cell growth was determined as in Figure 5, except that the integrated densities using ImageJ shown in parenthesis for the spots corresponding to the highest titer rare reported relative to that of PQN1 cells expressing CheX-TnN and CheX-TnC^{C156A}.

4.4.d Verification of *T. maritima* CheA^{P1P2}, CheY, and CheX Protein Interactions.
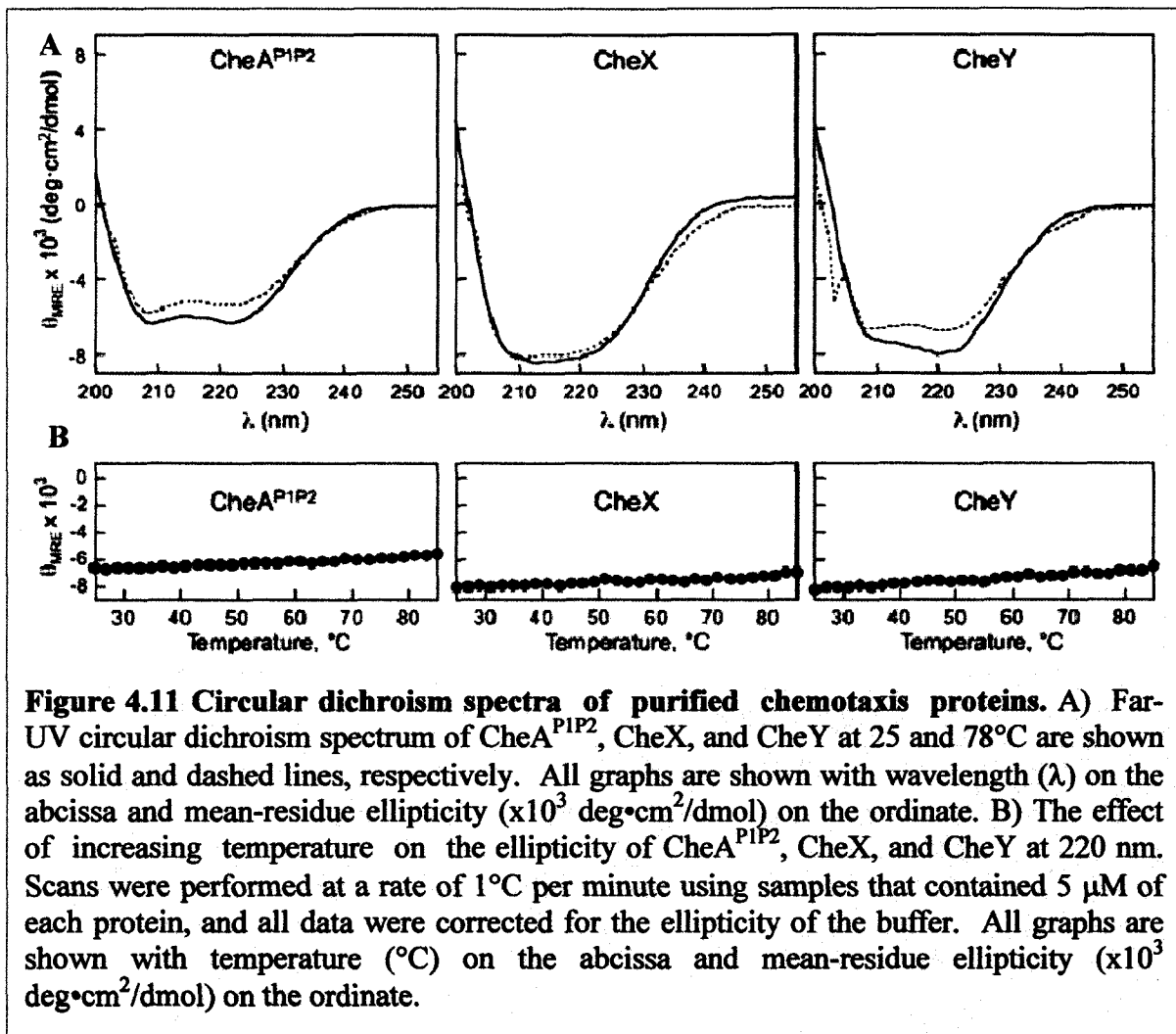
The results from PQN1 complementation analysis provide the first evidence for the homodimerization of CheX at thermophile growth temperatures and the formation of a CheA^{P1P2}–CheY heterodimer above 70°C. To verify that these interactions occur at 78°C, I purified all three chemotaxis proteins, characterized their thermostability, and evaluated their apparent molecular size at the temperature of the ht-PCA assay. SDS-PAGE analysis of each purified chemotaxis protein reveals that they are all ≥95% homogenous (Figure 4.10, next page). In addition, a comparison of the circular dichroism spectra of CheA^{P1P2}, CheX, and CheY measured at 25 and 78°C reveal only minor changes in ellipticity as temperature was increased to the level where the ht-PCA assay is implemented (Figure 4.11, on page 141). This indicates that each protein remains folded at the temperature where the AK$_{Tn}$-based PCA is conducted.
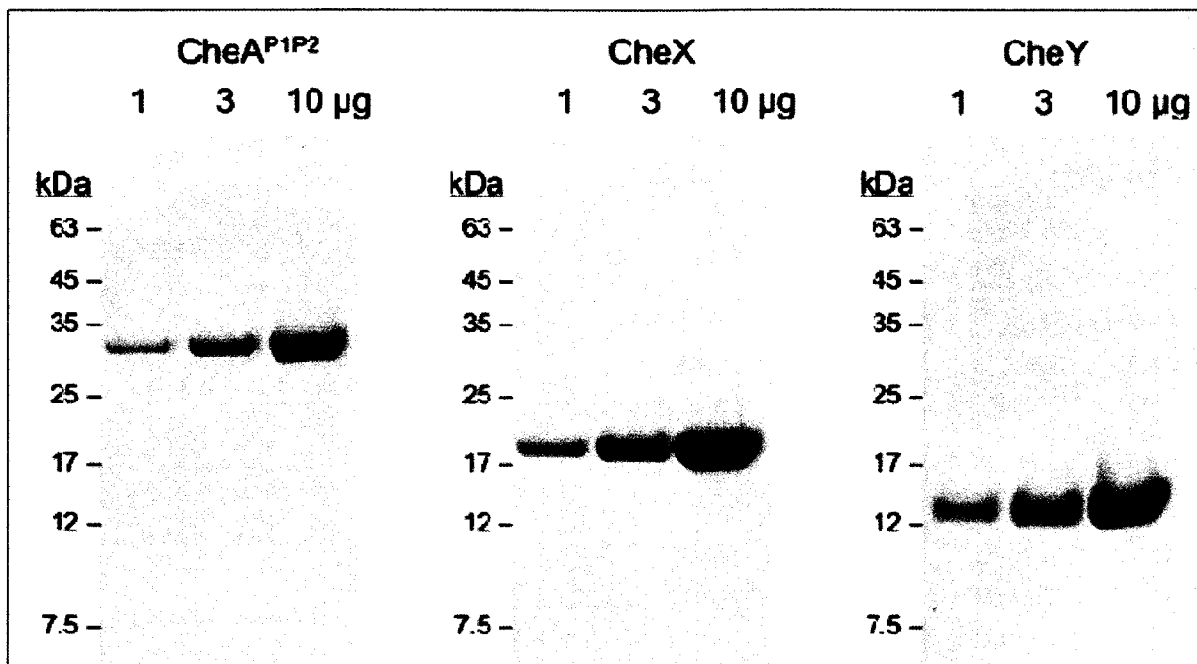
Figure 4.10 SDS-PAGE analysis of purified chemotaxis protein. A comparison of 1, 3, and 10 μg of *T. maritima* CheA^{P1P2}, CheX, and CheY indicate a high degree of purity for each protein. Based on the protein sequences, the molecular weights and A280 extinction coefficients (shown in parentheses) for the CheA^{P1P2} fragment, CheX, and CheY are 32 kDa (ϵ^{280} = 5120), 18.6 kDa (ϵ^{280} = 3840), and 15.4 kDa (ϵ^{280} = 2650), respectively.

Chapter 4. Design and Construction of an Adenylate kinase Protein Fragment Complementation System in *Thermus Thermophilus*

140

Figure 4.11 Circular dichroism spectra of purified chemotaxis proteins. A) Far-UV circular dichroism spectrum of CheA^{P1P2}, CheX, and CheY at 25 and 78°C are shown as solid and dashed lines, respectively. All graphs are shown with wavelength (λ) on the abcissa and mean-residue ellipticity (x10^3 deg•cm^2/dmol) on the ordinate. B) The effect of increasing temperature on the ellipticity of CheA^{P1P2}, CheX, and CheY at 220 nm. Scans were performed at a rate of 1°C per minute using samples that contained 5 μM of each protein, and all data were corrected for the ellipticity of the buffer. All graphs are shown with temperature (°C) on the abcissa and mean-residue ellipticity (x10^3 deg•cm^2/dmol) on the ordinate.

