This is to certify that the
dissertation entitled

Analysis of rigidity, stability, and activity in *Thermotoga neapolitana* adenylate kinase

presented by

Harini Krishnamurthy

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Cell and Molecular Biology Program

Major Professor's Signature

2/7/06

Date

ANALYSIS OF RIGIDITY, STABILITY, AND ACTIVITY IN *Thermotoga neapolitana* ADENYLATE KINASE

By

Harini Krishnamurthy

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Program in Cell and Molecular Biology

2006

ABSTRACT

ANALYSIS OF RIGIDITY, STABILITY, AND ACTIVITY IN *Thermotoga neapolitana* ADENYLATE KINASE

By

Harini Krishnamurthy

Proteins from hyperthermophiles are adapted to be stable and active at temperatures ≥ 80°C, conditions in which mesophilic proteins denature. Because the sequence and structural similarity between hyperthermophilic proteins and their mesophilic homologues is very high, it is not readily apparent as to what causes the enhanced stability of hyperthermophilic proteins. Understanding the mechanisms of protein thermostability has been one of the most intensely studied problems in the last three decades. The activity-temperature profiles of hyperthermophilic and mesophilic proteins show that the profile is right-shifted for hyperthermophiles. In other words, they attain similar activity as their mesophilic homologues at a much higher temperature and they are inactive at low temperatures. These facts have lead to the proposal of the 'rigidity' hypothesis regarding hyperthermophilic proteins. Hyperthermophilic proteins achieve high thermostability through increased structural rigidity. These proteins are highly rigid in their native structure that allows them to maintain their structural integrity at high temperatures. The increased rigidity of hyperthermophilic proteins freezes out fluctuations required for activity at room temperature making them inactive at low temperatures.

The studies described here use adenylate kinase from the hyperthermophilic bacteria *Thermotoga neapolitana* (TNAK) as a model system in which to test the rigidity hypothesis. What makes TNAK a really interesting model system is the fact that it is highly active at 30°C, an unusual property for a hyperthermophilic protein. Special attention was paid to the techniques used to investigate motions in TNAK. NMR and MD simulations, techniques that can access a wide range of timescales in atomic detail, were used to compare dynamics in TNAK with its mesophilic homologue from *Escherichia coli* (ECAK).

Results from ^{15}N NMR relaxation data shows that TNAK is uniformly more rigid than ECAK in the ps-ns timescales as well as μs timescale. Although, overall, TNAK has higher rigidity than ECAK, several residues in the AMP-binding and lid domains exhibit high flexibility in the ps-ns timescale. Residues in the hinge regions between the lid and core domains of TNAK exhibit flexibility in the ps-ns timescales as well μs timescales. H-D exchange data, which provide information on timescales greater than seconds, show that TNAK's lid and AMP-binding domains are more stabilized compared to ECAK. Again, in spite of this increased rigidity, several residues in these domains of TNAK show considerable local fluctuations.

The increased rigidity coupled with localized flexibility may explain the simultaneously high activity at low temperatures and stability at high temperatures of TNAK. The results also show that TNAK is not uniformly rigid across the entire structure at all timescales. Instead, TNAK domains are differentially stabilized at different timescales. Finally, the results show that the rigidity hypothesis may not hold true for all proteins.

To my Mother, Father, and Suba

ACKNOWLEDGEMENTS

I would like to start by sincerely thanking everyone who has helped me carry out this research. I thank Dr. Greg Zeikus, my thesis advisor, for believing in me and for giving me the independence to explore new scientific perspectives. It was a great fortune for me to have Dr. Claire Vieille in the lab who was a mentor and friend to me throughout these years. Without her support and guidance this project would not have seen the light of day. Both of them have been the driving force behind this interdisciplinary thesis and I look forward to using the skills and enthusiasm imparted to me in this lab for an exciting scientific career.

I would like thank my thesis guidance committee: Dr. Leslie Kuhn, Dr. Honggao Yan, Dr. Sheilagh Ferguson-Miller, Dr. Rawle Hollingsworth and Dr. Tom Schmidt for their many substantive suggestions and thoughtful discussions. I am especially grateful to Dr. Kuhn and Dr. Yan for always being available to answer my questions. I am grateful to Dr. Aizhou Liu for his invaluable support and guidance in NMR experiments. My thanks to Lishan Yao, who was instrumental in completing this research. We have both worked on the same techniques in our respective labs and our long discussions have helped me understand the theoretical aspects of NMR. I also extend my appreciation to the members of the Cell and Molecular Biology Program and the Department of Biochemistry and Molecular Biology for their help and support.

I thank all the present and past members of the Zeikus laboratory. Special thanks to Jake McKinlay for all the interesting discussions, both scientific and non-scientific, and for his friendship over the years. I extend an enthusiastic thank you to Maris

Laivenieks and Karla Fjeld for creating an enjoyable working environment and providing useful feedback. My thanks also go to Dr. Pil Kim, Dr. Suil Kang, and Dr. Yoshiharu Sasaki, all former members of the lab, for sharing their expertise and friendship.

But for all the great friends I have made as a graduate student, my life at MSU would not have been as enjoyable. I especially thank Ajith Vengellur, Clarisa Bejar, Anastasia Gridasova, Sandhya Payankulam, and Radhika Iyer for their close friendship and support. There are many more friends who I would like to thank for the great moments that I have had during these years but I am unable to due to lack of space. My most cherished time in graduate school has been the many good friends I have made from diverse cultures. All these people have infinitely enriched my life.

Finally, I would like to thank my family: mom, dad, and my sister, Subashini without whose support I would not be where I am today. They have provided me with the courage and inspiration to keep going when things seemed impossible. I also thank my uncles and aunts, especially Krishnan mama and Radha mami for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

$\langle S^2 \rangle$	mean S^2
2D	2-dimensional
3D	3-dimensional
AAAK	*Aquifex aeolicus* AK
AK	adenylate kinase
AMPbd domain	AMP-binding domain
Ap5A	diadenosine pentaphosphate
BGAK	*Bacillus globisporus* AK
BStAK	*Bacillus stearothermophilus* AK
BSubAK	*Bacillus subtilis* AK
CD	circular dichroism
DSC	differential scanning calorimetry
ECAK	*Escherichia coli* AK
ECAK*Ap5A	ECAK complexed with Ap5A
EcRNase	*E. coli* RNase H
GdnHcl	guanidine hydrochloride
H-D exchange	hydrogen-deuterium exchange
HSQC	heteronuclear single quantum coherence
ICP-AES	inductively coupled plasma-atomic emission spectroscopy
IPMDH	isopropyl malate dehydrogenase
MD	molecular dynamics
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
ns	nanoseconds

P values	protection factor values
PAR	4-(2-pyridylazo) resorcinol
pdb	protein data bank
PMPS	p-hydroxy-mercuriphenyl sulphonate
ps	picoseconds
RMSF	root mean square fluctuations
RNase	ribonulease
TNAK	*Thermotoga neapolitana* AK
TNAK*Ap5A	TNAK complexed with Ap5A
TtRNase	*Thermus thermophilus* Rnase H
μs	microseconds

Chapter 1

1.1 Hyperthermophiles: Definition and evolution

Bacteria that have an optimal growth temperature above 80 °C are termed hyperthermophiles. All but two of the hyperthermophilic genera are archaea, *Thermotogales* and *Aquificales* being the only bacteria (66, 67). These two genera are the most ancient of bacteria (6). They are the first to have diverged from the Archaea/Eukarya lineage, and are they are also the most slowly evolving of all extant life forms (67). These facts together led many scientists to conclude that life has first appeared on Earth under hyperthermophilic conditions.

This thesis focuses on enzymes isolated from hyperthermophilic bacteria. Pioneering work on thermophilic and hyperthermophilic bacteria was done by Dr Thomas D. Brock at Indiana University in the late 1960s, who studied life in the boiling hot springs of Yellowstone National Park, Wyoming. The first organism to be identified was *Sulfolobus acidocaldarius*, both a hyperthermophile and an acidophile (5). Since then, more than 70 hyperthermophilic species have been isolated (53) from terrestrial and marine hot environments, with the record of high temperature for life held since 2003, by an as-yet-unnamed microbe closely related *Pyrodictium occultum* (30),

Strain 121, which can grow at temperatures up to 121 °C. Previously, the most heat tolerant organism known was *Pyrolobus fumarii* discovered in 1997 from a thermal pool in Italy, which was reported to grow at temperatures up to 113 °C (4). All these findings raise the question: What is the upper limit of temperature for life? When T. D. Brock first discovered microorganisms in hot springs he found that they were not just surviving but growing and flourishing at these temperatures. Subsequent biochemical studies of thermophiles led to the discovery of thermostable protein-synthesis enzymes in *Thermus aquaticus* and the conclusion that thermophiles evolved enzymes that function optimally at their optimal growth temperature (74). Since high temperatures are required rather than tolerated by hyperthermophiles, proteins from these organisms (or hyperthermophilic proteins) are inherently adapted to be stable at high temperatures. Hyperthermophilic enzymes can therefore be used as model systems for studying molecular evolution and molecular mechanisms for protein thermostability and enzyme function. Such studies can lead to the development of new and efficient strategies for protein engineering and to a wide range of biotechnological applications.

1.2 Hyperthermophilic enzymes

Enzymes can be classified into three broad categories based on their thermostability properties:

1. Mesophilic: These are enzymes isolated from organisms that grow below 50 °C. They are typically active between 25 °C and 50 °C.

2. Thermophilic: These enzymes are isolated from organisms that grow optimally between 50 °C and 80 °C. They are typically stable between 60 °C and 80 °C.

3. Hyperthermophilic: These enzymes are isolated from hyperthermophiles, organisms that grow optimally above 80 °C. They are typically active and stable above 80 °C.

With the discovery of hyperthermophilic enzymes, studies were initiated to understand the unusual thermal properties of these enzymes. Initial biochemical and structural studies comparing homologous mesophilic and hyperthermophilic enzymes showed that:

1. Under their respective physiological conditions the molecular properties of hyperthermophilic and mesophilic enzymes are comparable (i.e., they maintain "corresponding states" (52) with respect to overall topology, solvation, activity, etc).

2. Their sequences are typically 40% to 85% similar (13, 60) and their 3D structures are super-imposable. Particularly, the core structures of hyperthermophilic proteins and their mesophilic homologues are similar. The high level of similarity in the core region suggests that even mesophilic proteins are packed very efficiently, not leaving much room for further improvements in stabilization. Sequence comparisons between hyperthermophilic proteins and their mesophilic homologues show that stabilizing interactions are often found in the less conserved regions of the structure (61).

3. When expressed in *Escherichia coli*, the majority of the hyperthermophilic enzymes retain all of their native biochemical properties including proper folding (19), thermostability, and optimal activity at high temperatures (57, 60). This fact has greatly facilitated their study, since hyperthermophilic organisms are hard to grow.

The discovery that hyperthermophilic enzymes can be expressed in mesophilic bacteria like *E. coli* without loss of their thermostability or activity highlighted the inherent nature of their thermostability. Most hyperthermophilic enzymes are intrinsically thermostable with the exception of a few enzymes that are stabilized by intracellular factors such as salts, cofactors, substrates, activators, and general stabilizers such as thermamine. Remarkably, the main differentiating feature between hyperthermophilic and mesophilic proteins is their temperature ranges for stability and activity. Subsequent to this discovery, several studies were conducted to identify the molecular mechanisms by which thermostability is achieved.

1.3 Mechanisms for protein thermostability

A powerful method for identifying molecular mechanisms that confer thermostability is comparing a number of families of homologous hyperthermophilic, thermophilic, and mesophilic proteins. As of 2005, there are 241 completed genomes listed in The Institute for Genomic Research website (www.tigr.org/tigr-scripts/CMR2/CMRGenomics.spl) out of which 18 belong to thermophilic and hyperthermophilic organisms. The structures of hundreds of hyperthermophilic proteins have been deposited in the Protein Data Bank

(pdb). The temperature range of stability for these proteins spans 70 °C to 200 °C, and their sizes vary from 53 residues (rubredoxin) to 419 residues (a monomeric unit of the hexameric glutamate dehydrogenase). Numerous studies have therefore focused on identifying the mechanisms of adaptation of proteins to high temperatures by- first comparing the structures of hyperthermophilic-mesophilic homologous proteins, and second by testing stability (3, 10, 28, 29, 34-36, 55, 56, 59).

Based on these studies, several mechanisms have been proposed to contribute to the greater stability of hyperthermophilic proteins, the most commonly cited being: (i) Optimal hydrophobic packing, including deletion or shortening of loops (48) to reduce conformational entropy; (ii) increase in the number of hydrophobic residues, which may increase the change in enthalpy at melting temperature (10, 47); (iii) increased occurrence of proline residues in turns or loops, increase in the number of charged residues, increase in helix-stabilizing residues such as Arg and decrease in helix-destabilizing residues such as Cys and Pro (10, 21, 48); (iv) increase in the number of hydrogen bonds (62) and salt bridges (34, 48, 69, 72). While there are several strategies to increase thermostability, nature seems to have employed electrostatic interactions (H-bonds and saltbridges) most frequently. The role of salt bridges as a stabilizing mechanism in proteins has long been controversial (22). However, given the high temperatures to which hyperthermophilic proteins are exposed, and given the fact that the desolvation penalty to form a salt bridge decreases with increasing temperature (the dielectric constant of water decreases with increasing temperature), salt-bridges could be stabilizing factors in these enzymes (34, 69). Sequence comparisons and interaction analyses of mesophilic-hyperthermophilic

enzyme pairs agree that no single molecular mechanism dominates the interaction pattern in hyperthermophilic proteins.

1.4 Protein stability

Protein stability can be described in terms of two properties: thermodynamic stability and kinetic stability. The thermodynamic stability of a simple two-state protein is described as:

$$Folded \overset{K_u}{\Leftrightarrow} Unfolded \qquad (1)$$

where K_u is the equilibrium constant for unfolding. The stability of a protein is simply the difference in the Gibbs free energy between the folded and unfolded states (ΔG_{stab}). The two most common ways of studying thermodynamic stability are via thermal denaturation using differential scanning calorimetry (DSC) and chemical denaturation followed by circular dichroism (CD) spectroscopy. Thermal denaturation experiments using DSC yield the enzyme's melting temperature (T_m, the temperature at which 50% of the enzyme is unfolded) and ΔG_{stab}. ΔG_{stab} can be determined only if the enzyme unfolds reversibly. Studies have shown that ΔG_{stab} is typically of the order of 5-15 kcal/mol for a globular protein, a number that corresponds to only a small number of weak interactions. Since the stabilizing enthalpic energy created by the noncovalent bonds in the secondary structures and the hydrophobic core competes with the loss in configurational entropy upon folding, $\Delta G_{stab,}$ in effect represents a small difference between large numbers. The

increase in the free energy of stabilization for a hyperthermophilic protein ($\Delta\Delta G_{stab} = \Delta G_{stab,\ hyperthermophilic} - \Delta G_{stab,\ mesophilic}$) is of the same order of magnitude as ΔG_{stab} indicating that minute structural alterations within the protein may be sufficient to cope with high temperatures. It has been shown that differences in ΔG_{stab} values as small as 3 to 6.5 kcal/mol can lead to increases in T_m of about 12 °C (31, 43).

Figure I-1 indicates the three theoretical ways in which a higher T_m can be achieved for a hyperthermophilic protein: (i) ΔG_{stab} of a hyperthermophilic protein may be up-shifted to higher ΔG_{stab} values (ii) the curvature of the stability curve (specified by the difference in the heat capacity between the folded and unfolded states, ΔC_p) may be broader, leading to a higher T_m, or (iii) the stability curve of the hyperthermophilic protein may be right-shifted with respect to the curve of the mesophilic protein. The most common stabilization mechanism among hyperthermophilic proteins is through a combination of up-shifting and broadening the ΔG_{stab} vs. temperature curve (33). Several examples have shown that the up-shifted stability curve in hyperthermophilic proteins is achieved by a greater enthalpy change at T_m. The greater enthalpy change is derived from a larger number of interactions (e.g., electrostatic, hydrophobic, cation-π) (24, 25, 33, 37).

Kinetic stability, which is expressed as the half-life ($t_{1/2}$) of the enzyme at defined temperatures, is a measure of how rapidly an enzyme unfolds. Kinetic stability of a protein depends on the energy barrier to unfolding. It is a particularly important

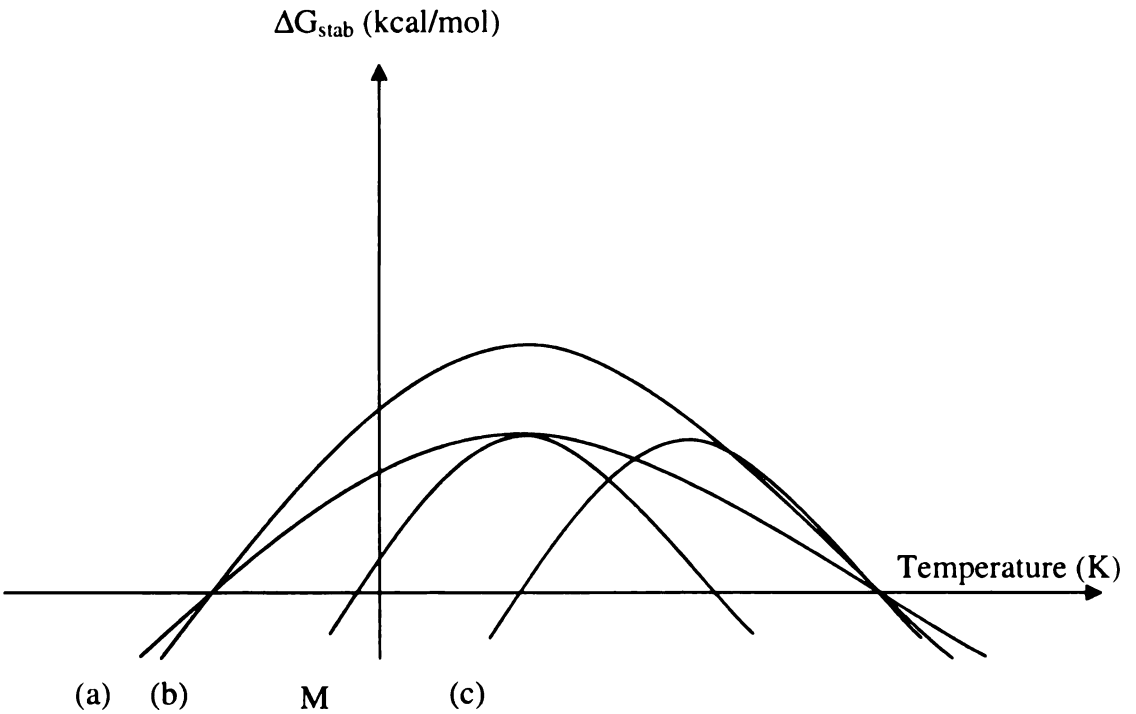


Figure I-1 Comparison of theoretical protein stability curves for hyperthermophilic and mesophilic proteins. Curve M represents a typical ΔG_{stab} vs. Temperature curve for a mesophilic protein. Curves (a), (b), and (c) represent the theoretical ΔG_{stab} vs. Temperature curves for hyperthermophilic proteins. (a) The ΔG_{stab} vs. Temperature curve of the hyperthermophilic protein is broadened. The hyperthermophilic and mesophilic proteins have the same temperature of maximum stability (T_s) and the same ΔG_{stab} at the T_s. (b) The stability curve of the hyperthermophilic protein is up-shifted and broadened (this appears to be the most common mechanism of stabilization in hyperthermophilic proteins). The T_s is the same for both proteins but the ΔG_{stab} value for the hyperthermophilic protein is higher. (c) the stability curve is right-shifted to higher temperatures but the ΔG_{stab} at the T_s are the same for the mesophilic and hyperthermophilic proteins.

consideration for proteins that denature irreversibly or slowly, as is often observed for hyperthermophilic proteins. Irreversible loss of protein folded structure is described by:

$$F \underset{k_2}{\overset{k_1}{\Leftrightarrow}} U \overset{k_3}{\rightarrow} I(nactive) \qquad (2).$$

It has been shown that several hyperthermophilic proteins are kinetically trapped (they unfold more slowly than their mesophilic homologues) and their higher stability is due to their slower unfolding rate constants (8, 9, 49). In the case of the *T. maritima* cold shock protein, entropic (and not enthalpic) factors contribute to its high thermostability (49). In the past two decades, several studies have presented evidence that residual entropy in folded proteins can significantly contribute to the free energy balance of protein stability (54). The only methods currently available to address these issues are molecular dynamic simulations and NMR spin relaxation rate measurements (38, 54, 71).

1.5 Rigidity, thermostability, and activity

A current working hypothesis is that, to be stable at high temperatures, hyperthermophilic enzymes are more rigid than their mesophilic homologues at mesophilic temperatures. This hypothesis stems from the observation that most hyperthermophilic enzymes are inactive at low temperatures (20 °C – 40 °C), the temperature range of optimum activity for their mesophilic homologues. Proteins from hyperthermophilic organisms are adapted to function at temperatures greater than 80°C, at which their mesophilic homologues are denatured. The 3D structures of many homologous hyperthermophilic-mesophilic

enzyme pairs are highly similar and their active site residues are conserved suggesting that the reduced catalytic rate of the hyperthermophilic enzyme at low temperatures does not arise from a dissimilar energy profile of the catalyzed reaction. Instead, the most popular theory invoked to explain the difference in activity at low temperatures involves enzyme motions required for activity. The surprising constancy in the catalytic rates of homologous mesophilic, thermophilic, and hyperthermophilic enzymes at the optimal growth temperatures of their source organism led G. N Somero to propose the "corresponding states" hypothesis (52). It is thought that homologous hyperthermophilic-mesophilic enzyme pairs have similar 'catalytically relevant' flexibility (corresponding states) at their respective optimum temperatures. The stability of hyperthermophilic enzymes is optimized to maintain corresponding states under a given set of environmental growth conditions. Enhanced thermal stability of hyperthermophilic enzymes would then be the result of enhanced conformational rigidity in their folded native state.

Two questions arise from this hypothesis: (i) Do hyperthermophilic enzymes have increased conformational rigidity as a consequence of their stability? (ii) Is the low activity of hyperthermophilic enzymes at ambient temperatures a consequence of their increased rigidity? Several studies have been conducted on mesophilic-hyperthermophilic enzyme pairs in the past two decades to answer these questions. I will present here a few representative examples that either support or do not support this hypothesis.

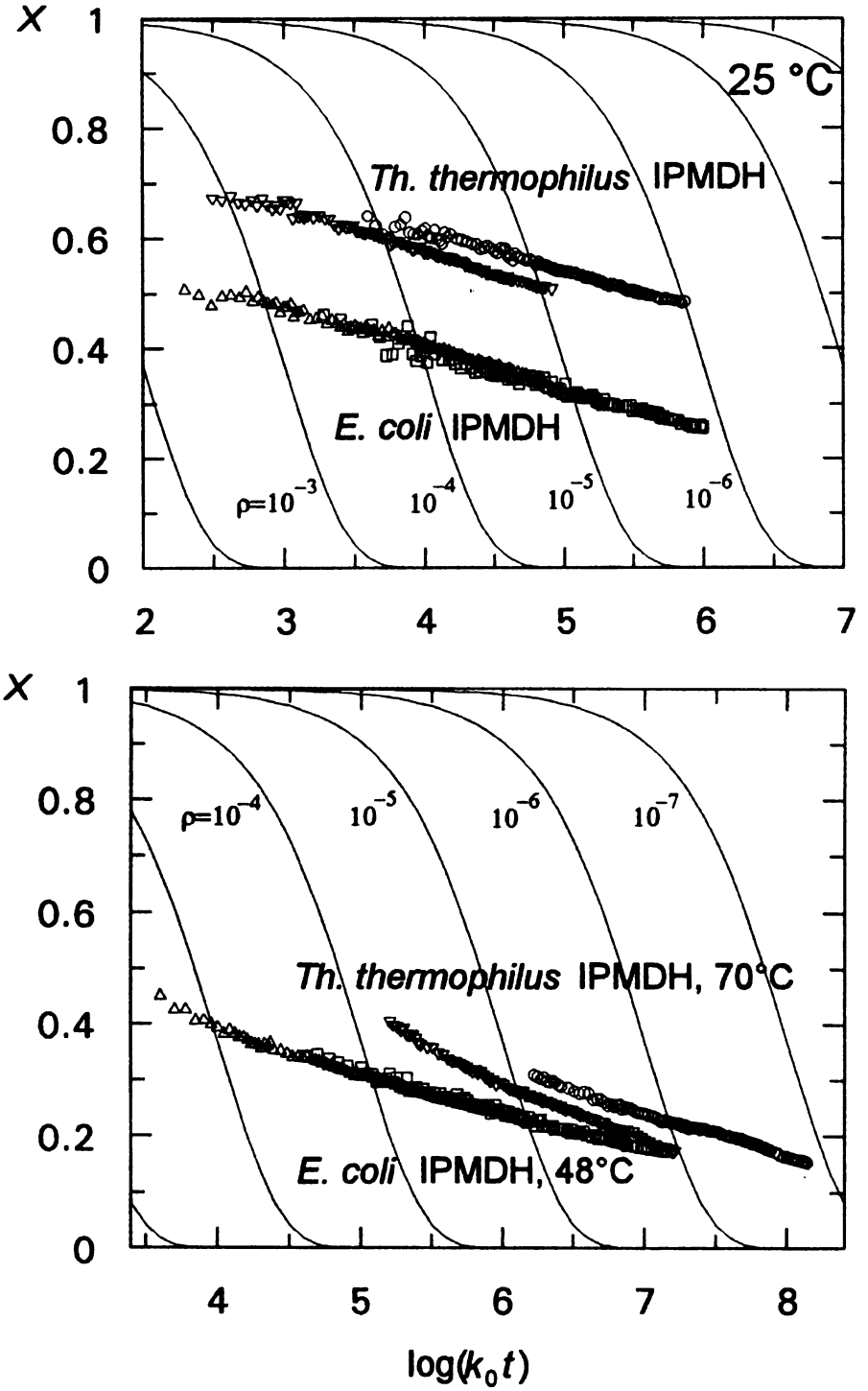
Supporting evidence:

a) Figure I-2 illustrates the slow hydrogen-deuterium (H-D) exchange in *Thermus thermophilus* 3-isopropyl malate dehydrogenase (IPMDH) compared to the faster exchange in its mesophilic *E. coli* homologue at 25°C, each at pD 7.15 and 8.15. At temperatures close to their respective activity optima (48 °C for *E. coli* IPMDH and 70 °C for *T. thermophilus* IPMDH), the exchange rates become super-imposable indicating similar degree of flexibility for the two enzymes and lending support to Somero's "corresponding states" concept. (73).

b) Low concentrations of denaturants such as guanidium hydrochloride increase the activity of hyperthermophilic enzymes at suboptimal temperatures (32, 64).

c) ^{15}N relaxation dispersion experiments on the thermophilic *Aquifex aeolicus* adenylate kinase (AAAK) and the mesophilic *E. coli* adenylate kinase (ECAK) show that the lid-opening rate is the rate-limiting step in catalysis in the two enzymes. AAAK has a much slower lid-opening rate at 20 °C than ECAK. This slow opening rate accounts fully for difference in the k_{cat} values between the two enzymes at 20 °C. This difference in the lid-opening rates between AAAK and ECAK suggests the possibility that the closed conformation in much more stabilized relative to the open conformation in AAAK than in ECAK due to more intramolecular interactions in AAAK (68).

Figure I-2 H-D exchange data for *E. coli* and *Thermus thermophilus* IPMDHs at 25 °C is shown in the upper panel. *E. coli* IPMDH, pD 7.15 (Δ) and 8.15 (□); *Th. thermophilus* IPMDH, pD 7.15 (∇) and 8.15 (○). X represents the percentage of unexchanged peptide hydrogens, t is the time, and k_0 is the chemical exchange rate constant. X at different tk_0 values is higher for the hyperthermophilic enzyme compared to that of the mesophilic IPMDH indicating greater structural rigidity for the hyperthermophilic enzyme. The lower panel shows the exchange data for the *E. coli* and *Th. Thermophilus* IPMDHs' respective temperature optima. *E. coli* IPMDH, 48°C, pD 7.15 (Δ) and 8.15 (□); *Th. thermophilus* IPMDH, 70°C, pD 7.15 (∇) and 8.15 (○). The exchange curves are similar indicating similar levels of flexibility for the mesophilic and hyperthermophilic enzymes. (Image reproduced from (73), copyright (1998) National Academy of Sciences, USA)


X
1
0.8
0.6
0.4
0.2
0
25 °C
Th. thermophilus IPMDH
E. coli IPMDH
ρ=10^-3
10^-4
10^-5
10^-6
2
3
4
5
6
7
X
1
0.8
0.6
0.4
0.2
0
ρ=10^-4
10^-5
10^-6
10^-7
Th. thermophilus IPMDH, 70°C
E. coli IPMDH, 48°C
4
5
6
7
8
log(k0t)

Evidence against the hypothesis

a) Using H-D exchange rate measurements in *Pyrococcus furiousus* rubredoxin, Hernández et al showed that (i) at 28°C, conformational opening for solvent access occurred in the millisecond timescale for the entire protein. This timescale corresponds to the turnover rate of many enzymes; (ii) distribution of amide protection factors were typical of those observed for a large number of mesophilic proteins; and (iii) all native hydrogen bonds involving amide hydrogens were disrupted within 1 s close to its temperature of maximal thermodynamic stability (23).

b) ^{15}N relaxation data on *E. coli* and *T. thermophilus* RNase H (EcRNase and TtRNase, respectively) show that the difference in mobility across multiple timescales is complex: backbone order parameters show that certain regions of the mesophilic EcRNase are more rigid on sub-nanosecond timescales, while other regions are more flexible. In contrast, the thermophilic TtRNase has a higher activation barrier for structural fluctuations in the substrate-binding handle region in the micro-seconds timescale (7).

Conflicting reports such as the above mentioned examples reflect the complex nature of the energy landscape of enzymes. A correlation between rigidity required for protein stability and flexibility required for activity does not have to exist per se. Moreover, there is no single measure of flexibility, since proteins exhibit motions ranging from picosecond (ps) to nanosecond (ns) timescales (36). Flexibility implies increased conformational entropy in the folded state, and should therefore lead to thermodynamic

stability of the folded state. Besides, to satisfy requirements for activity and stability imposed by their high physiological temperatures, hyperthermophilic enzymes may partition conformational flexibility between the active site and the rest of the protein differently from what occurs in mesophilic enzymes. Proteins show individual features of adaptation and more examples need to be studied to draw general conclusions.

The motional properties of a number of hyperthermophilic-mesophilic enzyme pairs have been compared using a variety of techniques, including neutron scattering (16, 17), tryptophan fluorescence (18), H-D exchange (24, 27, 73), and more recently NMR (7, 23) and MD simulations (11, 20, 36). Many of these techniques measure only the average global dynamics in a protein. Few proteins have been characterized on multiple time scales using site-specific, rather than global, measures of flexibility. NMR and molecular dynamics (MD) simulations are two techniques that are uniquely suited to report site-specific protein motions on multiple timescales. Both of these techniques are ideal tools to obtain a deeper understanding of the relationship between stability, activity, and flexibility in hyperthermophilic proteins.

1.6 Protein dynamics from NMR

The principal import of the corresponding states hypothesis is that stabilizing factors in hyperthermophilic proteins are associated with a decrease in structural flexibility. Conformational flexibility not only impacts protein stability, it also plays a significant role in catalytic activity. Therefore, a reasonable balance between rigidity and flexibility

of the protein structure is a key element for thermal adaptation. While a number of experimental techniques are available to probe motional properties of proteins, NMR spectroscopy has become popular because of the range of motional timescales that are accessible and the spatial resolution offered by isotopic labeling of proteins. NMR techniques for studying protein dynamics span more than 12 orders of magnitude in timescale, as seen in Figure I-3. Of relevance to this thesis are ^{15}N NMR relaxation and H-D exchange measurements to probe protein dynamics. An overview of these two techniques is presented in the following sections.

1.6.1 ^{15}N NMR relaxation

If a protein sample containing magnetically active nuclei (i.e., ^{1}H, ^{15}N, and/or ^{13}C) is placed in a static magnetic field, a net magnetization will be induced within the sample. The induction is not instantaneous; instead it accumulates at an exponential rate. If we allow such a magnetization to establish itself, and then apply a radio frequency pulse to disturb the equilibrium magnetization, we find that the system will relax back to the original state at the same exponential rate as the induction. The NMR relaxation process is therefore the propensity of the magnetic moment to re-establish its equilibrium polarization subsequent to its perturbation by an external pulse. There are two types of relaxation processes in NMR. The return of the z magnetization component to equilibrium (Boltzmann distribution) after perturbation is described by the time constant T_1 (i.e., longitudinal relaxation). The return of the xy magnetization component (due to loss of coherence or dephasing) is described by the time constant T_2 (i.e., transverse

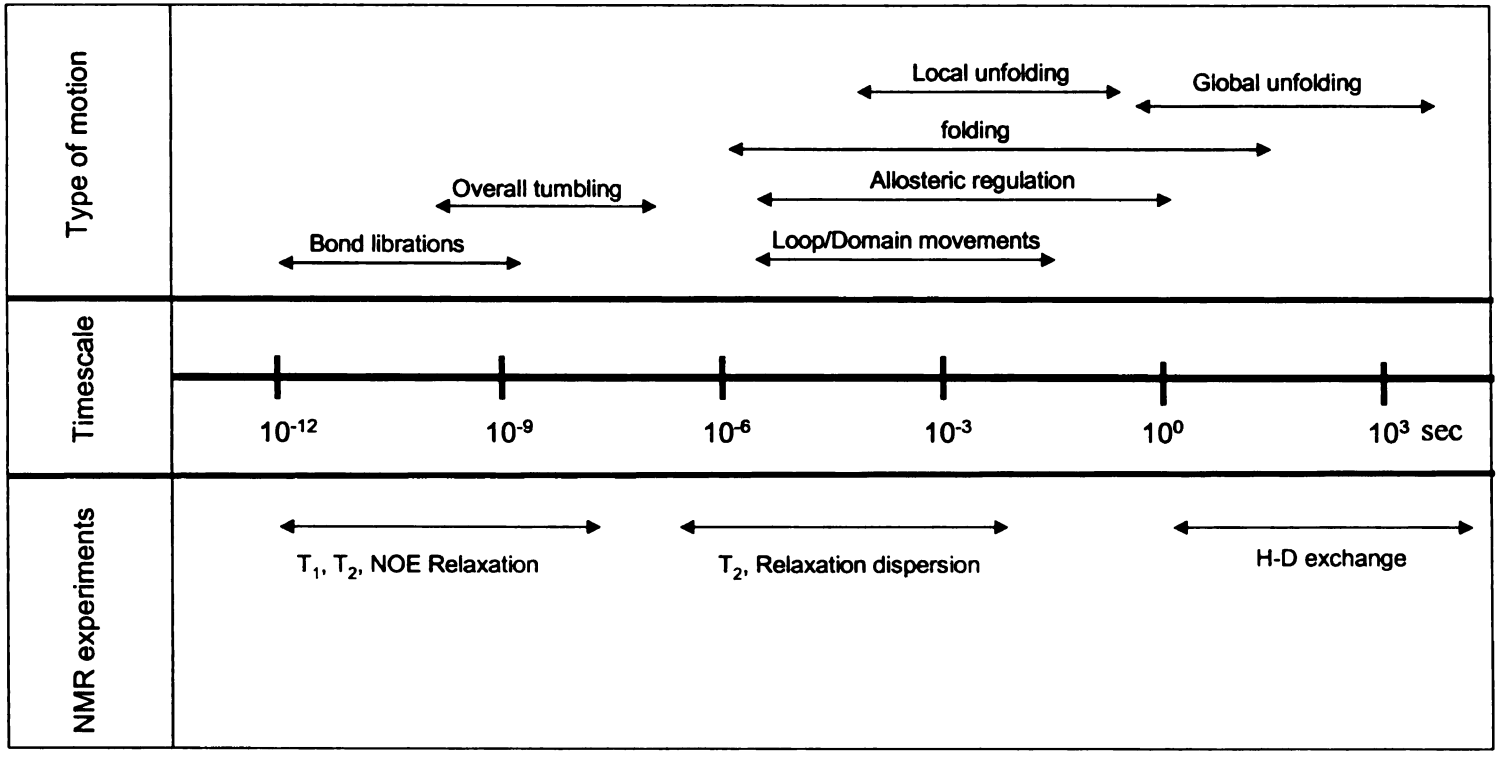


Figure I-3 The top panel in the figure shows the typical timescale of molecular motions. The bottom panel shows the NMR experiments that are sensitive to motions in different time regimes.

relaxation). The relaxation rates are defined by the fluctuating magnetic fields at the site of nuclear spins, the fluctuation being caused by internal motions in the protein.

A majority of studies of protein dynamics through NMR have been performed using the amide ^{15}N nucleus to probe backbone dynamics. The ^{15}N probe has a number of advantages: (i) ^{15}N-enriched protein samples are easy to produce; (ii) resolution offered by 2D heteronuclear correlation spectroscopy is high; (iii) procedures for reliable measurement of relaxation rates are well established; and (iv) interpretation of the relaxation data of a ^{15}N-enriched protein is made easy because relaxation of the ^{15}N-nucleus is primarily determined by dipolar interactions with the only other spin-½ nuclei, the directly bonded protons. All other ^{1}H nuclei will be too far away to make any significant contribution. Contrast this to ^{13}C uniform enrichment wherein most ^{13}C nuclei will be directly bonded to two other magnetically active C nuclei. The presence of additional spin-½ nuclei adds substantial complexity to the relaxation kinetics. Nevertheless, recent studies have made use of other backbone probes such as $^{13}C_\alpha$ (15, 70) and carbonyl spins (12, 14, 75) to obtain a more complete picture of protein backbone dynamics.

The extraction of motional parameters from ^{15}N NMR relaxation data involves two steps. First, the appropriate pulse sequences are chosen to accurately measure R_1 ($1/T_1$), R_2 ($1/T_2$), and the nuclear Overhauser effect (NOEs), which are related to the heteronuclear cross-relaxation rates. Secondly, an analytical method is employed to relate the experimental relaxation rates to the dynamics of the protein. The measured ^{15}N NMR relaxation rates are directly related to the rotational fluctuations of the ^{15}N-^{1}H bond vector with respect to an external magnetic field, B_0. The fluctuations of each NH vector

are described by the autocorrelation function, C(t), which describes how quickly the orientation of the NH dipoles change as a function of time, t. As t increases, the NH vector undergoes motions such that C(t) decreases. The manner in which C(t) decays is described by the spectral density function, $J(\omega)$, of the NH fluctuations. J(ω), the Fourier cosine transform of C(t) (65), gives the relative contribution that frequencies in the range of (ω, ω + dω) make towards C(t) for each NH vector. Essentially, it is J(ω) that gives us information about the NH vector dynamics. The orientational fluctuations of the NH vectors, due to the overall tumbling of the protein and internal motions, create time-varying fluctuating magnetic fields at each NH site, which lead to the relaxation process. Since these fields are driven by internal motions, the frequency distribution in these fields is what is described by J(ω). Since the form of J(ω) is dictated by the NH bond dynamics, the central aim of heteronuclear relaxation studies is to characterize the shapes of the individual spectral density functions. The relaxation rates are linear combinations of the spectral densities at 0, ω_H, ω_N, $\omega_H + \omega_N$, and $\omega_H - \omega_N$ (1). They are given by:

$$R_1 = \frac{d^2}{4}[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N) + c^2 J(\omega_N)] \quad (3)$$

$$R_2 = \frac{d^2}{8}[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H - \omega_N)] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)] + R_{ex} \quad (4)$$

$$NOE = 1 + (\frac{d^2}{4R_1})(\frac{\gamma_H}{\gamma_N})[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (5)$$

where $d = d = (\frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2})\langle r_{NH}^{-3} \rangle$, $c = \frac{\omega_N \sigma \Delta}{\sqrt{3}}$, μ_0 is the permeability of free space; h is the Planck's constant; γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N, respectively; r_{NH} = 1.02 Å is the mean NH bond length; and ω_H and ω_N are the Larmor frequencies of ^1H and ^{15}N, respectively. The phenomenological R_{ex} term in Eq. 4 represents conformational exchange and pseudo-first-order processes occurring on the microsecond (μs) to millisecond (ms) time scale.

Relaxation data analysis proceeds by directly analyzing the measured relaxation rate constants or by analyzing J(ω) values determined from the relaxation rate constants by spectral density mapping. In either case, determination of J(ω) proceeds by using specific physical models or by using "model-free" functional forms containing a limited number of parameters. There are many reasonable physical models for the motion of bond vectors in proteins, most of them decomposing the motions into global and local. The global motion can be isotropic or anisotropic, anisotropic motion being further categorized as axially symmetric or fully anisotropic. The models invoked for local motions are either jumping type of motions between specific sites or restricted wobbling in a cone type of motion. The latter model is the most widely used. Relaxation data are most commonly analyzed by using the model-free formalism of Lipari and Szabo (40, 41). In the model-free formalism, J(ω) for a single timescale is modeled as:

$$J(\omega) = \frac{2}{5}\{[\frac{S^2 \tau_m}{1+\omega^2 {\tau_m}^2}] + [\frac{(1-S^2)\tau}{1+\omega^2 \tau^2}]\} \qquad (6)$$

with $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$; where τ_m is the effective correlation time for global motion and τ_e is

the effective correlation time for the internal motion. For internal motions on two timescales, $J(\omega)$ is given by:

$$J(\omega) = \frac{2}{5}\{[\frac{S^2\tau_m}{1+\omega^2\tau_m{}^2}] + [\frac{1-S_f{}^2)\tau'_f}{1+\omega^2\tau'_f{}^2}] + [\frac{(S_f{}^2 - S^2)\tau'_s}{1+\omega^2\tau'_s{}^2}]\} \qquad (7)$$

where $\tau'_f = \tau_f\tau_m /(\tau_f + \tau_m)$; $\tau'_s = \tau_s\tau_m /(\tau_s + \tau_m)$; τ_f is the effective correlation time on the fast timescale ($\tau_f < 100$ ps); τ_s is the effective correlation time on the slow timescale ($\tau_f < \tau_s < \tau_m$); $S^2 = S_f^2S_s^2$, is the square of the order parameter characterizing the amplitude of internal motions ($0 \leq S^2 \leq 1$); and S_s^2 and S_f^2 are squares of the order parameters for internal motions on the slow and fast timescales, respectively . Motions characterized by S^2 are referred to as ps-ns timescale motions, those associated with S_f^2 are referred to as sub-ns timescale, while motions characterized by S_s^2 are in the ns time scale. As stated in the paper by Lipari and Szabo (40, 41), S^2 is a model-independent measure of the degree of the NH bond vector spatial restriction, in that the exact nature of the motion experienced by each NH bond vector does not have to be identified. On the other hand, τ_e depends on both the rate and amplitude of internal motions and it requires a physical model, such as diffusion in a cone model, for its interpretation. In summary, the model-free formalism supplies a set of motional parameters that are optimized in a non-linear least squares fitting procedure to reproduce the experimental relaxation data.

In the model-free formalism described above, the spectral density function at five frequencies (0, ω_H, ω_N, ω_H - ω_N, ω_H + ω_N) is being evaluated using only three experimental observables: T_1, T_2, and NOE, at a single magnetic field. Hence, Lipari and

Szabo made some simplifying assumptions about NH motions without the use of an explicit motional model. They assumed that the auto-correlation function, C(t), can be described as a sum of exponential decays, and that global and local motions are totally uncoupled. It is argued, with merit, that, since the information supplied by three relaxation parameters is not sufficient to accurately specify J(ω), the errors in the motional interpretations stem not only from experimental errors but also from potentially incorrect assumptions. In an attempt to do away with the assumptions in characterizing J(ω), an alternate method called spectral density mapping has been proposed (45, 46). This method increases the experimental sampling of J(ω) by measuring the relaxation data at multiple fields, thereby providing a clearer view of the shape of J(ω). Although spectral density mapping has been used in a number of studies, the model-free approach continues to dominate the current thinking in NMR relaxation data analysis. Work reported in this thesis uses the model-free analysis of relaxation data acquired at a single magnetic field. The reason for this is two-fold: (i) during the course of this work, an NMR spectrometer at 600 MHz was the only one available. (ii) Relaxation data reported in chapter 3 of this thesis require comparison with previously published work that used the same method. The use of model-free analysis is justified here due to the qualitative and comparative nature of the account.

