From the Department of Biosciences at Novum
Karolinska Institutet, Stockholm, Sweden

Structure Determination of *Methanocaldococcus jannaschii* Nucleoside Kinase

Linda Arnfors

Licentiate thesis

Stockholm 2005

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7140-447-3

Till farmor

ABSTRACT

All organisms can be divided into the tree kingdoms of life: Archaea, Bacteria and Eukarya. Archaeal organisms are often found to exhibit, compared to human perspectives, extreme environments, such as high temperature, high salt and acidic habitats. Many archaeal species are hyperthermophiles, i.e., microorganisms that grow optimally at temperatures above 80 °C. In hyperthermophiles, as in other organisms, metabolic reactions and other essential biological processes are catalyzed by enzymes, which are optimally active where mesophilic enzymes (optimally active at 20 – 50 °C) loose their structure and function. There appears to be no single mechanism for thermal stability. Instead, a combination of small – but relevant – structural changes, such as an increased number and optimal placement of ionic interactions and a reduction of the solvent-exposed hydrophobic surface, seem to be contributing to the enhanced thermotolerance. A better insight in what factors that make these enzymes able to retain their structure and function at high temperatures, will lead to an increased knowledge of protein stability, function and folding in general. This can in turn result in improved industrial processes and let us better understand diseases caused by incorrectly folded proteins, such as Alzheimer's, ALS and cystic fibrosis.

This licentiate thesis is mainly focused on the structure determination of *Methanocaldococcus jannaschii* nucleoside kinase (MjNK) by X-ray crystallography. Nucleoside kinase is a member of the ribokinase family, present in all domains of life. However, now structure is currently available for an archaeal representative of this protein family. The three-dimensional structure of MjNK will provide additional information on characteristics regarding thermal stability, enzyme mechanism and evolution of the ribokinase family. In addition, the stability of MjNK was studied by differential scanning calorimetry.

Nucleoside kinase is a homodimer with 34 kDa subunits. In the presence of ATP and Mg2+, MjNK is able to phosphorylate a wide range of nucleosides and shows the highest catalytic activity for cytidine, inosine, guanosine, and adenosine. The apparent melting temperature was 90 °C at pH 7.0. Moreover, the enzyme shows a kinetically dependent transition in 10 mM glycine pH 3.0. The three-dimensional structure of MjNK was determined by the multiple-wavelength anomalous dispersion technique using a Se-Met derivative. Additional crystal structures were determined for the apo-enzyme at 1.7 Å and MjNK in complex with an ATP-analogue and adenosine at 1.9 Å. Nucleoside kinase comprises one α/β domain and a smaller lid domain. The enzyme has an overall fold homologous to the members of the ribokinase superfamily and is assumed to catalyze the phosphorylation reaction similarly to the superfamily members. MjNK shares the highest structural similarity to ribokinase from *E. coli*. The structures of MjNK and ribokinase were compared regarding determinants for thermal stability. Relative to ribokinase, MjNK shows an increased charged and a decreased hydrophobic accessible surface area, a higher amount of charged residues as well as ionic networks and large aromatic clusters, characteristics that frequently are observed in enzymes from hyperthermophiles.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Alla levande organismer på jorden kan placeras in i någon av de tre huvudgrenarna på livets träd: arkéer, bakterier och eukaryoter. Människor, djur och växter är eukaryoter. *E. coli* är ett exempel på en bakterie. Det som karakteriserar arkéerna är att de ofta lever i extrema miljöer. Många arkéer är hypertermofiler – värmeälskande mikroorganismer som trivs bäst vid temperaturer över 80 grader. En del forskare tror att de första levande organismerna på jorden var just hypertermofiler eftersom de lever under samma förhållanden som rådde när de första livsformerna på jorden utvecklades. Den mest värmetåliga hypertermofilen som man känner till idag kan överleva 121 grader, samma temperatur som används för mikrobiell sterilisering.

Precis som för en mänsklig cell, så innehåller en hypertermofil cell hundratusentals olika proteiner. En del av dessa är enzymer som skyndar på biologiska reaktioner utan att själv förbrukas. Till skillnad från mänskliga proteiner som förstörs vid höga temperaturer så fungerar enzymer från hypertermofiler som allra bäst vid temperaturer nära vattnets kokpunkt. Att ett hönsägg stelnar när vi kokar det beror just på att proteinerna i ägget förstörs, eller denatureras. Tänk dig att proteinerna i ägget byts ut mot hypertermofila proteiner som klarar 110 grader, då skulle ägget aldrig stelna! Om vi bättre kan förstå vad som gör att de hypertermofila enzymerna kan behålla sin struktur och funktion vid dessa extrema temperaturer, då har vi större möjlighet att kunna förklara vad som gör ett protein stabilt. Denna kunskap kan i sin tur leda till en effektivisering av industriella processer och kanske framför allt bidra till en ökad förståelse av sjukdomar som orsakas av defekta proteiner, som till exempel Alzheimers, ALS och cystisk fibros.

Ett protein kan liknas vid ett uppknäppt pärlhalsband som formas när du knyter handen. Pärlorna representerar aminosyror länkade ihop till en proteinkedja som är veckad på ett specifikt sätt. Enzymer från hypertermofiler har ofta liknande övergripande veckning som motsvarande enzym från så kallade mesofiler – de mikroorganismer som lever i temperaturer mellan 20 och 50 grader. Däremot har man funnit skillnader mellan vilka aminosyror som bygger upp proteinet samt antalet och placeringen av olika interaktioner mellan proteinatomerna (kol, syre, kväve, svavel). Med en metod som heter röntgenkristallografi kan proteinkedjan identifieras och de atomer som bygger upp aminosyrorna kan lokaliseras.

Den här avhandlingen baseras i huvudsak på strukturbestämning med röntgen-kristallografi av enzymet nukleosidkinas från den hypertermofila arkéen *Methanocaldococcus jannaschii*. Nukleosidkinas fosforylerar – sätter en fosfatgrupp på – en rad olika nukleosider. Fosforylerade nukleosider, det vill säga nukleotider, är beståndsdelar i viktiga biologiska molekyler, som DNA. Nukleosidkinas har en struktur som är mycket lik den för ribokinas i *E. coli* och enzymerna katalyserar troligtvis sina substrat på liknande sätt. Nukleosidkinas påvisar många av de egenskaper som karakteriseras av proteiner från hypertermofiler. Till exempel så innehåller nukleosidkinas mycket fler interaktioner mellan plus- och minusladdade atomer och stabiliseras troligen också av stora kluster av aromatiska grupper.

LIST OF PUBLICATIONS

This thesis is based on the following publications referred to by their roman numerals in the text.

I. **Linda Arnfors**, Thomas Hansen, Winfried Meining, Peter Schönheit and Rudolf Ladenstein
Expression, purification, crystallization and preliminary X-ray analysis of a nucleoside kinase from the hyperthermophile *Methanocaldococcus jannaschii*.
Acta Cryst. (2005) **F61**, 591-594.

II. Thomas Hansen, **Linda Arnfors** and Peter Schönheit
PFK-B from *Methanocaldococcus jannaschii* (ORF MJ0406) represents a nucleoside kinase with broad substrate specificity.
Manuscript

III. **Linda Arnfors**, Thomas Hansen, Winfried Meining, Peter Schönheit and Rudolf Ladenstein
Crystal structure of *Methanocaldococcus jannaschii* nucleoside kinase – an archaeal member of the ribokinase family.
Manuscript

CONTENTS

LIST OF ABBREVIATIONS

AMP-PNP	Phosphoaminophosphonic acid-adenylate ester
AIR	Aminoimidazole riboside
ALS	Amyotrofisk lateralskleros
ASA	Accessible surface area
β-OG	β-Octyl glucoside
EcRK	*Escherichia coli* ribokinase
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
e.g.	Lat. exempli gratia = for example
Fru-6-P	Fructose-6-phosphate
KDG	2-keto-3-deoxygluconate
HMPP	4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate
i.e.	Lat. id est = that is
MjNK	*Methanocaldococcus jannaschii* nucleoside kinase
MO6	Magnesium ion coordinated by 6 water molecules
NMR	Nuclear magnetic resonance
PDB	Protein data bank
PEG	Polyethylene glycole
Pfk	Phosphofructokinase
RK	Ribokinase
rmsd	Root mean square deviation
SeMet	Selenium methionine
ThiD	Phosphomethyl pyrimidine kinase
THZ	4-methyl-5-β-hydroxyethylthiazole
YXKO	Open reading frame in the YXCO locus of *Bacillus subtilis*

1 INTRODUCTION

For a long period of time, organisms were divided into only two domains, prokaryotes (organisms lacking the cell nucleus) and eukaryotes (organisms comprising the cell nucleus). In the 1970's, Carl Woese suggested that there instead are three lines of evolution – eukaryotes, eubacteria and archaeabacteria (Woese & Fox, 1977). Eubacteria (greek eu = real) were believed to be the true bacteria. The name archaebacteria (greek archein = to begin) was given since these organisms live in environments which resemble those thought to have existed when life first started on Earth. Woese based his studies on ribosomal RNA, molecules that play an important role in the protein synthesis in all cells. The genes coding for ribosomal RNA are highly conserved among all forms of life and the sequences of the four bases cytosine, guanine, adenine, and thymine, which is replaced by uracil in RNA, are almost identical. Woese continued his analysis of the ribosomal RNA of eukaryotes, eubacteria and archaebacteria and showed that the organism groups were genetically so different that the archaebacteria branch was renamed Archaea and the eubacteria were simply termed Bacteria (Woese et al., 1990). Today the division of organisms into the three kingdoms of life: Eukarya, Bacteria, and Archaea is well established and accepted.

Humans, animals, plants and fungi, including *Saccharomyces cerevisiae*, also known as baker's yeast, are all eukaryotes. Among bacteria, one can find the widespread *Escherichia coli*, *Lactobacillus acidophilus* which is important to preserve a healthy intestinal microflora and which is also an additive in dairy, and the pathogen *Streptococcus pneumoniae* which causes severe diseases such as pneumonia and meningitis. The archaeal group is not as well known among people as the two others. Most archaeal species are extremophiles, i.e., organisms that live in, according to human perspectives, extreme environments. Archaeal organisms have been found to live in extreme temperature environments (thermophiles and psychrophiles), high salt concentrations (halophiles), extremes of pH (acidophiles and alkalophiles), and under other extreme conditions such as pressure (Danson & Hough, 1998). However, archaeal organisms have as well been found to thrive in eukaryotic habitats. Archaea have been detected in both animals and humans, but so far no pathogenic archaea have been found (Lange et al., 2005). The role of the archaeal microflora in animals has not yet been established.

According to comparisons of small subunit ribosomal RNA sequence, Archaea can be divided into the major phyla Crenarchaeota and Euryarchaeota (Forterre et al., 2002). Two additional phyla have been suggested, the Korachaeota (Barns et al., 1994) and the Nanoarchaeota (Huber et al., 2002). Euryarchaeota comprise halophiles, methanogens, some thermoacidophiles and some hyperthermophiles. So far, only hyperthermophilic organisms including *Sulfolobus*, *Aeropyrum* and *Thermoproteus*, have been identified in the Crenarchaeota lineage, but there is evidence for both mesophilic and psychrophilic Crenarchaeotes as well (Forterre et al., 2002). Several crenarchaeotic species are also acidophiles and live under very acidic conditions.

1.1 HYPERTHERMOPHILES LIKE IT HOT

Depending on their optimal growth temperature, microorganisms are classified into the following four groups (optimal growth temperature in parenthesis): psycrophiles (-5 to 20 °C), mesophiles (20-50 °C), thermophiles (50-80 °C), and hyperthermophiles (≥ 80 °C) (Karshikoff & Ladenstein, 2001). The word *thermophile* origins from ancient greek (thermos = heat; philein = to like) and has the meaning 'heat-lover'. Hyperthermophiles are thus 'extreme heat-lovers'. But hyperthermophilic organisms do not only love heat, they require high temperatures for survival. Most hyperthermophiles that have been described are archaea such as *Pyrococcus*, *Methanopyrus*, *Methanocaldococcus, Sulfolobus,* and *Aeropyrum*. The only hyperthermophilic bacteria known today are *Thermotoga* and *Aquifex*.

In the late 1960's, the first hyperthermophile *Sulfolobus acidocaldarius* (Brock et al., 1972), which also is an acidophile, was found in a hot, acidic spring in Yellowstone National Park, Wyoming. Since then, over fifty hyperthermophiles have been identified. Some have been found in volcanic vents and solfataric fields, but most of them have been found in marine hydrothermal vents, such as white and black smokers at the bottom of the Pacific Ocean (Stetter, 1996). The most thermophilic organism known is called 'strain 121' (Kashefi & Lovley, 2003), which is able to survive and even double in cell numbers after 24 hours at 121 °C (hence its name), the temperature used for sterilizing in autoclaves. This archaeon can even survive short periods at 130 °C. It is still unknown what the upper temperature limit for life is. It is worth noting, that at temperatures above 110°C, amino acids and metabolites are highly unstable and hydrophobic interactions become significantly weaker (Vieille & Zeikus, 2001; Yip et al., 1998). It would be unlikely to find microorganisms living at temperatures higher than 150 °C (Stetter, 1998). At this point, it would be impossible to avoid the breakdown of the chemical bonds keeping DNA and other vital molecules together.

Organisms living in hot environments need to stabilise their DNA and RNA against thermal denaturation (Grogan, 1998). The simplest approach is to have a high content of the bases guanine and cytosine, since the G-C base pair is more tightly bound and thus more stable at elevated temperatures, than is the A-T/U base pair. However, the genomes of hyperthermophiles are not always rich in G and C and hyperthermophiles rather stabilize their DNA by various other mechanisms, such as increased intracellular ion concentrations, cationic proteins, and supercoiling (Daniel & Cowan, 2000; Galtier & Lobry, 1997; Grogan, 1998). RNA molecules, on the other hand, seem to be stabilized by the stronger interactions of the G-C bond. The genes coding for ribosomal RNA and transfer RNA in hyperthermophiles have an increased GC content which show a strong correlation with the thermal stability (Galtier & Lobry, 1997).

Hyperthermophiles are believed to be the most descendent organisms, since they occupy the shortest and deepest branches and are found closest to the root of the phylogenetic trees. However, a study of slowly evolving sequences placed a non-thermophilic group at the base of the bacterial family tree (Brochier & Philippe, 2002).

Some researchers suggest that life instead began at low temperatures in shallow waters instead of in a hot environment (Whitfield, 2004).

The unique features of hyperthermophiles and other extreme organisms are taken into account for in the development of new industrial applications (Schiraldi et al., 2002). For instance, the hyperthermophilic and sulphur-reducing archaeon *Pyrococcus furiosus* is used in the process of rubber recycling (Bredberg et al., 2002).

1.1.1 The hyperthermophilic archaeon *Methanocaldococcus jannaschii*

Methanocaldococcus jannaschii is a hyperthermophilic euryarchaeon with an optimal growth temperature of 85 °C (Jones et al., 1983). It was originally isolated from a "white-smoker" chimney at 2,600 meters depth in the Pacific Ocean. *M. jannaschii* is a strict anaerob and grows at pressures higher than 200 atm. As the name implies, *M. jannaschii* is a methane producer. It grows chemolitoautotrophically and requires only H_2 and CO_2 for growth and methane formation.

The complete genome of *Methanocaldococcus jannaschii* was published in 1996 by Bult and collaborators (Bult et al., 1996). It was the third microorganism ever to have its genetic map entirely sequenced and the first archaeon. The genome of *M. jannaschii* consists of 1,738 genes where only 38 % could be assigned as recognizable protein-coding regions with high confidence. Most of the genes involved in energy production, cell division and metabolism in *M. jannaschii* share the highest similarity to bacterial genes, and the majority of the genes related to transcription, translation and replication share the highest similarity to eukaryotic genes.

1.2 ENZYMES FROM HYPERTHERMOPHILES

In hyperthermophiles, metabolic reactions and other important processes are catalyzed by enzymes that are optimally active under these extreme conditions. Enzymes from thermophilic organisms are of great interest for industrial processes. At elevated temperatures, chemical reactions often run at higher speed and there is a reduced risk of microbal contamination. Of particular importance is the commercial application of DNA polymerases from *Thermus aquaticus*, *Pyrococcus furiosus* and *Thermococcus litoralis* in the polymerase chain reaction (PCR), the rapid and efficient method for DNA amplification, which is used worldwide in e.g., forensic science, clinical medicine and food analysis.

An increased knowledge of structural, thermodynamic and kinetic factors responsible for function and stability of extreme enzymes is important for a better understanding of protein stability, function and folding in general (Karshikoff & Ladenstein, 2001). This information is essential in protein engineering, design of more stable enzymes for industrial processes or enzymes with a specific character of interest, as well as in structure prediction of biomolecules. Additional knowledge in this field could as well give a clue about the evolutionary path of life and make us better understand how life on Earth first started. Furthermore, a wider understanding of protein folding could possibly and hopefully be a piece of a puzzle to find a cure for diseases caused by incorrectly folded proteins, such as Parkinson's disease, Alzheimer's disease, ALS,

Huntington's disease, cystic fibrosis and the prion diseases mad-cow disease (BSE) and the human variant Creutzfeldt-Jacob disease (Dobson, 2003).