Dynamic light scattering (DLS) was used to assess the size and interactions of $CheA^{P1P2}$, CheX, and CheY. A comparison of the relative hydrodynamic diameters of these proteins determined using DLS at 78°C demonstrated that an equimolar mixture of $CheA^{P1P2}$ and CheY displays a hydrodynamic diameter (8.3±0.1 nm) that is greater than that observed with either $CheA^{P1P2}$ (7.1±0.2 nm) or CheY (4.7±0.1 nm) alone (Figure 4.12, next page), consistent with formation of a complex between these proteins as evidenced by the results from the assisted fragment complementation of PQN1. Furthermore, the hydrodynamic diameter of CheX at 78°C (6.3±0.2 nm) is similar in size to the longest axis of the dimeric form of CheX (6 nm) reported in the crystal structure (Park, Chao, et al., 2004), and bigger than the longest axis of the individual monomers (4.6 nm). Taken together with the finding that CheX retains secondary structure content similar to that at laboratory temperatures, this provides evidence that CheX self-associates as observed in the high-temperature PCA.
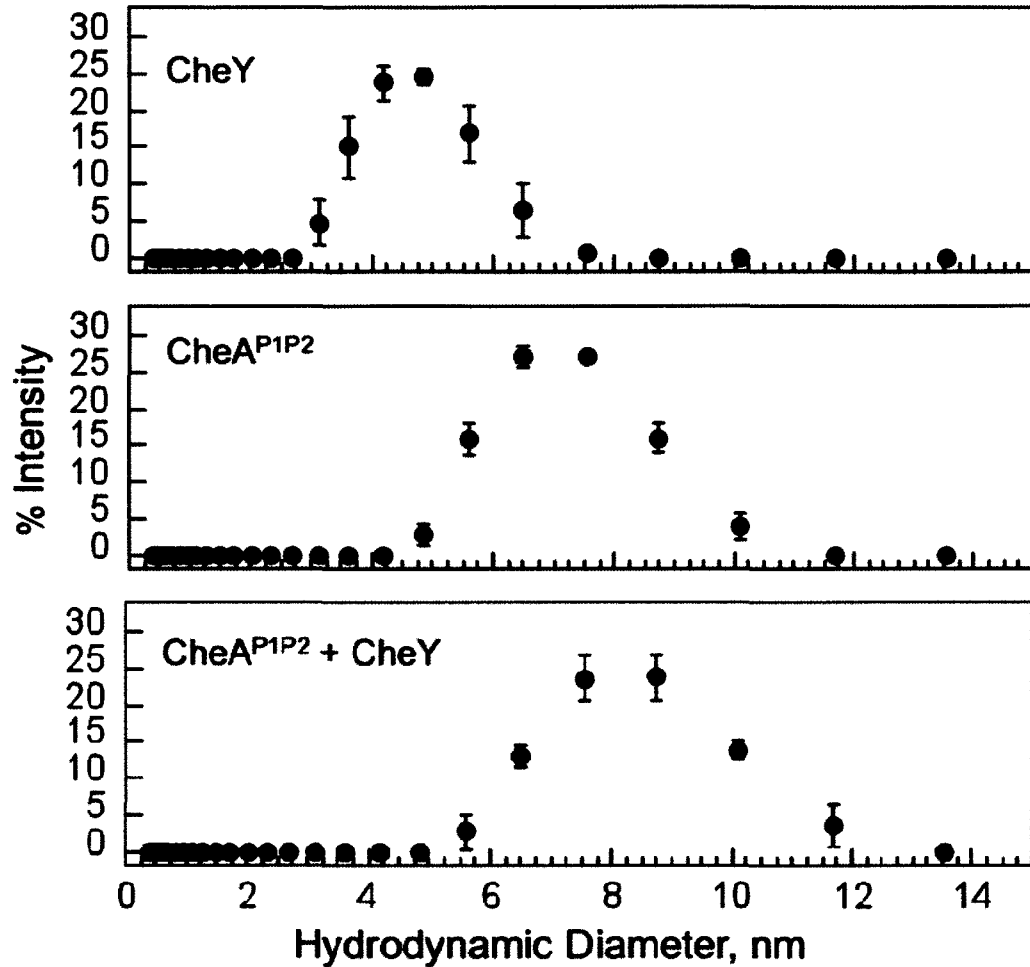
Figure 4.12 Dynamic light scattering analysis of purified chemotaxis proteins. The size distribution of CheA^{P1P2}, CheX, CheY, and a mixture of CheA^{P1P2} and CheY were determined at 78°C after a five minute incubation at that temperature. All proteins were dialyzed into PS buffer prior to analysis, and the data shown represent average values calculated from three separate experiments performed using 115 "M of each protein. The hydrodynamic diameters of CheA^{P1P2} (7.1 ±0.2 nm), CheX (6.3 ±0.2 nm), CheY (4.7 ±0.1 nm), and the mixture of CheA^{P1P2} and CheY (8.3±0.1 nm) were calculated using the Zetasizer software high-resolution analysis model.

4.5 Discussion and Summary.

In this chapter, I provide evidence that AK fragment complementation in $T.$ *thermophilus* is proportional to the thermostability of the protein being fragmented, as previously observed in Chapter 2 for $E.$ *coli*. Whereas AK_{Tn} fragments alone could complement PQN1 at 78°C, fragments with mutations in the cysteines that coordinate a thermostabilizing Zn^{+2} (Glaser et al., 1992; Vieille, Krishnamurthy, Hyung-Hwan, et al., 2003) could not support PQN1 growth at this temperature. However, AK_{Tn} fragments harboring the destabilizing C156A mutation were able to complement PQN1 when they were fused to chemotaxis proteins that homodimerize (CheX) and heterodimerize (CheA^{P1P2} and CheY). This indicates that covalent linkage to associating proteins can drive the reassembly of the AK_{Tn} fragments harboring the C156A mutation at 78°C, extending the temperature range where a split AK can be used as a PCA from mesophile to thermophile growth temperatures. The finding that TnN and TnC fragments cooperatively function at 78°C without the assistance of associating proteins also suggests that these polypeptides may be useful for creating a PCA that functions at even higher temperatures.

My studies examining the effects of $T.$ *maritima* chemotaxis proteins on AK_{Tn}-fragment complementation increases the maximum temperature (78°C) where an interaction has been observed between the kinase CheA and its target CheY . PQN1 growth was complemented by a vector that expresses AK_{Tn} fragments fused to the two-domain fragment CheA^{P1P2} and CheY but not by vectors that express AK_{Tn} fragments fused to only one of these chemotaxis proteins. This is consistent with the findings from dynamic light scattering measurements performed at 78°C, which revealed that a mixture of purified CheA^{P1P2} and CheY exhibits an average hydrodynamic diameter that is greater than either of the individual

proteins. Measurements of CheX-assisted fragment complementation also provides evidence that the CheX phosphatase self-associates at the temperature where it functions in *T. maritima* (Park, Chao, et al., 2004). PQN1 growth was complemented at 78°C by a vector that express CheX-TnN and CheX-TnC^{C156A} to a greater extent than vectors that express only one of the AK$_{Tn}$ fragments fused to CheX. However, this growth was less pronounced than that observed with AK$_{Tn}$ fragment fused to CheA^{P1P2} and CheY. This finding cannot not necessarily be interpreted as arising from a weaker interaction between two CheX molecules, since several issues could influence the magnitude of complementation observed in the assay. Homodimerization of CheX protein fusions is expected to result in only a fraction $(1/3^{rd})$ of the dimers containing the heterodimeric TnN/TnC^{C156A} complex required for AK complementation. In addition, CheX protein fusions may not accumulate to the same extent as CheA^{P1P2} and CheY fusions due to differences in translation and degradation, CheX could associate to a greater extent with *T. thermophilus* protein(s) in a manner that competes with self-association, and CheX protein fusions may be less compatible with my high-temperature PCA due to the proximity of their termini fused to TnN and TnC^{C156A} (Park, Beel, et al., 2004; Park, Chao, et al., 2004).

The high-temperature PCA described herein is expected to have several advantages in discovering interactions among natural and engineered proteins compared with previously described approaches, such as mesophilic two-hybrid (Usui et al., 2005) and PCA screens (Michnick et al., 2007). The AK$_{Tn}$-based PCA is implemented within a *T. thermophilus* HB8 strain growing at a temperature where many thermophiles and hyperthermophiles grow (Rothschild et al., 2001), making it better suited for discovering interactions among proteins that have evolved to fold and function at extreme temperatures (Abd Rahman et al., 1997;

Siddiqui et al., 1998; Goda et al., 2005). In addition, this selection is expected to be more sensitive than existing assays at detecting interactions among proteins whose free energy of binding increases as temperature decreases below the physiological growth range of thermotolerant microbes (Ogasahara et al., 2003; Esue et al., 2005). Unfortunately, the fraction of hyperthermophilic proteins that exhibit this behavior is not known because screens have not yet been performed at temperatures near the boiling point of water. Furthermore, the high-temperature PCA will be capable of assessing protein complex formation in the presence of specialized solutes that accumulate within thermolerant microbes (Empadinhas & da Costa, 2006), which have been proposed to affect the equilibrium binding of protein complexes (Ellis, 2001).

The AK_{Tn}-based PCA should also aid in future protein design efforts working to create artificial protein complexes that are stable at the temperatures where thermophiles and hyperthermophiles grow. This novel PCA can be used as a high-throughput screen to evaluate the association of proteins created through rational design (Huang et al., 2007) and directed evolution (Huang et al., 2008). In addition, this assay can be used to investigate the robustness of protein complexes to different mutational processes (Bloom et al., 2005; Drummond, 2005) at thermophile growth temperatures, and why some hyperthermophilic protein complexes weaken in affinity as temperature is decreased below the physiological growth range of hyperthermophiles (Ogasahara et al., 2003), whereas other do not (Park et al., 2004). The latter can be accomplished by mining libraries of engineered proteins for interactions that occur at 78°C and rescreening discovered variants for decreased association at 40°C, using a structurally-related PCA that was developed for use in *E. coli* at 40°C (Chapter 4 of this work). By comparing the sequences of the protein complexes discovered

with different responses to decreases in temperature, the biophysical origin of these effects can be elucidated, *e.g.,* the contribution of increased charged residues in thermophilic proteins (Suhre and Claverie, 2003a).

Chapter 5

Ongoing Work and Future Research

5.1   Overview of Described Work.

In the work presented herein, I have investigated how split protein systems are influenced by thermostability. Although previous studies have demonstrated that thermostable proteins are more robust to destabilization caused by random mutation and protein recombination, our studies are the first to examine if this phenomena extends to protein fission. My hypothesis was that fragmentation of homologous proteins at structurally-related sites will lead to similar destabilizing effects on protein homologs, and proteins with increased thermostability will *on average* be more likely to retain parent-like structure and function. This was investigated using a combination of *in vitro* biophysical characterization, library functional analysis, and protein engineering. The results presented in Chapters 2-4 of this thesis have established the following novel findings:

i.   *On average*, proteins that are more thermostable will have a greater fraction of functional split protein pairs than a less thermostable protein.

ii. Tuning of protein fragment stability in PCAs also extends to the destabilization of protein motifs (cofactor binding) that influence global protein stability.

iii. Adenylate kinases can be used as a PCA system to monitor protein-protein interactions at mesophilic as well as thermophilic temperatures.

iv. The HyperMuA transposase method can be used to quickly and easily generate split protein libraries in which all nearly possible split sites are sampled.

In Chapter 2, I demonstrate for the first time that split AK fragments can be used as a PCA to detect protein-protein interactions using AK proteins of varying thermostabilities homologously split at a single site. Split *B. subtilis* AK fragments harboring thermostabilizing mutations were shown to complement an AK complementation mutant strain ($CV2_{T7}$) of *E.coli* only when the fragments were fused to polypeptide tags that are known to interact. In contrast, homologous split AK fragments from the hyperthermophile *Thermotoga neapolitana* was able to complement without the assistance of interacting protein fusions. Subsequent biophysical analysis of the purified mesophilic and thermophilic protein fragments demonstrated that equimolar mixtures of AK_{Tn} fragments had greater catalytic activity than mixtures of homologous mesophilic AK_{Bs} fragments, by three orders of magnitude. Secondary structure analysis indicated that the C-terminal AK_{Tn} fragment retained significant structure, whereas all other fragments were largely unfolded. When mixed at equimolar concentrations, the AK_{Bs} fragments displayed no difference compared to the sum of the CD spectra for the individual fragments. However, the AK_{Tn} fragments displayed an increase in secondary structure when mixed together that is ~17% greater than that predicted from summation of the individual protein fragment spectra. These results indicate that the AK_{Tn} fragments are more structured and are able to recombine *in vitro* to recover more secondary structure in a cooperative manner. These results bolster my hypothesis that more thermostable proteins are robust to the destabilizing effects of protein backbone cleavage.