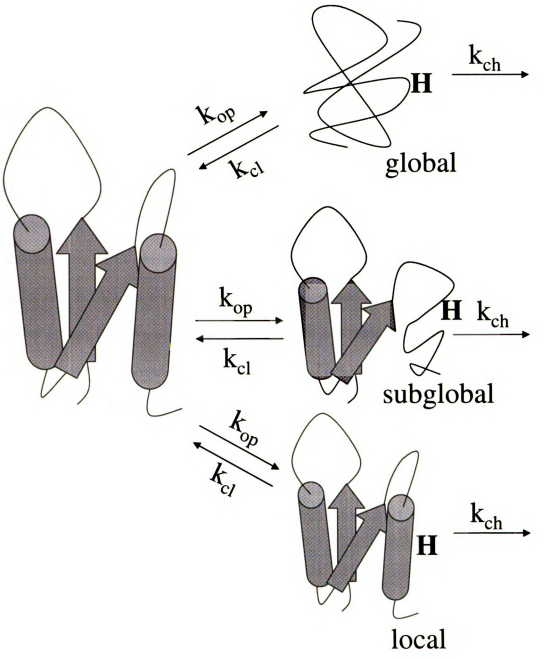
1.6.2 H-D exchange, protein stability, and flexibility

While ^{15}N NMR relaxation experiments provide information on the ps-ns and μs timescales, measurements of amide proton exchange can provide information on global

stability as well as on local conformational fluctuations that occur at timescales greater than milliseconds. Since amide NH groups are distributed uniformly throughout a protein, access to their exchange behavior can, in principle, provide detailed residue-level information on structure change, dynamics, and energetics of proteins. For many years, however, this level of resolution was not realized since the rate measurements relied on techniques that measure H-D rates in a summed, structurally unresolved way. Measurement of H-D exchange rates through NMR has made accessible the full power of site-resolved H-D exchange. In a typical H-D exchange experiment, ^{15}N-labeled protein is expressed and purified under conditions favoring the native state. A 2D hetero-nuclear single quantum correlation (HSQC) spectrum of the protein in undeuterated ($^{1}H_2O$) buffer is recorded, and the observed chemical shifts are assigned to specific backbone amides. An HSQC spectrum is then recorded for the protein in deuterated (D_2O) buffer. Several HSQC spectra are recorded at definite time intervals during which the protein is allowed to exchange with D_2O in the buffer. Because deuterated amide hydrogens will not produce a signal in HSQC, exchange is observed as decay in signal intensity for each amide proton as time progresses. An exponential fit is made to the signal intensity decay of the HSQC spectra, and the exchange rate for each amide proton is computed.

Under native conditions a protein will experience dynamic fluctuations that range from local breathing motions to global unfolding events. To exchange, protected amides must be exposed to the solvent by either solvent penetration into the protein interior or by transient opening that exposes the amide to the exterior. Potential opening events include fluctuations that expose single amides, cooperative partial unfolding, as well as complete unfolding. Figure I-4 illustrates the unfolding model for protein H-D exchange proposed

Figure I-4 The unfolding model for hydrogen exchange in a hypothetical protein. Three types of structural changes that cause transient separation of H-bond donor and acceptor groups and solvent accessibility can lead to hydrogen exchange: local fluctuations (bottom), partial unfolding (center), and complete unfolding (top). k_{cl} and k_{op} are the closing and opening rates, respectively, and k_{ch} represents the chemical exchange rate.


H
k_{ch}
k_{op}
k_{cl}
global
k_{op}
k_{cl}
H
k_{ch}
subglobal
k_{op}
k_{cl}
H
k_{ch}
local

by Linderstrøm-Lang (who first conceived the H-D exchange approach) and his colleagues (26). In this view, hydrogens involved in hydrogen bonds exchange during a small fraction of time when the H-bond donor and acceptor are transiently separated in a dynamic opening-closing reaction or breathing movements. These movements have the effect of exposing the amide hydrogens to the solvent during which the peptide group is attacked by either an OH^- or H_3O^+ ion. Local breathing motions can arise as a result of H-bonds breaking individually, as in the local pathway, in the absence of strong bond forces or as a result of allosteric or functional dynamics in the protein. Figure I-4 also considers large, more cooperative opening modes for global unfolding which can be measured by H-D exchange even in conditions favoring the native state. Proteins molecules in solution are required by thermodynamic principles to occupy all possible higher energy (relative to the native state) states according to the Boltzmann distribution and, over time, to cycle through these states. Thus, the protein molecule unfolds and refolds even under native conditions. This behavior is invisible to most techniques one can think of since they are swamped by signals from the majority population of the native state. H-D exchange is unique because the native state does not contribute significantly to the measured exchange rates. Instead, measurable exchange rates are produced by the cycling of protein molecules through high energy states.

The general scheme for amide hydrogen exchange with solvent as proposed by Linderstrøm-Lang (39) is:

$$Closed \underset{k_{cl}}{\overset{k_{op}}{\Leftrightarrow}} Open \xrightarrow{k_{ch}} Exchanged \qquad (8)$$

In the above scheme, closed represents the exchange-incompetent state, and open represents the open exchange-competent state with k_{op} and k_{cl} representing opening and closing rates, respectively. Once in the exchange-competent state, exchange of the amide proton with the solvent occurs with an exchange rate, k_{ch}. K_{ch} values, also called Bai factors, are typically measured using short model peptides, and they can be easily predicted from the protein sequence (2). In folded proteins, observed exchange rates are often much smaller than the predicted Bai factors. The extent of retardation of the exchange rate is expressed as the protection factor, P, the ratio of the random coil exchange rate constant k_{ch} to the observed rate constant k_{ex}. From Eq. (8), the observed rate constant, k_{ex}, can be expressed as:

$$k_{ex} = \frac{k_{op} k_{ch}}{k_{cl} + k_{ch}} \tag{9}$$

Depending on the relative rates of closure (k_{cl}) and chemical exchange (k_{ch}), Eq. (9) can be further simplified. EX1 type exchange occurs when $k_{ch} >> k_{cl}$ (the rate-determining step is the opening), and $k_{ex} = k_{op}$. EX1 is rarely observed in proteins under native conditions. In the other limiting case, $k_{ch} << k_{cl}$, opening events rarely result in exchange. This case is called the EX2 type exchange where Eq. 8 simplifies to:

$$k_{ex} = \frac{k_{op}}{k_{cl}} \times k_{ch} = K_{op} k_{ch} \tag{10}$$

where K_{op} is the equilibrium constant for opening and closing.

Based on the above equations, H-D exchange can report on the equilibrium (EX2) and kinetic (EX1) features of the protein. In the native folded protein, the protection

factor, P, can range from 1 to 10^{10} or even higher. Residues with small P values are generally associated with local breathing movements. Residues with large P are the slowest exchanging residues and are associated with global unfolding. Therefore, comparing the P values of backbone amides in homologous mesophilic-hyperthermophilic proteins can give per-residue information on the relative conformational stability of the two proteins.

1.7 Work presented in this thesis

The motivation for this thesis has been to verify the validity of the hypothesis that enhanced thermostability in proteins implies enhanced structural rigidity. Since many hyperthermophilic enzymes have reduced or no activity at temperatures far below their physiological growth temperatures (i.e., 20°C-40°C), it is thought that their reduced activity is a consequence of their enhanced structural rigidity. The emphasis of this project has been to use techniques such as NMR and molecular dynamics (MD) simulations that enable site-specific measures of flexibility and can access a wide time scale range, to answer the following questions that stem from the hypothesis:

1. Is protein rigidity at low temperatures a consequence of stability at high temperature?

2. Is reduced activity in hyperthermophilic enzymes a consequence of increased structural rigidity?

In this work, I describe studies investigating the relationship between dynamics, activity, and stability using adenylate kinase (AK) from the mesophile *Escherichia coli* (ECAK) and the hyperthermophile *Thermotoga neapolitana* (TNAK) as models. Adenylate Kinase was chosen as the model enzyme because of the following criteria: (i) it is a well studied enzyme and AKs from a number of organisms have been biochemically characterized; (ii) many crystal structures of AKs from several organisms particularly *E. coli* are known, both free form and in complex with various substrates and inhibitors (63); (iii) AKs are small monomeric enzymes, which make them accessible to NMR and MD simulations; (iv) AKs have two "lids" that close upon the substrates ATP and AMP to catalyze phosphoryl transfer to produce ADP. Hence, AKs undergo large conformational changes during substrate binding and product release; and (v) backbone dynamics of ECAK measured using ^{15}N NMR relaxation have been reported (50, 51, 58) which facilitates comparative studies of dynamics between ECAK and TNAK. As a first step in the project, TNAK was cloned, expressed, and its biochemical properties were investigated. Chapter 2 presents the results of biochemical characterization of the hyperthermophilic TNAK. The results show that while TNAK is highly stable, as is expected of a hyperthermophilic protein, it is as active as ECAK at 30 °C. This property of TNAK is unusual for a hyperthermophilic enzyme, and it makes TNAK a particularly interesting example for this project. TNAK and ECAK, therefore, make a unique hyperthermophilic-mesophilic enzyme pair, unlike other examples studied so far, to investigate the relationship between dynamics and stability and dynamics and catalysis.

Chapter 3 presents results from ^{15}N NMR relaxation studies on TNAK and the comparison of these results to those of ECAK. ^{15}N NMR relaxation studies are used as a

probe of backbone dynamics in the ps to ns as well as in the μs timescales. The dynamics of the two enzymes are correlated to the observed differences in stability and activity. Chapter 4 presents the results from H-D exchange studies on TNAK and ECAK that report on the dynamics in the minutes to day's timescale. These studies give insight into the structural distribution of stability in ECAK and TNAK. Chapter 5 presents the results from MD simulation studies of ECAK in complex with its substrates ATP and AMP. The results highlight important substrate-protein interactions that explain the observed substrate specificity. Interactions between charged residues in the binding site, the Mg^{2+} cation, and water molecules maintain the geometry and distances of the AMP α-phosphate and ATP β- and γ-phosphates in such a way as to support an associative reaction mechanism for phosphoryl transfer. Results from ECAK MD simulations are presented here as a starting point for further investigations comparing ECAK and TNAK using this technique. A summary and perspective of results is given in chapter 6. Potential future directions are also discussed, including MD simulations that will provide complementary information on the structural and energetic bases of TNAK thermostability.

References

1. **Abragam, A.** 1961. Principles of nuclear magnetism. Oxford University Press, London.

2. **Bai, Y. W., J. S. Milne, L. Mayne, and S. W. Englander.** 1993. Primary structure effects on peptide group hydrogen-exchange. Proteins - Struct. Funct. Genet. **17:**75-86.

3. **Berezovsky, I. N., E. I. Shakhnovich.** 2005. Physics and evolution of thermophilic adaptation. Proc. Nat. Acad. Sci. USA. **102:**12742-12747.

4. **Blochl, E., R. Rachel, S. Burggraf, D. Hafenbradl, H. W. Jannasch, and K. O. Stetter**. 1997. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 °C. Extremophiles. **1**:14-21.

5. **Brock TD, B. K., Belly RT, Weiss RL.** 1972. *Sulfolobus* - New genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Microbiol. **84**:54–68.

6. **Burrgraf, S., G. J. Olsen, K. O. Stetter, and C. R. Woese.** 1992. A phylogenetic analysis of *Aquifex pyrophilus*. Syst. Appl. Microbiol. **15:**352-356.

7. **Butterwick, J. A., J. P. Loria, N. S. Astrof, C. D. Kroenke, R. Cole, M. Rance, and A. G. Palmer.** 2004. Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. J. Mol. Biol. **339:**855-871.

8. **Cavagnero, S., D. A. Debe, Z. H. Zhou, M. W. Adams, and S. I. Chan.** 1998. Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry **37:**3369–3376.

9. **Cavagnero, S., Z. H. Zhou, M. W. W. Adams, and S. I. Chan.** 1998. Unfolding mechanism of rubredoxin from *Pyrococcus furiosus*. Biochemistry **37:**3377–3385.

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

11. **Colombo, G., and K. M. Merz.** 1999. Stability and activity of mesophilic subtilisin E and its thermophilic homolog: Insights from molecular dynamics simulations. J. Am. Chem. Soc. **121:**6895-6903.

12. **Cordier, F., B. Brutscher, and D. Marion.** 1996. Measurement of $^{13}C_{\alpha}$-(CO)-^{13}C cross-relaxation rates in ^{15}N-(IC)-^{13}C-labelled proteins. J. Biomol. NMR **7:**163-168.

13. **Davies, G. J., S. J. Gamblin, J. A. Littlechild, and H. C. Watson.** 1993. The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. Proteins: Struct. Funct. Genet. **15:**283-289.

14. **Dayie, K. T., G. Wagner.** 1997. Carbonyl carbon probe of local mobility in ^{13}C, ^{15}N-enriched proteins using high-resolution nuclear magnetic resonance. J. Am. Chem. Soc. **119:**7797-7806.

15. **Engelke, J., H. Ruterjans.** 1995. Determination of $^{13}C_{\alpha}$ relaxation times in uniformly $^{13}C/^{15}N$-enriched proteins. J. Biomol. NMR **5:**173-182.

16. **Fitter, J., R. Herrmann, N. Dencher, and A. Blume.** 2002. The role of protein dynamics for thermal adaptation of mesophilic and thermostable α-amylase. Biophys. J. **82:**304A-304A.

17. **Fitter, J., R. Herrmann, N. A. Dencher, A. Blume, and T. Hauss.** 2001. Activity and stability of a thermostable α-amylase compared to its mesophilic homologue: Mechanisms of thermal adaptation. Biochemistry **40:**10723-10731.

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

19. **Grattinger, M., A. Dankesreiter, H. Schurig, and R. Jaenicke.** 1998. Recombinant phosphoglycerate kinase from the hyperthermophilic bacterium *Thermotoga maritima:* catalytic, spectral and thermodynamic properties. J. Mol. Biol. **280:**525-**533**.

20. **Grottesi, A., M. A. Ceruso, A. Colosimo, and A. Di Nola.** 2002. Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. Proteins: Struct. Funct. Genet. **46:**287–294.

21. **Haney, P., J. Konisky, K. K. Koretke, Z. Luthey-Schulten, and P. G. Wolynes.** 1997. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*. Proteins **28:**117–130.

22. **Hendsch, Z. S., and B. Tidor.** 1994. Do salt bridges stabilize proteins - A continuum electrostatic analysis. Protein Sci. **3:**211-**226**.

23. **Hernandez, G., F. E. Jenney, Jr., M. W. W. Adams, and D. M. LeMaster.** 2000. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. Proc. Nat. Acad. Sci. USA **97:**3166-3170.

24. **Hollien, J., and S. Marqusee.** 1999. Structural distribution of stability in a thermophilic enzyme. Proc. Nat. Acad. Sci. USA **96:**13674-13678.

25. **Hollien, J., and S. Marqusee.** 1999. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. Biochemistry **38:**3831-3836.

26. **Hvidt, A., J. G., Linderstrøm-Lang K.** 1960. Hydrogen exchange in proteins. Ad. Prot. Chem. **21:**287-386.

27. **Fitter, J., and J. Heberle.** 2000. Structural equilibrium fluctuations in mesophilic and thermophilic α-amylase. Biophys. J. **79:**1629–1636.

28. **Jaenicke, R.** 1996. Glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: strategies of protein stabilization. FEMS Microbiol. Rev. **18:**215-224.

29. **Karshikoff, A., and R. Ladenstein.** 1998. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Engin. **11:**867-872.

30. **Kashefi, K., and D. R. Lovley.** 2003. Extending the Upper Temperature Limit for Life. Science **301:**934-934.

31. **Kawamura, S., Y. Kakuta, I. Tanaka, K. Hikichi, S. Kuhara, N. Yamasaki, and M. Kimura.** 1996. Glycine-15 in the bend between two α-helices can explain the thermostability of DNA binding protein HU from *Bacillus stearothermophilus*. Biochemistry **35:**1195-1200.

32. **Kujo, C., and T. Ohshima.** 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.

33. **Kumar, S., and R. Nussinov.** 2001. How do thermophilic proteins deal with heat? Cell. Mol. Life Sci. **58:**1216-1233.

34. **Kumar, S., C. J. Tsai, B. Y. Ma, and R. Nussinov.** 2000. Contribution of salt bridges toward protein thermostability. J. Biomol. Struct. Dyn.:79-85.

35. **Kumar, S., C. J. Tsai, and R. Nussinov.** 2000. Factors enhancing protein thermostability. Protein Engin. **13:**179-191.

36. **Lazaridis, T., I. Lee, and M. Karplus.** 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci. **6:**2589–2605.

37. **Li, W. T., R. A. Grayling, K. Sandman, S. Edmondson, J. W. Shriver, and J. N. Reeve.** 1998. Thermodynamic stability of archaeal histones. Biochemistry **37:**10563-10572.

38. **Li, Z. G., S. Raychaudhuri, and A. J. Wand.** 1996. Insights into the local residual entropy of proteins provided by NMR relaxation. Protein Sci. **5:**2647-2650.

39. Linderstrøm-Lang, K. 1958. Deuterium exchange and protein structure. In: Neubreger A, (ed.) Symposium on protein structure. London: Methuen.

40. **Lipari, G., and A. Szabo.** 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules .1. Theory and range of validity. J. Am. Chem. Soc. **104:**4546-4559.

41. **Lipari, G., and A. Szabo.** 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules .2. Analysis of experimental results. J. Am. Chem. Soc. **104:**4559-4570.

42. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, L. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima.* Nature **399:**323-329.

43. **Nicholls, A., K. A. Sharp, and B. Honig.** 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Sruct. Funct. Genet. **11:**281-296.

44. **Ogata, Y., E. I. Imai, H. Honda, K. Hatori, and K. Matsuno.** 2000. Hydrothermal circulation of seawater through hot vents and contribution of interface chemistry to prebiotic synthesis. Orig. Life Evol. Biosph. **30:**527-537.

45. **Peng, J. W., and G. Wagner.** 1995. Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. Biochemistry **34:**16733-16752.

46. **Peng, J. W., and G. Wagner.** 1992. Mapping of the spectral densities of N-H bond motions in eglin-c using heteronuclear relaxation experiments. Biochemistry **31:**8571-8586.

47. **Prevost, M., S. J. Wodak, B. Tidor, and M. Karplus.** 1991. Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96-Ala mutation in barnase. Proc. Nat. Acad. Sci. USA **88:**10880–10884.

48. **Russell R. J. M., J. M. C. Ferguson, D. W. Hough, M. J. Danson, G. L. Taylor GL.** 1997. The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. Biochemistry **36:**9983-9994.

49. **Schuler, B., W. H. Kremer, H. R. Kalbitzer, and R. Jaenicke.** 2002. Role of entropy in protein thermostability: Folding kinetics of a hyperthermophilic cold shock protein at high temperatures using ^{19}F NMR. Biochemistry **41:**11670-11680.

50. **Shapiro, Y. E., E. Kahana, V. Tugarinov, Z. C. Liang, J. H. Freed, and E. Meirovitch.** 2002. Domain flexibility in ligand-free and inhibitor-bound *Escherichia coli* adenylate kinase based on a mode-coupling analysis of ^{15}N spin relaxation. Biochemistry **41:**6271-6281.

51. **Shapiro, Y. E., M. A. Sinev, E. V. Sineva, V. Tugarinov, and E. Meirovitch.** 2000. Backbone dynamics of *Escherichia coli* adenylate kinase at the extreme stages of the catalytic cycle studied by ^{15}N NMR relaxation. Biochemistry **39:**6634-6644.

52. **Somero, G. N.** 1978. Temperature adaptation of enzymes: Biological optimization through structure-function compromises. Ann. Rev. Ecol. Systemat. **9:**1-29.

53. **Stetter, K. O.** 1996. Hyperthermophilic procaryotes. FEMS Microbiol. Rev. **18:**149-158.

54. **Stone, M. J., S. Gupta, N. Snyder, and L. Regan.** 2001. Comparison of protein backbone entropy and beta-sheet stability: NMR-derived dynamics of protein G B1 domain mutants. J. Am. Chem. Soc. **123:**185-186.

55. **Szilagyi, A., and P. Zavodszky.** 1995. Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: analysis based on homology modeling. Protein Engin. **8:**779-89.

56. **Szilagyi, A., P. Zavodszky.** 2000. Structural differences between mesophilic, moderately thermophilic, and extremely thermophilic protein subunits: results of a comprehensive survey. Structure **8:**493-504.

57. **Tomschy, A., R. Glockshuber, and R. Jaenicke.** 1993. Functional expression of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* in *Escherichia coli.* Authenticity and kinetic properties of the recombinant enzyme. Eur. J. Biochem. **214:**43-50.

58. **Tugarinov, V., Y. E. Shapiro, Z. Liang, J. H. Freed, and E. Meirovitch.** 2002. A novel view of domain flexibility in *E. coli* adenylate kinase based on structural mode-coupling ^{15}N NMR relaxation. J. Mol. Biol. **315:**155–170.

59. **Usher, K. C., A. F. A. De la Cruz, F. W. Dahlquist, R. V. Swanson, M. I. Simon, and S. J. Remington.** 1998. Crystal structures of CheY from *Thermotoga maritima* do not support conventional explanations for the structural basis of enhanced thermostability. Protein Sci. **7:**403-412.

60. **Vieille, C., J. M. Hess, R. M. Kelly, and J. G. Zeikus.** 1995. *xylA* cloning and sequencing and biochemical characterization of xylose isomerase from *Thermotoga neapolitana*. Appl. Environ. Microbiol. **61:**1867–1875.

61. **Vieille, C., and J. G. Zeikus.** 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. Rev. **65:1–43.**

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

63. **Vonrhein, C., G. J. Schlauderer, and G. E. Schulz.** 1995. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure **3:**483-490.

64. **Wilquet, V., J. A. Gaspar, M. van de Lande, M. Van de Casteele, C. Legrain, E. M. Meiering, N. Glansdorff.** 1998. Purification and characterization of recombinant *Thermotoga maritima* dihydrofolate reductase. Eur. J. Biochem. **255:**628-37.

65. **Wittebort, R. J., and A. Szabo.** 1978. Theory of NMR relaxation in macromolecules - restricted diffusion and jump models for multiple internal rotations in amino-acid side-chains. J. Chem. Phys. **69:**1722-1736.

66. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221-271.

67. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms - proposal for the domains Archaea, bacteria, and eucarya. Proc. Nat. Acad. Sci. USA **87:**4576-4579.

68. **Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern.** 2004. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat. Struc. Mol. Biol. **11:**945-949.

69. **Xiao, L., and B. Honig.** 1999. Electrostatic contributions to the stability of hyperthermophilic proteins. J. Mol. Biol. **289:**1435-44.

70. **Yamazaki, T., R. Muhandiram, and L. E. Kay.** 1994. NMR experiments for the measurement of carbon relaxation properties in highly enriched, uniformly ^{13}C, ^{15}N-labeled proteins - Application to $^{13}C_{\alpha}$ carbons. J. Am. Chem. Soc. **116:**8266-8278.

71. **Yang, D. W., and L. E. Kay.** 1996. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. J. Mol. Biol. **263:**369-382.

72. **Yip, K. S. P., T. J. Stillman, K. L. Britton, P. J. Artymiuk, P. J. Baker, S. E. Sedelnikova, P. C. Engel, A. Pasquo, R. Chiaraluce, V. Consalvi, R.**

Scandurra, and D. W. Rice. 1995. The structure of *Pyrococcus fFuriosus* glutamate-dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure **3:**1147-1158.

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

74. **Zeikus, J. G., B. T. D. Brock**.1971. Protein synthesis at high temperatures: aminoacylation of tRNA. Biochim. Biophys. Acta. **228:736-745**.

75. **Zeng, L., M. W. F. Fischer, and E. R. P. Zuiderweg.** 1996. Study of protein dynamics in solution by measurement of $^{13}C_{\alpha}$-(CO)-^{13}C NOE and (CO)-^{13}C longitudinal relaxation. J. Biomol. NMR **7:**157-162.

Chapter 2

Biochemical characterization of *Thermotoga neapolitana* adenylate kinase

Research presented in this chapter was published as the following manuscript: **C. Vieille, H. Krishnamurthy, H. H. Hyun, A. Savchenko, H. Yan, and J. G. Zeikus.** 2003. *Thermotoga neapolitana* adenylate kinase is highly active at 30°C. Biochem. J. **372:**577–585.

Work presented in this chapter was a cooperative effort by C. Vieille, H. Krishnamurthy, H. H. Hyun, and A. Savchenko. The *T. neapolitana* adenylate kinase (TNAK) gene was cloned by A. Savchenko. H. H. Hyun standardized the expression and purification protocols for TNAK. Studies of the effect of temperature on TNAK activity and stability shown in Figure II-2 were also done by H. H. Hyun. C. Vieille performed the sequence analysis of adenylate kinases summarized in Table II-1 and Figure II-1. C. Vieille also carried out the experiments presented in Figure II-5 showing the titration curves for TNAK Zn^{2+} with PMPS. H. Krishnamurthy performed the experiments to determine the effect of pH on TNAK activity and stability (Figure II-3), and the calorimetric and circular dichroism spectroscopy studies of TNAK (Figure II-4 and Table II-2). The kinetic parameters of TNAK at 30 °C (Table 3) and the Zn^{2+} content of TNAK through atomic emission spectroscopy (Table II-4) were also done by H. Krishnamurthy.

2.1 Abstract

The adenylate kinase (AK) gene from *Thermotoga neapolitana*, a hyperthermophilic bacterium, was cloned and overexpressed in *Escherichia coli*, and the recombinant enzyme was biochemically characterized. The *T. neapolitana* AK (TNAK) sequence indicates that this enzyme belongs to the long bacterial AKs. TNAK contains the four cysteine residues that bind Zn^{2+} in all Gram-positive AKs and in a few other Zn^{2+}-containing bacterial AKs. Atomic emission spectroscopy and titration data indicate a content of 1 mol of Zn^{2+}/mol of recombinant TNAK. The EDTA-treated enzyme has a

melting temperature (T_m =93.5 °C) 6.2 °C below that of the holoenzyme (99.7 °C), identifying Zn^{2+} as a stabilizing feature in TNAK. TNAK is a monomeric enzyme with a molecular mass of approximately 25 kDa. TNAK displays V_{max} and K_m values at 30 °C identical with those of the *E. coli* AK at 30 °C, and displays very high activity at 80 °C, with a specific activity above 8000 units/mg. The unusually high activity of TNAK at 30 °C makes it an interesting model to test the role of enzyme flexibility in activity.

2.2 Introduction

Proteins from hyperthermophilic organisms are highly thermostable and are optimally active at high temperatures (often 90 °C and above). Most studies agree that stabilization mechanisms (such as hydrophobic interactions, hydrogen bonds, and salt bridges) vary from one protein to another and that no single molecular mechanism is responsible for protein thermal stabilization. One hypothesis explaining the remarkable thermostability of hyperthermophilic enzymes is that these enzymes have enhanced conformational rigidity at low temperatures. According to this hypothesis, psychrophilic, mesophilic, thermophilic and hyperthermophilic enzymes with sequence similarity have comparable catalytic efficiencies (indicated by k_{cat}/K_m values) at their respective optimal temperatures, because optimal activity requires a certain degree of conformational flexibility at the active site. Owing to their increased rigidity at low temperatures, thermophilic and hyperthermophilic enzymes are only marginally active at these temperatures, and they gain flexibility required for optimal activity only at higher temperatures (1). Recent experimental results (e.g. amide hydrogen/deuterium exchange

and Fourier-transform infrared spectroscopy) and molecular dynamics simulations show that results vary from one protein to another. Some thermophilic enzymes are less flexible than their mesophilic counterparts (2–4), whereas others are equally, if not more, flexible than their mesophilic counterparts (5). The limited amount of experimental results and the type of data available (most of the experimental flexibility data are available only for the protein as a whole, rather than at the residue level) do not allow us to determine clearly whether rigidity at mesophilic temperatures is a key factor in protein thermostability. The time scale and amplitude of the conformational fluctuations that might lead to low thermostability are also not clear.

For these reasons, we decided to compare a new set of similar enzymes in terms of flexibility and thermostability. We chose the monomeric enzyme adenylate kinase (AK) because its small size (25 kDa) renders it accessible to NMR and molecular dynamics studies. AK (EC 2.7.4.3) catalyzes the reaction MgATP + AMP $\Leftrightarrow$ MgADP + ADP. Because it is the only enzyme forming ADP from AMP, this ubiquitous enzyme is essential for the maintenance of the cellular ATP pool, for ATP production and for all macromolecular syntheses (6). Over the past 27 years, numerous studies (e.g. X-ray crystallography, NMR spectroscopy and site-directed mutagenesis) have been performed to characterize the ATP- and AMP-binding sites of AK, to identify its catalytic mechanism and to understand the mechanisms underlying the conformational changes occurring during catalysis (reviewed in [7]).

Four AK families have been described previously (8, 9). The short AKs (family 1) represented by the mammalian cytosolic enzymes, and the long AKs (family 2), represented by the bacterial and mitochondrial enzymes differ by a 20 to 30-amino-acid

residue insertion that is present in the long enzymes. This insertion creates a separate domain that acts as a lid covering the active site (10, 11). Among the bacterial enzymes, two subgroups of AKs can be differentiated by the presence or absence of a structural Zn^{2+} in the lid (11, 12). The four Zn^{2+} ligands (typically four cysteine residues or three cysteine residues plus a carboxylic residue) are conserved in all the Zn^{2+}-containing AKs. (12). The third AK family represented by methanococcal and *Sulfolobus acidocaldarius* AKs does not show significant similarity to the other two AK families, with the exception of the P loop, a signature motif in nucleotide kinases (9,13,14). In contrast with the cytosolic and bacterial AKs, which are monomeric enzymes, methanococcal and sulfolobal AKs are trimeric (14). A fourth AK subfamily was described recently and it accounts at least for the *Mycobacterium capricolum* enzyme (15): this subfamily consists of short bacterial enzymes that are related to the long bacterial AKs more closely when compared with the short mammalian cytosolic AKs.

Despite the fact that monomeric AKs are well-studied small ubiquitous enzymes, and even though the *Escherichia coli* AK (ECAK) unfolds almost completely reversibly (16), monomeric AKs have rarely been studied in terms of thermostabilization mechanisms and in terms of their activity versus stability at high temperatures. Zhang et al. (17, 18) noticed that inactivation of the muscle cytosolic AK always occurred at lower temperatures and at lower urea concentrations compared with other significant conformational changes. These authors suggested that the active site was more flexible than the rest of the enzyme. They also suggested that with increasing temperature the active site reached flexibility levels incompatible with catalysis, while the rest of the enzyme was still folded. The stabilization provided by Zn^{2+} in Zn^{2+}-containing AKs was

studied by site-directed mutagenesis. A mutant ECAK engineered to contain a Zn^{2+} cation is significantly more stable than the wild-type enzyme (19). This mutant ECAK is still less stable than the *Bacillus stearothermophilus* AK (BSAK), indicating that the Zn finger-like structure in BSAK is not its sole stabilizing feature. This observation has been confirmed by two facts: (i) the *B. subtilis* AK, a Zn^{2+}-containing AK, is significantly less stable than BSAK, and (ii) it is not more stable than the wild-type ECAK (20).

No AKs have been characterized from hyperthermophilic bacteria. However, a wealth of AK structure information (e.g. X-ray and NMR data for ECAK and BSAK) is already available, which would facilitate comparative studies between mesophilic and hyperthermophilic bacterial AKs. For these reasons, we decided to characterize *Thermotoga neapolitana* AK (TNAK). In the present study, we show that, with a T_m close to 100 °C, TNAK is indeed a highly thermostable protein. The three enzymes ECAK, BSAK and TNAK thus represent a nice set of similar enzymes whose dynamic properties can be compared at the residue level using NMR and molecular dynamics approaches. With kinetic parameters identical with those of ECAK at 30 °C, TNAK is one of the rare hyperthermophilic enzymes that are highly active at mesophilic temperatures. These intriguing catalytic properties of TNAK will make the comparison of TNAK, ECAK, and BSAK in terms of flexibility, activity and thermostability particularly interesting.

2.3 Materials and Methods

2.3.1 Bacterial strains and plasmids

T. neapolitana strain 5068 [(21) and Deutsche Sammlung von Mikroorganismen und Zellkulturen catalogue, 1993] was used as the source of chromosomal DNA to construct the genomic library. *E. coli* strains XL1-BlueMRF' and XLOLR (Stratagene, La Jolla, CA, U.S.A.) were used as the host and the excision plating strains respectively for the *T. neapolitana* genomic library. *E. coli* strain CV2, which expresses the thermosensitive $Pro^{87} \rightarrow$ Ser AK mutant (22), was used to express the *T. neapolitana* genomic library. Strain CV2 grows at 30 °C, but it cannot grow at 42 °C. Strain DH5α (Life Technologies, Gaithersburg, MD, U.S.A.) was used for subcloning experiments. Strains HB101(DE3) and CV2(DE3) were built using the *ë*DE3 lysogenization kit (Novagen, Madison, WI, U.S.A.). Strain HB101(DE3) was used to overexpress the recombinant TNAK. *E. coli* strains were grown in Luria–Bertani (LB) or superbroth (SB) (23) medium, which contained 100 μg/ml ampicillin or 25 μg/ml kanamycin when necessary. Plasmid pCR2.1 (TA cloning kit; Invitrogen, Carlsbad, CA, U.S.A.) was used to clone PCR products. Plasmid pET23a(+) (Novagen) was the expression vector.

2.3.2 Library construction and screening

T. neapolitana genomic DNA was extracted as described previously (24). This genomic DNA was partially digested with *Sau*3A. *Sau*3A fragments (2–12 kb) were isolated on a 10–40% sucrose gradient and cloned into the *Bam*HI/alkaline phosphatase treated ZAP Express vector using the ZAP Express Predigested Gigapack II Gold Cloning kit

(Stratagene). Mass excision of the library from the phage vector was performed according to the manufacturer's instructions. *E. coli* strain CV2 was transformed with the excised library and transformants were screened for growth at 42 °C.

2.3.3 Manipulation of DNA

Plasmid DNA purification, restriction analysis, subclonings and PCRs were performed using conventional techniques (25, 26). Oligonucleotides used for PCRs and for sequencing were synthesized by the Michigan State University Macromolecular Facility (East Lansing, MI, U.S.A.). DNA was recovered from agarose gels with the Geneclean II kit (BIO 101, La Jolla, CA, U.S.A.). DNA sequences were determined on both strands using the ThermoSequenase radiolabelled terminator cycle sequencing kit (U.S. Biochemical Corp., Cleveland, OH, U.S.A.). Sequencing data were analyzed using version 8 of the Sequence Analysis Software Package of the Genetics Computer Group (University of Wisconsin, Madison, WI, U.S.A.) (27). The GenBank® accession no. for the sequence published in the present paper is AF494055.

2.3.4 Subcloning of the AK gene from *T. neapolitana* into an expression vector

The AK gene from *T. neapolitana* (*adk*) was amplified using plasmid pTNAK1 as the template and oligonucleotides 5′-GCATATGATGGCTTATCTGGTGTTTC-3′ (where sequence CATATG creates an *Nde*I site) and 5′-CGAATTCTTATCATTTATCACTCC-ACCC-3′ (where sequence GAATTC creates an *Eco*RI site) as primers. The amplified *adk* gene was cloned into plasmid pCR2.1, and its sequence was verified. The *adk* gene was then subcloned into *Nde*I and *Eco*RI sites of the vector pET23a(+) to create plasmid

pTNAK2. Since pTNAK2 was not well maintained in *E. coli*, the kanamycin-resistant Genblock (kanR; Amersham Biosciences, Piscataway, NJ, U.S.A.) was cloned into the unique *Eco*RI site of pTNAK2. HB101(DE3)(pTNAK2::kanR) cultures were grown in the presence of kanamycin.

2.3.5 AK purification

Fresh 4 ml HB101(DE3)(pTNAK2::kanR) precultures in LB kanamycin were used to inoculate four 1 litre flasks of SB kanamycin, which contained 1 mM $ZnCl_2$. TNAK production was induced with 1 mM isopropyl-β-D-thiogalactoside when cultures reached an attenuance D_{600} of 1. After a 20 h induction, the cultures were centrifuged at 5000 ***g*** for 10 min and resuspended in 150 ml of 50 mM Tris/HCl, pH 7.4 (buffer A). The bacterial suspension was incubated at 37 °C for 1 h with 0.6 mg/ml lysozyme and a few mg of DNase 1. Cells were then disrupted by two passages through a French pressure cell. The soluble and insoluble fractions were separated by centrifugation at 20 000 ***g*** for 20 min. The insoluble fraction was resuspended in approx. 80 ml of buffer A. Both fractions were heat-treated at 80 °C for 20 min, cooled on ice for 5 min, and centrifuged again at 20 000 *g* for 20 min. Pellets containing heat-denatured proteins were discarded. Supernatants from the heat-treated soluble and insoluble fractions were pooled together, and they were loaded (at a flow rate of 0.8 ml/min) on to an Affi-Gel Blue Gel column (2.6 cm×30 cm; Bio-Rad Laboratories, Hercules, CA, U.S.A.), pre-equilibrated with buffer A. After washing the column with 5 vol. of buffer A, proteins were eluted with a 500 ml linear 0–1 M NaCl gradient in buffer A. Fractions containing AK activity were pooled, dialyzed against three batches of buffer A, then concentrated in an ultrafiltration

cell equipped with a 10 kDa molecular mass cut-off membrane (Amicon, Beverly, MA, U.S.A.). The resulting 20 ml solution was loaded on to a Sephacryl-S100 HR column (2.4 cm×90 cm; Pharmacia, Uppsala, Sweden), pre-equilibrated with buffer A. Elution was performed with buffer A at a flow rate of 50 ml/h. Fractions containing AK activity were pooled, dialyzed against 20 mM Tris/HCl, pH 8.6 (buffer B), and concentrated by ultrafiltration as described above. This sample was loaded on to a DEAE– Sepharose Fast Flow (Amersham Biosciences) anion-exchange column (2.6 cm×30 cm; flow rate 1 ml/min) pre-equilibrated with buffer B. After washing the column with 5 vol. of buffer B, TNAK was eluted with a 500 ml linear 0–1.2 M NaCl gradient in buffer A.

2.3.6 Gel-permeation chromatography

The purified TNAK (0.8 mg in 0.3 ml of buffer A) was loaded on to a Sephadex G-75 column (1.0 cm×40 cm). Elution was performed with buffer A at a flow rate of 0.2 ml/min. The molecular mass markers used were from the Sigma kit for molecular masses in the range 65–66 kDa (catalogue no. MW-GF-70; Sigma, St. Louis, MO, U.S.A.).

2.3.7 AK assays

TNAK activity in the direction of ADP formation was determined by a coupled assay at 38 °C. The reaction mixture (1 ml) consisted of buffer A containing 250 mM KCl, 4 mM $MgCl_2$, 1 mM ATP, 1 mM AMP, 0.5 mM phosphoenolpyruvate, 0.25 mM NADH and 2.5 units each of pyruvate kinase (catalogue no. P-9136; Sigma) and lactate dehydrogenase (L-2500; Sigma). The reaction was started by the addition of TNAK, which was appropriately diluted with buffer A. NADH consumption was followed at 340 nm. At temperatures above 40 °C, TNAK activity was determined by an end-point

method. The reaction mixture (0.8 ml; buffer A containing 250 mM KCl, 4 mM $MgCl_2$, 1 mM ATP and 1 mM AMP) was incubated with TNAK for 2–5 min at the temperature of interest, and then cooled on ice. The reaction mixture (0.6 ml) was then transferred to 0.4 ml of buffer A, containing 250 mM KCl, 4 mM $MgCl_2$, 1.25 mM phosphoenolpyruvate, 0.625 mM NADH, 0.25 mM P^1,P^5-di(adenosine-5′) pentaphosphate ('Ap5A') and 2.5 units of lactate dehydrogenase. After measuring the absorbance at 340 nm, 2.5 units of pyruvate kinase were added and absorbance was measured again.

To determine the kinetic parameters of TNAK, assays were performed in the presence of 0.05–2 mM ATP or 0.025–1 mM AMP, using the end-point method described above. One unit of enzyme corresponds to 1 μmol of product formed per min under the given conditions.

To determine the effect of pH on TNAK activity, enzyme activity was measured in 50 mM sodium acetate (pH 3.7–5.5), sodium phosphate (pK_{a2}) (pH 6.0–7.0), Tris (pH 6.5–8.0), glycine (pH 8.5–10.5) and sodium phosphate (pK_{a3}) (pH 11.0–12.0) at 40 °C using the end-point method. All pH values were adjusted at room temperature (25 °C), and the $\Delta pK_a/\Delta t$ for the different buffers (28) were considered to calculate the pH values at 40 °C. The assays were performed at a constant ionic strength of 220 mM. The amount of KCl added to adjust the ionic strength I was based on the equation $I = \Sigma(Z_i^2[C_i])$, where Z_i is the number of charges of one of the ion and $[C_i]$ its concentration.

2.3.8 Kinetic stability assays

The kinetic thermostability of TNAK was determined by incubating the enzyme (at protein concentrations of 0.55 mg/ml and 5.5 μg/ml) in 0.1 M sodium phosphate (pH 7.5) at 75, 80, 90, 95, 100, 105, and 110 °C for different time periods. The stability of TNAK at 5.5 μg/ml was also tested at 90 °C in the presence of 1 mg/ml BSA. After heat inactivation, samples were immediately cooled on ice, then diluted 1000 times with 0.1 M Tris (pH 7.4). TNAK residual activity was determined with 10 μl of these diluted samples using the coupled spectrophotometric assay. Remaining activity (100%) corresponded to 1050 units/mg at 38 °C.

To determine the effect of pH on the kinetic stability of TNAK, TNAK (1.85 μg/ml) was incubated in 50 mM sodium acetate (pH 3.7–5.5), sodium phosphate (pK_{a2}) (pH 6.0–7.0), Tris (pH 6.5–8.0), glycine (pH 8.5–10.5) and sodium phosphate (pK_{a3}) (pH 11.0–12.0) [total ionic strength adjusted to 220 mM with KCl (calculated as above)] for 1 h at 50 °C. The residual activity of TNAK was assayed in 50 mM Tris (pH 7.4) using the coupled spectrophotometric method at 30 °C.

2.3.9 Microcalorimetry

Differential scanning calorimetry (DSC) was performed with an MC-2 microcalorimeter (Microcal, Northampton, MA, U.S.A.). Scans were performed with 2 mg/ml TNAK solutions in buffer A in the presence of different concentrations of high-purity guanidine hydrochloride (GdnHCl; Pierce, Rockford, IL, U.S.A.). The reference cell contained buffer A with the corresponding GdnHCl concentration. Thermal gradients were from 25

to 100 °C at a scanning rate of 1 °C/min. To minimize the background noise, a buffer-buffer scan was subtracted from the buffer–TNAK scan before data analysis. DSC scans of TNAK and the apoenzyme in the absence of the denaturant were performed at the Glasgow Biological Microcalorimetry Facility (University of Glasgow, Glasgow, U.K.). These scans were performed with 1.3 mg/ml TNAK in 50 mM Mops (pH 7.4), with a 1 °C/min thermal gradient from 25 to 125 °C.

2.3.10 Spectroscopic methods

Spectrophotometric titrations of Zn^{2+} in TNAK with *p*-hydroxymercuriphenyl sulphonate (PMPS) were performed in a Beckman DU650 spectrophotometer equipped with a Peltier system, using quartz microcuvettes (1 cm path length). The enzyme was extensively dialyzed against 40 mM Hepes/NaOH (pH 7.2) (buffer C) before titration. Buffer C, PMPS and 4-(2-pyridylazo) resorcinol (PAR) solutions were prepared using MilliQ water. Release of Zn^{2+} from TNAK was followed after adding successive samples of 1-10 mM PMPS to 300 μl of 10μM TNAK in buffer C, which contained 0.1 mM PAR. Absorbances at 250 and 500 nm were recorded 5 min after each PMPS addition. In each experiment, the first sample cuvette (10 μM TNAK in buffer C plus 0.1mM PAR) was used as the blank, so that absorbance values recorded thereafter were values of ΔA_{250} and ΔA_{500}. Initial values of A_{250} and A_{500} for other sample cuvettes were measured before adding the first PMPS sample. These initial absorbance values were later subtracted from the experimental readings to eliminate variations due to cuvette differences.

CD spectroscopy was performed on a JASCO J-170 spectropolarimeter (JASCO, Easton, MD, U.S.A.) equipped with a 1ml quartz cylindrical cuvette. The sample

solutions contained 0.03 mg/ml TNAK in 10 mM Tris (pH 7.4) in the presence of different concentrations of high-purity GdnHCl. Solutions were incubated overnight at room temperature for equilibration. For each GdnHCl concentration, a CD spectrum was recorded in the absence of the enzyme. This spectrum was subtracted from the enzyme spectrum before data were analyzed.

2.3.11 Other analytical procedures

Protein concentrations were determined using the Bio-Rad protein assay kit (Bio-Rad Laboratories) with BSA as the standard. SDS/PAGE was performed as described previously (29). Proteins were visualized by Coomassie Blue staining. Elemental analysis was performed by inductively coupled plasma-atomic emission spectroscopy (ICP-AES) at the Animal Health Diagnostics Laboratory, Michigan State University. The dialysis buffer was used as the negative control for ICP-AES.

2.4 Results

2.4.1 Cloning of the *adk* gene from *T. neapolitana* in *E. coli*

E. coli strain CV2 transformed with the *T. neapolitana* genomic phagemid library was plated on LB ampicillin agar and incubated at 42 °C. One colony was obtained after a 2 day incubation. The phagemid (pTNAK1) extracted from this colony contained a 3 kb insertion. Transformation of strain CV2 with pTNAK1 allowed CV2 to grow at 42 and

45 °C. The cell extract of CV2(pTNAK1) showed six times more AK activity than the cell extract of CV2 without the plasmid.

2.4.2 Nucleotide sequence of the *adk* gene from *T. neapolitana*

The pTNAK1 insert was sequenced using vector- and insert-specific primers. Two open reading frames, ORF1 and ORF2, were identified. ORF1 was truncated at its 5′-terminus. The two ORFs overlap by one nucleotide (results not shown). A BLAST search using the truncated peptide encoded by ORF1 as the probe showed significant similarity to SecY, a subunit of the SecAEGY preprotein translocase complex. A BLAST search using the peptide encoded by ORF2 as the probe showed significant similarity to multiple AKs. A potential ribosome binding site was identified 4–7 nt upstream of ORF2 (i.e. the *adk* gene). No potential promoter could be identified, suggesting that the *T. neapolitana adk* gene is co-transcribed with *secY*.