Enzymes from thermophiles and hyperthermophiles are *thermostable* in the sense of resistant to irreversible inactivation at high temperatures and *thermophilic* in the sense of optimally active at temperatures above 60 °C (Li et al., 2005). Enzymes from hyperthermophiles are usually intrinsically thermostable and require only in a few cases extrinsic factors, such as salts or polyamines, or posttranslational modifications, such as glycosylation, to be fully thermostable (Vieille & Zeikus, 2001). However, in hyperthermophiles a number of so-called compatible solutes have been found, which have a strong stabilization effect (Santos & da Costa, 2002). There is no single universal mechanism responsible for the remarkable stability of enzymes from hyperthermophiles (Karshikoff & Ladenstein, 2001; Matthews et al., 1974). Hyperthermostable enzymes contain the same amino acids and are composed of the same structural motifs, such as α-helices and β-sheets, as are enzymes from mesophiles (Fields, 2001; Li et al., 2005). Homologous enzymes from mesophiles and hyperthermophiles are often highly similar with amino acid sequences that typically share 40-85 % similarity and three-dimensional structures that in many cases are superposable (Vieille & Zeikus, 2001). In addition, the catalytic mechanism is often the same. Instead, it appears to be a combination of small, but relevant, structural changes that is the reason for the ability of thermostable enzymes to function at elevated temperatures. Site-directed mutagenesis experiments as well as comparisons of structure and stability of enzymes from hyperthermophiles and their mesophilic homologues have revealed some important factors that contribute to the enhanced

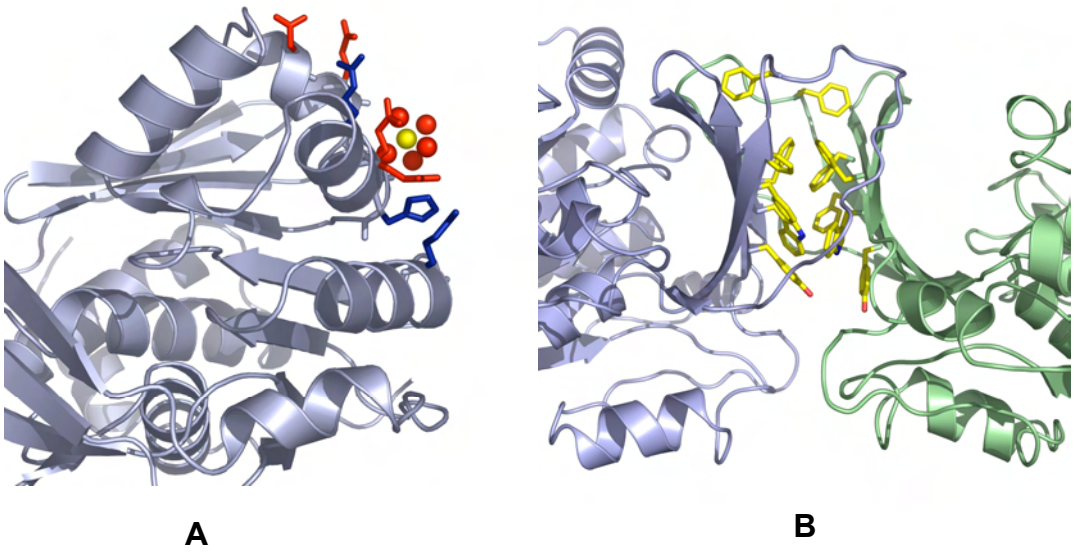

Figure 1. Ionic networks at the surface of the enzyme and an increased number of aromatic interactions might contribute to the thermal stability of an enzyme. A) the seven-member ionic network of *Methanocaldococcus jannaschii* nucleoside kinase on the protein surface and next to an magnesium ion (yellow sphere) coordinated by five waters (red spheres). Residues in red are negatively charged and those in blue are positively charged. B) In the dimeric interface of nucleoside kinase, six aromatic residues from each monomer form a twelve-member aromatic network, which most likely is one of the determinants for the thermal stability of this enzyme.

thermostability (Ladenstein & Antranikian, 1998). Hyperthermostable enzymes normally show, compared to enzymes from mesophiles, a better optimization of the overall conformational structure, such that they are often better packed and show an optimization of the hydrophobic effect (Yutani et al., 1987), a stabilization of the α-helix dipole (Nicholson et al., 1988), an increased intensity of low-frequency local fluctuations of the protein molecule (Wintrode et al., 2003), and an reduced entropy of unfolding (Matthews et al., 1987). Moreover, and perhaps of particular importance, hyperthermostable enzymes are characterized by additional and optimally placed intermolecular interactions, such as ion-pair interactions, hydrogen bonds, aromatic interactions, and disulfide bridges (Fig. 1) (Vogt & Argos, 1997). Stabilizing interactions in hyperthermostable enzymes are often found in the less conserved regions of the protein. In many cases the cores of mesostable and hyperthermostable enzyme homologues are highly similar (Vieille & Zeikus, 2001). Hyperthermostable enzymes often contain ion-pair interactions, typically large ionic networks, on the protein surface, a decreased hydrophobic solvent accessible surface (Wigley et al., 1987), and an increase in anchoring of "loose ends" (i.e., the N- and C-termini and loops) to the protein surface (Vieille & Zeikus, 2001). Some of the structural determinants for thermal stability will be described in more detail below.

1.2.1 Optimization of the hydrophobic effect

Hyperthermophilic enzymes have in many cases a better optimized packing than enzymes from mesophiles (Vieille & Zeikus, 2001). A better internal packing is often related to increased hydrophobicity and an optimization of the hydrophobic effect.

The hydrophobicity of a protein is defined as the ratio of the buried apolar surface area and the total apolar surface area of the protein (Li et al., 2005). For example, the introduction of a methyl group in a cavity increases the hydrophobic interactions with the surrounding hydrophobic residues and thus increases the hydrophobicity of the protein. An average increase in stability was estimated to 1.3 ± 0.5 kcal mol^{-1} for each additional methyl group that is buried in the molecule during protein folding (Pace, 1992). A larger and denser hydrophobic core results accordingly in a reduction of the size and/or number of cavities as well. The elimination of unnecessary cavities has been found to contribute to the thermal stability of some hyperthermostable enzymes (Ishikawa et al., 1993; Russell et al., 1997). However, a comparison of partial volumes, void volumes, and cavity volumes in 80 non-homologous mesophilic, 20 thermophilic and 4 hyperthermophilic enzymes showed no correlation between better packing and enhanced thermostability (Karshikoff & Ladenstein, 1998).

The hydrophobic effect is assumed to make the major contribution to molecular folding and thermostability (Goodenough & Jenkins, 1991). The hydrophobic effect results from the tendency of apolar groups to interact with each other in order to minimize their contacts to polar groups (Rose & Wolfenden, 1993). In the protein folding process, this reflects a propensity of apolar side chains to leave the solvent and form the hydrophobic core of the protein. Thus, an optimization of the hydrophobic effect results in a denser hydrophobic core and a reduction of the solvent-exposed hydrophobic surface. Hydrophobic residues at the protein surface are unable to participate in stabilizing interactions with the surrounding solvent. Owing to the

contribution of negative entropy to the free energy when highly ordered water molecules are formed around apolar groups, the surface hydrophobic groups are reducing the stability and solubility of the protein. Enzymes from hyperthermophiles often show an increase in the interior and a decrease of the exterior hydrophobic residues compared to mesophilic enzymes (Spassov et al., 1995). Many hyperthermostable enzymes have a significant reduction in hydrophobic accessible surface area (ASA) compared to their mesophilic homologues (Spassov et al., 1995). The decrease in hydrophobic ASA is balanced by an increase in mainly charged ASA but also polar ASA. Both the decrease in hydrophobic ASA and the increase in charged ASA are believed to contribute to the thermal stability of enzymes.

1.2.2 Electrostatic optimization

Electrostatic optimization is essential for a protein's adaptation to the environment and is therefore an important factor for enhanced thermal stability (Dominy et al., 2004; Karshikoff & Ladenstein, 2001; Perutz, 1978; Spassov et al., 1995). Many factors have been found to contribute to the electrostatic optimization in hyperthermophilic enzymes, such as an increased number and an optimal placement of ion-pairs and ionic networks, the cooperativity in ionic networks, a reduced electrostatic desolvation penalty, and an increase in the dielectric constant of the protein (Dominy et al., 2004; Elcock, 1998; Lebbink et al., 2002; Lebbink et al., 1999; Spassov et al., 1994; Xiao & Honig, 1999; Yip et al., 1995).

An ion-pair is an electrostatic charge-charge interaction between oppositely charged residues. Ion-pairs are also called salt bridges. Ionic networks are formed through interactions between two or more ion-pairs. The only residues that can contribute to an ion-pair interaction are the positively charged Arg, Lys, and His, and the negatively charged Asp and Glu.

In 1975, Perutz and Raidt suggested that electrostatic interactions are important for thermal stability (Perutz & Raidt, 1975). Recent genomic analyses have indeed revealed that the amount of charged residues, particularly Glu, Lys and Arg, of the overall amino acid sequence is often higher in hyperthermophilic enzymes than in equivalent enzymes from mesophiles (Cambillau & Claverie, 2000; Das & Gerstein, 2000). However, both experimental and theoretical studies on proteins have indicated that at room temperature ion-pairs only marginally stabilize and can even destabilize the folded state of proteins (Elcock, 1998; Hendsch & Tidor, 1994; Honig & Nicholls, 1995; Horovitz et al., 1990). Bringing two charged molecules together in the enzyme interior is associated with a large desolvation penalty, which is seldom completely compensated by energetically favourable interactions within the ion-pair and with the rest of the protein (Elcock, 1998). However, at increasing temperatures the electrostatic desolvation penalty is decreasing because of the decrease of the dielectric constant of water at high temperatures. This leads to a higher energy barrier to solvate an ion-pair and thereby incurs the formation of a salt bridge. The direct consequence is that ion-pairs might contribute to the thermal stability of a protein.

Relative to mesophilic homologues, hyperthermophilic enzymes have been shown to contain a higher number of ion-pairs, which often participate in ionic networks located

at the protein surface (Fig. 1a) as well as at inter-domain and, in the case of oligomeric proteins, inter-subunit interfaces (Korndörfer et al., 1995; Lim et al., 1997; Yip et al., 1995; Zhang et al., 2001). For example, *A. aeolicus* lumazine syntase was shown to have the largest charged ASA and the smallest hydrophobic ASA compared to other lumazine syntase structures (Zhang et al., 2001). Furthermore, the hexameric *Pyrococcus furiosus* glutamate dehydrogenase contains an intersubunit ionic network composed of 24 residues – belonging to four different subunits – connected by 18 ion-pairs (Yip et al., 1998). An ionic network is energetically more favourable compared to an equivalent number of isolated ion-pairs. For each new ion-pair formed within the network it is only one charged residue that must be buried and desolvated compared to two charged residues for isolated ion-pairs (Yip et al., 1995). The average stabilization contribution to the free energy by a surface ion-pair has been estimated to -1.0 kcal mol^{-1} (Perutz & Raidt, 1975).

Although it has been shown experimentally that substitutions of single amino acids to charged surface residues can enhance protein stability (Grimsley et al., 1999) and that ionic networks more frequently occur in hyperthermophilic proteins, it is not only the quantity of charged residues and ion-pair interactions that are decisive for the enhanced thermal stability. The minimization of repulsive contacts was demonstrated to be more important for the optimization of electrostatic interactions than the formation of salt bridges (Karshikoff & Ladenstein, 2001; Spassov et al., 1994). Enzymes from hyperthermophiles show besides an increased number of ion-pairs and surface networks, also a reduction of repulsive and unfavourable contacts. In addition, hyperthermostable enzymes are also characterized by an optimization of the charged residues, which is attained by minimizing the number of excess charges (Spassov et al., 1995).

In a study by Xiao and Honig, it was suggested that the favourable electrostatic interactions observed in hyperthermophilic enzymes may be achieved rather through an optimal placement of charged residues than by an increased number of charged amino acids, ion-pairs and ionic network in the protein (Xiao & Honig, 1999). Among the ion-pairs and ionic networks in the hyperthermophilic *Pyrococcus furiosus* glutamate dehydrogenase, some were found to enhance protein stability while other had a destabilizing effect. The contribution to the stabilization was found to strongly depend on the local environment of the particular network.

In the sense of protein stability, the cooperative effect of ion-pair interactions in a network means that a particular ion-pair interaction is strengthened if one of the ion-pair residues is involved in another ion-pair interaction. The engineered construction of a 16-residue ionic network in *Thermotoga maritima* glutamate dehydrogenase showed long-range cooperative enhancement of the ion-pairs within the network and it was suggested that the cooperative nature of the stabilizing interactions could result from the electrostatic fields generated by the residues involved in the network (Lebbink et al., 1999). The cooperative effect of stabilization in networks was later demonstrated in a study on a salt bridge triad in the NADP-domain of *T. maritima* glutamate dehydrogenase (Lebbink et al., 2002). From molecular dynamics simulations of two protein structure families, Dominy and co-workers found that the effect of dynamics on the charge distribution varies with thermostability (Dominy et

al., 2004). Particularly the dielectric response, which is related to the fluctuation of the dipole moment of a protein, was found to increase with thermostability. The dielectric response is a function of the number and distribution of the charges within a protein and has a direct impact on long-range electrostatic interactions (Dominy et al., 2004; Lebbink et al., 2002; Lebbink et al., 1999). Moreover, the dielectric response was shown to be closely connected to the complementary pairs of charged residues on the protein surface.

1.2.3 Increased number of hydrogen bonds

An increase in the number of hydrogen bonds is thought to be related to enhanced thermostability, since many hyperthermophilic enzymes contain more hydrogen bonds than their mesophilic counterparts (Vogt & Argos, 1997). It was estimated that a hydrogen bond contributes to the stabilization of the free energy by –1.6 kcal mol^{-1} (Pace et al., 1996). Hydrogen bonds are typically defined by a distance of less than 3 Å between the hydrogen donor and the hydrogen acceptor, and by a donor-hydrogen-acceptor angle below 90 °C (Vieille & Zeikus, 2001). Protein-solvent interactions are considered as an important factor responsible for protein stability (Karshikoff & Ladenstein, 2001). Hydrogen bonds do not only occur within the enzyme but are also present between the enzyme surface and the surrounding water molecules. The identification of hydrogen bonds is highly dependent on the distance cut-off and is also affected by the resolution limit of the crystal structure (Karlström et al., 2005; Vieille & Zeikus, 2001).

A strong correlation between thermostability and the number of charged-neutral hydrogen bonds has been observed for a few enzymes (Macedo-Ribeiro et al., 1996; Tanner et al., 1996). Charged-neutral hydrogen bonds occur between a side chain atom of a charged residue and either a main chain atom of any residue or a side chain atom of a neutral residue. These hydrogen bonds might be favoured over charged-charged hydrogen bonds (i.e., ion-pairs) and neutral-neutral hydrogen bonds for the following reasons: *1)* the desolvation penalty for forming a charged-neutral hydrogen bond is less than that for forming an ion-pair, and *2)* the enthalpy gained by forming a charged-neutral hydrogen bond is greater than forming a neutral-neutral hydrogen bond, because of the dipole moment (Tanner et al., 1996). Thus, charged residues in hyperthermophilic enzymes may not only stabilize through ion-pairs, but also through charged-neutral hydrogen bonds.

1.2.4 Aromatic interactions

The contribution of aromatic interactions to thermal stability (Fig. 1b) is not as well studied as other determinants, such as increased number of ion-pairs and ionic networks, better hydrogen bonding, and a reduced hydrophobic accessible surface. However, aromatic interactions are known to be important for structural stability of proteins (Burley & Petsko, 1985). Protein engineering studies have shown that the introduction of aromatic interactions into a protein increases the thermal stability (Burley & Petsko, 1985). One pair of aromatic interaction was estimated to stabilize a protein by –0.6 to –1.3 kcal mol^{-1} (Serrano et al., 1991).The common definition of aromatic-aromatic interactions implies a distance of less than 7 Å between the centres

of the two aromatic rings (Burley & Petsko, 1985). Aromatic clusters are interactions between three or more aromatic residues.

The aromatic residues comprising a pair-wise aromatic interaction can have three different geometric orientations with regard to their inter-planar angles: near-parallel face to face interaction (0°-30°), tilted geometry (30°-60°) and perpendicular T-shaped packing geometry (60°-90°). The T-shape packing is energetically most favourable and was shown to be the preferred geometry in aromatic pairs (Burley & Petsko, 1985). In order to study the contribution of aromatic interactions to the thermal stability of proteins, Kannan and Vishveshwara compared the crystal structures of homologous thermophilic and mesophilic enzymes from 24 protein families (Kannan & Vishveshwara, 2000). Their analysis showed that additional aromatic clusters or enlarged aromatic networks were present in 17 of the thermophilic enzymes, but were absent in the corresponding mesophilic homologues. An analysis of all pair-wise aromatic interactions (i.e., including both aromatic pairs and aromatic clusters) in the 24 thermophilic enzymes revealed that the pair-wise aromatic interaction was preferably in T-shaped or tilted geometry, rather than in parallel geometry (Kannan & Vishveshwara, 2000). In addition, the thermophilic enzymes were shown to have a larger number of pair-wise aromatic interactions compared to their mesophilic homologues. The additional clusters found in the thermophilic enzymes mostly involved pair-wise aromatic interactions and were more often located on the protein surface where they built up more rigid regions of the surface (Kannan & Vishveshwara, 2000). The additional aromatic cluster residues were usually replaced by Leu, Ile or Ser in the mesophilic enzyme.

The cation-π interaction is another type of interaction involving aromatic residues (Burley & Petsko, 1985; Ma & Dougherty, 1997). In this kind of interaction, the electron-rich π-ring of an aromatic residue interacts with a positive charge of a metal cation or the quaternary amine of the side chain of Arg, Lys or His. The cation-π interactions have been found in proteins, but its contribution to thermal stability has not been studied in detail. However, the arginine-phenylalanine pairing was found to be the only cation-π interaction that provided significant stabilization of a designed peptide structure. (Slutsky & Marsh, 2004).

1.3 THE RIBOKINASE FAMILY AND RIBOKINASE SUPERFAMILY

1.3.1 Background

Carbohydrates, or sugars, are an important energy source, such as glucose, and are essential molecules in nucleotide biosynthesis, such as ribose, in a cell. Nucleotides, for instance in the form of ATP, are energy-rich compounds that drive metabolic processes in all cells. They function as chemical signals, important mediators in cellular systems that respond to hormones and other extracellular stimuli, and are structural components in enzyme cofactors and metabolic intermediates. In addition and of great importance, nucleotides are the building blocks of DNA and RNA – the molecules carrying all genetic information (Lehninger et al., 1993).

When a sugar enters the cell it becomes phosphorylated by specific carbohydrate kinases. This is the first step in the sugar metabolism of the cell. The phosphorylated sugar is unable to cross the cell membrane and is thus trapped inside the cell. Glycolysis is a catabolic pathway in which glucose is metabolized and energy gained. Phosphofructokinase (Pfk) is a key regulatory enzyme in glycolysis and catalyzes the reaction of fructose-6-phosphate (Fru-6-P) to fructose-1,6-bisphosphate. *E. coli* comprises two Pfks, referred to as PfkA and PfkB. PfkA is the major Pfk and is a homotetramer with subunits of 35 kDa. The enzyme shows cooperative kinetics with respect to Fru-6-P and is allosterically controlled by phosphoenolpyruvate (Blangy et al., 1968). PfkB is the minor Pfk of *E. coli*. It is a homodimer consisting of subunits of about 33 kDa and shows neither cooperative kinetics nor inhibition by phosphoenolpyruvate (Babul, 1978; Kotlarz & Buc, 1981). The two Pfks of *E. coli* were found to belong to different protein families (Wu et al., 1991) – the PfkA and PfkB family of carbohydrate kinases – which are named after the *E. coli* enzymes. The PfkA family comprises ATP-dependent Pfks from higher eukaryotes and bacteria together with PP_i-dependent Pfks from bacteria, plants, and archaea. Homologous carbohydrate kinases, including PfkB of *E. coli,* constitute a separate family. Wu and collaborators termed this family the PfkB family (Wu et al., 1991). On the basis of their similar functions, sizes and sequences, the two protein families were believed to share a common evolutionary origin.

A few years after Wu's discovery, Bork and co-workers presented their results from amino acid sequence comparison of more than 60 carbohydrate kinases. Their research revealed that sugar kinases can be divided into three distinct non-homologous families named the hexokinase, ribokinase, and galactokinase families of carbohydrate kinases (Bork et al., 1993). The ribokinase family and the PfkB family are different names for the same protein family. The three families were believed to have different three-dimensional structures since there is no significant sequence similarity between the enzymes of the three families and conserved sequence patterns are different in both length and location. However, the enzymes have similar specificity and catalyze phosphorylation reactions of similar or identical substrates (fructokinases exist in both the hexokinase and ribokinase family). Accordingly, it was suggested that the similar enzymatic function of carbohydrate phosphorylation has evolved independently for each of the three families by convergent evolution.