In Chapter 3, I expanded this inquiry to determine if my hypothesis is applicable for all split sites among homologous AK proteins. To do this, I designed and implemented a novel HyperMuA transposase-based method which generates vector libraries coexpressing

protein fragment pairs that are randomly split along the length of a gene. $CV2_{T7}$ complementation analysis of 231 library clones establishes that randomly split AK_{Tn} have a greater number of fraction functional split sites (44%) than randomly split AK_{Bs} (only 6%). Complementation revalidation of each split clone acquired revealed that, on average, the split AK_{Tn} library clones complemented stronger than split AK library derived from the less thermostable AK_{Bs}. In addition, I also analyzed functional complementation at specific sites that are homologous between the two proteins, using engineered constructs. A careful analysis of split sites that occur at homologous positions between the two libraries clearly demonstrates that the hyperthermophilic AK_{Tn} is able to tolerate split sites identified in the mesophilic AK_{Bs}, but the converse is not true. Analysis of the where the split sites occur for the AK_{Bs} and AK_{Tn} proteins reveals that split sites for both othologs occur primarily at solvent exposed regions and are absent at regions that are proximal to the active site. However, the AK_{Tn} is able to tolerate fragmentation sites throughout the CORE domain and in the LID domain, which has been shown to play a role in stabilizing AK proteins (Perrier et al., 1994; Perrier, Burlacu-Miron, Bourgeois, et al., 1998). This is likely due to a greater fraction of AK_{Tn} fragment pairs retaining structure in contrast to the AK_{Bs}, although the exact biophysical underpinnings of this difference remain unknown. The conclusions from the library analysis experiments establish that overall, more thermostable AK orthologs should have a greater fraction functional of split protein pairs. These results further support the hypothesis that thermostable proteins which have molecular adaptations to a highly destabilizing environment are better able than less stable proteins to functionally survive deleterious structural events. This has been demonstrated previously for random mutation

and recombination, in this work we have presented evidence that this phenomena extends to protein fission as well.

To see if the split AK_{Tn} could function as a PCA at higher temperatures where the stabilities of the fragments are reduced, in Chapter 4 I described the construction and validation of a high-temperature PCA (ht-PCA). Due to the lack of a hyperthermophilic AK complementation strain, I first generated such a strain using *Thermus thermophilus* HB8, by genomic recombination. This novel mutant strain, named PQN1, is unable to grow at temperatures exceeding 75°C unless transformed with a plasmid expressing a functional AK. Initial tests indicated that coexpressing split AK_{Tn} fragments are able to complement PQN1 cells without the need of interacting protein fusions. I postulated that similar to thermostability, the stabilization of proteins by cofactor binding could also be used to tune the complementation strength of a PCA system. Subsequently, I mutated two residues of the AK_{Tn} tetra-cysteine motif that is known to coordinate a zinc ion. This motif occurs in the LID domain of AKs and has been shown in the past to contribute significantly to the stability of the protein (Perrier et al., 1994; Perrier, Burlacu-Miron, Bourgeois, et al., 1998). Mutation of these residues abolished the ability of the AK_{Tn} split fragments to complement PQN1 cells. However, when these mutant AK_{Tn} fragments are fused to hyperthermophilic chemotaxis proteins that are predicted to interact, the PQN1 complementation was restored. In addition, the predicted interaction of these hyperthermophilic chemotaxis proteins at the assay temperature of 78°C was verified by dynamic light scattering analysis of the purified proteins. These results demonstrate first that AKs can also be created for use at thermophilic temperatures, and second that structural motifs that promote protein stability can be targeted to tune the complementation strength of a PCA system. The ht-PCA constructed herein is the

first assay capable of detecting protein-protein interactions in a cell at thermophilic temperatures. In addition, the recombinant *T. thermophilus* complementation mutant PQN1 I have engineered is the first platform that allows the *in vitro* analysis of AK function at thermophilic temperatures.

5.2   Future Directions: Optimization of the ht-PCA.

The ht-PCA outlined in Chapter 4 would be of great utility in creating genome-wide maps of thermophilic protein-protein interactions, studying the effects of temperature on biomolecular interactions, and engineering oligomeric thermostable nanomaterials. To optimize the prototype ht-PCA system I have designed, a number of approaches can be taken to maximize the complementation activity and minimize false positive and/or false negative results. As with any assay platform, optimization and characterization of the ht-PCA is necessary before the full utility of the assay can be realized.

To precisely determine the dynamic range of our ht-PCA system, a set of interacting proteins with known association strengths should be tested as fusion tags. This will allow the determination of the threshold at which weakly interacting proteins are able to drive the complementation activity of the ht-PCA in PQN1 cells. In addition, this will also provide further validation of the ht-PCA for detecting protein-protein interactions. Controls to assess the level of background will be coexpression of these fusion proteins in pairs that are not predicted to interact.

Although the ht-PCA developed here was based upon the split site occurring between the D76 and C77 residues[16], there are also number of split-AK$_{Tn}$ fragment pairs that have been acquired in the library complementation study presented in Chapter 3. Many of these

---

[16] AK$_{Bs}$ numbering convention.

split sites may yield ht-PCAs with superior signal-to-noise, or they could be used to generate multiple ht-PCA sets that have different dynamic ranges. The split AK_{Tn} library has only been tested in the mesophilic $CV2_{T7}$ complementation strain at 40°C. I predict that testing in the PQN1 complementation strain at the higher temperature of 78°C will destabilize many of these split AK_{Tn} fragments and identify a number of pairs that are only able to complement when fused to interacting proteins. In addition, the performance of these alternative ht-PCAs can be tuned using the same strategy I have demonstrated in Chapter 4 – mutating residues in the zinc binding motif that contributes to AK stability.

5.3   Future Directions: Mapping of the *T. maritima* Proteome.

With a functioning ht-PCA, the PINs of various thermophilic and hyperthermophilic proteins can be assessed. To date, there has been no PIN acquired from hyper/thermophilic proteins in which the assay was performed at the high temperatures where these proteins interact. A high-confidence PIN for a thermophilic organism would allow, for the first time, identification of cryptic thermophilic proteins that may be of industrial importance, a deeper understanding of the global physiology of these unique organisms, and a systems-biology level analysis of extremophilic PIN structure.

Dr. Scott Lesley of the Joint Center for Structural Genomics (JCSG) at the Genomics Institute of the Novartis Research Foundation has spearheaded a project to obtain atomic-resolution structures of the *Thermotoga maritima* MSB8 proteome using high-throughput protein crystallography and structure determination (Lesley et al., 2002). Owning to their rigidity at lower temperatures, hyper/thermophilic proteins make excellent candidates for

protein crystallization and crystallography (Bae et al., 2004). The *T. maritima* ORF library

acquired from the JCSG can be easily cloned into our ht-PCA platform to acquire a genomic

ORF x genomic ORF ht-PCA library. High-throughput selection in the PQN1 strain will

yield identification of physiologically-relevant protein-protein interactions which can then be

scrutinized using methods which are lower-throughput but higher-resolution.

5.4   Future Directions: Investigating Emergent Properties of Networks.

Once a system with interacting elements is defined, one can apply the mathematics of

graph theory[17] to probe the network itself. Networks can be classified based upon the unique

properties of their topology (Barabási and Oltvai, 2004; Ma'ayan, 2009). As with other

biological networks, PINs are considered to be 'scale-free' and 'small-world' networks

(Goldberg and Roth, 2003; Nacher and Akutsu, 2007). In networks considered to be scale-

free, the degree of connections for all network nodes follows a power law[18] distribution.

Small-world structures are those in which the overall node-to-node distance is short and that

the nodes are highly clustered[19] (Watts and Strogatz, 1998). Such architectures seem to occur

in many naturally evolving networks, such as social networks, power grids, neural networks,

the internet, and ecosystem food chains. These networks occupy a space between highly

ordered networks and randomly connected networks. This unique network structure is

thought to arise from a 'preferential attachment' phenomena that in biological networks may

occur through gene duplication (Bhan et al., 2002; Pastor-Satorras et al., 2003). Clustering

analysis of such PINs also allows for the identification of modules (also called network

---

[17] Graph theory is the field of mathematics which provides an analytical framework for modeling pairwise interactions
between objects, created by Leonhard Euler in his 1735 proof titled 'Seven Bridges of Königsberg'.
[18] Exhibiting the properties of scale invariance.
[19] That is, containing highly interconnected 'hubs'.

motifs or subgraphs) that organize the proteome along functional, spatial, or temporal dimensions (Milo et al., 2002). In addition, the adaptive nature of living organisms has been tied to the property of 'robustness' that is inherent in this topology, whereby the PINs are able to maintain functional and dynamical robustness in the face of significant perturbations (Barkai and Leibler, 1997; Albert et al., 2000; Krylov et al., 2003). The continuing systems-level study of how emergent properties arise from network structures will shed light on the importance of these properties for defining the essential dynamics of life, define the functions of unknown proteins by the functional network motifs they belong to, and establish evolutionary relationships based upon comparisons of the network structures.

5.5   Conclusions and Outlook.

The results of the work presented in this dissertation can be broadly divided into original findings and novel technology platforms developed. I have established that protein thermostability is an important factor in designing PCA systems. I have established through *in vivo* functional complementation and *in vitro* biophysical experiments that thermostable proteins are more robust to disruption caused by backbone cleavage. These results are similar to that previously established for random mutation and recombination, indicating that the resistance of thermostable proteins to destabilizing influences is a wide-ranging phenomena. Analysis of the split protein libraries presented in Chapter 4 also support my hypothesis that the positive correlation between thermostability and split protein complementation can be generalized for other proteins. I have also demonstrated during the construction of the ht-PCA that protein stability induced by cofactor binding can also be targeted to fine tune the sensitivity of a PCA. Previous to the work herein, the only strategies for designing PCA

systems were by screening large random fragmentation libraries acquired from a long multi-enzymatic process (Besenmatter et al., 2007; Bloom et al., 2005; Bloom, Labthavikul, et al., 2006; Meyer et al., 2006). The results from my experiments point to the value of incorporating rational design into such libraries. The inclusion of parent proteins of varying thermostabilities should greatly expand the diversity of the fragmentation libraries and result in many unique solutions. Once unique protein fragment pairs are found, they can be tuned for complementation strength by either increasing stabilizing interactions in the fragments or destabilizing pre-existing protein motifs, as was demonstrated in the zinc-binding motif of AK_{Tn} in Chapter 4.

I have also developed two new technology platforms that will be of utility to the protein biology community. The ht-PCA described in Chapter 4 is the first assay capable of detecting hyper/thermophilic proteins *in vivo* and at the physiological temperatures at which they have evolved to function. This assay will allow the generation of high confidence protein-protein interaction datasets for the mapping of PINs of high temperature extremophiles. These systems level studies of thermophile protein networks will greatly accelerate the identification of thermostable enzymes that of great interest in various industries, as they exhibit longer shelf-life and greater resistance to organic solvents than their mesophilic counterparts (Coolbear et al., 1992; Lasa and Berenguer, 1993; Haki and Rakshit, 2003). Although a large number of hyper/thermophilic ORFs are unidentified, it would be possible to assign likely functions based on contextual analysis of what local protein networks these unknown proteins associate with. This 'guilt-by-association' approach would require high-confidence PIN maps of thermophiles, now a possibility with the ht-PCA I have developed. In addition, in the course of generating the randomly split libraries for

AK$_{Bs}$ and AK$_{Tn}$ in Chapter 3, I have conceived, designed, and validated a novel method for creating plasmid libraries that have a gene of interest split at random positions. This process is mediated by a MuA transposase variant (HyperMuA), and the final library consists of a single randomly inserted piece of DNA containing regulatory elements, allowing for the coexpression of the protein fragments from a single vector. In contrast to the current methods for random split protein library generation, which uses a suite of enzymes to obtain random insertions, the transposon-based method described here is a single enzymatic reaction, and will result in higher-efficiency coverage of all possible split sites.