2.4.3 Sequence comparisons

As shown in Table II-1 and Figure II-1, TNAK belongs to the family of long bacterial AKs. With the exception of *Thermotoga maritima* AK (TMAK) (with which it is 93.2% identical), TNAK is more similar to the archaeal *Archaeoglobus fulgidus* AK than to other bacterial AKs (Table II-1). TNAK also shows significant similarity to another archaeal AK, the *Pyrococcus abyssi* enzyme. These two archaeal AKs show much higher similarity to the long bacterial AKs, in general, than to the AKs from the archaea *S. acidocaldarius* and *Methanococcus igneus* (results not shown), suggesting that archaeal AKs do not represent a phylogenetically homogeneous enzyme group.

Table II-1 Percent identity (percent similarity) of TNAK with selected prokaryotic AKs.

Scores were obtained using the GAP software. Sequences were obtained from Genbank® and from genome sequences.

Enzyme source	*Aful	Aaeo	BSAK	ECAK	PDAK	*PAAK	*SAAK	*MIAK
TNAK	60.7	48.5	44.7	46.7	41.1	37.7	31.5	21.4
	(70.1)	(60.2)	(57.6)	(56.7)	(56.1)	(53.0)	(43.2)	(34.1)

*Archaeal species.

AFAK: *Archaeoglobus fulgidus* AK; AAAK: *Aquifex aeolicus* AK; BSAK: *Bacillus stearothermophilus* AK; ECAK *E. coli* AK; PDAK: *P. denitrificans* AK; PAAK: *Pyrococcus abyssi* AK; SAAK: *S. acidocaldarius* AK; MIAK: *Methanococcus igneus* AK.

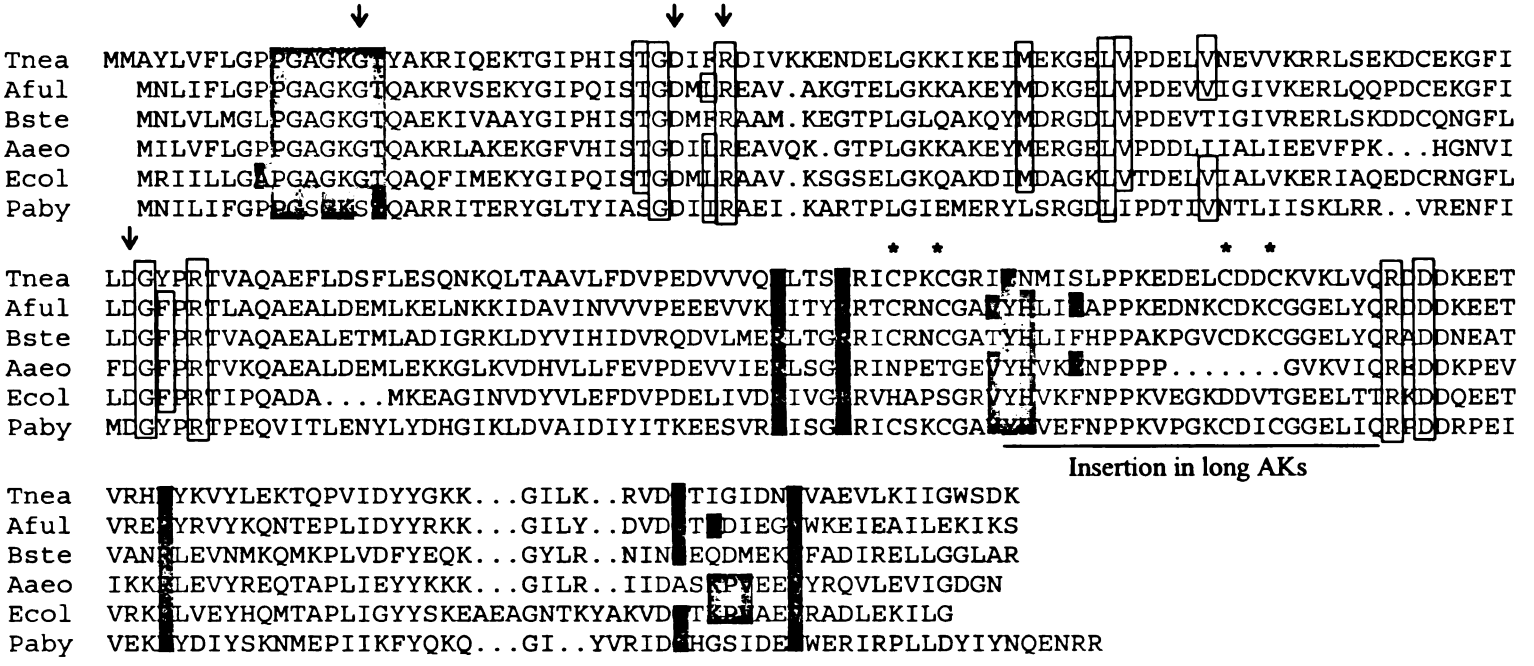


Figure II-1 Alignment of T. neapolitana (Tnea), A. fulgidus (Aful), B. stearothermophilus (Bste), A. aeolicus (Aaeo), E. coli (Ecol) and P. abyssi (Paby) AK sequences

Grey background and framed residues indicate residues interacting with ATP and AMP respectively (10). ↓, residues involved in Mg^{2+} binding (through water molecules) (11); *, cysteine residues involved in Zn^{2+} binding.

Most residues that bind ATP and AMP in ECAK (10) are conserved in TNAK as well as in the *A. fulgidus* and *P. abyssi* enzymes (Figure II-1). The ATP- and AMP-binding residues of ECAK that are not conserved in TNAK correspond either to ECAK residues that have one atom at a 4.5 Å distance from ATP or AMP or to residues that interact with the nucleotides through their backbone nitrogen or oxygen atoms. The four Zn^{2+}- liganding cysteine residues characterized in all AKs from Gram-positive bacteria (30) and in a few other bacterial AKs (31, 32) are present in TNAK (Figure II-1) and in TMAK (results not shown). These cysteine residues are also present in the *A. fulgidus* and *P. abyssi* enzymes (Figure II-1).

2.4.4 TNAK overexpression in *E. coli* and purification

We first attempted to overexpress TNAK in strain BL21(DE3) (Stratagene). The expression level of TNAK remained mediocre (approx. 5 mg/l of culture) even after induction with isopropyl-β-D-thiogalactoside. When we tried expressing TNAK in strain CV2(DE3), no overexpression was observed with or without induction. *E. coli* strain HB101(DE3) gave satisfying results: TNAK was not expressed at significant levels in the absence of induction, and approx. 50 mg/l TNAK was obtained after 20 h induction in SB medium.

During TNAK purification, a significant fraction of the recombinant enzyme (approx. 30%) was present in the pellet after centrifugation of the bacterial crude extract. Most of this enzyme was recovered in solution after resuspension of the pellet in buffer A, heat treatment at 80 °C and centrifugation. The heat treatment at 80 °C was a highly efficient purification step, and it allowed us to obtain the enzyme preparation without any

contaminating ECAK that was initially present in the crude extract. Our purification procedure yielded pure TNAK, as determined by SDS/PAGE (results not shown). The purified TNAK consisted of a single subunit of 25 kDa (as determined by SDS/PAGE), and it migrated as a 29 kDa protein during gel permeation on Sephadex G-75, indicating that TNAK is a monomeric enzyme, a trait typical of long bacterial AKs. The purified TNAK showed a specific activity of 1050 units/mg enzyme at 38 °C.

2.4.5 Effects of temperature and pH on TNAK activity and stability

The specific activity of TNAK, as determined by the end-point method, increased with temperature to its maximum (8400 units/ mg) at 80 °C, and then it started to decrease sharply (Figure II-2A). The Arrhenius plot of TNAK activity (ADP formation) against temperature was linear between 30 °C and 70 °C (Figure II-2B), with activation energy of 49 kJ/mol. This value is of the same order of magnitude as the activation energies for the reactions catalyzed by the muscle cytosolic AK (38 kJ/mol) (17), by ECAK (46 kJ/mol) and by BSAK (73 kJ/mol) (30). The kinetic thermostability of TNAK was tested at two different enzyme concentrations (Figures II-2C and II-2D). TNAK is significantly more stable at high concentrations or in the presence of BSA than at low concentrations. At both concentrations tested, inactivation is of first order, with activation energies of 185.6 kJ/mol at 0.55 mg/ml enzyme and 44.4 kJ/mol at 5.5 μg/ml enzyme. TNAK is optimally active between pH 7.4 and 8.0. It is more than 70% active at pH 7.0 and 8.5 (Figure II-3A). The enzyme is optimally stable at pH 8.3 and retains 80% activity at pH 7.4 and 9.3. TNAK is not at all stable at pH values below 5.0 (Figure II-3B).

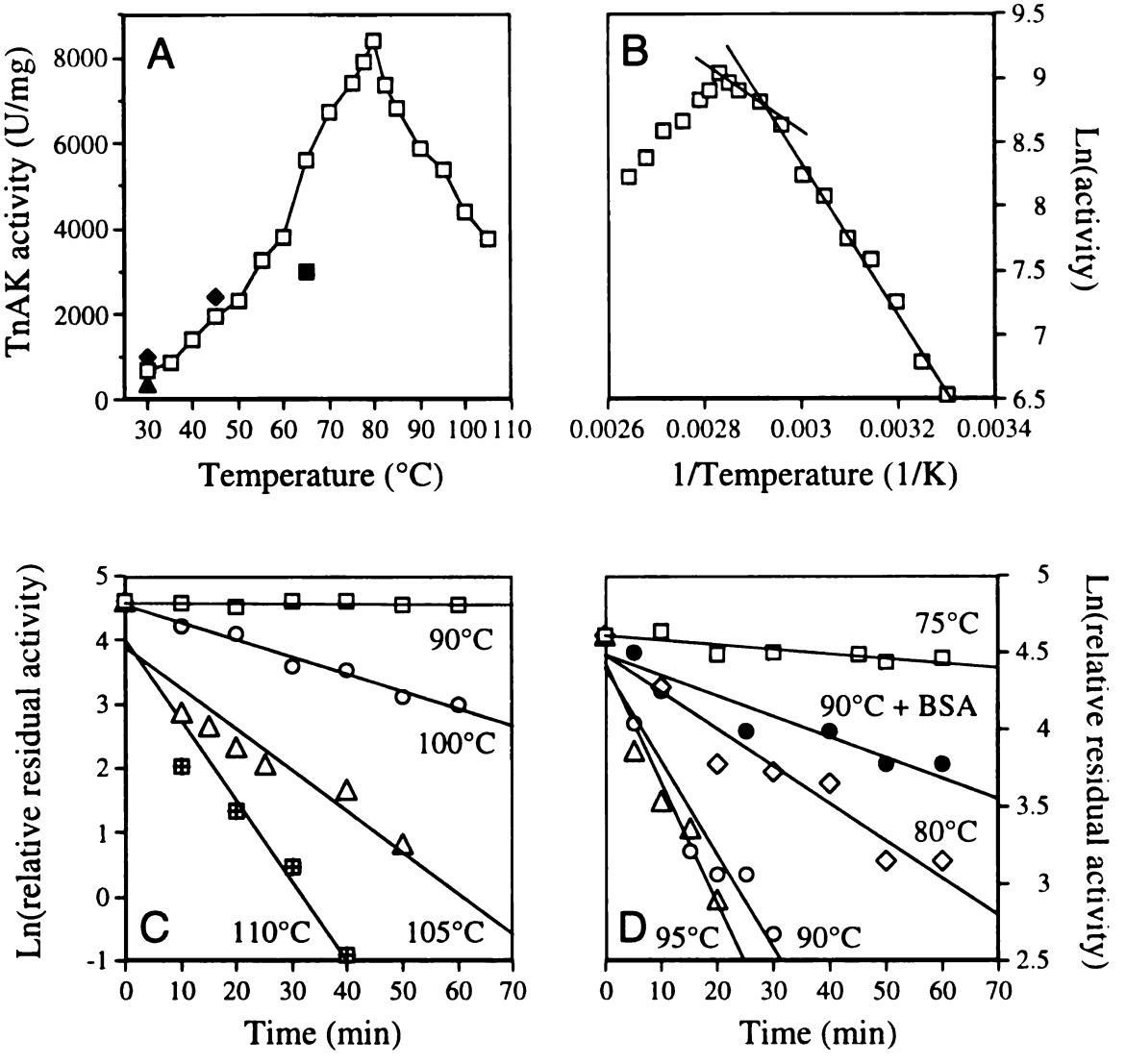


Figure II-2 Effect of temperature on TNAK activity and stability

(**A**) Specific activities of TNAK (□), ECAK (♦), B. subtilis AK (▲) and BSAK (■) at various temperatures (20, 30). (**B**) Arrhenius plot of TNAK activity. (**C**, **D**) TNAK concentrations were 0.55 mg/ml (**C**) and 5.5 μg/ml (**D**). (**C**) □, 90 °C; ○, 100 °C; Δ, 105 °C; ⊞, 110 °C. (**D**) □,75 °C; ●, 90 °C + BSA (BSA concentration=1 mg/ml); ◊, 80 °C; ○, 90 °C; Δ, 95 °C.

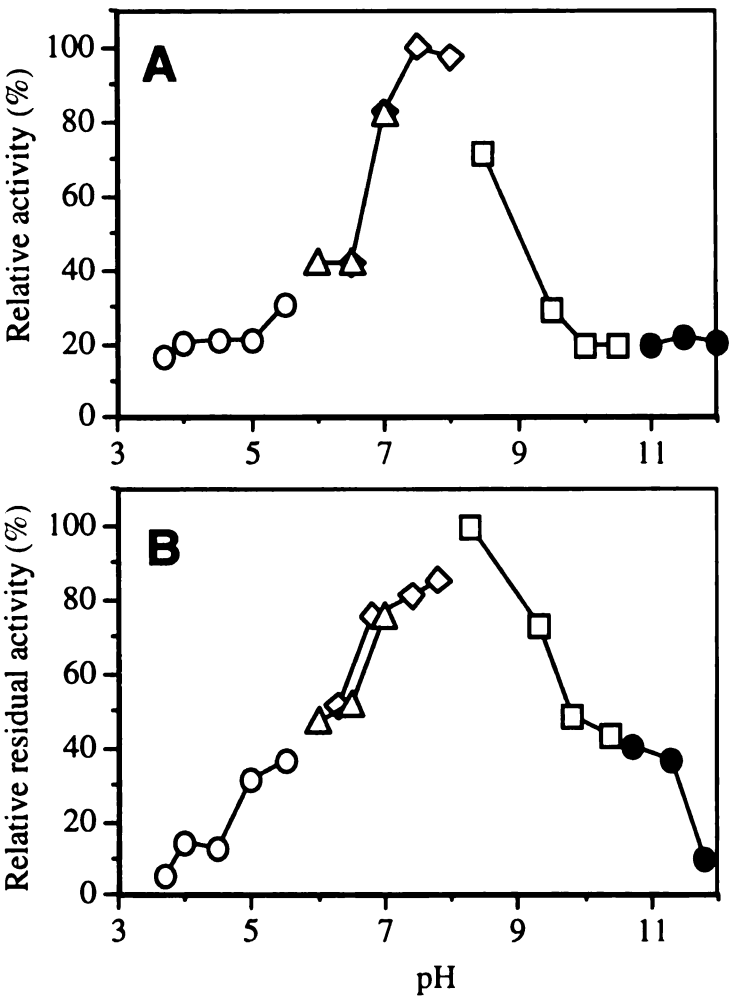


Figure II-3 Effect of pH on TNAK activity (A) and stability (B)

Buffers used: ○, sodium acetate; Δ, sodium phosphate (pK_{a2}); ◊, Tris; □, glycine; ●, sodium phosphate (pK_{a3}).

2.4.6 TNAK stability studied by calorimetry and CD spectroscopy

We did not detect any obvious melting transitions in our DSC experiments with TNAK in buffer A at temperatures below 100 °C (the upper temperature limit for our MC-2 microcalorimeter). For this reason, we followed TNAK thermal unfolding in the presence of GdnHCl. Melting was almost completely reversible at a GdnHCl concentration of 0.4 M, as indicated by the almost complete superimposability of the excess heat capacity curves obtained in two consecutive scans (results not shown). The percentage of refolding after cooling decreased with increasing GdnHCl concentration. At 2.6 M GdnHCl, almost no transition was seen during a second scan. For this reason, no thermodynamic analysis of the data was possible. In Figure II-4A, T_m of TNAK is shown as a function of GdnHCl concentration. A linear curve fit of the data (y=-10.216x+99.334 with r^2 +0.991) would predict a T_m of 99.3 °C in the absence of the denaturant. A second degree polynomial fit of the data (y=-1.315x2 -6.231x+ 97.125 with r^2 =0.996) would predict a T_m of 97.1 °C in the absence of the denaturant. A confirmatory DSC scan of TNAK in the absence of the denaturant was performed by the Glasgow Biological Microcalorimetry Facility on a calorimeter that allowed thermal gradients to be run up to 125 °C. At a scan rate of 1 °C/min, TNAK (1.3 mg/ml) showed a T_m of 99.7 °C. The melting transition observed in the first scan was not observed in the second scan, indicating that under these conditions, TNAK unfolded irreversibly. This T_m value is significantly above the values measured for ECAK, *B. subtilis* AK and BSAK (Table II-2) (20, 30). It also corresponds to the T_m predicted by linear extrapolation of the melting data obtained in the presence of GdnHCl.

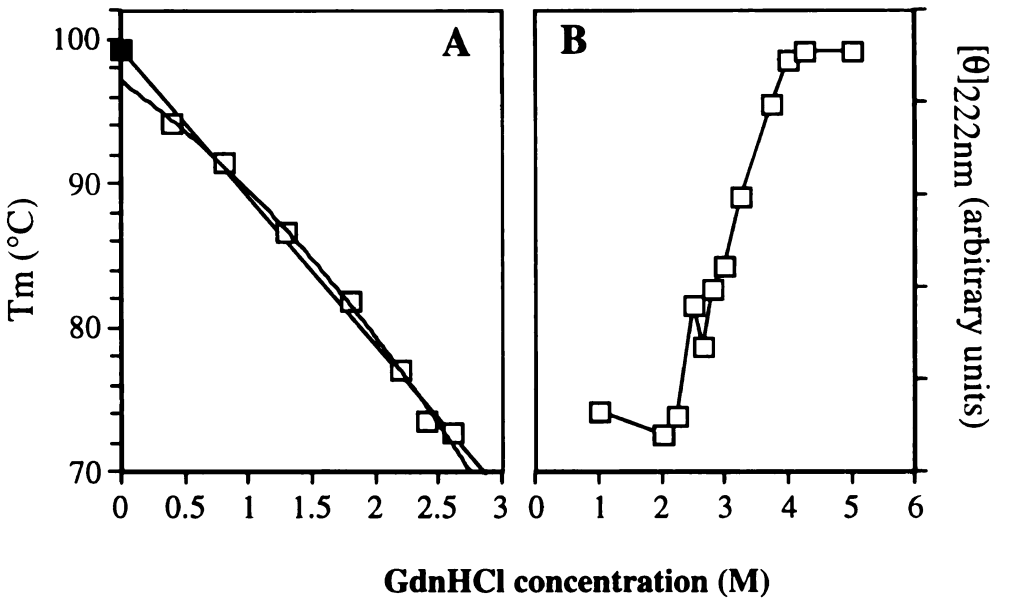


Figure II-4 TNAK unfolding in the presence of GdnHCl.

(A) T_m as a function of GdnHCl concentration, followed by DSC. TNAK concentration was 2 mg/ml (□). For T_m in the absence of GdnHCl, the TNAK concentration was 1.3 mg/ml (■). The two lines represent the linear curve fit and the second degree polynomial fit of the data (see the Results section). **(B)** TNAK (30 μg/ml) equilibrium unfolding induced by GdnHCl at 25 °C as determined by the ellipticity value at 222 nm.

Table II-2 T_m values of ECAK (19), B. subtilis AK, BSAK (30), and TNAK

Enzyme	Melting temperature (°C)	Reference
ECAK	51.8	(19)
ECAK quadruple Cys mutant containing Zn^{2+}	63.0	(19)
B subtilis AK	50.7	(20)
BSAK	74.5	(30)
Apo-BSAK	67.0	(30)
TNAK	99.1	This study
apo-TNAK	93.5	This study

Table II-3 Comparison of ECAK, BSAK, and TNAK kinetic parameters.

Enzyme	Temperature (°C)	K_M		V_{max}(ATP, AMP)	Reference
		ATP (μM)	AMP (μM)	(units/mg protein)*	
ECAK	30	51	38	1,020	(30)
BSAK	37	36	76	288	(30)
TNAK	30	49.4 ± 4.9	40.9 ± 6.6	1130	This study

*No S.D available for V_{max} (ATP, AMP), because it is the mean value calculated from V_{max} (ATP) and V_{max} (AMP).

The TNAK GdnHCl equilibrium unfolding profile obtained by CD spectroscopy at 25 °C is similar to that of other long bacterial AKs, with the following exception: GdnHCl concentration at the midpoint transition for TNAK (i.e. 3.1 M) is much higher than that for ECAK, *B. subtilis* AK and BSAK (0.98, 1.38 and 2.45 M respectively) (20, 30).

2.4.7 Kinetic properties of TNAK

Because TNAK activity at temperatures between 30 and 45 °C is comparable with that of ECAK (Figure II-2A), the kinetic parameters of TNAK were determined at 30 °C for comparison with those of ECAK. As shown in Table II-3, the kinetic parameters of TNAK in the direction of ADP formation were more or less identical with those of ECAK at the same temperature, namely 30 °C.

2.4.8 Determination of the Zn^{2+} content of TNAK

The four cysteine residues involved in Zn^{2+} binding in other AKs (30) are present in TNAK also. This suggests that TNAK also contains Zn^{2+}. The presence of Zn^{2+} in TNAK was tested by ICP-AES and by titration with the strongly dissociating thiol specific reagent PMPS. ICP-AES results (Table II-4) indicate that the recombinant TNAK contains 1 mol of Zn^{2+}/mol of enzyme. Treatments with 10 mM EDTA for 30 min at 60 °C were almost completely inefficient in removing Zn^{2+} from TNAK. EDTA treatment for 1 h at 75 or 80 °C was necessary to deplete approx. 75% Zn^{2+} from TNAK (Table II-4). Other treatments (i.e. with 10 mM phenanthroline or with 5M GdnHCl plus 10 mM

Table II-4 Zn^{2+} content in TNAK as determined by ICP-AES

Treatment	Temperature (°C)	Duration (min)	Zn^{2+} content in TnAK (mol/mol)	Remaining activity (%)
None	—	—	0.88	100
EDTA (10 mM)	20	30	0.85	n.d.*
	40	30	0.91	n.d.
	60	30	0.76	n.d.
	70	30	0.58	n.d.
	75	60	0.24	n.d.
	80	60	0.23	94.6
Phenanthroline (10 mM)	75	60	0.58	n.d.
GdnHCl (5M) + EDTA (10 mM)	20	60	0.26	n.d.

*n.d., not determined

EDTA) were not more efficient than the EDTA treatments. As has been observed with BSAK (30), reducing the Zn^{2+} content of TNAK does not affect its activity (Table II-4).

We confirmed the ICP-AES results by titrating Zn^{2+} in TNAK with PMPS at 25 °C. Formation of the PMPS–thiol chromophore (followed at 250 nm; Figure II-5) shows that PMPS is linearly incorporated into TNAK, up to approx. 4 mol of PMPS/mol of TNAK. ΔA_{250} becomes constant once the ratio [PMPS]/[TNAK] >4. PMPS binding is accompanied by the release of Zn^{2+}, which immediately forms $(PAR)_2Zn^{2+}$ complexes that can be detected at 500 nm. Parallel with what is observed at 250 nm, Zn^{2+} is linearly released from TNAK up to approx. 4 mol of PMPS/mol of TNAK (followed at 500 nm; Figure II-5).With a molar absorption coefficient of 6.6×10^4 $M^{-1}\cdot$ min-1 for $(PAR)_2Zn^{2+}$ at 20 °C (33), a complete release of Zn^{2+} from TNAK (i.e. 6.58 μM) would give a ΔA_{500} of 0.43. The value of ΔA_{500} reached at the plateau (i.e. 0.42) corresponded to 0.98 mol of Zn^{2+}/mol of enzyme, a result that agrees with the ICP-AES results.

2.4.9 Thermostability of the apo-TNAK

A quadruple cysteine mutant ECAK capable of binding Zn^{2+} had a T_m 11 °C above that of the wild-type ECAK (18). The apo-BSAK had a T_m 7.5 °C lower than that of the holo BSAK (26). These two results indicate that Zn^{2+} binding is a stabilizing feature in these two AKs. To test the role of Zn^{2+} binding in the thermodynamic stability of TNAK, we measured the T_m of apo-TNAK by DSC. To prepare the apoenzyme, TNAK was incubated at 80 °C for 1 h in the presence of 10 mM EDTA. It was then extensively dialysed against buffer A containing 2 mM EDTA and also against three batches of

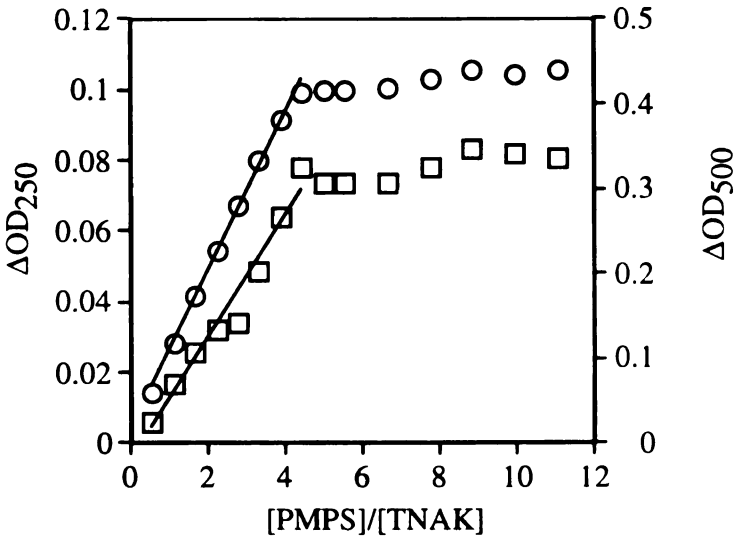


Figure II-5: Titration of Zn^{2+} in TNAK with PMPS

□, ΔA_{250} ('OD_{250}'); O: ΔA_{500} ('OD_{500}'). [PMPS] and [TNAK] were corrected for dilution before calculating the [PMPS]/[TNAK] ratios. ΔA_{250} and ΔA_{500} values were corrected for dilution.

metal-free buffer A. By DSC, the apoenzyme shows a T_m approx. 6 °C below that of the holoenzyme (Table II-2).

2.5 Discussion

As expected from TMAK sequence (32), TNAK belongs to the long bacterial AKs. Since some proteins that are known to be monomeric in mesophiles exist as multimers in hyperthermophiles (34–36), and since some AKs are trimeric (14), it was important for us to determine the quaternary structure of TNAK. The fact that TNAK is a monomeric enzyme that is highly similar to ECAK and BSAK (Table II-1, Figure II-1) makes ECAK, BSAK and TNAK a perfect set of similar enzymes whose dynamic properties can be compared at the residue level to correlate thermostability and activity with conformational dynamics.

Molecular mechanisms underlying the thermostability of TNAK are unknown. TNAK contains a significantly smaller number of alanine and glycine residues when compared with ECAK or BSAK (26 in TNAK versus 39 in ECAK and 40 in BSAK). Six alanine and glycine residues in ECAK are substituted with charged residues in TNAK. These six residues are located on the surface of ECAK [as visualized in the X-ray structure of ECAK (PDB code 1ake); results not shown]. Of the seven bulky hydrophobic residues of ECAK (i.e. Ile, Leu, Met or Val) substituted with charged residues in TNAK, six are exposed to the solvent (results not shown). Although these and other similar, not mentioned, observations suggest that TNAK might be stabilized by optimized

hydrophobic packing and by an increased number of surface salt bridges, these stabilization mechanisms remain highly hypothetical in the absence of supporting structural and experimental evidence (the crystallization of TNAK is in progress).

The similarity of TNAK to the archaeal *A. fulgidus* AK is an intriguing result. Such a high similarity level suggests that the two genes originate from a common ancestor. Although evidence exists suggesting that lateral gene transfer has occurred between thermophilic bacteria and archaea (37), the *adk* gene might not be among the genes that have been laterally transferred. In *A. fulgidus*, the *adk* gene is not preceded by *secY* (38), as it is in *T. neapolitana*. In addition, the *T. maritima adk* gene does not belong to one of the 15 genomic cluster regions that are most archaea-like (37).

As shown in Figure II-2(A) and Table II-3, TNAK is as active as mesophilic and thermophilic bacterial AKs between 30 and 65 °C. In particular, it is three times as active at 30 °C as BSAK is at 37 °C. This result is highly uncommon. Only three other hyperthermophilic enzymes have been characterized that are highly active at mesophilic temperatures: (i) the *T. maritima* L-isoaspartyl *O*-methyltransferase (39) is optimally active at 85 °C, with a specific activity that is at least ten times higher than that of mesophilic enzymes at 25–37 °C. When compared with similar mesophilic enzymes, the *T. maritima* L-isoaspartyl *O*-methyltransferase had the lowest K_m values for each of the methyl-accepting substrates examined at 37 °C. (ii) The *T. maritima* indoleglycerol phosphate synthase (40) is twice as efficient as the *E. coli* enzyme at 25 °C. This higher catalytic efficiency for the hyperthermophilic enzyme is due to a 70 times higher affinity for the substrate, which compensates for a 33 times lower k_{cat} value. (iii) The *T. maritima*

phosphoribosyl anthranilate isomerase (Tm-PRAI) (41) is almost four times more efficient at 25 °C than its *E. coli* counterpart, *E. coli* phosphoribosyl anthranilate isomerase (Ec-PRAI). Although the k_{cat} value of Tm-PRAI is almost ten times lower than that of Ec-PRAI, the affinity of Tm-PRAI for its substrate is almost 44 times higher than that of Ec-PRAI. For these three *T. maritima* enzymes, the enzyme activity is increased due to a very high affinity for their substrate. Their k_{cat}, however, remains typically one order of magnitude below that of the mesophilic protein. In TNAK, both K_m and k_{cat} values are identical with those of ECAK at 30 °C, making TNAK, again, a unique hyperthermophilic enzyme.

Hyperthermophilic enzymes are usually less active than their mesophilic counterparts at low temperatures. One explanation is that hyperthermophilic enzymes are too rigid at low temperatures to be highly active. Their flexibility increases with temperature, and reaches levels compatible with optimal activity only at temperatures close to the growth temperature of their source organism. This hypothesis and the fact that TNAK shows kinetic parameters at 30 °C identical with those of ECAK suggest that the conformational flexibility of the active site of TNAK should be comparable with that of the active site of ECAK at 30 °C. The small size of bacterial AKs (20–25 kDa) makes possible the use of techniques such as NMR relaxation studies or hydrogen/deuterium exchange coupled with NMR detection, to study flexibility at the atomic level. These techniques will be used to compare the molecular dynamics of TNAK and ECAK catalytic sites at 30 °C.

The facts that TNAK at 0.55 mg/ml is kinetically and thermodynamically stable at 90 °C and that the optimal temperature for TNAK activity is only 80 °C suggest at first that TNAK reversibly inactivates between 80 and 90 °C. On the other hand, TNAK enzymatic assays are typically performed using TNAK concentrations below 1 μg/ml. These concentrations are lower than the concentrations used for the kinetic thermostability assays or for DSC. The lower kinetic thermostability of TNAK at low concentrations suggests that the decrease in its activity at temperatures above 80 °C is due to an irreversible inactivation. This result is different from what Zhang et al. (17) concluded for the muscle cytosolic AK. These authors suggested that the active site was more flexible than the rest of the enzyme, and that the active site reached flexibility levels incompatible with catalysis with increasing temperature, while the rest of the enzyme was still folded.

We observed a linear relationship between T_m and GdnHCl concentration (Figure II-4A). It is interesting to note that the T_m values obtained in the absence of GdnHCl correspond to the T_m values predicted by extrapolation from the melting data obtained in the presence of GdnHCl. Although we cannot, at this time, explain this correlation thermodynamically, we observed similar linear relationships between T_m values and GdnHCl concentration for the *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase (42) and for the *T. neapolitana* xylose isomerase (V. Tchernajenko, C. Vieille and J.G. Zeikus, unpublished work). It is not surprising to observe a higher GdnHCl concentration for the midpoint of the unfolding transition of TNAK than for ECAK and BSAK. Although an increase in GdnHCl concentration at the midpoint of the unfolding transition of the enzyme is not an absolute rule for

hyperthermophilic proteins, similar observations have been made for many of them (43,44).

TNAK is kinetically unstable at pH values below 5.0. This result is different from that observed with the *E. coli* and chicken AKs, which are stable under acidic conditions. This property has been used for their purification: a crude cell extract is acidified, neutralized, and then denatured proteins are removed by centrifugation (17, 45, 46). It would be interesting to determine whether the acid-instability behavior of TNAK is due to its thermostability requirements.

The native *Paracoccus denitrificans* AK and TMAK apparently do not contain any Zn^{2+} (12). To determine whether these two enzymes contained any Zn^{2+}, *P. denitrificans* and *T. maritima* were grown in the presence of $^{65}ZnCl_2$; the AKs were purified from the bacterial crude extracts and tested for radioactivity (12). At no point have these two enzymes been tested for the presence of a metal other than Zn^{2+}. Since the recombinant *P. denitrificans* AK expressed in *E. coli* contains an equal amount of Fe^{2+} and Zn^{2+} (31), we cannot exclude the possibility that the native *P. denitrificans* AK contains Fe^{2+} rather than Zn^{2+}. The same possibility exists for the native TMAK. Purification and characterization of these two native enzymes would be necessary to determine whether they contain any metal at all. In any case, with a T_m of 93.5 °C, the apo-TNAK remains stable and active at the temperatures compatible with *T. neapolitana* growth (i.e. 75–80 °C). Thus although Zn^{2+} stabilizes TNAK, its presence in the enzyme is not absolutely required for TNAK's function *in vivo*.

As observed with *Pyrococcus furiosus* α-amylase (47), removing metals from hyperthermophilic enzymes often requires harsh treatments. Here, a 1 h EDTA treatment at 80 °C is not totally efficient in removing Zn^{2+} from TNAK. But in contrast with *P. furiosus* α-amylase, the Zn^{2+}-liganding cysteine residues of TNAK are accessible for equimolar reaction with PMPS at 25 °C.

2.6 References

1 **Jaenicke, R.** 1991 Protein stability and molecular adaptation to extreme conditions. Eur. J. Biochem. **202**:715–728

2 **D'Auria,S., Herman, P., Lakowicz, J. R., Tanfani, F., Bertoli, E., Manco, G. and Rossi, M.** 2000 The esterase from the thermophilic eubacterium Bacillus acidocaldarius: structural–functional relationship and comparison with the esterase from the hyperthermophilic archaeon *Archaeoglobus fulgidus*. Proteins **40**:473–481

3 **Manco, G., Giosu`e, 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

4 **Wrba, A., Schweiger, A., Schultes, V., Jaenicke, R. and Zavodsky, P.** 1990 Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. Biochemistry **29**:7584–7592

5 **Lazaridis, T., Lee, I. and Karplus, M.** 1997 Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci. **6**:2589–2605

6 **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

7 Yan, H. and Tsai, M. D. 1999 Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. Adv. Enzymol. Relat. Areas Mol. Biol. **73**:103–134

8 **Rusnak, P., Haney, P. and Konisky, J.** 1995 The adenylate kinases from a mesophilic and three thermophilic methanogenic members of the Archaea. J. Bacteriol. **177**:2977–2981

9 **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

10 **M¨uller, C. W. and Schulz, G. E.** 1992 Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap_5A refined at 1.9 Å resolution. A model for a catalytic transition state. J. Mol. Biol. **224**:159–177

11 **Berry, M. B. and Phillips, G. N. J.** 1998 Crystal structures of Bacillus stearothermophilus adenylate kinase with bound Ap5A, Mg^{2+} Ap_5A, and Mn^{2+} Ap_5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg^{2+} and Mn^{2+}. Proteins: Struct. Function Genet. **32**:276–288

12 **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

13 **Kath, T., Schmid, R. and Sch¨afer, G.** 1993 Identification, cloning, and expression of the gene for adenylate kinase from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. Arch. Biochem. Biophys. **307**:405–410

14 **Vonrhein, C., B¨onisch, H., Sch¨afer, G. and Schulz, G. E.** 1998 The structure of a trimeric archaeal adenylate kinase. J. Mol. Biol. **282**:167–179

15 **Munier-Lehmann, H., Burlacu-Miron, S., Craescu, C. T., Mantsch, H. H. and Schultz, C. P.** 1999 A new subfamily of short bacterial adenylate kinases with the *Mycobacterium tuberculosis* enzyme as a model: a predictive and experimental study. Proteins **36**:238–248

16 **Reinstein, J., Vetter, I. R., Schlichting, I., Rosch, P., Wittinghofer, A. and Goody, R. S.** 1990 Fluorescence and NMR investigations on the ligand binding properties of adenylate kinases. Biochemistry **29**:7440–7450

17 **Zhang, Y. L., Zhou, J. M. and Tsou, C. L. 1993** Inactivation precedes conformation change during thermal denaturation of adenylate kinase. Biochim. Biophys. Acta **1164**:61–67

18 **Zhang, Y. L., Zhou, J. M. and Tsou, C. L.** 1996 Sequential unfolding of adenylate kinase during denaturation by guanidine hydrochloride. Biochim. Biophys. Acta **1295**:239–244

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

20 **Perrier, V., Surewicz, W. K., Glaser, P., Martineau, L., Craescu, C. T., Fabian, H., Mantsch, H. H., B^arzu, O. and Gilles, A. M.** 1994 Zinc chelation and structural stability of adenylate kinase from *Bacillus subtilis*. Biochemistry **33**:9960–9967

21 **Belkin, S., Wirsen, C. O. and Jannasch, H. W.** 1986 A new sulfur-reducing, extremely thermophilic eubacterium from a submarine thermal vent. Appl. Environ. Microbiol. **51**:1180–1185

22 **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

23 **Dong, G., Vieille, C., Savchenko, A. and Zeikus, J. G.** 1997 Cloning, sequencing, and expression of the gene encoding extracellular α-amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. Appl. Environ. Microbiol. **63**:3569–3576

24 **Goldberg, J. B. and Ohman, D. E.** 1984 Cloning and expression in *Pseudomonas aeruginosa* of a gene involved in the production of alginate. J. Bacteriol. **158**:1115–1121

25 **Sambrook, J., Fritsch, E. F. and Maniatis, T.** 1989 Molecular Cloning: A Laboratory Manual, 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

26 **Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K.** 1993 Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York

27 **Devereux, J., Haeberli, P. and Smithies, O.** 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395

28 **Dawson, R. M., Elliott, D. C., Elliott, W. H. and Jones, K. M.** 1969 Data for Biochemical Research, 2nd edn, Oxford University Press, Oxford

29 **Laemmli, U. K.** 1970 Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685

30 **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

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

32 **Miura, K., Inouye, S., Sakai, K., Takaoka, H., Kishi, F., Tabuchi, M., Tanaka, T., Matsumoto, H., Shirai, M., Nakazawa, T. et al.** 2001 Cloning and characterization of adenylate kinase from *Chlamydia pneumoniae*. J. Biol. Chem. **276**:13490–13498

33 **Hunt, J. B., Neece, S. H. and Ginsburg, A.** 1985 The use of 4-(2-pyridylazo) resorcinol in studies of zinc release from *Escherichia coli* aspartate transcarbamoylase. Anal. Biochem. **146**:150–157

34 **Hess, D., Kruger, K., Knappik, A., Palm, P. and Hensel, R.** 1995 Dimeric 3-phosphoglycerate kinases from hyperthermophilic archaea. Cloning, sequencing and expression of the 3-phosphoglycerate kinase gene of *Pyrococcus woesei* in *Escherichia coli* and characterization of the protein. Structural and functional comparison with the 3-phosphoglycerate kinase of *Methanothermus fervidus*. Eur. J. Biochem. **233**:227–237

35 **Voorhorst, W. G., Eggen, R. I., Luesink, E. J. and De Vos, W. M.** 1995 Characterization of the celB gene coding for β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. J. Bacteriol. **177**:7105–7111

36 **Dams, T., Bohm, G., Auerbach, G., Bader, G., Schurig, H. and Jaenicke, R.** 1998 Homo-dimeric recombinant dihydrofolate reductase from *Thermotoga maritima* shows extreme intrinsic stability. Biol. Chem. **379**:367–371

37 **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.** 1999 Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature (London) **399**:323–329

38 **Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D. et al.** 1997 The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature (London) **390**:364–370

39 **Ichikawa, J. K. and Clarke, S.** 1998 A highly active protein repair enzyme from an extreme thermophile: the L-isoaspartyl methyltransferase from *Thermotoga maritima*. Arch. Biochem. Biophys. **358**:222–231

40 **Merz, A., Kn¨ochel, T., Jansonius, J. N. and Kirschner, K.** 1999 The Hyperthermostable indoleglycerol phosphate synthase from *Thermotoga maritima* is destabilized by mutational disruption of two solvent-exposed salt bridges. J. Mol. Biol. **288**:753–763

41 **Sterner, R., Kleemann, G. R., Szadkowski, H., Lustig, A., Hennig, M. and Kirschner, K.** 1996 Phosphoribosyl anthranilate isomerase from *Thermotoga maritima* is an extremely stable and active homodimer. Protein Sci. **5**:2000–2008

42 **Burdette, D. S., Tchernajencko, V. and Zeikus, J. G.** 2000 Effect of thermal and chemical denaturants on *Thermoanaerobacter ethanolicus* secondary-alcohol dehydrogenase stability and activity. Enzyme Microbial Technol. **27**:11–18

43 **Jaenicke, R.** 1991 Protein stability and protein folding. Ciba Found. Symp. **161**:206–216; discussion 217–221

44 **Beaucamp, N., Schurig, H. and Jaenicke, R.** 1997 The PGK–TIM fusion protein from *Thermotoga maritima* and its constituent parts are intrinsically stable and fold independently. Biol. Chem. **378**:679–685

45 **Heil, A., Muller, G., Noda, L., Pinder, T., Schirmer, H., Schirmer, I. and von Zabern, I.** 1974 The amino-acid sequence of sarcine adenylate kinase from skeletal muscle. Eur. J. Biochem. **43**:131–144

46 **Tagaya, M., Yagami, T., Noumi, T., Futai, M., Kishi, F., Nakazawa, A. and Fukui, T.** 1989 Site-directed mutagenesis of Pro17 located in the glycine-rich region of adenylate kinase. J. Biol. Chem. **264**:990–994

47 **Savchenko, A., Vieille, C., Kang, S. and Zeikus, J. G.** 2002 *Pyrococcus furiosus* α-amylase is stabilized by calcium and zinc. Biochemistry **41**:6193–6201

Chapter 3

Backbone dynamics in a hyperthermophilic adenylate kinase that is highly active at 30 °C

The research in Chapter 3 is presented in the form of a manuscript to be submitted soon.

3.1 Introduction

Hyperthermophilic archaea and bacteria thrive at temperatures above 80 °C. This adaptation implies that enzymes from these organisms are stable and functional at high temperatures. Indeed, enzymes from hyperthermophiles have developed unique structure-function properties of high stability and optimal activity at temperatures above 70 °C. Hyperthermophilic proteins have been intensively studied for over 30 years to understand their stability mechanisms. Such understanding is required both to gain knowledge of the physical principles underlying protein thermostability and because of thermostable enzymes' potential biotechnological applications. Thermostable enzymes are useful in industrial applications such as food and starch processing, detergent manufacturing, production of high fructose corn syrup, and organic syntheses (48).

Identifying the factors contributing to the enhanced thermostability of proteins from hyperthermophiles has been a long standing problem, and many studies have been directed towards comparing pairs of hyperthermophilic and mesophilic homologous proteins. Several mechanisms have been proposed to contribute to the greater stability of hyperthermophilic proteins (49), the most cited ones being enhanced hydrophobicity (18), better packing, shortening or deletion of loops (38), increased occurrence of proline residues (18, 52), increased helical content, and increased number of hydrogen bonds (H-bonds) and salt bridges (27, 38, 50, 55, 57). Among these, the only consistent trend is shown by salt bridges and hydrogen bonds, whose numbers increase in a majority of hyperthermophilic proteins (27). An indirect evidence for this trend is the finding that proteins encoded by known hyperthermophile genomes contain significantly larger

numbers of charged residues than the proteins encoded by mesophile genomes. The current understanding, though, is that a unifying mechanism for thermostability does not exist (25), and that different hyperthermophilic proteins use different thermal adaptation strategies.

Many hyperthermophilic enzymes are optimally active at elevated temperatures, and they are often inactive at low temperatures (i.e., around 20 °C – 37 °C) (9, 54, 58). There is also a surprising constancy in the catalytic rates of homologous mesophilic, thermophilic, and hyperthermophilic enzymes at the respective optimal growth temperatures of their source organisms. In other words, they maintain "corresponding states" with respect to their activity, topology, solvation, etc. in their respective physiological conditions (43). The reduced activity of hyperthermophilic enzymes at low temperatures appears to support the hypothesis that hyperthermophilic proteins are more rigid than their mesophilic homologues at ambient temperatures. It is thought that enzyme motions required for activity become too slow and too restrained in hyperthermophilic enzymes at low temperatures, and that these proteins gain the flexibility required for optimal activity only at higher temperatures. Lack of activity at low temperatures need not be a consequence of stability at high temperatures, though, as shown by a number of laboratory-evolved thermostable enzymes that are catalytically efficient at low temperatures (15). Also, there is no fundamental reason for linking stability and rigidity, since increased flexibility should lead to thermodynamic stabilization through increased conformational entropy of the folded state (28).