Since then, more knowledge has been gained about the ribokinase family and crystal structures of a number of members have been determined, but still there are unanswered questions for example regarding the evolution of the ribokinase family. The structures available today of both enzymes belonging to the ribokinase family and to the ribokinase superfamily are presented in Table 1.

1.3.2 Ribokinase family

The ribokinase family (Bork et al., 1993; Wu et al., 1991), or PfkB family, is a diverse protein family present in all three domains of life: Archaea, Bacteria and Eukarya. The ribokinase family is characterized by amino acid sequence analysis and represents the protein family (Pfam) PF00294. This family comprises a variety of carbohydrate and

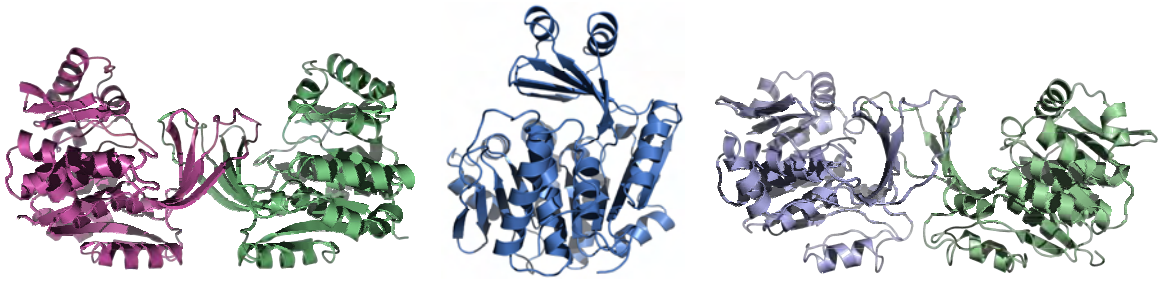

Figure 2. *E. coli* ribokinase (Sigrell et al., 1998; Sigrell et al., 1999) and human adenosine kinase (Mathews et al., 1998) were the first crystal structures of enzymes belonging to the ribokinase family. Both enzymes comprise an α/β domain and a smaller lid domain and share the overall monomeric fold characteristic for members of the ribokinase superfamily. However, ribokinase is active as a homodimer whereas adenosine kinase functions as a monomer. The structure of *Methanocaldococcus jannaschii* nucleoside kinase is very similar to *E. coli* ribokinase.

pyrimidine kinases such as ribokinase, adenosine kinase, fructokinase, 2-dehydro-3-deoxy-glucokinase, 1-phosphofructokinase, phosphomethylpyrimidine kinase, inosine-guanosine kinase and the minor 6-phosphofructokinase from *E. coli* (PfkB).

The first crystal structures within the ribokinase family were *E. coli* ribokinase (Sigrell et al., 1998) and human adenosine kinase (Mathews et al., 1998) (Fig. 2.). Today, there are five additional structures of characterized enzymes belonging to the ribokinase family. These are *Toxoplasma gondii* adenosine kinase (Schumacher et al., 2000), sheep brain pyridoxal kinase (Li et al., 2002), *Bacillus subtilis* 4-methyl-5-β-hydroxyethylthiazole (THZ) kinase (Campobasso et al., 2000), *Thermus thermophilus* 2-keto-3-deoxygluconate (KDG) kinase (Ohshima et al., 2004), and *E. coli* pyridoxamine kinase (Safo et al., 2004). In addition, crystal structures are available for the following five uncharacterized enzymes: PfkB carbohydrate kinase (Joint Center for Structural Genomics, unpublished results-a), possible 1-phosphofructokinase (Joint Center for Structural Genomics, unpublished results-b) and KDG kinase (Joint Center for Structural Genomics, unpublished results-c) from *Thermotoga maritima*, as well as *Thermus thermophilus* phosphomethyl pyrimidine kinase (ThiD) (Bagautdinov et al.) and the putative YXKO kinase (named after the open reading frame in the YXKO locus) from *Bacillus subtilis* (Zhang et al., 2002). At present there is no published structure of an archaeal member of the ribokinase family.

1.3.3 Ribokinase superfamily

The twelve structures of the members of the ribokinase family have a similar overall fold which is characteristic for the ribokinase superfamily (SCOP 53613) according to the SCOP (structural classification of proteins) database. The ribokinase superfamily comprises, besides the ribokinase family members, *Salmonella typhimurium* 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMPP) kinase (Cheng et al., 2002), *Pyrococcus horikoshii* hydroxyethylthiazole (Jeyakanthan & Tahirov) kinase protein and the ADP-dependent glucokinases from *Thermococcus litoralis* (Ito et al., 2001), *Pyrococcus furiosus* (Ito et al., 2003) and *Pyrococcus horikoshii* (Tsuge et al., 2002). The three glucokinases are archaeal representatives of the ribokinase superfamily,

however, they do not belong to the ribokinase family but instead comprise the ADP-specific phosphofructokinase/glucokinase family (Pfam, PF04587). In addition, *Salmonella typhimurium* aminoimidazole riboside (AIR) kinase (Zhang et al., 2004) was reported to belong to the ribokinase family and the ribokinase superfamily, but is not included in either Pfam or SCOP databases. However, amino acid sequence searches of AIR kinase in Pfam and SCOP reveal that the enzyme shares most similarity with the ribokinase family and its structure has an overall fold similar to the members of the ribokinase superfamily.

1.3.4 Characteristics of the ribokinase superfamily

Enzymes belonging to the ribokinase superfamily are phosphotransferases with a methyl hydroxyl group as acceptor for the phosphate group. *S. typhimurium* HMPP kinase is however able to catalyze both the phosphorylation of a hydroxymethyl pyrimidine (HMP) and the phosphorylation of the phosphomethyl group of HMPP. The ribokinase superfamily members share a similar overall fold. The enzyme core is an α/β fold composed of a central eight-stranded β-sheet and eight flanking α-helices. The location of the substrate binding site as well as certain conserved residues and regions are common characteristics among members of the ribokinase superfamily. The active site is positioned in a shallow groove along one edge of the β-sheet. The phosphate accepting hydroxyl group of the substrate and the γ-phosphate of ATP are close neighbours in the centre of the groove. The substrate and adenosine part of ATP are located at opposite ends of the groove. Two highly conserved residues in the enzymes are the aspartate and asparagine that form hydrogen bonds with the 2' and 3'-hydroxyl groups in the ribose ring of the substrate. One of the fingerprint regions of the ribokinase family is the glycine-glycine dipeptide sequence (Bork et al., 1993; Wu et al., 1991). A conformational shift of the GG dipeptide upon substrate binding is believed to be the reason for forming the closed conformation of the enzymes (Schumacher et al., 2000). The most highly conserved region for ribokinase superfamily members is the anion hole. It is made up by the main chain nitrogen amide atoms of the GAGD sequence and is located near the γ-phosphate. The anion hole neutralizes the accumulated negative charges during the phosphate group transfer (Sigrell et al., 1998). The aspartic residue in the anion hole is important for the reaction mechanism, since it forms a hydrogen bond with the phosphate accepting hydroxyl group and is believed to deprotonate the hydroxyl group as a first step in catalysis (Mathews et al., 1998; Sigrell et al., 1998). The only enzymes in the ribokinase superfamily that do not contain the aspartate residue in the anion hole are HMPP kinase and THZ kinase, which instead have a cysteine as a catalytic base (Campobasso et al., 2000; Cheng et al., 2002). A mutation of the cysteine to aspartate in THZ kinase actually increases the enzymatic activity 9-fold (Campobasso et al., 2000). Many family members do also contain a non-conserved arginine or lysine that is assumed to stabilize the transition state during the phosphate transfer.

1.3.5 Catalytic reaction

The reaction catalyzed by the ribokinase superfamily members is believed to be similar to the proposed reaction mechanism of *E. coli* ribokinase (Sigrell et al., 1998). The first step is the deprotonation of the methyl hydroxyl group in the substrate by the aspartate in the anion hole. Subsequently, the negatively charged hydroxyl group

Table 1. Members of the ribokinase (RK) family, additional members of the RK superfamily as well as *S. typhimurium* AIR kinase and *M. jannaschii* nucleoside kinase. For enzymes on grey background there are no published data but their structures are deposited in PDB.

Enzyme	PDB ID	Pfam[a]	SCOP superfamily[b]	Res./ Mon.[c]	Lid[d]	Mon./ Enz.[e]
human adenosine kinase	1BX4	PfkB/RK	RK	345	+	1
T. gondii adenosine kinase	1DGM	PfkB/RK	RK	363	+	1
E. coli ribokinase	1RKD	PfkB/RK	RK	309	+	2
sheep pyridoxal kinase	1LHP	PfkB/RK	RK	312	-	2
B. subtilis THZ kinase	1EKQ	HK	RK	272	-	3
T. thermophilus KDG kinase	1V1A	PfkB/RK	RK	309	+	6
E. coli pyridoxamine kinase	1VI9	PfkB/RK	RK	299	-	1
T. maritima PfkB kinase	1VK4	PfkB/RK	RK	298	+	1
T. maritima 1-Pfk	1O14	PfkB/RK	RK	331	+	2
T. maritima KDG kinase	1J5V	PfkB/RK	RK	351	+	2
T. thermophilus ThiD	1UB0	PfkB/RK	RK	258	-	2
B. subtilis YXKO kinase	1KYH	Carb. kinase	RK	276	-	4
S. typhimurium HMPP kinase	1JXH	-	RK	288	-	2
P. horikoshii THZ kinase	1V8A	HK	RK	265	-	3
T. litoralis glucokinase	1GC5	ADP-PFK/GK	RK	467	+	1
P. furiosus glucokinase	1UA4	ADP-PFK/GK	RK	455	+	1
P. horikoshii glucokinase	1L2L	ADP-PFK/GK	RK	457	+	1
S. typhimurium AIR kinase	1TZ6	hom. PfkB/RK[f]	hom. RK[f]	339	+	2
M. jannaschii nucleoside kinase	-	hom. PfkB/RK[f]	hom. RK[f]	302	+	2

[a] Protein family: PfkB/RK=ribokinase family: PF00294; Carb. kinase=carbohydrate family: PF01256; HK=hydroxyethylthiazole kinase: PF02110

[b] RK: SCOP 53613

[c] The number of residues per monomer

[d] +, presence of lid domain; -, absence of lid domain

[e] The number of monomers per biological enzyme or in the crystal structure if the biological function is not known.

[f] Homologous to the Pfam or SCOP superfamily.

makes a nucleophilic attack on the γ-phosphate of ATP through a transition state where the γ-phosphate binds pentacovalently to the five oxygen atoms. The magnesium ion is expected to be involved in orienting the γ-phosphate as well as in stabilizing the phosphate transfer (Mathews et al., 1998). The transition state is stabilized by the main chain nitrogens of the anion hole and the positively charged arginine or lysine. Finally, the phosphorylated substrate and ADP are formed and released from the enzyme.

1.3.6 Lid domain

Members of the ribokinase superfamily can generally be divided into two groups depending on the presence or absence of an additional lid domain. The first group has no lid and comprises only the α/β domain. These enzymes oligomerize in some way to create an active site at the interface of the subunits. To this group belong e.g., *B. subtilis* THZ kinase (Campobasso et al., 2000), sheep pyridoxal kinase (Li et al., 2002) and *E. coli* pyridoxamine kinase (Safo et al., 2004). Enzymes of the second group contain in addition to the α/β domain also a smaller protruding lid domain. This

group includes e.g., *E. coli* ribokinase (Sigrell et al., 1998; Sigrell et al., 1999), human and *T. gondii* adenosine kinase (Mathews et al., 1998; Schumacher et al., 2000), the three archaeal glucokinases (Ito et al., 2003; Ito et al., 2001; Tsuge et al., 2002) and *T. thermophilus* KDG kinase (Ohshima et al., 2004). Most of the enzymes comprising the lid domain undergo a conformational change upon substrate binding. The conformational change results in a movement of the lid domain that brings it closer to the α/β domain and is thereby covering the active site and shielding the substrate from the solvent.

The fold of the lid domain differs among the enzymes. *E. coli* ribokinase and the hexameric KDG kinase have lid domains containing 4 β-strands and connecting loops. In ribokinase, the lid domains of two subunits form a β-clasp, which builds up the dimeric interface. The two adenosine kinases and the three glucokinases are all monomers and have more complex lid domains. Both human and *Toxoplasma gondii* adenosine kinase comprise a lid domain with five β-strands and two α-helices. The lid domains of *T. litoralis* and *P. horikoshii* glucokinases are made up by four β-strands and four α-helices. *P. furiosus* glucokinase comprise an even more complex lid domain with five large and two small β-strands and four α-helices. The lid domain has been suggested to be a morphological marker of evolution within the ribokinase superfamily (Cheng et al., 2002; Zhang et al., 2004). Members of the ribokinase superfamily that lack the lid domain are believed to precede those comprising the lid in evolution. It seems that the lid domain since then has become structurally more advanced. A comparison of the secondary structure of the enzymes belonging to the ribokinase superfamily (including *S. typhimurium* AIR kinase) in 2004 resulted in the following evolutionary order (Zhang et al., 2004): 1) *B. subtilis* THZ kinase (Campobasso et al., 2000), 2) *S. typhimurium* HMPP kinase (Cheng et al., 2002), 3) sheep pyridoxal kinase (Li et al., 2002), 4) *S. typhimurium* AIR kinase and *T. thermophilus* KDG kinase (Ohshima et al., 2004; Zhang et al., 2004), 5) *E. coli* ribokinase (Sigrell et al., 1998; Sigrell et al., 1999), 6) human and *T. gondii* adenosine kinases (Mathews et al., 1998; Schumacher et al., 2000), and 7) *T. litoralis* glucokinase (Ito et al., 2001).

1.3.7 Differences in quaternary structures

Despite the similar overall fold of the ribokinase superfamily members, the enzymes show significant variation in their quaternary structures. For instance, human and *T. gondii* adenosine kinases are monomeric (Mathews et al., 1998; Schumacher et al., 2000), ribokinase and sheep pyridoxal kinase are dimeric (Kwok et al., 1987; Sigrell et al., 1997), *B. subtilis* THZ kinase is trimeric (Campobasso et al., 2000), *B. subtilis* YXKO kinase is tetrameric (Zhang et al., 2002), and *T. thermophilus* KDG kinase is hexameric (Inagaki et al., 2004).

2 AIM OF THE PROJECT

This licentiate thesis describes the structural determinants for thermal stability in enzymes and gives an overview of the background and present knowledge of the ribokinase family and ribokinase superfamily. In order to better understand the evolution of the ribokinase family, as well as of carbohydrate kinases in general, additional structures of enzymes belonging to this family are required. In this project, the three-dimensional structure of a hyperthermostable archaeal member of the ribokinase family was determined. This structure will provide additional information on aspects of thermal stability, catalytic mechanism and evolution.

In 1996, when Bult and coworkers presented the complete genome of *Methanocaldococcus jannaschii*, they described the gene product of MJ0406 as a putative ribokinase and a possible PfkB kinase (Bult et al., 1996). In this project, cloning and expression of the gene MJ0406 and purification of the gene product were developed and performed by Dr. Thomas Hansen in the group of Prof. Dr. Peter Schönheit at the Christian-Albrecht University in Kiel, Germany. Characterization studies were performed by Dr. Thomas Hansen.

The aim of my part of this project was to:

- crystallize nucleoside kinase as apo-enzyme as well as in complex with substrates (Paper I).
- determine the three-dimensional atomic structure of the enzyme, with and without bound substrates and co-factor, by X-ray crystallography (Paper III).
- analyse the stability and the kinetics of unfolding with differential scanning calorimetry (Paper II and incomplete results).
- evaluate and interprete the atomic structure regarding conserved regions, enzyme mechanism and thermal stability (Paper III).

3 METHODS USED IN THE PRESENT STUDY

3.1 STRUCTURE DETERMINATION BY X-RAY CRYSTALLOGRAPHY

X-ray crystallography is a common and widely used technique to determine the three-dimensional atomic structure of macromolecules. There are also other experimental methods available, such as nuclear magnetic resonance (NMR) spectroscopy and electron crystallography. In addition, homology modelling is a theoretical method, based on molecular dynamics calculations, also used for three-dimensional structure determination. However, the resulting structures tend to be less accurate than those obtained by experimental methods (Berman et al., 2000). All methods have advantages as well as disadvantages, but should be seen rather more complementary than competitive.

X-ray crystallography can provide high-resolution atomic structures of large macromolecules, which is not (yet) possible with NMR or electron crystallography. NMR is performed on a protein in aqueous solution, which allows the molecules to tumble and vibrate. In order to obtain resonances sharp enough for adequate resolution, the protein molecule must tumble rapidly. This limits the size of the molecule to about 30 kDa, however increased magnetic fields and new techniques have revealed structures of much larger proteins (Fiaux et al., 2002; Riek et al., 2002). One great advantageous with NMR is that this method measures fluctuations and the mobility of the molecule (Kay, 2005). The result of an NMR analysis is an ensemble of alternative models, in contrast to the unique, however time-averaged, model obtained by X-ray crystallography. Electron crystallography was originally developed for structure determination of two-dimensional crystals of membrane proteins (Unwin & Henderson, 1975). These structures could only be determined to low or medium resolution. Today, protein structures have been determined to atomic resolution by electron crystallography, e.g., bacteriorhodopsin (Henderson et al., 1990), the light harvesting complex (Kühlbrandt et al., 1994) and tubulin (Nogales et al., 1998). However, the lacking data in the third dimension is a severe obstacle for electron crystallography. Nevertheless, electron crystallography is a method under intense development and structure biologists will probably see a progress of this technique in the future.

Normally, one would believe that a structure obtained from a molecule in aqueous solution would be more reliable and accurate than a crystal structure. This is an argument frequently used by NMR spectroscopists. However, the protein crystals used for X-ray crystallography experiments are highly hydrated and contain large amounts of water. Besides, experiments performed with NMR spectroscopy require protein concentrations of 50-100 mg ml^{-1}, which are similar to that of a crystal. Proteins studied both by crystallography and NMR have revealed three-dimensional structures with good agreement in the overall fold. One example is *E. coli* thioredoxin with PDB-IDs 3TRX for the NMR structure and 1SRX or 1TDE for the crystal structure (Forman-Kay et al., 1991; Holmgren et al., 1975; Waksman et al., 1994).