The biophysics of split-protein systems is a poorly studied subject. In addition, although the genomes of many hyper/thermophiles have been determined, many aspects of their molecular biology and physiology are still mysteries. The work I have presented here are the first to shed light on the link between protein thermostability and the complementation ability of protein fragements. I hope that in the future, more biophysical experiments will be performed on split protein systems to allow a more complete understanding of this phenomenon. In the process of this pursuing this line of inquiry, I have also developed a new and facile method to generate randomly split protein libraries, which should greatly accelerate future attempts to engineer PCA systems for various applications. In addition, I have also constructed the first high-temperature protein-protein interaction assay platform that will allow the mapping of thermophilic PINs at their physiological temperatures, a feat that has yet to be accomplished but is now within reach.

References

Abd Rahman, R. N., Fujiwara, S., Takagi, M., Kanaya, S., and Imanaka, T. (1997). Effect of heat treatment on proper oligomeric structure formation of thermostable glutamate dehydrogenase from a hyperthermophilic archaeon. Biochem. Biophys. Res. Commun. *241*, 646-52.

Abramoff, M., Magelhaes, P., and Ram, S. (2004). Image Processing with ImageJ. Biophotonics Intl. *11*, 36-42.

Agrafioti, I., Swire, J., Abbott, J., Huntley, D., Butcher, S., and Stumpf, M. P. (2005). Comparative analysis of the *Saccharomyces cerevisiae* and *Caenorhabditis elegans* protein interaction networks. BMC Evol. Biol. *5*, 23.

Albers, Sonja-V. Albers, Vossenberg, J., Driessen, A., and Konings, W. (2001). Bioenergetics and solute uptake under extreme conditions. Extremophiles *5*, 285-294.

Albert, Jeong, and Barabasi (2000). Error and attack tolerance of complex networks. Nature *406*, 378-382.

Albert, R., and Wang, R. (2009). Discrete dynamic modeling of cellular signaling networks. Meth. Enzymol. *467*, 281-306.

Armengaud, J. (2009). A perfect genome annotation is within reach with the proteomics and genomics alliance. Curr. Opin. Microbiol. *12*, 292-300.

Arnold, F. H. (1998). Design by Directed Evolution. Acc. Chem. Res. *31*, 125-131.

Atkinson, D. E. (1969). Regulation of enzyme function. Ann. Rev. Microbiol. *23*, 47–68.

Averhoff, B. (2009). Shuffling genes around in hot environments: the unique DNA transporter of Thermus thermophilus. FEMS Microbiol. Rev. *33*, 611-626.

Bae, E. (2004). Structures and Analysis of Highly Homologous Psychrophilic, Mesophilic, and Thermophilic Adenylate Kinases. J. Biol. Chem. *279*, 28202-28208.

Bae, E., and Phillips, G. N. (2006). Roles of static and dynamic domains in stability and catalysis of adenylate kinase. Proc. Natl. Acad. Sci. U.S.A. *103*, 2132-7.

Bae, E., and Phillips, G. N. (2004). Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. J. Biol. Chem. *279*, 28202-8.

Baker-Austin, C., and Dopson, M. (2007). Life in acid: pH homeostasis in acidophiles. Trends Microbiol. *15*, 165-171.

Baldasseroni, F., and Pascarella, S. (2009). Subunit interfaces of oligomeric hyperthermophilic enzymes display enhanced compactness. Int. J. Biol. Macromol. *44*, 353-360.

Barabási, A., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101-113.

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. Nature *387*, 913-917.

Batra, P. P., Takeda, K., and Kreye, W. (1989). Studies on adenylate kinase (ATP:AMP phosphotransferase) of Mycobacterium marinum (ATCC 927). Acta Leprol. *7 Suppl 1*, 25-29.

Besenmatter, W., Kast, P., and Hilvert, D. (2007). Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. Proteins *66*, 500-6.

Bhan, A., Galas, D. J., and Dewey, T. G. (2002). A duplication growth model of gene expression networks. Bioinformatics *18*, 1486-1493.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453-1462.

Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C., and Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. U.S.A. *102*, 606–611.

Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. U.S.A. *103*, 5869-74.

Bloom, J. D., Raval, A., and Wilke, C. O. (2006). Thermodynamics of Neutral Protein Evolution. Genetics *175*, 255-266.

Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C., and Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. U.S.A.*102*, 606-11.

Bloom, J., Lu, Z., Chen, D., Raval, A., Venturelli, O., and Arnold, F. (2007). Evolution favors protein mutational robustness in sufficiently large populations. BMC Biology *5*, 29.

Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. U.S.A *103*, 5869-5874.

Borges, N., Ramos, A., Raven, N. D., Sharp, R. J., and Santos, H. (2002). Comparative study

of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes. Extremophiles *6*, 209-16.

Bouthier de la Tour, C., Portemer, C., Nadal, M., Stetter, K. O., Forterre, P., and Duguet, M. (1990). Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. J. Bacteriol. *172*, 6803-6808.

Brock, T. D., and Freeze, H. (1969). Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. J. Bacteriol. *98*, 289-297.

Brock, T. D., and Boylen, K. L. (1973). Presence of Thermophilic Bacteria in Laundry and Domestic Hot-Water Heaters. Appl. Microbiol. *25*, 72-76.

Brouns, S. J., Wu, H., Akerboom, J., Turnbull, A. P., de Vos, W. M., and van der Oost, J. (2005). Engineering a selectable marker for hyperthermophiles. J. Biol. Chem. *280*, 11422-31.

Brune, M., Schumann, R., and Wittinghofer, F. (1985b). Cloning and sequencing of the adenylate kinase gene (adk) of Escherichia coli. Nucleic Acids Res. *13*, 7139-51.

Burlacu-Miron, S., Perrier, V., Gilles, A. M., Pistotnik, E., and Craescu, C. T. (1998). Structural and energetic factors of the increased thermal stability in a genetically engineered Escherichia coli adenylate kinase. J. Biol. Chem. *273*, 19102–19107.

Butcher, E. C., Berg, E. L., and Kunkel, E. J. (2004). Systems biology in drug discovery. Nat. Biotech. *22*, 1253-1259.

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature *433*, 531-537.

Butterfield, Y. S. N., Marra, M. A., Asano, J. K., Chan, S. Y., Guin, R., Krzywinski, M. I., Lee, S. S., MacDonald, K. W. K., Mathewson, C. A., Olson, T. E., et al. (2002). An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. Nucl. Acids Res. *30*, 2460-2468.

Cabantous, S., Terwilliger, T. C., and Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. *23*, 102-7.

Cacciapuoti, G., Porcelli, M., Bertoldo, C., De Rosa, M., and Zappia, V. (1994). Purification and characterization of extremely thermophilic and thermostable 5'-methylthioadenosine phosphorylase from the archaeon Sulfolobus solfataricus. Purine nucleoside phosphorylase activity and evidence for intersubunit disulfide bonds. J. Biol. Chem. *269*, 24762-24769.

Cambillau, C., and Claverie, J. (2000). Structural and Genomic Correlates of Hyperthermostability. J. Biol. Chem. *275*, 32383-32386.

Cameron, D. M., Gregory, S. T., Thompson, J., Suh, M., Limbach, P. A., and Dahlberg, A. E. (2004). Thermus thermophilus L11 Methyltransferase, PrmA, Is Dispensable for Growth and Preferentially Modifies Free Ribosomal Protein L11 Prior to Ribosome Assembly. J. Bacteriol. *186*, 5819-5825.

Cavicchioli, R. (2006). Cold-adapted archaea. Nat. Rev. Microbiol. *4*, 331-343.

Cesareni, G., Ceol, A., Gavrila, C., Palazzi, L. M., Persico, M., and Schneider, M. V. (2005). Comparative interactomics. FEBS Lett. *579*, 1828-1833.

Chakravarty, S., and Varadarajan, R. (2002). Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. Biochemistry *41*, 8152-8161.

Charlier, D., and Droogmans, L. (2005). Microbial Life at high temperature, the challenges, the strategies. Cell. Molec. Life Sci. *62*, 2974-2984.

Chen, L., Brügger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H., et al. (2005). The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. J. Bacteriol. *187*, 4992-4999.

Chien, C. T., Bartel, P. L., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc. Natl. Acad. Sci. U.S.A *88*, 9578-9582.

Coolbear, T., Daniel, R. M., and Morgan, H. W. (1992). The enzymes from extreme thermophiles: bacterial sources, thermostabilities and industrial relevance. Adv. Biochem. Eng. Biotechnol. *45*, 57-98.

Counago, R., Chen, S., and Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. Mol. Cell *22*, 441-9.

Counago, R., and Shamoo, Y. (2005). Gene replacement of adenylate kinase in the gram-positive thermophile Geobacillus stearothermophilus disrupts adenine nucleotide homeostasis and reduces cell viability. Extremophiles *9*, 135-44.

Crameri, A., Raillard, S., Bermudez, E., and Stemmer, W. P. C. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature *391*, 288-291.

Criswell, A. R., Bae, E., Stec, B., Konisky, J., and Phillips, G. N. (2003). Structures of thermophilic and mesophilic adenylate kinases from the genus Methanococcus. J. Mol. Biol. *330*, 1087-1099.

Cronan, J. E., and Godson, G. N. (1972). Mutants of Escherichia coli with temperature-sensitive lesions in membrane phospholipid synthesis: genetic analysis of glycerol-3-phosphate acyltransferase mutants. Mol. Gen. Genet. *116*, 199-210.

D'Amico, S., Collins, T., Marx, J., Feller, G., and Gerday, C. (2006). Psychrophilic microorganisms: challenges for life. EMBO reports *7*, 385-389.

Daugherty, P. S., Chen, G., Iverson, B. L., and Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. Proc. Natl. Acad. Sci. U.S.A. *97*, 2029-2034.

Davlieva, M., and Shamoo, Y. (2009). Structure and biochemical characterization of an adenylate kinase originating from the psychrophilic organism Marinibacillus marinus. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun *65*, 751-756.

Deane, C. M., Salwinski, L., Xenarios, I., and Eisenberg, D. (2002). Protein Interactions: Two Methods for Assessment of the Reliability of High Throughput Observations. Mol. Cell. Proteomics *1*, 349-356.

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., et al. (1998). The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. Nature *392*, 353-358.

Delorenzo, V. (2008). Systems biology approaches to bioremediation. Curr. Opin. Biotechnol. *19*, 579-589.

Doi, N., and Yanagawa, H. (1999). Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. FEBS Lett. *453*, 305–307.

Dominy, B. N., Minoux, H., and Brooks, C. L. (2004). An electrostatic basis for the stability of thermophilic proteins. Proteins *57*, 128-141.