The dynamic properties of several hyperthermophilic-mesophilic enzyme pairs have been compared using a variety of techniques such as Fourier transform Infrared

Spectroscopy (58), fluorescence anisotropic decay (31), hydrogen-deuterium (H/D) exchange (21), neutron scattering (13, 14), NMR spectroscopy (5, 20), and molecular dynamics simulations (8, 16, 28). While several of these comparisons have found evidence supporting the hypothesis that hyperthermophilic enzymes are more rigid than their mesophilic counterparts (23, 28, 44, 46, 58), others have not (8, 12, 16, 20). Several of the techniques mentioned above are limited by their ability to measure only average global dynamics. Enzyme activity and stability involve different structural regions and different timescales. Hence, flexibility comparisons in hyperthermophilic-mesophilic enzyme pairs should use site-specific measures of dynamics spanning a wide timescale range. The limited amount of experimental data plus the scarcity of residue-level data available today do not yet allow us to clearly determine whether rigidity at mesophilic temperatures is a key factor in protein thermostability. The time scale and amplitude of the conformational fluctuations that might lead to low thermostability are also not clear.

The purpose of this study is to investigate the relationships between conformational flexibility, thermostability, and enzyme activity by comparing the backbone amide motions of adenylate kinases (AK) from the mesophile *Escherichia coli* (ECAK) and from the hyperthermophile *Thermotoga neapolitana* (TNAK) using ^{15}N NMR relaxation. We chose to study AK because it is monomeric, and because its small size (i.e., 25 kDa) renders it accessible to NMR and molecular dynamics studies. Backbone dynamics of both the ligand-free and ligand-bound forms of ECAK studied with ^{15}N NMR relaxation have already been reported (41, 42, 47).

AK is a ubiquitous, multi-domain enzyme that catalyzes the reversible transfer of the γ-phosphate group from ATP to AMP to form ADP. AKs have been extremely well

studied both biochemically and structurally with close to 20 crystal structures of various AKs present in the protein data bank. The crystal structure of ECAK shows three domains: a core, a lid, and an AMP-binding (AMPbd) domain (34, 35). A conserved loop, termed the P-loop, essential for nucleotide binding, is also present. In the absence of any ligands, AK adopts an open form. Based on the structures of ligand-free AKs and those of several AKs complexed with nucleoside monophosphates, nucleoside triphosphates, and inhibitors, it has been inferred that the lid and AMPbd domains undergo large conformational changes upon binding of the substrates, AMP, ATP, and of the cofactor Mg^{2+}. The enzyme ternary complex adopts a closed form in which the AMPbd domain closes over AMP and the lid closes over ATP, thereby expelling waters to prevent ATP and AMP hydrolyses (40).

TNAK is 40% identical to ECAK, with most of the ECAK ATP and AMP ligands (Figure III-1) being conserved in TNAK (35). TNAK also shares high sequence identity with several other bacterial AKs whose structures with various ligands are known (*Bacillus subtilis* AK [BSubAK, 46%], *Bacillus stearothermophilus* AK [BStAK, 45%], and *Bacillus globisporus* AK [BGAK, 43%]) (2, 3). Biochemical characterization of TNAK showed that it is optimally active at 80°C, and that it is a highly thermostable enzyme with a melting temperature (T_m) of 99.1 °C (48), more than 47 °C higher than the T_m of 51.8 °C measured for ECAK (37). To our surprise and in contrast to all other characterized hyperthermophilic enzymes, the kinetic parameters of TNAK at 30 °C are identical to those of ECAK (48). Figure III-2, showing the k_{cat} values of ECAK, TNAK,

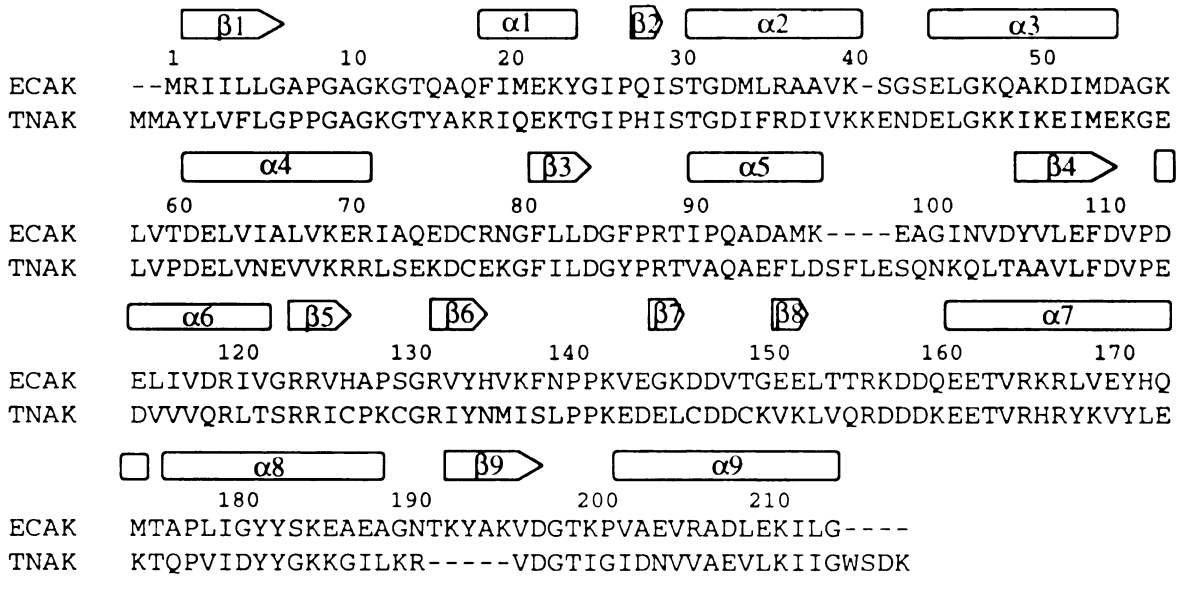


Figure III-1: Alignment of TNAK and ECAK sequences using ClustalW. Helices (bars) and β-strands (arrows) based on the ECAK crystal structure and TNAK 3D model are indicated above the sequence.

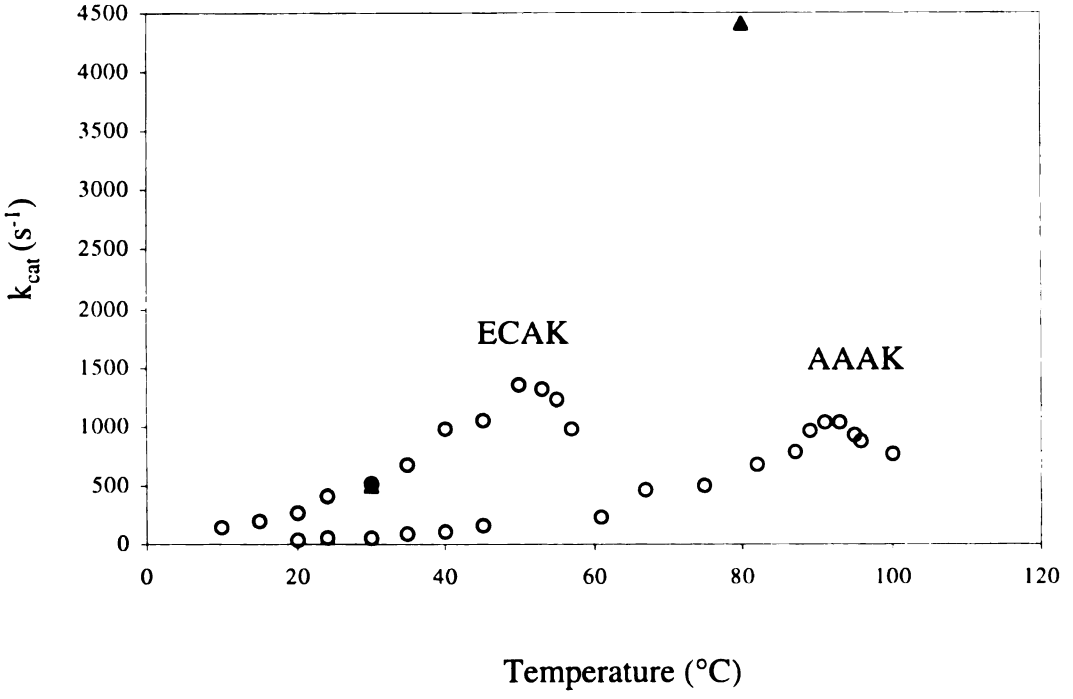


Figure III-2 Activity profiles of ECAK and AAAK (data kindly provided by Wolf-Watz, M and Kern, D) and k_{cat} values of TNAK at 30 °C and 80 °C (▲).

and the AK from another hyperthermophile, *Aquifex aeolicus* (AAAK) at various temperatures, highlights TNAK's unusual behavior. ^{15}N NMR dispersion experiments on AAAK and ECAK have shown the slower lid-opening rate of AAAK (compared to ECAK) to be responsible for its low activity at 20 °C (53). TNAK's intriguing catalytic properties make the comparison of TNAK and ECAK in terms of flexibility, activity, and thermostability particularly interesting and novel. This comparison will also be the third leg in the triangular comparison involving ECAK and AAAK. Using ^{15}N NMR relaxation, we address the question: Do the dynamics of TNAK at 30 °C explain in any way why this enzyme is active at 30 °C? If the hypothesis that enhanced thermostability is attained by enhanced structural rigidity is true, then one expects TNAK to present globally enhanced rigidity with regions of local flexibility that allow activity at low temperatures.

In the present work, we report the resonance assignments for TNAK in the free and inhibitor-bound forms. We investigate the motional properties of TNAK backbone amides in solution on multiple time scales, from ps to μs. Dynamics parameters of TNAK free and inhibitor-bound forms are obtained using ^{15}N relaxation data measured at 30 °C. Results for TNAK are compared to the 30 °C ^{15}N relaxation data for the free and inhibitor-bound forms of ECAK (42).

3.2 Materials and Methods

3.2.1 Sample preparation

Uniformly labeled recombinant TNAK was overexpressed in *E. coli* BL21(DE3) cells transformed with pTNAK2::Km (48). Cells were grown in M9 minimal medium (39) containing (i) 1 g/l of > 98% [^{15}N]-ammonium chloride and 2 g/l of 99% [$^{13}C_6$]-D-glucose for assignments and (ii) 1 g/l of > 98% [^{15}N]-ammonium chloride and unlabeled glucose for relaxation measurements. Protein expression was induced at an OD_{600nm} of 1.0-1.2 with 0.4 mM isopropyl-β-d-thiogalactopyranoside for 20 h at 37 °C. The previously reported protocol for TNAK purification (48) was modified as described below. Supernatant from the heat-treated soluble extract was fractionated with ammonium sulfate at 50% saturation to precipitate impurities. The supernatant was dialyzed against 20 mM Tris-HCl (pH 8.6) (buffer A) and loaded on a Q-sepharose Fast Flow (Amersham Biosciences, Uppsala, Sweden) anion exchange column (2.6 x 30 cm, flow rate 2 ml/min) pre-equilibrated with buffer A. After washing the column with five volumes of buffer A, TNAK was eluted with a 400 ml linear 0.2–0.7 M NaCl gradient in buffer A. Fractions containing TNAK were pooled, dialyzed against 20 mM Tris-HCl (pH 8.0) (buffer B) and loaded again on the Q-sepharose column pre-equilibrated with buffer B. TNAK was eluted with a 0-0.7 M NaCl gradient in 50 mM Tris-HCl, pH 7.4. Purified TNAK was concentrated in an ultrafiltration cell equipped with a 10,000 molecular weight cut-off membrane (Amicon, Beverly, MA), dialyzed against milliQ water, freeze-dried, and stored at -20 °C.

NMR samples of TNAK in the ligand-free form were prepared by dissolving an appropriate amount of freeze-dried TNAK in 320 µl of 50 mM sodium phosphate buffer (pH 7.0) containing 10 µM sodium azide, 6% D_2O/94% H_2O, and protease inhibitor cocktail (Roche, Indianapolis, IN). TNAK concentrations were 1.8 mM for assignment experiments and 1.5 mM for relaxation experiments. For the inhibitor-bound form of the enzyme, TNAK was dissolved in the same buffer, containing in addition, 8 mM $MgCl_2$ and an appropriate amount of the inhibitor P^1, P^5–di(adenosine-5') pentaphosphate (Ap5A). Final concentrations of TNAK and Ap5A were 1.8 mM and 5 mM (assignments) and 1.5 mM and 3 mM (relaxation), respectively.

3.2.2 NMR spectroscopy

Sequential assignments

Triple-resonance data were collected at 30 °C on a Varian Unity Inova 600 spectrometer equipped with triple-resonance probes at 14.1 T. ^{15}N-HSQC, gradient sensitivity-enhanced triple resonance CBCA(CO)NH, 3D HNCACB, HNCA, and ^{15}N–edited NOESY-HSQC data were collected using established pulse sequences (17, 33, 56). Data were processed using NMRPipe (11) and NMRView (24). A Gaussian window function in the direct dimension and a sine window function in the two indirect dimensions were applied. Linear prediction was used for the indirect evolution time periods.

Relaxation measurements

2D sensitivity-enhanced NMR experiments were used to measure ^{15}N longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and heteronuclear ^{15}N-{^{1}H}

steady-state NOEs at 30 °C and 14.1T (26). The T_1, T_2, and NOE data were acquired with spectral widths of 2200 x 8000 Hz (F_1 x F_2). The number of complex points was 160 x 2048 (t_1 x t_2) for T_1 and NOE and 132 x 2048 for T_2. T_1, T_2, and NOE measurements were performed with 16, 32, and 64 transients per t_1 experiment, respectively. Relaxation delays for R_1 experiments ranged from 0 to 1.8 ms and included nine unique time points. The experiments for three time points were repeated twice for error estimation in the measured peak heights. For R_2 measurements, nine unique time points were used; with parametric delays ranging from 0 to 170 ms. Experiments were repeated twice for three of the time points. NOE spectra were recorded with a 3-s pre-delay for the proton-saturated spectrum and 6.5 s for the proton-unsaturated spectrum. A 3.5-s saturation period was used for both TNAK and TNAK*Ap5A. All spectra were processed using NMRPipe (11) and NMRView (24).

Data processing

Resonance signal intensities were quantified using NMRView. R_1 and R_2 relaxation rates were obtained by fitting cross-peak intensities as a function of relaxation delay time to a two-parameter mono-exponential decay function using the Levenberg-Marquardt non-linear least squares fitting software CurveFit (A.G. Palmer, Columbia University). Uncertainties in the fitted relaxation rates were estimated using a jackknife algorithm. Experimental peak uncertainties were determined from the duplicate time points. Steady-state NOEs were calculated as the ratio of cross peak heights of the saturated and unsaturated protons. The average percentage errors for TNAK's R_1, R_2, and NOE were 2.7%, 1.6%, and 5%, respectively. The average percentage errors for TNAK*Ap5A's R_1, R_2, and NOE were 2.6%, 1.1%, and 4.6%, respectively.

Data analysis

The relaxation rates and NOEs were fitted to the Lipari-Szabo model-free formalism (6, 7, 29, 30) using the ModelFree 4.0 software (32, 36). Inertial tensor analyses of TNAK and TNAK*Ap5A were performed using TNAK and TNAK*Ap5A homology models (see *Structure coordinates*) as input PDB files. Hydrogen atoms were added using insightII. The inertia tensor calculations were performed using PDBINERTIA (from the ModelFree 4.0 package [32, 36]), and the hydrodynamic calculations were performed using the HYDRO suite of programs (10).

The R_2/R_1 ratios of a subset of rigid NH vectors (having negligible components of internal motions and/or exchange contributions to the ^{15}N relaxation rates) were selected for calculating the principal components and the orientation of the diffusion tensors of TNAK and TNAK*Ap5A. Residues were selected if their NOE values were > 0.7 and

$$(\langle T_2 \rangle - T_{2,n}) / \langle T_2 \rangle - (\langle T_1 \rangle - T_{1,n}) / \langle T_1 \rangle < 1.5 \text{ SD} \qquad (1)$$

where $\langle T_2 \rangle$ is average T_2 and $T_{2,n}$ is the T_2 value of residue n. SD is given as the standard deviation of the function $(\langle T_2 \rangle - T_{2,n}) / \langle T_2 \rangle - (\langle T_1 \rangle - T_{1,n}) / \langle T_1 \rangle$ (45). Components of the overall rotational diffusion tensor were determined using local diffusion approximation as implemented in the Quadric Diffusion software by Palmer et al (32, 36). To choose the best-fit model to describe the molecular rotational diffusion, F statistic testing was used to compare isotropic, axially symmetric, and anisotropic diffusion models.

Following model selection, individual NH data were fitted to each of the following five motional models as described (32). In the models described below, the

square of the generalized order parameter S^2 is given as $S^2 = S_f^2 S_s^2$, where S_f^2 refers to the amplitude of motions in the fast timescale (the corresponding internal correlation time $\tau_f < 200$ ps) and S_s^2 refers to amplitude of motions on the slow timescale with an internal correlation time τ_s (τ_s > ~200 ps). Thus, motions characterized by S^2 encompass ps-ns timescale motions. The value of S^2 ranges from 0 to 1, 0 indicating high flexibility of the NH vector and 1 indicating a completely restricted NH vector. Model 1 assumes that slow internal motions are negligible ($S_s^2 = 1$) and that fast internal motions are very fast (< 20 ps). These assumptions leave only the order parameter S^2 to describe motions on the fast time scale. Model 2 assumes only that the slow internal motions are slow ($S_s^2 = 1$). This assumption gives two parameters to be derived, namely S^2 and an effective internal correlation time, τ_e ($= \tau_f$), for fast motions. Model 3 yields, in addition to S^2, a R_{ex} terms that describes μs-ms time scale exchange contributions to R_2. Model 4 yields S^2 and τ_e for fast internal motions, as well as R_{ex}. Model 5 assumes only that τ_f tends to 0, which allows order parameters on two timescales: S_f^2 for fast internal motions and S_s^2 and the corresponding τ_s for internal motions slower than 500 ps but faster than the overall correlation time τ_c. For models 1 through 4, $S^2 = S_f^2$ ($S_s^2 = 1$).

The difference between parallel and perpendicular components of the ^{15}N chemical shift tensor ($\Delta\sigma$) was taken to be -160 ppm, and the N-H internuclear distance was assumed to be 1.02 Å (22). Fitting procedures, model selection criteria, and optimization procedures were performed as described (32). A model was selected by comparing the value of the χ^2 function at a critical value, α, of 0.1 with the simulated critical value. F-statistical testing (with $\alpha = 0.2$) was used to determine if the more

complicated model fitted the data better for each spin. Monte Carlo simulations were performed using 500 randomly distributed synthetic data to determine uncertainties in the fitted modelfree parameters. The internal modelfree parameters for each residue were further optimized simultaneously with the global correlation time. The procedure was iterated till the results converged.

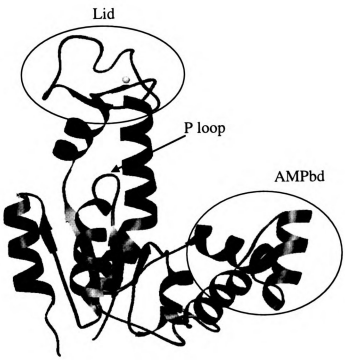
Structure coordinates

A 3D structure of TNAK is not available at this time. A comparison of bacterial AK structures (i.e., ECAK*Ap5A, BGAK*Ap5A, BsubAK*Ap5A, and BstAK*Ap5A) shows that the AK fold is highly conserved. The BGAK*Ap5A, BSubAK*Ap5A, and BStAK*Ap5A structures are highly similar to each other (RMSD is within 1.6 Å) while they differ from the ECAK structure mainly in the position of the lid domain and in the C-terminus. The lid domain is involved in opening and closing motions during catalysis and is therefore, likely to be mobile in solution. The C-terminus in ECAK*Ap5A consists of an α-helix while in BGAK*Ap5A, BStAK*Ap5A, and BSubAK*Ap5A, the helix is followed by a short stretch of residues that do not have a regular secondary structure.

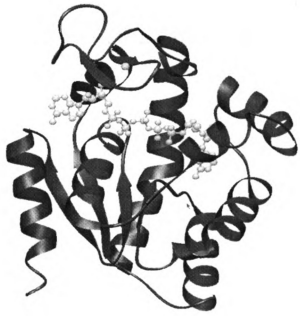
With 40% sequence identity or more, homology modeling can produce structures that are in many respects equivalent to a medium resolution crystal structure. 3D models of TNAK in both the open and closed forms were built by comparative modeling using the MODELER software (Figure III-3) (1). TNAK*Ap5A was modeled using the structures of BSubAK*Ap5A (pdb code 1P3J), BGAK*Ap5A (pdb code 1S3G), and BStAK*Ap5A (pdb code 1ZIN) as templates. BGAK, BSubAK, and BStAK share 43%,

Figure III-3: Ribbon diagrams of the homology-modeled structures of A) TNAK (open form) and B) TNAK*Ap5A (closed form). The lid and (AMPbd) domains and the P-loop are indicated in the open form. The figures were generated using the program Ribbons. Zn^{2+} in the lid domain and Ap5A are shown with a ball and stick representation.

A



B



46%, and 45% identity with TNAK, respectively, and all four AKs contain a structural zinc in their lid domain. TNAK was modeled using the structures of ECAK (pdb code 4AKE) and BStAK*Ap5A (pdb code 1ZIN) as templates. ECAK is the only bacterial AK whose structure is available in the open form. Because ECAK does not have the structural zinc in the lid, BSAK*Ap5A was used to model only the lid domain of TNAK.

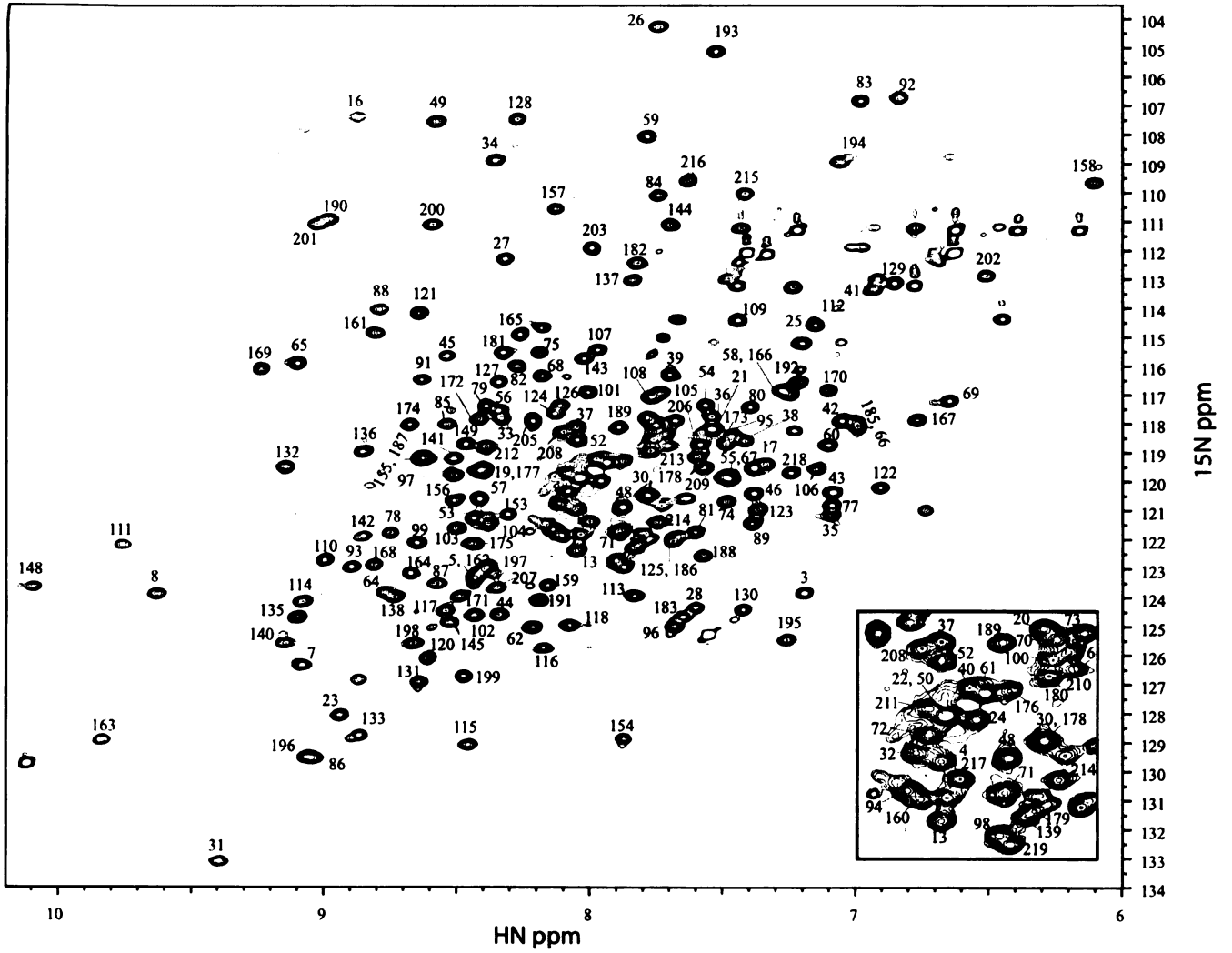
The 3D structure of AK consists of three domains and two pairs of bending regions, one between AMPbd domain and core (i.e., core-AMPbd hinges) and the other between lid and core (i.e., core-lid hinges). In ECAK, the AMPbd domain contains residues 31-72, the lid contains residues 119-156 and the core is formed by residues 1-28, 80-112, and 173-214. The core-AMPbd hinges involve residues 29, 30, and 73-79 and the core-lid hinges consist of residues 113-118 and 157-172 (19). The corresponding residues in TNAK domains are (i) core: 1-30, 83-119, and 180-220 (ii) AMPbd domain: 33-75 (iii) lid: 126-163 (iv) core-AMPbd hinges: 31, 32, and 76-82; and (v) core-lid hinges: 120-125 and 163-179.

3.3 Results

3.3.1 Resonance assignments of TNAK

TNAK contains 220 residues, 10 of which are prolines that are not detectable in a 2D ^{1}H-^{15}N HSQC (Figure III-4A). ^{1}H and ^{15}N backbone resonances of 197 out of the 210

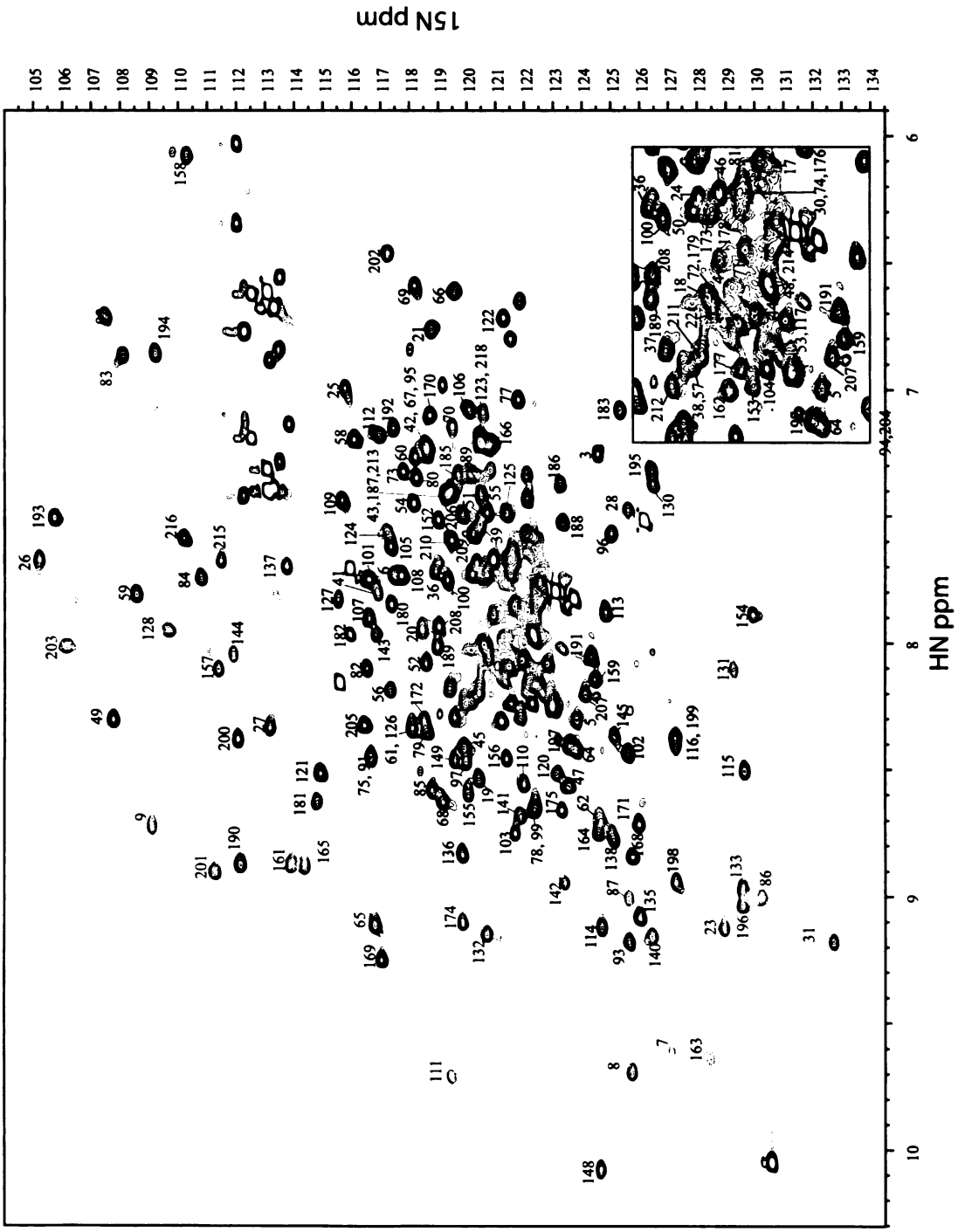
Figure III-4A: 600 MHz ^{1}H-^{15}N HSQC spectra of TNAK at 30 ^{o}C


HN ppm
15N ppm
10
9
8
7
6
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134

non-proline residues were assigned in ligand-free TNAK. Five residues in the P-loop, three in the lid domain, N-terminal residues M1 and M2, and C-terminal K220 are among the 13 residues whose correlations could not be detected in the HSQC spectrum. Among the 197 residues with detectable resonances, eight pairs of residues (i.e., L5 and Q122, Y178 and H30, I67 and V55, I22 and K50, A19 and V177, D155 and D187, L66 and V185, and K58 and D166) generated peaks that overlapped with each other. With TNAK*Ap5A, 193 peaks were seen in the HSQC spectrum (Figure III-4B) out of the expected 210. Five of the 17 missing residues are located in the P-loop, two are the N-terminal residues M1 and M2, and two are the C-terminal residues D219 and K220. Three correlations from residues in the lid domain and four from residues in the AMPbd domain are also missing. Of the 193 correlations that are detected in the HSQC spectrum, nine pairs of residues (i.e., L75 and R91, L61 and R126, R38 and E57, K72 and L179, V123 and S218, K78 and L99, K53 and D117, A94 and I204, and L48 and W217) and three triplets of residues (i.e., K43, D187, and K213; K42, V67, and Q95; and H30, R74, and K176) generated overlapping peaks. TNAK and TNAK*Ap5A HSQC spectra differed considerably from each other (Figures III-4A and III-4B), underlying the fact that large conformational changes are induced by inhibitor binding, as has already been observed for ECAK (42). Strip plots of HNCA spectra showing the sequential walk for residues 190-197 in TNAK and TNAK*Ap5A are shown in Figure III-4C.

α-helices can be readily detected in a 3D-NOESY-HSQC spectrum due to strong NOEs between consecutive amide hydrogens of α-helices. α-helices could be easily

Figure III-4B: 600 MHz ^{1}H-^{15}N HSQC spectra of TNAK*Ap5A at 30 °C


15N ppm
HN ppm

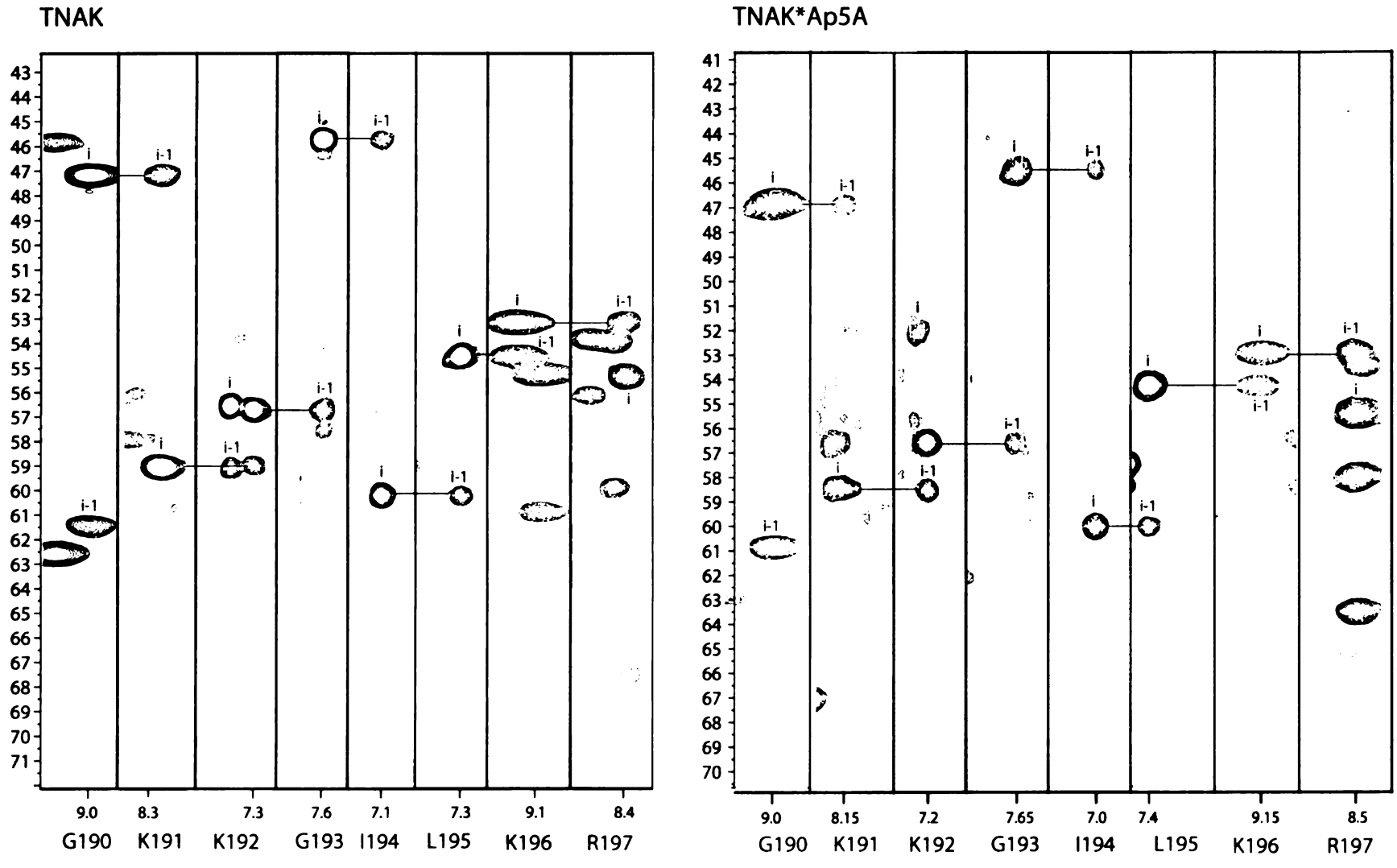


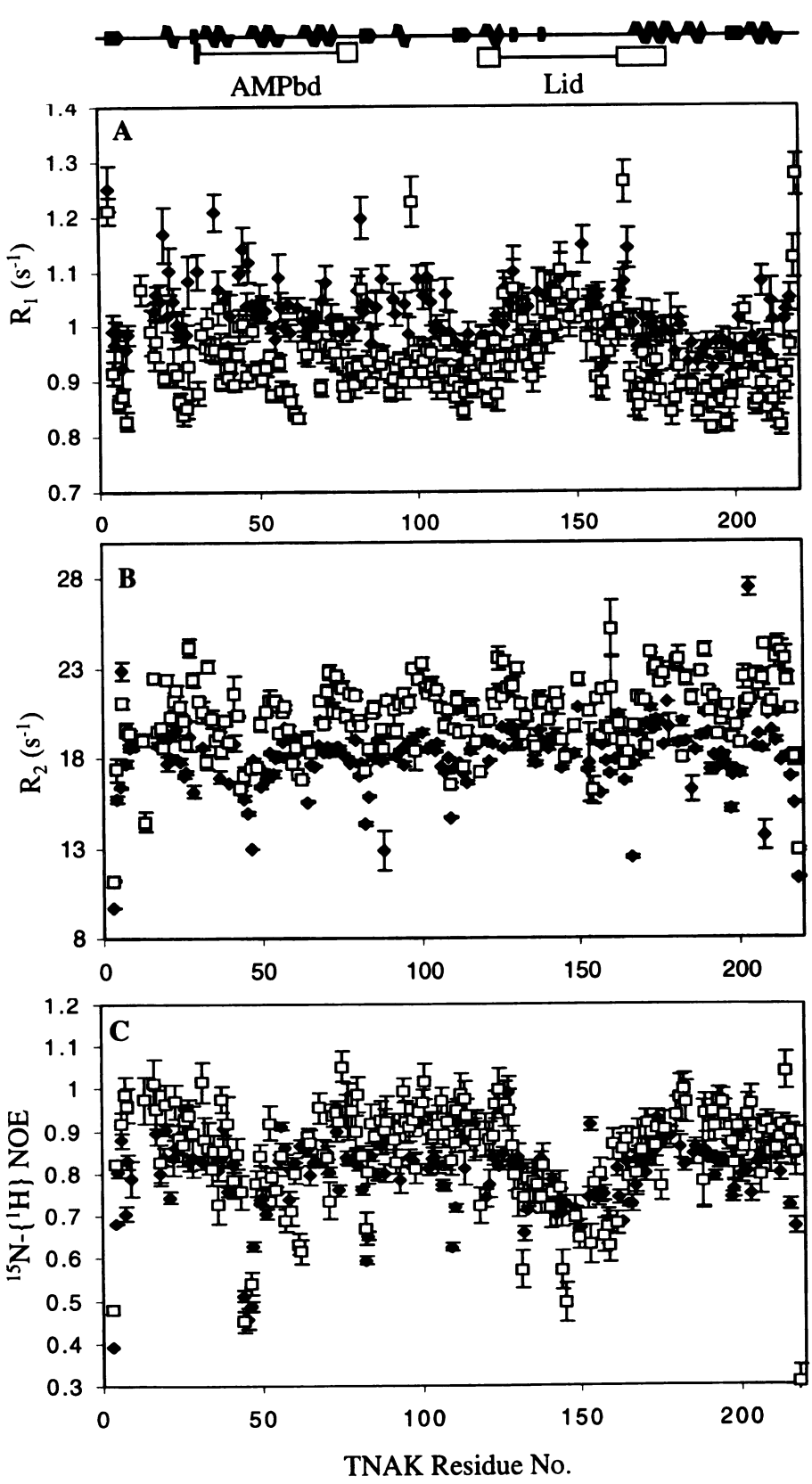
Figure III-4C: HNCA strip plots of TNAK and TNAK*Ap5A

discerned from the NOESY-HSQC spectra of TNAK and TNAK*Ap5A, and they correlated well with those predicted by the 3D models.

3.3.2 Relaxation data

Excluding residues with overlapping or weak correlation peaks, a total of 177 residues were used for TNAK's relaxation data analysis. For TNAK*Ap5A, 164 residues were used to the analysis after excluding 27 residues with overlapping correlations and 2 residues with weak correlations. The R_1, R_2 and ^{15}N-{^1H} NOE values of TNAK and TNAK*Ap5A residues are shown in Figures III-5A, III-5B, and III-5C, respectively. The 10% trimmed mean values of R_1, R_2, and ^{15}N-{^1H} NOE values of TNAK (TNAK*Ap5A) are 0.93 ± 0.023 s^{-1} (1.01 ± 0.024 s^{-1}), 20.53 ± 0.30 s^{-1} (18.32 ± 0.17 s^{-1}), and 0.87 ± 0.039 (0.81 ± 0.017), respectively. Low NOE values reflect the presence of considerable motions on the fast timescale (ps to ns). As expected from the movements of the lid and AMPbd domains triggered by substrates or inhibitor binding in AKs, and as expected from the low NOE values in the ECAK free enzyme, TNAK's lid and AMPbd domains show NOE values that are lower than those of the rest of the protein. Surprisingly, TNAK*Ap5A's lid and AMPbd domains also show low NOE values, suggesting that fast timescale motions are not suppressed in the lid and AMPbd domains upon inhibitor binding. This trend in NOE values is unlike that seen for ECAK, where the NOE values become uniformly high throughout the backbone upon inhibitor binding (42).

Figure III-5: Relaxation parameters of TNAK (□) and TNAK*Ap5A (♦). (A) R_1; (B) R_2; and (C) ^{15}N-{^1H} NOE. The location of TNAK secondary structures and domains are indicated on top of the figures. The open rectangles specify the location of the hinge regions framing the AMPbd and lid domains.


AMPbd
Lid
A
B
C
R_1 (s^{-1})
R_2 (s^{-1})
^{15}N-{^1H} NOE
TNAK Residue No.

3.3.3 Molecular rotational diffusion from hydrodynamic calculations and NMR data

The ratios of TNAK's principal moments were 1.0: 0.88: 0.529. Modeling TNAK as a prolate ellipsoid yielded the Perrin shape factors F_a= 0.834 and F_c=1.335. The R_2/R_1 ratios for 112 residues not subject to large amplitude fast internal motions or to slow timescale conformational exchange (see Materials and Methods) were used to estimate the components of the rotational diffusion tensor. The axially symmetric model agreed significantly better ($p < 0.025$) with the experimental data than the isotropic model did according to an F-statistic test (F = 18.4). The fitting improvement for the fully anisotropic model relative to the axially symmetric model is not significant (F = 0.021). The prolate model was selected as the best fit to the data. After selecting the appropriate models for the individual NH vectors and optimizing the parameters in modelfree calculations, the final τ_c was 14.88 ± 0.021 ns with a $D_{||}/D_{\perp}$ ratio of 1.2 ± 0.01. These values agreed well with the τ_c (15.26 ns) and $D_{||}/D_{\perp}$ (1.26) values predicted from the structure and molecular weight of TNAK using hydrodynamic calculations.

The inertia tensor for TNAK*Ap5A using the homology-modeled structure has principal moments with ratio 1: 0.88: 0.75. A rotational correlation time, τ_c of 13.16 ns and a $D_{||}/D_{\perp}$ ratio of 1.1 were calculated for TNAK*Ap5A with a molecular weight mass of 24.8 kDa using hydrodynamic calculations. R_2/R_1 data of 110 spins selected using the criteria described in Materials and Methods were used to estimate the diffusion tensor parameters. The axially symmetric and the fully anisotropic tensors fitted the data better than the isotropic tensor, but the two models could not be distinguished statistically

using an F-test. Hence, the simpler model of axially symmetric oblate model was chosen. The final τ_c and $D_{||}$ /$D_{\perp}$ values after optimizing the modelfree parameters were 13.56 ± 0.03 and 0.87 ± 0.009, respectively.

One must consider the possibility of protein self association at such high concentrations (1.8 mM), which can adversely affect the values of τ_c and the relaxation parameters. We do not believe that aggregation occurred in the TNAK and TNAK*Ap5a NMR samples for the following reasons: (i) The TNAK and TNAK*Ap5A τ_c values (14.88 ns and 13.6 ns, respectively) are close to the estimated τ_c based on hydrodynamic calculations (15.26 ns for TNAK and 13.16 ns for TNAK*Ap5A) and agree quite well with the molecular size and anisotropy of the protein; (ii) similar τ_c values were calculated for ECAK and ECAK*Ap5A (15.05 ns and 11.4 ns, respectively) from R2/R1 data (42). Since ECAK and TNAK are close in their molecular size and shape, an agreement in their τ_c values suggests monomers are predominant in both the TNAK and TNAK*Ap5A samples.