The process of determining a protein structure by X-ray crystallography can be quite straightforward but is in many cases time consuming. The structure determination process involves several steps (Fig. 3) which will be explained in more detail below. Firstly, the protein must be crystallized to provide single three-dimensional well-ordered crystals. This is a critical stage, since the attempts to produce suitable crystals often fail. The next step is data collection of the crystal. The incident X-ray beam is diffracted by the electrons of the molecules in the crystal. During a data collection experiment, the crystal is exposed to X-ray radiation while it is rotated by a small angle increment, and the X-ray diffraction pattern is recorded by a detector. This is repeated to cover a certain rotation angle segment, depending on the crystal symmetry. A diffraction pattern contains thousands of spots, or reflections. X-rays can be described by waves, characterized by wavelength, amplitude and a phase angle. The amplitude can directly be obtained from the intensity of the reflection. The phase has to be determined by any or combinations of the following phasing methods: isomorphous replacement, molecular replacement, and/or anomalous dispersion. Solving the 'phase problem' is the second critical stage in the protein structure determination process. If both the amplitudes and phase angles for each X-ray reflection, i.e., for each wave, are known, a three-dimensional image of the electron density distribution of the molecule – an electron density map – can be obtained. The electron density map is interpreted and an initial structure model of the protein is built into the electron density using computer graphics programs. The phase angles often need to be improved to get a clearer and more reliable electron density map. The initial model building is followed by cyclic refinement by which the difference between the observed and calculated structure is minimized. After each

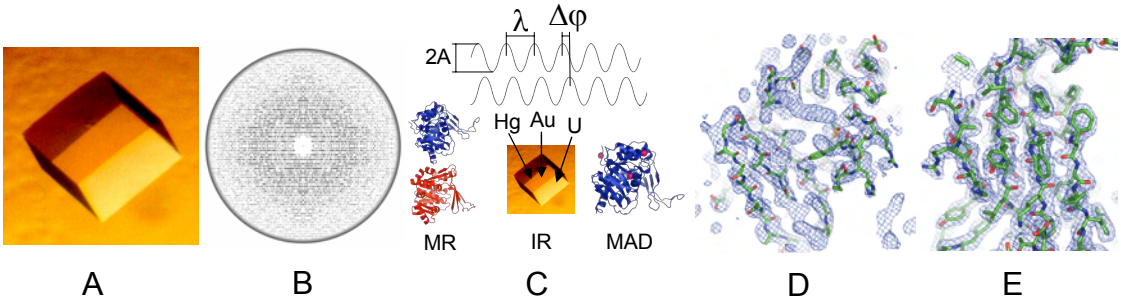

Figure 3. Structure determination of a macromolecule by X-ray crystallography, here illustrated for *Methanocaldococcus jannaschii* nucleoside kinase. Structure determination starts with crystallization of the enzyme. (A) Crystal obtained for nucleoside kinase co-crystallized with an ATP-analogue. The crystallization is followed by data collection and evaluation, here symbolized by a pseudo-precession image of the crystal data (B). (C) X-rays can be described by waves containing a wavelength (λ), amplitude (A) and phase (φ). The phase difference Δφ between the two waves in C is π/2. One critical step in the process of solving a structure is phase determination. Attempts were made to determine phases by molecular replacement (MR) and isomorphous replacement (IR), but were unsuccessful. The phases were eventually determined by multiple-wavelength anomalous dispersion (MAD) using a seleno-methionine derivative of nucleoside kinase. (D) Initial electron density map and the automatically built model obtained from RESOLVE (Terwilliger, 2003). After several cycles of model building and refinement, the electron density map was improved considerably. (E) Final model and electron density of nucleoside kinase.

refinement cycle and finally when the refinement procedure has converged to give the most reliable structure, the quality of the model is checked and validated.

3.1.1 Crystallization

Proteins and other biological macromolecules can be crystallized by a variety of techniques using different crystallization conditions (McPherson, 2004). The hanging drop and sitting drop vapour diffusion techniques are used most frequently, but other crystallization techniques are also possible, such as crystallization by dialysis via a membrane or a gel, liquid-liquid diffusion and crystallization under oil.

A general rule is that the purer the protein sample, the better are the chances to get good crystals. The protein is crystallized from an aqueous solution, called mother liquor, often containing some type of *salt* (e.g., $MgCl_2$), a *precipitant*, which can be a polymer (e.g., PEG 4000), an organic solvent (e.g., 2-propanol) or a particular salt at high concentration, and a *buffer* to keep the pH (e.g., Tris-HCl pH 8.0), and sometimes also a detergent (e.g., β-OG) or a reducing agent (e.g., DTT).

Crystals can only grow from a protein solution that has reached the solubility of the protein and thus is supersaturated. If the supersaturation is too high, the protein will precipitate and if the protein solution is undersaturated no crystal nucleation will occur. There are several approaches to create supersaturation from a crystallization condition that does not reach nucleation, such as by concentrating the protein sample, increasing the precipitant concentration, increasing/decreasing the concentration of salt, altering the pH as well as the temperature, or by adding a ligand (McPherson, 1999).

It is often useful to start the crystallization trials by using *random screens*, where the concentrations and type of the regents as well as pH are varied arbitrarily, or commercially available *crystallization screens*, which include certain conditions that have been proven most successful for macromolecules. Once nucleation is observed, the particular crystallization condition is optimized using a more systematic search with *grid screens*, where only two parameters (e.g., pH and precipitate concentration) are varied at a time in order to find the optimal condition for crystal nucleation and growth. Normally, protein crystals grow to their final size within two weeks, but in some cases it can take as long as half a year for nucleation.

3.1.2 Description of a protein crystal

Protein crystals are small, usually with edges shorter than 1 mm, compared to salt crystals, such as NaCl, which can have dimensions of several centimeters. Protein crystals have a high solvent content of normally 40-60 %, but it can be in the range of 30-90 % (McPherson, 1999). As a consequence, protein crystals are soft and sensitive to drying and are more fragile compared to salt crystals. Owing to the high solvent content, small molecules like substrates, cofactors, inhibitors, and heavy atoms can be diffused, or soaked, into protein crystals.

A crystal is a precisely ordered, three-dimensional array of objects. In biological crystals the objects are molecules, such as an enzyme (Fig. 4) (McPherson, 1999; McPherson, 2003). The smallest repeating unit of a crystal is the asymmetric unit. The asymmetric unit contains one or more protein monomers. A set of symmetry operators, termed the space group operators, applied to the asymmetric unit generates a set of identical asymmetric units that is building up a whole unit cell. The unit cell with its content is periodically repeated by translational symmetry along the unit cell axes forming the crystal lattice.

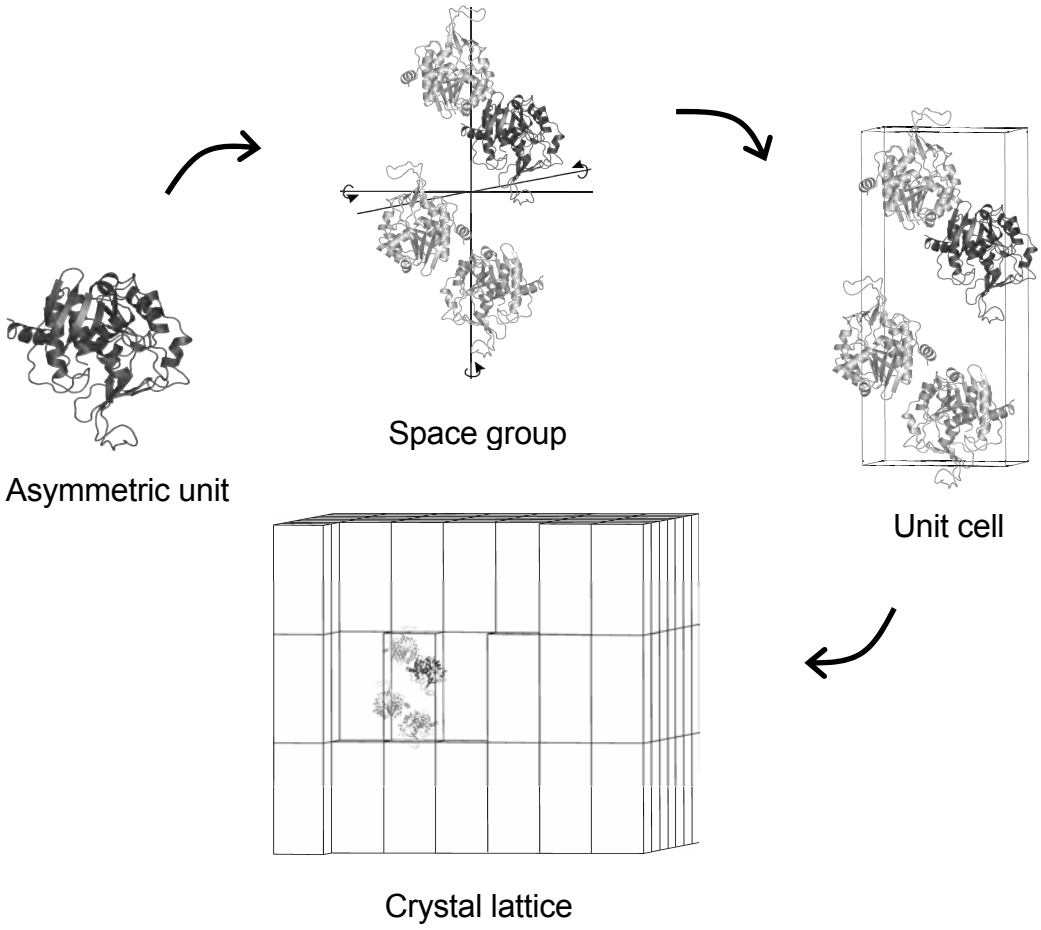

Figure 4. A schematic picture of an apo-enzyme crystal of *Methanocaldococcus jannaschii* nucleoside kinase. A crystal is created from the asymmetric unit by application of a set of symmetry operators called the space group operators (here, represented by $P2_12_12$). A small set of packed identical asymmetric units is obtained, which comprise the unit cell. Multiple more or less ordered unit cells form the crystal.

3.1.3 Data collection and processing

X-ray diffraction data can be collected either at an in-house X-ray generator equipped with a rotating anode that produces X-rays with a fixed wavelength, usually Cu Kα radiation with λ=1.542 Å, or at a synchrotron where the wavelength of the X-ray radiation is near 1 Å and highly intense and focused by mirror systems (Drenth, 1999). The characteristics of the synchrotron radiation result in smaller but more brilliant diffraction spots, which allow a larger number of reflections to be collected for a certain detector area. Therefore, diffraction data collected with synchrotron radiation result in higher resolution for a certain crystal compared to data collected at an in-house X-ray apparatus (Drenth, 1999). The resolution is an important parameter of the X-ray diffraction data. (Dauter, 1999). For instance, 4 Å is rather poor resolution for an enzyme structure, 2.5 Å is medium resolution and resolves atoms with great certainty, and 1.0 Å is high atomic resolution for a protein structure and allows observing even electron density for hydrogen atoms. In addition, the possibility of adjusting the wavelength at the synchrotron is required for anomalous dispersion experiments.

It is nowadays common practice to perform data collection under *cryo-conditions* (cold conditions) where the crystal is soaked for a short while in a solution containing a cryo-protectant, e.g., glycerol, ethylene glycol or a low molecular PEG, and subsequently flash-frozen in a stream of nitrogen holding a temperature of 100 K. The cryo-protectant solution has often a similar composition as the mother liquor. With the proper amount of cryo-protectant, the crystal will be vitrified, i.e., it assumes a "glassy" state which protects and conserves the crystal lattice. Working under cryo-conditions decreases the rate of radiation damage to the crystal. Thus, from one single crystal several data sets can be collected. This is highly advantageous in order to keep the isomorphism when collecting multiple-wavelength anomalous dispersion data.

The most common procedure for collecting diffraction data for proteins is the *rotation method* (Arndt et al., 1968; Arndt & Wonacott, 1977), where the crystal is rotated about a single axis, which is perpendicular to the direction of the X-ray beam. The diffraction pattern is in reciprocal relation to the real crystal lattice, which means that the longer the unit-cell dimension in the crystal, the shorter becomes the distance between the resulting diffraction spots. Thus, to get well separated reflections it is preferable to rotate around the longest crystal unit-cell dimension.

Certain data collection parameters must the optimized to obtain data with high quality and high completeness. The *crystal-to-detector-distance* must be adjusted to a position as close to the crystal as possible in order to capture the high resolution spots, but still not as close that the reflections cannot longer be separated. The *exposure time* (or *exposure frequency*) of the crystal needs to be long (high) enough to detect the weaker reflections at high resolution. However, a too long exposure time can result in too strong diffraction spots ("overloads") and increases the risk of radiation damage of the crystal. The exposure time is as well limited by the detector. Image-plate detectors are most frequently used at an in-house X-ray generator and require exposure times in the range of 4-8 minutes. The exposure time can be reduced to less than 30 seconds by using a charge-coupled device (CCD), which is typically used together with synchrotron radiation. The *oscillation angle* is the degree about which the crystal is

rotated during X-ray exposure. It is normally set to 1°, but if the diffraction spots still overlap, it needs to be smaller. The *rotation range* is the total angle segment of diffraction data collected and its size is dependent on the symmetry within the crystal. However, if the crystal is geometrically optimized in the X-ray beam, a rotation range of 180° will be sufficient to collect a complete data set for all macromolecular crystal symmetries. The completeness represents the number of measured reflections compared to the total number of reflections present at the specified resolution. The completeness should be near 100 % and is one of the most important characteristics of the X-ray data (Dauter, 1999). However, it is preferable to measure the intensities of the reflections more than once, i.e., to increase the *redundancy*, since this will improve the accuracy of the data.

After data collection, the data are processed through indexing and scaling using computer programs such as *DENZO* and *SCALEPACK* in the *HKL* package (Otwinowski & Minor, 1997). The data quality is usually judged by R_{merge}, based on the intensities (or rather $|F|^2$), as well as the ratio of the intensities to their uncertainties (I/σ). The resolution limit is often taken where I/σ becomes lower than about 2.0 and R_{merge} reaches values of 20-40 % (Dauter, 1999), however there are suggestions to always use all the data especially if maximum likelihood methods are used for refinement.

3.1.4 Patterson methods

The Patterson function, $P(uvw)$, (Patterson, 1935) is a Fourier transform of the intensities of the diffraction. Thus, it includes no phase angles and as a result, the Patterson function can be calculated without any previous knowledge of the structure. The Patterson cell has identical dimensions to the crystal unit cell. The Patterson map is a vector map and the coordinates of the peaks in a Patterson map correspond to all vectors between each atom pair in the unit cell. Two atoms with coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2) will result in two peaks in Patterson space with coordinates $(u_1, v_1, w_1 = x_2 - x_1, y_2 - y_1, z_2 - z_1)$ and $(u_2, v_2, w_2 = x_1 - x_2, y_1 - y_2, z_1 - z_2)$. Thus, for a crystal structure containing N non-hydrogen atoms the Patterson map will contain N^2 peaks. N of these peaks result in coordinates $(u, v, w = 0,0,0)$ since each atom produces a vector to itself with zero length. The Patterson origin peak is hence the most intense peak in the Patterson map. The remaining $N^2 - N = N(N-1)$ peaks are non-origin peaks. The Patterson map is always centrosymmetric since the vectors between atoms point in both directions. Moreover, the Patterson map contains the same rotational symmetries as real space. Symmetry elements can cause a concentration of peaks in certain planes, so-called "Harker planes" (Harker, 1936).

Correlations between a rotated Patterson function and the original Patterson function calculated from a data set are analyzed by the *self-rotation function* (Rossmann & Blow, 1962). This function is used to identify the rotational non-crystallographic symmetry between identical objects in the asymmetric unit.

By evaluating the Harker sections in a *native Patterson map*, the existence of pseudo-symmetries in the crystal can be established. A peak higher than 15 % of the origin peak in the native Patterson map might indicate that there is pseudo-translation

present in the unit cell, as is the case when a non-crystallographic symmetry axis is parallel to a crystallographic axis.

The Patterson function and its Harker sections are very useful in the process of locating a limited number of heavy atoms in a crystal made for phasing by isomorphous replacement (isomorphous difference Patterson maps) or the selenium atoms in a SeMet-derivative used in the anomalous dispersion technique (anomalous difference Patterson maps).

3.1.5 The phase problem

X-rays can be described by waves characterized by amplitude and a phase angle. The individual reflections in the diffraction pattern are the interference sum of the waves scattered by all of the atoms in the crystal in one particular direction (McPherson, 2003). Each reflection can be described by a structure factor, $\vec{F}(hkl)$, that contains information about the amplitude, $|F(hkl)|$ and phase angle, φ_{hkl}. Thus, the scattered wave can be described by using complex numbers:

$$\vec{F}(hkl) = |F(hkl)| \exp i\varphi_{hkl}, \qquad (1)$$

where *h, k,* and *l* are integer indices characterizing the diffracted X-ray beam. To be able to determine the protein structure, it is necessary to have information about both the amplitude and the phase angle for each scattered wave, i.e., for each individual reflection (*h, k, l*). The amplitude can be obtained from the intensity $I(hkl)$ of the reflection from the following relation:

$$I_{hkl} = \vec{F}_{hkl} \cdot \vec{F}_{-h-k-l} = |F_{hkl}| \exp(i\varphi_{hkl}) \cdot |F_{-h-k-l}| \exp(-i\varphi_{hkl}) = |F_{hkl}|^2. \qquad (2)$$

Thus, $|F(hkl)| = c \cdot \sqrt{I(hkl)}$ (c is a constant). However, as can be seen from equation 2, it is not possible to directly measure the phase angles φ_{hkl} of the structure factors from the recorded diffraction pattern. This is known as "the phase problem" of X-ray crystallography. In order to obtain the three-dimensional structure the phase problem needs to be solved. The phases can be obtained by different phasing methods. The ones described here are used for protein crystallography and embrace isomorphous replacement, molecular replacement, anomalous dispersion, and to some extent direct methods.

3.1.5.1 Isomorphous replacement

Isomorphous replacement is a phasing technique which is based on the intense scattering power of heavy atoms, such as gold, mercury, and uran, compared to the commonly occurring atoms in proteins (C, N, O, S, P) (Kendrew & Perutz, 1957). In the isomorphous replacement method, the protein crystal is soaked or co-crystallized in a solution containing a heavy atom compound, e.g., $HgCl_2$. It is important that the heavy-atom derivative is isomorphous, i.e., that it has the same unit cell and same orientation of the protein in the crystal, compared to the native protein crystal. Data sets for both native crystals and at least one heavy-atom derivative are collected. The method is specified as single isomorphous replacement (SIR) if only one heavy-atom

derivative is used. In the case of several heavy-atom derivatives the term multiple isomorphous replacement (MIR) is used.

The changes in intensity and thus in amplitude of a reflection of the native crystal, $|F_P|$, and for the derivative crystal, $|F_{PH}|$, are measurable. The isomorphous difference, $|F_H| \cong |F_{PH}| - |F_P|$, is used to locate the heavy-atom positions using Patterson or direct methods. For SIR, two phase angles are possible. The phase problem can anyhow be solved, sometimes without additional data or in combination with anomalous dispersion methods. The phase ambiguity with SIR is avoided in MIR experiments.

3.1.5.2 Molecular Replacement

Molecular replacement (Rossmann & Blow, 1962) can be used as a phasing technique to determine a structure when a homologous model already is available. According to the accumulated experience, the chances to determine the phases by molecular replacement becomes better by using a search model with a sequence identity higher than 25 % compared to the unknown model and an r.m.s. deviation smaller than 2.0 Å between the Cα atoms of the homology model and the final new structure (Taylor, 2003). To avoid model bias, non-conserved amino acids are usually mutated to alanines or, in the case of prolines, to glycines. The search process is often simplified if large loop regions or non-conserved regions are deleted from the search model.