Drummond, D. A. (2005). On the conservative nature of intragenic recombination. Proc. Natl. Acad. Sci. U.S.A. *102*, 5380-5385.

Dutta, S., Batori, V., Koide, A., and Koide, S. (2005). High-affinity fragment complementation of a fibronectin type III domain and its application to stability enhancement. Protein Sci. *14*, 2838-48.

Eggers, D. K., and Valentine, J. S. (2001). Crowding and hydration effects on protein conformation: a study with sol-gel encapsulated proteins. J. Mol. Biol. *314*, 911-922.

Egorova, K., and Antranikian, G. (2005). Industrial relevance of thermophilic Archaea. Curr. Opin. Microbiol. *8*, 649-655.

Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. *26*, 597-604.

Empadinhas, N., and da Costa, M. S. (2006). Diversity and biosynthesis of compatible solutes in hyper/thermophiles. Int. Microbiol *9*, 199-206.

Empadinhas, N., and da Costa, M. S. (2008). Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. Int. Microbiol. *11*, 151-161.

Esue, O., Cordero, M., Wirtz, D., and Tseng, Y. (2005). The Assembly of MreB, a Prokaryotic Homolog of Actin. Journal of Biological Chemistry *280*, 2628-2635.

Ethier, M., Lambert, J., Vasilescu, J., and Figeys, D. (2006). Analysis of protein interaction networks using mass spectrometry compatible techniques. Anal. Chim. Acta *564*, 10-18.

Fang, X., Cui, Q., Tong, Y., Feng, Y., Shan, L., Huang, L., and Wang, J. (2008). A stabilizing alpha/beta-hydrophobic core greatly contributes to hyperthermostability of archaeal [P62A]Ssh10b. Biochemistry *47*, 11212-11221.

Faraldo, M. M., de Pedro, M. A., and Berenguer, J. (1992). Sequence of the S-layer gene of Thermus thermophilus HB8 and functionality of its promoter in *Escherichia coli*. J. Bacteriol. *174*, 7458-62.

Ferber, D. M., Haney, P. J., Berk, H., Lynn, D., and Konisky, J. (1997). The adenylate kinase genes of M. voltae, M. thermolithotrophicus, M. jannaschii, and M. igneus define a new family of adenylate kinases. Gene *185*, 239-244.

Fiebitz, A., Nyarsik, L., Haendler, B., Hu, Y., Wagner, F., Thamm, S., Lehrach, H., Janitz, M., and Vanhecke, D. (2008). High-throughput mammalian two-hybrid screening for protein-protein interactions using transfected cell arrays. BMC Genomics *9*, 68.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Finley, R. L., and Brent, R. (1994). Interaction mating reveals binary and ternary connections between Drosophila cell cycle regulators. Proc. Natl. Acad. Sci. U.S.A *91*, 12980-12984.

Fu, P., and Panke, S. (2009). Systems Biology and Synthetic Biology (Wiley-Interscience).

Fukami-Kobayashi, K., Nosaka, M., Nakazawa, A., and Go, M. (1996). Ancient divergence of long and short isoforms of adenylate kinase: molecular evolution of the nucleoside monophosphate kinase family. FEBS Lett. *385*, 214-220.

Galarneau, A., Primeau, M., Trudeau, L., and Michnick, S. W. (2002). Beta-lactamase

protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. Nat. Biotechnol. *20*, 619-622.

Galtier, N., and Lobry, J. R. (1997). Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. *44*, 632-636.

Gandhi, T. K. B., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K. N., Mohan, S. S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B., et al. (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat. Genet. *38*, 285-293.

Gardy, J. L., Lynn, D. J., Brinkman, F. S., and Hancock, R. E. (2009). Enabling a systems biology approach to immunology: focus on innate immunity. Trends in Immun.*30*, 249-262.

Ge, M., Xia, X., and Pan, X. (2008). Salt bridges in the hyperthermophilic protein Ssh10b are resilient to temperature increases. J. Biol. Chem. *283*, 31690-31696.

Gegg, C. V., Bowers, K. E., and Matthews, C. R. (1997). Probing minimal independent folding units in dihydrofolate reductase by molecular dissection. Protein Sci. *6*, 1885-92.

Gershenson, A., Schauerte, J. A., Giver, L., and Arnold, F. H. (2000). Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. Biochemistry *39*, 4658-4665.

Giesecke, A. V., Fang, R., and Joung, J. K. (2006). Synthetic protein-protein interaction domains created by shuffling Cys2His2 zinc-fingers. Mol. Syst. Biol. *2*, 2006 2011.

Gigliobianco, T., Lakaye, B., Makarchikov, A. F., Wins, P., and Bettendorff, L. (2008). Adenylate kinase-independent thiamine triphosphate accumulation under severe energy stress in Escherichia coli. BMC Microbiol. *8*, 16.

Gilles, A. M., Glaser, P., Perrier, V., Meier, A., Longin, R., Sebald, M., Maignan, L., Pistotnik, E., and Barzu, O. (1994). Zinc, a structural component of adenylate kinases from gram-positive bacteria. J. Bacteriol. *176*, 520-523.

Gilles, A. M., Saint-Girons, I., Monnot, M., Fermandjian, S., Michelson, S., and Bârzu, O. (1986). Substitution of a serine residue for proline-87 reduces catalytic activity and increases susceptibility to proteolysis of Escherichia coli adenylate kinase. Proc. Natl. Acad. Sci. U.S.A *83*, 5798-5802.

Gilles, A. M., Sismeiro, O., Munier, H., Fabian, H., Mantsch, H. H., Surewicz, W. K., Craescu, C. C., Barzu, O., and Danchin, A. (1993). Structural and physico-chemical characteristics of Bordetella pertussis adenylate kinase, a tryptophan-containing

enzyme. Eur. J. Biochem. *218*, 921-927.

Giver, L., Gershenson, A., Freskgard, P., and Arnold, F. H. (1998). Directed evolution of a thermostable esterase. Proc. Natl. Acad. Sci. U.S.A. *95*, 12809-12813.

Glaser, M., Nulty, W., and Vagelos, P. R. (1975). Role of adenylate kinase in the regulation of macromolecular biosynthesis in a putative mutant of Escherichia coli defective in membrane phospholipid biosynthesis. J. Bacteriol. *123*, 128-136.

Glaser, P., Presecan, E., Delepierre, M., Surewicz, W. K., Mantsch, H. H., Barzu, O., and Gilles, A. M. (1992). Zinc, a novel structural element found in the family of bacterial adenylate kinases. Biochemistry *31*, 3038-3043.

Gliozzi, A., Paoli, G., DeRosa, M., and Gambacorta, A. (1983). Effect of isoprenoid cyclization on the transition temperature of lipids in thermophilic archaeabacteria. Biochim. Biophys. Acta. *735*, 234-242.

Goda, S., Kojima, M., Nishikawa, Y., Kujo, C., Kawakami, R., Kuramitsu, S., Sakuraba, H., Hiragi, Y., and Ohshima, T. (2005). Intersubunit interaction induced by subunit rearrangement is essential for the catalytic activity of the hyperthermophilic glutamate dehydrogenase from Pyrobaculum islandicum. Biochemistry *44*, 15304-13.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., et al. (1996). Life with 6000 genes. Science *274*, 546, 563-567.

Goldberg, D. S., and Roth, F. P. (2003). Assessing experimentally derived interactions in a small world. Proc. Natl. Acad. Sci. U.S.A *100*, 4372-4376.

González, J. M., Masuchi, Y., Robb, F. T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaoka, J., and Kato, C. (1998). Pyrococcus horikoshii sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. Extremophiles *2*, 123-130.

de Grado, M., Castan, P., and Berenguer, J. (1999). A high-transformation-efficiency cloning vector for Thermus thermophilus. Plasmid *42*, 241-5.

Gregory, S. G., Barlow, K. F., McLay, K. E., Kaul, R., Swarbreck, D., Dunham, A., Scott, C. E., Howe, K. L., Woodfine, K., Spencer, C. C. A., et al. (2006). The DNA sequence and biological annotation of human chromosome 1. Nature *441*, 315-321.

de Groot, A., Dulermo, R., Ortet, P., Blanchard, L., Guérin, P., Fernandez, B., Vacherie, B., Dossat, C., Jolivet, E., Siguier, P., et al. (2009). Alliance of Proteomics and Genomics to Unravel the Specificities of Sahara Bacterium Deinococcus deserti. PLoS Genet. *5*, e1000434.

Haase, G. H., Brune, M., Reinstein, J., Pai, E. F., Pingoud, A., and Wittinghofer, A. (1989). Adenylate kinases from thermosensitive Escherichia coli strains. J. Mol. Biol. *207*, 151-162.

Haki, G. D., and Rakshit, S. K. (2003). Developments in industrially important thermostable enzymes: a review. Bioresour. Technol. *89*, 17-34.

Haney, P. J., Stees, M., and Konisky, J. (1999). Analysis of thermal stabilizing interactions in mesophilic and thermophilic adenylate kinases from the genus Methanococcus. J. Biol. Chem. *274*, 28453–28458.

Henne, A., Brüggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., Johann, A., Lienard, T., Gohl, O., Martinez-Arias, R., et al. (2004a). The genome sequence of the extreme thermophile Thermus thermophilus. Nat. Biotechnol. *22*, 547-553.

Hensel, R., and König, H. (1988). Thermoadaptation of methanogenic bacteria by intracellular ion concentration. FEMS Microbiol. Lett. *49*, 75-79.

Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001). Dissection of a (betaalpha)8-barrel enzyme into two folded halves. Nat. Struct. Biol. *8*, 32-6.

Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S., and Kagamiyama, H. (1999). Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for Thermus thermophilus. J. Biochem. (Tokyo) *126*, 951-6.

Huang, J., Koide, A., Makabe, K., and Koide, S. (2008). Design of protein function leaps by directed domain interface evolution. Proc. Natl. Acad. Sci. U.S.A. *105*, 6578-6583.

Huang, P. S., Love, J. J., and Mayo, S. L. (2007). A de novo designed protein protein interface. Protein Sci. *16*, 2770-4.

Huber, R., Huber, H., and Stetter, K. O. (2000). Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. FEMS Microbiol. Rev *24*, 615-623.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. U.S.A. *98*, 4569-4574.

Jaenicke, R., and Böhm, G. (1998). The stability of proteins in extreme environments. Curr. Opin. Struct. Biol. *8*, 738-748.

Jeanthon, C. (2000). Molecular ecology of hydrothermal vent microbial communities. Antonie Van Leeuwenhoek *77*, 117-133.

Johnsson, N., and Varshavsky, A. (1994). Split ubiquitin as a sensor of protein interactions in vivo. Proc. Natl. Acad. Sci. U.S.A. *91*, 10340-4.

Kannan, N., and Vishveshwara, S. (2000). Aromatic clusters: a determinant of thermal stability of thermophilic proteins. Protein Eng. *13*, 753-761.

Kashefi, K., and Lovley, D. R. (2003). Extending the upper temperature limit for life. Science *301*, 934.

Kato, I., and Anfinsen, C. B. (1969). On the stabilization of ribonuclease S-protein by ribonuclease S-peptide. J. Biol. Chem. *244*, 1004-7.