3.3.4 Backbone internal dynamics

The relaxation parameters of 112, 9, 32, 3, and 21 TNAK residues were fit to motional models 1, 2, 3, 4, and 5, respectively. Contrast this with ECAK where the majority of the residues were fit to model 2 (with parameters S^2 and τ_e) and not model 1 (only S^2), indicating a higher rigidity for many residues in TNAK. For ECAK, 2, 109, 7, 48, and 21 residues were fit to models 1, 2, 3, 4, and 5, respectively. For TNAK*Ap5A the relaxation parameters of 84, 27, 35, 3, and 15 residues were fit to models 1, 2, 3, 4, and 5,

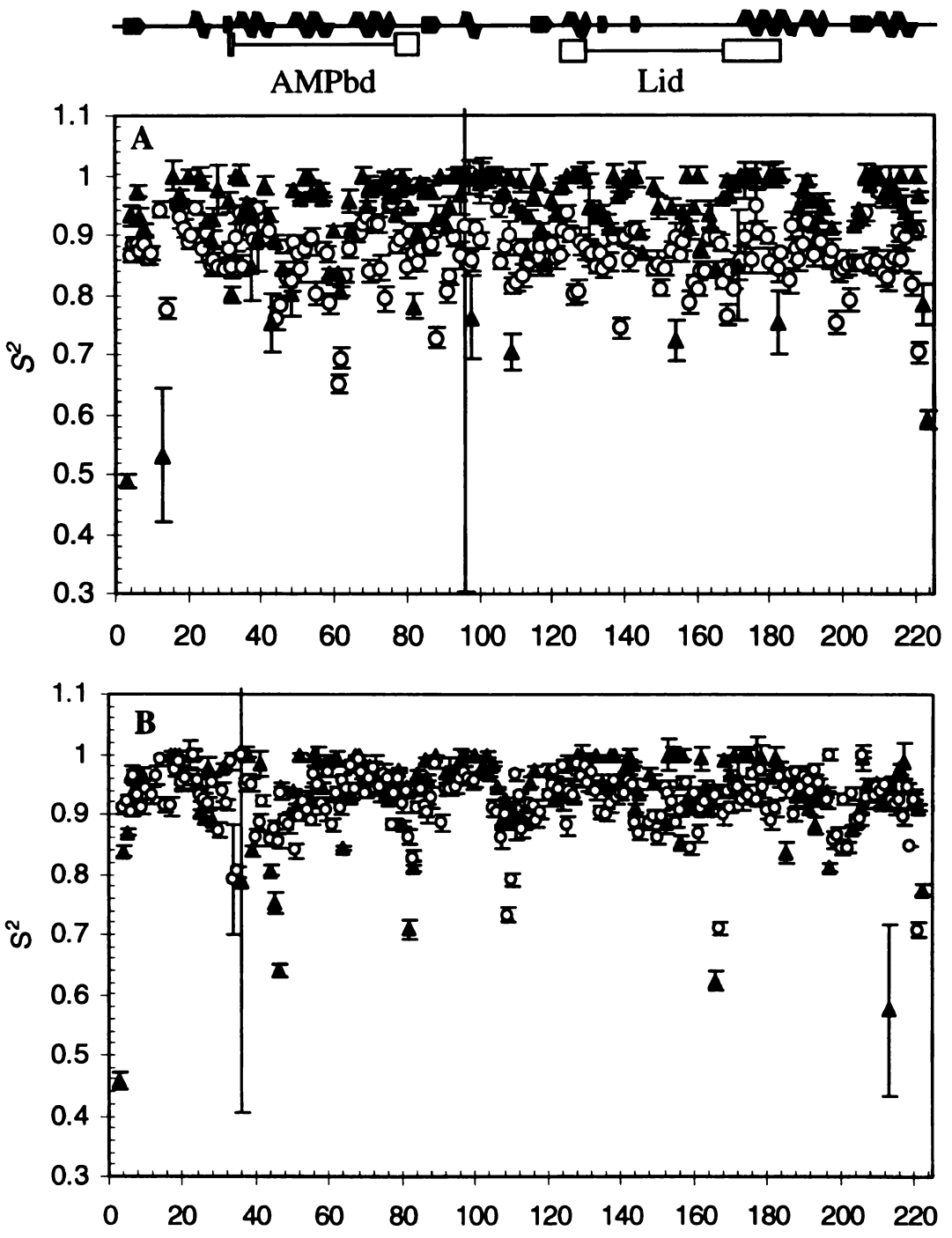
respectively, while in ECAK*Ap5A, 66, 59, 39, 5, and 30 residues were fit to models 1, 2, 3, 4, and 5, respectively.

Pico –nanosecond timescale dynamics

ECAK and TNAK's generalized order parameter (S^2) values are plotted as a function of residue number in Figure III-6A, while Figure III-6B compares ECAK*Ap5A and TNAK*Ap5A's S^2 values. A consistent increase in S^2 values is observed for most TNAK residues compared to the corresponding ECAK residues indicating that the amplitude of internal motions in the ps-ns timescale is larger in ECAK than TNAK. In contrast, this difference in S^2 values almost vanishes in the inhibitor-bound forms, ECAK*Ap5A and TNAK*Ap5A. Although the S^2 values of TNAK residues are uniformly higher, individual residues show significant deviation from this trend. Many of these residues are located in TNAK's AMPbd and lid domains. Residues in the region between 43 and 62 of AMPbd have similar S^2 values in ECAK and TNAK. The S^2 values of several residues in TNAK's lid domain are comparable to (i.e., residues 161, 162, 165-169, and 171) or lower than (i.e., residue 154) the S^2 values of the corresponding residues in ECAK. Residues 98, 109, and 182 in TNAK's core show lower S^2 values than the corresponding residues in ECAK's core. Interestingly, these residues are located in the loops of TNAK's core. One other residue that shows a lower S^2 value in TNAK than in ECAK is residue 82 situated in one of the core-lid hinges.

In TNAK*AP5A also several residues have lower S^2 values than the corresponding residues in ECAK*Ap5A. These residues are 44-46 and 64 in the AMPbd

Figure III-6: Order parameters, S^2, at 30 °C. (A) TNAK (▲) and ECAK (○); and (B) TNAK*Ap5A (▲) and ECAK*Ap5A (○). ECAK data (42) are plotted as per the sequence alignment shown in Figure III-1. TNAK secondary structures and domains are indicated on top of the figure. The open rectangles specify the location of the hinge regions framing the AMPbd and lid domains.


AMPbd
Lid
A
B
S2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
20
40
60
80
100
120
140
160
180
200
220
TNAK Residue No.

domain and residue 166 in the lid domain. In the core domain, residues 83 and 109 have similarly low S^2 values in ECAK*Ap5A and TNAK*Ap5A.

Mean S^2 values ($\langle S^2 \rangle$) for each domain and for each secondary structural element in ECAK, ECAK*Ap5A, TNAK, and TNAK*Ap5A categorized as belonging to the lid, AMPbd, or core domains, or to the different hinge regions are listed in Table III-1. A few salient points that arise from Table III-1 are: (i) the $\langle S^2 \rangle$ values of the core, AMPbd, and lid domains are similar to one another in each of the four enzyme forms indicating similar overall rigidity on the ps–ns timescale across the whole protein; (ii) the % increase in the overall $\langle S^2 \rangle$ values of the three domains in TNAK (compared to ECAK) are similar (between 7.7% and 9.5%), indicating again a uniform increase in the overall rigidity of TNAK. In other words, each domain contributes to a similar extent to TNAK's overall rigidity; and (iii) in contrast, the overall $\langle S^2 \rangle$ values of the three TNAK*Ap5A domains are similar to those of ECAK*Ap5A, indicating similar levels of rigidity on the ps-ns timescale for both enzymes.

At the level of individual secondary structures, significant differences between ECAK and TNAK $\langle S^2 \rangle$ values are seen only for β9 and α5 of the core, for α3 of AMPbd, for the loop region of the lid, and for α7 (belonging to one of the 2 core-lid hinges). In all these regions TNAK has a higher $\langle S^2 \rangle$ value than ECAK. The $\langle S^2 \rangle$ of ECAK*Ap5A's AMPbd loops and core helix α8 are actually higher than that of the corresponding regions in TNAK*Ap5A. Loops in the AMPbd domain have amongst the lowest $\langle S^2 \rangle$ values of any secondary structural elements in both TNAK and TNAK*Ap5A. This statement is true for ECAK as well, but not for ECAK*Ap5A. Interestingly, the

Table III-1: Average S^2 values, $\langle S^2 \rangle$, of the secondary structures in the various domains of ECAK, TNAK, ECAK*Ap5A, and TNAK*Ap5A.

[a] ECAK and ECAK*Ap5A data are from (42)

[b] Standard deviation is in parentheses

[c] Number of residues included in the average

[d] Regions defined in (19)

[e] β7 and β8 are not predicted in the TNAK and TNAK*Ap5A models

Domain	Secondary structure	ECAK[a] <S^2>[b]	N[c]	TNAK <S^2>[b]	N[c]	ECAK*Ap5A[a] <S^2>[b]	N[c]	TNAK*Ap5A <S^2>[b]	N[c]
Core	β1	0.88 (0.01)	6	0.94 (0.03)	4	0.92 (0.02)	6	0.91 (0.06)	6
	β2	0.89 (0.01)	4	0.96 (0.03)	4	0.92 (0.02)	4	0.94 (0.07)	5
	β3	0.86 (0.02)	6	0.93 (0.04)	6	0.91 (0.02)	6	0.93 (0.04)	5
	β9	0.84 (0.02)	6	0.96 (0.03)	6	0.88 (0.04)	6	0.93 (0.04)	6
	α1	0.90 (0.04)	5	0.95 (0.05)	4	0.95 (0.03)	5	0.95 (0.04)	6
	α5	0.89 (0.03)	8	0.99 (0.01)	5	0.95 (0.02)	8	0.99 (0.01)	3
	α8	0.89 (0.04)	10	0.97 (0.02)	4	0.95 (0.02)	11	0.92 (0.04)	9
	α9	0.86 (0.06)	12	0.97 (0.03)	11	0.91 (0.07)	13	0.91 (0.04)	9
	Loops	0.87 (0.06)	30	0.93 (0.07)	35	0.92 (0.06)	31	0.93 (0.07)	31
	Overall	0.873		0.947		0.92		0.93	
AMPbd	α2	0.90 (0.03)	10	0.92 (0.08)	10	0.91 (0.08)	10	0.93 (0.10)	5
	α3	0.86 (0.03)	10	0.95 (0.06)	8	0.90 (0.04)	10	0.95 (0.04)	9
	α4	0.89 (0.04)	11	0.98 (0.03)	10	0.96 (0.02)	11	0.95 (0.05)	10
	Loops	0.80 (0.09)	9	0.85 (0.04)	6	0.91 (0.04)	10	0.87 (0.16)	5
	Overall	0.865		0.932		0.92		0.933	
Lid	β5	0.86 (0.01)	4	0.94 (0.01)	3	0.96 (0.06)	4	0.97 (0.02)	3
	β6	0.83 (0.08)	3	0.98	2	0.94 (0.01)	4	0.97	2
	β7	0.88	1	-[e]		0.94	2	-[e]	
	β8	0.80	2	-[e]		0.90	2	-[e]	
	Loops	0.86 (0.04)	20	0.93 (0.07)	19	0.91 (0.03)	22	0.96 (0.04)	23
	Overall	0.854		0.935		0.921		0.948	
Core-AMPbd hinge regions[d]	Loops	0.88 (0.03)	8	0.92 (0.08)	9	0.92 (0.03)	9	0.91 (0.09)	7
Core-lid hinge regions[d]	α6	0.87 (0.05)	9	0.98 (0.02)	8	0.95 (0.03)	9	0.96 (0.03)	7
	α7	0.86 (0.05)	12	0.96 (0.08)	12	0.93 (0.02)	15	0.98 (0.02)	12
	Loops	0.87 (0.05)	3	0.91 (0.05)	3	0.86 (0.13)	3	0.85 (0.15)	4

significant increase seen in the ECAK's AMPbd loop $\langle S^2 \rangle$ value (from 0.79 in ECAK to 0.91 in ECAK*Ap5A) upon inhibitor binding is not seen between TNAK and TNAK*Ap5A. The AMPbd loop $\langle S^2 \rangle$ of TNAK*Ap5A (0.85) remains similar to that of TNAK (0.87), indicating similar dynamic characteristics of this loop in both forms. The standard deviation associated with TNAK's AMPbd loop $\langle S^2 \rangle$ value is quite high (0.16). It comes from the particularly low S^2 values of residues 45 and 46, and indicates substantial ps-ns timescale motion for these residues. In ECAK*Ap5A and TNAK*Ap5A, but not in ECAK and TNAK, the loops belonging to the lid-core hinges have amongst the lowest $\langle S^2 \rangle$ values. Notably, these hinge loops become more flexible upon inhibitor binding in TNAK ($\langle S^2 \rangle$ decreases from 0.91 in TNAK to 0.85 in TNAK*Ap5A) than in ECAK ($\langle S^2 \rangle$ is 0.87 in ECAK and 0.86 in ECAK*Ap5A).

As in ECAK, TNAK's AMPbd domain features the highest density of residues whose motions are best described by model 5, featuring motions in the slow (ns) and fast (sub-ns) timescales. In TNAK, another cluster of residues that show model 5 type motions is located in the lid and the hinge region following the lid. This cluster of residues does not show model 5 type in TNAK*Ap5A.

Microsecond timescale dynamics

Microsecond to ms timescale motions are identified by the parameter R_{ex}, which represents exchange contribution to T_2. R_{ex} indicates conformational exchange between conformational states that sense different chemical environments. R_{ex} is directly proportional to the chemical shift difference between the exchanging species.

Residues showing exchange contributions are illustrated in Figure III-7A (TNAK and ECAK) and Figure III-7B (TNAK*Ap5A and ECAK*Ap5A). In addition to the R_{ex} values shown in Figures III-7A and III-7B, the dynamics of several TNAK and TNAK*Ap5A P-loop residues could not be analyzed due to large line-widths attributable to exchange broadening. Forty five ECAK residues vs. 22 TNAK residues show exchange contributions at 30 °C. This difference is almost completely eliminated in the inhibitor-bound enzymes where only 17 ECAK*Ap5A residues show μs timescale motion vs. 14 in TNAK*Ap5A. Residues with $R_{ex} > 1.5\ s^{-1}$ are listed in Table III-2 for ECAK, TNAK, ECAK*Ap5A, and TNAK*Ap5A. The predominant differences in μs timescale motions between TNAK and ECAK lie in the core domain and in the loops of core-lid hinges. Specifically, the core β-strands have no exchange contribution in TNAK, while 3 of the 4 ECAK core β-strands have R_{ex} values above $1.5\ s^{-1}$. Interestingly, in both TNAK and ECAK, helices α8 and α9 are quite mobile with several residues showing μs timescale dynamics. Helix α9 forms the C-terminus in ECAK, while in the TNAK model it stretches from residues 204–215. The C-terminal residues 216–220 do not form any regular secondary structure in TNAK. This is seen also in BGAK*Ap5A and BsubAK*Ap5A. The reason for the presence of a large number of residues with R_{ex} values in the C-terminal region of both enzymes is not clear, although this could be simply because of a greater mobility of an unanchored C-terminus. Also of particular note is the presence of several residues with exchange contributions in ECAK and TNAK's core domain loops. Upon inhibitor binding, ECAK's core is significantly rigidified in the μs timescale, as also observed at the ps-ns timescale. Twenty of the 22 residues in ECAK's core β-strands and α-helices that had R_{ex} values in ECAK do not

Figure III-7: Exchange contributions, R_{ex} (s^{-1}) of (A) TNAK (▲) and ECAK (○), and (B) TNAK*Ap5A (▲) and ECAK*Ap5A (○) at 30 °C. ECAK data (42) are plotted as per the sequence alignment shown in Figure III-1. TNAK secondary structures and domains are indicated on top of the figure. The open rectangles specify the location of the hinge regions framing the AMPbd and lid domains.

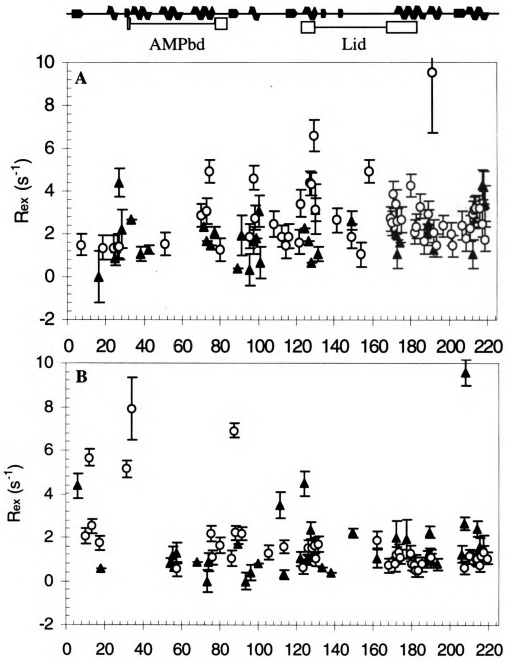

AMPbd
Lid
A
B
R_{ex} (s^{-1})
R_{ex} (s^{-1})
10
8
6
4
2
0
-2
0 20 40 60 80 100 120 140 160 180 200 220
TNAK Residue No.

Table III-2: Residues with R_{ex} values > 1.5 s^{-1} in the various domains of ECAK, TNAK, ECAK*Ap5A, and TNAK*Ap5A.

[a]ECAK and ECAK*Ap5A data are from (42).

[b]β8 is not predicted in the TNAK and TNAK*Ap5A models.

[c]Regions defined in (19)

Domain	Structure	ECAK[a] Residues	TNAK Residues	ECAK*Ap5A[a] Residues	TNAK*Ap5A Residues
Core	β1	5	-	-	6
	β3		-	84	-
	β4	105, 109	-	106	-
	β9	194	-	-	-
	α5	92, 94, 95	97	-	-
	α8	173, 175178, 180, 182, 184, 185	188, 189	-	189
	α9	202, 204, 205, 206, 207, 209, 210, 212	208, 212, 213, 214	-	209, 210, 212
	Loop	101, 176, 190, 200	27, 28, 91, 99, 100	8, 10, 11, 15, 85, 88	89, 111, 202, 203
AMPbd	α2	-	33	32	-
	α3	48	-	-	-
	α4	67, 70, 71	71, 73	-	-
Lid	α6	120, 121	126	119, 121	127
	β5	123	130	124	-
	β6	134	-	-	-
	β8	151	-[b]	-	-[b]
	Loop	122, 142	149	122, 155	149
Core-AMPbd hinge regions[c]	Loop	-	77	29, 72, 77	-
Core-lid hinge regions[c]	α6	114, 115	124	-	124
	α7	162, 163, 164, 165, 166, 168	172, 174	-	172, 177

have R_{ex} values in ECAK*Ap5A. In contrast, the core domain in TNAK*Ap5A shows similar rigidity as in TNAK. The core loops in ECAK*Ap5A and TNAK*AP5A have several residues showing μs timescale dynamics but the individual residues showing R_{ex} values are not the same in the free and bound-forms of the enzymes.

As would be expected from the involvement of this domain in conformational changes, several residues of ECAK and TNAK's AMPbd domains show μs timescale chemical exchange. While no residue in ECAK's core-AMPbd hinges shows μs timescale dynamics, only one residue in the corresponding region in TNAK (residue 77) shows μs timescale motions. Note that the AMPbd domain and its hinges become more rigid on the μs timescale with only a single residue from these regions (#32) showing any chemical exchange in TNAK*Ap5A. In contrast, several residues in ECAK*Ap5A's core-AMPbd hinges show significant mobility.

ECAK and TNAK's lid domains also feature several residues with exchange contribution. In the inhibitor-bound forms of ECAK and TNAK also, several residues (some of them the same as in the free forms) show significant R_{ex} values. The notable difference between ECAK*Ap5A and TNAK*Ap5A lies in the core-lid hinges. Whereas in the ECAK core-lid hinges, there are 8 residues with R_{ex} values, there are none in ECAK*Ap5A. In contrast the two residues that show R_{ex} values in TNAK's core-lid hinges (residues 124 and 172), also show R_{ex} values in TNAK*Ap5A. Note that lid opening has been shown to be the rate-limiting step in ECAK catalysis (53).

3.4 Discussion

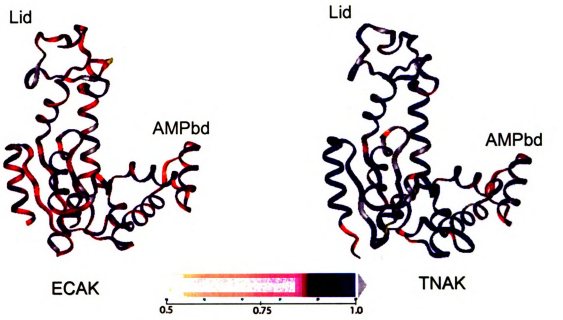
Much effort has been spent in the last two decades in investigating the relationships between flexibility, stability, and activity in hyperthermophilic enzymes. These investigations have been prompted by the fact that most hyperthermophilic enzymes are inactive at ambient temperatures (30 °C). It is clear from these studies that the energy landscape of enzymes is complex, with motions that span picoseconds to seconds and that vary spatially. ECAK and TNAK make a unique mesophilic-hyperthermophilic enzyme pair to study dynamics in relation to stability and activity owing to the unusually high activity of TNAK at 30 °C. We have reported here the difference in the motional properties of ECAK and TNAK over multiple timescales as determined by ^{15}N relaxation measurements acquired at 30 °C.

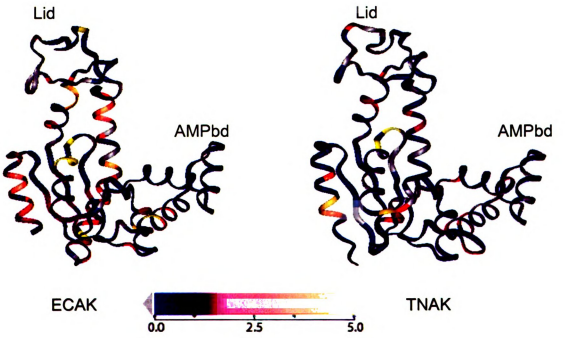
The differences in ECAK and TNAK motions are seen in the NOE data which primarily reflect high frequency (ns) local motions. The high frequency motions in ECAK's AMPbd and lid domains are largely suppressed upon inhibitor binding. In contrast, high frequency motions in TNAK*Ap5A's AMPbd and lid domains are as prevalent as in TNAK. This difference in high frequency local motions in ECAK*Ap5A and TNAK*Ap5A is likely to reflect the difference in the two enzymes motional properties in relation to activity.

Overall, TNAK is uniformly more rigid than ECAK at 30 °C in the ps to ns timescale (Figure III-8). Increased S^2 values in TNAK, particularly in the core, are

Figure III-8: Backbone ps-ns motions (S^2 values) summarized for TNAK and ECAK open forms. Residues with insufficient data are colored gray. ECAK data are from (42). Figures were generated using InsightII. Images in this dissertation are presented in color.

Figure III-9: μs timescale dynamics in ECAK and TNAK backbones open forms as per the scale shown. Residues go from rigid (dark blue) to flexible (yellow). Residues with insufficient data are colored gray. ECAK data are from (42). Figures were generated using InsightII.


Lid
AMPbd
Lid
AMPbd
ECAK
TNAK
0.5
0.75
1.0


Lid
AMPbd
Lid
AMPbd
ECAK
TNAK
0.0
2.5
5.0

indicative of increased enthalpic contribution to the thermodynamic stability of the native state (5), since the formation of enthalpically favorable interactions is expected to quench ps-ns timescale motions. The stability of TNAK's core is further evidenced by the lack of μs timescale exchange in this region unlike what is observed in ECAK's core (Figure III-9). A few residues in ECAK and TNAK core loops have low S^2 values in the free and inhibitor-bound forms. Perhaps, the low S^2 values of residues in the core loops of all four enzyme forms reflect a compensatory effect for the loss in conformational entropy in the rest of the core. ECAK*Ap5A's core is substantially more rigid than ECAK's core in the μs timescale also. Together, these results indicate a higher thermodynamic stability of the inhibitor-bound form of ECAK. Interestingly, the $\langle S^2 \rangle$ of the secondary structures in the core of TNAK*Ap5A are slightly lower than the corresponding $\langle S^2 \rangle$ in TNAK indicating a small increase in flexibility upon inhibitor binding. If significant, this flexibility increase could reflect an inhibitor binding-induced entropic contribution to TNAK*Ap5A stability.

Residues with low S^2 values in the TNAK and TNAK*Ap5A's AMPbd domains are coincidental with the dynamics of this region during catalysis. The loops in ECAK*Ap5A and TNAK*Ap5A's core-lid hinges have low $\langle S^2 \rangle$ values. Although the role of ps-ns timescale motions in slower functional processes is still not understood, these high frequency motions may facilitate the occurrence of slower motions (4). Therefore, the low $\langle S^2 \rangle$ values of the core-lid hinge loops in only the complex forms may have important implication for the lid-opening rates in the two enzymes. As pointed out by Wand (51), it is not necessary for the timescale of functionally relevant motions to be restricted by or to be correlated with functional rate constants.

Most of the residues in TNAK's lid and its hinge regions that show exchange contributions also show significant exchange contributions in TNAK*Ap5A. In contrast, while ten residues show significant R_{ex} values in the core-lid hinge regions in ECAK , these regions contain no residues with R_{ex} values in ECAK*Ap5A. An ^{15}N relaxation dispersion study comparing a thermophilic AK from *Aquifex aeolicus* (AAAK) and ECAK found that lid opening for product release was the rate-limiting step in the two enzymes (53). AAAK, like other thermophilic enzymes is nine-fold less active than ECAK at 20 °C. AAAK's lower catalytic activity at 20 °C was fully accounted for by the slower lid opening rate measured at the same temperature. The fact that TNAK is highly active at 30 °C suggests that the significant exchange contributions observed in TNAK*Ap5A's core-lid hinges reflect dynamics related to lid movements in catalysis. It would be interesting to measure lid opening rates in TNAK to confirm this postulate.

The results from this study show that TNAK, a multi-domain protein in which catalysis-related motions mainly involve hinge movements, is simultaneously highly active at low temperatures and stable at high temperatures by localizing flexibility to the hinge regions. The domains themselves are rigid without compromising activity at low temperatures. Also, localizing flexibility to the hinges does not endanger the stability of the protein. Thus, proteins showing catalytic hinge movements or a catalytic site between well defined, independent domains may be more amenable to rationally engineering both thermostability and high activity at low temperatures than single-domain proteins.

3.5 Conclusion

The increased S^2 values of TNAK indicate increased rigidity consistent with TNAK's increased stability. The lack of μs timescale dynamics in the TNAK core also indicates increased stability of this enzyme. The nature of the change in μs timescale dynamics from TNAK to TNAK*Ap5A indicate that lid opening may not be restricted in this hyperthermophilic enzyme, consistent with its activity at ambient temperatures.

3.6 References

1. **Šali, A., and T. L. Blundell.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234:**779-815.

2. **Bae, E., and G. N. Phillips Jr.** 2004. Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. J. Biol. Chem. **279:**28202-28208.

3. **Berry, M. B., and G. N. J. Phillips.** 1998. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap_5A, Mg^{2+} Ap(5)A, and Mn^{2+} Ap(5)A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg^{2+} and Mn^{2+}. Proteins: Struct. Function, Genet. **32:**276–288.

4. **Brooks III, C. L., M. Karplus, and B. M. Pettitt.** 1988. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. *In* I. Prigogine and S. A. Rice (ed.), Advances in Chemical Physics. John Wiley & Sons, New York.

5. **Butterwick, J. A., J. P. Loria, N. S. Astrof, C. D. Kroenke, R. Cole, M. Rance, and A. G. Palmer.** 2004. Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. J. Mol. Biol. **339:**855-871.

6. **Clore, G. M., P. C. Driscoll, P. T. Wingfield, and A. M. Gronenborn.** 1990. Analysis of the backbone dynamics of interleukin-1β using two-dimensional inverse detected heteronuclear ^{15}N-^{1}H NMR spectroscopy. Biochemistry **29:**7387-7401.

7. **Clore, G. M., A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn.** 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. J. Am. Chem. Soc **112:**4989-4991.

8. **Colombo, G., and K. M. Merz.** 1999. Stability and activity of mesophilic subtilisin E and its thermophilic homolog: Insights from molecular dynamics simulations. J. Am. Chem. Soc. **121:**6895-6903.

9. **D'Amico, S., J. C. Marx, C. Gerday, and G. Feller.** 2003. Activity-stability relationships in extremophilic enzymes. J. Biol. Chem. **278:**7891-7896.

10. **de la Torre, J. G., and V. A. Bloomfield.** 1981. Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. Q Rev. Biophys. **14:**81-139.

11. **Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax.** 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6:**277-293.

12. **Fitter, J., and J. Heberle.** 2000. Structural equilibrium fluctuations in mesophilic and thermophilic α-amylase. Biophys. J. **79:**1629-1636.

13. **Fitter, J., R. Herrmann, N. Dencher, and A. Blume.** 2002. The role of protein dynamics for thermal adaptation of mesophilic and thermostable α-amylase. Biophys. J. **82:**304A-304A.

14. **Fitter, J., R. Herrmann, N. A. Dencher, A. Blume, and T. Hauss.** 2001. Activity and stability of a thermostable α-amylase compared to its mesophilic homologue: Mechanisms of thermal adaptation. Biochemistry **40:**10723-10731.

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

16. **Grottesi, A., M. A. Ceruso, A. Colosimo, and A. Di Nola.** 2002. Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. Proteins **46:**287–294.

17. **Grzesiek, S., and A. Bax.** 1992. Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. J. Am. Chem. Soc. **114:**6291-6293.

18. **Haney, P., J. Konisky, K. K. Koretke, Z. Luthey-Schulten, and P. G. Wolynes.** 1997. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*. Proteins **28:**117–130.

19. **Hayward, S.** 2004. Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. J. Mol. Biol. **339:**1001-1021.

20. **Hernandez, G., F. E. Jenney, Jr., M. W. Adams, and D. M. LeMaster.** 2000. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. Proc. Natl. Acad. Sci. USA **97:**3166-70.

21. **Hiller, R., Z. H. Zhou, M. W. Adams, and S. W. Englander.** 1997. Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 °C. Proc. Natl. Acad. Sci. USA **94:**11329–11332.

22. **Hiyama, Y., C. H. Niu, J. V. Silverton, A. Bavoso, and D. A. Torchia.** 1988. Determination of N-15 Chemical-Shift Tensor Via N-15-H-2 Dipolar Coupling in Boc-Glycylglycyl[N-15]Glycine Benzyl Ester. J. Am. Chem. Soc. **110:**2378-2383.

23. **Hollien, J., and S. Marqusee.** 1999. Structural distribution of stability in a thermophilic enzyme. Proc. Natl. Acad. Sci. USA **96:**13674-13678.

24. **Johnson, B. A., and R. A. Blevins.** 1994. NMRView: A computer program for the visualization and analysis of NMR data. J. Biomol. NMR **4:**603-614.

25. **Karshikoff, A., and R. Ladenstein.** 1998. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Engin. **11:**867-872.

26. **Kay, L. E., P. Keifer, and T. Saarinen.** 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J. Am. Chem. Soc **114:**10663-10665.

27. **Kumar, S., C. J. Tsai, and R. Nussinov.** 2000. Factors enhancing protein thermostability. Protein Engin. **13:**179-191.

28. **Lazaridis, T., I. Lee, and M. Karplus.** 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci. **6:**2589–2605.

29. **Lipari, G., and A. Szabo.** 1982. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules .1. Theory and range of validity. J. Am. Chem. Soc **104:**4546-4559.

30. **Lipari, G., and A. Szabo.** 1982. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules .2. Analysis of experimental results. J. Am. Chem. Soc **104:**4559-4570.

31. **Manco, G., E. Giosue, S. D'Auria, P. Herman, G. Carrea, and M. Rossi.** 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.

32. **Mandel, A. M., M. Akke, and A. G. Palmer, III.** 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. J. Mol. Biol. **246:**144-163.

33. **Muhandiram, D. R., and L. E. Kay.** 1994. Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity. J. Magn. Reson. Series B. **103:**203-216.

34. **Muller, C. W., G. J. Schlauderer, J. Reinstein, and G. E. Schulz.** 1996. Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. Structure **4:**147-156.

35. **Muller, C. W., and G. E. Schulz.** 1992. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap_5A refined at 1.9 Å resolution - a model for a catalytic transition-state. J. Mol. Biol. **224:**159-177.

36. **Palmer, A. G., M. Rance, and P. E. Wright.** 1991. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. J. Am. Chem. Soc. **113:**4371-4380.

37. **Perrier, V., S. Burlacu-Miron, S. Bourgeois, W. K. Surewicz, and A. M. Gilles.** 1998. Genetically engineered zinc-chelating adenylate kinase from

Escherichia coli with enhanced thermal stability. J. Biol. Chem. **273:**19097-19101.

38. **Russell, R. J., J. M. Ferguson, D. W. Hough, M. J. Danson, and G. L. Taylor.** 1997. The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. Biochemistry **36:**9983-9994.

39. **Sambrook, J., F. E. Fritsch, and T. Maniatis.** 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

40. **Schulz, G. E., C. W. Muller, and K. Diederichs.** 1990. Induced-fit movements in adenylate kinases. J. Mol. Biol. **213:**627–630.

41. **Shapiro, Y. E., E. Kahana, V. Tugarinov, Z. C. Liang, J. H. Freed, and E. Meirovitch.** 2002. Domain flexibility in ligand-free and inhibitor-bound *Escherichia coli* adenylate kinase based on a mode-coupling analysis of N-15 spin relaxation. Biochemistry **41:**6271-6281.

42. **Shapiro, Y. E., M. A. Sinev, E. V. Sineva, V. Tugarinov, and E. Meirovitch.** 2000. Backbone dynamics of *Escherichia coli* adenylate kinase at the extreme stages of the catalytic cycle studied by ^{15}N NMR relaxation. Biochemistry **39:**6634-6644.

43. **Somero, G. N.** 1978. Temperature adaptation of enzymes: Biological optimization through structure-function compromises. Ann. Rev. Ecol. System. **9:**1-29.

44. **Svingor, A., J. Kardos, I. Hajdu, A. Nemeth, and P. Zavodszky.** 2001. A better enzyme to cope with cold - Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic IPMDHs. J. Biol. Chem. **276:**28121-28125.

45. **Tjandra, N., S. E. Feller, R. W. Pastor, and A. Bax.** 1995. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. J. Am. Chem. Soc **117:**12562 - 12566.

46. **Tsai, A. M., T. J. Udovic, and D. A. Neumann.** 2001. The inverse relationship between protein dynamics and thermal stability. Biophys. J . **81:**2339-2343.

47. **Tugarinov, V., Y. E. Shapiro, Z. Liang, J. H. Freed, and E. Meirovitch.** 2002. A novel view of domain flexibility in *E. coli* adenylate kinase based on structural mode-coupling ^{15}N NMR relaxation. J. Mol. Biol. **315:**155–170.

48. **Vieille, C., H. Krishnamurthy, H. H. Hyun, A. Savchenko, H. Yan, and J. G. Zeikus.** 2003. *Thermotoga neapolitana* adenylate kinase is highly active at 30 °C. Biochem. J. **372:**577–585.

49. **Vieille, C., and J. G. Zeikus.** 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. Rev. **65:**1–43.

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

51. **Wand, A. J.** 2001. On the dynamic origins of allosteric activation. Science **293:**U1-U1.

52. **Watanabe, K., Y. Hata, H. Kizaki, Y. Katsube, and Y. Suzuki.** 1997. The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. J. Mol. Biol. **269:**142-153.

53. **Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern.** 2004. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nature Structural & Molecular Biology **11:**945-949.

54. **Wrba, A., A. Schweiger, V. Schultes, R. Jaenicke, and P. Zavodsky.** 1990. Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. Biochemistry **29:**7584-7592.

55. **Xiao, L., and B. Honig.** 1999. Electrostatic contributions to the stability of hyperthermophilic proteins. J. Mol. Biol. **289:**1435-1444.

56. **Yamazaki, T., W. Lee, C. H. Arrowsmith, D. R. Muhandiram, and L. E. Kay.** 1994. A suite of triple-resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. J. Am. Chem. Soc **116:**11655-11666.

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

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

Chapter 4

Stability and flexibility in a mesophilic and a hyperthermophilic adenylate kinase

The research in Chapter 4 is presented in the form of a manuscript to be submitted soon.

4.1 Introduction

Proteins from hyperthermophiles are stable and active at high temperatures, conditions in which their mesophilic homologues denature. Since the three-dimensional (3-D) structure of mesophilic and hyperthermophilic homologous proteins are very similar and since their sequence similarity typically ranges between 40% and 85%, the mechanisms responsible for the dramatic stabilization of hyperthermophilic proteins at high temperatures is not readily apparent. To understand the molecular mechanisms that confer thermostability, families of homologous mesophilic-hyperthermophilic proteins have been compared both by biochemical experimentation as well as through computational methods (6, 16, 22, 23, 41, 44). These studies have proposed several mechanisms that could contribute to the stability of hyperthermophilic enzymes at high temperatures. These mechanisms include increased numbers of hydrogen bond (H-bond) interactions (45), presence of salt-bridge networks (23, 33, 48, 49), deletion or shortening of loops, and better packing (33), to name a few. Attempts have been made to introduce specific mutations, as predicted by these studies, in mesophilic enzymes to increase their thermostability to the level of their hyperthermophilic homologues (8, 20, 21, 42). Although such mutations have often increased mesophilic enzymes' thermostability, achieving the level of thermostability of their hyperthermophilic counterparts through rational design has been impossible. This lack of success can be attributed to a lack of understanding the extent to which each interaction contributes to thermostability and, most importantly, how these stabilizing interactions are distributed in the hyperthermophilic protein.

One hypothesis explaining the remarkable stability of hyperthermophilic enzymes is that these enzymes have enhanced conformational rigidity at low temperatures. Indirect evidence for this hypothesis comes from the fact that many hyperthermophilic enzymes are inactive at low temperatures (20-40 °C) but have similar levels of catalytic efficiency as their mesophilic homologue at the optimal growth temperature of their source organism. It is argued that enzymes require a certain degree of conformational flexibility for optimal activity. The presumed rigidity of hyperthermophilic enzymes is used to explain their inactivity at low temperatures, and it is thought that these enzymes require high temperatures to gain the requisite flexibility for activity (17). This hypothesis thereby links local conformational fluctuations and global thermodynamic stability. It implies that an increase in conformational flexibility decreases stability of the native folded state. According to this hypothesis, stability and activity are fundamentally linked, and gain in one would result in a loss of the other. However, as argued by Lazaridis et al (24), protein rigidity and stability do not have to be correlated per se. Conformational flexibility in proteins can span several timescales, from picoseconds (ps) to seconds or more and could be spatially localized.

Native state hydrogen-deuterium (H-D) exchange experiments have become an important tool to study conformational flexibility in proteins. H-D exchange experiments allow one to compare the structural distribution of stability in hyperthermophilic-mesophilic enzyme pairs. Comparing the backbone amide protection factors in the two proteins can also provide insights into the stabilities of individual sites in the proteins. H-D exchange studies with varying denaturant concentrations or temperature are a common technique to determine the thermodynamics of protein stability (2, 4). Several studies

have shown that a subset of the most slowly exchanging residues exchange only during global unfolding events. The free energy of exchange of these residues corresponds to the free energy of unfolding (1, 3). H-D exchange also has emerged as a tool for determining the kinetics of protein conformational changes as well as the global unfolding kinetics (9, 39, 40). Several studies have used H-D exchange experiments to compare mesophilic-hyperthermophilic enzyme pairs. H-D exchange rates in these proteins have been measured using a variety of techniques such as Fourier transformed Infrared Spectroscopy (FTIR) (26, 50), mass spectrometry (25, 26), and NMR (12, 19) . Although results from several of these studies suggest that the hyperthermophilic enzyme is more rigid than its mesophilic homologue, such studies have primarily used techniques such as FTIR that provide a structurally averaged measure of amide exchange rates. NMR methods for following H-D exchange, on the other hand, allow the full potential of site-resolved H-D exchange to be realized. In an H-D exchange study of *Thermus thermophilus* and *Escherichia coli* RNase H (TtRNase and EcRNase, respectively) using NMR, it was found that the general stability distribution was similar for TtRNase and EcRNase. Region-specific ΔG of unfolding increased by approximately 41-42% going from EcRNase to TtRNase, suggesting that each region of the protein contributes the same relative stabilization in TtRNase. The authors also showed that under conditions where the two enzymes have similar global stability (i.e., 0 M guanidine-HCl for EcRNase and 0.88 M Guanidine-HCl for TtRNase), values of region-specific ΔG of unfolding are very similar for the 2 proteins (14). Conversely, an H-D exchange comparison study of *Pyrococcus furiosus* and *Clostridium pasteurianum* rubredoxins demonstrated a more spatially localized increase in rigidity in the

hyperthermophilic protein with some regions, in fact, exhibiting higher flexibility than the mesophilic protein (12, 13).

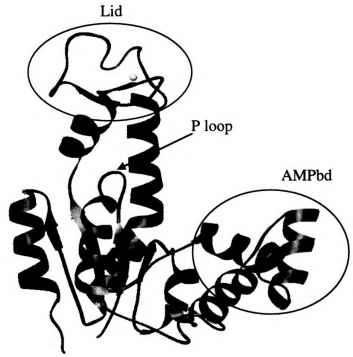
In the work presented here, we use native state H-D exchange to compare the rigidity distribution in the hyperthermophilic *Thermotoga neapolitana* adenylate kinase (TNAK) and the mesophilic *Escherichia coli* adenylate kinase (ECAK) through determination of residue-specific protection factors (P) in the two proteins. Adenylate kinase (AK) is an ubiquitous enzyme that catalyses the transfer of ATP's γ-phosphate to AMP to form ADP. The ECAK structure reveals three domains: the core, the lid, and the AMPbd domains (27, 28). Crystal structures of ECAK in the free form and in complex with various substrates and inhibitors also show that the lid and AMPbd domains are significantly displaced upon ligand-binding (35, 36, 46). These two domains close over the substrates, presumably to protect them from hydrolysis, to adopt the 'closed' conformation. Following phosphoryl transfer, the lid and AMPbd domains open up to release the products, and the enzyme adopts the so-called 'open' conformation.

Although TNAK and ECAK share 40% sequence identity and the active site residues are conserved, TNAK is a highly thermostable enzyme with a melting temperature (T_m) of 99.1°C, more than 47°C higher than the Tm of 51.8°C measured for ECAK (43). TNAK, though, is an unusual hyperthermophilic enzyme in that it is as active as ECAK at 30 °C (43). Since a crystal structure of TNAK is not available, we modeled TNAK's structure in the free form as well as in the ligand-bound form using the software MODELLER 8v1 (1, 10). The models of TNAK and TNAK bound to the inhibitor, diadenosine pentaphosphate (i.e., AP_5A) (referred to as TNAK*Ap5A) are shown in Figure IV-1. The domains in ECAK are defined as AMPbd: residues 31-72; lid:

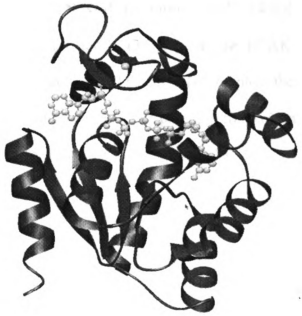
Figure IV-1: Ribbon diagrams of the homology modeled structures of A) TNAK (open form) and B) TNAK*Ap5A (closed form). The lid and AMP binding (AMPbd) domains are indicated in the open form. The figures were generated using the program Ribbons. Zn^{2+} in the lid domain and Ap5A are shown with a ball and stick representation.

.

A



B



residues 119-156; core-AMPbd hinges: residues 29, 30, and 73-79; core-lid hinges: 113-118 and 157-172; and core: rest of the residues (11). The domain limits used here are different from those described by (36), where the respective AMPbd and lid domains are described as residues 30-59 and 122-159. The definitions used here seem to better match the ECAK crystal structures.

TNAK domain definitions are based on sequence alignment with ECAK and on the model are (i) AMPbd: residues 33-75; (ii) lid: residues 126-163; (iii) core-AMPbd hinges: residues 30, 31, and 76-82; (iv) core-lid hinges: residues 120-125 and 164-179; and (v) core: rest of the residues.

To obtain comprehensive information on the flexibilities of ECAK and TNAK in the two forms and at various timescales, we performed ^{15}N NMR relaxation analyses of TNAK and TNAK*Ap5A as well as H-D exchange analyses of ECAK, ECAK*Ap5A, TNAK, and TNAK*Ap5A. ^{15}N NMR relaxation data for ECAK in the open and closed forms have been reported previously (37, 38) and the ECAK and TNAK ^{15}N NMR relaxation results are compared in Chapter 3. To complete the flexibility mapping of ECAK and TNAK at various timescales, we present a comparison H-D exchange data for ECAK, ECAK*Ap5A, TNAK, and TNAK*Ap5A at 30 °C.

4.2 Materials and Methods

4.2.1 Protein samples

TNAK was over-expressed in *E. coli* BL21(DE3) cells containing the expression plasmid pTNAK2::Km (43). TNAK was uniformly ^{15}N labeled by growing the cells in M9 minimal medium (34) containing 1g/l of > 98% [^{15}N]-ammonium chloride. Protein expression was induced at an OD_{600nm} of 1.0-1.2 with 0.4 mM isopropyl-β-d-thiogalactopyranoside (IPTG) for 20 h at 37°C. The previously reported protocol for TNAK purification (43) was modified as described below. Supernatant from the heat-treated soluble extract was fractionated with ammonium sulfate at 50% saturation to precipitate impurities. The supernatant was dialyzed against 20 mM Tris-HCl (pH 8.6) (buffer A) and loaded on a Q-sepharose Fast Flow (Amersham Biosciences, Uppsala, Sweden) anion exchange column (2.6 × 30 cm, flow rate 2 ml/min) pre-equilibrated with buffer A. After washing the column with five volumes of buffer A, proteins were eluted with a 400 ml linear 0.2–0.7 M NaCl gradient in buffer A. Fractions containing TNAK (as determined by OD_{280nm} and SDS-PAGE) were pooled, dialyzed against 20 mM Tris-HCl (pH 8.0) (buffer B) and loaded again on the Q-sepharose column pre-equilibrated with buffer B. TNAK was eluted with a 0-0.7 M NaCl gradient in 50 mM Tris-HCl, pH 7.4. Purified TNAK was concentrated in an ultrafiltration cell equipped with a 10,000 molecular weight cut-off membrane (Amicon, Beverly, MA), dialyzed against milliQ water whose pH was adjusted to the experimental value to maintain the ionization state of the protein, freeze-dried, and stored at -20°C.