By comparing the Patterson function of the homology model with the Patterson function calculated from the data, the homology model is first rotated and subsequently translated to a position $\mathbf{x}_{new}$ of the unit cell given by the crystal data. This can be explained mathematically by

$$\mathbf{x}_{new} = \mathbf{R} \cdot \mathbf{x}_{old} + \mathbf{t}, \tag{3}$$

where $\mathbf{x}_{old}$ is the starting model. $\mathbf{R}$ is the rotation matrix which rotates the molecule into the new orientation and $\mathbf{t}$ is the translation vector required to place the oriented model into the position in the unit cell. A good correlation between the homology model and the data is described by a high correlation coefficient, which is calculated in different ways for various programs. Molecular replacement can be performed with the programs AMoRe (Navaza, 1994) and MolRep (Vagin & Teplyakov, 1997) as implemented in the CCP4 suite (Collaborative Computational Project, 1994).

3.1.5.3 Anomalous dispersion

Anomalous dispersion techniques have become the method of choice for phase determination. Anomalous scattering occurs at the absorption edge when the incident photon of the X-ray beam has sufficient energy to cause an expulsion of an inner shell electron that results in a re-emitted photon (fluorescence) with a somewhat reduced wavelength and altered by 90° in phase (McPherson, 2003; Taylor, 2003). All atom types give rise to anomalous scattering, but the effect is negligible for light atoms with weakly bound electrons, such as most protein atoms (C, N, O). The chemical elements used for phase determination using anomalous dispersion may be selenium in a seleniomethionine containing recombinant protein, iron in heme-containing proteins or other prosthetic metal atoms such as zink and copper. It can also be any other

anomalous scatterer, such as uranium or mercury in a heavy-atom derivative. The anomalous signal is varied by altering the wavelength of the X-rays. The anomalous scattering in combination with the 'conventional' scattering produces a breakdown in Friedel's law, i.e., the intensities $I(hkl)$ and $I(-h-k-l)$ are no longer equal. The difference between the amplitudes $|F(hkl)|$ and $|F(-h-k-l)|$ is the anomalous difference and it can be used to locate the anomalous scatterers by using Patterson or direct methods in the same way as the isomorphous difference.

For multiple-wavelength anomalous dispersion (MAD) experiments, data are collected at several, typically three, wavelengths. As a consequence, MAD experiments must be performed at a tuneable synchrotron radiation source. It has now been increasingly possible to solve the phase problem with the single-wavelength anomalous dispersion (SAD) method, where data are collected at only one wavelength, preferably at the absorption peak, and the phase ambiguity is broken using density modification (Dodson, 2003). Anomalous dispersion can be combined with SIR experiments, leading to SIRAS (single isomorphous replacement with anomalous scattering), to resolve the phase ambiguity, as well as with MIR data, termed MIRAS (multiple isomorphous replacement with anomalous scattering), in order to improve phases. One great advantage with the anomalous dispersion technique is that it usually preserves the isomorphism since all data can be collected from one single frozen crystal.

3.1.5.4 Direct methods

Direct methods, or *ab initio* methods, can usually only be applied on molecules containing less than 100 non-hydrogen atoms and require accurate X-ray diffraction data that extend to atomic resolution, i.e., 1.0 Å resolution or better (McPherson, 2003; Taylor, 2003). Diffraction data from larger proteins at this resolution are only rarely available. However, direct methods are used routinely to find heavy atom substructures in isomorphous replacement or constellations of anomalous scatterers in anomalous dispersion experiments with programs such as Shake-and-Bake (Miller et al., 1994), SHELXD (Schneider & Sheldrick, 2002) and SHARP (de La Fortelle & Bricogne, 1997).

3.1.6 Phase improvement

Experimentally determined phases are rarely sufficiently accurate to give a completely interpretable electron density map. The experimental phases often need to be improved using a variety of density modification methods. The main techniques used to improve the phases and modify the electron density are solvent flattening (Leslie et al., 1988; Wang, 1985), histogram matching and non-crystallographic averaging (Rossmann & Blow, 1962).

Solvent flattening is a powerful method to refine phases. Owing to the high thermal motion and disorder of the solvent molecules, the electron density in the solvent regions is flat at medium resolution (Rossmann & Arnold, 2001). Flattening of the solvent regions, by removing negative electron density in the solvent regions, will reduce the noise in the map and thereby improve the phases. Accordingly, the protein density will appear more clearly.

Histogram matching varies the electron density values of a map to make it agree with the expected values of an ideal electron density map. The density histogram of a map is the probability distribution of electron density values. The initial electron density map can be improved by adjusting density values in a systematic way to make its histogram match the ideal histogram (Rossmann & Arnold, 2001; Taylor, 2003).

Non-crystallographic symmetry averaging can be used when the asymmetric unit contains more than one copy of the same molecule. In this technique, regions of electron density, which are related by non-crystallographic symmetry, are averaged. The phase improvement obtained by non-crystallographic averaging increases with the number of molecules in the asymmetric unit (Rossmann & Arnold, 2001; Taylor, 2003).

Density modification is a cyclic procedure. It starts with a modification of the electron density map. Subsequently, the modified electron density is back-transformed to yield modified phases, which are combined with the experimental phases. Then it starts all over again and a new electron density map is calculated from the combined phases until convergence. Density modification with all the above techniques can be run using the programs DM (Cowtan & Zhang, 1999) and RESOLVE (Terwilliger, 2000).

3.1.7 Model building and refinement

After the initial phases have been determined and improved, the model is iteratively rebuilt and refined. The protein model is built into the electron density either automatically by various programs, such as RESOLVE (Terwilliger, 2000) and ARP/WARP (Lamzin, 1993; Perrakis, 1997), or manually in e.g., O (Jones et al., 1991). During the refinement process, the atomic positions in the model are improved by minimizing the difference between the calculated structure factors $F_{calc}(hkl)$ derived from the model and the observed structure factors $F_{obs}(hkl)$ for all the reflections (Drenth, 1999). Normally, the *parameters* to be refined for each non-hydrogen atom in the model are three coordinate values (x, y, z), one isotropic temperature B factor[§] parameter or six parameters for an anisotropic B factor, and the occupancy parameter. The *observations* include everything known about the crystal structure prior to refinement, such as unit cell parameters, structure factor amplitudes and experimental phases. By applying *constraints* to a model during the refinement the effective number of parameters is reduced whereas *restraints* effectively increase the number of observations. The larger the ratio of observations to parameters, the better is the quality of the refined model and the more easily the refinement process converges.

3.1.7.1 Least squares and maximum likelihood

The main refinement techniques used in protein X-ray crystallographic refinement are based on the least-squares function or the maximum-likelihood function (Drenth,

§ The temperature B factor, or Debye-Waller factor, represents an atom's vibration around its center. A high B factor indicates either disorder or thermal motion. An isotropic B factor means that the actual atom position is only dependent on the distance to the averaged atom position. An anisotropic B factor is described by an atoms vibration within an ellipsoid centred at the atomic coordinate. Even if all atoms within a crystal move anisotropically, they are refined as isotropic if the crystal diffracts to lower than approximately 2 Å resolution.

1999). Least squares and maximum likelihood are implemented in REFMAC (Murshudov, 1997) and CNS (Brünger et al., 1998).

The *least-squares* refinement method poses the problem 'Given these data, what are the parameters of the model that give the minimum variance of the observations?' (Rossmann & Arnold, 2001). During least-squares refinement, the observations are fixed and the parameters are varied such that $|F_{calc}(hkl)|$ come as close as possible to $|F_{obs}(hkl)|$. The least-squares function is described by

$$Q = \sum_{hkl} w(hkl)\left(|F_{obs}(hkl)| - |F_{calc}(hkl)|\right)^2, \qquad (4)$$

where $w(hkl)$ is the weight given to an observation. The goal in least-squares refinement is to minimize this function. The minimum of Q is obtained by varying the atomic parameters, which determine $|F_{calc}(hkl)|$, in the model. In refinement using least squares, it is assumed that the observed and calculated structure factors have exactly the same phase, which means that the only error is the structure factor amplitude of an observation $|F_{obs}(hkl)|$. If the intensities for some area of the structure are weak or incomplete, the $|F_{calc}(hkl)|$ will be small too. This is the main reason to some of the problems of least-squares refinement using poor starting models. Conventional full-matrix least-squares refinement can only be used for high resolution data of around 1 Å.

A more suitable method for proteins with low or medium resolution data is the *maximum likelihood* refinement. (Rossmann & Arnold, 2001; Tronrud, 2004). Maximum likelihood fronts the problem 'Given this model, what is the probability that the given set of data would be observed?' (Rossmann & Arnold, 2001). The model is adjusted to maximize the probability of the given observations. The likelihood of a model given a set of observations is the product of the probabilities of all of the observations given the model. The likelihood of a model is expressed by

$$L_{total} = \prod_{hkl} p\left[|F_{obs}(hkl)|; F_{calc}(hkl)\right], \qquad (5)$$

where p is the probability. The most reliable set of parameters is obtained if the likelihood L_{total} is maximized, i.e., when the parameters are most consistent with the data.

Unlike least squares, maximum likelihood can accommodate observations with uncertainties and model parameters whose values are expected to have the same uncertainties. In comparison to least-squares, which function is rather simple and is used in approximately the same way in different programs, each implementation of maximum likelihood has its own set of assumptions and depends on how errors are distributed as well as the consequence of these errors (Tronrud, 2004).

3.1.7.2 Specific refinement methods

Refinement by *simulated annealing* involves torsion-angle parameterization in which the atomic coordinates are replaced by torsion angles. This results in a reduction of the total number of parameters and subsequently an increase in the observation-to-parameter ratio. Simulated annealing of a protein starts at a high temperature, e.g., 5000 K, to allow great freedom to the model, followed by a slow-cooling process until the model is trapped in the global energy minimum. Simulated annealing can be performed in CNS (Brünger et al., 1998).

Rigid body refinement is applied to a protein model when its internal structure is known, but when the orientation and location of the atoms need to be refined. During the refinement process, the protein structure is divided into different parts that are treated as rigid bodies and are refined with respect to each other. Thus, this refinement technique does not refine individual atomic parameters (Drenth, 1999). Rigid body refinement is often applied after molecular replacement, where the initial model is treated as one or several rigid bodies.

Non-crystallographic symmetry refinement should be applied when the asymmetric unit of a crystal contains more than one copy of the same molecule. Basically, the number of parameters in the model decreases with the number of identical molecules in the asymmetric unit (Rossmann & Arnold, 2001).

TLS B factor parameterization (Schomaker & Trueblood, 1968) applies anisotropic movement to a group of atoms described with isotropic B factors, but involves much fewer parameters than for a model with anisotropic B factors for each atom. In TLS refinement, the motion of a group of atoms is described by three matrices representing *1)* the translation, i.e., the translational vibration of the group, *2)* the libration (= wobbling) of the group around a fixed point, and *3)* the unison of the translation and libration, which is described by screw motions. One model can be composed by several TLS groups, however, if there are too many small TLS groups the number of parameters will not be sufficiently reduced.

Rigid body refinement, non-crystallographic symmetry refinement and TLS refinement can be applied using the refinement program REFMAC (Murshudov, 1997).

3.1.8 Validation of the protein model

The protein structure model should after each step of refinement be checked and validated in order to avoid model bias and to follow convergence of the process. The quality of a protein model can be judged by the crystallographic reliability factor, R_{cryst}, and the free reliability factor, R_{free} (Brünger, 1992). R_{cryst} is calculated from both calculated and experimentally observed structure factor amplitudes, whereas R_{free} is calculated from a small portion, usually 5-10 %, of only observed amplitudes, which are not used in refinement. R_{free} was introduced to get an unbiased measurement of the quality of a structure since R_{cryst} was often overestimated and reached too low values. Changes to the model should affect R_{cryst} and R_{free} similarly. Generally, the

lower the R-factors, the more reliable is the model. The final R_{cryst} should be close to or below 20 % and R_{free} a few percent higher than R_{cryst} (Drenth, 1999).

The validation programs SFCHECK (Vaguine et al., 1999) and PROCHECK (Laskowski et al., 1993) perform comprehensive analysis of a protein's molecular geometry, which includes bond lengths, bond angles, torsion angles, conformations, and planarity of groups of atoms (e.g., for ring systems). It is also of importance to check the intra- and intermolecular geometry, such as hydrogen bonds, van der Waals contacts and π-π stacking interactions of aromatic side chains. When the protein model is complete and finished it can be deposited to the protein data bank (Berman et al., 2000) to make it accessible to the scientific community.

3.2 STABILITY AND UNFOLDING STUDIES BY DIFFERENTIAL SCANNING CALORIMETRY

Differential scanning calorimetry, DSC, is an experimental technique which is used to measure the heat energy uptake that takes place in a sample during controlled increase/decrease of the temperature. DSC is used to determine thermal transition temperatures or absolute thermodynamic data for thermally induced transitions of various kind e.g., protein unfolding, conformational transitions and ligand binding.

DSC measures the heat capacity of a sample as a function of temperature. The differential scanning calorimeter contains a pair of cells – the sample cell and the reference cell – that are placed in at thermostated chamber. The excess heat capacity, C_p, is obtained by constantly increasing, "scanning", the temperature and measuring

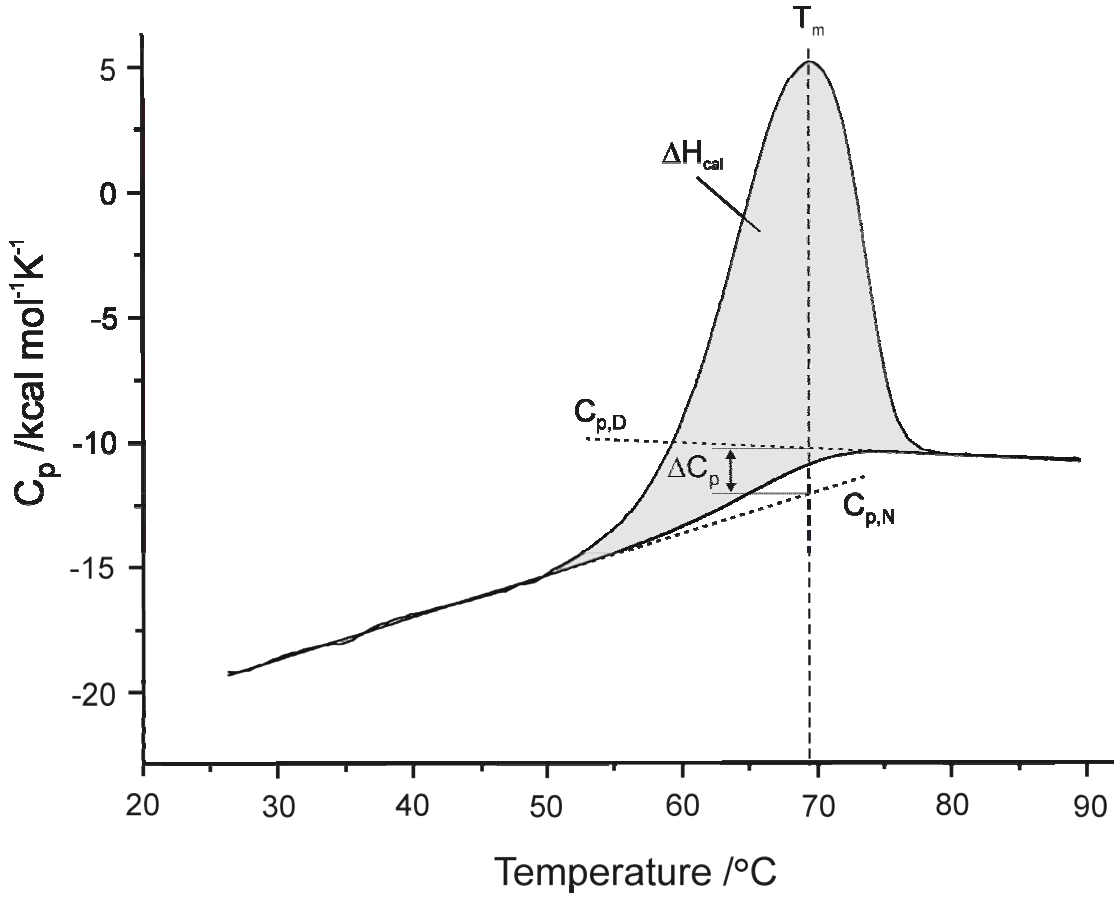

Figure 5. A calorimetric trace from a DSC scan performed on *Methanocaldococcus jannaschii* nucleoside kinase in 10 mM glycine pH 3.0. DSC measures the excess heat capacity, C_p, of a molecule as a function of temperature. The melting temperature, T_m, is the temperature at which the unfolding reaction is half complete and at which C_p has its maximum value. The difference in heat capacity of the native, $C_{p,N}$, and denatured $C_{p,D}$ states, , is given by $\Delta C_p = C_{p,D} - C_{p,N}$. The area under the curve corresponds to the calorimetric enthalpy, ΔH_{cal}. For this experiment an apparent T_m of 70 °C, an apparent ΔC_p of 13 kcal mol^{-1}K^{-1} and an apparent ΔH_{cal} of 142 kcal mol^{-1} was measured.

the difference in heat absorbed in the sample cell compared to the reference cell (Fig. 5). In the beginning of the calorimetric trace the protein is in native state. As the experiment proceeds, the temperature increases and the protein denatures, i.e., looses its functional conformation. The difference in heat capacity, ΔC_p, between the native, $C_{p,N}$, and denatured, $C_{p,D}$, states of the protein molecule is related to variations in the conformational energy. The scan rate in a DSC experiment is typically 1 K min^{-1}. The characteristic peak of the calorimetric trace, or the thermogram, provides the melting temperature, T_m, at which the reaction is half complete. The area under the curve represents the total calorimetric enthalpy change, ΔH_{cal}, for the entire process. By knowing the protein concentration, normally 1-2 mg ml^{-1}, the excess molar heat capacity, $C_{p,mol}$, can be obtained.

Thermodynamic stability of an enzyme is defined by the free energy of stabilization, ΔG_{stab}, which is the difference in free energy between the folded and unfolded states, $G_{unfolded} - G_{folded}$ (Fig. 6) (Vieille & Zeikus, 2001). Kinetic stability depends on the height of the energy barrier between the folded and unfolded states which corresponds to the activation energy of unfolding, E_a (Vieille & Zeikus, 2001). For globular mesostable proteins, ΔG_{stab} is typically between 5 and 15 kcal mol^{-1} at 25 °C (Vieille & Zeikus, 2001). The difference in ΔG_{stab} values between hyperthermostable and mesostable enzymes is often small and usually 5 to 20 kcal mol^{-1} higher for the hyperthermostable enzyme (Vieille & Zeikus, 2001). However, such a small increase as 3.0-6.5 kcal mol^{-1} in ΔG_{stab} can account for an increase of 12 °C in melting temperature (Nicholls et al., 1991). Sometimes, only minor changes are required in the enzyme structure to enhance its thermal stability, such as an introduction of an additional ion-pair interaction optimally placed within an ionic network.