Kawai, M., Kidou, S., Kato, A., and Uchimiya, H. (1992). Molecular characterization of cDNA encoding for adenylate kinase of rice (Oryza sativa L.). Plant J. *2*, 845-854.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., et al. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain7. DNA Res. *8*, 123-140.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., et al. (1998a). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3. DNA Res. *5*, 55-76.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., et al. (1998b). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3. DNA Res. *5*, 55-76.

Kelly, R. M., and Adams, M. W. (1994). Metabolism in hyperthermophilic microorganisms. Antonie Van Leeuwenhoek *66*, 247-270.

Kikuchi, A., and Asai, K. (1984). Reverse gyrase--a topoisomerase which introduces positive superhelical turns into DNA. Nature *309*, 677-81.

Kim, J., Bates, D. G., Postlethwaite, I., Heslop-Harrison, P., and Cho, K. (2008). Linear time-varying models can reveal non-linear interactions of biomolecular regulatory networks using multiple time-series data. Bioinformatics *24*, 1286-1292.

Kishi, F., Tanizawa, Y., and Nakazawa, A. (1987). Isolation and characterization of two types of cDNA for mitochondrial adenylate kinase and their expression in Escherichia coli. J. Biol. Chem. *262*, 11785-11789.

Kitano, H. (2002). Systems Biology: A Brief Overview. Science *295*, 1662-1664.

Kladova, A. V., Gavel, O. Y., Mukhopaadhyay, A., et al. (2009). Cobalt-, zinc- and iron-bound forms of adenylate kinase (AK) from the sulfate-reducing bacterium Desulfovibrio gigas: purification, crystallization and preliminary X-ray diffraction analysis. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 65, 926-929.

Kladova, A. V., Gavel, O. Y., Zhadan, G. G., Roig, M. G., Shnyrov, V. L., and Bursakov, S. A. (2009). Zinc-, cobalt- and iron-chelated forms of adenylate kinase from the Gram-negative bacterium Desulfovibrio gigas. Int. J. Biol. Macromol 45, 524-531.

Kohen, A., Cannio, R., Bartolucci, S., and Klinman, J. P. (1999). Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. Nature 399, 496-499.

Kojoh, K., Matsuzawa, H., and Wakagi, T. (1999). Zinc and an N-terminal extra stretch of the ferredoxin from a thermoacidophilic archaeon stabilize the molecule at high temperature. Eur. J. Biochem. 264, 85-91.

Konrad, M. (1988). Analysis and in vivo disruption of the gene coding for adenylate kinase (ADK1) in the yeast Saccharomyces cerevisiae. J. Biol. Chem. 263, 19468-19474.

Koskinen, P. E., Beck, S. R., Örlygsson, J., and Puhakka, J. A. (2008). Ethanol and hydrogen production by two thermophilic, anaerobic bacteria isolated from Icelandic geothermal areas. Biotech. Bioeng. 101, 679-690.

Krishnamurthy, H., Munro, K., Yan, H., and Vieille, C. (2009a). Dynamics in Thermotoga neapolitana Adenylate Kinase: 15N Relaxation and Hydrogen−Deuterium Exchange Studies of a Hyperthermophilic Enzyme Highly Active at 30 °C†. Biochemistry 48, 2723-2739.

Krylov, D. M., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2003). Gene Loss, Protein Sequence Divergence, Gene Dispensability, Expression Level, and Interactivity Are Correlated in Eukaryotic Evolution. Genome Res. 13, 2229-2235.

Kujo, C., and Ohshima, T. (1998). Enzymological characteristics of the hyperthermostable NAD-dependent glutamate dehydrogenase from the archaeon Pyrobaculum islandicum and effects of denaturants and organic solvents. Appl. Environ. Microbiol. 64, 2152-2157.

Kumar, S., Sham, Y. Y., Tsai, C. J., and Nussinov, R. (2001). Protein folding and function: the N-terminal fragment in adenylate kinase. Biophysical J. 80, 2439–2454.

Kumar, S., Tsai, C., and Nussinov, R. (2000). Factors enhancing protein thermostability. Protein Eng. 13, 179-191.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 390, 249-256.

Lasa, I., and Berenguer, J. (1993). Thermophilic enzymes and their biotechnological potential. Microbiologia *9*, 77-89.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., and Davis, R. W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. U.S.A *94*, 13057-13062.

Legrain, P., Wojcik, J., and Gauthier, J. M. (2001). Protein–protein interaction maps: a lead towards cellular functions. TRENDS in Genetics *17*, 346–352.

Lennon, N. J., Lintner, R. E., Anderson, S., Alvarez, P., Barry, A., Brockman, W., Daza, R., Erlich, R., Giannoukos, G., Green, L., et al. (2010). A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. Genome Biol. *11*, R15.

Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H. E., McMullan, D., Shin, T., et al. (2002). Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. U.S.A *99*, 11664-11669.

Litowski, J. R., and Hodges, R. S. (2002). Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. J. Biol. Chem. *277*, 37272-9.

Liu, R., Xu, H., Wei, Z., Wang, Y., Lin, Y., and Gong, W. (2009). Crystal structure of human adenylate kinase 4 (L171P) suggests the role of hinge region in protein domain motion. Biochem. Biophys. Res. Commun. *379*, 92-97.

Löfdahl, P., Nord, O., Janzon, L., and Nygren, P. (2009). Selection of TNF-alpha binding affibody molecules using a beta-lactamase protein fragment complementation assay. Nat. Biotech. 26(5):251-9.

Longtin, R. (2005). An Integrated Approach: Systems Biology Seeks Order in Complexity (Journal of the National Cancer Institute, Vol. 97, No. 7,\copyright Oxford University Press 2005, all rights reserved.).

Luker, K. E., Gupta, M., and Luker, G. D. (2009). Imaging chemokine receptor dimerization with firefly luciferase complementation. FASEB J. *23*, 823-834.

Luker, K. E., Gupta, M., Steele, J. M., Foerster, B. R., and Luker, G. D. (2009). Imaging ligand-dependent activation of CXCR7. Neoplasia *11*, 1022-1035.

Lynn, D. J., Singer, G. A. C., and Hickey, D. A. (2002). Synonymous codon usage is subject to selection in thermophilic bacteria. Nucl. Acids Res. *30*, 4272-4277.

Ma'ayan, A. (2009). Insights into the organization of biochemical regulatory networks using graph theory analyses. J. Biol. Chem. *284*, 5451-5455.

Magliery, T. J., Wilson, C. G., Pan, W., Mishler, D., Ghosh, I., Hamilton, A. D., and Regan, L. (2005). Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. J. Am. Chem. Soc. *127*, 146-57.

Maheshwari, R., Bharadwaj, G., and Bhat, M. K. (2000). Thermophilic Fungi: Their Physiology and Enzymes. Microbiol. Mol. Biol. Rev. *64*, 461-488.

Manco, G., Giosuè, E., D'Auria, S., Herman, P., Carrea, G., and Rossi, M. (2000). Cloning, overexpression, and properties of a new thermophilic and thermostable esterase with sequence similarity to hormone-sensitive lipase subfamily from the archaeon Archaeoglobus fulgidus. Arch. Biochem. Biophys. *373*, 182-192.

Marguet, E., and Forterre, P. (1994). DNA stability at temperatures typical for hyperthermophiles. Nucleic Acids Res. *22*, 1681-6.

Martins, L., Huber, R., Huber, H., Stetter, K., Da Costa, M., and Santos, H. (1997). Organic Solutes in Hyperthermophilic Archaea. Appl. Environ. Microbiol. *63*, 896-902.

Massant, J., and Glansdorff, N. (2004). Metabolic channelling of carbamoyl phosphate in the hyperthermophilic archaeon Pyrococcus furiosus: dynamic enzyme-enzyme interactions involved in the formation of the channelling complex. Biochem. Soc. Trans. *32*, 306-9.

Massant, J., Verstreken, P., Durbecq, V., Kholti, A., Legrain, C., Beeckmans, S., Cornelis, P., and Glansdorff, N. (2002a). Metabolic channeling of carbamoyl phosphate, a thermolabile intermediate: evidence for physical interaction between carbamate kinase-like carbamoyl-phosphate synthetase and ornithine carbamoyltransferase from the hyperthermophile Pyrococcus furiosus. J. Biol. Chem. *277*, 18517-22.

Massant, J., Verstreken, P., Durbecq, V., Kholti, A., Legrain, C., Beeckmans, S., Cornelis, P., and Glansdorff, N. (2002b). Metabolic Channeling of Carbamoyl Phosphate, a Thermolabile Intermediate. J. Biol. Chem. *277*, 18517-18522.

Mathrani, I. M., and Ahring, B. K. (1992). Thermophilic and alkalophilic xylanases from several Dictyoglomus isolates. Appl. Microbiol. Biotech. *38*, 23-27.

Maugini, E., Tronelli, D., Bossa, F., and Pascarella, S. (2009). Structural adaptation of the subunit interface of oligomeric thermophilic and hyperthermophilic enzymes. Comput. Biol. Chem. *33*, 137-148.

Mayr, L. M., and Bojanic, D. (2009). Novel trends in high-throughput screening. Curr. Opin. Pharmacol. *9*, 580-588.

Mazzocchi, F. (2008). Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. EMBO reports *9*, 10.

Menon, A. L., Poole, F. L., Cvetkovic, A., Trauger, S. A., Kalisiak, E., Scott, J. W., Shanmukh, S., Praissman, J., Jenney, F. E., Wikoff, W. R., et al. (2009). Novel Multiprotein Complexes Identified in the Hyperthermophilic Archaeon Pyrococcus furiosus by Non-denaturing Fractionation of the Native Proteome. Mol. Cell. Proteomics *8*, 735-751.

Meyer, M. M., Hochrein, L., and Arnold, F. H. (2006). Structure-guided SCHEMA recombination of distantly related beta-lactamases. Protein Eng. Des. Sel. *19*, 563-70.

Meyer, M. M., Silberg, J. J., Voigt, C. A., Endelman, J. B., Mayo, S. L., Wang, Z. G., and Arnold, F. H. (2003). Library analysis of SCHEMA-guided protein recombination. Protein Sci. *12*, 1686-93.

Michael, S., Auld, D., Klumpp, C., Jadhav, A., Zheng, W., Thorne, N., Austin, C. P., Inglese, J., and Simeonov, A. (2008). A robotic platform for quantitative high-throughput screening. Assay Drug Dev. Technol. *6*, 637-657.

Michnick, S. (2003). Protein fragment complementation strategies for biochemical network mapping. Curr. Op. Biotech. *14*, 610-617.

Michnick, S. W., Remy, I., Campbell-Valois, F. X., Vallée-Bélisle, A., and Pelletier, J. N. (2000). Detection of protein-protein interactions by protein fragment complementation strategies. Meth. Enzymol. *328*, 208-230.

Michnick, S. W. (2003). Protein fragment complementation strategies for biochemical network mapping. Curr. Opin. Biotechnol. *14*, 610-7.

Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I., and Stefan, E. (2007). Universal strategies in research and drug discovery based on protein-fragment complementation assays. Nat. Rev. Drug Discov. *6*, 569-82.