The *E. coli adk* gene was amplified from *E. coli* K12 genomic DNA and cloned into the pET24a(+) *Nde*1 and *Xho*1 sites to yield plasmid, pECAK1. Once cloned, the *E. coli adk* gene was sequenced by the MSU Genomic Technology Support facility to verify that no mutation had been introduced during the amplification step. Uniformly ^{15}N-labeled ECAK was obtained by growing *E. coli* BL21(DE3) (pECAK1) cells in M9 minimal medium (34) containing 1 g/L ^{15}N-ammonium chloride (> 98%) as the sole nitrogen source. ECAK expression was induced by adding 0.4 mM IPTG at an OD_{600nm} of 1.0-1.2. The cells were harvested after 20 hours of induction at 37 °C and lysed by two passages through a French pressure cell. The soluble and insoluble fractions were separated by centrifugation at 20,000 x g for 20 min. ECAK was purified as described (31) with the modification that the protein was eluted from an Affi-gel Blue gel affinity column (Biorad, Hercules, CA) using a 0-2 mM NaCl gradient in 50 mM Tris-HCl, pH 7.4. For the ion-exchange chromatography, Q-sepharose Fast Flow (Amersham Biosciences, Uppsala, Sweden) column was used instead of DEAE. The purified protein was extensively dialyzed against de-ionized water with the pH adjusted to the experimental value, lyophilized, and stored at -20 °C.

4.2.2 NMR sample preparation and data collection

The lyophilized protein samples were dissolved in 50 mM sodium phosphate buffer in D_2O at the desired pH to initiate exchange. The buffer was prepared by mixing appropriate amounts of sodium diphosphate and sodium monophosphate in D_2O to obtain the desired pH* (i.e., the pH meter reading in D_2O without adjusting for the deuterium isotope effect). The final NMR sample contained 2 mM protein (TNAK or ECAK) and 10 μM sodium azide in 320 μl deuterated buffer. To prepare the inhibitor-bound forms of

the proteins (ECAK*Ap5A or TNAK*Ap5A), Ap5A and $MgCl_2$ in D_2O were also added to the mixture at 5 mM and 8 mM final concentrations, respectively. The following procedure was followed while setting up an exchange experiment to minimize the dead time: (i) a mock sample identical to that used in the exchange experiment was used to tune the NMR probe and adjust the shim settings; (ii) all acquisition parameters were determined using the mock sample; (iii) the exchange experiments were initiated by adding the lyophilized protein in 320 µl of 50 mM sodium phosphate buffer in D_2O pre-heated to 30 °C. Ap5A and $MgCl_2$ were added immediately to prepare the inhibitor-bound protein samples; (iv) the solution was then briefly mixed and centrifuged to remove insoluble protein, and the sample was quickly transferred to a shigemi tube; (v) the lock signal was monitored until it stabilized, shim settings were readjusted with the exchange sample, and data acquisition was initiated. The temperature for the exchange experiments was 30 °C. The exchange experiments were performed at two different pH values for each protein conformation. The pH* values measured at the end of the exchange experiments were 7.0 and 7.6 (TNAK), 7.1 and 7.8 (ECAK), 7.0 and 7.5 (TNAK*Ap5A), and 7.0 and 7.6 (ECAK*Ap5A). For each exchange experiment, the start time was taken to be the time of protein-buffer mixing. The dead time of the experiment (i.e., from the time that the protein and buffer were mixed to the start of spectrum recording) was approximately 4-5 min. The last spectrum recorded was approximately 46 hours after initiating the exchange.

The ^{1}H-^{15}N HSQC spectra were collected on a Varian Unity Inova 600 MHz spectrometer equipped with a triple resonance probe. Spectral widths were 8,000 Hz and 1,900 Hz in the ^{1}H and ^{15}N dimensions, respectively. With 32 x 1956 (t_1 x t_2) complex

points and four transients per t_1 increment, the total acquisition time for each spectrum was 5 min and 18 sec.

4.2.3 H-D exchange rate analysis

Data were processed using NMRPipe (7) and NMRView (18). A Gaussian window function in the direct dimension and a sine window function in the indirect dimension were applied. Phase and baseline corrections were used in all cases. The HSQC spectra were zero-filled to 2048 x 512 data points in ^{1}H and ^{15}N dimensions, respectively. The H-D exchange rates were quantified using the intensity of the ^{1}H-^{15}N amide cross-peaks in a series of HSQC spectra recorded at different times. For each amide, i, the time dependency of peak intensity was fit to a single exponential decay function equation, $I(t) = I_0(t) + I_i \exp(-k_{ex}t)$, using the program CURVEFIT (Arthur Palmer group, Columbia University). CURVEFIT uses a least-squares fitting algorithm to yield the exchange rate, k_{ex}. The intrinsic rate constant (Bai factor), k_{ch}, for each amide proton at 30 °C and a specific pH was predicted from the protein sequence using the program HXPrep (courtesy Zhonggyi Zhang), which employs the procedure described in (Bai et al, 1993). The calculated k_{ch} values and the measured k_{ex} values were used to calculate the protection factors. For the theory underlying EX1 and EX2 exchange mechanisms, see (2, 4, 47).

4.3 Results and Discussion

4.3.1 The exchange mechanism at the experimental pH* values

Native state H-D exc hange in TNAK and ECAK at 30 °C was followed by collecting ^{1}H-^{15}N HSQC NMR spectra at different time points for 46 h. Exchange experiments were performed at two pH values for each protein. TNAK is most stable between pH 7.0 and pH 8.5, and its activity at pH 6.5 drops to 50% of its optimal activity at pH 7.5.(43). For these reason, we chose 7.0 and 8.0 as the target experimental pH values for the two enzymes. The final pH* of the deuterated samples were 7.0 and 7.6 for TNAK and 7.1 and 7.8 for ECAK. Twenty four of the ECAK amide protons with measurable exchange rates showed EX2 exchange at 30 °C, with higher exchange rates at pH* 7.8 than at pH 7.1. The ratio of exchange rates for these amide protons at pH* 7.8 and pH* 7.1 was consistent with a ten-fold increase in exchange rate for every pH unit increase (i.e., the pH-dependent increase expected for EX2 exchange mechanism). Fourteen ECAK amide protons showed similar exchange rates at the two pH* values, indicating that these protons likely exchange via the EX1 mechanism. Similarly, twenty TNAK amide protons did not show any difference in their exchange rates, within error, at pH* 7.0 and pH* 7.6 indicating an EX1 exchange mechanism for these protons. All other TNAK amide protons with measurable exchange rates (i.e., 23) showed EX2 exchange in these conditions. The pH dependence of exchange rates was also determined for ECAK*Ap5A and TNAK*Ap5A. The pH* values of the deuterated samples were 7.0 and 7.5 for TNAK*Ap5A and 7.0 and 7.6 for ECAK*Ap5A. As for the ligand-free proteins, 19 (out of 47) ECAK*Ap5A amide protons and 28 (out of 41) TNAK*Ap5A amide protons did

not show any pH-dependent exchange rate variation. Since H-D exchange is base-catalyzed, EX2 is generally the dominant exchange mechanism between pH 4.0 and pH 7.0 (3, 19, 29). The EX1 mechanism is usually observed at alkaline pH values (usually above pH 9.0), where k_{ch} exceeds k_{cl}. Since the experiments reported here were performed between pH 7.0 and 8.0, some residues are likely to show a mixture of EX1 and EX2 mechanisms or predominantly EX1 mechanism of exchange.

4.3.2 Determination of P values

For both EX1 and EX2 mechanisms, the exchange rate can be measured by fitting a single exponential function to peak intensities, and the protection factors P can be calculated using equation [1]:

$$P = k_{ch}/k_{ex} = k_{cl}/k_{op} = 1/K_{eq} \qquad [1]$$

where k_{op} and k_{cl} are the rate constants for conformational opening and closing in the scheme [2] (2, 4, 47):

$$Closed \underset{k_{cl}}{\overset{k_{op}}{\Leftrightarrow}} Open \xrightarrow{k_{ch}} Exchanged \qquad [2]$$

The P values reported and discussed in this work were calculated from the exchange data collected at pH* 7.0. P factors were calculated for only those amide protons with measurable exchange rates, since some exchanged within the dead time of the experiment while some others remain unexchanged during the time course of the experiments (~46 hours). Exchange data for 75, 70, 60, and 93 residues in ECAK, TNAK, ECAK*Ap5A,

and TNAK*Ap5A, respectively, could not be obtained. The missing exchange data can be attributed to two reasons: (i) not all residues could be detected in the HSQC spectrum during the initial NMR assignments of ECAK (202 assigned [37]), ECAK*Ap5A (191 assigned [37]), TNAK (197 assigned [Chapter 3]) and TNAK*Ap5A (193 assigned [Chapter 3]); (ii) since HSQC spectra for the exchange experiments were recorded for only 5 min, resolution was compromised and several crosspeaks overlapped. H-D exchange rates for the overlapped peaks were not quantified.

The percentages reported in this paragraph were calculated based on the total number of residues having exchange data. Overall, TNAK has a significantly higher percentage of unexchanged residues (20%) than ECAK (12%). Conversely, a smaller percentage of TNAK residues (51%) exchange in the dead time of the experiments than in ECAK (64%). If unexchanged residues are indicative of high conformational rigidity, and if very fast exchange indicates high flexibility, these results suggest that TNAK is more rigid than ECAK. Although the increase in percentage of protected amides in ECAK*Ap5A relative to ECAK is modest (only 3%), the percentage of unexchanged amide protons increases significantly (by 7%) from ECAK to ECAK*Ap5A, implying increased rigidity for inhibitor-bound form of ECAK. The increase in the percentages of unexchanged amides (2%) and amides with some protection factor (3%) from TNAK to TNAK*Ap5A are smaller than between ECAK and ECAK*Ap5A. Still, they suggest that TNAK*Ap5A is at least as rigid, if not more rigid, than TNAK. Not surprisingly, TNAK*Ap5A is overall more rigid than ECAK*Ap5A (54% of the amides are protected in TNAK*Ap5A vs. 46% in ECAK*Ap5A). The following sections discuss in detail the distribution of protected residues in the four enzyme forms.

4.3.3 The slowest exchanging amide hydrogens are in the ECAK and TNAK cores

The log of P values computed for the ECAK and TNAK residues with measurable amide proton exchange rates are shown in Figures IV-2A and IV-2B. Residues whose amide protons exchanged within the dead time of the NMR experiments were arbitrarily given a logP value of 0.1, indicating rapid exchange, to distinguish them from amide protons whose exchange rates could not be quantified. Residues whose amide protons did not exchange within the time period of the experiment were arbitrarily assigned a logP value of 7.

Figures IV-2A and IV-2B show that more amide protons in TNAK than in ECAK are unexchanged or have some level of protection from exchange. The domain distribution of protected amides in all four enzyme forms is summarized in Table IV-1. In ECAK, 59% of the core residues show some level of protection, while only 21.4% and 3% of the AMPbd and lid residues, respectively any protection. This clear demarcation between a relatively rigid core and very flexible AMPbd and lid domains is absent in TNAK, where residues with moderate to high protection are distributed more uniformly throughout the protein. Table IV-1 suggests that not all regions of the protein contribute to the higher stability of TNAK to the same extent. Indeed, the percentages of ECAK and TNAK core amides which show some level of protection are very similar, indicating similar rigidity in the two enzyme cores. The increase in rigidity occurs predominantly in TNAK's AMPbd and lid domains, which show 2.2-fold (AMPbd) and 18-fold (lid) more amide protons that are protected compared to ECAK.

The slowest exchanging amide protons in native proteins are typically involved in

Figure IV-2: Sequence dependence of log of protection factors, P (■) in (A) ECAK (B) TNAK (C) ECAK*Ap5A and (D) TNAK*Ap5A. S^2 values (○) for ECAK and ECAK*Ap5A (35) and for TNAK and TNAK*Ap5A (Chapter 3) are shown. Residues that exchanged within the experimental dead time have been assigned an arbitrary logP of 0.1 to distinguish them from residues whose exchange rate could not be quantified. Residues that did not exchange during the time course of the experiments have been assigned an arbitrary logP value of 7. A cartoon indicating the location of secondary structures and domains is indicated above the data.

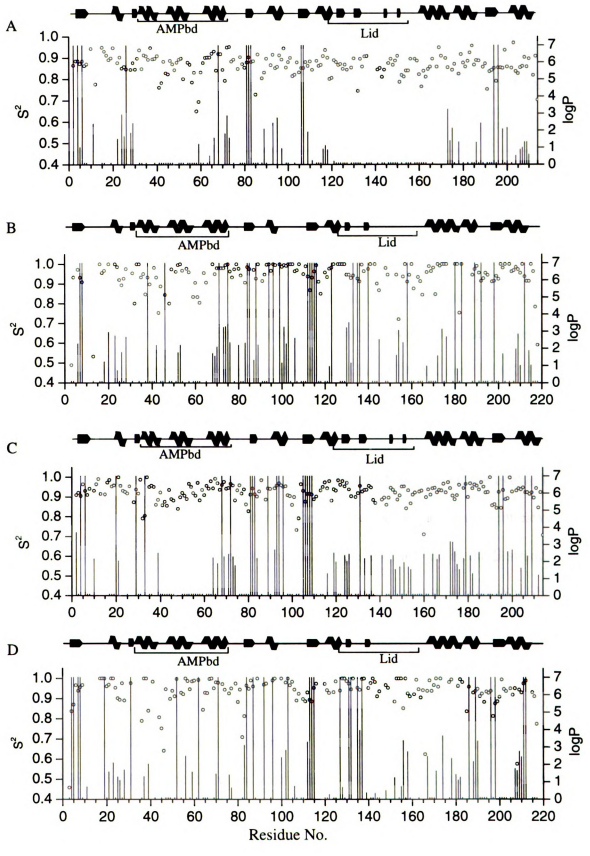

A
AMPbd
Lid
B
AMPbd
Lid
C
AMPbd
Lid
D
AMPbd
Lid
S²
logP
Residue No.

Table IV-1 Distribution of protected amides in the free and inhibitor forms of ECAK and TNAK[a].

	ECAK		TNAK		ECAK*Ap5A		TNAK*Ap5A	
	A	B	A	B	A	B	A	B
Core	59	16.6	56	22.5	56.5	29	60	29
AMPbd	21.4	3.6	41	13.8	25	9.3	37	14
Lid	3.7	0	46	21.4	51.6	3.2	61	22

[a]Percentages were calculated based on the number of residues in each domain that have exchange data

A: % residues with some protection
B: % unexchanged amide protons

well-ordered H-bond networks, and/or they are buried and have restricted solvent access. Exchange at these amides typically takes place only during global unfolding events. Although the percentage of protected amides are similar in ECAK and TNAK' core, the only suggestion that TNAK's core is more stabilized than ECAK's core comes from comparison of the percentage of unexchanged amide protons. Only 17% of ECAK core's amide protons are unexchanged compared to 23% in TNAK's core. Of the twelve ECAK amide protons that did not exchange during the time course of the experiment eleven are located in the core domain. Of these eleven amide protons, ten are in the β strands, and one (Ile26) is in a loop. The only unexchanged amide proton not belonging to the core (Val68) is located in helix α_4 of ECAK's AMPbd domain. Using the ECAK crystal structure (pdb code: 4AKE), we computed the best possible H-bond network in the ECAK free form using the Optimal H-bonding Network module of the software WHAT IF (15, 32). The WHAT IF results indicate that all unexchanged ECAK amide protons are involved in H-bonds. Particularly, the core amide protons form main chain H-bonds with

other core residues, highlighting the importance of these interactions in the stability and rigidity of the core domain.

Of the 30 TNAK amide protons that did not exchange, 21 belong to core domain residues. Although, as in ECAK, a majority of these 21 residues belong to the core β-strands, eight reside in the core loops and α-helices, indicating an additionally rigidified core domain in TNAK. Each of these 21 core amide protons is involved in at least one H-bond with the main chain of another core residue. Some additionally participate in H-bonds with side chains.

That TNAK's lid domain is much more rigid than ECAK's lid is evidenced not only by the percentage of protected amides but also from the number of unexchanged amides in this domain. There are five unexchanged amide protons in TNAK's lid while there are none in ECAK's lid. TNAK's lid domain has four cysteines (i.e., Cys133, Cys136, Cys153, and Cys156) that bind a Zn^{2+}, forming a Zn-finger-like structure. The unexchanged TNAK lid residues that are highly protected include Cys133, Lys135, Cys136, Tyr140, and Cys158. Of these residues, only Cys 133, Tyr 140, and Cys 158 are involved in H-bonds according to our WHAT IF analysis. In contrast, the corresponding ECAK residues, which are also involved in H-bonds are not protected. The fact that these TNAK amide protons remain unexchanged could also be a consequence of solvent inaccessibility. The presence of these unexchanged residues indicates a highly stabilized lid domain in TNAK relative to ECAK. In ECAK, the Zn-finger-like structure is replaced by an H-bond network that may not be as stabilizing as the Zn-finger. The structural role of Zn^{2+} as a stabilizing factor has been studied previously by site-directed mutagenesis by creating a quadruple cysteine mutant ECAK (C_4-ECAK) capable of binding Zn^{2+}.

Calorimetric studies showed that C_4-ECAK is significantly more stable (T_m increases from 51.8 °C to 63 °C) than the wild-type ECAK (5, 30). Correspondingly, removal of Zn^{2+} from TNAK reduces its melting temperature by nearly 6°C to 93.5°C (43). Therefore, while Zn^{2+} is not the sole contributor to TNAK's thermostability, it may play an important role in stabilizing the lid.

In addition to the core and lid domains, TNAK's AMPbd domain also has four residues (Arg38, Asp46, Val71, and Leu75) belonging to helices α_2, α_3, and α_4 that have unexchanged amide protons. These amide protons are involved in the same kind of H-bonds as the corresponding residues in ECAK. In spite of this similarity, only the amide proton of ECAK Val68 exchanges slowly. Moderate to no protection at all is observed for the other three corresponding ECAK AMPbd amide protons (i.e., Arg36, Ser43, and Ile72). The results from the H-D exchange analyses of ECAK and TNAK are summarized in Figures IV-3A and IV-3B.

4.3.4 Inhibitor-binding stabilizes ECAK's AMPbd and lid domains

Figures IV-2C and IV-2D show the plots of residue number vs. logP values for ECAK*Ap5A and TNAK*Ap5A. Again, the logP values of unexchanged amide protons are set to 7 and those of amide protons that exchanged within the dead time of the experiments are set to 0.1.

Compared to ECAK, ECAK*Ap5A has 10% more residues that show protection against hydrogen exchange. A total of 25 residues in ECAK*Ap5A show no exchange during the time course of the experiment. The core in ECAK*Ap5A is

Figure IV-3 Protection factors summarized for (A) ECAK, (B) TNAK, (C) ECAK*Ap5A, and (D) TNAK*Ap5A backbones as per the scales shown. Residues that did not exchange within the time course of the experiments are colored blue while those that exchanged within the dead time of the experiments are colored red. Residues whose exchange rates could not be quantified are colored grey. Figures were generated using insight II.

A
B
C
D
0.0 0.4 0.7 1.1 1.4 1.8 2.2 2.5 2.9 3.2 3.6
0.0 0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0

somewhat more rigidified with 29% of amide protons unexchanged, compared to ~17% in ECAK. All the core residues whose amide protons were unexchanged in the free form continue to be protected in the bound form. In addition, two amide protons (Arg206 and Ala209) in the C-terminal core helix (α_9), are also unexchanged.

The most predominant difference between ECAK and ECAK*Ap5A is in the lid domain (Table IV-1). While all but one ECAK lid residues exchange during the dead time of the experiment, ~52% of ECAK*Ap5A lid residues have some level of protection, and Arg131 remains unexchanged. The Arg131 amide proton has one extra H-bond with His126 that is absent in ECAK. Only 3 of 16 partially protected amides have H-bonds that are absent in the free enzyme form. The enhanced protection of these residues could be because they are buried upon inhibitor binding, thereby limiting solvent accessibility.

In contrast to the lid domain, the AMPbd domain shows only marginal stabilization in ECAK*Ap5A and this stabilization is limited to a few residues in helices α2 and α4. Residues between 40 and 63 exchange within the dead time of the experiment indicating the high flexibility of this region in the inhibitor-bound form also. The increase in the percentage of protected amides in the AMPbd domain is due to the unexchanged amide protons of Asp33, Val68, and Arg71. It is interesting to note that none of these amides form H-bonds with Ap5A, and that many of their intra-molecular H-bonds are already present in ECAK. Their protection in the inhibitor-bound form could result from an increase in the stabilization of the AMPbd domain that limits solvent accessibility.

Overall, the inhibitor-bound form of ECAK has protection factors for more residues than the free form (Table IV-1) indicating increased conformational stability of ECAK*Ap5A.

In TNAK*Ap5A, the percentage of amide protons that show any protection from exchange only marginally increases compared to TNAK (~5%). Several of TNAK's core residues remain unexchanged in TNAK*Ap5A as well, due to many of the same H-bond interactions that are preserved in both conformational states. Even though fewer residues have exchange data in TNAK*Ap5A (127 residues) than TNAK (150 residues), the protection distribution is highly similar in the cores of TNAK and TNAK*Ap5A. While the AMPbd domain is not additionally stabilized in TNAK*Ap5A, the lid domain is more rigid in TNAK*Ap5A than in TNAK. Except for Cys136, none of the newly protected amides forms any new (i.e., not found in TNAK) intra-molecular H-bonds nor do they form H-bonds with Ap5A.

For both the AMPbd and lid domains, the distribution of unexchanged amide protons is different in the two TNAK forms. For example, lid residue 158 that was unexchanged in TNAK is only marginally stabilized in TNAK*Ap5A, and residue 137 that has no protection against exchange in TNAK is fully protected in TNAK*Ap5A. Similarly, the AMPbd residue 62 that has no protection in TNAK is fully protected in TNAK*Ap5A. None of TNAK*Ap5A's unexchanged lid and AMPbd residues seem to H-bond with the inhibitor.

Taken together, the results show that while ECAK is significantly rigidified upon inhibitor binding, the same is not true for TNAK. The only common factor in these two enzymes is the significant increase in the rigidity of the lid domains, especially for

ECAK. The AMPbd domain is only marginally stabilized in ECAK, while in TNAK there is almost no change upon inhibitor binding. In both enzymes, but particularly in ECAK, several AMPbd amide protons exchange within the dead time of the experiments indicating considerable local fluctuations in this domain.

Taken together, the results show that not only is TNAK not uniformly stabilized compared to ECAK, the two enzymes are not uniformly stabilized upon inhibitor binding. In the free form, the lid and AMPbd domains are considerably more rigid in TNAK than ECAK while in the closed forms of both enzymes; the lid domain becomes more rigid upon inhibitor binding. Results from the H-D exchange analysis of ECAK*Ap5A and TNAK*Ap5A are summarized in figures IV-3C and IV-3D.

4.3.5 Hinge regions

The AMPbd-core hinge regions consist of nine residues of which 5, 6, 8, and 4 residues have exchange data in ECAK, TNAK, ECAK*Ap5A, and TNAK*Ap5A, respectively. The lid-core hinge regions consist of a total of 22 residues and 13, 16, 14, and 13 of these residues have exchange data in ECAK, TNAK, ECAK*Ap5A, and TNAK*Ap5A, respectively. As seen from Figure IV-2, amide protons of 3 out of the 5 residues with exchange data in ECAK's AMPbd-core hinge residues and 4 out of the 6 residues with exchange data in TNAK's AMPbd-core hinges have no protection from exchange (i.e., have a logP value of 0.1 in Figure IV-2). The remaining residues in this region in both the enzymes have logP values less than 3.0. In the lid-core hinges, the amide protons of 10 out of 13 residues in ECAK and 11 out of 16 residues in TNAK exchange within the dead time of the experiments. As was noted for the AMPbd domain, amide protons of the

remaining residues in the lid-core hinges of both enzymes have logP values less than 3.0. In the inhibitor-bound form of ECAK, 5 out of 8 amide protons in the AMPbd-core hinges and 8 out of 14 amide protons in the lid-core hinges have no protection from exchange. Corresponding numbers for TNAK*Ap5A are: 1 out of 4 in the AMPbd-core hinges and 8 out of 13 in the lid-core hinges. Interestingly in both ECAK*Ap5A and TNAK*Ap5A, a single residue (Ile 29 in ECAK*Ap5A and Ile31 in TNAK*Ap5A) in the AMPbd-core hinge remains unexchanged during the experiments. This Ile is H-bonded to an Asp (Aps84 in ECAK*Ap5A and Asp 87 in TNAK*Ap5A) that is important for coordinating Mg^{2+}.Since fast exchanging residues indicate high flexibility, especially if not solvent exposed, the high density of unprotected residues in the hinges indicate high flexibility for the AMPbd-core and lid-core hinge regions in both ECAK and TNAK in the free and inhibitor bound forms.

4.3.6 Hydrogen exchange and ^{15}N NMR relaxation

The backbone dynamics of TNAK and TNAK*Ap5A in the ps-ns timescale at 30 °C have been characterized using ^{15}N NMR relaxation rate measurements (Chapter 3). Backbone dynamics of ECAK and ECAK*Ap5A at 30 °C studied with ^{15}N NMR relaxation have been reported previously (38). They were used for comparison with the TNAK and TNAK*Ap5A results (Chapter 3). On the ps-ns timescale, TNAK is uniformly more rigid than ECAK, particularly in the core domain. This rigidity difference is abolished in ECAK*Ap5A and TNAK*Ap5A, due a rigidity increase in ECAK upon inhibitor binding. Although TNAK is more rigid than ECAK, several residues in the TNAK AMPbd and lid domains exhibit lower than average S^2 values consistent with the higher flexibility of these domains (Chapter 3). To determine if the trends for S^2 values and the

trends for P values from hydrogen-exchange coincide, the ^{15}N NMR relaxation order parameters and the logP values from amide hydrogen exchange for ECAK, TNAK, ECAK*Ap5A, and TNAK*Ap5A are plotted in Figures IV-2A, IV-2B, IV-2C, and IV-2D, respectively.

Comparing the ^{15}N relaxation dynamics and H-D exchange data on ECAK and TNAK brings up some interesting differences in the two enzyme cores. The average ^{15}N relaxation order parameters ($\langle S^2 \rangle$) show that ECAK and TNAK's cores are only marginally more rigid in the ps-ns timescale than the rest of the proteins. The $\langle S^2 \rangle$ values of the core, AMPbd, and lid domains increase by a similar amount (8.5%, 7.7%, and 9.5%, respectively) going from ECAK to TNAK, suggesting a uniform stabilization across the protein fold of the hyperthermophilic enzyme in the ps-ns timescale. The H-D exchange data, on the other hand, show that ECAK's core is significantly more rigid than the rest of the protein with the rigidity decreasing in the order core > AMpbd >> lid. Moreover, TNAK's core is similar in rigidity to that of ECAK in the ≥ ms timescale. The only suggestion that TNAK's core is more rigid than that of ECAK resides in the fact that 22.5% of TNAK core amides remain unexchanged, against only 16.6% in ECAK's core. TNAK's increased rigidity is most significant in the lid domain, followed by the AMPbd domain. Hence, unlike on the ps-ns timescales, the increase in rigidity of TNAK on the ≥ ms timescale is localized rather than uniform. In the μs timescale (determined by the exchange contribution (R_{ex}) to the transverse relaxation rate [T_2]), many more residues in ECAK's core, particularly the β-strands, have R_{ex} values than in TNAK's core, indicating a higher rigidity of the TNAK core in the μs timescale.

It is evident from both the ^{15}N NMR relaxation and H-D exchange data that ECAK's AMPbd domain is characterized by local fluctuations. For example, the S^2 values of several residues (i.e., for residues 43-45, 61, and 62) are lower than average. This domain also contains contiguous residues showing model 5-type, sub-ns to ns timescale motions ([38], not shown]. In the H-D exchange data also, ECAK residues 30 through 65 are completely unprotected, their amide protons exchanging within the dead time of the exchange experiments, indicating exchange through local fluctuations. Although overall, TNAK's AMPbd domain is more rigid than ECAK's AMPbd domain, it also has several residues showing low S^2 values (i.e., 32, 43, 46, 48, and 59-62), the S^2 value being similar in both enzymes in many cases (Chapter 3). The H-D exchange data corroborates these results because several amide protons in the region between 33 and 65 in TNAK's AMPbd domain are unprotected or show only moderate protection. Hence ECAK and TNAK's AMPbd domains feature dynamics in ps to $\geq$ ms timescales.

ECAK's lid domain, on the other hand, does not show any significant trend in S^2 values. Several residues, though, show R_{ex} values indicating significant μs timescale dynamics. The lid domain is also the least protected from exchange in ECAK indicating pervasive local fluctuations in the $\geq$ ms timescales. Residues in TNAK's lid also show R_{ex} values, although these are fewer in number in TNAK than in ECAK. H-D exchange data show the TNAK lid to be more rigid than ECAK's lid (likely due to the presence of Zn^{2+}), although several residues in this region show only moderate protection. Hence, ECAK and TNAK's lids are likely to be characterized by motions mainly in timescales greater than μs.

4.4 Conclusions

Here we have compared the conformational stability of a mesophilic and a hyperthermophilic AK using H-D exchange monitored by NMR. Protection factors derived from the exchange rates have been used as a measure of conformational dynamics in the two proteins. The core domains of ECAK and TNAK are similarly stable as indicated by the percentage of protected amides in the two proteins. The increase in the rigidity of TNAK, compared to ECAK is not uniform across all domains. Instead the lid and AMPbd domains are proportionally more rigid in TNAK. This result is different from the one obtained from H-D exchange analyses of EcRNase and TtRNase, where the relative stability increase of all TtRNase subdomains/substructures is the same (14). These results reinforce the fact that stabilization mechanisms vary from one protein family to another, and that they are not necessarily uniformly distributed around the structure. The dynamics in ECAK and TNAK domains also are not similar across all timescales. While the AMPbd domain in both enzymes shows fluctuations ranging from ps to ≥ ms timescale, motions in ≥ μs timescales are predominant in the lid domains of ECAK and TNAK. The pattern of rigidity increase in TNAK is also different at different timescales. Whereas it is uniformly more rigid than ECAK in the ps-ns timescales, only its lid and AMPbd domains are more rigid in ≥ ms timescales. The presence of Zn^{2+} in TNAK's lid domain plays an important role in the local stability of that domain. In essence, the increased rigidity of TNAK is not uniform across all the domains nor is it uniform at all timescales.

The fact that the lid and AMPbd domains are rigid structures in TNAK does not preclude them from having flexible regions and flexible hinges. This result is important in light of the fact that TNAK is as active as ECAK at 30°C. The hinge regions are important for the movement of the lid and AMPbd domains, and the domains themselves can be rigid without compromising domain closing and opening. TNAK, therefore, is an excellent example of a hyperthermophilic enzyme where rigidity for stability and flexibility for activity are partitioned in a manner that makes it simultaneously active at low temperatures and stable at high temperatures.

Although the protection factors are used here as reports of the conformational stability of ECAK and TNAK, the energetics of the stability of the two enzymes have not been determined. Future directions of this research include measurement of dependence of H-D exchange rates in ECAK and TNAK on denaturant concentrations and temperature. These measurements can be used to fully determine residue-specific free energies of unfolding.

4.5 References

1. **Šali., A., and T. L. Blundell.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234:**779-815.

2. **Arrington, C. B., and A. D. Robertson.** 2000. Kinetics and thermodynamics of conformational equilibria in native proteins by hydrogen exchange. Methods Enzymol. **323:**104-124.

3. **Arrington, C. B., and A. D. Robertson.** 1997. Microsecond protein folding kinetics from native-state hydrogen exchange. Biochemistry **36:**8686-8691.

4. **Bai, Y. W., J. J. Englander, L. Mayne, J. S. Milne, and S. W. Englander.** 1995. Thermodynamic parameters from hydrogen exchange measurements. Methods Enzymol. **259:**344-356.

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

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

7. **Delaglio., F., S. Grzesiek., G. W. Vuister., G. Zhu., J. Pfeifer., and A. Bax.** 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6:**277-293.

8. **Eidsness, M. K., K. A. Richie, A. E. Burden, D. M. Kurtz, and R. A. Scott.** 1997. Dissecting contributions to the thermostability of *Pyrococcus furiosus* rubredoxin: β-sheet chimeras. Biochemistry **36:**10406-10413.

9. **Ferraro, D. M., and A. D. Robertson.** 2004. EX1 hydrogen exchange and protein folding. Biochemistry **43:**587-594.

10. **Fiser, A., R. K. Do, and A. Sali.** 2000. Modeling of loops in protein structures. Protein Sci. **9:**1753-1773.

11. **Hayward, S.** 2004. Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. J. Mol. Biol. **339:**1001-1021.

12. **Hernandez, G., F. E. Jenney, Jr., M. W. Adams, and D. M. LeMaster.** 2000. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. Proc. Natl. Acad. Sci. USA **97:**3166-3170.

13. **Hernandez, G., and D. M. LeMaster.** 2001. Reduced temperature dependence of collective conformational opening in a hyperthermophile rubredoxin. Biochemistry **40:**14384-14391.

14. **Hollien, J., and S. Marqusee.** 1999. Structural distribution of stability in a thermophilic enzyme. Proc. Natl. Acad. Sci. USA **96:**13674-13678.

15. **Hooft, R. W. W., C. Sander, and G. Vriend.** 1996. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins: Struct., Funct., Genet. **26:**363-376.

16. **Jaenicke, R.** 1996. Glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: strategies of protein stabilization. FEMS Microbiol. Rev. **18:**215-224.

17. **Jaenicke, R.** 1991. Protein stability and molecular adaptation to extreme conditions. Eur. J. Biochem. **202:**715–728.

18. **Johnson, B. A., and R. A. Blevins.** 1994. NMRView: A computer program for the visualization and analysis of NMR data. J. Biomol. NMR **4:**603-614.

19. **Kahsai, M. A., E. Martin, S. P. Edmondson, and J. W. Shriver.** 2005. Stability and flexibility in the structure of the hyperthermophile DNA-binding protein Sac7d. Biochemistry **44:**13500-13509.

20. **Kawamura, S., Y. Kakuta, I. Tanaka, K. Hikichi, S. Kuhara, N. Yamasaki, and M. Kimura.** 1996. Glycine-15 in the bend between two α-helices can explain the thermostability of DNA binding protein HU from *Bacillus stearothermophilus*. Biochemistry **35:**1195-1200.

21. **Kawamura, S., I. Tanaka, N. Yamasaki, and M. Kimura.** 1997. Contribution of a salt bridge to the thermostability of DNA binding protein HU from *Bacillus stearothermophilus* determined by site-directed mutagenesis. J. Biochem. **121:**448-455.

22. **Kumar, S., and R. Nussinov.** 2001. How do thermophilic proteins deal with heat. Cell Mol. Life Sci. **58:**1216-1233.

23. **Kumar, S., C. J. Tsai, and R. Nussinov.** 2000. Factors enhancing protein thermostability. Protein Engin. **13:**179-191.

24. **Lazaridis, T., I. Lee, and M. Karplus.** 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci. **6:**2589–2605.

25. **Liang, Z. X., T. Lee, K. A. Resing, N. G. Ahn, and J. P. Klinman.** 2004. Thermal-activated protein mobility and its correlation with catalysis in thermophilic alcohol dehydrogenase. Proc. Natl. Acad. Sci. USA **101:** 9556-9561.

26. **Liang, Z. X., I. Tsigos, T. Lee, V. Bouriotis, K. A. Resing, N. G. Ahn, and J. P. Klinman.** 2004. Evidence for increased local flexibility in psychrophilic

alcohol dehydrogenase relative to its thermophilic homologue. Biochemistry **43:**14676-14683.

27. **Muller, C. W., G. J. Schlauderer, J. Reinstein, and G. E. Schulz.** 1996. Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. Structure **4:**147-156.

28. **Muller, C. W., and G. E. Schulz.** 1992. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution - a model for a catalytic transition-state. J. Mol. Biol. **224:**159-177.

29. **Neira, J. L., P. Sevilla, M. Menendez, M. Bruix, and M. Rico.** 1999. Hydrogen exchange in ribonuclease A and ribonuclease S: Evidence for residual structure in the unfolded state under native conditions. J. Mol. Biol. **285:**627-643.

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

31. **Reinstein, J., M. Brune, and A. Wittinghofer.** 1988. Mutations in the nucleotide binding loop of adenylate kinase of *Escherichia coli.* Biochemistry **27:**4712–4720.

32. **Rodriguez, R., G. Chinea, N. Lopez, T. Pons, and G. Vriend.** 1998. Homology modeling, model and software evaluation: three related resources. Bioinformatics **14:**523-528.

33. **Russell, R. J. M., J. M. C. Ferguson, D. W. Hough, M. J. Danson, and G. L. Taylor.** 1997. The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. Biochemistry **36:**9983-9994.

34. **Sambrook, J., E. F. Fritsch, and T. Maniatis (ed.).** 1989. Molecular cloning: A laboratory manual., 2nd edition ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

36. **Schulz, G. E., C. W. Muller, and K. Diederichs.** 1990. Induced-fit movements in adenylate kinases. J. Mol. Biol. **213:**627–630.

37. **Shapiro, Y. E., E. Kahana, V. Tugarinov, Z. C. Liang, J. H. Freed, and E. Meirovitch.** 2002. Domain flexibility in ligand-free and inhibitor-bound *Escherichia coli* adenylate kinase based on a mode-coupling analysis of ^{15}N spin relaxation. Biochemistry **41:**6271-6281.

38. **Shapiro, Y. E., M. A. Sinev, E. V. Sineva, V. Tugarinov, and E. Meirovitch.** 2000. Backbone dynamics of *Escherichia coli* adenylate kinase at the extreme

stages of the catalytic cycle studied by ^{15}N NMR relaxation. Biochemistry **39:**6634-6644.

39. **Sivaraman, T., C. B. Arrington, and A. D. Robertson.** 2001. Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. Nat. Struct. Biol. **8:**331-333.

40. **Sivaraman, T., and A. D. Robertson.** 2001. Kinetics of conformational fluctuations by EX1 hydrogen exchange in native proteins. Meth. Molec. Biol. **168:**193-214.

41. **Szilagyi, A., and P. Zavodszky.** 1995. Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: analysis based on homology modelling. Protein Engin. **8:**779-789.

42. **Tomschy, A., G. Bohm, and R. Jaenicke.** 1994. The effect of ion-pairs on the thermal stability of D-glyceraldehyde 3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*. Protein Engin. **7:**1471-1478.

43. **Vieille, C., H. Krishnamurthy, H. H. Hyun, A. Savchenko, H. Yan, and J. G. Zeikus.** 2003. *Thermotoga neapolitana* adenylate kinase is highly active at 30°C. Biochem. J. **372:**577–585.

44. **Vieille, C., and J. G. Zeikus.** 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. Rev. **65:**1–43.

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

46. **Vonrhein, C., G. J. Schlauderer, and G. E. Schulz.** 1995. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure **3:**483-490.

47. **Wildes, D., and S. Marqusee.** 2004. Hydrogen-exchange strategies applied to energetics of intermediate processes in protein folding. Methods Enzymol. **380**: 328-349.

48. **Xiao, L., and B. Honig.** 1999. Electrostatic contributions to the stability of hyperthermophilic proteins. J. Mol. Biol. **289:**1435-1444.

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

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

Chapter 5

Molecular dynamics simulation of *Escherichia coli* adenylate kinase complexed with its substrates

Research presented in this chapter was published as the following manuscript:

H, **Krishnamurthy, H. F. Lou, A. Kimple, C. Vieille, and R. I. Cukier.** (2005). Associative mechanism for phosphoryl transfer: A molecular dynamics simulation of Escherichia coli adenylate kinase complexed with its substrates. Proteins-Structure Function and Bioinformatics **58:**88-100.

5.1 Abstract

The ternary complex of *Escherichia coli* adenylate kinase (ECAK) with its substrates adenosine monophosphate (AMP) and Mg-ATP, which catalyzes the reversible transfer of a phosphoryl group between adenosine triphosphate(ATP) and AMP, was studied using molecular dynamics. The starting structure for the simulation was assembled from the crystal structures of ECAK complexed with the bisubstrate analog diadenosine pentaphosphate (AP5A) and of *Bacillus stearothermophilus* adenylate kinase complexed with AP_5A, Mg^{2+} and 4 coordinated water molecules, and by deleting one phosphate group from AP5A. The interactions of ECAK residues with the various moieties of ATP and AMP were compared to those inferred from NMR, X-ray crystallography, site-directed mutagenesis, and enzyme kinetic studies. The simulation supports the hypothesis that hydrogen bonds between AMP's adenine and the protein are at the origin of the high nucleoside monophosphate (NMP) specificity of AK. The ATP adenine and ribose moieties are only loosely bound to the protein, while the ATP phosphates are strongly bound to surrounding residues. The coordination sphere of Mg^{2+}, consisting of 4 waters and oxygens of the ATP β- and γ-phosphates, stays approximately octahedral during the simulation. The important role of the conserved Lys13 in the P loop in stabilizing the active site by bridging the ATP and AMP phosphates is evident. The influence of Mg^{2+}, of its coordination waters, and of surrounding charged residues in maintaining the geometry and distances of the AMP α-phosphate and ATP β- and γ-phosphates are sufficient to support an associative reaction mechanism for phosphoryl transfer.

5.2 Introduction

Adenylate kinases (AKs) are small, monomeric phosphoryl transferases that catalyze the reversible transfer of a phosphoryl group from adenosine triphosphate (ATP) to adenosine monophosphate (AMP) by nucleophilic attack on the γ-phosphate of ATP (1). The net reaction is schematized as

$$Mg^{2+}\text{-ATP} + AMP \Leftrightarrow Mg^{2+}\text{-ADP} + ADP \qquad \text{(Scheme 1)}$$

AKs are composed of 3 domains: the core, LID, and AMP-binding (AMP-bd) domains (Figure V-1). In the absence of substrates, AKs are in an open form. In the presence of ATP, AMP, and Mg^{2+}, the LID and AMP-bd undergo major conformational rearrangements, resulting in the enzyme closing to form the ternary complex and expelling waters to prevent ATP and AMP hydrolysis (2). The best-known AK families are the short AKs represented by the mammalian cytosolic enzymes, and the long AKs represented by bacterial and mitochondrial enzymes. The long AKs contain a 20- to 30-residue insertion inside the LID domain (3, 4). Among the short AKs, only the porcine cytosolic AK has been crystallized in the absence of substrate [Protein Data Bank (PDB) code: 3ADK] (5). On the other hand, the long AKs from *Escherichia coli* (ECAK), *Bacillus stearothermophilus*, *Saccharomyces cerevisiae*, and bovine mitochondria have been crystallized with various substrate–inhibitor combinations or with the 5 phosphate bisubstrate analog, diadenosine pentaphosphate [i.e., AP_5A] (6–15). The crystal structures of the *E. coli* and bovine mitochondrial AKs in the absence of any ligands are also known (10, 16). ATP binds to the catalytic site as the Mg^{2+}ATP complex, in which the catalytically essential Mg^{2+} is coordinated to ATP's β- and -γ-phosphates and to 4 water

molecules. Thus, the functional catalytic assembly is a ternary complex. In this work, we focus on ECAK, since a wealth of kinetic (17), NMR (18–22), and mutagenesis (23–28) data are available for this enzyme. Crystallography can provide structures of enzymes complexed with products, with inhibitors that resemble substrates, and even with inhibitors that model transition states. In this work, we create a model of the ECAK – Mg-ATP - AMP ternary reaction complex based on the ECAK · AP_5A structure [PDB code: 1AKE] (6) and on the *B.stearothermophilus* AK (BSAK) structure crystallized with AP_5A and Mg^{2+} [PDB code: 1ZIO] (11). Constructing a reaction complex model allows one to study the stabilization of a possible catalytically competent complex that is more realistic than a model inferred from a crystal structure containing substrate mimics. Thus, the ternary complex model will be used as the starting point of a molecular dynamics (MD) simulation that can assess how ATP and AMP are maintained in a catalytically competent configuration with the aid of the surrounding protein, the Mg^{2+} ion, and its 4 coordinating waters. The highly charged complex suggests a focus on strong hydrogen bonds (H-bonds) and salt bridges that can serve the purpose of maintaining a catalytically relevant geometry. Two simulations of AKE have appeared previously. One, a 300-ps study of ECAK complexed with AP_5A, was carried out in the gas and solvated states (29). A number of residues that would be candidates for strong interaction with a Mg^{2+} ion were suggested in this study. The other simulation used a weighted masses MD method to explore the nature of the open conformation of apo-ECAK that could be reached starting from a number of closed form (with different bound substrates) crystallographic structures (30). The similarities and differences between various experimental conclusions based mainly on studies using AP_5A, and the

simulation of the ECAK · Mg-ATP · AMP structure are analyzed. While there are some differences in details, the suggested distinction between the high specificity of AKs for the AMP substrate and the lower specificity for ATP is supported by the simulation. Particular attention is devoted to investigating the interactions of the conserved Lys13 (the numbering refers to ECAK) located at the end of the P loop, with the AMP and ATP phosphate groups. The simulation shows that Lys13 does play an important role in substrate stabilization by bridging the AMP and ATP phosphates. There is great interest, in general, in the mechanism of phosphoryl transfers, whether it is an associative or a dissociative mechanism, these being extremes of a continuum of mechanisms qualitatively defined by the phosphorous–phosphorous distance, which increases in going from an associative to a dissociative mechanism (25, 31–34). There is evidence (32) that a dissociative mechanism is favored in solution, while an associative mechanism is favored in enzymes. The nature of the transition state (or reaction intermediate) also is an issue, with the most frequently suggested geometry invoking a pentavalent phosphorous (25, 31–34). As we discuss, the simulation carried out herein maintains the AMP and ATP terminal phosphates at a distance that supports an associative phosphoryl transfer mechanism.