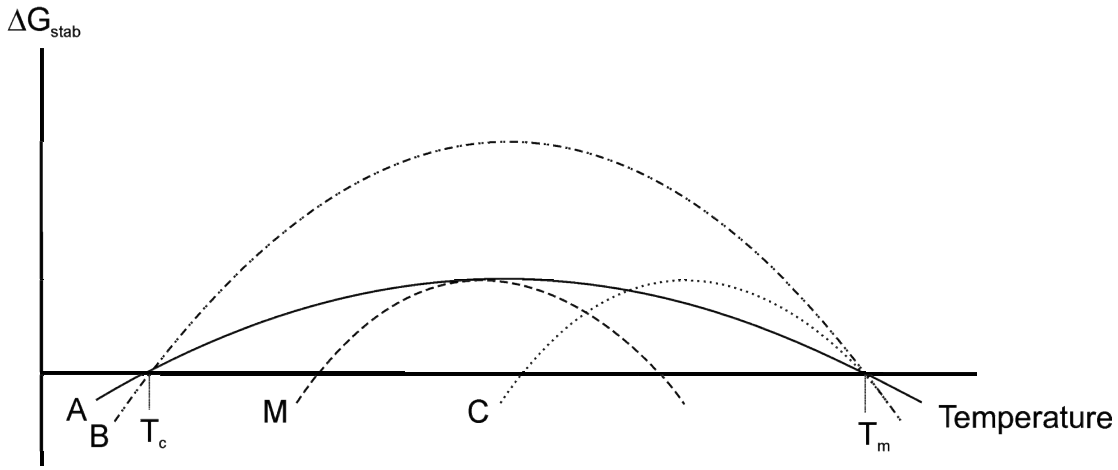

Figure 6. Theoretical free energy-temperature profiles for proteins from a mesophile (M) and three hyperthermophiles (A, B, C). T_c and T_m are for curves A and B. T_c represents the temperature where cold denaturation of the protein occurs. The hyperthermostable proteins represented by curve A and B have the same T_c and T_m, but different values of the ΔG_{stab}. The protein giving curve A has the same maximal ΔG_{stab} as the mesophilic protein, but a decreased T_c and an increased T_m. The mesostable protein and the hyperthermostable protein characterized by curve C show the same appearance of their stability curves and thus the same ΔG_{stab}, however curve C is shifted towards higher temperatures which results in increased T_c and T_m values. Modified from Vieille and Zeikus (2001).

The criteria and thermodynamic/kinetic quantities that can be obtained for a two-state reversible process and a two-state irreversible process will be described and discussed in the two following sections. An unfolding process can of course as well follow a non-two-state process. The elucidation of such data is much more complicated and requires correct unfolding models.

3.2.1 Two-state reversible process

A process is considered reversible if the protein can be completely refolded to its native conformation after denaturation. For a reversible process, a reheating of the sample in a second scanning experiment will reproduce the first thermogram. In a two-state process, the protein is either in native state (N) or in denatured state (D), and intermediate states are not significantly populated at equilibrium (Sturtevant, 1987). A two-state reversible process can be expressed by the chemical equation

$$\mathrm{N} \underset{\leftarrow}{\overset{\mathrm{K_U}}{\rightleftharpoons}} \mathrm{D} \tag{6}$$

where K_U is the equilibrium constant of unfolding. A fundamental goal of DSC experiments is to determine the thermodynamic stability of an enzyme, which corresponds to the free energy change of unfolding, ΔG_U, in a reversible process. For a two-state reversible transition, ΔG_U is the difference in free energy between the native and denatured state. As a function of temperature T, ΔG_U can be expressed by

$$\Delta G_U(T) = \Delta H_U(T) - T\Delta S_U(T) = -RT \ln K_U \tag{7}$$

where $\Delta H_U(T)$ is the enthalpy change of unfolding, $\Delta S_U(T)$ is the entropy change of unfolding, and R is the universal gas constant. The application of equation 7 requires that the protein exists in only two thermodynamically stable states and that the native structure breaks down cooperatively as a whole. As a consequence of the enthalpic and/or entropic stabilizations occurring in a hyperthermostable enzyme, the appearance of the plot of ΔG_U versus T will differ from that of its mesostable counterpart (Fig. 6) (Vieille & Zeikus, 2001). A derivation of equation 7 results in a slope that equals $-\Delta S_U$ as is shown in equation 8.

$$\frac{\partial \Delta G_U(T)}{\partial T} = -\Delta S_U \tag{8}$$

As a consequence, when ΔG_U has its maximum value, i.e., when $\Delta S_U = 0$, the protein is only affected by enthalpic contributions. The curvature of the stability curve is related to the change in heat capacity by

$$\frac{\partial^2 \Delta G_U(T)}{\partial T^2} = -\frac{\Delta C_p}{T}. \tag{9}$$

The temperatures at which the curve intersects with the x-axis, i.e., where $\Delta G_U(T) = 0$, represent the temperature for cold denaturation, T_c, and the melting temperature, T_m.

For a two-state process, the calorimetric enthalpy, ΔH_{cal}, obtained from the thermogram equals the van't Hoff enthalpy, ΔH_{vH} (Naghibi et al., 1995). ΔH_{vH} is defined by

$$\frac{\partial \ln K}{\partial (1/T)} = \frac{\Delta H_{vH}}{R}. \tag{10}$$

The enthalpy is a function of temperature and is expressed by

$$\Delta H_U(T) = \Delta H_U(T_m) + \int_{T_m}^{T} \Delta C_p(T)\,\mathrm{d}T. \tag{11}$$

If ΔC_p is assumed to be constant and not dependent on temperature, one can derive the following expressing:

$$\Delta H_U(T) = \Delta H_U(T_m) + (T - T_m)\Delta C_p. \tag{12}$$

By knowing the $\Delta H_U(T_m)$ from the calorimetric trace, the entropy change during the unfolding transition can be obtained from

$$\Delta S_U(T_m) = \frac{\Delta H_U(T_m)}{T_m}. \tag{13}$$

The entropy as a function of temperature is expressed by

$$\Delta S_U(T) = \frac{\Delta H_U(T_m)}{T_m} + \int_{T_m}^{T} \Delta C_p(T)\,\mathrm{d}\ln T. \tag{14}$$

With the assumption of constant ΔC_p, the expression can be simplified to

$$\Delta S_U(T) = \frac{\Delta H_U(T_m)}{T_m} + \Delta C_p \ln\frac{T}{T_m}. \tag{15}$$

Thus, $\Delta G_U(T)$ can be calculated from the melting temperature, the heat capacity difference and the calorimetric enthalpy using the following modified Gibbs-Helmholtz equation (Becktel & Schellman, 1987):

$$\Delta G_U(T) = \Delta H(T_m) - T\Delta S(T_m) + \Delta C_p\left((T - T_m) - T\ln\frac{T}{T_m}\right). \tag{16}$$

In conclusion, from a two-state reversible process the following thermodynamic quantities can be obtained: T_m, ΔC_p, $C_{p,\max}$, $\Delta H_U(T)$, ΔH_{vH}, $\Delta S_U(T)$, and $\Delta G_U(T)$.

3.2.2 Two-state irreversible process

In many cases, the second thermogram shows no excess heat capacity and it is concluded that the protein was irreversibly unfolded during the first heating scan (Freire et al., 1990). No thermodynamic quantities, except for the thermodynamically insignificant ΔH_{cal}, can be obtained from an irreversible process. Irreversible processes, in comparison to equilibrium reversible processes, are kinetically controlled and are described by rate equations. In an irreversible transition, the molecules in the final state at any temperature depend on the time required to reach that temperature, and consequently, the calorimetric traces are scan rate dependent. Therefore, the scan rate must be included into the analysis of thermograms corresponding to irreversible processes. For a complete analysis of an irreversible transition, several DSC experiments need to be performed within a range of different scan rates.

The simplest model for an irreversible denaturation is the two-state irreversible model. For this model, the calorimetric traces are characterized by two features. Firstly, they are shifted to higher temperatures at increasing scan rates and secondly, they are asymmetrical in shape with a tailing effect at the low temperature side of the transition peak. The protein undergoes a conformational change from the native state (N) to a final unfolded denatured state (D) upon heating. The transition can be described by a first-order kinetic process

$$\mathrm{N} \xrightarrow{\mathrm{k}} \mathrm{D}, \tag{17}$$

where k is the first-order rate constant. A more realistic model of an irreversible thermal denaturation of a protein might be

$$\mathrm{N} \underset{\mathrm{k_{-1}}}{\overset{\mathrm{k_1}}{\rightleftharpoons}} \mathrm{D} \xrightarrow{\mathrm{k_2}} \mathrm{I} \tag{18}$$

where I is a final irreversible state (Lumry & Eyring, 1954). The protein molecules in native and denatured states are in equilibrium. However, most of the protein molecules in the denatured state will be converted to an irreversible state, if all three processes are of first-order and if $k_2 >> k_{-1}$ at any time. Consequently, the concentration of protein in denatured state will always be low and the equilibrium between native and denatured protein cannot be counted for. Thus, the irreversible process from N to I is kinetically controlled by the much slower conversion from N to D, such that

$$\mathrm{N} \xrightarrow{\mathrm{k_1}} \mathrm{I}. \tag{19}$$

This reaction would give the same result as the model N→ D.

The rate constant k as a function of temperature is given by the Arrhenius equation

$$k = A \cdot e^{-\frac{E_A}{RT}}, \tag{20}$$

where A is a frequency factor and E_A the activation energy required for transferring the protein from native state to denatured state. Since the process is irreversible, the concentrations of protein in native state [N] and denatured state [D] are changed over time according to

$$-\frac{\mathrm{d}[N]}{\mathrm{d}t} = \frac{\mathrm{d}[D]}{\mathrm{d}t} = k[N]. \tag{21}$$

Equation 21 can be expressed in mole fractions, $P_N = [N]/C_{tot}$ and $P_U = [U]/C_{tot}$, where C_{tot} is the total protein concentration. By introducing the scan rate $\alpha = \mathrm{d}T/\mathrm{d}t$ the following scan rate dependent equation is obtained:

$$-\frac{\mathrm{d}P_N}{\mathrm{d}T} = \left(\frac{A}{\alpha} \cdot e^{-\frac{E_A}{RT}}\right) P_N . \tag{22}$$

After integration of both sides of equation 22, the temperature dependence of [N] and [D] can be expressed by

$$P_N = e^{-F(T)/\alpha} \qquad P_U = 1 - e^{-F(T)/\alpha}, \tag{23}$$

where $F(T)$ is the following function of temperature:

$$F(T) = \int_{T_0}^{T} k \cdot \mathrm{d}T = \int_{T_0}^{T} A \cdot e^{-\frac{E_A}{RT}} \cdot \mathrm{d}T . \tag{24}$$

The excess apparent molar heat capacity is related to $-\mathrm{d}P_N/\mathrm{d}T$ (Freire et al., 1990) by

$$C_p = -\Delta H_{cal} (\mathrm{d}P_N/\mathrm{d}T). \tag{25}$$

As mentioned before, the only thermodynamic quantity that can be obtained from a two-state irreversible process is the apparent ΔH_{cal}. However, the kinetic parameters – rate constants, activation energy, and frequency factor – can be revealed from the expressions developed by Sanchez-Ruiz and collaborators (Sanchez-Ruiz et al., 1988). They presented four expressions to independently derive the activation energy. The simplest expression is the Arrhenius plot, $\ln k$ versus $1/T$, where the inclination corresponds to $-E_A / R$.

4 RESULTS AND DISCUSSION

4.1 BIOCHEMICAL PROPERTIES OF *METHANOCALDOCOCCUS JANNASCHII* NUCLEOSIDE KINASE (PAPER II AND III)

Nucleoside kinase (MjNK) from the hyperthermophilic archaeon *Methanocaldococcus jannaschii* is optimally active at 85 °C, the same temperature as the optimal growth temperature of *M. jannaschii*. The enzyme is active as a homodimer with a subunit size of 34 kDa. In the presence of ATP and Mg^{2+}, MjNK is able to phosphorylate a wide range of substrates and preferably nucleosides (Paper III, Scheme 1). It has the highest catalytic activity for inosine, cytidine and guanosine, but phosphorylates to some extent also adenosine, thymidine, uridine, xanthosine, and ribose (Paper II, Table 2). All nucleosides (and nucleotides) comprise a ribose sugar group which in the 1'-position is linked to the corresponding base. Nucleoside kinase catalyzes the phosphorylation of the 5'-hydroxyl group of ribose to yield the nucleotide nucleoside-5'-phosphate. Nucleoside kinase is homologous to members of the ribokinase family as well as the ribokinase superfamily.

4.2 CRYSTALLIZATION AND PRELIMINARY CRYSTALLOGRAPHIC ANALYSIS (PAPER I)

Methanocaldococcus jannaschii nucleoside kinase was crystallized as apo-enzyme (MjNK-apo) (Paper I, Fig. 1a) as well as in complex with the non-hydrolyzable ATP-analogue AMP-PNP and Mg^{2+} (MjNK-A) (Paper I, Fig. 1b). Some of the MjNK-A crystals were soaked in a solution containing fructose-6-phosphate (MjNK-AF). All crystals were grown at 20 °C using the vapour diffusion method. MjNK-apo crystals were obtained from a mother liquor containing 10 % (w/v) PEG3500 and 0.15 M magnesium acetate. MjNK-A was crystallized from a solution containing 15 % (w/v) PEG4000, 0.10 M magnesium chloride and 0.1 M Tris-HCl pH 8.5.

High-resolution synchrotron data were collected at cryo-conditions for the three crystal forms. The cryo-protectant solutions for each crystal form contained the mother-liquor composition with the addition of 30 % ethylene glycol for MjNK-apo and 20 % ethylene glycol for MjNK-A and MjNK-AF. The apo enzyme crystals diffracted to 1.70 Å and the two complex crystals each to 1.93 Å.

MjNK-apo crystallized in space group $P2_12_12$ with unit cell axes a=64.1 Å, b=148.1 Å and c=41.1 Å. The two complex forms crystallized in space group $P2_12_12_1$ and had approximately the same unit cell axis lengths, which were a=64.2 Å, b=83.2 Å and c=147.0 Å for MjNK-A and a=64.0 Å, b=83.1 Å and c=146.8 Å for MjNK-AF. The unit cell parameters of the apo enzyme crystal and the complex crystals are almost identical except for the b axes of the complexes, which are twice as long as the c axis for MjNK-apo. Consequently, the complex crystal forms have a doubled unit cell volume and thus contain twice the number of monomers compared to the apo-enzyme crystal. This was further confirmed by the calculated Matthew's coefficients (Matthews, 1968) (MjNK-apo: 2.88 $Å^3$ Da^{-1} for 4 monomers/unit cell; MjNK-A: 2.90 $Å^3Da^{-1}$ for 8 monomers/unit cell; MjNK-AF: 2.88 $Å^3$ Da^{-1} for 8 monomers/unit

cell) which indicated that MjNK-apo contained one monomer per asymmetric unit and the complexes a dimer per asymmetric unit. It was assumed that the biological dimer was generated by a two-fold crystallographic symmetry operation in the apo-enzyme crystals. The stereographic plots of the self-rotation function at κ=180° for either of the complexes, showed no additional two-fold symmetry except for crystallographic symmetry (Paper I, Fig. 2). This indicates that the two-fold non-crystallographic axis might be parallel to a crystallographic axis. This was confirmed from a native Patterson map which revealed a pseudo-origin peak at $(u, v, w =$ 0.0, 0.5, 0.0*)* with an intensity of 25 % of the Patterson origin peak (Paper I, Fig. 3). Accordingly, the two-fold non-crystallographic symmetry axis relating the two subunits in the asymmetric unit is parallel to the y-axis in the unit cell.

4.3 DIFFERENTIAL SCANNING CALORIMETRY (PAPER II AND INCOMPLETE RESULTS)

Differential scanning calorimetry was performed on nucleoside kinase in 20 mM sodium phosphate buffer pH 7.0 between 20 and 120 °C and with a scan rate of 1 K min^{-1}. The observed apparent T_m was 90 °C (Fig. 7). As for many proteins, the process was irreversible and resulted in aggregation of the protein after the transition. In order to find a condition at which the denaturation was reversible or at least at which the calorimetric trace of an irreversible process was interpretable, different buffer conditions a well as scan rates were tested. A thermogram with good appearance was observed for nucleoside kinase in 10 mM glycine pH 3.0, however, the enzyme appeared still irreversibly denatured in the second thermogram. DSC scans were performed at this condition at the following scan rates: 0.58, 1.00, 1.50, 1.75, and 2.00 K min^{-1} (Fig. 8). A plot of the apparent T_m values obtained at the different scan rates against the scan rate, showed that the transition temperatures could be fitted to a second degree polynomial, which indicates that the T_m values are dependent on the scan rates, i.e., the process is kinetically controlled (Table 2, Fig. 9). The calorimetric enthalpy was found to be larger than the van't Hoff enthalpy. Thus, the process did

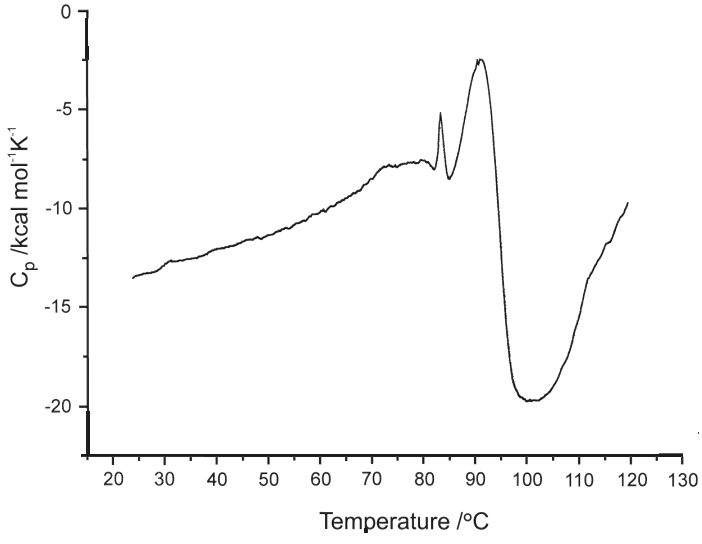

Figure 7. Differential scanning calorimetry thermogram for nucleoside kinase in 20 mM sodium phosphate buffer pH 7.0. The resulting apparent T_m was 90 °C. The transition N → D was irreversible.

not follow a two-state irreversible model and was instead assumed to follow a non-two-state irreversible model. Such a process involves one or several intermediate states which may have similar T_m values as the native protein. As a consequence, the area under the calorimetric trace is larger than that for the corresponding two-state irreversible process, which results in the difference between the calorimetric and van't Hoff enthalpies.