Middendorf, M., Ziv, E., and Wiggins, C. H. (2005). Inferring network mechanisms: the Drosophila melanogaster protein interaction network. Proc. Natl. Acad. Sci. U S A *102*, 3192-3197.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824-827.

Missimer, J. H., Steinmetz, M. O., Baron, R., Winkler, F. K., Kammerer, R. A., Daura, X., and van Gunsteren, W. F. (2007). Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. Protein Sci. *16*, 1349-1359.

Mizuuchi, K. (1983). In vitro transposition of bacteriophage Mu: a biochemical approach to a

novel replication reaction. Cell *35*, 785-794.

Monnot, M., Gilles, A. M., Girons, I. S., Michelson, S., Barzu, O., and Fermandjian, S. (1987). Circular dichroism investigation of Escherichia coli adenylate kinase. J. Biol. Chem. *262*, 2502-2506.

Monteiro, P. T., Ropers, D., Mateescu, R., Freitas, A. T., and de Jong, H. (2008). Temporal logic patterns for querying dynamic models of cellular interaction networks. Bioinformatics *24*, i227-233.

Muff, T. J., Foster, R. M., Liu, P. J. Y., and Ordal, G. W. (2007). CheX in the Three-Phosphatase System of Bacterial Chemotaxis. J. Bacteriology *189*, 7007-7013.

Nacher, J. C., and Akutsu, T. (2007). Recent progress on the analysis of power-law features in complex cellular networks. Cell Biochem. Biophys *49*, 37-47.

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., et al. (2002). Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. DNA Res. *9*, 123-130.

Naylor, T. (2004). Finding extraterrestrial sites for thermophiles. Biochem. Soc. Trans. *32*, 165-167.

Neelon, K., Schreier, H. J., Meekins, H., Robinson, P. M., and Roberts, M. F. (2005). Compatible solute effects on thermostability of glutamine synthetase and aspartate transcarbamoylase from Methanococcus jannaschii. Biochim. Biophys. Acta *1753*, 164-173.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., et al. (1999a). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature *399*, 323-329.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., et al. (1999b). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature *399*, 323-329.

Neylon, C. (2004). Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucl. Acids Res. *32*, 1448-1459.

Nguyen, P. Q., Liu, S., Thompson, J. C., and Silberg, J. J. (2008). Thermostability promotes the cooperative function of split adenylate kinases. Protein Eng. Des. Sel. *21*, 303-310.

Noll, K. M., and Vargas, M. (1997). Recent advances in genetic analyses of hyperthermophilic archaea and bacteria. Arch. Microbiol. *168*, 73-80.

Noma, T., Fujisawa, K., Yamashiro, Y., Shinohara, M., Nakazawa, A., Gondo, T., Ishihara, T., and Yoshinobu, K. (2001). Structure and expression of human mitochondrial adenylate kinase targeted to the mitochondrial matrix. Biochem. J. *358*, 225-232.

Nyfeler, B., and Hauri, H. (2007). Visualization of protein interactions inside the secretory pathway. Biochem. Soc. Trans. *35*, 970-973.

Nyfeler, B., Michnick, S. W., and Hauri, H. P. (2005). Capturing protein interactions in the secretory pathway of living cells. Proc. Natl. Acad. Sci. U S A *102*, 6350-5.

Ogasahara, K., Ishida, M., and Yutani, K. (2003). Stimulated Interaction between α and β Subunits of Tryptophan Synthase from Hyperthermophile Enhances Its Thermal Stability. J. Biol. Chem. *278*, 8922-8928.

Oshima, T. (2007). Unique polyamines produced by an extreme thermophile, Thermus thermophilus. Amino Acids *33*, 367-372.

Oshima, T., and Imahori, K. (1974). Description of Thermus thermophilus (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. Int. J. Syst. Bacteriol. *24*, 102-112.

Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999). Combinatorial protein engineering by incremental truncation. Proc. Natl. Acad. Sci. U S A *96*, 3562-7.

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837-846.

Pantazaki, A. P. (2002). Biotechnologically relevant enzymes from *Thermus thermophilus*. Appl. Microbiol. Biotech. *58*, 1-12.

Park, S. Y., Beel, B. D., Simon, M. I., Bilwes, A. M., and Crane, B. R. (2004). In different organisms, the mode of interaction between two signaling proteins is not necessarily conserved. Proc. Natl. Acad. Sci. U S A *101*, 11646-51.

Park, S. Y., Chao, X., Gonzalez-Bonet, G., Beel, B. D., Bilwes, A. M., and Crane, B. R. (2004). Structure and Function of an Unusual Family of Protein Phosphatases The Bacterial Chemotaxis Proteins CheC and CheX. Mol. Cell *16*, 563–574.

Pasek, S., Risler, J. L., and Brezellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics *22*, 1418-23.

Pastor-Satorras, R., Smith, E., and Solé, R. V. (2003). Evolving protein interaction networks

through gene duplication. J. Theor. Biol. *222*, 199-210.

Paulmurugan, R., Umezawa, Y., and Gambhir, S. S. (2002). Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. Proc. Natl. Acad. Sci. U S A *99*, 15608-13.

Pazos, F., Valencia, A., and DeLorenzo, V. (2003). The organization of the microbial biodegradation network from a systems-biology perspective. EMBO reports *4*, 994–999.

Pelletier, J. N., Campbell-Valois, F. X., and Michnick, S. W. (1998). Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. Proc. Natl. Acad. Sci. U S A 95(21):12141-6.

Pentecost, A. (1996). High temperature ecosystems and their chemical interactions with their environment. Ciba Found. Symp *202*, 99-108; discussion 108-111.

Perner, M., Bach, W., Hentscher, M., Koschinsky, A., Garbe-Schönberg, D., Streit, W. R., and Strauss, H. (2009). Short-term microbial and physico-chemical variability in low-temperature hydrothermal fluids near 5 degrees S on the Mid-Atlantic Ridge. Environ. Microbiol. *11*, 2526-2541.

Perrier, V., Burlacu-Miron, S., Boussac, A., Meier, A., and Gilles, A. M. (1998). Metal chelating properties of adenylate kinase from Paracoccus denitrificans. Protein Eng. *11*, 917-923.

Perrier, V., Burlacu-Miron, S., Bourgeois, S., Surewicz, W. K., and Gilles, A. M. (1998). Genetically engineered zinc-chelating adenylate kinase from Escherichia coli with enhanced thermal stability. J. Biol. Chem. *273*, 19097–19101.

Perrier, V., Surewicz, W. K., Glaser, P., Martineau, L., Craescu, C. T., Fabian, H., Mantsch, H. H., Barzu, O., and Gilles, A. (1994). Zinc Chelation and Structural Stability of Adenylate Kinase from Bacillus subtilis. Biochemistry *33*, 9960-9967.

Perugino, G., Valenti, A., D'amaro, A., Rossi, M., and Ciaramella, M. (2009). Reverse gyrase and genome stability in hyperthermophilic organisms. Biochem. Soc. Trans. *37*, 69-73.

Petursdottir, S. K., Bjornsdottir, S. H., Hreggvidsson, G. O., Hjorleifsdottir, S., and Kristjansson, J. K. (2009). Analysis of the unique geothermal microbial ecosystem of the Blue Lagoon. FEMS Microbiol. Ecol. *70*, 425.

Pham, V. D., Palden, T., and DeLong, E. F. (2007). Large-scale screens of metagenomic libraries. J. Vis. Exp., 201.

Phizicky, E. M., and Fields, S. (1995). Protein-protein interactions: methods for detection

and analysis. Microbiol. Rev. *59*, 94-123.

Pikuta, E. V., Hoover, R. B., and Tang, J. (2007). Microbial extremophiles at the limits of life. Crit. Rev. Microbiol. *33*, 183-209.

de Prat Gay, G., Ruiz-Sanz, J., and Fersht, A. R. (1994). Generation of a family of protein fragments for structure-folding studies. 2. Kinetics of association of the two chymotrypsin inhibitor-2 fragments. Biochemistry *33*, 7964-70.

Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schächter, V., et al. (2001). The protein-protein interaction map of Helicobacter pylori. Nature *409*, 211-215.

Ramaley, R. F., and Bitzinger, K. (1975). Types and Distribution of Obligate Thermophilic Bacteria in Man-Made and Natural Thermal Gradients. Appl. Microbiol. *30*, 152-155.

Remy, I., and Michnick, S. W. (2006). A highly sensitive protein-protein interaction assay based on Gaussia luciferase. Nat Methods *3*, 977-9.

Remy, I., and Michnick, S. W. (1999). Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. Proc. Natl. Acad. Sci. U S A *96*, 5394-9.

Reysenbach, A. L., and Cady, S. L. (2001). Microbiology of ancient and modern hydrothermal systems. Trends Microbiol. *9*, 79-86.

Rial, D. V., and Ceccarelli, E. A. (2002). Removal of DnaK contamination during fusion protein purifications. Protein Expr. Purif. *25*, 503-7.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends in Genetics *16*, 276-7.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotech. *17*, 1030-1032.

Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M., and Marqusee, S. (2003). Role of residual structure in the unfolded state of a thermophilic protein. Proc. Natl. Acad. Sci. U S A *100*, 11345-9.

Robinson-Rechavi, M., Alibés, A., and Godzik, A. (2006a). Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritima. J. Mol. Biol. *356*, 547-557.

Robinson-Rechavi, M., Alibés, A., and Godzik, A. (2006b). Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons

from structural genomics of Thermotoga maritima. J. Mol. Biol. *356*, 547-557.

Rose, T., Brune, M., Wittinghofer, A., Le Blay, K., Surewicz, W. K., Mantsch, H. H., Bârzu, O., and Gilles, A. M. (1991). Structural and catalytic properties of a deletion derivative (delta 133-157) of Escherichia coli adenylate kinase. J. Biol. Chem. *266*, 10781-10786.

Rossi, F., Charlton, C. A., and Blau, H. M. (1997). Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. Proc. Natl. Acad. Sci. U.S.A *94*, 8405-8410.

Rothschild, L. J., and Mancinelli, R. L. (2001). Life in extreme environments. Nature *409*, 1092-101.

Rundqvist, L., Adén, J., Sparrman, T., Wallgren, M., Olsson, U., and Wolf-Watz, M. (2009). Noncooperative folding of subdomains in adenylate kinase. Biochemistry *48*, 1911-1927.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487-491.

Saint Girons, I., Gilles, A. M., Margarita, D., Michelson, S., Monnot, M., Fermandjian, S., Danchin, A., and Bârzu, O. (1987). Structural and catalytic characteristics of Escherichia coli adenylate kinase. Journal of Biological Chemistry *262*, 622-629.

Santos, H., and da Costa, M. S. (2001). Organic solutes from thermophiles and hyperthermophiles. Meth. Enzymol *334*, 302-315.

Schadt, E. E., and Lum, P. Y. (2006). Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. J. of Lipid Res. *47*, 2601-2613.

Schadt, E. E., Zhang, B., and Zhu, J. (2009). Advances in systems biology are enhancing our understanding of disease and moving us closer to novel disease treatments. Genetica *136*, 259-269.