5.3 Methods

5.3.1 Starting structure

The 1.9 Å[a] resolution crystal structure of the ECAK · Ap_5A complex (7) [PDB code: 1AKE] was used as the starting point for constructing an ECAK · AMP · Mg^{2+}ATP ternary complex, which will be denoted as 1AKE*. The 1AKE PDB file contains the coordinates of the 2 enzyme–inhibitor complexes (A and B) present in the crystal asymmetric unit. We used complex B to build 1AKE*. AP_5A (Figure V-1) is considered to be a bisubstrate mimic of the reactant state, since ECAK catalyzes the reaction of ATP and AMP to produce two ADPs (Scheme 1). ATP and AMP coordinates were obtained by eliminating AP_5A's δ-PO_4 group (see Figure V-1). While the 1AKE structure provides the AP_5A coordinates needed to create well-positioned ATP and AMP molecules within the protein, it does not contain the catalytic Mg^{2+} and its 4 waters of coordination. To create the starting structure for the MD simulation, we imported the coordinates of Mg^{2+} and its coordinated waters from the crystal structure of the BSAK · Mg^{2+}-AP_5A complex [PDB code: 1ZIO] (11). The 1ZIO and 1AKE structures were superimposed by .fitting 1ZIO's AP_5A molecule onto that of 1AKE. 1ZIO's Mg^{2+} and its 4 coordination waters were then merged with the ECAK · AMP · ATP structure to create 1AKE*. This structure contains Mg^{2+} positioned between nonbridging oxygens of ATP's β-and γ-phosphates and the 4 Mg^{2+} coordinating waters, as displayed in Figure V-1.

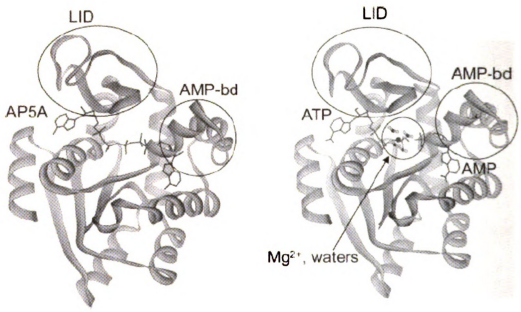


Figure V-1. Structures of 1AKE (**left**) and 1AKE* (**right**). ECAK is shown in ribbon representation. AP_5A, ATP, and AMP are in stick representation; the Mg^{2+} atom and its 4 coordinating waters are in ball-and-stick representation. ATP and AMP were created by deleting AP5A's δ-PO_4. Mg^{2+} and its coordinating waters were imported from the 1ZIO structure.

5.3.2 Force field

The GROMOS96 force-field was used for the MD simulations (35). This force field is a united atom force field with explicit polar hydrogens, and hydrogens on the phenylalanine, tryptophan, and tyrosine rings. The GROMOS96 force field contains parameters for all protein residues, as well as for ATP, but it does not include parameters for AMP. Parameters for AMP were created based on the ATP parameters (35) The force field is identical up to atom 23 (A05* as numbered on ATP). The parameters from 32 (APG) to 36 (AH3PG) in ATP are then used for AMP's monophosphate group. The ionization states of all residues were set to reflect a neutral pH. Histidines were assumed to be neutral, with the Nδ atom protonated, and the lysine, arginine, aspartate, and glutamate residues were assumed to be charged. To assess whether this charge assignment is reasonable in the binding pocket, where numerous positively charged residues bind the substrate phosphates, we adopted the following procedure. The accessible surface area (ASA) of each residue was evaluated using the Surface Racer program (36), for the residues in the 1AKE* complex and in the 1AKE* structure in the absence of the substrates. For each residue, the ASA in the presence of substrate was subtracted from the ASA in the absence of substrate. The nonzero ASA difference values then single out those residues that are buried at the protein–substrate interface. The charged residues buried at the protein–substrate interface, Lys13, Arg36, Asp84, Arg88, Arg119, Arg123, Arg156, and Asp158, have a net charge of +4. These charges are almost compensated, through various interactions discussed below, by the net charge of +2 from ATP (-3, with 1 terminal oxygen protonated), AMP (-1), and Mg^{2+} (+2). A look at the charged residues buried at the protein–substrate interface shows that, with the exception

of Arg123, all the ionizable groups on positively charged residues are engaged in charge–charge interactions with the substrate phosphates. All these charge–charge interactions are stable throughout the 3 ns simulation period (see below). The side-chain of Asp84 is engaged in stable H bonds with 2 of the waters coordinated to Mg^{2+}, namely, 219 and 221. Arg123 forms a salt bridge with Asp159, which is also stable during the simulation. Note that Asp159, along with its neighboring Asp158 at the C-terminal end of the LID domain, are conserved in all AKs implicating a crucial role for them in AKs. These 2 aspartates form salt bridges with Arg123 and Arg156 side-chains only in the substrate bound form. Müller and Schulz (7) suggested that salt bridges Arg123–Asp159 and Arg156–Asp158 (formed with the aid of ATP and AMP, which help in positioning the side-chains) are instrumental in the conformational change leading to LID closure. Thus, the conventional ionization states assumed in this study are appropriate, since almost all of the charged groups buried in the 1AKE* complex have their charges satisfied either through protein–substrate, or protein–protein interactions.

5.3.3 Molecular dynamics

The MD simulation was carried out using CUKMODY (37), a code designed for the efficient simulation of proteins and other large solutes. A combination of a cell index method with linked lists (38) and a Verlet neighbor list (39) is used to provide linear scaling with the number of atoms in the pair list routine, essential for the large systems considered here. For the Verlet neighbor list, the outer distance was r_l = 11.84 Å and the inner distance was r_c = 10.42 Å. The pair list was updated whenever any atom moved a distance greater than 0.5*(r_l - r_c), leading to updates roughly every 25 steps. Electrostatic interactions are evaluated using the charge-group method, to be consistent with the

parametrization of the GROMOS force field. The SHAKE algorithm (39) was used to constrain bond lengths permitting a 2 fs timestep. Periodic boundary conditions were used. The simulation was carried out at constant number, volume, and temperature (NVT), with velocity scaling to control the temperature at 303 K. The runs were all performed on a PC with dual 1.6GHz AMD Athlon processors. A 1 ns simulation takes about 2 weeks on a single processor.

5.3.4 Equilibration

The 1AKE* system was equilibrated according to the following protocol. A simulation box of side 59.19 Å was filled with an equilibrated sample of 6912 waters, and the 1AKE* structure, totaling 2162 atoms, was centered in the simulation cell. The waters that overlapped any 1AKE* atom were discarded if $r_{Oj} < \sigma_{Oj}$, where r_{Oj} is the distance between water oxygen (atom O) and protein atom (atom j), and σ_{Oj} is the van der Waals distance parameter. This procedure eliminated 1470 waters. One-body forces, with force constant k = 30 kcal/mol/Å^2, were used on all protein and ligand atoms for the initial 50,000 steps. This procedure allowed the waters to equilibrate with the protein and with each other. The van der Waals length (σ) and well depth (ϵ) values for the water oxygen-Mg^{2+} interactions were designed to maintain coordination throughout the simulations. To enforce the initial water coordination of Mg^{2+}, the parameters listed in Table V-1 were used to equilibrate the system and then modified as indicated for the duration of the simulation.

After the initial 50,000 steps with one-body forces on, the one-body force constant was linearly reduced to 0 over 30,000 steps. The system was then run for

Table V-1. Water-Mg^{2+} Van der Waals parameters for the Mg^{2+} coordination waters during the MD equilibration and simulation periods.

	Equilibrate with		Run with	
	$\sigma^{(a)}$ (Å)	$\varepsilon^{(a)}$ (K)	$\sigma^{(a)}$ (Å)	$\varepsilon^{(a)}$ (K)
Water 218	2.16137	50,000.0	2.07364	54,273.3
Water 219	2.07814	20,000.0	2.07364	54,273.3
Water 220	1.88086	20,000.0	2.07364	54,273.3
Water 221	2.19517	50,000.0	2.07364	54,273.3

[(a)]The van der Waals interaction between atoms i and j is

$V = 4\varepsilon_{ij}\left[\left(r_{ij}/\sigma_{ij}\right)^{12} - \left(r_{ij}/\sigma_{ij}\right)^{6}\right]$ where $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$

another 20,000 steps to make sure that the system's strain was sufficiently reduced to initiate unconstrained dynamics. A 0.5-ns trajectory was run at 273 K, after which the temperature was linearly brought to 303 K over 10,000 steps. After the temperature had reached 303 K, the MD production trajectory proceeded for 3 ns.

5.3.5 H-bond–salt Bridge analysis

Interactions of AMP and ATP with 1AKE* residues were analyzed through H-bond analysis. Donor and acceptor atoms within a maximum distance of 3.5 Å and a minimum angle of 120° were considered to form H-bonds. The data were analyzed every 0.8 ps. Interactions between a charged phosphate group of the substrates and the side-chain of a positively charged residue (arginine or lysine) were considered to be salt bridge–like. Oppositely charged groups within 4.5Å of each other were considered to be saltbridged.

5.4 Results and Discussion

5.4.1 Validation of simulation

The radius of gyration of the protein increases very slowly during the first 740 ps, with an average of 17.18 ± 0.09 Å during that period and, subsequently, suddenly increases and stabilizes to an average of 17.60 ± 0.11 Å during the remainder of the simulation, indicating that the overall dimension of the protein is stable. (The radius of gyration of 1AKE in the crystal structure is 16.635 Å.) The RMSD (i.e., root-mean-square deviation from the starting 1AKE* structure) for all C_α's increases slowly from 1.5 to 3.1 Å during

the first 800 ps of the production trajectory; then, between 1100 and 1300 ps, it increases more sharply to finally plateau between 4.0 and 5.0 Å during the rest of the simulation. The RMSD for the C_α's of the core domain (residues 1–30, 60–120, and 160–214) increases very slowly to 2.3 Å during the first 1 ns of the production trajectory, then plateaus between 2.1 and 2.75 Å before finally increasing after 2800 ps. The RMSD for the C_α's of the core plus the AMP-bd domains (residues 1–120 and 160–214) increases during the first 800 ps at the same rate as the RMSD for all C_α's, then oscillates between 2.5 and 3.4 Å during the next 2000 ps, before finally increasing after 2800 ps. These results indicate that the deviation from the crystal structure is not distributed uniformly along the sequence, and that most of the large deviations (in particular after the first 1200 ps) occur in the LID domain. To compare residue fluctuations in 1AKE* with those in the 1AKE crystal, RMSF values (i.e., positional root-mean square fluctuations from the average structure) were calculated for all C_α carbons as $RMSF_n = \langle (r_n - \langle r_n \rangle)^2 \rangle^{1/2}$, with $\langle r_n \rangle$ being the average position of the C_α carbon of residue n over time, and r_n being the position at every step. Simulation RMSF values are compared with 1AKE crystallographic RMSF values in Figure V-2. Crystallographic RMSF values were calculated from the B values using the equation $RMSF = \sqrt{3B/8\pi^2}$. Regions with low RMSF values (e.g., helices α_4 and α_8, and strand β_4) in the simulation coincide well with the regions with low crystallographic RMSF values. In contrast, the regions showing high simulation RMSF values and high crystallographic RMSF values do not always match each other. Residues 41–48 (the loop between helices α_2 and α_3), and 132–140 (the loop between strands βL_2 and βL_3) show high simulation RMSF values that were not expected

Figure V-2: Comparison of 1AKE crystallographic B-factors (●) and simulation RMSF values (○) for all Cα atoms. Residues involved in crystal contacts (+ signs) are indicated at the bottom of the graph. Secondary structure elements are shown as published (7). AMP-bd and Lid domains are indicated by two-headed arrows.

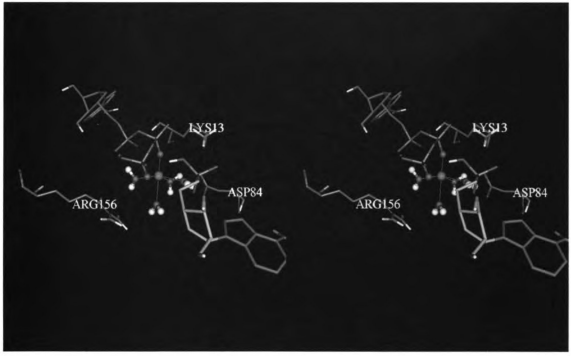
Simulation RMSF (Å)
Crystallographic RMSF (Å)
AMP-bd
Lid

from their average crystallographic RMSF values. Residues 74–79 (the loop between helix α_4 and strand β_3) show the opposite trend: Although they have the highest crystallographic RMSF values in 1AKE (1.72 ± 0.17 Å for 74–79 C_α's against an average B factor of 1.001 ± 0.19 for all C_α's), these residues show simulation RMSF values (2.16 ± 0.49 Å) only slightly above the average for the protein (1.93 ± 0.94 Å). The low crystallographic RMSF values associated with residues 42–48 in 1AKE might be due to the fact that residues 44, 45, 47, and 48 are involved in crystal contacts (7). In the simulation, the absence of crystal contacts results in a larger conformational freedom for this region of the protein. In this respect, it is interesting to note that residues 46–48 adopt different secondary structures in 1AKE complexes I and II (7). Müller et al. (10) later noticed that residues 44–48 are part of helix α_3 in the apo-enzyme (PDB code: 4AKE). These crystallographic observations suggest that this region is highly susceptible to conformational changes, and they agree with the large RMSF values observed for residues 41–48 in our simulation. Because residues 74–79 have the highest crystallographic RMSF values in 1AKE, and because these residues have lower crystallographic RMSF values in the apo-enzyme, it was suggested that this loop is part of an energy counterweight in ECAK (10). The low RMSF values observed for residues 74–79 in our simulation do not support this energy counterweight hypothesis. Another possibility is that the high crystallographic RMSF values of residues 74 79 are an artifact of crystallization. In 1AKE, this loop is involved in crystal contacts: Glu75 forms a H-bond and salt bridge with Asp76 and Arg78 of another molecule, respectively (7). These

interactions might force this loop into a conformation that is not 100% occupied or that would not be favored in solution. Some of these differences may be attributable not only to the different conditions involved in crystalline versus liquid states but, of course, also to the 1AKE* model, where $AP_5A$is replaced by the ATP-AMP-Mg^{2+}-4(H_2O) complex. The highly charged ATP and AMP phosphate chains and Mg^{2+}can exert significant forces that extend far into the protein.Thus, differences can arise from this replacement.The integrity of the ATP-AMP-Mg^{2+}4(H_2O) complex as the simulation proceeds is well maintained. as displayed in Figure V-3. The Mg^{2+} keeps its ligation to the ATP γ- and β-PO_4 oxygens, and the 4 waters are coordinated to the Mg^{2+}, forming an octahedral coordination sphere. The distance between the ATP γ-PO_4 and AMP α-PO_4 is reasonable for a catalytically competent configuration, and residues thought to be key for catalysis are positioned close to the reactant complex, as we detail below.

The stereochemical quality of 5 snapshots of 1AKE* taken between 900 and 1500 ps was checked using the software PROCHECK v 3.5.4. The results show that 74–79% of the residues are in the most favored regions of the Ramachandran plot, and 17–23% are in the additional allowed regions (around 97% total). These numbers are comparable to values obtained from solution structures of proteins from NMR. These observations suggest that the protein does not go through disallowed dihedral conformations during the simulation.

Figure V-3: The ATP-AMP-Mg^{2+}-4(H_20) complex and several of the (charged) residues that are within 7Å radius of Mg^{2+} from a snapshot around 2.5 ns. The key catalytic residues LYS13, ARG156, and ASP84, are displayed. Residues SER30, GLY14, GLY12, and GLY10 are also within this radius but are not displayed for clarity.

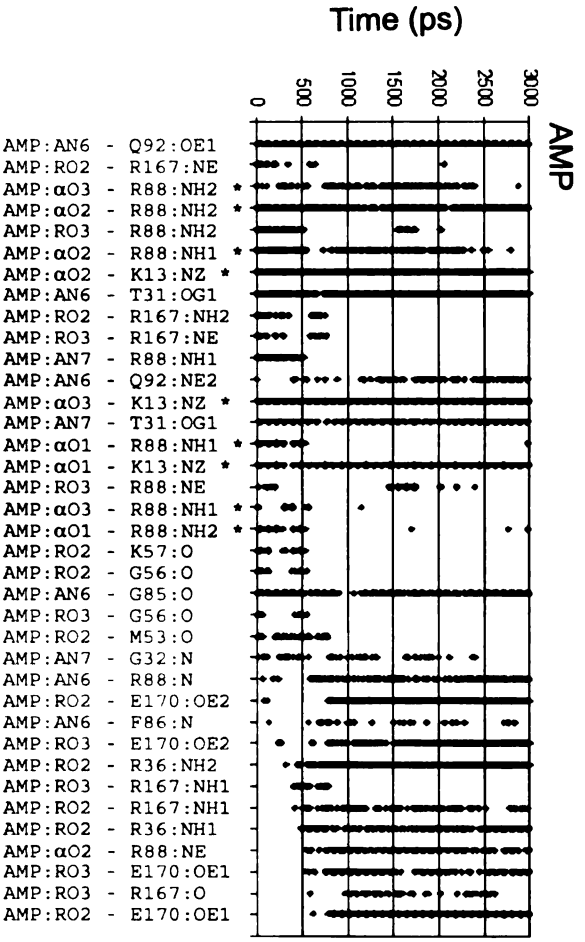

LYS13
ASP84
ARG156
LYS13
ASP84
ARG156

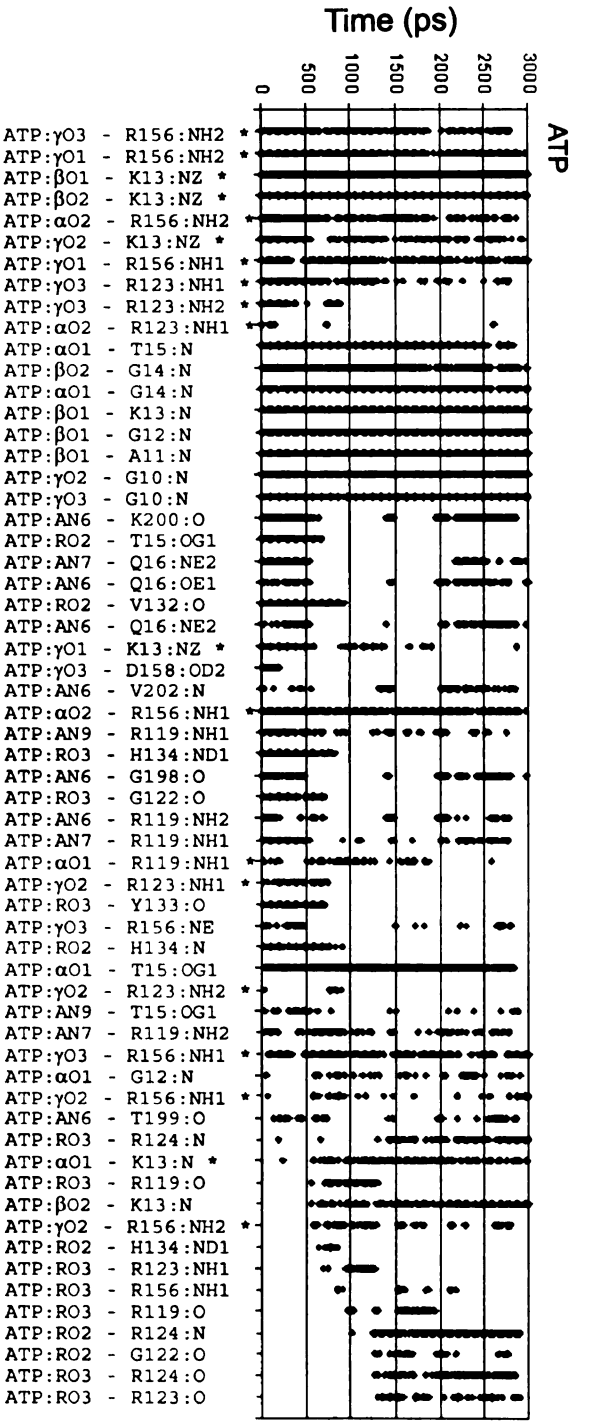
5.4.2 Substrate specificity

5.4.2.1 AMP binding site

AKs, in general, are known to have higher specificity for their nucleoside monophosphate (NMP) substrate than for their nucleoside triphosphate (NTP) substrate.40–43 Extensive mutagenesis studies and structural data have shown that AKs control AMP specificity through several interactions with highly conserved residues (24, 44, 45). The H-bonding pattern between AMP and ECAK was analyzed over the complete 3 ns simulation (Figure V-4). Figure V-5 shows all the interactions of AMP and ATP with ECAK residues that are present during the 3 ns simulation. In the 1AKE crystal structure of the ECAK · AP_5A complex (7), AMP's adenine interacts with ECAK through 5 strong H-bonds (2 H-bonds to backbone atoms and 3 H-bonds to side-chains) and several other less polar interactions. A strictly conserved glutamine residue, Gln92, plays a critical role both in catalysis and AMP specificity (43, 44, 46) through H-bonds with AMP's adenine N6 the orientation of Gln92's carboxamide is fixed by additional H-bonds to a backbone amide (from Arg88) and a conserved carboxylate (Asp61). In the simulation, plots (not shown) of the H-bonds between Asp61, Arg88, and Gln92 versus time show that Gln92's side-chain is indeed fixed by very stable H-bonds. Specifically, Gln92:NE2 H-bonds with Asp61 side-chain oxygens OD1 and OD2, and Gln92:OE1 H-bonds with Arg 88:N. Two other H-bonds between Thr31:OG1, Gly85:O, and the adenine N6 that are stable over the trajectory also help in discriminating against guanosine monophosphate (GMP), which lacks the NH2 group. Additionally, Thr31:OG1 forms a stable H-bond with AMP's adenine N7. Other H-bonds like adenine N6 to Phe86:N and adenine N7 to Gly32:N are

Figure V-4: Hydrogen bonds and salt bridges between the protein and ATP and between the protein and AMP during the course of the simulation. Donor and acceptor atoms within a maximum distance of 3.5 Å and a minimum angle of 120° were considered to form H-bonds. Charged donor and acceptor groups within a maximum distance of 4.5 Å were considered to form salt bridges (indicated by a *). H-bond and salt bridge analysis was performed every 0.8 ps (i.e., 3750 time points). Shown are only the H-bonds and salt bridges that are present at least 1.3% of the time (i.e., in at least 50 time points). In the designation of ATP and AMP, R denotes Ribose, A denotes Adenine and α, β and γ denote phosphates.


Time (ps)
0
500
1000
1500
2000
2500
3000
AMP
AMP:AN6 - Q92:OE1
AMP:RO2 - R167:NE
AMP:αO3 - R88:NH2 *
AMP:αO2 - R88:NH2 *
AMP:RO3 - R88:NH2
AMP:αO2 - R88:NH1 *
AMP:αO2 - K13:NZ *
AMP:AN6 - T31:OG1
AMP:RO2 - R167:NH2
AMP:RO3 - R167:NE
AMP:AN7 - R88:NH1
AMP:AN6 - Q92:NE2
AMP:αO3 - K13:NZ *
AMP:AN7 - T31:OG1
AMP:αO1 - R88:NH1 *
AMP:αO1 - K13:NZ *
AMP:RO3 - R88:NE
AMP:αO3 - R88:NH1 *
AMP:αO1 - R88:NH2 *
AMP:RO2 - K57:O
AMP:RO2 - G56:O
AMP:AN6 - G85:O
AMP:RO3 - G56:O
AMP:RO2 - M53:O
AMP:AN7 - G32:N
AMP:AN6 - R88:N
AMP:RO2 - E170:OE2
AMP:AN6 - F86:N
AMP:RO3 - E170:OE2
AMP:RO2 - R36:NH2
AMP:RO3 - R167:NH1
AMP:RO2 - R167:NH1
AMP:RO2 - R36:NH1
AMP:αO2 - R88:NE
AMP:RO3 - E170:OE1
AMP:RO3 - R167:O
AMP:RO2 - E170:OE1


Time (ps)
0
500
1000
1500
2000
2500
3000
ATP
ATP:γO3 - R156:NH2 *
ATP:γO1 - R156:NH2 *
ATP:βO1 - K13:NZ *
ATP:βO2 - K13:NZ *
ATP:αO2 - R156:NH2 *
ATP:γO2 - K13:NZ *
ATP:γO1 - R156:NH1 *
ATP:γO3 - R123:NH1 *
ATP:γO3 - R123:NH2 *
ATP:αO2 - R123:NH1 *
ATP:αO1 - T15:N
ATP:βO2 - G14:N
ATP:αO1 - G14:N
ATP:βO1 - K13:N
ATP:βO1 - G12:N
ATP:βO1 - A11:N
ATP:γO2 - G10:N
ATP:γO3 - G10:N
ATP:AN6 - K200:O
ATP:RO2 - T15:OG1
ATP:AN7 - Q16:NE2
ATP:AN6 - Q16:OE1
ATP:RO2 - V132:O
ATP:AN6 - Q16:NE2
ATP:γO1 - K13:NZ *
ATP:γO3 - D158:OD2
ATP:AN6 - V202:N
ATP:αO2 - R156:NH1 *
ATP:AN9 - R119:NH1
ATP:RO3 - H134:ND1
ATP:AN6 - G198:O
ATP:RO3 - G122:O
ATP:AN6 - R119:NH2
ATP:AN7 - R119:NH1
ATP:αO1 - R119:NH1 *
ATP:γO2 - R123:NH1 *
ATP:RO3 - Y133:O
ATP:γO3 - R156:NE
ATP:RO2 - H134:N
ATP:αO1 - T15:OG1
ATP:γO2 - R123:NH2 *
ATP:AN9 - T15:OG1
ATP:AN7 - R119:NH2
ATP:γO3 - R156:NH1 *
ATP:αO1 - G12:N
ATP:γO2 - R156:NH1 *
ATP:AN6 - T199:O
ATP:RO3 - R124:N
ATP:αO1 - K13:N *
ATP:RO3 - R119:O
ATP:βO2 - K13:N
ATP:γO2 - R156:NH2 *
ATP:RO2 - H134:ND1
ATP:RO3 - R123:NH1
ATP:RO3 - R156:NH1
ATP:RO3 - R119:O
ATP:RO2 - R124:N
ATP:RO2 - G122:O
ATP:RO3 - R124:O
ATP:RO3 - R123:O

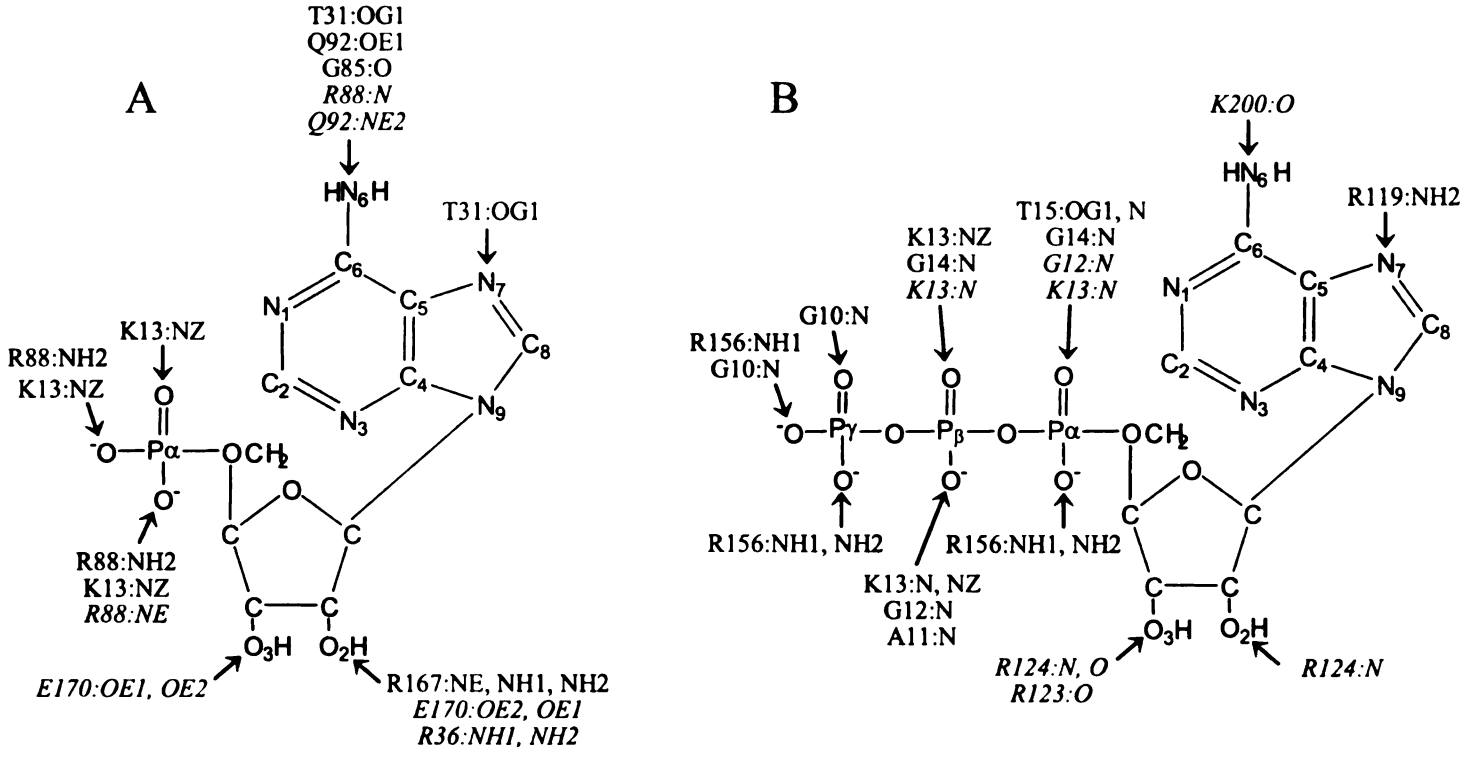


Figure V-5: Stable H-bond and salt bridge interactions between ECAK residues and AMP (A) and between ECAK residues and ATP (B) during the 3-ns simulation. Data were extracted from Figure 4. Regular characters: interactions present during the whole 3 ns; Italics: interactions appearing and remaining relatively stable after 500 ps or 1 ns. ATP and AMP are drawn with the -1 charge on each phosphate group split between 2 of the oxygens to indicate that the formal charge is shared.

on and off during the simulation (Figure V-4). Another residue, Arg88, forms transient H-bonds with the AMP adenine N6 and N7, as well as with AMP's α-PO_4, and with the AMP ribose O3 (Figure V-4). These bonds are possible because Arg88's side-chain is positioned parallel to the π rings of AMP's adenine. In 1AKE (9), Arg88 occupies 2 different conformations in the 2 molecules of the asymmetric unit. In the BSAK · AP_5A complex (11), Arg88 adopts yet another configuration, suggesting that specific interactions between Arg88 and AMP are not what make this residue essential in AMP binding. Arg88 is highly conserved in AKs, and its function in substrate binding and/or catalysis is the subject of ongoing discussions.

Only 1 good (2.68 Å) H-bond between AMP ribose and Lys57:O is seen in the 1AKE crystal structure. Interactions between Lys57:O and Gly56:O and the ribose are present only during the first 500 ps of the simulation (Figure V-4). However, a H-bond between the side-chain of Arg167 (involving NH1, NH2, or NE) and the ribose is present during the entire simulation (Figure V-4). In 1AKE, Arg167 is within H-bonding distance of the fourth (δ) phosphate in AP_5A. Examination of snapshots from the simulation show that the side-chain of Arg167 moves toward the ribose moiety early in the simulation. The difference in Arg167 orientation between the 1AKE crystal structure and our simulation may be due to the presence in 1AKE of the extra phosphate in AP_5A. In 1AKE, Glu170's side-chain points away from the ribose. Starting after 700 ps of our simulation, Glu170 (OE1 and OE2) forms a stable, double H-bond with AMP ribose O2 and O3 (Figure V-4). Whether this double H-bond is significant for AMP is open to question, because Glu170 is replaced by an alanine or a valine in the

B. stearothermophilus, *Thermotoga neapolitana*, yeast, and bovine heart mitochondrial AKs.

Several arginines (Arg36, Arg88, Arg119, Arg123, Arg131, and Arg167) are situated in the active site cleft surrounding AP_5A's phosphate groups in 1AKE, presumably to help stabilize the negatively charged phosphates. Some of these arginines have been shown to participate in phosphate binding and catalysis (7, 47–50). In 1AKE, AP_5A's ϵ-PO_4 (that corresponds to AMP's α-PO_4) is well fixed through interactions with R36:NH2, R88:NH2, and R156:NH2, the latter interaction being weaker than the first two (7). AP5A's ϵ-PO_4 sits between the head groups of Arg88 and Arg36. Chemical modification of ECAK with phenylglyoxal completely inactivates the enzyme (24). Reinstein et al (24) showed that this complete inactivation is obtained by modification of a single arginine: Arg88. Mutant R88G is 1000 times less active than the wild-type ECAK, the Michaelis constant (K_m) values for ATP and AMP increasing up to 5-fold and 85-fold, respectively. The simulation results show that R88:NH2 and R88:NE have very stable interactions with AMP's α-PO_4. These interactions are salt bridge–like considering the charges of the groups involved; hence, they are much stronger than H-bonds. In a previously published 300-ps simulation of the ECAK · AP_5A complex, the interactions between Arg88 and AP_5A's ϵ-PO_4, analyzed as H-bonds, were not preserved (29). In the 1AKE* simulation, H-bonds involving Arg36:NH1 and NH2, and Glu170:OE1 and OE2 appear after 500 ps. The Arg36 H-bonds are to the ribose, whereas in the 1AKE crystal structure, Arg36 H-bonds an ϵ-PO_4 oxygen. It seems that the removal of AP_5A's δ-PO_4 to create ATP and AMP (with the resulting loss of 2 negative charges) weakens some of the charged and polar interactions between phosphates and arginines and rearranges the

H-bond and salt bridge–like interactions. The presence of Mg^{2+} can also influence the distribution of charged interactions between the protein Arg residues and the phosphate groups. Permanent H-bonds are formed between AMP's α-PO_4 and the Lys13:NZ. Lys13, situated in the P loop of all AKs, has been described as the "invariant lysine" due to dramatic loss of activity by chemical modification. In summary, interactions of Thr31, Arg88, Gly85, and Gln92 with the adenine N6 are very persistent providing the possibility of discrimination. Arg88 and Lys13 are crucial for binding the α-PO_4 and positioning it suitably to receive a phosphate group from ATP.

5.4.2.2 ATP binding site

The ATP-binding site is formed by residues in the P-loop (residues 7–15), part of the LID domain, and residues in the C-terminal helix with its connecting loop region (residues 198–214). The ATP H-bonds found in the simulation are shown in Figure V-4, and the stable interactions are indicated in Figure V-5. Similar to other AKs, ECAK is less specific for ATP than for AMP (42), and most of the tight interactions between the protein and Mg-ATP involve the ATP phosphates. In the ECAK · Ap_5A complex, ATP's adenine moiety is only loosely bound to the protein through a single H-bond (7) between Lys200:O and ATP's adenine N6. This H bond leads to a preference for ATP over guanosine triphosphate (GTP). The simulation shows (Figure V-4) that this H-bond breaks and then reforms during the 3 ns, due to .uctuations in the position of ATP's adenine. The position of Lys200 itself does not deviate much from the crystal structure, and the H-bond between Lys200:O and adenine N6 is formed whenever the adenine moiety of ATP moves to the appropriate position.

In the crystal structure, the ATP adenine is parallel to the Arg119 guanidinium group, suggesting a cation–π interaction (7). The backbones of Pro201 and Val202 are in van der Waals contact distance of the adenine atoms. The Arg119–ATP interaction is not stable during the simulation, owing again to the large movements of ATP's adenine. N7 of the adenine ring forms an H-bond with Arg119:NH1 and NH2. Although this bond is relatively infrequent (occurring only about 8% of the time), it occurs on and off during the simulation (Figure V-4), concurrent with the movement of the adenine ring. A similar unstable H-bond is also seen between Val 202:O and ATP:AN6. Two new H-bonds between ATP:AN6 and either Tyr199:O or Gly198:O also form more or less in synchronization with the movement of the adenine ring. Taken together, none of the H-bonds are strong enough to hold ATP's adenine in place. Moreover, these interactions (with the exception of the H-bond involving Lys200) are independent of the base (it can be either adenine or guanine). While Lys200 is not conserved among the AKs, Arg119 is conserved in all known NMP kinases. These observations concur with experimental data confirming that ATP adenine is not tightly bound to the enzyme, and that it can be substituted with other NTPs such as GTP. In ECAK, the ATP adenine is quite solvent accessible, and so is Arg119; these observations probably lead to the extensive sampling of conformation space observed for ATP adenine in our MD simulation. They might also explain why the R128A mutation (equivalent to R119A in ECAK) only moderately perturbed the kinetic parameters of the chicken muscle AK1 enzyme (28).

Only one strong H-bond between the ATP ribose O3 and Tyr133:O exists in the crystal structures of ECAK · Ap_5A and ECAK · AMP · AMPPNP. It is thought that this H-bond plays a significant role in the closing of the LID over ATP, since it is one of the

few interactions between the LID and AP_5A (2). The H-bond between Tyr133 and ribose is present only during the first 750 ps of our simulation. Even during this time, the distance between Tyr133:O and ribose:O3 is close to 3.5 Å, making it a very weak interaction and hence unlikely to be a driving force in LID closure. A weak H-bond involving His134:O and ATP's ribose O2 (at 3.48 Å in 1AKE) is present until 1000 ps. Since His134 lies in the LID domain that shows considerable backbone flexibility during the simulation (Figure V-2), this H-bond is present only during the first nanosecond. An H-bond between ATP ribose O3 and Arg124:N appears after ~1290 ps. This H-bond is fairly stable, occurring 78% of the time. Taken together, the ATP ribose, like the adenine, is also not very tightly bound to the enzyme.

In contrast to ATP ribose and adenine, the ATP phosphates have extensive contacts with the protein. The triphosphate-binding site is a highly charged pocket formed by the LID and the core. It is lined with several arginines and the P loop (glycine-rich), and it is often described as a giant anion hole (51). Phosphate binding, orientation, and transfer are mediated by Arg36, Arg88, Arg123, Arg156, Arg167, and Lys13, all interacting with the P-loop backbone (9). As in the case of AMP, interactions between the phosphate oxygens and the positively charged side-chains of arginines and lysines are salt bridge–like in character and hence much stronger than H-bonds.

In 1AKE, α-PO_4 shows 3 good H-bonds. It is connected in the P loop to Thr15:N (at 2.85 Å), Thr15:OG1 (at 2.71 Å), and to Gly12:N (at 2.98 Å). The α-PO_4:O2 is also within H-bonding distance from Arg123:NH2 (at 2.91 Å). Thr23 (Thr15 in ECAK) in chicken muscle AK1 displays strong nuclear Overhauser effects (NOEs) with the

adenosine moiety of Mg-ATP (52), although it is within H-bonding distance from the ATP α-PO_4 in the AK1 crystal structure. Phosphate stereochemistry experiments in AK1 also suggest that Thr23 directly interacts with α-PO_4 (53). This result agrees with the 1AKE structure, in which Thr15:OG1 forms 2 stable H-bonds with α-PO_4. Based on these observations, it was suggested that in AK1, Thr23 plays a role in catalysis through direct interaction with ATP (53). Although Thr23 is conserved in all known AKs, the mild perturbations of mutant T23A's kinetic parameters (k_{cat} and K_m) are insufficient to confirm a functional role for Thr23 (53). In our simulation, Thr15:OG1 and Thr15:N form permanent H-bonds with α-PO_4 (Figures. V-4 and V-5) in accord with its suggested catalytic role. Thr15 also forms transitory H-bonds with ATP adenine N6, N7, and N9, in keeping with its proximity to ATP adenine, as seen in the solution structure of AK1 and in the crystal structure of the yeast AK · Mg-Ap_5A complex (13). These H-bonds are weaker, lasting for at most 500 ps.

The H-bond between Gly12:N and ATP's α-PO_4 seen in the 1AKE structure is present only rarely during the simulation. The on–off nature of this H-bond in the simulation originates from fluctuations in the position of the phosphate relative to Gly12. Instead, we observe a stable H-bond between ATP α-PO_4 and Gly14:N throughout the simulation. The salt bridge–like interaction with Arg123:NH1 is present only for 100 ps, with the side-chain of Arg123 moving away from ATP early in the simulation. The ATP phosphate chain adopts a slightly different position in the equilibrated 1AKE* compared to that in 1AKE. α-PO_4 is farther away from Arg123 in 1AKE* than in 1AKE, leading to a weaker salt bridge–like interaction. The side-chain of Arg123 is thus able to fluctuate

and move away from α-PO_4 during the simulation. This side-chain, however, moves within 4.5 Å of ATP γ-PO_4, thus creating another potential salt bridge interaction. Note that in the 1AKE crystal structure, Arg123:NH1 is 3.94 Å from a γ-PO_4 oxygen providing a bifurcated situation that, when the extra AP_5A PO_4 is removed, most likely leads to the observed repositioning of Arg123 in the simulation. In our simulation, the charge compensation for α-PO_4 is provided by its salt bridge interaction with Arg156:NH1, which is present about 50% of the time (V- 4), rather than its salt bridge with Arg123:NH1. Another H-bond not found in 1AKE forms between ATP α-PO_4 and Lys13:N after 600 ps, and it is present during most of the remaining 2.4 ns. Arg119:NH1 is within 4.5 Å of α-PO_4 until about 1750 ps. Even during this time, the interaction is weak and intermittent.

In 1AKE, β-PO_4 is very well connected to the P loop through as many as 4 strong H-bonds with Ala11:N, Gly12:N, Lys13:N, and Gly14:N, and through a salt bridge with Lys13:NZ. Almost all the H-bonds involving ATP β-PO_4 have a strong presence throughout the simulation (Figure V-4). The invariant P-loop lysine, Lys13, forms a permanent H-bond with Ala8:O during the simulation (not shown), thus acting as a bridge between Ala8 and ATP phosphate chain. The stability of this H-bond underlines the importance of the bridge in maintaining the integrity of the P-loop, since Ala8 and Lys13 are at opposite ends of the P-loop. Another H-bond between Lys13:NZ and Gly7:O is also present throughout the simulation, confirming the structural importance of Lys13. At no point is there an H-bond between the side-chain of Arg123 and β-PO_4 as

seen in 1AKE. With a donor -acceptor distance of 3.3 Å in 1AKE, the H-bond between Arg123:NH2 and β-PO_4 is weak even in the crystal structure.

The to-be-transferred γ-PO_4 is loosely bound, with fewer interactions with the active site residues than α-PO_4 and β-PO_4 in 1AKE, only forming salt bridge-like interactions with Lys13:NZ and Arg123:NH2, and weakly interacting with Arg156:NH2. In the BSAK · Mg-Ap_5A complex, γ-PO_4 is within H-bonding distance of Lys13, and within 4.5 Å of the side-chains of Arg123 (NH1 and NH2) and Arg156. These H-bonds have donor–acceptor distances of 3.3 Å or more, indicating weaker interactions with the protein as compared to ATP's α-PO_4 and β-PO_4. In our simulation, the salt-bridges between γ-PO_4 and Arg156:NH1 and NH2 are present as in the crystal structure. The interactions with Lys13 and Arg123 are not stable in the simulation. They are replaced by other interactions arising from the presence of Mg^{2+}, and from the enhanced conformational range available to γ-PO_4, which is no longer covalently attached to AP_5A's δ-PO_4. γ-PO_4 is now free to rotate and, with its negative charges, it strongly interacts with Mg^{2+}. An additional H-bond not seen in the crystal structures is seen in our simulation between Gly10:N and γ-PO_4. This H-bond is stable throughout the simulation and it is the only link between γ-PO_4 and the P loop.

In summary, we see about the same interactions between the ATP phosphate chain and the P loop (main-chain and side-chains) as in the crystal structure, but we see more and stronger interactions with Arg156 than with Arg123, for both β-PO_4 and γ-PO_4. Without exception, all permanent H-bonds between ATP and the protein in 1AKE* in the

simulation involve the 3 ATP phosphate groups (Figures V-4 and V-5), the ATP adenine and ribose moieties being only loosely bound to the protein.

5.4.2.3 Role and conformation of active site residues around Mg^{2+}

Kinases require a divalent cation, usually Mg^{2+}, for catalysis. The ATP β- and γ-phosphates are coordinated to Mg^{2+}, and the true substrate for AKs is Mg^{2+}-ATP. Mg^{2+} serves to neutralize the highly negative charge on the ATP triphosphate, thereby reducing electrostatic repulsion of the transferred phosphate and increasing the efficiency of nucleophilic attack. In AKs, conserved serine, threonine, or aspartate residues participate in binding the Mg^{2+} cation, coordinating it either directly in the first coordination sphere or indirectly through water molecules.