It is of importance that the calorimetric enthalpies, in contrast to the transition temperatures, are independent of scan rate and thus have a constant value for different scan rates (Vogl et al., 1997). For kinetically dependent transitions, DSC experiments at different scan rates are needed. A total number of five or six DSC experiments performed at different scan rates take up to two weeks to perform. At least three, and for good reproducibility five, DSC experiments performed at the same scan rate should agree in appearance as well as in apparent T_m, ΔC_p and ΔH_{cal}. This requires a minimum of six weeks for performing all DSC experiments. Nucleoside kinase slowly degraded in the buffer even when kept at 4 °C. During a period of two weeks, the native protein was partially denatured, resulting in a lower concentration of native protein (= starting material) in the DSC experiment. This in turn led to a concentration dependent difference in calorimetric enthalpies between the varied scan rates (Table 2). Such a problem can be avoided by freezing the sample and buffers in portions suitable for one DSC experiment.

Additional information of the assembly state of the enzyme at pH 3.0 is needed in order to be able to propose a correct model for the irreversible process before the thermograms can be analyzed. This would preferably be established by ultracentrifugation. Assumptions must be made for which and how many intermediate states that could be involved. A possible model is

$$N_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2\,N \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} 2\,D \xrightarrow{k_3} I \qquad (26)$$

where N_2 represents the native dimer, N represents the protein in native state and D the denatured state. I is the irreversibly unfolded protein. The first and second transitions are rate dependent. However, the process can instead follow the model

$$N_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} D_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} 2\,D \xrightarrow{k_3} I \qquad (27)$$

which differs from the first one in that the dimer (N_2) is denatured before it is disassembled into monomers. However, if the ultracentrifugation data reveals that MjNK is a monomer at pH 3.0, neither of these two assumptions are correct.
If the DSC traces anyhow show bad consistency, folding experiments could instead be performed by fluorescence spectroscopy. By using fluorescence spectroscopy,

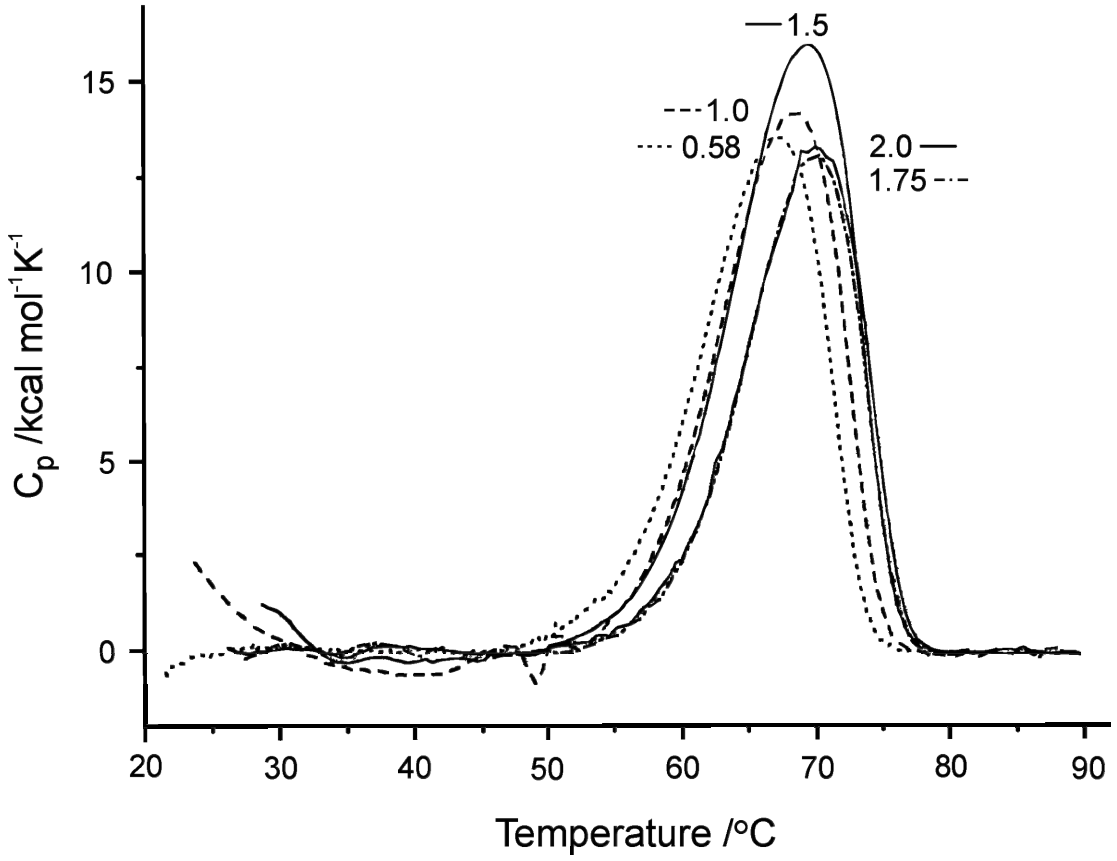

Figure 8. Variation of the scan rate of the transition curves obtained for nucleoside kinase in 10 mM glycine pH 3.0. The number closest to each of the transition peaks correspond to the scan rate in K min^{-1} of the respective curve. The respective line style is indicated in front of the numbers. For clarity, the transition curves obtained at 1.75 and 2.0 K min^{-1} are overlapping.

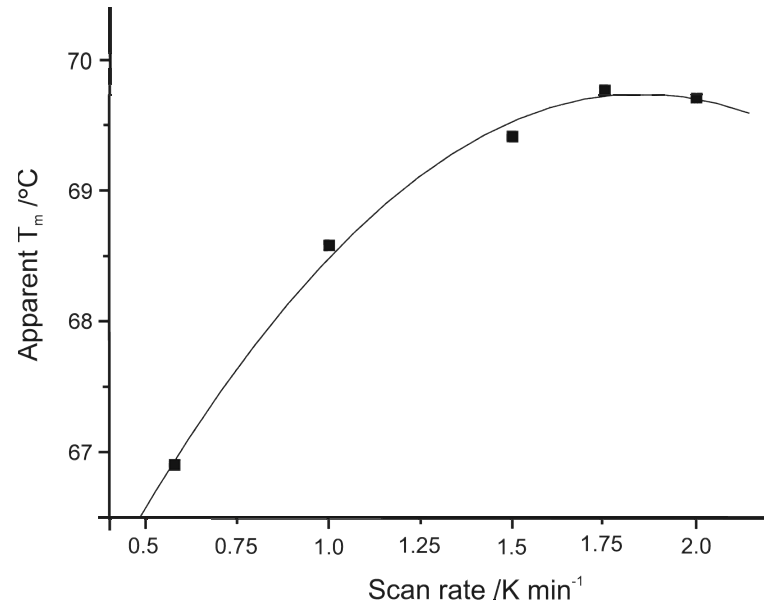

Figure 9. The apparent T_m values as a function of scan rate of nucleoside kinase in 10 mM glycine pH 3.0. The transition temperatures are fitted to a second degree polynomial, which indicates the dependence of T_m on scan rate and thereby that the process is kinetically controlled.

Table 2. Variations of apparent thermodynamic transition parameters with scan rate for nucleoside kinase in buffer 10 mM glycine pH 3.0.

scan rate (K min^{-1})	app. T_m (°C)	ΔH_{cal} (kJ mol^{-1})	ΔH_{vH} (kJ mol^{-1})	$\Delta H_{cal} / \Delta H_{vH}$
0.58	66.9	659	333	1.98
1.00	68.6	646	354	1.83
1.50	69.4	769	341	2.26
1.75	69.8	593	366	1.62
2.00	69.7	594	363	1.64

equilibrium folding/unfolding can be studied at lower temperatures by using chemical denaturants, such as urea or guanidinium hydrochloride (Royer, 1995). In addition, this method often requires a much smaller concentration of protein, which increases the possibilities of refolding.

4.4 STRUCTURE DETERMINATION AND ANALYSIS (PAPER III)

The three-dimensional crystal structure of *Methanocaldococcus jannaschii* nucleoside kinase was determined at 2.2 Å by the multiple-wavelength anomalous dispersion technique using a selenomethionine derivative. Two crystal structures of MjNK were determined: unliganded enzyme and MjNK in complex with adenosine, the non-hydrolyzable ATP-analogue AMP-PNP, and Mg^{2+} (MjNK/adenosine/AMP-PNP complex). The two complex crystal forms described in paper I, MjNK-A and MjNK-AF, were found to be identical since Fru-6-P did apparently not bind to the enzyme. Instead, an adenosine molecule, possibly an impurity from the purchased AMP-PNP material or a hydrolyzation product, bound to the active site. The apo-enzyme model was refined at 1.7 Å to a final R_{cryst} of 23.5 % and an R_{free} of 25.3 %. The model of the MjNK/adenosine/AMP-PNP complex was refined from the MjNK-AF data at 1.9 Å to a final R_{cryst} of 24.3 % and an R_{free} of 28.1 %.

Nucleoside kinase exists as a homodimer (monomer A and B in the complex) in both crystal structures (Paper III, Fig. 1). The enzyme has an overall monomer fold characteristic for the members of the ribokinase superfamily. MjNK comprises both the α/β domain and the smaller lid domain. The α/β domain is made up of a central β-sheet that contains nine mainly parallel β-strands flanked by five α-helices on one side and four α-helices on the opposite side. The protruding lid domain is made up by four β-strands (β3↓β2↑β6↑β7↓) and connecting loops.

The dimer is formed mainly through hydrophobic interactions between the lid domains from each monomer. In addition, there is a large aromatic cluster consisting of in total twelve residues in the dimer interface (Paper III, Fig. 2). Nucleoside kinase can be in two conformational conformers, i.e., with the lid either opened or closed. The enzyme most likely undergoes an induced-fit conformational shift upon substrate binding. This has been observed for structural homologues, e.g., ribokinase from *E. coli*. Strands β2A, β6A, β7A and loopβ2Bβ3B of the lid domains fold back over the substrate binding site of monomer A, and the opposite for monomer B. In the apo-enzyme structure, the monomers are in the open conformation. In the MjNK/adenosine/AMP-PNP complex, the two monomers show different conformations (Paper III, Fig. 3). Monomer A exists in the closed conformation and

contains a tightly bound adenosine. Monomer B is in the open conformation and contains an adenosine molecule that is bound with lower occupancy. This binding behaviour could reflect that only one substrate molecule can be phosphorylated at a time. Moreover, Asn33B in the lid domain is forming hydrogen bonds with the adenosine bound in monomer A. Owing to the different conformations of the monomers, Asn33A is too far from the active site of monomer B. This interaction is not possible if both monomers are closed. The relevance of this residue is however unclear. The α/β domain retains its conformation upon lid closure. The lid domain, however, undergoes a small intradomain structural change. A superposition of the homodimers (apo-enzyme and complex) shows that the two α/β domains in each dimer remain at the same relative position and that a conformational change is only observed in the lid domains (Paper III, Fig. 4).

The MjNK dimer contains one active site per monomer. The active site is located in a shallow groove along one edge of the central β-sheet as is characteristic for the enzymes of the ribokinase superfamily. In the active site of monomer A, clear density was observed for adenosine, a magnesium ion coordinated by six waters (MO6) and the ADP-part of the ATP-analogue (Paper III, Fig. 5). In the active site of monomer B, density was observed for AMP-PNP, but was only partially visible for adenosine and MO6. Each monomer of the MjNK/adenosine/AMP-PNP complex contains one additional magnesium ion coordinated by five water molecules (MO5) that is located at the interface between the dimers in the crystal. The location of the substrates and cofactor can be described by a V formation, where the magnesium is positioned in the nib with adenosine on one side and AMP-PNP on the opposite side. There are many interactions, mainly hydrogen bonds, between MjNK and adenosine. The water molecules of MO6 form hydrogen bonds with the enzyme, adenosine, as well as with AMP-PNP. Only a few interactions occur between AMP-PNP and the enzyme. In MjNK, the conserved aspartate and asparagine involved in hydrogen bonding with the 2' and 3' hydroxyl groups of ribose are represented by Asp17 and Asn47. Asp247 corresponds to the highly conserved and catalytically important aspartate and is involved in hydrogen bonding to the 5' hydroxyl group of the ribose ring. In addition, the pyrimidine ring of the adenine base of adenosine forms planar aromatic π-π stacking interactions with the benzene ring of Phe113. The electrophilic magnesium ion located in the active site is coordinated by both the β- and γ-phosphates in monomer B. These interactions result in a decrease of the negative charge of the phosphate groups, which in turn alleviates the leaving of ADP. In addition to the conserved Asp17, Asn47 and Asp247, MjNK also contains the other residues and conserved regions that are characteristic for the enzymes belonging to the ribokinase superfamily. These include the GG dipeptide sequence (G42G43 in MjNK), the anion hole (GAGD, 244-247 in MjNK) as well as a non-conserved arginine or lysine (Arg250 in MjNK) that stabilizes the transition state during phosphate transfer.The catalytic reaction mechanism of MjNK is likely to be similar to that of other ribokinase superfamily members, since the position of substrate and ATP-analogue and important catalytic residues are conserved.

A structure comparison of nucleoside kinase and available structures within the PDB, revealed that the monomer structure of MjNK is most similar to *E. coli* ribokinase (rmsd 1.48 Å for 263 Cα atoms), but shares also a high structural homology to

S. typhimurium AIR kinase (rmsd 1.65 Å for 232 Cα atoms) and human adenosine kinase (rmsd 1.74 Å for 239 Cα atoms) (Paper III, Fig. 6). Since the hyperthermostable nucleoside kinase and ribokinase from *E. coli* show a similar overall fold and dimerize in the same way, some structural properties, which are putative determinants for thermostability, were compared. Factors that were looked upon were secondary structure differences, accessible surface area, ionic interactions, hydrogen bonding and aromatic interactions. A summary of the results is presented in Paper III, Table 3.

The structures of MjNK and ribokinase are indeed very similar and there exist only a few differences between their structures (Paper III, Fig. 7). The secondary structure content of MjNK and ribokinase are approximately the same (58.9 % and 61.2 %, respectively). However, a big difference is found between the lid domains (β3↓β2↑β6↑β7↓ and connecting loops) since MjNK possesses a much shorter strand β3 and instead a long ordered loop between β2 and β3, which forms dimer interactions with β7 of the other monomer. Moreover, MjNK has shorter loop regions as well as an insertion of a small α-helix in one loop that could contribute to the enhanced thermal stability of MjNK.

The accessible surface area of MjNK is somewhat larger than that for ribokinase (25,639 $Å^2$ and 24,704 $Å^2$, respectively). Nucleoside kinase shows, compared to EcRK, a clear decrease in hydrophobic ASA and an increase in charged ASA; two structural characteristics, which often are observed for enzymes from hyperthermophiles. The polar ASA between the two enzymes are similar.

Nucleoside kinase has an increased number of ionic interactions compared to ribokinase. At a cutoff radius of 6 Å, the MjNK dimer contains a higher amount of ion-pairs and multiple-member networks than ribokinase. Nucleoside kinase comprises an inter-subunit network located in the dimer interface that does not exist in ribokinase. A seven-member network is positioned at the enzyme surface of monomer A, near the magnesium ion MO5 and the crystallographically related dimer (Paper III, Fig. 8). The electrostatic interactions seem to be highly optimized in terms of excess charges in nucleoside kinase, which has a net charge of zero at pH 7. For comparison, the net charge at pH 7 is -9 for ribokinase.

There are no obvious differences in hydrogen bonding between MjNK and EcRK and consequently no conclusions could be drawn about increased hydrogen bonding as a contribution to the thermal stability of nucleoside kinase.

Nucleoside kinase contains a much higher amount of aromatic residues than ribokinase and the majority of these are involved in aromatic pairs or aromatic network clusters. A twelve-member aromatic network is formed through aromatic interactions between six residues from each lid domain in the homodimer (Paper III, Fig. 2). The aromatic interactions in MjNK are considered to be an important factor for the thermal stability of the enzyme.

5 ACKNOWLEDGEMENTS

Many people have in one way or another contributed to this thesis. I would like to thank...

My supervisor **Prof. Rudolf Ladenstein** for your pedagogical ability to make complex and difficult subjects easier, for sharing your genuine interest and knowledge in science, and for always taking your time to discuss various problems.

My secondary supervisor **Dr. Winfried Meining**, for deep scientific discussions and for being my neighboring "hacker", teaching me binary numbers, the grep-command and other computer-related things.

Prof. Dr. Peter Schönheit, für die gute Zusammenarbeit während dieser Jahre. Ich bin sehr dankbar, dass Sie mich nach Kiel und in die Universität eingeladen haben.

Dr. Thomas Hansen für die gute Zusammenarbeit und die ergiebigen Diskussionen über Pfk's, Forchung und das Leben. Vielen Dank für die Gastfreundschaft von Dir, **Ulrike**, und dem lieben **Tobias** und **Florian** während meines Besuchs in Kiel.

Micke, for being my third supervisor! Many, many thanks for all your help and a lot of fun during these years.

Boel, för att du är så snäll och lugn och omtänksam och en underbar vän. Tack för alla samtal och för att du lyssnar.

Present and former members of the X-ray group: **Wei** for nice discussions in English, Chinese and Swedish; **Katja** for all good chats and laughs; **Heiko** for nice and quiet mornings; **Xiaofeng** for helping out with computers and other problems; **Ida** for your great enthusiasm and positive attitude; **Petras**, Spasíba!; **Bin** for learning me how to eat roasted duck; **Luca**, it was very nice to get to know you and I wish you all the best in Sweden, **Andrey** and **Assen**.

Katarina, Sophia, Jianxin, Ana, Jan, Peter C, Tobbe, Johan, Anna-Karin, Kimberly, Peter H, Caroline, Priya and all other colleagues at CSB and the rest of Novum that have helped me and spread some joy during my postgraduate studies.

Erik, for keeping the computer network running and that you always so kindly are helping those (=me) with the computer gene knocked-out...

Lotta Hambraeus, for good supervison, discussions and help during my teaching in Thermodynamics.

Underbara **Ceciliakören** och **Rutger**, vad vore livet utan er och musiken? Och tack **Berit** för att du hjälper till att anordna alla våra resor, vårluncher och julmiddager (framför allt jordgubbs-/hallonsylten/marmeladen!).

Carin, vi har så mycket gemensamt fastän det skiljer nära 50 år mellan oss. Tack för alla trevliga stunder med fika och vackert pianospel!

Jenni, du är en sann vän! Det är alltid lika roligt att prata med dig och förhoppningsvis kan vi träffas lite oftare nu.

Anna 1, **Anna 2**, **Mattias**, **Thomas**, och solstrålarna **Saga**, **Adam** och **Liv** för midsommarfirande, kräftskivor, julmys och allt däremellan!

Kerstin, tack för all din värme och omsorg. Jag blir så lugn och glad med dig. Och tack hela Christers släkt för att ni har tagit emot mig med öppna armar.

Mormor, vi har så roligt ihop när vi ses och jag saknar dig så mycket. Nu har jag äntligen studerat färdigt...

Mamma och **pappa**, för att ni alltid finns där och ställer upp för mig!

Min bäste store lille **Bror**, för att du är den bästa lillebror en storasyster kan ha.

Dumbo, Ninus, Kräket, Steken – du förstår ju hur som helst inte, men du är i alla fall det allra gulligaste lilla djuret och så himla bra på att få mig glad! 

Christer, du är min själsfrände! Tack för allt stöd, all hjälp och allt roligt under de här åren; du är verkligen bäst!