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

Schlauderer, G. J., Proba, K., and Schulz, G. E. (1996). Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. J. Mol. Biol. *256*, 223–227.

Schulz, G. E. (1987). Structural and functional relationships in the adenylate kinase family.

Cold Spring Harb. Symp. Quant. Biol. *52*, 429-439.

Schulz, G. E., Müller, C. W., and Diederichs, K. (1990). Induced-fit movements in adenylate kinases. J. Mol. Biol. *213*, 627-630.

Schulz, G. E., Schiltz, E., Tomasselli, A. G., Frank, R., Brune, M., Wittinghofer, A., and Schirmer, R. H. (1986). Structural relationships in the adenylate kinase family. Eur. J. Biochem. *161*, 127-132.

Schulz, G. E., Muller, C. W., and Diederichs, K. (1990). Induced-fit movements in adenylate kinases. J. Mol. Biol. *213*, 627-30.

Schulz, G. (1987). Structural and Functional Relationships in the Adenylate Kinase Family. Cold Spring Harbor Symposia on Quantitative Biology *52*, 429-439.

Segerer, A. H., Burggraf, S., Fiala, G., Huber, G., Huber, R., Pley, U., and Stetter, K. O. (1993). Life in hot springs and hydrothermal vents. Orig. Life. Evol. Biosph. *23*, 77-90.

Shetty, R. P., Endy, D., and Knight, T. F. (2008). Engineering BioBrick vectors from BioBrick parts. J. Biol. Eng. *2*, 5.

Shiba, K., and Schimmel, P. (1992). Functional assembly of a randomly cleaved protein. Proc. Natl. Acad. Sci. U S A *89*, 1880-4.

Shima, S., Tziatzios, C., Schubert, D., Fukada, H., Takahashi, K., Ermler, U., and Thauer, R. K. (1998). Lyotropic-salt-induced changes in monomer/dimer/tetramer association equilibrium of formyltransferase from the hyperthermophilic Methanopyrus kandleri in relation to the activity and thermostability of the enzyme. Eur. J. Biochem. *258*, 85-92.

Siddiqui, M. A., Fujiwara, S., Takagi, M., and Imanaka, T. (1998). In vitro heat effect on heterooligomeric subunit assembly of thermostable indolepyruvate ferredoxin oxidoreductase. FEBS Lett. *434*, 372-6.

Siderius, M., Van Wuytswinkel, O., Reijenga, K. A., Kelders, M., and Mager, W. H. (2000). The control of intracellular glycerol in Saccharomyces cerevisiae influences osmotic stress response and resistance to increased temperature. Mol. Microbiol. *36*, 1381-1390.

Singh, A., Mai, D., Kumar, A., and Steyn, A. J. (2006). Dissecting virulence pathways of Mycobacterium tuberculosis through protein-protein association. Proc. Natl. Acad. Sci. U S A *103*, 11346-51.

Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999). Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure

of Bacillus Ak.1 protease at 1.8 A resolution. J. Mol. Biol. *294*, 1027-1040.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., et al. (1997). Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. J. Bacteriol. *179*, 7135-7155.

Snel, B., Bork, P., and Huynen, M. (2000). Genome evolution. Gene fusion versus gene fission. Trends in Genetics *16*, 9-11.

Somero, G. N. (1978). Temperature Adaptation of Enzymes: Biological Optimization Through Structure-Function Compromises. Annu. Rev. Ecol. Syst. *9*, 1-29.

Sommer, P., Georgieva, T., and Ahring, B. K. (2004). Potential for using thermophilic anaerobic bacteria for bioethanol production from hemicellulose. Biochem. Soc. Trans. *32*, 283-289.

Song, S., Inouye, S., Kawai, M., Fukami-Kobayashi, K., Gõ, M., and Nakazawa, A. (1996). Cloning and characterization of the gene encoding Halobacterium halobium adenylate kinase. Gene *175*, 65-70.

Spotts, J. M., Dolmetsch, R. E., and Greenberg, M. E. (2002). Time-lapse imaging of a dynamic phosphorylation-dependent protein-protein interaction in mammalian cells. Proc. Natl. Acad. Sci. U S A *99*, 15142-7.

Stewart, R. C., and Van Bruggen, R. (2004). Association and Dissociation Kinetics for CheY Interacting with the P2 Domain of CheA. J. Mol. Biol. *336*, 287-301.

Stolworthy, T. S., and Black, M. E. (2001). The mouse guanylate kinase double mutant E72Q/D103N is a functional adenylate kinase. Protein Eng. *14*, 903-909.

Suhre, K., and Claverie, J. M. (2003a). Genomic correlates of hyperthermostability, an update. J. Biol. Chem. *278*, 17198-202.

Suhre, K., and Claverie, J. (2003b). Genomic correlates of hyperthermostability, an update. J. Biol. Chem. *278*, 17198-17202.

Tanner, J. J., Hecht, R. M., and Krause, K. L. (1996). Determinants of enzyme thermostability observed in the molecular structure of Thermus aquaticus D-glyceraldehyde-3-phosphate dehydrogenase at 25 Angstroms Resolution. Biochemistry *35*, 2597-2609.

Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Molina, M. M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S. W. (2008). An in vivo map of the yeast protein interactome. Science's STKE *320*, 1465.

Taylor, M. P., Eley, K. L., Martin, S., Tuffin, M. I., Burton, S. G., and Cowan, D. A. (2009). Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. Trends in Biotechnology *27*, 398-405.

Tehei, M., and Zaccai, G. (2005). Adaptation to extreme environments: macromolecular dynamics in complex systems. Biochim. Biophys. Acta. *1724*, 404-410.

Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainshtein, B. K., Frömmel, C., Höhne, W. E., and Wilson, K. S. (1990). Crystal structure of thermitase at 1.4 A resolution. J. Mol. Biol. *214*, 261-279.

Terui, Y., Ohnuma, M., Hiraga, K., Kawashima, E., and Oshima, T. (2005). Stabilization of nucleic acids by unusual polyamines produced by an extreme thermophile, Thermus thermophilus. Biochem. J. *388*, 427-433.

Thoma, R., Hennig, M., Sterner, R., and Kirschner, K. (2000). Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from Thermotoga maritima. Structure *8*, 265-276.

Thompson, M. J., and Eisenberg, D. (1999). Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J. Mol. Biol. *290*, 595-604.

Thummes, K., Schäfer, J., Kämpfer, P., and Jäckel, U. (2007). Thermophilic methanogenic Archaea in compost material: Occurrence, persistence and possible mechanisms for their distribution to other environments. Syst. and Appl. Microb. *30*, 634-643.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., et al. (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature *388*, 539-547.

de la Tour, C. B., Kaltoum, H., Portemer, C., Confalonieri, F., Huber, R., and Duguet, M. (1995). Cloning and sequencing of the gene coding for topoisomerase I from the extremely thermophilic eubacterium, Thermotoga maritima. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression *1264*, 279-283.

Trebbi, B., Dehez, F., Fowler, P. W., and Zerbetto, F. (2005). Favorable entropy of aromatic clusters in thermophilic proteins. J. Phys. Chem. B *109*, 18184-18188.

Uetz, P., and Hughes, R. E. (2000). Systematic and large-scale two-hybrid screens. Curr. Opin. Microbiol. *3*, 303-308.

Usui, K., Katayama, S., Kanamori-Katayama, M., Ogawa, C., Kai, C., Okada, M., Kawai, J., Arakawa, T., Carninci, P., Itoh, M., et al. (2005). Protein-protein interactions of the hyperthermophilic archaeon Pyrococcus horikoshii OT3. Genome Biol *6*, R98.

Vidal, M., and Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. Nucleic

Acids Res. *27*, 919-929.

Vidi, P., and Watts, V. J. (2009). Fluorescent and bioluminescent protein-fragment complementation assays in the study of G protein-coupled receptor oligomerization and signaling. Mol. Pharmacol. *75*, 733-739.

Vieille, C., Krishnamurthy, H., Hyun, H. H., Savchenko, A., Yan, H., and Zeikus, J. G. (2003). Thermotoga neapolitana adenylate kinase is highly active at 30 degrees C. Biochem J *372*, 577-85.

Vieille, C., Krishnamurthy, H., Hyung-Hwan, H., Savchenko, A., Honggao, Y. A. N., and Zeikus, J. G. (2003). Thermotoga neapolitana adenylate kinase is highly active at 30° C. Biochem. J *372*, 577–585.

Villalobos, V., Naik, S., and Piwnica-Worms, D. (2008). Detection of protein-protein interactions in live cells and animals with split firefly luciferase protein fragment complementation. Methods Mol. Biol *439*, 339-352.

Villoslada, P., Steinman, L., and Baranzini, S. E. (2009). Systems biology and its application to the understanding of neurological diseases. Ann Neurol *65*, 124–139.

Vogt, G., Woell, S., and Argos, P. (1997). Protein thermal stability, hydrogen bonds, and ion pairs. J. Mol. Biol *269*, 631-643.

van de Vossenberg, J. L., Driessen, A. J., and Konings, W. N. (1998). The essence of being extremophilic: the role of the unique archaeal membrane lipids. Extremophiles *2*, 163-170.

Wang, X., Minasov, G., and Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol *320*, 85-95.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature *393*, 440-442.

Wilson, C. G., Magliery, T. J., and Regan, L. (2004). Detecting protein-protein interactions with GFP-fragment reassembly. Nat Methods *1*, 255-62.

Wood, E., Ghane, F., and Grogan, D. (1997). Genetic responses of the thermophilic archaeon Sulfolobus acidocaldarius to short-wavelength UV light. J. Bacteriol. *179*, 5693-5698.

Wu, F. T. H., Stefanini, M. O., Mac Gabhann, F., and Popel, A. S. (2009). Modeling of growth factor-receptor systems from molecular-level protein interaction networks to

whole-body compartment models. Meth. Enzymol *467*, 461-497.

Wycuff, D. R., and Matthews, K. S. (2000). Generation of an AraC-araBAD promoter-regulated T7 expression system. Anal. Biochem. *277*, 67-73.

Xu, Y., and Glansdorff, N. (2002). Was our ancestor a hyperthermophilic procaryote? Comp. Biochem. Physiol., Part A Mol. Integr. Physiol *133*, 677-688.

Yamada, T., and Bork, P. (2009). Evolution of biomolecular networks: lessons from metabolic and protein interactions. Nat. Rev. Mol. Cell Biol *10*, 791-803.

Young, K. H. (1998). Yeast two-hybrid: so many interactions, (in) so little time. Biol. Reprod. *58*, 302-311.

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

Zhou, H., Li, J., Peng, X., Meng, J., Wang, F., and Ai, Y. (2009). Microbial diversity of a sulfide black smoker in main endeavour hydrothermal vent field, Juan de Fuca Ridge. J. Microbiol. *47*, 235-247.

Zhu, D., and Qin, Z. S. (2005). Structural comparison of metabolic networks in selected single cell organisms. BMC Bioinformatics *6*, 8.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., and van 't Riet, K. (1990). Modeling of the Bacterial Growth Curve. Appl. Environ. Microbiol. *56*, 1875-1881.