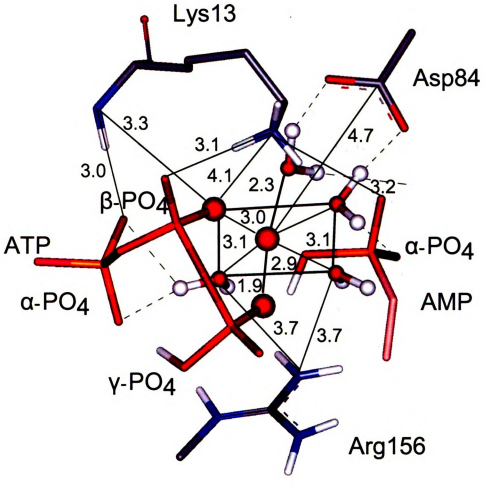
In porcine muscle AK · CoATP, AK · CoGTP, and AK · CoGDP, ^{31}P NMR relaxation rates indicate that Co^{2+} is directly coordinated to only β-PO_4 (GDP) or β-PO_4 and γ-PO_4 [ATP and GTP] (54). In addition, α-PO_4 is too far from Co^{2+} to be in the first coordination sphere, but also too close to be in the second coordination sphere (not enough room for a water molecule). These results suggest that ATP-Mg^{2+} binds AK as a β,γ bidentate complex rather than an α,β,γ tridentate complex. In the structure of the BSAK · Mg^{2+}-AP_5A complex, Mg^{2+} is coordinated to the phosphates corresponding to ATP's β- and γ-phosphates, and to 4 water molecules numbered 300–303 (11). Mg^{2+}-AP_5A is anchored to BSAK through several H-bond networks mediated by the 4 Mg^{2+}-coordinating water molecules. Waters 300 and 303 help orient the α-PO_4 of AMP. Water 300 is also positioned to interact with Arg36 and Asp33, thus linking Mg^{2+}-AP_5A to the protein. Water 302 is within H-bond distance of ATP's 3 phosphate groups,

reinforcing the linkage between the phosphate chain and Mg^{2+}. Water 301 H-bonds to Asp84:OD2 and Gly14:N, and it is the only water that does not interact with the phosphates. It is evident that this H-bond network is important for orienting the donor and acceptor phosphates and aiding the ternary complex to proceed to the transition state. Based on their 300 ps simulation results of ECAK · AP_5A, Kern et al. predicted that 3 aspartate residues, Asp84, Asp33, and Asp110, are possible candidates for binding the Mg^{2+} complex (29). Ser30, Thr31, and Thr89 were also suitably placed and were available to interact with Mg^{2+}. Orientation of the side-chains of these residues and the position of the phosphate chain left just enough room for the Mg^{2+} ion.

In the simulation, Mg^{2+} remains coordinated to 4 waters (218–221) as dictated by the force field (see the Methods section). The water–Mg^{2+} distances are short in the crystal structure, indicating that these waters are ligated to the ion. Then, one can choose to modify the force field by inserting specific "covalent" bonds, or by using noncovalent but relatively strong interactions (55), as we have done. Even though no such interactions were introduced between Mg^{2+} and ATP β- and γ-phosphates, we find that throughout the simulation Mg^{2+} is also coordinated to ATP's β-PO_4 through its O2 oxygen, and to ATP's γ-PO_4 through its O1 oxygen. As shown in Figure V-6, the coordination is close to octahedral. Furthermore, the additional flexibility introduced by having ATP/AMP versus AP_5A in the structure, as in the 1ZIO crystal structure, still preserves the octahedral geometry.

A second coordination sphere for Mg^{2+} includes Asp84: OD1 and OD2 that have waters 219 and 221 (equivalent to waters 301 and 303 in BSAK) screening this direct salt

Figure V-6: Mg^{2+} coordination sphere. The four Mg^{2+}-coordinated waters, the β-PO_4:O2, and the γ-PO_4:O1 form an essentially octahedral first coordination sphere. The multi-functional role of Lys13 in anchoring ATP and AMP is evident. Asp84 is in the second coordination sphere and is H-bonded to 2 of the Mg^{2+}-coordinated waters. Arg156 is almost in hydrogen bonding distance of the other two Mg^{2+}-coordinated waters and may be positioned to participate in catalysis. See text for further discussion. Distances are in angstroms.


Lys13
Asp84
3.3
3.1
4.7
3.0
4.1
2.3
3.2
β-PO4
3.0
ATP
3.1
3.1
α-PO4
2.9
1.9
α-PO4
AMP
3.7
3.7
γ-PO4
Arg156

bridge–like interaction. The H-bonds between water molecules 219 and 221 and the side-chain of Asp84 are very stable over the 3 ns trajectory (not shown), indicating the importance of Asp84 in binding the Mg^{2+} water complex. In AK1, mutation of the corresponding Asp93 to alanine decreased the k_{cat} 650-fold (56). The D93A mutant showed no structural perturbation as evidenced by NMR analysis, indicating a local effect. Moreover, the AK1 D93A mutant had a markedly lower affinity for Mg^{2+} (56). Similar effects were also observed for the ECAK D84A mutant (26). Considering the fact that Mg^{2+} plays a dual structural and chemical role in catalysis, these results suggest that Asp84 is important in the structural role of Mg^{2+}. Specifically, in the AK1 D93A mutant, Mg^{2+} may be unable to orient the phosphate chain for the transfer (56).

Throughout the simulation, Asp84's main- and side-chains form a stable network of H-bonds. Besides the above-mentioned H-bonds with waters 219 and 221, Asp84 forms stable H-bonds with Ser30 and Thr31. The H-bond between Asp84 and Thr31 backbone atoms is particularly significant, since Thr31 is also stably linked to AMP's adenine. Thus, these interactions anchor AMP to the Mg^{2+} complex involving the ATP phosphate chain. Asp84 interacts with Ser30 and Ile29 through strong H-bonds and also forms a persistent salt bridge with Lys13:NZ in the P-loop. Additionally, Ser30 is by itself coordinated to the Mg^{2+} ion through water molecule 219.

As in BSAK, water molecule 220 (water 302 in 1ZIO), is within H-bonding distance of ATP α-, β-, and γ-phosphates, strengthening the Mg^{2+}–phosphate complex and properly orienting the ATP phosphate chain. Interactions not seen in BSAK but transiently present in the 1AKE* simulation include H-bonds between Arg156 and waters 220 and 218. Arg156 has been implicated in the stabilization of the transition state.

Another rather weak H-bond is also seen between Gly14:N and water molecule 219. A corresponding H-bond is present in 1ZIO involving Gly14 and water molecule 301.

Lysine, with its backbone amide, long and flexible side-chain, and charged head group, is well-suited for a multifunctional role. These features are apparent in Figure V-6, where Lys13 stabilizes the active site by forming H-bonds and salt bridge–like interactions with both ATP and AMP using its flexibility and multifunctionality. The other interactions of Lys13 are discussed in previous sections.

5.4.3 Catalytic Mechanism

The 1AKE* simulation can be used to suggest whether the mechanism of phosphoryl transfer is associative or dissociative, at least from the perspective of geometrical requirements. The consensus is that in enzymes in general (31–34, 57), and in AKs in particular (25, 43), an associative mechanism is operative. The geometric criterion for an associative mechanism is that the transition state has sufficiently short entering and leaving group distances to support reasonably large fractional bond numbers, *n*. For example, if an S_N2 mechanism were operative ($n = ½$) the P–O distance would be 1.91 Å, while if a fully associative mechanism were operative ($n = 1$), the P–O distance would be 1.73 Å, the P–O single-bond distance.

In Figure V-7, the ATP γ–P to AMP α–P distance is plotted during the 3 ns simulation time. For most of the time, this distance ranges between 4.5 Å and 5.0 Å, about 1 Å greater than the 3.8 Å van der Waals P–P contact distance. In the absence of the Mg^{2+} cation and Lys13, with its ability to span ATP and AMP, it would be difficult to support the close approach of these (negatively charged) phosphate groups. The P–P

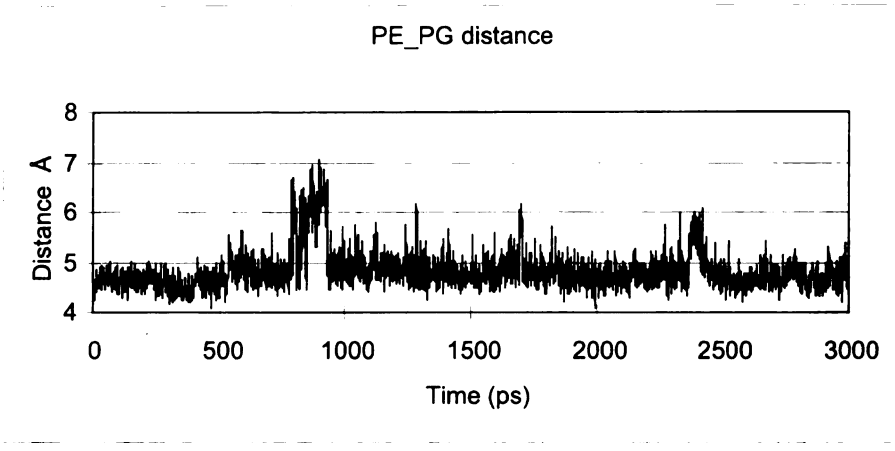


Figure V-7: Distance between ATP γ–P and AMP α–P atoms during the 3-ns simulation

distance suggests that an associative mechanism could be operative in AKE, as illustrated in Figure V-8. The atoms ATP β–P, Lys13:NZ and AMP α–P form an essentially equilateral triangle with apex Lys13:NZ and equal legs ATP β-P–Lys13:NZ and Lys13:NZ–AMP α-P. The atoms Lys13:NZ, ATP β–P, ATP γ–P, and AMP α–P form a plane, where the ATP β-P–AMP α-P distance is approximately twice that of the ATP β P–ATP γ-P distance. Thus, there would be room to rotate the ATP γ-PO_4 into a line formed by ATP β–P, ATP γ–P, and AMP α–P and produce a trigonal bipyrimidal phosphorane transition state, as suggested by Reinstein et al (25). In this configuration, Lys13:NZ would be positioned above the 3 apical oxygens, those in the plane perpendicular to the ATP β–P, ATP γ–P, and AMP α–P line. The Mg^{2+} ($4H_2O$) complex is below the above-mentioned plane, with Mg^{2+} equidistant from the oxygens (shown as spheres) associated with ATP β–P and ATP γ–P. From these geometrical considerations, the role played by Mg^{2+} ($4H_2O$) and Lys13 in maintaining the proximity and orientation of the ATP and AMP phosphates appropriate for phosphoryl transfer is clarified.

5.5 Conclusions

MD simulation of the ECAK · Mg–ATP · AMP ternary reactant complex that we constructed shows general agreement with conclusions drawn from the analysis of the crystallographic and other studies of ECAK complexed with AP_5A. The strong repulsive interaction between the terminal, negative phosphate groups that results from the split of AP_5A into ATP and AMP is counteracted by the presence of Mg^{2+} and, presumably,

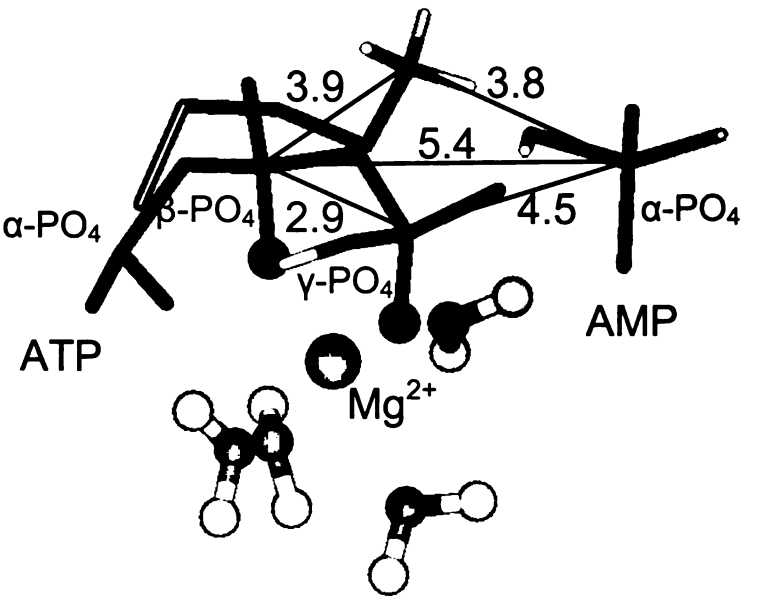


Figure V-8: Goemetry of Mg2+ association with its four coordinating waters, ATP and AMP phosphates, and Lys13. Distances are in Å. The displayed distances allow room for the ATP γ-PO4 to rotate to form an ATP β-P, ATP γ-P and AMP α-P line and produce a trigonal bipyrimidal phosphorane transition state.

some of the positively charged surrounding residues. The AMP adenine N6 strongly interacts with the protein through H-bonds, which forms the basis for NMP specificity. In contrast, and also in agreement with conclusions drawn from the crystallographic data (6), the ATP adenine and ribose moieties are only loosely bound to the protein, while the phosphates interact strongly with nearby residues. Octahedral coordination of the Mg^{2+} by 4 waters and by oxygens of ATP β- and γ-phosphates is maintained throughout the simulation. The conserved Lys13 in the P loop bridges the ATP and AMP phosphates, a phenomenon that relies on the lysine side-chain's unique properties of flexibility and H-bond–salt bridge capability. In addition to Lys13, the Mg^{2+}, its coordination waters, and some surrounding charged residues maintain the AMP α-phosphate and ATP β- and γ phosphates in a configuration suggesting that phosphoryl transfer occurs by an associative mechanism in AK. The simulation results could provide a starting point for a combined quantum mechanical–molecular mechanical simulation of the reaction mechanism for phosphoryl transfer.

5.6 References

1. **Walsh C.** 1979. Enzymatic Reaction Mechanisms. San Francisco: W. H. Freeman and company 978 p.

2. **Schulz GE, Muller CW, Diederichs K.** 1990. Induced-fit movements in adenylate kinases. J Mol Biol **213**:627–630.

3. **Schulz GE.** 1992. Induced-fit movements in adenylate kinases. Faraday Discuss **93**:85–93.

4. **Vieille C, Krishnamurthy H, Hyun HH, Savchenko A, Yan H, Zeikus JG.** 2003. *Thermotoga neapolitana* adenylate kinase is highly active at 30°C. Biochem J 372:577–585.

5. **Dreusicke D, Karplus PA, Schulz GE.** 1988. Refined structure of porcine cytosolic adenylate kinase at 2.1 Å resolution. J Mol Biol **199**:359–371.

6. **Müller CW, Schulz GE.** 1988. Structure of the complex of adenylate kinase from *Escherichia coli* with the inhibitor P1,P5-di(adenosine-5'-)pentaphosphate. J Mol Biol **202**:909–912.

7. **Müller CW, Schulz GE.** 1992. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. J Mol Biol **224**:159–177.

8. **Müller CW, Schulz GE.** 1993. Crystal structures of two mutants of adenylate kinase from *Escherichia coli* that modify the Gly-loop. Proteins **15**:42–49.

9. **Berry MB, Meador B, Bilderback T, Liang P, Glaser M, Phillips GN, Jr.** 1994. The closed conformation of a highly flexible protein: the structure of *E. coli* adenylate kinase with bound AMP and AMPPNP. Proteins **19**:183–198.

10. **Müller CW, Schlauderer GJ, Reinstein J, Schulz GE.** 1996. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure **4**:147–156.

11. **Berry MB, Phillips GNJ.** 1998. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap_5A, Mg^{2+} Ap_5A, and Mn^{2+} Ap_5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg^{2+} and Mn^{2+}. Proteins: Struct Function, Genet **32**:276–288.

12. **Abele U, Schulz GE.** 1995. High-resolution structures of adenylate kinase from yeast ligated with inhibitor Ap5A, showing the pathway of phosphoryl transfer. Protein Sci **4**:1262–1271.

13. **Egner U, Tomasselli AG, Schulz GE.** 1987. Structure of the complex of yeast adenylate kinase with the inhibitor P1,P5-di(adenosine-5'-)pentaphosphate at 2.6 Å resolution. J Mol Biol **195**:649–658.

14. **Diederichs K, Schulz GE.** 1990. Three-dimensional structure of the complex between the mitochondrial matrix adenylate kinase and its substrate AMP. Biochemistry **29**:8138–8144.

15. **Schlauderer GJ, Proba K, Schulz GE.** 1996. Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. J Mol Biol **256**:223–227.

16. **Schlauderer GJ, Schulz GE.** 1996. The structure of bovine mitochondrial adenylate kinase: Comparison with isoenzymes in other compartments. Protein Sci **5**:434–441.

17. **Sheng XR, Li X, Pan XM.** 1999. An iso-random Bi Bi mechanism for adenylate kinase. J Biol Chem **274**:22238–22242.

18. **Sinev MA, Sineva EV, Ittah V, Haas E.** 1996 Domain closure in adenylate kinase. Biochemistry.**35**:6425–6437.

19. **Lin Y, Nageswara Rao BD.** 2000. Structural characterization of adenine nucleotides bound to *Escherichia coli* adenylate kinase. 1. Adenosine conformations by proton two-dimensional transferred nuclear Overhauser effect spectroscopy. Biochemistry **39**:3636–3646.

20. **Lin Y, Nageswara Rao BD.** 2000. Structural characterization of adenine nucleotides bound to *Escherichia coli* adenylate kinase. 2. ^{31}P and ^{13}C relaxation measurements in the presence of cobalt(II) and manganese(II). Biochemistry **39**:3647–3655.

21. **Shapiro YE, Sinev MA, Sineva EV, Tugarinov V, Meirovitch E.** 2000. Backbone dynamics of *Escherichia coli* adenylate kinase at the extreme stages of the catalytic cycle studied by (15)N NMR relaxation. Biochemistry **39**:6634–6644.

22. **Shapiro YE, Kahana E, Tugarinov V, Liang ZC, Freed JH, Meirovitch E.** 2002. Domain flexibility in ligand-free and inhibitor-bound *Escherichia coli* adenylate kinase based on a mode-coupling analysis of N-15 spin relaxation. Biochemistry **41**:6271–6281.

23. **Reinstein J, Brune M, Wittinghofer A.** 1988. Mutations in the nucleotide binding loop of adenylate kinase of *Escherichia coli*. Biochemistry **27**:4712–4720.

24. **Reinstein J, Gilles AM, Rose T, Wittinghofer A, Saint Girons I, Bârzu O, Surewicz WK, Mantsch HH.** 1989. Structural and catalytic role of arginine 88 in

Escherichia coli adenylate kinase as evidenced by chemical modification and site-directed mutagenesis. J Biol Chem **264**:8107–8112.

25. **Reinstein J, Schlichting I, Wittinghofer A.** 1990. Structurally and catalytically important residues in the phosphate binding loop of adenylate kinase of *Escherichia coli*. Biochemistry **29**:7451–7459.

26. **Rose T, Glaser P, Surewicz WK, Mantsch HH, Reinstein J, Le Blay K, Gilles AM, Barzu O.** 1991. Structural and functional consequences of amino acid substitutions in the second conserved loop of *Escherichia coli* adenylate kinase. J Biol Chem **266**:23654–23659.

27. **Rose T, Brune M, Wittinghofer A, Le Blay K, Surewicz WK, Mantsch HH, Barzu O, Gilles AM.** 1991. Structural and catalytic properties of a deletion derivative (D133-157) of *Escherichia coli* adenylate kinase. J Biol Chem **266**:10781–10786.

28. **Tsai MD, Yan HG.** 1991. Mechanism of adenylate kinase: site-directed mutagenesis versus X-ray and NMR. Biochemistry **30**:6806–6818.

29. **Kern P, Brunne RM, Folkers G.** 1994. Nucleotide-binding properties of adenylate kinase from *Escherichia coli*: a molecular dynamics study in aqueous and vacuum environments. J Comput Aided Mol Des **8**:367–838.

30. **Elamrani S, Berry MB, Phillips GN, McCammon JA.** 1996. Study of global motions in proteins by weighted masses molecular dynamics: adenylate kinase as a test case. Proteins **25**:79–88.

31. **Knowles JR.** 1980. Enzyme-catalyzed phosphoryl transfer reactions. Ann Rev Biochem **49**:877–919.

32. **Hollfelder F, Herschlag D.** 1995. The nature of the transition-state for enzyme-catalyzed phosphoryl transfer - hydrolysis of O-aryl phosphorothioates by alkaline-phosphatase. Biochemistry **34**:12255–12264.

33. **Matte A, Tari LW, Delbaere LTJ.** 1998. How do kinases transfer phosphoryl groups? Structure **6**:413–419.

34. **Lahiri SD, Zhang GF, Dunaway-Mariano D, Allen KN.** 2003. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. Science **299**:2067–2071.

35. **van Gunsteren WF, Billeter SR, Eising AA, Hünnenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG.** 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Zürich: Vdf Hochschulverlag AG an der ETH Zürich.

36. **Tsodikov OV, Record MTJ, Sergeev YV.** 2002. A novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. J Comput Chem **23**: 600–609.

37. **Cukier RI, Seibold SA.** 2002. Molecular dynamics simulations of prostaglandin endoperoxide II synthase-1: Role of water and the mechanism of compound I formation from hydrogen peroxide". J Phys Chem B **106**:12031–12044.

38. **Hockney RW, Eastwood JW.** 1981. Computer Simulation Using Particles. New York: McGraw-Hill. 540 p.

39. **Allen MP, Tildesley DJ.** 1987. Computer Simulation of Liquids. Oxford: Clarendon Press. 385 p.

40. **Tomasselli AG, Noda LH.** 1980. Mitochondrial ATP:AMP phosphotransferase from beef heart: purification and properties. Eur J Biochem **103**:481–491.

41. **Ito Y, Tomasselli AG, Noda LH.** 1980. ATP:AMP phosphotransferase from baker's yeast. Purification and properties. Eur J Biochem **105**:85–92.

42. **Saint Girons I, Gilles AM, Margarita D, Michelson S, Monnot M, Fermandjian S, Danchin A, Barzu O.** 1987. Structural and catalytic characteristics of *Escherichia coli* adenylate kinase. J Biol Chem **262**:622–629.

43. **Yan H, Tsai MD.** 1999. Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. Adv Enzymol Relat Areas Mol Biol **73**:103-134.

44. **Beichner S, Byeon I-JL, Tsai M-D.** 1996. Structure-function relationship iof adenylate kinase: Glu-101 in AMP specificity. In: Kaumaya PTP, Hodges RS, editors. Peptides: Chemistry, Structure and Biology. Kingswinford, England: Mayflower Sicentific Ltd.; p 721–723.

45. **Okajima T, Tanizawa K, Yoneya T, Fukui T.** 1991. Role of leucine 66 in the asymmetric recognition of substrates in chicken muscle adenylate kinase. J Biol Chem **266**:11442–11447.

46. **Okajima T, Tanizawa K, Fukui T.** 1993. Site-directed random mutagenesis of AMP-binding residues in adenylate kinase. J Biochem **114**:627–633.

47. **Kim HJ, Nishikawa S, Tokutomi Y, Takenaka H, Hamada M, Kuby SA, Uesugi S.** 1990. In vitro mutagenesis studies at the arginine residues of adenylate kinase. A revised binding site for AMP in the X-ray-deduced model. Biochemistry **29**:1107–1111.

48. **Yan HG, Dahnke T, Zhou BB, Nakazawa A, Tsai MD.** 1990. Mechanism of adenylate kinase. Critical evaluation of the X-ray model and assignment of the AMP site. Biochemistry **29**:10956–10964.

49. **Yan HG, Shi ZT, Tsai MD.** 1990. Mechanism of adenylate kinase. Structural and functional demonstration of arginine-138 as a key catalytic residue that cannot be replaced by lysine. Biochemistry **29**:6385–6392.

50. **Dahnke T, Shi Z, Yan H, Jiang RT, Tsai MD.** 1992. Mechanism of adenylate kinase. Structural and functional roles of the conserved arginine-97 and arginine-132. Biochemistry **31**:6318–6328.

51. **Dreusicke D, Schulz GE.** 1986. The glycine-rich loop of adenylate kinase forms a giant anion hole. FEBS Lett **208**:301–304.

52. **Byeon IJ, Yan H, Edison AS, Mooberry ES, Abildgaard F, Markley JL, Tsai MD.** 1993. Mechanism of adenylate kinase. 1H, 13C, and 15N NMR assignments, secondary structures, and substrate binding sites. Biochemistry **32**:12508–12521.

53. **Shi Z, Byeon IJ, Jiang RT, Tsai MD.** 1993. Mechanism of adenylate kinase. What can be learned from a mutant enzyme with minor perturbation in kinetic parameters? Biochemistry **32**:6450–6458.

54. **Ray BD, Rosch P, Rao BD.** 1988. ^{31}P NMR studies of the structure of cation-nucleotide complexes bound to porcine muscle adenylate kinase. Biochemistry **27**:8669–8676.

55 **Banci L.** 2003. Molecular dynamics simulations of metalloproteins. Curr Opin Chem Biol **7**:143–149.

56. **Yan HG, Tsai MD.** 1991. Mechanism of adenylate kinase. Demonstration of a functional relationship between aspartate 93 and Mg^{2+} by site-directed mutagenesis and proton, phosphorus-31, and magnesium-25 NMR. Biochemistry **30**:5539–5546.

57. **Mildvan AS.** 1997. Mechanisms of signaling and related enzymes. Proteins-Structure Function and Genetics **29**:401-416.

Chapter 6

Conclusions and future directions

The popular hypothesis regarding the stability of hyperthermophilic enzymes at high temperatures postulates that these enzymes have enhanced conformational rigidity in their folded native state. The enhanced rigidity is responsible for the low catalytic activity of these enzymes at low temperatures. To date, no other physical explanation regarding the constancy in the maximal catalytic rates of hyperthermophilic and mesophilic enzyme pairs at their respective source organisms' optimal growth temperatures, has gained as much attention. This chapter summarizes how TNAK may not conform fully to this hypothesis, and how TNAK, and more enzymes like it should be studied to deepen our understanding of the link between flexibility, stability, and activity.

The hypothesis implies that, as the temperature increases, the thermal fluctuations in an enzyme increases, leading to increased catalytic activity up to the optimal rate that can be measured. Beyond the temperature for optimal activity, the increased fluctuations lead to instability in the enzyme structure, loss of activity, and global unfolding. This

simplistic view of conformational dynamics links local conformational fluctuations and global stability and thereby links thermodynamic stability with maximal catalytic activity. A practical outcome of this hypothesis would be that thermostability and enhanced catalytic activity, desirable properties in many enzymes with industrial applications, cannot be achieved. However, in contrast to this expectation, several enzymes have been evolved in the laboratory for enhanced thermostability while maintaining low temperature activity (2, 12), suggesting that the low activity of hyperthermophilic proteins at low temperatures has its basis in evolution rather than in any physico-chemical limitation. Moreover, TNAK which is as active as ECAK at 30°C is itself a good example of a naturally evolved enzyme that contradicts this expectation of a tradeoff between activity and stability. ECAK and TNAK therefore, make a unique and interesting enzyme pair to study the interplay between flexibility, activity, and stability.

On the basis of physical principles, linking structural rigidity to global stability is overly simplistic. Changes along the protein sequence can change flexibility and stability in different ways. As an instance, it has been shown that salt-bridges can be thermodynamically destabilizing at room temperature due to the desolvation penalty associated with their formation (3). Introducing a mutation in a protein interior to form a salt-bridge may result in increased rigidity but also cause thermodynamic instability. On the other hand, a cavity-inducing mutation in the interior of a protein is likely to increase flexibility and decrease stability. Flexibility depends on the energy landscape of proteins near their native states. Proteins are complex systems with different modes of fluctuations that span from ps to greater than ms timescales. These fluctuations contribute differently to activity and stability. For example, hinge-bending domain movements as seen in AKs

need not be correlated to the stability of the domains themselves. Hence, flexibility for activity is not necessarily coupled to flexibility for stability, unless active site residues are involved in both. The first step towards understanding the interplay between flexibility, stability, and activity is to recognize the range of motions, their amplitude and their spatial distribution in the proteins under investigation. In this study, TNAK was used as a model system to investigate dynamics at every residue, in the ps, ns, and μs timescales as well as longer timescales (in the order of minutes to days), to understand how dynamics on these timescale relate to activity and stability.

6.1 Dynamics in ECAK and TNAK on the ps-ns and μs timescales

^{15}N NMR relaxation was used to investigate ps to ns and μs timescale backbone dynamics of TNAK in its open and closed conformations. These results were compared to the results from a similar study on ECAK (10). The following observations were made from the comparison:

(i) The average S^2 values of the domains showed that no single domain was more rigid or flexible that the others in both ECAK and TNAK, in their free and complexed forms. In other words, overall, the core, the lid, and the AMPbd domains have similar ps-ns timescale flexibility in all four enzyme forms.

(ii) TNAK is more rigid than ECAK on the ps-ns timescale and the increase in rigidity is uniform across the entire protein.

(iii) Although having overall similar flexibility to the core and lid domains, the AMPbd domain is singled out as having many residues lower S^2 values, i.e.,

showing higher amplitude of ps-ns motions. This is true both for ECAK and TNAK. Many TNAK residues in this region had S^2 values comparable to the corresponding residues in ECAK indicating similar flexibility as ECAK.

(iv) ECAK becomes significantly more rigid upon inhibitor binding. Specifically, the average S^2 values of the inhibitor-bound forms of ECAK and TNAK become very similar.

(v) The core-lid hinge loops in ECAK*Ap5A and TNAK*Ap5A have low average S^2 values compared to the rest of the protein.

(vi) On the μs timescale also, TNAK is more rigid than ECAK. μs timescale motions are more pervasive in ECAK, present in the core, lid, and AMPbd domains, and the hinges. In TNAK however, the core β-strands are specifically excluded from having motions on this timescale. μs timescale motions are found only for residues belonging to the core helices and loops, the AMPbd and lid domains, and the hinges of these domains.

While TNAK's increased rigidity is attributable to increased enthalpic contributions from intramolecular interactions, what does the higher flexibility of AMPbd, lid, and hinge residues in the ps to ns timescale signify?

Analyses of order parameters in several enzymes have shown that ps-ns timescale motions are modulated in the free enzyme, in the presence of substrates and inhibitors and in different mutant forms. Some well studied examples include the HIV protease in complex with various inhibitors (9). Low order parameters were observed for residues in the loop preceding the flap that closes over the active site in an inhibitor-bound form of HIV-protease. The authors suggested that the ps-ns timescale motion in these residues

could be important for substrate-binding and product release (7) Similar studies of different complexes of *E. coli* dihydrofolate reductase (DHFR) representing different stages of the catalytic cycle provide insight into how ps-ns timescale motions are modulated at the different stages of the catalytic cycle and how these motions are related to catalysis (8, 9). Motions on the ps-ns timescale of key loops in DHFR complexes were implicated in dynamics related to substrate accessibility to the active site and in product release. These examples clearly show that ps-ns timescale motions play an important role in enzyme function, modulation, and substrate recognition. These observations therefore, raise the question: what is the timescale of kinetically relevant motions? While it is tempting to ascribe functional significance to only μs-ms timescale motions since this timescale often corresponds to the catalytic turn-over numbers, the timescale of kinetically relevant motions need not be restricted by the timescale of the measured catalytic rates. If a catalytic event is not strongly correlated to a functionally relevant motion, then this motion can occur on a much faster timescale. In this case, several such dynamic events may be required for a much slower catalytic transition to occur.

The facts that (i) several residues with low S^2 values are found mainly in the AMPbd residues of ECAK and TNAK (ii) that the average S^2 values of the core-lid hinge loops in ECAK*Ap5A and TNAK*Ap5A are lower than the rest of the protein and (iii) that residues in TNAK's core-lid hinges in both the free and inhibitor-bound forms show μs timescale motions, suggest that these motions are perhaps linked to catalysis related conformational transitions. Moreover, several TNAK residues in these regions have similar S^2 values as the corresponding ECAK residues. Taken together, the data suggests that flexibility required for activity may be similar in ECAK and TNAK. Specifically, the

lid-opening rate for product release in TNAK may be comparable to that of ECAK. Since lid-opening rate has been shown to be rate-limiting in ECAK catalysis (11), similar lid-opening rates in TNAK and ECAK may explain their similar activities at 30°C. This then would be in contrast to the result for another thermophilic AK, the *Aquifex aeolicus* AK (AAAK), whose lid-opening rate (k_{op}) was measured using ^{15}N NMR relaxation dispersion studies (11). It was found that the k_{op} of AAAK was slower than the k_{op} for ECAK at 20°C. The difference in the k_{op} for the two enzymes explained well the difference in their catalytic rates measured at 20°C.

In summary, the ^{15}N NMR relaxation study of TNAK has shown that TNAK is more rigid than ECAK, in accordance with the hypothesis. However, contrary to the expectation of the hypothesis, it maintains high activity at low temperatures by localizing flexibility to regions (such as hinges) that are important for inducing conformational changes. Since it is the hinges that are important for motions for activity, localizing flexibility to the hinges while keeping the lid and AMPbd domains rigid does not compromise TNAK's activity. Conversely, localizing flexibility in the hinges does not compromise the stability of the protein.

Future directions

The ^{15}N NMR relaxation method is only an indirect method for detecting μs timescale motions. Deviation of observed T_2 values from the values predicted from T_1 and NOE values is used as evidence for motion in the μs timescale and is described by the parameter R_{ex}. Since R_{ex} is directly proportional to the chemical shift difference of the exchanging nuclei in their different conformational states, absence of R_{ex} does not mean

there is no conformational exchange for a particular residue. If the chemical environment sensed by a nucleus in the different conformational states is similar, then no R_{ex} value for that residue will be observed within experimental error. Moreover, nothing can be said about the rate of conformational exchange from R_{ex} values. It was speculated from the results of the ^{15}N NMR relaxation analyses, that TNAK's lid-opening rate at 30°C may be comparable to that of ECAK, since several residues in the hinge between TNAK's lid and core domains showed R_{ex} values. To confirm this theory, accurate values of k_{op} and k_{cl} (lid-closing rate) can be obtained using ^{15}N NMR relaxation dispersion experiments.

6.2 H-D exchange rate analyses of ECAK and TNAK

H-D exchange is a powerful tool to examine protein dynamics on longer timescales (minutes-days) in the native state. To complement the ^{15}N NMR relaxation measurements of TNAK dynamics on the ps-μs timescale, H-D exchange experiments were used to investigate slower motional modes. H-D exchange rates of the open and closed forms of ECAK and TNAK were measured and compared. The purpose of the H-D experiments was to compare the distribution of structural rigidity in the two enzymes, in the two conformational states. Protection factors, P, were used as a measure of stability and residue-specific P values were calculated. The results show that:

(i) ECAK is not uniformly rigid. ECAK's core is much more rigid than its AMPbd and lid domains.

(ii) TNAK's core is only somewhat more rigid than its AMPbd and lid domains

(iii) The increase in rigidity in TNAK is significant only for the lid and AMPbd domains. TNAK's core is only marginally more rigid than ECAK's core.

(iv) Zn^{2+} likely plays an important role in stabilizing TNAK's lid.

(v) In both ECAK and TNAK, the lid is significantly more stabilized upon inhibitor binding than the other domains.

(vi) Although the AMPbd and lid domains of TNAK are more rigid than ECAK, there are several residues in these domains and their hinges that exchange within the dead time of the experiments indicating significant local fluctuations.

These results indicate that TNAK is not uniformly more rigid than ECAK at all timescales. While in the ps-ns and μs timescales, TNAK is uniformly more rigid than ECAK, in the slower timescales of H-D exchange (≥ s) only TNAK's lid and AMPbd are more rigid than those of ECAK. This result highlights the overly simplistic nature of the commonly accepted hypothesis. Proteins are characterized by motional modes on a wide range of timescales. Hence one cannot talk about flexibility in general, as the hypothesis does, but one must specify the timescale. Also, the hypothesis implies a uniform increase in rigidity across the structure of a hyperthermophilic protein. While this is true for the ps-ns timescale for TNAK, it is not true for timescales greater than seconds. Although TNAK's AMPbd and lid domains are more rigid than ECAK's, there are regions in these domains and their hinges that are characterized by local fluctuations. Flexibility for activity and rigidity for stability are partitioned in TNAK in a manner that makes it simultaneously highly active at room temperature and stable at high temperatures.

H-D exchange rates were used to calculate residue-specific ΔG_{HX} values for *Thermus thermophilus* and *E. coli ribonuclease* (TtRNase and EcRNase) (4). The results showed that the pattern of stability distribution for the two enzymes were the same with ΔG_{HX}

values of TtRNase residues being on average higher than those of EcRNase. Moreover, each region of the protein contributes the same relative stability to each protein. This is in contrast to TNAK, where the lid and AMPbd domains contribute proportionally more to its stability. Therefore the mechanism of stabilization is different for different proteins. Proteins have individual features of adaptation and hence it may not be possible to find a stabilization mechanism that is common to all hyperthermophilic proteins.

Future directions

While H-D exchange results reported in chapter 4 provide insight into stability distribution in ECAK and TNAK, it would be more useful to measure the energetics of stability in the two enzymes. To obtain protein stability parameters such as free energy, enthalpy, and entropy of unfolding, the dependence of H-D exchange rates on increasing denaturant concentration or increasing temperature has to be determined.

As summarized in Chapter 1, exchange in proteins takes place through three mechanisms. By the first mechanism, exchange occurs due to global unfolding. For residues that undergo exchange due to global unfolding, the $\Delta G_{HX} = \Delta G_{unf}$ at all denaturant concentrations, [D], and ΔG_{HX} varies linearly with [D]. The plot of ΔG_{HX} vs. [D] yields a slope called the m value. The m value is roughly proportional to the amount of surface area exposed in the opening reaction. In native state H-D exchange experiments, low [D] is used such that less than 1% of the protein molecules are unfolded. The presence of denaturant simply modulates the equilibrium population of folded and unfolded molecules. The stability of the protein (ΔG_{unf}) under native

conditions is determined by extrapolating the ΔG_{HX} value to zero concentration of the denaturant.

Other than the slow exchanging amides, some amides exchange due to partial unfolding mechanisms and their exchange rate is consequently faster than the slowest exchanging amides. This is the second mechanism of exchange. For these residues, at low [D], $\Delta G_{HX} < \Delta G_{unf}$ and the m value is lower than the m value calculated for residues exchanging via global unfolding. That is, these residues are less dependent on the denaturant for exchange when [D] is low. As [D] increases, exchange in these amides becomes dominated by global unfolding and the m values at high [D] become consistent with the m values of the slowly exchanging amides. Generally, one can find clusters of residues with similar ΔG_{HX}, allowing one to identify specific structural elements that unfold cooperatively.

Finally, for residues exchanging via the third mechanism, no dependence on [D] is seen. These amides undergo exchange through non-cooperative events such as local fluctuations involving single amides.

Analysis of exchange rates in ECAK and TNAK at different denaturant concentrations or temperature will allow one to delineate residues in the exchange core (the most slowly exchanging amides), structural elements that undergo cooperative, local unfolding and residues that are involved in local fluctuations. The ΔG_{HX} for the residues in the two proteins can be measured and compared, giving us a more detailed view of the level of structural stability level in different regions of ECAK and TNAK.

6.3 MD simulation of ECAK and TNAK

Like NMR, molecular dynamics simulation can provide details, at the atomic level, about motions in a protein as a function of time. The rapid increase in computing power and advances in potential energy functions describing the forces in a protein have reached a point where protein simulations can provide important dynamic aspects that can withstand critical examination by experimentalists Thus, simulations can now be designed to answer specific questions about the dynamics of a model system, often more readily than experimentation. Chapter 5 presented results from the MD simulation of ECAK in complex with its substrates, AMP, ATP, and Mg^{2+}. These simulations were carried out in collaboration with Dr. Robert. I. Cukier, Department of Chemistry, Michigan State University. The results presented in this work showed the importance of H-bonds and salt-bridges between AMP's adenine and phosphate moieties and ECAK. These bonds were shown to be the basis for the high specificity of AKs for AMP. The results also showed ATP, AMP, Mg^{2+} and its coordinating waters, and the active site side-chains to maintain an octahedral geometry throughout the simulation, thus lending strong support to an associative mechanism for phosphoryl transfer.

The simulation results of ECAK complexed with AMP and Mg·ATP show good agreement with experimental data, validating the fact that the AK system can be accurately simulated. Several *in silico* experiments can now be designed to compare the dynamics of ECAK and TNAK. Before TNAK simulations can be designed, crystal structures of TNAK in the free and substrate-bound forms are necessary. The accuracy of MD simulations depends on the quality of structural coordinates provided as input, and

structures obtained from homology modeling may not yield reliable results. Attempts to crystallize TNAK in the free form and in complex with various substrates are currently underway.

Future directions

With the crystal structure of TNAK in hand, simulations investigating the kinetic and thermodynamic stabilities of ECAK and TNAK can be performed. This would provide another means by which to compare the enzymes and provide a deeper understanding of the interplay between flexibility, stability, and activity. The following two MD simulation methods are of particular interest and are already being carried out on ECAK by Dr. Cukier's group:

1. Essential dynamics to study motions of AMPbd and lid domains in the two AKs.
2. Study the potential of mean force (PMF) along the reaction coordinate of conformational change involving the AMPbd and lid domains.

Thermodynamic stability of a protein is characterized by an ensemble of degenerate substates spanning a large conformational space (1). Due to thermal fluctuations, protein molecules experience dynamical processes that allow transitions between these substates non-periodically (5). These transitions are thought to be responsible for the wide timescale-range of motions observed in proteins. MD simulations can capture many of these motions, limited only by the length of the simulation. To extract the most important motion modes, principal component analysis (PCA) or essential dynamics (ED) is used to analyze the trajectories from an MD simulation. ED is a statistical method that can identify slow motion modes in proteins and, thereby, facilitate the study of long timescale

dynamics. These motional modes are sufficient to describe most of the fluctuations in the system under investigation. ED analyses of *Pyrococcus furiosus* and *Desulfovibrio vulgaris* rubredoxins (PfRd and DvRd, respectively) have revealed three main clusters of conformations for the two proteins (6). These clusters are much better defined in PfRd than in DvRd, indicating the mesophilic DvRd to be more flexible than the hyperthermophilic PfRd. ED analysis can also be used to define a reaction coordinate for conformational changes, which is required for another set of simulations called constrained MD simulations described below.

The second method of simulation of interest, called constrained MD simulations, uses the reaction coordinate for conformational closing/opening for ECAK and TNAK obtained from ED analysis, to determine the PMF along the reaction coordinate. The aim of this simulation is to calculate and compare the free energy barrier for conformational opening/closing for ECAK and TNAK. For a thermophilic AK such as AAAK, which has a much slower lid-opening rate than ECAK at 20°C, this simulation may reveal a higher energy barrier to opening for AAAK than ECAK. TNAK, on the other hand, has similar activity as ECAK at 30°C. It is expected, therefore, that the barrier to lid-opening for ECAK and TNAK may be similar.

In conclusion, NMR investigation of TNAK dynamics has shown that TNAK is able to maintain high activity at low temperatures and stability at high temperatures by partitioning rigidity for stability and flexibility for activity in its structure in an intricate manner in different timescales. At the same time, TNAK's activity at 30°C is about 10 times lower than at 80°C, a fact that may still be attributed to its high rigidity at 30°C. Although the prevailing rigidity hypothesis has existed for over 3 decades, and many

mesophilic-hyperthermophilic enzyme pairs have been studied to test its validity, it is only now that the hypothesis can be tested thoroughly. Developments in NMR and MD simulation techniques have opened the possibility of a more careful analyses of the dynamics in many more such protein pairs in atomic detail. Proteins have individual features of adaptation and several examples have to be studied to deepen our understanding of the role of flexibility in stability and activity.

6.4 References

1. **Frauenfelder, H., S. G. Sligar, and P. G. Wolynes.** 1991. The Energy landscapes and motions of proteins. Science **254:**1598-1603.

2. **Giver, L., A. Gershenson, P. O. Freskgard, and F. H. Arnold.** 1998. Directed evolution of a thermostable esterase. Proc. Natl. Acad. Sci. USA **95:**12809-12813.

3. **Hendsch, Z. S., and B. Tidor.** 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci. **3:**211-226.

4. **Hollien, J., and S. Marqusee.** 1999. Structural distribution of stability in a thermophilic enzyme. Proc. Natl. Acad. Sci. USA **96:**13674-13678.

5. **Krumhansl, J. A.** 1987. Presented at the Proceedings in Life Sciences, Austin.

6. **Lazaridis, T., I. Lee, and M. Karplus.** 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci. **6:**2589–2605.

7. **Nicholson, L. K., T. Yamazaki, D. A. Torchia, S. Grzesiek, A. Bax, S. J. Stahl, J. D. Kaufman, P. T. Wingfield, P. Y. S. Lam, P. K. Jadhav, C. N. Hodge, P. J. Domaille, and C. H. Chang.** 1995. Flexibility and function in HIV-1 protease. Nature Structural Biology **2:**274-280.

8. **Osborne, M. J., J. Schnell, S. J. Benkovic, H. J. Dyson, and P. E. Wright.** 2001. Backbone dynamics in dihydrofolate reductase complexes: Role of loop flexibility in the catalytic mechanism. Biochemistry **40:**9846-9859.

9. **Sawaya, M. R., and J. Kraut.** 1997. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence. Biochemistry **36:**586-603.

10. **Shapiro, Y. E., M. A. Sinev, E. V. Sineva, V. Tugarinov, and E. Meirovitch.** 2000. Backbone dynamics of *Escherichia coli* adenylate kinase at the extreme stages of the catalytic cycle studied by ^{15}N NMR relaxation. Biochemistry **39:**6634-6644.

11. **Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern.** 2004. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nature Structural & Molecular Biology **11:**945-949.

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