6 REFERENCES

Arndt, U. W., Crowther, R. A., and Mallett, J. F. (1968). A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. *J. Sci. Instrum.* **1**, 510-516.

Arndt, U. W., and Wonacott, A. J. (1977). The Rotation Method in Crystallography: Data Collection from Macromolecular Crystals (Amsterdam, North-Holland Publishing Company).

Babul, J. (1978). Phosphofructokinases from Escherichia coli. Purification and characterization of the nonallosteric isozyme. *J. Biol. Chem.* **253**, 4350-4355.

Bagautdinov, B., Kuramitsu, S., Yokoyama, S., Miyano, M., and Tahirov, T. H. unpublished results.

Barns, S. M., Fundyga, R. E., Jeffries, M. W., and Pace, N. R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl. Acad. Sci. USA* **91**, 1609-1613.

Becktel, W. J., and Schellman, J. A. (1987). Protein stability curves. *Biopolymers* **26**, 1859-1877.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242.

Blangy, D., Buc, H., and Monod, J. (1968). Kinetics of the allosteric interactions of phosphofructokinase from Escherichia coli. *J. Mol. Biol.* **31**, 13-35.

Bork, P., Sander, C., and Valencia, A. (1993). Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci.* **2**, 31-40.

Bredberg, K., Andersson, B. E., Landfors, E., and Holst, O. (2002). Microbial detoxification of waste rubber material by wood-rotting fungi. *Bioresour. Technol.* **83**, 221-224.

Brochier, C., and Philippe, H. (2002). Phylogeny: a non-hyperthermophilic ancestor for bacteria. *Nature* **417**, 244.

Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**, 54-68.

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54 (Pt 5)**, 905-921.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273**, 1058-1073.

Burley, S. K., and Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23-28.

Cambillau, C., and Claverie, J. M. (2000). Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.* **275**, 32383-32386.

Campobasso, N., Mathews, II, Begley, T. P., and Ealick, S. E. (2000). Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from Bacillus subtilis at 1.5 A resolution. *Biochemistry* **39**, 7868-7877.

Cheng, G., Bennett, E. M., Begley, T. P., and Ealick, S. E. (2002). Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from Salmonella typhimurium at 2.3 A resolution. *Structure (Camb)* **10**, 225-235.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763.

Cowtan, K. D., and Zhang, K. Y. (1999). Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72**, 245-270.

Daniel, R. M., and Cowan, D. A. (2000). Biomolecular stability and life at high temperatures. *Cell. Mol. Life Sci.* **57**, 250-264.

Danson, M. J., and Hough, D. W. (1998). Structure, function and stability of enzymes from the Archaea. *Trends Microbiol.* **6**, 307-314.

Das, R., and Gerstein, M. (2000). The stability of thermophilic proteins: a study based on comprehensive genome comparison. *Funct. Integr. Genomics* **1**, 76-88.

Dauter, Z. (1999). Data-collection strategies. *Acta Crystallogr. D Biol. Crystallogr.* **55 (Pt 10)**, 1703-1717.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.

Dodson, E. (2003). Is it jolly SAD? *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1958-1965.

Dominy, B. N., Minoux, H., and Brooks, C. L., 3rd (2004). An electrostatic basis for the stability of thermophilic proteins. *Proteins* **57**, 128-141.

Drenth, J. (1999). Principles of Protein X-ray Crystallography, Springer-Verlag).

Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489-502.

Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002). NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207-211.

Fields, P. A. (2001). Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **129**, 417-431.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T., and Gronenborn, A. M. (1991). High-resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. *Biochemistry* **30**, 2685-2698.

Forterre, P., Brochier, C., and Philippe, H. (2002). Evolution of the Archaea. *Theor. Popul. Biol.* **61**, 409-422.

Freire, E., van Osdol, W. W., Mayorga, O. L., and Sanchez-Ruiz, J. M. (1990). Calorimetrically determined dynamics of complex unfolding transitions in proteins. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 159-188.

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

Goodenough, P. W., and Jenkins, J. A. (1991). Protein engineering to change thermal stability for food enzymes. *Biochem. Soc. Trans.* **19**, 655-662.

Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843-1849.

Grogan, D. W. (1998). Hyperthermophiles and the problem of DNA instability. *Mol. Microbiol.* **28**, 1043-1049.

Harker, D. (1936). The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag3AsS3, and pyrargyrite, Ag3SbS3. *J. Chem. Phys.* **4**, 381-390.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899-929.

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

Holmgren, A., Soderberg, B. O., Eklund, H., and Branden, C. I. (1975). Three-dimensional structure of Escherichia coli thioredoxin-S2 to 2.8 A resolution. *Proc. Natl. Acad. Sci. USA* **72**, 2305-2309.

Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144-1149.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990). Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**, 1031-1044.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63-67.

Inagaki, E., Ukita, Y., Kumei, M., Kajihara, Y., and Tahirov, T. H. (2004). Crystallization and preliminary crystallographic analysis of 2-keto-3-deoxygluconate kinase from Thermus thermophilus. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 761-763.

Ishikawa, K., Nakamura, H., Morikawa, K., and Kanaya, S. (1993). Stabilization of Escherichia coli ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* **32**, 6171-6178.

Ito, S., Fushinobu, S., Jeong, J. J., Yoshioka, I., Koga, S., Shoun, H., and Wakagi, T. (2003). Crystal structure of an ADP-dependent glucokinase from Pyrococcus furiosus: implications for a sugar-induced conformational change in ADP-dependent kinase. *J. Mol. Biol.* **331**, 871-883.

Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H., and Wakagi, T. (2001). Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Structure (Camb)* **9**, 205-214.

Jeyakanthan, J., and Tahirov, T. H. unpublished results.

Joint Center for Structural Genomics, unpublished results-a.

Joint Center for Structural Genomics, unpublished results-b.

Joint Center for Structural Genomics, unpublished results-c.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47 (Pt 2)**, 110-119.

Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R., and Wolfe, R. S. (1983). *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136**, 254-261.

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

Karlström, M., Stokke, R., Steen, I. H., Birkeland, N. K., and Ladenstein, R. (2005). Isocitrate dehydrogenase from the hyperthermophile Aeropyrum pernix: X-ray

structure analysis of a ternary enzyme-substrate complex and thermal stability. *J. Mol. Biol.* **345**, 559-577.

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

Karshikoff, A., and Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem. Sci.* **26**, 550-556.

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

Kay, L. E. (2005). NMR studies of protein structure and dynamics. *J Magn Reson* **173**, 193-207.

Kendrew, J. C., and Perutz, M. F. (1957). X-ray studies of compounds of biological interest. *Annu. Rev. Biochem.* **26**, 327-372.

Korndörfer, I., Steipe, B., Huber, R., Tomschy, A., and Jaenicke, R. (1995). The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium Thermotoga maritima at 2.5 A resolution. *J. Mol. Biol.* **246**, 511-521.

Kotlarz, D., and Buc, H. (1981). Regulatory properties of phosphofructokinase 2 from Escherichia coli. *Eur. J. Biochem.* **117**, 569-574.

Kwok, F., Scholz, G., and Churchich, J. E. (1987). Brain pyridoxal kinase dissociation of the dimeric structure and catalytic activity of the monomeric species. *Eur. J. Biochem.* **168**, 577-583.

Kühlbrandt, W., Wang, D. N., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614-621.

Ladenstein, R., and Antranikian, G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv. Biochem. Eng. Biotechnol.* **61**, 37-85.

Lamzin, V. S. (1993). Automated refinement of protein models. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 129-147.

Lange, M., Westermann, P., and Ahring, B. K. (2005). Archaea in protozoa and metazoa. *Appl. Microbiol. Biotechnol.* **66**, 465-474.

Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049-1067.

Lebbink, J. H., Consalvi, V., Chiaraluce, R., Berndt, K. D., and Ladenstein, R. (2002). Structural and thermodynamic studies on a salt-bridge triad in the NADP-binding domain of glutamate dehydrogenase from Thermotoga maritima: cooperativity and electrostatic contribution to stability. *Biochemistry* **41**, 15524-15535.

Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R., and de Vos, W. M. (1999). Engineering activity and stability of Thermotoga maritima glutamate dehydrogenase. II: construction of a 16-residue ion-pair network at the subunit interface. *J Mol Biol* **289**, 357-369.

Lehninger, A. L., Nelson, D. L., and Cox, M. M. (1993). Principles of Biochemistry, 2nd edn, Worth Publishers).

Leslie, A. G., Moody, P. C., and Shaw, W. V. (1988). Structure of chloramphenicol acetyltransferase at 1.75-A resolution. *Proc. Natl. Acad. Sci. U S A.* **85**, 4133-4137.

Li, M. H., Kwok, F., Chang, W. R., Lau, C. K., Zhang, J. P., Lo, S. C., Jiang, T., and Liang, D. C. (2002). Crystal structure of brain pyridoxal kinase, a novel member of the ribokinase superfamily. *J. Biol. Chem.* **277**, 46385-46390.

Li, W. F., Zhou, X. X., and Lu, P. (2005). Structural features of thermozymes. *Biotechnol. Adv.* **23**, 271-281.

Lim, J. H., Yu, Y. G., Han, Y. S., Cho, S., Ahn, B. Y., Kim, S. H., and Cho, Y. (1997). The crystal structure of an Fe-superoxide dismutase from the hyperthermophile Aquifex pyrophilus at 1.9 A resolution: structural basis for thermostability. *J. Mol. Biol.* **270**, 259-274.

Lumry, R., and Eyring, H. (1954). *J. Phys. Chem.* **58**, 110-???

Ma, J. C., and Dougherty, D. A. (1997). The Cation-Pi Interaction. *Chem. Rev.* **97**, 1303-1324.

Macedo-Ribeiro, S., Darimont, B., Sterner, R., and Huber, R. (1996). Small structural changes account for the high thermostability of 1[4Fe-4S] ferredoxin from the hyperthermophilic bacterium Thermotoga maritima. *Structure (Camb)* **4**, 1291-1301.

Mathews, II, Erion, M. D., and Ealick, S. E. (1998). Structure of human adenosine kinase at 1.5 A resolution. *Biochemistry* **37**, 15607-15620.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA* **84**, 6663-6667.

Matthews, B. W., Weaver, L. H., and Kester, W. R. (1974). The conformation of thermolysin. *J. Biol. Chem.* **249**, 8030-8044.

McPherson, A. (1999). Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press).

McPherson, A. (2003). Introduction to Macromolecular Crystallography, Wiley-Liss, Inc.).

McPherson, A. (2004). Introduction to protein crystallization. *Methods* **34**, 254-265.

Miller, R., Gallo, S. M., Khalak, H. G., and Weeks, C. M. (1994). SnB: crystal structure determination via shake-and-bake. *J. Appl. Crystallogr.* **27**, 613-621.

Murshudov, G. N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255.

Naghibi, H., Tamura, A., and Sturtevant, J. M. (1995). Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc. Natl. Acad. Sci. USA* **92**, 5597-5599.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A Found. Crystallogr.* **50**, 157-163.

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

Nicholson, H., Becktel, W. J., and Matthews, B. W. (1988). Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. *Nature* **336**, 651-656.

Nogales, E., Wolf, S. G., and Downing, K. H. (1998). Structure of the alpha beta tubulin dimer by electron crystallography. *Nature* **391**, 199-203.

Ohshima, N., Inagaki, E., Yasuike, K., Takio, K., and Tahirov, T. H. (2004). Structure of Thermus thermophilus 2-Keto-3-deoxygluconate kinase: evidence for recognition of an open chain substrate. *J. Mol. Biol.* **340**, 477-489.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Pace, C. N. (1992). Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* **226**, 29-35.

Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *Faseb J.* **10**, 75-83.

Patterson, A. L. (1935). A direct method for the determination of the components of interatomic distances in crystals. *Z. Krist.* **90**, 517-542.

Perrakis, A. (1997). wARP: Improvement and Extension of Crystallographic Phases by Weighted Averaging of Multiple-Refined Dummy Atomic Models. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 448-455.

Perutz, M. F. (1978). Electrostatic effects in proteins. *Science* **201**, 1187-1191.

Perutz, M. F., and Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**, 256-259.

Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002). Solution NMR techniques for large molecular and supramolecular structures. *J Am Chem Soc* **124**, 12144-12153.

Rose, G. D., and Wolfenden, R. (1993). Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 381-415.

Rossmann, M. G., and Arnold, E., eds. (2001). International Tables for Crystallography F - Crystallography of Biological Macromolecules (Kluwer Academic Publishers).

Rossmann, M. G., and Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24-31.

Royer, C. A. (1995). Fluorescence spectroscopy. *Methods Mol. Biol.* **40**, 65-89.

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

Safo, M. K., Musayev, F. N., Hunt, S., di Salvo, M. L., Scarsdale, N., and Schirch, V. (2004). Crystal structure of the PdxY Protein from Escherichia coli. *J. Bacterio* **186**, 8074-8082.

Sanchez-Ruiz, J. M., Lopez-Lacomba, J. L., Cortijo, M., and Mateo, P. L. (1988). Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry* **27**, 1648-1652.

Santos, H., and da Costa, M. S. (2002). Compatible solutes of organisms that live in hot saline environments. *Environ Microbiol* **4**, 501-509.

Schiraldi, C., Giuliano, M., and De Rosa, M. (2002). Perspectives on biotechnological applications of archaea. *Archaea* **1**, 75-86.

Schneider, T. R., and Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772-1779.

Schomaker, V., and Trueblood, K. N. (1968). On the rigid-body motion of molecules in crystals. *Acta. Crystallogr. B* **24**, 63-76.

Schumacher, M. A., Scott, D. M., Mathews, II, Ealick, S. E., Roos, D. S., Ullman, B., and Brennan, R. G. (2000). Crystal structures of Toxoplasma gondii adenosine kinase

reveal a novel catalytic mechanism and prodrug binding. *J. Mol. Biol.* **298**, 875-893.

Serrano, L., Bycroft, M., and Fersht, A. R. (1991). Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *J. Mol. Biol.* **218**, 465-475.

Sigrell, J. A., Cameron, A. D., Jones, T. A., and Mowbray, S. L. (1997). Purification, characterization, and crystallization of Escherichia coli ribokinase. *Protein Sci.* **6**, 2474-2476.

Sigrell, J. A., Cameron, A. D., Jones, T. A., and Mowbray, S. L. (1998). Structure of Escherichia coli ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insights into a new family of kinase structures. *Structure (Camb)* **6**, 183-193.

Sigrell, J. A., Cameron, A. D., and Mowbray, S. L. (1999). Induced fit on sugar binding activates ribokinase. *J. Mol. Biol.* **290**, 1009-1018.

Slutsky, M. M., and Marsh, E. N. (2004). Cation-pi interactions studied in a model coiled-coil peptide. *Protein Sci.* **13**, 2244-2251.

Spassov, V. Z., Karshikoff, A. D., and Ladenstein, R. (1994). Optimization of the electrostatic interactions in proteins of different functional and folding type. *Protein Sci.* **3**, 1556-1569.

Spassov, V. Z., Karshikoff, A. D., and Ladenstein, R. (1995). The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci.* **4**, 1516-1527.

Stetter, K. O. (1996). Hyperthermophiles in the history of life. *Ciba Found. Symp.* **202**, 1-10; discussion 11-18.

Stetter, K. O. (1998). Extremophiles: Microbial life in Extreme Environments. In Hyperthermophiles: Isolation, classification and properties (Wiley-Liss), pp. 1-24.

Sturtevant, J. M. (1987). Biochemical Applications of Differential Scanning Calorimetry. *Ann. Rev. Phys. Chem.* **38**, 463-488.

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

Taylor, G. (2003). The phase problem. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1881-1890.

Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56 (Pt 8)**, 965-972.

Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 38-44.

Tronrud, D. E. (2004). Introduction to macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2156-2168.

Tsuge, H., Sakuraba, H., Kobe, T., Kujime, A., Katunuma, N., and Ohshima, T. (2002). Crystal structure of the ADP-dependent glucokinase from Pyrococcus horikoshii at 2.0-A resolution: a large conformational change in ADP-dependent glucokinase. *Protein Sci.* **11**, 2456-2463.

Unwin, P. N., and Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* **94**, 425-440.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *Appl. Cryst.* **30**, 1022-1025.

Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D Biol. Crystallogr.* **55 (Pt 1)**, 191-205.

Waksman, G., Krishna, T. S., Williams, C. H., Jr., and Kuriyan, J. (1994). Crystal structure of Escherichia coli thioredoxin reductase refined at 2 A resolution. Implications for a large conformational change during catalysis. *J. Mol. Biol.* **236**, 800-816.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90-112.

Whitfield, J. (2004). Origins of life: born in a watery commune. *Nature* **427**, 674-676.

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

Wigley, D. B., Clarke, A. R., Dunn, C. R., Barstow, D. A., Atkinson, T., Chia, W. N., Muirhead, H., and Holbrook, J. J. (1987). The engineering of a more thermally stable lactate dehydrogenase by reduction of the area of a water-accessible hydrophobic surface. *Biochim. Biophys. Acta* **916**, 145-148.

Wintrode, P. L., Zhang, D., Vaidehi, N., Arnold, F. H., and Goddard, W. A., 3rd (2003). Protein dynamics in a family of laboratory evolved thermophilic enzymes. *J. Mol. Biol.* **327**, 745-757.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**, 5088-5090.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576-4579.

Vogl, T., Jatzke, C., Hinz, H. J., Benz, J., and Huber, R. (1997). Thermodynamic stability of annexin V E17G: equilibrium parameters from an irreversible unfolding reaction. *Biochemistry* **36**, 1657-1668.

Vogt, G., and Argos, P. (1997). Protein thermal stability: hydrogen bonds or internal packing? *Fold. Des.* **2**, S40-46.

Wu, L.-F., Reizer, A., Reizer, J., Cai, B., Tomich, J. M., and Saier, J., M. H. (1991). Nucleotide sequence of the Rhodobacter capsulatus fruK gene, which encodes fructose-1-phosphate kinase: evidence for a kinase superfamily including both phosphofructokinases of Escherichia coli. *J. Bact.* **173**, 3117-3127.

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

Yip, K. S., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D., and Rice, D. W. (1998). Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur. J. Biochem.* **255**, 336-346.

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

Yutani, K., Ogasahara, K., Tsujita, T., and Sugino, Y. (1987). Dependence of conformational stability on hydrophobicity of the amino acid residue in a series

of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. *Proc. Natl. Acad. Sci. USA* **84**, 4441-4444.

Zhang, R. G., Grembecka, J., Vinokour, E., Collart, F., Dementieva, I., Minor, W., and Joachimiak, A. (2002). Structure of Bacillus subtilis YXKO--a member of the UPF0031 family and a putative kinase. *J. Struct. Biol.* **139**, 161-170.

Zhang, X., Meining, W., Fischer, M., Bacher, A., and Ladenstein, R. (2001). X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *J. Mol. Biol.* **306**, 1099-1114.

Zhang, Y., Dougherty, M., Downs, D. M., and Ealick, S. E. (2004). Crystal structure of an aminoimidazole riboside kinase from Salmonella enterica: implications for the evolution of the ribokinase superfamily. *Structure (Camb)* **12**, 1809-1821.